

**SOPORTE A LA CLASIFICACIÓN DIAGNÓSTICA DEL PARKINSON A PARTIR DE
HALLAZGOS RADIOLÓGICOS Y LA VALORACIÓN CLÍNICA DESDE UN MODELO DE
APRENDIZAJE A GRAN ESCALA.**

**ÁNGEL FERNANDO ORTEGA GARCÍA
ESTEFANÍA OSMA LLANES**

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

2026

**SOPORTE A LA CLASIFICACIÓN DIAGNÓSTICA DEL PARKINSON A PARTIR DE
HALLAZGOS RADIOLÓGICOS Y LA VALORACIÓN CLÍNICA DESDE UN MODELO DE
APRENDIZAJE A GRAN ESCALA.**

**ÁNGEL FERNANDO ORTEGA GARCÍA
ESTEFANÍA OSMA LLANES**

Trabajo de grado para optar al título de ingenieros de sistemas.

Director:

Fabio Martínez Carrillo

Ph.D en Ingeniería de Sistemas y Computación

Codirector:

Franklin Samuel Sierra Jerez

Magíster en Ingeniería de Sistemas e Informática

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

2026

Agradecimientos

Ángel Fernando Ortega García

Este triunfo se lo dedico a mis padres y hermanos, mi apoyo incondicional; mi familia, mi base, mi guía, mi ejemplo y mi motor principal. Que Dios los bendiga y los guarde por muchos años. Muchas gracias por todo. A mis amigos, mis rivales y mis aliados, quienes son mi motivación para ser cada día mejor y con quienes puedo contar para cualquier desafío. *“El hierro con hierro se afila; y así el hombre afila el rostro de su amigo.”* A la Universidad Industrial de Santander y a todos los “senseis” que conocí a lo largo de este hermoso camino. Finalmente, al grupo de investigación *Bivl²ab*, a las personas que conocí allí, a mi codirector, Franklin Sierra, y a mi director, el profesor Fabio Martínez Carrillo, un verdadero ejemplo de profesor, investigador y guía, por su infinita paciencia y su constante disposición a compartir su conocimiento desde la humildad y el amor por el mismo.

Estefanía Osma Llanes

Agradezco principalmente a Dios por guiarme y permitirme culminar este camino. A mi padre, Luis Fernando Osma Larrota, y a mi madre, Maribel Llanes Solano, por su amor, apoyo incondicional y por creer en mí. A mis hermanos, Kevin Fernando, Julián Alberto, Marisol y Dora Juliana; a mi sobrino Julián Esteban; y a toda mi familia, por ser siempre un pilar fundamental en mi vida. A mis amigos, con quienes compartí risas y lágrimas a lo largo de este proceso. A mi codirector, Franklin Sierra, y a mi director, Fabio Martinez, por su orientación, paciencia y por motivarme a crecer en el mundo de la ciencia. A mi compañero Ángel, por su apoyo y compromiso en este proceso de aprendizaje. Y un agradecimiento especial a quien me acompañó en cada paso de este recorrido, apoyándome, creciendo y aprendiendo a mi lado.

Índice general

	página
INTRODUCCIÓN	11
1. FUNDAMENTOS Y TRABAJO PREVIO	15
1.1. ENFERMEDAD DE PARKINSON	15
1.1.1. Imagen de resonancia magnética (MRI)	16
1.1.2. Tomografía computarizada por emisión de fotón único (SPECT)	17
1.1.3. Información clínica complementaria	19
1.2. APRENDIZAJE MULTIMODAL PROFUNDO	21
1.3. ARQUITECTURAS GRANDES DE VISIÓN Y DE LENGUAJE	23
1.3.1. Arquitectura BERT.	23
1.3.2. Modelo fundacional SAM-Med3D	24
1.4. TRABAJO PREVIO	27
2. PROBLEMA DE INVESTIGACIÓN	30
2.1. PREGUNTA DE INVESTIGACIÓN	31
2.1.1. Hipótesis	31
3. OBJETIVOS	32
3.1. OBJETIVO GENERAL	32
3.2. OBJETIVOS ESPECÍFICOS	32
4. ENFOQUE PROPUESTO	33
4.1. Descriptor Textual	33
4.2. Descriptor Visual	36
4.3. APRENDIZAJE MULTIMODAL MEDIANTE UNA REGLA CONTRASTIVA	37

4.4. CONFIGURACIÓN EXPERIMENTAL	38
4.4.1. Los datos.	38
4.4.2. Configuración de la arquitectura.	39
5. EVALUACIÓN Y RESULTADOS	48
5.1. SELECCIÓN DE REPRESENTACIONES VISUALES EN SAMMED3D	48
5.2. COMPARACIÓN DE ESTRATEGIAS DE FUSIÓN MULTIMODAL	49
5.3. ANÁLISIS DE DISCRIMACIÓN EN ESTADIOS NO OBSERVADOS	51
5.4. ESTRATEGIAS DE FUSIÓN	53
5.5. EVALUACIÓN DEL COMPONENTE CLÍNICO TEXTUAL	54
6. CONCLUSIONES Y TRABAJO FUTURO	56
BIBLIOGRAFÍA	57
APÉNDICE DE PRODUCTOS ACADÉMICOS	65

Índice de figuras

	página
Figura 1. Corte axial de MRI y SPECT en dos pacientes control y dos pacientes con EP.	18
Figura 2. Tipos de estrategias para la clasificación multimodal mediante la fusión de características.	22
Figura 3. Diagrama simplificado de BERT en la tarea de <i>Masked Language Modeling</i> . .	24
Figura 4. Arquitectura simplificada de <i>SAM-Med3D</i>	26
Figura 5. Flujo multimodal contrastivo propuesto para la estratificación de la EP.	34
Figura 6. Comparación antes y después de la extracción de cráneo.	41
Figura 7. Comparación antes y después de la corrección de intensidades con N4 Bias Field Correction.	42
Figura 8. Alineación de la imagen MRI al espacio de referencia MNI.	42
Figura 9. Visualización del atlas Harvard–Oxford y su parcelación anatómica.	44
Figura 10. Regiones anatómicas seleccionadas del atlas Harvard–Oxford.	45
Figura 11. Ejemplos representativos de predicción del modelo.	50
Figura 12. Análisis de confianza del modelo a través de diferentes niveles de severidad clínica.	52

Índice de cuadros

	página
Tabla 1. Variables clínicas utilizadas para generar los reportes clínicos.	35
Tabla 2. Descripción del conjunto de datos: grupos leve y moderado.	39
Tabla 3. Selección de <i>embedding</i> visual	48
Tabla 4. Métricas de clasificación para MRI: línea base multimodal vs. enfoque pro- puesto (SCL).	49
Tabla 5. Métricas de clasificación para SPECT: línea base multimodal vs. enfoque pro- puesto (SCL).	51
Tabla 6. Resultados de estrategias de fusión a nivel de decisión y comparación con el método propuesto.	53
Tabla 7. Métricas de clasificación en el conjunto de prueba para el enfoque propuesto y ablaciones del texto clínico (macro-promediadas).	54
Tabla 8. Aporte clínico de los componentes utilizados en el descriptor textual.	55

LISTA DE ANEXOS

página

Apéndice A. Productos Académicos 65

RESUMEN

TÍTULO: Soporte a la clasificación diagnóstica del Parkinson a partir de hallazgos radiológicos y la valoración clínica desde un modelo de aprendizaje a gran escala. *

AUTORES: ÁNGEL FERNANDO ORTEGA GARCÍA, ESTEFANÍA OSMA LLANES **

PALABRAS CLAVE: Enfermedad de Parkinson (EP), modelos multimodales, modelos fundacionales, aprendizaje contrastivo, neuroimágenes, información clínica, ClinicalBERT, SamMed3D.

DESCRIPCIÓN: La enfermedad de Parkinson (EP) es la segunda enfermedad neurodegenerativa más común a nivel mundial, afectando aproximadamente a 11 millones de personas. Su diagnóstico se basa principalmente en la observación clínica de alteraciones motoras, lo que conduce a detecciones tardías y a un alto grado de subjetividad, con tasas de error reportadas cercanas al 30 %. En la actualidad, las neuroimágenes han sido incorporadas como herramientas complementarias al diagnóstico, permitiendo analizar cambios estructurales y funcionales del cerebro incluso en estadios tempranos. Sin embargo, su interpretación continúa siendo predominantemente observacional y presenta variabilidad en los hallazgos radiológicos, alcanzando diagnósticos erróneos en hasta el 25 % de los casos. Es por ello que han emergido herramientas computacionales para el soporte diagnóstico usando la caracterización, modelamiento y aprendizaje de hallazgos radiológicos. Sin embargo, estos modelos suelen ser entrenados en entornos controlados y sin explotar la complementariedad con la información clínica.

Este trabajo de grado implementó un modelo de aprendizaje a gran escala para la clasificación de patrones parkinsonianos usando información radiológica y clínica. La información radiológica se representó mediante el modelo fundacional SamMed3D, mapeando regiones asociadas al putamen, núcleo caudado, tálamo y tronco encefálico. La información clínica fue codificada mediante ClinicalBERT. La integración de ambas modalidades fue lograda mediante una arquitectura contrastiva supervisada para clasificar distintos estadios de la EP.

La evaluación se realizó sobre una cohorte de 184 pacientes de la base de datos Parkinson's Progression Markers Initiative (PPMI). El enfoque propuesto alcanzó una precisión de 85 % y una puntuación macro-F1 de 83 %, superando enfoques basados exclusivamente en imágenes.

* Trabajo de investigación

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D. Codirector: Franklin Samuel Sierra Jerez.

Abstract

TITLE: A Large-Scale Learning Model for Diagnostic Classification of Parkinson's Disease Based on Radiological Findings and Clinical Assessments *

AUTHORS: ÁNGEL FERNANDO ORTEGA GARCÍA, ESTEFANÍA OSMA LLANES **

KEYWORDS: Parkinson's disease (PD), multimodal models, foundational models, contrastive learning, neuroimaging, clinical data, ClinicalBert, SamMed3D.

DESCRIPTION: Parkinson's disease (PD) is the second most common neurodegenerative disease worldwide, affecting approximately 11 million people. Its diagnosis is mainly based on the clinical observation of motor symptoms, which leads to late detection and a high degree of subjectivity, with reported error rates close to 30%. At present, neuroimaging has been incorporated as a complementary diagnostic tool, allowing the analysis of structural and functional changes in the brain even in early stages. However, its interpretation remains predominantly observational and shows variability in radiological findings, resulting in misdiagnoses in up to 25% of cases. For this reason, computational tools have emerged to support diagnosis through the characterization, modeling, and learning of radiological findings. However, these models are usually trained in controlled environments and do not exploit the complementarity of clinical information.

This thesis implemented a large-scale learning model for the classification of parkinsonian patterns using radiological and clinical information. Radiological information was represented using the SamMed3D foundation model, mapping regions associated with the putamen, caudate nucleus, thalamus, and brainstem. Clinical information was encoded using ClinicalBERT. The integration of both modalities was achieved through a supervised contrastive architecture for the classification of different PD stages.

The evaluation was conducted on a cohort of 184 patients from the Parkinson's Progression Markers Initiative (PPMI) database. The proposed approach achieved an accuracy of 85% and a macro-F1 score of 83%, outperforming approaches based exclusively on imaging data.

* Research work

** Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez Carrillo, Ph.D. Co-advisor: Franklin Samuel Sierra Jerez.

INTRODUCCIÓN

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo crónico y progresivo asociado con la pérdida gradual de células dopaminérgicas en la sustancia negra compacta (mesencéfalo), que afecta principalmente al sistema nervioso central ¹. El diagnóstico estándar de la EP se basa principalmente en la evaluación clínica del desempeño motor y cognitivo del paciente. Sin embargo, estas afectaciones motoras suelen evidenciarse cuando ya ha ocurrido una pérdida de dopamina estriatal de entre el 35 % y el 45 %, lo que limita el diagnóstico y el tratamiento oportunos ².

En el contexto del análisis de neuroimágenes, las imágenes de resonancia magnética T1 son una herramienta estándar en neurorradiología para el estudio estructural y volumétrico cerebral. Las regiones asociadas a la enfermedad, evaluadas a partir de este tipo de imágenes, incluyen el estriado (putamen y núcleo caudado), el tronco encefálico, el mesencéfalo, el hipocampo, el lóbulo temporal y la amígdala. No obstante, los hallazgos morfométricos basados en T1 son heterogéneos y dichas alteraciones estructurales tienden a ser más evidentes en fases avanzadas y en fenotipos asociados con deterioro cognitivo ^{3,4}. Actualmente, el análisis de este tipo de imágenes de resonancia magnética T1 se emplea principalmente como protocolo de triaje para diferenciar la enfermedad de Parkinson de otras patologías que

¹ William Dauer y Serge Przedborski. "Parkinson's disease: mechanisms and models". En: *Neuron* 39.6 (2003), págs. 889-909.

² N. Heng et al. "Striatal dopamine loss in early Parkinson's disease: Systematic review and novel analysis of dopamine transporter imaging". En: *Movement Disorders Clinical Practice* 10.4 (2023), págs. 539-546. DOI: 10.1002/mdc3.13658.

³ Beatrice Heim et al. "Magnetic resonance imaging for the diagnosis of Parkinson's disease". En: *Journal of neural transmission* 124.8 (2017), págs. 915-964.

⁴ Trina Mitchell et al. "Emerging neuroimaging biomarkers across disease stage in Parkinson disease: a review". En: *JAMA neurology* 78.10 (2021), págs. 1262-1272.

presentan trastornos motores similares ⁵. Incluso en la práctica clínica, esta evaluación continúa siendo en gran medida subjetiva y no proporciona información funcional directa sobre las vías dopaminérgicas ^{6,7}. De acuerdo con la literatura, la interpretación visual por parte de expertos puede asociarse con tasas de falsos positivos de hasta el 20 % en la diferenciación de otros parkinsonismos ⁸.

Herramientas computacionales han surgido como una alternativa para apoyar la inspección radiológica y contribuir a la estratificación de la enfermedad de Parkinson (EP) ^{9,10}. En general, estos enfoques generan descriptores que codifican patrones radiológicos asociados con hallazgos estructurales y funcionales de la EP ^{11,12}. Muchos de estos métodos se ba-

-
- ⁵ Lydia Chougar et al. "The role of magnetic resonance imaging for the diagnosis of atypical parkinsonism". En: *Frontiers in Neurology* 11 (2020), pág. 665.
- ⁶ K Morgen et al. "Structural brain abnormalities in patients with Parkinson disease: a comparative voxel-based analysis using T1-weighted MR imaging and magnetization transfer imaging". En: *American journal of neuroradiology* 32.11 (2011), págs. 2080-2086.
- ⁷ Schekeb Aludin y Lars-Patrick Andreas Schmill. "MRI signs of Parkinson's disease and atypical parkinsonism". En: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. Vol. 193. 12. Georg Thieme Verlag KG. 2021, págs. 1403-1410.
- ⁸ Chadrick Dewey et al. "Structural MRI ratios fail to distinguish progressive supranuclear palsy from Parkinson disease in individual patients". En: *Neurology: Clinical Practice* 13.3 (2023), e200157.
- ⁹ S. Priyadharshini et al. "A comprehensive framework for Parkinson's disease diagnosis using explainable artificial intelligence empowered machine learning techniques". En: *Alexandria Engineering Journal* 107 (2024), págs. 568-582.
- ¹⁰ M. Wenzel et al. "Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics". En: *European Journal of Nuclear Medicine and Molecular Imaging* 46 (2019), págs. 2800-2811.
- ¹¹ Yiwei Pan et al. "A multi-atlas-based [18F] 9-fluoropropyl-(+)-dihydrotrabenazine positron emission tomography image segmentation method for Parkinson's disease quantification". En: *Frontiers in Aging Neuroscience* 14 (2022), pág. 902169.
- ¹² Dafa Shi et al. "Classification of Parkinson's disease using a region-of-interest-and resting-state functional magnetic resonance imaging-based radiomics approach". En: *Brain Imaging and Behavior* 16.5 (2022), págs. 2150-2163.

san en la segmentación de regiones cerebrales específicas, como los ganglios basales, el tronco encefálico y el mesencéfalo ^{13,14}. En este contexto, los atlas cerebrales desempeñan un papel fundamental para delimitar regiones de interés y estimar descriptores volumétricos que permitan clasificar la población con EP ^{15,10}. Sin embargo, estos métodos suelen centrarse en muestras en estadios avanzados para cuantificar cambios estructurales, lo que limita su utilidad en fases tempranas. Otros enfoques emplean cortes o volúmenes completos de imágenes MRI-T1 para aprender automáticamente patrones espaciales mediante modelos de clasificación entrenados con sujetos control y pacientes diagnosticados con EP ^{16,17}. Aunque estos métodos han demostrado capacidad para diferenciar entre población sana y pacientes con EP, presentan una sensibilidad limitada para estratificar entre distintos estadios de la enfermedad, lo cual es un aspecto clave para el diagnóstico temprano y la planificación del tratamiento.

