RIEMANNIAN GEOMETRIC REPRESENTATION WITH ATTENTION MECHANISMS FOR PARKINSONIAN CLASSIFICATION OF OCULOMOTOR PATTERNS.

LUIS FERNANDO CELIS MANTILLA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2026

RIEMANNIAN GEOMETRIC REPRESENTATION WITH ATTENTION MECHANISMS FOR PARKINSONIAN CLASSIFICATION OF OCULOMOTOR PATTERNS.

LUIS FERNANDO CELIS MANTILLA

Research work in fulfillment of the requirements for the degree of:
Magíster en Ingeniería de Sistemas e Informática

Advisor:
Fabio Martínez Carrillo, Ph.D.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2026

ACKNOWLEDGEMENTS

The author expresses his acknowledgement:

I would like to begin by thanking the *Americana, Champiñón, Carnes, Hawaiana*, and *Criolla* pizzas, as well as the salchipapas and hamburgers. Each of these flavors reflects the support and hard work of my parents, Lilina and Luis, who kept me company throughout this process and remind me every day that hard work always finds a way to pay off.

I would also like to express my sincere gratitude to Professor Fabio Martínez for his guidance and trust throughout the development of this work. I also thank the members of the BivL^{2}ab group for the learning opportunities that made this research possible. It was an environment of symmetrical collaboration, where each exchange of ideas was, definitively, a positive one, contributing not only to this work but also to my personal and professional growth.

CONTENTS

page

LIST OF FIGURES

page

LIST OF TABLES

page

LIST OF APPENDICES

page

ABSTRACT

TITLE: RIEMANNIAN GEOMETRIC REPRESENTATION WITH ATTENTION MECHANISMS FOR PARKINSONIAN CLASSIFICATION OF OCULOMOTOR PATTERNS. *

AUTHOR: LUIS FERNANDO CELIS MANTILLA **

KEYWORDS: GEOMETRIC DEEP LEARNING, PARKINSON'S DISEASE, GEOMETRIC ATTENTION MECHANISM, RIEMANNIAN DEEP LEARNING.

DESCRIPTION: Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder worldwide, expressed by the progressive loss of dopaminergic neurotransmitters, resulting in motor disturbances. Characterization and analysis of movement disorders play a crucial role in clinical diagnosis, yet the absence of a definitive biomarker poses challenges in disease characterization. Smooth pursuit eye movement (SPEM) analysis has emerged as a potential PD biomarker, even evaluated to detect motor abnormalities at early stages. However, traditional SPEM recording methodologies often require intrusive procedures and specialized protocols, limiting the dynamic description to coarse eye trajectories. We hypothesize that SPEM patterns encompass a diverse range of movement, characterized by intricate spatio-temporal relationships, which may be related to PD, even at early stages. This work introduces a geometrical attention architecture that learns SPEM discriminative patterns to classify PD patients. The proposed approach explores self- and cross-attention variations, coded to obtain oculomotor SPEM patterns. The proposed approach follows a non-invasive and flexible protocol to record oculomotor sequences, summarizing patterns in symmetric positive definite (SPD) matrices. From such a compact SPD representation it is possible to learn the second-order most significant for the proposed self-attention and cross-attention architectures, respectively. The proposed approach was validated in a study involving 25 patients with PD and 25 control subjects, achieving an AUC-ROC of 97.0 ± 4.0 and 98.4 ± 3.2 for the self- and cross-attention strategies, respectively.

---

* Research work

** Facultad de Ingeniería fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Fabio Martínez Carrillo, Ph.D.

RESUMEN

TÍTULO: REPRESENTACIÓN GEOMÉTRICA RIEMANNIANA CON MECANISMOS DE ATENCIÓN PARA LA CLASIFICACIÓN PARKINSONIANA BASADA EN PATRONES OCULOMOTORES. *

AUTOR: LUIS FERNANDO CELIS MANTILLA **

PALABRAS CLAVE: APRENDIZAJE PROFUNDO GEOMÉTRICO, ENFERMEDAD DE PARKINSON, MECANISMO DE ATENCIÓN GEOMÉTRICO, APRENDIZAJE PROFUNDO RIEMANNIANO.

DESCRIPCIÓN: La enfermedad de Parkinson (EP) es el segundo trastorno neurodegenerativo más prevalente a nivel mundial, caracterizado por la pérdida progresiva de neurotransmisores dopaminérgicos que provoca alteraciones motoras. La caracterización y el análisis de los trastornos del movimiento desempeñan un papel crucial en el diagnóstico clínico; sin embargo, la ausencia de un biomarcador definitivo plantea desafíos en la caracterización de la enfermedad. El análisis del movimiento ocular de persecución suave (SPEM) ha surgido como un posible biomarcador de la EP e incluso se ha evaluado para detectar anormalidades motoras en etapas tempranas. No obstante, las metodologías tradicionales de videos de SPEM suelen requerir procedimientos intrusivos y protocolos especializados, lo que limita la descripción dinámica a trayectorias oculares de baja resolución. Hipotetizamos que los patrones de SPEM abarcan una amplia gama de movimientos caracterizados por complejas relaciones espaciotemporales, potencialmente asociadas con la EP incluso en sus etapas iniciales. Este trabajo introduce una arquitectura de atención geométrica capaz de aprender patrones discriminativos de SPEM para la clasificación de pacientes con EP. Se realizó un análisis detallado de las variaciones de autoatención y atención cruzada para codificar los patrones oculomotores de SPEM. El enfoque propuesto sigue un protocolo no invasivo y flexible para registrar secuencias oculomotoras, que se resumen en matrices definidas positivas simétricas (SPD). A partir de esta representación SPD compacta, el modelo aprende las características de segundo orden más significativas que sustentan las arquitecturas de autoatención y atención cruzada propuestas. El enfoque propuesto fue validado en un estudio que incluyó a 25 pacientes con EP y 25 sujetos control, alcanzando un AUC-ROC de 97.0 ± 4.0 y 98.4 ± 3.2 para las estrategias de autoatención y atención cruzada, respectivamente.

---

* Trabajo de investigación

** Facultad de Ingeniería Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.

INTRODUCTION

Neurological diseases are a significant source of disability globally, being the Parkinson's disease (PD) the second most common neurodegenerative disorder, with a current prevalence of 6 million people and an expected growth that projects a total of 267 cases per 100000 population, affecting globally in 2050 [1][2]. PD is associated with the loss of dopaminergic cells, affecting voluntary movement control and resulting in motor manifestations such as tremors, rigidity, slowness, and changes in posture [3]. Today, clinical diagnosis is primarly based on the observation of motor symptoms, often validated using standard scales (UPDRS and H & Y [4]), which support diagnosis and treatments [3]. Nonetheless, such scales are expert-dependent and based on coarse motor alterations, failing to diagnose early is key for effective and personalized treatments that slow down symptoms and disability.

Recently, alterations in oculomotor function have been recognized as potential early biomarkers, sensitive even before the onset of traditional motor symptoms, offering important insights into early disease detection or progression [5]. Some studies have analyzed eye movements during smooth pursuit eye movement (SPEM) tasks, identifying abnormal patterns and suggesting

---

[1] Jian Zhang et al. "Global, regional and national temporal trends in Parkinson's disease incidence, disability-adjusted life year rates in middle-aged and older adults: a cross-national inequality analysis and Bayesian age-period-cohort analysis based on the global burden of disease 2021". In: *Neurological Sciences* 46.4 (2025), pp. 1647–1660.

[2] Dongning Su et al. "Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021". In: *bmj* 388 (2025).

[3] Werner Poewe et al. "Parkinson disease". In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21.

[4] Roongroj Bhidayasiri and Pablo Martinez-Martin. "Clinical assessments in Parkinson's disease: scales and monitoring". In: *International review of neurobiology* 132 (2017), pp. 129–182.

[5] Karen Frei. "Abnormalities of smooth pursuit in Parkinson's disease: A systematic review". In: *Clinical parkinsonism & related disorders* 4 (2021), p. 100085.

them as potential early biomarkers for Parkinson's diagnosis [6] [7]. During SPEM assessment, these abnormal patterns include saccadic alterations, which produce a loss of gaze control and decreasing accuracy when tracking moving stimuli [8]. Additionally, diplopia results in a decrease in the eyes' ability to converge and focus on objects, while impaired vergence affects the perception of dimension and depth [9][10]. Reduced SPEM speed compared to the target stimulus, along with a delay in initiating the eye task, has indicated ocular bradykinesia in some PD cohorts [11]. Furthermore, the prevalence of these anomalies may vary depending on the direction of eye movement [12]. Despite SPEM's potential to improve PD quantification, the wide range of movement patterns may present challenges in accurately capturing relevant and discriminant disease-related abnormalities. Moreover, brief saccades or binocular divergence are often excluded during SPEM evaluations, overlooking potential deficits in binocular coordination when tracking a moving target [9]. Nonetheless, current diagnostic setups and clinical scales have demonstrated sensitivity by relying on motor symptoms that appear in advanced stages. These delays in diagnosis and appropriate treatment impact the prognosis of patients

---

[6] Han Li et al. "Abnormal eye movements in Parkinson's disease: From experimental study to clinical application". In: *Parkinsonism & Related Disorders* (2023), p. 105791.

[7] Martin Gorges, Elmar H Pinkhardt, Jan Kassubek, et al. "Alterations of eye movement control in neurodegenerative movement disorders". In: *Journal of ophthalmology* 2014 (2014).

[8] Yasuo Terao, Hideki Fukuda, and Okihide Hikosaka. "What do eye movements tell us about patients with neurological disorders?—An introduction to saccade recording in the clinical setting—". In: *Proceedings of the Japan Academy, Series B* 93.10 (2017), pp. 772–801.

[9] Chia-Chien Wu et al. "Eye movement control during visual pursuit in Parkinson's disease". In: *PeerJ* 6 (2018), e5442.

[10] Yue Ran Sun et al. "Monitoring eye movement in patients with Parkinson's disease: what can it tell us?" In: *Eye and Brain* (2023), pp. 101–112.

[11] Hiroshi Shibasaki, Sadatoshi Tsuji, and Yoshigoro Kuroiwa. "Oculomotor abnormalities in Parkinson's disease". In: *Archives of Neurology* 36.6 (1979), pp. 360–364.

[12] Elmar H Pinkhardt et al. "Eye movement impairments in Parkinson's disease: possible role of extradopaminergic mechanisms". In: *BMC neurology* 12 (2012), pp. 1–8.

[13]. Consequently, eye movement analysis has emerged as a valuable tool to provide objective and noninvasive disease quantification, offering detailed spatial and temporal information on sensory and cognitive functions [14].

Conventional methods today typically assess ocular movements using sophisticated devices to diagnose Parkinson's disease (PD). For example, electro-oculography (EOG) measures the electrical potential difference between the cornea and retina, while video-oculography (VOG) tracks the pupil's position using a high-resolution camera [15]. However, the complexity of equipment used in oculomotor analysis methods, based on EOG and VOG, limits their use in clinical routines. They are invasive, covering the ocular region, prone to calibration errors, and can affect task performance due to inflexible protocols. Additionally, they often overlook spontaneous manifestations that may vary among patients in magnitude, frequency, and other hidden patterns [15]. Furthermore, these data capture methods have undergone analysis using deep learning strategies, yielding significant findings in disease characterization [16]. However, these deep representations demand a substantial volume of data, a feature that is often unrealistic in clinical centers.

This work introduces a novel Riemannian cross-attention architecture to learn compact SPEM geometric representations, following a non-invasive video protocol, to discriminate between PD patients and control subjects. The proposed method captures complex SPEM dynamics of left and right eye video sequences, separately, through 3D convolutional backbones, which learn rich spatio-temporal representations. Both eyes features are then independently encoded into

---

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

[14] Chrystalina A Antoniades and Miriam Spering. "Eye movements in Parkinson's disease: from neurophysiological mechanisms to diagnostic tools". In: *Trends in Neurosciences* (2023).

[15] Agostina J Larrazabal, CE García Cena, and César Ernesto Martínez. "Video-oculography eye tracking towards clinical applications: A review". In: *Computers in biology and medicine* 108 (2019), pp. 57–66.

[16] Jie Mei, Christian Desrosiers, and Johannes Frasnelli. "Machine learning for the diagnosis of Parkinson's disease: a review of literature". In: *Frontiers in aging neuroscience* 13 (2021), p. 633752.

compact symmetric positive definite (SPD) matrices, correlating disease-related patterns. These SPD matrices are geometric descriptors that are robust to noise, variability, and outliers. Hence, the proposed Riemannian attention mechanism integrates geometric descriptors from both eyes to highlight significant relationships while preserving geometric properties. This mechanism leverages both local and global information from the compact SPD representations. The proposed approach was evaluated in a cohort of 50 patients and demographically matched controls was included to validate the proposed approach, using a non-invasive video setup to capture eye movements during a SPEM task.

Academic products Next publications were included part of reported research on this work.

- **Celis, L. F.**, Olmos, J., & Martínez, F. (2024, November). A Geometric Attention Mechanism to Classify Parkinsonism Smooth Pursuit Patterns. In Ibero-American Conference on Artificial Intelligence (pp. 99-109).

- **Celis, L. F.**, Olmos, J., Manzanera, A., & Martínez, F. (2025). Learning a geometric deep representation to classify Parkinson smooth pursuit patterns. Pattern Analysis and Applications, 28(3), 1-14.

- **Celis, L. F.**, Olmos, J., & Martínez, F. (2025, November). EGA: Enhanced Geodesic Attention for SPEM oculomotor Parkinson Quantification (2025).

1. FUNDAMENTALS AND CURRENT WORKS

1.1. SPEM oculomotor patterns

Parkinson's disease (PD) is produced by the decrease of dopamine production in the substantia nigra, affecting the functions of the basal ganglia responsible for managing involuntary movements [3]. These motor abnormalities are only perceptible at advanced stages, impacting in progression of the disease, and even being subjective regarding neurologist analysis [17].

