Biomedical Signal Processing and Control 70 (2021) 103069

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Neurodegenerative disease detection and severity prediction using deep learning approaches

Çağatay Berke Erdaş [a,*], Emre Sümer [a], Seda Kibaroğlu [b]

[a] Department of Computer Engineering, Faculty of Engineering, Başkent University, Ankara, Turkey
[b] Department of Neurology, Faculty of Medicine, Başkent University, Ankara, Turkey

ARTICLE INFO

Keywords:
Neurodegenerative disease
Classification of NDD
Prediction of NDD
ConvLSTM
3D CNN

ABSTRACT

Neurodegenerative diseases (NDDs) such as amyotrophic lateral sclerosis (ALS), Huntington's disease (HD), and Parkinson's disease (PD) can manifest themselves anatomically by degeneration in the brain as well as motor symptoms. The motor symptoms can affect walking dynamics in a disease-specific fashion; characteristically they disrupt gait. As the severity of the disease increases, walking ability deteriorates. We examined the effect of NDDs such as ALS, HD, and PD on gait and developed a convolutional long short-term memory (ConvLSTM) and three-dimensional convolutional learning network (3D CNN)-based approach to detecting neurodegenerative conditions and predicting disease severity.

1. Introduction

Neurodegenerative diseases (NDDs) are characterized by the degeneration of nerve cells in the substantia nigra, which is responsible for dopamine secretion in the brain. This degeneration affects dopamine production, resulting in movement disorders. In amyotrophic lateral sclerosis (ALS), Huntington's disease (HD), and Parkinson's disease (PD)—the major NDDs—involuntary movements occur due to the degeneration but manifest themselves in different ways. In PD, the most common movement-related clinical findings are bradykinesia, rigidity, resting tremor, postural instability, and difficulty in walking. In HD, the characteristic clinical findings include chorea, stumbling and clumsiness, and difficulty in movement. Muscle atrophy is typical of ALS, but muscle cramps and twitches may also make walking difficult [1–5]. Walking ability varies according to disease severity with all three conditions.

Examining the gait characteristics of patients is useful in diagnosing NDDs and estimating the severity of disease in a given patient. Correct detection of NDDs is the most important criterion for patients to access the right treatments, drugs, and care. Furthermore, the treatment of motor symptoms must be adjusted according to the severity of the disease. Over- or underdosing leads to incomplete suppression of the motor symptoms. Additionally, it changes their presentation, sometimes leading to misdiagnoses that seriously reduce patients' quality of life. Artificial intelligence-based models using gait data can aid in disease detection and prediction of disease severity. Some such models have been developed [6–11], but most are based on the processing and manipulation of one-dimensional gait data for disease detection [12–15].

Zeng et al. [12] introduced a new method for classifying NDDs: a deterministic learning theory using gait dynamics. Using a supercentral pattern generator model, a time series of swing and stance intervals of left and right feet were used to model the gait dynamics of patients with NDDs and healthy control subjects. The gait dynamics underlying the gait patterns were approximated by radial fundamental function (RBF) neural networks.

Daliri [13] developed an automated diagnostic system using gait dynamics. A feature selection strategy based on a genetic algorithm was used to select for important diagnostic features of NDDs. The selected features were classified using a boost vector machine into two groups: features characteristic of healthy people and those typical of patients with NDDs. It has been stated that the double support interval is the most important gait parameter in the diagnosis of NDDs.

Baratin et al. [14], proposed an automatic classification scheme for NDDs in which gait features were integrated into a wavelet space by using wavelet transform. The wavelet features were classified using the support vector machine method. The authors showed that asymmetry between gait features could be used for the detection of NDDs.

Gupta et al. [15] developed an efficient classification method using a new set of gait features for gait intervals identified using auto-

* Corresponding author.
E-mail addresses: berdas@baskent.edu.tr (Ç.B. Erdaş), esumer@baskent.edu.tr (E. Sümer), skibaroglu@baskent.edu.tr (S. Kibaroğlu).

https://doi.org/10.1016/j.bspc.2021.103069
Received 28 February 2021; Received in revised form 1 July 2021; Accepted 8 August 2021
Available online 18 August 2021
1746-8094/

correlation and cross-correlation features. Mutual information (MI) analysis was used, and 500 features were generated for the classification phase. A rule-based classifier technique based on a single decision-tree classifier was used to classify NDDs.

A handful of researchers have attempted to predict disease severity in Parkinson's disease [16–19]. Besides these, a few studies have investigated deep learning-based methods for the diagnosis of NDDs [20–22].

Zhao et al. [20] proposed a two-layer long short-term memory (LSTM) deep learning method for the detection of NDDs using walking data obtained by force-sensitive sensors. The model was based on the detection of ALS, HD, PD diseases using attributes of walking dynamics. To provide sufficient foot-changing knowledge, blocks of twenty samples were used for the experiments.

Lin et al. [21] sought a machine-learning solution to detecting NDDs such as ALS, HD and PD. They used blocks containing ten samples of walking dynamics obtained by force-sensitive sensors. Each block had an overlap of two-thirds of the block size. This increased the number of blocks to avoid missing any walking patterns. The blocks were converted into 2D with the recurrence plot technique, and PCA was used for feature enhancement. The features extracted were used to feed the AlexNet CNN deep learning algorithm, and the NDDs were classified.

Paragliola and Coronato [22] took a deep learning approach to the detection of NDDs, focusing on whole gait dynamic data rather than a block-based pattern recognition process. They clustered all the data obtained from each patient. Since the number of samples was different for each person, the data set of the subject with the maximum number of samples (310) was taken as the basis. Clusters with fewer instances were extended to 310 samples by applying zero padding to ensure proper operation of the long short-term memory (LSTM) and convolutional neural network (CNN) deep learning methods.