Además, el entrenamiento de estos modelos tradicionales suele requerir grandes volúmenes de datos, lo cual representa un desafío en entornos clínicos reales. Considerando la naturaleza multifactorial y heterogénea de la EP, su análisis requiere un enfoque que complemente la información neurológica con escalas motoras y cognitivas. Sin embargo, los enfoques actuales basados exclusivamente en imágenes suelen descartar información clínica disponible

¹³ Takuro Shiiba et al. "Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting Parkinson's disease". En: *EJNMMI research* 12.1 (2022), pág. 39.

¹⁴ Hongyun Tan, Binbin Luo y Chao Cong. "Using Deep Learning to differentiate between Parkinson's Disease and Multiple System Atrophy based on PET and MRI images". En: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2023, págs. 3181-3187.

¹⁵ Chang Li et al. "Automatic diagnosis of Parkinson's disease using artificial intelligence base on routine T1-weighted MRI". En: *Frontiers in Medicine* 10 (2024), pág. 1303501.

¹⁶ Milton Camacho et al. "Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets". En: *NeuroImage: Clinical* 38 (2023), pág. 103405.

¹⁷ Çağatay Berke Erdaş y Emre Sümer. "A deep learning method to detect Parkinson's disease from MRI slices". En: *SN Computer Science* 3.2 (2022), pág. 120.

en la práctica rutinaria, como escalas de valoración motora, datos demográficos e incluso indicadores de progresión. Esta información puede complementar el diagnóstico, aumentar la sensibilidad para diferenciar estadios y orientar la identificación de patrones de imagen asociados con cambios sutiles de la enfermedad ¹⁸. La ausencia de métodos multimodales que integren información radiológica y clínica limita una visión holística de la EP y la identificación de relaciones clave dentro del proceso neurodegenerativo ¹⁹.

Es por tanto que el presente trabajo de grado propone un enfoque multimodal basado en aprendizaje contrastivo supervisado, el cual integra hallazgos de resonancia magnética tridimensional (MRI-T1) con escalas clínicas cognitivas y motoras en formato textual, con el objetivo de apoyar la estratificación de la EP. Desde la rama textual, se obtuvo una representación mediante ClinicalBERT a partir de información clínica y reportes motores. En cuanto al descriptor radiológico, se propuso extraer una representación embebida a partir del modelo fundacional SamMed3D, procesando imágenes MRI-T1 junto con un prompt espacial que delimita regiones asociadas con la EP. Dicho prompt se definió automáticamente alineando cada muestra con el atlas Harvard-Oxford (HarvardOxford sub maxprob thr25 1mm) ²⁰, recuperando regiones correspondientes al putamen, núcleo caudado, tálamo y tronco encefálico. Finalmente, ambas ramas se integraron mediante una pérdida contrastiva supervisada, con el fin de alinear las representaciones embebidas y construir un espacio geométrico con capacidad discriminativa suficiente para separar los distintos estadios de la enfermedad.

¹⁸ Konrad Szewczyk-Krolikowski et al. "The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort". En: *Parkinsonism & related disorders* 20.1 (2014), págs. 99-105.

¹⁹ Ronald B Postuma et al. "MDS clinical diagnostic criteria for Parkinson's disease". En: *Movement disorders* 30.12 (2015), págs. 1591-1601.

²⁰ <https://identifiers.org/neurovault.image:1700>

1. FUNDAMENTOS Y TRABAJO PREVIO

1.1. ENFERMEDAD DE PARKINSON

La enfermedad de Parkinson (EP) es el segundo trastorno neurodegenerativo más común a nivel mundial, registrando en el 2021 una cifra de más de 11 millones de personas. A pesar de que hoy en día se desconoce su causa principal, esta enfermedad del sistema nervioso central ha evidenciado una alta correlación con la pérdida paulatina de neuronas dopaminérgicas en la sustancia negra (SN) ubicada en el mesencéfalo ¹. En esta región es donde se da la producción de dopamina, un neurotransmisor esencial para el control de los movimientos voluntarios y el estado cognitivo/anímico ²¹. Por tanto, niveles anómalos de este neurotransmisor ocasionan desórdenes tanto motores como no motores²². Particularmente, dentro de los síntomas motores se destacan la bradicinesia (lentitud en la ejecución de movimientos voluntarios), temblor, rigidez muscular, inestabilidad postural y alteraciones en la marcha ²³. Por otra parte, algunos síntomas no motores incluyen alteraciones cognitivas, trastornos del sistema nervioso reflejados en los estados de ánimo y del sueño, anosmia, entre otros ²⁴. El diagnóstico de la EP se basa principalmente en la evaluación observacional de estos síntomas durante la realización de acciones específicas, donde la severidad suele determinarse mediante escalas de referencia como la *Unified Parkinson's Rating Scale* (UPDRS) y

²¹ Thomas B Stoker y Julia C Greenland. "Parkinson's disease: pathogenesis and clinical aspects [internet]". En: (2018).

²² Saad Latif et al. "Dopamine in Parkinson's disease". En: *Clinica chimica acta* 522 (2021), págs. 114-126.

²³ Sigurlaug Sveinbjornsdottir. "The clinical symptoms of Parkinson's disease". En: *Journal of neurochemistry* 139 (2016), págs. 318-324.

²⁴ Abdul Qayyum Rana et al. "Parkinson's disease: a review of non-motor symptoms". En: *Expert review of neurotherapeutics* 15.5 (2015), págs. 549-562.

la *Hoehn and Yahr scale* (H&Y)²⁵. Sin embargo, dicha valoración resulta altamente subjetiva y tardía, conllevando a errores diagnósticos, especialmente en las fases tempranas de la enfermedad, donde los síntomas pueden pasar inadvertidos²⁶. Además, se ha identificado que el diagnóstico suele evidenciarse cuando una proporción considerable de las neuronas dopaminérgicas ya ha sufrido daño, dificultando una intervención temprana y efectiva²⁷. En este orden de ideas, las técnicas de neuroimagen, como la resonancia magnética (MRI, por sus siglas en inglés) y la tomografía computarizada por emisión de fotón único (SPECT, por sus siglas en inglés), han sido empleadas dentro de la rutina clínica. Este tipo de neuroimágenes han permitido el análisis estructural y funcional del cerebro, que junto con la información clínica recolectada durante el diagnóstico observacional, ha permitido una mejor caracterización de la EP²⁸.

1.1.1. Imagen de resonancia magnética (MRI) Este tipo de neuroimagen constituye el primer tipo de neuroimagen posterior al análisis observacional por parte del experto. Particularmente, secuencias como T1 y T2 aportan información relevante para estudiar diferencias estructurales y volumétricas del cerebro frente a otras patologías con sintomatología motora semejante²⁹. No obstante, la valoración visual de estas imágenes puede ser tediosa y sub-

²⁵ Christopher G Goetz et al. "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease". En: *Movement disorders* 19.9 (2004), págs. 1020-1028.

²⁶ Jayakrishna S Madabushi et al. "Parkinson's disease: diagnostic challenges amidst transdiagnostic and overlapping mental health symptoms". En: *Cureus* 15.3 (2023).

²⁷ Janina Grosch, Jürgen Winkler y Zacharias Kohl. "Early degeneration of both dopaminergic and serotonergic axons—a common mechanism in Parkinson's disease". En: *Frontiers in cellular neuroscience* 10 (2016), pág. 293.

²⁸ Jaskeerat Gujral et al. "PET, SPECT, and MRI imaging for evaluation of Parkinson's disease". En: *American Journal of Nuclear Medicine and Molecular Imaging* 14.6 (2024), pág. 371.

²⁹ Usman Saeed et al. "Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: current and emerging concepts". En: *Translational neurodegeneration* 6 (2017), págs. 1-25.

jetiva, especialmente en estadios tempranos, donde las diferencias frente a sujetos control suelen ser sutiles o incluso inapreciables ^{29,30}.

Adicionalmente, algunas alteraciones en regiones como la sustancia negra han sido estudiadas mediante secuencias MRI especializadas, como SWI-MRI (*Susceptibility Weighted Imaging in MRI*), que permiten analizar cambios asociados a la concentración de hierro y otros hallazgos sutiles de relevancia diagnóstica ²⁹. A pesar de ello, la interpretación de estos hallazgos continúa siendo difícil incluso para expertos, reportándose errores en la diferenciación de otros parkinsonismos de hasta el 20 % ³¹.

1.1.2. Tomografía computarizada por emisión de fotón único (SPECT) La SPECT es una técnica imagenológica nuclear que permite valorar funcionalmente la integridad del sistema dopaminérgico mediante el uso de radiofármacos específicos en el estriado cerebral ³². Esta modalidad suele emplearse cuando el análisis estructural mediante MRI resulta insuficiente para un diagnóstico formal de la EP. En SPECT, la integridad dopaminérgica suele evidenciarse mediante dos regiones simétricas brillantes conocidas como el “signo de la coma”. La alteración o ausencia de este patrón se asocia con la presencia de EP, incluso en estadios iniciales ³³. En la figura 1 se muestra el “signo de la coma” en un paciente sano frente a un paciente EP, donde la asimetría es evidente.

A pesar de estos grandes beneficios, la SPECT presenta una serie de limitaciones. En pri-

³⁰ Jörg B Schulz et al. “Magnetic resonance imaging–based volumetry differentiates idiopathic Parkinson’s syndrome from multiple system atrophy and progressive supranuclear palsy”. En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.1 (1999), págs. 65-74.

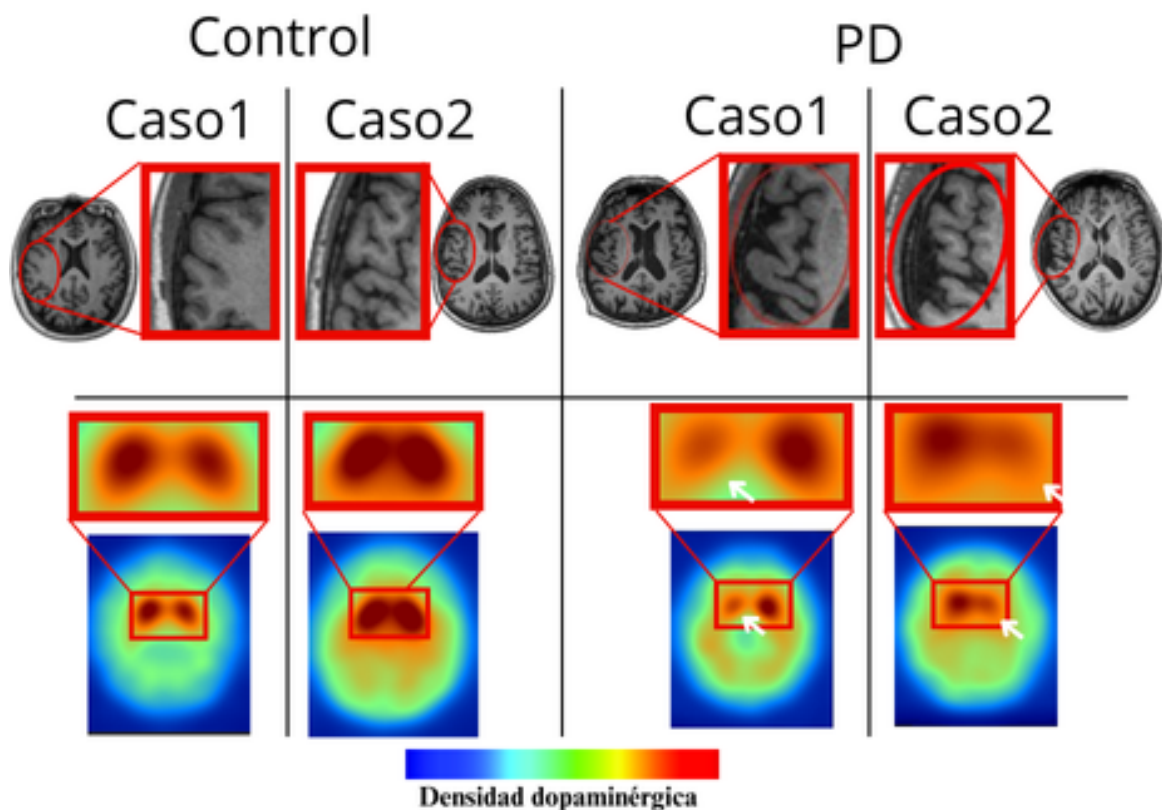
³¹ Shuo Gao et al. “Early detection of Parkinson’s disease through multiplex blood and urine biomarkers prior to clinical diagnosis”. En: *npj Parkinson’s Disease* 11.1 (2025), pág. 35.

³² Mehrzad Cohenpour y Haim Golan. “Nuclear neuroimaging of dopamine transporter in Parkinsonism–role in routine clinical practice”. En: *Harefuah* 146.9 (2007), págs. 698-702.

³³ Ling Wang et al. “SPECT Molecular Imaging in Parkinson’s Disease”. En: *BioMed Research International* 2012.1 (2012), pág. 412486.

mera medida, su uso requiere de sustancias radioactivas que pueden ser perjudiciales en estudios longitudinales ³⁴. Dado lo anterior y a sus costes operacionales, esta modalidad es altamente restringida a centros especializados y académicos ³⁴. Adicionalmente, dada su naturaleza funcional, no existe una referencia estructural que permita establecer las zonas cerebrales que presentan la mayor alteración de concentración dopaminérgica, representando esto un problema aún por abordar en la comunidad científica ³⁴.

Figura 1. Corte axial de MRI y SPECT en dos pacientes control y dos pacientes con EP, respectivamente. La diferencia volumétrica (óvalo rojo) puede ser difícil de identificar incluso por parte de expertos. Nótese también una asimetría en la actividad dopaminérgica producida por la EP (señalada con flecha blanca).



³⁴ Tohid Mortezaadeh et al. "Imaging modalities in differential diagnosis of Parkinson's disease: opportunities and challenges". En: *Egyptian journal of radiology and nuclear medicine* 52 (2021), págs. 1-12.

1.1.3. Información clínica complementaria La información clínica corresponde al conjunto de datos recolectados durante el proceso de diagnóstico. Dada la naturaleza heterogénea de la enfermedad y la ausencia de test definitivos para su diagnóstico, el análisis de la información clínica podría ser clave para identificar características motoras relevantes y sugerir diagnósticos alternos ³⁵. Esta información sigue escalas de referencia que permiten cuantificar el grado de afectación del paciente a la hora de desempeñar actividades específicas. Mediante dichas escalas se ha observado una correlación positiva entre ciertas manifestaciones motoras (como la bradicinesia) y el deterioro neurológico a la hora de desempeñar pruebas específicas ^{36,37}

Entre estas escalas, se encuentran la *Unified Parkinson's Disease Rating Scale* (UPDRS), la *Hoehn & Yahr* (H&Y), y la *Modified Bradykinesia Rating Scale* (MBRS), como las de amplio uso clínico. Mediante estas escalas se ha logrado valorar la progresión de la enfermedad, la severidad de los síntomas, la respuesta al tratamiento y la relación motora con el daño cognitivo ¹⁹. A continuación, se menciona brevemente cada una de ellas.