Typically, the PD is carried out from gait, tremor, and other locomotion variables, following standard scales, such as the H&Y, [4]. These scales coarsely stratify the disease by detecting patterns such as bradykinesia, referring to the slowness of movement, bradykinesia that is the most characteristic clinical feature of PD, encompassing difficulties related to planning, initiating, executing movements, and the development of sequential and simultaneous tasks [18]. In addition to motor disability, freezing refers to a motor block that usually affects the legs during walking but can also affect the arms and eyelids. Also, the resting tremor is the most common and easily recognized symptom of PD, typically unilateral and often affecting the distal parts of extremities, such as the fingers or toes. Additionally, it may manifest in areas like the lips, chin, jaw, and legs. In addition to resting tremor, many PD patients also experience postural tremors, which can be more prominent and disabling than resting tremors and may

---

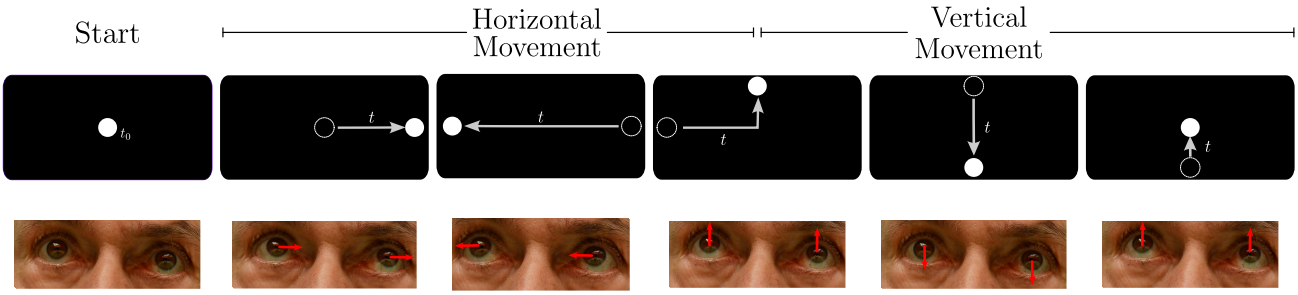
[17] Renee M Hendricks, Mohammad T Khasawneh, et al. "An investigation into the use and meaning of Parkinson's disease clinical scale scores". In: *Parkinson's Disease* 2021 (2021).

[18] Alfredo Berardelli et al. "Pathophysiology of bradykinesia in Parkinson's disease". In: *Brain* 124.11 (2001), pp. 2131–2146.

even serve as the initial indication of the disease [19] [20]. A main drawback of such analysis is the poor sensitivity to detect the disease at early stages [13]. Also, there exist different phenotypes of the disease, which result in remarked variability among motor patterns for different subjects[20]. PD patients may report different patterns, with different magnitudes and from different onset motor alterations.

Recently, ocular movements have emerged as a potential biomarker of PD with advantages regarding the sensitivity to detect abnormalities [9]. In general, the ocular movements are a function of the human body sensitive to neural alterations, offering insights into sensory and cognitive processes [21]. Nowadays, some works evidence an existing relation between neurodegenerative processes and eye motion alterations [5]. This provides an objective and non-invasive method for assessing the disease, potentially supporting it as an early biomarker.

Figure 1. Smooth Pursuit Eye Movement. The upper section of the image shows the target stimulus trajectories during the eye-tracking task, with the dotted circle indicating the starting point and arrows showing the movement. The lower section displays a patient's gaze following these trajectories.



Source: Author's own work

Particularly, Smooth pursuit eye movement (SPEM) is the task, where the subject maintains

[19] Christian Duval. "Rest and postural tremors in patients with Parkinson's disease". In: *Brain research bulletin* 70.1 (2006), pp. 44–48.

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

[21] Ileok Jung and Ji-Soo Kim. "Abnormal eye movements in parkinsonism and movement disorders". In: *Journal of movement disorders* 12.1 (2019), p. 1.

the gaze on a moving target in the visual field, while trying to describe vertical or horizontal trajectories [22] as illustrated in Figure 1. The detection of pattern abnormalities from SPEM has been identified and described in prevalent Parkinsonism cases [12]. The patterns that are observable and correlated with PD are:

- **Saccades.** Rapid eye movements are designed to shift the gaze from one point in the visual field to another, often resulting in a loss of control and accuracy. At early-stage Parkinson's disease primarily affects specific brain areas involved in controlling eye movements, and then it is reported a reduced ability to suppress undesired eye movements. Particularly, for PD, the substantia nigra is limited, a brain region dedicated to sends dopamine signals to the striatum, providing insights into how Basal Ganglia dysfunction affects saccade performance [8].

- **Strabismus and impaired vergence.** These non-specific manifestations include diplopia, blurred vision, or reading difficulty. The prevalence of strabismus in PD could be attributed to the critical role of dopamine in the vergence pathway. Disruption of this mechanism may lead to a decrease in the ability of the eyes to converge and focus on close objects, affecting the perception of dimension and depth [10].

- **Kinematic abnormalities.** These patterns are detected during gaze, including speed impairment during SPEM, manifesting at specific time instances. This indicates lower speed compared to the target stimulus, with a possible delay in initiating the eye task, possibly indicative of ocular bradykinesia [11].

---

[22] Silvia Marino et al. "Quantitative analysis of pursuit ocular movements in Parkinson's disease by using a video-based eye tracking system". In: *European Neurology* 58.4 (2007), pp. 193–197.

1.2. Attention mechanism and geometric learning

This work is focused on the development of attention mechanisms that cover more than local patterns, but operate on geometric representations. For this, in the next subsections, we bring a description of the attention mechanism, as well as, an introduction to geometric learning in Riemannian manifolds.

1.2.1. Attention mechanisms The conventional convolutional architectures are based on extracting local dependencies, producing pixel-wise features that are covariant concerning translation, which have found many applications, particularly in computer vision [23]. Nonetheless, these convolutional representations remain limited to capturing non-local dependencies that may enrich modeling key relationships. For that, Attention models provide the capability of weighting the most relevant non-local features of input sequences, finding main applications in natural language processing. Methodologically, attention models have focused on improving convolutional architectures according to spatiality, focusing on channels and image non-locality [24].

The standard self-attention mechanism involves three parallel non-linear transformations denoted as key ($\mathbf{K} = \mathbf{w}_k\mathbf{X}$), query ($\mathbf{Q} = \mathbf{w}_q\mathbf{X}$), and value ($\mathbf{V} = \mathbf{w}_v\mathbf{X}$), where $\mathbf{X}$ represents the input and $\mathbf{w}_k$, $\mathbf{w}_q$, and $\mathbf{w}_v$ are the respective weight matrices. Implementation within a computational neural network architecture necessitates a similarity measurement function between the query and key representations, capturing non-local correlations across all representation elements. Subsequently, attention scores are derived through a nonlinear operation, commonly a softmax function, applied to the similarity values, yielding attention weights. The value

---

[23] Michael M Bronstein et al. "Geometric deep learning: going beyond euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42.

[24] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. "A review on the attention mechanism of deep learning". In: *Neurocomputing* 452 (2021), pp. 48–62.

projection computation entails a weighted summation of the values using their corresponding attention scores.

$$attention(\mathbf{X}) = softmax\left(\frac{\mathbf{Q}\cdot\mathbf{K}^{\top}}{\sqrt{d_Q}}\right)\cdot\mathbf{V}. \tag{1}$$

From the standard attention module, several important variations have emerged. For instance, Cross-attention models aim to integrate information from two different sources, $\mathbf{X}$ and $\mathbf{Y}$, rather than focusing exclusively on a single source. In these models, the main difference lies in the handling of queries, keys, and values. Queries are extracted from one modality, while keys and values are obtained from the other modality. Other attention modules have integrated convolution branches to preserve spatial representation [25]. Additionally, some proposals leverage transformer models to extract long-range interactions through attention mechanisms, enhancing semantic features and spatial dependencies for segmentation tasks [26].

1.2.2. SPD background Symmetric Positive Definite (SPD) matrices play a key role in this work, as they are used to encoding SPEM features. Specifically, observations $\{M_i\}_{i=1}^{n}$, are coded in an SPD matrix ($\boldsymbol{X}\in\mathbb{R}^{n\times n}$) measuring relationships between vector pairs, and capturing second-order statistical relationships. This matrix has two key properties: symmetry ($\boldsymbol{X}=\boldsymbol{X}^{\top}$) and positive definiteness, guaranteed that $\boldsymbol{v}^{\top}\boldsymbol{X}\boldsymbol{v}>0$, for all $\boldsymbol{v}\in\mathbb{R}^{n}\setminus\mathbf{0}_n$, with strictly positive eigenvalues [27]. The space of SPD matrices of dimension n denoted as $\mathcal{S}_n^{++}$ constitutes a Riemannian manifold, a topological space endowed with a Riemannian metric

---

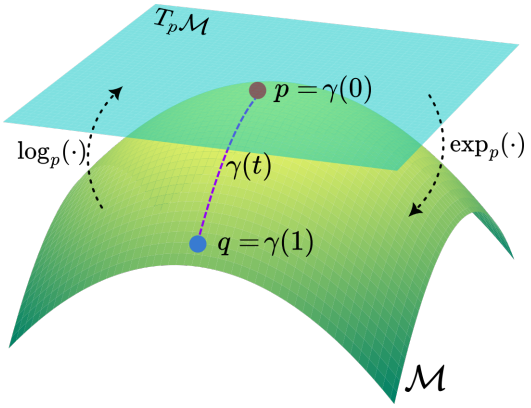
[25] S Woo et al. *Cbam: convolutional block attention module. In proceedings of the European conference on computer vision (ECCV): 3-19*. 2018.

[26] Olivier Petit et al. "U-net transformer: Self and cross attention for medical image segmentation". In: *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer. 2021, pp. 267–276.

[27] George T Gilbert. "Positive definite matrices and Sylvester's criterion". In: *The American Mathematical Monthly* 98.1 (1991), pp. 44–46.

structure. This manifold has an associated Euclidean tangent space at each point $p \in \mathcal{S}_n^{++}$, noted as $T_p\mathcal{S}_n^{++}$. This space provides a vector structure enabling Euclidean operations, such as distances and statistical measures, to the manifold structure [28]. There are two principal mapping functions for this purpose: the logarithmic map $\log_p : \mathcal{S}_n^{++} \to T_p\mathcal{S}_n^{++}$, which maps each point in $\mathcal{S}_n^{++}$ to a tangent vector associated with p, and its inverse, the exponential map, $\exp_p : T_p\mathcal{S}_n^{++} \to \mathcal{S}_n^{++}$, which maps the tangent space back to the manifold.

Figure 2. The tangent plane $T_P\mathcal{M}$ at a point P of a manifold $\mathcal{M}$, along with the Riemannian mappings of the logarithm $\log(\cdot)$ that map the points of the manifold to the tangent plane and its exponential $\exp(\cdot)$ inverse function that facilitate the transition between the manifold and the Euclidean plane, as well as their respective inverses. Purple pointed line $\gamma(t)$ representing the unique geodesic between q, p points.



Source: Author's own work

The distance measure imposes a geometric structure on the descriptor space within the Riemannian manifold. A geodesic curve $\gamma : [a, b] \to \mathcal{S}_n^{++}$ on this manifold generalizes a straight line in Euclidean space, representing the shortest path between two points. Moreover, the SPD manifold is geodesically complete, meaning a unique geodesic can connect any two points. Given $P_1, P_2 \in \mathcal{S}_n^{++}$, the geodesic curve can be defined as $\gamma(t) = \mathbf{P}_2^{\frac{1}{2}}(\mathbf{P}_2^{-\frac{1}{2}}\mathbf{P}_1\mathbf{P}_2^{-\frac{1}{2}})^t\mathbf{P}_2^{\frac{1}{2}}$. Starting from a point P_1 in the direction of vector $V \in T_{P_1}\mathcal{S}_n^{++}$, the geodesic extends until reaching a point P_2, expressed as $\gamma_V(1) = \exp_{P_1}(V) = P_2$. The length of this path defines the Riemannian dis-

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

tance, often calculated using the Log-Euclidean distance: $d_{LE}(P_1, P_2) = \|\log(P_1) - \log(P_2)\|_F$, where $||\cdot||_F$ is the Frobenius norm [29]. An illustration of Riemannian mappings and the construction of a geodesic is shown in Figure 2. The affine-invariant Riemannian distance and Log-Euclidean distance are geodesic distances, described in Table 1 and corresponding to two different Riemannian metrics on this manifold [29].

Table 1. For given $A, B \in S_n^{++}$, commonly used distances. Here, $\|\cdot\|_F$ denotes the matrix Frobenius norm.

Euclidean distance	$d_E(A, B) = \|A - B\|_F$
Affine-invariant Riemannian distance	$d_{AI}(A, B) = \|\log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\|_F$
Log-Euclidean distance	$d_{LE}(A, B) = \|\log(A) - \log(B)\|_F$

1.2.3. Learning with SPD Representations

SPD-Net In the literature has been proposed a Riemannian net to learn SPD matrices, following a standard deep learning strategy [30]. This net assumes Riemannian requirements to preserve the geometric structure of SPD embeddings, while carrying out the processing of information along SPD net layers. These learned SPD embeddings belong to the non-Euclidean structure of the space where the compact descriptors $\{\mathbf{X}\}$ lie [30]. At each layer of the SPD-net, the SPD matrix $\mathbf{X}_{k-1}$ is processed using a bilinear map (*BiMap*) $\mathbf{X}_k = \mathbf{W}_k \mathbf{X}_{k-1} \mathbf{W}_k^\top$, where $\mathbf{X}_{k-1} \in S_{d_{k-1}}^{++}$ represents the SPD matrix of dimension d_{k-1} used in the previous layer. Again, this process results in a dimensionality reduction ($d_k < d_{k-1}$).

