

**CARACTERIZACIÓN DE PATRONES DE LA ENFERMEDAD DE PARKINSON
USANDO MODELOS DE APRENDIZAJE DE MÁQUINA**

LUIS CARLOS GUAYACÁN CHAPARRO

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

2021

**CARACTERIZACIÓN DE PATRONES DE LA ENFERMEDAD DE PARKINSON
USANDO MODELOS DE APRENDIZAJE DE MÁQUINA**

LUIS CARLOS GUAYACÁN CHAPARRO

**Trabajo de investigación en cumplimiento de los requisitos para la obtención
del título de:
Magíster en Matemática Aplicada**

**Director:
Fabio Martínez Carrillo
Ph.D en Ingeniería de Sistemas y Computación**

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

2021

AGRADECIMIENTOS

El autor expresa su agradecimiento a la fundación FAMPAS (*Fundación del Adulto Mayor y Parkinson de Santander*) y a la institución local de la tercera edad *Asilo San Rafael* por facilitar el registro del dataset propuesto en este trabajo.

Por otra parte, este trabajo no hubiera sido posible sin el esfuerzo del profesor Fabio Martínez, quien en todo momento estuvo aportando valiosas ideas y ayuda en el desarrollo de este proyecto. Muchas gracias a él por su paciencia y las muchas horas que dedicó a aportar ideas, consejos y dirección.

CONTENIDO

	pág.
INTRODUCCIÓN	11
1. ANÁLISIS DE LA MARCHA PARKINSONIANA	16
2. PROBLEMA DE INVESTIGACIÓN	22
3. OBJETIVOS PROPUESTOS	24
4. MÉTODO 1: REPRESENTACIÓN CINEMÁTICA USANDO TRAYECTO- RIAS DE MOVIMIENTO	25
4.1. ANÁLISIS REGIONAL	25
4.2. TRAYECTORIAS DENSAS DE MOVIMIENTO	29
4.3. REPRESENTACIÓN CINEMÁTICA REGIONAL	31
4.4. CODIFICACIÓN Y PREDICCIÓN GLOBAL	33
5. METODO 2: REPRESENTACIÓN CONVOLUCIONAL USANDO MAPAS DE SALIENCIA	35
5.1. APRENDIZAJE DE LA SECUENCIA VOLUMÉTRICA DE LA MARCHA .	36
5.2. PATRONES PARKINSONIANOS A PARTIR DE MAPAS VOLUMÉTRICOS DE SALIENCIA	39
6. MONTAJE EXPERIMENTAL	43
6.1. CONJUNTO DE DATOS PROPUESTO	43
6.2. CONFIGURACIÓN DEL MÉTODO DE TRAYECTORIAS	45
6.2.1 Comparación entre <i>DensePose</i> y un algoritmo de clustering	46
6.3. CONFIGURACIÓN DE LA RED CONVOLUCIONAL 3D	47

7. EVALUACIÓN Y RESULTADOS	48
7.1. REPRESENTACIÓN CINEMÁTICA	48
7.2. REPRESENTACIÓN CONVOLUCIONAL	53
7.2.1 Mapas de Saliencia Relacionados con el Párkinson	56
7.2.2 Factores de Confusión	58
8. DISCUSIÓN	60
9. CONCLUSIONES Y TRABAJO FUTURO	64
BIBLIOGRAFÍA	66
ANEXOS	73

LISTA DE FIGURAS

	pág.
Figura 1. Esquema del método de representación cinemática.	26
Figura 2. Construcción del descriptor a nivel de trayectoria.	32
Figura 3. Esquema del método de representación convolucional.	36
Figura 4. Pseudodeconvolución durante la retropropagación	42
Figura 5. Muestra del conjunto de datos utilizado.	44
Figura 6. Agrupación de trayectorias usando DensePose.	50
Figura 7. Descriptor cinemático de movimiento para la cabeza y el torso. . . .	52
Figura 8. Descriptor cinemático de movimiento global.	53
Figura 9. Curva ROC para secuencias RGB y flujo óptico.	55
Figura 10 Matrices de confusión para datos RGB y flujo óptico	55
Figura 11 Mapas de saliencia para pacientes con la enfermedad de Parkinson y sujetos control.	56
Figura 12 Comparación entre el método propuesto y Gradcam	58

LISTA DE TABLAS

	pág.
Tabla 1. Rendimiento de los distintos métodos de clasificación para cada clúster mediante la descripción de la velocidad.	49
Tabla 2. Evaluación cinemática por grupo utilizando el método de clasificación de mejor rendimiento.	49
Tabla 3. Resumen del resultado de la clasificación cuantitativa para las secuencias RGB en bruto y de flujo óptico.	54

LISTA DE ANEXOS

	pág.
Anexo A. Productos Académicos	73

RESUMEN

TÍTULO: CARACTERIZACIÓN DE PATRONES DE LA ENFERMEDAD DE PARKINSON UTILIZANDO MODELOS DE APRENDIZAJE DE MÁQUINA. *

AUTOR: LUIS CARLOS GUAYACÁN CHAPARRO **

PALABRAS CLAVE: ENFERMEDAD DE PARKINSON, REPRESENTACIÓN DE PATRONES, ANÁLISIS DE MARCHA, MARCADORES CLÍNICOS.

DESCRIPCIÓN: La enfermedad de Parkinson (EP) carece de un diagnóstico definitivo, siendo la observación de los patrones de movimiento el principal análisis para caracterizar el progreso de la enfermedad y planificar los tratamientos de los pacientes. Entre las observaciones de la EP, los patrones de movimiento de la marcha, como la longitud de los pasos y el control postural, apoyan la caracterización de la enfermedad. Sin embargo, este análisis es usualmente cuantificado con protocolos basados en marcadores, reduciendo la descripción de la marcha y omitiendo patrones locales durante la locomoción. Este trabajo introduce dos aproximaciones computacionales para el soporte del análisis y diagnóstico del párkinson. Primero se propuso un método computacional sin marcadores basado en una caracterización cinemática completa de los segmentos corporales relacionados con las alteraciones motoras de la EP. En un segundo aporte de este trabajo se obtuvo una representación convolucional 3D de la marcha en un contexto de clasificación automática de la EP, destacando patrones espacio-temporales aprendidos para esta tarea. Estos enfoques permiten obtener modelos que clasifican secuencias sin marcadores y describen las principales regiones espacio-temporales asociadas a patrones anormales relacionados con la EP. Estos métodos se evaluaron en un conjunto de 11 sujetos control y 11 pacientes con EP. La representación cinemática fue validada en un esquema de clasificación utilizando diferentes modelos preentrenados, logrando una precisión media de 99,62 % para las regiones de los miembros inferiores y la cabeza. Mientras, para la representación convolucional los mapas de saliencia resultantes destacaron los patrones de las extremidades inferiores de los pacientes con EP, y de la cabeza y el tronco de los sujetos control, logrando una precisión de 94,89 %.

* Trabajo de grado

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

ABSTRACT

TITLE: PARKINSON'S DISEASE PATTERN CHARACTERIZATION USING MACHINE LEARNING MODELS. *

AUTHOR: LUIS CARLOS GUAYACÁN CHAPARRO. **

KEYWORDS: PARKINSON'S DISEASE, PATTERN REPRESENTATION, GAIT ANALYSIS, CLINICAL MARKERS.

DESCRIPTION: Parkinson's disease (PD) lacks a definitive diagnosis, being the observation of movement patterns the main analysis to characterize disease progression and design patients' treatments. Among PD observations, gait movement patterns, such as step length and postural control, support disease characterization. However, this analysis is usually performed with marker-based protocols, reducing the description of gait and missing local patterns during locomotion. This work introduces: 1) A markerless computational method based on a complete kinematic characterization of body segments related to PD motor impairments. 2) A 3D convolutional representation of gait in an automatic PD classification context, highlighting spatio-temporal patterns learned for this task. These approaches obtain models that classify markerless sequences and describe the main spatiotemporal regions associated with abnormal PD-related patterns. These methods were evaluated on a set of 11 control subjects and 11 PD patients. The kinematic representation was validated in a classification scheme using different pretrained models, achieving a mean accuracy of 99,62 % for the lower limb and head regions. Meanwhile, for the convolutional representation the obtained saliency maps highlighted the patterns of the lower limbs of PD patients, and the head and trunk of control subjects, achieving an accuracy of 94,89 %.

* Degree work

** Faculty of Sciences. School of Physics. Advisor: Fabio Martínez.

INTRODUCCIÓN

La enfermedad de Parkinson (EP) es el segundo trastorno neurodegenerativo más común y afecta a más de 6.2 millones de personas en todo el mundo. Según la Organización Mundial de la Salud, esta cifra aumentará a más de 12 millones para el 2030 ¹. La EP se caracteriza por la pérdida progresiva de neuronas productoras de dopamina, un neurotransmisor vital en la ejecución de los movimientos voluntarios. Como consecuencia, los pacientes con EP suelen presentar una serie de trastornos motores, como lentitud de movimientos (bradicinesia), inestabilidad postural y rigidez corporal, entre otros ².

Para estimar el estadio del paciente, la evolución de la enfermedad y medir una determinada respuesta al tratamiento sintomático, la estratificación de la enfermedad se apoya concretamente en observaciones relacionadas con trastornos de movimiento. Estos patrones anormales han sido estratificados y agrupados en diferentes escalas, como por ejemplo, la escala H&Y (Hoehn-Yahr), la UPDRS (la Escala Unificada de la Enfermedad de Parkinson) o la MDS-UPDRS (la Escala Unificada de la Sociedad de Trastornos del Movimiento de la Enfermedad de Parkinson) ³. Sin embargo, estas escalas dependen en gran medida del análisis subjetivo realizado por el médico especialista, lo que da lugar a una gran variabilidad en la estratificación

¹ VOS, Theo, *et al.* "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016". En: *The Lancet* 390.10100 (2017), págs. 1211-1259.

² DAVIS III, Roy B, *et al.* "A gait analysis data collection and reduction technique". En: *Human movement science* 10.5 (1991), págs. 575-587.

³ GOETZ, Christopher G, *et al.* "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results". En: *Movement disorders: official journal of the Movement Disorder Society* 23.15 (2008), págs. 2129-2170.

de la escala (en Post et al. (2005) ⁴ se informa de un bajo nivel de consenso en un estudio con 50 pacientes y seis evaluadores diferentes). Esto, junto con la falta de biomarcadores que apoyen plenamente el análisis médico, hace que el diagnóstico de la enfermedad de Parkinson sea un proceso altamente subjetivo. De hecho, se estima que la tasa de falsos positivos para el diagnóstico de la EP es de alrededor del 35 % en el diagnóstico inicial y del 24 % en el diagnóstico final ⁵.

Tradicionalmente, los análisis cuantitativos de los trastornos del movimiento basados en marcadores han contribuido a mitigar esa subjetividad. Estos protocolos de diagnóstico permiten estimar y seguir la posición de puntos de referencia corporales, pero a costa de una considerable simplificación de la complejidad de la marcha, siguiendo pocos puntos de interés. Además, estos métodos requieren protocolos estrictos y su naturaleza invasiva puede afectar la naturalidad de los movimientos del paciente. En particular, en el caso de los pacientes con EP, estos enfoques pasan por alto los trastornos de movimiento sutiles y localizados que puedan estar fuertemente correlacionados con la enfermedad. Por otro lado, estos análisis suelen centrarse en el estudio de las diferentes medidas de los movimientos de las extremidades inferiores, restando importancia a la información aportada por el resto de segmentos corporales ⁶ ⁷. Además, los métodos tradicionales ignoran los patrones

⁴ POST, Bart, *et al.* "Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?" En: *Movement disorders: official journal of the Movement Disorder Society* 20.12 (2005), págs. 1577-1584.

⁵ SPLETE, HEIDI. "Avoid Diagnostic Pitfalls For Parkinson's Disease". En: *Internal Medicine News* 39.17 (2006), págs. 31-31.

⁶ PISTACCHI, Michele, *et al.* "Gait analysis and clinical correlations in early Parkinson's disease". En: *Functional neurology* 32.1 (2017), pág. 28.

⁷ NAGHAVI, Nader y WADE, Eric. "Prediction of freezing of gait in Parkinson's disease using statistical inference and lower-limb acceleration data". En: *IEEE transactions on neural systems and rehabilitation engineering* 27.5 (2019), págs. 947-955.

ocultos de la EP, centrándose en proporcionar mediciones generales de la marcha e ignorando los patrones de Parkinson que surgen de la inestabilidad producida por la enfermedad. Es por esto que se requieren herramientas más especializadas para cuantificar adecuadamente los patrones parkinsonianos globales y localizados, así como para descubrir nuevas relaciones dinámicas que apoyen el diagnóstico y el seguimiento de la enfermedad.

Recientemente, se han propuesto métodos sin marcadores para capturar, describir y cuantificar los patrones de la marcha parkinsoniana dentro de un dominio específico. Por ejemplo, algunos enfoques comunes se basan en siluetas y sensores RGB-D para aproximar poses dinámicas a partir de la estimación de las articulaciones en cada fotograma, y permiten estimar medidas típicas como la longitud de los pasos, la cadencia o la velocidad global ⁸ ,⁹. Por otra parte, algunos trabajos han mostrado la relevancia de los movimientos de la cabeza, el cuello y las extremidades superiores durante la marcha en la caracterización de la EP, demostrando que dichos patrones proporcionan información complementaria y no son el resultado de un simple movimiento reflejo ¹⁰ ,¹¹. Así mismo, algunos enfoques han propuesto el uso de la extracción de características de vídeo para calcular índices o informa-

⁸ VERLEKAR, Tanmay T; SOARES, Luís D y CORREIA, Paulo L. "Automatic classification of gait impairments using a markerless 2D video-based system". En: *Sensors* 18.9 (2018), pág. 2743.

⁹ REN, Peng, *et al.* "Multivariate Analysis of Joint Motion Data by Kinect: Application to Parkinson's Disease". En: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.1 (2019), págs. 181-190.

¹⁰ BUCKLEY, Christopher, *et al.* "Quantification of upper body movements during gait in older adults and in those with Parkinson's disease: Impact of acceleration realignment methodologies". En: *Gait & posture* 52 (2017), págs. 265-271.

¹¹ BARBE, Michael T, *et al.* "Gait and upper limb variability in Parkinson's disease patients with and without freezing of gait". En: *Journal of neurology* 261.2 (2014), págs. 330-342.

ción estadística que describan la marcha de la EP ¹². Estos enfoques construyen representaciones de movimiento aparente, a partir del vídeo, que podrían ser útiles para predecir y cuantificar patrones anormales relacionados con la enfermedad. Además, los enfoques de aprendizaje profundo convolucional han aprendido características de extremo a extremo para obtener una predicción global sobre diferentes enfermedades relacionadas con la marcha.

Este trabajo introduce dos esquemas computacionales para la caracterización, cuantificación y representación de patrones de la marcha parkinsoniana:

- **Representación cinemática:** Siguiendo los esquemas clásicos de análisis, en una primera aproximación sin marcadores, se propuso una estrategia para la caracterización cinemática completa de los segmentos corporales relacionados con las alteraciones motoras de la EP. Inicialmente, el enfoque propuesto realiza una segmentación regional y dinámica de los segmentos corporales, permitiendo individualizar cabeza, tronco y extremidades. Los patrones cinemáticos se capturan en cada una de las regiones usando un conjunto de trayectorias de movimiento por píxel. Todas las trayectorias puntuales se resumen en un descriptor de la marcha propuesto mediante histogramas polares que codifican la distribución espacial de diferentes cinemáticas. Estos histogramas constituyen una representación del movimiento del segmento corporal que permite distinguir la población de control y la de EP, y podría utilizarse para evaluar la capacidad de los segmentos para predecir comportamientos anormales asociados a la enfermedad.
- **Representación convolucional:** En un segundo enfoque propuesto en el marco de este trabajo, se propuso una representación convolucional 3D de la mar-

¹² SUGANDHI, K; WAHID, Farha Fatina y RAJU, G. "Feature Extraction Methods for Human Gait Recognition—A Survey". En: *International Conference on Advances in Computing and Data Sciences*. Springer. 2016, págs. 377-385.

cha en un contexto de clasificación automática de la EP, destacando patrones espacio-temporales aprendidos para esta tarea. Este enfoque aprovecha una red neuronal convolucional (CNN, por sus siglas en inglés) que utiliza características volumétricas y jerárquicas construidas a partir de capas convolucionales 3D, permitiendo procesar naturalmente secuencias de vídeo, capturando patrones posturales y de movimiento relevantes propios de la enfermedad. El enfoque propuesto inicia entonces con el aprendizaje de una representación de vídeo de la enfermedad bajo un esquema de clasificación supervisada. A continuación, se obtiene un conjunto de mapas de saliencia representativos mediante un proceso de retropropagación de la capa final de la CNN al espacio de vídeo de entrada. Estos mapas constituyen una representación explicable que descubre las principales regiones espacio-temporales, cuando la red centra la atención en asociar patrones anormales. Esta representación puede apoyar potencialmente el diagnóstico y el seguimiento de la EP siendo utilizada como un foco de atención para los médicos. Además, el análisis de estas regiones salientes durante la marcha podría conducir a una comprensión más profunda de la enfermedad al revelar patrones ocultos asociados a esta patología. Curiosamente, durante la evaluación del enfoque propuesto, los mapas asociados a los pacientes de párkinson realizan las regiones espaciales relacionadas con los pies durante la fase de apoyo simple.