Thus, a limited number of deep learning-based studies [20–22] have addressed NDD detection, but with a focus on grouping samples. There have been no attempts to predict disease severity in NDDs nor to detect disease and predict severity simultaneously. In this study, we transformed one-dimensional gait features collected by ground reaction force (GRF) sensors into two and three dimensions to feed convolutional long short-term memory (ConvLSTM) and three-dimensional convolutional neural network (3D CNN) models and used the models to study classification and regression problems (disease detection and disease-severity prediction, respectively) in NDDs. With these methods, a single sample of gait dynamics was sufficient to detect disease and estimate its severity.

2. Methods

2.1. General overview

We studied disease detection and disease severity prediction problems in NDDs by applying ConvLSTM and 3D CNN deep learning methods to one-dimensional gait data. The one-dimensional gait data was converted to QR codes to give it a two-dimensional structure, making possible the use of a ConvLSTM model. Three-dimensional tensors were extracted using ConvLSTM, and a 3D CNN model was fed with these tensors. For disease detection, which is a classification problem, 5 different subproblems were created (Table 1). Since a disease-specific system is used to grade severity for each disease, three different subproblems were created, with data belonging only to the disease and its control group (Table 2).

To apply ConvLSTM, which takes 2-dimensional data as input, one-dimensional gait data was represented in two dimensions by transforming the data into QR codes. Each sample obtained from the original GRF sensor data was converted into QR codes sequentially (Fig. 1). Python was used for this conversion, choosing QR version 10 and medium error correction capability (Table 3).

In artificial intelligence, disease detection in NDDs is a classification problem, and the prediction of disease severity is a regression problem. We sought solutions for these problems using deep learning models. ConvLSTM was used for two purposes: as a decision-maker for the classification and regression problems and as an extractor to output three-dimensional data. ConvLSTM, by its nature, takes 2-dimensional data as input and extracts 3-dimensional tensors as output; these tensors are ideal for 3D CNN use. Through 3D CNN, the gait features converted into tensors were used to find solutions for the both classification and regression problems. Fig. 2 summarizes the use of ConvLSTM and 3D CNN.

2.2. Convolutional LSTM

ConvLSTM blends the LSTM and CNN approaches [26]. In essence, ConvLSTM takes advantage of the sequential data approach of the LSTM method and the pattern detection feature of CNNs. ConvLSTM departs from the basic CNN + LSTM strategy in that the convolution structures are extended both to the input-to-state transition and the transitions from state to state [27].

The ConvLSTM architecture designed for this study works with 100 × 100 QR codes. Each of the 100 × 100 single-channel QR codes in the QR data set was transformed into a 10 × 10 feature map. Starting with an initial ConvLSTM layer, 64 feature maps were obtained without distorting the original dimensions given as input. After that, this feature map, which had a shape of 64 × 100 × 100, was decreased to 64 × 33 × 33 by max pooling. This 64 × 33 × 33 feature map was first decreased with a ConvLSTM layer to 64 × 31 × 31 and then to 64 × 10 × 10 with a max-pooling layer. At this point, the architecture had 2 separate dense layers containing 6400 units and N units. The N value was 2 or 4 depending on the number of classes in the classification problem, and 1 was used in the regression problem. The 'mini-batch' dimension was 64, the learning rate was 0.01, and the optimizer was "Adam," and 30 epochs were run.

2.3. 3D CNN

To improve performance in both the classification and regression tasks, a 3D CNN was fed with the outputs of ConvLSTM. ConvLSTM generates a tensor that has a 3D cell shape as output. When provided as input to the 3D CNN structure, those 3D tensors match perfectly.

The 3D convolutional layer was added to the ConvLSTM architecture just before the flatten layer. For the hyperparameters of the 3D convolutional layer, we selected a kernel size of 32, a filter size of (1,3,3), and "same" as the padding. These hyperparameters do not change the feature map dimensions output by the ConvLSTM architecture. The hyperparameters and working logic used in the other stages were the same as those used in the ConvLSTM architecture.

Table 1
Classification subproblems.

Subproblem	Explanation
NDD	Each disease and control group are located
NDD vs Control	All diseases are grouped under a single label against control group
ALS vs Control	Only ALS patients are classified against the control group
HD vs Control	Only Huntington patients are classified against the control group
PD vs Control	Only Parkinson's patients are classified against the control group

Table 2
Regression subproblems.

Subproblem	Explanation	Severity Evaluation Scale
ALS & Control	Only ALS patients and control group	Duration [23]
HD & Control	Only HD patients and control group	UHDRS-TFC [24]
PD & Control	Only PD patients and control group	Hoehn & Yahr [25]

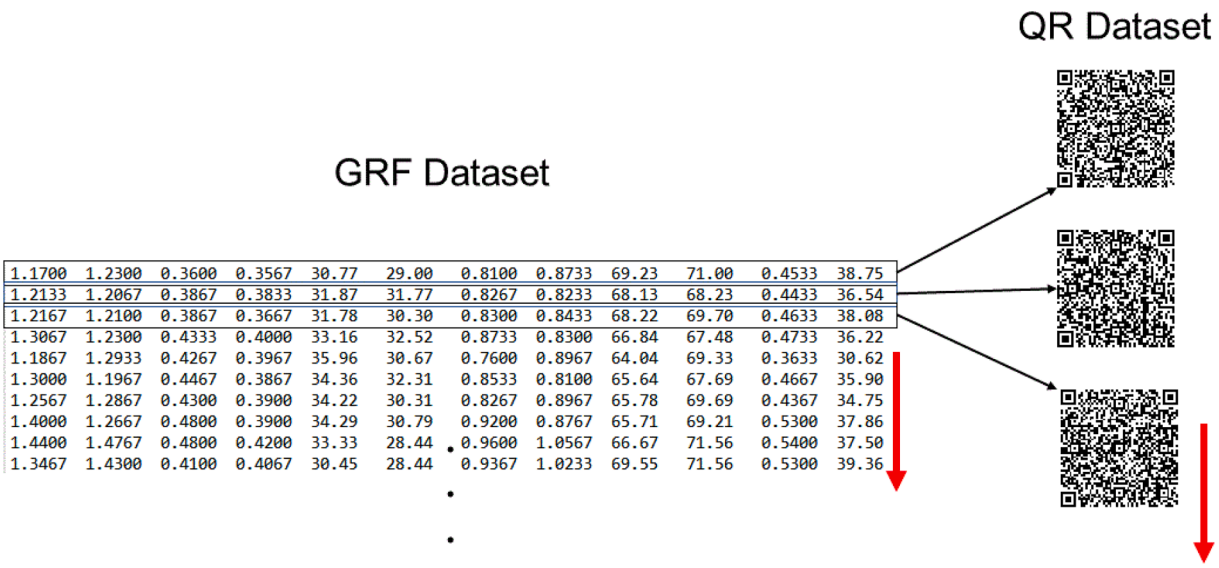

Fig. 1. The QR conversion.

