Feasibility of differentiating gait in Parkinson's disease and spinocerebellar degeneration using a pose estimation algorithm in two-dimensional video

Katsuki Eguchi[a,b,*], Hiroaki Yaguchi[a], Hisashi Uwatoko[a], Yuki Iida[b], Shinsuke Hamada[b], Sanae Honma[b], Asako Takei[b], Fumio Moriwaka[b], Ichiro Yabe[a]

[a]Department of Neurology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Kita 15, Nishi 7, Kita-ku, Sapporo, 060-8638, Japan

[b]Hokuyukai Neurological Hospital, 4-30, 2jo, 2cho-me, Nijuyonken, Nishi-ku, Sapporo, 063-0802, Japan

***Corresponding author:** Katsuki Eguchi

Department of Neurology, Faculty of Medicine and Graduate School of Medicine

Hokkaido University, Kita 15, Nishi 7, Kita-ku, Sapporo 060-8638, Japan

E-mail: k198762@gmail.com

Abbreviations 1

PD, Parkinson's disease; SCD, spinocerebellar degeneration; MSA, multiple system atrophy; HCA, hereditary cerebellar ataxia; SAOA, sporadic adult-onset ataxia; MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; H-Y stage, Hoehn and Yahr stage; SARA, Scale for the Assessment and Rating of Ataxia; BMI, body mass index; DNN, deep neural network; AUC, area under the receiver operating characteristic curve

Abstract

Background: Although pose estimation algorithms have been used to analyze videos of patients with Parkinson's disease (PD) to assess symptoms, their feasibility for differentiating PD from other neurological disorders that cause gait disturbances has not been evaluated yet. We aimed to determine whether it was possible to differentiate between PD and spinocerebellar degeneration (SCD) by analyzing video recordings of patient gait using a pose estimation algorithm.

Methods: We videotaped 82 patients with PD and 61 patients with SCD performing the timed up-and-go test. A pose estimation algorithm was used to extract the coordinates of 25 key points of the participants from these videos. A transformer-based deep neural network (DNN) model was trained to predict PD or SCD using the extracted coordinate data. We employed a leave-one-participant-out cross-validation method to evaluate the predictive performance of the trained model using accuracy, sensitivity, and specificity. As there were significant differences in age, weight, and body mass index between the PD and SCD groups, propensity score matching was used to perform the same experiment in a population that did not differ in these clinical characteristics.

Results: The accuracy, sensitivity, and specificity of the trained model were 0.86, 0.94, and 0.75 for all participants and 0.83, 0.88, and 0.78 for the participants extracted by propensity score matching.

Conclusion: The differentiation of PD and SCD using key point coordinates extracted from gait videos and the DNN model was feasible and could be used as a collaborative tool in clinical practice and telemedicine.

Keywords

Parkinson’s disease, spinocerebellar degeneration, ataxia, gait disturbance, machine learning

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease and is characterized by resting tremor, bradykinesia, muscle rigidity, and responsiveness to dopaminergic treatment [1,2]. The prevalence of PD has increased worldwide over the last 30 years [3], making accurate diagnosis and treatment increasingly important. From the early stages of the disease, patients exhibit an abnormal gait pattern characterized by shortened stride length, increased stride length variability, and decreased walking speed [4,5]. In addition, decreased arm swing and increased asymmetry of limb movements are more specific features of gait in PD and may be the first motor symptoms [6]. Therefore, gait analysis has the potential to be effective in the diagnosis of PD.

Various objective assessment methods have been used to analyze the gait of patients with PD. Motion capture is one of the most commonly used methods to evaluate gait [7]. This technique quantitatively measures spatiotemporal parameters of gait, such as stride length, gait velocity, and cadence, and has been reported to reveal significant differences in these parameters between patients with PD and healthy controls [4]. Another widely practiced method is the analysis of data obtained from wearable devices [7,8]. The advantage of wearable devices is that real-life data of patients can be obtained over time, and attempts have been made to use measurement devices, such as

accelerometers in smartphones and smartwatches [9,10]. However, these methods have limitations. Because motion capture systems are usually expensive and usable only in laboratory environments, they have not yet been applied in clinical practice. Wearable devices can be used outside of hospitals and laboratories; however, wearing these devices could change patients' movement patterns, and therefore, the results may not reflect gait in natural settings [11]. Additionally, PD is common among the elderly, and the advanced stage of the disease is accompanied by cognitive decline; it is thought that some patients may not be able to equip wearable devices appropriately.

An alternative method for evaluating gait in PD is to analyze two-dimensional (2D) videos by applying pose estimation algorithms. Pose estimation is a computer vision technique that detects human joints or key points in images and videos [12]. One of the advantages of this technique is that it does not require patients to wear a measurement device, making it a noninvasive and simple way to evaluate gait. Additionally, as 2D video cameras are available in smartphones and are widely used, they are easily accessible at a low cost. In previous studies, it was reported that using pose estimation algorithms to analyze videos of patients with PD allowed for an assessment of the severity of resting tremor [13], bradykinesia of the extremities [13,14], dyskinesia [15], and gait [16]. In addition, it has been reported that it was possible to differentiate

patients with PD from healthy controls by analyzing gait videos using a pose estimation algorithm [17]. Therefore, the analysis of gait videos using pose estimation algorithms is expected to be useful in differentiating PD from other neurological disorders that cause gait disturbance; however, few studies have verified this. If video analysis with pose estimation algorithms is shown to be useful in differentiating gait disorders, it could be used as a noninvasive, easy-to-use diagnostic aid for PD.

Apart from parkinsonian gait, ataxic gait, which is characterized by wide step width, trunk instability, and increased variability of temporal-spatial characteristics, is one of the most common gait disorders caused by neurological diseases [18]. Since neurologists can usually distinguish between parkinsonian and ataxic gait by examining the patient's gait, we hypothesized that these distinctions could also be made by using gait videos and a pose estimation algorithm. In this study, we aimed to assess whether it is possible to differentiate PD from spinocerebellar degeneration (SCD), which is a typical disease presenting with ataxic gait, using gait videos and a pose estimation algorithm. We considered that such an investigation would be a milestone in the development of a method to differentiate PD from other disorders that cause gait disturbances using gait videos and pose estimation algorithms.

2. Materials and methods

2.1 Participants

Eighty-two patients with PD and 61 patients with SCD were recruited for participation in this study. The diagnosis of PD was based on the United Kingdom Parkinson's Disease Society Bank Criteria [19]. Patients with multiple system atrophy (MSA), hereditary cerebellar ataxia (HCA), or sporadic adult-onset ataxia (SAOA) were included. The diagnosis of MSA was based on consensus diagnostic criteria [20]. Only patients with the cerebellar variant (MSA-C) were included. Patients with HCA were defined as those who presented with ataxic signs on neurological examination and had a family history of similar ataxia syndromes in at least one first- or second-degree relative. The diagnostic criteria for SAOA were as follows: (1) presenting with progressive ataxia, (2) disease onset after 20 years of age, (3) no similar disorders in first- or second-degree relatives, (4) no established cause of ataxia (no ischemia, hemorrhage, or tumor of the posterior fossa; no alcohol abuse; no chronic intake of anticonvulsant drugs; no other toxic causes; no malignancies; no evidence of ataxia onset in association with encephalitis, sepsis, hyperthermia, or heat stroke; normal thyroid function), (5) no subacute ataxia onset, (6) and no probable MSA based on the consensus diagnostic criteria [20]. The exclusion criteria were as follows: (1) inability

to walk without manual assistance, (2) inability to follow instructions, (3) a history of central nervous system disorders other than PD or SCD, and (4) symptomatic musculoskeletal diseases (such as acute bone fracture, spinal canal stenosis, and osteoarthritis). All procedures were conducted after obtaining written informed consent from the participants, and the study was approved by the Institutional Review Board of Hokkaido University Hospital (approval number: 022-0205).