Estos enfoques permiten obtener modelos que clasifican secuencias sin marcadores y representan las principales regiones espacio-temporales asociadas a patrones anormales relacionados con la EP. Este modelos fueron evaluados en 176 secuencias de vídeo de la marcha de 22 pacientes, es decir, 11 pacientes por clase y 8 vídeos por paciente.

1. ANÁLISIS DE LA MARCHA PARKINSONIANA

El análisis de la marcha se realiza en general a partir de enfoques basados en marcadores, siguiendo protocolos como: Sistema 3D Mocap, Vicon, Cámaras IR, entre otros ¹³. Estos protocolos utilizan marcadores pasivos (puntos de referencia reflectantes), marcadores activos (LEDs) o dispositivos electrónicos (sensores portátiles) colocados en el cuerpo del paciente con el fin de capturar los patrones de movimiento durante la locomoción ¹⁴. Estos métodos son invasivos, alteran el gesto natural y simplifican mucho la dinámica de la locomoción a un conjunto reducido de trayectorias registradas a lo largo del ciclo de la marcha. En cuanto a la caracterización del párkinson, estos enfoques pueden perder, además, la sensibilidad para capturar, caracterizar y modelar movimientos como el temblor y la distonía (contracción simultánea de músculos opuestos). Por lo tanto, se han propuesto varios métodos sin marcadores para complementar el análisis clásico de la marcha, pero también para cubrir caracterizaciones nuevas y específicas de los patrones anormales relacionados con la enfermedad de Parkinson, capturados a partir de secuencias de vídeo. A partir del diseño de nuevos descriptores ha sido posible modelar y comprender patrones dinámicos específicos de la marcha. En las siguientes secciones se resumen los principales enfoques relacionados con los métodos sin marcadores, que podrían utilizarse en el modelado del párkinson.

¹³ KOUR, Navleen; ARORA, Sakshi, *et al.* "Computer-Vision Based Diagnosis of Parkinson's Disease via Gait: A Survey". En: *IEEE Access* 7 (2019), págs. 156620-156645.

¹⁴ SUTHERLAND, David H. "The evolution of clinical gait analysis: Part II Kinematics". En: *Gait & posture* 16.2 (2002), págs. 159-179.

ENFOQUES DE ESTIMACIÓN DE ARTICULACIONES

El análisis clásico de la marcha se basa en la estimación de las articulaciones a lo largo de la locomoción, lo que permite aproximar las configuraciones dinámicas de la postura y también medir los cambios particulares durante las fases específicas de la marcha. En este sentido, muchos de los enfoques sin marcadores han intentado emular estos sistemas calculando el esqueleto y las articulaciones, como primitivas para recuperar las características biomecánicas. Por ejemplo, Khan et. al.¹⁵ esqueletizan siluetas binarias para obtener características como el ángulo de inclinación de la postura y la variabilidad de la zancada y luego calculan la similitud coseno con respecto a un patrón de marcha ideal. Asimismo, Das et. al.¹⁶ generan esqueletos para extraer características de la marcha, que después se utilizan para clasificar automáticamente las enfermedades de la marcha. Para la clasificación, la combinación del Modelo Oculto de Markov (HMM) y la Máquina de Soporte Vectorial (SVM, por sus siglas en inglés) ha mostrado resultados prometedores. Además, algunos trabajos han aprovechado los sensores RGB-D (kinect), añadiendo información de profundidad, para estimar con mayor precisión las articulaciones, en comparación con el sistema de Vicon¹⁷,¹⁸. Con respecto a estos sistemas comerciales, la aproximación RGB-D puede seguir los movimientos clínicos relevantes con la mis-

¹⁵ KHAN, Taha; WESTIN, Jerker y DOUGHERTY, Mark. "Motion cue analysis for parkinsonian gait recognition". En: *The open biomedical engineering journal* 7 (2013), pág. 1.

¹⁶ DAS, Deepjoy. "Human gait classification using combined HMM & SVM hybrid classifier". En: *2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV)*. IEEE. 2015, págs. 169-174.

¹⁷ LI, Qiannan, *et al.* "Classification of gait anomalies from kinect". En: *The Visual Computer* 34.2 (2018), págs. 229-241.

¹⁸ TAN, Dawn, *et al.* "Automated analysis of gait and modified timed up and go using the Microsoft Kinect in people with Parkinson's disease: associations with physical outcome measures". En: *Medical & biological engineering & computing* 57.2 (2019), págs. 369-377.

ma precisión espacial pero con algunas limitaciones para recuperar los movimientos pequeños. Estos métodos permiten capturar el movimiento de las articulaciones sin afectar a la naturalidad de la marcha de los pacientes. Sin embargo, se limitan a emular las medidas tradicionales basadas en marcadores, simplificando la caracterización de la marcha y perdiendo información espacial y cinemática relevante.

ENFOQUES DE EXTRACCIÓN DE CARACTERÍSTICAS

Algunos enfoques sin marcadores se han interesado en modelar variables típicas de la marcha y características de vídeo no comunes para construir nuevas representaciones y diferenciar entre un conjunto de enfermedades, o niveles de diagnóstico. Por ejemplo, Chien-Wen *et al.* (2009) codifican la media y la covarianza de las siluetas, para construir descriptores de componentes principales que luego se combinan con el análisis lineal discriminante para clasificar diferentes patologías de la marcha ¹⁹. Asimismo, Shaw *et al.* (2014) propusieron un estudio para inspeccionar las características de la marcha de 16 pacientes con EP y 16 sujetos control mediante siluetas y utilizando el modelo oculto de Markov para la detección de la enfermedad ²⁰. Por otra parte, Bhaskar *et al.* (2017) propusieron un sistema con características óptimas utilizando el análisis de la varianza combinado con la selección incremental de características . Estas características seleccionadas se introdujeron como entrada a diferentes métodos de aprendizaje automático. Estos clasificadores individuales se

¹⁹ CHO, Chien-Wen , *et al.* "A vision-based analysis system for gait recognition in patients with Parkinson's disease". En: *Expert Systems with applications* 36.3 (2009), págs. 7033-7039.

²⁰ SHAW, Laxmi. "HMM based Parkinson's detection by analysing symbolic postural gait image sequences". En: *Int. J. Tech. Res. Appl* 2.4 (2014), págs. 211-216.

combinaron para producir un modelo de fusión de clasificadores ²¹.

Recientemente, algunos métodos de cuantificación del párkinson se han centrado en la medición de las aceleraciones de la cabeza y la pelvis, que a su vez se correlacionan con otras variables dinámicas y permiten predecir la EP ²². Por ejemplo, Buckley *et al.* (2015) exploró las aceleraciones de la parte superior del cuerpo para medir el control postural, y reportó una mayor magnitud en la dirección medio-lateral en personas con EP ²³. Tales hallazgos evidencian la importancia de medir el control postural en la EP, sin embargo, los sistemas de captura corporales afectan la naturalidad de la marcha del paciente y conducen a una caracterización corrupta o poco fiable de la enfermedad. Además, estos métodos limitan la exploración de otras cinemáticas, potencialmente adecuadas para la representación de la EP, como la distancia al suelo, la reducción del balanceo de los brazos y el arrastre de los pies.

Estos enfoques basados en características han logrado resultados significativos en el reconocimiento de patrones parkinsonianos, sin embargo, los resultados obtenidos con estos métodos, usualmente, no permiten una correlación clara con los síntomas y no contribuyen mucho a la comprensión de la enfermedad.

²¹ SEMWAL, Vijay Bhaskar, *et al.* "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification". En: *Multimedia tools and applications* 76.22 (2017), págs. 24457-24475.

²² BUCKLEY, Christopher, *et al.* "Upper body accelerations as a biomarker of gait impairment in the early stages of Parkinson's disease". En: *Gait & posture* 71 (2019), págs. 289-295.

²³ BUCKLEY, Christopher, *et al.* "Attenuation of upper body accelerations during gait: piloting an innovative assessment tool for Parkinson's disease". En: *BioMed research international* 2015 (2015).

ENFOQUES BASADOS EN REDES NEURONALES CONVOLUCIONALES

Recientemente, los enfoques de aprendizaje profundo han superado esquemas clásicos del aprendizaje automático y otras estrategias informáticas para aprender y modelar patrones complejos que representan instancias de alto nivel, como la enfermedad. Más concretamente, las CNN permiten un modelado de extremo a extremo, que abarca el descubrimiento de características visuales, así como, la clasificación de instancias. Estas CNN construyen características a diferentes niveles de complejidad y logran una representación invariable de las acciones.

El reconocimiento de acciones en secuencias de vídeo trata de asociar patrones de vídeo con instancias particulares de actividad humana. Esta área podría proyectarse potencialmente en el análisis automático, el apoyo a la clasificación y a la predicción de enfermedades. Por ejemplo, en Simonyan *et al.* (2014)²⁴ se propuso una arquitectura de CNN que codifica secuencias de video en bruto (RGB) y de flujo óptico en dos ramas convolucionales. Sin embargo, ese trabajo demostró ser poco descriptivo temporalmente, ya que apilando de 5 a 10 imágenes de flujo óptico los resultados no marcaban una diferencia significativa. Además, se han implementado arquitecturas neuronales recurrentes²⁵ para modelar patrones no lineales que podrían estar asociados a secuencias de movimiento complejas. Estos enfoques podrían ser interesantes para descubrir relaciones no lineales entre fases de la marcha diferentes y distantes. Sin embargo, estos enfoques codifican dichas relaciones temporales a partir de vectores codificados que pierden la representación espacial de los vídeos, lo que podría ser determinante para examinar las configuraciones posturales del

²⁴ SIMONYAN, Karen y ZISSERMAN, Andrew. "Two-stream convolutional networks for action recognition in videos". En: *Advances in neural information processing systems*. 2014, págs. 568-576.

²⁵ DONAHUE, Jeffrey, *et al.* "Long-term recurrent convolutional networks for visual recognition and description". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, págs. 2625-2634.

paciente.

Las CNN también se han implementado para codificar patrones relacionados con la enfermedad de Parkinson utilizando otros sensores. Por ejemplo se han obtenido nuevas representaciones de aprendizaje a partir de señales de acelerómetros ²⁶, utilizando características de audio y de voz ²⁷ e incluso modelando cambios de contraste en secuencias de resonancia magnética ²⁸. Estos modelos han permitido cuantificar representaciones que permiten una mejor comprensión de la enfermedad. Sin embargo, estas estrategias dependen de protocolos sofisticados y muchas veces son estrategias invasivas.

²⁶ OH, Shu Lih, *et al.* "A deep learning approach for Parkinson's disease diagnosis from EEG signals". En: *Neural Computing and Applications* (2018), págs. 1-7.

²⁷ GUNDUZ, Hakan. "Deep learning-based Parkinson's disease classification using vocal feature sets". En: *IEEE Access* 7 (2019), págs. 115540-115551.

²⁸ MARTINEZ-MURCIA, Francisco Jesús, *et al.* "A 3D convolutional neural network approach for the diagnosis of Parkinson's disease". En: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2017, págs. 324-333.

2. PROBLEMA DE INVESTIGACIÓN

La enfermedad de Parkinson es el segundo desorden neurodegenerativo más común en la población, evidenciado macroscópicamente por alteraciones motoras. A pesar de los múltiples esfuerzos por entender el origen y desarrollo de esta enfermedad, hasta el día de hoy, no existe una prueba y/o biomarcadores definitivos que puedan soportar su diagnóstico. En la rutina clínica, el análisis de patrones anormales de movimiento, se basan típicamente de forma observacional, basados exclusivamente en el conocimiento del experto. Esta tarea subjetiva conlleva a múltiples limitaciones, tales como: error en el diagnóstico hasta del 25 % y diagnóstico únicamente efectivo en estadios avanzados de la enfermedad.

Modelos basados en marcadores y sensores corporales permiten mitigar la subjetividad en la valoración del párkinson. Sin embargo estas aproximaciones son modelos bastante simplificados, que toman en cuenta pocas trayectorias (asociadas a las articulaciones), para describir procesos complejos, como la locomoción. Así estos métodos emulan las observaciones clásicas de los expertos y rara vez tienen en cuenta la correlación entre ellos. Además la condición invasiva de estos métodos pueden llegar a afectar la naturalidad de los movimientos del paciente, interfiriendo en la captura de las señales que se desean estudiar.

Recientemente, el análisis de video ha permitido desarrollar descriptores que, basados en información espacio-temporal, permiten encontrar mayor información dinámica relacionada con cambios geométricos de estructuras de interés durante el tiempo. Lo anterior, sumado con algoritmos de aprendizaje de máquina, han permitido predecir y modelar patrones visuales asociados con la enfermedad de Parkinson. Estas aproximaciones aún son dependientes de la apariencia y patrones visuales, que pueden inferir con la independencia del modelamiento de patrones dinámicos de la enfermedad. Además, algunas de estas inferencias de las secuencias de vi-

deo, presentan relaciones dependientes de la posición en el video, lo cual puede afectar la efectividad en la descripción de procesos complejos de la enfermedad. El desarrollo de descriptores embebidos en espacios de baja dimensionalidad, basado en el modelos de aprendizaje de máquina, puede conducir al análisis de nuevos marcadores relacionados con la enfermedad.

Pregunta de Investigación: ¿Cómo detectar patrones motores de la enfermedad de Parkinson a partir de representaciones obtenidas de secuencias de video?

3. OBJETIVOS PROPUESTOS

Objetivo General

Proponer un modelo de aprendizaje para la representación de patrones relacionados con el párkinson, registrados en secuencias de video.

Objetivos Específicos

- Seleccionar un dataset con secuencias de video de pacientes control y pacientes de párkinson en las que se registren patrones motores afectados por la enfermedad.
- Seleccionar un modelo de representación de patrones espacio-temporales presentes en las secuencias de video.
- Codificar características espacio-temporales que representen la EP mediante la implementación del modelo previamente seleccionado.
- Evaluar la estrategia desarrollada en el conjunto de videos seleccionado.