Table 3
Gait features.

Index	Feature
1	Left Stride Interval (sec)
2	Right Stride Interval (sec)
3	Left Swing Interval (sec)
4	Right Swing Interval (sec)
5	Left Swing Interval (% of stride)
6	Right Swing Interval (% of stride)
7	Left Stance Interval (sec)
8	Right Stance Interval (sec)
9	Left Stance Interval (% of stride)
10	Right Stance Interval (% of stride)
11	Double Support Interval (sec)
12	Double Support Interval (% of stride)

2.4. Performance Evaluation

All of the experiments conducted were carried out using the k-fold cross-validation technique, in which the data set is divided into k equal parts with randomly determined samples. While one part is used for testing, the remaining parts are used for training. This process continues until each part has been used for testing, so that every part (and therefore every sample) is used for both testing and training [28]. The K value for this study was 10.

In order to measure the performance of the classification and regression experiments, accuracy, F_1 score, precision, and recall metrics were used for the classification tasks [28]. For the regression tasks, the correlation coefficient (R), coefficient of determination (R^2), mean absolute error (MAE), median absolute error (MedAE), root mean squared error (MSE), and root mean squared error (RMSE) metrics were used [29].

The accuracy-of-classification metric measures how often the model classifies a sample correctly. Accuracy is the number of correctly predicted samples out of all the samples. As shown in equation (1), accuracy can be described as the number of true positives and true negatives divided by the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

To fully assess the effectiveness of a model, precision and recall metrics should be included in addition to accuracy. These metrics are even more important for imbalanced data sets.

Precision can be expressed as the ratio of the true positives to all the positives (TP and FP), as shown in equation (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall can be described as a ratio of the true positives to the true positives plus false negatives, as shown in equation (3). It expresses the likelihood of the model identifying true positives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F_1 score is a metric used to measure the relationship (balance) between score precision and recall. Along with accuracy, it is often a good performance measure for imbalanced data sets. F_1 score, which is mathematically defined as the harmonic mean of precision and recall, can be formulated as in equation (4).

$$F_1 = 2*\frac{Precision*Recall}{Precision + Recall} \tag{4}$$

The correlation coefficient (R) is the most commonly used metric for regression problems. It is used to identify a relationship between two values, such as predicted and actual. R can range between −1 and 1. The value 0 indicates an absence of correlation, 1 a perfect positive correlation, and −1 a perfect negative correlation.

$$R = \frac{S_{ap}}{S_a S_p} \tag{5}$$

$S_{ap} = \frac{\sum_{i=1}^{n}(a_i-\bar{a})(p_i-\bar{p})}{n-1}$, $S_a = \frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1}$, $S_p = \frac{\sum_{i=1}^{n}(p-\bar{p})^2}{n-1}$, $\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$, $\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$. where a represents actual value and p stands for predicted value.

The coefficient of determination (R^2) is a metric used to describe how much variability a vector can have, based on its relationship to another factor. The coefficient of determination, commonly known as goodness of fit, takes values between 0 and 1. A value of 1 signifies a perfect fit, and 0 indicates that there is no fit. The formula for the coefficient of determination is shown in equation (6).

$$R^2(a,p) = 1 - \frac{\sum_{i=1}^{n}(a_i - p_i)^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2} \tag{6}$$

Mean absolute error (MAE) is the measure of the difference between two continuous variables. MAE is the average vertical distance between

