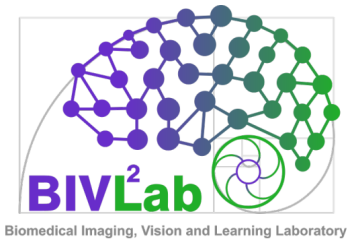

DEEP REPRESENTATIONS IN RIEMANNIAN MANIFOLDS FOR THE QUANTIFICATION OF PARKINSONIAN OCULOMOTOR PATTERNS

JUAN ANDRÉS OLMOS ROJAS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA
BUCARAMANGA

2022

DEEP REPRESENTATIONS IN RIEMANNIAN MANIFOLDS FOR THE QUANTIFICATION OF PARKINSONIAN OCULOMOTOR PATTERNS

JUAN ANDRÉS OLMOS ROJAS

Research work in partial fulfillment of the requirements for the degree of:
Magíster en Matemática Aplicada

Advisor:
Fabio Martínez Carrillo
Ph.D in Systems and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA
BUCARAMANGA
2022

ACKNOWLEDGEMENTS

The author expresses his acknowledgement:

First of all, to Professor Fabio Martinez, director of the BIVL2ab research group. Acknowledgment for his guidance and reception during the years of development of this project. It was an invaluable academic and professional guidance. Also, thank Professors Antoine Manzanera for their collaboration and advice in research works. In general, to the research group for allowing the development of this project. And, of course, to my colleagues for their support and constant feedback.

In the process, personal support from my parents, girlfriend and family was also very important. I thank them for their trust and respect for my master's research process. Especially to my parents for their constant support and messages of confidence to keep moving forward towards better achievements. Also, to my girlfriend for supporting me in the best and most difficult moments. In my family, to my guide and friend Lilia, thank you.

Finally, thanks to "Decanato de Ciencias", "Fisicomecanica", "Movilidad UIS", "Vicerrectoria academica" for supporting my participation in local and international academic activities.

CONTENTS

page

LIST OF FIGURES

page

LIST OF TABLES

page

LIST OF APPENDICES

page

ABSTRACT

TITLE: DEEP REPRESENTATIONS IN RIEMANNIAN MANIFOLDS FOR THE QUANTIFICATION OF PARKINSONIAN OCULOMOTOR PATTERNS [*]

AUTHOR: JUAN ANDRÉS OLMOS ROJAS [**]

KEYWORDS: PARKINSON'S DISEASE, OCULOMOTOR PATTERNS, SPD MATRICES, RIEMANNIAN NON-LINEAR LEARNING, RIEMANNIAN MANIFOLD.

DESCRIPTION: Parkinson's disease (PD) is the second most common neurodegenerative disorder, mainly characterized by motor alterations because of the non-controlled degeneration of dopamine neurotransmitters. Today, there is no definitive biomarker for an early diagnosis and progression characterization. Recently, anomaly oculomotor patterns have shown promising evidence to quantify PD, even at early stages. However, current capture setups require sophisticated protocols, limiting the analysis to coarse measures that poorly exploit alterations and restrict their standard use in clinical environments. Although computational-based deep learning strategies today bring a robust alternative by discovering in video sequences hidden patterns associated to the disease, these approaches depend on large training data volumes to cover the variability. This work introduces a novel strategy that exploits data geometry within a deep Riemannian manifold, quantifying and discovering oculomotor PD patterns. From a fixational oculomotor task, we perform non-invasive analysis using only video sequences. Oculomotor information is encoded as symmetric positive (SPD) matrices that capture second-order statistics from deep representations computed by convolutional schemes. These symmetric matrices then form an embedded representation, which is decoded by a deep Riemannian learning scheme that preserves the SPD geometrical structure and discriminates Parkinsonian patients w.r.t a control population. The proposed strategy was evaluated on eye fixational recordings in a population of 13 Parkinson's patients and 13 controls, achieving an accuracy above 98%. The proposed strategy was able to differentiate the disease at different stages and also demonstrates coherent results from explainable maps propagated from output probabilities.

---

[*] Research work

[**] Facultad de Ciencias. Escuela de Física. Advisor: Fabio Martínez Carrillo, Ph.D.

RESUMEN

TÍTULO: REPRESENTACIONES PROFUNDAS EN VARIEDADES RIEMANNIANAS PARA LA CUANTIFICACIÓN DE PATRONES OCULOMOTORES PARKINSONIANOS [*]

AUTOR: JUAN ANDRÉS OLMOS ROJAS [**]

PALABRAS CLAVE: ENFERMEDAD DE PARKINSON, PATRONES OCULOMOTORES, MATRICES SPD, APRENDIZAJE PROFUNDO NO-LINEAL, VARIEDAD RIEMANNIANA

DESCRIPCIÓN: La enfermedad de Parkinson (PD) es el segundo trastorno neurodegenerativo más común, caracterizado principalmente por alteraciones motoras debido a la degeneración no controlada de los neurotransmisores de la dopamina. En la actualidad, no existe un biomarcador definitivo para el diagnóstico precoz y la caracterización de la progresión. Recientemente, los anomalías de patrones oculomotoras han mostrado evidencias prometedoras para cuantificar la PD, incluso en etapas tempranas. Sin embargo, las configuraciones de captura actuales requieren protocolos sofisticados, limitando el análisis a medidas gruesas que explotan pobremente las alteraciones y restringen su uso estándar en ambientes clínicos. Aunque las estrategias de aprendizaje profundo basadas en computación aportan hoy una alternativa robusta al descubrir en las secuencias de vídeo patrones ocultos asociados a la enfermedad, estos enfoques dependen de grandes volúmenes de datos de entrenamiento para cubrir la variabilidad. Este trabajo introduce una estrategia novedosa que explota la geometría de los datos en una variedad profunda Riemanniana, cuantificando y descubriendo patrones oculomotores de la PD. A partir de una tarea oculomotora de fijación, realizamos un análisis no invasivo utilizando únicamente secuencias de vídeo. La información oculomotora se codifica como matrices simétricas positivas (SPD) que capturan estadísticas de segundo orden a partir de representaciones profundas computadas por esquemas convolucionales. Estas matrices simétricas forman entonces una representación embebida, que es decodificada por un esquema de aprendizaje profundo Riemanniano que preserva la estructura geométrica SPD y discrimina a los pacientes con Parkinson con respecto a una población de control. La estrategia propuesta fue evaluada en grabaciones de fijación ocular en una población de 13 pacientes de Parkinson y 13 controles, logrando una precisión superior al 98%. La estrategia propuesta fue capaz de diferenciar la enfermedad en diferentes etapas y también demuestra resultados coherentes de mapas de explicabilidad propagados a partir de las probabilidades de salida.

---

[*] Trabajo de investigación

[**] Facultad de Ciencias. Escuela de Física. Director: Fabio Martínez Carrillo, Ph.D.

INTRODUCTION

Neurological diseases are currently the major cause of disability across the world. Parkinson's disease (PD) is the second most common neurodegenerative disorder affecting between 2-3% of the population over 65 years of age worldwide [1]. Besides, PD reports a prevalence around 22%, being the neurological disorder of fastest growth world-around [21]. The cause of this disease is partially understood, being the principal neurodegeneration features related to the disruption of the dopamine neurotransmitters that control voluntary movement, producing in consequence alterations in the patient's movement. Currently, this disease has no cure, but the early and personalized treatment planning is fundamental to slow down motor symptoms and disabilities. Nonetheless, today there is no definitive disease biomarker and the diagnosis is commonly subject to observational analysis, reporting errors up to 24% [23].

In the literature have been done multiple efforts to characterize and measure motor disabilities correlated with PD. However, the motion patterns related to tremor in hands, disabilities during gait and trunk rigidity [4], are mostly captured at an advanced stage of the disease. More recently, different studies have experimentally supported the hypothesis of a strong correlation

---

[1] E Ray Dorsey et al. "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016". In: *The Lancet Neurology* 17.11 (2018), pp. 939–953.

[2] Werner Poewe et al. "Parkinson disease". In: *Nature Reviews Disease Primers* 3.1 (Mar. 2017), pp. 1–21.

[3] Eduardo Tolosa et al. "Challenges in the diagnosis of Parkinson's disease". In: *The Lancet Neurology* 20.5 (2021), pp. 385–397.

[4] Rimona S. Weil et al. "Visual dysfunction in Parkinson's disease". In: *Brain: A Journal of Neurology* 139.11 (Nov. 2016), pp. 2827–2843.

of oculomotor patterns with PD, even at early stages [4][5][6]. These patterns, however, are collected from sophisticated capture devices that simplify eye dynamic to global displacement trajectories, making it difficult to address the wide range of disease evolution.

Hence, computational approaches have emerged as an alternative to support the quantification of motor patterns directly from video analysis. More recently, deep learning approaches have revealed determining advantages to discover hidden patterns and to characterize kinematic descriptors in the modeling of gait videos, eye movement disorders, cerebrospinal data and tracking eye movement, among many others [7][8]. These deep strategies however require a huge amount of training information to deal with observation variability and the quantification of proper hidden variables demand deeper architectures to discriminate the data [9][7]. These requirements are rarely realistic in clinical scenarios with large pattern variability, and where annotated examples are difficult to get, above all in the task of discovering motor anomalies associated with PD. To avoid such challenging training scheme, some works have used a collection of convolutional responses extracted from the first layers of a Convolutional Neural Network (CNN) architecture previously trained on a general natural image classification problem [10]. Subsequently, different works propose pooling methods to compact these representations in low dimension descriptors

---

[5] Florence Chan et al. "Deficits in Saccadic Eye-Movement Control in Parkinson's Disease". In: *Neuropsychologia* 43 (2005), pp. 784–796.

[6] George T. Gitchel, Paul A. Wetzel, and Mark S. Baron. "Pervasive Ocular Tremor in Patients With Parkinson Disease". In: *Archives of Neurology* (2012).

[7] Wu Wang et al. "Early detection of Parkinson's disease using deep learning and machine learning". In: *IEEE Access* 8 (2020), pp. 147635–147646.

[8] Luis Carlos Guayacán, Edgar Rangel, and Fabio Martínez. "Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 3688–3691.

[9] Peihua Li et al. "Is second-order information helpful for large-scale visual recognition?" In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2070–2078.

[10] Isail Salazar et al. "A convolutional oculomotor representation to model Parkinsonian fixational patterns from magnified videos". In: *Pattern Analysis and Applications* 24.2 (2021), pp. 445–457.

using Symmetric Positive Definite (SPD) matrices that summarize feature statistics [10][11]. However, these matrices belong to a Riemannian manifold, making it necessary to design proper methods regarding the geometric structure [12][13]. Despite current efforts, learning methods on the manifold still use shallow learning and machine learning algorithms need to be redesigned to take into account data geometry [12].

This work introduces a novel digital biomarker that captures discriminatory PD eye patterns, following a Riemannian deep representation, that compactly codes in symmetric matrices second-order statistics of deep convolutional features. For this purpose, each video sequence is transformed in spatio-temporal slices that recover tiny tremor patterns, during ocular fixation experiment. From this input, an end-to-end learning process is herein proposed through a hybrid deep network, whose first layers are convolutional (CNN) and last layers are Riemannian (SPDNet). The CNN module projects the input slices to deep features, which are then summarized in symmetric positive embedding matrices, allowing to exploit feature correlations that are related to visual observations. Hence, these embedding matrices feed the Riemannian module, that focus on non-linear learning, while preserving geometry of input SPD data, achieving a discrimination between Parkinsonian and control classes. The result is a method able to discriminate between oculomotor patterns of patients diagnosed with the disease and a control population. In a further validation, the proposed approach evidenced sufficient sensitivity to discriminate among the different stages of PD. Moreover, this work introduces an interpretability strategy that recovers attention maps, highlighting important spatiotemporal regions that support PD prediction.

---

[11] Dinesh Acharya et al. "Covariance pooling for facial expression recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 367–374.

[12] Zhiwu Huang and Luc Van Gool. "A Riemannian network for SPD matrix learning". In: *Thirty-First AAAI Conference on Artificial Intelligence* 31.1 (2017).

[13] Michael M. Bronstein et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418.

1. FUNDAMENTALS AND PREVIOUS WORK

1.1. Parkinsonian oculomotor analysis

The study of PD progression is highly correlated with the reduction in dopamine concentration and related cells [2]. This neurodegenerative process affects motor functions and motor skills [7][14]. Therefore, an early detection and proper characterization of the disease are fundamental to planning treatments and control the rapid degeneration of the disease. Eye motor-based patterns have been recently presented as a potential PD biomarker with high sensitivity for an early detection of the disease [15][16]. Actually, some works evidence an existing relation between neurodegenerative processes and eye locomotion alterations [17][16][18]. Moreover, some studies showed prevalence of ocular instability on almost every Parkinson patient [6][19].

Particularly, these oculomotor abnormalities are related with slowness of movement (bradykinesia) of ocular pursuit, impaired gaze, abnormalities in rapid movements, disturbed smooth

---

[14] Joseph Jankovic. "Parkinson's disease: clinical features and diagnosis". In: *Journal of neurology, neurosurgery & psychiatry* 79.4 (2008), pp. 368–376.

[15] CA Antoniades and C Kennard. "Ocular motor abnormalities in neurodegenerative disorders". In: *Eye* 29.2 (2015), pp. 200–207.

[16] Michael R MacAskill and Tim J Anderson. "Eye movements in neurodegenerative diseases". In: *Current opinion in neurology* 29.1 (2016), pp. 61–68.

