



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica y Automática Industrial

TRABAJO FIN DE GRADO

**ESTUDIO DE LA EVOLUCIÓN DEL
DECLIVE COGNITIVO EN SUJETOS CON
PARKINSON NOVO A PARTIR DE
MODELOS PREDICTIVOS Y
PROGRESIVOS**

Eva Gómez Barahona

Tutor: Carlos Platero Dueñas

Departamento: Ingeniería Eléctrica, Electrónica, Automática y Física
Aplicada

Madrid, Septiembre, 2023



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica y Automática Industrial

TRABAJO FIN DE GRADO

ESTUDIO DE LA EVOLUCIÓN DEL
DECLIVE COGNITIVO EN SUJETOS CON
PARKINSON NOVO A PARTIR DE
MODELOS PREDICTIVOS Y
PROGRESIVOS

Firma Autor

Firma Cotutor (si lo hay)

Firma Tutor

Copyright ©2023. Eva Gómez Barahona

Esta obra está licenciada bajo la licencia Creative Commons

Atribución-NoComercial-SinDerivadas 3.0 Unported (CC BY-NC-ND 3.0). Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.es> o envíe una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, EE.UU.

Todas las opiniones aquí expresadas son del autor, y no reflejan necesariamente las opiniones de la Universidad Politécnica de Madrid.

Título: Estudio de la evolución del declive cognitivo en sujetos con Parkinson novo a partir de modelos predictivos y progresivos

Autor: Eva Gómez Barahona

Tutor: Carlos Platero Dueñas

EL TRIBUNAL

Presidente:

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Trabajo Fin de Grado el día de de ... en, en la Escuela Técnica Superior de Ingeniería y Diseño Industrial de la Universidad Politécnica de Madrid, acuerda otorgarle la CALIFICACIÓN de:

VOCAL

SECRETARIO

PRESIDENTE

Agradecimientos

Deseo expresar mi más sincero agradecimiento a mis padres y a mi hermano por su apoyo incondicional a lo largo de estos cuatro años. Gracias por hacerme sacar una sonrisa incluso en los momentos más difíciles, por estar siempre a mi lado, por entenderme y, muchas veces, aguantarme. Por confiar en mí incluso cuando yo no lo hacía, por enseñarme a nunca rendirme y luchar por lo que verdaderamente quiero. Todo este trabajo no habría sido posible sin vosotros al lado.

A Lucía y a Marta, mis dos pilares durante estos cuatro años. Gracias por acompañarme siempre en los momentos buenos y en los difíciles, por vuestro apoyo y confianza incondicional en mí, y por vuestro cariño.

Finalmente, a mi tutor Carlos Platero, gracias por la orientación y apoyo a lo largo del proyecto. Por la manera de transmitirnos los conocimientos sobre este área y conseguir que disfrutemos tanto. Me siento muy agradecida de haber podido sumergirme en el área de la inteligencia artificial y el análisis de datos y haber podido disfrutar tanto a lo largo de este trabajo.

Gracias a todas esas personas que me han acompañado a lo largo de este camino. Sin ellos, nada de esto sería posible.

Resumen

La enfermedad de Parkinson (Parkinson's Disease (PD)) es uno de los trastornos neurodegenerativos que afecta tanto el control motor como diversos síntomas no motores.

Este proyecto se centra en la investigación del declive cognitivo en individuos recién diagnosticados con la enfermedad de Parkinson. En una primera parte de la investigación, se analizarán diferentes variables y biomarcadores, responsables del desarrollo de la PD y el declive cognitivo. Posteriormente, mediante modelos progresivos y predictivos, se procederá al análisis de estas características con el fin de obtener un análisis de la progresión cognitiva en individuos con enfermedad de Parkinson.

El objetivo es comprender la trayectoria de la enfermedad y desarrollar herramientas que permitan anticipar y personalizar intervenciones terapéuticas. El estudio contribuirá a mejorar la comprensión de la progresión de la enfermedad y proporcionar información para mejorar la atención y el tratamiento de pacientes con PD.

Palabras clave: Enfermedad de Parkinson, PPMI, modelos predictivos, biomarcadores, declive cognitivo, MEANCAU, edad.

Abstract

Parkinson's Disease (PD) is one of the neurodegenerative disorders that affects both motor control and various non-motor symptoms.

