CLASIFICACIÓN DE PATRONES PARKINSONIANOS DESDE UNA REPRESENTACIÓN MULTIMODAL QUE INCLUYE PUNTOS DE REFERENCIA DE MARCHA Y ROSTRO

FAIBER STIVEN ANGARITA MENDOZA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2025

CLASIFICACIÓN DE PATRONES PARKINSONIANOS DESDE UNA REPRESENTACIÓN MULTIMODAL QUE INCLUYE PUNTOS DE REFERENCIA DE MARCHA Y ROSTRO

FAIBER STIVEN ANGARITA MENDOZA

Trabajo de Grado presentado en cumplimiento de los requisitos para optar al título de:
Ingeniero de Sistemas

Director:
Fabio Martínez Carrillo
Doctor en Ingeniería de Sistemas y Computación

Codirector:
John Edinson Archila Valderrama
Magíster en Ingeniería Electronica

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2025

AGRADECIMIENTOS Y DEDICATORIA

Agradezco profundamente al director del grupo de investigación BIVL^{2}ab, el profe Fabio Martínez Carrillo, por su dedicación incansable, por las innumerables horas de trabajo invertidas y por la guía constante que orientó el desarrollo de este proyecto de investigación.

Mi gratitud también para el muchacho John, quien fue un segundo guía esencial en este proceso. Su apoyo permanente y sus valiosas retroalimentaciones fueron fundamentales para la evolución y la solidez de este trabajo.

Al grupo de investigación BIVL^{2}ab, gracias por acogerme con respeto y por brindarme un espacio de crecimiento en esta noble familia académica. En especial, extiendo mi agradecimiento a Jean, al profesor Olmos, al profesor Guayacán y a Fernando, por sus enseñanzas y compañía.

Agradezco a la Universidad Industrial de Santander y a la Escuela de Ingeniería de Sistemas por brindarme la oportunidad de formarme como profesional en esta universidad que tanto aprecio.

Dedico este trabajo a mi padre Jose Angarita, con quien compartía el sueño de convertirme algún día en ingeniero de sistemas. Agradezco también a mis hermanos y a Stelita, por su amor y apoyo incondicional.

Finalmente, reconozco con orgullo el papel de mi propia voluntad, que nunca se rindió a pesar de las dificultades de este arduo pero hermoso camino que es la investigación.

CONTENIDO

pág.

INTRODUCCIÓN **10**

1. FUNDAMENTOS Y TRABAJOS PREVIOS **12**

1.1. PARKINSON: UNA AFECTACIÓN MULTIFACTORIAL 12

1.1.1 Afecciones motoras en la marcha 13

1.1.2 Afecciones asociadas a la hipomimia 16

1.2. REPRESENTACIONES POSTURALES ASOCIADAS AL MOVIMIENTO 18

1.3. ANÁLISIS MULTIMODAL COMPUTACIONAL 20

1.3.1 La covarianza como una estrategia de fusión 21

1.3.2 Análisis multimodal basado en representaciones profundas 22

1.3.3 Conceptos de variedades Riemannianas 23

1.4. MÉTODOS COMPUTACIONALES PARA LA CUANTIFICACIÓN MOTORA DEL PARKINSON 26

2. PROBLEMA DE INVESTIGACIÓN **30**

3. OBJETIVOS **32**

3.1. OBJETIVO GENERAL 32

3.2. OBJETIVOS ESPECÍFICOS 32

4. MÉTODO PROPUESTO **33**

4.1. PUNTOS DE INTERÉS Y DESCRIPCIÓN ESPACIO-TEMPORAL 33

4.2. DESCRIPTORES DE COVARIANZA PARA MARCHA Y ROSTRO 38

4.3. APRENDIZAJE GEOMÉTRICO DE REPRESENTACIONES MÁS COMPACTAS 40

4.4. DESCRIPTOR MULTIMODAL D_M Y CLASIFICACIÓN RIEMANNIANA 42

5. DISEÑO EXPERIMENTAL **46**
5.1. DATOS 46
5.2. CONFIGURACIÓN EXPERIMENTAL DE LA ARQUITECTURA PROPUESTA . 47

6. EVALUACIÓN Y RESULTADOS **50**
6.1. CLASIFICACIÓN DEL PARKINSON DESDE LA MARCHA 50
6.2. CLASIFICACIÓN DEL PARKINSON DESDE LA EXPRESIÓN FACIAL 51
6.3. FUSIÓN DE LAS MODALIDADES DE MARCHA Y EXPRESIÓN FACIAL . . . 53

7. CONCLUSIONES Y TRABAJO FUTURO **59**

BIBLIOGRAFÍA **62**

LISTA DE FIGURAS

pág.

Figura 1. Comparación de Pacientes Parkinson y control 14
Figura 2. Ilustración de la bradicinesia presente en la marcha. 15
Figura 3. Ilustración de la rigidez presente en la marcha. 16
Figura 4. Modelos esqueléticos para el análisis de la postura. 19
Figura 5. Arquitectura Blazeface 20
Figura 6. Ejemplo de la estimación postural de Mediapipe con un fotograma del dataset de pacientes. 21
Figura 7. estrategias de fusión en el aprendizaje profundo 24
Figura 8. Esquema de la estrategia propuesta. 34
Figura 9. Cuaterniones y Trayectorias 38
Figura 10. Covarianzas entre poses 39
Figura 11 Arquitectura de la fusión temprana: Concatenando canales 43
Figura 12 Arquitectura de la fusión intermedia: Sumando covarianzas 44
Figura 13 Arquitectura de la fusión tardía 45
Figura 14. Protocolo de grabación 48
Figura 15. Evolución de las métricas al variar el parámetro λ en la fusión tardía. . . 57

LISTA DE TABLAS

pág.

Tabla 1. Resultados de clasificación entre Parkinson y control a partir de representaciones de cuaterniones. 51
Tabla 2. Resultados de clasificación utilizando diferentes representaciones cinemáticas 53
Tabla 3. Resultados de clasificación utilizando diferentes estrategias de fusión de características. 54
Tabla 4. Resultados de clasificación en la fusión temprana e intermedia en sus dos diferentes configuraciones respectivamente. 56

RESUMEN

TÍTULO: Clasificación de patrones parkinsonianos desde una representación multimodal que incluye puntos de referencia de marcha y rostro [*]

AUTOR: Faiber Stiven Angarita Mendoza [**]

PALABRAS CLAVE: Enfermedad de Parkinson, clasificación, multimodal, puntos de referencia.

DESCRIPCIÓN: El Parkinson es una enfermedad multifactorial neurodegenerativa que es caracterizada por afecciones motoras como la bradicinesia (lentitud de los movimientos), inestabilidad postural, hipomimia, desordenes del habla y rigidez, como consecuencia del creciente déficit de dopamina. Estos síntomas son variables en intensidad y frecuencia, por lo que representan un desafío para el diagnóstico, estratificación y seguimiento de la enfermedad. De hecho, se han reportado errores diagnósticos del 47% para un médico general y hasta un 8% para especialistas en desordenes de movimiento. Recientemente, se han propuesto métodos computacionales para soportar la cuantificación y análisis de patrones durante la marcha y la expresión facial, pero abordados de forma independiente, perdiendo el carácter de análisis multimodal. En este trabajo de investigación se desarrolló una estrategia multimodal que, desde un conjunto de puntos de referencia, tanto posturales (en marcha), como de referencia de gestos de rostros, permite dar soporte al diagnóstico del Parkinson, bajo una tarea de clasificación. El método extrajo puntos de interés desde una red profunda, que permite hacer estimaciones sin marcadores. Luego estos puntos de interés fueron utilizados para construir descriptores correlativos, que permitan una discriminación entre una población de Parkinson y una población control. La metodología propuesta fue ajustada y validada con conjuntos de 580 videos, registrados desde 11 pacientes diagnosticados con la enfermedad y 18 pacientes control. El método desarrollado, bajo un esquema de validación cruzada (k=5), logró obtener una exactitud de 92% ± 0.02, sensibilidad de $84\% \pm 0.05$, precisión de $94\% \pm 0.06$, *F1-score* de 89% ± 0.03 y AUC de $90\% \pm 0.03$, logrando superar esquemas unimodales.

---

[*] Trabajo de investigación

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, PhD en ingeniería de sistemas y computación, análisis de imágenes y análisis de vídeo. Codirector: John Edinson Archila Valderrama, Magíster en Ingeniería Electronica

ABSTRACT

TITLE: Classification of Parkinsonian patterns from a multimodal representation that includes gait and facial landmarks [*]

AUTHOR: Faiber Stiven Angarita Mendoza [**]

KEYWORDS: Parkinson's disease, classification, multimodal, landmarks.

DESCRIPTION: Parkinson's disease is a multifactorial neurodegenerative disorder characterized by motor impairments such as bradykinesia (slowness of movement), postural instability, hypomimia, speech disorders, and rigidity, all resulting from the progressive depletion of dopamine. These symptoms vary in intensity and frequency, which makes them a challenge for the diagnosis, stratification, and monitoring of the disease. In fact, diagnostic errors have been reported as high as 47% for general practitioners and up to 8% for specialists in movement disorders. Recently, computational methods have been proposed to support the quantification and analysis of gait and facial expression patterns, but these have been addressed independently, losing the multimodal analysis aspect. This work proposes the extraction of both postural reference points and facial gesture reference points to build correlational descriptors, which enable discrimination between a Parkinson's population and a control group. The low-dimensional representation of these reference points, both for gait and facial expressions, will be used to extract dynamic information for each motion trajectory. These trajectories will then be grouped and summarized into covariance representations, which will allow the determination of patterns such as coordination and instability associated with the disease. The proposed methodology will be fine-tuned and validated using video datasets recorded in an experimental scheme with control participants and Parkinson's patients. The results obtained were promising, achieving an accuracy of $92\% \pm 0.02$, a recall of $84\% \pm 0.05$, a precision of $94\% \pm 0.06$, an F1-score of $89\% \pm 0.03$ and AUC of $90\% \pm 0.03$ for the fusion methodology with the best classification metrics, demonstrating the effectiveness of the method in discriminating patterns related to Parkinson's disease.

---

[*] Research work

[**] Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez Carrillo, PhD. Computer and systems engineering, medical image analysis and video analysis. Co-advisor: John Edinson Archila Valderrama, Master in Electronic Engineering.

INTRODUCCIÓN

La enfermedad de Parkinson (EP) es una enfermedad multifactorial neurodegenerativa, siendo la segunda más frecuente en el mundo. Esta enfermedad afecta al 1% de la población mundial mayor de 60 años, convirtiéndose en una de las principales causas de discapacidad neurológica [1]. En la actualidad, el diagnóstico de esta enfermedad se basa en la observación del neurólogo y su respectiva estratificación, basándose en escalas motoras como H&Y (*Hoehn and Yahr Scale*) y la MDS-UPDRS (*Movement Disorders Society - Unified Parkinson's Disease Rating Scale*) parte III[2]. Sin embargo, esta medición es subjetiva, dependiendo de la pericia del especialista[3]. A su vez, esta enfermedad reporta diversos fenotipos funcionales, es decir, personas con alta variabilidad en cuanto a los síntomas de origen, la escala de afectación y la evolución de los mismos. Así, la enfermedad tiene una alta variabilidad, lo que dificulta un diagnóstico apropiado. De hecho, reportes en la literatura han estimado errores diagnósticos entre un 6% y un 25%, dependiendo del nivel de experticia de los especialistas [4].

Diferentes enfoques computacionales se han propuesto para ser herramientas de apoyo

---

[1] Tanya SIMUNI and Rajesh PAHWA. *Parkinson's disease*. Oxford University Press, USA, 2009.

[2] E. ROVINI; C. MAREMMANI, and F. CAVALLO. "How wearable sensors can support parkinson's disease diagnosis and treatment: a systematic review". In: *Frontiers in neuroscience* 11 (2017), p. 555.

[3] Michela TINELLI; Panos KANAVOS, and Federico GRIMACCIA. "The value of early diagnosis and treatment in Parkinson's disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in Parkinson's disease". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 16.1 (2016), pp. 41–51. DOI: 10.1586/14737167.2016.1121768.

[4] Michela TINELLI; Panos KANAVOS, and Federico GRIMACCIA. "The value of early diagnosis and treatment in Parkinson's disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in Parkinson's disease". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 16.1 (2016), pp. 41–51. DOI: 10.1586/14737167.2016.1121768.

al diagnóstico mediante la clasificación de modalidades individuales como la marcha, la expresión facial, el temblor en las manos, entre otros [5]. Sin embargo, estos enfoques son eficaces a medida que el síntoma dominante se refleja en la modalidad bajo estudio. Por ende, un paciente cuyo síntoma dominante es difícilmente reflejado en la modalidad analizada conducirá a errores en la predicción del diagnóstico. Como consecuencia, surge la necesidad de avanzar en el análisis multimodal, analizando varias modalidades e identificando síntomas cardinales que permitan la caracterización de los pacientes. En la literatura se han reportado enfoques multimodales, que han evidenciado mayor robustez para caracterizar los pacientes con Parkinson. Sin embargo, el uso de múltiples modalidades resulta desafiante debido a la dificultad de capturar información con múltiples fuentes de movimiento, y el uso de arquitecturas que permitan operar en estos escenarios. Además, estos enfoques deben considerar referencias clínicas sobre regiones relevantes de análisis, para incluirlas dentro de los modelos, lo cual permita una mayor adaptación como herramienta de soporte en la rutina clínica.

En este trabajo de investigación se desarrolló una estrategia de aprendizaje profundo para clasificar patrones de marcha y expresión facial asociados al Parkinson. Para ello, se implementó una arquitectura capaz de extraer puntos de referencia postural y marcadores faciales, los cuales fueron utilizados para construir descriptores espaciotemporales en una variedad Riemanniana. En este trabajo se modelaron representaciones geométricas compactas para la caracterización de los dos patrones de interés. Luego se exploraron, implementaron y propusieron mecanismos multimodales, que preservando la geometría de los descriptores de covarianza, permitieron obtener una clasificación más certera para discriminar entre sujetos con Parkinson y controles, validando así la metodología propuesta.

---

[5] A. W. MICHELL et al. “Biomarkers and Parkinson’s disease”. In: *Brain* 127.8 (June 2004), pp. 1693–1705. DOI: 10.1093/brain/awh198. eprint: https://academic.oup.com/brain/article-pdf/127/8/1693/842701/awh198.pdf.

1. FUNDAMENTOS Y TRABAJOS PREVIOS

1.1. PARKINSON: UNA AFECTACIÓN MULTIFACTORIAL

La enfermedad de Parkinson (EP) es la segunda enfermedad neurodegenerativa más frecuente en el mundo, afectando principalmente a personas mayores de 60 años, a nivel mundial tiene una prevalencia mayor al 1% en pacientes mayores de 65 años. La prevalencia en América Latina oscila entre 30 y 200 casos por cada 100,000 habitantes, con una tendencia al alza debido al envejecimiento de la población. Además, se ha reportado que los hombres tienen entre 1.5 y 2 veces más probabilidades de desarrollar la enfermedad en comparación con las mujeres.

En cuanto al origen y la descripción fisiopatológica, la enfermedad del Parkinson esta asociada con la pérdida progresiva de los niveles de dopamina, causado por la degeneración de las células dopaminérgicas. Estas células actúan como los neuroreceptores encargados de coordinar funciones motoras y cognitivas en el cerebro. Particularmente, los niveles bajos de dopamina provocan una disfunción de los ganglios basales, lo que genera déficits motores.[6] Esto conduce a una reducción progresiva en la iniciación y control del movimiento, manifestándose en bradicinesia, rigidez y alteraciones en los reflejos posturales. Con el avance de la enfermedad, se observan complicaciones motoras como fluctuaciones en la respuesta al tratamiento, discinesias inducidas por la medicación y un mayor riesgo de caídas debido a la inestabilidad postural.[7]

En la actualidad, el diagnóstico de esta enfermedad se fundamenta en la observación de anomalías motoras, cognitivas, entre otras muestras de deterioro causadas por la

---

[6] D BERGERON et al. “Sir William Osler and the evolving neurological sciences”. In: *Lancet* 11 (2012), pp. 999–1004.

