Original research

DIGITAL HEALTH

Digital Health
Volume 8: 1–14

DOI: 10.1177/20552076221075147
journals.sagepub.com/home/dhj
SAGE

CNN-based severity prediction of neurodegenerative diseases using gait data

Çağatay Berke Erdaş[1], Emre Sümer[1] and Seda Kibaroğlu[2]

Abstract

Neurodegenerative diseases occur because of degeneration in brain cells but can manifest as impairment of motor functions. One of the side effects of this impairment is an abnormality in walking. With the development of sensor technologies and artificial intelligence applications in recent years, the disease severity of patients can be estimated using their gait data. In this way, decision support applications for grading the severity of the disease that the patient suffers in the clinic can be developed. Thus, patients can have treatment methods more suitable for the severity of the disease. The presented research proposes a deep learning-based approach using gait data represented by a Quick Response code to develop an effective and reliable disease severity grading system for neurodegenerative diseases such as amyotrophic lateral sclerosis, Huntington's disease, and Parkinson's disease. The two-dimensional Quick Response data set was created by converting each one-dimensional gait data of the subjects with a novel representation approach to a Quick Response code. This data set was regressed with the convolutional neural network deep learning method, and a solution was sought for the problem of grading disease severity. Further, to demonstrate the success of the results obtained with the novel approach, native machine learning approaches such as Multilayer Perceptron, Random Forest, Extremely Randomized Trees, and K-Nearest Neighbours, and ensemble machine learning methods, such as voting and stacking, were applied on one-dimensional data. Finally, the results obtained on the prediction of disease severity by testing one-dimensional gait data with a convolutional neural network architecture that operates on one-dimensional data were included. The results showed that, in most cases, the two-dimensional convolutional neural network approach performed the best among all methods.

Keywords

Neurodegenerative diseases, gait, CNN, regression, QR

Submission date: 5 July 2021; Acceptance date: 3 January 2022

Introduction

Millions of people worldwide are affected by neurodegenerative diseases. Neurodegenerative disorders emerge whenever neurons in the brain lose authority over time and gradually perish. Although some of the physical or mental symptoms related to neurodegenerative diseases can be reduced by treatment, there is currently no cure to prevent the progression of the disease and no proven cure. Neurodegenerative diseases consist of many different sub-diseases according to the region where this degeneration occurs in the brain nerve cells and the affected ability.[1–6]

Amyotrophic lateral sclerosis (ALS) is a group of rare neurological diseases that involve nerve cells dedicated to the authority of muscle movement. ALS is among a larger group of disorders called neurodegenerative diseases, defined as motor neuron disorders caused by progressive degeneration and mortality of neurons. The ALS disease is progressive, which means that the symptoms get worse over time. There is currently no cure for ALS and no appropriate treatment to stop or avert the disease's progression.[7–10]

[1]Department of Computer Engineering, Faculty of Engineering, Başkent University, Ankara, Turkey
[2]Department of Neurology, Faculty of Medicine, Başkent University, Ankara, Turkey

Corresponding author:
Çağatay Berke Erdaş, Department of Computer Engineering, Faculty of Engineering, Başkent University, Bağlıca Kampüsü Fatih Sultan Mahallesi Eskişehir Yolu 18.km TR, Ankara 06790, Turkey.
Email: berdas@baskent.edu.tr

Huntington's disease (HD) is a progressive brain disease that leads to unmanaged movements, emotional issues, and deterioration of cognition skills. Huntington's disease characteristic findings are the loss of brain nerve cells. Huntington's disease, which is one of the diseases in the neurodegenerative disease group like ALS, currently has no cure or treatment which can halt, slow, or reverse the progression of the disease.[11–13]

Parkinson's disease (PD) is a brain disease that causes shaking, fatigue, and discomfort gait movement, balancing, and coordinating issues. Parkinson's symptoms generally get progressively worse over time. As the conditions progress, people may find it difficult to walk and talk. Parkinson's disease, like ALS and HD, is included in the group of neurodegenerative diseases.[14–18]

Diagnosing ALS, HD, and PD during a medical examination is an important process to increase the quality of life of patients in the fight against the disease. Also, after the diagnosis of the diseases, the follow-up of the relevant patients and the dosage of the drugs given to relieve the symptoms they experience are crucial. To complete this follow-up, various grading scales measure the severity of the disease to determine its progress. These scales are defined differently for each disease.

The revised ALS functional rating scale (ALSFRS-R) is an appealing primary predictor in clinical trials of ALS since it is valid, simple to handle, reduces dropout, lowers cost, and interacts with survival.[19] This scale, which includes ALSFRS elements to improve the ability to evaluate respiratory symptoms, is an examination that grades the degree of disability of ALS patients to function independently in daily living activities. To test bulbar function, motor function, and respiratory function, it consists of 12 items, and each item is scored from 0 (unable) to 4 (normal).[20]

For HD, the Unified Huntington Disease Rating Scale (UHDRS) is the gold standard clinical assessment method. UHDRS is a clinical rating scale for evaluating four areas of clinical performance and capacity (cognitive function, motor function, functional capacity, and behavioural abnormalities).[21] To assess the severity of limitation in functional ability and motor symptoms in HD, the UHDRS-TFC can be used. The UHDRS-TFC is a component of a multi-factor rating scale initially developed to examine both HD patients and individuals at risk for HD.[22] The UHDRS-TFC scale varies from 13 (normal) to 0 (severe disability). A higher score implies better functionality.

Various scales are available to measure the severity of Parkinson's symptoms and their effects on daily life. The Hoehn and Yahr scale (HY) is the gold standard in clinical evaluation and is widely used in research settings to measure motor and non-motor symptoms. The Unified Parkinson's Disease Rating Scale[23] is another common scale. The HY scale assesses the progression of Parkinson's symptoms and the extent of impairment.[24] For functional symptoms in PD, the HY scale is the most used and widely accepted staging system. It is used more frequently as it assesses the limitations of daily activities and non-motor symptoms. HY ranges from 0 (normal) to 5 (wheelchair-bound or bedridden unless aided.) A lower score implies better functionality.

Determining the severity of neurodegenerative diseases is a costly, time-consuming, and lengthy process that requires a specialist doctor. Despite the mentioned difficulties, it is critical to carry out this task for patients to live a more effective life, participate in daily activities, and improve their quality of life. More importantly, the correct determination of the severity of the disease means that the patient receives the correct dose of the drug. During the determination of disease severity, differences based on the patient's statement and the doctor's interpretation may be misleading in determining the correct severity. At this stage, an artificial intelligence-supported system, which is carried out entirely on gait data, can provide a more objective assessment and reduce the cost of traditional methods.

In this study, for disease severity prediction for neurodegenerative diseases such as ALS, HD, and PD, (i) one-dimensional (1D) raw gait data, (ii) two-dimensional (2D) Quick Response (QR) transformed gait data were fed by various machine and deep learning models. In addition to native machine learning models, such as Multilayer Perceptron, Random Forest (RF), Extremely Randomized Trees, K-Nearest Neighbours fed with 1D raw gait data. It was aimed to increase the regression performance obtained by the voting and stacking methods. Also, by using 1D raw gait data, a solution to the problem of disease severity prediction was sought using 1D convolutional neural network (CNN) deep learning modelling. Thanks to our method that provides a 2D representation of gait data by converting it into QR codes, the created QR dataset has been sought to solve the related regression problem by feeding the 2D dimensional CNN. The CNN was designed as a novelty to work with two dimensions and can exhibit all its advantages. Although there are different grading methods in the literature, within the scope of this study, to grade the disease, the duration for ALS, UHDRS-TFC for HD, and HY for PD were used.