4. MÉTODO 1: REPRESENTACIÓN CINEMÁTICA USANDO TRAYECTORIAS DE MOVIMIENTO

En este capítulo se introduce una estrategia sin marcadores para observar y analizar la cinemática regional de los principales segmentos corporales involucrados durante la marcha. Este enfoque permite emular análisis clínicos, encontrar relaciones de movimiento relacionados con la enfermedad de Parkinson. En la metodología propuesta, inicialmente se obtiene una segmentación temporal de la cabeza, el torso y las extremidades superiores e inferiores. A continuación, se mapean las trayectorias de movimiento sobre cada segmento corporal segmentado, obteniendo una descripción de los movimientos locales durante la marcha. Entonces, se calculan descriptores cinemáticos representados como histogramas polares (*dartboards*) de las trayectorias que convergen en un mismo segmento. La representación de cada segmento cinemático puede ser tratado independientemente con respecto a la capacidad de predecir y agrupar a los pacientes diagnosticados con EP. En las siguientes subsecciones se presenta una descripción de cada paso de este método. También la figura 2 bosqueja de forma general los principales pasos propuestos en este método. *El contenido de este capítulo ha sido integrado en un artículo que se encuentra en revisión en la revista Biomedical Physics & Engineering Express.*

4.1. ANÁLISIS REGIONAL

La enfermedad de Parkinson se diferencia de otras patologías neurológicas por la asimetría corporal lateral en los síntomas, y la focalización de las alteraciones mo-

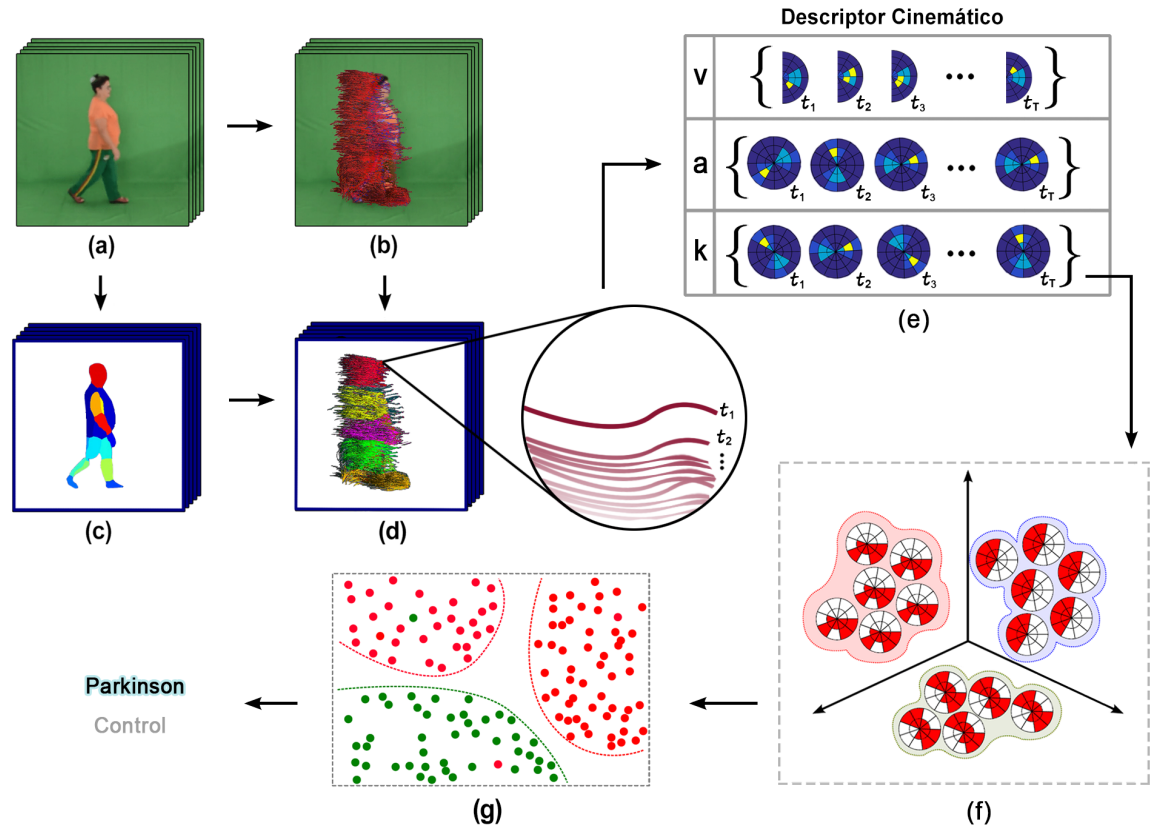


Figura 1. **a)** Secuencia de vídeo original. **b)** Estimación de trayectorias de movimiento densas. **c)** Máscaras de segmentación corporal. **d)** Trayectorias densas agrupadas por segmento corporal. **e)** Para cada trayectoria t_i , la distribución espacial del desplazamiento $2D$ (velocidad), su derivada (aceleración) y la curvatura (orientada hacia la aceleración radial) se codifica en histogramas cinemáticos polares. **f)** Bolsa de palabras sobre un espacio multidimensional. **g)** Esquema de clasificación binaria.

toras en partes específicas del cuerpo, especialmente en las etapas tempranas ²⁹. Sin embargo, gran parte de las estrategias propuestas sin marcadores sólo realizan una caracterización global durante la marcha, perdiendo la descripción de los movimientos localizados. De hecho, la cabeza, el tronco y las extremidades superiores e

²⁹ CHOE, Bo-Young, *et al.* "Neuronal laterality in Parkinson's disease with unilateral symptom by in vivo 1H magnetic resonance spectroscopy". En: *Investigative radiology* 33.8 (1998), págs. 450-455.

inferiores describen patrones parkinsonianos específicos que deben ser tratados de forma independiente durante el análisis cinemático debido a la dinámica particular, la asimetría y la respuesta de partes específicas del cuerpo durante la locomoción^{30, 11}. Por ejemplo, se ha demostrado que los movimientos anormales de la cabeza producidos por la inestabilidad del tronco son significativos para describir la enfermedad de párkinson³¹ pero siguen siendo poco explorados. Teniendo en cuenta estas consideraciones, el primer paso de este trabajo fue obtener una segmentación dinámica densa para aproximar segmentos anatómicos³².

En la actualidad, los enfoques de aprendizaje profundo han logrado una demarcación automática fiable de múltiples objetos mediante una segmentación semántica supervisada, es decir, una segmentación densa que asigna una etiqueta de clase a cada píxel de la imagen. En general, estos enfoques aprenden a partir de imágenes segmentadas manualmente a través de un marco convolucional de extremo a extremo que proporciona un mapa de características sobre la localización del objeto y el tamaño correspondiente³³. Para ello, la respuesta en el mapa de características de salida se proyecta en la imagen de entrada para establecer los objetos más probables de diferentes tamaños, que posteriormente se delimitan como propuestas de objetos o regiones de interés (RoI, por sus siglas en inglés). Más recientemente, se propuso un enfoque de segmentación de máscaras (Mask R-CNN) con una ra-

³⁰ UTTI, Ryan J, *et al.* "Parkinson disease: handedness predicts asymmetry". En: *Neurology* 64.11 (2005), págs. 1925-1930.

³¹ KAVANAGH, Justin; BARRETT, Rod y MORRISON, Steven. "The role of the neck and trunk in facilitating head stability during walking". En: *Experimental brain research* 172.4 (2006), pág. 454.

³² ALP GÜLER, Rıza; NEVEROVA, Natalia y KOKKINOS, Iasonas. "Densepose: Dense human pose estimation in the wild". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, págs. 7297-7306.

³³ REN, Shaoqing, *et al.* "Faster r-cnn: Towards real-time object detection with region proposal networks". En: *Advances in neural information processing systems*. 2015, págs. 91-99.

ma adicional para predecir máscaras de segmentación en cada Rol y devolver una máscara delineada de tales objetos ³⁴. Esta segunda rama tiene cuatro capas convolucionales adicionales que son muestreadas por capas convolucionales de transposición, llevando las dimensiones de la máscara de características a $S \times S \times K$, donde S es el tamaño espacial y K es el número de clases.

A partir de estos principios, Güler *et. al.* (2018) propusieron *DensePose* ³², un enfoque que obtiene regiones de segmentos corporales específicos. En este trabajo se implementó *DensePose* para obtener una máscara de pose general, pero también partes de segmentación del cuerpo durante la marcha. El modelo fue entrenado con unas 50.000 imágenes con el esquema *Mask R-CNN* para aprender la correspondencia entre las imágenes de entrenamiento y las etiquetas de los segmentos corporales. *DensePose* consigue, en tiempo real, una segmentación robusta de las partes del cuerpo en presencia de fondos complejos, oclusiones y variaciones de escala. En nuestro trabajo, sólo se utiliza el canal que corresponde a las etiquetas de segmentación de las partes del cuerpo para agrupar las trayectorias de movimiento. La segmentación de la silueta corporal en secciones de movimiento rígidas facilita la descripción de patrones con significado semántico y sintomático. Además, es eficaz para diferenciar cada lado, es decir, las extremidades derechas de las izquierdas, permitiendo la identificación de patrones de EP asimétricos. En la Figura 1.c se ilustra un ejemplo de segmentación dinámica de partes del cuerpo obtenida en nuestro trabajo.

³⁴ HE, Kaiming, *et al.* "Mask r-cnn". En: *Proceedings of the IEEE international conference on computer vision*. 2017, págs. 2961-2969.

4.2. TRAYECTORIAS DENSAS DE MOVIMIENTO

El flujo óptico es la herramienta más conocida y estándar para capturar patrones de movimiento de los vídeos. En resumen, calcula el movimiento aparente entre dos fotogramas consecutivos, lo que da lugar a un campo vectorial en el que cada muestra vectorial representa el desplazamiento local de cada píxel ³⁵, ³⁶, ³⁷. A partir de este campo de movimiento es posible recuperar patrones espacio-temporales que pueden ser representativos de algunos comportamientos humanos concretos. Sin embargo, estos patrones podrían ser insuficientes para estudiar las características de la marcha, que suele estudiarse a lo largo de un conjunto de movimientos semiperiódicos conocidos como ciclos de marcha (desarrollados a través de múltiples fotogramas). Con mayor razón, en el caso de la enfermedad de Parkinson, donde los pacientes suelen mostrar lentitud en sus movimientos ³⁸, el flujo óptico podría ser insuficiente para capturar patrones cinemáticos relevantes.

Por lo tanto, una forma natural de ampliar el flujo óptico es seguir el movimiento local a través del vídeo, lo que da lugar a largas trayectorias de puntos, que representan los movimientos de los pacientes durante la marcha ³⁹. Así pues, las trayectorias densas son primitivas de movimiento local basadas en la concatenación de vecto-

³⁵ SUN, Shuyang, *et al.* "Optical flow guided feature: A fast and robust motion representation for video action recognition". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 1390-1399.

³⁶ SUN, Deqing, *et al.* "Learning optical flow". En: *European Conference on Computer Vision*. Springer. 2008, págs. 83-97.

³⁷ SEVILLA-LARA, Laura, *et al.* "On the integration of optical flow and action recognition". En: *German Conference on Pattern Recognition*. Springer. 2018, págs. 281-297.

³⁸ MARSDEN, CD. "Slowness of movement in Parkinson's disease". En: *Movement Disorders: Official Journal of the Movement Disorder Society* 4.S1 (1989), S26-S37.

³⁹ WANG, Heng y SCHMID, Cordelia. "Action Recognition with Improved Trajectories". En: *IEEE International Conference on Computer Vision*. Sydney, Australia, 2013.

res de flujo óptico, que rastrean numerosos puntos a lo largo de la secuencia de vídeo, lo que permite una representación más robusta de los patrones de movimiento ⁴⁰ (véase la Figura 1.b). Dado que las trayectorias abarcan un gran número de fotogramas, facilitan la descripción de estructuras temporalmente complejas de la enfermedad de Parkinson. El muestreo uniformemente denso sobre el vídeo de estas trayectorias proporciona una rica descripción regional de los movimientos de los pacientes durante la locomoción. La extracción de trayectorias densas comienza con el cálculo del flujo denso de Farnebäck, basado en una expansión polinómica ⁴¹. En este método el valor de vecindad (nivel de gris) f de cada píxel de la imagen se aproxima mediante una expresión polinómica cuadrática, definida como: $f(\mathbf{x}) \simeq \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, donde $\mathbf{x} = (x, y)^T$ son las posiciones de los píxeles, $\mathbf{A}$, $\mathbf{b}$ y c son los términos desconocidos cuadrático (matriz), lineal (vector) y constante (escalar), respectivamente. Luego, el desplazamiento del campo vectorial $\mathbf{d}$ se obtiene a partir de los coeficientes del polinomio en fotogramas consecutivos f_t y f_{t+1} , de la forma: $\mathbf{d} = -\frac{1}{2} \mathbf{A}_t^{-1} (\mathbf{b}_{t+1} - \mathbf{b}_t)$. A continuación, el flujo denso calculado se suaviza utilizando un filtro mediano 2d sobre el campo de desplazamiento. En la práctica, los puntos $\mathbf{x}_t = (x_t, y_t)$ se muestrean densamente en una cuadrícula regular, se rastrean en el tiempo t a diferentes escalas espaciales y se redondean a valores enteros tras el suavizado de la mediana. Los puntos coincidentes se concatenan finalmente para formar una trayectoria $(\mathbf{x}_{t_0}, \mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \dots, \mathbf{x}_{t_n})$.

⁴⁰ WANG, Heng, *et al.* "Action recognition by dense trajectories". En: *CVPR 2011*. IEEE. 2011, págs. 3169-3176.

⁴¹ FARNEBÄCK, Gunnar. "Two-frame motion estimation based on polynomial expansion". En: *Scandinavian conference on Image analysis*. Springer. 2003, págs. 363-370.

4.3. REPRESENTACIÓN CINEMÁTICA REGIONAL

Partiendo de la base de que los síntomas motores de la enfermedad de Parkinson pueden caracterizarse adecuadamente mediante el cálculo de la cinemática regional dada por la velocidad y sus derivadas, se propone aquí una representación cinemática a partir de trayectorias densas. Para ello, se propone un esquema de histogramas polares (*dartboards*) para codificar el comportamiento cinemático de la marcha. La idea principal de este descriptor es que la cinemática local durante la marcha puede ser descrita por la distribución de los movimientos de las partes del cuerpo en términos de sus magnitudes y direcciones. Cada trayectoria puede verse como una serie de vectores de movimiento concatenados. Como cualquier vector, se describen a partir de su magnitud y orientación, y dentro de cada trayectoria, se conforma un histograma de ocurrencia de vectores en un histograma polar cuantificado (como se representa en la Figura 2). Por último, para mejorar la precisión, los histogramas locales se normalizan utilizando la norma L2. Esta codificación permite aproximar una representación cinemática estadística de los movimientos patológicos de la EP durante la marcha. El uso de diferentes medidas cinemáticas puede ser fundamentado de acuerdo a los aspectos de la enfermedad, como:

- La bradicinesia o lentitud de movimiento, uno de los principales signos de la EP, puede evaluarse mediante patrones de velocidad. Este trastorno del movimiento no está causado por la debilidad muscular, sino por la transmisión incorrecta de las señales neuromusculares. La bradicinesia puede afectar a una extremidad, a un lado del cuerpo o a todo el cuerpo, lo que puede definir una firma de locomoción particular para cada paciente. Para describir cinemáticamente la bradicinesia, sea $\vec{v} = \mathbf{v}(t)$ el desplazamiento 2D de un elemento del vídeo entre los fotogramas t y $t + 1$. Construimos una cuantificación de medio histograma polar (half-dartboard) del vector $\vec{v}$ en cuanto a su magnitud y

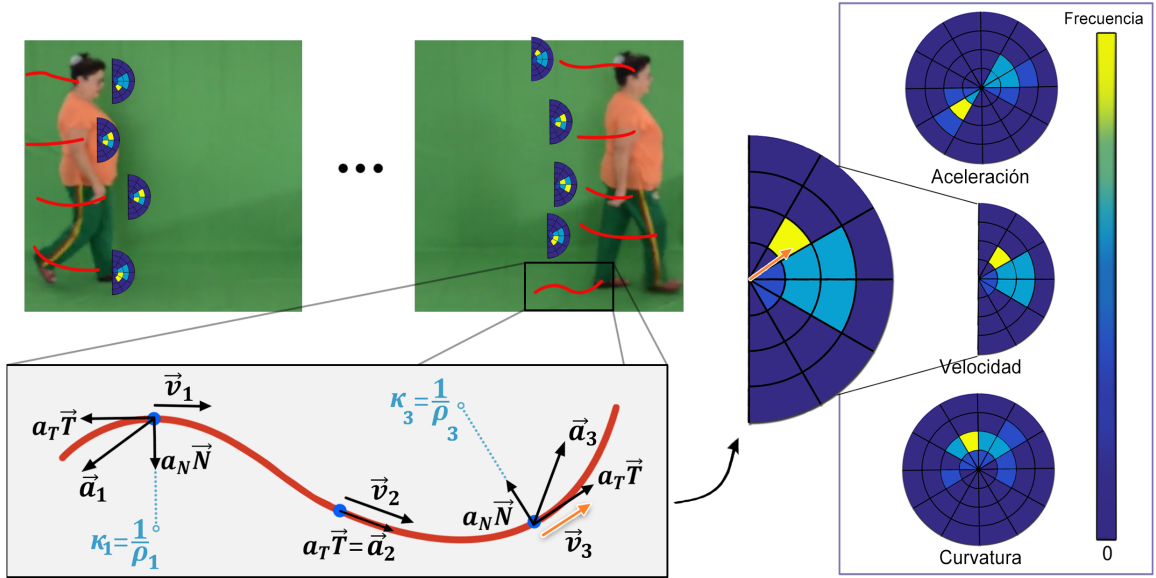


Figura 2. Construcción del descriptor a nivel de trayectoria.

dirección, ya que se supone que la marcha tiene la misma dirección para toda la secuencia, como se representa en la Figura 2.