[17] Eric R Kandel et al. "The Control of Gaze". In: *Principles of Neural Science, Fifth Edition*. McGraw-Hill Education, 2014, pp. 894–916.

[18] Merel S Ekker et al. "Ocular and visual disorders in Parkinson's disease: common but frequently overlooked". In: *Parkinsonism & related disorders* 40 (2017), pp. 1–10.

[19] George T Gitchel et al. "Experimental support that ocular tremor in Parkinson's disease does not originate from head movement". In: *Parkinsonism & related disorders* 20.7 (2014), pp. 743–747.

ocular pursuit movements, and ocular tremor [20][18][21]. Nonetheless, today there exist controversy around the hypothesis of a vestibule-ocular reflex in the eyes caused by the head oscillation and the interference on the natural motion of the eyes [22]. Some recent works evidenced reliability of the eye as clinical biomarker, but major analysis are demanded to clarify the source of such patterns [23][19]. The most common oculomotor movements where Parkinson is persistent are illustrated in Figure 1 and described thereafter.

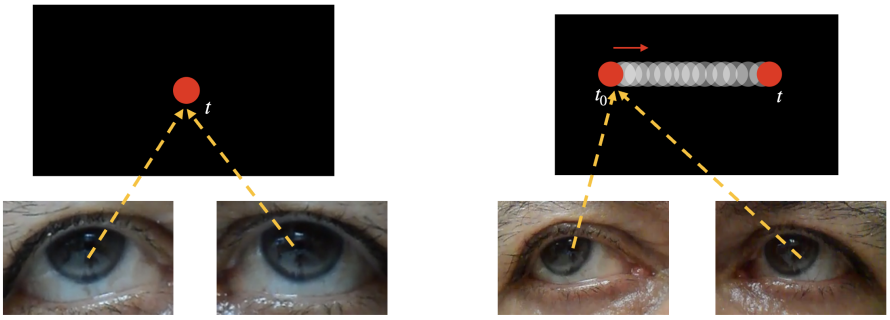

Figure 1. On the left, diagram representing the oculomotor fixation task, where subjects maintain the gaze on a fixed stimulus. On the right, diagram of the smooth pursuit task, where the participant follows a dot stimulus that moves in one direction.

1.1.1. Ocular fixation During the fixation task, participants are invited to observe a fixed stimulus in front of them (as illustrated in Figure 1). This oculomotor behavior captures tiny motions that appears during gaze fixation on a stationary mark of interest [23]. During fixation, there

---

[20] Martin Gorges, Hans-Peter Müller, and et al. "The association between alterations of eye movement control and cerebral intrinsic functional connectivity in Parkinson's disease". In: *Brain imaging and behavior* 10.1 (2016), pp. 79–91.

[21] Pierpaolo Turcano et al. "Early ophthalmologic features of Parkinson's disease: a review of preceding clinical and diagnostic markers". In: *Journal of neurology* 266.9 (2019), pp. 2103–2111.

[22] Diego Kaski et al. "Ocular tremor in Parkinson's disease is due to head oscillation". In: *Movement Disorders* 28.4 (2013), pp. 534–537.

[23] Tabish A Saifee et al. "Tremor of the eyes in Parkinson's disease: Merely a measure of the head movement". In: *Parkinsonism & related disorders* 20.12 (2014), pp. 1447–1448.

are different continuous movements: "eye drift", "microsaccades", and "microtremors". Ocular drift is a slow meandering movement that produces larger shifts in eye position. Microsaccades are largest fixational movements and comprise fast movements with frequencies of 1-2Hz and amplitude $< 0.2°$. Finally, micro tremors have been characterized as oscillatory movements of high frequency (40-100Hz) and little amplitude ($\sim 0.002°$). Parkinson abnormalities are related with the combination of these three movements that produces ocular instability during fixation, generating a dynamic process of gazing movements [24]. For this reason, ocular fixation has been used to discover abnormal patterns related to PD that potentially constitute a biomarker. In addition, some works have evidenced reduction of ocular movement width and fundamental frequency between 4 and 7Hz for parkinsonian population [2564]. In a more recent study, Gitchel et al. showed the presence of micro tremors in a group of 120 people with Parkinson compared to the fixation's stability of 60 healthy people [6]. In this work, we are dedicated to characterize ocular fixational patterns to find correspondences with Parkinson disease and its progression.

1.1.2. Ocular smooth pursuit This oculomotor pattern is recovered from the continuous eye movement, following the smooth movement of a stimulus in a vertical or horizontal direction (as observed in Figure 1). During these experiments, PD-related disorders include saccadic interruptions and a low velocity compared with the target velocity [26]. Thus, the sensitivity to follow the 'target object' decreases during the pursuit. Since neural pathways cause control system abnormalities, oculomotor changes are a sensitive neurological biomarker that may track the disease progression from the beginning [24].

---

[24] Richard J Krauzlis, Laurent Goffart, and Ziad M Hafed. "Neuronal control of fixation and fixational eye movements". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1718 (2017).

[25] Christian Duval and Anne Beuter. "Fluctuations in Tremor at Rest and Eye Movements during Ocular Fixation in Subjects with Parkinson's Disease". In: *Parkinsonism & Related Disorders* 4 (1998), pp. 91–97.

[26] Kikuro Fukushima et al. "Impaired smooth-pursuit in Parkinson's disease: normal cue-information memory, but dysfunction of extra-retinal mechanisms for pursuit preparation and execution". In: *Physiological reports* 3.3 (2015).

1.2. Learning from SPD matrices

1.2.1. SPD matrix descriptors In computer vision, learning representation and descriptors become large sets of high dimensional vector representations [27]. Various approaches have emerged to generate more compact data from vectors, allowing reliable estimations and dealing, for instance, with online predictions. An example is the covariance matrices that encode the existing correlations between the vectors in a lower-dimensional matrix. From a set of vectors $X = \{x_i\}_{i=1}^{m}$, the covariance matrix is an SPD matrix that encodes pairwise covariance between these vectors. In this way, it saves second-order statistical relationships, resulting in robustness to noise and allowing the coding of different features or descriptions of the data [27]. These representations have been successfully applied for diverse domains and applications, such as, face recognition [28], fast detection and classification tasks [29], matching descriptors [3031], and encoding deep learning outputs [28]. It should be noted that these matrices are symmetric positive definite (SPD) and requires a special treatment due to their geometric structure.

1.2.2. Underlying geometry of SPD matrices Symmetric positive definite (SPD) matrices (denoted also as S_{++}^{N}) belongs to a curved (non-flatten) space Riemannian manifold. Particularly, a property of these matrices is that all their eigenvalues are strictly positive. For

---

[27] Hà Quang Minh and Vittorio Murino. "Covariances in computer vision and machine learning". In: *Synthesis Lectures on Computer Vision* 7.4 (2017), pp. 1–170.

[28] Naima Otberdout et al. "Deep covariance descriptors for facial expression recognition". In: *arXiv preprint arXiv:1805.03869* (2018).

[29] Oncel Tuzel, Fatih Porikli, and et al. "Region Covariance: A Fast Descriptor for Detection and Classification". In: *ECCV*. 2006, pp. 589–600.

[30] Hedi Tabia et al. "Covariance Descriptors for 3D Shape Matching and Retrieval". In: *Proceedings of ICCVPR*. 2014.

[31] Yinghao Cai, Valtteri Takala, and Matti Pietikainen. "Matching groups of people by covariance descriptor". In: *20th ICPR*. IEEE. 2010, pp. 2744–2747.

instance, the space of SPD matrices of dimension 2×2 given by:

$$S^2_{++} = \left\{ \begin{bmatrix} x & y \\ y & z \end{bmatrix} : xz - y^2 \geq 0, x \geq 0, z \geq 0 \right\},$$

can be illustrated in $\mathbb{R}^3$ as in Figure 2. From here we can observe that geometrically, SPD matrices constitute a curved convex cone as illustrated. Therefore, to manipulate data in this space, have been introduced different Riemannian metrics because the Euclidean primitives lead to problems of sub-optimality and propagation of numerical errors [27]. Since S^N_{++} is a geodesically complete Riemannian manifold with non-positive curvature (see [2732]), any two points on S^N_{++} can be always joined with a unique geodesic, the shortest path on this space. Therefore, usual operations can be defined and result stable in this space [33]. For instance, Pennec in [33], introduces a computing framework on this manifold using an *affine-invariant* Riemannian metric and a set of basic statistical tools on the manifold.

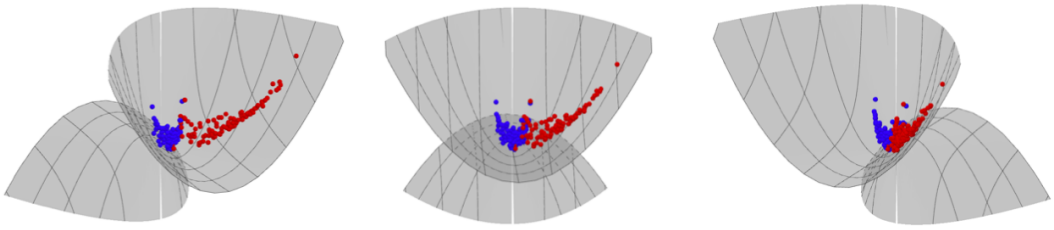

Figure 2. Representation of S^2_{++} with synthetic SPD data.

An important related structure to the manifold is the tangent space, denoted by $T_P S^N_{++}$. It is an Euclidean space tangent to S^N_{++} in the point P. In the light of develop a computing framework for SPD matrices, the tangent space has been used to aggregate a vector structure to the manifold and to calculate statistic measures on this space, e.g., has been used to compute the Riemannian mean of SPD matrices [33]. Considering that, given $P \in S^N_{++}$ there are two important

---

[32] Serge Lang. *Fundamentals of Differential Geometry*. Vol. 191. Springer, 1999.

[33] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. "A Riemannian framework for tensor computing". In: *International Journal of computer vision* 66.1 (2006), pp. 41–66.

maps, the exponential $\exp_P$ and their inverse, the logarithm map $\log_P$, which maps vectors from the tangent space $T_P S^N_{++}$ to S^n_{++} and elements from S^n_{++} to $T_P S^n_{++}$ respectively [32]. See Figure 3 for an illustration of these maps.

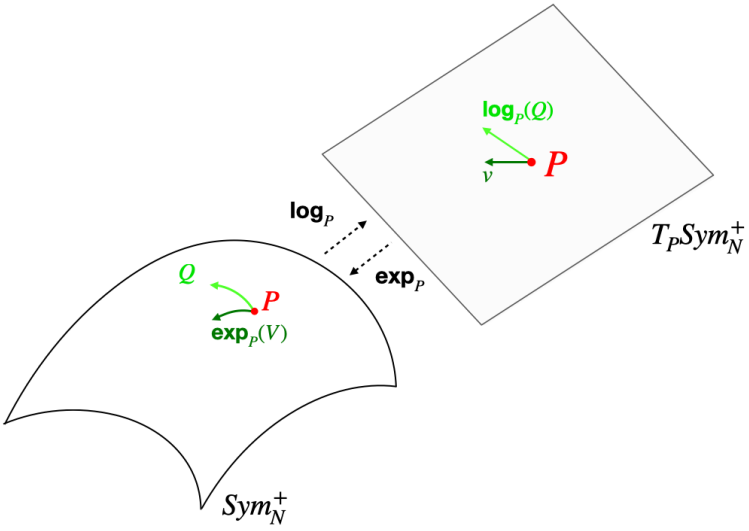

Figure 3. Riemannian maps to operate in the space of Riemannian manifold of SPD matrices. Exponential ($\exp_P$) and logarithm ($\log_P$) map elements from the manifold into the Euclidean tangent space and from the tangent space into the Riemannian manifold, respectively.

There exist different metrics that approximate distances between points at Riemmanian space. Here, we adopt the log-Euclidean metric to build a vectorial-space structure for the manifold, mapping the SPD data to the Euclidean tangent space [34]. From this, was defined the log-euclidean distance (d_{le}) as

$$d_{le}(C_i, C_j) = ||\log(C_i) - \log(C_j)||,$$

where log is the algorithm map in the identity matrix and $||\cdot||$ is the usual Frobenius norm. This metric is cost effective compared with other approximations and the most common metric

[34] Vincent Arsigny, Pierre Fillard, and et al. "Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices". In: *SIAM Journal on Matrix Analysis and Applications* (2007), pp. 328–347.

adopted in strategies to generate efficient algorithms in the manifold of SPD matrices. [35].

1.2.3. Learning algorithms Machine learning methods are algorithms capable of learning from a data set labeled $\mathbb{D} = \{x, y\}_{i=1}^{N}$. Where y is namely a categorical variable for classification problems and continues for regressions. Usually, the inputs are high-dimensional vectors ($x \in \mathbb{R}^n$). To learn, the model minimize a loss $L(y, \hat{y})$ from the output $\hat{y}$ generated by the parameters θ of the model [3637]. Nowadays, the most common methodology, in both learning schemes, are the deep neural networks which provide a series of non-linear units (neurons), grouped at different levels of processing (layers). Non-linear transition maps $\sigma(\cdot)$ are designed to transfer data layer by layer using linear combinations. Finally, the estimation of the model is established as a composition of these transformations.

$$f(x, \theta) = \sigma_L(W_{L-1} * \sigma_{L-1}(W_{L-1} * \cdots \sigma_1(W_1 * x + b)) + b).$$

There are multiple operators $*$ to compute data transformation. The most common are lineal combination for dense neural networks (DNN) and convolution-like operators for convolutional neural networks (CNN). For classic networks, θ is a set of Euclidean matrices and vectors ($\subset \mathbb{R}^n$), where the gradient descent algorithm is used to adjust the optimum value of representation. Briefly, this algorithm updates θ recursively, minimizing $L(y, \hat{y})$, which is an error measure or a probability. As a result, we have a model that discriminates from the learned θ and that depends on:

- Input data $\mathbb{D} = \{x, y\}_{i=1}^{N}$.