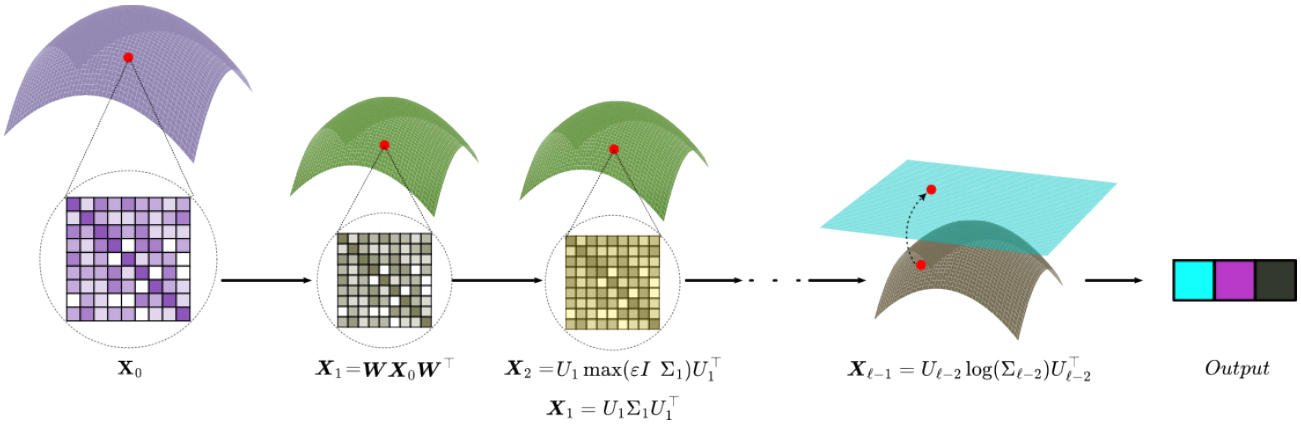
The selection of weights $\mathbf{W}_k \in \mathbb{R}^{d_k \times d_{k-1}}$ is crucial because they must be semi-orthogonal to properly generate an SPD matrix. In addition, these should be maintained in the Stiefel manifolds, where they belong for an appropriate weight upgrading during the learning process [??].

---

[29] Hà Quang Minh and Vittorio Murino. *Covariances in computer vision and machine learning*. Springer, 2018.

[30] Zhiwu Huang and Luc Van Gool. "A riemannian network for spd matrix learning". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017.

Figure 3. The standard architecture for SPD information processing begins with an SPD input projected by Bimap to reduce dimensionality while considering the intrinsic geometry. This is followed by a rectification layer to avoid non-positivity errors. After stacking several *BiRe* blocks, the data is projected back to the Euclidean plane using the logarithm, allowing the application of conventional strategies.



Source: Adapted from [31]

After that, there is a rectification eigenvalue (*ReEig*) layer applied as regularization to SPD matrices from the eigenvalue decomposition $\mathbf{X}_{k-1} = \mathbf{U}_{k-1}\mathbf{\Sigma}_{k-1}\mathbf{U}_{k-1}^\top$. This process computes the maximum between each eigenvalue (element of the diagonal) and a non-negative rectification threshold ε. Hence, to avoid non-positiveness errors, it is defined as the rectification as: $\mathbf{X}_k = \mathbf{U}_{k-1} \max(\varepsilon I, \mathbf{\Sigma}_{k-1})\mathbf{U}_{k-1}^\top$. This *ReEig* is the counterpart of the *ReLU* unit used in the Euclidean layers and avoids non-positiveness errors during training. The *BiMap* and *ReEig* layers constitute a block denoted as *BiRe* block, which can be stacked at the top of another *BiRe* block, and so on. After the k^{th} Riemannian layer, the *LogEig* is introduced to project the SPD matrices onto a Euclidean space, using the logarithmic Riemannian map given by $\mathbf{X}_{k+1} = \log(\mathbf{X}_k) = \mathbf{U}_k \log(\mathbf{\Sigma}_k)\mathbf{U}_k^\top$. Here, the matrix logarithmic function was calculated again over the eigenvalues. The resulting $\mathbf{X}_{k+1}$ allows us to operate using classical neural network layers, which are designed on such Euclidean space.

SPD-CNN Another approach, inspired by classical convolutional frameworks, employs SPD kernels to perform convolution-like operations directly on SPD matrices, enabling the aggrega-

tion of local geometric relationships [32]. For a multi-channel SPD input $\mathbf{X}_{k-1} \in \mathbb{R}^{C \times d_{k-1} \times d_{k-1}}$, where C is the number of channels and d_{k-1} is the matrix dimension, the convolutional SPD module applies C' learnable SPD kernels $\mathbf{W}_k \in \mathbb{R}^{C \times d_k \times d_k}$. Each kernel is constructed from a learnable base matrix $\mathbf{V}$ and made SPD using the formulation $\mathbf{W}_k = \mathbf{V}^\top \mathbf{V} + \epsilon \mathbf{I}_{d_k}$, where ϵ is a small positive scalar that ensures positive definiteness. This layer performs local convolution while preserving the SPD structure, producing an output $\mathbf{X}_k \in \mathbb{R}^{C' \times (d_{k-1}-d_k+1) \times (d_{k-1}-d_k+1)}$, which remains SPD. After convolution, a non-linear activation is applied by normalizing each matrix as $\frac{\mathbf{X}}{\|\mathbf{X}\|_F}$, where $\|\cdot\|_F$ denotes the Frobenius norm, to ensure numerical stability. Element-wise functions such as $\exp(\cdot)$, $\sinh(\cdot)$, $\cosh(\cdot)$, or $ReLU$ are then applied, all of which preserve the SPD property [32] [33]. These steps define a single SPD convolutional block.

1.3. Related works

Oculomotor Parkinson patterns. Ocular movements are a sensitive function to detect neural alterations related to sensory and motor actions during exercises of anticipated target tracking. In particular, Smooth Pursuit Eye Movements (SPEM) refer to controlled and voluntary eye movements that track visual targets moving with predictable velocities and directions [21] [34]. Regarding SPEM, oculomotor PD alterations are associated with saccades pursuit, a reduced ability to suppress undesired eye movements, produced by substantia nigra dysfunction [8]. Additionally, SPEM abnormalities report slower target tracking and delayed initiation, possibly reflecting ocular bradykinesia [11]. For PD, these impairments have been assessed from eye-tracking technologies. Nonetheless, global measure poses limitations on the characterization

---

[32] Tong Zhang et al. "Deep manifold-to-manifold transforming network". In: *2018 25th IEEE international conference on image processing (ICIP)*. IEEE. 2018, pp. 4098–4102.

[33] Tingting Dan et al. "Learning brain dynamics of evolving manifold functional MRI data using geometric-attention neural network". In: *IEEE transactions on medical imaging* 41.10 (2022), pp. 2752–2763.

[34] Christoph Helmchen et al. "Role of anticipation and prediction in smooth pursuit eye movement control in Parkinson's disease". In: *Movement Disorders* 27.8 (2012), pp. 1012–1018.

and discovery of spatio-temporal markers that could be associated with the disease [35] [7]. For instance, such approaches may fail to capture complex behaviors, including unilateral impairments or binocular coordination that has revealed an impaired convergence [10].

More recently, computational approaches have gained relevance for characterizing oculomotor behavior. For example, Brien *et al.* used a video-based setup to record eye movements from 104 PD patients and 106 healthy controls performing pro- and anti-saccade tasks [36]. They extracted features from voluntary saccades, blinks, and fixations to train classifiers such as support vector machines (SVM), logistic regression, and random forests, which showed sensitivity to disease severity. Similarly, Chang *et al.* Following an eye-tracking video setup, recorded SPEM and gaze analyzing pupil trajectories through frequency analysis, and employed an SVM to carry out the classifications [37]. In the same line, Jansson *et al.* proposed a mathematical model based on probability density functions to detect anomalies associated with PD SPEM [38]. However, the knowledge based on machine learning incurs on a dependency of sufficient amount of data to fix learning functions. Besides, the feature based on coarse trajectories features without considering potential relationships may underestimate PD descriptors and associations with disease that could offer new insights into PD during SPEM tasks.

Geometric SPD attention mechanisms. Encoding high-dimensional representations into compact geometric descriptors is prominent in computer vision. Unlike classical CNNs that

---

[35] Chrystalina A Antoniades and Miriam Spering. "Eye movements in Parkinson's disease: from neurophysiological mechanisms to diagnostic tools". In: *Trends in Neurosciences* 47.1 (2024), pp. 71–83.

[36] Donald C Brien et al. "Classification and staging of Parkinson's disease using video-based eye tracking". In: *Parkinsonism & Related Disorders* 110 (2023), p. 105316.

[37] Zhuoqing Chang et al. "Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and machine learning". In: *Scientific reports* 10.1 (2020), p. 18641.

[38] Daniel Jansson et al. "Stochastic anomaly detection in eye-tracking data for quantification of motor symptoms in Parkinson's disease". In: *Signal and Image Analysis for Biomedical and Life Sciences* (2015), pp. 63–82.

focus on local information, geometric methods can measure the geodesic affinities of these representations to uncover metric patterns of data. [23]. Geometric learning techniques based on the Riemannian manifold of SPD matrices adjust representations from non-Euclidean data by progressively compressing information within the geometric space [??] [32].

Recently, attention mechanisms over Riemannian manifolds have gained relevance to enhance the learning of covariate patterns. Particularly, Pan. *et al.* compute covariance matrices of temporal embedding signals, which in turn are projected onto a lower-dimensional manifold using bilinear mapping to obtain query, key, and value representations. The geometric space is leveraged to measure the similarity between query and key using the Log-Euclidean distance. The resulting similarity matrix is weighted via a row Softmax, and the value projections are weighted accordingly to decode spatiotemporal features and track brain dynamics [39]. In EEG signal processing, SPD matrices are constructed from spatiotemporal convolutional outputs and then mapped to produce query, key, and value representations. In such case, Log-Euclidean distances are also used to compute similarity between queries and keys, building a similarity matrix of Riemannian distances [40]. As an alternative to bilinear mapping, a geometric attention block was introduced following SPD kernels that capture local relationship patterns. This approach combines geometric attention layers with convolutional SPD layers, enabling learning from multichannel inputs and compact, lower-dimensional manifold representations, serving as an alternative to traditional manifold learning while preventing degeneration [33]. Also, in vision transformers, after multi-head attention projects sequences into geometric domains such as Euclidean space, SPD, and Grassmannian manifolds. Self-attention similarity distances at each space are concatenated and weighted, creating a comprehensive representation that captures intricate relationships and leverages geometric and statistical properties for image classification

---

[39] Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. "MAtt: A manifold attention network for EEG decoding". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 31116–31129.

[40] Bin Lu et al. "Manifold attention-enhanced multi-domain convolutional network for decoding motor imagery intention". In: *Knowledge-Based Systems* 296 (2024), p. 111904.

[41]. Similarly, Ma *et al.* proposed a geometric architecture to analyze Alzheimer's disease that relies on two-modal interactions and manifold learning, using bilinear mapping to project both modalities onto the same manifold. Log-Euclidean operations are employed to exploit SPD manifold properties, with the cross-attention mechanism computing log-Euclidean distances between modalities to measure similarity while preserving the manifold's geometric structure [42]. Although attention mechanisms effectively enhance discriminative patterns, adapting them to SPD contexts while maintaining geometric integrity remains challenging. These approaches often depend on global measures, potentially overlooking the nuanced, latent information embedded within the matrices.

---

[41] Dimitrios Konstantinidis et al. "Multi-manifold attention for vision transformers". In: *IEEE Access* (2023).

[42] Junbo Ma and et al. "Multimodality Alzheimer's disease analysis in deep Riemannian manifold". In: *Information Processing & Management* (2022).

2. RESEARCH PROBLEM

Parkinson's is the second most prevalent neurological disorder worldwide, and today, there is no definitive method for diagnosis and effective management of this disease. Currently, characterizing oculomotor patterns has revealed essential traits for early Parkinson's diagnosis, potentially serving as a disease biomarker. However, clinical assessments require invasive devices that need rigorous calibration. Alternatively, learning spatiotemporal patterns from deep learning representations has evidenced potential capabilities to find and discriminate abnormal locomotor patterns in diverse domains and applications. Despite promising results, the ability of these models to detect subtle and meaningful patterns within eye movement tracking data may be restricted to training conditions with a huge amount of labeled data to adjust representations, a common limitation in clinical scenarios. More recently, deep geometric learning has emerged as a promising alternative for data-sparse scenarios, being, among others, invariant to outlier artifacts and robust to noisy data. Nonetheless, these representations report yet limitations to learning deep representations and even highlighting key relationships that may support the discrimination of spatiotemporal between Parkinsonian and control observations.

2.1. Research questions

How to design a geometrical attention mechanism to learn key relationships that contribute to the classification of Parkinsonian oculomotor patterns?

2.2. Hypothesis

Designing and learning attention mechanisms to stand out second-order relationships can improve detection and characterization of oculomotor patterns associated with Parkinson's disease.

3. OBJECTIVES

General objective

- To propose a geometrical attention mechanism to classify parkinsonian oculomotor patterns from SPEM oculomotor patterns coded in symmetric positive definite matrices (SPD).

