A multimodal gait and ocular geometric representation to generate a Parkinson progression report

John Archila, Ivan Peña, Luis Celis, Juan Olmos, Antoine Manzanera, Fabio Martínez

Available online at www.sciencedirect.com

ScienceDirect

(2025) –22

Journal Logo

www.elsevier.com/locate/procedia

A multimodal gait and ocular geometric representation to generate a Parkinson progression report.

John Archila[a], Ivan Peña[c], Luis Celis[a], Juan Olmos[a,b], Antoine Manzanera[b], Fabio Martínez[a]

[a]*Biomedical Imaging, Vision and Learning Laboratory (BivL^{2}ab), Universidad Industrial de Santander (UIS). Santander, Colombia*
[b]*Unité d'Informatique et d'Ingénierie des Systèmes (U2IS), ENSTA Paris, Institut Polytechnique de Paris. Palaiseau, France*
[c]*Grupo de renovación educativa de la medicina interna (GERMINA), Universidad Industrial de Santander (UIS). Santander, Colombia*

Abstract

Parkinson's disease (PD) is a progressive neurological condition, primarily associated with a deficiency in dopamine neurotransmitters, generating premotor, motor control, emotional, and executive dysfunctions. The characterization and PD diagnosis are principally based on the analysis of observed motion alterations, such as slowed movements (bradykinesia), wrong posture, and freezing of gait. Computational methods to support some of these observations remain limited to distinguishing between PD patients and control patients on the basis of protocols at advanced stages, according to current clinical PD guidelines. This work introduces a multi-item PD progression support *statistically consistent* with both the modified Hoehn and Yahr (H&Y) scale and the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) part III, which is based on a multimodal geometric representation that combines gait and oculomotor markerless video sequences. The proposed representation evaluates gait autonomy, posture impairment, freezing, bradykinesia, bilateral gait issues, and ocular bradykinesia. To do so, a three dimensional (3D) convolutional neural network (CNN) first captures spatiotemporal PD patterns. Then, a Riemannian network learns second order relationships among observed patterns, which are further fused at early or intermediate stages to output multi-item PD predictions. *In a retrospective study with 13 control subjects and 19 diagnosed Parkinson's patients, the proposed approach (early fusion) achieved F1-scores of 95% for bilateral impairment, 94% for gait autonomy, 81% for freezing of gait, 92% for wrong posture, 91% for gait bradykinesia, and 94% for ocular bradykinesia in Parkinson's patients.* The proposed approach is a promising tool to support routine and standard clinical PD analysis.

Keywords: Parkinson progression report, multimodal geometrical end-to-end neural networks.

1. Introduction

PD is the second most disabling neurodegenerative disease worldwide, affecting 6.1 million people [1]. There is no definitive biomarker for diagnosis and considering the wide phenotyping range of the disease, there is no truly effective treatment [2]. PD is associated with dopamine deficiency and leads to sleep disturbances, mood disorders, and primarily motor impairments [3]. In particular, motor impairments are the main biomarkers of the disease, including bradykinesia (slowness of movement), postural instability, and gait abnormalities [4]. The multifactorial nature of PD makes it difficult to carry out appropriate diagnosis and monitoring. In fact, the large variety of motor symptoms cannot be fully characterized by a single type of observation. Consequently, employing multiple observations offers complementary insights, enhancing both diagnostic accuracy and understanding disease's progression.

Currently, PD characterization diagnosis and follow-up are based on observational scales such as the modified H&Y scale and MDS-UPDRS. The modified H&Y scale is based on observation of motor disorders during gait only, allowing the coarse classification of disturbance levels between zero and five (strongest motor manifestations) [5]. Alternatively, the MDS-UPDRS Part III scale considers various motion modes, such as finger tapping, facial expression, and gait quantifying items observed by a specialist (from the minimum (0) to the severe (4) stage) [6, 7].

Owing to the observational nature of the scales, characterization is highly dependent on specialist's expertise. Additionally, at early stages, motor symptoms are very subtle and barely noticeable, reporting diagnostic errors of 25% for neurologists with little experience in motor observation and between 6% and 8% for expert specialists in movement disorder analysis [8]. To mitigate the low sensitivity, clinical guidelines suggest reporting other motor observations. For example, oculomotor patterns exhibit alterations at early stages of the disease, even in the prodromal phases [9]. In particular, smooth eye movements provide fine-grained spatiotemporal information and insight into sensory-cognitive processes. They offer a unique opportunity to investigate how the basal ganglia interact with other neural structures associated with eye movement [10]. Some experimental works have shown that kinematic features such as pupil reaction, saccadic jerks, or velocity gains can be used to distinguish PD patients and control subjects [11, 12]. Specifically, eye movement has been deemed crucial for identifying early stages, whereas gait has been considered essential for detecting cardinal symptoms such as bradykinesia, and key symptoms such as freezing and postural instability. However, eye movements are not easily assessed in clinical settings and current protocols require specialized laboratory equipment, such as video-oculography, along with additional training for specialists.

Computational methods have been used to support the characterization, diagnosis and follow up of PD patients [13, 14, 15]. These methods are principally based on the analysis and quantification of gait, which makes it difficult to identify motor impairments that are not prominent in this modality. *Motor impairments in PD affect locomotion and eye movements, but at different stages and in a complementary manner. While gait disturbances such as bradykinesia, postural instability, and freezing of gait are characteristic of advanced stages of the disease, abnormal eye movement patterns may appear even in the prodromal phases. Evaluating both modalities allows for capturing a broader spectrum of motor manifestations, facilitating a better differentiation between patients and control subjects. Moreover, integrating these signals offers a multi-factorial perspective on the impact of the disease, providing valuable information to improve diagnostic accuracy and the characterization of motor phenotypes. Recent multimodal methods have considered integrating gait and ocular motions. A method was developed to quantify both gait and eye movement via an infrared eye tracker along with an inertial gait device to measure the kinematics of both modalities. This approach selects the most statistically significant variables and develops a multivariate logistic regression model incorporating saccade velocity, maximum trunk sway, and mean angular velocity during turns. Such a method had the goal of distinguishing motor patterns between PD patients and control subjects [16]. Another approach combines eye fixation and gait sequences through the calculation of the covariance of deep features, followed by random forest classification, achieving notable sensitivity within the studied dataset [17]. These methods can coarsely differentiate between PD patients and control subjects, which supports global discrimination of PD subjects. However, in clinical routine, the neurologist bases the diagnosis on standard scales, which are subjective and expert-dependent. Although such methods provide a global PD prediction, it is difficult to understand the complex support of each modality, and which of sub-items are affected for each patient. In consequence, there is not a natural support of clinical routine, where experts requires supervision and support of specific items to identify and planning personalized treatments, for instance, support the prediction of fine motor impairments such as bradykinesia, gait autonomy, freezing of gait, and postural abnormalities, which are key symptoms for clinical assessment. Additionally, these models rely on handcrafted features or deep learning representations that lack a unified metric structure, which limits their ability to encode complex motor relationships.*

This work proposes an end-to-end multimodal geometric approach that enables the prediction, from gait and eye movement video sequences, of six clinical items commonly observed in patients with Parkinson's disease. The model aims to complement the traditional evaluation performed by specialists by providing a computational tool that quantifies motor impairments related to bradykinesia (ocular and gait-related),

posture, freezing of gait, bilateral gait impairment, and gait autonomy. As a main contribution, a novel architecture is introduced, based on early and intermediate fusion of spatiotemporal features extracted through 3D convolutions, integrated into a geometric representation on the manifold of symmetric positive definite matrices. This approach introduces, for the first time, a multimodal fusion strategy that combines oculomotor and gait patterns to simultaneously predict multiple clinically relevant items, including ocular bradykinesia, a symptom with high diagnostic sensitivity that has rarely been considered in computational studies. Taken together, this proposal represents an objective, non-invasive, and automated tool with potential impact in clinical practice, by facilitating patient monitoring, the evaluation of therapeutic interventions, and the identification of motor impairments, thus contributing to personalized medicine and a deeper understanding of disease progression. The paper is structured as follows. Section 2 provides an overview of the literature on Parkinson's disease with a focus on motor patterns. Section 3 outlines the collected data and provides a brief summary of the protocol used to capture the two modalities under semicontrolled conditions. Section 4 elaborates on the proposed convolutional-Riemannian fusion methods. Section 5 presents the results of the clinometric progression report. Finally, Section 6 discusses the advantages and limitations of geometrically integrating motion modalities to generate a clinometric report on Parkinson's disease.

2. Related works

2.1. Parkinsonian gait analysis

Parkinson gait impairments are evident when the prefrontal cortex is affected by dopamine deficiency [18]. One common motor-related disorder is bradykinesia, expressed as slowness of the body and limitations of spontaneous movements [19]. Other reported impairments include freezing of gait, described as significant impediments to the forward movement of the feet [20], and posture changes, which involve a forward stooped stance accompanied by cervical and thoracic spine flexion [21]. Classically, the evaluation of these symptoms is carried out visually by experts, and may significantly differ from one specialist to another, due the substantial variability in symptoms [22].

Recently, analysis of video sequences has proven to be a highly valuable and noninvasive strategy for capturing Parkinsonian gait patterns. For example, there exist strategies that estimate postures at each frame and compute reference joints, allowing the measurement gait patterns related to: step length, walking speed, arm swing magnitude, and velocity [23]. Alternatively, silhouette sequences from patients' videos have been extracted and summarized to distinguish between Parkinsonian gait and control subjects [24]. These studies, however, missed gait features correlated with PD that are now known to be relevant for the diagnosis and monitoring of the disease. Video classification strategies may also be biased by external factors related to the acquisition process, to the detriment of motion information.

Other studies have characterized the level of gait independence following specific item (3.10) of the MDS-UPDRS Part III [25, 26]. These methods compute landmarks that, from 2D [25] and 3D poses [26], together with kinematic variables, allow for the stratification of the disease stage (normal, slight, mild, and moderate levels). Some studies have also considered the stratification of gait across multiple scales to estimate distance and angle calculations between limbs and use regressions to generate a report based on the gait items of the UPDRS and the SAS scale [27]. However, stratification on the basis of solely the gait modality may hinder the prediction of patients in early stages, where gait-related motor impairments are difficult to perceive.

2.2. Eye motion analysis

Ocular movement impairments are recent valuable disease indicators since they often correlate with neuronal deficits caused by the loss of dopamine. Furthermore, they can manifest before recurrent symptoms of the disease, such as gait disturbances and incorrect posture [28]. Various studies have demonstrated that ocular movements can be notably affected by dopamine deficiency [29, 30]. In smooth pursuit tasks, alterations in PD are linked to foveal motion and the capacity to follow linear trajectories and changes in direction [9]. For PD patients, a decrease in tracking speed and an increase in the frequency of involuntary movements have been observed [9, 30]. As a consequence, these oculomotor patterns have been proposed

as promising biomarkers for Parkinson's disease, facilitating differentiation between healthy individuals and those in the early or intermediate stages of the condition, even in the absence of additional motor symptoms [31]. Despite the importance of eye movement characterization, scales used in clinical routines such as the modified H&Y and MDS-UPDRS part III do not consider this type of movement. This may be due to the current sophisticated setup used to capture such movements and the difficulty establishing a consensus between specialists to define a scale of affect from these observations [32, 33]. To reduce the complexity of calibration and facilitate use in routine clinical practice, computer vision-based methods have been proposed [34, 35, 36]. For example, kinematic features such as amplitudes and reaction times have been quantified. Then, using an ensemble of classifiers, the overall diagnosis is predicted through binary classification [34]. Alternatively, kinematic features are quantified to calculate the statistical significance between stages [35] or for multistage classification on the MDS-UPDRS [36].

2.3. PD multimodal characterization

Multimodal computational approaches have been also reported in the literature, which consider several symptoms in patients, enabling the ability to support the diagnosis between control subjects and Parkinson's patients [37, 38, 16, 17]. For example, voice frequency and handwriting features have been classified independently using ensemble classifiers. The output diagnosis is considered Parkinson if at least one modality predicts the disease [37]. However, voice and handwriting modalities may primarily appear in the intermediate stages of the disease, being limited for patients who predominantly experience gait disturbances rather than tremors. Other approaches have considered other motion modes to leverage the complementarity of different motor impairments, such as gait, speech, and tapping modalities through the use of sensors such as gyroscopes and accelerometers. These signals have been used to train deep convolutional models and achieve discrimination between patients and controls subjects [38].

Recently, gait and eye movement modalities have been coded from an infrared eye tracker and an inertial gait device to estimate statistically significant kinematic variables for each modality. To differentiate motor patterns between PD patients and control subjects, a multivariate logistic regression model was fixed using the resulting kinematics associated with saccade velocity, maximum trunk sway, and average turn angular velocity [16]. Another strategy combined eye fixation and gait sequences by computing the covariance of deep features, followed by random forest classification. Such an approach nonetheless learns complementary and multimodal patterns to discriminate PD [17]. In general, previous approaches have carried out binary classification, which is insufficient to support the clinical and personalized characterization of PD.