[7] Carrillo-Ruiz José D JOSÉ CÁT Rodrigo BJ. “Interpretación neuroanatómica de los principales síntomas motores y no-motores de la enfermedad de Parkinson”. In: *Rev Mex Neurociencia* 1 (2010), pp. 1–10.

enfermedad. Entre estos patrones anormales, las afecciones motoras y las afecciones causadas por la hipomimia (anormalidades gestuales) son patrones de principal relevancia para su caracterización. En los patrones motores, la bradicinesia y la inestabilidad postural son los más relevantes en el diagnóstico clínico[8].

Además, en estadios tempranos de la EP, pueden aparecer alteraciones en el habla, como dificultades en la articulación de palabras y variaciones en el tono de voz, lo que genera un habla monótona y la llamada "cara de máscara" o hipomimia [9]. A pesar de estos patrones anormales, asociados con la enfermedad, existen otras enfermedades con parkinsonismos que pueden generar confusiones en el diagnóstico y procedimientos inadecuados en los pacientes. Además, la EP es multifactorial y reporta diferentes fenotipos, lo que indica que existe una alta variabilidad en cuanto a la aparición, tipo y escala de los síntomas. Ejemplos de estas alteraciones son ilustradas en la Figura 1 [10][11]. En las siguientes subsecciones se darán más detalles de estos síntomas.

1.1.1. Afecciones motoras en la marcha En la EP, las afecciones motoras son uno de los principales biomarcadores utilizados para soportar el diagnóstico de la enfermedad. Estos patrones pueden ser analizados durante ejercicios de locomoción, pudiendo establecer escalas de afección, que en algunos pacientes, es temprano. A continuación, se detallan algunas de estas afecciones motoras:

---

[8] K BERGANZO et al. “Síntomas no motores y motores en la enfermedad de Parkinson y su relación con la calidad de vida y los distintos subgrupos clínicos”. In: *Neurología* 31.9 (2016), pp. 585–591.

[9] Teresa MAYCAS-CEPEDA et al. “Hypomimia in Parkinson’s disease: what is it telling us?” In: *Frontiers in Neurology* 11 (2021), p. 603582.

[10] Jan RUSZ et al. “Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson’s disease”. In: *npj Parkinson’s Disease* 7.1 (2021), p. 98.

[11] L RICCIARDI; A DE ANGELIS, et al. “Hypomimia in Parkinson’s disease: an axial sign responsive to levodopa”. In: *European Journal of Neurology* 27.12 (2020), pp. 2422–2429.

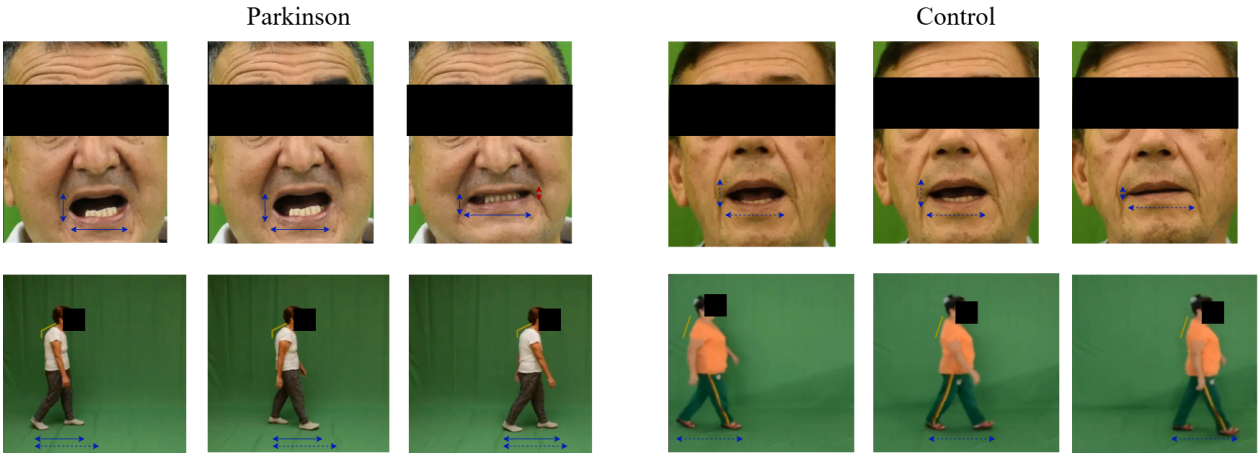


Figura 1. Comparación de Pacientes Parkinson y control. La línea amarilla representa la afectación postural. Las manos presentan amplitud reducida lo que genera la no alienación con su opuesto. La línea discontinua azul muestra cómo debería ser la apertura aproximada del paso, por ende, se observa una reducción de la amplitud de los movimientos, la línea roja muestra el tamaño de cierre correcto de la boca.

- Bradicinesia: Es la lentitud que puede presentar un paciente en el momento en el que este realiza un movimiento voluntario. Se puede expresar como disminución del braceo (mediante la marcha). Adicionalmente, puede haber marcha a pasos cortos, arrastrando los pies y lentitud generalizada en todos sus movimientos, lo que genera que los pacientes tarden más en realizar actividades cotidianas[12]. La bradicinesia es un síntoma cardinal. En un estudio realizado a 136 pacientes se obtuvo que un 80% de pacientes presentaban bradicinesia[13]. A continuación, se ilustra en la Figura 2 el patrón asociado a la bradicinesia en pacientes con la EP.

- Rigidez: Se le conoce como la forma de hipertonicidad o incremento del tono muscular observado en movimientos pasivos del paciente, tanto para músculos flexores

[12] Carolina LEÓN-JIMÉNEZ. "Síndrome rígido acinético". In: *Revista de Medicina Clínica* 3.2 (2019), pp. 104–108.

[13] Sofía GARRIDO-ELUSTONDO et al. "Capacidad de detección de patología psiquiátrica por el médico de familia". In: *Atención Primaria* 48.7 (2016), pp. 449–457.

Figura 2. Ilustración de la bradicinesia presente en la marcha reflejada como la lentitud de los movimientos en este ejemplo se observa en la distancia reducida del paso además de la postura encorvada.

como extensores sin la presencia de resistencia al movimiento pasivo de las articulaciones. Puede variar su intensidad de leve a mayor; cuando éste es mayor, la resistencia al movimiento pasivo se identifica durante todo el movimiento y se caracteriza por no ser suave y presentar fuertes cambios en la aceleración. Su identificación se centra en la observación de las dificultades que un paciente pueda tener en la realización de los siguientes movimientos mediante la marcha: movilización de articulaciones, desplazamientos de muñecas y tobillos, movimientos de codos y rodillas. Mediante la expresión facial, este patrón se observa en la dificultad para gesticular[14]. La rigidez, por ejemplo, en el estudio de 136 pacientes, el 60% de los pacientes presentaron este síntoma [15]. En la figura 3 se ilustran patrones de rigidez que pueden ser observados en pacientes con la EP.

---

[14] Carolina LEÓN-JIMÉNEZ. "Síndrome rígido acinético". In: *Revista de Medicina Clínica* 3.2 (2019), pp. 104–108.

[15] Nancy Bertado RAMÍREZ et al. "Datos clave para el diagnóstico clínico de enfermedad de Parkinson". In: *Revista Mexicana de Neurociencia* 10.5 (2009), pp. 340–343.

Figura 3. Ilustración de la rigidez reflejada en los brazos semiflexionados pegados al cuerpo durante la marcha.

1.1.2. Afecciones asociadas a la hipomimia Diferentes estudios proponen que hasta el 90% de los pacientes presentan un deterioro vocal muy marcado en las características fonatorias y articulatorias en el ejercicio de la comunicación [9]. Entre los síntomas más comunes se encuentran la reducción de la intensidad vocal, voz con aspereza, temblor exagerado, articulación imprecisa de vocales y consonantes, así como alteraciones en la temporalidad del habla, lo que disminuye la comprensión en la comunicación [16][10][11].
Además, en los estadios avanzados de la EP, la hipomimia no solo afecta la comunicación verbal, sino que también repercute en la interacción social, ya que los pacientes experimentan una disminución en la capacidad de expresar emociones a través de gestos faciales. Esto puede generar dificultades en la percepción social y la respuesta emocional de quienes los rodean.

- Bradicinesia: Se manifiesta como una reducción progresiva y lenta de los movimientos faciales automáticos y voluntarios. Esto incluye una disminución en la frecuencia

[16] Michal NOVOTNỲ et al. “Automatic evaluation of articulatory disorders in Parkinson’s disease”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.9 (2014), pp. 1366–1378.

del parpadeo, pérdida de la gesticulación espontánea (como sonrisas, elevación de cejas o movimientos labiales), y dificultad para cambiar la expresión facial en respuesta a estímulos emocionales o interacciones sociales.

- Hipofonía: Disminución del volumen de la voz, lo que hace que el habla sea difícil de escuchar.
- Disprosodia: Alteraciones en la entonación y ritmo del habla, que pueden hacer que el tono sea monótono o poco expresivo.
- Disartria hipocinética: Articulación imprecisa de los sonidos del habla debido a la reducción de la movilidad en los músculos orofaciales.
- Dificultad en la deglución (disfagia): Puede afectar tanto la deglución de líquidos como de sólidos, lo que aumenta el riesgo de aspiración y complicaciones respiratorias.
- Seborrea facial: Exceso de producción sebácea que puede dar lugar a un rostro brillante, frecuentemente observado en pacientes con EP avanzada.
- Alteraciones en la mirada: Se reduce la amplitud y velocidad de los movimientos oculares (hipocinesia ocular), lo que puede afectar la capacidad de seguimiento visual y generar una mirada fija o ausente. Además, se ha observado una disminución en la frecuencia del parpadeo espontáneo, alteraciones en los movimientos sacádicos de movimiento suave y fijación, lo que puede afectar actividades diarias como la lectura o la interacción social. En estadios avanzados de la enfermedad, los pacientes pueden presentar apraxia oculomotora, dificultando la ejecución voluntaria de movimientos oculares, lo que refuerza la apariencia de una mirada rígida y carente de expresión.

1.2. REPRESENTACIONES POSTURALES ASOCIADAS AL MOVIMIENTO

La representación de la postura humana ha sido ampliamente estudiada en el análisis de movimiento. Por ejemplo, un método tradicional es la captura de movimiento (Mocap), que registra datos espacio-temporales del cuerpo humano para su posterior representación digital [17]. Sin embargo, su precisión depende del uso de marcadores o sensores especializados, lo que exige personal capacitado, protocolos complejos y condiciones controladas. Además, la colocación de marcadores puede ser tediosa, afectar la comodidad del paciente y alterar su marcha natural [18].Para superar estas limitaciones, han surgido sistemas basados en cámaras RGBD, que incorporan información de profundidad mediante sensores infrarrojos. Estas tecnologías han sido aplicadas en entornos clínicos para la caracterización de enfermedades motoras, aunque su fiabilidad se ve comprometida por la distancia al sensor y la interferencia de la luz solar en exteriores.[19]

Alternativamente, recientes avances en análisis de video han permitido calcular y estimar estas posturas, aprendiendo la mejor ubicación de las posturas, sin el requerimiento de sensores o marcadores en las articulaciones. Estos modelos estructurales permiten estimar la postura de un individuo a partir de la ubicación espacial de sus articulaciones[20]. A continuación, se describe la estrategia más conocida para representar posturas sin el uso de marcadores. Diferentes modelos de captura de posturas se observan en la Figura

---

[17] Fabián Vicente HORNA LÓPEZ and Luis Mifguel TARÍS RAMOS. "Diseño e implementaicón de un sistema alternativo de captura de movimiento para efectos visuales. Caso práctico: Sacrilegio del 4 de mayo de 1897". B.S. thesis. 2013.

[18] Mariana Hernández GONZÁLEZ-MONJE and Norberto MALPICA. "Sistemas basados en vídeo". In: *MANUAL SEN DE* (2021).

[19] Juan SALVATORE; Jorge OSIO, and Martín MORALES. "Detección de objetos utilizando el sensor Kinect". In: *Guayaquil, Ecuador, LACCEI* (2014).

[20] Juan Camilo LOZANO CARRILLO. "Pose temporal estimation in markerless normal human gait integrating kinematic patterns and segmented video". In: *Departamento de Imágenes Diagnósticas* ().

4.

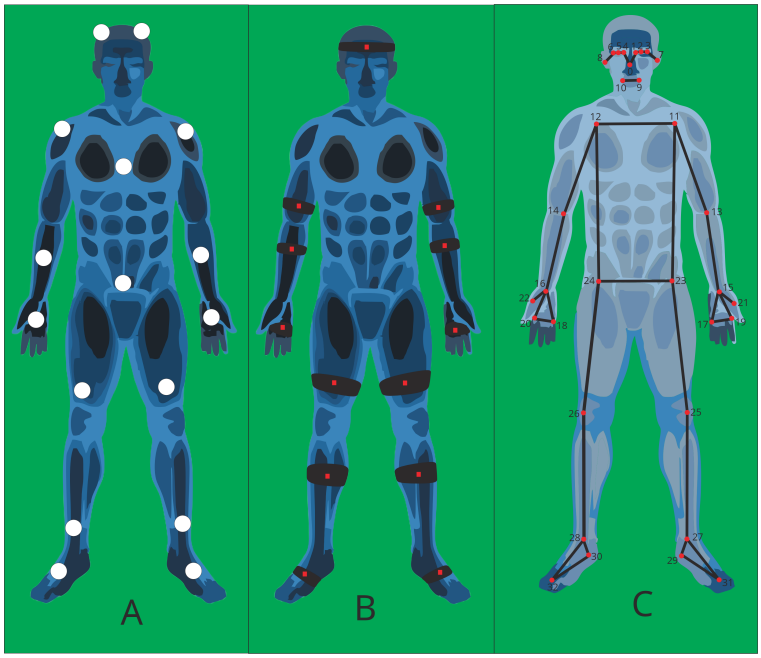


Figura 4. En A tenemos un modelo de tipo MOCAP el cual captura el movimiento por medio de marcadores reflectantes, en B el modelo inercial que captura el movimiento mediante sensores electrónicos y en C la representación del esqueleto humano de Mediapipe usando algoritmos de visión por computador.

MediaPipe: una herramienta para la estimación de posturas. El uso de técnicas de aprendizaje profundo ha permitido avances significativos en la estimación de posturas a partir de secuencias de video, proporcionando una representación más precisa de las cinemáticas asociadas a un gesto específico. Mediapipe es un "marco" computacional diseñado para la inferencia en tiempo real de poses humanas, permitiendo la detección de cuerpo, pies, manos y expresiones faciales a partir de datos de entrada en formato RGB (ver Figura 6). A continuación, veremos la arquitectura convolucional para generar los puntos de interés del rostro.

- BlazeFace: es un modelo convolucional ligero optimizado para detección de rostros y emplea convoluciones separables en profundidad para reducir la carga computa-

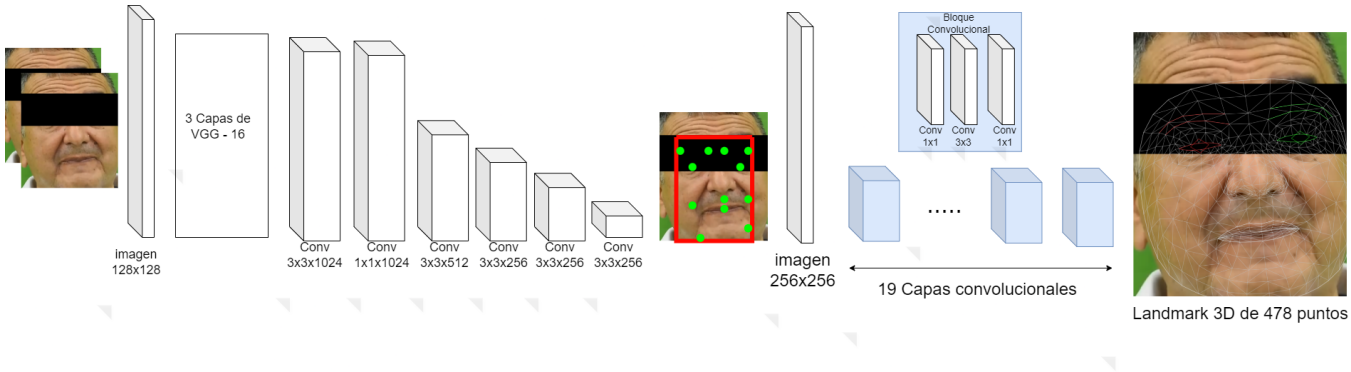


Figura 5. Etapa 1: Arquitectura convolucional (BlazeFace) con una adaptación de la VGG-16 con 3 capas convolucionales para la detección de rostros en Mediapipe. Adaptado de arxiv.org/pdf/1907.05047 Etapa 2: Arquitectura convolucional (FaceMesh y Blendshape) para la generación del landmark 3D con 478 puntos de interés. Adaptado de arxiv.org/pdf/1801.04381

cional permitiendo detecciones en tiempo real. El modelo predice puntos de interés faciales mediante aprendizaje supervisado, estableciendo correspondencias espaciales con la estructura anatómica.