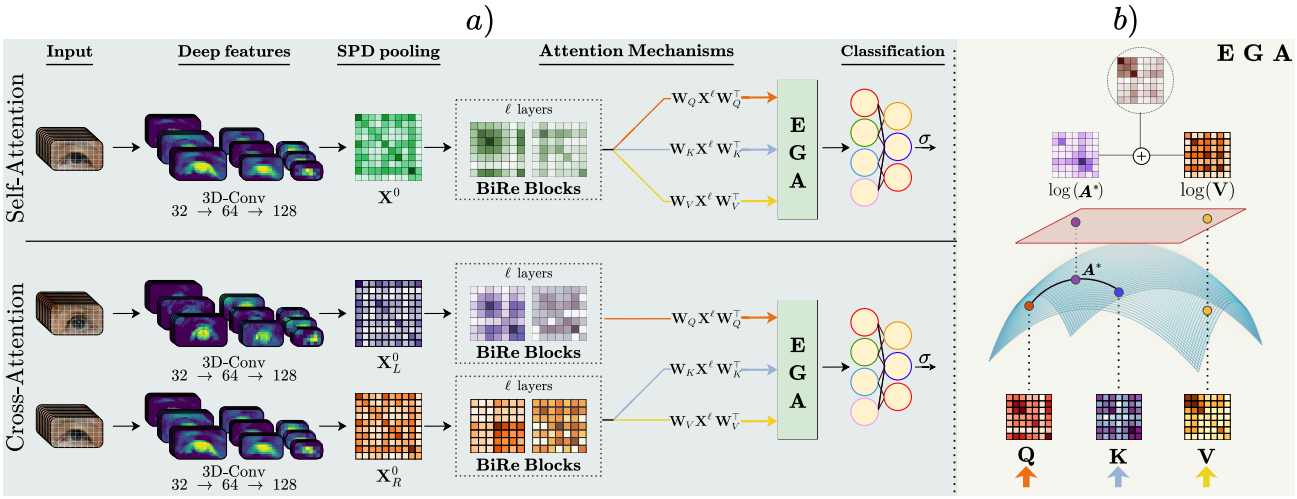
Specific objectives

- To define a dataset with markerless videos that record oculomotor tasks for a population diagnosed with Parkinson and a control population.
- To design a geometrical attention mechanism to learn and stand out covariate patterns from SPEM video observations
- To develop a deep representation that integrates an attention mechanism to represent spatiotemporal patterns from SPD matrices
- To validate the proposed architecture regarding the capability to classify Parkinsonian versus control oculomotor patterns.

4. PROPOSED APPROACH

In this work, we designed an attention mechanism strategy that leverages intrinsic geometry to highlight relationships from SPD matrices that encode spatio-temporal disease-related patterns using non-intrusive SPEM video recordings. Each video sequence forms a convolutional bank of spatiotemporal features to capture multi-scale and multi-term spatio-temporal patterns. Next, the information is summarized into a single SPD matrix that encodes second-order information and is subsequently fed to a Riemannian network to generate SPD geometrical embeddings, preserving the structure of the Riemann manifold. Finally, the geometric attention mechanisms are used to highlight disease-related patterns and carry out the classification. Figure 4 summarizes the proposed pipeline.

Figure 4. In **(a)**, the general pipeline of the proposed approach is shown. First, a bank of deep representations is generated from an input smooth pursuit eye movement (SPEM) video, capturing the spatio-temporal patterns. Then, an SPD matrix is calculated to summarize the relationships between the volumetric deep representations. The subsequent step involves the implementation of a Riemannian module for SPD learning, and the proposed Enhanced Geodesic Attention (EGA), which allows to highlight relevant covariances being illustrated in **(b)**. The scheme is trained in an end-to-end manner.



Source: Author's own work

4.1. Dataset

The study included 50 participants: 25 control subjects (average age 73.6 ± 6.3 years) and 25 individuals diagnosed with Parkinson's disease (average age 71.6 ± 9.7 years). PD patients were categorized by disease progression using the Hoehn-Yahr scale: Five patients were at stage 2.5, six at stage 3, two at stage 4, and twelve without a reported scale.The data collection process involved a video recording of participants' eye regions while performing an SPEM task.

Figure 5. A non-invasive capture configuration is depicted, showcasing a patient engaged in the task of smooth pursuit tracking a stimulus projected onto a screen.

Source: Author's own work

First, participants were invited to stand in front of a screen positioned at a distance of 1 m and at the same height as the eyes. On this screen, a white dot (stimulus) was projected, performing a continuous smooth movement over a 15-second duration, first 8 seconds, horizontal movements were performed, followed by 7 seconds of vertical movements. During this SPEM task, participants were instructed to follow the stimulus with their gaze, while their faces were recorded by a camera with a resolution of 60 frames per second (FPS) positioned below the screen level. After video capture, a template-matching technique was used to properly crop the eye region, resulting in a square area of 128×128 for each eye and each video sequence. Consequently, two video volumes were generated per participant, one for each eye. Written informed consent was obtained from all participants.

Table 2. Dataset Description A total of 50 participants, including 25 individuals diagnosed with Parkinson's disease (PD) and 25 healthy controls. Demographic variables include age, gender, and disease stage according to the Hoehn–Yahr scale.

Variable	Description	Stats
Participants	Total sample size	50 subjects (25 PD / 25 Controls)
Age (years)	mean ± std	71.6 ± 9.7 (PD) 73.6 ± 6.3 (Ctrl)
Gender	(F / M / Unknown)	(8 / 17 / 0) (PD) (16 / 7 / 2) (Ctrl)
Hoehn–Yahr stage	(2.5 / 3 / 4 / Unknown)	(5 / 6 / 2 / 12)

4.2. 3D Spatio-temporal convolutions to SPD representation

PD expresses multifactorial symptoms with asymmetric manifestation and is highly diverse in magnitude across the affected population [43]. Considering such fact, this work utilizes a 3D convolutional network (3D-Conv) for feature extraction, consisting of stacked 3D convolutional layers with kernels that consider temporal and spatial dimensions, along with 3D max-pooling. For an input video $\mathbf{I} \in \mathbb{R}^{C \times T \times W \times H}$, whether from the left (L) or the right (R) eye, and where C is the number of channels, T is the temporal dimension, W is the width, and H is the height, the 3D-Conv hierarchically reduces $\mathbf{I}$ into a bank of spatio-temporal deep features $\mathbf{F} = ENC(\mathbf{I}) \in \mathbb{R}^{N \times T' \times W' \times H'}$, where N is the number of features, while T', W', and H' are the temporal and spatial resulted dimensions respectively ($T > T', W > W'$, and $H > H'$). Here, $\mathbf{F} = \{\mathbf{F}^{(1)}, \ldots, \mathbf{F}^{(N)}\}$ decomposes the input eye video sequence information to build a deep representation that captures several magnitudes and discrete patterns for oculomotor abnormalities.

4.2.1. Geometric SPD learning The resulting bank of features $\mathbf{F}$ is structured as a tensor that exhibits highly informative representations with some redundancy about eye dynamics. Given this, compact descriptors were computed from the feature bank $\mathbf{F}$. For this, we first

[43] RA Armstrong. "Oculo-visual dysfunction in Parkinson's disease". In: *Journal of Parkinson's disease* 5.4 (2015), pp. 715–726.

reshaped $\mathbf{F}$ into a matrix $\mathbf{M} \in \mathbb{R}^{N \times d}$, where each row corresponds to a vectorized feature map, then each feature resulting in a vector of dimension $d = T' \times W' \times H'$.

From here, we compute $\mathbf{X}^0 = \frac{1}{d}\mathbf{M}\mathbf{M}^\top$, where $\mathbf{X}^0[i,j]$ represents the similarity between the $\mathbf{F}^{(i)}$-th and $\mathbf{F}^{(j)}$-th features. Here, $\mathbf{X}^0$ forms a symmetric and positive definite (SPD) matrix of dimension $N \times N$, and represents a point on a Riemannian manifold. This reduction of features into an SPD matrix is referred to as SPD pooling (see Figure 4). Subsequently, a specialized SPD learning module is employed to learn discriminative similarities and lower-dimensional representations while preserving the geometric structure of SPD descriptors. Specifically, we used the deep Riemannian SPD network (SPD-NET) modules [??], starting with a Bilinear Mapping (*BiMap*) layer to process the SPD descriptors as:

$$\mathbf{X}^\ell = \left(\mathbf{W}^\ell\right) \mathbf{X}^{\ell-1} \left(\mathbf{W}^\ell\right)^\top, \tag{2}$$

where $\mathbf{X}^{\ell-1} \in \mathcal{S}^{++}_{d_{\ell-1}}$ is the SPD matrix from the previous layer $(\ell-1)$, and $\mathbf{W}^\ell \in \mathbb{R}^{d_\ell \times d_{\ell-1}}$ are learnable weights on the Stiefel manifold that allow reducing the dimensionality $d_\ell < d_{\ell-1}$, while the resulting matrix $\mathbf{X}^\ell$ remains SPD [??]. To prevent non-positive errors, the *ReEig* layer applies a non-linear transformation using spectral decomposition:

$$\mathbf{X}^\ell = \left(\mathbf{U}^{\ell-1}\right) \max(\varepsilon I, \mathbf{\Sigma}^{\ell-1}) \left(\mathbf{U}^{\ell-1}\right)^\top, \tag{3}$$

where $\mathbf{U}^{\ell-1}$ are the eigenvectors, and $\mathbf{\Sigma}^{\ell-1}$ a diagonal matrix with eigenvalues. Both layers (*BiMap* and *ReEig*) can be sequentially stacked to form a *BiRe* block. After ℓ-*BiRe* blocks, the *LogEig* layer is finally used at the end to project the SPD matrices onto the identity tangent space $T_I\mathcal{S}^{++}_\ell$, using the logarithmic Riemannian map given by

$$\mathbf{X}^\ell = \log(\mathbf{X}^{\ell-1}) = \left(\mathbf{U}^{\ell-1}\right) \log\left(\mathbf{\Sigma}^{\ell-1}\right) \left(\mathbf{U}^{\ell-1}\right)^\top, \tag{4}$$

which has the structure of Euclidean Space.

4.2.2. Riemannian Attention Mechanisms Traditional attention mechanisms capture local relational patterns within a global context. Recently, these methods have been extended to geometric-based approaches, designing attention modules into non-Euclidean learning schemes such as the Riemannian manifold. In this case, attention aims to extract compact, discriminative features from SPD representations while preserving their intrinsic geometric structure. Previous works introduced geometric attention mechanisms that compute similarity measure between Key ($\mathbf{K}$) and Query ($\mathbf{Q}$) using Riemannian distances [39]. This yields a global scalar value, equivalent to the geodesic length between $\mathbf{K}$ and $\mathbf{Q}$ [42], which may oversimplify the complex relationships encoded in SPD matrices. Typically, these methods rely on projected SPD matrices, reducing attention to a single scalar distance and limiting the learning of richer descriptors. To address this limitation, we propose the *EGA* module, a Riemannian attention mechanism that highlights relevant components through local-geometric projections in the spectral domain of SPD descriptors. Both self- and cross-attention mechanisms are explored, as detailed in the following subsections.

Riemannian Self-Attention. In this work, from the initial SPD descriptor $\mathbf{X}^0$ are computed three *BiRe* projections to encode a geometric self-attention. These three projections are named the Key ($\mathbf{K}$), the Query ($\mathbf{Q}$), and the value ($\mathbf{V}$) coded as:

$$(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{W}_{\{Q,K,V\}} \mathbf{X}^{\ell} \mathbf{W}_{\{Q,K,V\}}^{\top} \tag{5}$$

The weights $\mathbf{W}_K, \mathbf{W}_Q, \mathbf{W}_V$ are selected so that the query, key, and value SPD matrices lie on the same manifold, *i.e*, have the same dimension. To address global scalar measures, reported in the literature [42], in this work we computed the midpoint of the geodesic (γ) joining $\mathbf{K}$ and $\mathbf{Q}$, to achieve a unified geometric representation that integrates information from $\mathbf{K}$ and $\mathbf{Q}$ on the manifold, preserving the SPD matrix structure rather than reducing it to a single scalar.

Explicitly, this matrix is calculated as:

$$\mathbf{A} = \gamma\left(\frac{1}{2}\right) = \mathbf{Q}^{\frac{1}{2}}(\mathbf{Q}^{-\frac{1}{2}}\mathbf{K}\mathbf{Q}^{-\frac{1}{2}})^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}. \tag{6}$$

From this, we applied a Softmax function to the eigenvalues of $\mathbf{A}$, to normalize the principal directions of information scores from its spectral decomposition, emphasizing the most relevant components. The guarantee of SPD structure preservation is formally defined as:

Proposition 1. *Let* $\mathbf{A} \in S_{d_\ell}^{++}$ *with spectral decomposition* $\boldsymbol{A} = \boldsymbol{U}\text{diag}(\boldsymbol{\Sigma})\,\boldsymbol{U}^\top$, *where* $\boldsymbol{\Sigma} = (\lambda_1, \ldots, \lambda_{d_\ell})^\top$. *Define* $\mathbf{A}^* = \mathbf{U}\,\text{diag}(\text{Softmax}(\boldsymbol{\Sigma}))\,\mathbf{U}^\top$. *Then,* $\mathbf{A}^* \in S_{d_\ell}^{++}$.

Proof. Since $\mathbf{A}$ is an SPD matrix ($\in S_{d_\ell}^{++}$), their eigenvalues are all positive, i.e., $\lambda_i > 0$ for all i. Then, after applying the Softmax function to the eigenvalues $\text{Softmax}(\lambda_1, \lambda_2, \ldots, \lambda_{d_\ell})$ yields a diagonal matrix $\boldsymbol{\Sigma}^* = \text{diag}(\lambda_1^*, \lambda_2^*, \ldots, \lambda_{d_\ell}^*)$, where $\lambda_i^* = \frac{e^{\lambda_i}}{\sum_{j=1}^{d_\ell} e^{\lambda_j}} > 0$. Therefore, the transformed matrix $\mathbf{A}^* = \mathbf{U}\boldsymbol{\Sigma}^*\mathbf{U}^\top$ remains symmetric and positive. Since $\mathbf{U}$ is an orthogonal matrix, it preserves symmetry in $\boldsymbol{\Sigma}^*$, and the softmax-transformed eigenvalues will correspond to the eigenvalues of $\mathbf{A}^*$. Thus, as these values are positive, the operation $\mathbf{A}^* = \text{Softmax}(\mathbf{A})$ yields a new SPD matrix. □

The calculated $\mathbf{A}^* = \text{Softmax}(\mathbf{A})$ entails the proposed geometric attention matrix. It highlights the relative importance of each eigenvalue by emphasizing dominant components of the spectral decomposition along the most discriminative directions. This is achieved through the Softmax operation, which normalizes their contributions and amplifies relative differences. As a result, the eigenvalues form a probability distribution that prioritizes those capturing large variance components, often associated with the most relevant SPD patterns. In addition, geometrically, attention matrix $\mathbf{A}^*$ is learned to be a new element in SPD manifold, leveraging the latent space of queries and keys SPD descriptors. Finally, we project this matrix together with $\mathbf{V}$ into an Euclidean space using the Riemannian logarithm (*LogEig*) layer, in order to fuse attention and value SPD representations in an additive Euclidean form. That is:

$$\mathrm{EGA}(\mathbf{X}^{\ell}) = \log(\mathbf{A}^{*}) + \log(\mathbf{V}) \tag{7}$$

The resulting matrix $\mathbf{X}^{\ell+1} = \mathrm{EGA}(\mathbf{X}^{\ell})$ is an element in the tangent space of the Riemannian manifold that geometrically weights key and query while recalibrating the value descriptor information.