Related studies

The studies added to the literature in the recent past have proven that the act of gait in various tasks is meaningful and informative. Specifically, the studies in the field of fall detection have revealed the importance of the phenomenon of gait.[25,26]

Sadeeh et al.[25] proposed a prototype of a patient-specific fall prediction and detection system with a single sensor. It would notify the patient to act if a fall case is predicted, but if a fall incident occurs, it would communicate with the health care providers. The detected fall events would be alarmed to the health care providers through Clouds to provide immediate help to the fallen elder. In another study,

according to the findings of Sadeeh et al.,[26] a mono triaxial accelerometer sensor bonded to a person's thigh is used to predict and detect fall incidents and saves their data for in-depth follow-up by a healthcare professional. If a fall incident is predicted, the patient will be informed, and the incidence of a fall occurrence would also send an alert to the concerned healthcare professionals via the internet.

With the use of gait data in different fields, several studies have been conducted to detect neurodegenerative diseases and rate disease severity. These studies were generally conducted on a single disease such as ALS, HD, or PD.

Du et al.[27] represented an evaluation of the longitudinal support vector regression (LSVR) to predict ALS's severity from demographic data. They advocate a unique longitudinal machine learning algorithm that monitors subjects by estimating additional weighted parameters at various time stamps. It enables the assessment of the SVR hyperplane parameters as well as the temporal trend parameters. The apparent need for methods that combine longitudinal modelling and machine learning paradigms motivated their development of LSVR.

Zhang et al.[28] assessed the usage of numerous electromyography (EMG) interference indicators for the discovery of surface EMG variations in ALS, thereby assisting the diagnosis of the disease. A newly developed clustering index, the kurtosis of the surface EMG amplitude distribution, and the kurtosis of the crossing rate were also among the indicators. Their research only focused on distinguishing ALS patients from neurologically intact patients. It remains to be seen whether the combined approach improves discrimination between myopathic and neurologically intact subjects. Furthermore, it is unclear whether the surface EMG examination can differentiate ALS from other neurogenic disorders.

Benassar et al.[29] aimed to apply machine learning and signal processing techniques to the development and evaluation of a low-cost, objective automated system for assessing the impairment of upper limb movement in HD. They introduce a framework for unbiased and ongoing assessment of motor impairment in HD patients while performing a novel upper limb task. The system focused on data gathered from tri-axial accelerometers worn during in the Money Box Test (MBT), a recently suggested assessment of bilateral, upper motor function. The recorded accelerometer signals were processed using signal processing and machine learning methods to generate an automatic Movement Impairment Score intended to reflect the degree of movement impairment during the MBT performance.

Gaßner et al.[30] aimed at objectively identifying the characteristics of the gait in HD patients utilizing a sensor-based gait analysis. The goal of this research was to compare gait characteristics in HD patients to age- and gender-matched controls using mobile sensor-based gait analysis. To clinically validate these objective parameters, gait parameters were correlated with clinical scores total motor score (TMS) and total functional capacity (TFC). The main finding showed that gait variability parameters representing disease-characteristic irregularity of gait were the most relevant parameters correlating with TMS and TFC.

Açıcı et al.[31] proposed to derive from their gait signals a set of time-domain and frequency-domain characteristics that would be useful in distinguishing normal and diseased individuals. They discuss a specific dilemma in which they seek a solution to classify whether an individual has PD or not, using signals obtained from a set of ground reaction force (GRF) sensors padded under the foot. They introduce an ensemble classification system that uses a variety of time-domain and frequency-domain features derived from pre-processed sensor signals.

Aşuroğlu et al.[32] introduced a machine learning model fed by GRF data collected from these gait sensors. They offered a hybrid model, named Locally Weighted Random Forest (LWRF), that provided for regression analysis on the numerical characteristics derived from an input signal. They offer a computational solution for quantifying PD motor symptoms. Their methodology, in particular, can deliver effective prognosis solutions based on gait analysis using multiple foot-worn sensors that measure GRF. To that end, they provide an LWRF regression model to predict the precise value of the severity of PD symptoms. The proposed local weighting scheme contributes to the elimination of the effects of interpatient variability in gait patterns.

As in the study by Saadeh et al.,[33] studies which tested a model on more than one disease and were successful are very limited in the literature. The main objectives of their study are to investigate and distinguish the gait dynamics of neurodegenerative disease patients and compare them to healthy individuals. They specifically evaluate how the three different types of neurodegenerative diseases (PD, HD, and ALS) impair the patient's ability to control the movement of their feet. They propose a wearable gait dynamics detection system that uses foot sensors implanted in the patient's shoe to analyse gait dynamics and identify the corresponding neurodegenerative disease. They propose a unique wearable neurodegenerative disease detection framework for ALS, HD, and PD. The suggested framework is based on minimal features, allowing for lower power and area requirements while maintaining overall detection accuracy. This framework is appropriate for the next generation of miniaturized wearable devices and can detect neurodegenerative disease at an early stage, potentially reducing its severity.

Materials and methods

Dataset

Within the scope of the studies, a data set made public by PhysioNet was used to test the algorithms developed so far. The data set created by Hausdorff et al.[34] consists of a total of 15,092 gait samples obtained from 64 participants consisting of 13 ALS patients, 20 Huntington patients, 15

Parkinson's patients, and 16 healthy individuals. Using force-sensitive resistors the raw data were acquired, with the output roughly proportional to the force underneath the foot. Stride-to-stride measurements of footfall contact instances were derived from these indicators.[35] At the end of this process, a total of 13 features are obtained and represented in Table 1.

Table 1. The features of gait.

Index	Contents
0	Elapsed Time (sec)
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)

The duration feature containing time step information was not used in this study, apart from the original data set. The 12 features mentioned are demonstrated on the right leg and shown in Figure 1, which includes phases of the normal gait cycle. The description criteria are the same as the right leg for the features obtained with the left leg.

The data set also contains demographic data on each subject, which can be seen in Table 2, such as age, gender, weight, height, body mass index, walking speed, and disease severity or duration. Disease severity or duration values vary according to the diseases. For ALS, the time (months) since the disease was diagnosed, UHDRS-TFC value for HD and HY for PD were used. For control, a value of 0 was assigned if it would be used in the ALS and Parkinson subsets, and a value of 13 if it would be used in the Huntington subset, based on the rules in the disease severity/duration disease scales. Data were obtained from power-sensitive pads located on the shoes of the subjects. At a measurement rate of 300 Hz, these sensors measured for the force differences utilized to the ground throughout walking. According to the information documentation, individuals were instructed to walk along the 77-m corridor for five minutes at their normal pace.[34]

Data are stated as mean ± standart deviation. ALS: amyotrophic lateral sclerosis, HD: Huntington's disease, PD: Parkinson's disease, Control: healthy control, BMI: body mass index.

General framework

The overview of the proposed method can be seen in Figure 2. Accordingly, studies conducted can be grouped under two different subtitles as 1D and 2D representation.