2.2 Gait Video Recording

The video recording environment is shown in Fig. 1. We videotaped the participants performing the timed up-and-go test (round-trip of 5 m). The recordings were conducted in a flat hallway, and videos were recorded from the frontal view. After video recording began, the participants were instructed to stand up from their chairs, walk directly toward the camera, turn around, walk directly away again, and sit back down on their chairs. The camera was placed on a tripod located 8 m from the chair and 3 m from the turning point. Videos were recorded using a consumer-grade video camera (HDR-CX470, Sony Corporation, Tokyo, Japan) at 30 fps with a resolution of 1280 × 720 pixels in MP4 format. If needed, the participants were permitted to use a cane or handrail during video recording. Video recordings of patients with PD were performed

in the on-drug state. Three videos were captured in each patient using the same steps. The patients were allowed appropriate rest between video recordings.

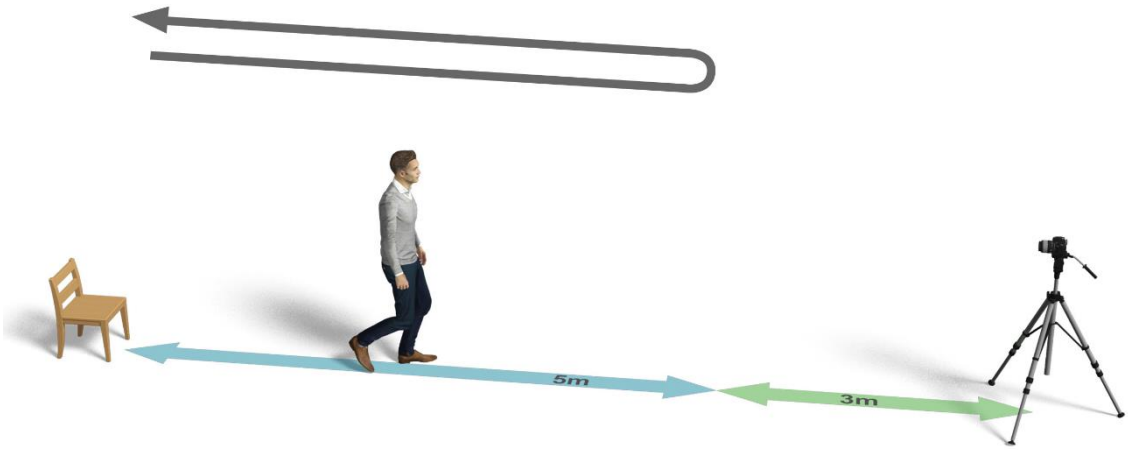

Fig. 1 Environment in which gait videos were recorded

Gait video recordings were conducted in a flat hallway. Before the video recording began, the patient waited on a chair. The camera was placed on a tripod located 8 m from the chair. After the video recording began, the participant was instructed to stand up from the chair, walk directly toward the camera, turn 3 m before the camera, walk directly away again, and sit back down on the chair.

2.3 Clinical Assessment

The patients with PD were assessed using the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) part III [21]

and Hoehn and Yahr (H–Y) stage, whereas those with SCD were assessed using the Scale for the Assessment and Rating of Ataxia (SARA) [22]. The participants' height, weight, and body mass index (BMI) were also measured.

2.4 Key Point Extraction with Pose Estimation Algorithm

We used the OpenPose library [23], a deep learning-based pose estimation algorithm, to extract the key points of the patients from the recorded videos. OpenPose extracts the 2D coordinates of 25 key points of a person in each frame when analyzing a video (i.e., 50 values were output per frame). The key points were as follows: eye, ear, shoulder, elbow, wrist, hip, knee, ankle, heel, big toe, little toe (both sides of the aforementioned parts were considered key points), nose, neck, and middle hip. As an example, the key points extracted from two participants' gait videos are shown in Supplementary Videos 1 and 2. Given a video length of T seconds, and because the frame rate of the videos was 30 fps, a matrix measuring 30 $T \times 50$ in size was obtained. The extracted key point coordinates were used as inputs in the deep neural network (DNN) model (see **Creating datasets and training the model** section).

2.5 DNN Model

We attempted to differentiate PD from SCD by inputting the key point coordinates, which were extracted using the pose estimation algorithm, into a transformer-based DNN model. The transformer architecture is characterized by a succession of self-attention layers that compute the distance between each pair of elements within the input sequence [24]. This architecture excels in handling serial data and was originally proposed for natural language processing. Recently, it was reported that a DNN model based on the transformer architecture performed effectively when used to analyze time series data [25]. Since the output of OpenPose is time series data of patients' key points during gait videos, we considered that the transformer architecture may successfully extract features that distinguish PD from SCD based on the key point data; therefore, we employed a transformer-based DNN model in this study. The detailed structure of the DNN model used in this study is described in Supplementary Materials and Methods.

2.6 Creating Datasets and Training the Model

Due to the relatively small number of participants, we employed a leave-one-participant-out cross-validation method. First, one participant was selected as the test patient. Subsequently, the data of the key point coordinates obtained from the remaining participants were randomly split into the training (80%) and validation (20%) sets, with

the ratio of patients with PD and SCD maintained. The training set was used to train the model, whereas the validation set was used to improve hyperparameters, such as the number of training epochs. The model predicted whether the test patient had PD or SCD using the parameters that exhibited the best prediction performance in the validation data. This procedure was repeated so that all patients were selected as the test patient once. Of note, there was no patient overlap between the training patients and the test patient.

For preprocessing, the x- and y-coordinates of the 25 key points were normalized by subtracting the mean and dividing by the standard deviation, which were both computed for all the coordinates of all frames contained in the training dataset. The coordinates of key points that were not recognized by OpenPose, resulting in missing values, were interpolated with the value "zero."

We clipped the time series data of the key point coordinates in each video every 5 s and used the resulting 5-s data as input for the DNN model. The number of generated 5-s clips from one video was dependent on its length. These 5-s clips were labeled as PD if they were generated from a video of a patient with PD or as SCD if they were from a patient with SCD. Since a 5-s video contained 150 frames, the input into the model was a matrix measuring 150 × 50 in size. The training data were first divided into 7.5-s data

(225 frames) with a 50% overlap. Subsequently, for data augmentation, 5 s of consecutive data were randomly extracted from each 7.5-s clip and used as the input data for the model (Fig. 2). In the case of videos for the validation and test data, the coordinate data were separated every 5 s with no duplication (Fig. 3). In addition, we applied the following data augmentation techniques to the training data: rotation (up to an angle of 3º), horizontal flip, translation (up to 50 pixels for the x-axis and 25 pixels for the y-axis), and mixup [26]. Details of each data augmentation technique are described in Supplementary Materials and Methods.

The model was trained to predict whether a label on the input coordinate data clip was from a patient with PD or SCD. For each input, the model outputs two values: the probability that the input data were obtained from a patient with PD and the probability that they were obtained from a patient with SCD. The prediction error was calculated as the binary cross-entropy loss between the model-predicted values and labels previously assigned to the input data. Subsequently, the model parameters were updated to reduce the binary cross-entropy loss value. The Adam optimizer was used for parameter optimization. We set the weight decay to 0.0001 and the mini-batch size to 64 and trained the model for 50 epochs. The learning rate was set to 0.001 for epochs 1 to 25 and to 0.0001 for epochs 26 to 50.