This project focuses on investigating cognitive decline in individuals newly diagnosed with Parkinson's Disease. In the initial phase of the research, different variables and biomarkers responsible for the development of PD and cognitive decline will be analyzed. Subsequently, through progressive and predictive models, an analysis of these characteristics will be conducted to obtain insights into the cognitive progression in individuals with Parkinson's disease.

The goal is to comprehend the disease trajectory and develop tools that enable the anticipation and customization of therapeutic interventions. The study will contribute to enhancing the understanding of disease progression and providing information to enhance the care and treatment of PD patients.

Keywords: Parkinson's Disease, PPMI, predictive models, biomarkers, cognitive decline, MEANCAU, age.

Índice general

Agradecimientos	IX
Resumen	XI
Abstract	XIII
Índice	XVIII
1. Introducción	1
1.1. Neurociencia	1
1.2. La enfermedad de Parkinson	1
1.2.1. Síntomas	2
1.2.2. Factores de riesgo	3
1.2.3. Diagnóstico	4
1.2.4. Tratamiento	4
1.2.5. Impacto socioeconómico	5
1.3. Motivación del proyecto	6
1.4. Estructura del documento	6
2. Estado del arte	9
2.1. Clinical variables and biomarkers in prediction of cognitive impairment [41]	9
2.1.1. Introducción	9
2.1.2. Materiales	9
2.1.3. Métodos	10
2.1.4. Resultados	11
2.2. Parkinson disease-associated cognitive impairment [4]	12
2.2.1. Introducción	12
2.2.2. Epidemiología	12
2.2.3. Fisiopatología	13
2.2.4. Neuropatología	14
2.2.5. Tratamientos para el deterioro cognitivo en la PD	16
2.3. Diagnosing PD-MCI by MDS TaskForce Criteria [21]	17
2.3.1. Introducción	17
2.3.2. Dominios cognitivos	17
2.3.3. Métodos	18
2.3.4. Resultados	19
2.3.5. Conclusión	19

2.4.	Toward a biological definition of Alzheimer's disease [25]	20
2.4.1.	Introducción	20
2.4.2.	Enfermedad de Alzheimer	20
2.4.3.	Biomarcadores de la AD	20
2.4.4.	Estados	21
2.4.5.	Perfiles y categorías según los biomarcadores	21
2.5.	Brain Amyloid Contribution to Cognitive Dysfunction in Early PD [17]	22
2.5.1.	Introducción	22
2.5.2.	Métodos	22
2.5.3.	Resultados	23
2.5.4.	Conclusiones	24
2.6.	Longitudinal survival analysis and two-group comparison for MCI to AD [34]	25
2.6.1.	Introducción	25
2.6.2.	Materiales	26
2.6.3.	Métodos	26
2.6.4.	Resultados	29
2.6.5.	Conclusiones	29
2.7.	Estimating long-term multivariate progression from short-term data [13]	30
2.7.1.	Introducción	30
2.7.2.	Supuestos del modelo	31
2.7.3.	Resultados	31
2.8.	Aportación a la investigación	32
2.8.1.	Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study	32
2.8.2.	Parkinson disease-associated cognitive impairment	32
2.8.3.	Diagnosing PD-MCI by MDS TaskForce Criteria: How Many and Which Neuropsychological Tests?	32
2.8.4.	NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease	33
2.8.5.	Brain Amyloid Contribution to Cognitive Dysfunction in Early-Stage Parkinson's Disease: The PPMI Dataset	33
2.8.6.	Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease	33
2.8.7.	Estimating long-term multivariate progression from short-term data	34
3.	Datos empleados, herramientas y materiales	35
3.1.	Introducción	35
3.2.	Datos empleados	35
3.2.1.	Test neuro-psicológicos utilizados en PPMI	36
3.3.	Herramientas	39
3.3.1.	MATLAB	39
3.3.2.	R	40
3.3.3.	FileZilla	40