Alternatively, some multimodal studies group patients into a stage of the disease according to an observation scale, providing an estimate of its progression level. Recent works have proposed stratified predictions via common rating scales [39, 40]. For example, voice and gait have been analyzed using microphones, gyroscopes, and accelerometers. Graphs were constructed from these signals for further classification via a random forest. This classification identifies levels as low, intermediate, and severe [39]. Alternatively, handwriting, gait, and speech have been measured by accelerometers, gyroscopes, and tablets. These signals are subsequently represented by frequencial features and have been used to train deep convolutional models to classify control subjects from early, intermediate and advanced stages [40]. Although stratification provides more specialized information than binary prediction does, it still lacks more relevant diagnostic details. Therefore, identifying and characterize PD symptoms, such as bradykinesia in patients is vital, as it is a core symptom of the disease. Additionally, other motor impairments, such as posture and gait disturbances, are crucial in characterizing the diagnosis [41]. Nonetheless, current approaches only consider global classification score without additional output about motion items.

3. Multimodal data

The study involved 13 *control subjects (average age* 72.2 ± 6.1 *years) and* 19 *PD patients (average age* 72.3 ± 7.4 *years). A expert neurologist supported the PD progression severity. The specialist conducted an evaluation of six motor impairments, including an item that predicts bilateral gait impairment based on observations from the modified H&Y scale. Furthermore, four items according to the MDS-UPDRS*

Part III scale were also evaluated (gait autonomy, freezing, posture, bradykinesia), and an additional and complementary detection of ocular bradykinesia. In general, according to the modified H&Y scale there are five PD patients in stage 1, five PD patients in stage 1.5, five PD patients in stage 2, and four PD patients in stage 3.

Regarding the data acquisition, each participant underwent 10 gait recordings (5 in which the individual moved from left to right and 5 in the opposite direction) and 10 smooth ocular motion recordings (5 for the right eye and 5 for the left eye) with a conventional camera, Nikon D3200 with spatial resolution of 1280×720. *To ensure participant safety, recordings were only conducted if the patient was under the effect of medication. Before beginning, the patient was asked whether they had taken their medication in the morning, as the recordings were conducted in the early hours to minimize the risk of falls. Although the videos were generally of high quality, controlled lighting conditions were maintained throughout the study. To preserve the symmetry of laterality, 8 out of the 10 recordings per modality were selected. Some recordings were discarded when participants got out of focus during the gait or ocular motion task.* During the study, the participants were instructed to adhere to the following movement guidelines:

- **Gait.** The participants were invited to walk in a straight line while the camera captured them from a sagittal view. Locomotion was recorded over a 5-meter walk, with an average video duration of 5 seconds per recording. Each video had a spatial resolution of 520×520 pixels and a temporal resolution of 60 frames per second (fps). *The recording environment was standardized in terms of lighting and flooring to reduce variability. Participants were asked to walk at a self-selected pace, and up to three trials were performed to ensure consistency. The camera was placed at a fixed height of 1 meter and a distance of 5 meters from the center of the walking path.* Figure 1 illustrates the postural configuration during the gait exercise, realized in front of a uniform green background. *In the Figure 2 is illustrated the gait locomotion for two PD-affected patients, in different stages of the disease. As observed, patients with unilateral impairment tend to exhibit more evident motor difficulties on one side of the body, whereas those with bilateral impairment experience a greater reduction in step amplitude and more pronounced postural alterations. These differences can influence model predictions, as motor patterns vary significantly depending on disease progression. These views illustrate the diversity of clinical manifestations captured in the dataset, facilitating the interpretation of the model's results.*

- **Smooth ocular motion.** In this scenario, patients were instructed to maintain their gaze on a spotlight projected onto a screen with a dark background. The monitor's height was adjusted to align the center of the screen with the center of the pupillary plane, as depicted in Figure 1. The motion of the spotlight was controlled both horizontally (from right to left and vice versa) and vertically (from top to bottom and vice versa). *Recordings were performed with uniform ambient lighting to minimize reflections or pupil dilation effects.* The recorded video was subsequently manually cropped to a region of interest around the eye, with dimensions of 210×140 pixels. *In the figure 3 is illustrated the ocular motion for a control and for a pD patient is illustrated, synchronized according to the spotlight projected onto a screen. As observed, the PD patient shows a delay in tracking the point, evidencing an anomaly in the correct execution of the exercise.*

Sixteen videos were obtained for each participant, encompassing both gait and ocular smooth motion exercises. For both modalities, the same camera, a conventional Nikon D3200, was utilized, providing a spatial resolution of 1280×720. Consequently, the entire dataset for this study comprises a total of 512 videos. This study received approval from the Ethics Committee, and written informed consent was obtained from each participant.

4. Proposed approach

This work presents an end-to-end multimodal and geometric architecture, that supports a multiple item PD prediction, allowing support clinical PD characterization. The proposed approach starts from a convolutional 3D representation, and the resulting deep features are summarized in symmetric positive definite

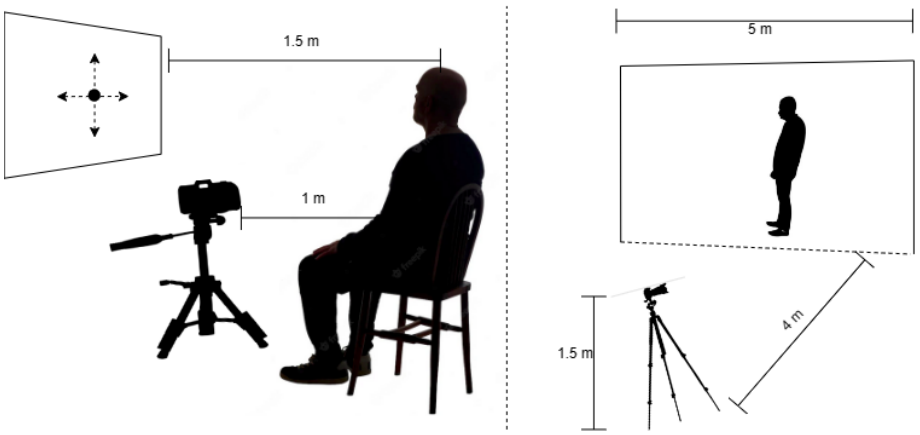

Fig. 1. *Gait and ocular smooth motion acquisition setups for our markerless video dataset. On the left, it is represented how oculomotor video sequences are recorded while the patient maintains their gaze on a spotlight projected onto a screen. In Right is illustrated how walking videos are recorded from a sagittal view.*

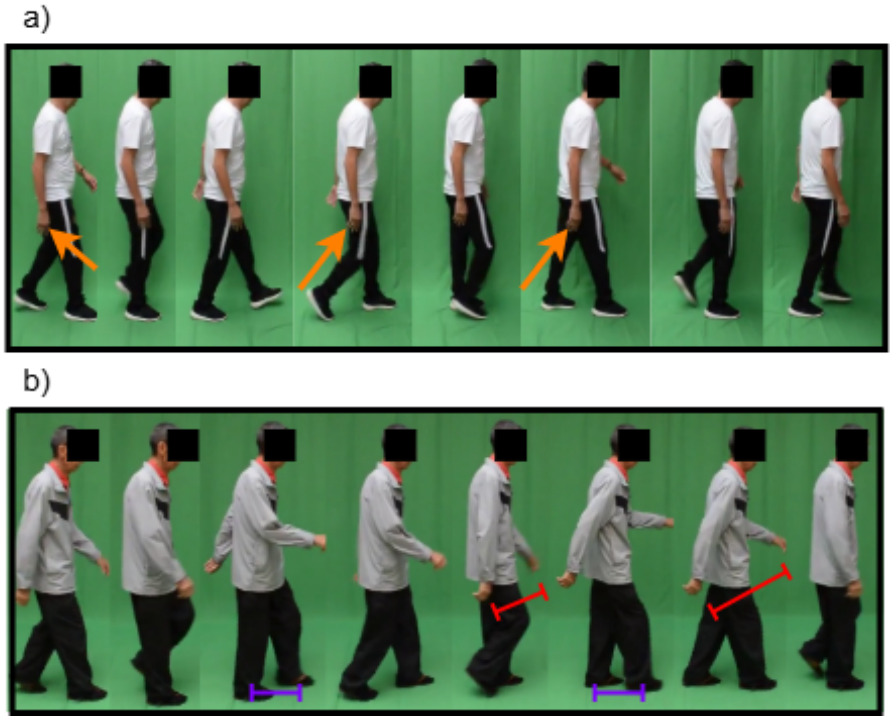

Fig. 2. *Two patients during gait recording. The patient with unilateral impairment is shown in Figure a), where the right arm exhibits reduced mobility (see orange arrows), and coordination with the left arm is impaired. The patient with advanced impairment is shown in Figure b). Arm swing amplitude is uncoordinated (see red lines), and there is also a noticeable reduction in step length relative to arm swing amplitude (see purple lines).*

(SPD) matrices to deal with potential scarcity in training data. The SPD matrices are then processed by geometric layers whose function is to fuse PD modalities. The fusion of both modalities occurs at either the early or intermediate level (Figure 4). At the early fusion level, a unique SPD matrix is computed from

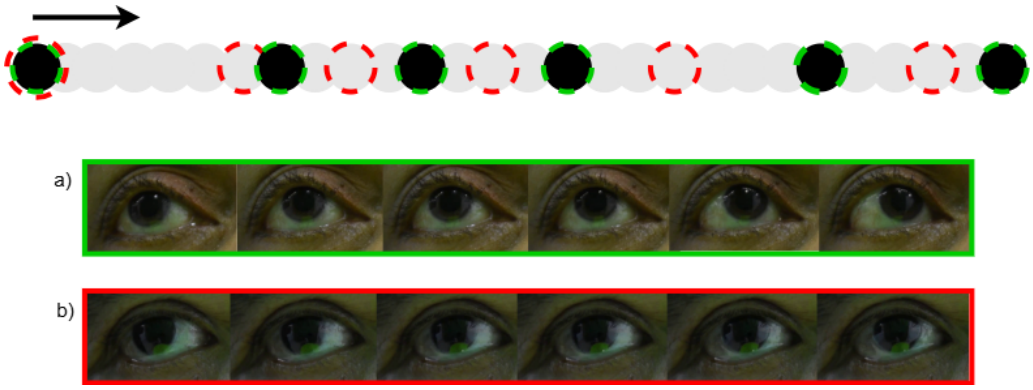

Fig. 3. *Example of smooth eye movement. In green: smooth eye movement of a control subject, who appropriately follows the point on the screen (see Figure a). In red: eye movement of a patient who has difficulty tracking the point on the screen. The patient's reflexes and eye movements are generally slower compared to those of the control subject (see Figure b).*

the two motion modes, while at the intermediate level, an SPD matrix is computed for each modality, from which are learnt new geometric representations, that are subsequently fused in the dense layers to output multiple predictions to support motor scales. The details of the proposed approach are described in the next subsections.