---

[35] Ruiping Wang et al. "Covariance discriminative learning: A natural and efficient approach to image set classification". In: *2012 ICCVPR*. IEEE. 2012, pp. 2496–2503.

[36] Ian Goodfellow, Yoshua Bengio, and et al. *Deep learning*. 2016.

[37] Geert Litjens, Thijs Kooi, and et al. "A Survey on Deep Learning in Medical Image Analysis". In: *Medical Image Analysis* (2017), pp. 60–88.

- The intermediate transformation functions of the neural network $\sigma(\cdot)$.
- Non-lineal operation of transformations $*$.
- The error or distance defined for the prediction $L(y, \hat{y})$.

To work with SPD descriptors, new transformation functions must be considered to generate this type of SPD structures in intermediate layers of the models to measure relationships between deep semantic representations. This would yield a new learning structure that can quantify the PD from its wide range of motor alterations. However, computing a non-Euclidean SPD data structure carries the challenge of designing appropriate learning methods to take advantage of the underlying geometry and statistical information of the SPD matrices [12].

1.3. Quantification of oculomotor alterations

Ocular movement properties, such as speed, acceleration, amplitude, and frequency are relevant measurements to the associated diagnosis of altered locomotion patterns. Quantifying these movement alterations is a challenging task even for expert physicians. Today, there are different tools to support such oculomotor quantification, allowing support diagnosis and analysis of different movement variables. However, most of the existing methods capture ocular patterns from sophisticated and intrusive devices, that simplify the eye dynamics to displacement trajectories, which in turn limits the understanding of the disease progression [19]. In literature, the most accurate assessing methods for eye movement are electro-oculography, scleral search coil system [6], and video-oculography (VOG), with the latter (VOG) being the most widely used in recent years. Nonetheless, the needed equipment covers the entire eye region, is invasive, expensive, and requires calibration and technical configuration procedures. In consequence, researchers have proposed oculomotor examination methods that rely primarily on common video

sequences [38]. For example, there are video-based methods to quantify oculomotor alterations that use algorithms for data exploration and classification support [39].

Nowadays, computer-aided diagnosis systems have integrated machine learning mechanisms, allowing, among others, the analysis of motor information to guide PD classification. In fact, deep learning strategies have revealed advantages to quantify PD patterns even for early diagnosis using eye movement [78]. However, these strategies require a huge amount of labeled data to properly learn discriminative patterns, which is a limitation in the medical domain. In consequence, recent works propose to take magnified eye fixation sequences and represent them as a collection of convolutional responses (deepfeatures) projected by the initial layers of CNN architectures. Then, in order to obtain a low dimensional and robust descriptor, deepfeatures are compacted and summarized using covariance descriptors (SPD matrices). To classify them, they are mapped to the tangent space, and classified using machine learning methods [10]. However, these approach is dependent on the quality of deepfeatures and the lower dimensional analysis to properly codify SPD representations. To overcome such issues, the study of the SPD manifold geometry encourages the design of new learning methods in the manifold. For instance, there have been proposed novel algorithms such as online dictionary learning algorithms to find a representation basis for a set of SPD matrices [40]. Similarity functions have also been introduced to learn a discriminative mapping of SPD matrices to the tangent plane [3541]. More recently, *Huang et al.* [12] have proposed a Riemannian network for SPD matrix deep non-linear

---

[38] Hsin-Yu Lai et al. "Enabling saccade latency measurements with consumer-grade cameras". In: *25th ICIP*. 2018, pp. 3169–3173.

[39] Artur Szymański et al. "Building Intelligent Classifiers for Doctor-Independent Parkinson's Disease Treatments". In: *Information Technologies in Medicine*. Advances in Intelligent Systems and Computing. 2016, pp. 267–276.

[40] Shengping Zhang et al. "Online dictionary learning on symmetric positive definite manifolds with vision applications". In: *29th AAAI*. 2015.

[41] Sadeep Jayasumana et al. "Kernel methods on the Riemannian manifold of symmetric positive definite matrices". In: *Proceedings of ICCVPR*. 2013, pp. 73–80.

learning, the SPDnet. To carry out the learning, the SPDnet introduces the *BiMap*, *ReEig* and *LogEig* transformation layer. These layers forward SPD matrices through the network preserving their geometric structure and calculating more compact and discriminate matrices. Other works propose adding structures to the network, such as SPD pooling layers [42], covariance pooling layers [43], Riemannian batch normalization [44]. These learning SPD representations may be key to discover new PD oculomotor patterns recorded in videos by capturing the geometry of data with a limited amount of information.

---

[42] Rui Wang, Xiao-Jun Wu, and Josef Kittler. "SymNet: A Simple Symmetric Positive Definite Manifold Deep Learning Method for Image Set Classification". In: *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[43] Kaicheng Yu and Mathieu Salzmann. "Second-order convolutional neural networks". In: *arXiv preprint arXiv:1703.06817* (2017).

[44] Daniel Brooks et al. "Riemannian batch normalization for SPD neural networks". In: *arXiv preprint arXiv:1909.02414* (2019).

2. Research Problem

Several studies have shown that oculomotor abnormalities are a potential biomarker of the Parkinson disease even for early diagnosis. However, current quantification protocols are based on capture devices that simplify eye dynamics to global trajectories. Computer-aided diagnosis methods based on Deep learning techniques have demonstrated outstanding performance in quantifying challenging patterns directly from video sequences. However, these strategies require deeper architectures and a large training dataset to deal with observation variability. The study of intermediate representations that compact deep representations in low dimensional descriptors may alleviate such restrictions, allowing to capture discriminative oculomotor patterns. For instance, SPD matrices allow to summarize high dimensional observations using statistic measures, which may be useful to understand spatio-temporal patterns from relatively few samples. However, these matrices are in a Riemmanian space and their modeling from deep learning strategies requires a proper geometric formulation.

Research Question

How much does SPD learning representations contribute to the characterization of Parkinson's disease-related oculomotor alterations?

3. OBJECTIVES

General Objective

- To propose a Riemannian deep network architecture for the quantification of Parkinsonian oculomotor patterns using symmetric positive definite matrices (SPD).

Specific Objectives

- To develop a Riemannian SPD network that learn from SPD matrices preserving geometrical data structure.
- To design convolutional layers that map Euclidean input activations into SPD matrices.
- To define an end-to-end scheme to learn oculomotor patterns through a convolutional-SPD Riemannian network.
- To validate the proposed architecture regarding the discrimination of Parkinsonian oculomotor patterns with respect to a control population.

4. DATASET

A retrospective study was conducted with 13 patients diagnosed with PD (average age of 72.3 ± 7.4) and 13 control subjects (average age of 72.2 ± 6.1). To include inter-subject variability, the study incorporates PD subjects with different disease degree progression. The modified Hoehn-Yahr rating scale was used to categorize PD patients with the aid of a physical therapist. A total of five patients were categorized in stage 2.5, six patients in stage 3, and two in stage 4.
To record oculomotor patterns, the participants were invited to observe a fixed stimulus in front of them (distance between the monitor and the subject was 1.5m). The monitor display was positioned at the same height as the eyes to avoid extra effort from subjects. From 1m ahead and at a lower altitude, we recorded the upper face of the participants, see Figure 4.A) for illustration. Then, over the monitor was projected a white spot circle. Participants received the instruction to fix their gaze on the white spot during the recording. We used a standard optical camera with a temporal resolution of 60 fps to record the eye behavior. The recovered sequences were cropped in video samples of individual eyes for each patient, where each one constitutes a fixation period of five seconds, and the spatial resolution was cropped to 210×140 pixels, centering the first frame to the center of the pupil.
The dataset was approved by an Ethic Committee. Written informed consent was obtained for every participant. To our knowledge, there is no public dataset containing more eye movement videos for Parkinson stage prediction.

Video slicing computation. To recover tremor observations from each video recording, spatio-temporal information was 'coded as video slices (see Figure 4.B)). For this purpose, each video is considered a volume $\{I(x, y, t)\}_{x=1,y=1,t=1}^{W,H,N}$, with spatial and temporal dimension $W \times H$ and N (number of frames) respectively. To obtain the slices, we choose four radial directions $\theta \in \{0°, 45°, 90°, 135°\}$ around the center and cut the video volume along each direction (see Figure 4). This way, each 2d slice image $S_\theta(x, t)$ records temporal variations along the θ direc-

tion. The resulting image slices record subtle eye displacements, capturing potential differential oculomotor patterns related to the disease.

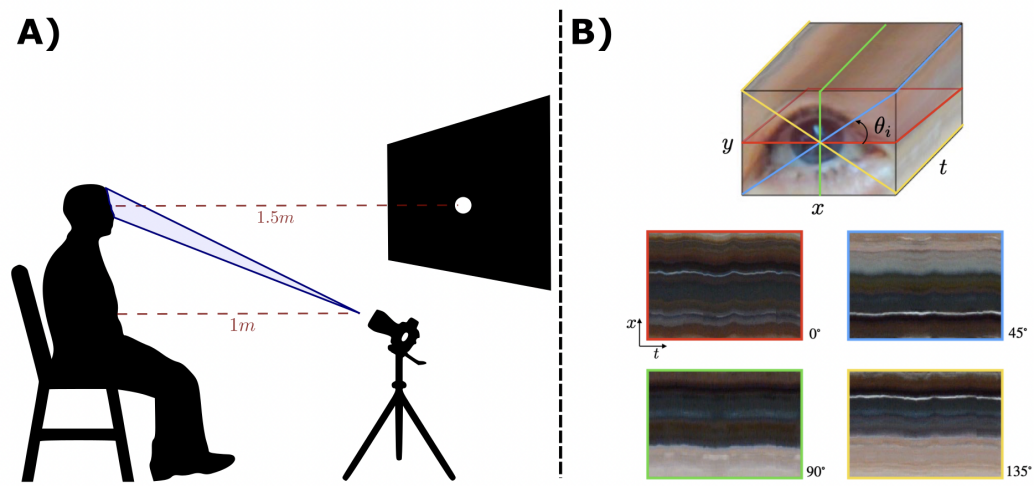

Figure 4. A). Recording set up of the oculomotor fixation task. B) Video slices representation of eye movements from an ocular fixation task.

Extra dataset. To evaluate the generalization of the proposed approach, an *extra set of videos* were captured using similar semi-controlled conditions but from other institutions. We used a chin rest to improve head stabilization in an updated protocol. A total of 6 PD patients (average age of 70.16 ± 6.8) were recorded. This dataset has undergone the same preprocessing as the previous sequences and then slice samples were fed to the trained architecture, only for test purposes. This allows evaluation of the proposed method's generalization under different conditions, such as illumination changes and the recording angle. The entire dataset, including the test data, was approved by an Ethic Committee, and each participant filled out written informed consent. To our knowledge, no public dataset contains more eye movement videos for Parkinson's prediction.

5. PROPOSED APPROACH

We hypothesize that symmetric positive embedding vectors are key to discover new PD digital biomarkers from deep oculomotor fixational patterns. In this work, we designed an end-to-end Riemannian deep strategy (ConvSPD network) that uses spatio-temporal video slices observations of a fixational experiment. Each slice is overcompletely represented by convolutional deep features, which in turn are summarized in symmetric embedding vectors and finally exploited by a Riemannian module to carry out the classification task. Figure 5 summarizes the proposed pipeline. The complete content of this section has been published in the "International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI)" [45] and corresponds to an extra work developed because of an invitation to extend this work to the "Pattern recognition letters" journal.

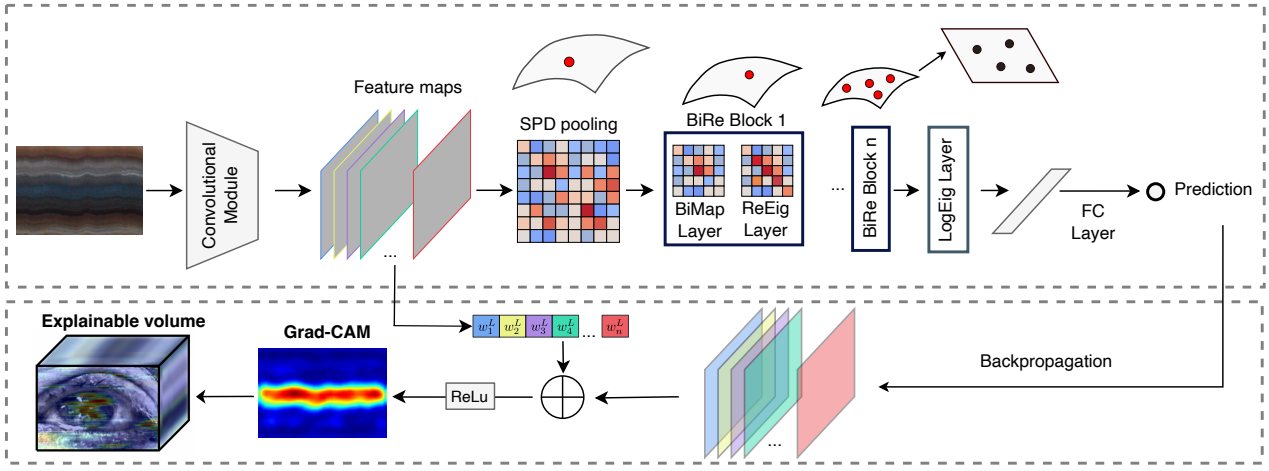

Figure 5. Pipeline and network architecture. Top: the architecture of the proposed Convolutional / SPD (ConvSPD) model. Bottom: interpretation of the model using GradCAM maps from the convolutional module and an explainable video reconstruction.