Riemannian Cross-Attention Attention mechanisms have been applied to integrate complementary information from multiple sources, enriching learned representations [44]. In the context of PD, asymmetric motor symptoms reflect uneven cerebral degeneration. As the disease progresses, this asymmetry may diminish, indicating varying degeneration rates [45]. Particularly, in the early stages, oculomotor alterations are frequently observed, including often impaired binocular coordination during SPEM or gaze fixation, as well as deficits in convergence [10].
To capture such asymmetries or binocular impairments during SPEM, we propose to integrate information from the left (L) and right (R) eye video inputs using SPD representations and a cross-attention module. For this purpose, each input video, left ($\mathbf{I}_L$) and right ($\mathbf{I}_R$) is processed through an encoder, producing feature banks for each eye: $\mathbf{F}_L = ENC_L(\mathbf{I}_L) = \{\mathbf{F}_L^{(1)}, \cdots, \mathbf{F}_L^{(N)}\}$ and $\mathbf{F}_R = ENC_R(\mathbf{I}_R) = \{\mathbf{F}_R^{(1)}, \cdots, \mathbf{F}_R^{(N)}\}$. Then, each bank of features is summarized into an SPD descriptor $\mathbf{X}_L^0$ and $\mathbf{X}_R^0$, respectively, each one summarizing the spatio-temporal deep representations of each eye. After individually processing of each SPD eye descriptor through *BiRe* blocks, the resulting lower-dimensional left-eye SPD matrix ($\mathbf{X}_L^{\ell}$) is projected to calculate the SPD query

---

[44] Vandana Rajan, Alessio Brutti, and Andrea Cavallaro. "Is cross-attention preferable to self-attention for multi-modal emotion recognition?" In: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2022, pp. 4693–4697.

[45] P Riederer and J Sian-Hülsmann. "The significance of neuronal lateralisation in Parkinson's disease". In: *Journal of Neural Transmission* 119 (2012), pp. 953–962.

$$\mathbf{Q}_L = \mathbf{W}_{L,Q}\mathbf{X}_L^{\ell}\mathbf{W}_{L,Q}^{\top}, \tag{8}$$

and the Key and Value SPD descriptors are computed from the right eye SPD matrix ($\mathbf{X}_R^{\ell}$) as

$$\mathbf{K}_R = \mathbf{W}_{R,K}\mathbf{X}_R^{\ell}\mathbf{W}_{R,K}^{\top}, \quad \mathbf{V}_R = \mathbf{W}_{R,V}\mathbf{X}_R^{\ell}\mathbf{W}_{R,V}^{\top}. \tag{9}$$

All projection weights $\mathbf{W}_*$ are constrained to have the same dimensions to ensure compatibility on the same SPD manifold. Based on this, the cross-attention from the left to the right SPD matrix is computed as:

$$\mathrm{EGA}(\mathbf{X}_L^{\ell}, \mathbf{X}_R^{\ell}) = \log(\mathbf{A}^*) + \log(\mathbf{V}_R) \tag{10}$$

The calculated geometric attention matrix $\mathbf{A}^*$ results from the matrix $\mathbf{A}$, the midpoint of the geodesic connecting $\mathbf{Q}_L$ with $\mathbf{K}_R$. Similarly to the EGA self-attention configuration (Equation 7), the sum of the attention matrix and value is performed in Euclidean space using the *LogEig* layer, fusing cross-modal information. This operation integrates relevant relationships between both eyes, facilitating the learning of oculomotor abnormalities and laterality differences.

Multihead Riemannian Attention Attention mechanisms have proven effective in enhancing convolutional architectures by focusing on spatial and channel-based features, allowing the model to emphasize the most relevant image regions [24]. Among configurations, in both self- and cross-attention, additional relationships can be exploited by computing multiple mechanisms at the same processing level, where independent heads capture diverse interrelationships on the SPD manifold. So we can be further extended to a multi-head framework.

In this setup, multiple attention heads are defined by learning distinct sets of projection weights ($\mathbf{W}_Q^h, \mathbf{W}_K^h, \mathbf{W}_V^h$), each producing SPD queries, keys, and values ($\mathbf{Q}^h, \mathbf{K}^h, \mathbf{V}^h$). Each head computes its geodesic-based matrix $\mathbf{A}^h$ as Equation 6, and output $\mathrm{EGA}_h(\mathbf{X}_\ell) = \log(\mathbf{A}^{h^*}) + \log(\mathbf{V}^h)$. These outputs are concatenated in Euclidean space $\big[\mathrm{EGA}_1(\mathbf{X}^{\ell}), \mathrm{EGA}_2(\mathbf{X}^{\ell}), \ldots, \mathrm{EGA}_h(\mathbf{X}^{\ell})\big]$, enabling the attention mechanisms to capture diverse geometric representations.

5. EXPERIMENTAL SETUP

The proposed solution was validated across several configurations and thoroughly compared with state-of-the-art solutions. Both analyses supported the performance of the proposed approach and its respective components. The next subsection summarizes the above approach configurations and the baseline methods considered during evaluation.

5.1. Implementations Details and Configurations

The main components of the proposed approach (EGA-Net) were validated according to several configurations, detailed as follows:

3D convolutional backbone This component was coded as a 3D-CNN backbone with three convolutional layers (with kernels, $k = 3 \times 3 \times 3$), resulting in 128 deep features $37 \times 16 \times 16$. Then, an SPD pooling is operated to recover an SPD matrix of dimension 128×128.

Geometric attention From the SPD pooling, we stacked $BiRe_i$ blocks ($i \in 0, 1, 2, 3$) to obtain SPD matrices of size 64×64, 32×32, and 16×16, respectively.
Independent *BiRe* blocks are then generated the $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ branches, producing matrices of 32×32, 16×16, and 8×8 for $i = 1, 2, 3$. The case $i = 0$ corresponded to no *BiRe* blocks, allowing evaluation of the attention mechanism under direct dimensional reductions (e.g., 64×64, 32×32, 16×16) instead of progressive compression. The resulting attention matrix $\mathbf{A}^*$ was projected onto the Euclidean space (with the same dimension as the attention output). We also experimented with stacked attention blocks using multiple heads ($H \in 1, 2, 3, 4$) to assess the benefits of parallel processing. Both self- and cross-geometric attention mechanisms were evaluated under these configurations.

Classification A dense layer of 128 was followed by a single output neuron with a sigmoid activation function. Training used a fixed learning rate of $\alpha_1 = 10^{-2}$ for the SPD modules and $\alpha_2 = 10^{-3}$ for the convolutional backbone, with binary cross-entropy loss and the Adam optimizer. Model validation employed a 5-fold cross-validation scheme, with each fold including five control patients and five with Parkinson's disease. Performance was evaluated by averaging accuracy, recall, precision, specificity, F1-score, and AUC-ROC across folds.

5.2. Baseline Analysis

For baseline, we included convolutional, geometric, and attention-based approaches, described as follows:

5.2.1. Convolutional strategies

- *3D-CNN* [46]. Configured with 32 volumetric features in the first layer, 64 in the second, and 128 in the final layer.
- *ConvNeXt* [47]. For this, a weight inflation was implemented to adapt the network for 3D video input [48]. Each volumetric input was then processed through three ConvNeXt layers, producing 96, 192, and 356 features. Each layer employed $7 \times 7 \times 7$ kernels, residual connections, layer normalization, and fully connected layers with GELU activation.

---

[46] Roman Solovyev, Alexandr A Kalinin, and Tatiana Gabruseva. "3D convolutional neural networks for stalled brain capillary detection". In: *Computers in Biology and Medicine* 141 (2022), p. 105089. DOI: 10.1016/j.compbiomed.2021.105089.

[47] Zhuang Liu et al. "A convnet for the 2020s". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 11976–11986.

[48] Joao Carreira and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6299–6308.

- *ViViT* [49]. The vision video transformer divides each video into non-overlapping patches of size $3 \times 8 \times 8$, with 3 channels and an 8×8 spatial dimension. These patches are projected into 128-dimensional embeddings. The transformer employs multi-head self-attention, with three heads across four stacked layers.

5.2.2. SPD Geometrical models Accordingly, we employ 3D-CNN as a feature extractor to further integrating SPD-based mechanisms.

- *SPD-NET* [??]. This approach from CNN bank of features, employs two *BiRe* blocks (64×64, 32×32). A rectification threshold $\epsilon = 10^{-3}$ was utilized for the *ReEig* layer. The *LogEig* layer was employed to project them onto the tangent space, without attention.

- SPD-CNN [32]. This model employs SPD kernels, starting with an SPD input $\mathbf{X}^0$, followed by a non-linear activation that normalizes each matrix as $\frac{\mathbf{X}_*}{|\mathbf{X}_*|_F}$, where $|\cdot|_F$ denotes the Frobenius norm, to prevent numerical degeneration. An element-wise exponential activation is then applied. We implemented three convolutional modules with kernels of size 15×15, 7×7, and 5×5. The resulting SPD output corresponds to the upper-triangular vectorization of a 67×67 matrix, yielding $\frac{67(67+1)}{2}$ dimensions. This representation is subsequently projected onto the tangent space through a *LogEig* layer.

geometric attention methods

- Geometric Attention Mechanism (GAM) [33]: Three consecutive SPD-CNN layers with kernels of size 15×15 and 7×7 produce an output matrix of size 74×74. The final block refines the representation through SPD convolution, normalization, and an additional attention layer. The resulting upper-triangular matrix is then flattened and projected through a dense layer of dimension $\frac{62(63+1)}{2}$.

---

[49] Anurag Arnab et al. "Vivit: A video vision transformer". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 6836–6846.

- Cross Modal Riemannian Network (CMRN) [42]: The model employed scalar-based geometric attention with three parallel branches through *BiRe* layers, producing attention matrices of size 64×64. For the left-eye input $\mathbf{X}_L^\ell$, it generates $\mathbf{Q}_L, \mathbf{K}_L, \mathbf{V}_L$, and for the right-eye input $\mathbf{X}_R^\ell$, it produces $\mathbf{Q}_R, \mathbf{K}_R, \mathbf{V}_R$, each of the dimension 32×32. A similarity measure is computed using the Frobenius norm between the query of one eye and the key of the other. The resulting scalar values are rescaled via $\text{Softmax}(\lambda_L, \lambda_R) = \lambda_L^*, \lambda_R^*$ and used to weight the value representations in the Euclidean tangent space as $\lambda_L^* \cdot \log(\mathbf{V}_R) + \lambda_R^* \cdot \log(\mathbf{V}_L)$. This output is vectorized by extracting its upper-triangular part and passed through a dense layer with 128 units, followed by a final unit.

6. EVALUATION AND RESULTS

6.1. An Ablation component analysis

As part of the component analysis in this work, we evaluated the effect of geometrical coding before the EGA mechanism, as well as the number of attention heads required to support classification. Firstly, we measured the impact of using a different number of *BiRe* blocks before EGA. Figure 6 shows how progressively stacking *BiRe* blocks before the EGA enhances the discriminative capability. Commonly, approaches using these geometric components (*BiRe* blocks) do not benefit from stacking multiple blocks [50]. In this sense, our results suggest that EGA leverages compact geometric representations in deeper SPD networks, enhancing the discriminative performance. Also, the output dimension of 1-*BiRe* depth networks (32x32, 16x61, and 8x8) and 2-BiRe depth networks (16x16 and 8x8) was varied, demonstrating that the 3-*BiRe* depth network effectively achieves the highest performance.