4.1. 3D Convolutional Representation

During the initial phase, the 3D-CNN builds a latent representation of spatiotemporal patterns, enabling the emergence of relevant gait and ocular movement patterns across several hierarchical levels. Given a video sequence $\boldsymbol{S}$, each layer $\ell \in \{1, \ldots, L\}$ computes image transformations sequentially, using a bank of K_ℓ learnable 3D convolution filters Ψ_k^ℓ and scalar biases b_k^ℓ, commonly followed by non-linear activation functions $\mathbf{a}_k^\ell$ and pooling operation Π_ℓ, projecting in this way the input information onto a set of deep features by the following sequential process:

$$\Phi_k^\ell = \Pi_\ell \left(\mathbf{a}_k^\ell \left(\boldsymbol{\Phi}^{\ell-1} * \Psi_k^\ell + b_k^\ell \right) \right)$$

with $\boldsymbol{\Phi}^\ell = \left\{ \Phi_k^\ell \right\}_{k=1}^{K_\ell}$, $\boldsymbol{\Phi}^0 = \mathbf{S}$, and the resulting bank of 3D representations $\boldsymbol{\Phi}^L$ is the final result of the feature extraction process of the 3D-CNN, yielding a three-dimensional feature bank $\boldsymbol{\Phi}^{\mathbf{L}} = \left\{ \boldsymbol{F}^i \right\}_{i=1}^{N}$ (with $N = K_L$), which encompasses a set of attributes. This bank of features is calculated for each motion mode, allowing to capture Parkinsonian spatiotemporal patterns encountered in gait and ocular smooth motion. *In Figure 5 is illustrated a sample of deep features for gait and eye motion. Also in the last column illustrated the corresponding SPD descriptor for both videos. In both cases, the learned deep features capture spatio-temporal edge patterns, which may result in discriminative for classification tasks. Also, the SPD descriptor evidence some patterns, which potentially may help to discriminate item classes, considered in this work.*

4.2. Geometrical fusion level

4.2.1. Early Riemannian fusion

The spatiotemporal patterns associated with Parkinson's disease, derived from ocular movement modalities (initial impairments) and gait modalities (intermediate and advanced impairments), can be summarized in a single matrix that calculates the covariance between the output volumes of both modalities. In this context, a Riemannian signature is obtained for each patient, considering impairments at different stages of the disease. Let $\boldsymbol{\Phi}_{\mathbf{L}}^{\mathbf{e}}$ *be the eye motion feature bank* $\boldsymbol{\Phi}_{\mathbf{L}}^{\mathbf{e}} = \left\{ \boldsymbol{F}^i \right\}_{i=1}^{N}$ *and, be* $\boldsymbol{\Phi}_{\mathbf{L}}^{\mathbf{g}}$ *the gait feature bank* $\boldsymbol{\Phi}_{\mathbf{L}}^{\mathbf{g}} = \left\{ \boldsymbol{G}^i \right\}_{i=1}^{N}$*, the first fusion alternative aims to retain and enhance the information captured by the*

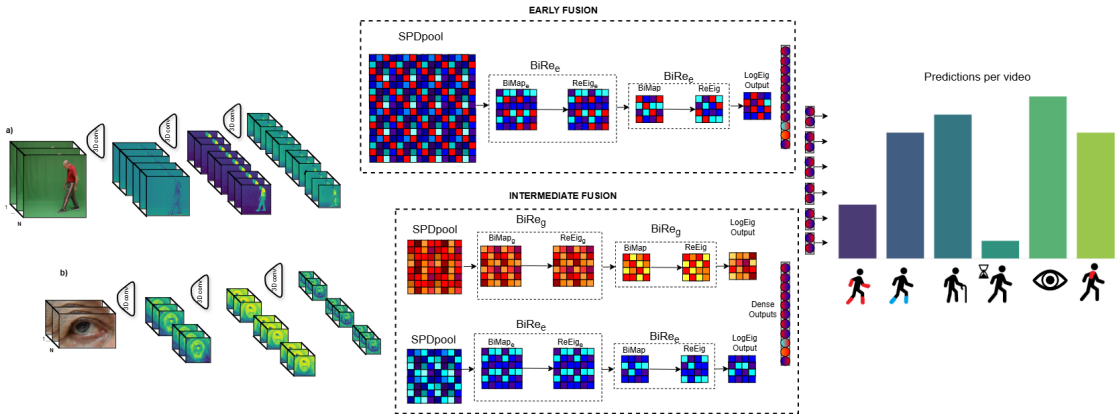

Fig. 4. End-to-end multimodal and geometric architectures, that support multiple-item PD prediction. Two approaches were performed. Early Fusion: Gait and ocular smooth motion modalities are processed by the 3D-CNN and fused early through SPD pooling. Then further compact representations are learned by two BiRe blocks and mapping to the Euclidean space is performed to predict a symptom report of patients. Intermediate Fusion: Gait and Ocular smooth motion modalities are independently processed by their own 3D-CNN and Riemannian networks. Then their respective mappings to the Euclidean space are concatenated and input to the dense layers whose outputs form the predictions for the symptom report.

3D convolutional deep features of both modalities. To do this, a covariance-like symmetric positive definite (SPD) matrix is calculated from a flattened version of the eye and gait feature maps ($WH \times 2N$), as: $\boldsymbol{R} = \left[vec(\boldsymbol{F}^1), \cdots, vec(\boldsymbol{F}^N), vec(\boldsymbol{G}^1), \cdots, vec(\boldsymbol{G}^N)\right]$. *In such a case, $W \times H$ are the spatial dimensions of all feature maps. Then the $2N \times 2N$ SPD pooled matrix, input of the Riemannian network, is defined as:* $\boldsymbol{C^0} = \boldsymbol{R}^T\boldsymbol{R}$. *The computation of the SPD matrix $\boldsymbol{C^0}$ in the early fusion framework yields a second-order statistical representation that captures the covariance structure of spatiotemporal features from both ocular and gait modalities. This matrix encodes intermodal and interfeature dependencies, standing out dominant patterns and latent interactions associated with early Parkinsonian impairments. The symmetric positive definite nature of matrix $\boldsymbol{C^0}$ requires its processing to be done within the Riemannian manifold, to respect the non-Euclidean geometry of the data and enable a faithful characterization of relevant features that may be lost if processed in Euclidean spaces.*

To perform the geometrical learning, the SPD matrix $\boldsymbol{C^0}$ is transformed into a more compact SPD matrix using a sequence of Riemannian blocks, which are composed of two layers, the bilinear mapping (BiMap) layer and the Rectification Eigenvalues (ReEig) layer. Here, the BiMap computes: $C^\ell = W_\ell C^{\ell-1} W_\ell^T$, with $W_\ell \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$ is the weight matrix transformation ($d_\ell << d_{\ell-1}$) [42]. After a BiMap layer, the Riemannian block is completed with a ReEig layer, which, similarly to the ReLU function for Euclidean networks, applies a rectification threshold, by computing $C^{\ell+1} = U^\ell \max(\varepsilon I, \Sigma^\ell)(U^\ell)^T$, where (Σ^ℓ, U^ℓ) respectively represent the diagonal matrix of eigenvalues and the matrix of eigenvectors of C^ℓ, and ε is a scalar threshold. At the end of the Riemannian module, a Riemannian logarithm map **log** is computed to map the resulting Riemannian descriptor back onto the Euclidean space [42]. Finally, this embedding is vectorized, concatenated, and thereafter processed by two dense layers $\mathcal{D}$. In the output layer, a softmax layers computes the probability of predictions of different PD symptoms and motor affectations to differentiate between PD patients and control subjects. The early fusion approach is illustrated in Figure 4.

4.2.2. Intermediate fusion from independent geometrical branches

Alternatively, gait and ocular movement can be identified with two types of motion: one with coarse granularity, involving limb coordination and balance of the trunk and head, and the other with fine granularity, which is based on smooth ocular movements in the vertical and horizontal directions. In this context, it is of particular interest that each movement independently learns new geometric representations within its respective branch. Finally, a subsequent fusion occurs in the Euclidean space through the dense layer. In the

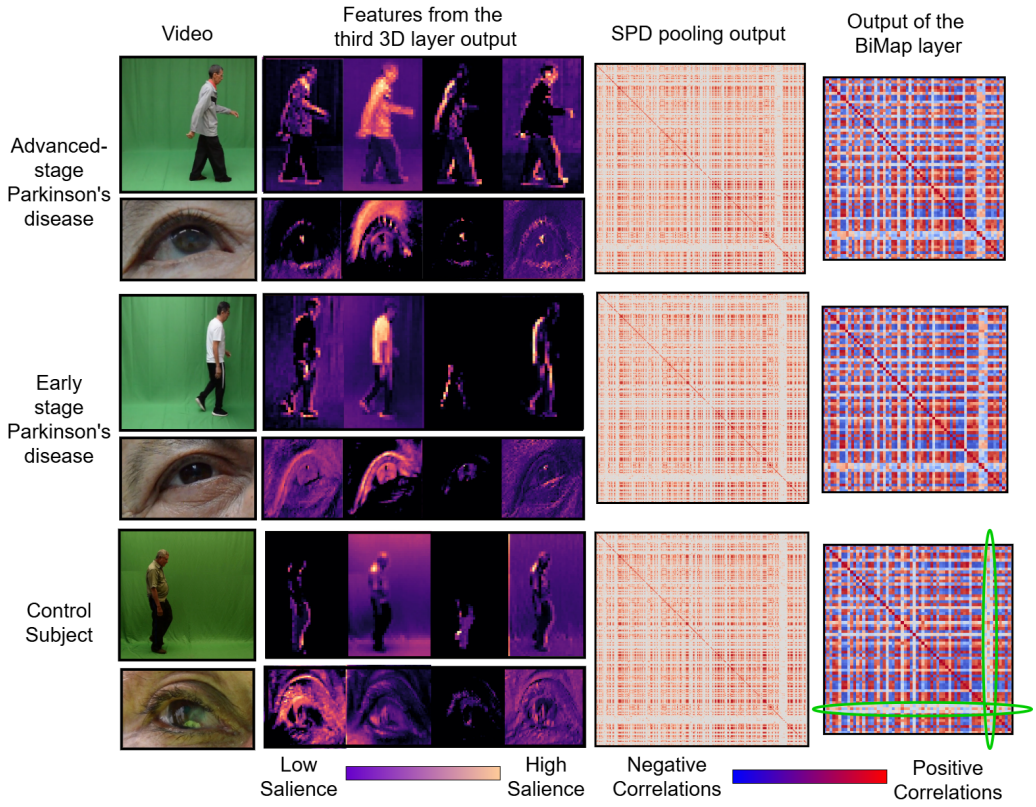

Fig. 5. *Intermediate representations from the multimodal geometric architecture for three subjects: advanced-stage Parkinson's disease (top), early-stage Parkinson's disease (middle), and a healthy control (bottom). Spatiotemporal features extracted from gait and eye movement videos are projected into symmetric positive definite (SPD) matrices through SPD pooling, and the output is generated by the BiMap layer. More structured and intense patterns in the advanced PD group suggest a progression-related synchronization between motor and ocular features, while controls show more diffuse, unstructured correlations, with larger areas of neutral gray in the output of the BiMap layer.*

intermediate fusion, for each feature bank of gait $\mathbf{\Phi}_{\mathbf{L}}^{\mathbf{g}} = \left\{\boldsymbol{G}^i\right\}_{i=1}^{N}$, and of ocular movement $\mathbf{\Phi}_{\mathbf{L}}^{\mathbf{e}} = \left\{\boldsymbol{F}^i\right\}_{i=1}^{N}$, their respective $WH \times N$ rectangular matrices $\boldsymbol{R_g} = \left[vec(\boldsymbol{G}^1), \cdots, vec(\boldsymbol{G}^N)\right]$ and $\boldsymbol{R_e} = \left[vec(\boldsymbol{F}^1), \cdots, vec(\boldsymbol{F}^N)\right]$ are built. Each modality calculates its own $N \times N$ SPD pooling matrix $\boldsymbol{C_g^0} = \boldsymbol{R_g}^T \boldsymbol{R_g}$ and $\boldsymbol{C_e^0} = \boldsymbol{R_e}^T \boldsymbol{R_e}$. *These matrices $\boldsymbol{C_g^0}$ and $\boldsymbol{C_e^0}$ serve as modality-specific second-order descriptors that encode intra-modal covariance patterns, enabling the independent learning of geometric relationships within each domain. Such decomposition not only enhances the interpretability of the learned representations but also allows each branch to specialize in modeling the distinct phenotypic signatures associated with either oculomotor or gait impairments. In such a case ($\boldsymbol{C_g^0}$, $\boldsymbol{C_e^0}$) approximates a normal distribution from features coded from each representation, without considering mean components that may overlap between PD and control classes. We expect that the learning of compact representation for each modality, along geometrical branches, may effectively code discriminating patterns from each modal observation.* Next, for each branch, the geometric layers Bimap, ReEig, and the logarithmic mapping are computed as detailed in the previous subsection. Finally, these embeddings are vectorized and concatenated, to be processed by the same two dense layers $\mathcal{D}$. In the output layers, a Softmax layer computes the probability of discrimination between 6 motor affectations from control and PD patients. The intermediate fusion approach is illustrated in Figure 4.

4.3. Motor report prediction: A multitask learning

A main interest of this work is to develop a computational tool with the ability to support PD characterization, following standard scales under cardinal, postural, balance and mobility symptoms. Hence, the final geometrical layer C^L is mapped to a Euclidean space $\mathbf{log}(C^L)$ allowing operation over standard deep operations. In both early and intermediate representations, the respective $\mathbf{log}(C^L)$ is further processed by a dense layer (D). *This strategy allows the model to learn impairment-specific feature representations while benefiting from shared information across related tasks. In such a sense, the proposed approach is designed to support item characterization of standard scales, allowing for detection of specific impairments that globally bring a PD deficit score. Even, we hypothesize that learning about local and multiple impairments over the same backbone may benefit learning of representation.*. The global motor loss L is defined as:

$$L = \underbrace{\gamma_1 L_{gb} + \gamma_2 L_{ob}}_{\text{Cardinal}} + \underbrace{\gamma_3 L_{wp} + \gamma_4 L_{bg}}_{\text{Posture and Balance}} + \underbrace{\gamma_5 L_{fg} + \gamma_6 L_{ga}}_{\text{Mobility}}$$

where γ_i is an importance weight for each considered motor symptom. *The cardinal symptoms, gait bradykinesia* (L_{gb}) and ocular bradykinesia (L_{ob}), *are both* associated with slowness of movement, a reduction in amplitude (hypokinesia), the absence of movement (akinesia), or a progressive and repetitive reduction in motion. The motor loss also considers changes in posture and balance. Specifically, here, we consider labels associated with bilateral gait impairment (L_{bg}) and the wrong posture (L_{wp}) that involve observations associated with postural distortion (forward flexion of the trunk and symmetrical disruption of walking patterns decreasing velocity and increasing shuffling). Additionally, the freezing of gait L_{fg} and gait autonomy L_{ga} were considered mobility symptoms. These observations aim to characterize impaired autonomy, typically for starting or sustaining locomotion. All loss functions are based on cross-entropy. The rationale for using six independent predictions is to drive the model to learn specific features for each impairment. This is particularly important in the context of Parkinson's disease, as a patient may exhibit ocular bradykinesia without necessarily having gait impairment or an incorrect posture.