For 1D representation solutions, 15,092 1D samples containing 12 features were regressed with machine

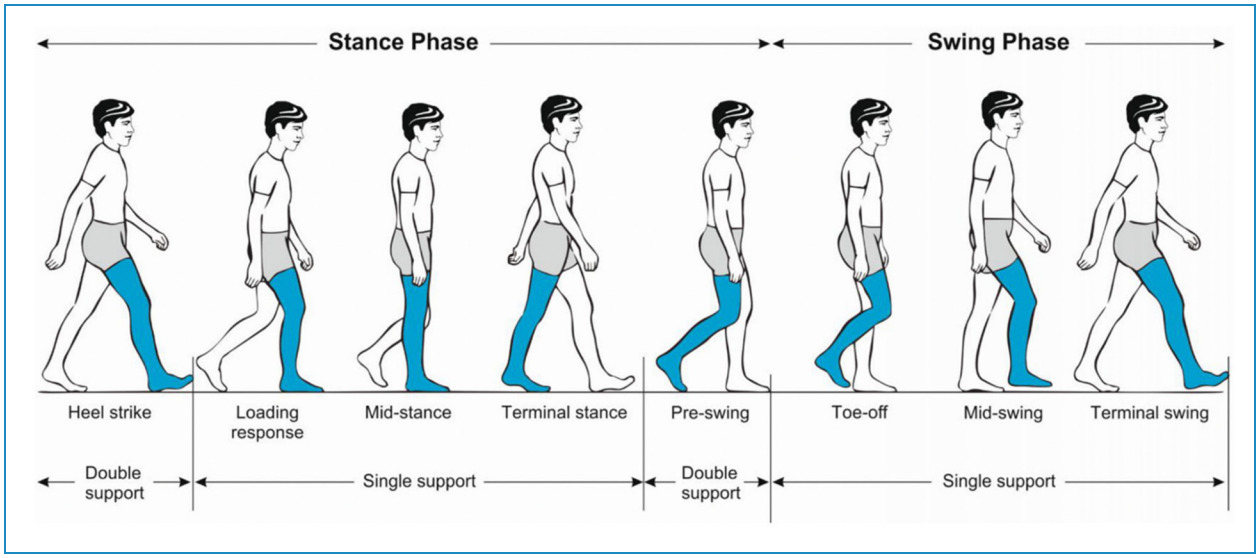

Figure 1. Phases of the normal gait cycle.[36]

Table 2. Demographic data of the subjects.

	ALS (n = 13)	HD (n = 20)	PD (n = 15)	Control (n = 16)
Age (years)	55.62 ± 12.83	47.37 ± 12.51	67.20 ± 10.69	38.69 ± 18.73
Gender (Male %)	66.7%	76.9%	30.0%	12.5%
Weight (kg)	77.12 ± 21.15	73.47 ± 16.24	75.07 ± 16.90	66.81 ± 11.08
Height (m)	1.7446 ± 0.950	1.8437 ± 0.089	1.87 ± 0.152	1.833 ± 0.087
BMI (kg/m^2)	25.21 ± 5.35	21.55 ± 4.44	21.21 ± 2.64	19.87 ± 2.71
Walking speed (m/s)	1.054 ± 0.218	1.15 ± 0.349	0.999 ± 0.202	1.354 ± 0.160
Disease severity or duration	19.474 ± 17.817	6.9 ± 3.837	2.8 ± 0.862	0-13

ALS: amyotrophic lateral sclerosis; HD: Huntington's disease; PD, Parkinson's disease.

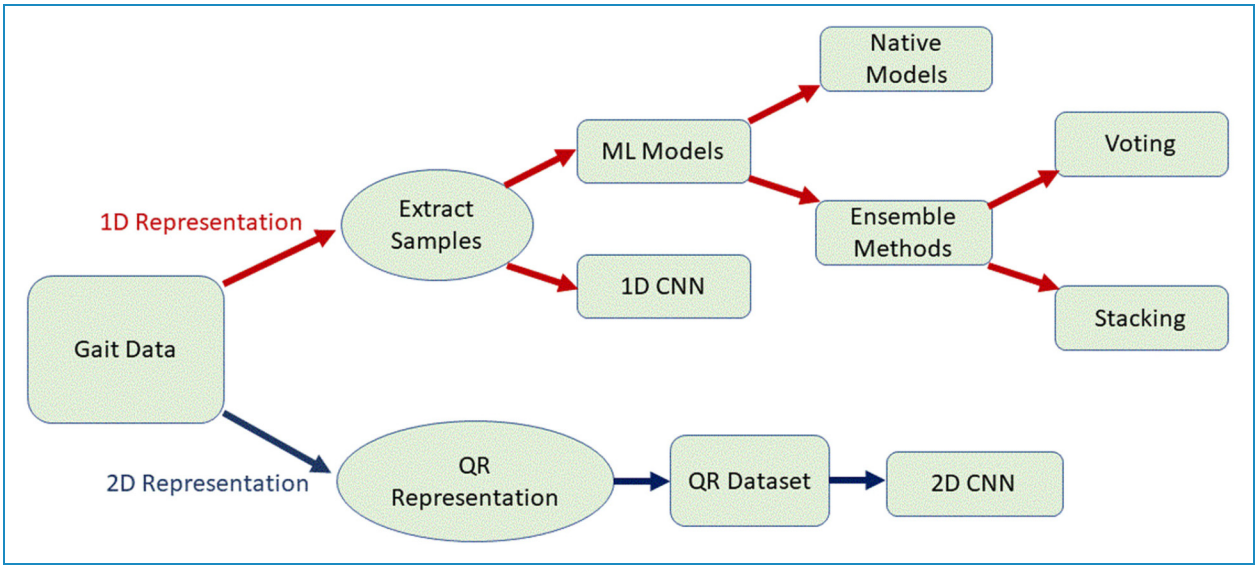

Figure 2. General framework of the proposed method.

learning algorithms such as Multilayer Perceptron, Extremely Randomized Trees, RF, and K-Nearest Neighbors (KNNs) as well as 1D CNN deep learning algorithm. To increase the performance of the native models, voting and stacking methods were applied with two native models showing the best performance for each sub-problem.

For the 2D CNN solution, all features except Elapsed Time from the gait samples obtained with the GRF sensor were first converted into QR codes and their 2D representation was achieved. Then, the solution to the regression problem was sought by using the obtained QR codes. Each of the 15,092 samples (2550 from ALS, 4846 from Huntington patients, 3620 from the PD, and 4076 from the control group) in the data set were converted to QR codes of 100×100 in the version 10 standard and thus a new QR code data set was created. The QR code generation process was completed using PYPI's QRcode library, which is an extension of Python for each line after all the gait data were concatenated. During this creation process, the version was selected as 10 standards, the error correction parameter was specified as 'm'. An illustration of the QR dataset creation is shown in Figure 3. Note that in Figure 3, each line corresponds to 1D data of GRF, and its QR codes generated are illustrated.

Methods

In this study, regression methods were used as a solution to the problem of grading the severity of neurodegenerative diseases. In this context, the disease severity grades of Parkinson's patients belonging to the HY scale, which is one of the international scales, and The TFC degree of Huntington patients, which is included in the UHDRS

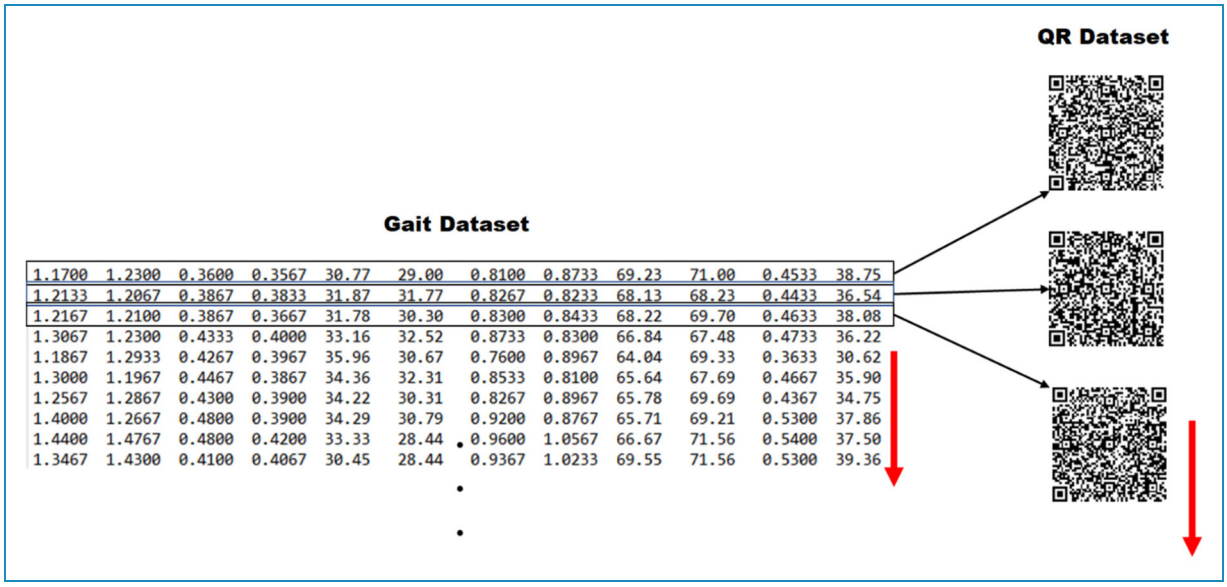

Figure 3. Quick Response (QR) dataset creation.

scale that shows the symptom severity, was tried to be estimated. Due to there is no information about disease degree for subjects with ALS, the number of months since the diagnosis of the disease has been tried to be estimated. Since ALS is a disease with progressive characteristics, there is no known treatment and available drugs only slow the progression of the disease. The number of months since diagnosis can be associated with the severity of the disease. The methods used in the experiments conducted to predict the disease severity of these diseases are grouped under three subheadings: Pure Machine Learning Regressors methods, Ensemble methods.