La *Unified Parkinson's Disease Rating Scale* (UPDRS) se compone de cuatro secciones principales:

- **Parte I:** síntomas no motores, como alteraciones cognitivas, del sueño, del ánimo, apatía o fatiga.
- **Parte II:** función motora en actividades cotidianas como comer, escribir, vestirse y ca-

³⁵ Benjamin Shickel et al. "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis". En: *IEEE journal of biomedical and health informatics* 22.5 (2017), págs. 1589-1604.

³⁶ Asha Kishore et al. "Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: differential effects on bradykinesia". En: *Movement Disorders: Official Journal of the Movement Disorder Society* 22.3 (2007), págs. 328-333.

³⁷ Francois JG Vingerhoets et al. "Which clinical sign of Parkinson's disease best reflects the nigrostriatal lesion?" En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 41.1 (1997), págs. 58-64.

minar.

- **Parte III:** evaluación motora objetiva realizada por el especialista, incluyendo temblor, rigidez, bradicinesia, marcha e inestabilidad postural.
- **Parte IV:** complicaciones del tratamiento dopaminérgico, como fluctuaciones motoras y discinesias.

La escala Hoehn & Yahr proporciona un marco global para evaluar la progresión de la enfermedad en una escala ordinal de 0 a 5:

- **0:** asintomático.
- **1:** afectación unilateral.
- **2:** afectación bilateral sin deterioro del equilibrio.
- **3:** afectación moderada con inestabilidad postural.
- **4:** afectación grave, aún capaz de caminar o estar de pie sin ayuda.
- **5:** postrado en cama o en silla de ruedas, salvo asistencia.

La *Modified Bradykinesia Rating Scale* (MBRS) es una escala diseñada para valorar los efectos de la bradicinesia, uno de los principales síntomas motores de la EP. Durante esta evaluación se analiza el desempeño del paciente al ejecutar tareas motoras de manera unilateral y bilateral. Particularmente, estos ejercicios corresponden a los ítems 23 a 26 de la escala UPDRS, en los que se evalúan el golpeteo de dedos, el agarre de manos y los movimientos de pronación-supinación en extremidades superiores e inferiores ³⁶.

Mediante su ejecución se evalúan tres componentes: velocidad, amplitud y ritmo. Cada uno de estos componentes recibe una puntuación de 0 a 4, siendo 0 la ausencia de bradicinesia y 4 la incapacidad casi total para realizar el movimiento. Estudios previos sugieren que la escala MBRS es más sensible a identificar los diferentes aspectos de la bradicinesia que la valoración completa de la UPDRS ³⁶.

Además, la valoración de estas tareas motoras han mostrado una correlación frente a hallazgos en imágenes funcionales consideradas como referencia para detectar cambios y evaluar la progresión de la enfermedad ³⁸. Por otra parte, también se ha demostrado que la bradicinesia es el resultado directo de lesiones dopaminérgicas nigroestriatales que afectan la función motora a través del circuito ganglios basales–tálamo–corteza ³⁷.

1.2. APRENDIZAJE MULTIMODAL PROFUNDO

La caracterización multimodal se refiere a la integración de información proveniente de diversas fuentes, como datos textuales, visuales, clínicos o radiológicos ³⁹. En particular, el aprendizaje multimodal busca construir modelos capaces de procesar y relacionar información de múltiples dominios, capturando correspondencias entre modalidades y proyectándolas en un espacio de representación común ^{39,40}.

Particularmente, se definen secuencias de entrada correspondientes a cada modalidad, representadas como vectores de características $x^{(r)}$ y $x^{(c)}$, de dimensiones r y c , respectivamente. Dado un conjunto de etiquetas Y , se entrena un modelo multimodal f que mapea las entradas $x^{(r)}$ y $x^{(c)}$ hacia el espacio de etiquetas, es decir, $f(x^{(r)}, x^{(c)}) \mapsto Y$.

Típicamente, se pueden construir dos arquitecturas unimodales para $x^{(r)}$ y $x^{(c)}$, denotadas como f_r y f_c , respectivamente, donde la predicción final del modelo multimodal se define como:

$$\hat{y} = f(x^{(r)}, x^{(c)}) = g(f_r(x^{(r)}), f_c(x^{(c)}))$$

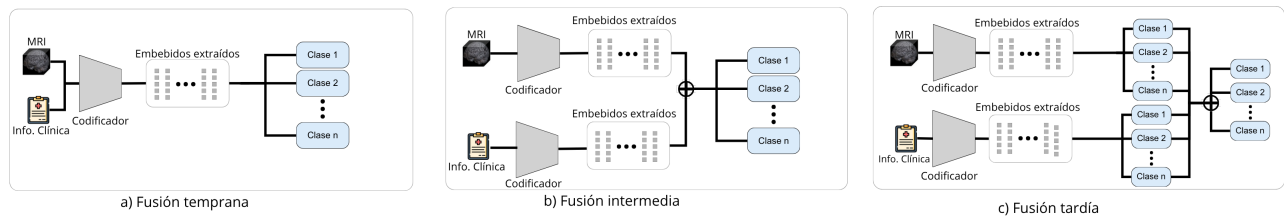
³⁸ Dustin A Heldman et al. “The modified bradykinesia rating scale for Parkinson’s disease: reliability and comparison with kinematic measures”. En: *Movement Disorders* 26.10 (2011), págs. 1859-1863.

³⁹ Tadas Baltrušaitis, Chaitanya Ahuja y Louis-Philippe Morency. “Multimodal machine learning: A survey and taxonomy”. En: *IEEE transactions on pattern analysis and machine intelligence* 41.2 (2018), págs. 423-443.

⁴⁰ Hao Wu et al. “Unified visual-semantic embeddings: Bridging vision and language with structured meaning representations”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, págs. 6609-6618.

donde g representa la operación de fusión de características, y \hat{y} corresponde a la etiqueta de clase predicha por el modelo. La fusión multimodal puede realizarse mediante distintas estrategias, que varían según el punto en que se integran las modalidades y puede clasificarse en fusión temprana, fusión intermedia o fusión tardía. En la fusión temprana las modalidades se combinan desde la entrada o en las primeras capas del modelo, de tal forma que el sistema aprende una representación conjunta desde el inicio. En la fusión intermedia cada modalidad se procesa primero por separado para obtener representaciones propias para luego ser combinadas en una etapa interna del modelo. Finalmente, la fusión tardía genera su propia predicción para cada modalidad y la decisión final se obtiene combinando esas salidas ³⁹.

Figura 2. Tipos de estrategias para la clasificación multimodal mediante la fusión de características.



Recientemente, el aprendizaje contrastivo es una estrategia de entrenamiento que busca construir representaciones discriminativas aproximando en el espacio latente aquellas muestras que comparten información semántica y separando aquellas que pertenecen a clases o contextos distintos ^{41,42}. El aprendizaje contrastivo define pares positivos y pares negativos, los pares positivos corresponden a muestras que comparten una representación cercana, mientras que los pares negativos representan representaciones distintas que deben separarse. En escenarios multimodales, el aprendizaje contrastivo permite alinear fuentes de información en un espacio compartido. De esta manera, una imagen puede acercarse a su descripción textual

⁴¹ Aaron van den Oord, Yazhe Li y Oriol Vinyals. "Representation learning with contrastive predictive coding". En: *arXiv preprint arXiv:1807.03748* (2018).

⁴² Ting Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations". En: *arXiv preprint arXiv:2002.05709* (2020).

correspondiente cuando ambas representan la misma clase, facilitando la construcción de representaciones conjuntas más robustas y discriminativas.

Un ejemplo exitoso en el estado del arte, usando modelos contrastivos, son los modelos fundacionales. Estos modelos a gran escala aprenden representaciones que pueden adaptarse a tareas específicas sin necesidad de reentrenar completamente la arquitectura ^{43,44}. Típicamente, estos trabajos involucran diferentes modos, mediante reglas contrastivas, identificando ventajas potenciales en el uso novedoso de información multimodal, alineando la información en espacios embebidos. A continuación se describe brevemente ejemplos de modelos sobresalientes en texto e imágenes.

1.3. ARQUITECTURAS GRANDES DE VISIÓN Y DE LENGUAJE

1.3.1. Arquitectura BERT. Esta arquitectura es un modelo basado en *transformers* bidireccionales que aprende representaciones profundas del lenguaje mediante tareas auto-supervisadas como el enmascaramiento de tokens (*Masked Language Modeling*) y la predicción de la siguiente oración (*Next Sentence Prediction*) ⁴⁵. Su arquitectura permite capturar dependencias contextuales tanto de izquierda a derecha como de derecha a izquierda, generando representaciones semánticas ricas y transferibles. La Figura 3 ilustra de manera simplificada el flujo de procesamiento de BERT en la tarea de *Masked Language Modeling*. En este ejemplo, una secuencia de tokens de entrada contiene un token enmascarado (*MASK*), el cual es proyectado inicialmente a una representación embebida. Posteriormente, toda la

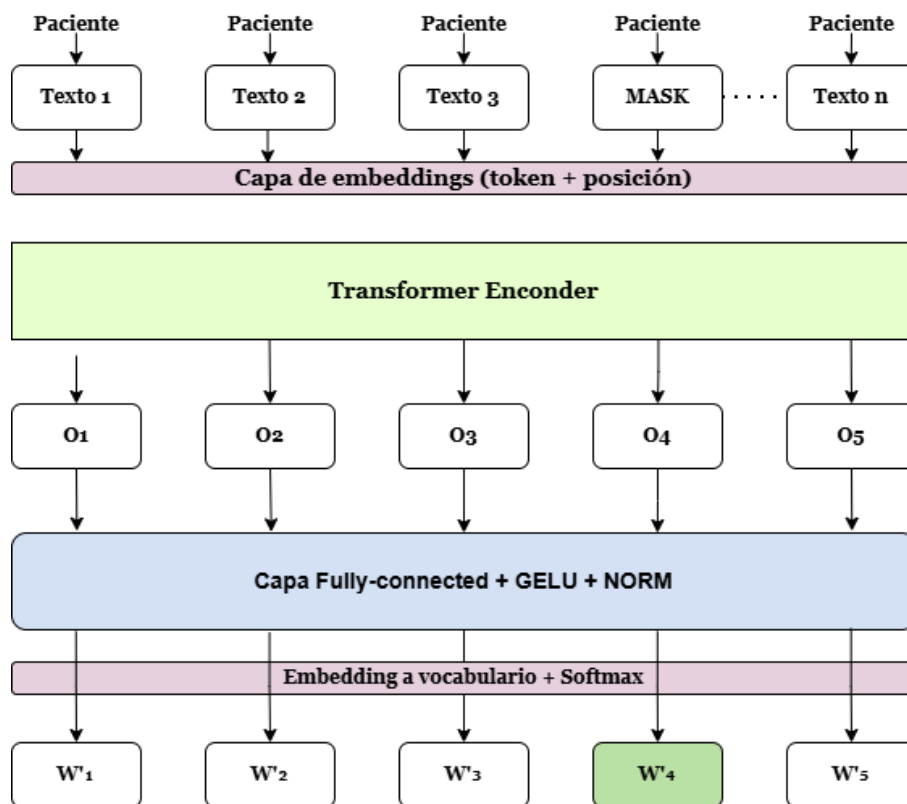
⁴³ Rishi Bommasani et al. "On the opportunities and risks of foundation models". En: *arXiv preprint arXiv:2108.07258* (2021).

⁴⁴ Michael Moor et al. "Foundation models for generalist medical artificial intelligence". En: *Nature* 616.7956 (2023), págs. 259-265.

⁴⁵ Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". En: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019, págs. 4171-4186.

secuencia es procesada por un codificador *Transformer Encoder*, que genera representaciones contextualizadas para cada posición y finalmente, estas representaciones pasan por una capa completamente conectada seguida de una activación GELU y una normalización, para luego proyectarse nuevamente al vocabulario mediante una capa *softmax*, permitiendo predecir el token originalmente oculto.

Figura 3. Diagrama simplificado de BERT en la tarea de *Masked Language Modeling*.



1.3.2. Modelo fundacional SAM-Med3D . El modelo SAM-Med3D es una adaptación del modelo *Segment Anything* (SAM) orientada específicamente a la segmentación automática y generalizable en imágenes médicas volumétricas. SAM-Med3D fue ajustado sobre volúmenes médicos de diferente naturaleza, como MRI, CT y PET, entre otros. Esta capacidad le permite delinear distintas estructuras anatómicas y patológicas en entornos donde

la anotación manual suele ser costosa, subjetiva y limitar la escalabilidad de los enfoques tradicionales de segmentación ⁴⁶. Su arquitectura, al igual que la del modelo SAM, consta de tres componentes principales. En primera medida, un codificador de imagen 3D codifica en una representación compacta el volumen de entrada (embedding de imagen) a partir de diferentes bloques de atención basados en *transformers* adaptados para el análisis de características volumétricas. Al mismo tiempo, un codificador de *prompts* se encarga de refinar la máscara de segmentación empleando puntos de referencia previamente marcados en el espacio 3D. Posteriormente, un decodificador de máscara 3D combina las representaciones de entrada (volumen y *prompt*) para generar una o varias máscaras de segmentación y, en un proceso iterativo, reutiliza el codificador de *prompts* y el decodificador de máscara para refinar la máscara generada.

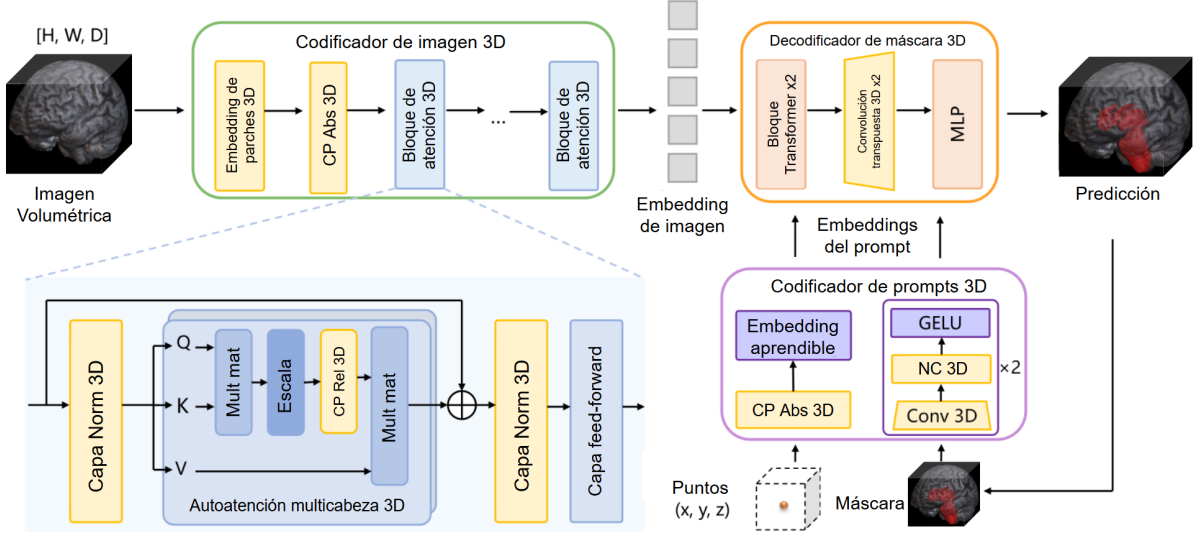
De esta manera, SAM-Med3D logra segmentaciones anatómicas sobre diferentes modalidades médicas, mostrando resultados competitivos sin importar la diversidad de los datos médicos empleados durante su entrenamiento. Este hecho resulta útil en aplicaciones donde los datos y sus anotaciones asociadas suelen ser escasos y costosos de conseguir, ofreciendo flexibilidad para ajustarse a nuevas tareas clínicas ⁴⁷.