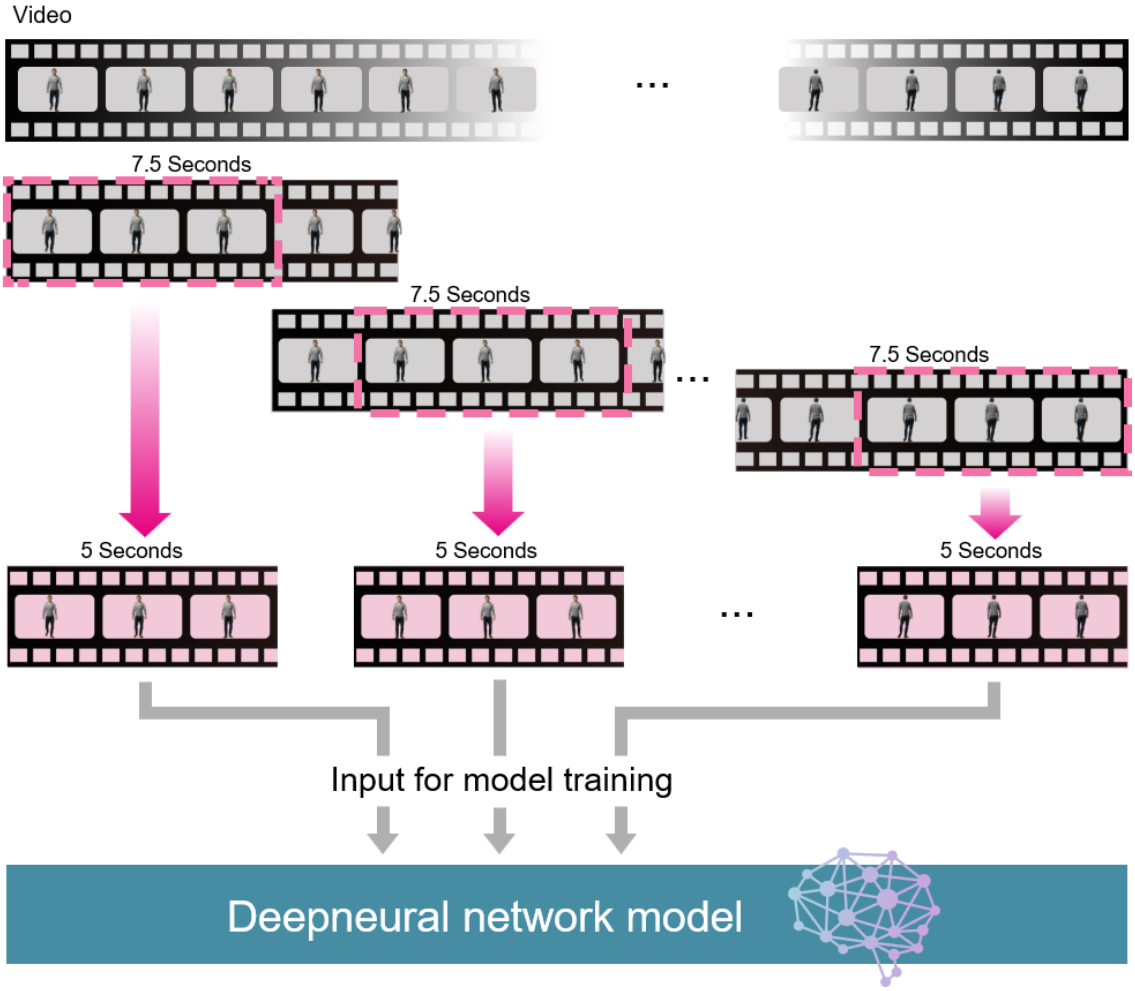

Fig. 2 Input data for model training

For training, the time series data of the key point coordinates in each video were clipped into 7.5-s clips (225 frames) with 50% overlap. Subsequently, continuous 5-s clips of data were randomly extracted from each 7.5-s clip and were used as input data for the model.