---

[45] Juan Olmos, Antoine Manzanera, and Fabio Martínez. "An Oculomotor Digital Parkinson Biomarker from a Deep Riemannian Representation". In: *International Conference on Pattern Recognition and Artificial Intelligence*. Springer. 2022, pp. 677–687.

5.1. Fixational patterns from Convolutional Riemannian representations

5.1.1. Convolutional Module and Symmetric Pooling Representation. A convolutional scheme is here introduced as the first part of the proposed approach to represent slices of oculomotor sequences. More precisely, this CNN module learns to extract early to mid-level (textural) features, to capture relevant patterns in eye micro-movements during fixational experiment. The convolutional representation is structured in several layers, hierarchically organized to progressively increase the time $\times$ space receptive field, as well as the semantic level w.r.t. Parkinsonian classification.

In the deepest layer, the corresponding D feature maps form a tensor $X \in \mathbb{R}^{D \times W \times H}$, where (W, H) are the dimensions of the feature maps, resulting from the successive convolutional and pooling processes along the representation. This embedding tensor X, is then summarized in a special symmetric positive embedding matrix by computing second-order statistics from such description [1146]. To this end, we implement a special pooling layer (*SPDpool*) that summarizes the information from the last layer into a symmetric positive embedding matrix (SPD matrix). In general SPD matrices aim to compute second order statistics to recover similarities among features. For instance, by taking the inner product of all pairs of feature vectors, some works propose correlation volumes to compute visual similarities between pixels [47]. In this work, if we suppose that $(k-1)$ is the index of the previous layer, i.e, the last convolutional layer, we re-organize the embedding tensor X in a matrix named X_{k-1} with D rows and $W \times H$ columns. Then we compute the outer product of X_{k-1} with itself. Then, the proposed SPDpool layer calculates: $X_k = f_{SPDpool}(X_{k-1}) = \frac{1}{W \times H} X_{k-1} X_{k-1}^T$. This way, the (i, j) element of X_k is the inner product (correlation) between the i-th feature map and the j-th feature map.

---

[46] Qilong Wang et al. "Deep CNNs meet global covariance pooling: Better representation and generalization". In: *IEEE transactions on pattern analysis and machine intelligence* (2020).

[47] Zachary Teed and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow". In: *European conference on computer vision*. Springer. 2020, pp. 402–419.

This provides second order statistics, and has been used to measure statistical discrepancy [48]. Here, the resultant symmetric positive embedding $D \times D$ matrix allows to capture the most statistically relevant relationships in the previous CNN module, with respect to the Parkinson classification task.

5.1.2. Riemannian Module Structure In the same way as the SPD Gram matrix was used in [49] and [50] to model texture and pictorial style respectively, we may consider the Parkinsonian stage of the patient as a stationary parameter, in the sense that it will affect the ocumolotor action independently on the temporal position (phase) of the movement. In short, the oculomotor task (fixation or tracking) represents the "content" (layout) of the input slices, whereas the stage of the disease represents their "style". However, unlike Gatys et. al. [50], who use Gram matrices in their loss function, we choose to go on working with stationary features that may be interpreted in terms of distributions, by completing the CNN module by a Riemannian SPD network.

Then, once obtained the SPD embedding, we maintain the deep representation while taking into account the Riemmanian geometry of SPD matrices. We then base the next processing layers on the SPDNet framework, that carries out a deep non-linear representation [12]. This requires the codification of special layers, such as: *BiMap*, *ReEig*, and *LogEig*. We now describe the respective Riemmanian layers and their particular learning procedure.

- *The BiMap layer* is a Riemannian layer designed to generate a new bank of more compact

---

[48] Yanghao Li et al. "Demystifying neural style transfer". In: *arXiv preprint arXiv:1701.01036* (2017).

[49] Leon Gatys, Alexander S Ecker, and Matthias Bethge. "Texture Synthesis Using Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015.

[50] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. "Image Style Transfer Using Convolutional Neural Networks". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2414–2423.

and discriminative SPD matrices using a bilinear mapping

$$X_k = W_k X_{k-1} W_k^T. \tag{1}$$

Here, $X_{k-1} \in S_{++}^{d_{k-1}}$ is the input $d_{k-1} \times d_{k-1}$ SPD matrix of the previous layer $(k-1)$, and $W_k \in \mathbb{R}_*^{d_k \times d_{k-1}}$ is the transformation matrix (connection weight), and $X_k \in S_{++}^{d_k}$ is the output SPD matrix. This layer is used after apply the *SPDpool* layer. Similarly to the CNN, the sizes of the SPD matrices decrease from layer to layer, i.e. $d_k < d_{k-1}$. In addition, the connection weight should be an orthogonal full-rank matrix to generate a consistent X_k SPD matrix. That implies these weights matrices lie in a compact *Stiefel manifold*, denoted as $St(d_k, d_{k-1}) = \{W \in \mathbb{R}^{d_{k-1} \times d_k} | WW^T = \mathbb{I}_{d_{k-1}}\}$ [12]. An illustration of the Stiefel manifold is illustrated in Figure 6. Additionally, as will be introduced in Section 5.1.3 this constrain aid the learning process to achieve optimal solutions on $St(d_k, d_{k-1})$ and avoid data degeneration problems [12].

- *The ReEig layer* was inspired by rectified linear units (ReLU) of CNNs. It is composed of a non-linear function to improve the training process by rectifying the SPD matrices:

$$X_k = U_{k-1} \max(\varepsilon I, \Sigma_{k-1}) U_{k-1}^T,$$

where U_{k-1} and Σ_{k-1} come from the eigen decomposition $X_{k-1} = U_{k-1}\Sigma_{k-1}U_{k-1}^T$. Here, $\varepsilon \in \mathbb{R}_+^*$ is a rectification threshold, I is the identity matrix and Σ_{k-1} the diagonal matrix of the eigenvalues of X_{k-1}. This operation tunes up the eigenvalues avoiding non-positiveness and improving the discriminative performance. The BiMap and ReEig layers constitute the main layers of the SPD network, and similar to the ReLU activation function in Conv networks, the BiMap is followed by a ReEig. We then refer as *BiRe block* to the concatenation of these layers.

- *The LogEig layer* results from the necessity to project SPD matrices back to Euclidean

space where the classifiers are designed. In Riemannian manifolds we can attach each point to a flat tangent space with a vector space structure. This structure ease classic Euclidean computations. To map elements here is used the Riemannian logarithm map. To facilitate the computing and work on the same tangent space, we map the SPD matrix X_{k-1} onto the tangent space at the identity by

$$X_k = \log(X_{k-1}) = U_{k-1} \log(\Sigma_{k-1}) U_{k-1}^T.$$

In this way, the output matrix X_k is an Euclidean squared matrix that facilitates the computation of the following output layers [12]. These layers correspond to the Euclidean output layers: Flatten layer, a fully connected layer, and the final output layer is a softmax operation.

5.1.3. Learning Scheme Once the layers are defined, the output of the proposed ConvSPD network can be seen as a composition of functions $f = f^{(l)} \circ f^{(l-1)} \circ \cdots \circ f^{(1)}$, where each function $f^{(k)}$ applies a transformation to the previous data (X_{k-1}), that is $X_k = f^{(k)}(X_{k-1})$. To learn the parameters of these layers (convolutional, and Riemmanian) a backpropagation (BP) algorithm is computed. This strategy uses the chain rule to compute the gradients such as

$$\frac{\partial L^{(k)}}{\partial \Theta_k} = \frac{\partial L^{(k+1)}}{\partial X_k} \frac{\partial f^{(k)}}{\partial \Theta_k}, \tag{2}$$

where Θ_k represents the parameters (or weights) of the k-th layer. The usual procedure to update these parameters in order to achieve a local minimum is the gradient descent algorithm

$$\Theta_k = \Theta_k - \alpha \frac{\partial L^{(k)}}{\partial \Theta_k}, \tag{3}$$

where α denotes the learning rate. Nowadays, there is a wide range of optimization strategies with advantages in terms of computational cost and numerical stability. In this work, an stochastic optimization algorithm using adaptive moment estimation (*Adam* algorithm) is im-

plemented for the non-Riemannian layers. However, two issues arise using euclidean gradients and traditional BP within the Riemannian layers. The first issue is that the eigen decomposition of Eig layers (*ReEig* and *LogEig*) has not been well-solved by the traditional BP. And the second consist of the update the connection weights in the *BiMap* layers cannot yield valid orthogonal matrices [12].

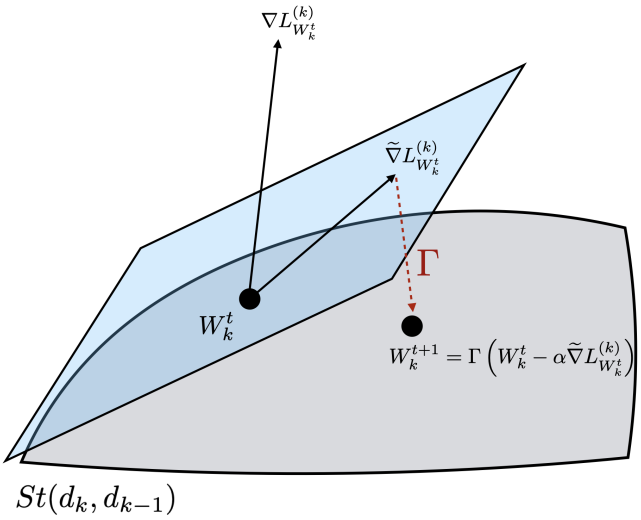

Figure 6. Connection weights of the Riemannian *BiMap* layer are in a *Stiefel* manifold $St(d_k, d_{k-1})$. Therefore to update them is necessary to project the usual Euclidean gradient $\nabla L^{(k)}_{W^t_k}$ into the tangent component of $St(d_k, d_{k-1})$. Then, the optimization step is taken along this direction and the result is mapped back to the manifold using the retraction operation Γ. The result is an updated weight $W^{t+1}_k \in St(d_k, d_{k-1})$ that allows a proper processing of the SPD data.

To address the first problem, structured derivatives are necessary to compute the gradients in the ReEig and LogEig layers [51]. Because these layers do not have parameters, we are interested in compute $\frac{\partial L^{(k)}}{\partial X_{k-1}}$. For that, a virtual layer *k'* is introduced after of the Eig layer. And now the gradient in this layer is given by

---

[51] Catalin Ionescu, Orestis Vantzos, and Cristian Sminchisescu. "Matrix backpropagation for deep networks with structured layers". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 2965–2973.

$$\frac{\partial L^{(k)}}{\partial X_{k-1}} = 2U_{k-1}\left(P^T \circ \left(U_{k-1}^T \frac{\partial L^{(k')}}{\partial U_{k-1}}\right)_{sym}\right) U_{k-1}^T + U_{k-1}\left(\frac{\partial L^{(k')}}{\partial \Sigma_{k-1}}\right)_{diag} U_{k-1}^T,$$

where $\circ$ is the Hadamard product, $A_{sym} = \frac{1}{2}\left(A + A^T\right)$, A_{diag} is a diagonal matrix with the diagonal elements of A and

$$P(i,j) = \begin{cases} \frac{1}{\lambda_i - \lambda_j}, & i \neq j. \\ 0, & i = j \end{cases},$$

where λ_i are the eigenvalues of Σ_{k-1}. Now, the gradients of the virtual layer $\frac{\partial L^{(k')}}{\partial U_{k-1}}$ and $\frac{\partial L^{(k')}}{\partial \Sigma_{k-1}}$ depend of the Eig layer. In the *ReEig* layer, they are computed as

$$\frac{\partial L^{(k')}}{\partial U_{k-1}} = 2\left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym} U_{k-1} \max\left(\epsilon I, \Sigma_{k-1}\right),$$

$$\frac{\partial L^{(k')}}{\partial \Sigma_{k-1}} = Q U_{k-1}^T \left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym} U_{k-1},$$

where Q is the diagonal matrix

$$Q(i,i) = \begin{cases} 1, & \Sigma_{k-1}(i,i) > \epsilon, \\ 0, & \Sigma_{k-1}(i,i) \leq \epsilon. \end{cases}.$$

In the case of the *LogEig* layer, this gradients are computed as

$$\frac{\partial L^{(k')}}{\partial U_{k-1}} = 2\left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym} U_{k-1} \log\left(\Sigma_{k-1}\right),$$

$$\frac{\partial L^{(k')}}{\partial \Sigma_{k-1}} = \Sigma_{k-1}^{-1} U_{k-1}^T \left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym} U_{k-1}.$$

For the second issue, as was mentioned in Section 5.1.2 the learning depends on the computation of the steepest descent direction within the Stiefel manifold of the connection weights [12]. Given a connection weight $W_k \in St(d_k, d_{k-1})$ in the t-th descent step, their Euclidean gradient w.r.t the loss function $\nabla L_{W_k^t}^{(k)}$ (see equation 2) is not consistent with the geometry of the *Stiefel* manifold

(see Figure 6). For that, it is necessary to calculate the tangent component of $St(d_k, d_{k-1})$, which is defined as the subtraction of the Euclidean gradient $\nabla L^{(k)}_{W^t_k}$ and the normal component: $\widetilde{\nabla} L^{(k)}_{W^t_k} = \nabla L^{(k)}_{W^t_k} - \nabla L^{(k)}_{W^t_k}(W^t_k)^\top W^t_k$. Now, the tangent component $\widetilde{\nabla} L^{(k)}_{W^t_k}$ can be seen as the direction to update the connection weight W^t_k. Finally, the retraction operation Γ is used to map this weight into the $St(d_k, d_{k-1})$ via

$$W^{t+1}_k = \Gamma\left(W^t_k - \alpha\widetilde{\nabla} L^{(k)}_{W^t_k}\right). \tag{4}$$

The result is an updated weight W^{t+1}_k using a learning rate α and a the gradient descent algorithm defined in equation 3. This process is illustrated in Figure 6. The retraction operation is defined as the Q matrix of the QR decomposition [52]. In this way, the Riemannian weights are updated and the loss can be propagated through the successive convolutional layers to update them using the *Adam* optimizer.