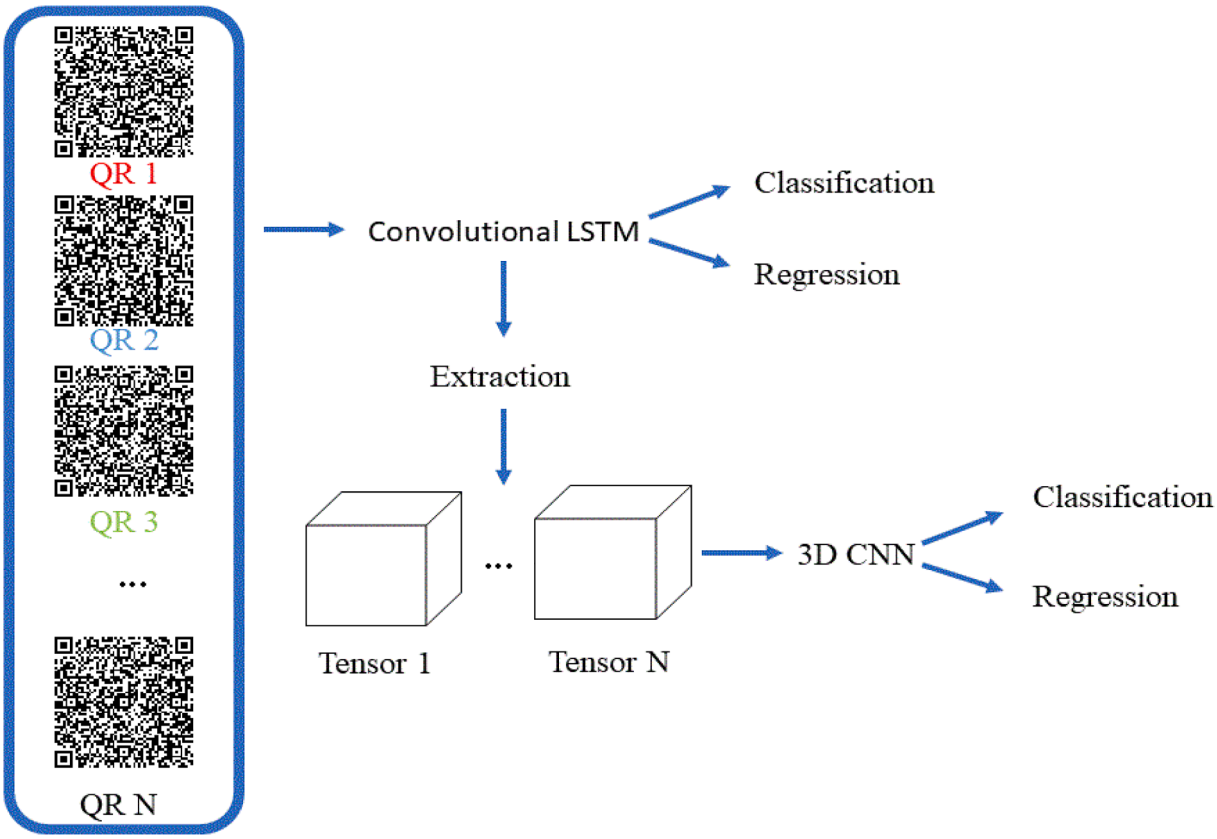

Fig. 2. Use of ConvLSTM and 3D CNN.

each actual value and the line that best fits the data. It is a linear score that measures the average magnitude of errors in a range of estimates without considering their direction. All individual errors are weighted equally on the mean. Equation (7) is the formula for the MAE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|a_i - p_i| \tag{7}$$

Median absolute error (MedAE) is a regression metric that minimizes the distortions caused by excessive and potentially erroneous outliers. MedAE is found by taking the median value of the absolute difference of all of the actual values and the values predicted by the model corresponding to the actual values. Its formula is given in equation (8).

$$MedAE = median(|a_i - p_i|, ..., |a_n - p_n|) \tag{8}$$

The mean square error (MSE) indicates how close a regression curve is to a set of points. It measures the performance of a machine learning mode. The predictor is always positive, and predictors with an MSE value close to zero perform better. Errors are squared before averaging, which imposes a high penalty on significant errors [30]. Eq. (9) shows the formula for the MSE.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)2 \tag{9}$$

The root mean square error (RMSE) is a quadratic metric that measures the magnitude of the error. It is often used to find the distance between predicted and actual values. It indicates that how dense that data is around the line that best fits the data. The formula for the RMSE is shown in equation (10).

$$RMSE = \sqrt[2]{MSE} \tag{10}$$

3. Materials

The data set used to test the methods proposed in this study was published by Hausdorff et al. [31–32] and has been made available on an open access basis by PhysioNet. There are 15,092 samples in this data set from 64 subjects: 13 patients with ALS, 20 patients with HD, 15 patients with PD, and 16 healthy individuals (controls). Each sample contains 12 gait characteristics (i.e., features) as well as timestamp information extracted from the raw data collected by GRF sensors. For this study, the timestamp information was removed, and experiments were carried out using the 12 gait features (Table 1). The gait features are illustrated in Fig. 3.

4. Results

Classification experiments were conducted for the five subproblems listed in Table 1, taking the diseases and their respective control groups one by one. For the regression experiments, solutions were sought for the three subproblems listed in Table 2.

Table 4 shows the classification results obtained with ConvLSTM. In the NDD subproblem, containing four different classes, the accuracy, F1 score, precision, and recall values were 0.8944, 0.8945, 0.8946, and 0.8933, respectively. In the NDD versus control subproblem, these values were 0.9633, 0.9532, 0.9561, and 0.9605, respectively. When the subproblems that featured a single disease and a control group were examined, the accuracy, F1 score, precision, and recall values for ALS patients, compared with controls, were 0.9768, 0.9755, 0.9762, and 0.9747, respectively. They were 0.9469, 0.9465, 0.9466, and 0.9463, for patients with HD, compared with controls, and 0.9505, 0.9552, 0.9513, and 0.9512, respectively, for patients with PD, compared with controls.

Table 5 provides the regression results obtained with ConvLSTM. The measured R, R^2, MAE, MedAE, MSE, and RMSE values for the ALS and controls subproblem were 0.9394, 0.8824, 2.4636, 1.3506, 20.6687,