Multitask learning enables the model to capture unique impairment-specific patterns while leveraging complementary information. By independently optimizing each impairment prediction, the model can learn distinctive yet interrelated features, improving classification accuracy across multiple symptoms. To further enhance interpretability, we introduce a motor impairment heat-map as an additional output. This visualization highlights the affected laterality of each patient, aiding specialists in identifying predominant impairments and facilitating personalized assessments. Figure 6 illustrates the predicted motor impairments for a PD patient and a control subject. The heatmap rows correspond to different impairments (bilateral impairment, freezing, gait autonomy, bradykinesia in gait, ocular bradykinesia, and poor posture), while the columns represent the eight multimodal predictions obtained per subject: the first four columns correspond to left-side predictions (gait and left eye), and the last four columns correspond to right-side predictions (gait and right eye). In this example, the PD patient exhibits greater impairment on the left side, demonstrating the model's capability to capture asymmetries.

4.4. Experimental setup

The proposed approach was validated under 5-fold cross-validation method taking at each iteration, 26 subjects for model training and 6 subjects for testing. For these experiments, the Parkinsonian patients (and the control participants) who were correctly identified were considered true positives (TP) and true negatives (TN), respectively. A set of metrics was subsequently utilized to comprehensively evaluate the performance of the method in its different configurations. The metrics calculated here are accuracy (*acc*), sensitivity (*sen*), precision (*prec*), specificity (*spec*) and the F1-score ($F1$-s). In this way, each metric is calculated for each binary label of classification report. The labels were based on predominant motor impairments: a laterality gait affectation (H&Y scale) and specific items of the MDS-UPDRS part III scale (gait autonomy, freezing of gait, posture impairment, and gait bradykinesia. Additionally, non-standardized observations such as ocular bradykinesia were quantified. Regarding the parameter tuning, the following parameters were implemented:

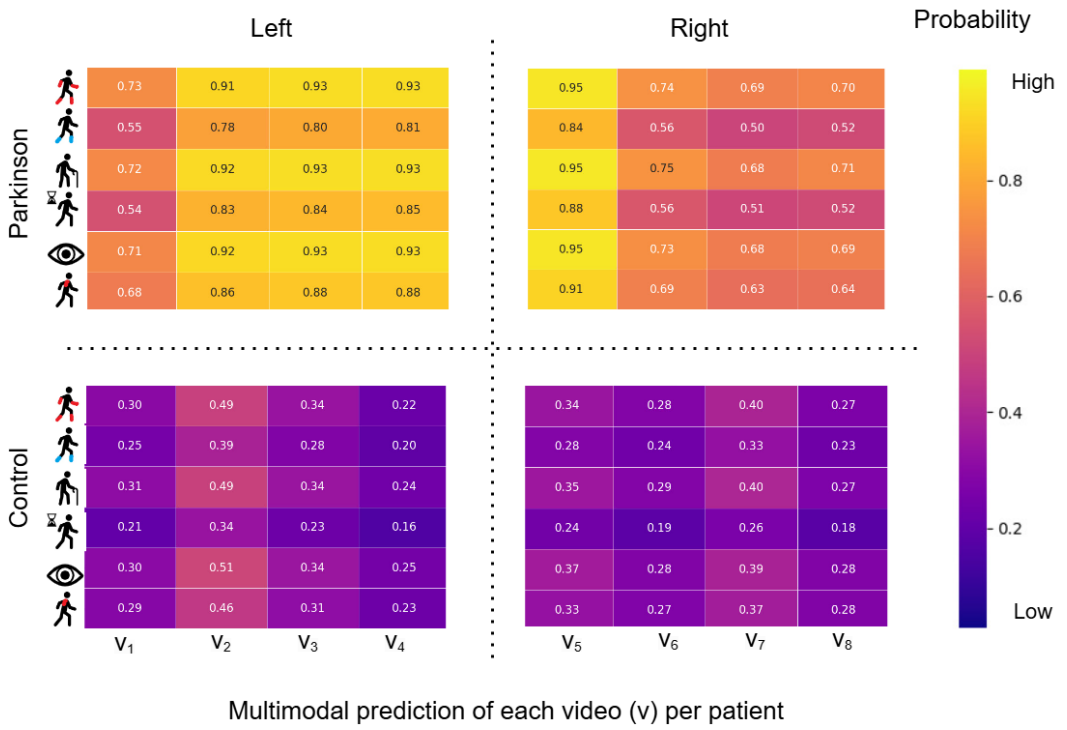

Fig. 6. Graphical representation of the report. Above: motor report of each of the impairments of a patient on the right side and on the left side. Below: motor report of a control subject with right-side and left-side impairments. The rows represent the patient's impairments, listed from top to bottom as follows: bilateral impairment, freezing, gait autonomy, bradykinesia in gait, ocular bradykinesia, and poor posture. The columns represent the 8 multimodal predictions obtained for a subject (one per video).

- **3D Convolutional Representation.** Experiments for each level of fusion were conducted with 1, 2 or 3 3D convolutional layers, with 32, 64, and 128. output channels, respectively.

- **Geometrical Representation:** Experiments for each level of fusion were conducted with 1, 2 BiRe layers, with dimensions of 64×64 and 32×32. output matrices, respectively.

- **Variation of the SPD Pooling Matrix Dimension:** The output channels of the last 3D convolutional layer (N channels for gait and M channels for ocular motion) are related to the input matrix of the geometric phase. In early fusion, the input SPD matrix, which considers the combination of both modalities, has dimensions of $(N + M) \times (N + M)$. In intermediate fusion, the dimensions of the 2 matrices are $N \times N$, and $M \times M$ since SPD pooling is performed independently for each branch. Consequently, depending on the depth of the 3D-CNN, the resulting matrix from early fusion can have dimensions of (64, 64), (128, 128), or (256, 256). For intermediate fusion, the SPD pool matrix for each branch can have dimensions of (32, 32), (64, 64), or (128, 128).

- ***Preprocessing Characteristics:*** *Before training, all videos were normalized to approximately 5 seconds by subsampling frames equidistantly throughout the video. The gait recordings captured the complete displacement of the patient. Regarding ocular motion, initial recordings included the entire facial expression, but the eye region was manually cropped while maintaining a spatial resolution of* 210×140 *pixels. The labeling process was conducted manually by a neurologist trained in standardized evaluation scales. Each ocular video received a single annotation, while each gait video was*

assigned five distinct labels corresponding to motor impairments: ocular bradykinesia, gait bradykinesia, bilateral gait impairment, postural instability, freezing of gait, and gait autonomy.

- ***Training Characteristics:*** *The model was trained with an early stopping criterion based on the minimum number of epochs required for loss reduction. To prevent overfitting, a dropout and a regularization technique were applied to the 3D convolutional layers. Additionally, data augmentation techniques, specifically horizontal flipping, were implemented during training.*

5. Evaluation and Results

This work introduced a geometric multimodal architecture to generate a motor scale report about Parkinsonism affectations observed from markerless videos of oculomotor tasks and gait. Hence, we consider two fusion alternatives of eye and gait observations, analyze geometric components, and extend the analysis to multiple stages of the disease. The next subsections describe the results of these considerations.

5.1. Results from Early fusion

Firstly, we evaluated the convolutional (conv3D) representation to form SPD descriptors. Table 1 summarizes the experiments using 1 and 3 conv3D layers both considering 1 BiRe layer. The results consistently demonstrate that using three layers yields better performance for Ocular and Gait Bradykinesia, Wrong posture, Bilateral impairment and Gait autonomy (average accuracy and F1-score gain of 6%, 5%, respectively). In contrast, Freezing shows gain of 3% in F1-s with experiment of one 3D conv layer.

Early fusion (ablation study) Number of 3D conv layers		**Cardinal Symptoms**		**Posture and balance symptoms**		**Gait and mobility symptoms**	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
1 layer 3D conv	**Prec**	0.85 ± 0.07	0.86 ± 0.17	0.85 ± 0.13	0.82 ± 0.07	0.85 ± 0.19	0.81 ± 0.07
	Sen	0.86 ± 0.14	0.89 ± 0.08	0.92 ± 0.06	0.94 ± 0.06	0.90 ± 0.11	0.94 ± 0.06
	Spec	0.73 ± 0.18	0.75 ± 0.32	0.66 ± 0.37	0.60 ± 0.33	0.76 ± 0.38	0.60 ± 0.33
	Acc	0.81 ± 0.12	0.83 ± 0.13	0.83 ± 0.11	0.84 ± 0.07	0.83 ± 0.17	0.83 ± 0.07
	F1-s	0.85 ± 0.09	0.85 ± 0.08	0.87 ± 0.06	0.87 ± 0.05	**0.85 ± 0.10**	0.87 ± 0.05
3 layers 3D conv	**Prec**	0.90 ± 0.13	0.88 ± 0.14	0.91 ± 0.10	0.90 ± 0.13	0.85 ± 0.17	0.90 ± 0.13
	Sen	0.96 ± 0.06	0.87 ± 0.07	0.95 ± 0.06	0.96 ± 0.06	0.82 ± 0.15	0.95 ± 0.07
	Spec	0.83 ± 0.21	0.86 ± 0.17	0.86 ± 0.16	0.82 ± 0.20	0.87 ± 0.15	0.84 ± 0.20
	Acc	**0.90 ± 0.10**	**0.86 ± 0.11**	**0.91 ± 0.10**	**0.90 ± 0.10**	0.84 ± 0.12	**0.89 ± 0.11**
	F1-s	**0.92 ± 0.08**	**0.87 ± 0.10**	**0.93 ± 0.08**	**0.92 ± 0.07**	0.82 ± 0.12	**0.92 ± 0.08**

Table 1. Performance metrics for the two best experiments associated with *Early* Fusion models (1 or 3 3D convolutional layers).

Secondly, we evaluated the geometric learning components that provide SPD patterns in the multiclassification report. Table 2 summarizes the experiments using 1 and 2 BiRe layers, being consistently better the use of two BiRe layers for Ocular and Gait Bradykinesia, Bilateral impairment, and Gait autonomy (with average accuracy and F1-score gain of 4% and 3%, respectively). Freezing shows similar metrics across both methods. In contrast, Wrong posture shows gain of 1% in F1-s on the experiment with 1 BiRe layer. In such a sense, the configuration the configuration with 3 3D CNN layers and 2 BiRe blocks results the most consistent representation in early fusion.

5.2. Results from Intermediate fusion

For intermediate fusion representation, we also first evaluated the convolutional representation to form SPD descriptors. Table 3 summarized the achieved results at different items, evidencing a superior performance for representation with three 3D convolutional layers for Gait and Ocular Bradykinesia, Wrong posture, Bilateral impairment and Freezing (average accuracy and F1-score gain of 4%, 3%, respectively). In contrast, one 3D convolutional layer was more consistent for the Wrong posture (gain of 2% and 5% in accuracy and F1-score). All experiments included 1 BiRe layer.

Early fusion (ablation study) Number of BiRe layers		Cardinal Symptoms		Posture and balance symptoms		Gait and mobility symptoms	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
2 layers BiRe 3 layers 3D conv	**Prec**	0.92 ± 0.14	0.89 ± 0.13	0.90 ± 0.12	0.93 ± 0.14	0.84 ± 0.20	0.93 ± 0.14
	Sen	0.97 ± 0.05	0.94 ± 0.06	0.96 ± 0.04	0.98 ± 0.04	0.82 ± 0.14	0.97 ± 0.05
	Spec	0.89 ± 0.21	0.86 ± 0.17	0.85 ± 0.18	0.89 ± 0.21	0.85 ± 0.17	0.89 ± 0.21
	Acc	**0.93 ± 0.10**	**0.90 ± 0.08**	0.91 ± 0.10	**0.93 ± 0.10**	0.83 ± 0.13	**0.93 ± 0.10**
	F1-s	**0.94 ± 0.08**	**0.91 ± 0.07**	0.92 ± 0.08	**0.95 ± 0.08**	0.81 ± 0.13	**0.94 ± 0.08**
1 layer BiRe 3 layers 3D conv	**Prec**	0.90 ± 0.13	0.88 ± 0.14	0.91 ± 0.10	0.90 ± 0.13	0.85 ± 0.17	0.90 ± 0.13
	Sen	0.96 ± 0.06	0.87 ± 0.07	0.95 ± 0.06	0.96 ± 0.06	0.82 ± 0.15	0.95 ± 0.07
	Spec	0.83 ± 0.21	0.86 ± 0.17	0.86 ± 0.16	0.82 ± 0.20	0.87 ± 0.15	0.84 ± 0.20
	Acc	0.90 ± 0.10	0.86 ± 0.11	0.91 ± 0.10	0.90 ± 0.10	0.84 ± 0.12	0.89 ± 0.11
	F1-s	0.92 ± 0.08	0.87 ± 0.10	**0.93 ± 0.08**	0.92 ± 0.07	0.82 ± 0.12	0.92 ± 0.08

Table 2. Performance metrics for the two experiments associated with *Early* Fusion models (1 or 2 BiRe layers with three 3D convolutional layers).