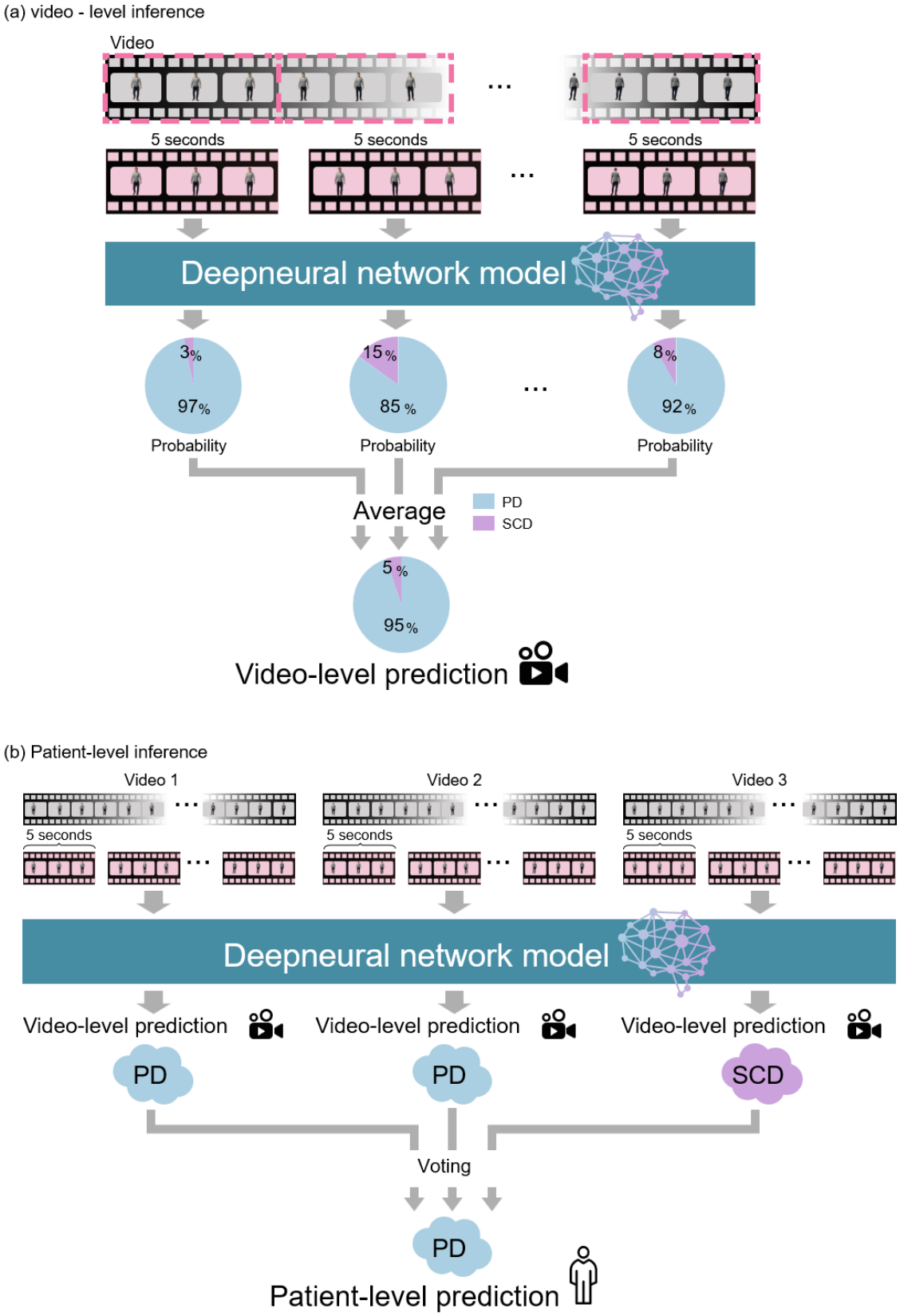

Fig. 3 Inference of the trained model

(a) For inference, the coordinate data from the test patient were separated every 5 s with no duplication. The model outputted values for the probabilities of PD and SCD for each input 5-s clip. The average of the outputs for all 5-s clips obtained from one gait video defined the video-level prediction of the model as PD if the average probability for PD exceeded 0.5; otherwise, it was defined as SCD. (b) Three gait videos were recorded in each patient, and video-level predictions were obtained for each of the three videos. Subsequently, voting was conducted on the three predictions, and the majority label was defined as the patient-level prediction of the model.

PD, Parkinson's disease; SCD, spinocerebellar degeneration

2.7 Evaluation of the Prediction Performance of the Model

After training the model, prediction performance was evaluated using test patient data with the best predictive parameters from the validation data. We generated multiple input data from one gait video by separating the key point coordinate data into 5-s clips. The model then output the probabilities of PD and SCD for each input data clip. Therefore, the average of the outputs for all 5-s clips obtained from one gait video defined the video-level prediction of the model as PD if the average probability for PD exceeded 0.5; otherwise, it was defined as SCD (Fig. 3a). The video-level predictive

performance of the model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC). Because we recorded three gait videos for each patient, voting was conducted on the resulting three video-level predictions, and the majority label was defined as the patient-level prediction of the model (Fig. 3b). The accuracy, sensitivity, and specificity were calculated using patient-level predictions for all patients.

2.8 Statistical Analysis

The sex ratios of the patients with PD and SCD were compared using a two-proportional Z-test. The age, disease duration, height, weight, and BMI of the two groups were compared using Student's *t*-test. The threshold for statistical significance was set at $p<0.05$.

3. Results

3.1 Characteristics of the Participants

The patients' demographic and clinical data are presented in Table 1. There was no significant difference in the sex ratio between patients with PD and those with SCD. The mean age of patients with PD was significantly higher than that of those with SCD.

The mean weight and BMI were significantly lower in the PD group than in the SCD group. The breakdown of SCD diagnoses was as follows: MSA-C, 9; HCAs, 42 (spinocerebellar ataxia (SCA) 6, 17; SCA 3, 8; SCA 2, 5; SCA 31, 4; SCA 1, 3; patients without identified pathologic gene, 3; SCA 36, 1; and dentatorubral-pallidoluysian atrophy, 1); and SAOA, 10. The distribution of gait sub-scores obtained using the SARA (item 1) and MDS-UPDRS part III (item 3.10) is shown in Fig. 4. Notably, the distribution of the severity of gait disturbance differed between the PD and SCD groups. In the SCD group, approximately one-third of the patients had a SARA item 1 score of 5, and a larger proportion of patients had severe gait disturbance. By contrast, the severity of gait disturbance was relatively evenly distributed among patients with PD ranging from mild to severe in the PD group.

	PD (n=82)	**SCD (n=61)**	**p-value**
Female (%)	46.3 (n=38)	49.2 (n=30)	0.74
Age (year)	69.8 ± 8.5	63.5 ± 9.4	< 0.001

Disease duration (year)	8.5 ± 5.5	9.4 ± 7.3	0.40
Height (cm)	160.3 ± 9.0	161.6 ± 8.5	0.40
Weight (kg)	55.6 ± 12.1	60.0 ± 11.4	0.032
BMI (kg/m^2)	21.5 ± 3.7	22.9 ± 3.5	0.028
Hoehn and Yahr stage	2.8 ± 1.0	NA	NA
MDS-UPDRS part III	31.8 ± 12.8	NA	NA
SARA	NA	12.4 ± 5.0	NA

Table 1 Demographic and clinical characteristics of the participants

PD, Parkinson's disease; SCD, spinocerebellar degeneration; BMI, body mass index; MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia

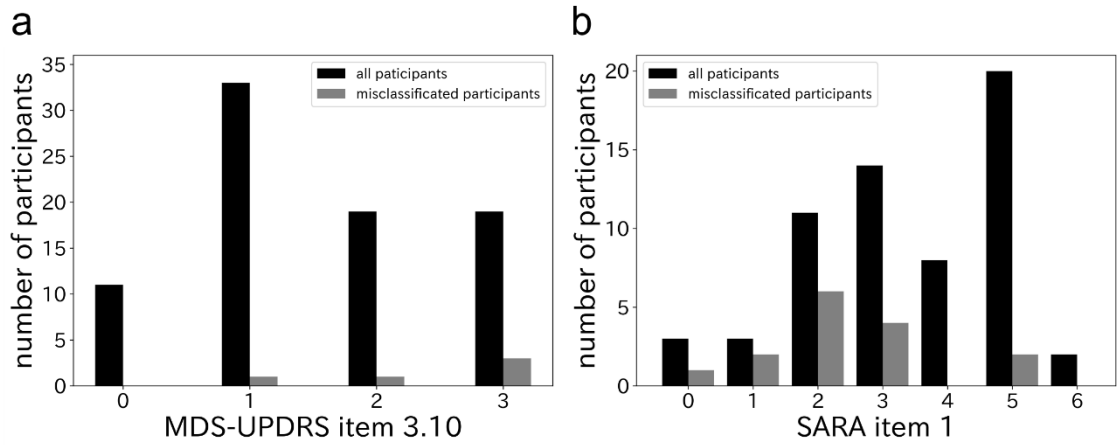

Fig. 4 Distribution of gait sub-scores obtained using the MDS-UPDRS and SARA

(a) Distribution of the score of MDS-UPDRS item 3.10 for all participants and for patients in whom the model failed to correctly predict the disease among those with Parkinson's disease. (b) Distribution of the score of SARA item 1 for all participants and patients in whom the model failed to correctly predict the disease among those with spinocerebellar degeneration. The black and gray bars indicate all participants and those with misidentified disease, respectively.

MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia

3.2 Predictive Performance of the Trained Model

The AUC of the video-level prediction for all the recorded gait videos was 0.91. The ROC curves for the video-level predictions are shown in Fig. 5. For patient-level predictions, the accuracy, sensitivity, and specificity for all participants were 0.86, 0.94, and 0.75, respectively. Overall, the model correctly predicted PD and SCD in 123 of 143 patients. Fig. 6 shows a confusion matrix representing the relationship between the true labels and model predictions in all participants.

The number of patients for whom the model failed to correctly predict the disease in relation to each gait sub-score obtained using the SARA and MDS-UPDRS is shown in Fig. 4. In the SCD group, model misclassification was more common in patients with mild gait disturbance (SARA item 1 $\leq$3), while in the PD group, misclassification was more common in patients with severe gait disturbance (MDS-UPDRS item 3.10 = 3). This result may be attributed to the larger proportion of patients with severe gait disturbance in the SCD group.