La Figura 4 evidencia el funcionamiento general de la arquitectura SAM-Med3D. Matemáticamente, la entrada puede expresarse como un volumen $V \in \mathbb{R}^{H \times W \times D}$, donde H , W y D representan la altura, el ancho y la profundidad del volumen, respectivamente. Este volumen es procesado por un codificador 3D que transforma V en un embedding semántico $f \in \mathbb{R}^{h \times w \times d \times c}$, en el cual las dimensiones espaciales h , w y d son menores que H , W y D , mientras que la dimensionalidad de características c se incrementa. A continuación, los *prompts* del usuario, que corresponden a puntos tridimensionales (x, y, z) y, en

⁴⁶ Yichi Zhang, Zhenrong Shen y Rushi Jiao. "Segment anything model for medical image segmentation: Current applications and future directions". En: *Computers in Biology and Medicine* (2024), pág. 108238.

⁴⁷ Haoyu Wang et al. "SAM-Med3D: a vision foundation model for general-purpose segmentation on volumetric medical images". En: *IEEE Transactions on Neural Networks and Learning Systems* (2025).

Figura 4. Arquitectura simplificada de *SAM-Med3D*, compuesta por un codificador de imagen 3D, un codificador de *prompts* 3D y un decodificador de máscara 3D.



etapas de refinamiento, también a la máscara generada en la iteración previa, son procesados por un codificador de *prompts* 3D y se convierten en embeddings e_p . Estos embeddings e_p se combinan con el embedding volumétrico f para formar una representación unificada $z = \text{Combine}(f, e_p)$. Esta representación z constituye la entrada de un decodificador de máscara 3D, que genera un conjunto de máscaras $\{M_i\}$ junto con sus respectivos puntajes de confianza $\{s_i\}$, es decir, $(M_i, s_i) = \text{Decoder}(z)$. El modelo devuelve un conjunto de máscaras segmentadas M , cada una asociada a su puntaje de confianza s_i . Además, existe un mecanismo de refinamiento iterativo de máscaras. Esto se logra reutilizando el codificador de *prompts* 3D no solo para procesar puntos, sino también para procesar la máscara generada en cada iteración. Esto permite que el modelo corrija y ajuste la segmentación de forma progresiva. De manera general, el proceso se puede expresar como:

$$\{(M_i, s_i)\}_{i=1}^N = \text{SAM-Med3D}(V, \text{prompt}),$$

donde N es el número de máscaras generadas. Este modelo es flexible y permite segmentar diferentes estructuras en volúmenes médicos tridimensionales a partir de puntos de referen-

cia.

1.4. TRABAJO PREVIO

Hoy en día, las imágenes estructurales y funcionales permiten diferenciar enfermedades a partir de la extracción de características que posteriormente pueden ser analizadas mediante estrategias clásicas de *Machine Learning*. En los últimos años, la aplicación de modelos de aprendizaje profundo ha representado un avance importante en la detección automática de la enfermedad de Parkinson (EP) mediante el análisis de imágenes médicas y datos clínicos, mejorando su caracterización y apoyo diagnóstico ^{48,49}. Estos modelos de aprendizaje profundo aprovechan la extracción jerárquica de características, principalmente a través de capas convolucionales, las cuales identifican patrones espaciales como texturas, formas y relaciones locales, que luego son procesados por capas superiores. A medida que la información se propaga a través de la red, emergen representaciones semánticas de mayor nivel, que finalmente conducen a la predicción del modelo.

Las imágenes MRI y SPECT se utilizan comúnmente para identificar patrones asociados con la EP. En particular, modelos basados en redes convolucionales han mostrado mejoras en tareas de clasificación al combinar imágenes SPECT con datos clínicos, aunque aún presentan limitaciones para interpretar hallazgos radiológicos complejos ⁵⁰. También se han explorado enfoques híbridos que combinan distintas estrategias de aprendizaje. Por ejemplo, Pahuja y Prasad propusieron un modelo que integra un autoencoder, una CNN y una LSTM (*Long*

⁴⁸ Vincenzo Dentamaro et al. "Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database". En: *Scientific Reports* 14.1 (2024), pág. 20941.

⁴⁹ Xingyu Gao et al. "Brain status transferring generative adversarial network for decoding individualized atrophy in Alzheimer's disease". En: *IEEE Journal of Biomedical and Health Informatics* 27.10 (2023), págs. 4961-4970.

⁵⁰ Jun-En Ding, Chien-Chin Hsu y Feng Liu. "Parkinson's disease classification using contrastive graph cross-view learning with multimodal fusion of SPECT images and clinical features". En: (2024), págs. 1-5.

Short-Term Memory) para combinar características extraídas de imágenes SPECT con información clínica y biológica en una tarea de clasificación binaria de la EP ⁵¹. Estos resultados muestran que la integración entre modalidades puede aportar información complementaria útil para la caracterización de la enfermedad.

En paralelo, se han desarrollado estrategias basadas específicamente en imágenes estructurales de resonancia magnética. Estos enfoques buscan capturar cambios anatómicos sutiles a partir de descriptores morfométricos, radiómicos o representaciones profundas derivadas de cortes o volúmenes completos de MRI. Aunque estos métodos han mostrado capacidad para diferenciar pacientes con EP de sujetos sanos, una limitación importante es que la MRI estructural convencional suele reflejar cambios heterogéneos y poco específicos durante las fases tempranas de la enfermedad, lo cual restringe su sensibilidad para tareas más finas, como la estratificación entre estadios clínicos ^{48,9}.

Más recientemente, se han propuesto soluciones basadas en arquitecturas *transformer* y modelos a gran escala. Por ejemplo, Yu *et al.* combinaron información imagenológica proveniente de MRI con información textual y genética para desarrollar un método de clasificación binaria en el contexto de una enfermedad neurodegenerativa ⁵². En el caso particular de la EP, Reyes *et al.* propusieron un enfoque basado en modelos de lenguaje para analizar cuestionarios autorreportados y apoyar tareas de clasificación diagnóstica ⁵³. Este tipo de enfoques evidencian que los modelos de lenguaje y las representaciones contextualizadas pueden aportar valor en escenarios clínicos donde la información textual es relevante.

A pesar de estos avances, el uso de enfoques multimodales que aprovechen simultánea-

⁵¹ Gunjan Pahuja y Bhanu Prasad. "Deep learning architectures for Parkinson's disease detection by using multi-modal features". En: *Computers in Biology and Medicine* 146 (2022), pág. 105610.

⁵² Qi Yu et al. "A transformer-based unified multimodal framework for Alzheimer's disease assessment". En: *Computers in Biology and Medicine* 180 (2024), pág. 108979.

⁵³ Diego Machado Reyes et al. "Language Modeling Screens Parkinson's Disease with Self-reported Questionnaires". En: *medRxiv* (2024).

mente la complementariedad entre información radiológica estructural e información clínica textual mediante modelos a gran escala sigue siendo limitado. En muchos trabajos, la información clínica se incorpora como variables tabulares o descriptores simplificados, lo que puede restringir la riqueza semántica del contexto clínico. Además, gran parte de la literatura se ha concentrado en tareas de clasificación binaria, dejando relativamente inexplorada la estratificación entre estadios de severidad.

En este sentido, aún existe campo de investigación en el desarrollo de estrategias multimodales que integren imágenes de resonancia magnética estructural con representaciones clínicas, especialmente bajo esquemas de alineación contrastiva que permitan construir espacios latentes más discriminativos para apoyar la estratificación de la enfermedad de Parkinson.

2. PROBLEMA DE INVESTIGACIÓN

La enfermedad de Parkinson es la segunda enfermedad neurodegenerativa más común a nivel mundial ⁵⁴, afectando a cerca de 11 millones de personas y se espera un incremento en su incidencia de un 20 % para el 2040 ⁵⁴. Aunque el origen de la EP es desconocido, se ha identificado que la pérdida progresiva de las neuronas dopaminérgicas causa las alteraciones motoras y cognitivas, constituyéndose en el principal biomarcador para el diagnóstico de la EP. Actualmente, el diagnóstico de la EP se basa en la valoración observacional del paciente a la hora de desarrollar tareas motoras y cognitivas, siguiendo escalas de referencia. Sin embargo, este análisis es altamente subjetivo, reportándose una tasa de diagnósticos errados cercana al 30 %. Además, este diagnóstico es tardío, cuando cerca del 80 % de las neuronas dopaminérgicas han sufrido algún daño ^{55,56}. En este sentido, las técnicas de neuroimagen han surgido como herramientas de análisis estructural y funcional del cerebro para soportar el diagnóstico. A pesar de esto, su análisis sigue siendo observacional registrando tasas de error cercanas al 25 % ⁵⁷. Por tanto, estrategias computacionales surgen como mecanismos de soporte clínico empleando estrategias que van desde lo clásico como algoritmos de *machine learning*, hasta aplicaciones de aprendizaje profundo logrando identificar y diferenciar de manera prometedora patrones relacionados a la enfermedad

⁵⁴ Jose A Obeso et al. "Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy". En: *Movement disorders* 32.9 (2017), págs. 1264-1310.

⁵⁵ M. C. Shih et al. "Parkinson's disease and dopamine transporter neuroimaging: a critical review". En: *Sao Paulo Medical Journal* 124 (2006), págs. 168-175.

⁵⁶ Irene Litvan et al. "Movement Disorders Society Scientific Issues Committee report: SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders." En: *Movement disorders: official journal of the Movement Disorder Society* 18.5 (2003), págs. 467-486.

⁵⁷ Sven R Suwijn et al. "The diagnostic accuracy of dopamine transporter SPECT imaging to detect nigrostriatal cell loss in patients with Parkinson's disease or clinically uncertain parkinsonism: a systematic review". En: *EJNMMI research* 5 (2015), págs. 1-8.

frente a poblaciones control de referencia ⁵⁸. A pesar de estos avances, dichos enfoques han sido ajustados sobre escenarios controlados, siendo propensos a sobreajustes durante su fase de entrenamiento. Adicionalmente, estos enfoques priorizan el aprendizaje sobre patrones radiológicos bien definidos, dejando de lado las valoraciones clínicas que podrían complementar la representación de la enfermedad. Dado lo anterior, surge como pregunta de investigación:

2.1. PREGUNTA DE INVESTIGACIÓN

¿Cómo integrar información clínica e imagenológica en un modelo a gran escala para soportar el diagnóstico de la enfermedad de Párkinson?

2.1.1. Hipótesis La integración de información clínica y radiológica puede mejorar la precisión de la caracterización de la enfermedad de Parkinson.

⁵⁸ R Prashanth et al. "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning". En: *International journal of medical informatics* 90 (2016), págs. 13-21.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Implementar un modelo de aprendizaje a gran escala para la clasificación de patrones parkinsonianos usando información radiológica y clínica.

3.2. OBJETIVOS ESPECÍFICOS

- Seleccionar una cohorte de pacientes con EP y sujetos control que cuenten con neuroimágenes MRI y SPECT junto con información clínica complementaria.
- Implementar un modelo a gran escala que permita capturar los hallazgos radiológicos MRI y SPECT para diferenciar entre poblaciones con la EP y controles de referencia.
- Implementar un modelo que desde la información clínica apoye la estratificación de los pacientes clasificados con Parkinson.
- Integrar la información radiológica y clínica complementaria para clasificar entre estadios de la EP y población control.
- Validar la capacidad del modelo a gran escala respecto a la tarea de clasificación de pacientes con la EP como un soporte al diagnóstico clínico.

4. ENFOQUE PROPUESTO

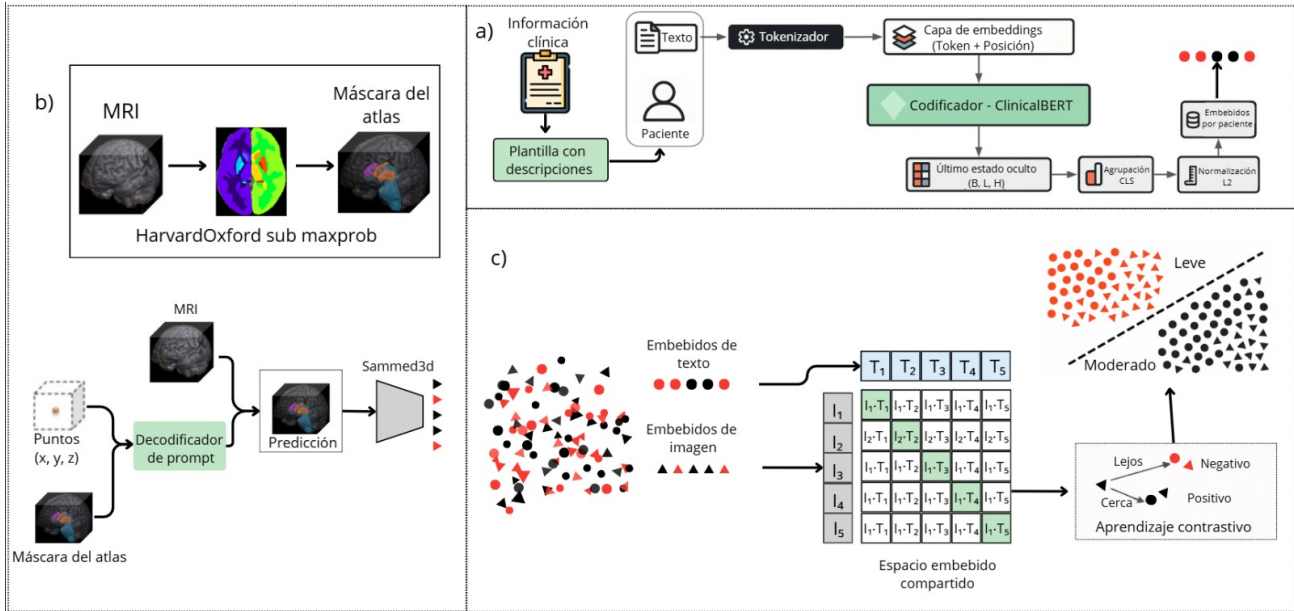
En este trabajo se implementó una representación profunda a gran escala para la clasificación de patrones parkinsonianos usando información radiológica y clínica. El enfoque propuesto integró información estructural obtenida a partir de volúmenes de MRI ponderada en T1 con información clínica derivada de escalas observacionales. El modelo multimodal proyecta ambas fuentes de información en un espacio embebido compartido, donde es posible capturar relaciones complementarias entre hallazgos radiológicos y manifestaciones clínicas.

En particular, la rama de imagen empleó regiones volumétricas seleccionadas mediante máscaras obtenidas de un atlas cerebral, las cuales se utilizaron como guía anatómica para extraer representaciones embebidas a partir de un modelo fundacional de segmentación. De forma paralela, la rama clínica codificó reportes construidos a partir de escalas motoras y variables demográficas, utilizando una representación basada en ClinicalBERT. Posteriormente, los pares multimodales imagen–texto correspondientes a una misma clase se alinearon mediante una estrategia de aprendizaje contrastivo supervisado, permitiendo la construcción de un descriptor conjunto que favoreció la separación entre estadios de severidad. El flujo general del método se resume en la Figura 5. En las siguientes subsecciones se detallan los pasos del método propuesto

4.1. INFORMACIÓN CLÍNICA CODIFICADA COMO DESCRIPTOR TEXTUAL

La caracterización inicial de la enfermedad de Parkinson se fundamenta en evaluaciones clínicas observacionales, escalas motoras estandarizadas y factores demográficos asociados al paciente. Aunque estas escalas por sí solas pueden presentar limitaciones diagnósticas, constituyen una fuente altamente informativa para complementar el análisis radiológico, especialmente en estadios tempranos donde los cambios estructurales pueden ser sutiles.

Figura 5. Visión general del flujo multimodal contrastivo propuesto para la estratificación de la EP, alineando *embeddings* estructurales de MRI y representaciones clínicas textuales en un espacio embebido compartido.