Machine learning regressors. One of the well-researched types of artificial neural networks is multilayer perceptron. Multilayer perceptron, which can be shortened as MLP, is a supervised, feed-forward, and hetero associative paradigm. In MLP, the weighted sum of the inputs and bias term is passed to the activation stage to estimate the actual value via a transfer function, and the units are organized in a layered as according to feed-forward neural network topology. MLP consists of three layers as input, hidden, and output and the connections between these layers. The number of neurons used in these layers, the weight of the connections, and the activation function can affect the working performance of the MLP.[37,38]

This research includes two random decision tree-based averaging algorithms, such as RF and Extremely Randomized Trees algorithms.[39] In both algorithms, the perturb-and-combine technique which is specially developed for trees is used. This technique is aimed to create a diverse classifier collection by adding randomness to the classifier structure. The ensemble's prediction is given as the average prediction of the individual classifiers.[40]

The methods based on the decision tree can serve as precise and effective means of analyzing motion signals. In particular, the RF approach, an advanced implementation of decision trees, has outperformed many powerful learning techniques which have proven themselves to be popular in other fields.[41] Each tree in the ensemble is constructed from a sample drawn with a replacement from the training set in RFs. Besides, during the construction of a tree, when splitting each node, the best split is found either from all input features or from a random size subset. These two parameters of randomness aim to reduce the forest estimator's variance. Nevertheless, individual decision trees generally produce an increased variance and attempt to overfit. In forests, the injected randomness yields decision trees with prediction errors rather than decoupled. Some errors can be avoided by taking an average of those predictions. By integrating diverse trees, RFs obtain a decreased variance, often at the expense of a small rise in bias. In practice, the reduction in variance is always important, resulting in a better overall model.[42]

Extremely Randomized Trees, which can be named Extra Trees (XT), generates numerous trees and splits nodes using random subsets of features. Randomness expands further in the way splits are calculated in extremely randomized trees.[43] Parallel with RFs, a subset of candidate features is used randomly, but thresholds are drawn at random for each candidate feature in place of searching for the most discriminatory thresholds, and the greatest one of these randomly created thresholds is chosen as the splitting principle. This typically helps the variance of the model to be minimized a little more, at the cost of a somewhat higher increase in bias.[44]

K-Nearest Neighbors is such an approach and tends to work reasonably well for large training sets, considering

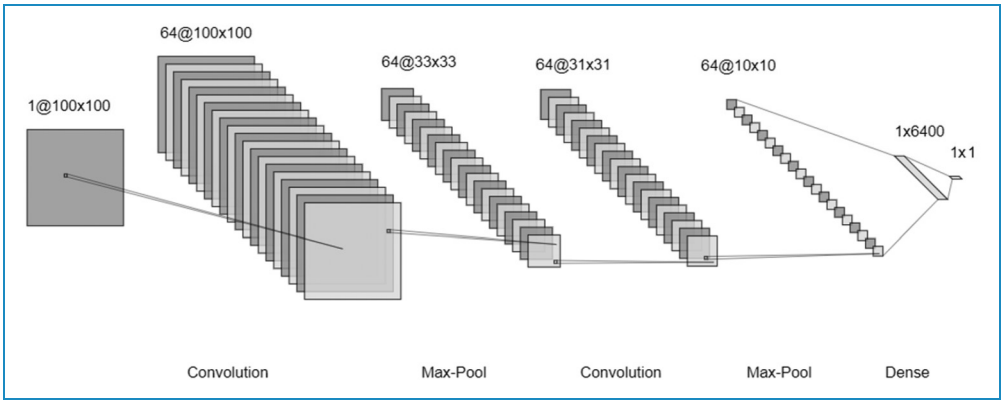

Figure 4. The two-dimensional (2D) convolutional neural network (CNN) architecture used for this study.

its simplicity. In essence, it relies only on a simple concept behind all predictions: experiments with similar characteristics would appear to have similar effects. In the training, Nearest Neighbor approaches allot a predicted value to a new observation based on the plurality or mean of its k 'Nearest Neighbors'. One of the key points of this approach is to choose the most optimal value for K. Choosing the wrong K value may degrade performance relatively. Since it decreases average noise, a greater K value is more accurate; but the consensus is that the various boundaries within the property domain are blurred. The correct selection of distance functions used to measure the relative proximity of the samples to each other is vital as well as the K value. Euclidean, Manhattan, and Minkowski are some of the distance functions that can be used.[45]

A CNN is one of the most popular algorithms for deep learning. It is a type of machine learning where a model learns to do classification tasks from pictures, video, text, or audio. CNNs are particularly useful for identifying patterns in images to recognize objects, faces, and scenes.[46] It can work directly on image files, uses patterns of image classification, and eliminates the need for manual feature extraction. For image recognition and pattern detection, CNN offers the most appropriate architecture.[47–49] There may be hundreds of layers in a CNN that detect and learn about various properties of an image or data. In each training step, filters are applied at various resolutions, and each layer's output is used as the input to the next layer. CNN architecture layers perform operations that modify data with data-specific learning characteristics. Convolution, Pooling, and Fully Connected are four of the most common layers. Convolution brings the images of the input into a sequence of convolution filters, each of which activates those image properties. Activation delivers faster and more effective preparation by mapping negative values to zero and retaining positive values. Thus, only active features are advanced to the next layer.[50] Pooling simplifies the output by conducting non-linear downsampling, thereby reducing the number of parameters that the network must learn. Transforming the 2D feature maps into a 1D feature vector is the job of the fully connected layer. These processes are replicated on tens or hundreds of layers. Each layer attempts to learn and identify various properties. After these layers, a single neuron added allows the CNN deep learning algorithm to predict the actual value.[51]

In this study, two different CNN architectures such as 1D and 2D were used in the experiments. In 1D CNN architecture, 12×1 data were processed by a 'Conv1D' layer with 512 filters using 'Relu' as activation, and the resulting feature maps passed through 'Flatten' and 'Dropout' layers and transferred to 'Dense' layers. Value estimation is made for the data passing through three separate 'Dense' layers with 2048, 1024, and 1 neuron, respectively.

The 2D CNN architecture designed for this study is shown in Figure 4. Each of the 100×100 single-channel QR codes in the QR data set was transformed into a 10×10 feature map using the architecture shown in Figure 4. With the convolution layer at the beginning, 64 feature maps were obtained without distorting the original dimensions given as input. After that, this feature map with a shape of $64 \times 100 \times 100$ has been decreased to $64 \times 33 \times 33$ by using the max pool layer. This feature map of $64 \times 33 \times 33$ sizes was first decreased with a convolutional layer to $64 \times 31 \times 31$ sizes and then to $64 \times 10 \times 10$ with a max-pooling layer. After this stage, the relevant architecture was supported with two separate dense layers containing 6400 units and 1 unit. Besides, the 'mini-batch' dimension is 64, the learning rate is 0.01, and the optimizer is 'adam', Both 1D and 2D CNN architecture have been run as 30 epochs.