5.2. Spatio-temporal explainable maps

In this work, we seek to discover the underlying characteristics of spatial and temporal regions that may explain particular disease prediction complement analysis by finding new digital disease biomarkers. To do so, we implement an interpretability module into the proposed architecture based on a Gradient-weighted class activation mapping (GradCAM)[53] (see in Figure 5). This provides a visual explanation output as a new spatio-temporal slice that informs about the contribution of each space $\times$ time location w.r.t a particular prediction. Particularly, for an input slice $S_\theta(x, t)$ of dimension $W \times H$, the model produces a bank of feature maps in their last convolutional layer $\{F^k(i_x, i_t); 1 \leq k \leq K, 1 \leq i_x \leq W', 1 \leq i_t \leq H'\}$, where K, W', H'

---

[52] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[53] Ramprasaath R Selvaraju et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

respectively refer to the number of channels, the number of space samples and the number of time samples of the considered layer.

Then, the output PD prediction probability p is related w.r.t changes at this convolutional level, taking $\tilde{p} = \max\{p, 1-p\}$ and computing gradients of $\tilde{p}$ with respect to $\{F^k(i_x, i_t)\}$. Following a backpropagation, the chain rule including Riemannian derivatives, we calculate global average of gradients of features from this layer, as follows:

$$w_k = \frac{1}{W' \times T'} \sum_{i_x=1}^{W'} \sum_{i_t=1}^{T'} \frac{\partial \tilde{p}}{\partial F^k(i_x, i_t)}.$$

Here w_k is interpreted as a weight that quantifies the global importance of the k-th feature map w.r.t the model prediction $\tilde{p}$. Then, finally, explainability maps are calculated as a weighted average of the feature maps, rectified by the ReLU function:

$$F_{GradCAM} = max\left(0, \sum_{k=1}^{K} w_k F^k(i_x, i_t)\right).$$

The resultant maps are valuable to complement prediction assistance, helping to visually identify discriminative patterns from the video slices.

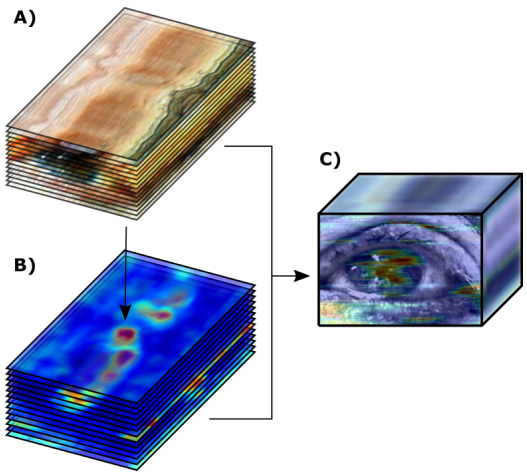

Figure 7. **A)** Video slices along the vertical axis with $\theta = 0°$ direction. **B)** Corresponding 2D GradCAM heatmaps. **C)** Explainable video volume.

In clinical scenarios, such tool can be used to support the diagnosis, marking important regions that suggest parkinsonian abnormalities. Further, we reconstruct an explainable output video from the stack of 2D GradCAM slices. For that, we sliced the video along the same direction but at different heights. For example, slices along the vertical axis with $\theta = 0°$ direction are illustrated in Figure 7.A). This way, we obtained a set of 2D slices from which we can calculate a corresponding set of 2D GradCAM maps. We placed these maps on top of each other in a corresponding order to construct an explicability volume as illustrated in Figure 7.B). Finally, we overlap the original volume in grayscale and the GradCAM volume to visualice relevant regions on the video, see Figure 7.C).

6. EXPERIMENTAL SETUP

6.1. Network Configuration.

The proposed model was adjusted according to the following configuration. For the convolutional module, the weights were initialized from a pre-trained ResNet-18 architecture [54], which has a total of 8 convolutional blocks. Regarding the Riemannian module, we implement BiRe blocks (a BiMap layer, followed by a ReEig layer), which reduce the input dimension by half. A total of four models were herein evaluated using the following architecture:

- A convolutional module using only the first $m = \{2, 4, 6, 8\}$ convolutional blocks, which outputs 64, 128, 256 and 512 feature maps of size 53×75, 27×38, 14×19 and 7×10 respectively.

- A pooling layer using $f_{SPDpool}$. The output of the convolutional module fed this layer. Assuming that the features come from the $\{2, 4, 6, 8\}th$ block, we obtain SPD matrices of size 64×64, 128×128, 256×256, and 512×512 respectively.

- The Riemannian module. We study deeper and shallow modules with 3 BiRe blocks (ConvSPD mth-3BiRe models) and 1 BiRe block (mth-3BiRe models). For all ReEig layers, a rectification threshold $\epsilon = 10^{-4}$ was set. In order to output a PD probability estimation, the final layers correspond to one *LogEig* layer, a flatten layer, and a fully connected layer with a softmax function.

This ConvSPD hybrid model is trainable end-to-end starting from video slices while learning convolutional and Riemannian structures. Each ConvSPD mth-nBiRe model was trained during 50 epochs. The CNN module and final layers were adjusted with a learning rate of 10^{-4} and

---

[54] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

an Adam optimizer, while the Riemannian module weights were adjusted with a learning rate 10^{-3}, using the optimization procedure of the equation 4.

6.2. Validation

The proposed strategy was validated using a 5-fold cross-validation scheme. Each fold was configured with 21 subjects for training ($840 = 21 \times 40$ samples) and five (approximately $200 = 5 \times 40$ samples) for testing. The test sets contain at least two patients with PD in these fold splits. There is only one fold with 20 subjects for training and 6 for testing. In this fold, the test set comprised three patients with PD and three Control. Each model's performance was measured using the average sensitivity, accuracy, and F1-score. The flowchart of the evaluation strategy is shown in Figure 8.

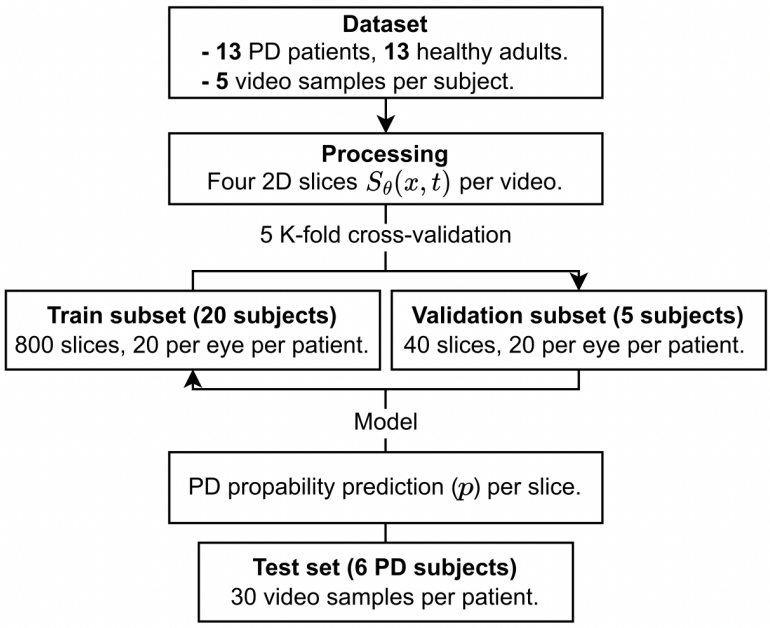

Figure 8. Flowchart of the study. From the fixation video sequences, video volumes are sliced along a direction to calculate 2D slices as part of the processing. Then, they feed the network to obtain a prediction. The model is evaluated using 5-fold cross-validation. In addition, it is tested on an extra dataset.

During training and validation, we also recover the resultant Riemmanian representation, en-

coded in SPD embedding matrices. We compute distances among these matrices, that represent points in the learned Riemmanian manifold, to measure the discriminatory capability of such representation. We use the Riemannian distance $d_R(x, y) = ||\log_x(y)||_x$, where x, y are embedding points, $\log(\cdot)$ is the Riemannian logarithm map and $||\cdot||_x$ an *affine-invariant* norm [33]. With the class-labeled data, we consider two disjoint sets P (Parkinsonian) and C (control) in the manifold, respectively. Therefore we define the separation metric as the Riemannian distance between P and C, as $d_R(P, C) = \frac{1}{|P||C|} \sum_{x \in P} \sum_{y \in C} d_R(x, y)$. From such analysis, we validated the extensive capabilities of the model to support observational tools, from embedding spaces by associating the motor patterns of a particular case, regarding the neighborhood labeled data in a particular space.

7. EVALUATION AND RESULTS

The proposed approach was exhaustively validated concerning the discrimination between Parkinsonian fixation patterns and control subjects' fixation recordings. Firstly, an ablation study was carried out to measure the impact of depth convolutional and Riemannian BiRe blocks. In such sense, the ConvSPD mth-nBiRe models were designed, using $m \in \{2, 4, 6, 8\}$ convolutional blocks and $n \in \{1, 3\}$ BiRe blocks. Such study was carried out according to the classification task, measuring the sensitivity, accuracy and F1-Score of each model. Table 1 summarizes the achieved results from different mth-nBiRe configurations. In all metrics it is visible that intermediate and deep blocks aid the network to achieve better results. The best results achieve an accuracy of 98.2%($\pm$2.2), 98.6($\pm$1.4), and 98.6($\pm$1.6) for 4th-3BiRe, 6th-1BiRe, and 6th-3BiRe model respectively. In general, all the model configurations achieve outstanding scores above 95%. This way, mid-level features maps and deep geometrical representations support the ConvSPD model to discriminate Parkinson from control subjects by observing tiny microtremor patterns into fixational exercises. Moreover, the results of ConvSPD 4th-3Block and 6th-nBlock enhance the trade-off between the size and the number of features, where small features (ConvSPD 8th-nBlock) turn out to be statistically insufficient and few features (ConvSPD 2nd-nBlock) limit the measurement of similarities. Also, it is worth noting that the proposed strategy evidence robustness to recording variations since the recording protocol is weakly controlled, without a strict requirement of camera configuration or head position.

As comparison with standard convolutional networks, Table 2-bottom summarizes the achieved results for proposed architectures implemented only with convolutional layers (fully convolutional Conv-mth models). In such a case, the SPDpooling and Riemannian layers were replaced with a global average pooling layer. The implemented convolutional models Conv-mth include $m \in \{2, 4, 6, 8\}$th blocks. The performance of these models increase in models that comprise more layers, achieving an accuracy of 93.2%($\pm$6.8) with the Conv-2nd model and above 96% for Conv-6th and Conv-8th models. As observed, these results allow measuring the contribution

Table 1. An ablation study of the proposed ConvSPD mth-nBiRe models varying the number of $m \in 2, 4, 6, 8$ convolutional blocks and the number of $n \in 1, 3$ BiRe blocks. The sensitivity, accuracy, and F1-Score metrics describe the performance of the models. The last column represents the number of parameters of each model in millions (M).

Model	Sensitivity	Accuracy	F1-score	Parameters
2nd-1BiRe	95.7 ± 7.0	96.5 ± 4.2	96.6 ± 3.7	0.161M
2nd-3BiRe	92.4 ± 9.6	95.5 ± 5.4	95.3 ± 5.1	0.163M
4th-1BiRe	95.8 ± 5.5	97.2 ± 3.1	97.2 ± 2.7	0.689M
4th-3BiRe	**97.4** ± 3.8	**98.2** ± 2.2	98.2 ± 1.8	0.698M
6th-1BiRe	**98.5** ± 2.2	**98.6** ± 1.4	98.6 ± 1.5	2.796M
6th-3BiRe	**98.0** ± 2.8	**98.6** ± 1.6	98.7 ± 1.3	2.831M
8th-1BiRe	97.1 ± 5.8	97.2 ± 3.5	97.4 ± 3.2	11.22M
8th-3BiRe	96.6 ± 6.0	97.3 ± 3.9	97.6 ± 3.5	11.36M

provided by the SPD modules, representing an accuracy gain of about 0.9-3.3% of ConvSPD models regarding fully convolutional models. Moreover, the highest gain is observed comparing the ConvSPD 2nd-1BiRe and Conv-2nd model. Additionally, the results of the best convolutional model (Conv-8th) were achieved with a more complex network of approximately 11.7M parameters.

Table 2. Comparison of the proposed architecture and fully convolutional models. Best ConvSPD mth-nBiRe models in top. And fully convolutional Conv-mth models in bottom.