Con el fin de aprovechar esta información, en este trabajo se generaron reportes sintéticos por sujeto a partir de índices cuantificados obtenidos durante la valoración clínica. Un primer grupo de variables correspondió al examen motor MDS-UPDRS Parte III, el cual evaluó alteraciones relacionadas con el habla, rigidez, bradicinesia, marcha y temblor, además de un puntaje global. Un segundo grupo incluyó la escala de Hoehn & Yahr (NHY), ampliamente utilizada para estadificación clínica de la enfermedad²⁵. También se incorporó la escala Modified Bradykinesia Rating Scale (MBRS), que proporcionó una evaluación más detallada de componentes de bradicinesia durante tareas motoras estándar³⁸. Adicionalmente, se consideraron variables demográficas y relacionadas con el estado del tratamiento.

La Tabla 1 resume los grupos de variables utilizados en la construcción de los reportes clínicos.

A partir de estas variables, se generó un reporte clínico sintético por sujeto. Para ello, se utilizó una plantilla en la que los nombres clave de las columnas de los datos tabulares se

Tabla 1. Conjunto de variables clínicas utilizadas para generar los reportes clínicos. Las variables originales del registro tabular fueron organizadas por grupo clínico y asociadas a nombres descriptivos para su posterior incorporación en una plantilla textual estructurada.

Grupo	Ítems (nombres descriptivos)
MDS-UPDRS (Parte III)	Habla, Expresión facial, Rigidez (cuello, brazo derecho, brazo izquierdo, pierna derecha, pierna izquierda), golpeteo de dedos (mano derecha, mano izquierda), Agilidad de piernas (derecha, izquierda), Levantarse de la silla, Marcha, Congelamiento de la marcha, Estabilidad postural, Postura, Bradicinesia global, Temblor postural (mano derecha, mano izquierda), Temblor cinético (mano derecha, mano izquierda), Amplitud del temblor en reposo (brazo derecho, brazo izquierdo, pierna derecha, pierna izquierda, labios/mandíbula), Constancia del temblor, Puntaje total MDS-UPDRS III.
MBRS	Golpeteo de pie (izquierdo, derecho), Pronación–supinación (mano izquierda, mano derecha), Movimientos de manos (izquierda, derecha), Puntaje MBRS.
Ítems complementarios	Tratamiento para Parkinson, Estado funcional actual, Horas desde la última medicación para EP, Medición ON/OFF, Estimulación cerebral profunda, Estadio Hoehn & Yahr (NHY), Sexo, Edad en la visita, Definición de cohorte.

reemplazaron por sus descripciones y se combinaron con el valor correspondiente. De esta manera, cada sujeto quedó representado mediante un texto estructurado y dicho texto se codificó posteriormente con una variante del modelo fundacional BERT llamado ClinicalBERT para obtener una representación contextualizada del registro clínico.

Particularmente, la arquitectura ClinicalBERT adapta el preentrenamiento al dominio clínico utilizando notas médicas y terminología especializada, lo que permite obtener representaciones contextualizadas más adecuadas para tareas sobre texto clínico⁵⁹. En este trabajo se utilizó el estado oculto final del token [CLS] del ClinicalBERT como un *embedding* global por sujeto f_T , seguido de normalización ℓ_2 , generando así el descriptor clínico final.

Esta representación permitió incorporar información semántica clínica que posteriormente guió y complementó el análisis de patrones estructurales en MRI.

⁵⁹ Kexin Huang, Jaan Altosaar y Rajesh Ranganath. “ClinicalBERT: Modeling clinical notes and predicting hospital readmission”. En: *arXiv preprint arXiv:1904.05342* (2019).

4.2. SECUENCIAS RADIOLÓGICAS COMO DESCRIPTOR VISUAL

Las secuencias estructurales de MRI ponderadas en T1 constituyen una herramienta ampliamente disponible en la práctica clínica, ya que permite analizar regiones cerebrales relevantes en el circuito motor afectado por la enfermedad de Parkinson. No obstante, los cambios estructurales asociados a la enfermedad suelen ser heterogéneos y difíciles de identificar visualmente, especialmente en estadios tempranos.

En este trabajo, se empleó una estrategia basada en modelos fundacionales para extraer descriptores compactos a partir de regiones anatómicas de interés. Para ello, se utilizó la parcelación del atlas *HarvardOxford sub maxprob thr25 1mm*, selecting regiones clave como el putamen, el caudado, el tálamo y el tronco encefálico. Estas estructuras son relevantes debido a su papel en la vía nigroestriatal dopaminérgica y en los circuitos estriado–talámico–corticales⁶⁰, así como en estadios neuropatológicos tempranos descritos en EP⁶¹. Las máscaras obtenidas a partir del atlas actuaron como *prompts* anatómicos para un modelo fundacional 3D basado en SamMed3D⁴⁷.

Particularmente, el modelo fundacional SamMed3D fue ajustado previamente con la capacidad de realizar múltiples segmentaciones en diferentes secuencias y múltiples dominios⁴⁶. En su arquitectura se reconocen el codificador de imagen 3D, definida como $V \in \mathbb{R}^{H \times W \times D}$, donde H , W y D representan la altura, el ancho y la profundidad del volumen, respectivamente. Este encoder retorna un vector embebido $f \in \mathbb{R}^{h \times w \times d \times c}$, en el cual las dimensiones espaciales h , w y d se obtienen a partir de diferentes bloques de atención; además, incorpora un codificador de *prompts* que emplea puntos de referencia, los cuales provienen de las máscaras extraídas desde atlas, tales como el putamen, el caudado, el tálamo y el tronco encefá-

⁶⁰ Jose A Obeso et al. "The basal ganglia in Parkinson's disease: current concepts and unexplained observations". En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 64.S2 (2008), S30-S46.

⁶¹ Heiko Braak et al. "Staging of brain pathology related to sporadic Parkinson's disease". En: *Neurobiology of aging* 24.2 (2003), págs. 197-211.

lico. Cada máscara se convierte en un embebido e_p . Estos embeddings e_p se combinan con el embedding volumétrico f para formar una representación unificada $z = \text{Combine}(f, e_p)$. En este trabajo, los embeddings $f \in \mathbb{R}^{h \times w \times d \times c}$ de cada región de interés del atlas, para cada imagen particular, se obtuvieron desde el codificador de imagen f_I (características volumétricas generales), durante la fusión con el codificador de *prompts* f_P , y en el decodificador de máscaras f_M , donde se incorporaron restricciones semánticas de la tarea. Estas representaciones permitieron capturar patrones locales asociados a la progresión de la enfermedad.

4.3. APRENDIZAJE MULTIMODAL MEDIANTE UNA REGLA CONTRASTIVA

Una vez obtenidos los *embeddings* clínicos y radiológicos, ambas modalidades fueron proyectadas en un espacio latente compartido mediante funciones de codificación aprendidas. En este espacio, se define como pares positivos aquellas representaciones correspondientes a una misma clase, incluso cuando provienen de modalidades distintas, mientras que representaciones de sujetos diferentes se consideran pares negativos.

Para ello, se empleó una función de pérdida contrastiva supervisada basada en similitud coseno, definida en la Ec. 1 ⁶². Esta formulación optimiza la cercanía entre pares positivos en relación con el resto de muestras del lote, promoviendo simultáneamente la separación respecto a pares negativos.

$$\mathcal{L}_{\text{out}}^{\text{sup}} = \sum_{i \in \mathcal{I}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(\frac{z_i \cdot z_p}{\tau}\right)}{\sum_{a \in A(i)} \exp\left(\frac{z_i \cdot z_a}{\tau}\right)} \quad (1)$$

Donde $P(i)$ representa el conjunto de muestras positivas para el ancla i , $A(i)$ corresponde al resto de muestras en el lote incluyendo tanto pares positivos como negativos, y τ es el parámetro de temperatura. Asimismo, z_i denota la representación embebida del ancla, mientras que z_p y z_a corresponden a las representaciones de las muestras positivas y

⁶² Prannay Khosla et al. "Supervised contrastive learning". En: *Advances in neural information processing systems* 33 (2020), págs. 18661-18673.

del conjunto total considerado en el denominador, respectivamente. Estas representaciones son normalizadas, de modo que el producto punto utilizado en la formulación equivale a la similitud coseno. De esta forma, el modelo logró correlacionar patrones estructurales locales presentes en MRI con información clínica semántica derivada de escalas observacionales, construyendo un espacio multimodal discriminativo para la estratificación de la EP.

En secciones previas, las representaciones embebidas de las modalidades fueron denotadas como f_I , f_P y f_M para imágenes y f_T para datos clínicos. En el contexto de la formulación de la pérdida contrastiva supervisada (Ec. 1), estas representaciones se unifican bajo la notación general z , de modo que z_i representa el embedding correspondiente a una muestra ancla.

Como resultado, el modelo aprende una representación multimodal alineada en la que la información estructural derivada de MRI y descriptores clínicos se integran en un mismo espacio, favoreciendo la discriminación entre distintos estadios de severidad clínica ya que a diferencia de otros enfoques basados en clasificación directa, esta formulación permite aprovechar relaciones relativas entre múltiples muestras dentro de cada lote, enriqueciendo la estructura del espacio latente.

4.4. CONFIGURACIÓN EXPERIMENTAL

4.4.1. Los datos. La información neurológica y clínica utilizada en este trabajo fue extraída de la base de datos Parkinson's Progression Markers Initiative (PPMI), un estudio multicéntrico, longitudinal y de acceso abierto orientado a la identificación de biomarcadores de progresión en la enfermedad de Parkinson (<https://www.ppmi-info.org/>).

A partir de esta cohorte, se recolectaron volúmenes de resonancia magnética estructural ponderada en T1, los cuales fueron preprocesados con el fin de garantizar consistencia anatómica entre sujetos. Este preprocesamiento incluyó un procedimiento de *skull stripping* para eliminar tejido no cerebral, un remuestreo isotrópico a una resolución de 1 mm^3 , y un suavizado espacial para reducir ruido de alta frecuencia.

Como referencia comparativa, también se incluyó un subconjunto de sujetos con imágenes funcionales SPECT. En este caso, los volúmenes funcionales fueron co-registrados con la MRI estructural correspondiente, con el objetivo de asegurar alineación anatómica entre ambas modalidades. Posteriormente, se emplearon parcelaciones cerebrales basadas en atlas para apoyar análisis regionales y la selección de estructuras relevantes.

Por otra parte, la información clínica fue incorporada a partir de variables estructuradas obtenidas durante la valoración médica y convertida en descripciones textuales cortas por sujeto, siguiendo la estrategia descrita en la Sección anterior.

La cohorte final considerada para el experimento estructural incluyó un total de 184 sujetos, distribuidos en 127 casos clasificados como *leves* y 57 como *moderados*. La Tabla 2 resume las características demográficas y clínicas principales de la población analizada.

Tabla 2. Descripción del conjunto de datos: grupos leve vs. moderado definidos por etiquetas derivadas de MBRS.

Variable	Leve (127)	Moderado (57)
train/test	99 / 28	44 / 13
Edad (media±desv.)	66.98 ± 9.61	65.59 ± 9.39
NP3TOT (media±desv.)	17.75 ± 9.53	37.07 ± 11.04
NHY (0/1/2)	0 / 27 / 82	0 / 2 / 43
NHY (3/4/5)	14 / 3 / 1	6 / 5 / 1
Género (F/M)	47 / 80	23 / 34

4.4.2. Configuración de la arquitectura. El modelo propuesto integra información multimodal a partir de imágenes de resonancia magnética (MRI) y variables clínicas estructuradas. La arquitectura se compone de tres módulos principales: (i) preprocesamiento de imágenes, (ii) codificación clínica y (iii) fusión multimodal contrastiva.

- **Preprocesamiento de imágenes (MRI)**

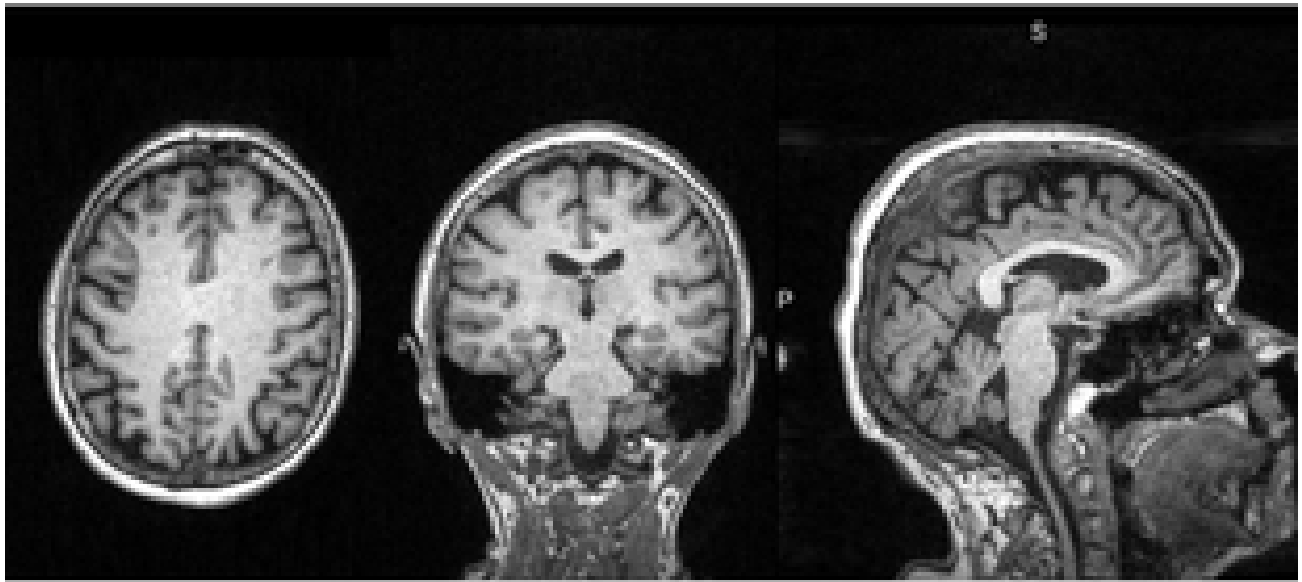
Para el procesamiento de las imágenes MRI se siguió una secuencia de pasos con el objetivo de normalizar y preparar los volúmenes para su uso en el modelo. Este proceso incluye remuestreo, extracción de cráneo, corrección de intensidades, registro

a un espacio de referencia común y parcelación anatómica mediante atlas.

- **Resampling:** Para normalizar el tamaño de cada voxel de la imagen a 1mm^3 se hizo un *resampling* o reescalado, de tal forma que los voxeles tengan la misma dimensión en todas las direcciones a nivel espacial. Para ello, la imagen se carga utilizando la librería *nibabel* y se aplica una interpolación trilineal, la cual consiste en estimar nuevos valores a partir de los vóxeles vecinos más cercanos. Este proceso permite homogeneizar la resolución espacial de los volúmenes antes de las etapas posteriores de alineación y análisis.
- **Skull stripping:** El *skull stripping* o extracción de cráneo consiste en eliminar la zona exterior de la imagen correspondiente a la estructura ósea, conservando únicamente el tejido cerebral. En la Figura 6 se muestra una comparación antes y después de este procedimiento, evidenciando la supresión de estructuras no cerebrales en vistas anatómicas representativas.
- **Bias field correction:** Para homogeneizar los valores de intensidad de la imagen, se usa el *bias field correction*, ya que permite reducir variaciones no deseadas. Para esto, se utiliza el algoritmo *N4 Bias Field Correction*, el cual está implementado en la librería *SimpleITK*. Como se observa en la Figura 7, este paso reduce diferencias suaves de intensidad y mejora la consistencia visual del tejido cerebral.
- **Registro:** Se realizó un registro que consiste en la alineación de cada imagen a un *template* en espacio MNI, con el objetivo de que todas las imágenes tengan la misma referencia espacial y este paso se realizó utilizando la librería *ants*. La Figura 8 ilustra el resultado del alineamiento al espacio de referencia, permitiendo que las estructuras anatómicas queden en posiciones comparables entre sujetos.
- **Parcelación con atlas:** Finalmente, se realizó una parcelación anatómica del volumen cerebral mediante un atlas probabilístico. En este trabajo se empleó la librería *FSL* junto con el atlas Harvard–Oxford (*sub-maxprob-thr25-1mm*) para obtener

Figura 6. Comparación antes y después del proceso de extracción de cráneo, mostrando la eliminación del tejido no cerebral y la conservación de las estructuras encefálicas relevantes para el análisis.