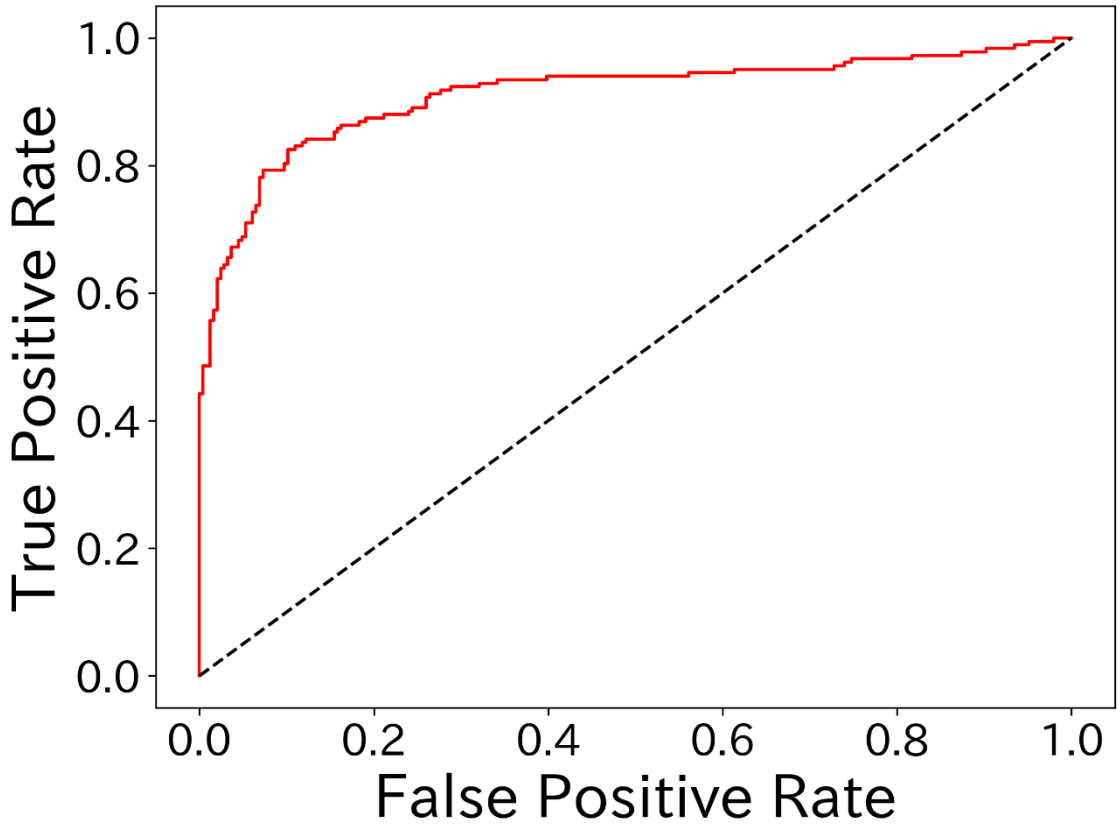

Fig. 5 Receiver operating characteristic curve of video-level prediction

The solid line represents the receiver operating characteristic curve of the video-level prediction of the trained model.

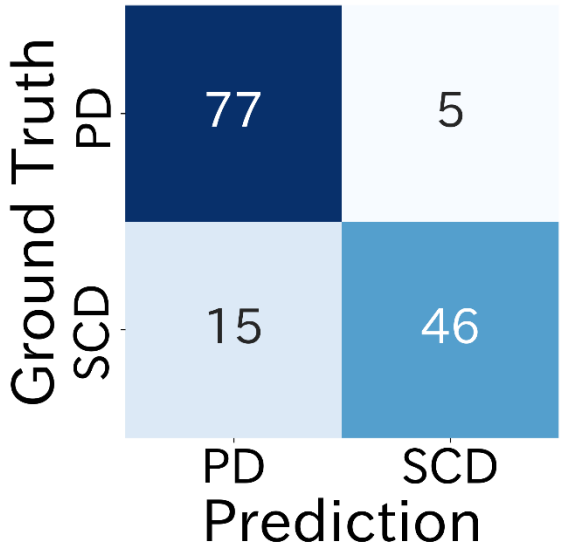

Fig. 6 Confusion matrix of the true label and prediction of the trained model

PD, Parkinson's disease; SCD, spinocerebellar degeneration

3.3 Additional Experiment with Propensity Score Matching

The aforementioned results suggested that the model can differentiate between PD and SCD using mere gait disturbance severity rather than gait characteristics. Additionally, there were significant differences in the age, weight, and BMI between the PD and SCD groups. We considered that these sources of bias may favor the model's disease differentiation potential, and the model's predictive performance may be overestimated. To balance clinical characteristics, we conducted propensity score matching [27]. We classified the severity of gait disturbance into four common levels: no gait disturbance (score of 0 in MDS-UPDRS item 3.10 and of 0 in SARA item 1), mild (score of 1 in

MDS-UPDRS item 3.10 and of 1 or 2 in SARA item 1), moderate (score of 2 in MDS-UPDRS item 3.10 and of 3 or 4 in SARA item 1), and requirement of assistance devices (score of 3 in MDS-UPDRS item 3.10 and of 5 or 6 in SARA item 1). Propensity scores were calculated using logistic regression, including age, sex, height, weight, BMI, and gait severity level, as described above. Each patient with SCD was matched with one patient with PD according to nearest neighbor matching using a caliper with 0.2 standard deviations. We performed the same experiment described in the Methods section (see 'Creating datasets and training the model' and 'Evaluating the prediction performance of the model') using data from the extracted patients.

After propensity score matching, 41 patients with SCD and 41 patients with PD were included. The characteristics and demographic data of the extracted patients are presented in Table 2. The AUC of the video-level prediction for all the gait videos of the extracted patients was 0.85. The ROC curves for the video-level predictions in the extracted patients are shown in Fig. 7. The average accuracy, sensitivity, and specificity of the model for all the extracted patients were 0.83, 0.88, and 0.78, respectively. Overall, the model correctly predicted PD and SCD in 68 of the 82 patients. Fig. 8 shows a confusion matrix representing the relationship between the true labels and model predictions. The distribution of gait sub-scores obtained using the SARA (item 1)

and MDS-UPDRS part III (item 3.10) in all extracted patients and in patients for whom the model failed to correctly predict the disease is shown in Fig. 9. In this analysis, there was no clear correlation between the severity of gait disturbance and rate of misclassification.

	PD (n=41)	**SCD (n=41)**	**p-value**
Female (%)	41.5 (n=17)	43.9 (n=18)	0.82
Age (year)	67.5 ± 9.4	66.7 ± 9.8	0.46
Disease duration (year)	8.6 ± 6.0	9.7 ± 7.9	0.48
Height (cm)	161.3 ± 8.7	161.7 ± 8.1	0.83
Weight (kg)	59.4 ± 12.6	58.8 ± 10.2	0.84

BMI (kg/m^2)	22.7 ± 3.8	22.4 ± 3.0	0.73
Hoehn and Yahr stage	3.0 ± 1.2	NA	NA
MDS-UPDRS part III	34.3 ± 14.7	NA	NA
SARA	NA	11.1 ± 4.7	NA

Table 2 Demographic and clinical characteristics of the participants extracted by propensity score matching

PD, Parkinson's disease; SCD, spinocerebellar degeneration; BMI, body mass index; MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia

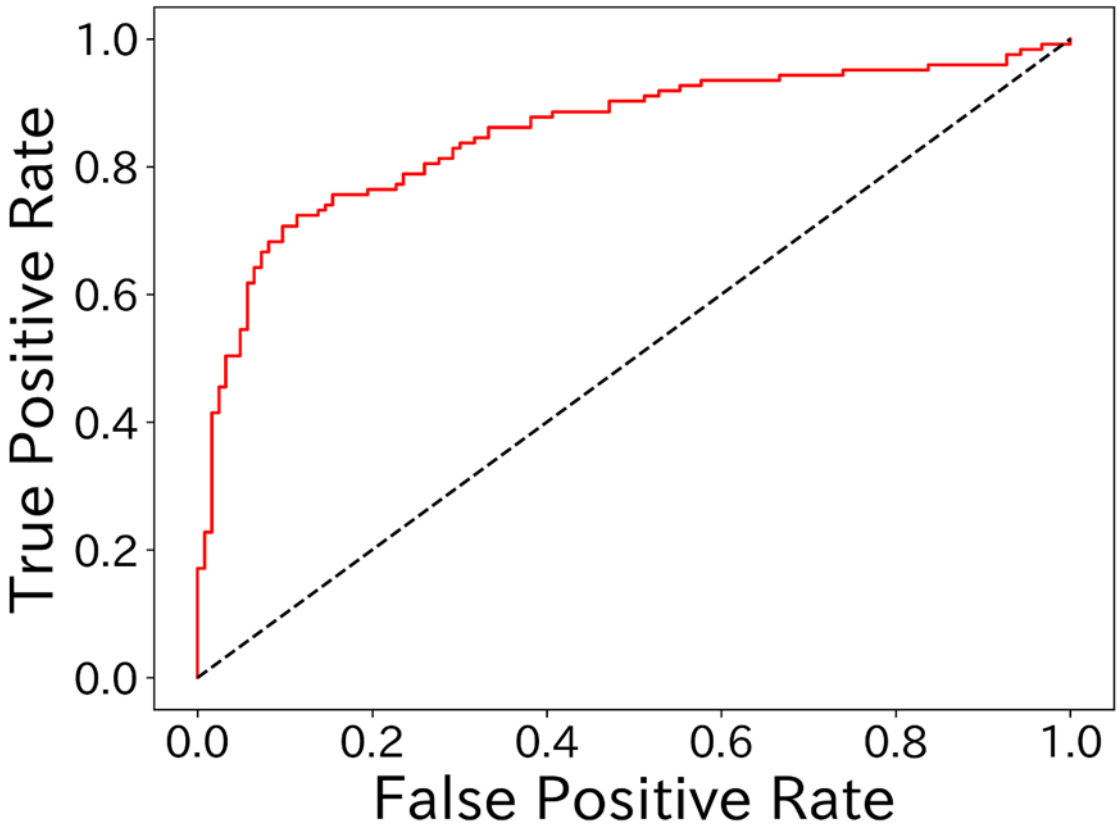

Fig. 7 Receiver operating characteristic curve of video-level prediction in patients extracted with propensity score matching

The solid line represents the receiver operating characteristic curve of the video-level prediction of the trained model.

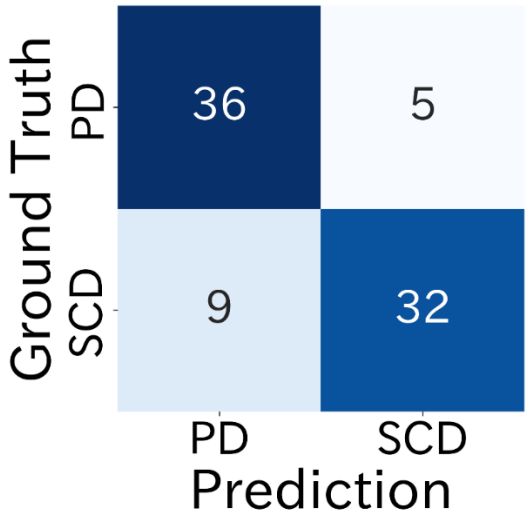

Fig. 8 Confusion matrix of the true label and prediction of the trained model in patients extracted with propensity score matching

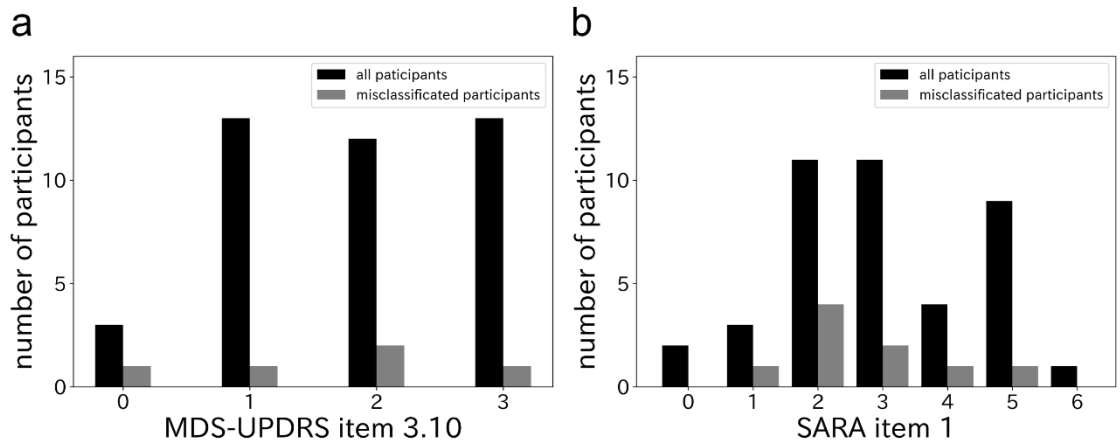

Fig. 9 Distribution of gait sub-scores obtained using the MDS-UPDRS and SARA in the participants extracted by propensity score matching

(a) Distribution of the score of MDS-UPDRS item 3.10 for all extracted participants and for patients in whom the model failed to correctly predict the disease among those with

Parkinson’s disease. (b) Distribution of the score of SARA item 1 for all extracted participants and for patients in whom the model failed to correctly predict the disease among those with spinocerebellar degeneration. The black and gray bars indicate all participants and those with misidentified disease, respectively.

MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; SARA, Scale for the Assessment and Rating of Ataxia

4. Discussion

In this study, we used a pose estimation algorithm to analyze patient gait videos in order to extract key point coordinates. Thereafter, we fed the extracted coordinates into a transformer-based DNN model to solve the binary classification task between PD and SCD. The proposed method was successful in classifying PD and SCD with an accuracy of 0.86, a sensitivity of 0.94, and a specificity of 0.75 in all participants and with an accuracy of 0.83, a sensitivity of 0.88, and a specificity of 0.78 in the participants extracted by propensity score matching. Although the performance of the DNN model was not satisfactory enough for it to be able to replace clinicians, it could be a collaborative tool in clinical practice and telemedicine.

Although few studies have reported the ability to differentiate PD and SCD using gait data, there was an attempt to classify PD and SCD using gait data obtained from a pressure-sensitive walkway [28]. The aforementioned study used recurrent gait units, which is a DNN model, to discriminate patients with PD and ataxia using gait data and reported that the model's predictive performance was as follows: AUC, 0.819; sensitivity, 0.826; and specificity, 0.781. Although a direct comparison of results is difficult due to the different number of patients and severity of disease, the method proposed in the current study exhibited a predictive performance comparable to that reported in this previous study, even in the participants extracted by propensity score matching. Additionally, measuring gait using a pressure-sensitive walkway requires specific measuring equipment and must be performed in a medical or research institution, similar to motion capture devices. Gait videos can be easily recorded, accessed, and interpreted without additional sensors. Therefore, our method may be an easy-to-use screening tool for differentiating neurological disorders based on the patient's gait.