Ensemble methods. The voting approach is a model of ensemble machine learning that incorporates the predictions of various other models. It is an approach that can be used to maximize the efficiency of the model, preferably achieving better performance than any other model used in the ensemble. The voting operates by unifying the predictions from multiple models. It can be useful for a set of effective

models to balance out their weaknesses.[52] All voting experiments conducted within the scope of this study were carried out with two of the native models with the best performance. Voting in regression is performed as the average of the outputs of the two models with the highest performance.

Stacking is one of the ensemble methods in the field of machine learning. The main principle of stacking is to train numerous models, generally, with varying types of machine learning algorithms, named as base learners, instead of selecting the best model or vote models, all base learner models aggregated final model, named as meta learner, to estimate predicted value. The meta learner executes the final prediction, fed by base learner outputs.[53] With two of the native models with the best performance, all stacking experiments performed within the framework of this study were carried out. In the related experiments, unlike voting, the stacking process works by feeding the linear regression model of the outputs of the two models with the highest performance.

Results

Performance evaluation

A ten-fold cross-validation setup assessed all the regression models. In the cross-validation technique, the dataset is split into ten equivalent sections. One of these sections is reserved for testing, while the remaining nine sections are used for training by the regression process.[49] The technique used six different metrics to evaluate the performance, correlation coefficient (R), R^2 score (coefficient of determination), mean absolute error (MAE), Median Absolute Error (MedAE), mean squared error (MSE), and root mean squared error (RMSE).

The coefficient of correlation, also known as Pearson's correlation coefficient, denoted as R, is the degree of relationship between two values actual and predicted.[32] If both actual and predicted values moving in unison, there is a correlation between them. If one tends to increase while the other tends to decrease, there are opposites between them. The coefficient of correlation can range from 1 to −1. 1 indicates that the two variables are moving in unison. They rise and fall together and have a perfect correlation. −1 means that the two variables are perfect opposites. 1 indicates that there is a perfect correlation between these two variables, while −1 indicates means that there are perfect opposites between these two variables. As can be seen in equation 1, the Coefficient of Correlation is equal to the division of the covariance of the actual and predicted value by the product of the standard deviations of the actual value and the estimated value. Normally, the R-value, which can vary between −1 and 1, is considered as strong when its absolute value is greater than 0.8, and as weak when its absolute value is less than 0.5.[54]

$$R = \frac{S_{ap}}{S_a S_p} \tag{1}$$

where $S_{ap} = \frac{\sum_{i=1}^{n} (actual_i - \overline{actual})\ (predicted_i - \overline{predicted})}{N-1}$,

$S_a = \frac{\sum_{i=1}^{n} (actual_i - \overline{actual})^2}{N-1}$, $S_P = \frac{\sum_{i=1}^{n} (predicted_i - \overline{predicted})^2}{N-1}$,

and $\overline{actual} = \frac{1}{N}\sum_{i=1}^{n} actual_i$, $\overline{predicted} = \frac{1}{N}\sum_{i=1}^{n} predicted_i$

R^2 score, denoted as R^2, computes the coefficient of determination. R^2 is a metric used for evaluating a regression model's performance.[55] It reflects the proportion of variance (of y) that the independent variables in the model have clarified. It offers an indicator of fit quality and thus a measure of how well-unseen samples, through the proportion of explained variance, are likely to be predicted by the model. R^2 must always be between 0 and 1 but higher values indicate a better result.[55] If $predicted_i$ is the predicted value of i th sample and $actual_i$ is the corresponding true value for whole n samples, the calculated R^2 is specified as equation 2:

$$R^2(a, p) = 1 - \frac{\sum_{i=1}^{n} (actual_i - predicted_i)^2}{\sum_{i=1}^{n} (actual_i - \overline{actual})^2} \tag{2}$$

where $\overline{actual} = \frac{1}{N}\sum_{i=1}^{n} actual_i$.

There are many evaluation criteria for the R^2 value, in which the values vary from the fields used in the literature.[53] Generally, If the R^2 value is greater than 0.5 and less than 0.7, it is considered as medium effect size (moderate), and when it is greater than 0.7, it is considered a strong effect size (substantial).[56]

Mean absolute error, which can be abbreviated as MAE, is one of the most successful loss functions used for regression models. MAE is generally defined as the absolute difference between the actual value and the value that the model predicts.[57] For outlier values, the

Table 3. Results obtained with native models on ALS & control.

Regressor	R	R^2	MAE	MedAE	MSE	RMSE
MLP	0.61	0.37	6.29	3.25	110.98	10.53
XT	0.73	0.53	4.01	0.14	85.11	9.23
RF	0.73	0.53	4.02	0.09	82.39	9.08
KNN	0.69	0.39	5.07	1.18	123.69	11.12

ALS: amyotrophic lateral sclerosis; KNN: K-Nearest Neighbor; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; RF: Random Forest; RMSE: root mean squared error; XT: Extra Trees.

Table 4. Results obtained with ensemble models on ALS & control.

Ensembler	Models	R	R^2	MAE	MedAE	MSE	RMSE
Voting	XT & RF	0.73	0.54	4.00	0.13	81.04	9.00
Stacking	XT & RF	0.46	0.21	7.64	5.29	128.50	11.34

ALS: amyotrophic lateral sclerosis; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; RF: Random Forest; RMSE: root mean squared error; XT: Extra Trees.

Table 5. Results obtained with native models on HD & control.

Regressor	R	R^2	MAE	MedAE	MSE	RMSE
MLP	0.37	0.14	3.48	3.22	17.19	4.15
XT	0.49	0.24	2.93	2.32	15.19	3.90
RF	0.52	0.27	2.09	2.32	14.63	3.82
KNN	0.74	0.54	3.65	2.00	30.77	5.55

HD: Huntington's disease; ; KNN: K-Nearest Neighbor; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; RF: Random Forest; RMSE: root mean squared error; XT: Extra Trees.

MAE is steadier and does not aim to penalize the errors as heavily as MSE or RMSE. MAE is a linear score that means that whole the variations between individuals are weighted equally.[57] The MAE formula can be observed in equation 3:

$$MAE = \frac{1}{N}\sum_{i=1}^{n} |actual_i - predicted_i| \tag{3}$$

Median Absolute Error, which can be truncated as MedAE, is one of the regression metrics that minimizes the penalties caused by excessive and potentially erroneous outliers.[58] In essence, MedAE is robust to outliers. MedAE can be defined as taking the median value of the absolute difference of all actual values and the values predicted by the model corresponding to the actual values. MedAE formula is given in equation 4:

$$MedAE = median(\ |actual_i - predicted_i|,\ \ldots,\ |actual_n - predicted_n|\ ) \tag{4}$$

Mean Square Error, which can be abbreviated as MSE, is one of the most used regression task metrics. In the field of artificial intelligence, MSE is outlined as the mean of the square of the difference between actual and predicted values. In many researches, it is favoured because the errors are squared first before averaging, which imposes a high penalty on significant errors.[59] Formally, MSE can be denned as equation 5:

$$MSE = \frac{1}{N}\sum_{i=1}^{n} (actual_i - predicted_i)^2 \tag{5}$$

Root Mean Squared Error, which can be shortened as RMSE, is generally defined as the square root of the average squared difference between the model's expected target value and value. RMSE is the square root of the MSE.[60] RMSE is useful in situations where massive errors are not desired, just like MSE. The RMSE formula is shown in equation 6:

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n} (actual_i - predicted_i)^2}{N}} \tag{6}$$

Empirical results

The studies conducted within the scope of this study were divided into three groups: (i) Native and Ensemble machine learning models, (ii) 1D CNN approach, and (iii) 2D CNN technique after using novel 2D representation which converts one-dimensional data into QR codes.