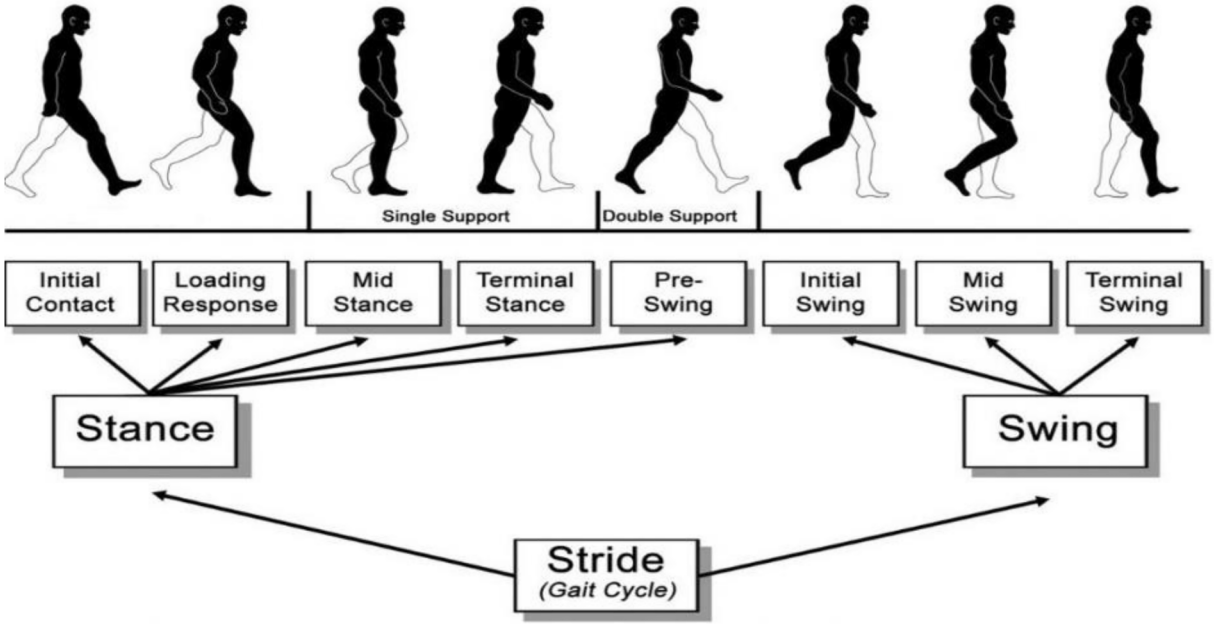

Fig. 3. The Gait Cycle.

Table 4
Classification results obtained with ConvLSTM.

	Accuracy	F_1 Score	Precision	Recall
NDD	0.8944	0.8945	0.8946	0.8933
NDD vs Control	0.9633	0.9532	0.9561	0.9605
ALS vs Control	0.9768	0.9755	0.9762	0.9747
HD vs Control	0.9469	0.9465	0.9466	0.9463
PD vs Control	0.9505	0.9552	0.9513	0.9512

Table 5
Regression results obtained with ConvLSTM.

	R	R^2	MAE	MedAE	MSE	RMSE
ALS & Control	0.9394	0.8824	2.4636	1.3506	20.6687	4.5462
HD & Control	0.9008	0.8115	1.1256	0.6362	3.7619	1.9395
PD & Control	0.9202	0.8267	0.3419	0.1991	0.3417	0.5845

and 4.5462, respectively. These values were 0.9008, 0.8115, 1.1256, 0.6362, 3.7619, and 1.9395, respectively, for the HD and controls subproblem and 0.9202, 0.8267, 0.3419, 0.1991, 0.3417, and 0.5845, respectively for the PDN and controls subproblem.

Table 6 shows the classification results obtained with the 3D CNN that was fed by ConvLSTM. The accuracy, F_1 score, precision and recall values obtained for the NDD subproblem were 0.8605, 0.8608, 0.8623, and 0.8579, respectively. These values were 0.9573, 0.9456, 0.9472, and 0.9442, respectively, for the patients with NDD compared with controls. When the subproblems that compared a single disease and a control group were examined, accuracy, F_1 score, precision, and recall values were 0.9768, 0.9755, 0.9256, and 0.9751, respectively, for the ALS patients compared with the controls. They were 0.9291, 0.9235, 0.929, and 0.9281, respectively, for the HD patients compared with the controls, and 0.9404, 0.9401, 0.9406, and 0.9397, respectively, for the patients with PD, compared with the controls.

Table 6
Classification results obtained with 3D CNN fed by ConvLSTM.

	Accuracy	F_1 Score	Precision	Recall
NDD	0.8605	0.8608	0.8623	0.8579
NDD vs Control	0.9573	0.9456	0.9472	0.9442
ALS vs Control	0.9768	0.9755	0.9256	0.9751
HD vs Control	0.9291	0.9235	0.929	0.9281
PD vs Control	0.9404	0.9401	0.9406	0.9397

Table 7 shows the regression results obtained with the 3D CNN that was fed by ConvLSTM. The R, R^2, MAE, MedAE, MSE, and RMSE values for the ALS and control subproblem were 0.9431, 0.8895, 2.5548, 1.5456, 19.4201, and 4.4068, respectively. They were 0.9042, 0.8175, 1.193, 0.6667, 3.6404, and 1.9079, respectively, for the HD and control subproblem and 0.9214, 0.849, 0.3542, 0.2228, 0.3365, and 0.58, respectively, for the PD and control subproblem.

For the NDD detection (classification) problem, a comparison of the proposed study with previous studies on the same data set is given in Table 8. The studies differed in their classifiers, feature dimensions, and feature extraction and assessment techniques. The machine-learning classifiers included deterministic learning theory, support vector machines, and decision trees, and the deep-learning classifiers were 2D and 3D CNN, LSTM, and ConvLSTM. In the machine-learning models, one-dimensional features were extracted using a RBF, a genetic algorithm, a discrete wavelet transform, and MI-based selection. In the deep-learning models, two-dimensional features were extracted using sample blocking, recurrence plots, and sample clustering. Three approaches were used for assessment: 10-fold cross-validation (10-fold CV), leave-one-out-cross-validation (LOOCV), and the split method (in which a certain proportion of the data set is reserved for training and the rest of the set is used for testing). Our method differed from other deep learning approaches in that we obtained 2D features from a single sample using QR code transformation. Then we extracted 3D tensors by using ConvLSTM fed with the QR codes. We sought a three-dimensional solution using the 3D tensors.

The multiclass NDD classification problem, in which all of the diseases and control groups were present, has been addressed only once before [23], with an accuracy of 0.61. We obtained an accuracy value of

Table 7
Regression results obtained with 3D CNN fed by ConvLSTM.

	R	R^2	MAE	MedAE	MSE	RMSE
ALS & Control	0.9431	0.8895	2.5548	1.5456	19.4201	4.4068
HD & Control	0.9042	0.8175	1.193	0.6667	3.6404	1.9079
PD & Control	0.9214	0.849	0.3542	0.2228	0.3365	0.58

Table 8
A comparison of the proposed study with previous studies using the same data set.