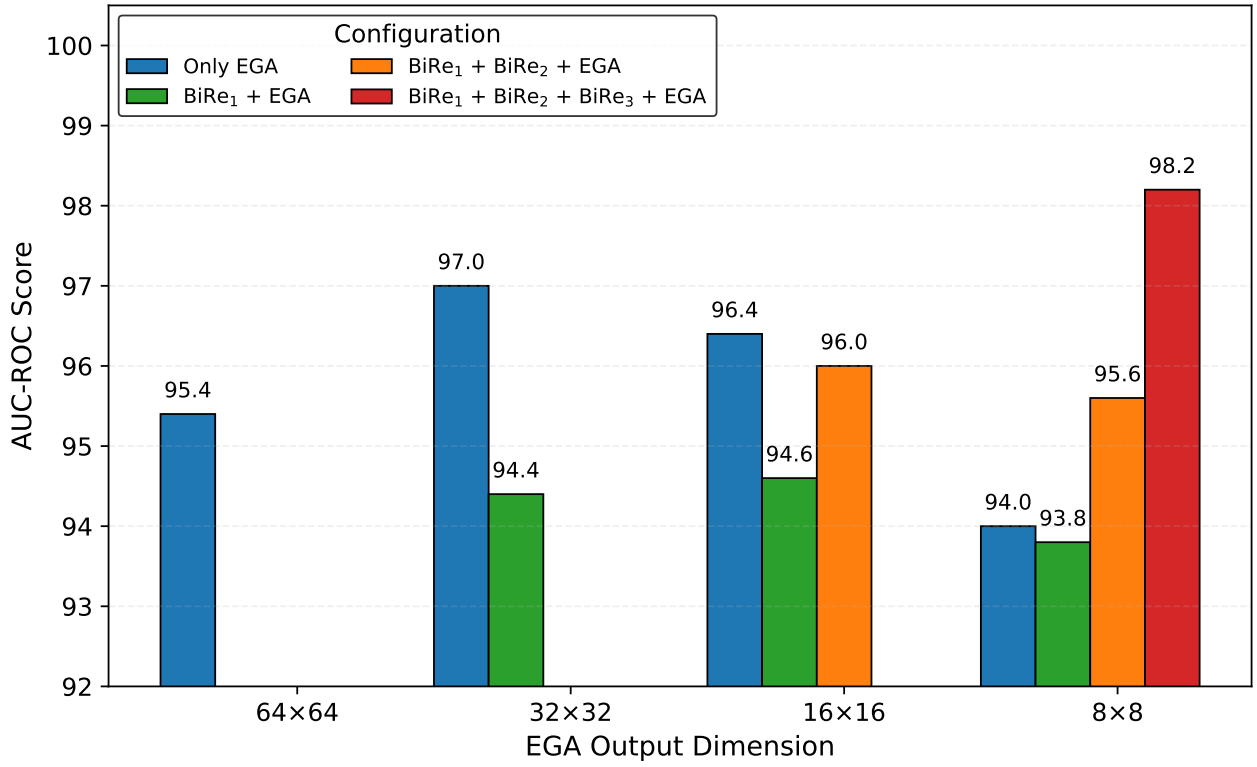
Secondly, the method was evaluated by configuring the proposed EGA with multiple attention mechanisms operating in parallel, that is, by incorporating multiple attention heads. Table 3 reports the performance with different numbers of heads, evidencing that two heads were sufficient to achieve the best AUC-ROC score. With four heads, we observed a similar AUC-ROC along with notably superior accuracy and recall scores. In consequence, this configuration with four heads was selected for the following EGA self-attention experiments.

A similar analysis was conducted for the EGA cross-attention configuration, where the query was derived from the right eye and the key and value from the left (R2L). As shown in Table 4, the best performance was achieved with a single head (an AUC-ROC score of 98.4 $\pm$ 3.2). Increasing the number of heads raised the number of learnable parameters but did not improve

---

[50] Juan Olmos, Antoine Manzanera, and Fabio Martínez. "Riemannian SPD learning to represent and characterize fixational oculomotor Parkinsonian abnormalities". In: *Pattern Recognition Letters* 177 (2024), pp. 157–163.

Figure 6. AUC-ROC performance across different output dimensions after EGA and model configurations progressively considering more *BiRe* blocks before EGA.



Source: Author's own work

Table 3. Ablation study of the number of heads over the proposed approach EGA self-attention

# of heads	*Accuracy*	*Recall*	*AUC*
1	88.0 ± 10.3	96.0 ± 8.0	94.2 ± 4.7
2	90.0 ± 5.5	92.0 ± 7.5	**97.0 ± 3.5**
3	89.0 ± 7.3	94.0 ± 4.9	96.6 ± 3.4
4	**93.0 ± 7.5**	**96.0 ± 4.9**	97.0 ± 4.0

performance. Hence, we selected this configuration with a single head for the following EGA cross-attention experiments.

Table 4. Results of the ablation study for the number of heads using EGA cross-attention R2L configuration.

# of heads	*Accuracy*	*Recall*	*AUC-ROC*
1	92.0 ± 4.0	**96.0 ± 8.0**	**98.4 ± 3.2**
2	88.0 ± 7.5	92.0 ± 9.8	96.0 ± 5.1
3	90.0 ± 8.9	96.0 ± 8.0	97.6 ± 4.8
4	**94.0 ± 8.0**	96.0 ± 8.0	95.2 ± 7.8

6.2. State-of-the-art Comparisons

This work conducted an exhaustive comparison with approaches dedicated to spatio-temporal input analysis and geometrical approximations. As an initial analysis of standard video-based methods, we considered ConvNext, ViViT, and 3D-CNN.

Table 5 reports the results of these three approaches compared to the proposed geometrical representation. As observed, the *ViViT*, an attention-based architecture, shows clear limitations in distinguishing patients with Parkinson's disease, with a low recall score of 60.0 ± 22.8. In contrast, ConvNeXt and 3D-CNN achieved more consistent results, with 3D-CNN offering a slight advantage in AUC-ROC, scoring 92.8 ± 9.8. The suboptimal performance of *ViViT* likely indicates its challenges in capturing temporal dynamics. From the experimental setup of this work. The proposed EGA demonstrated that employing an attention mechanism applied to low-dimensional representations, which capture spatio-temporal patterns, resulted in

Table 5. Comparison with spatio-temporal architectures. Three state-of-the-art methods were evaluated for comparison. The highest scores are highlighted in **bold**.

Model	*Accuracy*	*Recall*	*Precision*	*Specificity*	*F1*	*AUC-ROC*
ConvNext	82.0 ± 10.3	**86.0 ± 4.9**	81.0 ± 13.1	78.0 ± 16.0	83.2 ± 9.0	87.4 ± 9.2
ViViT	68.0 ± 9.3	60.0 ± 22.8	80.1 ± 17.9	76.0 ± 30.1	63.6 ± 12.4	76.0 ± 15.3
3D-CNN	83.0 ± 12.5	82.0 ± 11.7	85.0 ± 15.5	84.0 ± 16.2	83.1 ± 12.3	92.8 ± 9.8
EGA - *Self Attention (Proposed)*	**93.0 ± 7.5**	**96.0 ± 4.9**	**91.7 ± 10.5**	**90.0 ± 12.6**	**93.5 ± 6.8**	**97.0 ± 4.0**

Table 6. Comparison with SPD and geometric attention-based architectures. Three state-of-the-art methods were evaluated for comparison. The highest scores are highlighted in **bold**.

Model	*Accuracy*	*Recall*	*Precision*	*Specificity*	*F1*	*AUC-ROC*
SPD-NET	84.0 ± 17.7	94.0 ± 12.0	84.8 ± 18.5	74.0 ± 37.7	87.0 ± 11.7	93.6 ± 11.8
SPD-CNN	83.0 ± 6.8	80.0 ± 16.7	89.1 ± 12.2	86.0 ± 18.5	82.0 ± 8.4	93.4 ± 4.5
GAM	78.0 ± 15.4	68.0 ± 22.3	83.5 ± 18.3	88.0 ± 11.7	74.2 ± 19.9	92.4 ± 9.8
EGA - *Self Attention (Proposed)*	**93.0 ± 7.5**	**96.0 ± 4.9**	**91.7 ± 10.5**	**90.0 ± 12.6**	**93.5 ± 6.8**	**97.0 ± 4.0**

enhanced performance, surpassing the ConvNext, ViViT, and 3D-CNN methods. Notably, an improvement of +4.2% in AUC-ROC over the 3D-CNN was observed.

In a second state-of-the-art analysis, Table 6 presents the results achieved by geometric attention-based state-of-the-art methods, alongside a comparison with the proposed method. SPD-based models specifically enhance the characterization of ocular dynamics by leveraging the relationships between deep convolutional features of eye movement videos, achieving a higher AUC-ROC score than the convolutional and transformer methods reported in Table 5. Notably, the proposed SPD-based geometric attention module achieved an AUC-ROC score of 97.0±4.0. In contrast, alternative SPD architectures, such as *SPD-Net* and *SPD-CNN*, yielded lower AUC-ROC scores. The kernel-based geometric attention method (*GAM*) achieved the lowest performance, with an AUC-ROC score of 92.4±9.8 and a reduced F1 performance of 74.2±19.9. Overall, the proposed EGA self-attention mechanism exploits geometric structure more effectively, achieving the highest recall and precision.

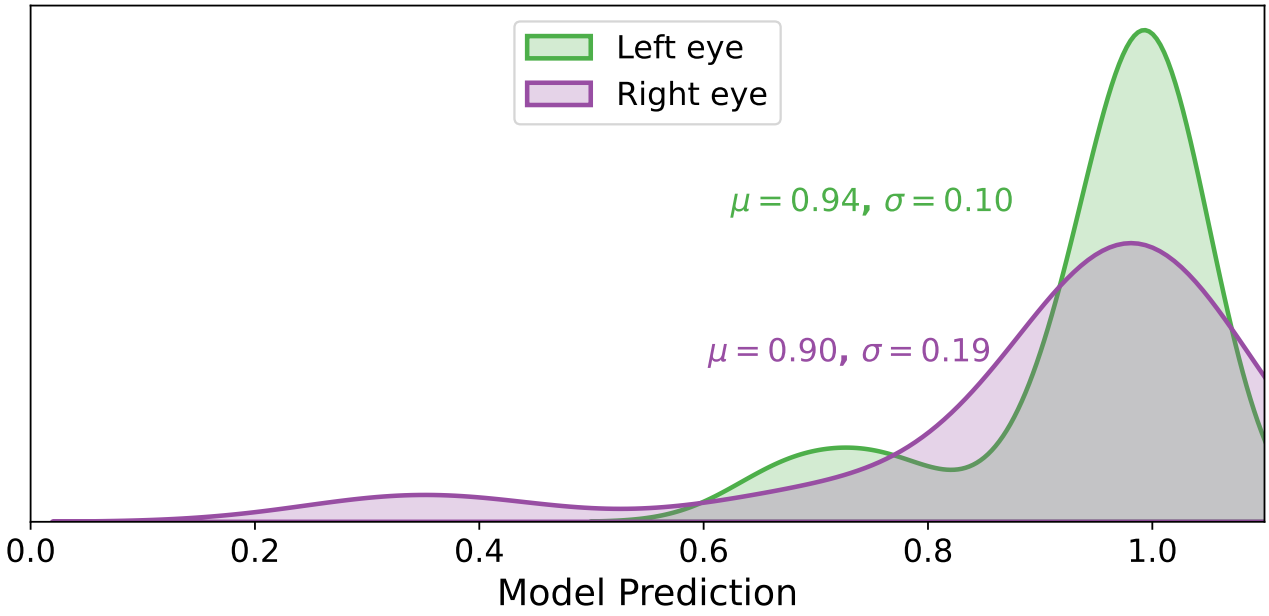
Since each patient produced two predictions (one per eye), we also analyzed the probability distributions of the PD group generated by the proposed EGA approach, as shown in Figure 7. First, we observe that both distributions from the right and left eyes are close to 1, with means of 0.9 ± 0.1 and 0.94 ± 0.19, respectively, indicating high confidence in the proposed method. Here, the right eye exhibited greater variability in prediction scores. This pattern may be associated with early unilateral impairment in the studied PD population, where the right eye exhibited more variation while the left eye consistently showed higher confidence. The tendency for unilateral impairment is well documented in Parkinson's lateralization, not only in general PD-related motor impairments but also in SPEM, particularly through vergence impairments

Table 7. Comparison of cross-attention architectures. A state-of-the-art method was evaluated and EGA cross-attention configurations considering the right eye as query (R2L) and the left eye as query (L2R). Highest scores are highlighted in **bold**.

Model	*Accuracy*	*Recall*	*Precision*	*Specificity*	*F1*	*AUC-ROC*
CMRN	92.0 ± 7.5	92.0 ± 9.8	92.7 ± 9.0	92.0 ± 9.8	92.0 ± 7.5	92.8 ± 10.9
EGA - Cross-Attention (L2R)	92.0 ± 7.5	88.0 ± 9.8	**96.0 ± 8.0**	**96.0 ± 8.0**	91.6 ± 7.6	96.8 ± 6.4
EGA - Cross-Attention (R2L)	**92.0 ± 4.0**	**96.0 ± 8.0**	90.0 ± 8.2	88.0 ± 9.8	**92.3 ± 3.9**	**98.4 ± 3.2**

during stimulus tracking [10].

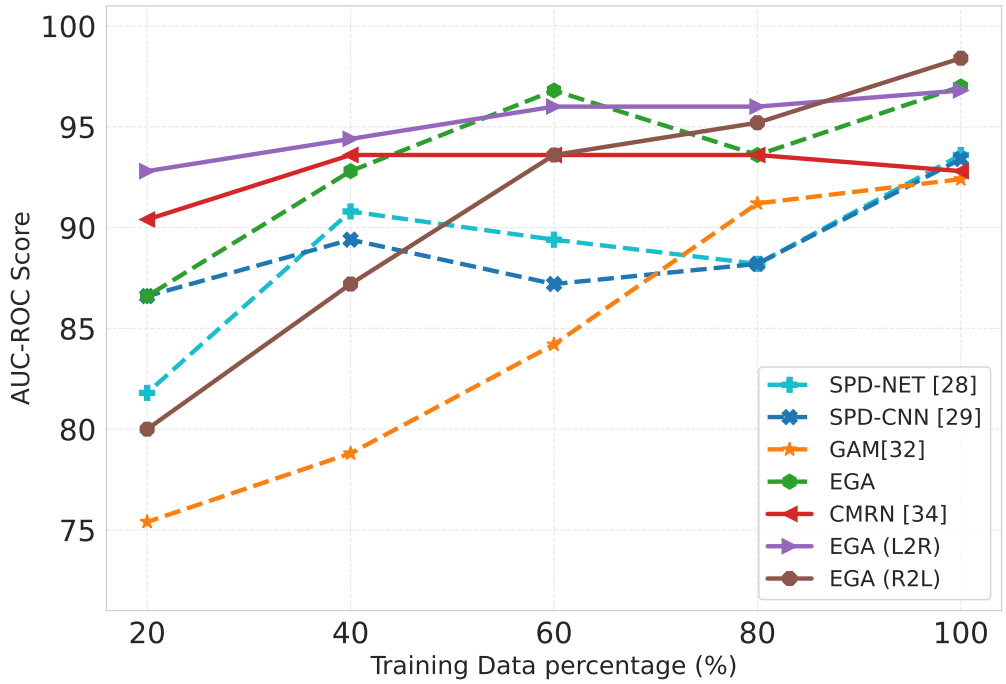
Figure 7. Distribution of left and right eye predictions for Parkinson's patients for the proposed EGA self-attention.