- FaceMesh y BlendshapeV2: FaceMesh es un modelo convolucional diseñado para el seguimiento y reconstrucción tridimensional de la estructura facial mediante 478 puntos de referencia, sin necesidad de sensores de profundidad. Luego, BlandshapeV2 complementa a FaceMesh al estimar coeficientes de mezcla facial a partir de los landmarks detectados, permitiendo la generación y manipulación de expresiones faciales dinámicas y realistas. La combinación de ambos modelos optimiza la representación de gestos faciales en tiempo real. (ver Figura 5).

1.3. ANÁLISIS MULTIMODAL COMPUTACIONAL

El Parkinson es una enfermedad multifactorial caracterizada por múltiples síntomas y alteraciones del movimiento que pueden manifestarse en distintas etapas. Por ello, el apoyo computacional para el diagnóstico de la enfermedad de Parkinson (EP) debe formularse desde una perspectiva multimodal, que permita integrar diversas fuentes de información

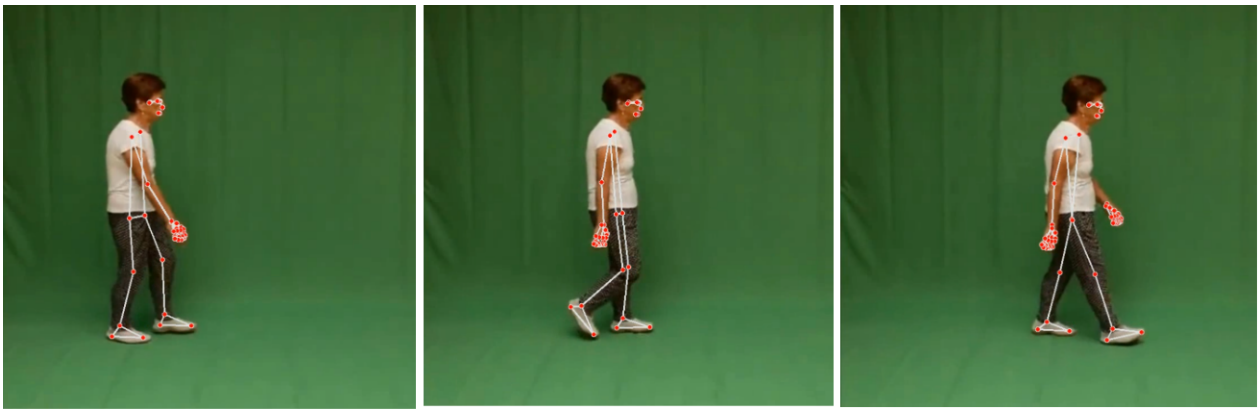


Figura 6. Ejemplo de la estimación postural de Mediapipe con un fotograma del dataset de pacientes.

sobre el movimiento y las características específicas de cada paciente evaluado. En esta sección, se presenta una revisión breve de los enfoques matemáticos y computacionales más comunes para desarrollar dichas estrategias, abarcando desde métodos estadísticos hasta representaciones multimodales profundas.

1.3.1. La covarianza como una estrategia de fusión Desde un análisis estadístico, la covarianza es un instrumento inherente para calcular la correlación lineal entre fuentes de información, constituyendo una estrategia multimodal natural para representar múltiples observaciones. Por ello, los enfoques multimodales basados en covarianza han sido ampliamente propuestos para evidenciar diferentes tendencias entre características, permitiendo, en particular en aplicaciones de la enfermedad de Parkinson (EP), resaltar discordancias en los movimientos, patrones arrítmicos y temblores. En el análisis de secuencias de video, la descripción estadística comienza calculando, para cada "frame" t, una matriz de covarianza espacial C_t relativo al conjunto de n vectores de características $F_t = \{f_{(1,t)}, \ldots, f_{(n,t)}\}$, la matriz de covarianza es calculada como

$$C_t(i,j) = \mathbb{E}\left((f_{(i,t)} - \mathbb{E}(f_{(i,t)}))(f_{(j,t)} - \mathbb{E}(f_{(j,t)}))\right) = \mathbb{E}\left(f_{(i,t)} f_{(j,t)}\right) - \mathbb{E}(f_{(i,t)})\mathbb{E}(f_{(j,t)})$$

donde la esperanza matemática $\mathbb{E}$ es calculada sobre todas las p dimensiones del vector de características $f_{(i,t)} \in \mathbb{R}^p$. De este modo, se obtiene una representación altamente compacta para cada video, permitiendo modelar patrones complejos desde una variedad de baja dimensionalidad temporal. De hecho, las matrices de covarianza pertenecen a un espacio de semi-cono, donde la dimensión real de la matriz de covarianza está dada por $\dim(C) = \frac{n(n+1)}{2}$ donde n es el número de características.

1.3.2. Análisis multimodal basado en representaciones profundas Actualmente, las representaciones profundas constituyen el estado del arte en numerosas aplicaciones de visión por computadora y análisis de imágenes. En términos generales, estos modelos construyen una representación jerárquica a partir de múltiples relaciones no lineales, ajustadas mediante la propagación del error, calculado sobre una gran cantidad de datos de entrenamiento. Estas representaciones han sido implementadas para abordar problemas multimodales desde diversas perspectivas, pero convergiendo en un espacio latente común construido a partir de diferentes modalidades. En términos generales, las estrategias multimodales profundas pueden categorizarse en fusión temprana, intermedia y tardía, como se ilustra en la Figura 7. En los siguientes ítems se darán más detalles de las mismas.

- En la Fusión temprana, las características originales se concatenan y el bloque de datos resultante se utiliza como entrada unimodal. En este sentido, la arquitectura de aprendizaje profundo no discrimina de qué modalidades provienen las características. Además, las estrategias de fusión temprana son sensibles a las posibles diferencias en las tasas de muestreo de las modalidades. Adicionalmente, la combinación intrincada de todas las modalidades en un solo descriptor dificulta interpretar la contribución de cada modalidad.

- En la fusión intermedia, las características se aprenden siguiendo ramas independientes y se combinan a partir de las respuestas de características densas y pro-

fundas previamente aprendidas. La principal ventaja de las estrategias de fusión intermedia radica en su flexibilidad para determinar la profundidad óptima de fusión entre las modalidades. De este modo, las arquitecturas de aprendizaje profundo se adaptan bien a la fusión intermedia, ya que permiten la combinación de representaciones unimodales al conectarlas a una capa compartida, facilitando la correspondencia entre representaciones jerárquicas.

- En la fusión tardía, los modelos se entrenan por separado y pueden optimizarse para aprender la probabilidad de la clase dada una modalidad x_i. La forma más sencilla de agregar las decisiones de los modelos entrenados en cada modalidad es promediando las probabilidades obtenidas mediante las funciones softmax para cada clase. Este enfoque asume una contribución equitativa de cada modelo entrenado, ya que no realiza ponderación de los resultados. Para evitar esta suposición, los valores de los pesos se calculan de manera que minimicen el error del conjunto de modalidades. Alternativamente, los pesos de las probabilidades pueden interpretarse como hiperparámetros ajustados en el conjunto de validación dentro de rangos específicos, con el objetivo de optimizar el desempeño en la tarea de clasificación.

1.3.3. Conceptos de variedades Riemannianas Cuando los datos exhiben no linealidad, la descripción matemática del espacio de datos a menudo debe alejarse de la conveniente estructura lineal de los espacios vectoriales euclidianos. La no linealidad impide la existencia de una estructura global de espacio vectorial. Sin embargo, es posible preservar ciertas propiedades matemáticas del caso euclidiano que permitan modelar con precisión los datos. En muchos casos, una de estas propiedades es la similitud local con un espacio vectorial euclidiano. En otras palabras, mediante una aproximación de segundo orden, el espacio de datos puede ser linealizado en regiones limitadas, evitando así la distorsión que introduciría forzar un modelo lineal en todo el espacio. El concepto

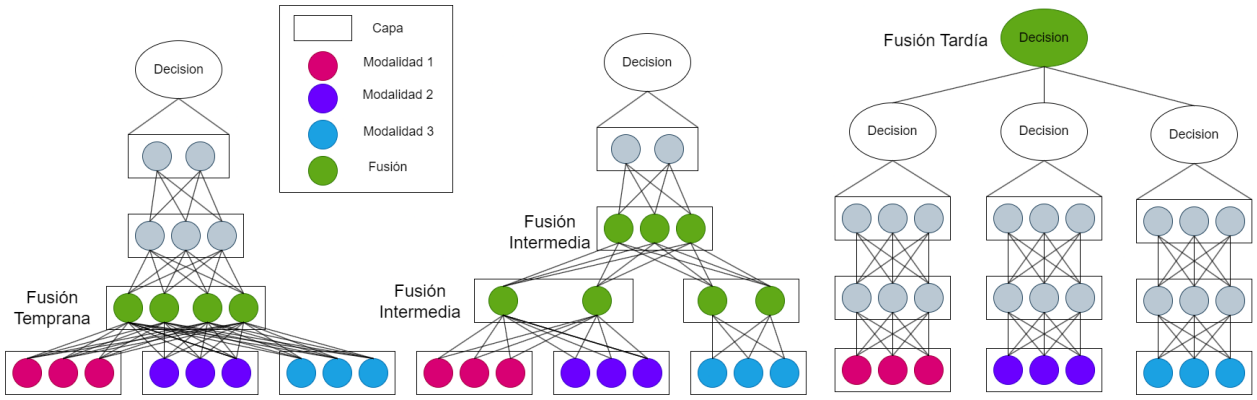


Figura 7. Estrategias de fusión en aprendizaje profundo. Las capas marcadas en verde son compartidas entre las modalidades y aprenden representaciones conjuntas. (Izquierda) En las estrategias de fusión temprana, se concatenan los arreglos de entrada de todas las modalidades. Luego, no se aprenden representaciones independientes por modalidad. (Centro) Las estrategias de fusión intermedia aprenden representaciones independientes y las fusionan en capas posteriores, ya sea específicamente en una capa o de manera gradual. (Derecha) Las estrategias de fusión tardía combinan las decisiones finales proporcionadas por cada representación aprendida por modalidad.

de similitud local con espacios euclidianos nos lleva precisamente al contexto de las variedades. Las variedades riemannianas se caracterizan por la existencia de mapas locales entre la variedad y el espacio euclidiano. Estos mapas preservan la estructura: en el caso de las variedades riemannianas, incluyen productos internos locales que codifican la geometría no euclidiana [21].

Una variedad M es un conjunto de puntos que localmente, pero no globalmente, se asemeja a un espacio euclidiano. Cuando el espacio euclidiano tiene dimensión finita, se puede relacionar sin pérdida de generalidad con $\mathbb{R}^d$ para algún $d > 0$.

Métrica Riemanniana: Una métrica riemanniana se define como una colección de productos escalares $< ., . >$ que varían suavemente en cada espacio tangente $T_x\mathrm{M}$ en los puntos x de la variedad. Para cada x, dicho producto escalar es un mapeo bilineal definido

[21] Stefan SOMMER; Tom FLETCHER, and Xavier PENNEC. “Introduction to differential and Riemannian geometry”. In: *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, 2020, pp. 3–37.

positivo $< ., . > x : Tx\mathbf{M} \times T_x\mathbf{M} \to \mathbb{R}$. Como consecuencia, el producto interno induce una norma $|\cdot|_x$.

Longitud de curva y distancia riemanniana: Si consideramos una curva $\gamma(t)$ en la variedad, podemos calcular el vector velocidad $\dot{\gamma}(t)$ y la norma $|\dot{\gamma}(t)|$, es decir, la velocidad instantánea. Para calcular la norma, se necesita una métrica riemanniana en el punto $\gamma(t)$. La longitud de la curva se obtiene integrando la norma a lo largo de la trayectoria, como:

$$I_a^b(\gamma) = \int_a^b \left\|\dot{\gamma}(t)_{\gamma(t)}\right\| dt = \int_a^b (\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)})^{\frac{1}{2}} dt$$

La longitud de la curva se define en el segmento $\gamma(a)$ a $\gamma(b)$. La distancia entre dos puntos de una variedad riemanniana corresponde a la mínima longitud que une dichos puntos, conocida como geodésica.

Mapas exponencial y logarítmico: Para proyectar datos y realizar transformaciones entre el espacio euclidiano y la variedad riemanniana, se emplean las operaciones exponencial y logarítmica. Sea x un punto en la variedad y v un vector en el espacio tangente T_xM en dicho punto. Existe una única geodésica $\gamma x, v(t)$ que parte desde $x = \gamma(x, v)(0)$ con vector tangente $v = \dot{\gamma}(x, v)(0)$.

Esto permite mapear los vectores del espacio tangente a la variedad, de modo que el vector $v \in T_xM$ se proyecta sobre la variedad siguiendo la geodésica y se denomina mapeo exponencial en el punto x: $Exp_x : T_xM \to M$, esto es: $Exp_x : v \to Exp_v = \gamma(x, v)$. De manera recíproca, $\vec{xy}$ o $log_x(y)$ se conoce como mapeo logarítmico, definido como el vector más corto, medido mediante una métrica riemanniana, tal que: $y = Exp_x(\vec{xy})$.

1.4. MÉTODOS COMPUTACIONALES PARA LA CUANTIFICACIÓN MOTORA DEL PARKINSON

En pacientes con enfermedad de Parkinson (EP), la marcha se ve afectada por patrones relacionados con la inestabilidad postural, los temblores, la rigidez y la bradicinesia [22]. Para la caracterización de la marcha, en la literatura se han propuesto alternativas que utilizan sensores inerciales (IMU, por sus siglas en inglés) [23] o sensores de fuerza ubicados en las suelas de los pies [24]. Estas alternativas clásicas producen señales temporales dedicadas principalmente al análisis de las extremidades inferiores [25]. Sin embargo, estos enfoques pueden ser intrusivos, alterar el carácter de la marcha y carecer de sensibilidad. Otros enfoques basados en sensores IMU pueden implicar múltiples variables cinemáticas para considerar modelos más complejos que incluyan múltiples partes del cuerpo, pero dicho enfoque sigue siendo intrusivo y requiere sistemas de calibración sofisticados [26]. Como alternativa, el uso de cámaras comunes simplifica considerablemente el protocolo clínico, reduce considerablemente la incomodidad de los pacientes y potencialmente

---

[22] Ana Beatriz Ramalho Leite SILVA et al. "Premotor, nonmotor and motor symptoms of Parkinson's disease: a new clinical state of the art". In: *Ageing Research Reviews* 84 (2023), p. 101834.

[23] Elham RASTEGARI; Sasan AZIZIAN, and Hesham ALI. "Machine learning and similarity network approaches to support automatic classification of parkinson's diseases using accelerometer-based gait analysis". In: (2019).

[24] Lazzaro di BIASE et al. "Parkinson's disease wearable gait analysis: kinematic and dynamic markers for diagnosis". In: *Sensors* 22.22 (2022), p. 8773.

[25] Qinghui WANG; Wei ZENG, and Xiangkun DAI. "Gait classification for early detection and severity rating of Parkinson's disease based on hybrid signal processing and machine learning methods". In: *Cognitive Neurodynamics* (2022), pp. 1–24.

[26] Dante TRABASSI et al. "Machine learning approach to support the detection of Parkinson's disease in IMU-based gait analysis". In: *Sensors* 22.10 (2022), p. 3700.

mejora la evaluación de patrones anormales relacionados con la EP [27] [28]. Por ejemplo, en la literatura se propuso una red neuronal convolucional en 3D para clasificar y resaltar regiones destacadas de las extremidades y otras partes del cuerpo [29]. Sin embargo, estos mapas de prominencia solo proporcionan información cualitativa y aún deben correlacionarse con los valores cuantitativos de las etapas para una interpretación más completa de la enfermedad. Alternativamente, se han utilizado estrategias de análisis de video para extraer el esqueleto de una persona [30]. En este trabajo la naturalidad del movimiento no se ve afectada, pero esta simplificación requiere de una acertada selección de puntos de interés para evitar una reducción en la caracterización de los patrones de movimiento.
El análisis de la expresión facial y su nivel de expresividad ha sido abordado mediante representaciones del movimiento para evaluar la capacidad de expresar e imitar emociones como tristeza, ira y disgusto[31]. Recientemente, enfoques computacionales han mejorado la caracterización de la hipomimia empleando redes neuronales convolucionales (CNN) adaptadas a partir de estimaciones del movimiento facial, como puntos de refer-

---

[27] Rachneet KAUR et al. "A Vision-Based Framework for Predicting Multiple Sclerosis and Parkinson's Disease Gait Dysfunctions—A Deep Learning Approach". In: *IEEE Journal of Biomedical and Health Informatics* 27.1 (2022), pp. 190–201.