	Classifier	Feature Extraction	Feature Dimension	Assessment Technique	Accuracy
Zeng et. al. [12]	Deterministic Learning Theory	Radial Basis Function	1D	10-fold CV	NDD = NA NDD vs Control = 0.9375 ALS vs Control = 0.8966 HD vs Control = 0.8333 PD vs Control = 0.9710
Daliri et al. [13]	Support Vector Machine	Genetic Algorithm	1D	%50 Train %50 Test	NDD = NA NDD vs Control = 0.9063 ALS vs Control = 0.9679 HD vs Control = 0.9028 PD vs Control = 0.8933
Baratin et al. [14]	Support Vector Machine	Discrete Wavelet Transform	1D	%85 Train %15 Test	NDD = NA NDD vs Control = 0.8040 ALS vs Control = 0.8620 HD vs Control 0.8610 PD vs Control = 0.8710
Gupta et al. [15]	Decision Tree	Mutual Information	1D	LOOCV	NDD = NA NDD vs Control = 0.8750 ALS vs Control = 0.9620 HD vs Control = 0.8850 PD vs Control = 0.9230
Zhao et al. [20]	Dual channel LSTM	Sample Blocking	2D	LOOCV	NDD = NA NDD vs Control = 0.9504 ALS vs Control = 0.9725 HD vs Control = 0.9225 PD vs Control = 0.9680
Lin et al. [21]	CNN	Recurrence Plot	2D	LOOCV	NDD = NA NDD vs Control = 0.9893 ALS vs Control = 0.1000 HD vs Control = 0.9498 PD vs Control = 0.9421
Paragliola and Coronato [23]	LSTM & CNN	Sample Clustering	2D	%80 Train %20 Test	NDD = 0.6100 NDD vs Control = 0.8600 ALS vs Control = 0.9000 HD vs Control = 0.8200 PD vs Control = 0.95
Our Proposed ConvLSTM Approach	ConvLSTM	QR transform	2D	10-fold CV	NDD = 0.8944 NDD vs Control = 0.9633 ALS vs Control = 0.9768 HD vs Control = 0.9469 PD vs Control = 0.9505
Our Proposed 3D CNN Approach	3D CNN	Extract 3D tensors with ConvLSTM	3D	10-fold CV	NDD = 0.8605 NDD vs Control = 0.9573 ALS vs Control = 0.9768 HD vs Control = 0.9291 PD vs Control = 0.9404

0.8944 using our ConvLSTM approach and a value of 0.8605 using our 3D CNN approach. For the NDD versus control subproblem, the accuracy of machine learning methods working on 1D data varied between 0.8040 and 0.9375, and deep learning methods using multiple 2D samples had accuracy values between 0.8600 and 0.9893. In contrast, we obtained an accuracy of 0.9633 with ConvLSTM and 0.9573 with 3D CNN for this subproblem. For the ALS versus control subproblem, the machine learning methods had accuracy values between 0.8620 and 0.9620 and the values for the deep learning methods were between 0.9 and 1.0. Our accuracy values were 0.9768 and 0.9573 with ConvLSTM and 3D CNN, respectively. In the HD versus control subproblem, the accuracy values obtained with the machine learning methods varied between 0.8333 and 0.9028; with the deep learning methods, the minimum accuracy value was 0.8200 and the maximum was 0.9498. For this subproblem, our accuracy values were 0.9469 with ConvLSTM and 0.9291 with 3D CNN. Finally, for the PD versus control subproblem, accuracy was between 0.8710 and 0.9710 with the machine learning methods and between 0.9421 and 0.9680 with the deep learning methods. With our proposed ConvLSTM method, we achieved an accuracy value of 0.9505; the accuracy of 3D CNN was 0.9404.

5. Discussion

When the results in Tables 4 and 7 are examined, it can be observed that the ConvLSTM model performs better than the 3D CNN model for the disease detection and severity prediction problems. ConvLSTM inherently takes advantage of CNN's pattern capturing and LSTM's sequential data approach. It can extract important features from the data as CNN does and, like LTSM, make predictions for current data based on previous predictions. Since gait is an ongoing action, previous samples

of gait act on the current sample. In earlier studies, multiple samples were grouped into blocks to capture the ongoing action. The ConvLSTM method proposed in this study captures ongoing actions with a single sample by taking advantage of LSTM's sequential data approach. The main reason why the classification and regression performance of the 3D CNN method is lower than the ConvLSTM approach is that it cannot fully process the information that it is fed by the tensors obtained using the extractor feature of ConvLSTM.

When the results of the disease-detection classification problem were analyzed, ConvLSTM performed as well as 3D CNN in the ALS versus control subproblem and surpassed 3D CNN in all other subproblems. The reason why 3D CNN was as accurate as ConvLSTM in the ALS versus control subproblem may be because motor symptoms are more severe in ALS patients than PD and HD patients. Because the severe motor symptoms will affect the gait dynamics of patients with ALS, they are easier to identify with 3D CNN.

In the disease-severity prediction regression problem, the results achieved with ConvLSTM and 3D CNN were similar overall, and the ideal model changed based on the performance metric that was used. Since the regression problem was more difficult than the classification problem, the advantage of ConvLSTM decreased in this part of the study. ConvLSTM stood out in the average error criteria, including MAE and MedAE, and 3D CNN was superior in criteria without outlier tolerance, such as MSE and RMSE. ConvLSTM can be used when the mean error must be minimized (e.g., for the disease-severity estimation regression problem), and 3D CNN can be used in cases where it is desirable to minimize outliers. The greater impact of ALS, compared with HD and PD, on motor functions, may have increased the regression performance in the ALS and control subset. The Total Functional Capacity domain of the Unified Huntington's Disease Rating Scale (UHDRS-TFC), which was used to evaluate the patients in the HD and control subproblem, gives the highest score to healthy individuals and the lowest to patients with the most severe symptoms, unlike the scales used to assess patients with the other two diseases. Unlike the walking difficulties experienced by patients with ALS and PD, those characteristic of HD patients may not be observed at all times. The defining gait characteristic of HD is a sudden and slow motion twitching due to excessive bending of the knee and lifting of the leg. As this occurs sporadically, it may have affected the regression success in the HD and control subproblem.

The models used for both the disease-detection classification problem and the disease-severity estimation regression problem produced similar results. This can be explained by the saturation of the data set. The difference between the two models may be greater when addressing different problems and using different data sets. We predict that using 3D CNN fed by ConvLSTM for classification problems and ConvLSTM for regression problems will yield better performance results.