Antes



Después

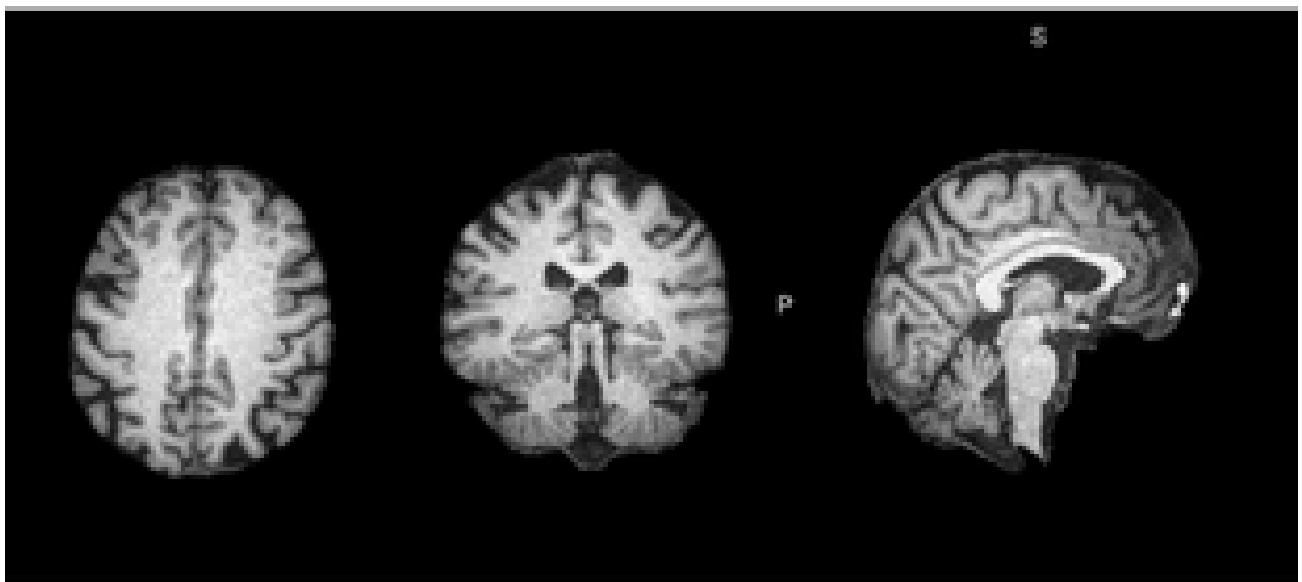


Figura 7. Comparación antes y después de la corrección de intensidades mediante N4 Bias Field Correction, resaltando la reducción de inhomogeneidades y la mejora en la uniformidad del contraste tisular.

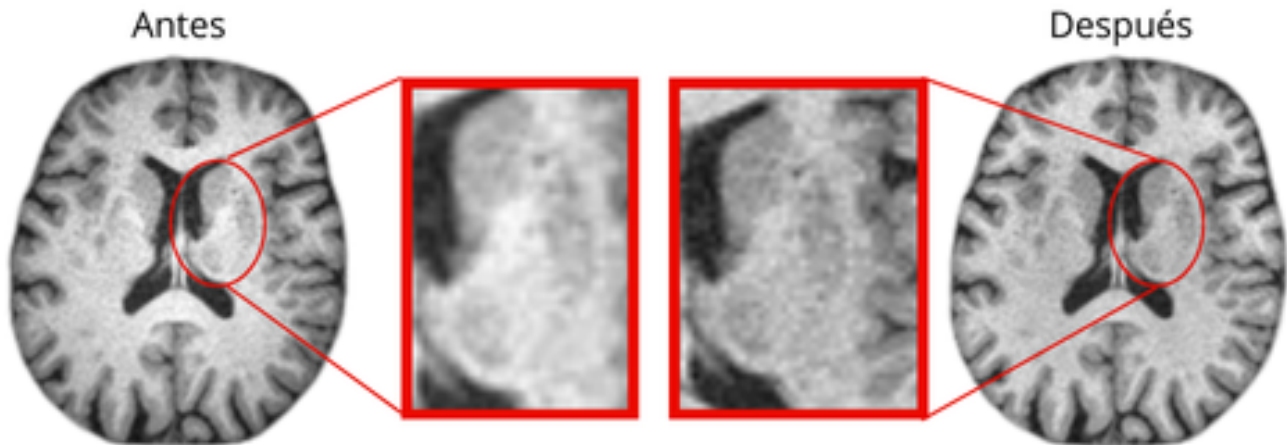
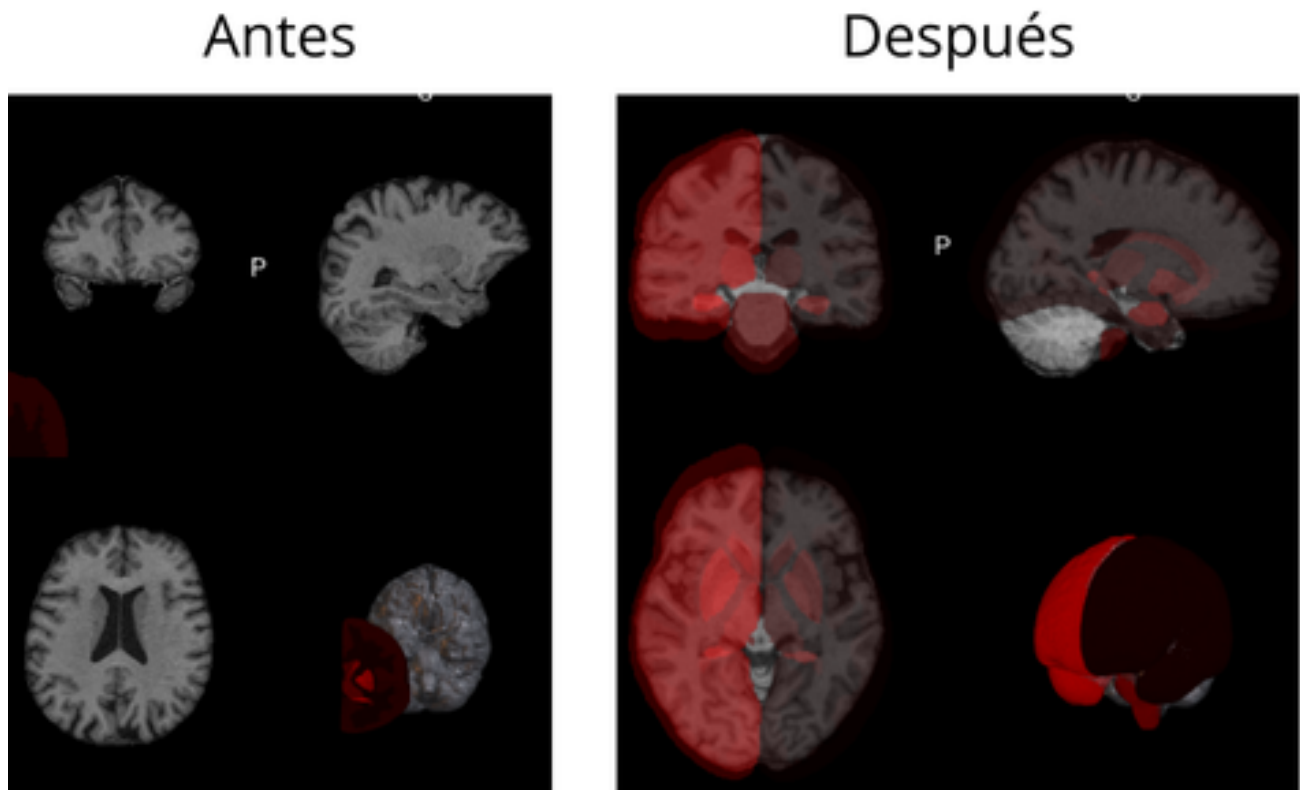


Figura 8. Alineación de la imagen MRI al espacio de referencia MNI. Se muestra la correspondencia espacial obtenida tras el registro, la cual permite comparar regiones anatómicas entre sujetos en un sistema común.



las regiones de interés utilizadas en el modelo.

Como se muestra en la Figura 9, el atlas define una parcelación del cerebro en múltiples regiones anatómicas en el espacio MNI. La visualización incluye cortes ortogonales (axial, coronal y sagital), así como una representación tridimensional del volumen cerebral, permitiendo observar la distribución espacial de las diferentes estructuras.

A partir de este atlas, se seleccionaron un total de cuatro regiones anatómicas: el núcleo caudado, el putamen, el tálamo y el tronco encefálico, por su relevancia en los circuitos anatómicos y funcionales asociados con la enfermedad de Parkinson. Estas máscaras fueron utilizadas posteriormente como guías anatómicas dentro del modelo.

En la Figura 10 se presenta una visualización de las regiones seleccionadas sobre el volumen cerebral. La figura muestra, de izquierda a derecha y de arriba hacia abajo, el cerebro completo como referencia anatómica, la máscara del núcleo caudado, la máscara del tronco encefálico, la máscara del putamen, la máscara del tálamo y una vista conjunta con la superposición de todas las regiones de interés. Esta representación permite identificar de manera explícita la localización anatómica de cada estructura y justificar su uso como soporte espacial en la extracción de características.

- **Modelo fundacional (SamMed 3D).** Durante el proceso de *fine-tuning* del modelo SAM-Med3D, se empleó la configuración experimental por defecto definida en la implementación original.

En particular, se utilizó un esquema de optimización con decaimiento de la tasa de aprendizaje tipo *MultiStep*, donde la tasa inicial se fijó en 8×10^{-4} y se redujo por un factor de 0.1 en las épocas 120 y 180. El entrenamiento se llevó a cabo durante un total de 200 épocas.

Figura 9. Visualización del atlas Harvard–Oxford (*sub-maxprob-thr25-1mm*) con vistas ortogonales (axial, coronal y sagital) junto con una representación tridimensional del volumen cerebral, ilustrando la parcelación anatómica en múltiples regiones.

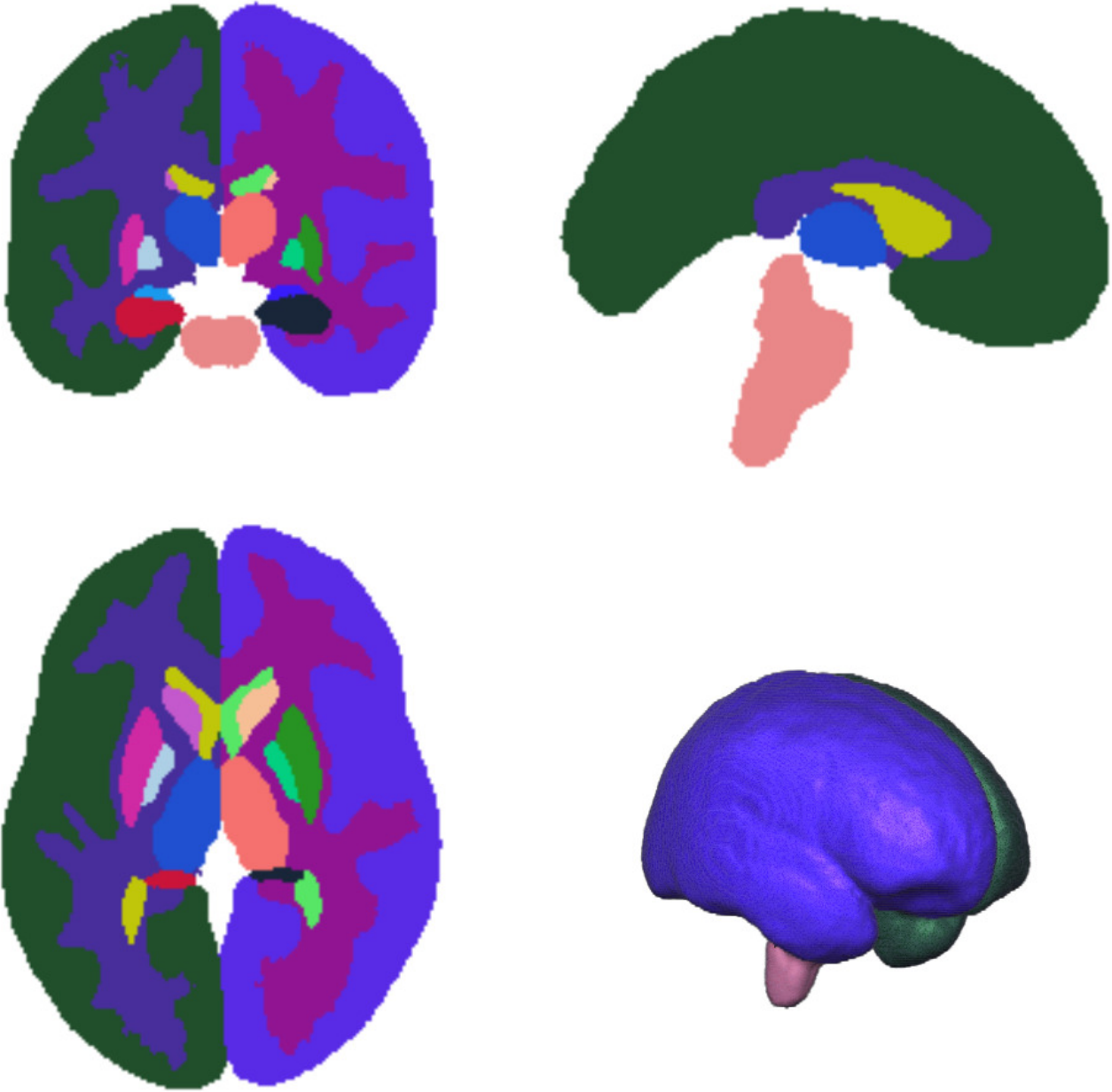
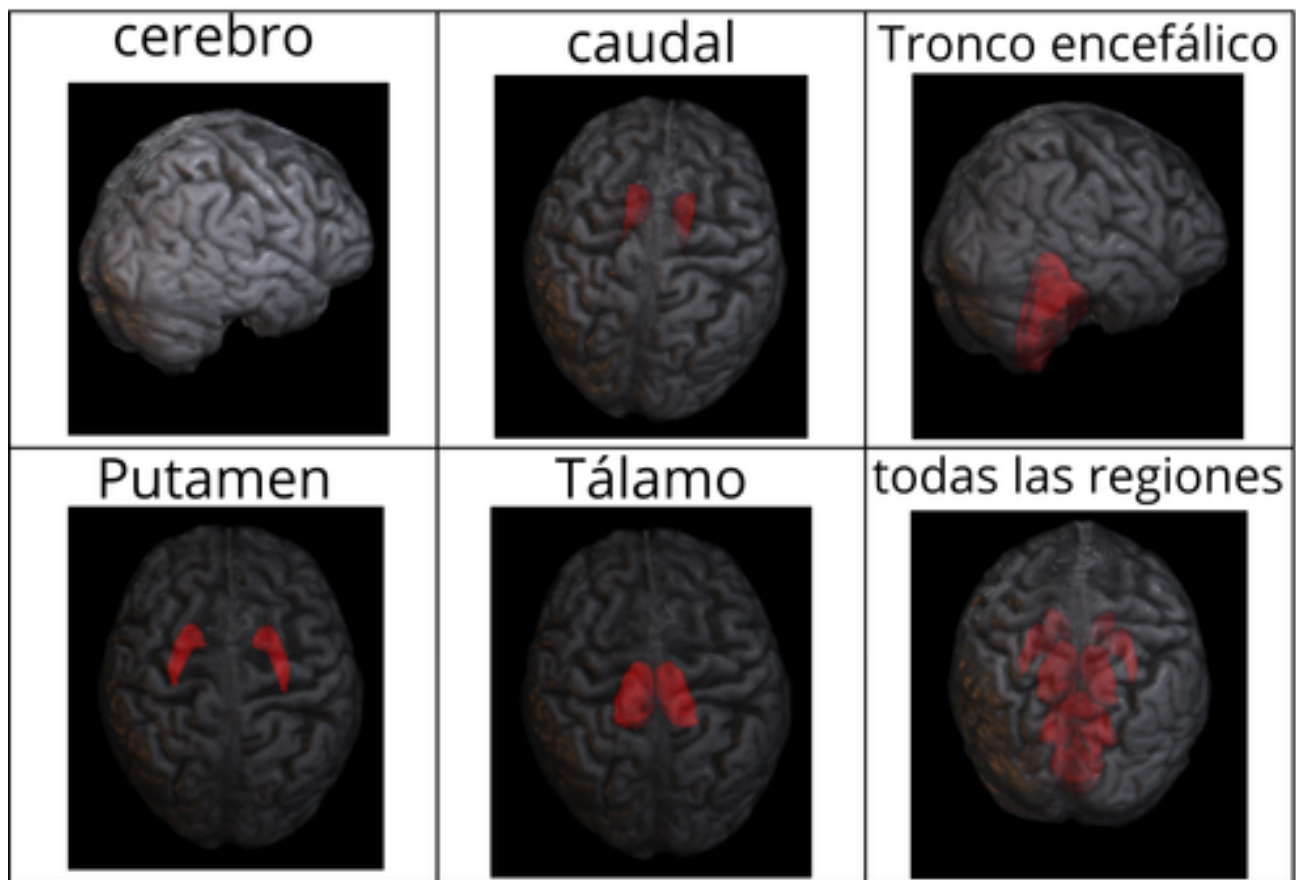


Figura 10. Visualización de las regiones anatómicas seleccionadas a partir del atlas Harvard–Oxford (*sub-maxprob-thr25-1mm*). Se muestran las máscaras correspondientes al núcleo caudado, putamen, tálamo y tronco encefálico, así como una vista de su superposición sobre el volumen cerebral completo. Estas cuatro regiones de interés fueron utilizadas como guías anatómicas para la extracción de representaciones en el modelo.