Model	Sen	Acc	F1
6th-1BiRe	**98.5** ± 2.2	**98.6** ± 1.4	98.6 ± 1.5
Conv2nd	92.8 ± 10.8	93.2 ± 6.8	93.7 ± 5.8
Conv4th	95.3 ± 7.0	95.3 ± 4.5	95.7 ± 3.6
Conv6th	94.9 ± 10.7	96.6 ± 6.2	96.7 ± 5.9
Conv8th	95.2 ± 9.0	96.4 ± 5.1	96.6 ± 4.7

The convergence of the proposed approach was also analyzed to measure the stability of the SPD component together with convolutional representation. Figure 9 shows the effective convergence performance for all models. Deeper models present a rapid and stable convergence as expected, thanks to the wide receptive field. The ability of the Riemannian module to improve generalization is also observed in compact models with fewer semantics features. The inference time of the proposed descriptor is determinant to be implemented in clinical scenarios, and

routine validation of PD patients. Inference time refers to the time that takes the model to forward the video slices data and produce a prediction. We observed during evaluations that all the proposed ConvSPD architectures compute the inference in acceptable time (< 1 sec). Additionally, we can see how the ConvSPD 4th-3BiRe model and 6th-nBiRe models with almost 20 and 4 times fewer parameters respectively, outperforms the fully convolutional model Conv8th. The implementation of clinical routine requires a trade-off between computational inference time, parameters of the representation to define computational architecture, but also sufficient precision to support diagnosis and following.

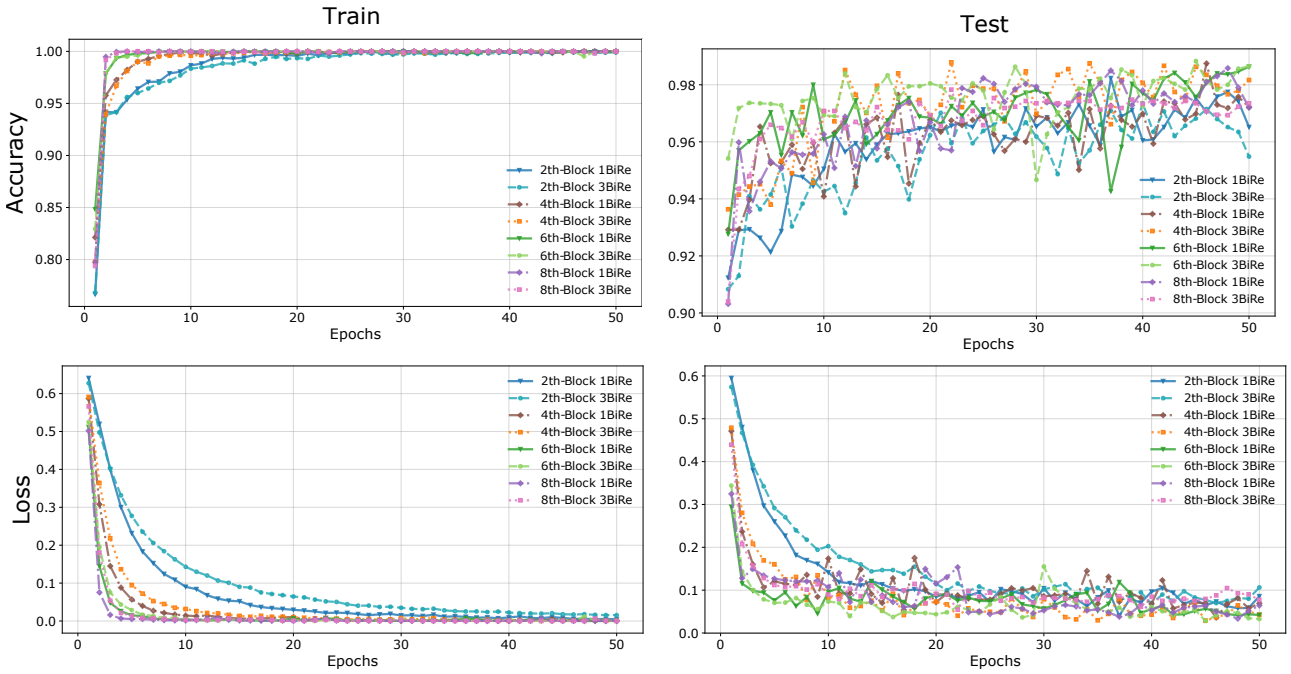

Figure 9. Convergence of the models. At the left, during the training, and at the right during validation.

As visual analysis, we recover the feature maps from PD patients of the 2nd convolutional Block at the first and last epochs. For this last experiment, instead of carrying out a cross validation, two random subjects were chosen for each class, and the rest was used for training a ConvSPD *4th-3BiRe* model one time. In Figure 10.a) is illustrated an average of these maps, observing the evolution of the model. Specifically, the model focus on the center of slices. Similarly,

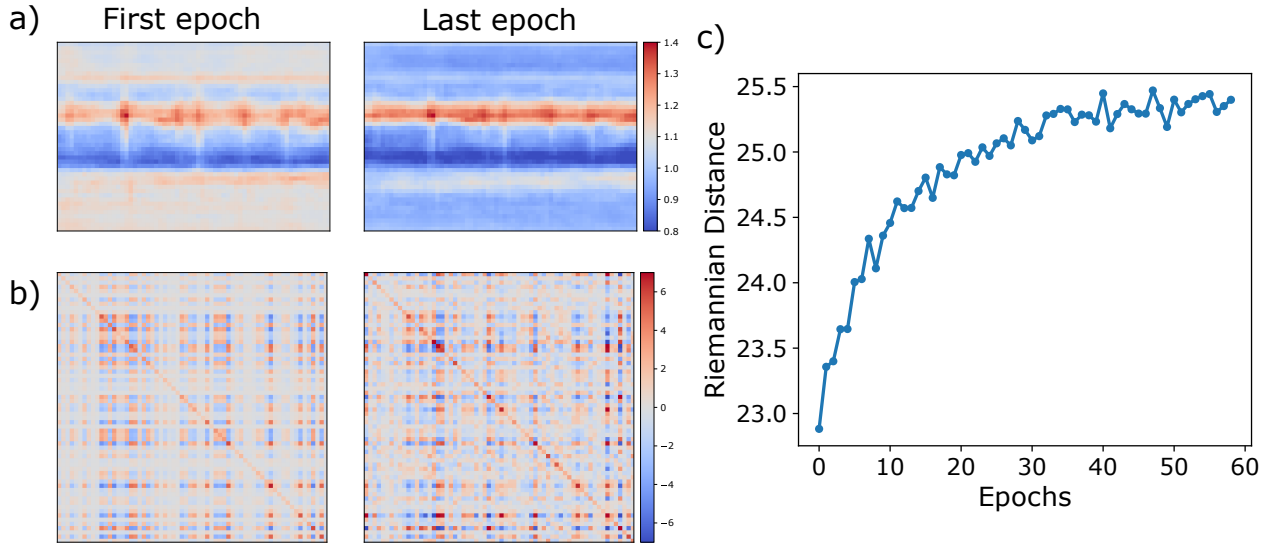

Figure 10. Analysis of ConvSPD structures. We take sample activations extracted from the ConvSPD 4th-3Bire model. **a)** Average activations from the second block of the convolutional module on the first (left panel) and last (right panel) epoch. **b)** SPD matrices from the third BiRe block of the Riemannian module on the first (left panel) and last (right panel) epoch. **c)** Evolution of the average Riemannian distance between Parkinson and Control subjects on the test sets.

in Figure 10.b) we illustrate the average of the SPD matrices extracted from the third BiRe block. As observed in this figure, the Riemannian module contributes to focus attention on the eye region, while supporting the selection of features that major contributes to the prediction. Finally, we measure the average Riemannian distance between the two classes in the test set during training, see Figure 10 .c). Here can be seen a growing trend of the distance. In such cases, there exists major support of the Riemannian structures to the discrimination between control and PD patients.

In a second experiment, we studied the capability of each model to differentiate among different disease stages. For this purpose, we stratified the data into Control, Stage 2.5, Stage 3, and Stage 4 subsets. Then, we recover the output probabilities about the disease class for each input sample into the validation test. This experiment was run with all ConvSPD models, and the model that best produced different distributions for the binary classification was the ConvSPD4th-3BiRe.

In this net, in the binary task, 14 PD samples were misclassified as normal; 11 correspond to a patient at stage 3. Figure 11 illustrates the violin plots of each disease subset and the control population from the Parkinson output probability. As expected, a clear discrimination regarding control patterns exists, evidencing a normal distribution around zero mean, with some outliers. Additionally, we plot distributions for Parkinson's stages and implemented the Kolmogorov-Smirnov (KS) statistical test to validate the capability of the proposed approach to separate classes. The KS test is a non-parametric method that measures the agreement between two distributions. In such a way, it allows testing the hypothesis that two sample probability distributions are not different. In such case, the proposed approach (configuration ConvSPD4th-3BiRe) finds significative differences between early stage 2.5 and 3 ($p < 0.07$), stage 2.5 and 4 ($p < 10^{-3}$), and between stage 3 and 4 ($p < 10^{-4}$). This discrimination may be associated with the way, shape, and clustering in which the central information of the different populations was distributed. Interestingly, the proposed approach evidences a capability to separate among Parkinson levels, even considering that the training regime was a binary task (Parkinson vs. Control).

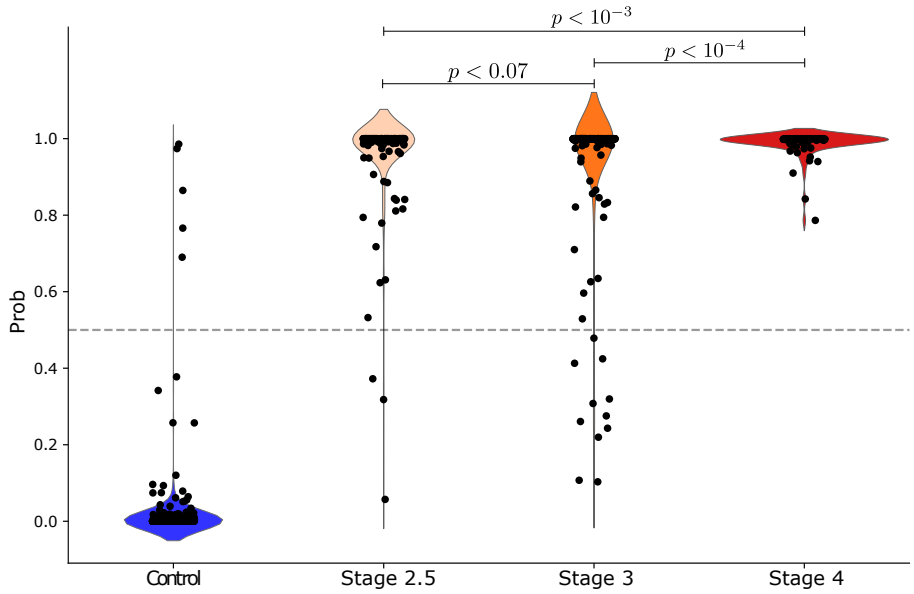

Figure 11. Stratification of distributions from the model's prediction. A statistical test measures the agreement between distributions. The proposed method find differences between PD stages.

We also validated the proposed approach concerning the capability to support observational analysis by generating synthetic videos from explainable probability maps. To achieve this, we densely split each video into slices at 0° in one-pixel steps. These slices achieve the highest probability predictions into trained ConvSPD4th-3BiRe net, obtaining a 98.85 (95% CI: [97.7,100])). Thus, we computed explainable maps from GradCAM to spatially recover regions with major association with output predictions. For a particular video, this block of explainable maps was used to reconstruct a synthetic video that stands out regions with significant disease correlation, activated during the mapping process of a particular net. An example of this procedure is shown in Figure 7).

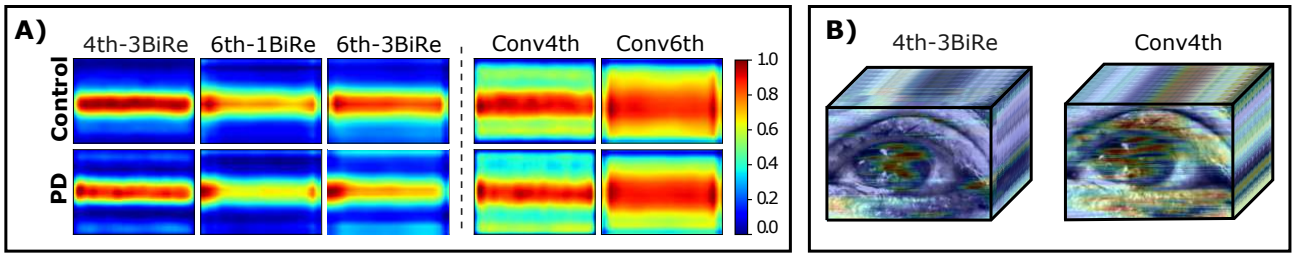

Figure 12. A) Comparison of average GradCAM heatmaps over horizontal slices ($\theta = 0°$). In the top, computed from the last layer of the convolutional module of the ConvSPD models. At the bottom, from the last convolutional layer of the Conv models. **B)** Synthethic video reconstruction from explainable maps.