[28] Peipei LIU et al. "Quantitative assessment of gait characteristics in patients with Parkinson's disease using 2D video". In: *Parkinsonism & Related Disorders* 101 (2022), pp. 49–56.

[29] Luis C GUAYACÁN and Fabio MARTÍNEZ. "Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network". In: *Journal of biomedical informatics* 123 (2021), p. 103935.

[30] Mohamed CHERIET et al. "Multi-speed transformer network for neurodegenerative disease assessment and activity recognition". In: *Computer Methods and Programs in Biomedicine* 230 (2023), p. 107344.

[31] YANG LIQIONG et al. "Changes in facial expressions in patients with Parkinson's disease during the phonation test and their correlation with disease severity". In: *Computer Speech & Language* 72 (2022), p. 101286.

encia y cuantificación del desplazamiento[32][33]. Algunos estudios integran características estáticas (fotogramas) y dinámicas (videos) derivadas de representaciones profundas y puntos faciales[34][35]. No obstante, la mayoría de estos enfoques se basan en expresiones faciales naturales o inducidas, no consideran ejercicios lingüísticos practicados frecuentemente en la rutina clínica. De manera reciente, también se han introducido métodos computacionales para cuantificar patrones motores a partir de diversas fuentes de información, utilizando modalidades de habla y actividades de escritura codificadas en patrones de frecuencia, que luego se asignan a clasificadores clásicos para discriminar entre la población Parkinson y control [36]. Sin embargo, esta estrategia modela cada modalidad del movimiento de forma independiente, perdiendo información de co-ocurrencia que podría enriquecer los descriptores de la enfermedad de Parkinson. La adición de la modalidad de marcha a la escritura y al habla se realizó para categorizar la enfermedad de Parkinson de acuerdo con la escala MDS-UPDRS [37]. Estas modalidades se modelaron individualmente con redes neuronales convolucionales, y los embebidos resultantes se concatenaron y se

---

[32] Avner ABRAMI et al. “Automated computer vision assessment of hypomimia in Parkinson disease: proof-of-principle pilot study”. In: *Journal of medical Internet research* 23.2 (2021), e21037.

[33] Bo JIN et al. “Diagnosing Parkinson disease through facial expression recognition: video analysis”. In: *Journal of medical Internet research* 22.7 (2020), e18697.

[34] Karen SIMONYAN and Andrew ZISSERMAN. “Very deep convolutional networks for large-scale image recognition. arXiv 2014”. In: *arXiv preprint arXiv:1409.1556* 1409 (2014).

[35] Luis F GOMEZ et al. “Improving parkinson detection using dynamic features from evoked expressions in video”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 1562–1570.

[36] Hung N PHAM et al. “Multimodal detection of Parkinson disease based on vocal and improved spiral test”. In: *2019 International Conference on System Science and Engineering (ICSSE)*. IEEE. 2019, pp. 279–284.

[37] Juan Camilo VÁSQUEZ-CORREA et al. “Comparison of user models based on GMM-UBM and i-vectors for speech, handwriting, and gait assessment of Parkinson’s disease patients”. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 6544–6548.

utilizaron para predecir la enfermedad [38]. Otros enfoques han utilizado las modalidades de expresión facial y habla (asociadas con hipomimia y disartria en la enfermedad de Parkinson), cuantificadas como características diseñadas a mano para clasificar la enfermedad de Parkinson [39]. Además, se han integrado representaciones multimodales convolucionales con descriptores de covarianza para detectar la enfermedad de Parkinson [40]. Sin embargo, este enfoque utiliza únicamente la covarianza como método de agrupación y no aprovecha el potencial de la geometría de Riemann subyacente.

---

[38] Juan Camilo VÁSQUEZ-CORREA et al. "Multimodal assessment of Parkinson's disease: a deep learning approach". In: *IEEE journal of biomedical and health informatics* 23.4 (2018), pp. 1618–1630.

[39] Justyna SKIBIŃSKA and Jiri HOSEK. "Computerized analysis of hypomimia and hypokinetic dysarthria for improved diagnosis of Parkinson's disease". In: *Heliyon* 9.11 (2023).

[40] J. ARCHILA; A. MANZANERA, and F. MARTINEZ. "A multimodal Parkinson quantification by fusing eye and gait motion patterns, using covariance descriptors, from non-invasive computer vision". In: *Computer Methods and Programs in Biomedicine* 215 (2022), p. 106607. DOI: 10.1016/j.cmpb.2021.106607.

2. PROBLEMA DE INVESTIGACIÓN

La enfermedad de Parkinson es el segundo trastorno neurodegenerativo más frecuente a nivel mundial, afectando a 150 por cada 100.000 habitantes[41]. Esta enfermedad es de ámbito multifactorial con múltiples afectaciones motoras, que varían en intensidad y frecuencia y pueden manifestarse en diferentes estadios. Sin embargo, no existe un biomarcador definitivo que permita un diagnóstico temprano, lo que dificulta la implementación de estrategias de tratamiento y seguimiento personalizadas.

El diagnóstico se basa en la observación de actividades motrices, lo que lo hace altamente dependiente de la experticia del especialista. Como consecuencia, los errores en el diagnóstico pueden variar entre el 40% y el 6% dependiendo de si el evaluador es un médico general o un especialista en desórdenes del movimiento[42].

Esta variabilidad en la evaluación clínica limita la caracterización global de la población con Parkinson, ya que los pacientes pueden desarrollar discapacidades motoras en diferentes regiones del cuerpo y con manifestaciones diversas, como la bradicinesia al caminar o la rigidez facial.

Los enfoques unimodales presentan limitaciones cuando los síntomas motores predominantes en un paciente no son perceptibles a través de la modalidad analizada, generando sesgos en la evaluación.

La caracterización de los síntomas motores a partir de videos plantea múltiples desafíos. Las manifestaciones del Parkinson pueden observarse en modalidades focalizadas, como la expresión facial, o en modalidades generales, como la marcha. Sin embargo, la captura

---

41 Valery L FEIGIN et al. “Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015”. In: *The Lancet Neurology* 16.11 (2017), pp. 877–897.

42 TINELLI; KANAVOS, and GRIMACCIA, see n. 4.

de estos movimientos en video introduce variabilidad en la resolución espacial y temporal, lo que dificulta la comparación directa entre modalidades.

Además, la heterogeneidad en la progresión de la enfermedad implica que un mismo paciente puede presentar alteraciones motoras con diferentes grados de severidad en distintos momentos, lo que hace que la identificación de patrones comunes en la población sea un reto significativo. La falta de mecanismos que permitan relacionar información capturada desde perspectivas y escalas temporales distintas genera incertidumbre en la evaluación de la enfermedad.

El desafío central radica en cómo abordar la integración de estas modalidades sin introducir distorsiones que comprometan la precisión en la caracterización de la enfermedad. La dependencia del diagnóstico clínico tradicional y la ausencia de enfoques que consideren de manera simultánea distintas manifestaciones motoras siguen siendo barreras en la comprensión y monitoreo del Parkinson. Considerando los retos anteriores surge la pregunta de investigación:

¿Cómo aprender una estrategia que integre puntos de interés de la expresión facial y la marcha con la capacidad para discriminar patrones parkinsonianos?

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Desarrollar una estrategia de aprendizaje profundo para clasificar patrones de la marcha y la expresión facial asociados al Parkinson.

3.2. OBJETIVOS ESPECÍFICOS

- Seleccionar un conjunto de videos de marcha y expresión facial en una población con Parkinson y una población control para el desarrollo, ajuste y validación de la metodología desarrollada.
- Implementar una arquitectura profunda para extraer puntos de referencia postural durante la marcha y marcadores espaciales del rostro durante un ejercicio de comunicación.
- Desarrollar descriptores espaciotemporales que permitan codificar y clasificar patrones de movimiento desde los puntos de referencia recuperados en ejercicios de marcha y comunicación.
- Validar la metodología propuesta en cuanto a la capacidad de discriminar patrones asociados al Parkinson y a la población control.

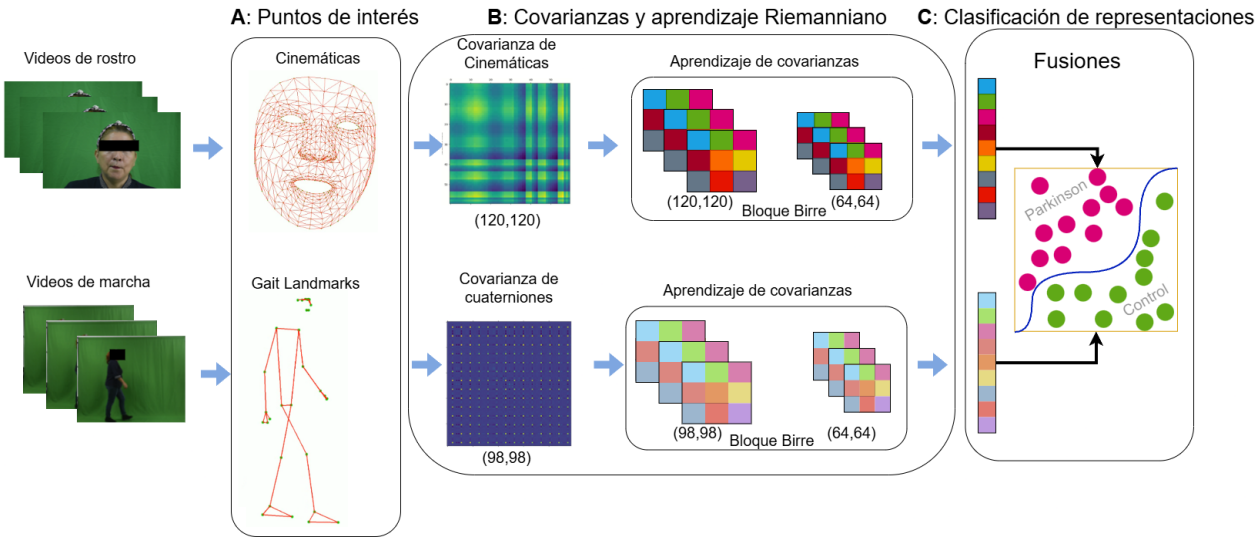
4. MÉTODO PROPUESTO

Este trabajo presenta una estrategia de aprendizaje geométrico multimodal para detectar patrones motores asociados a la enfermedad de Parkinson mediante la integración de la expresión facial y la marcha. Este método inicialmente extrae puntos de interés espaciotemporales, como base de la representación, sobre los cuales se calculan cinemáticas que describen el comportamiento durante ejercicios de locomoción (en marcha) y habla (rostro). Las características desde los puntos de interés se compactan en descriptores de covarianza, que permiten inferir patrones de coordinación y asociaciones entre puntos para llevar a cabo los movimientos que producen la marcha y el habla. Estos descriptores son simétricos y positivos, coexistiendo en una variedad geométrica de Riemman. Entonces, vectores compactos que mantengan esta geometría son aprendidos desde representaciones Riemanianas, permitiendo extraer cactacterísticas con alto valor asociado al carcáter discriminatorio entre el Parkinson y pacientes control. Luego, se realiza la clasificación multimodal mediante estrategias de fusión temprana, intermedia y tardía utilizando redes densas seguidas de funciones softmax para obtener la probabilidad final del modelo. En la figura 8 se ilustra el método propuesto y los componentes que lo conforman. En las siguientes secciones se da un mayor detalle de cada una de las partes.

4.1. PUNTOS DE INTERÉS Y DESCRIPCIÓN ESPACIO-TEMPORAL

En este trabajo estamos interesados en fusionar patrones multimodales que permitan brindar un soporte diagnóstico a la enfermedad del Parkinson. Para ello, se decidió seleccionar ejercicios de marcha y habla, los cuales en la literatura han sido ampliamente referenciados como biomarcadores de la enfermedad, incluso con hallazgos en estadios

Figura 8. Esquema de la estrategia propuesta. (A) Se obtienen características faciales y de marcha por medio de Mediapipe mediante la extracción de los puntos de referencia, (B) se extraen las matrices y se computan las covarianzas de las caracteristicas, (C) se realiza la clasificación usando diferentes métodos de integración.



tempranos y para la mayoría de fenotipos de la enfermedad[43]. Para la caracterización de estas dos modalidades, como representación primaria, se decidió codificar y extraer puntos de interés (*landmarks*) durante los ejercicios de marcha y habla. Para este trabajo, los puntos de interés en las dos modalidades consideradas fueron obtenidas desde el framework de MediaPipe[44]. MediaPipe permite extraer puntos de referencia posturales durante la marcha, así como también puntos de interés en el rostro durante el habla, sin el requerimiento de calibraciones específicas y sin la necesidad de marcadores invasivos que restringen la ejecución natural de cada uno de los ejercicios.

---

43 L RICCIARDI; A DE ANGELIS, et al. "Hypomimia in Parkinson's disease: an axial sign responsive to levodopa". In: *European Journal of Neurology* 27.12 (2020), pp. 2422–2429; Avner ABRAMI et al. "Automated computer vision assessment of hypomimia in Parkinson disease: proof-of-principle pilot study". In: *Journal of medical Internet research* 23.2 (2021), e21037.

44 Ivan GRISHCHENKO et al. "Attention mesh: High-fidelity face mesh prediction in real-time". In: *arXiv preprint arXiv:2006.10962* (2020); Valentin BAZAREVSKY et al. "Blazeface: Sub-millisecond neural face detection on mobile gpus". In: *arXiv preprint arXiv:1907.05047* (2019).

En nuestro enfoque, los puntos de interés detectados, son definidos como un conjunto de posiciones que varían espacial y temporalmente, tanto para la marcha $\mathbf{P^m} = \{\{\mathbf{p_1^{m,i}}\}_{i=1...n} \ldots \{\mathbf{p_T^{m,i}}\}_{i=1...n}\}$, como para el rostro, durante los ejercicios del habla $\mathbf{P^r} = \{\{\mathbf{p_1^{r,i}}\}_{i=1...n} \ldots \{\mathbf{p_T^{r,i}}\}_{i=1...k}\}$. En ambos casos i representa el iterador del número de puntos de interés, los cuales van describiéndose en T cuadros de video. Cada uno de estos j puntos $\mathbf{p_i^{(m,r),j}} = (x, y, z, t)$ esta descrito con respecto a las coordenadas del video, espacialmente en cada cuadro (x, y, z) y con respecto al tiempo en el que transcurre el cuadro t. A partir de estos puntos de interés, a continuación, explicaremos el descriptor obtenido en cada modalidad para codificar patrones con información motora de las poses y expresiones faciales Parkinsonianas.

Marcha: Durante la locomoción, inicialmente las coordenadas espaciales de la postura $\mathbf{P^m}$, para cada cuadro t, se mapeó a una representación en cuaterniones $Q_t^{m,i}$. Los cuaterniones han demostrado ser una representación ideal de sistemas dinámicos, que permiten ser robustos a variaciones en las posiciones, así como también brindan una capacidad de descripción dinámica de las variaciones angulares entre los segmentos de la postura. Estos cuaterniones han sido ampliamente usados en la literatura para la descripción de posturas [454647]. Los puntos de interés $\mathbf{p_i^j} = (x, y, z, t)$ se convertirán en representaciones $Q_t^{m,i} = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ donde w, x, y, z son números reales y $\mathbf{i}, \mathbf{j}, \mathbf{k}$ representan unidades imaginarias, se pueden construir este tipo de representaciones apartir de $Q_t^{m,i} = \cos\frac{\theta^{m,i}}{2} + \vec{u}^{m,i}\sin\frac{\theta^{m,i}}{2}$ donde $\vec{u}^{m,i} = (u_x^{m,i}, u_y^{m,i}, u_z^{m,i})$ es el eje de rotación uni-

---

45 Maxime DEVANNE et al. “Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation”. In: *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE. 2017, pp. 529–534.