Intermediate fusion (ablation study) Number of 3D conv layers		Cardinal Symptoms		Posture and balance symptoms		Gait and mobility symptoms	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
2 layers 3D conv	**Prec**	0.84 ± 0.13	0.83 ± 0.20	0.83 ± 0.17	0.80 ± 0.13	0.80 ± 0.24	0.84 ± 0.13
	Sen	0.91 ± 0.16	0.84 ± 0.11	0.95 ± 0.10	0.89 ± 0.19	0.71 ± 0.27	0.90 ± 0.16
	Spec	0.71 ± 0.23	0.75 ± 0.3	0.68 ± 0.32	0.62 ± 0.26	0.86 ± 0.17	0.72 ± 0.23
	Acc	0.82 ± 0.15	0.79 ± 0.15	**0.83 ± 0.18**	0.77 ± 0.18	0.77 ± 0.16	0.82 ± 0.14
	F1-s	0.86 ± 0.12	0.81 ± 0.12	**0.88 ± 0.13**	0.82 ± 0.14	0.72 ± 0.16	0.86 ± 0.11
3 layers 3D conv	**Prec**	0.89 ± 0.14	0.86 ± 0.16	0.84 ± 0.17	0.88 ± 0.14	0.81 ± 0.22	0.89 ± 0.14
	Sen	0.90 ± 0.17	0.83 ± 0.12	0.83 ± 0.21	0.88 ± 0.17	0.78 ± 0.24	0.89 ± 0.18
	Spec	0.81 ± 0.23	0.84 ± 0.21	0.80 ± 0.21	0.80 ± 0.23	0.86 ± 0.17	0.81 ± 0.23
	Acc	0.84 ± 0.15	**0.84 ± 0.13**	0.81 ± 0.20	**0.83 ± 0.15**	**0.80 ± 0.16**	0.84 ± 0.15
	F1-s	0.87 ± 0.12	**0.84 ± 0.12**	0.83 ± 0.19	**0.86 ± 0.12**	**0.77 ± 0.21**	0.87 ± 0.12

Table 3. Performance metrics for the two best experiments associated with *Intermediate* Fusion models (2 or 3 3D convolutional layers).

Regarding geometrical learning components, Table 4 summarizes the achieved results in the intermediate representation. Interestingly, adding one more geometric layer shows an average gain of 5% in accuracy and F1-score for all impairments. In the context of intermediate fusion the best configuration was achieved with three 3D convolutional layers and two BiRe blocks.

Intermediate fusion (ablation study) Number of BiRe layers		Cardinal Symptoms		Posture and balance symptoms		Gait and mobility symptoms	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
2 layer BiRe 3 layers 3D conv	**Prec**	0.90 ± 0.13	0.88 ± 0.15	0.92 ± 0.10	0.91 ± 0.13	0.84 ± 0.19	0.90 ± 0.12
	Sen	0.96 ± 0.08	0.88 ± 0.15	0.91 ± 0.11	0.95 ± 0.08	0.79 ± 0.15	0.90 ± 0.12
	Spec	0.83 ± 0.20	0.85 ± 0.18	0.88 ± 0.15	0.85 ± 0.19	0.86 ± 0.17	0.83 ± 0.19
	Acc	**0.89 ± 0.10**	**0.87 ± 0.12**	**0.90 ± 0.11**	**0.90 ± 0.09**	**0.82 ± 0.12**	**0.90 ± 0.09**
	F1-s	**0.92 ± 0.08**	**0.88 ± 0.11**	**0.91 ± 0.09**	**0.93 ± 0.07**	**0.80 ± 0.12**	**0.92 ± 0.07**
1 layer BiRe 3 layers 3D conv	**Prec**	0.89 ± 0.14	0.86 ± 0.16	0.84 ± 0.17	0.88 ± 0.14	0.81 ± 0.22	0.89 ± 0.14
	Sen	0.90 ± 0.17	0.83 ± 0.12	0.83 ± 0.21	0.88 ± 0.17	0.78 ± 0.24	0.89 ± 0.18
	Spec	0.81 ± 0.23	0.84 ± 0.21	0.80 ± 0.21	0.80 ± 0.23	0.86 ± 0.17	0.81 ± 0.23
	Acc	0.84 ± 0.15	0.84 ± 0.13	0.81 ± 0.20	0.83 ± 0.15	0.80 ± 0.16	0.84 ± 0.15
	F1-s	0.87 ± 0.12	0.84 ± 0.12	0.83 ± 0.19	0.86 ± 0.12	0.77 ± 0.21	0.87 ± 0.12

Table 4. Performance metrics for the two experiments associated with *Intermediate* Fusion models (1 or 2 BiRe layers with three 3D convolutional layers).

5.3. Multimodal and geometric contribution

During validation, the proposed multimodal approach was also compared regarding unimodal version using three 3D convolutional layers with one or two BiRe block. In unimodal approaches, a label prediction is made for eye movement (ocular bradykinesia) and five labels are predicted based on the gait modality. As reported in Table 5, standing out scores were achieved for the sensitivity of ocular movement as well as for predicting bilateral impairment (85%) and gait autonomy (83%). However, the other metrics report a remarked reduction as observed in the F1-score. For the unimodal gait model, there is not significative

Modalities		Cardinal Symptoms		Posture and balance symptoms		Gait and mobility symptoms	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
Unimodal 3 3D Conv-1 BiRe	**Prec**	0.78 ± 0.13	0.81 ± 0.22	0.78 ± 0.20	0.78 ± 0.14	0.78 ± 0.26	0.78 ± 0.16
	Sen	**1**	0.76 ± 0.14	0.81 ± 0.19	0.85 ± 0.18	0.72 ± 0.26	0.83 ± 0.17
	Spec	0.65 ± 0.15	0.74 ± 0.32	0.6 ± 0.36	0.59 ± 0.33	0.80 ± 0.27	0.59 ± 0.33
	Acc	0.85 ± 0.12	0.75 ± 0.18	0.72 ± 0.19	0.73 ± 0.16	0.74 ± 0.19	0.72 ± 0.16
	F1-s	0.87 ± 0.08	0.76 ± 0.15	0.77 ± 0.15	0.79 ± 0.11	0.70 ± 0.21	0.78 ± 0.11
Unimodal 3 3D Conv-2 BiRe	**Prec**	0.81 ± 0.15	0.83 ± 0.15	0.77 ± 0.17	0.82 ± 0.15	0.76 ± 0.26	0.83 ± 0.15
	Sen	0.82 ± 0.23	0.81 ± 0.15	0.80 ± 0.25	0.83 ± 0.21	0.71 ± 0.27	0.82 ± 0.22
	Spec	0.64 ± 0.30	0.76 ± 0.24	0.63 ± 0.29	0.67 ± 0.30	0.77 ± 0.27	0.68 ± 0.31
	Acc	0.73 ± 0.15	0.79 ± 0.10	0.73 ± 0.18	0.75 ± 0.13	0.72 ± 0.18	0.74 ± 0.14
	F1-s	0.77 ± 0.12	0.80 ± 0.09	0.76 ± 0.17	0.79 ± 0.10	0.69 ± 0.22	0.78 ± 0.11
Intermediate Fusion 3 3D conv-2 Bires	**Prec**	0.90 ± 0.13	0.88 ± 0.15	0.92 ± 0.10	0.91 ± 0.13	0.84 ± 0.19	0.90 ± 0.12
	Sen	0.96 ± 0.08	0.88 ± 0.15	0.91 ± 0.11	0.95 ± 0.08	0.79 ± 0.15	0.90 ± 0.12
	Spec	0.83 ± 0.20	0.85 ± 0.18	0.88 ± 0.15	0.85 ± 0.19	0.86 ± 0.17	0.83 ± 0.19
	Acc	0.89 ± 0.10	0.87 ± 0.12	0.90 ± 0.11	0.90 ± 0.09	0.82 ± 0.12	0.90 ± 0.09
	F1-s	0.92 ± 0.08	0.88 ± 0.11	0.91 ± 0.09	0.93 ± 0.07	0.80 ± 0.12	0.92 ± 0.07
Early Fusion 3 3D Conv-2 BiRes	**Prec**	0.92 ± 0.14	0.89 ± 0.13	0.90 ± 0.12	0.93 ± 0.14	0.84 ± 0.20	0.93 ± 0.14
	Sen	0.97 ± 0.05	0.94 ± 0.06	0.96 ± 0.04	0.98 ± 0.04	0.82 ± 0.14	0.97 ± 0.05
	Spec	0.89 ± 0.21	0.86 ± 0.17	0.85 ± 0.18	0.89 ± 0.21	0.85 ± 0.17	0.89 ± 0.21
	Acc	0.93 ± 0.10	0.90 ± 0.08	0.91 ± 0.10	0.93 ± 0.10	0.83 ± 0.13	0.93 ± 0.10
	F1-s	**0.94 ± 0.08**	**0.91 ± 0.07**	**0.92 ± 0.08**	**0.95 ± 0.08**	**0.81 ± 0.13**	**0.94 ± 0.08**

Table 5. Comparison between Unimodal and Multimodal approaches for different item predictions

improvements to include additional BiRe blocks, a fact associated to information with low variance from only one modality, which limit the learning of more complex patterns.

Contrary, fusing the modalities yields improvements over the best unimodal results. For instance, in the intermediate fusion, the F1 score for the motor impairments shows an average improvement of 13%, while accuracy increases by 12%. Early fusion presents slightly higher gains, with improvements of 15% and 13%, respectively demonstrating that fusing in the Riemannian manifold yields the best results. In Table 5, it is also observed that the difference in performance between unimodal and multimodal models is more pronounced for mobility symptoms such as Autonomy in Gait and Freezing of Gait, where multimodal models outperform unimodal ones. The F1-scores across table indicate consistent improvements with multimodal fusion strategies, particularly Early Fusion, which demonstrates the strongest overall performance for predicting motor impairments.

In addition, we compared the proposed approach with a 3D convolutional architecture. The outputs of the two branches are then fused through concatenation and a dense layer for the subsequent prediction of the six items. Table 6 summarizes the results obtained using this baseline model, compared to the Riemannian early fusion approach. The contribution of the geometric phase is evident, increasing most of the various metrics by 40% to 60%. However, the purely 3D convolutional fusion also achieves outstanding sensitivity for some items, such as Ocular Bradykinesia and Walking autonomy 80%, and highlights specificity in the Freezing of gait 80%.

The proposed approach was also compared with classical machine learning approaches that have been previously used for Parkinson's disease classification [17, 43, 44]. To this end, pretrained deep features were extracted, and the Riemannian mean was computed for each video descriptor [17]. Subsequently, the ocular movement descriptor was concatenated with the gait descriptor, and the classification task was performed for the six motor impairments using Logistic Regression (LG), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MP). The results are presented in Table 7. Overall, the proposed multimodal geometric method outperforms classical approaches across all motor impairments. In particular, it achieves superior performance in predicting ocular bradykinesia (F1-score of 94%), freezing of gait (F1-score of 81%), and gait independence (F1-score of 94%). While all metrics are higher for the geometric approach, Random Forest (RF) yielded the best results among the classical methods [17, 44]. For the remaining classical approaches, the best performance was associated with the classification of bilateral gait impairment, whereas the lowest results were observed in the classification of freezing of gait, making it the most challenging motor impairment to predict across all methods, including ours.