Machine learning models were tested as Native and Ensemble with three different subsets formed from the data set, namely ALS & Control, HD & Control, PD & Control.

Table 3 shows the regression results obtained by native machine learning regressors such as MLP, XT, RF, and KNN belonging to the ALS & Control subset. Accordingly, the best performance from these four regressors belongs to the RF, which surpasses the XT by a small margin. The results obtained with RF are 0.73, 0.53, 4.02, 0.09, 82.39, and 9.08 for R, R^2, MAE, MedAE, MSE, and RMSE, respectively.

It is aimed to increase the success of these two successful regressors by voting and stacking with XT and RF, which have the best results. As can be seen in Table 4, stacking

Table 6. Results obtained with ensemble models on HD & control.

Ensembler	Models	*R*	R^2	MAE	MedAE	MSE	RMSE
Voting	KNN & RF	0.32	0.10	3.04	1.92	17.92	4.23
Stacking	KNN & RF	0.54	0.29	4.48	4.48	25.70	5.07

HD: Huntington's disease; KNN: K-Nearest Neighbor; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; RF: Random Forest; RMSE: root mean squared error.

Table 7. Results obtained with native models on PD & control.

Regressor	*R*	R^2	MAE	MedAE	MSE	RMSE
MLP	0.68	0.46	0.87	0.70	1.20	1.10
XT	0.78	0.61	0.58	0.28	0.86	0.93
RF	0.78	0.62	0.86	0.28	0.86	0.93
KNN	0.72	0.52	0.72	0.44	1.08	1.04

KNN: K-Nearest Neighbor; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; PD, Parkinson's disease; RF: Random Forest; RMSE: root mean squared error; XT: Extra Trees.

failed, although the results obtained with voting with the other two native models slightly exceeded. Therefore, the best performance in the subset of ALS & Control groups was obtained by voting with XT and RF, and *R*, R^2, MAE, MedAE, MSE, and RMSE values were measured as 0.73, 0.54, 4.0, 0.13, 81.04, and 9.002, respectively.

The regression results obtained by MLP, XT, RF, and KNN from the machine learning regression models belonging to the HD & Control subset can be observed in Table 5. The best performance of these four regressors belongs to KNN, which surpasses other models by a large margin. The best performance in the subset of HD & Control groups was obtained by KNN, and *R*, R^2, MAE, MedAE, MSE, and RMSE values were measured 0.74, 0.54, 3.65, 2.00, 30.77, and 5.55, respectively.

The objective is to boost the performance of these two effective regressors such as KNN and RF by voting and stacking. As can be seen in Table 6, both voting and stacking failed. Native models are more successful in this sub-problem.

In Table 7, the regression results obtained by MLP, XT, RF, and KNN from the machine learning regression models belonging to the PD & Control subset can be observed. Accordingly, the one with the best performance of these four regressors is the RF that surpasses a slightly different XT.

The aim is to boost the achievement of these two successful regressors by voting and stacking XT and RF with both the best results. As can be seen in Table 8, stacking struggled, but while the results obtained by voting with the other two native models were slightly exceeded. The best performance in the subset of PD & Control groups was obtained by voting with XT and RF, and *R*, R^2, MAE, MedAE, MSE, and RMSE values were measured as 0.79, 0.62, 0.59, 0.30, 0.86, and 0,92, respectively.

Parameters in the methods used within the scope of the experiments are set according to the 'grid search' logic. In this context, as a result of the studies, in all experiments, while the number of neurons in the hidden layer in MLP is determined as 100 and the activation function as 'tanh'. In XT, which can reach 100 random trees, Entropy and best feature are chosen as the separation criteria in all experiments. Some parameters used in all experiments varied in RF and KNN. According to this, while the division criterion in RF is selected as Entropy in all experiments, the number of trees to be randomly generated is set to be 250, 450, 350 for ALS & CON, HD & CON, PD & CON, respectively. On the other hand, for KNN, the determined k value – distance measurement pairs are selected as 11 and Minkowski, 3 and Minkowski, 3 and Euclidean for ALS, HD, and PD, respectively.

Another method used in experiments on 1D data is the 1D CNN deep learning model. 1D CNN was fed with subsets in the dataset including ALS & Control, HD & Control, and PD & Control, just like ML models, and a solution to the regression-based disease severity prediction problem was sought. Accordingly, the results obtained using the subsets mentioned in Table 9 are given. Considering the results obtained, the performance achieved in any subset did not exceed the performance success achieved with the machine learning model. Although the results obtained in the HD & Control subset are by no means close to the most successful results achieved with ML, the results obtained from ALS & Control, and PD & Control subsets are relatively closer to ML performance.

Using the proposed novel method, which enables 1D data to be transformed into QR codes, expanding the size and reproducing the data, the results obtained with the 2D CNN run on the QR data are given in Table 10. Looking at the results, the performance achieved on the ALS & Control, HD & Control, PD & Control subsets can be observed.

As can be clearly seen in the results obtained within the scope of all experiments, the performance obtained with the

Table 8. Results obtained with ensemble models on PD & control.

Ensembler	Models	Features	R	R^2	MAE	MedAE	MSE	RMSE
Voting	XT & RF	0.79	0.62	0.58	0.29	0.85	0.92	0.79
Stacking	XT & RF	0.63	0.39	1.03	0.89	1.35	1.16	0.63

MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; PD, Parkinson's disease; RF: Random Forest; RMSE: root mean squared error; XT: Extra Trees.

Table 9. 1D CNN performance achieved for all subsets.

Subset	R	R^2	MAE	MedAE	MSE	RMSE
ALS & Control	0.61	0.37	6.25	3.48	110.26	10.50
HD & Control	0.39	0.15	3.27	3.05	16.90	4.11
PD & Control	0.7	0.49	0.79	0.56	1.14	1.07

ALS: amyotrophic lateral sclerosis; CNN, convolutional neural network; HD: Huntington's disease; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; PD, Parkinson's disease; RMSE: root mean squared error; 1D, one-dimensional.

Table 10. 2D CNN performance achieved for all subsets.

Subset	R	R^2	MAE	MedAE	MSE	RMSE
ALS & Control	0.88	0.79	3.96	2.55	37.48	6.12
HD & Control	0.83	0.69	1.86	1.44	6.19	2.49
PD & Control	0.79	0.62	0.71	0.58	0.85	0.92

ALS: amyotrophic lateral sclerosis; ; CNN, convolutional neural network; HD: Huntington's disease; MAE: mean absolute error; MedAE: Median Absolute Error; MSE: mean squared error; PD, Parkinson's disease; RMSE: root mean squared error; 2D, two-dimensional.

2D CNN deep learning model fed with data converted into QR codes by using the proposed novel approach was at least as successful as other methods such as ML and 1D CNN in all subsets.

If further elaborated, the best results obtained in the experiments based on the 1D principle on the ALS & Control subset are 0.73, 0.54, 4.0, 0.13 respectively, for the values of R, R^2, MAE, MedAE, MSE, RMSE via voting with XT and RF; while R, R^2, MAE, MedAE, MSE, and RMSE values obtained with 2D CNN were measured as 0.88, 0.79, 3.96, 2.55, 37.48, and 6.12, respectively. The performance difference between the two models can be clearly observed.

Similarly, the best results achieved in the experiments based on the 1D principle on the HD & Control subset are 0.74, 0.54, 3.65, 2.00, 30.77, and 5.55, for the values of R, R^2, MAE, MedAE, MSE, and RMSE via KNN; while R, R^2, MAE, MedAE, MSE, and RMSE values obtained with 2D CNN were measured as 0.83, 0.69, 1.86, 1.44, 6.19, and 2.49, respectively. The difference is noticeable, although not as great as in the ALS & Control subset.