3.3.4.	PuTTY	40
3.3.5.	Xming	41
3.3.6.	Tmux	41
3.3.7.	Magerit-3	41
3.3.8.	LaTeX	41
4.	Métodos	43
4.1.	Introducción	43
4.2.	Análisis de los sujetos al inicio del seguimiento	43
4.3.	Perfiles AT en sujetos PD y tiempos de conversión y censura	45
4.3.1.	Perfiles AT en sujetos PD	45
4.3.2.	Tiempos de conversión y censura	46
4.4.	Análisis estadístico	47
4.4.1.	Modelos predictivos	47
4.4.2.	Modelos de progresión con GRACE	50
4.5.	Modelos de progresión de forma masiva con Magerit	53
4.5.1.	Boostrap	56
5.	Resultados	59
5.1.	Introducción	59
5.2.	Análisis de los sujetos al inicio del seguimiento	59
5.2.1.	Análisis de sujetos HC y PD al inicio del seguimiento	59
5.2.2.	Análisis de sujetos PD-CU PD-MCI y PDD al inicio del seguimiento	61
5.3.	Perfiles AT y tiempos de conversión en pacientes PD-novo	62
5.3.1.	Perfiles AT	62
5.3.2.	Tiempos de conversión	63
5.4.	Análisis estadístico	64
5.4.1.	Modelos predictivos	64
5.4.2.	Modelos de progresión con GRACE	66
5.5.	Modelos de progresión de forma masiva con Magerit	71
5.5.1.	Boostrap	73
6.	Discusiones	75
6.1.	Introducción	75
6.2.	Análisis de los sujetos al inicio del seguimiento	75
6.2.1.	Análisis de los sujetos HC y PD	75
6.2.2.	Análisis	75
6.3.	Perfiles AT y tiempos de conversión y censura	76
6.4.	Análisis estadístico	78
6.4.1.	Modelos predictivos	78
6.4.2.	Modelos de progresión con GRACE	80
6.5.	Modelos de progresión de forma masiva con Magerit	81
6.5.1.	Boostrap	83
6.6.	Conclusiones	84
6.7.	Desarrollos futuros	85

A. Análisis estadístico aplicado a las ciencias de la salud.	87
A.1. Introducción	87
A.2. Conceptos básicos	87
A.2.1. Inferencia estadística	87
A.2.2. Intervalos de confianza	88
A.2.3. Estadísticos y p-valor	89
A.3. Calidad del experimento	90
A.4. Análisis de la varianza	91
A.5. Análisis de supervivencia	95
A.5.1. Método de Cox	96
A.5.2. Modelo de regresión de Cox extendido	96
A.5.3. Clasificación logística	97
A.5.4. Variables normalizadas	97
B. Uso de Magerit	99
B.1. Introducción	99
B.2. Acceso	99
B.3. Ejecución de tareas	99
B.4. Mandatos SLURM	100
B.5. Definición de trabajos	100
B.6. Directorio SCRATCH	101
B.7. Redirección de salidas y notificaciones	101
C. Diccionario de siglas y acrónimos	103
Bibliografía	105

Índice de figuras

1.1. Imágenes de la activación dopaminérgica mediante la técnica de imagen DaTSCAN de una persona sana respecto de un paciente PD respectivamente.	2
2.1. Visión general del método propuesto.	27
3.1. Logo PPMI.	35
3.2. Logo MATLAB.	39
3.3. Logo lenguaje R.	40
3.4. Logo FileZilla.	40
3.5. Logo PuTTY.	40
3.6. Logo Xming.	41
3.7. Logo Tmux.	41
3.8. Logo CesVisMa.	42
3.9. Logo LaTeX.	42
4.1. Método de validación cruzada para el desarrollo de posibles modelos predictivos.	48
5.1. Perfiles AT al principio del estudio de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.	63
5.2. Perfiles AT tras el análisis longitudinal de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.	63
5.3. Histograma con los tiempos de conversión de los sujetos PD-novo y su evolución en los 5 primeros años de estudio: a) Conversión de sujetos PD-CU a PD-MCI o PDD, b) Distinción entre sujetos convertidos a PD-MCI y PDD c) Tiempo de censura de los sujetos PD-CU.	64
5.4. Matriz de correlación de la pendiente según el criterio SITE con Age.	67
5.5. Matriz de correlación del intercepto según el criterio SITE con Age.	67
5.6. Trayectorias de los biomarcadores del vector con Age . Trayectorias a corto plazo en verde para sujetos PD-CU, azul para PD-MCI y rojo para PDD. Trayectoria a largo plazo representada en negro.	68
5.7. Distribución de los sujetos PD según el tiempo de conversión a partir del vector con Age como covariable	68
5.8. Matriz de correlación de la pendiente según el criterio SITE sin Age.	69
5.9. Matriz de correlación del intercepto según el criterio SITE sin Age.	69