- Algunos síntomas típicos de la EP, como el temblor, y otros como la discinesia (movimientos involuntarios e incontrolables) o la distonía (contracción simultánea de músculos opuestos) se caracterizan por la variación repentina de la velocidad o la dirección de los movimientos, es decir, por patrones de aceleración. Además, la inestabilidad del tronco y la cabeza provoca la aparición de movimientos compensatorios bruscos para mantener el equilibrio y evitar las caídas. Además, los síntomas parkinsonianos suelen variar en periodos cortos de tiempo. Una forma razonable de describir cinemáticamente estos síntomas es calcular la aceleración local de los segmentos del cuerpo: $\vec{a} = \frac{d\vec{v}}{dt}$. De forma análoga a la codificación de la velocidad (pero simétrica para admitir las aceleraciones negativas), se realiza una cuantificación en un histograma polar de los vectores de aceleración en términos de magnitud y dirección.

- Por último, la rigidez hace que los músculos se tensionen debido a su incapacidad para relajarse, hecho que podría representarse a partir de las trayectorias midiendo sus patrones de curvatura. Este síntoma típico puede impedir que los músculos se estiren y relajen como deberían. Como consecuencia, la rigidez puede experimentarse como: músculos rígidos y/o inflexibles, la postura encorvada comúnmente asociada a la EP, dolor y calambres musculares, dificultad para girar, reducción del balanceo de los brazos, entre otros. La información de la curvatura de los movimientos de los segmentos del cuerpo puede describir todos estos síntomas. Matemáticamente se puede escribir como $\kappa = \frac{a_N}{\|\vec{v}\|^2}$, donde $a_N = \frac{\|\vec{v} \times \vec{a}\|}{\|\vec{v}\|}$. Como la dirección también es necesaria para su codificación en el histograma polar, utilizamos un vector con dirección normal al desplazamiento (es decir, dirección de la aceleración radial, que apunta hacia el centro del círculo osculante a la trayectoria), y con magnitud dada por la curvatura.

4.4. CODIFICACIÓN Y PREDICCIÓN GLOBAL

Para cada segmento corporal las unidades cinemáticas se agrupan por patrones de ocurrencia. En este trabajo se han explorado tres versiones diferentes de descriptores regionales, que se describen a continuación:

- En la primera versión, para cada región, se ejecuta un algoritmo de agrupación en los descriptores a nivel de trayectoria, para obtener N grupos utilizando un enfoque de k-means. El centroide de cada grupo se considera entonces como información atómica representativa. El descriptor de cada región se define entonces como el conjunto de N centroides. Esta versión se denomina *KC* en la sección de resultados.
- La segunda versión es un enfoque de bolsa de palabras, en el que se construye

un diccionario de N palabras para cada región, utilizando un algoritmo de agrupación en los descriptores a nivel de trayectoria extraídos para un segmento corporal concreto. A continuación, cada descriptor de región se construye como un histograma de ocurrencia sobre el diccionario regional. Esta versión se denomina *RD* en la sección de resultados.

- La tercera versión también utiliza un enfoque de bolsa de palabras, para formar descriptores regionales, pero esta vez se construye un único diccionario global, y el mismo diccionario se utiliza para codificar cada uno de los descriptores regionales por histograma de ocurrencias. Esta versión se denomina *GD* en la sección de resultados.

Estos histogramas temporales por regiones constituyen los descriptores finales que permiten describir las partes rígidas durante la locomoción, y destacan la cinemática que describe estadísticamente el tipo de marcha. Cada descriptor resultante se mapea a una estrategia de aprendizaje para realizar una clasificación automática. En este trabajo se implementó una Máquina de Soporte Vectorial (SVM, por sus siglas en inglés), una técnica ampliamente utilizada en muchos problemas de aprendizaje supervisado gracias a un atractivo balance entre precisión y coste computacional⁴². Aquí las clases son *Párkinson* y *Control*, y los hiperplanos óptimos las separan utilizando una fórmula clásica de margen máximo. Se realizó un análisis de sensibilidad de los parámetros (γ, C) utilizando un esquema de validación cruzada k-fold y seleccionando los parámetros con el mayor número de verdaderos positivos. Como esquema de clasificación alternativo, también se evaluó el *Random Forest* (RF) que permite una representación discreta del problema de clasificación⁴³.

⁴² CHANG, Chih-Chung y LIN, Chih-Jen. "LIBSVM: a library for support vector machines". En: *ACM transactions on intelligent systems and technology (TIST)* 2.3 (2011), pág. 27.

⁴³ BREIMAN, Leo. "Random forests". En: *Machine learning* 45.1 (2001), págs. 5-32.

5. METODO 2: REPRESENTACIÓN CONVOLUCIONAL USANDO MAPAS DE SALIENCIA

Como segundo enfoque propuesto, en este capítulo se presenta una representación profunda de la marcha de la EP, que explota las relaciones espacio-temporales durante la locomoción. Esta representación puede constituir un potencial biomarcador de imagen de la enfermedad. Una contribución principal de este trabajo es entonces un conjunto de secuencias de imágenes que representan las regiones espacio-temporales de atención, que permiten destacar las principales regiones asociadas con la EP. En primer lugar, se consigue una representación robusta mediante una red neuronal profunda que recupera las características volumétricas del vídeo a partir del aprendizaje de un conjunto de kernels 3D, organizado jerárquicamente en diferentes capas. Esta red convolucional permite obtener una representación robusta del movimiento de la marcha, explotando las relaciones temporalmente largas, y por lo tanto capturar eficazmente los gestos parkinsonianos descriptivos durante la marcha. Esta arquitectura se evaluó a partir de dos entradas diferentes, es decir, valores de píxeles de vídeo y campos vectoriales de flujo óptico que describen la marcha en términos de mapas densos de velocidad instantánea.

Bajo un esquema supervisado, la red convolucional tiene la capacidad de predecir y clasificar cada marcha como parkinsoniana o control. Una vez entrenada la red, se propuso una retropropagación para obtener mapas volumétricos que intentan explicar una decisión particular tomada por la red al clasificar un vídeo. En el caso de los vídeos parkinsonianos, estos mapas representan las regiones en las que la red centra la atención para localizar patrones anormales. El análisis de estas regiones destacadas de la marcha podría conducir a una comprensión más profunda de la enfermedad y convertirse en una herramienta de apoyo al diagnóstico. El proceso de la estrategia propuesta se ilustra en la figura 3. *Una primera versión del con-*

tenido de este capítulo fue presentada en la conferencia EMBC-2020 (42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society), Montreal, y una segunda versión se encuentra en revisión en la revista Biomedical Informatics.

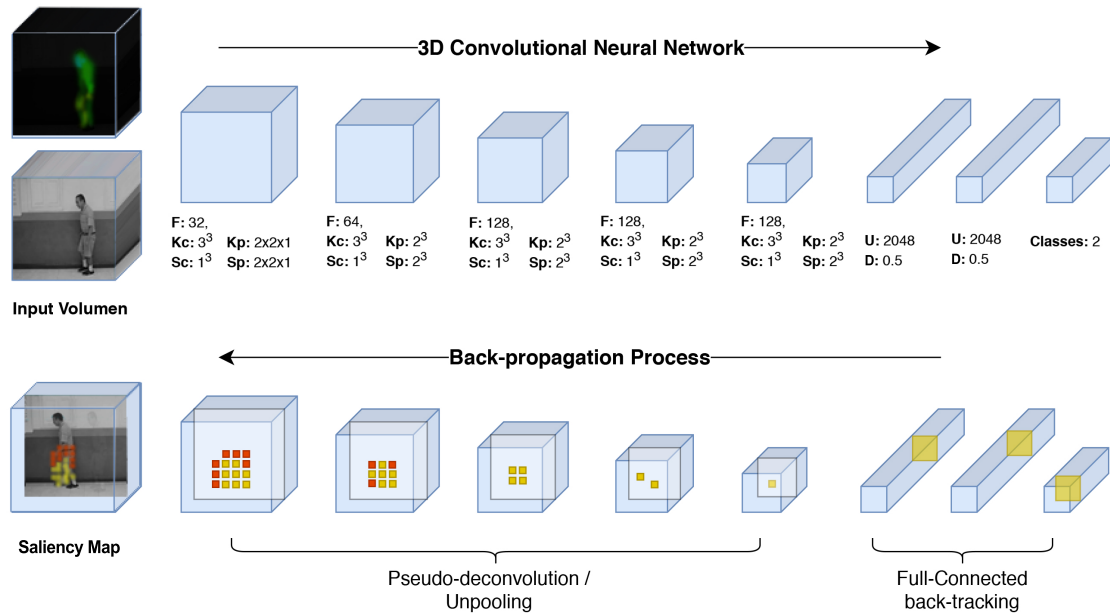


Figura 3. Esquema del enfoque propuesto. Arriba: se implementa un esquema de convolución 3D para capturar patrones de movimiento espacio-temporales relacionados con la EP. Esta arquitectura tiene la capacidad de clasificar automáticamente el párkinson con respecto a los sujetos de control. Abajo: un esquema de retropropagación permite obtener las regiones más destacadas de la enfermedad.

5.1. APRENDIZAJE DE LA SECUENCIA VOLUMÉTRICA DE LA MARCHA

Los patrones de locomoción predominantes y las principales fases de la marcha se describen a lo largo de ciclos, que pueden durar una media de 1,07 segundos para

las personas sanas ⁴⁴. En el caso de los pacientes con EP, estos patrones de locomoción pueden cambiar su periodicidad y los principales marcadores temporales de la marcha pueden describirse en períodos de tiempo más largos. Para el análisis de vídeo, los enfoques recientes han intentado capturar la estructura espacio-temporal de los movimientos y aprender una representación de la acción con redes neuronales convolucionales. Sin embargo, estas representaciones suelen aprenderse a nivel de fotograma, o codificando intervalos de tiempo muy cortos, lo que no permite modelar patrones de locomoción complejos ni recuperar las principales etapas de la marcha.

En este trabajo, se introduce una estrategia de convolución 3D para resaltar los gestos anormales de la marcha parkinsoniana en secuencias de vídeo. Implementamos una representación de vídeo temporalmente larga a partir de volúmenes espacio-temporales. ⁴⁵. Esta arquitectura considera convoluciones espaciales y temporales para aprender localmente características espacio-temporales relacionadas con la EP a lo largo de los vídeos, permitiendo procesar naturalmente las secuencias de vídeo y capturar patrones posturales y de movimiento relevantes. Bajo el supuesto de que la información más relevante se encuentra en el tiempo y no en el espacio (por ejemplo, los movimientos de bradicinesia se desarrollan en decenas de fotogramas), la arquitectura en cada capa aumenta la longitud temporal de las representaciones mientras que disminuye la resolución espacial con respecto a los enfoques anteriores de 3D-CNN. Además, la simplicidad de este esquema de 3D-CNN mantiene las activaciones trazables a través de la red. La 3D-CNN aquí implementada

⁴⁴ MURRAY, M Pat; DROUGHT, A Bernard y KORY, Ross C. "Walking patterns of normal men". En: *JBJS* 46.2 (1964), págs. 335-360.

⁴⁵ VAROL, Gül; LAPTEV, Ivan y SCHMID, Cordelia. "Long-term temporal convolutions for action recognition". En: *IEEE transactions on pattern analysis and machine intelligence* 40.6 (2017), págs. 1510-1517.

puede entenderse en dos etapas diferentes. En primer lugar, una serie de capas convolucionales, cada una de las cuales es seguida por un proceso de no linealización (*pooling*) que aprende las características de la marcha en diferentes niveles de complejidad. La segunda etapa describe una serie de capas densas (*full-connected*) a partir de las cuales se obtiene una representación embebida de las características convolucionales aprendidas de alto nivel. Estas dos etapas se describen a continuación:

- **Representación convolucional 3D:** En la primera etapa de la red, cada capa L consiste en una capa convolucional 3D seguida de una capa de max-pooling. Esta etapa permite una representación robusta de los patrones espacio-temporales, aprendiendo las características salientes de la marcha en diferentes niveles y asociadas a entidades específicas. Sea $\mathbf{I}(\mathbf{x})_t$ cada imagen en la secuencia de video, en el tiempo t , donde $\mathbf{x} = (x, y)^T$ son los píxeles espaciales y W y H representa la anchura y la altura de la imagen, respectivamente. Para cada capa L , estas arquitecturas computan transformaciones lineales progresivamente, seguidas de no linealidades contractivas, proyectando la información sobre un conjunto de q filtros aprendidos, expresados como: $\Psi_L = \{\psi_i\}_{i=1}^q$. De este modo, la imagen $\mathbf{I}(\mathbf{x})_t$ se filtra inicialmente en la primera capa por el conjunto Ψ_1 , obteniendo una representación: $\Phi_1(\mathbf{x}) = \sum_{g=1 \dots q} \mathbf{I}(\mathbf{x})_t * \psi_g$, donde ψ_g es cada filtro de convolución aprendido de forma independiente. La representación resultante $\Phi_1(\mathbf{x})$ se convoluciona sucesiva y progresivamente con los filtros de las siguientes $N - 1$ capas $\{\Psi_i\}_{i=2}^N$. La representación convolucional está definida por las representaciones de cada capa: $\{\Phi_i(\mathbf{x})\}_{i=1}^N$.

Las respuestas de las primeras capas de esta red generan características de bajo nivel y generalizadas, como los bordes y las texturas. Estas características de bajo nivel se combinan para construir representaciones más complejas

en las capas convolucionales posteriores. Finalmente, las características encontradas en la última capa convolucional corresponden a características de nivel superior que pueden representar la semántica de los datos. La red propuesta en este trabajo está compuesta por cinco capas convolucionales 3D como se representa en la Figura 3.

- Representación embebida: La representación jerárquica y convolucional aprendida es embebida en una red totalmente conectada que aprende combinaciones no lineales de estas características obteniendo una representación vectorial continua de baja dimensionalidad. En estos vectores embebidos, las neuronas tienen conexiones completas con todas las activaciones de la capa anterior, lo que permite, entre otras cosas, clasificar la EP. Formalmente, estas capas pueden describirse de la siguiente forma:

$$X_L = X_{L-1} \odot W_{L-1} + B_{L-1}$$

donde X es la entrada de la capa, W son los pesos y B es el sesgo asociado. En este trabajo, esta etapa consta de tres capas densas. La última capa corresponde a la *softmax* cuyo número de unidades es equivalente al número de clases. En este trabajo, este tipo de representaciones puede caracterizar los patrones parkinsonianos, registrados en secuencias de vídeo, lo que puede conducir al análisis de las relaciones visuales relacionadas con la enfermedad en un espacio compacto embebido.

5.2. PATRONES PARKINSONIANOS A PARTIR DE MAPAS VOLUMÉTRICOS DE SALIENCIA

Uno de los principales aportes de este trabajo fue el análisis de los patrones espaciotemporales parkinsonianos obtenidos a partir de la representación convolucional,

entendidos como las principales regiones volumétricas que contribuyen a la clasificación de una determinada enfermedad. Para ello, un primer esquema que se propuso fue la obtención de mapas de saliencia de atención mediante el retroceso de las activaciones desde la representación vectorial embebida hasta las primeras capas volumétricas. Para ello, se analiza la activación neuronal que contribuye principalmente en el valor más alto de la capa posterior en un esquema de predicción inversa. Dado que el enfoque propuesto tiene capas embebidas y convolucionales 3D, se llevaron a cabo dos esquemas de seguimiento (back-tracking) diferentes:

- **Embedded back-tracking:** En esta primera etapa los vectores embebidos son rastreados hacia atrás seleccionando el índice mayor, a lo largo de las capas totalmente conectadas, en la capa L que señala el vector más relevante en la capa $L - 1$. Formalmente, esta relación puede expresarse como:

$$E_{L-1} = \arg \max_j (X_{L-1} \odot W_{L-1}^{j, E_L})$$

donde X es la entrada de la capa $L - 1$ con forma $1 \times n$ y W son los pesos de la capa $L - 1$ con forma $n \times m$ (n y m representan los tamaños de entrada y salida de la capa $L - 1$, respectivamente). En W , los j, E_L son la posición de la fila y la columna, respectivamente. Por ejemplo, el índice de probabilidad más significativo en la capa softmax (E_L) apunta al índice máximo en el vector correspondiente de la capa anterior (E_{L-1}). Por último, el índice obtenido con el seguimiento hacia atrás en la capa de aplanamiento (*flatten*) se codifica en índices volumétricos (*unflattening*).