Source: Author's own work

Motivated by this, we proposed a variation of the EGA architecture to integrate multimodal (left and right eye) information through a cross-attention mechanism. In Table 7, we explored the proposed approach, but in the cross-attention configuration. Additionally, we include a state-of-the-art cross-attention geometric method (*CMRN*) [42]. Compared with *CMRN*, the proposed approach yielded improvements of +4.0 (L2R) and +5.6 (R2L) points in AUC-ROC, demonstrating that both configurations enhanced discriminative capability. However, a closer

Figure 8. Performance of SPD-based architectures across varying training data sizes. AUC-ROC scores varying the training data percentage. Self-attention and cross-attention methods are represented by dotted and continuous lines, respectively.

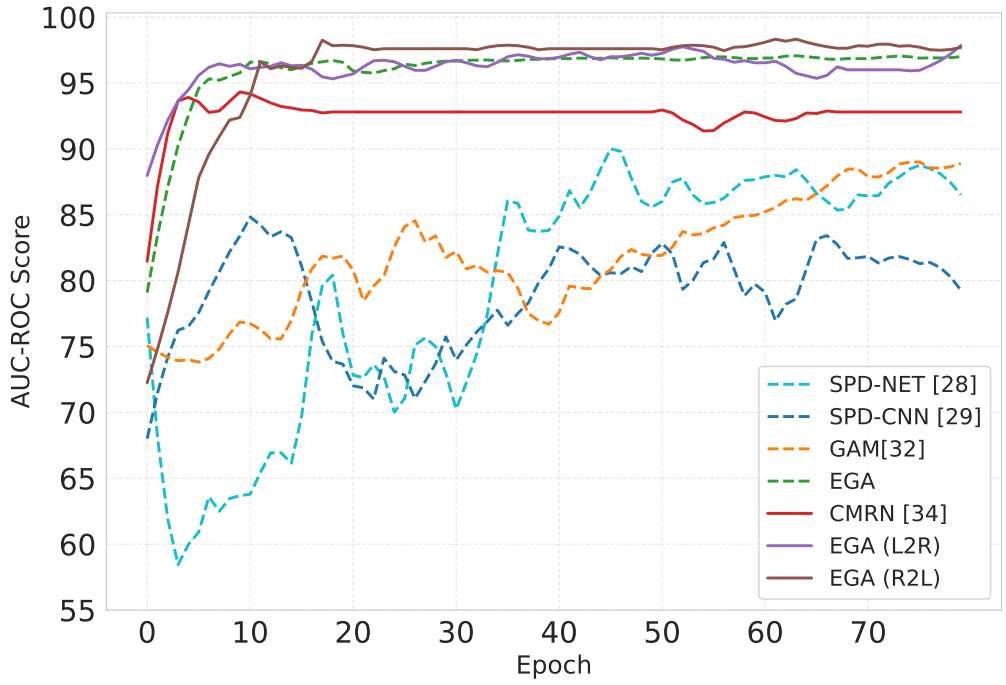


Source: Author's own work

examination of both configurations, the query direction determines how complementary features are emphasized. When guided by the right eye as a query (R2L), the model achieves the highest performance, particularly for PD cases (recall: 96.0±8.0), though at the cost of increased control misclassifications (specificity: 88.0 ± 9.8). Inversely, the left-to-right (L2R) configuration yields higher precision and specificity, aligning more closely with control cases but reducing sensitivity to classify PD.

In an additional experiment, we evaluated the overall performance of the proposed methods in comparison to SPD architectures by examining the evolution of the AUC-ROC score under conditions of reduced training data, as well as the progression of AUC-ROC across epochs (see Figure 8). In terms of training data reduction, we evaluated scenarios using 80%, 60%, 40%, and 20% of the training data. With respect to cross-attention methods, the proposed EGA

Figure 9. Model AUC-ROC performance across training epochs for SPD-based architectures. AUC-ROC score evolution across epochs, illustrating the convergence behavior of the different models. Self-attention and cross-attention methods are represented by dotted and continuous lines, respectively.



Source: Author's own work

(L2R) demonstrated robust performance across all percentages. In contrast, the EGA (R2L) model, despite achieving the best overall scores using 100% of the training data, showed a decline at 40% and 20% of training data. Such results may be attributed to the reduction in the right eye samples, which indicate significant impairments associated with PD. While the CMRN scalar-based attention maintained consistent performance when reducing training data, it did not surpass EGA (L2R). In self-attention methods, the EGA substantially outperformed state-of-the-art attention GAM, as well as the SPD-based methods SPD-NET and SPD-CNN. Notably, GAM was the least effective method in scenarios where less than 60% of the training data was available. Furthermore, in data-scarce scenarios with only 20% and 40% of the data, the cross-attention methods EGA (L2R) and CMRN outperformed the self-attention methods. In the analysis of AUC-ROC score evolution across epochs (right panel Figure 9), the proposed

EGA demonstrates clear superiority over GAM, SPD-NET, and SPD-CNN. These methods present late convergence, on average after 35 epochs, while EGA converges rapidly, approximately after 10 epochs. This highlights the instability and sensitivity in the training dynamics of SPD-NET, SPD-CNN, and GAM methods. Overall, all cross-attention methods exhibited superior performance compared to self-attention methods, except for EGA self-attention. Notably, the proposed EGA (R2L) achieved the highest performance, surpassing both the EGA (L2R) and CMRN methods on average after 15 epochs. Although CMRN exhibited a rapid convergence, its performance was inferior to both EGA cross-attention configurations. In general, all attention-based methods, except for GAM, demonstrated rapid convergence at around 10 epochs, while other methods required more epochs to converge and ultimately achieved lower scores.

7. DISCUSSION

Today, despite that there is no definitive biomarker of PD, there exist several observational patterns that allows to coarsely capture PD phenotypic. Many of these observations are monitored and followed from standardized scales, but with limited sensitivity for early diagnosis, and therefore impacting in guided and personalized protocols of the disease. In consequence, the search and discovery of ocular PD patterns have been valuable to characterize disease and cover a wide range of abnormal motion phenotypes, even in early stages. Particularly, in smooth pursuit eye movements (SPEM), video recordings reveal a diverse set of oculomotor patterns associated with PD. This work introduced the Enhanced Geodesical Attention (EGA) architecture, designed to capture the intricate dynamics of SPEM from left and right eye video sequences by learning to emphasize informative relationships encoded within SPD matrices into a Riemannian space. We explored self-attention and cross-attention alternatives, finding that the best results were achieved by considering both eyes within a cross-attention mechanism. This approach yielded the highest classification performance, with an AUC-ROC of 97.0 ± 4.0, using 4 geometric attention heads in parallel.

Existing works in attention mechanisms from Riemannian SPD descriptors primarily rely on distance computations between attention components (key and query), overlooking the geometric space of these components and the potential of operating directly at the matrix level [40]. By doing so, we demonstrated that using the proposed EGA helps to exploit and highlight the information encoded in SPD matrices, rather than abruptly reducing them to an attention score represented by a scalar value. The proposed EGA utilizes the computation of geodesics in SPD space to carry out the attention process, maintaining and leveraging the intrinsic geometry throughout the mechanism. Interestingly, incorporating the EGA module into classic SPD networks with multiple *BiRe* blocks enabled the development of deeper SPD networks while preserving compact and more discriminative representations. The proposed EGA architecture admits longer Riemannian representations unlike conventional SPD networks, which often suffer

from statistical degradation, where stacking multiple *BiRe* layers does not necessarily lead to performance improvements [50], [51]. In this sense, the proposed EGA offers a novel alternative to the state-of-the-art SPD networks in addressing this limitation.

Traditional deep learning architectures for processing video sequences, such as 3D convolutional neural networks (*ConvNext* and *3D-CNN*) or transformer-based methods (*ViViT*), face limitations associated with their large number of trainable parameters and layers to capture spatial and temporal patterns occurring during videos, requiring extensive data. In our problem, the best alternative was the *3D-CNN* with an AUC-ROC of 92.8 ± 9.8. An alternative has been the incorporation of SPD modules to leverage the most informative relationships within deep spatio-temporal representations into SPD matrices, enabling the capture of more detailed patterns with less data. Including the proposed EGA, which has been demonstrated to improve further the performance advantages in capturing patterns that may exhibit limitations related to PD behavior, such as the temporal concurrence of abnormalities or spatial variations, particularly associated with PD, such as saccades [52]. Compared to the *3D-CNN,* the EGA surpassed it with an AUC-ROC of 97.0 ± 4.0, a gain of 4.2%. This demonstrated that the proposed attention mechanisms enhance traditional networks in scenarios with limited data, thereby improving their performance in distinguishing PD from control cases.

Attention mechanisms enable models to integrate information from one source or modality while focusing on another. In Parkinson's disease, early stages are often characterized by unilateral impairment, leading to an asymmetric progression of the disease [45]. This suggests that jointly considering both left and right eye information may provide a more accurate characterization of the disease [10]. Motivated by this, we proposed a novel cross-attention mechanism (EGA R2L) that integrates binocular information. Notably, this approach surpassed the self-attention

---

[51] Rui Wang et al. "Dreamnet: A deep riemannian manifold network for spd matrix learning". In: *Proceedings of the Asian conference on computer vision*. 2022, pp. 3241–3257.

[52] Agostina J Larrazabal, CE García Cena, and César Ernesto Martínez. "Video-oculography eye tracking towards clinical applications: A review". In: *Computers in biology and medicine* 108 (2019), pp. 57–66.

baseline, achieving an AUC-ROC of 98.4 ± 3.2. This result underscores the robustness of the proposed geometric attention method in computing an attention matrix capable of effectively summarizing relevant information from an SPD descriptor that encapsulates PD patterns from one eye, as well as another SPD descriptor that encapsulates PD patterns from the opposite eye, improving the performance of only considering information from a single eye.

Furthermore, even with the utilization of only half the data compared to self-attention mechanisms, the EGA (L2R) configuration maintained competitive performance. This result demonstrates the robustness and efficiency in data-limited conditions, a recognized challenge in PD research [13]. The results demonstrated that, in fact, the cross-attention EGA method achieved the highest score using only 20% of the training data, with an AUC-ROC score of 92.8 ± 6.4 surpassing existing SPD networks and alternative state-of-the-art geometric attention mechanisms, such as *GAM* and *CMRN* [40], [42].

This underscores the importance of considering binocular information to capture global eye dynamics, rather than relying solely on a monocular configuration. Furthermore, it highlights the advantage of the proposed attention mechanisms in computing a matrix-level attention representation, contrary to *GAM* and *CMRN* methods, which reduce the attention to a single similarity scalar measure. The analysis of learning evolution across epochs demonstrated the proposed geometric attention mechanisms' ability to rapidly converge, identifying relevant discriminative patterns and achieving superior performance. We observed differences in performance when changing eyes order in the EGA cross-attention mechanisms. We found a difference in model prediction in the right eyes of the PD population in this study, which led us to hypothesize that the right eye may exhibit greater variability of abnormalities in the dataset, possibly linked to disease lateralization. This variability can affect model learning, because using the right eye as a query (R2L), it needs to capture a wider range of patterns compared to when the left eye is used as the query (L2R). Under reduced training data, samples from the right eye may not sufficiently capture the full range of abnormalities, resulting in lower performance compared to the alternative cross-attention method (EGA L2R).

In general, the proposed approach captured and coded subtle binocular asymmetries through cross-attention, offering additional insight into early lateralized disease progression and complementing traditional neurological assessments. Besides, the approach was designed over scarce conditions, with limited data, which may be suitable for potential applicability in real-world clinical settings. Even such characterization may be useful in complementary schemes with standard motion scales, bringing additional support for disease stratification. Although the proposed method achieved competitive results, it faces limitations, mainly related to the relatively small cohort size and the lack of longitudinal validation, which may limit its generalization to broader clinical contexts. Additionally, the selection of the eye as a query during training could constrain performance, as one may better capture disease manifestations than the other. This asymmetry may result in uneven learning across eyes.

8. CONCLUSIONS AND FUTURE WORK

This study introduced a novel enhanced geodesic attention (EGA) Riemannian attention architecture for learning compact geometric representations of SPEM video sequences to classify PD patients from healthy controls. The proposed approach provides an alternative to conventional oculomotor-based diagnosis. The presented strategy enables the integration of lateralized impairments observed in PD patients, showing that incorporating binocular information through the proposed EGA mechanism yields consistent gains in distinguishing the two populations, even with limited training data.

Future work should validate the proposed method on larger cohorts, ideally incorporating clinical data such as expert-driven labels regarding PD lateralization and disease staging, to enhance both performance and interpretability of the learned features. It is also essential to continue exploring alternative cross-attention mechanisms, not only for integrating information from both eyes, but also for potentially incorporating additional sources of information, such as gait and speech, to further capture the multifactorial nature of PD.

BIBLIOGRAPHY

Antoniades, Chrystalina A and Miriam Spering. "Eye movements in Parkinson's disease: from neurophysiological mechanisms to diagnostic tools". In: *Trends in Neurosciences* (2023) (cit. on p. 13).

— "Eye movements in Parkinson's disease: from neurophysiological mechanisms to diagnostic tools". In: *Trends in Neurosciences* 47.1 (2024), pp. 71–83 (cit. on p. 24).

Armstrong, RA. "Oculo-visual dysfunction in Parkinson's disease". In: *Journal of Parkinson's disease* 5.4 (2015), pp. 715–726 (cit. on p. 31).

Arnab, Anurag et al. "Vivit: A video vision transformer". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 6836–6846 (cit. on p. 39).

Berardelli, Alfredo et al. "Pathophysiology of bradykinesia in Parkinson's disease". In: *Brain* 124.11 (2001), pp. 2131–2146 (cit. on p. 15).