In the analysis of all participants, the model tended to misclassify SCD as PD in patients with mild gait disturbance and PD as SCD in patients with severe gait disturbance. Furthermore, patients with SCD were significantly younger and had a

higher weight and BMI than those with PD. These findings raised concerns that the model distinguishes between PD and SCD based on body shape, age-related changes, and the severity of gait disturbance rather than on the gait features themselves. Therefore, we performed propensity score matching to assess whether the model could differentiate between PD and SCD in populations that did not differ in these characteristics. Although this reduced the model's performance slightly compared to that in the original dataset, there was no clear correlation between severity of gait disturbance and misclassification. This result suggested that the model distinguished between PD and SCD by extracting the patients' gait characteristics rather than by simply basing the prediction on gait disturbance severity or body shape. The reduction of the model's performance in the extracted patients could be because of the reduced amount of data as well as the alignment of clinical characteristics. Generally, the performance of DNN models is expected to improve with an increase in the amount of training data. By increasing the amount of data, the model's performance in an age- and dysarthria severity-matched group of patients could approach that observed in the original data. Further studies with larger numbers of patients are needed to validate this viewpoint.

We used gait videos in which the participants were recorded from the frontal view. To assess the laterality of the patients' symptoms, examiners are instructed to analyze the patients' gait from frontal view using the MDS-UPDRS [21]. In the evaluation of gait using the SARA, the patient’s gait is also assessed from the frontal view [29]. However, an accurate measurement of depth is not possible in 2D videos. Therefore, several important gait parameters, such as stride length and arm swing, cannot be quantitatively measured from the frontal view. Previous studies analyzing gait videos with pose estimation algorithms used videos recorded from the side view to assess these parameters quantitatively [17,30]. However, recording gait videos from the side view requires a larger space and is less convenient than recording them from the frontal view [31]. Instead of quantitatively measuring gait parameters from the key point coordinates, we attempted to differentiate the gait of patients with PD and SCD by feeding them into a DNN model. One advantage of DNN models is that they can automatically extract features from the input data, which are useful for solving specific tasks. The proposed method enabled the extraction of gait features from key point coordinates without the need for quantitative measurements or selection of gait parameters. Additionally, this method can evaluate a patient’s gait in a limited space,

such as a hallway. Therefore, we believe that our method can be easily implemented as a screening tool.

This study had several limitations. First, because DNN models are generally highly complex mathematical models and process high-dimensional data, it is difficult to understand which features of the key point coordinates were used by the model to differentiate PD and SCD. We expect the model to distinguish between PD and SCD by extracting important gait parameters, such as stride length and step width, from the key point coordinates, but this is difficult to ascertain due to the complexity of the model's computational process. The black box nature of the model is the greatest limitation of this study. To ensure that the model's predictions are trustworthy, future studies should identify features in the data that the DNN model focuses on to differentiate diseases. Second, healthy controls were not enrolled in this study. Therefore, we could not verify whether the model could differentiate between the presence and absence of gait disturbances. Furthermore, it would be more clinically useful if in addition to differentiating between PD and SCD, the model could differentiate other gait-affecting disorders. Further studies, including healthy controls and patients with other target diseases, are necessary to elucidate this. Notably, the addition of healthy controls and patients with other gait-affecting disorders needs to re-train the model using new data

obtained from such participants. Third, this was a single-center study, and external verification should be performed by conducting multicenter studies in order to assess the generalizability of the proposed method. Finally, we did not assess the cognitive function of the participants and were unable to evaluate the effect of cognitive function on participants' gait.

5. Conclusions

In summary, this study demonstrated the feasibility of differentiating between PD and SCD by using a pose estimation algorithm and a transformer-based DNN model to analyze gait videos. To use the proposed method in clinical practice, it is desirable for it to be able to differentiate between healthy controls and patients with other diseases that cause gait disturbances in addition to PD and SCD. Further studies including these participants are required to address this issue. This study can be used as a starting point for further development of machine learning approaches for differentiating gait disorders based on gait videos.

Declarations of Interest: None.

Funding: This work was supported by JSPS KAKENHI [grant numbers JP22K20843, MHLW]; a grant from the Research Committee on the Medical Basis of Motor Ataxias, Health and Labor Sciences Research Grants, The Ministry of Health, Labor and Welfare, Japan [grant number 23FC1010]; and Grants-in Aid from the Research Committee of CNS Degenerative Diseases, Research on Policy Planning and Evaluation for Rare and Intractable Diseases, Health, Labour and Welfare Sciences Research Grants, The Ministry of Health, Labour, and Welfare, Japan [grant number JPMH23FC1008]. The funding source had no involvement in designing or conducting the study, nor in the preparation of this article.

Author contributions:

K. Eguchi: Writing – original draft, Visualization, Validation, Data curation, Methodology, Formal analysis, Conceptualization. **H. Yaguchi:** Writing – review & editing, Supervision, Conceptualization. **H. Uwatoko:** Writing – review & editing, Supervision. **Y. Iida:** Data curation. **S. Hamada:** Data curation. **S. Honma:** Data

curation. **A. Takei:** Data curation, Supervision. **F. Moriwaka:** Data curation, Supervision. **I. Yabe:** Writing – review & editing, Supervision, Conceptualization.

Data statement: Raw video data are not available for public access because of patient privacy concerns. Other anonymous data are available from the corresponding author on reasonable request.

References

[1] R.L. Nussbaum, C.E. Ellis, Alzheimer's disease and Parkinson's disease, N. Engl. J. Med. 348 (2003) 1356–1364. https://doi.org/10.1056/NEJM2003ra020003.

[2] L.V. Kalia, A.E. Lang, Parkinson's disease, Lancet. 386 (2015) 896–912. https://doi.org/10.1016/S0140-6736(14)61393-3.

[3] W.A. Rocca, The future burden of Parkinson's disease, Mov. Disord. 33 (2018) 8–9. https://doi.org/10.1002/mds.27114.

[4] M. Pistacchi, M. Gioulis, F. Sanson, E. De Giovannini, G. Filippi, F. Rossetto, S. Zambito Marsala, Gait analysis and clinical correlations in early Parkinson's disease, Funct. Neurol. 32 (2017) 28–34. https://doi.org/10.11138/fneur/2017.32.1.028.

[5] G. Albani, V. Cimolin, A. Fasano, C. Trotti, M. Galli, A. Mauro, "Masters and servants" in parkinsonian gait: a three-dimensional analysis of biomechanical changes sensitive to disease progression, Funct. Neurol. 29 (2014) 99–105.

[6] A. Mirelman, H. Bernad-Elazari, A. Thaler, E. Giladi-Yacobi, T. Gurevich, M. Gana-Weisz, R. Saunders-Pullman, D. Raymond, N. Doan, S.B. Bressman, K.S. Marder, R.N. Alcalay, A.K. Rao, D. Berg, K. Brockmann, J. Aasly, B.J. Waro, E. Tolosa, D. Vilas, C. Pont-Sunyer, A. Orr-Urtreger, J.M. Hausdorff, N. Giladi, Arm swing as a potential new

prodromal marker of Parkinson's disease, Mov. Disord. 31 (2016) 1527–1534. https://doi.org/10.1002/mds.26720.