- **Pseudo-deconvolución/Unpooling.** En este proceso, los índices más relevantes obtenidos en el proceso de *Embedded back-tracking* se proyectan a capas volumétricas y se trazan hasta el vídeo de entrada. Esto permite obtener una representación 3D de la información espacio-temporal de salida respecto

a lo aprendido por la red neuronal. En general, se realiza un *unpooling* para cada capa seguido de un proceso de pseudo-deconvolución. Para el proceso de *unpooling* basta con leer la ubicación de los valores máximos de *pooling*, previamente almacenados durante el entrenamiento de la red. El conjunto de índices de *unpooling* en la capa L se denota como $U_L = \{(t_i, h_i, w_i)\}_{i=1}^n$ donde t , h y w representan las coordenadas de profundidad, altura y ancho de cada índice, respectivamente. Cada índice de *unpooling* representa una región en la activación de la capa anterior. Para el proceso de pseudodeconvolución, esta región se multiplica elemento por elemento con el conjunto de kernels, y los resultados del argumento máximo se devuelven como las posiciones 3D más relevantes en la capa anterior. Este proceso se realiza sucesivamente hasta la imagen de entrada, revelando la posición más importante que contribuye a la etiqueta de clasificación. Formalmente, esto puede expresarse de la siguiente forma:

$$D_L = \bigcup_{u \in U_L} \bigcup_{\psi \in \Psi_L} \arg \max(X_L^u \odot \psi)$$

Donde Ψ_L es el conjunto de kernels convolucionales de la capa L y X^u es el segmento $3 \times 3 \times 3$ del tensor X centrado en la posición u . A continuación, este proceso se repite de forma iterativa para estimar puntos adicionales con menor relevancia pero con influencia en el proceso de clasificación. Estos puntos se unen para formar los mapas de saliencia que representan las principales regiones volumétricas relacionadas con la enfermedad, como se muestra en la figura 4.

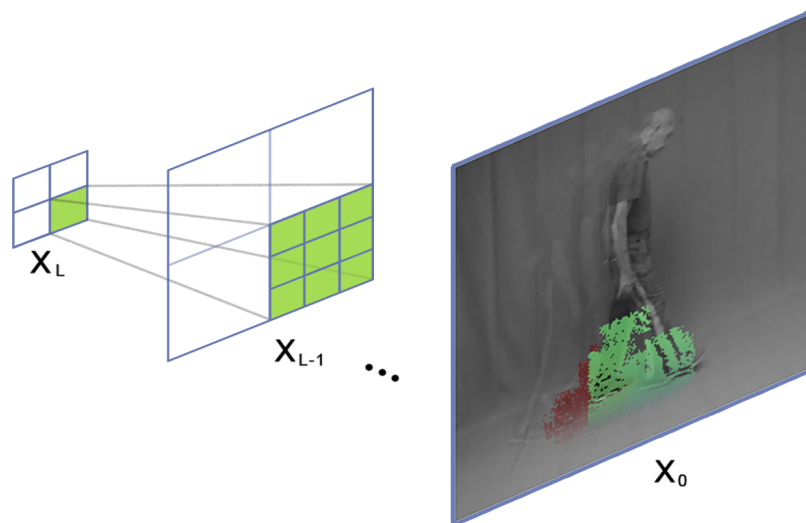


Figura 4. Pseudodeconvolución durante la retropropagación. La imagen ilustra cómo, a través del proceso de pseudodeconvolución, cada punto obtenido durante la etapa de Embedding back-tracking en la capa X_L representa una región completa en el cuerpo del paciente en la secuencia original X_0 .

6. MONTAJE EXPERIMENTAL

6.1. CONJUNTO DE DATOS PROPUESTO

El conjunto de datos consiste en un grupo de vídeos capturados con una cámara convencional en condiciones semicontroladas, de 22 participantes, siendo 11 sujetos de control y 11 pacientes diagnosticados con la enfermedad de Parkinson. Este conjunto de datos está balanceado en edad, es decir, la edad media es de 72.2 ± 4.0 para los sujetos de control y 72.6 ± 7.7 para los pacientes de EP.

Los pacientes con EP se encontraban en estadios de la enfermedad entre 2.5 y 3.0 según la escala de Hoehn y Yahr (estos valores fueron estimados por un fisioterapeuta). Cada sujeto del estudio fue grabado 8 veces mientras realizaba una caminata natural sin marcadores, 4 veces hacia la derecha y 4 veces hacia la izquierda, para un total de 176 secuencias de video. Todo el conjunto de vídeos se grabó en interiores, con una cámara estática, los pacientes vistiendo su propia ropa y frente a un fondo verde (para garantizar homogeneidad en los datos, teniendo en cuenta que los videos fueron obtenidos en 2 lugares diferentes). Las secuencias originales se grabaron a 60 fps y la duración media es de 4 segundos. Sin embargo, para los experimentos de la representación convolucional los vídeos se redujeron a una longitud de 60 fotogramas, se recortaron a 500x500 píxeles y se redimensionaron a 256x256 píxeles. Todas las secuencias de vídeo se grabaron en la fundación *FAM-PAS (Fundación del Adulto Mayor y Parkinson de Santander)* y en la residencia de ancianos *Asilo San Rafael*. Este estudio fue aprobado por el Comité de Ética de la Universidad Industrial de Santander y cuenta con un consentimiento informado por escrito.

Flujo Óptico: el análisis de patrones de movimiento como un campo denso de velo-

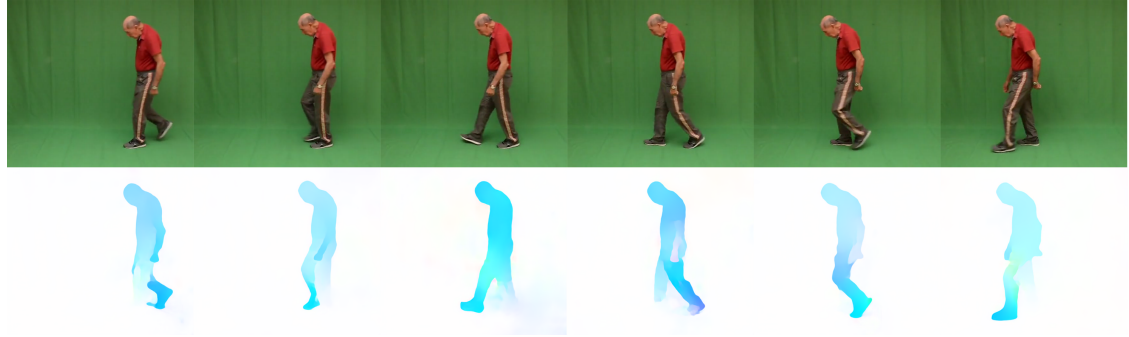


Figura 5. Primera fila: Imagen de la secuencia de vídeo original de un paciente con EP. Segunda fila: Flujo óptico calculado con FlowNet2.0.

cidad han sido una línea común en múltiples enfoques de reconocimiento de acciones. En este trabajo implementamos *FlowNet 2.0*, una aproximación de flujo óptico basada en CNNs con la que se ha conseguido una gran mejora de calidad respecto a los métodos tradicionales ⁴⁶. Este flujo óptico es una mejora de un enfoque anterior de CNN llamado *FlowNet* en el que los autores simplemente apilan dos imágenes de entrada secuencialmente adyacentes y las alimentan a través de la red. También presentaron otra variación en la que primero producen representaciones de las dos imágenes por separado, y luego las combinan juntas en una *capa de correlación*, y aprenden la representación superior en conjunto. Luego, *FlowNet 2.0* introdujo varias mejoras significativas, utilizando una arquitectura apilada e introduciendo una subred especializada dedicada a capturar y codificar pequeños movimientos. En concreto, el *FlowNet 2.0* implementa una arquitectura apilada en la que la primera red obtiene las imágenes I_1 y I_2 como entrada. Las redes siguientes obtienen I_1 , I_2 y el flujo estimado anterior $w_i = (u_i, v_i)^T$ donde i denota el índice de la red en la pila. Además, deforman la segunda imagen $I_2(x, y)$ mediante $\tilde{I}_{2,i}(x, y) = I_2(x + u_i, y + v_i)$

⁴⁶ ILG, Eddy, et al. "FlowNet 2.0: Evolution of optical flow estimation with deep networks". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, págs. 2462-2470.

e incorporan el error $e_i = \|\tilde{I}_{2,i} - I_1\|$. Como resultado, *Flownet 2.0* presentó una gran mejora tanto en calidad como en velocidad (véase un ejemplo en la figura 5). Para la representación convolucional, nuestra hipótesis es que este enfoque puede ser eficaz para recuperar los patrones de movimiento asociados a la enfermedad de Parkinson, capturando el movimiento global de las piernas pero también centrando la atención en las pequeñas variaciones probablemente asociadas al temblor.

6.2. CONFIGURACIÓN DEL MÉTODO DE TRAYECTORIAS

El enfoque propuesto se evaluó con trayectorias de movimiento de longitud fija de 16 fotogramas, calculadas en una cuadrícula espaciada por 5 píxeles y rastreadas en 8 escalas espaciales diferentes. Esta configuración se estableció como un compromiso entre la precisión y el coste computacional. Una vez obtenido el descriptor de movimiento final a partir de cualquiera de los tres esquemas de codificación, se implementaron dos clasificadores diferentes: El SVM y el RF. Se llevó a cabo una evaluación de los mejores parámetros para afinar el mejor rendimiento de la clasificación. Además, se comparó el rendimiento del método de agrupación de trayectorias en DensePose con un clustering clásico de k-means. Por último, se comprobó la relevancia del enfoque regional propuesto comparándolo con una clasificación global sobre todo el cuerpo sin agrupación.

Para la validación, se utilizó el método de clasificación *leave-one-patient-out* (validación cruzada k-fold), en el que en cada iteración, un paciente se deja fuera para la prueba y los restantes (21 sujetos en nuestro experimento particular) se utilizan para el entrenamiento. Los pacientes con EP clasificados correctamente se contaron como verdaderos positivos (TP) y los pacientes de control correctos como verdaderos negativos (TN). Los descriptores globales de cada región se evaluaron por separado y se identificaron de la siguiente manera $C1$ - cabeza y torso, $C2$ - parte superior de los brazos, $C3$ - parte inferior de los brazos y las manos, $C4$ - parte superior

de las piernas, y *C5* - parte inferior de las piernas y los pies, lo que corresponde a los principales movimientos rígidos de los segmentos del cuerpo desde un punto de vista sagital, como se representa en la Figura 6. El descriptor final, para todos los experimentos, se fijó en 200 elementos de longitud. Este valor se obtuvo a partir de un ajuste global previo.

6.2.1. Comparación entre *DensePose* y un algoritmo de clustering Para el primer enfoque propuesto, una de las principales contribuciones es la agrupación regional realizada por *DensePose* durante las secuencias de vídeo de locomoción. Para evaluar el beneficio real de la segmentación de *DensePose*, también utilizamos una estrategia de agrupación alternativa utilizando un algoritmo de clustering. Utilizamos el algoritmo *k*-means, que agrupa los objetos en *k* clusters siguiendo un criterio de trayectoria de vecindad en un espacio multidimensional euclidiano. A continuación, se definió un conjunto de cinco grupos, que corresponden aproximadamente a segmentos espaciales del cuerpo, como una división alternativa de la región. Para llevar a cabo la agrupación de *k*-means, todas las trayectorias se alinearon temporalmente y luego se calculó una matriz de similitud de afinidad entre los puntos correspondientes de cada trayectoria. La distancia de similitud aquí implementada para comparar las trayectorias se describe en ⁴⁷ y se expresa como:

$$d(a, b) = \left(\max_{n \in [1, L]} d_{spatial}[n] \right) \left(\frac{1}{L} \sum_{n=1}^L d_{velocity}[n] \right), \quad (1)$$

donde *L* es la longitud de las trayectorias,

$d_{spatial}[n] = ||\mathbf{x}_a[n] - \mathbf{x}_b[n]||$ es la distancia espacial entre los puntos de la trayectoria y $d_{velocity}[n] = ||\mathbf{v}_a[n] - \mathbf{v}_b[n]||$ es la distancia entre velocidades puntuales. Dicha

⁴⁷ RAPTIS, Michalis; KOKKINOS, Iasonas y SOATTO, Stefano. “Discovering discriminative action parts from mid-level video representations”. En: (*CVPR*), 201. IEEE. 2012, págs. 1242-1249.

distancia penaliza las trayectorias que están espacialmente alejadas y asocia las trayectorias con patrones de velocidad similares. Una matriz de afinidad $\mathbf{W}$ se obtiene luego por $\mathbf{W}(a, b) = \exp(-d(a, b)/1000)$ calculada en cada par de trayectorias. La matriz de afinidad tiene una dimensión de $T \times T$, donde T es el número de trayectorias en el vídeo. Para garantizar la compactación del grupo espacial, las trayectorias que no están espacialmente próximas se fuerzan a cero (utilizamos la condición $\max(d_{spatial}) \geq 120$ pixeles en nuestros experimentos).

6.3. CONFIGURACIÓN DE LA RED CONVOLUCIONAL 3D

La arquitectura propuesta sigue la configuración de capas que se ilustra en la Figura 3. En general, la estrategia tiene un total de cinco capas convolucionales 3D. La disposición de los kernels varía entre 32 y 128. Al final, dos vectores densos permiten una representación embebida de los vídeos. Durante el proceso de aprendizaje, se utilizó el descenso de gradiente estocástico para minimizar la función de pérdida. El mejor rendimiento se obtuvo con 10 y 20 *epochs* para las secuencias RGB y de flujo óptico, respectivamente, con un tamaño de *batch* de 2. Todos los experimentos se inicializaron con un *momentum* de 0.9, *dropout* de 0.5, *weight decay* de 5×10^{-3} , *learning rate* de 10^{-3} y *batch normalization*. Se logró una convergencia más rápida utilizando una reducción del *learning rate* y del *weight decay* (un factor de 10^{-1}) en los epochs cinco y nueve. La Figura 3 muestra una vista más detallada de la configuración del modelo.

La evaluación de este enfoque también se llevó a cabo del esquema de validación cruzada *leave-one-patient-out*. Para obtener los mapas de saliencia, el proceso de retropropagación se llevó a cabo tres veces para obtener una localización con tres grados de relevancia.

7. EVALUACIÓN Y RESULTADOS

7.1. REPRESENTACIÓN CINEMÁTICA

La Figura 6 ilustra los resultados observacionales logrados por el enfoque propuesto. En este caso, la segmentación del cuerpo (segunda fila) y las trayectorias de movimiento recuperadas (tercera fila) pueden utilizarse como herramienta de observación durante la evaluación clínica. En la cuarta fila se ilustra una visualización integrada, en la que la información cinemática (de las trayectorias locales) tiene información espacial sobre la localización del cuerpo. A partir de este último gráfico, los patrones de movimiento expresados a partir de las trayectorias en una parte específica del cuerpo podrían sugerir asociaciones con la enfermedad de Parkinson. Las tablas 1 y 2 presentan un estudio detallado de cada uno de los componentes envueltos en el enfoque propuesto, utilizando diferentes configuraciones, componentes de referencia y evaluando la capacidad de los descriptores de movimiento, codificados en cada segmento del cuerpo, para clasificar automáticamente los patrones de DP. Para cada vídeo, se extrajo un conjunto de trayectorias cinemáticas, y para cada trayectoria se calcularon histogramas radiales con un número de bins establecido en 4 (cada bin de 6 píxeles de ancho) y 18 bins angulares (9 para la velocidad) de 20 grados. También se evaluó el mejor descriptor de codificación regional, respecto a los tres esquemas definidos: la bolsa de palabras (BoW) con diccionarios globales (GD), BoW con diccionarios regionales (RD) y centroides de K-means (KC). Cada descriptor regional se evaluó con la segmentación de partes del cuerpo, obtenida a partir de dos estrategias diferentes 1) utilizando el enfoque de Densepose (DP) y 2) utilizando una agrupación espacial de trayectorias llamado k-means (KM). Además, para analizar la correlación regional de los patrones de movimiento con el DP, la evaluación se realizó de forma independiente para cada uno de los segmentos

Tabla 1. Resumen del rendimiento de los distintos métodos de clasificación para cada clúster mediante la descripción de la velocidad. Diccionario Global (GD), Diccionario Regional (RD), K-means Centroids (KC).