Figure 12 summarizes the achieved results from slices and synthetic videos, from output predictions of Parkinsonian and control fixational patterns. The explainable maps and synthetic videos were also recovered from the fully convolutional models as a baseline. As expected, the maps obtained using geometrical representations (ConvSPD models) show evident importance in the center region, *i.e.,* the response of eye microtremor. Additionally, Figure 12-B illustrates the synthetic videos reconstructed from the recovered explainable slices for both: geometrical and fully convolutional strategies. As expected, the videos recovered from the proposed approach focus on oculomotor patterns, which in time, consider blinking patterns that correlate with disease. Moreover, the convolutional strategies represent sparse attention patterns without coherence in many regions responsible for producing predictions. Therefore, recovered videos

may support observational analysis of the disease, as well as support numerical prediction outputs of the net regarding the spatio-temporal coherence. Finally, concerning the capability of the proposed approach with respect to the generalization and robustness to operate in different conditions, we evaluate the best model over the extra dataset's six PD patients. Here all samples were well predicted with an average PD prediction of 99.94%.

8. DISCUSSION

This work presented a geometric deep learning-based strategy to characterize oculomotor fixational patterns following a convolutional representation coded in a compact descriptor that operates in a Riemannian manifold. The proposed approach reveals a stable convergence, generalization of the architecture, and relative robustness for learning complex patterns from limited training data samples. The method shows an outstanding performance, achieving 98.2% of accuracy using four convolutional blocks and three Riemannian BiRe structures (4th-3BiRe model) in a real scenario with 13 PD patients and 13 control subjects. From an ablation study, we observed that intermediate layers aid the network to capture spatio-temporal patterns associated to PD. Since shallow models have feature maps with less semantics, these results show the ability of the Riemannian module to improve the network's generalization.

We also note that more stacked BiRe blocks only improve models that use intermediate convolutional representations. That might suggest that the learning of the Riemannian module depends on the quality of the SPDlayer encoded statistics. Compared with fully convolutional models, the proposed method achieves higher scores and with a better confidence, where the gain of ConvSPD models was larger comparing shallow models. Besides, ConvSPD models found a clear difference between the Control subset and each subset of the Parkinson stages (below $p < 0.001$).

Developing a deep learning strategy requires consistent networks where each architecture module performs the expected processing. In this sense, a visual analysis of the convolutional layers showed that the proposed model detects areas of interest in the slices where the eye movement occurs. Furthermore, we found that the Riemannian module effectively enriches the relationships between features while reducing the importance of others. Besides, this geometric module shows the ability to learn Riemannian mappings that separate the SPD matrices, which traduces in a discriminative method.

In fact, in a clinical context, a mandatory factor to include tools in standard analysis protocols is

the capability to explain outputs to support decisions coherently. More than output predictions, the strategies should be able to explain the particular model decision in terms of particular inputs. Hence, a complementary contribution of the proposed strategy is the adaptation of an explainability mechanism to retrieve the main eye spatio-temporal regions that principally contribute to a particular disease prediction. As expected, the attention maps retrieve a more coherent representation that may support observational analysis to discover abnormal locomotor patterns associated for a particular PD-diagnosed patient. For the controls, the area of these maps is wider and more regular than PD slices, suggesting that the method capture disease-related spatiotemporal abnormalities. Furthermore, the videos mark skin regions such as the eye rims and eyebrows. Contrary, although convolutional schemes bring competitive scores, their retrieved grad-cam maps are sparse without attention in the eye region.

Additionally, a stage disease discrimination sensibility was here validated from the geometrical approach, trained from a binary classification task. Here the 4th-3BiRe model produced statistically different probability distributions among PD stages. Interestingly, for advanced stages, the model easily predicts all samples closely spread around higher scores. However, for the early stages (stage 2.5 and 3), the distributions were more sparse and challenging to distinguish. That can be explained by the no clear progression through H&Y rating scale. Indeed, more than 60% of experts suggests that there is a non-linearity through stages [55].

Finally, the model achieves a remarkable quantitative and qualitative generalization over extra data that guarantee head stability, supporting the hypothesis of eye fixational patterns [19]. That is, PD quantification is achieved by ocular abnormalities but not by head motion. In fact, the proposed strategy outperforms an approach reported in the literature with single measures that are mapped to standard machine learning strategies, which achieves an accuracy of 87.7%, 10% less than the proposed method [10].

---

[55] Christopher G Goetz 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". In: *Movement disorders* 19.9 (2004), pp. 1020–1028.

9. CONCLUSIONS AND FUTURE WORK

This work introduced a novel PD digital biomarker from fixational oculomotor patterns using a geometric deep-learning strategy. For this purpose, video sliced data was provided to represent ocular fixation abnormalities. The whole deep representation integrates convolutional and Riemannian modules to discriminate between Parkinson and Control population effectively. In a clinical scenario with few observations, the proposed method produces stable results considering inter-patient variability and can accurately predict in a reasonable inference time. In addition, the presented work differentiates PD stages and generates explainable maps that retrieve the spatio-temporal regions, in the input video sequences, with a major contribution to output predictions. This tool may effectively support PD diagnosis and the progression of motor alterations. Future works include an exhaustive study about the generalization capability of the proposed strategy in different clinical centers under more flexible capture protocols and with larger datasets. The study of unbalanced scenarios is necessary for a robust training of the introduced strategy and the adaptation to receive complete video sequences and process raw volumetric inputs. An exhaustive evaluation of other eye movements is also of great concern for future approaches, with aims for progression monitoring and follow-up of patient's treatments.

BIBLIOGRAPHY

Absil, P-A, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009 (cit. on p. 35).

Acharya, Dinesh et al. "Covariance pooling for facial expression recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 367–374 (cit. on pp. 13, 29).

Antoniades, CA and C Kennard. "Ocular motor abnormalities in neurodegenerative disorders". In: *Eye* 29.2 (2015), pp. 200–207 (cit. on p. 14).

Arsigny, Vincent, Pierre Fillard, and et al. "Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices". In: *SIAM Journal on Matrix Analysis and Applications* (2007), pp. 328–347 (cit. on p. 19).

Bronstein, Michael M. et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418 (cit. on p. 13).

Brooks, Daniel et al. "Riemannian batch normalization for SPD neural networks". In: *arXiv preprint arXiv:1909.02414* (2019) (cit. on p. 23).

Cai, Yinghao, Valtteri Takala, and Matti Pietikainen. "Matching groups of people by covariance descriptor". In: *20th ICPR*. IEEE. 2010, pp. 2744–2747 (cit. on p. 17).

Chan, Florence et al. "Deficits in Saccadic Eye-Movement Control in Parkinson's Disease". In: *Neuropsychologia* 43 (2005), pp. 784–796 (cit. on p. 12).

Dorsey, E Ray et al. “Global, regional, and national burden of Parkinson’s disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”. In: *The Lancet Neurology* 17.11 (2018), pp. 939–953 (cit. on p. 11).

Duval, Christian and Anne Beuter. “Fluctuations in Tremor at Rest and Eye Movements during Ocular Fixation in Subjects with Parkinson’s Disease”. In: *Parkinsonism & Related Disorders* 4 (1998), pp. 91–97 (cit. on p. 16).

Ekker, Merel S et al. “Ocular and visual disorders in Parkinson’s disease: common but frequently overlooked”. In: *Parkinsonism & related disorders* 40 (2017), pp. 1–10 (cit. on pp. 14, 15).

Fukushima, Kikuro et al. “Impaired smooth-pursuit in Parkinson’s disease: normal cue-information memory, but dysfunction of extra-retinal mechanisms for pursuit preparation and execution”. In: *Physiological reports* 3.3 (2015) (cit. on p. 16).

Gatys, Leon, Alexander S Ecker, and Matthias Bethge. “Texture Synthesis Using Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015 (cit. on p. 30).

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. “Image Style Transfer Using Convolutional Neural Networks”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2414–2423 (cit. on p. 30).

Gitchel, George T., Paul A. Wetzel, and Mark S. Baron. “Pervasive Ocular Tremor in Patients With Parkinson Disease”. In: *Archives of Neurology* (2012) (cit. on pp. 12, 14, 16, 21).

Gitchel, George T et al. “Experimental support that ocular tremor in Parkinson’s disease does not originate from head movement”. In: *Parkinsonism & related disorders* 20.7 (2014), pp. 743–747 (cit. on pp. 14, 15, 21, 49).

Goetz, Christopher G 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”. In: *Movement disorders* 19.9 (2004), pp. 1020–1028 (cit. on p. 49).

Goodfellow, Ian, Yoshua Bengio, and et al. *Deep learning*. 2016 (cit. on p. 20).

Gorges, Martin, Hans-Peter Müller, and et al. “The association between alterations of eye movement control and cerebral intrinsic functional connectivity in Parkinson’s disease”. In: *Brain imaging and behavior* 10.1 (2016), pp. 79–91 (cit. on p. 15).

Guayacán, Luis Carlos, Edgar Rangel, and Fabio Martínez. “Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 3688–3691 (cit. on pp. 12, 22).

He, Kaiming et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778 (cit. on p. 38).

Huang, Zhiwu and Luc Van Gool. “A Riemannian network for SPD matrix learning”. In: *Thirty-First AAAI Conference on Artificial Intelligence* 31.1 (2017) (cit. on pp. 13, 21, 22, 30–34).

Ionescu, Catalin, Orestis Vantzos, and Cristian Sminchisescu. “Matrix backpropagation for deep networks with structured layers”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 2965–2973 (cit. on p. 33).

Jankovic, Joseph. “Parkinson’s disease: clinical features and diagnosis”. In: *Journal of neurology, neurosurgery & psychiatry* 79.4 (2008), pp. 368–376 (cit. on p. 14).

Jayasumana, Sadeep et al. “Kernel methods on the Riemannian manifold of symmetric positive definite matrices”. In: *Proceedings of ICCVPR*. 2013, pp. 73–80 (cit. on p. 22).

Kandel, Eric R et al. “The Control of Gaze”. In: *Principles of Neural Science, Fifth Edition*. McGraw-Hill Education, 2014, pp. 894–916 (cit. on p. 14).

Kaski, Diego et al. “Ocular tremor in Parkinson’s disease is due to head oscillation”. In: *Movement Disorders* 28.4 (2013), pp. 534–537 (cit. on p. 15).

Krauzlis, Richard J, Laurent Goffart, and Ziad M Hafed. “Neuronal control of fixation and fixational eye movements”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1718 (2017) (cit. on p. 16).

Lai, Hsin-Yu et al. “Enabling saccade latency measurements with consumer-grade cameras”. In: *25th ICIP*. 2018, pp. 3169–3173 (cit. on p. 22).

Lang, Serge. *Fundamentals of Differential Geometry*. Vol. 191. Springer, 1999 (cit. on pp. 18, 19).

Li, Peihua et al. “Is second-order information helpful for large-scale visual recognition?” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2070–2078 (cit. on p. 12).

Li, Yanghao et al. “Demystifying neural style transfer”. In: *arXiv preprint arXiv:1701.01036* (2017) (cit. on p. 30).

Litjens, Geert, Thijs Kooi, and et al. “A Survey on Deep Learning in Medical Image Analysis”. In: *Medical Image Analysis* (2017), pp. 60–88 (cit. on p. 20).

MacAskill, Michael R and Tim J Anderson. "Eye movements in neurodegenerative diseases". In: *Current opinion in neurology* 29.1 (2016), pp. 61–68 (cit. on p. 14).

Minh, Hà Quang and Vittorio Murino. "Covariances in computer vision and machine learning". In: *Synthesis Lectures on Computer Vision* 7.4 (2017), pp. 1–170 (cit. on pp. 17, 18).

Olmos, Juan, Antoine Manzanera, and Fabio Martínez. "An Oculomotor Digital Parkinson Biomarker from a Deep Riemannian Representation". In: *International Conference on Pattern Recognition and Artificial Intelligence*. Springer. 2022, pp. 677–687 (cit. on p. 28).

Otberdout, Naima et al. "Deep covariance descriptors for facial expression recognition". In: *arXiv preprint arXiv:1805.03869* (2018) (cit. on p. 17).

Pennec, Xavier, Pierre Fillard, and Nicholas Ayache. "A Riemannian framework for tensor computing". In: *International Journal of computer vision* 66.1 (2006), pp. 41–66 (cit. on pp. 18, 40).

Poewe, Werner et al. "Parkinson disease". In: *Nature Reviews Disease Primers* 3.1 (Mar. 2017), pp. 1–21 (cit. on pp. 11, 14).

Saifee, Tabish A et al. "Tremor of the eyes in Parkinson's disease: Merely a measure of the head movement". In: *Parkinsonism & related disorders* 20.12 (2014), pp. 1447–1448 (cit. on p. 15).

Salazar, Isail et al. "A convolutional oculomotor representation to model Parkinsonian fixational patterns from magnified videos". In: *Pattern Analysis and Applications* 24.2 (2021), pp. 445–457 (cit. on pp. 12, 13, 22, 49).

Selvaraju, Ramprasaath R et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626 (cit. on p. 35).

Szymański, Artur et al. "Building Intelligent Classifiers for Doctor-Independent Parkinson's Disease Treatments". In: *Information Technologies in Medicine*. Advances in Intelligent Systems and Computing. 2016, pp. 267–276 (cit. on p. 22).

Tabia, Hedi et al. "Covariance Descriptors for 3D Shape Matching and Retrieval". In: *Proceedings of ICCVPR*. 2014 (cit. on p. 17).

Teed, Zachary and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow". In: *European conference on computer vision*. Springer. 2020, pp. 402–419 (cit. on p. 29).

Tolosa, Eduardo et al. "Challenges in the diagnosis of Parkinson's disease". In: *The Lancet Neurology* 20.5 (2021), pp. 385–397 (cit. on p. 11).