Since there are no comparable studies of the disease-severity estimation (regression) problem, only the results obtained for the classification problem could be compared with those of previous studies (Table 8). Because the studies differ in their sample numbers, feature dimensions, and assessment techniques, it is difficult to make direct comparisons. Nevertheless, both of the methods we tested provided above average performance compared with previous studies. This indicates that they may be sufficient for disease detection and estimation of disease severity using a single gait dynamics sample.

6. Conclusion

NDDs manifest themselves through the degeneration of nerve cells. Due to this degeneration, movement disorders may occur. They can be continuous, and they tend to increase during different activities. The different NDDs have different effects on gait. In addition, patients with more severe disease have more severe and frequent abnormal and involuntary movements when walking.

With advances in the field of artificial intelligence, many clinical decision-support systems have been developed. Obtaining data with sensors has become easier as the sensors have become smaller and consume less energy. This has led to the development of artificial intelligence-based approaches to the detection of NDDs and disease-severity prediction using gait dynamics data obtained from GRF sensors. Because of the dynamics of gait and the structure of the sensors, the data obtained are one-dimensional. Hence, these studies generally use machine learning for feature extraction and/or manipulation. Instead, we employed a ConvLSTM model, which combines the advantages of LSTM, a deep-learning model that is highly successful at handling time-series data, and a CNN, a deep-learning model that captures patterns in the data it receives as input. Both a multiclass problem (the NDD subproblem) and binary classification problems (NDD versus control, ALS versus control, HD versus control, and PD versus control), were solved using the proposed ConvLSTM method. In addition, the prediction performance achieved in regression subproblems (ALS and control, HD and control, and PD and control) was satisfactory and correlated with the classification performance. Although the classification and regression results obtained with 3D CNN did not surpass those obtained with ConvLSTM, 3D CNN has a disadvantage: it lacks the ability to use information that it has acquired from previous data.

The QR code technique that transforms unidimensional data into two dimensions can be used together with a 2D CNN deep-learning approach, which has advantages in pattern capture and feature learning, in future studies to detect NDDs and predict disease severity. Furthermore, we expect that the proposed QR code representation technique will be applicable to other research topics. In theory, all one-dimensional samples could be represented in 2 dimensions using the QR code technique. Therefore, two-dimensional techniques such as 2D CNN and ConvLSTM could be applied to that data. Moreover, all two-dimensional data can be converted into three-dimensional tensors using the extraction feature of ConvLSTM and, thus, can be processed with 3D CNN.

For detection of NDDs and prediction of their severities utilizing the motor symptoms, different data other than gait can also be used. In this context, as stated in study by Nishad et al. [33], signals from muscles can be considered in future studies to detect diseases and determine the severity of diseases.

CRediT authorship contribution statement

Çağatay Berke Erdaş: Methodology, Software, Writing - original draft, Funding acquisition. **Emre Sümer:** Supervision, Investigation, Project administration, Writing - review & editing. **Seda Kibaroğlu:** Conceptualization, Investigation, Writing - review & editing.

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.

Acknowledgment

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Project 120E380. The authors would like to thank TUBITAK for funding this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2021.103069.

References

[1] M.G. Erkkinen, M.-O. Kim, M.D. Geschwind, Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases, Cold Spring Harbor Perspect. Biol. 10 (4) (2018) a033118, https://doi.org/10.1101/cshperspect.a033118.

[2] M.-T. Heemels, Neurodegenerative diseases, Nature 539(7628) (2016) 179–179. doi: 10.1038/539179a.

[3] A. Ghanemi, Unifying the Common Concepts Shared by Neurodegenerative Diseases, J. Neurol. Stroke 2 (5) (2015), https://doi.org/10.15406/jnsk10.15406/jnsk.2.510.15406/jnsk.2015.02.00065.

[4] B.N. Dugger, D.W. Dickson, Pathology of Neurodegenerative Diseases, Cold Spring Harbor Perspect. Biol. 9 (7) (2017) a028035, https://doi.org/10.1101/cshperspect.a028035.

[5] S. Bilgin, The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects, Biomed. Signal Process. Control 31 (2017) 288–294, https://doi.org/10.1016/j.bspc.2016.08.016.

[6] O.C. Yurdakul, M.S.P. Subathra, S.T. George, Detection of Parkinson's Disease from gait using Neighborhood Representation Local Binary Patterns, Biomed. Signal Process. Control 62 (2020) 102070, https://doi.org/10.1016/j.bspc.2020.102070.

[7] M.A. Myszczynska, P.N. Ojamies, A.M.B. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J.D. Holbrook, L. Ferraiuolo, Applications of machine learning to diagnosis and treatment of neurodegenerative diseases, Nat. Rev. Neurol. 16 (8) (2020) 440–456, https://doi.org/10.1038/s41582-020-0377-8.

[8] S. Marziyeh Ghoreshi Beyrami, P. Ghaderyan, A robust, cost-effective and non-invasive computer-aided method for diagnosis three types of neurodegenerative diseases with gait signal analysis, Measurement 156 (2020) 107579. doi: 10.1016/j.measurement.2020.107579.

[9] Vincenzo Dentamaro, Donato Impedovo, Giuseppe Pirlo, Gait Analysis for Early Neurodegenerative Diseases Classification Through the Kinematic Theory of Rapid Human Movements, IEEE Access 8 (2020) 193966–193980, https://doi.org/10.1109/Access.628763910.1109/ACCESS.2020.3032202.

[10] Peyvand Ghaderyan, Seyede Marziyeh Ghoreshi Beyrami, Neurodegenerative diseases detection using distance metrics and sparse coding: A new perspective on gait symmetric features, Comput. Biol. Med. 120 (2020) 103736, https://doi.org/10.1016/j.compbiomed.2020.103736.