The use of a 3D convolutional neural network and Riemannian-based geometric modeling is motivated by

Computational Approaches		Cardinal Symptoms		Posture and balance symptoms		Gait and mobility symptoms	
Configuration	**Metrics**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy in Gait**
Early Fusion 3 3D conv-2 BiRes	**Prec**	0.92 ± 0.14	0.89 ± 0.13	0.90 ± 0.12	0.93 ± 0.14	0.84 ± 0.20	0.93 ± 0.14
	Sen	0.97 ± 0.05	0.94 ± 0.06	0.96 ± 0.04	0.98 ± 0.04	0.82 ± 0.14	0.97 ± 0.05
	Spec	0.89 ± 0.21	0.86 ± 0.17	0.85 ± 0.18	0.89 ± 0.21	0.85 ± 0.17	0.89 ± 0.21
	Acc	0.93 ± 0.10	0.90 ± 0.08	0.91 ± 0.10	0.93 ± 0.10	0.83 ± 0.13	0.93 ± 0.10
	F1-s	**0.94 ± 0.08**	**0.91 ± 0.07**	**0.92 ± 0.08**	**0.95 ± 0.08**	**0.81 ± 0.13**	**0.94 ± 0.08**
3 3D conv layers	**Prec**	0.48 ± 0.26	0.30 ± 0.24	0.33 ± 0.27	0.23 ± 0.29	0.10 ± 0.2	0.48 ± 0.26
	Sen	0.80 ± 0.4	0.60 ± 0.48	0.60 ± 0.48	0.40 ± 0.48	0.20 ± 0.4	0.80 ± 0.40
	Spec	0.20 ± 0.40	0.40 ± 0.48	0.40 ± 0.48	0.60 ± 0.48	0.80 ± 0.40	0.20 ± 0.40
	Acc	0.58 ± 0.10	0.46 ± 0.06	0.50 ± 0.09	0.48 ± 0.13	0.50 ± 0.10	0.58 ± 0.10
	F1-s	0.59 ± 0.30	0.40 ± 0.32	0.42 ± 0.35	0.29 ± 0.3	0.13 ± 0.26	0.59 ± 0.30

Table 6. Geometrical early fusion architecture compared with 3D CNN based fusion, using 3 convolutional layers

the spatiotemporal nature of the input data and the need to preserve the non-Euclidean structure of the features extracted from both ocular and gait videos. Traditional machine learning models, although competitive, rely on flattened representations that overlook the intrinsic temporal dynamics and geometry of the video descriptors. In contrast, the 3D CNN is capable of directly capturing spatiotemporal patterns, which are critical for identifying subtle impairments such as ocular bradykinesia and freezing of gait. Furthermore, the geometric component enables a principled treatment of feature distributions via symmetric positive definite (SPD) matrices, enhancing the expressiveness of the representations. The superior results obtained with our method confirm that this architectural design is well suited to model the complex motor patterns involved in Parkinson's disease.

5.4. Stratification of the disease

Finally, we examined the ability of early fusion to distinguish between various stages of the disease. The proposed method was developed and trained for binary PD/Control classification, but provides a PD probability output, allowing us to analyze the behavior of samples labeled with different PD stages. Consequently, we are interested in evaluating stratification effectiveness on the basis of the patients' stage labels. Figure 7 displays the output probabilities for the 6 predictions. a) ocular bradykinesia, b) bradykinesia in gait, c) wrong posture, d) bilateral affectation of gait (H&Y scale), e) freezing, and f) Autonomy. To assess significant differences, we utilized the Kolmogorov–Smirnov (KS) test to validate the effectiveness of the proposed method in distinguishing between classes. The KS test is a non-parametric technique that evaluates the alignment between two distributions. It allows us to check whether the distributions of two samples are significantly different or not.

We analyzed the probability distribution of cardinal symptoms, such as bradykinesia at the ocular level and during gait. The prediction based on ocular bradykinesia was significantly different ($\rho < 10^{-44}$) between the control group and the PD group, as shown in the graph (a) of figure 7. For bradykinesia during gait (graph b), significant differences are observed between the normal group (without problems) and all other stages (slight, mild, and moderate) of the disease. However, there were no significant differences between the mild and moderate stages ($\rho > 0.2$).

Postural impairment (graph (c) of the figure 7) presents significant differences across the various stages in the dataset: normal, slight and mild, all with ρ values less than 0.05. For the prediction based on bilateral gait impairment (item (d) from figure 7; modified H&Y scale), there were significant differences ($\rho < 0.05$) between all stages included in the dataset (0, 1, 1.5, 2, and 3). With exception of stages 1.5 and 2 ($\rho > 0.5$), or between 2 and 3 ($\rho > 0.2$). Interestingly, freezing of gait (graph (e) of Figure 7) shows significant differences across the various stages in the dataset (all ρ values less than 0.05): normal, slight, mild, and moderate. In graph (f) of Figure 7, which shows the prediction of gait autonomy (gait item, MDS-UPDRS Part III), significant differences are observed between each of the Parkinson's stages in the dataset (normal,slight, mild, moderate). However, there are no statistical differences between the slight and mild stages ($\rho > 0.2$).

Computational Approaches		Cardinal Symptoms		Posture and balance Symptoms		Gait and mobility symptoms	
Config	**Metr**	**Bradykinesia Ocular**	**Bradykinesia in Gait**	**Wrong Posture**	**Bilateral Afectation (H&Y)**	**Freezing of Gait**	**Autonomy of Gait**
Early Fusion 3 3D conv 2 Bires	**Prec**	0.92 ± 0.14	0.89 ± 0.13	0.90 ± 0.12	0.93 ± 0.14	0.84 ± 0.20	0.93 ± 0.14
	Sen	0.97 ± 0.05	0.94 ± 0.06	0.96 ± 0.04	0.98 ± 0.04	0.82 ± 0.14	0.97 ± 0.05
	Spec	0.89 ± 0.21	0.86 ± 0.17	0.85 ± 0.18	0.89 ± 0.21	0.85 ± 0.17	0.89 ± 0.21
	Acc	0.93 ± 0.10	0.90 ± 0.08	0.91 ± 0.10	0.93 ± 0.10	0.83 ± 0.13	0.93 ± 0.10
	F1-s	**0.94 ± 0.08**	**0.91 ± 0.07**	**0.92 ± 0.08**	**0.95 ± 0.08**	**0.81 ± 0.13**	**0.94 ± 0.08**
Logistic Regression (LR)	**Prec**	0.66 ± 0.07	0.77 ± 0.17	0.76 ± 0.07	0.87 ± 0.13	0.60 ± 0.14	0.79 ± 0.22
	Sen	0.69 ± 0.26	0.71 ± 0.17	0.72 ± 0.21	0.69 ± 0.17	0.50 ± 0.22	0.68 ± 0.25
	Spec	0.67 ± 0.17	0.76 ± 0.17	0.74 ± 0.09	0.85 ± 0.14	0.72 ± 0.06	0.81 ± 0.18
	Acc	0.66 ± 0.08	0.74 ± 0.14	0.73 ± 0.10	0.76 ± 0.07	0.62 ± 0.10	0.75 ± 0.16
	F1-s	0.64 ± 0.11	0.73 ± 0.15	0.73 ± 0.13	0.75 ± 0.10	0.54 ± 0.19	0.72 ± 0.21
Random Forest (RF)	**Prec**	0.71 ± 0.28	0.81 ± 0.17	0.79 ± 0.14	0.85 ± 0.10	0.75 ± 0.20	0.80 ± 0.18
	Sen	0.74 ± 0.25	0.85 ± 0.13	0.93 ± 0.07	0.95 ± 0.08	0.74 ± 0.16	0.75 ± 0.22
	Spec	0.66 ± 0.34	0.74 ± 0.28	0.68 ± 0.25	0.75 ± 0.20	0.70 ± 0.26	0.73 ± 0.28
	Acc	0.67 ± 0.20	0.79 ± 0.10	0.81 ± 0.13	0.87 ± 0.10	0.72 ± 0.17	0.74 ± 0.10
	F1-s	0.68 ± 0.17	**0.81 ± 0.06**	**0.85 ± 0.09**	**0.89 ± 0.08**	0.73 ± 0.14	0.74 ± 0.12
Support Vector Machine (SVM)	**Prec**	0.54 ± 0.33	0.55 ± 0.17	0.66 ± 0.23	0.72 ± 0.21	0.59 ± 0.20	0.68 ± 0.22
	Sen	0.26 ± 0.20	0.61 ± 0.18	0.62 ± 0.24	0.68 ± 0.23	0.43 ± 0.09	0.66 ± 0.25
	Spec	0.61 ± 0.34	0.50 ± 0.15	0.61 ± 0.23	0.65 ± 0.19	0.67 ± 0.22	0.67 ± 0.23
	Acc	0.41 ± 0.09	0.55 ± 0.16	0.62 ± 0.19	0.67 ± 0.23	0.56 ± 0.12	0.66 ± 0.21
	F1-s	0.25 ± 0.16	0.58 ± 0.17	0.63 ± 0.20	0.69 ± 0.20	0.49 ± 0.11	0.66 ± 0.22
Multilayer Perceptron (MP)	**Prec**	0.66 ± 0.17	0.66 ± 0.07	0.77 ± 0.17	0.77 ± 0.09	0.64 ± 0.13	0.70 ± 0.11
	Sen	0.84 ± 0.23	0.76 ± 0.23	0.75 ± 0.22	0.63 ± 0.14	0.61 ± 0.09	0.79 ± 0.16
	Spec	0.55 ± 0.28	0.57 ± 0.17	0.69 ± 0.26	0.70 ± 0.16	0.66 ± 0.18	0.57 ± 0.33
	Acc	0.66 ± 0.12	0.66 ± 0.07	0.73 ± 0.16	0.67 ± 0.03	0.63 ± 0.09	0.70 ± 0.09
	F1-s	0.69 ± 0.11	0.69 ± 0.08	0.74 ± 0.15	0.68 ± 0.08	0.62 ± 0.09	0.73 ± 0.07

Table 7. Comparison of the geometrical early fusion architecture with classical machine learning methods.

6. Discussion and concluding remarks

Today, the evaluation of patients with PD is carried out via scales such as modified H&Y scale and MDS-UPDRS, but depending on the expertise of specialists [5, 7]. This work introduces an end-to-end geometric multimodal approach that combines gait and eye movement patterns to support multiple-scale motor impairments classification. For instance, bradykinesia (both ocular and during gait), but also to symptoms related to posture and balance, such as incorrect posture and bilateral gait impairment, as well symptoms associated with mobility impairments, such as freezing of gait and gait autonomy. The proposed approach explored the early and intermediate covariate fusion of 3D convolutional inputs, showing the robustness of the geometrical approach to predict scale items of the modified Hoehn and Yahr scale, four items from the MDS-UPDRS Part III, and an observation by a specialist related to-eye-movement. Particularly, early fusion configuration shows greater robustness, improving by approximately 3% in F1-score and accuracy for cardinal symptoms such as ocular and gait bradykinesia, as well as bilateral impairment and gait autonomy. *The improved performance of early fusion over intermediate fusion can be attributed to the use of a global second-order covariance representation. Unlike intermediate fusion, which computes modality-specific SPD matrices independently, early fusion constructs a single SPD matrix that captures both intra-modal (within each modality) and inter-modal (across modalities) correlations. This unified representation enables the model to learn complex co-activation patterns between ocular and gait features—patterns that are clinically relevant given the progressive and concurrent motor and oculomotor impairments in Parkinson's disease. By integrating information at an earlier stage, the model leverages subtle yet meaningful cross-modal dependencies, leading to a more expressive and discriminative representation of the underlying pathology.* The intermediate fusion approach considering three 3D conv layers and 2 BiRe blocks shows gains between 11% and 12% in F1 score compared with the best unimodal approaches. Alternatively, fusing in the Riemannian manifold (early fusion) considering three 3D conv layers and 2 Bires, reports average gains from 15% in F1 score and 13% in accuracy compared to the best unimodal results. *These results highlight*

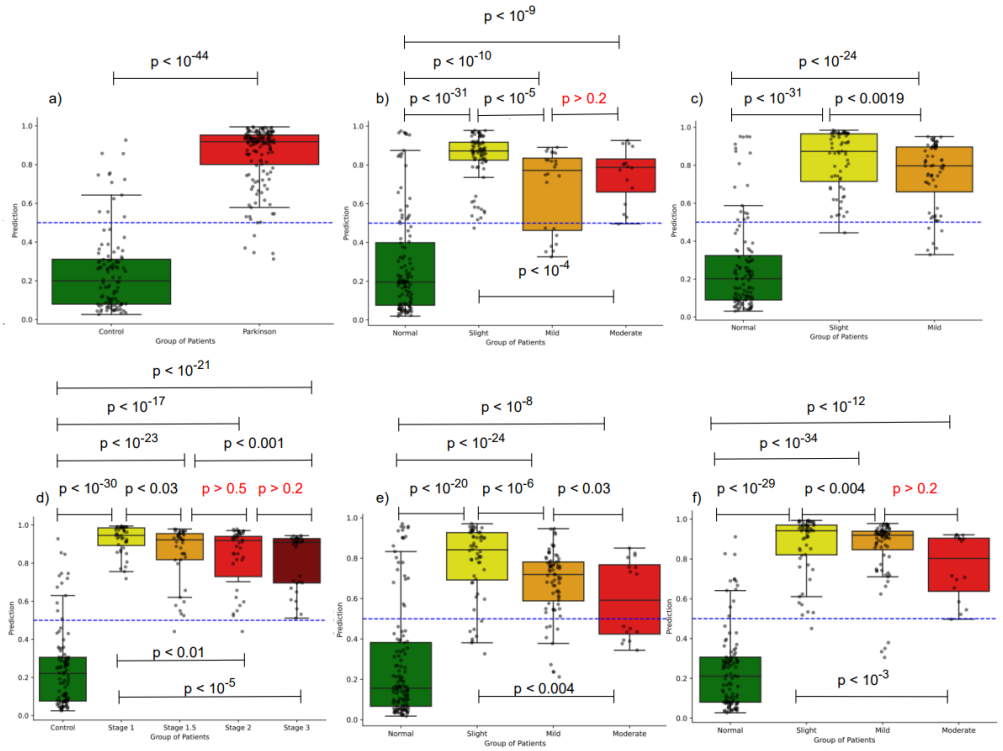

Fig. 7. Distributions of Parkinson probability outputs by the early fusion binary classifier, for the different stage groups. The PD patients were categorized into different stages of the disease on the basis of a) ocular bradykinesia, b) bradykinesia in gait, c) wrong posture, d) bilateral affectation of gait (H&Y scale), e) freezing of gait, f) autonomy of gait. The values of ρ that appear in red correspond to insignificant differences between the two distributions.

the advantage of multimodal integration, as fusing information from both gait and eye movement modalities significantly outperforms unimodal baselines. This demonstrates that complementary information from different motor subsystems can be effectively combined to model the complex and heterogeneous symptomatology of Parkinson's disease. The early fusion of gait and eye patterns, through covariance representation, allows to capture of primary correspondences and second-order discriminative patterns, from multimodal analysis. Besides, from such a fused descriptor, which approximates a normal multimodal distribution, it is possible to learn new geometrical patterns, preserving data structure but finding new relationships to encode compact and major discriminative descriptors.