At the PD & Control subset, the best performance with 1D data was measured with voting using XT and RF as 0.79, 0.62, 0.58, 0.29, 0.85, and 0.92 for R, R^2, MAE, MedAE, MSE, and RMSE, respectively; with 2D CNN, these values were calculated as 0.79, 0.62, 0.7, 0.58 0.85, and 0.92, respectively. As can be seen, although there were no obvious differences in this subset as in the other subsets, 2D CNN fed with the QR dataset achieved almost the same performance as the other models.

Discussion

When the results obtained are analyzed, it is observed that the superiority of 2D CNN over other methods in ALS & Control and HD & Control subsets is clearly observed, in the PD & Control subset, it has been observed that 1D methods have the same performance as XT and RF-based voting that has the best performance among all 1D-based approaches. Considering the results obtained in terms of Coefficient of Correlation with 2D CNN, the R values obtained for ALS & Control, HD & Control, and PD & Control were measured as 0.88, 0.83, and 0.79, respectively. In addition, the coefficient of determination, in other words, R^2 value was calculated as 0.79, 0.69, and 0.62 for the corresponding subsets, respectively. Based on the conditions of a successful evaluation of the criteria described in the performance evaluation section, R and R^2 values obtained for each subset can be considered successful. When MAE, MedAE, MSE, RMSE values are examined, the reason for the higher values in the ALS & Control subset compared to the other subsets can be explained as the related PD values vary between 0 and 4, HD values vary between 0 and 13, while ALS values change between 0 and 54. The range of values to be predicted and the size of the value itself cause the error to be larger in possible errors.

In addition, it is believed that the proposed QR representation method will be suitable for other research topics. Thus, by converting 1D data to two dimensions, the 2D-CNN deep learning method, which has strengths such

as pattern capture and feature learning, can be used. Moreover, since the QR transformation is the representation of the data as a binary, this structure can be fed with a feature extractor consisting of image metrics and these metrics can be operated with machine learning algorithms for classification and regression.

Conclusion

Predicting the severity of neurodegenerative diseases is an important and challenging task, as well as diagnosis. Grading the severity of neurodegenerative diseases depends on the subjectivity of the clinicians and the variability of the patient's response, albeit slight. Moreover, the correct prediction of severity plays a vital role in the treatment of the respective diseases. Since neurodegenerative diseases manifest themselves as lack or excess of movement, these diseases' severity can be predicted by using sensors placed in certain parts of the body without the need for any specialist supervision. The main purpose of this study is to estimate the severity of the disease experienced by people suffering from neurodegenerative diseases such as ALS, HD, and PD using gait data. Although doing this, the novel representation method, which converts 1D gait data into two-dimensional QR codes, is used to maximize the performance of the CNN deep learning algorithm and, moreover, to work more optimally for pattern recognition and feature extraction features. In addition, machine learning methods such as MLP, XT, RF, and KNN were applied on 1D gait data to compare the prediction results obtained with 2D CNN and to obtain possible promising results. Two different ensemble methods such as voting and stacking have been tried in order to improve the results obtained by machine learning methods. Moreover, to make a complete comparison, a solution to the problem of disease rating was sought by processing data about a CNN architecture that can operate on one dimension.

In future studies, the success of the proposed method will be retested by adding samples with other neurodegenerative disease labels to the dataset. In addition, classification and regression performances will be observed after applying the QR representation method to 1D data in different domains, making it suitable for deep learning methods that work in two dimensions such as CNN.

Acknowledgements: 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.

Conflicting interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Contributorship: All authors have made substantive contributions to this study. All authors endorsed the conclusions and approved the final version of the manuscript.

Ethical approval: NA

Funding: This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) [grant number:120E380].

Guarantor: ÇBE.

ORCID iD: Çağatay Berke Erdaş https://orcid.org/0000-0003-3467-9923

References

1. Moraru L, Moldovanu S, Dimitrievici LT, et al. Texture anisotropy technique in brain degenerative diseases. *Neural Comput Applic* 2018; 30: 1667–1677.
2. Kaur H, Malhi AK and Pannu HS. Machine learning ensemble for neurological disorders. *Neural Comput Applic* 2020; 32: 12697–12714.
3. Ren H, Yang Y, Gu C, et al. A patient suffering from neurodegenerative disease may have a strengthened fractal gait rhythm. *IEEE Trans Neural Syst Rehabil Eng* Sept. 2018; 26: 1765–1772.
4. Nahar N, Hossain MS and Andersson K. A machine learning based fall detection for elderly people with neurodegenerative disorders. In: Mahmud M, Vassanelli S, Kaiser MS and Zhong N (eds) *Brain informatics. BI 2020. Lecture notes in computer science, vol 12241*. Cham: Springer, 2020, pp. 194–203.
5. Marinescu RV, et al. Disease knowledge transfer across neurodegenerative diseases. In: Shen D, et al. (ed) *Medical image computing and computer assisted intervention – MICCAI 2019. MICCAI 2019. Lecture notes in computer science, vol 11765*. Cham: Springer, 2019.
6. Noor MBT, Zenia NZ, Kaiser MS, et al. Detecting neurodegenerative disease from MRI: a brief review on a deep learning perspective. In: Liang P, Goel V and Shan C (eds) *Brain informatics. BI 2019. Lecture notes in computer science, vol 11976*. Cham: Springer, 2019, pp. 115–125.
7. Gómez P, Londral ARM, Gómez A, et al. Monitoring ALS from speech articulation kinematics. *Neural Comput Applic* 2020; 32: 15801–15812.
8. Khamparia A, Singh A, Anand D, et al. A novel deep learning-based multi-model ensemble method for the prediction of neuromuscular disorders. *Neural Comput Applic* 2020; 32: 11083–11095.
9. Ramli AA, et al. BWCNN: blink to word, a real-time convolutional neural network approach. In: Song W, Lee K, Yan Z, Zhang LJ and Chen H (eds) *Internet of things – ICIOT 2020. ICIOT 2020. Lecture notes in computer science, vol 12405*. Cham: Springer, 2020, pp. 133–140.
10. Huang CH, Liao YT, Taniar D, et al. Computing a weighted jaccard Index of electronic medical record for disease prediction. In: Fujita H, Fournier-Viger P, Ali M and Sasaki