[7] D. Sethi, S. Bharti, C. Prakash, A comprehensive survey on gait analysis: history, parameters, approaches, pose estimation, and future work, Artif. Intell. Med. 129 (2022) 102314. https://doi.org/10.1016/j.artmed.2022.102314.

[8] E. Rovini, C. Maremmani, F. Cavallo, How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review, Front. Neurosci. 11 (2017) 555. https://doi.org/10.3389/fnins.2017.00555.

[9] A.Y. Meigal, L.I. Gerasimova-Meigal, S.A. Reginya, A.V. Soloviev, A.P. Moschevikin, Gait characteristics analyzed with smartphone IMU sensors in subjects with parkinsonism under the conditions of 'dry' immersion, Sensors (Basel). 22 (2022) 7915. https://doi.org/10.3390/s22207915.

[10] A.L. Silva de Lima, T. Hahn, L.J.W. Evers, N.M. de Vries, E. Cohen, M. Afek, L. Bataille, M. Daeschler, K. Claes, B. Boroojerdi, D. Terricabras, M.A. Little, H. Baldus, B.R. Bloem, M.J. Faber, Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease, PLOS ONE. 12 (2017) e0189161. https://doi.org/10.1371/journal.pone.0189161.

[11] A. Darwish, A.E. Hassanien, Wearable and implantable wireless sensor network solutions for healthcare monitoring, Sensors (Basel). 11 (2011) 5561–5595. https://doi.org/10.3390/s110605561.

[12] J. Stenum, K.M. Cherry-Allen, C.O. Pyles, R.D. Reetzke, M.F. Vignos, R.T. Roemmich, Applications of pose estimation in human health and performance across the lifespan, Sensors (Basel). 21 (2021) 7315. https://doi.org/10.3390/s21217315.

[13] K.W. Park, E.J. Lee, J.S. Lee, J. Jeong, N. Choi, S. Jo, M. Jung, J.Y. Do, D.W. Kang, J.G. Lee, S.J. Chung, Machine learning-based automatic rating for cardinal symptoms of Parkinson disease, Neurology. 96 (2021) e1761–e1769. https://doi.org/10.1212/WNL.0000000000011654.

[14] J.H. Shin, J.N. Ong, R. Kim, S.M. Park, J. Choi, H.J. Kim, B. Jeon, Objective measurement of limb bradykinesia using a marker-less tracking algorithm with 2D-video in PD patients, Parkinsonism Relat. Disord. 81 (2020) 129–135. https://doi.org/10.1016/j.parkreldis.2020.09.007.

[15] M.H. Li, T.A. Mestre, S.H. Fox, B. Taati, Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation, J. Neuroeng. Rehabil. 15 (2018) 97. https://doi.org/10.1186/s12984-018-0446-z.

[16] M. Lu, K. Poston, A. Pfefferbaum, E.V. Sullivan, L. Fei-Fei, K.M. Pohl, J.C. Niebles, E. Adeli, Vision-based estimation of MDS-UPDRS gait scores for assessing Parkinson's disease motor severity, in: *Med Image Comput Comput Assist Interv* 12263:637–647, 2020 637–647. https://doi.org/10.1007/978-3-030-59716-0_61.

[17] P. Liu, N. Yu, Y. Yang, Y. Yu, X. Sun, H. Yu, J. Han, J. Wu, Quantitative assessment of gait characteristics in patients with Parkinson's disease using 2D video, Parkinsonism Relat. Disord. 101 (2022) 49–56. https://doi.org/10.1016/j.parkreldis.2022.06.012.

[18] P. Mahlknecht, S. Kiechl, B.R. Bloem, J. Willeit, C. Scherfler, A. Gasperi, G. Rungger, W. Poewe, K. Seppi, Prevalence and burden of gait disorders in elderly men and women aged 60–97 years: a population-based study, PLOS ONE. 8 (2013) e69627. https://doi.org/10.1371/journal.pone.0069627.

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

[20] S. Gilman, G.K. Wenning, P.A. Low, D.J. Brooks, C.J. Mathias, J.Q. Trojanowski, N.W. Wood, C. Colosimo, A. Dürr, C.J. Fowler, H. Kaufmann, T. Klockgether, A. Lees, W. Poewe, N. Quinn, T. Revesz, D. Robertson, P. Sandroni, K. Seppi, M. Vidailhet,

Second consensus statement on the diagnosis of multiple system atrophy, Neurology. 71 (2008) 670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15.

[21] C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M.B. Stern, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A.E. Lang, A. Lees, S. Leurgans, P.A. LeWitt, D. Nyenhuis, C.W. Olanow, O. Rascol, A. Schrag, J.A. Teresi, J.J. van Hilten, N. LaPelle, Movement Disorder Society UPDRS Revision Task Force, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results, Mov. Disord. 23 (2008) 2129–2170. https://doi.org/10.1002/mds.22340.

[22] T. Schmitz-Hübsch, S.T. du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J.S. Kang, B. Kremer, C. Mariotti, B. Melegh, M. Pandolfo, M. Rakowicz, P. Ribai, R. Rola, L. Schöls, S. Szymanski, B.P. van de Warrenburg, A. Dürr, T. Klockgether, R. Fancellu, Scale for the assessment and rating of ataxia: development of a new clinical scale, Neurology. 66 (2006) 1717–1720. https://doi.org/10.1212/01.wnl.0000219042.60538.92.

[23] Z. Cao, G. Hidalgo, T. Simon, S.E. Wei, Y. Sheikh, OpenPose: realtime multi-person 2D pose estimation using Part Affinity fields, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 172–186. https://doi.org/10.1109/TPAMI.2019.2929257.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need. Presented at: Proceedings of the Advances in Neural Information Processing Systems pp: 1-11. (2017)

[25] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence 2021; 35. 35 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325.

[26] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond Empirical Risk Minimization, arXiv Preprint arXiv (2017) 1710.09412, doi:10.48550/arXiv.1710.09412.

[27] C.W. Pattanayak, D.B. Rubin, E.R. Zell, Propensity score methods for creating covariate balance in observational studies, Rev. Esp. Cardiol. 64 (2011) 897–903. https://doi.org/10.1016/j.recesp.2011.06.008.

[28] J. Song, J. Kim, M.J. Lee, J.H. Ahn, D.Y. Lee, J. Youn, M.J. Chung, Z. Kim, J.W. Cho, Differential diagnosis between Parkinson’s disease and atypical parkinsonism

based on gait and postural instability: artificial intelligence using an enhanced weight voting ensemble model, Parkinsonism Relat. Disord. 98 (2022) 32–37. https://doi.org/10.1016/j.parkreldis.2022.04.003.

[29] M. Grobe-Einsler, A.T. Amin, J. Faber, H. Völkel, M. Synofzik, T. Klockgether, Scale for the Assessment and Rating of Ataxia (SARA): development of a training tool and certification program, Cerebellum Published online (2023). https://doi.org/10.1007/s12311-023-01543-3.

[30] Ł. Kidziński, B. Yang, J.L. Hicks, A. Rajagopal, S.L. Delp, M.H. Schwartz, Deep neural networks enable quantitative movement analysis using single-camera videos, Nat. Commun. 11 (2020) 4054. https://doi.org/10.1038/s41467-020-17807-z.

[31] T. Khan, A. Zeeshan, M. Dougherty, A novel method for automatic classification of Parkinson gait severity using front-view video analysis, Technol. Health Care. 29 (2021) 643–653. https://doi.org/10.3233/THC-191960.