	C1		C2		C3		C4		C5	
	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM
DP-GD	94.73	92.05	94.73	92.61	98.86	94.32	99.62	94.89	99.62	97.73
DP-RD	98.83	93.18	98.69	90.91	98.86	88.64	98.86	84.66	99.62	97.73
DP-KC	77.22	72.02	75.02	72.32	68.82	70.20	71.20	66.26	67.29	73.22
KM-GD	96.56	96.59	97.73	96.59	98.86	94.32	97.73	96.59	98.86	90.34
KM-RD	93.13	79.66	93.07	91.02	92.90	89.32	93.45	89.32	93.40	87.05
KM-KC	72.02	70.52	71.20	66.82	70.80	67.29	72.54	66.26	72.48	66.62

Tabla 2. Evaluación cinemática por grupo utilizando el método de clasificación de mejor rendimiento (DP-RD con RF).

	C1	C2	C3	C4	C5
velocity	98.83	98.69	98.86	98.86	99.62
acceleration	90.31	88.28	91.45	88.28	89.39
curvature	99.62	97.80	96.74	93.89	94.30

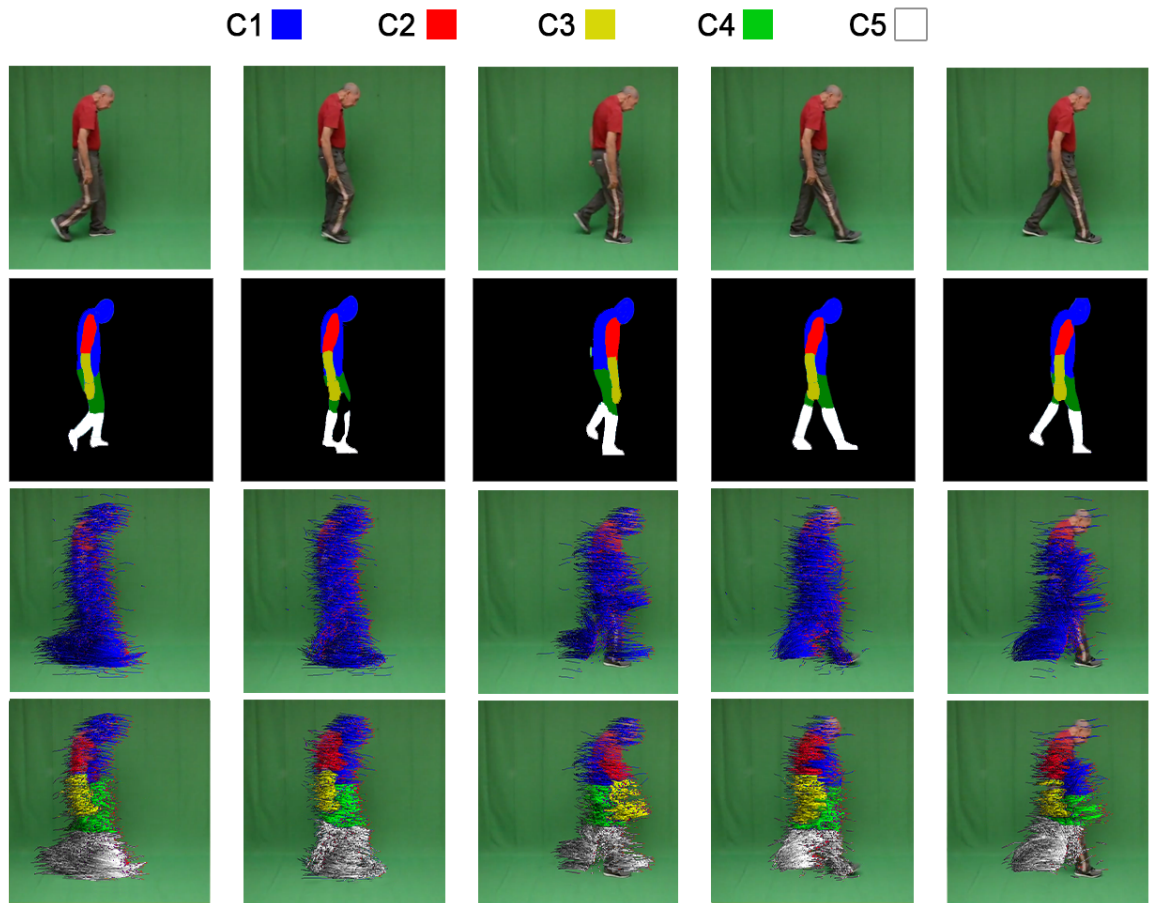


Figura 6. Primera fila: Secuencia de vídeo original. Segunda fila: Máscara DensePose, donde C1 corresponde a la cabeza y el tronco; C2, a los brazos; C3, a los antebrazos y a las manos; C4 a los muslos, y C5 a las piernas y los pies. Tercera fila: Trayectorias densas de movimiento. Cuarta fila: Trayectorias de movimiento agrupadas.

considerados.

Cada descriptor de vídeo se clasificó utilizando Random Forest (RF) o Support Vector Machine (SVM). Como se muestra en la Tabla 1 el enfoque propuesto tiene un rendimiento estable con respecto a las diferentes configuraciones, y utilizando diferentes clasificadores. Cabe destacar que la mejor agrupación espacial se consigue a partir de *DensePose* (DP), utilizando un diccionario regional (RD) y un RF. El mejor rendimiento de Random Forest puede explicarse por una construcción más fina de

los límites del modelo basada en el conjunto de árboles discretos. Como se esperaba, las regiones de las extremidades inferiores proporcionan los mejores resultados, con un mayor rendimiento discriminativo para clasificar los patrones de DP, alcanzando una precisión media del 99,62 %. (ver la Tabla 1).

Resulta interesante que el enfoque propuesto explota la descripción cinemática de varias regiones del cuerpo, que podría ser determinante en el diagnóstico de la EP. Por ejemplo, la región de la cabeza alcanza por sí misma una precisión media del 94,73 %, utilizando la mejor configuración del enfoque. La región C3 (que incluye la parte inferior de los brazos y las manos) también alcanza una notable precisión del 98,86 %. Estas regiones pueden integrarse en un análisis clínico para comprender mejor una determinada etapa y el progreso de la enfermedad, hecho que podría ser clave para entender las primeras etapas de la enfermedad.

En la Tabla 2 se lleva a cabo un segundo experimento, a partir del cual se analiza la mejor configuración de clasificación respecto a las tres cinemáticas diferentes calculadas para cada cluster. El objetivo es determinar los patrones cinemáticos más relevantes en relación con la enfermedad. Los patrones de velocidad parecen tener mayor capacidad de descripción en los miembros inferiores, lo que de hecho está relacionado con un patrón de EP bien conocido. De hecho, el uso de sólo los patrones de velocidad obtiene una puntuación casi perfecta en la clasificación. En cuanto a los patrones de cabeza, la curvatura codificada a partir de las trayectorias podría estar fuertemente correlacionada con la EP. Esto puede estar relacionado con la rigidez de la cabeza y parte del tronco. Por último, la velocidad y la curvatura alcanzan el mejor rendimiento de clasificación con una precisión media del 99,62 % para C5 y C1, respectivamente. También cabe destacar la región C3, relacionada con las manos, que es la más discriminante cuando el descriptor se forma a partir de la aceleración únicamente. Esto podría estar relacionado con el temblor de la mano y sugiere combinar patrones complementarios para diferenciar mejor entre

los pacientes de EP y los de control.

La Figura 7 muestra un análisis más detallado del ajuste de los parámetros del histograma polar. Para este propósito se eligió el cluster C1, debido a la captura de dos de los biomarcadores más relevantes para la EP, la rigidez del movimiento de la cabeza y la estabilidad del tronco. Confirmando experimentos anteriores, todas las cinemáticas obtienen un rendimiento razonable, y la curvatura reporta la mayor precisión. La selección de los parámetros parece depender de cada cinemática. Por un lado, en general, tanto la curvatura como la velocidad presentan buenos resultados, por encima del 95 %. La aceleración, por su parte, varía en torno al 90 % de precisión y alcanza su valor máximo con 8 bins radiales y un ángulo de 15 grados, logrando una puntuación del 93 %.

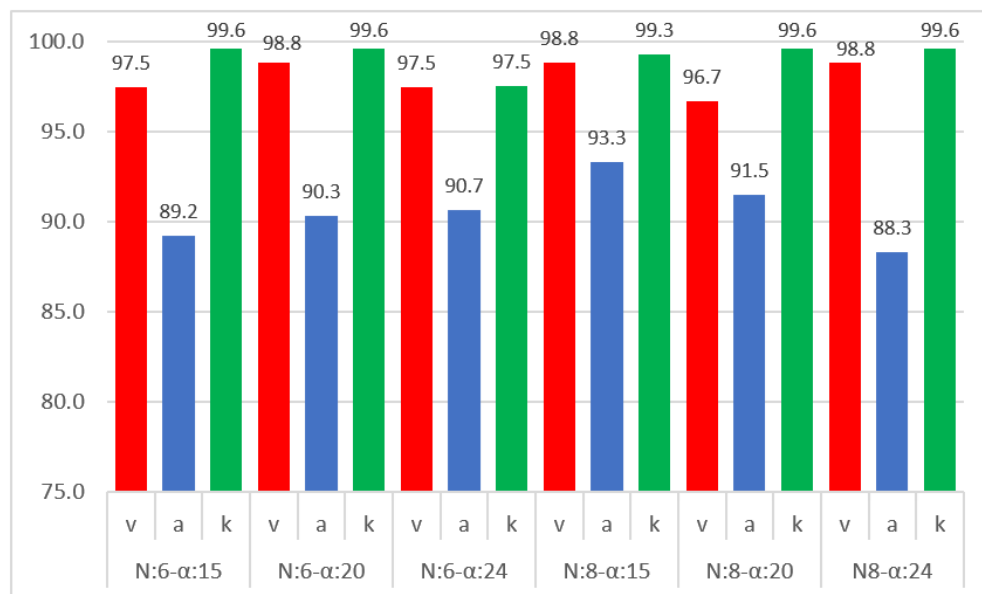


Figura 7. Evaluación de la cinemática. Precisión para el cluster C1 (cabeza y torso) para diferentes valores de los parámetros donde N representa el número de bins radiales y α representa el ángulo de cada bin angular.

En un último experimento, la representación del movimiento a partir de trayectorias densas se obtuvo sin tener en cuenta los segmentos corporales. El propósito es

tanto evaluar el interés de la segmentación de la parte del cuerpo como comprobar la capacidad de un patrón global para representar la EP. La Figura 8 resume los resultados obtenidos, para diferentes cinemáticas, y utilizando diferentes configuraciones del histograma polar. Resulta que la representación global del movimiento consigue un resultado notable, donde la velocidad en general tiene el mejor y más estable rendimiento. Curiosamente, el análisis del descriptor de movimiento sobre regiones corporales específicas (experimentos anteriores) da como resultado una mejor representación, incluso considerando que se cuenta con menos trayectorias para realizar el análisis.

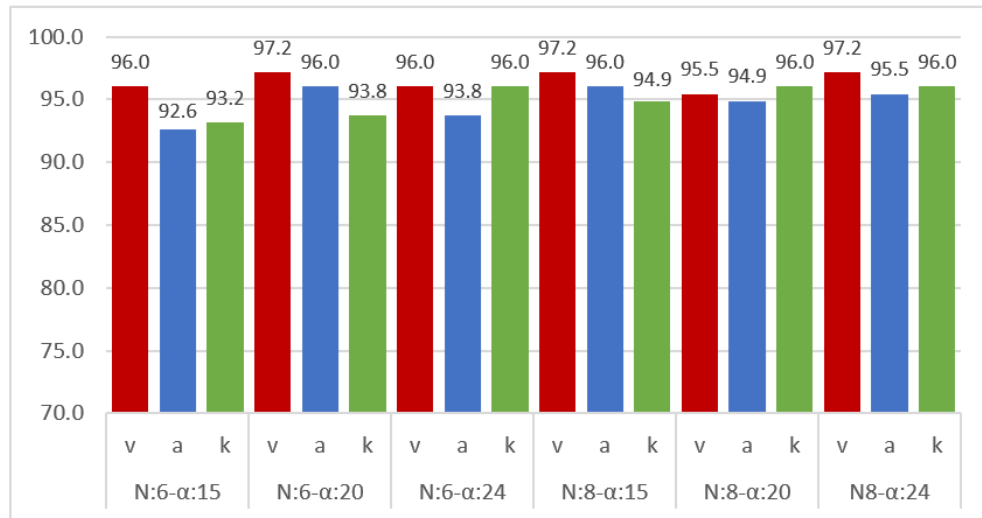


Figura 8. Descriptor de movimiento global. Precisión para todo el conjunto de trayectorias sin segmentar el cuerpo en regiones, y con variación de los parámetros del histograma polar donde N representa el número de bins radiales y α representa el ángulo de cada bin angular.

7.2. REPRESENTACIÓN CONVOLUCIONAL

En cuanto al segundo enfoque propuesto, a continuación nos permitimos presentar la validación realizada. Una primera evaluación de esta estrategia se dedicó a medir la capacidad de reconocer a los pacientes con EP con respecto a la población

control. La Tabla 3 muestra los resultados cuantitativos de la clasificación de la EP frente al control a partir de las secuencias RGB en bruto y el flujo óptico. Podemos observar que las secuencias RGB superan sistemáticamente al flujo óptico alcanzando una puntuación de clasificación del 94.89 % sobre el 84.66 % del flujo óptico. Curiosamente, para el flujo óptico la precisión (0.87) fue mayor que la sensibilidad (0.78), en contraste con las secuencias RGB con una precisión y una sensibilidad de 0.91 y 1.0, respectivamente. Sin embargo, cabe señalar que en nuestro caso un falso positivo no tendría tantas repercusiones desfavorables como un falso negativo, en el que, al menos, se retrasaría el diagnóstico y el tratamiento del paciente. Así pues, la sensibilidad es una medida más significativa para evaluar el método propuesto.

Tabla 3. Resumen del resultado de la clasificación cuantitativa para las secuencias RGB en bruto y de flujo óptico (OF).

	Exactitud (%)	Precisión	Sensibilidad	F2-score
RGB	94.89	0.91	1.00	0.78
OF	84.66	0.87	0.78	0.70

Teniendo en cuenta este hecho, se realizó un análisis más detallado considerando la métrica F2-score. Esta medida pondera más la sensibilidad que la precisión, y puede ser expresada como, $F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{sensibilidad}}{(\beta^2 \cdot \text{precision}) + \text{sensibilidad}}$, con $\beta = 2$, obteniendo 0,70 y 0,78 para las secuencias de flujo óptico y RGB, respectivamente.

Además, se cuantificó la curva ROC, como se ilustra en la Figura 9 mostrando un buen equilibrio entre la sensibilidad y la especificidad para el modelo propuesto. En este caso, el enfoque propuesto alcanza un área bajo la curva (AUC) de 0.95 para las secuencias RGB y de 0.91 para el flujo óptico. En la Figura 10 se ilustran las correspondientes matrices de confusión. Teniendo en cuenta que hay 8 vídeos de cada paciente, es importante destacar que el modelo para las imágenes RGB clasificó completamente mal a un solo paciente (señalado en rojo en la Figura 11). Este

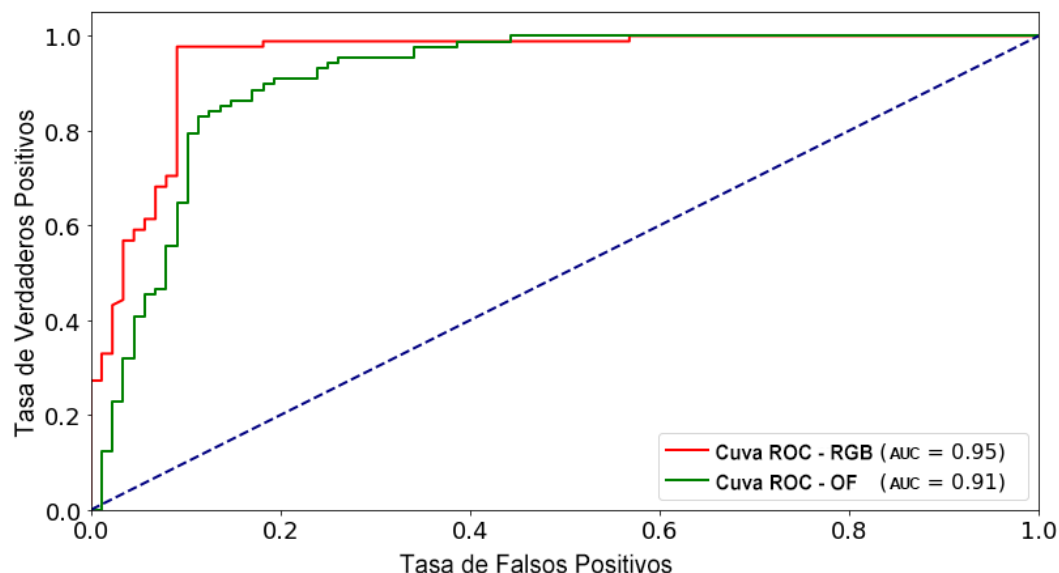


Figura 9. Curva ROC para secuencias RGB y flujo óptico (OF). Los vídeos en bruto superan al flujo óptico con un valor de AUC de 0.95 frente a un 0.91.