46 Timo KUHLGATZ et al. “On Stair Walk Recognition Using a Single Magnetometer-free IMU and Deep Learning”. In: *Current Directions in Biomedical Engineering*. Vol. 10. 4. De Gruyter. 2024, pp. 404–407.

47 Katerina FRAGKIADAKI et al. “Recurrent network models for human dynamics”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4346–4354.

tario. Alternativamente, un cuaternión se puede representar de una forma más explicita $Q_t^{m,i} = \cos\frac{\theta^{m,i}}{2} + u_x \sin\frac{\theta^{m,i}}{2}\mathbf{i} + u_y \sin\frac{\theta^{m,i}}{2}\mathbf{j} + u_z \sin\frac{\theta^{m,i}}{2}\mathbf{k}$. Para obtener la representación de cuaternios $Q_t^{m,i}$, entre dos segmentos corporales, por ejemplo entre el brazo y el antebrazo, debemos realizar las siguientes operaciones:

- **Primer paso:** Dados 3 puntos de interés $\mathbf{p_i^{:j}} = (x, y, z, t)$ se pueden hallar dos **vectores** $\vec{v}_1^{m,i}, \vec{v}_2^{m,i}$ restando la posición espacial de 3 puntos de interés $\mathbf{p_1^{:j}}, \mathbf{p_2^{:j}}, \mathbf{p_3^{:j}}$.
- **Segundo paso:** Es necesario normalizar los vectores $\vec{v}_1^{m,i}, \vec{v}_2^{m,i}$ con respecto a la estatura de cada participante, podemos restar cada vector con la posición absoluta de la columna y dividiendo por la longitud de la columna(desde la vértebra lumbar L1 hasta la vértebra torácica T1), para esto usamos $\vec{v}^{m,i} = \frac{(\vec{v}^{m,i} - p^{j,columna})}{L_{columna}}$.
- **Tercer paso:** Calcular el ángulo entre los vectores normalizados, para esto se usa $\theta^{m,i} = \cos^{-1}\left(\frac{\vec{v}_1^{m,i} \cdot \vec{v}_2^{m,i}}{\|\vec{v}_1^{m,i}\| \, \|\vec{v}_2^{m,i}\|}\right)$
- **Cuarto paso:** Calcular el vector unitario de rotación, para esto podemos usar $\vec{u}_{rotacin}^{m,i} = \vec{v}_1^{m,i} \times \vec{v}_2^{m,i}$
- **Quinto paso:** Construir el cuaternión $Q_t^{m,i}$ a partir de cada vector de rotación $\vec{u}_i^{m,i}$ y angulo entre los vectores $\theta_i^{m,i}$ reemplazando cada valor en la forma explícita del cuaternión.
- **Sexto paso:** Aplicar la operación logarítmica al cuaternion $Q_t^{m,i}$ con el fin de obtener una representación en el espacio euclidiano.

$$\log(Q_t^{m,i}) = \left(\theta^{m,i}\frac{u_x}{\|\vec{u}\|}, \theta^{m,i}\frac{u_y}{\|\vec{u}\|}, \theta^{m,i}\frac{u_z}{\|\vec{u}\|}\right)$$

Finalmente, para cada uno de los cuaterniones $Q_t^{m,i}$ le concatenaremos la posición relativa al cuaternio $\mathbf{p_2^{:j}}$. De esta manera obtenemos una colección de cuaterniones $Q = \{Q_t^{m,i}; p_t^{:j} | i = 1, \ldots, t = 1, \ldots, T\}$ que es una representación más informativa, representando el movimiento incluyendo la rotación. En la figura 9 podemos ver los cuaterniones construidos.

Rostro: En este ejercicio del habla, para cada fotograma t se construyó una representación cinemática P, V, A usando las coordenadas espaciales del rostro (conjuntos de puntos de interés). Se concatenaron primero los valores del eje x, seguidos por los valores del eje y, formando una única representación para cada instante de tiempo.

- **Primera representación:** Se obtuvo una colección de puntos de interés faciales en cada fotograma, definida como:

$$\mathbf{P} = \left\{ \{\mathbf{p_1^{r,i}}\}_{i=1\ldots n} \ldots \{\mathbf{p_T^{r,i}}\}_{i=1\ldots n} \right\}$$

  donde cada punto $\mathbf{p_t^{r,i}} = (x, y, t)$ representa la posición bidimensional de un punto de interés facial en el instante t. La representación final de la posición se construyó concatenando todas las posiciones x y luego todas las posiciones y:

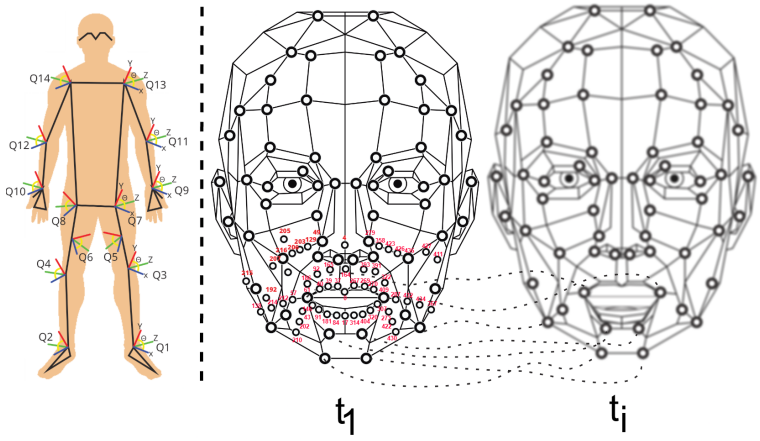
$$\mathbf{P}_t = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]^T .$$

- **Segunda representación:** Se calculó la velocidad de los puntos de interés faciales a partir de la diferencia entre posiciones en cuadros consecutivos: $\mathbf{V} = \left\{ \{\mathbf{v_1^{r,i}}\}_{i=1\ldots n} \ldots \{\mathbf{v_T^{r,i}}\}_{i=1\ldots n} \right\}$, donde la velocidad de cada punto de interés i en el instante t se define como: $v_t^{r,i} = \frac{\mathbf{p_t^{r,i}} - \mathbf{p_t^{r,i+1}}}{\Delta t}$. La representación final de la velocidad se obtuvo concatenando todas las velocidades x y luego todas las velocidades y: $\mathbf{V}_t = [v_{x_1}, v_{x_2}, \ldots, v_{x_n}, v_{y_1}, v_{y_2}, \ldots, v_{y_n}]^T$.

- **Tercera representación:** Se obtuvo la aceleración de los puntos de interés faciales a partir de la diferencia entre velocidades en cuadros consecutivos: $\mathbf{A} = \left\{ \{\mathbf{a_1^{r,i}}\}_{i=1\ldots n} \ldots \{\mathbf{a_T^{r,i}}\}_{i=1\ldots n} \right\}$, donde la aceleración de cada punto de interés i en el instante t se calcula como: $a_t^{r,i} = \frac{\mathbf{v_t^{r,i}} - \mathbf{v_t^{r,i+1}}}{\Delta t}$. La representación final de la aceleración se construyó concatenando todas las aceleraciones x y luego todas las aceleraciones y: $\mathbf{A}_t = [a_{x_1}, a_{x_2}, \ldots, a_{x_n}, a_{y_1}, a_{y_2}, \ldots, a_{y_n}]^T$

Estos tres tipos de representaciones proporcionan una descripción más completa de los movimientos faciales durante los ejercicios del habla. Así obtenemos tres colecciones de cinemáticas P, V, A que pueden ser complementarias en la representación motora. En la figura 9 podemos ver las trayectorias del rostro en el ejercicio de la comunicación.

Figura 9. A la izquierda los cuaterniones construidos a partir de 14 tripletas de puntos de interés y a la derecha las trayectorias del rostro a partir de 60 puntos de interés en el ejercicio de la comunicación.



4.2. DESCRIPTORES DE COVARIANZA PARA MARCHA Y ROSTRO

En este trabajo se calcularon los cuaterniones y las cinemáticas faciales mapeados a matrices a partir de puntos de referencia (ver Figura 10). Estos cuaterniones y cinemáticas faciales representan las características que se codificarán en descriptores correlativos de movimiento. Es importante resaltar que un patrón fundamental en Parkinson es la coordinación local de puntos claves, articulares, durante el desarrollo de un movimiento. Es por ello, que los descriptores correlativos resultan fundamentales para hallar aproximaciones de coordinación que puedan ser discriminativos al Parkinson. Entonces, en este trabajo se definieron características de movimiento para modelar las dependencias estadísticas, proporcionando una representación compacta, de segundo orden de los patrones de movimiento. A continuación, detallamos como se calcularon estos descriptores

en cada modalidad.

Marcha: En esta modalidad, se calculó la covarianza $\mathbf{C_Q} = \frac{1}{T-1}\sum_{t=1}^{T}(Q_t^{m,i} - \bar{Q})(Q_t^{m,i} - \bar{Q})^T$, donde

- $Q_t^{m,i}$ representa el cuaternión asociado al sujeto m, en la dimensión i (una de las cuatro componentes del cuaternión) en el instante de tiempo t.
- $\bar{Q}$ es la media de los cuaterniones a lo largo del tiempo.
- t denota el instante de tiempo dentro de la secuencia de movimiento.
- T es el tiempo total.
- $\mathbf{C_Q}$ es la matriz de covarianza de los cuaterniones, que captura la relación estadística entre las diferentes componentes de los movimientos a lo largo del tiempo.

En este descriptor de covarianza planeamos entonces corelaccionar posturas entre cuadros, para describir la marcha. Una ilustración se puede observar en la figura 10.

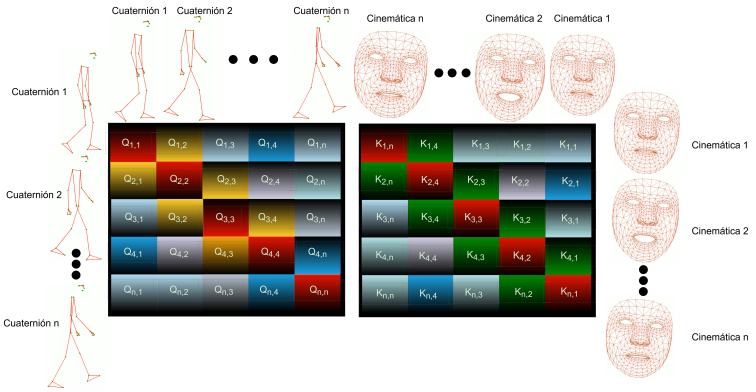


Figura 10. Ilustración de las covarianzas entre poses $\mathbf{C_Q}$ y $\mathbf{C_k}$ a lo largo del tiempo T.

Rostro: Para el análisis de la expresión facial, se construyeron covarianzas a partir de tres representaciones cinemáticas k: posición $(\mathbf{M}_p)$, velocidad $(\mathbf{M}_v)$ y aceleración $(\mathbf{M}_a)$. De manera general, siendo $(\mathbf{M}_k)$ alguna de las tres cinemáticas. Entonces el descriptor de covarianza es definido como: $\mathbf{C_k} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{M}_k^t - \bar{\mathbf{M}}_k)(\mathbf{M}_k^t - \bar{\mathbf{M}}_k)^T$, donde:

- $\mathbf{M}_k^t$ representa la matriz de características de la representación cinemática k (posición, velocidad o aceleración) en el instante de tiempo t.
- $\bar{\mathbf{M}}_k$ es la media de la matriz de características a lo largo del tiempo.
- k denota la representación cinemática específica utilizada: $k \in \{p, v, a\}$ para posición, velocidad y aceleración, respectivamente.
- T es el tiempo total.
- $\mathbf{C_k}$ es la matriz de covarianza para la cinemática k, que modela la variabilidad y las relaciones estadísticas de las características faciales a lo largo del tiempo.

Estos descriptores de covarianza pueden no solo aproximar coordinación entre puntos de interés, sino aproximar otros patrones anormales de movimiento asociados a la enfermedad. En particular, la afectación de los músculos cigomáticos y orbiculares genera limitaciones en la gesticulación verbal, dificultando la comunicación y manifestándose en la hipomimia facial, un rasgo distintivo del trastorno conocido como "cara de máscara". Estos patrones podrían resaltarse en los descriptores calculados, teniendo en cuenta los valores cinemáticos correlacionados. De manera análoga, en la marcha se evidencian signos patológicos como rigidez, alteraciones posturales y bradicinesia, este último considerado un síntoma cardinal en el diagnóstico diferencial del Parkinson. Del mismo modo, estas matrices de covarianza en la marcha pueden exhibir estos patrones y ser claves para la diferenciación con respecto a observaciones de una población control.

4.3. APRENDIZAJE GEOMÉTRICO DE REPRESENTACIONES MÁS COMPACTAS

Una vez obtenidos las matrices de covarianza para cada modalidad, se puede aprender un espacio geométrico de estas, para obtener descriptores que extraigan las principales relaciones de las matrices, sean más compactas y permitan una mayor captura de información discriminativa con respecto a la clasificación. Para abordar este desafío, se empleó la red SPDNet, un modelo de aprendizaje geométrico diseñado para procesar

matrices de covarianza, reduciendo su dimensionalidad mientras preserva la estructura intrínseca de los datos. Esta red está compuesta por tres capas fundamentales: BiMap, ReEig y LogEig [48]. Además, en siguientes subsecciones se definirá el Descriptor Multimodal D_M el cual es la representación combinada en la fusión temprana.

Transformación Bilineal (BiMap) . La capa BiMap implementa una transformación bilineal que proyecta las matrices de covarianza a una representación más compacta: $C_{\text{salida}} = W C_{\text{entrada}} W^\top, \quad W \in \mathbb{R}^{M \times N}$, donde:

- C_{salida}: Es la matriz de covarianza transformada, obtenida tras la proyección bilineal.
- W: Es la matriz de proyección de dimensiones $M \times N$, utilizada para reducir la dimensionalidad.
- C_{entrada}: Es la matriz de covarianza original antes de la transformación, que representa la marcha o la expresión facial.
- $W^\top$: Es la transpuesta de la matriz de proyección, necesaria para garantizar la preservación de propiedades geométricas.

Esta transformación garantiza que la matriz resultante conserve sus propiedades de simetría y positividad definida (SPD). Sin embargo, la reducción de dimensionalidad puede introducir errores computacionales y numéricos, lo que requiere mecanismos adicionales para verificar la estabilidad de las representaciones obtenidas.

Rectificación espectral (ReEig) . La capa ReEig descompone cada matriz de covarianza en autovalores y autovectores mediante la descomposición espectral con el objetivo de preservar las características geométricas de la covarianza: $\hat{C} = U \Sigma U^\top, \quad \hat{C} \in \mathbb{R}^{n \times n}, \quad U \in \mathbb{R}^{n \times n}, \quad \Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, donde:

---

[48] 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.

- $\hat{C}_{Movimiento}$: Es la matriz de covarianza que representa la variabilidad en la marcha o la expresión facial.
- U: Es la matriz ortogonal cuyas columnas son los autovectores de $\hat{C}_{Movimiento}$.
- Σ: Es la matriz diagonal cuyos elementos son los autovalores λ_i de $\hat{C}_{Movimiento}$.
- diag$(\lambda_1, \lambda_2, \ldots, \lambda_n)$: Es la representación diagonal de los autovalores, que conserva la información esencial de la matriz de covarianza.

Proyección Tangente (LogEig) : Finalmente, la capa LogEig proyecta las matrices de covarianza al espacio tangente en el punto de referencia de la variedad de Riemann, permitiendo una representación más adecuada para la clasificación en espacios euclidianos. Esta transformación se define aplicando el logaritmo natural a los autovalores de la matriz de covarianza: $\log(\hat{C}_{Movimiento}) = U \log(\Sigma) U^\top, \quad \log(\Sigma) = (\log(\lambda_1), \log(\lambda_2), \ldots, \log(\lambda_n))$, donde:

- $\log(\hat{C}_{Movimiento})$: Es la transformación de la matriz de covarianza del espacio Riemanniano al plano tangente.
- U: Es la matriz de autovectores obtenidos en la descomposición espectral.
- $\log(\Sigma)$: Es la matriz diagonal donde se aplica el logaritmo natural a cada autovalor de $\hat{C}_{Movimiento}$.
- $\log(\lambda_1), \log(\lambda_2), \ldots, \log(\lambda_n)$: Son los autovalores transformados mediante el logaritmo natural, permitiendo su representación en un espacio euclidiano.