Las imágenes volumétricas fueron redimensionadas a un tamaño de 128^3 , permitiendo un balance entre resolución espacial y consumo de memoria. Se utilizó también un tamaño de lote de 12 muestras; sin embargo, para mitigar limitaciones de memoria, se implementó una estrategia de acumulación de gradientes con 20 pasos, incrementando efectivamente el tamaño de lote sin aumentar el uso de memoria de manera directa.

El proceso de optimización se realizó con una tasa de regularización (weight decay) de 0.1, favoreciendo la generalización del modelo.

- **Codificación clínica (ClinicalBERT).** Para la codificación de la información clínica se utilizó el modelo ClinicalBERT, basado en la arquitectura BERT con aproximadamente 110 millones de parámetros, preentrenado sobre textos clínicos.

Las variables clínicas estructuradas se transformaron en una descripción textual por sujeto (*prompt*), la cual fue utilizada como entrada al modelo. Cada texto fue tokenizado y se extrajo el embedding asociado al token [CLS], obteniendo una representación global de dimensión 768. Finalmente, cada embedding fue normalizado para su posterior fusión en el espacio multimodal.

- **Fusión contrastiva multimodal.** Para fusionar las modalidades, se realizó una proyección de los embeddings clínicos y de imagen a un espacio compartido mediante capas completamente conectadas.

En este espacio se empleó una función de pérdida contrastiva basada en similitud coseno, siguiendo un esquema tipo CLIP. Se construyó una matriz de similitud entre embeddings de texto e imagen mediante producto punto normalizado.

Los pares positivos corresponden a las representaciones de texto e imagen de la misma clase, mientras que el resto de combinaciones se consideran pares negativos. La función de pérdida se formuló como una entropía cruzada sobre estas similitudes.

El entrenamiento se realizó mediante el optimizador Adam utilizando aprendizaje por lotes. Adicionalmente, se incorporó una pérdida de clasificación supervisada sobre las

etiquetas clínicas, combinada con la pérdida contrastiva mediante un factor de ponderación.

5. EVALUACIÓN Y RESULTADOS

En este capítulo se presentan los principales resultados obtenidos con el enfoque propuesto, que incluye la validación necesaria para analizar el aporte de cada componente del modelo multimodal.

5.1. SELECCIÓN DE REPRESENTACIONES VISUALES EN SAMMED3D

En primer lugar, se realizó un estudio de los componentes del procesamiento visual del enfoque con el fin de identificar qué nivel de representación extraída desde la arquitectura SamMed3D resulta más discriminativo para la tarea de estratificación. En particular, se evaluaron tres puntos de extracción de *embeddings*: (1) una representación temprana obtenida desde el codificador de imagen llamada p_I , (2) una representación intermedia posterior a la etapa de fusión con los *prompts* llamada f_P , y (3) una representación de alto nivel semántico obtenida desde el decodificador de máscaras llamada f_M .

La Tabla 3 resume los resultados alcanzados. Como puede observarse, los *embeddings* extraídos en etapas tempranas del codificador preservaron información estructural y resultaron más efectivos para la clasificación, alcanzando el mejor desempeño global.

Tabla 3. Selección de *embedding* visual

Experimento	Prec.	Rec.	F1	Acc.
Codificación temprana f_I	0.83	0.83	0.83	0.85
Codificación intermedia f_P	0.74	0.73	0.74	0.78
Alto nivel semántico f_M	0.81	0.80	0.80	0.80

En este análisis, uno de los hallazgos más relevantes fue que las características extraídas en las primeras capas del modelo SamMed3D fueron las más útiles para la clasificación. Esto indica que, en el contexto de la enfermedad de Parkinson, los cambios estructurales del cerebro son muy sutiles, y se capturan mejor en etapas tempranas del procesamiento,

antes de que el modelo se enfoque en tareas más generales como segmentación.

En la figura 11 se presentan cuatro ejemplos representativos que ilustran el comportamiento del modelo a nivel individual. Se muestran dos pacientes en estadio leve correctamente clasificados, un paciente en estadio moderado correctamente clasificado y un paciente en estadio moderado clasificado erróneamente como leve.

5.2. COMPARACIÓN DE ESTRATEGIAS DE FUSIÓN MULTIMODAL

Posteriormente, se compararon distintas estrategias de integración multimodal. En particular, se evaluó una línea base supervisada basada en la concatenación directa de *embeddings* clínicos y radiológicos, entrenada mediante entropía cruzada estándar, frente al enfoque contrastivo propuesto.

En el escenario con MRI, la Tabla 4 muestra una diferencia notable entre ambos esquemas. Mientras que la línea base presentó un desempeño limitado, el entrenamiento mediante pérdida contrastiva supervisada permitió una mejora significativa en precisión, recall, F1 macro-promediado y exactitud.

Tabla 4. Métricas de clasificación para MRI: línea base multimodal vs. enfoque propuesto (SCL).

Modelo	Prec.	Rec.	F1	Acc.
Línea base multimodal	0.47	0.36	0.58	0.30
Enfoque propuesto (SCL)	0.83	0.83	0.83	0.85

En adición a lo anterior, el enfoque propuesto fue evaluado utilizando secuencias funcionales SPECT como modalidad radiológica alternativa con el propósito de analizar si la estrategia de alineación multimodal propuesta conserva su capacidad discriminativa cuando se sustituye la MRI estructural por una modalidad funcional más directamente relacionada con la integridad dopaminérgica.

Los resultados reportados en la Tabla 5 muestran que el modelo contrastivo también mejoró el desempeño en este escenario, alcanzando métricas balanceadas.

Figura 11. Ejemplos representativos de predicción del modelo. Se muestran cuatro pacientes del conjunto de prueba: (i) un falso positivo correspondiente a un paciente en estadio moderado clasificado erróneamente como leve, (ii) un verdadero positivo correspondiente a un paciente en estadio leve clasificado correctamente, (iii) un verdadero positivo correspondiente a un paciente en estadio leve clasificado correctamente y (iv) un verdadero negativo correspondiente a un paciente en estadio moderado clasificado correctamente. Cada fila corresponde a la vista axial, la reconstrucción tridimensional, las regiones segmentadas y un resumen textual de la información clínica.





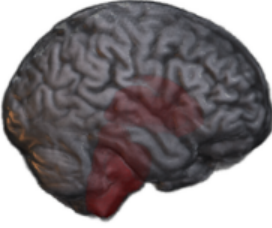
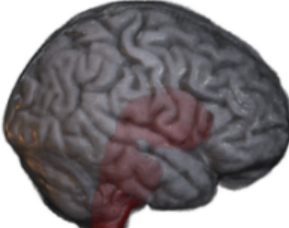
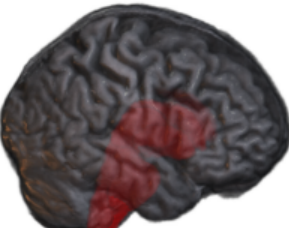
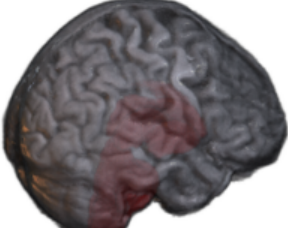
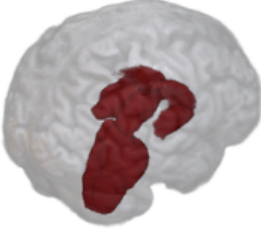
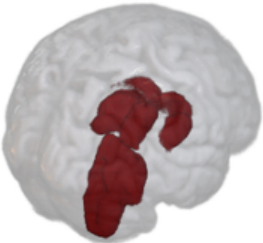
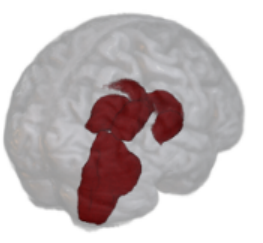
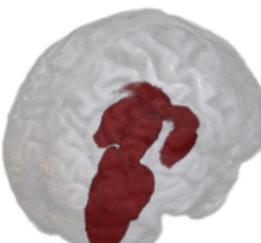
Falso positivo (3557)	Verdadero positivo (3108)	Verdadero positivo (3311)	Verdadero negativo (3128)
			
			
			
Parkinson's patient. PATNO: 3557. speech impairment: 3.0. neck rigidity: 2.0. right foot tapping: 3.0. left foot tapping: 3.0....	Parkinson's patient. PATNO: 3108. speech impairment: 0.0. neck rigidity: 0.0. right foot tapping: 0.0. left foot tapping: 1.0.	Parkinson's patient. PATNO: 3311. speech impairment: 1.0. neck rigidity: 1.0. right foot tapping: 1.0. left foot tapping: 0.0...	Parkinson's patient. PATNO: 3128. speech impairment: 1.0. neck rigidity: 2.0. right foot tapping: 2.0. left foot tapping: 2.0...

Tabla 5. Métricas de clasificación para SPECT: línea base multimodal vs. enfoque propuesto (SCL).

Modelo	Prec.	Rec.	F1	Acc.
Línea base multimodal	0.34	0.50	0.40	0.68
Enfoque propuesto (SCL)	0.79	0.73	0.75	0.80

Cabe resaltar que el desempeño obtenido con MRI estructural resultó comparable e incluso superior al alcanzado con SPECT, lo cual es particularmente relevante desde el punto de vista clínico, dado que la MRI es una modalidad ampliamente disponible, mientras que SPECT se restringe a centros especializados. Aunque SPECT es una modalidad altamente sensible para detectar alteraciones dopaminérgicas asociadas con la enfermedad de Parkinson, el enfoque propuesto en este trabajo se apoyó en regiones anatómicas de interés definidas mediante atlas y en un esquema de alineación espacial que favorece la caracterización estructural del cerebro. Bajo estas condiciones, la MRI estructural parece ofrecer una base más consistente para la obtención de descriptores discriminativos.

5.3. ANÁLISIS DE DISCRIMACIÓN EN ESTADIOS NO OBSERVADOS

Como análisis complementario, se proyectaron muestras correspondientes a estadios avanzados de la enfermedad dentro de la arquitectura binaria entrenada únicamente con casos leves y moderados. Este experimento permitió explorar propiedades de confianza y generalización bajo escenarios con datos limitados.

Para complementar la interpretación visual de la Figura 12, se aplicó una prueba estadística de Mann–Whitney U entre las clases leve y grave. El objetivo fue determinar si las diferencias observadas entre los niveles de severidad clínica eran estadísticamente significativas. En este contexto, un valor de $p = 0,0002$ indica que la diferencia entre los grupos comparados es improbable que se deba al azar, lo que sugiere que la confianza del modelo cambia de manera consistente con el nivel de severidad clínica. Esta prueba fue seleccionada debido a que no asume normalidad en los datos y es adecuada para comparar distribuciones entre

futuros que analicen la evolución del Parkinson a lo largo del tiempo.

5.4. ESTRATEGIAS DE FUSIÓN

Asimismo, se evaluaron dos estrategias de fusión a nivel de decisión basadas en la combinación de las predicciones de múltiples modelos unimodales, con el fin de comparar su desempeño frente al método propuesto basado en alineación contrastiva multimodal. La primera estrategia corresponde a un esquema de *majority voting* mediante promedio simple de probabilidades *soft voting*, en el cual cada modelo aporta el mismo peso a la decisión final. Este enfoque alcanzó una exactitud de 0.60, mostrando un desempeño limitado especialmente en la detección de la clase positiva.

La segunda estrategia introduce una fusión ponderada entre modalidades, asignando un 50 % del peso a las predicciones derivadas del texto y el 50 % restante a la información proveniente de imágenes (promediada entre las distintas regiones anatómicas). Este enfoque logró una exactitud de 0.80 y evidenció una mejora consistente en métricas como precisión, recall y F1 macro-promediado.

Tabla 6. Resultados de estrategias de fusión a nivel de decisión y comparación con el método propuesto.

Estrategia	Prec.	Rec.	F1	Acc.
<i>Majority voting</i>	0.48	0.49	0.44	0.60
Fusión ponderada	0.82	0.75	0.76	0.80
Método propuesto (SCL + MRI)	0.83	0.83	0.83	0.85

As se observa en la Tabla 6, el método propuesto superó a ambas estrategias de fusión a nivel de decisión. En particular, alcanzó una exactitud de 0.8537, superior al 0.80 obtenido con la fusión ponderada y al 0.60 del esquema de *majority voting*. Estos resultados sugieren que la mejora no proviene únicamente de combinar predicciones finales, sino de aprender una representación conjunta entre la información clínica y radiológica. Esto favorece una mejor integración entre modalidades y una separación más clara entre los distintos estadios

de severidad.

5.5. EVALUACIÓN DEL COMPONENTE CLÍNICO TEXTUAL

En esta consideración, se evaluó el impacto de diferentes componentes del reporte clínico mediante un estudio de ablación textual. En particular, se consideraron dos variantes: (i) eliminando las variables asociadas a MBRS y (ii) eliminando los ítems del examen motor MDS-UPDRS Parte III.

Los resultados presentados en la Tabla 7 muestran una reducción significativa en el desempeño al remover cualquiera de estos componentes, lo que comprueba que las escalas motoras aportan información complementaria clave para guiar la alineación multimodal y mejorar la estratificación.

Tabla 7. Métricas de clasificación en el conjunto de prueba para el enfoque propuesto y ablaciones del texto clínico (macro-promediadas).

Experimento	Prec.	Rec.	F1	Acc.
Enfoque propuesto	0.83	0.83	0.83	0.85
Sin MBRS	0.62	0.62	0.60	0.61
Sin MDS-UPDRS (Parte III)	0.56	0.56	0.56	0.58

La reducción observada al eliminar cualquiera de estos componentes confirma la importancia de incorporar escalas motoras dentro de la representación multimodal, especialmente la eliminación de variables asociadas a MBRS o de los ítems del examen motor MDS-UPDRS Parte III produjo una caída notable en el desempeño del modelo.