Thereafter, we discuss in detail how each considered impairment can be support from the proposed approach concerning the analysis of PD. For instance, for body bradykinesia, where individuals exhibit slower movement initiation and execution [45], we observed that early fusion approach significantly differed between groups of subjects without bradykinesia (normal), slight, and mild. Nonetheless, there is not statistical difference between the mild and moderate stages (Figure 7, graph b), probably because patients were recorded under the effect of medication (mainly levodopa). Regarding posture, as observed in Figure 7, graph c), the output PD probabilities three groups are significantly different: normal, slight, and mild. The values closest to one correspond to the slight stage, with all patients being under the effect of medication. Also, the probability distributions for bilateral gait disturbances [46] have limitations to differentiate among intermediate stages (1.5 to 2 and 2 to 3), but it is compelling between stages (0 to 1.5). These results may be justified because observation of stage 1.5 involves unilateral and axial impairment. However, the video recording is only in the sagittal plane, making it difficult to identify axial impairment. Additionally, the evaluation of Stage 3 involves a retropulsion test, which is not recorded and included in the dataset, making

it challenging for the computational method to differentiate between Stage 2 and Stage 3.

In addition to classification accuracy, the results provide insights into the ability of the proposed architecture to stratify disease severity. The use of the Kolmogorov–Smirnov (KS) test confirms that the output probability distributions are significantly different across various stages of Parkinson's disease for multiple symptoms. This analysis was included to place special emphasis on the statistical significance of the results, ensuring that the observed differences between groups are not due to random variability. In particular, the model demonstrates strong discriminative power for ocular bradykinesia, postural impairment, and freezing of gait. These findings support the clinical relevance of the learned representations and their potential application for disease monitoring. The statistical significance of these results (p-values < 0.05 in most contrasts) reinforces the robustness of the approach and its applicability beyond binary classification. To enhance multiple output PD analysis, the proposoed approach analyzed and supported freezing gait episodes [47]. In Figure 7, graph e), the probability distributions for Freezing of gait, reporting significant p-values in all evaluated stages: normal, slight, mild, and moderate. Aditionally, this work quantify the ability to walk independently and safely[48]. In Figure 7, graph f), the probability distributions are showing significant p-values between the stages, except between the mild and moderate stages. To even improve standard scales, with patterns reported at early stages, in this study was also considered the ocular bradykinesia related with a delay in initiating movement, as well as a reduction in the amplitude and speed of movement (ocular bradykinesia)[49]. In Figure 7, graph (a), significant distributions with p-values close to zero can be observed, differentiating between the control group and the Parkinson's group. Interestingly, through the early fusion approach, outstanding accuracy results of 92% are obtained.

Regarding PD observations, state-of-the-art approaches principally report unimodal models to predict local impairments, but with restrictions to *include* the multisystemic nature of the disease [50, 51]. Alternative a multimodal approach has included finger tapping, hand movements, and rapid alternating movements, via inertial sensors, predicting and quantifying bradykinesia, using kinematic features [52]. Besides, triaxial gyroscopes and triaxial accelerometers of smartwatches have used to record hand movements, and gait to support quantification of bradykinesia [53]. These approaches however are limited to only detect bradykinesia alterations, losing complementarity with other key motor alterations in PD patients. Other multimodal approaches have used gait, hand movements, and limb tapping movements to quantify rigidity and postural stability in Parkinson's patients [54]. In contrast, the proposed approach not only predicts bradykinesia and postural instability, also, predicts other motor impairments observed in clinical routines: bilateral impairment, freezing of gait, and gait autonomy. Additionally, it can predict ocular bradykinesia, which, although not observed in clinical standards, has been shown in numerous studies to be sensitive in detecting Parkinson's disease, particularly in its early stages. *All patients were recorded under the effects of medication to ensure their safety during data acquisition; however, this introduces a limitation in result interpretation, as medication can mitigate certain motor symptoms, making it more challenging to distinguish between different disease stages. Additionally, the progression of Parkinson's disease is heterogeneous, and differentiating between adjacent stages, such as mild and moderate bradykinesia, can be difficult.* The proposed approach achieves an average F1-score range for the six predicted impairments from a minimum of 81% (freezing of gait) to a maximum of 95% (Bilateral affectation) through early fusion. Intermediate fusion yields an average F1-score ranging from a minimum of 80% (freezing of gait) to a maximum of 93% (bilateral affectation). As a contribution of this work, the analysis of the binary probability distributions reveals that the predicted motor impairments that significantly differ across all stages observed by the expert specialist are: ocular bradykinesia, postural instability, and freezing of gait. *In this study, classification relied on expert evaluation by a neurologist, introducing potential subjectivity in data annotation. Moreover, the H&Y scale utilizes the retropulsion test to classify patients at stage 3 by assessing postural stability through a controlled push, but this test was omitted from data acquisition to minimize fall risk, relying instead on gait and posture assessments to infer postural stability. While these factors pose certain limitations, future research could address them by evaluating patients under both ON and OFF medication conditions to analyze potential variations in model predictions, incorporating more detailed kinematic analyses to enhance classification accuracy, and exploring alternative assessments for postural stability to improve disease staging.*

The proposed method is capable of discriminating and quantifying motor impairments between Parkinson's patients and control subjects using spatiotemporal features and multimodal geometric configurations.

The descriptors provided by the covariance representation are highly compact, with an input size of 256×256 and an output size of 64 × 64 (with 2 BiRe blocks) in the LogEig layer during early fusion, and input size of 128 × 128 with an output size of 32 × 32 (with 2 BiRe blocks) in the LogEig layer for each branch modality during intermediate fusion, enabling efficient learning. Increasing the depth of geometric layers in multimodal approaches enables learning more compact and discriminative representations for distinguishing between patients and control subjects.

The quantification of motor impairments through multimodal geometric strategies enhances diagnostic accuracy by capturing complementary information that includes diverse symptoms, such as cardinal symptoms and those related to balance and mobility, providing a comprehensive assessment of the patient and offering a detailed profile of motor status. This is particularly useful for cases in which certain symptoms may not be evident or are subtle, such as ocular bradykinesia. With this proposed approach, the physician can make informed decisions regarding treatment or follow-up based on objective data. Potentially, this approach can facilitates the evaluation of therapeutic interventions' effectiveness, adjustment according to changes in the patient's symptom profile, and evaluation on disease progression. Furthermore, the proposed approach opens new perspectives in supporting the identification of Parkinson's phenotypes (such as the akinetic-rigid phenotype and the postural instability and gait difficulty phenotype), helping to personalize treatment and improve diagnosis, as to verify how some phenotypes respond better to specific types of intervention. By automating the evaluation of complex motor symptoms and generating reports, the proposed approach can supports the specialist's observations. *From a clinical translation standpoint, the proposed framework has the potential to be integrated in diagnosis workflows, providing support for different items of standard scales. Also, the analysis of recorded videos open the possibility to extend evaluation in telemedicine scenarios. Its reliance on conventional video recordings, combined with the interpretability of the multimodal outputs, offers a practical and scalable solution that could assist neurologists in monitoring motor fluctuations over time. With further validation on larger and more diverse cohorts, the system could be deployed as a clinical decision-support tool, enhancing the objectivity and consistency of motor assessments across different healthcare environments. Further investigations could explore advanced feature extraction strategies to improve the resilience and generalization of models, especially in noisy or less controlled recording environments [43, 55, 56]. Enhancing the quality and discriminative power of features may not only improve classification performance but also provide more robust indicators of subtle or early-stage motor impairments, thus reinforcing the clinical utility of these computational methods in longitudinal monitoring and early diagnosis.*

Our proposed multimodal approach, which integrates ocular and gait movement analysis through 3D convolutional networks and a Riemannian geometric phase using markerless video sequences from independent cameras, is closely aligned with the future directions outlined in recent literature [29]. the use of camera-based video acquisition per modality offers a pathway toward the development of lightweight, portable solutions suitable for real-world clinical settings, in line with the ongoing efforts to transition from controlled laboratory setups to scalable, field-ready technologies [29].

Now the proposed approach needs to be evaluated in a larger cohort of patients to determine, among others, statistical significance within the affected population and the efficacy in several PD phenotypes. Additionally, integrating other modalities, such as hand movements, would allow the assessment of additional items from the MDS-UPDRS, providing the quantification of other key impairments, such as postural or resting tremor. In same way, it is demanding to study the performance of the proposed approach in a population in states where the patient is not under the effects of medication. In this regard, the assessment of ON-OFF effects would allow for models that are more adaptable to the patients' everyday situations. A larger validation and the inclusion of complementary variables can enrich the study , determining the advantages and limitations in a clinical context. However, the current owner dataset is limited to the population of 19 *PD patients, and* 13 *control subjects. Even, to the best of our knowledge, there is not exist public and external datasets with same recording and modality contidions to extend the validation of the proposed approach. Furthermore, future work should evaluate the performance of the proposed model in context with other disease variables and including a larger cohort of patients. In this way, we expect that future works can determine the capability of the model to support the development of personalized treatment strategies and assistive technologies for continuous disease monitoring.*

Acknowledgments

To the Ministry of Science, Technology and Innovation of Colombia by the Ecos-Nord Colombian-French project: *Caracterización de movimientos anormales del parkinson desde patrones oculomotores, de marcha y enfoques multimodales basados en visión computacional*, with code 92694. This work also benefited from the support of ECOS-Nord Colombian-French project number C23MS01 "Computer Vision Multimodal Models for Characterising Motor Disorders related to Parkinson Disease".

Statements & Declarations

Funding

This work was partially funded To the Ministry of Science Technology and Innovation of Colombia (MINCIENCIAS) for the support with the ECOSNORD project: *Caracterización de movimientos anormales del Parkinson desde patrones oculomotores, de marcha y enfoques multimodales basados en visión computacional*. Code 92694.

Competing Interests:

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

Author Contributions:

The authors contributed equally to this work.

Ethics approval:

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Industrial University of Santander and written informed consent was obtained from each participant.

Availability of data and materials:

The data set can be shared with the prior signed consent of the requester. Please contact Fabio Martínez at famarcar@saber.uis.edu.co

References

[1] Z. Ou, J. Pan, S. Tang, D. Duan, D. Yu, H. Nong, Z. Wang, Global trends in the incidence, prevalence, and years lived with disability of parkinson's disease in 204 countries/territories from 1990 to 2019, Frontiers in public health 9 (2021) 776847.

[2] E. Dorsey, T. Sherer, M. S. Okun, B. R. Bloem, The emerging evidence of the parkinson pandemic, Journal of Parkinson's disease 8 (s1) (2018) S3–S8.

[3] A. B. R. L. Silva, R. W. G. de Oliveira, G. P. Diógenes, M. F. de Castro Aguiar, C. C. Sallem, M. P. P. Lima, L. B. de Albuquerque Filho, S. D. P. de Medeiros, L. L. P. de Mendonça, P. C. de Santiago Filho, et al., Premotor, nonmotor and motor symptoms of parkinson's disease: A new clinical state of the art, Ageing Research Reviews 84 (2023) 101834.

[4] J. Burtscher, E. M. Moraud, D. Malatesta, G. P. Millet, J. F. Bally, A. Patoz, Exercise and gait/movement analyses in treatment and diagnosis of parkinson's disease, Ageing Research Reviews (2023) 102147.