4.4. DESCRIPTOR MULTIMODAL D_M Y CLASIFICACIÓN RIEMANNIANA

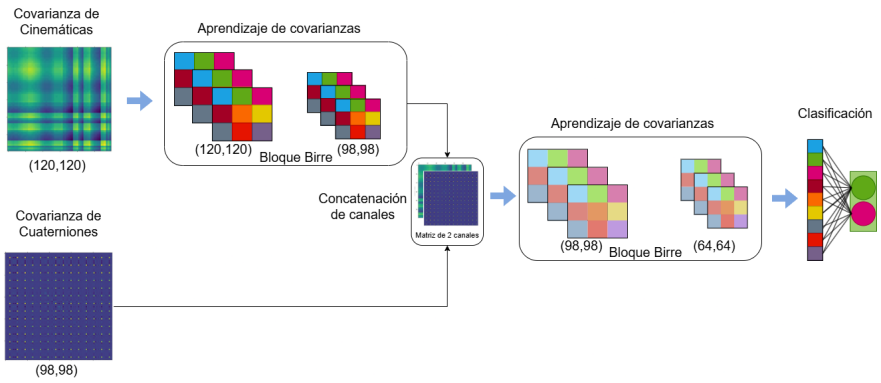
En este trabajo se exploraron diferentes esquemas de integración que permitan aprovechar de forma efectiva la complementariedad de las modalidades involucradas. La forma en que se fusionan estas representaciones tiene un impacto directo en la capacidad del modelo para capturar patrones discriminativos. Por ello, se evaluaron tres enfoques de

fusión de covarianzas (temprano, intermedio y tardío) que varían en el momento en que se realiza la integración de la información y en la naturaleza de la interacción entre las modalidades. A continuación, se describe brevemente cada una de estas estrategias:

Fusión temprana: implica la combinación directa de las representaciones de ambas modalidades antes del aprendizaje riemanniano. Esta estrategia busca capturar de forma intrínseca las correlaciones intermodales dentro del espacio de Riemann, aprovechando la estructura geométrica común durante el entrenamiento.

En esta configuración particular de fusión, se siguió el esquema mostrado en la figura 11. En este caso se aprendió una representación Riemmaniana desde una componente BIRE y el descriptor fue fusionado con el descriptor SPD, formado desde los cuaterniones, en el proceso de la marcha. En este caso, el descriptor multimodal $D_M = \{\hat{C}_q; \hat{C}_k\}$ se obtiene al concatenar las matrices resultantes en cada modalidad, las cuales son apiladas en dos canales, y en el siguiente paso se realiza un aprendizaje conjunto multimodal, usando la geometría de Riemman. Después de aprovechar este aprendizaje en conjunto se procede a mapear el descriptor a un espacio Euclidiano para proceder con la respectiva clasificación. En este último paso, se toma la triangular superior del vector D_M para evitar redundancias y hacer la clasificación.

Figura 11. Arquitectura con los mejores resultados en la fusión temprana: Concatenando canales

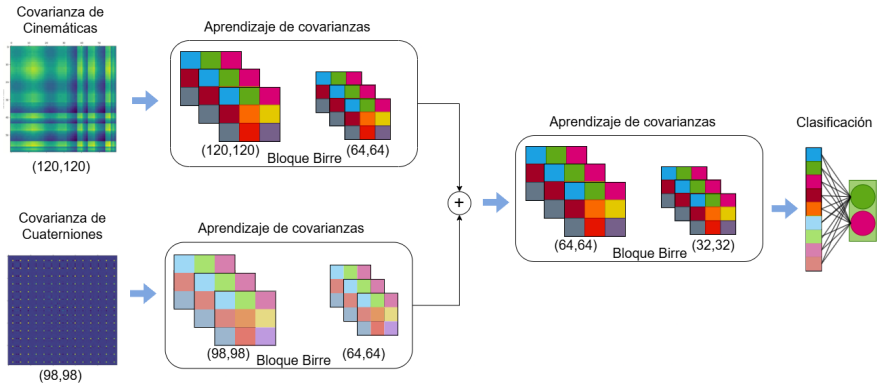


Fusión intermedia: en este enfoque, las características de cada modalidad son primero

transformadas por separado en el espacio riemanniano. Posteriormente, se realiza su integración en una única representación conjunta antes de la etapa de clasificación, permitiendo conservar propiedades geométricas específicas de cada modalidad.

En la figura 12 se ilustra el esquema de esta fusión. En este caso, en cada modalidad se aprenden descriptores geométricos en ramas independientes, explotando la información en cada rama. Entonces, en el nivel de fusión el vector multimodal se obtiene como $D_M = \{\hat{C}_q \bigoplus \hat{C}_k\}$, donde las matrices resultantes son sumadas para luego aprender un componenete BIRE con el fin de aprender nuevas representaciones mas compactas, finalmente las covarianzas son proyectadas al espacio euclidiano para realizar su respectiva clasificación. Luego de la proyección euclidiana, se toma la parte superior de la matriz, evitando redundancia y permitiendo la clasificación $(\overset{\triangle}{D_M})$

Figura 12. Arquitectura con los mejores resultados en la fusión intermedia: Sumando covarianzas



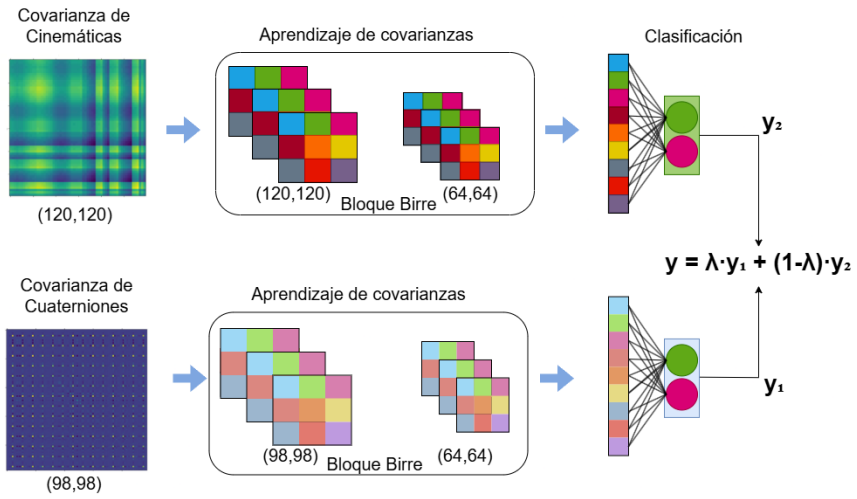
Fusión tardía: preserva la independencia total de ambas fuentes de información hasta el momento de la clasificación. En esta última fase se realiza una combinación de las predicciones mediante el uso de un parámetro que pondera la importancia de cada una de las modalidades en la predicción final del modelo.

En la figura 13 se ilustra la metodología empleada en la fusión tardía.

Una vez son aprendidos los descriptores compactos en el espacio de Riemman, por cada

modalidad (marcha y expresión facial), estos se mapean a capas Euclidianas respectivamente. En este caso de las dos covarianzas mapeadas, por cada modalidad, es obtenido la triangular superior para realizar la clasificación $\{\overset{\triangle}{C}_q, \overset{\triangle}{C}_k\}$. Desde estas clasificaciones por ramas independientes se obtiene una clasificación Parkinson, por cada modalidad y_1, y_2. Entonces, en la fusión tardía se ponderan de forma lineal las dos predicciones, variando los valores de importancia de cada modalidad, según un parámetro λ. Así, la integración tardía es definida como $y = \lambda y_1 + (1 - \lambda) y_2$.

Figura 13. Arquitectura de la fusión tardía variando el parámetro λ



5. DISEÑO EXPERIMENTAL

5.1. DATOS

El conjunto de datos empleado en el presente estudio está conformado por grabaciones de la marcha y de la expresión facial de un total de 29 participantes, distribuidos en 11 pacientes con diagnóstico de enfermedad de Parkinson (edad promedio: 65 ± 7.7 años) y 18 sujetos control (edad promedio: 67 ± 9.7 años). Es importante destacar que, dentro del grupo de pacientes con enfermedad de Parkinson, uno de ellos no se encontraba bajo tratamiento con Levodopa en el momento de la adquisición de los datos, mientras que los demás pacientes sí recibían dicho tratamiento. Para cada participante, se registraron 10 videos correspondientes a la marcha y a la expresión facial, capturados desde un plano sagital y frontal, respectivamente. Durante el proceso de adquisición de datos, a los participantes se les proporcionaron las siguientes instrucciones de movimiento:

- **Marcha:** En este ejercicio, se instruyó a los participantes a caminar en línea recta mientras la cámara capturaba su desplazamiento desde una vista sagital. La marcha fue registrada a lo largo de un trayecto de 5 metros, con una duración promedio de 5 segundos por grabación. Cada video presenta una resolución espacial de 1920×1080 píxeles y una resolución temporal de 60 fotogramas por segundo, garantizando una captura detallada de los patrones de movimiento.

  La Figura 14 ilustra el protocolo de grabación durante el ejercicio de marcha, en el cual se utilizó un fondo uniforme de color verde con el fin de optimizar el análisis de los datos. Esta configuración experimental tiene como propósito facilitar la evaluación de alteraciones motoras asociadas a la enfermedad de Parkinson, en particular la bradicinesia y la hipocinesia, caracterizadas por una reducción en la velocidad y amplitud de los movimientos.

- **Rostro:** Para la grabación de la expresión facial, cada sujeto debía pronunciar de manera sostenida las vocales A, E, I, O, U durante 3 segundos, manteniendo una postura frontal hacia la cámara y repitiendo el protocolo 10 veces. Se utilizó una cámara montada sobre un trípode a 45 grados y un entorno semicontrolado para minimizar artefactos de iluminación y maximizar la calidad del material grabado. Cada video tiene una resolución espacial de 1920 × 1080 píxeles y una resolución temporal de 60 fotogramas por segundo. La Figura 14 muestra el protocolo de grabación durante el ejercicio de la comunicación, esto con el fin de analizar la bradicinesia facial y la poca expresividad al articular palabras.

Se adquirieron un total de 20 videos para cada participante, abarcando tanto los ejercicios de marcha como los de expresión facial. En total, el conjunto de datos está compuesto por 290 videos por modalidad, lo que permite una caracterización del comportamiento motor y expresivo de los sujetos. Como resultado, el conjunto de datos completo para este estudio comprende un total de 580 videos. En ambas modalidades, utilizamos la misma cámara, una Nikon D3500 convencional, que ofrecía una resolución espacial de 1920 × 1080 píxeles. Este estudio fue aprobado por el Comité de Ética de la Universidad Industrial de Santander (UIS) y contó con el consentimiento informado por escrito de todos los participantes. La recopilación de los datos fue posible gracias a la colaboración de la Fundación del Adulto Mayor y Parkinson en Santander (FAMPAS) y del grupo de investigación Biomedical Imaging, Vision and Learning Laboratory ($BIVL^2ab$).

5.2. CONFIGURACIÓN EXPERIMENTAL DE LA ARQUITECTURA PROPUESTA

Para evaluar el desempeño del método propuesto, se empleó una validación cruzada 5-fold sobre el conjunto de datos. En cada iteración, el modelo fue entrenado utilizando el 80% de los datos y evaluado con el 20% restante, garantizando que cada subconjunto de validación fuera utilizado exactamente una vez a lo largo del proceso. En estos experimentos, los pacientes con diagnóstico de enfermedad de Parkinson correctamente iden-

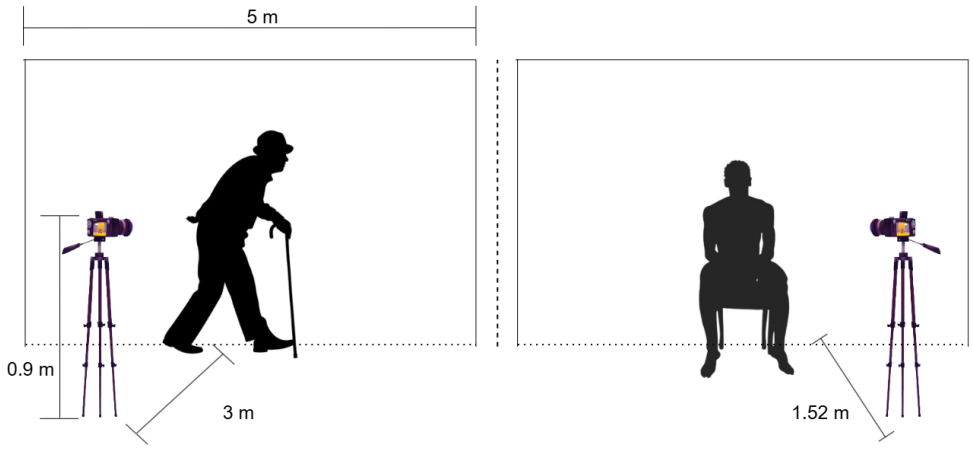


Figura 14. Protocolo de grabación usado en este estudio.

tificados fueron considerados verdaderos positivos (TP), mientras que los sujetos control correctamente clasificados fueron contabilizados como verdaderos negativos (TN). A fin de realizar una evaluación integral del rendimiento del modelo en sus distintas configuraciones, se calcularon un conjunto de métricas de desempeño. Las métricas incluidas en este trabajo son: la exactitud ($\frac{TP+TN}{TP+FP+FN+TN}$), sensibilidad ($\frac{TP}{TP+FN}$), precisión ($\frac{TP}{TP+FP}$), puntaje F1 ($\frac{2\times prec\times sen}{prec+sen}$) y área bajo la curva ROC (AUC): $\int_0^1 TPR(FPR)\,d(FPR)$. En particular, la precisión y la sensibilidad fueron fundamentales para evaluar la correcta identificación de patrones asociados a la enfermedad de Parkinson a partir de las representaciones obtenidas mediante matrices de covarianza. La combinación de estas métricas proporcionó una evaluación integral del rendimiento del modelo, asegurando un análisis detallado de su capacidad de generalización.

En cuanto al ajuste de los parámetros del modelo, el modelo usó matrices de covarianza de entrada $C_{dimk} = (120, 120)$ y $C_{dimq} = (98, 98)$ con el fin de obtener representaciones de salida s con tamaños más compactos $Cs_{dimk} = (64, 64)$ y $Cs_{dimq} = (64, 64)$. Esta red fue entrenada con el optimizador *Adam* y se usó la entropía cruzada binaria como función de pérdida. Los experimentos se realizaron con los siguientes parámetros: 400 épocas en cada experimento, tasa de aprendizaje de 0.0003, tamaño del lote de 64, valor del

épsilon en 0.0005, el cual se aplica en la capa de regularización, momentum en 0.9 y se usó la estrategia de detención temprana para evitar el sobreajuste. Además, se hicieron las siguientes pruebas:

- **Cantidad de puntos de referencia (landmarks)**. Para el rostro se usaron 60 y para la marcha se consideraron 33 puntos de interés. Ver figura 9
- **Representaciones**. Para la marcha se usaron cuaterniones con y sin posición relativa. Para el rostro se usó una representación basada en cinemáticas como la posición, velocidad y aceleración.
- **Número de bloques birre**. En el entrenamiento de cada uno de los enfoques unimodales y multimodales se probaron 1,2 y 3 bloques Birre, esto para aprender representaciones geométricas más eficientes, con el fin de compactar cada una de las representaciones a dimensiones más pequeñas como (64x64),(48x48),(36x36),(24x24) y (12x12).
- **Capas Finales Euclidianas**

  Todos los experimentos se realizaron con una única capa densa con 2 neuronas, cada una para cada clase. En la salida, las predicciones del modelo son procesadas por la función *softmax* con el fin de obatener una probabilidad por clase.

6. EVALUACIÓN Y RESULTADOS

El método desarrollado permite el análisis de representaciones multimodales derivadas de la marcha y la expresión facial para la identificación de patrones característicos de la enfermedad de Parkinson. Para ello, se diseñó una arquitectura basada en geometría riemanniana, la cual modela la dinámica motora a través de matrices de covarianza, proporcionando una representación estructurada y compacta de la información multimodal. En las siguientes secciones abordaremos los resultados unimodales (marcha y expresión facial) y multimodales, haciendo un énfasis en los diferentes tipos de fusión que se presentaron en secciones anteriores.