Desde el punto de vista clínico, estos resultados sugieren que ambas escalas aportan información complementaria. La MBRS contribuye con una caracterización más sensible de alteraciones motoras finas, particularmente asociadas a la bradicinesia, mientras que la MDS-UPDRS Parte III proporciona una evaluación más global del estado motor del paciente. Estos resultados demuestran que incorporar información clínica junto con neuroimágenes permite reflejar de manera más explícita cómo se manifiestan los síntomas en el paciente.

Tabla 8. Aporte clínico de los componentes utilizados en el descriptor textual.

Componente	Información que aporta	Relevancia para la estratificación
MBRS	Evalúa bradicinesia a partir de velocidad, amplitud y ritmo en tareas motoras específicas.	Aporta una medida más precisa de la afectación motora, especialmente útil para distinguir cambios sutiles entre niveles de severidad.
MDS-UPDRS Parte III	Resume el examen motor global, incluyendo rigidez, temblor, marcha, postura y bradicinesia.	Proporciona una visión más amplia del compromiso clínico, permitiendo contextualizar la severidad motora general del paciente.

6. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se propuso un marco multimodal basado en aprendizaje contrastivo para la estratificación de la enfermedad de Parkinson, el cual integra representaciones radiológicas obtenidas mediante un modelo fundacional 3D (SamMed3D) con representaciones clínicas derivadas de descripciones textuales codificadas con ClinicalBERT.

Los resultados obtenidos demuestran que el enfoque propuesto permite una mejor alineación entre la información estructural y clínica, superando el desempeño de una línea base supervisada basada en concatenación directa de características. En particular, el modelo propuesto logra mitigar el sesgo hacia la clase mayoritaria y recuperar capacidad discriminativa en la clase minoritaria, evidenciando un comportamiento más balanceado en la tarea de clasificación.

Los experimentos realizados sobre la cohorte PPMI mostraron una exactitud de 0.85 y mejoras superiores al 30 % frente a enfoques unimodales, lo cual respalda la capacidad del modelo para capturar información complementaria entre escalas clínicas e imágenes estructurales. Estos hallazgos sugieren que la alineación contrastiva multimodal constituye una estrategia efectiva para generar representaciones más discriminativas y clínicamente relevantes en la evaluación de severidad de la EP. Además, los análisis de fusión muestran que la forma en que se combinan las modalidades influye de manera importante en el desempeño del modelo.

Como trabajo futuro, se plantea la evaluación del modelo en cohortes más amplias y heterogéneas, así como la validación externa en datos provenientes de otros centros clínicos y la incorporación de información longitudinal para estudiar de manera explícita la progresión de la enfermedad a lo largo del tiempo. Además, se propone profundizar en el análisis de interpretabilidad del modelo, analizando cómo determinados patrones estructurales y variables clínicas influyen en la decisión final de clasificación.

BIBLIOGRAFÍA

- Aludin, Schekeb y Lars-Patrick Andreas Schmill. "MRI signs of Parkinson's disease and atypical parkinsonism". En: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. Vol. 193. 12. Georg Thieme Verlag KG. 2021, págs. 1403-1410 (vid. pág. 12).
- Baltrušaitis, Tadas, Chaitanya Ahuja y Louis-Philippe Morency. "Multimodal machine learning: A survey and taxonomy". En: *IEEE transactions on pattern analysis and machine intelligence* 41.2 (2018), págs. 423-443 (vid. págs. 21, 22).
- Bommasani, Rishi et al. "On the opportunities and risks of foundation models". En: *arXiv preprint arXiv:2108.07258* (2021) (vid. pág. 23).
- Braak, Heiko et al. "Staging of brain pathology related to sporadic Parkinson's disease". En: *Neurobiology of aging* 24.2 (2003), págs. 197-211 (vid. pág. 36).
- Camacho, Milton et al. "Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets". En: *NeuroImage: Clinical* 38 (2023), pág. 103405 (vid. pág. 13).
- Chen, Ting et al. "A Simple Framework for Contrastive Learning of Visual Representations". En: *arXiv preprint arXiv:2002.05709* (2020) (vid. pág. 22).
- Chougar, Lydia et al. "The role of magnetic resonance imaging for the diagnosis of atypical parkinsonism". En: *Frontiers in Neurology* 11 (2020), pág. 665 (vid. pág. 12).
- Cohenpour, Mehrzad y Haim Golan. "Nuclear neuroimaging of dopamine transporter in Parkinsonism—role in routine clinical practice". En: *Harefuah* 146.9 (2007), págs. 698-702 (vid. pág. 17).

- Dauer, William y Serge Przedborski. "Parkinson's disease: mechanisms and models". En: *Neuron* 39.6 (2003), págs. 889-909 (vid. págs. 11, 15).
- Dentamaro, Vincenzo et al. "Enhancing early Parkinson's disease detection through multi-modal deep learning and explainable AI: insights from the PPMI database". En: *Scientific Reports* 14.1 (2024), pág. 20941 (vid. págs. 27, 28).
- Devlin, Jacob et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". En: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019, págs. 4171-4186 (vid. pág. 23).
- Dewey, Chadrick et al. "Structural MRI ratios fail to distinguish progressive supranuclear palsy from Parkinson disease in individual patients". En: *Neurology: Clinical Practice* 13.3 (2023), e200157 (vid. pág. 12).
- Ding, Jun-En, Chien-Chin Hsu y Feng Liu. "Parkinson's disease classification using contrastive graph cross-view learning with multimodal fusion of SPECT images and clinical features". En: (2024), págs. 1-5 (vid. pág. 27).
- Erdaş, Çağatay Berke y Emre Sümer. "A deep learning method to detect Parkinson's disease from MRI slices". En: *SN Computer Science* 3.2 (2022), pág. 120 (vid. pág. 13).
- Gao, Shuo et al. "Early detection of Parkinson's disease through multiplex blood and urine biomarkers prior to clinical diagnosis". En: *npj Parkinson's Disease* 11.1 (2025), pág. 35 (vid. pág. 17).
- Gao, Xingyu et al. "Brain status transferring generative adversarial network for decoding individualized atrophy in Alzheimer's disease". En: *IEEE Journal of Biomedical and Health Informatics* 27.10 (2023), págs. 4961-4970 (vid. pág. 27).

- Goetz, Christopher G et al. "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease". En: *Movement disorders* 19.9 (2004), págs. 1020-1028 (vid. págs. 16, 34).
- Grosch, Janina, Jürgen Winkler y Zacharias Kohl. "Early degeneration of both dopaminergic and serotonergic axons—a common mechanism in Parkinson's disease". En: *Frontiers in cellular neuroscience* 10 (2016), pág. 293 (vid. pág. 16).
- Gujral, Jaskeerat et al. "PET, SPECT, and MRI imaging for evaluation of Parkinson's disease". En: *American Journal of Nuclear Medicine and Molecular Imaging* 14.6 (2024), pág. 371 (vid. pág. 16).
- Heim, Beatrice et al. "Magnetic resonance imaging for the diagnosis of Parkinson's disease". En: *Journal of neural transmission* 124.8 (2017), págs. 915-964 (vid. pág. 11).
- Heldman, Dustin A et al. "The modified bradykinesia rating scale for Parkinson's disease: reliability and comparison with kinematic measures". En: *Movement Disorders* 26.10 (2011), págs. 1859-1863 (vid. págs. 21, 34).
- Heng, N. et al. "Striatal dopamine loss in early Parkinson's disease: Systematic review and novel analysis of dopamine transporter imaging". En: *Movement Disorders Clinical Practice* 10.4 (2023), págs. 539-546. DOI: 10.1002/mdc3.13658 (vid. pág. 11).
- Huang, Kexin, Jaan Altosaar y Rajesh Ranganath. "ClinicalBERT: Modeling clinical notes and predicting hospital readmission". En: *arXiv preprint arXiv:1904.05342* (2019) (vid. pág. 35).
- Khosla, Prannay et al. "Supervised contrastive learning". En: *Advances in neural information processing systems* 33 (2020), págs. 18661-18673 (vid. pág. 37).

- Kishore, Asha et al. "Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: differential effects on bradykinesia". En: *Movement Disorders: Official Journal of the Movement Disorder Society* 22.3 (2007), págs. 328-333 (vid. págs. 19, 20).
- Latif, Saad et al. "Dopamine in Parkinson's disease". En: *Clinica chimica acta* 522 (2021), págs. 114-126 (vid. pág. 15).
- Li, Chang et al. "Automatic diagnosis of Parkinson's disease using artificial intelligence base on routine T1-weighted MRI". En: *Frontiers in Medicine* 10 (2024), pág. 1303501 (vid. pág. 13).
- Litvan, Irene et al. "Movement Disorders Society Scientific Issues Committee report: SIC Task Force appraisal of clinical diagnostic criteria for parkinsonian disorders." En: *Movement disorders: official journal of the Movement Disorder Society* 18.5 (2003), págs. 467-486 (vid. pág. 30).
- Madabushi, Jayakrishna S et al. "Parkinson's disease: diagnostic challenges amidst transdiagnostic and overlapping mental health symptoms". En: *Cureus* 15.3 (2023) (vid. pág. 16).
- Mitchell, Trina et al. "Emerging neuroimaging biomarkers across disease stage in Parkinson disease: a review". En: *JAMA neurology* 78.10 (2021), págs. 1262-1272 (vid. pág. 11).
- Moor, Michael et al. "Foundation models for generalist medical artificial intelligence". En: *Nature* 616.7956 (2023), págs. 259-265 (vid. pág. 23).
- Morgen, K et al. "Structural brain abnormalities in patients with Parkinson disease: a comparative voxel-based analysis using T1-weighted MR imaging and magnetization transfer imaging". En: *American journal of neuroradiology* 32.11 (2011), págs. 2080-2086 (vid. pág. 12).

- Mortezazadeh, Tohid et al. "Imaging modalities in differential diagnosis of Parkinson's disease: opportunities and challenges". En: *Egyptian journal of radiology and nuclear medicine* 52 (2021), págs. 1-12 (vid. pág. 18).
- Obeso, Jose A et al. "Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy". En: *Movement disorders* 32.9 (2017), págs. 1264-1310 (vid. pág. 30).
- Obeso, Jose A et al. "The basal ganglia in Parkinson's disease: current concepts and unexplained observations". En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 64.S2 (2008), S30-S46 (vid. pág. 36).
- Oord, Aaron van den, Yazhe Li y Oriol Vinyals. "Representation learning with contrastive predictive coding". En: *arXiv preprint arXiv:1807.03748* (2018) (vid. pág. 22).
- Pahuja, Gunjan y Bhanu Prasad. "Deep learning architectures for Parkinson's disease detection by using multi-modal features". En: *Computers in Biology and Medicine* 146 (2022), pág. 105610 (vid. pág. 28).
- Pan, Yiwei et al. "A multi-atlas-based [18F] 9-fluoropropyl-(+)-dihydrotrabenazine positron emission tomography image segmentation method for Parkinson's disease quantification". En: *Frontiers in Aging Neuroscience* 14 (2022), pág. 902169 (vid. pág. 12).
- Postuma, Ronald B et al. "MDS clinical diagnostic criteria for Parkinson's disease". En: *Movement disorders* 30.12 (2015), págs. 1591-1601 (vid. págs. 14, 19).
- Prashanth, R et al. "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning". En: *International journal of medical informatics* 90 (2016), págs. 13-21 (vid. pág. 31).

- Priyadharshini, S. et al. "A comprehensive framework for Parkinson's disease diagnosis using explainable artificial intelligence empowered machine learning techniques". En: *Alexandria Engineering Journal* 107 (2024), págs. 568-582 (vid. págs. 12, 28).
- Rana, Abdul Qayyum et al. "Parkinson's disease: a review of non-motor symptoms". En: *Expert review of neurotherapeutics* 15.5 (2015), págs. 549-562 (vid. pág. 15).
- Reyes, Diego Machado et al. "Language Modeling Screens Parkinson's Disease with Self-reported Questionnaires". En: *medRxiv* (2024) (vid. pág. 28).
- Saeed, Usman et al. "Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: current and emerging concepts". En: *Translational neurodegeneration* 6 (2017), págs. 1-25 (vid. págs. 16, 17).
- Schulz, Jörg B et al. "Magnetic resonance imaging–based volumetry differentiates idiopathic Parkinson's syndrome from multiple system atrophy and progressive supranuclear palsy". En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.1 (1999), págs. 65-74 (vid. pág. 17).
- Shi, Dafa et al. "Classification of Parkinson's disease using a region-of-interest-and resting-state functional magnetic resonance imaging-based radiomics approach". En: *Brain Imaging and Behavior* 16.5 (2022), págs. 2150-2163 (vid. pág. 12).
- Shickel, Benjamin et al. "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis". En: *IEEE journal of biomedical and health informatics* 22.5 (2017), págs. 1589-1604 (vid. pág. 19).
- Shih, M. C. et al. "Parkinson's disease and dopamine transporter neuroimaging: a critical review". En: *Sao Paulo Medical Journal* 124 (2006), págs. 168-175 (vid. pág. 30).

- Shiiba, Takuro et al. "Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting Parkinson's disease". En: *EJNMMI research* 12.1 (2022), pág. 39 (vid. pág. 13).
- Stoker, Thomas B y Julia C Greenland. "Parkinson's disease: pathogenesis and clinical aspects [internet]". En: (2018) (vid. pág. 15).
- Suwijn, Sven R et al. "The diagnostic accuracy of dopamine transporter SPECT imaging to detect nigrostriatal cell loss in patients with Parkinson's disease or clinically uncertain parkinsonism: a systematic review". En: *EJNMMI research* 5 (2015), págs. 1-8 (vid. pág. 30).
- Sveinbjornsdottir, Sigurlaug. "The clinical symptoms of Parkinson's disease". En: *Journal of neurochemistry* 139 (2016), págs. 318-324 (vid. pág. 15).
- Szewczyk-Krolkowski, Konrad et al. "The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort". En: *Parkinsonism & related disorders* 20.1 (2014), págs. 99-105 (vid. pág. 14).
- Tan, Hongyun, Binbin Luo y Chao Cong. "Using Deep Learning to differentiate between Parkinson's Disease and Multiple System Atrophy based on PET and MRI images". En: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2023, págs. 3181-3187 (vid. pág. 13).
- Vingerhoets, Francois JG et al. "Which clinical sign of Parkinson's disease best reflects the nigrostriatal lesion?" En: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 41.1 (1997), págs. 58-64 (vid. págs. 19, 21).

- Wang, Haoyu et al. "SAM-Med3D: a vision foundation model for general-purpose segmentation on volumetric medical images". En: *IEEE Transactions on Neural Networks and Learning Systems* (2025) (vid. págs. 25, 36).
- Wang, Ling et al. "SPECT Molecular Imaging in Parkinson's Disease". En: *BioMed Research International* 2012.1 (2012), pág. 412486 (vid. pág. 17).
- Wenzel, M. et al. "Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics". En: *European Journal of Nuclear Medicine and Molecular Imaging* 46 (2019), págs. 2800-2811 (vid. págs. 12, 13).
- Wu, Hao et al. "Unified visual-semantic embeddings: Bridging vision and language with structured meaning representations". En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, págs. 6609-6618 (vid. pág. 21).
- Yu, Qi et al. "A transformer-based unified multimodal framework for Alzheimer's disease assessment". En: *Computers in Biology and Medicine* 180 (2024), pág. 108979 (vid. pág. 28).
- Zhang, Yichi, Zhenrong Shen y Rushi Jiao. "Segment anything model for medical image segmentation: Current applications and future directions". En: *Computers in Biology and Medicine* (2024), pág. 108238 (vid. págs. 25, 36).

APÉNDICE DE PRODUCTOS ACADÉMICOS

Anexo A. Productos Académicos

Conferencias:

A. Ortega-García, E. Osma, F. Sierra G. Castellanos, F. Martínez. “Parkinson stratification from a multimodal contrastive model merging scale-reports and segmentation Foundational findings”. 48th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2026.

Estado: Aceptado.