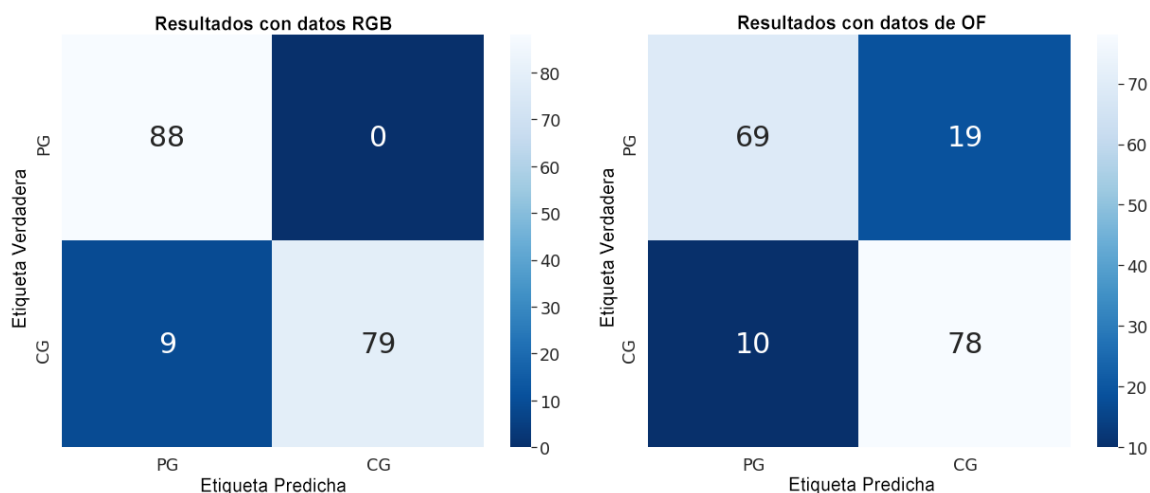


Figura 10. Matrices de confusión para datos RGB (izquierda) y de flujo óptico (derecha).

sujeto corresponde a una paciente de la clase control que muestra una reducción en algunas características de la marcha, como la velocidad, la cadencia, la zancada y el ángulo del pie, excepto la base de sustentación que parece ser mayor. Estos patrones de marcha podrían estar relacionados con otras morbilidades asociadas

de la paciente.

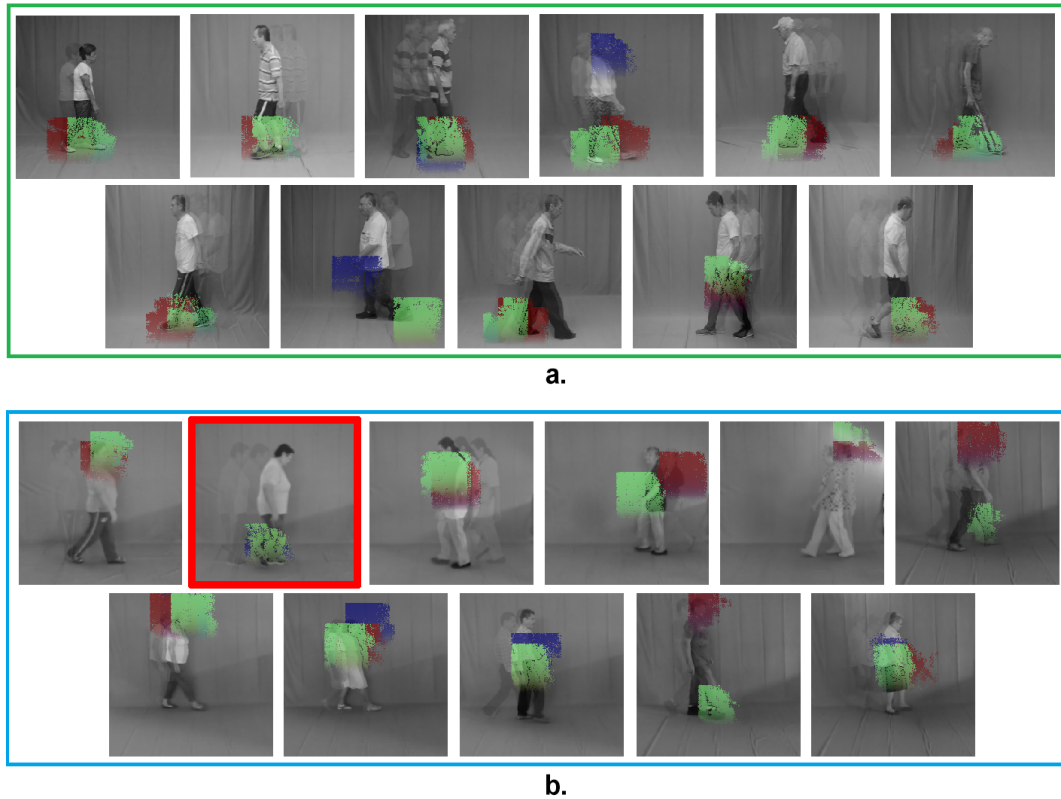


Figura 11. Mapas de saliencia para pacientes con la enfermedad de Parkinson (a) y sujetos control (b).

7.2.1. Mapas de Saliencia Relacionados con el Párkinson Una segunda evaluación del enfoque propuesto para caracterizar a los pacientes de EP estaba relacionada con la capacidad de detectar regiones destacadas asociadas a los patrones de la enfermedad. En este caso, seguimos un principio de explicabilidad del modelo, que trata de retropropagar una determinada decisión sobre las capas anteriores para detectar las regiones principales asociadas a dicho resultado. La Figura 11 muestra un ejemplo por paciente de los vídeos de atención explicables obtenidos de pacientes con EP y sujetos de control. Los colores rojo, verde y azul representan el

grado de importancia de cada región (de mayor a menor). El método es capaz de localizar y resaltar patrones espacio-temporales en el cuerpo del paciente con precisión. Curiosamente, en la mayoría de las secuencias de la marcha parkinsoniana, el método propuesto destaca la parte inferior del cuerpo, mientras que en las marchas de control el método destaca la parte superior del cuerpo. Esto sugiere que el modelo encuentra características espacio-temporales anormales en las piernas de los pacientes con EP durante la fase de apoyo simple (posiblemente relacionadas con la longitud de la zancada y la bradicinesia) e identifica características saludables relacionadas con el movimiento de la cabeza y la postura recta de los sujetos de control. Estos patrones constituyen una información relevante para apoyar potencialmente las observaciones de los médicos.

En el estado del arte, se han propuesto varios métodos para el cálculo de mapas de saliencia en redes neuronales convolucionales, donde el mapa de calor es una combinación ponderada de mapas de características. Por ejemplo, *GradCam*⁴⁸ calcula el gradiente de la puntuación para una clase dada con respecto a los mapas de características de una capa convolucional. Estos gradientes retropropagados se agrupan en un promedio global para obtener los pesos de importancia. Sin embargo, estos métodos requieren que no haya una reducción significativa de la dimensionalidad en las capas convolucionales. Además, estos métodos procesan una imagen a la vez y, hasta donde sabemos, no existe ningún método para el cálculo directo de mapas de saliencia en volúmenes espacio-temporales. Por último, como se muestra en la Figura 12, los mapas obtenidos por nuestro método son convincentes en cuanto a la asociación con los patrones de movimiento de la enfermedad. Por el contrario, los mapas resultantes de GradCam se centran principalmente en el fondo y normalizan las probabilidades de las regiones a lo largo de la escena. En este sentido,

⁴⁸ RS, Ramprasaath, *et al.* "Grad-CAM: visual explanations from deep networks via gradient-based localization". En: *Int J Comput Vis* 128 (2019).

esta línea base carece de la capacidad para explicar visualmente una predicción. Los mapas conseguidos con el enfoque propuesto siguen siendo coherentes en el tiempo y pueden ser una herramienta visual útil para el apoyo al diagnóstico médico.

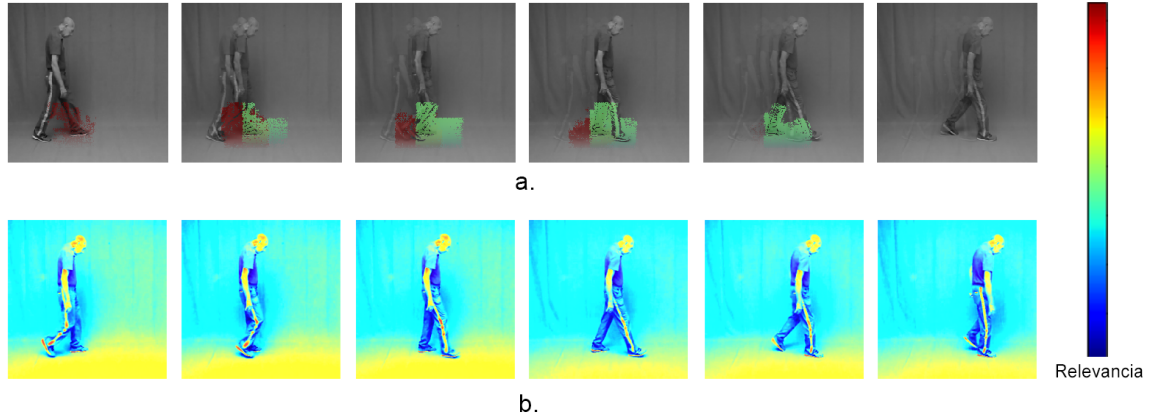


Figura 12. Comparación entre el método propuesto (a) y los mapas de saliencia de Grad-Cam (b).

7.2.2. Factores de Confusión El conjunto de datos consta de 12 hombres y 10 mujeres distribuidos de la siguiente manera: 2 mujeres y 9 hombres en el grupo de EP y 8 mujeres y 3 hombres en el grupo de control. Por lo tanto se llevo a cabo un análisis de factores de confusión con respecto a los resultados obtenidos. El sexo y la edad se evaluaron como posibles variables de confusión mediante el cálculo de las proporción de probabilidades crudas y ajustadas de un proceso de regresión logística. En este modelo, las variables predictoras x_i se relacionan linealmente con las probabilidades logarítmicas del resultado y . Para 3 predictores, esto se expresa formalmente de la siguiente forma:

$$\ln \left(\frac{P(y = 1)}{1 - P(y = 1)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

donde y es la probabilidad predicha de EP, x_1 es la etiqueta binaria verdadera, x_2 es el sexo y x_3 es la edad de cada participante. Para la proporción de probabilidad cruda (cOR), el coeficiente de regresión β_1 se calcula sin considerar el efecto de los otros coeficientes, es decir, $\beta_2 = \beta_3 = 0$. Para la proporción de probabilidad ajustada (aOR), β_1 se calcula ahora junto con los otros coeficientes. La estimación de la proporción de probabilidad se define entonces como la función exponencial e^{β_1} ⁴⁹. Las cOR y aOR obtenidas fueron 9.38 y 8.13, respectivamente ($p < 0,05$). La proporción de probabilidades refleja una asociación significativamente positiva entre el diagnóstico de la enfermedad y las probabilidades predichas por el método propuesto. Dicha asociación no está confundida por el sexo o la edad, lo que indica que estas variables no influyen en los resultados de la evaluación.

⁴⁹ SZUMILAS, Magdalena. "Explaining odds ratios". En: *Journal of the Canadian academy of child and adolescent psychiatry* 19.3 (2010), pág. 227.

8. DISCUSIÓN

En este trabajo se presentaron dos aproximaciones computacionales, basadas en aprendizaje de máquina, para la representación de patrones parkinsonianos en video. Los dos enfoques logran efectivamente caracterizar una población con la EP con respecto a una población con similar demografía pero sin un diagnóstico de la enfermedad.

En una primera aproximación se logró hacer un análisis cinemático detallado de cada región corporal, durante la marcha. Este trabajo introduce un novedoso enfoque sin marcadores que explota trayectorias densas, agrupadas y analizadas en cada segmento corporal, que permiten caracterizar con precisión las alteraciones de movimiento relacionadas con la enfermedad de Parkinson. Este modelo es capaz de obtener una representación de movimiento multinivel, partiendo de una descripción de cada trayectoria por su distribución de cinemática local, calculando histogramas polares que codifican características de velocidad, aceleración y curvatura. En un segundo nivel de representación del movimiento, utilizando la segmentación del cuerpo proporcionada por *DensePose*, el método utiliza la agrupación de trayectorias para obtener patrones de locomoción regionales. Finalmente, esos patrones de movimiento se utilizan en un marco de aprendizaje supervisado para predecir y clasificar automáticamente condiciones parkinsonianas. El enfoque propuesto, en un conjunto de datos con 22 pacientes, demostró un alto poder de discriminación con respecto a la clasificación de la enfermedad. De hecho, durante la evaluación, y utilizando únicamente los segmentos de la parte inferior del cuerpo, el enfoque propuesto alcanza una puntuación media de clasificación del 99,62 %. Este resultado se obtuvo utilizando únicamente la velocidad como cinemática básica. Curiosamente, el enfoque propuesto captura adecuadamente los patrones de movimiento de la cabeza y el tronco, logrando para ambas regiones un 94,73 % de precisión en la

clasificación. Este hecho resulta muy relevante, ya que trabajos recientes han determinado los patrones de cabeza y tronco como principales biomarcadores de la enfermedad. De hecho, el método propuesto explota el aumento del temblor y la inestabilidad postural durante la locomoción, lo que podría ser clave para modelar los patrones en las primeras etapas de la enfermedad.

En la rutina clínica, los patrones parkinsonianos (y de alteración del movimiento en general) se capturan cuantitativamente a partir del análisis clásico de la marcha, que incluye la aproximación de la dinámica corporal a partir de un conjunto de marcadores. Esta metodología de marcadores tiene protocolos estándar que permiten una fácil interpretación del análisis de un paciente en particular. Sin embargo, este enfoque es invasivo, altera el gesto natural y su análisis dinámico general puede limitar la descripción de la EP. De hecho, varios patrones de movimiento como la inestabilidad del tronco, así como los movimientos del temblor de la mano y de la cabeza, podrían no ser detectados o asociados a patrones de control. Por lo tanto, se necesitan enfoques novedosos que capturen el movimiento natural y recuperen una descripción dinámica mayor, para caracterizar mejor la EP. Por ejemplo, Verlekar *et al.*⁵⁰ lograron un 98,8 % al clasificar la marcha control y tres deficiencias emuladas en nueve personas masculinas y una femenina. Lograron una diferenciación de los pies derecho e izquierdo utilizando un método basado en la silueta. Por otra parte, nuestro método muestra una complementariedad cinemática en la descripción regional del movimiento. La aceleración, por ejemplo, resulta especialmente útil para caracterizar los patrones de movimiento de la parte superior del cuerpo. Esto ya

⁵⁰ VERLEKAR, Tanmay T.; SOARES, Luís Ducla y CORREIA, Paulo Lobato. "Automatic Classification of Gait Impairments Using a Markerless 2D Video-Based System". En: *Sensors*. 2018.

se ha sugerido en otros trabajos como Latt *et. al.*⁵¹. Ellos modelan la aceleración utilizando el cuadrado medio de las señales del torso de los pacientes con EP con historial de caídas y de los sujetos de control sin historial de caídas, y concluyen que los patrones de aceleración de la parte superior del cuerpo durante la marcha difieren entre las personas mayores con y sin EP. Sin embargo, este enfoque realizó un análisis visual de las medidas y no evaluó su utilidad en la predicción de la enfermedad.