6.1. CLASIFICACIÓN DEL PARKINSON DESDE LA MARCHA

El análisis de la marcha se enfocó en evaluar la capacidad discriminativa de las representaciones generadas a partir de cuaterniones, las cuales fueron modeladas mediante matrices de covarianza. Dichas matrices permiten capturar la variabilidad estadística de los patrones dinámicos asociados a la marcha, proporcionando una representación más compacta, estructurada y robusta para las tareas de clasificación. Esta estrategia busca maximizar la eficiencia en la codificación de la información motora relevante, facilitando la identificación de alteraciones propias de la enfermedad de Parkinson.

Para la versión unimodal de la clasificación de Parkinson desde la marcha, las matrices de los cuaterniones, se utilizó una red geométrica básica, con un bloque Birre, y una proyección euclidiana que permite hacer una clasificación con una capa densa. En la Tabla 1 se presentan las métricas de desempeño obtenidas para la modalidad de marcha (Cuaterniones con y sin la posición relativa).

Los resultados de la tabla 1 muestran que la representación basada en cuaterniones con información de posición logra una mayor capacidad discriminativa para diferenciar entre

Cuaterniones	Exactitud (%)	Precisión (%)	Sensibilidad (%)	F1 (%)	AUC (%)
Con Posición relativa	86 ± 2.7	89 ± 6.7	76 ± 7.0	87 ± 11.1	85 ± 3.9
Sin Posición relativa	86 ± 2.9	83 ± 6.6	79 ± 4.4	81 ± 3.8	84 ± 2.8

Tabla 1. Resultados de clasificación entre Parkinson y control a partir de representaciones de cuaterniones.

sujetos con Parkinson y controles. Esto sugiere que la inclusión de la posición en la codificación de los patrones motores permite capturar mejor las alteraciones cinemáticas propias de la enfermedad. En particular, la mayor precisión y el aumento del AUC al 85% indican que esta representación encapsula de manera más efectiva los déficits en la variabilidad del paso, la asimetría en la marcha y la reducción en la amplitud de los movimientos, características distintivas en la progresión del Parkinson.

Por otro lado, la representación sin información de posición presenta una disminución en la precisión, puntaje F1 y AUC, lo que sugiere que las alteraciones motoras en la enfermedad no solo se reflejan en los patrones de orientación angular del cuerpo, sino también en la trayectoria espacial de los desplazamientos. Esto refuerza la importancia de modelar conjuntamente la dinámica posicional y rotacional para una caracterización más completa de la marcha en pacientes con Parkinson. También se realizó un experimento sin el mapeo de cuaterniones, usando las posiciones relativas en cada video, logrando un F1 de 81 ± 7.2 y un AUC de 85 ± 5.5, lo cual resulta ser competitivo, con respecto a los cuaterniones. Sin embargo, estas representaciones que dependen de las posiciones en los videos pueden ser sensibles a cambios en el registro de los videos, así como también pueden perder generalización del enfoque.

6.2. CLASIFICACIÓN DEL PARKINSON DESDE LA EXPRESIÓN FACIAL

En cuanto al análisis de expresión facial, en este trabajo se enfocó en realizar una caracterización cinemática de los puntos de interés, estrechamente relacionados con el ejercicio del habla. Este análisis se enfocó en evaluar la capacidad de las representaciones

espacio-temporales de los movimientos faciales y gestuales para discriminar patrones motores asociados a la enfermedad de Parkinson. La afectación de la motricidad facial es una manifestación característica de la enfermedad y puede reflejarse en una reducción de la expresividad (hipomimia), así como en alteraciones en la velocidad y coordinación de los movimientos musculares

En este estudio, se analizaron diferentes representaciones cinemáticas del movimiento facial: posición, velocidad y aceleración, con el fin de identificar cuál de estos enfoques permite una mejor caracterización del deterioro motor. La Tabla 2 muestra los resultados obtenidos en cada caso. Los resultados indican que la posición del rostro y sus componentes lograron capturar mejor las diferencias entre sujetos con Parkinson y controles, obteniendo un AUC del 86%. Esto sugiere que las alteraciones en la motricidad facial se expresan más en cambios estáticos y estructurales, como la falta de variabilidad en la postura de los músculos faciales y la reducción en la amplitud del movimiento.

Por otro lado, la representación basada en aceleración presentó el peor desempeño, con valores significativamente bajos de precisión y sensibilidad. Esto podría deberse a la mayor variabilidad en los datos, ya que los movimientos faciales en pacientes con Parkinson tienden a ser irregulares y de menor amplitud, dificultando la extracción de patrones consistentes en los cambios acelero métricos. En el caso de la velocidad, los resultados fueron intermedios, lo que sugiere que la ralentización de los movimientos faciales es un marcador relevante, pero insuficiente por sí solo para discriminar entre clases. La combinación de estas representaciones cinemáticas podría mejorar la caracterización de la enfermedad, ya que cada una aporta información complementaria sobre la progresión del deterioro motor en la expresión facial.

Otro de los experimentos realizados fue el mapeo de los puntos de interés a cuaterniones sin posición relativa, tomando como punto de partida de los vectores $\vec{v_1}$ y $\vec{v_2}$ el punto de la nariz(4), este experimento logró un puntaje F1 de 50 ± 16.6 y un AUC de 62 ± 9.8, lo que resulta ser una representación más competitiva que la velocidad y la aceleración pero no

más que la posición de los puntos de interés. Sin embargo este tipo de representaciones en el rostro que intentan modelar movimientos 3D de muscúlos faciales no son intuitivos en el análisis facial, además de ser difícil de justificar este tipo de representaciones por la falta de movimiento en la tercera dimensión dada la naturaleza del movimiento del rostro. Por este motivo se decidió seleccionar la representación basada explícitamente en cinemáticas, sin incluir el análisis con los cuaterniones.

Representación	**Exactitud (%)**	**Precisión (%)**	**Sensibilidad (%)**	**F1 (%)**	**AUC (%)**
Posición	69 ± 7.1	64 ± 13.7	57 ± 20.3	57 ± 13.0	86 ± 8.1
Velocidad	56 ± 5.5	41 ± 5.4	58 ± 33.6	44 ± 19.7	56 ± 6.5
Aceleración	58 ± 6.2	35 ± 6.8	31 ± 21.0	25 ± 15.4	34 ± 7.1

Tabla 2. Resultados de clasificación utilizando diferentes representaciones cinemáticas (posición, velocidad y aceleración).

6.3. FUSIÓN DE LAS MODALIDADES DE MARCHA Y EXPRESIÓN FACIAL

En este trabajo se exploró la fusión de patrones de marcha y de expresión facial, bajo la hipótesis que su complemento puede cubrir mejor los diferentes fenotipos de la enfermedad, logrando en consecuencia mayor capacidad de clasificación de la enfermedad. La integración multimodal permite combinar la información extraída de la marcha y la expresión facial, logrando mejorar la discriminación de la enfermedad de Parkinson a partir de patrones representacionales. En este estudio, se implementaron diferentes estrategias de fusión, las cuales permiten procesar de manera simultánea ambas modalidades y capturar relaciones complementarias entre ellas. En la tabla 3 se reportan los resultados obtenidos en cada una de las estrategias de fusión, seleccionando el mejor ajuste en cada tipo.

Los resultados obtenidos sugieren que la fusión tardía con $\lambda = 0.3$ proporciona el mejor desempeño en términos de exactitud (92%) y *AUC* (90%), lo que indica que al combinar ambas modalidades al final del procesamiento mejora la capacidad del modelo para in-

Fusión	Exactitud (%)	Precisión (%)	Sensibilidad (%)	F1 (%)	AUC (%)
Temprana	89 ± 6.1	88 ± 12.7	83 ± 20.1	86 ± 13.0	88 ± 8.1
Intermedia	86 ± 2.7	89 ± 6.7	76 ± 7.0	87 ± 11.1	85 ± 3.9
Tardía ($\lambda = 0.3$)	92 ± 2.8	94 ± 6.5	84 ± 5.4	89 ± 3.8	90 ± 3.0

Tabla 3. Resultados de clasificación utilizando diferentes estrategias de fusión de características.

tegrar patrones complementarios. Esta estrategia logra capturar de una mejor manera la relación entre bradicinesia facial, postura y marcha, mejorando el AUC 90% del modelo. Alternativamente, la fusión temprana presenta métricas ligeramente menores comparados a la fusión tardía (Precisión 88%) y AUC (88%). En este sentido, aprender nuevas representaciones geométricas con una matriz de entrada de dos canales (uno para cada modalidad) y una matriz de salida más compacta con un solo canal donde se desempeña la fusión temprana permite capturar interacciones entre ambas modalidades. Con respecto a la fusión intermedia, se reportó un valor de AUC ligeramente menor comparado a la fusión temprana, en este sentido la suma de las covarianzas de ambas modalidades para después aprender una nueva representación que es clasificada mediante una capa densa y la función softmax. Sin embargo, en términos motores, esta estrategia es la que menos permite que el modelo aprenda patrones conjuntos entre rigidez facial y alteraciones en la locomoción, características claves en la enfermedad de Parkinson.
Un análisis detallado de cada fusión evaluada en este trabajo se reporta a continuación.

Fusion temprana: En esta fusión se aprendió la representación de la expresión facial mediante las capas Bimap y ReEig reduciendo el tamaño de la covarianza de (120x120) a (98x98) la cual es la dimensión original de la marcha. Para el primer experimento se concatenaron las matrices de ambas modalidades en una única matriz de dos canales con el fin de aprender una nueva representación multimodal. En una segunda adaptación, se aprendió la representación de rostro de tamaño (98x98), la cual a su vez se suma con la covarianza de marcha con el fin de obtener una covarianza multimodal para después ser compactada por 1 bloque birre en donde las características finales obtenidas serán

clasificadas. La tabla 4 resume los resultados obtenidos para estas dos configuraciones. Como se puede observar hay un rendimiento competitivo en la fusión temprana con una sensibilidad del 83% y un AUC del 88%, en la configuración donde se concatenan los canales, lo que permitió al modelo aprender una representación multimodal que integra tanto las características de la marcha como las de la expresión facial. Este enfoque permite que el modelo capture las interacciones entre ambas modalidades desde el inicio del proceso de clasificación. Sin embargo, la sensibilidad y el AUC obtenidos son ligeramente inferiores a los de la fusión tardía, lo que sugiere que, aunque la fusión temprana permite la integración de la información al inicio del aprendizaje, podría no ser suficiente para capturar de forma óptima las relaciones más complejas entre los patrones de marcha y la rigidez facial, características clave en la enfermedad de Parkinson.

Fusion Intermedia: El enfoque de fusión intermedia busca generar representaciones de igual tamaño (64x64) para ambas modalidades (marcha y expresión facial) de manera independiente con el fin de concatenar o sumar las covarianzas de cada modalidad para aprender una representación multimodal más compacta y clasificarla. Como se observa en la tabla 4, la configuración con la suma de las covarianzas muestra ser ligeramente inferiores al de la fusión temprana, con una sensibilidad de 76% y un AUC de 85%. Esta configuración, con respecto a la fusión temprana, no parece ser tan efectiva para capturar la interacción entre las características motrices y faciales. En particular, la fusión intermedia no permite que el modelo aprenda adecuadamente los patrones conjuntos de rigidez facial y alteraciones en la locomoción, que son cruciales para la detección de la EP. Esto sugiere que la simple combinación de las covarianzas de cada modalidad no es suficiente para capturar relaciones complejas entre las diferentes características de la enfermedad, lo que puede explicar la menor capacidad discriminativa de este enfoque.

Fusión Tardía: La fusión tardía se caracteriza por la combinación de las modalidades de marcha y expresión facial al final del proceso de clasificación, lo que permite que el modelo capture relaciones más complejas, de forma independiente y especializada en ambas

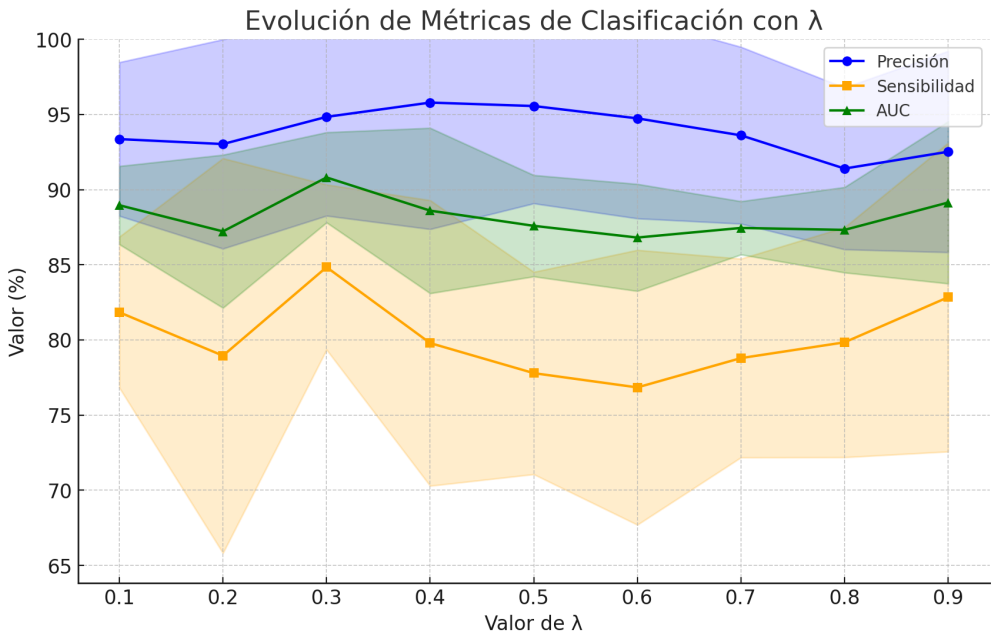
Fusión	Configuración	Exactitud (%)	Precisión (%)	Sensibilidad (%)	F1 (%)	AUC (%)
Temprana	Concatenando canales	89 $\pm$ 6.1	88 $\pm$ 12.7	83 $\pm$ 20.1	86 $\pm$ 13	88 $\pm$ 8.1
	Sumando covarianzas	88 $\pm$ 3.2	88 $\pm$ 5.9	83 $\pm$ 11.2	85 $\pm$ 5	87 $\pm$ 4.3
Intermedia	Concatenando características	85 $\pm$ 2.7	86 $\pm$ 7.2	75 $\pm$ 7.6	80 $\pm$ 4.2	83 $\pm$ 3.2
	Sumando covarianzas	86 $\pm$ 2.7	89 $\pm$ 6.7	76 $\pm$ 7.0	87 $\pm$ 11.1	85 $\pm$ 3.9

Tabla 4. Resultados de clasificación en la fusión temprana e intermedia en sus dos diferentes configuraciones respectivamente.

modalidades. En esta configuración, las modalidades se procesan de manera independiente hasta llegar a la etapa final de la clasificación, donde se integran sus respectivas representaciones. La tabla 3 muestra que el enfoque tardío tiene el mejor desempeño con una sensibilidad del 84% y un AUC del 90%, destacando como la estrategia más eficaz en términos de exactitud y discriminación entre pacientes con Parkinson y controles. Un aspecto clave de la fusión tardía es el peso relativo asignado a cada modalidad. Al observar la figura 15 podemos analizar que el mejor resultado de clasificación se obtiene cuando la marcha aporta un 30%, mientras que la expresión facial tiene una ponderación mayor del 70%. Esta mayor contribución de la expresión facial en el resultado resalta la relevancia de las alteraciones del rostro, como la bradicinesia y la falta de expresión, en la clasificación del Parkinson, lo que indica que estas características tienen un papel crucial en la identificación de los pacientes. El elevado AUC y la sensibilidad obtenidos con esta estrategia refuerzan la efectividad de esta fusión sobre las demás, pues al integrar las modalidades al final del proceso de clasificación, el modelo logra captar de manera más eficiente los patrones representacionales que podrían no ser tan fácilmente identificables en etapas previas del procesamiento. Este enfoque, por lo tanto, permite una clasificación más precisa y robusta, destacando la capacidad para mejorar el rendimiento global del modelo.