J (eds) *Trends in artificial intelligence theory and applications. Artificial intelligence practices. IEA/AIE 2020. Lecture notes in computer science, vol 12144*. Cham: Springer, 2020, pp. 445–456.
11. Hett K, Giraud R, Johnson H, Patch-Based abnormality maps for improved deep learning-based classification of Huntington's disease. In: Martel AL, et al. (ed) *Medical image computing and computer assisted intervention – MICCAI 2020. MICCAI 2020. Lecture notes in computer science, vol 12267*. Cham: Springer, 2020, pp. 636–645.
12. Li H, et al. Generalizing MRI subcortical segmentation to neurodegeneration. In: Kia SM, et al. (ed) *Machine learning in clinical neuroimaging and radiogenomics in neuro-oncology. MLCN 2020, RNO-AI 2020. Lecture notes in computer science, vol 12449*. Cham: Springer, 2020, pp. 139–147.
13. Perez M, Jin W, Le D, et al. Classification of Huntington disease using acoustic and lexical features. *Interspeech* 2018: 1898–1902. https://doi.org/10.21437/interspeech.2018-2029
14. Nagasubramanian G and Sankayya M. Multi-variate vocal data analysis for detection of Parkinson disease using deep learning. *Neural Comput Applic* 2020; 33: 4849–4864. https://doi.org/10.1007/s00521-020-05233-7
15. Oh SL, Hagiwara Y, Raghavendra U, et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput Applic* 2020; 32: 10927–10933.
16. Cai J, Liu A, Mi T, et al. Dynamic graph theoretical analysis of functional connectivity in Parkinson's disease: the importance of fiedler value. *IEEE J Biomed Health Inform* 2019; 23: 1720–1729.
17. Brewer BR, Pradhan S, Carvell G, et al. Application of modified regression techniques to a quantitative assessment for the motor signs of Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng* 2009; 17: 568–575.
18. Salarian A, Russmann H, Vingerhoets FJG, et al. Ambulatory monitoring of physical activities in patients With Parkinson's disease. *IEEE Trans Biomed Eng* 2007; 54: 2296–2299.
19. Kollewe K, Mauss U, Krampfl K, et al. ALSFRS-R score and its ratio: a useful predictor for ALS-progression. *J Neurol Sci* 2008; 275: 69–73.
20. Gordon PH, Miller RG and Moore DH. ALSFRS-R. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2004; 5: 90–93.
21. Winder JY, Achterberg WP, Gardiner SL, et al. Longitudinal assessment of the Unified Huntington's Disease Rating Scale (UHDRS) and UHDRS– for Advanced Patients (UHDRS–FAP) in patients with late stage Huntington's disease. *Eur J Neurol* 2015; 25: 780–785.
22. Huntington Study Group. Unified Huntington's disease rating scale: reliability and consistency. *Mov Disord* 1996; 11: 136–142.
23. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008; 23: 2129–2170.
24. Hoehn MM and Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967; 17: 427–442.
25. Saadeh W, Butt SA and Altaf MAB. A patient-specific single sensor IoT-based wearable fall prediction and detection system. *IEEE Trans Neural Syst Rehabil Eng* 2019; 27: 995–1003.
26. Saadeh W, Altaf MAB and Altaf MSB. A high accuracy and low latency patient-specific wearable fall detection system. *IEEE EMBS Int Conf Biomed Health Inform* 2017: 441–444.
27. Du W, Cheung H, Johnson CA, et al. A longitudinal support vector regression for prediction of ALS score. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). November 2015. https://doi.org/10.1109/bibm.2015.7359912.
28. Zhang X, Barkhaus PE, Rymer WZ, et al. Machine learning for supporting diagnosis of amyotrophic lateral sclerosis using surface electromyogram. *IEEE Trans Neural Syst Rehabil Eng* 2014; 22: 96–103.
29. Bennasar M, Hicks YA, Clinch SP, et al. Automated assessment of movement impairment in Huntington's disease. *IEEE Trans Neural Syst Rehabil Eng* 2018; 26: 2062–2069.
30. Gaßner H, Jensen D, Marxreiter F, et al. Gait variability as digital biomarker of disease severity in Huntington's disease. *J Neurol* 2020; 267: 1594–1601.
31. Açıcı K, Erdaş ÇB, Aşuroğlu T, et al. A random forest method to detect Parkinson's disease via gait analysis. In: Boracchi G, Iliadis L, Jayne C and Likas A (eds) *Engineering applications of neural networks. EANN 2017. Communications in computer and information science, vol 744 pp. 609-619*. Cham: Springer, 2017, pp. 609–619.
32. Aşuroğlu T, Açıcı K, Berke Erdaş Ç, et al. Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybern Biomed Eng* 2018; 38: 760–772.
33. Saadeh W, Altaf MAB and Butt SA. A wearable neurodegenerative diseases detection system based on gait dynamics. *2017 IFIP/IEEE Int Conf Very Large Scale Integr* 2017; 2017: 1–6.
34. Hausdorff JM, Lertratanakul A, Cudkowicz ME, et al. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *J Appl Physiol* 2000; 88: 2045–2053.
35. Hausdorff JM, Mitchell SL, Firtion R, et al. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. *J Appl Physiol* 1997; 82: 262–269.
36. Pirker W and Katzenschlager R. Gait disorders in adults and the elderly. *Wien Klin Wochenschr* 2016; 129: 81–95.
37. Ghanem WAHM and Jantan A. A new approach for intrusion detection system based on training multilayer perceptron by using enhanced Bat algorithm. *Neural Comput Applic* 2020; 32: 11665–11698.
38. Peyghami MR and Khanduzi R. Predictability and forecasting automotive price based on a hybrid train algorithm of MLP neural network. *Neural Comput Applic* 2012; 21: 125–132.
39. Yang F-J. An extended idea about decision trees. 2019 international conference on computational science and computational intelligence (CSCI). *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. December 2019: 349–354. https://doi.org/10.1109/csci49370.2019.00068
40. Gavankar SS and Sawarkar SD. Eager decision tree. 2017 2nd international conference for convergence in technology

(I2CT). *2017 2nd International Conference for Convergence in Technology (I2CT)*. April 2017: 837–840. https://doi.org/10.1109/i2ct.2017.8226246
41. Rajora M, Rathod M and Naik NS. Stroke prediction using machine learning in a distributed environment. In: Goswami D and Hoang TA (eds) *Distributed computing and internet technology. ICDCIT 2021. Lecture notes in computer science, vol 12582*. Cham: Springer, 2021, pp. 89–97.
42. Berk RA. Random forests. In: statistical learning from a regression perspective. In: *Springer series in statistics*. New York, NY: Springer, 2008, pp. 205–258.
43. Mishra G, Sehgal D and Valadi JK. Quantitative structure activity relationship study of the anti-hepatitis peptides employing random forest and extra tree regressors. *Bioinformation* 2017; 13: 60–62.
44. John V, Karunakaran NM, Guo C, et al. Free space, visible and missing lane marker estimation using the PsiNet and extra trees regression. *2018 24th International Conference on Pattern Recognition (ICPR)*. August 2018: 189–194. https://doi.org/10.1109/icpr.2018.8546108
45. Zhang M-L and Zhou Z-H. ML-KNN: a lazy learning approach to multi-label learning. *Pattern Recognit* 2007; 40: 2038–2048.
46. Bharati P and Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: a survey. In: Das A, Nayak J, Naik B, Pati S and Pelusi D (eds) *Computational intelligence in pattern recognition. Advances in intelligent systems and computing, vol 999*. Singapore: Springer, 2020, pp. 657–668.
47. Mahmoud M, Monshi A, Poon J, et al. Covidxraynet: optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. *Comput Biol Med* 2021; 133.
48. Sun L, Wang Z, Pu H, et al. Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection. *Comput Biol Med* 2019; 133.
49. Rosati R, Romeo L, Silvestri S, et al. Faster R-CNN approach for detection and quantification of DNA damage in comet assay images. *Comput Biol Med* 2020; 133.
50. Roy SK, Krishna G, Dubey SR, et al. HybridSN: exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 2020; 17: 277–281.
51. Dong P, Zhang H, Li GY, et al. Deep CNN-based channel estimation for mmWave massive MIMO systems. *IEEE J Sel Top Signal Process* 2019; 13: 989–1000.
52. Cao J, Lin Z, Huang G-B, et al. Voting based extreme learning machine. *Inf Sci (Ny)* 2012; 185: 66–77.
53. Džeroski S and Ženko B. Is combining classifiers with stacking better than selecting the best One? *Mach Learn* 2004; 54: 255–273.
54. Henseler J, Ringle CM and Sinkovics RR. The use of partial least squares path modeling in international marketing. In: *Advances in international marketing*. Bingley: Emerald Group Publishing Limited, 2009, pp.277–319.
55. Barrett JP. The coefficient of determination—some limitations. *Am Stat* 1974; 28: 19–20.
56. Hair F, Sarstedt M, Hopkins L, et al. Partial least squares structural equation modeling (PLS-SEM). *Eur Bus Rev* 2014; 26: 106–121.
57. Willmott C and Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 2005; 30: 79–82.
58. Chung N, Zhang XD, Kreamer A, et al. Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J Biomol Screening* 2008; 13: 149–158.
59. Köksoy O. Multiresponse robust design: mean square error (MSE) criterion. *Appl Math Comput* 2006; 175: 1716–1729.
60. Chai T and Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geosci Model Dev* 2014; 7: 1247–1250.