En un segundo enfoque logrado en este trabajo se utilizó un enfoque de aprendizaje profundo para englobar el modelamiento de la marcha y cuantificar tanto las características volumétricas de movimiento como obtener una predicción con respecto a la condición del paciente registrado. Este trabajo introduce una novedosa representación de la marcha sin marcadores que explota la capacidad de las redes neuronales convolucionales 3D para aprender características espacio-temporales que permiten visualizar biomarcadores clave de la marcha parkinsoniana. Este modelo implementa convoluciones temporalmente largas que hacen posible el uso de secuencias de vídeo lo suficientemente extensas como para capturar al menos un ciclo completo de la marcha. Esto es relevante porque los patrones anormales pueden ocurrir principalmente en etapas específicas del ciclo en cada paciente. El modelo entrenado es capaz de predecir y clasificar los patrones anormales logrando una diferenciación precisa de la marcha control.

Además, la principal aportación de este trabajo es la obtención de mapas salientes a partir de un proceso de retropropagación a través del modelo previamente entrenado. Este proceso consta de dos pasos principales: Primero, el rastreo de los valores máximos de activación en las capas densas anteriores. Segundo, un proceso de

⁵¹ LATT, Mark D, *et al.* "Acceleration patterns of the head and pelvis during gait in older people with Parkinson's disease: a comparison of fallers and nonfallers". En: *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 64.6 (2009), págs. 700-706.

pseudo-deconvolución que finalmente consigue una proyección sobre el vídeo de las características más importantes aprendidas durante el entrenamiento para la clasificación. Los vídeos de atención recuperados muestran patrones espacio-temporales anormales en las extremidades inferiores de los pacientes con EP durante la fase de apoyo simple. En este sentido, las características anormales más discriminativas podrían estar relacionadas con la longitud de la zancada, la cadencia, la distancia del pie con respecto al suelo o la bradicinesia. En el mismo conjunto de datos con 22 pacientes, el enfoque propuesto logró un alto poder de discriminación con respecto a la clasificación de la enfermedad de Parkinson, con una puntuación media de clasificación del 94.89 % para las secuencias RGB en bruto y del 84.66 % para el flujo óptico.

Los dos métodos aquí presentados logran clasificar secuencias de video sin marcadores y representar las principales regiones espacio-temporales asociadas a patrones anormales, relacionados con la EP, durante la marcha. En términos cuantitativos, la representación cinemática supera a la convolucional con una precisión en la clasificación de 99.62 % contra un 94.89 %. Esto puede deberse, por una parte, a la gran cantidad de información temporal y cinemática que contienen las trayectorias densas y por otra parte a la cantidad de datos, de lo cual depende en gran medida el desempeño de las redes neuronales convolucionales. Los dos métodos logran identificar patrones de movimiento anormales a nivel regional. Mientras que la representación cinemática lo hace por medio de una segmentación controlada, permitiendo la caracterización cinemática de cada segmento corporal, la representación convolucional lo hace de manera automática, encontrando regiones espacio-temporales que pueden estar asociadas directamente a síntomas parkinsonianos ya conocidos, o bien, a patrones ocultos aún inexplorados.

9. CONCLUSIONES Y TRABAJO FUTURO

En la rutina clínica, la EP se captura cuantitativamente a partir del análisis clásico de la marcha, que incluye la aproximación de la dinámica corporal a partir de un conjunto de marcadores. Sin embargo, estos métodos basados en marcadores requieren protocolos estrictos que afectan la naturalidad de los movimientos del paciente y son insuficientes como herramientas de diagnóstico de la enfermedad de Parkinson. Esto, junto con la falta de biomarcadores que apoyen plenamente el análisis médico, hace que el diagnóstico de la EP sea un proceso altamente subjetivo. En este trabajo se presentaron dos enfoques computacionales que permiten cuantificar patrones locomotores y soportar tareas de diagnóstico y seguimiento de pacientes. En un primer enfoque, el análisis cinemático por segmentos coportales resultó ser eficaz para determinar patrones anormales asociados con la condición del Parkinson. De hecho, también se evidenció una alta correlación con la enfermedad usando segmentos corporales independientes, permitiendo soportar análisis clínicos de unilateralidad de la enfermedad. En tal sentido, esta estrategia puede ser introducida en el protocolo clínico para complementar y apoyar las evaluaciones observacionales. Cabe destacar que el enfoque propuesto sólo requiere 200 de valores escalares para representar un descriptor global de la marcha, lo cual es una muestra de eficiencia computacional.

Por otra parte, en una segunda propuesta utilizando aprendizaje convolucional profundo sobre secuencias enteras de video, se logró clasificar patrones parkinsonianos, pero además se aprendieron características en el video que pueden aportar en el análisis personalizado de cada paciente. El enfoque propuesto demostró ser eficaz para recuperar los mapas de saliencia como indicadores de patrones anormales, lo que podría apoyar las observaciones de los médicos y eventualmente ser un biomarcador decisivo de la EP. Aunque existen otros métodos para calcular los

mapas de saliencia en imágenes, nuestro método permite calcular los mapas de saliencia directamente en vídeos, aprovechando los patrones espacio-temporales aprendidos por la red convolucional 3D. Además, nuestro método parece ser más intuitivo y preciso temporalmente. En este sentido, esta estrategia puede introducirse en los protocolos de diagnóstico para complementar y apoyar las evaluaciones clásicas. Los pacientes con EP que participaron en este estudio fueron caracterizados en la segunda y tercera etapas de la enfermedad. En futuros trabajos se probará este descriptor en estadios tempranos de la enfermedad y se realizará un estudio longitudinal para caracterizar la evolución de la enfermedad.

BIBLIOGRAFÍA

ALP GÜLER, Rıza; NEVEROVA, Natalia y KOKKINOS, Iasonas. "Densepose: Dense human pose estimation in the wild". En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, págs. 7297-7306 (vid. págs. 27, 28).

BARBE, Michael T, *et al.* "Gait and upper limb variability in Parkinson's disease patients with and without freezing of gait". En: *Journal of neurology* 261.2 (2014), págs. 330-342 (vid. págs. 13, 27).

BREIMAN, Leo. "Random forests". En: *Machine learning* 45.1 (2001), págs. 5-32 (vid. pág. 34).

BUCKLEY, Christopher, *et al.* "Attenuation of upper body accelerations during gait: piloting an innovative assessment tool for Parkinson's disease". En: *BioMed research international* 2015 (2015) (vid. pág. 19).

— "Quantification of upper body movements during gait in older adults and in those with Parkinson's disease: Impact of acceleration realignment methodologies". En: *Gait & posture* 52 (2017), págs. 265-271 (vid. pág. 13).

— "Upper body accelerations as a biomarker of gait impairment in the early stages of Parkinson's disease". En: *Gait & posture* 71 (2019), págs. 289-295 (vid. pág. 19).

CHANG, Chih-Chung y LIN, Chih-Jen. "LIBSVM: a library for support vector machines". En: *ACM transactions on intelligent systems and technology (TIST)* 2.3 (2011), pág. 27 (vid. pág. 34).

CHO, Chien-Wen, *et al.* "A vision-based analysis system for gait recognition in patients with Parkinson's disease". En: *Expert Systems with applications* 36.3 (2009), págs. 7033-7039 (vid. pág. 18).

CHOE, Bo-Young, *et al.* "Neuronal laterality in Parkinson's disease with unilateral symptom by in vivo ¹H magnetic resonance spectroscopy". En: *Investigative radiology* 33.8 (1998), págs. 450-455 (vid. pág. 26).

DAS, Deepjoy. "Human gait classification using combined HMM & SVM hybrid classifier". En: *2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV)*. IEEE. 2015, págs. 169-174 (vid. pág. 17).

DAVIS III, Roy B, *et al.* "A gait analysis data collection and reduction technique". En: *Human movement science* 10.5 (1991), págs. 575-587 (vid. pág. 11).

DONAHUE, Jeffrey, *et al.* "Long-term recurrent convolutional networks for visual recognition and description". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, págs. 2625-2634 (vid. pág. 20).

FARNEBÄCK, Gunnar. "Two-frame motion estimation based on polynomial expansion". En: *Scandinavian conference on Image analysis*. Springer. 2003, págs. 363-370 (vid. pág. 30).

GOETZ, Christopher G, *et al.* "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results". En: *Movement disorders: official journal of the Movement Disorder Society* 23.15 (2008), págs. 2129-2170 (vid. pág. 11).

GUNDUZ, Hakan. "Deep learning-based Parkinson's disease classification using vocal feature sets". En: *IEEE Access* 7 (2019), págs. 115540-115551 (vid. pág. 21).

HE, Kaiming, *et al.* "Mask r-cnn". En: *Proceedings of the IEEE international conference on computer vision*. 2017, págs. 2961-2969 (vid. pág. 28).

ILG, Eddy, *et al.* "Flownet 2.0: Evolution of optical flow estimation with deep networks". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, págs. 2462-2470 (vid. pág. 44).

KAVANAGH, Justin; BARRETT, Rod y MORRISON, Steven. "The role of the neck and trunk in facilitating head stability during walking". En: *Experimental brain research* 172.4 (2006), pág. 454 (vid. pág. 27).

KHAN, Taha; WESTIN, Jerker y DOUGHERTY, Mark. "Motion cue analysis for parkinsonian gait recognition". En: *The open biomedical engineering journal* 7 (2013), pág. 1 (vid. pág. 17).

KOUR, Navleen; ARORA, Sakshi, *et al.* "Computer-Vision Based Diagnosis of Parkinson's Disease via Gait: A Survey". En: *IEEE Access* 7 (2019), págs. 156620-156645 (vid. pág. 16).

LATT, Mark D, *et al.* "Acceleration patterns of the head and pelvis during gait in older people with Parkinson's disease: a comparison of fallers and nonfallers". En: *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 64.6 (2009), págs. 700-706 (vid. pág. 62).

LI, Qiannan, *et al.* "Classification of gait anomalies from kinect". En: *The Visual Computer* 34.2 (2018), págs. 229-241 (vid. pág. 17).

MARSDEN, CD. "Slowness of movement in Parkinson's disease". En: *Movement Disorders: Official Journal of the Movement Disorder Society* 4.S1 (1989), S26-S37 (vid. pág. 29).

MARTINEZ-MURCIA, Francisco Jesús, *et al.* “A 3D convolutional neural network approach for the diagnosis of Parkinson’s disease”. En: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2017, págs. 324-333 (vid. pág. 21).

MURRAY, M Pat; DROUGHT, A Bernard y KORY, Ross C. “Walking patterns of normal men”. En: *JBJS* 46.2 (1964), págs. 335-360 (vid. pág. 37).

NAGHAVI, Nader y WADE, Eric. “Prediction of freezing of gait in Parkinson’s disease using statistical inference and lower-limb acceleration data”. En: *IEEE transactions on neural systems and rehabilitation engineering* 27.5 (2019), págs. 947-955 (vid. pág. 12).

OH, Shu Lih, *et al.* “A deep learning approach for Parkinson’s disease diagnosis from EEG signals”. En: *Neural Computing and Applications* (2018), págs. 1-7 (vid. pág. 21).

PISTACCHI, Michele, *et al.* “Gait analysis and clinical correlations in early Parkinson’s disease”. En: *Functional neurology* 32.1 (2017), pág. 28 (vid. pág. 12).

POST, Bart, *et al.* “Unified Parkinson’s disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?” En: *Movement disorders: official journal of the Movement Disorder Society* 20.12 (2005), págs. 1577-1584 (vid. pág. 12).

RAPTIS, Michalis; KOKKINOS, Iasonas y SOATTO, Stefano. “Discovering discriminative action parts from mid-level video representations”. En: *(CVPR), 2012*. IEEE. 2012, págs. 1242-1249 (vid. pág. 46).

REN, Peng, *et al.* "Multivariate Analysis of Joint Motion Data by Kinect: Application to Parkinson's Disease". En: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.1 (2019), págs. 181-190 (vid. pág. 13).

REN, Shaoqing, *et al.* "Faster r-cnn: Towards real-time object detection with region proposal networks". En: *Advances in neural information processing systems*. 2015, págs. 91-99 (vid. pág. 27).

RS, Ramprasaath, *et al.* "Grad-CAM: visual explanations from deep networks via gradient-based localization". En: *Int J Comput Vis* 128 (2019) (vid. pág. 57).

SEMWAL, Vijay Bhaskar, *et al.* "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification". En: *Multimedia tools and applications* 76.22 (2017), págs. 24457-24475 (vid. pág. 19).

SEVILLA-LARA, Laura, *et al.* "On the integration of optical flow and action recognition". En: *German Conference on Pattern Recognition*. Springer. 2018, págs. 281-297 (vid. pág. 29).

SHAW, Laxmi. "HMM based Parkinson's detection by analysing symbolic postural gait image sequences". En: *Int. J. Tech. Res. Appl* 2.4 (2014), págs. 211-216 (vid. pág. 18).

SIMONYAN, Karen y ZISSERMAN, Andrew. "Two-stream convolutional networks for action recognition in videos". En: *Advances in neural information processing systems*. 2014, págs. 568-576 (vid. pág. 20).

SPLETE, HEIDI. "Avoid Diagnostic Pitfalls For Parkinson's Disease". En: *Internal Medicine News* 39.17 (2006), págs. 31-31 (vid. pág. 12).

SUGANDHI, K; WAHID, Farha Fatina y RAJU, G. "Feature Extraction Methods for Human Gait Recognition—A Survey". En: *International Conference on Advances in Computing and Data Sciences*. Springer. 2016, págs. 377-385 (vid. pág. 14).

SUN, Deqing, *et al.* "Learning optical flow". En: *European Conference on Computer Vision*. Springer. 2008, págs. 83-97 (vid. pág. 29).

SUN, Shuyang, *et al.* "Optical flow guided feature: A fast and robust motion representation for video action recognition". En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 1390-1399 (vid. pág. 29).

SUTHERLAND, David H. "The evolution of clinical gait analysis: Part II Kinematics". En: *Gait & posture* 16.2 (2002), págs. 159-179 (vid. pág. 16).

SZUMILAS, Magdalena. "Explaining odds ratios". En: *Journal of the Canadian academy of child and adolescent psychiatry* 19.3 (2010), pág. 227 (vid. pág. 59).

TAN, Dawn, *et al.* "Automated analysis of gait and modified timed up and go using the Microsoft Kinect in people with Parkinson's disease: associations with physical outcome measures". En: *Medical & biological engineering & computing* 57.2 (2019), págs. 369-377 (vid. pág. 17).

UITTI, Ryan J, *et al.* "Parkinson disease: handedness predicts asymmetry". En: *Neurology* 64.11 (2005), págs. 1925-1930 (vid. pág. 27).

VAROL, Gül; LAPTEV, Ivan y SCHMID, Cordelia. "Long-term temporal convolutions for action recognition". En: *IEEE transactions on pattern analysis and machine intelligence* 40.6 (2017), págs. 1510-1517 (vid. pág. 37).

VERLEKAR, Tanmay T; SOARES, Luís D y CORREIA, Paulo L. “Automatic classification of gait impairments using a markerless 2D video-based system”. En: *Sensors* 18.9 (2018), pág. 2743 (vid. pág. 13).

VERLEKAR, Tanmay T.; SOARES, Luís Ducla y CORREIA, Paulo Lobato. “Automatic Classification of Gait Impairments Using a Markerless 2D Video-Based System”. En: *Sensors*. 2018 (vid. pág. 61).

VOS, Theo, *et al.* “Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”. En: *The Lancet* 390.10100 (2017), págs. 1211-1259 (vid. pág. 11).

WANG, Heng y SCHMID, Cordelia. “Action Recognition with Improved Trajectories”. En: *IEEE International Conference on Computer Vision*. Sydney, Australia, 2013 (vid. pág. 29).

WANG, Heng, *et al.* “Action recognition by dense trajectories”. En: *CVPR 2011*. IEEE. 2011, págs. 3169-3176 (vid. pág. 30).

ANEXOS

Anexo A. Productos Académicos

- GUAYACÁN, Luis C.; RANGEL, Edgar y MARTÍNEZ, Fabio. "Towards understanding spatio-temporal parkinsonian patterns from salient regions of a 3D convolutional network". En: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3688-3691). DOI: 10.1109/EMBC44109.2020.9175717.
- GUAYACÁN, Luis C. y MARTÍNEZ, Fabio. "Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network". En: Journal of Biomedical Informatics. DOI: 10.1016/j.jbi.2021.103935.
- GUAYACÁN, Luis C. y MARTÍNEZ, Fabio. "Quantification of Parkinsonian kinematic patterns in body-segment regions during locomotion". En: Medical Engineering & Physics Journal.
Estado: En revisión.