[5] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, et al., Movement disorder society task force report on the hoehn and yahr staging scale: status and recommendations the movement disorder society task force on rating scales for parkinson's disease, Movement disorders 19 (9) (2004) 1020–1028.

[6] 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, et al., Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results, Movement disorders: official journal of the Movement Disorder Society 23 (15) (2008) 2129–2170.

[7] C. Goetz, W. Poewe, B. Dubois, A. Schrag, M. Stern, A. Lang, et al., Mds-unified parkinson's disease rating scale (mds-updrs), Available from the International Parkinson and Movement Disorder Society website: https://www. movementdisorders. org/MDS/MDS-Rating-Scales/MDS-Unified-Parkinsons-Disease-Rating-Scale-MDS-UPDRS. htm.
[8] M. Tinelli, P. Kanavos, F. Grimaccia, The value of early diagnosis and treatment in parkinson's disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in parkinson's disease.
[9] K. Frei, Abnormalities of smooth pursuit in parkinson's disease: A systematic review, Clinical parkinsonism & related disorders 4 (2021) 100085.
[10] C. A. Antoniades, M. Spering, Eye movements in parkinson's disease: from neurophysiological mechanisms to diagnostic tools, Trends in Neurosciences.
[11] P. Tsitsi, M. N. Benfatto, G. Ö. Seimyr, O. Larsson, P. Svenningsson, I. Markaki, Fixation duration and pupil size as diagnostic tools in Parkinson's disease, Journal of Parkinson's Disease 11 (2) (2021) 865–875.
[12] J. Zhang, B. Zhang, Q. Ren, Q. Zhong, Y. Li, G. Liu, X. Ma, C. Zhao, Eye movement especially vertical oculomotor impairment as an aid to assess parkinson's disease, Neurological Sciences 42 (2021) 2337–2345.
[13] E. Rastegari, S. Azizian, H. Ali, Machine learning and similarity network approaches to support automatic classification of Parkinson's diseases using accelerometer-based gait analysis, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.
[14] L. di Biase, L. Raiano, M. L. Caminiti, P. M. Pecoraro, V. Di Lazzaro, Parkinson's disease wearable gait analysis: Kinematic and dynamic markers for diagnosis, Sensors 22 (22) (2022) 8773.
[15] Q. Wang, W. Zeng, X. Dai, Gait classification for early detection and severity rating of Parkinson's disease based on hybrid signal processing and machine learning methods, Cognitive Neurodynamics (2022) 1–24.
[16] H. Li, W. Ma, C. Li, Q. He, Y. Zhou, A. Xie, Combined diagnosis for parkinson's disease via gait and eye movement disorders, Parkinsonism & Related Disorders 123 (2024) 106979.
[17] J. Archila, A. Manzanera, F. Martinez, A multimodal Parkinson quantification by fusing eye and gait motion patterns, using covariance descriptors, from non-invasive computer vision, Computer Methods and Programs in Biomedicine (2021) 106607.
[18] C. H. Hawkes, K. Del Tredici, H. Braak, A timeline for parkinson's disease, Parkinsonism & related disorders 16 (2) (2010) 79–84.
[19] A. Berardelli, J. C. Rothwell, P. D. Thompson, M. Hallett, Pathophysiology of bradykinesia in parkinson's disease, Brain 124 (11) (2001) 2131–2146.
[20] L. Goh, C. G. Canning, J. Song, L. Clemson, N. E. Allen, The effect of rehabilitation interventions on freezing of gait in people with parkinson's disease is unclear: a systematic review and meta-analyses, Disability and Rehabilitation 45 (19) (2023) 3199–3218.
[21] A. Langer, D. Roth, A. Santer, J. Gruber, L. Wizany, S. Hasenauer, R. Pokan, P. Dabnichki, M. Treven, S. Zimmel, et al., Climb up! head up! climbing improves posture in parkinson's disease. a secondary analysis from a randomized controlled trial, Clinical Rehabilitation 37 (11) (2023) 1492–1500.
[22] J. S. Perlmutter, Assessment of parkinson disease manifestations, Current protocols in neuroscience Chapter 10 (2009) Unit10.1. doi:10.1002/0471142301.ns1001s49.
URL https://europepmc.org/articles/PMC2897716
[23] Y. Yang, P. Liu, Y. Sun, N. Yu, J. Wu, J. Han, A video-based method to classify abnormal gait for remote screening of parkinson's disease, in: 2021 40th Chinese Control Conference (CCC), IEEE, 2021, pp. 3357–3362.
[24] L. Gong, J. Li, M. Yu, M. Zhu, R. Clifford, A novel computer vision based gait analysis technique for normal and parkinson's gaits classification, in: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), IEEE, 2020, pp. 209–215.
[25] S. Rupprechter, G. Morinan, Y. Peng, T. Foltynie, K. Sibley, R. S. Weil, L.-A. Leyland, F. Baig, F. Morgante, R. Gilron, et al., A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease, Sensors 21 (16) (2021) 5437.
[26] 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: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23, Springer, 2020, pp. 637–647.
[27] A. Sabo, S. Mehdizadeh, A. Iaboni, B. Taati, Estimating parkinsonism severity in natural gait videos of older adults with dementia, IEEE journal of biomedical and health informatics 26 (5) (2022) 2288–2298.
[28] Y. Bansal, R. Grover, Ocular disorders in parkinson's disease: A review, Journal of Clinical Ophthalmology and Research 12 (2) (2024) 172–176.
[29] A. Sekar, M. T. Panouillères, D. Kaski, Detecting abnormal eye movements in patients with neurodegenerative diseases–current insights, Eye and Brain (2024) 3–16.
[30] R. Armstrong, Oculo-visual dysfunction in parkinson's disease, Journal of Parkinson's disease 5 (4) (2015) 715–726.
[31] H. Li, X. Zhang, Y. Yang, A. Xie, Abnormal eye movements in parkinson's disease: From experimental study to clinical application, Parkinsonism & Related Disorders (2023) 105791.
[32] Z. Zeng, L. Tao, H. Zhu, Y. Zhu, L. Meng, J. Fan, C. Chen, W. Chen, A robust gaze estimation approach via exploring relevant electrooculogram features and optimal electrodes placements, IEEE Journal of Translational Engineering in Health and Medicine.
[33] O. Bredemeyer, S. Patel, J. J. FitzGerald, C. A. Antoniades, Oculomotor deficits in Parkinson's disease: Increasing sensitivity using multivariate approaches, Frontiers in Digital Health 4.
[34] D. C. Brien, H. C. Riek, R. Yep, J. Huang, B. Coe, C. Areshenkoff, D. Grimes, M. Jog, A. Lang, C. Marras, et al., Classification and staging of parkinson's disease using video-based eye tracking, Parkinsonism & Related Disorders 110 (2023) 105316.
[35] J. Reiner, L. Franken, E. Raveh, I. Rosset, R. Kreitman, E. Ben-Ami, R. Djaldetti, Oculometric measures as a tool for assessment

of clinical symptoms and severity of parkinson's disease, Journal of Neural Transmission 130 (10) (2023) 1241–1248.
[36] N. A. Koch, P. Voss, J. M. Cisneros-Franco, A. Drouin-Picaro, F. Tounkara, S. Ducharme, D. Guitton, É. de Villers-Sidani, Eye movement function captured via an electronic tablet informs on cognition and disease severity in parkinson's disease, Scientific Reports 14 (1) (2024) 9082.
[37] H. N. Pham, T. T. T. Do, K. Y. Jie Chan, G. Sen, A. Y. K. Han, P. Lim, T. S. Loon Cheng, Q. H. Nguyen, B. P. Nguyen, M. C. H. Chua, Multimodal detection of Parkinson disease based on vocal and improved spiral test, in: 2019 International Conference on System Science and Engineering (ICSSE), 2019, pp. 279–284. doi:10.1109/ICSSE.2019.8823309.
[38] J. Prince, F. Andreotti, M. De Vos, Multi-source ensemble learning for the remote prediction of parkinson's disease in the presence of source-wise missing data, IEEE Transactions on Biomedical Engineering 66 (5) (2018) 1402–1411.
[39] K. D. Raj, G. J. Lal, E. Gopalakrishnan, V. Sowmya, J. R. Orozco-Arroyave, A visibility graph approach for multi-stage classification of parkinson's disease using multimodal data, IEEE Access (2024).
[40] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, E. Nöth, Multimodal assessment of Parkinson's disease: a deep learning approach, IEEE journal of biomedical and health informatics 23 (4) (2018) 1618–1630.
[41] S. Caproni, C. Colosimo, Diagnosis and differential diagnosis of parkinson disease, clinics in geriatric medicine 36 (1) (2020) 13–24.
[42] Z. Huang, L. Van Gool, A riemannian network for spd matrix learning, in: Association for the Advancement of Artificial Intelligence (AAAI), 2017.
[43] S. Wan, Y. Liang, Y. Zhang, M. Guizani, Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones, IEEE Access 6 (2018) 36825–36833.
[44] W. S. Lim, S.-I. Chiu, M.-C. Wu, S.-F. Tsai, P.-H. Wang, K.-P. Lin, Y.-M. Chen, P.-L. Peng, Y.-Y. Chen, J.-S. R. Jang, et al., An integrated biometric voice and facial features for early detection of parkinson's disease, npj Parkinson's Disease 8 (1) (2022) 145.
[45] D. M. Herz, P. Brown, Moving, fast and slow: behavioural insights into bradykinesia in parkinson's disease, Brain 146 (9) (2023) 3576–3586.
[46] A. Mirelman, P. Bonato, R. Camicioli, T. D. Ellis, N. Giladi, J. L. Hamilton, C. J. Hass, J. M. Hausdorff, E. Pelosin, Q. J. Almeida, Gait impairments in parkinson's disease, The Lancet Neurology 18 (7) (2019) 697–708.
[47] F. Zhang, J. Shi, Y. Duan, J. Cheng, H. Li, T. Xuan, Y. Lv, P. Wang, H. Li, Clinical features and related factors of freezing of gait in patients with parkinson's disease, Brain and behavior 11 (11) (2021) e2359.
[48] D. Santos García, T. de Deus Fonticoba, C. Cores Bartolomé, L. Naya Ríos, L. García Roca, C. Martínez Miró, H. Canfield, S. Jesús, M. Aguilar, P. Pastor, et al., Predictors of loss of functional independence in parkinson's disease: results from the coppadis cohort at 2-year follow-up and comparison with a control group, Diagnostics 11 (10) (2021) 1801.
[49] Y.-m. Sun, Z.-y. Wang, Y.-y. Liang, C.-w. Hao, C.-h. Shi, Digital biomarkers for precision diagnosis and monitoring in parkinson's disease, NPJ Digital Medicine 7 (1) (2024) 218.
[50] S.-M. Fereshtehnejad, S. R. Romenets, J. B. Anang, V. Latreille, J.-F. Gagnon, R. B. Postuma, New clinical subtypes of parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes, JAMA neurology 72 (8) (2015) 863–873.
[51] B. R. Bloem, J. Marinus, Q. Almeida, L. Dibble, A. Nieuwboer, B. Post, E. Ruzicka, C. Goetz, G. Stebbins, P. Martinez-Martin, et al., M easurement instruments to assess posture, gait, and balance in p arkinson's disease: Critique and recommendations, Movement Disorders 31 (9) (2016) 1342–1355.
[52] D. J. Park, J. W. Lee, M. J. Lee, S. J. Ahn, J. Kim, G. L. Kim, Y. J. Ra, Y. N. Cho, W. B. Jeong, Evaluation for parkinsonian bradykinesia by deep learning modeling of kinematic parameters, Journal of Neural Transmission 128 (2021) 181–189.
[53] L. Sigcha, B. Domínguez, L. Borzì, N. Costa, S. Costa, P. Arezes, J. M. López, G. De Arcas, I. Pavón, Bradykinesia detection in parkinson's disease using smartwatches' inertial sensors and deep learning methods, Electronics 11 (23) (2022) 3879.
[54] L.-Y. Ma, W.-K. Shi, C. Chen, Z. Wang, X.-M. Wang, J.-N. Jin, L. Chen, K. Ren, Z.-L. Chen, Y. Ling, et al., Remote scoring models of rigidity and postural stability of parkinson's disease based on indirect motions and a low-cost rgb algorithm, Frontiers in Aging Neuroscience 15 (2023) 1034376.
[55] G. Vashishtha, S. Chauhan, M. Sehri, J. Hebda-Sobkowicz, R. Zimroz, P. Dumond, R. Kumar, Advancing machine fault diagnosis: A detailed examination of convolutional neural networks, Measurement Science and Technology 36 (2) (2024) 022001.
[56] A. Al Fahoum, A. Zyout, Wavelet transform, reconstructed phase space, and deep learning neural networks for eeg-based schizophrenia detection., International Journal of Neural Systems 34 (9) (2024) 2450046–2450046.