Turcano, Pierpaolo et al. "Early ophthalmologic features of Parkinson's disease: a review of preceding clinical and diagnostic markers". In: *Journal of neurology* 266.9 (2019), pp. 2103–2111 (cit. on p. 15).

Tuzel, Oncel, Fatih Porikli, and et al. "Region Covariance: A Fast Descriptor for Detection and Classification". In: *ECCV*. 2006, pp. 589–600 (cit. on p. 17).

Wang, Qilong et al. "Deep CNNs meet global covariance pooling: Better representation and generalization". In: *IEEE transactions on pattern analysis and machine intelligence* (2020) (cit. on p. 29).

Wang, Rui, Xiao-Jun Wu, and Josef Kittler. "SymNet: A Simple Symmetric Positive Definite Manifold Deep Learning Method for Image Set Classification". In: *IEEE Transactions on Neural Networks and Learning Systems* (2021) (cit. on p. 23).

Wang, Ruiping et al. "Covariance discriminative learning: A natural and efficient approach to image set classification". In: *2012 ICCVPR*. IEEE. 2012, pp. 2496–2503 (cit. on pp. 20, 22).

Wang, Wu et al. “Early detection of Parkinson’s disease using deep learning and machine learning”. In: *IEEE Access* 8 (2020), pp. 147635–147646 (cit. on pp. 12, 14, 22).

Weil, Rimona S. et al. “Visual dysfunction in Parkinson’s disease”. In: *Brain: A Journal of Neurology* 139.11 (Nov. 2016), pp. 2827–2843 (cit. on pp. 11, 12, 16).

Yu, Kaicheng and Mathieu Salzmann. “Second-order convolutional neural networks”. In: *arXiv preprint arXiv:1703.06817* (2017) (cit. on p. 23).

Zhang, Shengping et al. “Online dictionary learning on symmetric positive definite manifolds with vision applications”. In: *29th AAAI*. 2015 (cit. on p. 22).

APPENDICES

Anexo A. Academic Products

Journals

- O. Mendoza, J. Olmos, F. Martínez. "A Local Volumetric Covariance Descriptor for Markerless Parkinsonian Gait Pattern Quantification". Multimedia Tools and Applications. 2022.
  Status: Published.

- S. Niño, J. Olmos, J. Galvis, F. Martínez "Parkinsonian gait patterns quantification from principal geodesic analysis". Pattern Analysis and Applications. 2022.
  Status: Published.

- J. Olmos, B. Valenzuela, F. Martínez "Quantification of Parkinsonian unilateral involvement from ocular fixational patterns using a deep video representation". Journal of Neuroscience Methods. 2022.
  Status: Under review.

- J. Olmos, A. Manzanera, F. Martínez "Riemannian learning on the SPD manifold to represent and characterize fixational oculomotor Parkinsonian abnormalities". Pattern Recognition Letters. 2023.
  Status: Special Section invitation. Manuscript prepared for submission.

Conference papers

- J. Olmos, A. Manzanera, F. Martínez. "An Oculomotor Digital Parkinson Biomarker from a Deep Riemannian Representation". International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI). 2022.
  Status: Published.

Second Prize, best paper award.

Invitation to extend the work in a special section of "Pattern Recognition Letters" journal.

- J. Olmos, F. Martínez. "A Riemannian Deep Learning Representation to Describe Gait Parkinsonian Locomotor Patterns". 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022.
  Status: Published.

- J. Olmos, J. Galvis, and F. Martínez, "Gait patterns coded as Riemannian mean covariances to support Parkinson disease diagnosis". 17th Iberoamerican Conference on Artificial Intelligence (IBERAMIA). 2022.
  Status: Accepted/Presented.

- J. Olmos, J. Galvis, and F. Martínez, "A geometric mean algorithm of symmetric positive definite matrices". II Conferencia Colombiana de Matemáticas Aplicadas e Industriales (MAPI2). 2022.
  Status: Accepted/Presented.
  Invitation to publish the work in the "Revista Colombiana de Matemáticas". .

Anexo B. Informed Consent

Versión 0.1
Código: _____

ESCUELA DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD INDUSTRIAL DE SANTANDER - Laboratorios Vive Digital

CONSENTIMIENTO INFORMADO

9/10/2020

Proyecto: Cuantificación de patrones locomotores para el diagnóstico y seguimiento remoto en zonas de difícil acceso
Responsables: Fabio Martínez Carrillo, Gabriel Rodrigo Pedraza Ferreira.

Con base en los reglamentos establecidos en la Resolución Nº 008430 del 4 de octubre de 1993 por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud en Colombia y según el artículo 15 relacionado con el Consentimiento Informado usted deberá conocer de forma completa y clara los aspectos de la investigación que se desarrollará. Usted ha sido convocado para este proyecto por cumplir con los requisitos de inclusión para la grabación de un conjunto de datos en vídeo de movimientos en marchas anormales y patológicas de manera natural. Por tal motivo se le invita formalmente a que participe del estudio teniendo en cuenta los siguientes criterios de inclusión:

- Ser mayor de edad.
- Poder realizar 4 ciclos de marcha (8 pasos) de manera seguida

De acuerdo con lo anterior y en cumplimiento de estos criterios, por favor indique con una X en una de las siguientes opciones qué tipo de participante es usted:

___ **Persona control:** Es aquella persona que no presenta ninguna dificultad motora, implicando que no ha sido diagnosticada de ninguna enfermedad que afecte su movimiento natural.

___ **Persona con afectación en la marcha:** Es aquella persona que tiene algún tipo de patología diagnosticada la cual pueda afectar su marcha.

Tenga en cuenta que su participación en este proyecto es ***absolutamente voluntaria***. Por favor lea con cuidado el documento y haga todas las preguntas que desee hasta su total comprensión.

JUSTIFICACIÓN

Usted está invitado a participar en este proyecto para crear un conjunto de datos usados en el análisis de patrones de movimiento en marchas normales y patológicas registradas en vídeo. Los movimientos se esperan analizar en vídeo, registrados de manera natural mediante una cámara convencional buscando capturar el movimiento natural de mínimo 3 ciclos de marcha (6 pasos) en línea recta. Para investigaciones futuras los datos recolectados serán usados únicamente para fines de investigación y sus respectivas publicaciones.

OBJETIVO

Desarrollar una estrategia computacional para el registro, cuantificación y representación de patrones de la marcha en ambientes semi-controlados como soporte de procesos de asistencia a la telerehabilitación.

DESCRIPCIÓN

El proceso de captura será el mismo para personas control como para personas con algún tipo de afectación en sus patrones de marcha. Esto debido a que el estudio considera que los datos han sido capturados bajo las mismas condiciones en ambos tipos de población, en miras de una adecuada evaluación estadística. Esta convocatoria se limita a población con edad mayor a 18 años.

9/10/2020

A cada participante, en presencia de sus acompañantes, familiar o un testigo, se le entregará para su lectura. Si decide participar, podrá proceder a firmarlo. Seguidamente se registran sus datos personales. La grabación de vídeos tomará un tiempo aproximado de 10 minutos. Ante la cámara, usted deberá hacer caminata de 3 ciclos de marcha (6 pasos) equivalente a 8 metros aproximadamente, esta caminata deberá ser repetida 3 veces con el fin de poder capturar cualquier tipo de variación en los patrones de marcha.

Para la sesión de grabación es necesario que usted como los investigadores cumplan con unas medidas de bioseguridad para evitar el contagio del COVID-19, es por esto que a continuación presentamos a usted unas recomendaciones dadas por el Comité de Protocolos de Bioseguridad de la Universidad Industrial de Santander para el desarrollo de la sesión:

1. Mantener el distanciamiento mínimo de dos metros entre personas
2. Todas las personas usan los elementos de protección personal (EPPs) y el uso de careta las personas que participan.
3. El uso de gel antibacterial o alcohol al ingreso y salida del sitio de prueba de cada persona que participa en el proyecto, si la persona tiene contacto con alguna superficie deberá igualmente desinfectar sus manos.
4. Verificar que las personas que participan en el proyecto tengan una afiliación al Sistema de Seguridad Social.

Al participar en este estudio, usted no recibirá ningún tipo de subvención económica o material ni deberá aportar herramienta alguna para la intervención. Al finalizar la investigación, usted podrá recibir los resultados obtenidos de la captura. Este material será presentado a usted por los investigadores cuando culmine la actividad.

Las inquietudes adicionales que surjan en relación con el desarrollo e implicaciones del proyecto podrán ser aclaradas por Fabio Martínez Carrillo, Profesor de la Escuela de Ingeniería de Sistemas e Informática, a quien puede contactar en el teléfono 3103054041, o mediante correo electrónico dirigido a famarcar@saber.uis.edu.co; o directamente en su oficina en la Universidad Industrial de Santander (sede principal) ubicada en la Cra. 27 #9 Ciudad Universitaria, Edificio de Laboratorios Pesados, oficina 231; o al teléfono 577- 6344000 extensión 2110.

RIESGOS

De acuerdo con el Artículo 11 de la Resolución No. 8430 de 4 de octubre de 1993, esta investigación se considera sin riesgo para el participante dado que el estudio únicamente emplea el registro de datos a través de un procedimiento común de captura de vídeos por medio de cámaras ordinarias. De tal forma, ninguno de los métodos utilizados es invasivo o penetra la piel. si durante la captura de los vídeos usted experimenta cualquier tipo de malestar, la grabación será suspendida de inmediato y se le ubicará en estado de reposo. De presentarse cualquier incidente relacionado con esta jornada donde el participante requiere valoración médica inmediata será remitido al servicio de urgencias del Hospital Universitario de Santander o si es su decisión al servicio de urgencias de la entidad donde se encuentre afiliado al sistema de seguridad social. Durante este proceso será acompañado por el investigador principal.

9/10/2020

DERECHO A RETIRARSE

Su participación en este estudio es autónoma y voluntaria, en donde podrá actuar acorde a sus principios personales. Si usted decide no participar, no implicará sanción alguna. Además, usted cuenta con el derecho a negarse a responder a preguntas concretas si así lo desea. También puede optar por retirarse en cualquier momento y toda su información será descartada y eliminada.

CONFIDENCIALIDAD

Los resultados de las pruebas y la información que usted nos ha dado son de carácter absolutamente confidencial, de manera que solamente usted y el investigador principal tendrán acceso a estos datos.

Una copia de los registros con la información de cada participante será archivada por el investigador principal y a cada registro se le asignará un número con el cual se identificará y codificará para su ingreso a la base de datos durante la sistematización de la información. Por lo anterior, los nombres de los participantes no serán divulgados en forma alguna; y cuando los resultados de este estudio sean publicados en revistas o congresos científicos, la información personal de los participantes será debidamente anonimizada previamente.

A menos que Usted dé una autorización específica cuando la ley lo permita, sus resultados personales no estarán disponibles para terceras personas como empleadores, organizaciones gubernamentales, compañías de seguros o instituciones educativas. Esto también aplica a su cónyuge, a otros miembros de su familia. Sin embargo, con el objetivo de realizar un manejo adecuado de los datos, un miembro del Comité de Ética de la Universidad Industrial de Santander podrá consultar sus datos y su registro. Por lo anterior, atentamente se le invita a participar en el estudio y si está de acuerdo, se le solicita su nombre y la firma en las casillas abajo descritas.

AUTORIZACIÓN PARA EL USO DE LA INFORMACIÓN EN ESTUDIOS FUTUROS

Dentro del equipo de investigación al que pertenecen los investigadores responsables (Grupo de Investigación BIVL^{2}ab - *Biomedical Imaging, Vision and Learning Laboratory*) de la Universidad Industrial de Santander, se espera seguir utilizando la información registrada en este estudio para el desarrollo de estudios futuros y derivados. Por lo tanto, al firmar este consentimiento usted puede autorizar al investigador principal a ceder su información a otros investigadores de su equipo de investigación, con previa aprobación del Comité de Ética de la Universidad Industrial de Santander para realizar los estudios mencionados. Por favor marcar con una X si autoriza o no autoriza, y firmar en caso de si autorizar.

9/10/2020

___ Si autorizo
___ No autorizo

_____

Firma Participante
Nombre:
amerite)
C.C.

Huella digital
(En caso que se

_____

Firma Investigador Principal
Nombre:
amerite)

Huella digital
(En caso que se

Yo_____,
identificado con _____ N°_____ de _____ al firmar este consentimiento el día ___ de _____ del _____, acepto participar de manera voluntaria en el presente estudio y autorizo la grabación de mis vídeos y el uso de mis datos individuales para los análisis requeridos. He leído y entendido la información registrada en este documento y mis dudas fueron aclaradas. Entiendo que soy libre de retirarme del estudio. Por otro lado, se me ha garantizado justicia, equidad, autonomía en la participación y la confidencialidad en el manejo de toda la información recolectada, teniendo en cuenta que los resultados del procesamiento de dicha información podrán ser divulgados con fines científicos, mediante presentaciones en congresos o publicaciones en revistas científicas, con la debida protección de mi identidad

_____ Huella digital
Firma Participante (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Profesional Salud (Testigo 1) (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Acompañante Captura de datos (Testigo 2) (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Investigador Principal (En caso que se amerite)
Nombre:

9/10/2020

Contacto Comité de Ética: Para preguntas o aclaraciones acerca de los aspectos éticos de ésta investigación pueden comunicarse con el Comité de Ética en Investigación Científica de la Universidad Industrial de Santander (CEINCI-UIS), o con cualquiera de los miembros del Comité, al teléfono 6344000 Extensión 3808 ó al correo comitedetica@uis.edu.co.