[11] Kirti Raj Bhatele, Classification of Neurodegenerative Diseases Based on VGG 19 Deep Transfer Learning Architecture: A Deep Learning Approach, Biosci. Biotechnol. Res. Commun. 13 (4) (2020) 1972–1980, https://doi.org/10.21786/bbrc10.21786/bbrc/13.410.21786/bbrc/13.4/51.

[12] Wei Zeng, Cong Wang, Classification of neurodegenerative diseases using gait dynamics via deterministic learning, Inf. Sci. 317 (2015) 246–258.

[13] Mohammad Reza Daliri, Automatic diagnosis of neuro-degenerative diseases using gait dynamics, Measur.: J. Int. Measur. Confederation 45 (7) (2012) 1729–1734.

[14] E. Baratin, L. Sugavaneswaran, K. Umapathy, C. Ioana, S. Krishnan, Wavelet-based characterization of gait signal for neurological abnormalities, Gait Posture 41 (2) (2015) 634–639.

[15] Kartikay Gupta, Aayushi Khajuria, Niladri Chatterjee, Pradeep Joshi, Deepak Joshi, Rule based classification of neurodegenerative diseases using data driven gait features, Health Technol. 9 (4) (2019) 547–560.

[16] T. Aşuroğlu, K. Açıcı, Ç. Berke Erdaş, M. Kılınç Toprak, H. Erdem, H. Oğul, Parkinson's disease monitoring from gait analysis via foot-worn sensors, Biocybernet. Biomed. Eng. 38 (3) (2018) 760–772, https://doi.org/10.1016/j.bbe.2018.06.002.

[17] M. Anila, Diagnosis of Parkinson's Disease using Artificial Neural Network, Int. J. Emerg. Trends Eng. Res. 8 (7) (2020) 3700–3707, https://doi.org/10.30534/ijeter/2020/131872020.

[18] Imanne El Maachi, Guillaume-Alexandre Bilodeau, Wassim Bouachir, Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait, Expert Syst. Appl. 143 (2020) 113075, https://doi.org/10.1016/j.eswa.2019.113075.

[19] Nooshin Abbasi, Seyed-Mohammad Fereshtehnejad, Yashar Zeighami, Kevin Michel-Herve Larcher, Ronald B. Postuma, Alain Dagher, Predicting severity and prognosis in Parkinsons disease from brain microstructure and connectivity, NeuroImage: Clin. 25 (2020) 102111, https://doi.org/10.1016/j.nicl.2019.102111.

[20] A. Zhao, J. Dong, J. Li, L. Qi, H. Yu, LSTM for diagnosis of neurodegenerative diseases using gait data, in: H. Yu, J. Dong (Eds.), Ninth International Conference on Graphic and Image Processing (ICGIP 2017), SPIE, 2018.

[21] C.-W. Lin, T.-C. Wen, F. Setiawan, Evaluation of vertical ground reaction forces pattern visualization in neurodegenerative diseases identification using deep learning and recurrence plot image feature extraction, Sensors (Basel, Switzerland) 20 (14) (2020) 3857.

[22] G. Paragliola, A. Coronato, A deep learning-based approach for the classification of gait dynamics in subjects with a neurodegenerative disease, in: Advances in Intelligent Systems and Computing, Springer International Publishing, Cham, 2021, pp. 452–468.

[23] A. Chio, G. Logroscino, O. Hardiman, R. Swingler, D. Mitchell, E. Beghi, On Behalf Of The Eurals Consortium, Prognostic factors in ALS: A critical review, Amyotrophic Lateral Sclerosis (2008) 1–14, https://doi.org/10.1080/17482960802566824.

[24] Douglas R. Langbehn, Steven Hersch, Clinical Outcomes and Selection Criteria for Prodromal Huntingtons Disease Trials, Mov. Disord. 35 (12) (2020) 2193–2200, https://doi.org/10.1002/mds.v35.1210.1002/mds.28222.

[25] I. Arcolin, M. Giardini, M. Godi, S. Corna, How spatiotemporal gait variables change along the different stages of Hoehn and Yahr scale in Parkinsons disease? Parkinsonism Related Disord. 79 (2020) e49, https://doi.org/10.1016/j.parkreldis.2020.06.195.

[26] S.A. Rahman, D.A. Adjeroh, Deep Learning using Convolutional LSTM estimates Biological Age from Physical Activity, Sci. Rep. 9 (1) (2019), https://doi.org/10.1038/s41598-019-46850-0.

[27] Çağatay Berke Erdaş, Selda Güney, Human Activity Recognition by Using Different Deep Learning Approaches for Wearable Sensors, Neural Process. Lett. 53 (3) (2021) 1795–1809, https://doi.org/10.1007/s11063-021-10448-3.

[28] Ç.B. Erdaş, I. Atasoy, K. Açıcı, H. Oğul, Integrating Features for Accelerometer-based Activity Recognition, Procedia Comput. Sci. 98 (2016) 522–527, https://doi.org/10.1016/j.procs.2016.09.070.

[29] A. Botchkarev, Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio, SSRN Electron. J. (2018), https://doi.org/10.2139/ssrn.3177507.

[30] Abhay Upadhyay, Manish Sharma, Ram Bilas Pachori, Rajeev Sharma, A Nonparametric Approach for Multicomponent AM–FM Signal Analysis, Circ. Syst. Signal Process. 39 (12) (2020) 6316–6357.

[31] J.M. Hausdorff, A. Lertratanakul, M.E. Cudkowicz, A.L. Peterson, D. Kaliton, A. L. Goldberger, Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis, J. Appl. Physiol. 88 (6) (2000) 2045–2053, https://doi.org/10.1152/jappl.2000.88.6.2045.

[32] Ary L. Goldberger, Luis A.N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, H. Eugene Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation [Online] 101 (23) (2000), https://doi.org/10.1161/01.CIR.101.23.e215.

[33] Anurag Nishad, Abhay Upadhyay, Ram Bilas Pachori, U. Rajendra Acharya, Automated classification of hand movements using tunable-Q wavelet transform based filter-bank with surface electromyogram signals, Future Generation Comput. Syst. 93 (2019) 96–110.