Bhidayasiri, Roongroj and Pablo Martinez-Martin. "Clinical assessments in Parkinson's disease: scales and monitoring". In: *International review of neurobiology* 132 (2017), pp. 129–182 (cit. on pp. 11, 15).

Brien, Donald C et al. "Classification and staging of Parkinson's disease using video-based eye tracking". In: *Parkinsonism & Related Disorders* 110 (2023), p. 105316 (cit. on p. 24).

Bronstein, Michael M et al. "Geometric deep learning: going beyond euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42 (cit. on pp. 18, 25).

Carreira, Joao and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6299–6308 (cit. on p. 38).

Chang, Zhuoqing et al. "Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and machine learning". In: *Scientific reports* 10.1 (2020), p. 18641 (cit. on p. 24).

Dan, Tingting et al. "Learning brain dynamics of evolving manifold functional MRI data using geometric-attention neural network". In: *IEEE transactions on medical imaging* 41.10 (2022), pp. 2752–2763 (cit. on pp. 23, 25, 39).

Duval, Christian. "Rest and postural tremors in patients with Parkinson's disease". In: *Brain research bulletin* 70.1 (2006), pp. 44–48 (cit. on p. 16).

Frei, Karen. "Abnormalities of smooth pursuit in Parkinson's disease: A systematic review". In: *Clinical parkinsonism & related disorders* 4 (2021), p. 100085 (cit. on pp. 11, 16).

Gilbert, George T. "Positive definite matrices and Sylvester's criterion". In: *The American Mathematical Monthly* 98.1 (1991), pp. 44–46 (cit. on p. 19).

Gorges, Martin, Elmar H Pinkhardt, Jan Kassubek, et al. "Alterations of eye movement control in neurodegenerative movement disorders". In: *Journal of ophthalmology* 2014 (2014) (cit. on pp. 12, 24).

Helmchen, Christoph et al. "Role of anticipation and prediction in smooth pursuit eye movement control in Parkinson's disease". In: *Movement Disorders* 27.8 (2012), pp. 1012–1018 (cit. on p. 23).

Hendricks, Renee M, Mohammad T Khasawneh, et al. “An investigation into the use and meaning of Parkinson’s disease clinical scale scores”. In: *Parkinson’s Disease* 2021 (2021) (cit. on p. 15).

Huang, Zhiwu and Luc Van Gool. “A riemannian network for spd matrix learning”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017 (cit. on pp. 21, 22, 25, 32, 39).

Jankovic, Joseph. “Parkinson’s disease: clinical features and diagnosis”. In: *Journal of neurology, neurosurgery & psychiatry* 79.4 (2008), pp. 368–376 (cit. on p. 16).

Jansson, Daniel et al. “Stochastic anomaly detection in eye-tracking data for quantification of motor symptoms in Parkinson’s disease”. In: *Signal and Image Analysis for Biomedical and Life Sciences* (2015), pp. 63–82 (cit. on p. 24).

Jung, Ileok and Ji-Soo Kim. “Abnormal eye movements in parkinsonism and movement disorders”. In: *Journal of movement disorders* 12.1 (2019), p. 1 (cit. on pp. 16, 23).

Konstantinidis, Dimitrios et al. “Multi-manifold attention for vision transformers”. In: *IEEE Access* (2023) (cit. on p. 26).

Larrazabal, Agostina J, CE García Cena, and César Ernesto Martínez. “Video-oculography eye tracking towards clinical applications: A review”. In: *Computers in biology and medicine* 108 (2019), pp. 57–66 (cit. on p. 13).

— “Video-oculography eye tracking towards clinical applications: A review”. In: *Computers in biology and medicine* 108 (2019), pp. 57–66 (cit. on p. 50).

Li, Han et al. “Abnormal eye movements in Parkinson’s disease: From experimental study to clinical application”. In: *Parkinsonism & Related Disorders* (2023), p. 105791 (cit. on p. 12).

Liu, Zhuang et al. “A convnet for the 2020s”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 11976–11986 (cit. on p. 38).

Lu, Bin et al. “Manifold attention-enhanced multi-domain convolutional network for decoding motor imagery intention”. In: *Knowledge-Based Systems* 296 (2024), p. 111904 (cit. on pp. 25, 49, 51).

Ma, Junbo and et al. “Multimodality Alzheimer’s disease analysis in deep Riemannian manifold”. In: *Information Processing & Management* (2022) (cit. on pp. 26, 33, 40, 45, 51).

Marino, Silvia et al. “Quantitative analysis of pursuit ocular movements in Parkinson’s disease by using a video-based eye tracking system”. In: *European Neurology* 58.4 (2007), pp. 193–197 (cit. on p. 17).

Mei, Jie, Christian Desrosiers, and Johannes Frasnelli. “Machine learning for the diagnosis of Parkinson’s disease: a review of literature”. In: *Frontiers in aging neuroscience* 13 (2021), p. 633752 (cit. on p. 13).

Minh, Hà Quang and Vittorio Murino. *Covariances in computer vision and machine learning*. Springer, 2018 (cit. on p. 21).

Niu, Zhaoyang, Guoqiang Zhong, and Hui Yu. “A review on the attention mechanism of deep learning”. In: *Neurocomputing* 452 (2021), pp. 48–62 (cit. on pp. 18, 36).

Olmos, Juan, Antoine Manzanera, and Fabio Martínez. “Riemannian SPD learning to represent and characterize fixational oculomotor Parkinsonian abnormalities”. In: *Pattern Recognition Letters* 177 (2024), pp. 157–163 (cit. on pp. 41, 50).

Pan, Yue-Ting, Jing-Lun Chou, and Chun-Shu Wei. "MAtt: A manifold attention network for EEG decoding". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 31116–31129 (cit. on pp. 25, 33).

Pennec, Xavier, Pierre Fillard, and Nicholas Ayache. "A Riemannian framework for tensor computing". In: *International Journal of computer vision* 66 (2006), pp. 41–66 (cit. on p. 20).

Petit, Olivier et al. "U-net transformer: Self and cross attention for medical image segmentation". In: *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer. 2021, pp. 267–276 (cit. on p. 19).

Pinkhardt, Elmar H et al. "Eye movement impairments in Parkinson's disease: possible role of extradopaminergic mechanisms". In: *BMC neurology* 12 (2012), pp. 1–8 (cit. on pp. 12, 17).

Poewe, Werner et al. "Parkinson disease". In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21 (cit. on pp. 11, 15).

Rajan, Vandana, Alessio Brutti, and Andrea Cavallaro. "Is cross-attention preferable to self-attention for multi-modal emotion recognition?" In: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2022, pp. 4693–4697 (cit. on p. 35).

Riederer, P and J Sian-Hülsmann. "The significance of neuronal lateralisation in Parkinson's disease". In: *Journal of Neural Transmission* 119 (2012), pp. 953–962 (cit. on pp. 35, 50).

Shibasaki, Hiroshi, Sadatoshi Tsuji, and Yoshigoro Kuroiwa. "Oculomotor abnormalities in Parkinson's disease". In: *Archives of Neurology* 36.6 (1979), pp. 360–364 (cit. on pp. 12, 17, 23).

Solovyev, Roman, Alexandr A Kalinin, and Tatiana Gabruseva. "3D convolutional neural networks for stalled brain capillary detection". In: *Computers in Biology and Medicine* 141 (2022), p. 105089. DOI: 10.1016/j.compbiomed.2021.105089 (cit. on p. 38).

Su, Dongning et al. "Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021". In: *bmj* 388 (2025) (cit. on p. 11).

Sun, Yue Ran et al. "Monitoring eye movement in patients with Parkinson's disease: what can it tell us?" In: *Eye and Brain* (2023), pp. 101–112 (cit. on pp. 12, 17, 24, 35, 45, 50).

Terao, Yasuo, Hideki Fukuda, and Okihide Hikosaka. "What do eye movements tell us about patients with neurological disorders?—An introduction to saccade recording in the clinical setting—". In: *Proceedings of the Japan Academy, Series B* 93.10 (2017), pp. 772–801 (cit. on pp. 12, 17, 23).

Tolosa, Eduardo and et al. "Challenges in the diagnosis of Parkinson's disease". In: *The Lancet Neurology* 20.5 (2021), pp. 385–397 (cit. on pp. 13, 16, 51).

Wang, Rui et al. "Dreamnet: A deep riemannian manifold network for spd matrix learning". In: *Proceedings of the Asian conference on computer vision*. 2022, pp. 3241–3257 (cit. on p. 50).

Woo, S et al. *Cbam: convolutional block attention module. In proceedings of the European conference on computer vision (ECCV): 3-19*. 2018 (cit. on p. 19).

Wu, Chia-Chien et al. "Eye movement control during visual pursuit in Parkinson's disease". In: *PeerJ* 6 (2018), e5442 (cit. on pp. 12, 16).

Zhang, Jian et al. "Global, regional and national temporal trends in Parkinson's disease incidence, disability-adjusted life year rates in middle-aged and older adults: a cross-national

inequality analysis and Bayesian age-period-cohort analysis based on the global burden of disease 2021". In: *Neurological Sciences* 46.4 (2025), pp. 1647–1660 (cit. on p. 11).

Zhang, Tong et al. "Deep manifold-to-manifold transforming network". In: *2018 25th IEEE international conference on image processing (ICIP)*. IEEE. 2018, pp. 4098–4102 (cit. on pp. 23, 25, 39).

APPENDICES

Anexo A. Academic Products

Journals

- **Celis, L. F.**, Olmos, J., Manzanera, A., & Martínez, F. (2025). Learning a geometric deep representation to classify Parkinson smooth pursuit patterns. Pattern Analysis and Applications, 28(3), 1-14.

  Status: Publised

- **Celis, L. F.**, Olmos, J., & Martínez, F. (2025, November). EGA: Enhanced Geodesic Attention for SPEM oculomotor Parkinson Quantification. IEEE Journal of Biomedical and Health Informatics (2025).

  Status: Sent

Conference papers

- **Celis, L. F.**, Olmos, J., & Martínez, F. (2024, November). A Geometric Attention Mechanism to Classify Parkinsonism Smooth Pursuit Patterns. In Ibero-American Conference on Artificial Intelligence (pp. 99-109).

  Status: Published/Presented.

  Best paper award.

Collaborations

- Archila, J., Peña, I. **Celis, L.**, Olmos, J., A. Manzanera & Martínez, F. 2025. A multimodal gait and ocular geometric representation to generate a Parkinson progression report. In Engineering Applications of Artificial Intelligence (2025).

  Status: Published

Anexo B. Ethical Approval

In accordance with the principles established in the Declaration of Helsinki, the CIOMS Guidelines and in resolution 8430 of October 4, 1993 of the Ministry of Health, because this research is considered without risk and in compliance with the aspects mentioned in Article 6 of said resolution, this study will be developed in accordance with the following criteria:

- Autonomy will be guaranteed and will not be violated because the information will be collected from the health care process. The information collected is covered by the informed consent of the care process of the respective institutions.
- The principle of non-maleficence will not be affected, the possibility of maleficence will be minimized, given that it is an observational, analytical study, where no harm will be done, no changes will be made in the therapeutic schemes.
- The confidential and sensitive information and the privacy of the patients will be protected, only the personnel who will collect the information in the collection forms (CRF) will know the identification number to be able to register the necessary data.
- The principle of justice will not be affected, since individuals will not be exposed to a situation of real or potential risk and no advantage will be taken of any situation of legal vulnerability or subordination of the subjects due to this research.
- This study will not have a direct effect on the principle of Beneficence, since it is an observational study.
- This type of study has been previously carried out in human beings without generating a potential or real harm to the evaluated group.
- The data intended to be produced in the present study cannot be obtained from simulations, mathematical formulas or research in animals, or other samples.
- All the investigators have a certificate of good practices for the protection of the research participants.

- The beginning of the data collection will be done after having received the approval of the responsible Institution.
- The researchers participating in this study are qualified in their academic training and have no economic, legal or personal conflicts of interest associated with this research problem.

Informed Consent

Because this research project is in the no risk category, it is considered that informed consent can be deferred. Because no interventions will be performed on patients and data will be collected from the process of care that is within the institution's informed consent.

Confidentiality

The information obtained will have absolute confidentiality and will be used for academic purposes. The registration of each participant's information will be filed only by the principal investigator, and each record will be assigned to each participant's ID number and for entry into the database. Therefore, the names or personal data of the participants will not be disclosed in any form; when the results of this study are published in scientific journals or congresses, the personal identification data will be omitted.

The data of each participant will not be available to third parties such as employers, governmental organizations or insurance companies, unless each participant in accordance with the specifications governed by the Law grants an authorization for the disclosure of the data. However, to ensure proper management of the data, a member of the Ethics Committee of the Universidad Industrial de Santander may consult them and verify their registration to ensure compliance with ethical aspects.

Treatment of personal data

In compliance with the provisions of the Statutory Law 1581 of 2012, its Regulatory Decree 1377 of 2013 and the Rector's Resolution 1227 of 2013, the Universidad Industrial de Santander adopts the national policy for the treatment of personal data, which will be informed to all holders

of data collected or obtained in the future in the exercise of academic, cultural, commercial or labor activities arising from this research project. In this sense, the principal investigator of the project states that he guarantees the rights of privacy, intimacy and good name of the research subjects in the treatment of their personal data, and consequently, all his actions will be governed by the principles of legality, purpose, freedom, truthfulness or quality, transparency, restricted access and circulation, security and confidentiality. The above implies that all persons who, in the development of the different activities of the project, may provide any type of information or personal data may know, update, rectify or delete it.

The data collected will be stored and guarded in servers of the Biomedical Imaging, Vision and Learning Laboratory (BIVL2ab) research group, with access codes for the users where the information will be stored. By virtue of the anonymization of the data, they will remain permanently in the storage devices provided for the project, where they can be consulted by the researchers since these data do not imply any association with the patient.