En una comparación del trabajo propuesto se comparó con un trabajo reportado en el

Figura 15. Evolución de las métricas al variar el parámetro λ en la fusión tardía.



estado del arte, que incluye (Archila *et al.*[49]) que hace la evaluación sobre un modelo multimodal, pero integrando la marcha con movimientos oculares. Este trabajo usa redes pre-entrenadas, cuyas activaciones son codificadas en matrices de covarianza. El trabajo de Archila *et al.*[49] obtiene una precisión del 94%, sensibilidad del 92%, y un F1-score de 93% en la clasificación de pacientes. Estos resultados los logró haciendo un análisis en un estudio con 19 pacientes con EP y 13 sujetos control. En nuestro caso, se utilizaron las expresiones del rostro y la marcha, alcanzando en la mejor configuración una precisión del 92%, sensibilidad del 84%, y un F1-score de 89% en la clasificación de pacientes. En nuestro trabajo se realizó el estudio con 11 pacientes diagnosticados con la EP y 18 sujetos control. Si bien los resultados en ambos casos son competitivos, el trabajo propuesto muestra que los descriptores gestuales son competitivos y hacen parte integral

[49] John ARCHILA; Antoine MANZANERA, and Fabio MARTÍNEZ. "A Riemannian multimodal representation to classify parkinsonism-related patterns from noninvasive observations of gait and eye movements". In: *Biomedical Engineering Letters* 15.1 (2025), pp. 81–93.

del análisis del Parkinson. Por lo tanto, los trabajos futuros deben incluir otros sintomas para fortalecer el análisis de la enfermedad.

7. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se desarrolló una estrategia de aprendizaje Riemanniano multimodal para integrar la información de la marcha y la expresión facial con el fin de clasificar patrones parkinsonianos. Mediante la herramienta Mediapipe se extraen puntos de interés espaciotemporales, a partir de un conjunto de datos de 580 videos de 29 pacientes (11 pacientes diagnosticados con Parkinson y 19 sujetos control), con el fin de construir representaciones que describen el movimiento de la marcha y el rostro. En esta investigación se introdujeron descriptores compactos basados en cuaterniones para la representación de las rotaciones tridimensionales de las articulaciones durante la marcha, lo que permitió una clasificación eficaz entre sujetos con Enfermedad de Parkinson (EP) y sujetos controles. Los resultados obtenidos evidencian la efectividad de este enfoque, alcanzando una precisión del 89 % en la tarea de clasificación, lo que resalta su potencial para la identificación de patrones motores característicos de la EP. Adicionalmente, se desarrolló un método de análisis cinemático basado en 60 puntos de referencia faciales, centrado en los músculos clave involucrados en la comunicación. Esta representación permitió discriminar con precisión entre pacientes con Parkinson y sujetos controles, obteniendo un AUC del 86%. La modelización del comportamiento dinámico facial a partir de estos puntos de referencia se posiciona como una herramienta de gran utilidad para evaluar alteraciones en la comunicación oral asociadas a la patología.

En el contexto de enfoques multimodales, se exploraron y propusieron estrategias que integran los descriptores basados en cuaterniones con las cinemáticas a partir de los puntos de interés faciales, lo que resultó en una mejora significativa de todas las métricas más específicamente en la precisión con un 94%, sensibilidad de 84% y 90% de AUC en la discriminación entre clases en comparación con los enfoques unimodales, demostrando así que el análisis multimodal es una estrategia eficaz en el análisis de la enfermedad. La combinación de estas modalidades representa una vía prometedora para optimizar los

modelos de clasificación en el diagnóstico de la EP, al capturar tanto los patrones motores de la marcha como las alteraciones en la expresión facial. Los resultados obtenidos en este estudio destacan la relevancia de la fusión tardía con ponderación diferencial como estrategia óptima para la clasificación de sujetos con Enfermedad de Parkinson (EP) y controles. En particular, la combinación de características con un peso del 30% para la marcha y un 70% para la expresión facial alcanzó el mejor desempeño en cada una de las métricas, al comparar los mejores resultados unimodales de marcha se obtiene una ganancia de 6% en la exactitud, 5% en la precisión, 5% en la sensibilidad, 2% en el puntaje F1 y 5% en el AUC. Este hallazgo resalta la importancia relativa de las alteraciones en la expresión facial en la caracterización de la EP, alineándose con estudios previos que han señalado la hipomimia como un biomarcador clave en la progresión de la enfermedad.

Desde una perspectiva médica, estos resultados subrayan la necesidad de evaluar no solo las alteraciones en la marcha, que han sido tradicionalmente el foco del diagnóstico motor de la EP, sino también las disfunciones faciales como un componente esencial en la caracterización de la enfermedad. La mayor contribución de la expresión facial en el modelo sugiere que las alteraciones en la comunicación no verbal pueden proporcionar información más temprana y sensible sobre la progresión de la enfermedad que los déficits en la marcha. Este enfoque multimodal permite una evaluación más completa de los síntomas motores de la EP y abre nuevas oportunidades para el desarrollo de herramientas de diagnóstico automatizadas y estrategias de intervención temprana.

A pesar de los resultados prometedores obtenidos en este estudio, existen varias limitaciones que deben considerarse. Trabajos futuros deberán validar mecanismos de atención y codificación de la información en representaciones que incorporen un análisis espacial y de procesamiento de video, considerando los valores relevantes de la información. También, el tamaño reducido del conjunto de datos podría afectar la generalización del modelo a poblaciones más amplias, por lo que futuros trabajos deberán centrarse en la

recolección de un mayor número de muestras para fortalecer la robustez y validez de los hallazgos. Además, la mayoría de los pacientes con Enfermedad de Parkinson incluidos en este estudio estaban bajo el efecto de la medicación, lo que podría haber modificado la expresión de los síntomas motores y faciales. Estudios futuros deberán considerar la evaluación en diferentes estados de tratamiento, incluyendo medicación ON y OFF, para analizar el impacto de estos factores en la discriminación de los sujetos. Asimismo, la ausencia de información auditiva en la modalidad de expresión facial impide la evaluación de alteraciones en la prosodia, las cuales constituyen un componente relevante en la afectación comunicativa de los pacientes con Parkinson. Por ello, futuras investigaciones podrían integrar análisis acústicos del habla y exploraciones multimodales más completas que incluyan señales de audio y variables clínicas adicionales, con el fin de desarrollar modelos de clasificación más precisos y clínicamente interpretables.

BIBLIOGRAFÍA

ABRAMI, Avner et al. “Automated computer vision assessment of hypomimia in Parkinson disease: proof-of-principle pilot study”. In: *Journal of medical Internet research* 23.2 (2021), e21037 (cit. on pp. 28, 34).

ARCHILA, J.; MANZANERA, A., and MARTINEZ, F. “A multimodal Parkinson quantification by fusing eye and gait motion patterns, using covariance descriptors, from noninvasive computer vision”. In: *Computer Methods and Programs in Biomedicine* 215 (2022), p. 106607. DOI: 10.1016/j.cmpb.2021.106607 (cit. on p. 29).

ARCHILA, John; MANZANERA, Antoine, and MARTÍNEZ, Fabio. “A Riemannian multimodal representation to classify parkinsonism-related patterns from noninvasive observations of gait and eye movements”. In: *Biomedical Engineering Letters* 15.1 (2025), pp. 81–93 (cit. on p. 57).

BAZAREVSKY, Valentin et al. “Blazeface: Sub-millisecond neural face detection on mobile gpus”. In: *arXiv preprint arXiv:1907.05047* (2019) (cit. on p. 34).

BERGANZO, K et al. “Síntomas no motores y motores en la enfermedad de Parkinson y su relación con la calidad de vida y los distintos subgrupos clínicos”. In: *Neurología* 31.9 (2016), pp. 585–591 (cit. on p. 13).

BERGERON, D et al. “Sir William Osler and the evolving neurological sciences”. In: *Lancet* 11 (2012), pp. 999–1004 (cit. on p. 12).

BIASE, Lazzaro di et al. “Parkinson’s disease wearable gait analysis: kinematic and dynamic markers for diagnosis”. In: *Sensors* 22.22 (2022), p. 8773 (cit. on p. 26).

CHERIET, Mohamed et al. “Multi-speed transformer network for neurodegenerative disease assessment and activity recognition”. In: *Computer Methods and Programs in Biomedicine* 230 (2023), p. 107344 (cit. on p. 27).

DEVANNE, Maxime et al. “Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation”. In: *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE. 2017, pp. 529–534 (cit. on p. 35).

FEIGIN, Valery L et al. “Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015”. In: *The Lancet Neurology* 16.11 (2017), pp. 877–897 (cit. on p. 30).

FRAGKIADAKI, Katerina et al. “Recurrent network models for human dynamics”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4346–4354 (cit. on p. 35).

GARRIDO-ELUSTONDO, Sofía et al. “Capacidad de detección de patología psiquiátrica por el médico de familia”. In: *Atención Primaria* 48.7 (2016), pp. 449–457 (cit. on p. 14).

GOMEZ, Luis F et al. “Improving parkinson detection using dynamic features from evoked expressions in video”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 1562–1570 (cit. on p. 28).

GONZÁLEZ-MONJE, Mariana Hernández and MALPICA, Norberto. “Sistemas basados en vídeo”. In: *MANUAL SEN DE* (2021) (cit. on p. 18).

GRISHCHENKO, Ivan et al. “Attention mesh: High-fidelity face mesh prediction in real-time”. In: *arXiv preprint arXiv:2006.10962* (2020) (cit. on p. 34).

GUAYACÁN, Luis C and MARTÍNEZ, Fabio. “Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network”. In: *Journal of biomedical informatics* 123 (2021), p. 103935 (cit. on p. 27).

HORNA LÓPEZ, Fabián Vicente and TARÍS RAMOS, Luis Mifguel. “Diseño e implementaicón de un sistema alternativo de captura de movimiento para efectos visuales. Caso práctico: Sacrilegio del 4 de mayo de 1897”. B.S. thesis. 2013 (cit. on p. 18).

HUANG, Zhiwu and VAN GOOL, Luc. “A riemannian network for spd matrix learning”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017 (cit. on p. 41).

JIN, Bo et al. “Diagnosing Parkinson disease through facial expression recognition: video analysis”. In: *Journal of medical Internet research* 22.7 (2020), e18697 (cit. on p. 28).

JOSÉ CÁT Rodrigo BJ, Carrillo-Ruiz José D. “Interpretación neuroanatómica de los principales síntomas motores y no-motores de la enfermedad de Parkinson”. In: *Rev Mex Neurociencia* 1 (2010), pp. 1–10 (cit. on p. 12).

KAUR, Rachneet et al. “A Vision-Based Framework for Predicting Multiple Sclerosis and Parkinson’s Disease Gait Dysfunctions—A Deep Learning Approach”. In: *IEEE Journal of Biomedical and Health Informatics* 27.1 (2022), pp. 190–201 (cit. on p. 27).

KUHLGATZ, Timo et al. “On Stair Walk Recognition Using a Single Magnetometer-free IMU and Deep Learning”. In: *Current Directions in Biomedical Engineering*. Vol. 10. 4. De Gruyter. 2024, pp. 404–407 (cit. on p. 35).

LEÓN-JIMÉNEZ, Carolina. “Síndrome rígido acinético”. In: *Revista de Medicina Clínica* 3.2 (2019), pp. 104–108 (cit. on pp. 14, 15).

LIQIONG, YANG et al. “Changes in facial expressions in patients with Parkinson’s disease during the phonation test and their correlation with disease severity”. In: *Computer Speech & Language* 72 (2022), p. 101286 (cit. on p. 27).

LIU, Peipei et al. “Quantitative assessment of gait characteristics in patients with Parkinson’s disease using 2D video”. In: *Parkinsonism & Related Disorders* 101 (2022), pp. 49–56 (cit. on p. 27).

LOZANO CARRILLO, Juan Camilo. “Pose temporal estimation in markerless normal human gait integrating kinematic patterns and segmented video”. In: *Departamento de Imágenes Diagnósticas* () (cit. on p. 18).

MAYCAS-CEPEDA, Teresa et al. “Hypomimia in Parkinson’s disease: what is it telling us?” In: *Frontiers in Neurology* 11 (2021), p. 603582 (cit. on pp. 13, 16).

MICHELL, A. W. et al. “Biomarkers and Parkinson’s disease”. In: *Brain* 127.8 (June 2004), pp. 1693–1705. DOI: 10.1093/brain/awh198. eprint: https://academic.oup.com/brain/article-pdf/127/8/1693/842701/awh198.pdf (cit. on p. 11).

NOVOTNỲ, Michal et al. “Automatic evaluation of articulatory disorders in Parkinson’s disease”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.9 (2014), pp. 1366–1378 (cit. on p. 16).

PHAM, Hung N et al. “Multimodal detection of Parkinson disease based on vocal and improved spiral test”. In: *2019 International Conference on System Science and Engineering (ICSSE)*. IEEE. 2019, pp. 279–284 (cit. on p. 28).

RAMÍREZ, Nancy Bertado et al. “Datos clave para el diagnóstico clínico de enfermedad de Parkinson”. In: *Revista Mexicana de Neurociencia* 10.5 (2009), pp. 340–343 (cit. on p. 15).

RASTEGARI, Elham; AZIZIAN, Sasan, and ALI, Hesham. “Machine learning and similarity network approaches to support automatic classification of parkinson’s diseases using accelerometer-based gait analysis”. In: (2019) (cit. on p. 26).

RICCIARDI, L; DE ANGELIS, A, et al. “Hypomimia in Parkinson’s disease: an axial sign responsive to levodopa”. In: *European Journal of Neurology* 27.12 (2020), pp. 2422–2429 (cit. on pp. 13, 16, 34).

ROVINI, E.; MAREMMANI, C., and CAVALLO, F. “How wearable sensors can support parkinson’s disease diagnosis and treatment: a systematic review”. In: *Frontiers in neuroscience* 11 (2017), p. 555 (cit. on p. 10).

RUSZ, Jan et al. “Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson’s disease”. In: *npj Parkinson’s Disease* 7.1 (2021), p. 98 (cit. on pp. 13, 16).

SALVATORE, Juan; OSIO, Jorge, and MORALES, Martín. “Detección de objetos utilizando el sensor Kinect”. In: *Guayaquil, Ecuador, LACCEI* (2014) (cit. on p. 18).

SILVA, Ana Beatriz Ramalho Leite et al. “Premotor, nonmotor and motor symptoms of Parkinson’s disease: a new clinical state of the art”. In: *Ageing Research Reviews* 84 (2023), p. 101834 (cit. on p. 26).

SIMONYAN, Karen and ZISSERMAN, Andrew. “Very deep convolutional networks for large-scale image recognition. arXiv 2014”. In: *arXiv preprint arXiv:1409.1556* 1409 (2014) (cit. on p. 28).

SIMUNI, Tanya and PAHWA, Rajesh. *Parkinson’s disease*. Oxford University Press, USA, 2009 (cit. on p. 10).

SKIBIŃSKA, Justyna and HOSEK, Jiri. “Computerized analysis of hypomimia and hypokinetic dysarthria for improved diagnosis of Parkinson’s disease”. In: *Heliyon* 9.11 (2023) (cit. on p. 29).

SOMMER, Stefan; FLETCHER, Tom, and PENNEC, Xavier. “Introduction to differential and Riemannian geometry”. In: *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, 2020, pp. 3–37 (cit. on p. 24).

TINELLI, Michela; KANAVOS, Panos, and GRIMACCIA, Federico. “The value of early diagnosis and treatment in Parkinson’s disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in Parkinson’s disease”. In: *Expert Review of Pharmacoeconomics & Outcomes Research* 16.1 (2016), pp. 41–51. DOI: 10.1586/14737167.2016.1121768 (cit. on pp. 10, 30).

TRABASSI, Dante et al. “Machine learning approach to support the detection of Parkinson’s disease in IMU-based gait analysis”. In: *Sensors* 22.10 (2022), p. 3700 (cit. on p. 26).

VÁSQUEZ-CORREA, Juan Camilo et al. “Comparison of user models based on GMM-UBM and i-vectors for speech, handwriting, and gait assessment of Parkinson’s disease patients”. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 6544–6548 (cit. on p. 28).

VÁSQUEZ-CORREA, Juan Camilo et al. “Multimodal assessment of Parkinson’s disease: a deep learning approach”. In: *IEEE journal of biomedical and health informatics* 23.4 (2018), pp. 1618–1630 (cit. on p. 29).

WANG, Qinghui; ZENG, Wei, and DAI, Xiangkun. “Gait classification for early detection and severity rating of Parkinson’s disease based on hybrid signal processing and machine learning methods”. In: *Cognitive Neurodynamics* (2022), pp. 1–24 (cit. on p. 26).