5.10. Trayectorias de los biomarcadores del vector sin Age . Trayectorias a corto plazo en verde para sujetos PD-CU, azul para PD-MCI y rojo para PDD. Trayectoria a largo plazo representada en negro.	70
5.11. Distribución de los sujetos PD según el tiempo de conversión a partir del vector sin Age como covariable	70
5.12. Trayectorias a corto y largo plazo de las características del vector con mayor rendimiento.	73
5.13. Distribución de los sujetos PD según el tiempo de conversión a partir del vector con Age como covariable	73
6.1. Perfiles AT al principio del estudio de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.	77
6.2. Perfiles AT tras el análisis longitudinal de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.	77
6.3. Distribución de los sujetos PD según el tiempo de conversión y recta de regresión del vector	82
6.4. Matriz de correlación del intercepto según el criterio SITE con Age.	82
6.5. Matriz de correlación de la pendiente según el criterio SITE con Age.	82
6.6. Trayectorias de test neuropsicológicos y biomarcadores tras análisis bootstrap para perfiles A+ y A- en AD	84
6.7. Trayectorias de test neuropsicológicos y biomarcadores tras análisis bootstrap para perfiles A+ y A- en PD	84
A.1. Representación de un intervalo de confianza de 95.5%	89
A.2. Curvas ROC de calidades perfecta, buena y mala respectivamente	91
A.3. Visión general del método propuesto.	92
A.4. Gráfico interactivo donde se muestran los intervalos de comparación.	95
A.5. Tabla de contingencia para el estudio del sexo de los pacientes sometidos a estudio.	95

Índice de tablas

1.1. Escala de Hoehn y Yahr	4
3.1. Pacientes PD del experimento.	36
5.1. Características demográficas y clínicas de los pacientes con PD y HC al inicio del seguimiento.	60
5.2. Características demográficas y clínicas de los pacientes con PD-CU, PD-MCI y PDD al inicio del seguimiento.	61
5.3. Modelos univariantes con Age como covariable	65
5.4. Modelos univariantes sin Age como covariable	65
5.5. Modelos multivariantes con Age como covariable	65
5.6. Modelos multivariantes sin Age como covariable	65
5.7. Características de los vectores seleccionados, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.	71
5.8. Valores de diferentes características según el tiempo onset del vector.	71
5.9. Valores del test MoCA y la edad según el tiempo onset del vector.	71
5.10. Tiempo entre el inicio del deterioro cognitivo y demencia según las puntuaciones del test MoCA y el biomarcador $A\beta$ para los diferentes vectores seleccionados con la edad.	72
5.11. Características de los vectores seleccionados, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.	72
5.12. Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.	72
5.13. Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.	72
5.14. Tiempo entre el inicio del deterioro cognitivo y demencia según las puntuaciones del test MoCA y el biomarcador $A\beta$ para los diferentes vectores seleccionados sin la edad.	72
6.1. Porcentajes de los perfiles AT en el estudio al inicio del estudio y tras el análisis longitudinal.	78
6.2. Características del mejor vector, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.	81
6.3. Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.	82
6.4. Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.	83

Capítulo 1

Introducción

1.1. Neurociencia

El sistema motor es la parte del sistema nervioso central encargado del movimiento. La estructura de este se basa en las células nerviosas llamadas **neuronas**.

La neurona tiene como función principal la generación de señales eléctricas denominadas potenciales de acción cuyo objetivo principal es transmitir información de forma rápida a largas distancias. Para ello esta célula se divide en tres partes fundamentales:

- Dendritas: ramas que salen del cuerpo de la neurona. Se encargan de la transmisión de información en dirección centrípeta (hacia el cuerpo celular).
- Cuerpo celular: parte central encargada de mantener la información genética (ADN) con el núcleo, además de generar la energía necesaria para que la información se puede transmitir de manera correcta.
- Axón: es la prolongación principal de la neurona que se caracteriza por establecer la sinapsis neuronal.

Estas células se caracterizan por establecer miles de conexiones, permitiendo así recibir y enviar mensajes a grandes velocidades. En el cuerpo celular se produce la filtración de la información recibida, eliminando los impulsos débiles y no útiles. La parte del mensaje seleccionado transcurre por el axón, permitiendo así llegar a su destino.

Es aquí donde se produce la **sinapsis**, espacio que existe entre una neurona y otra célula. Aquí los impulsos son transmitidos a partir de neurotransmisores, liberados por la neurona que tiene la información. La otra célula receptora tiene como función recibir estas sustancias químicas.

1.2. La enfermedad de Parkinson

La enfermedad de Parkinson (PD: Parkinson's Disease) es un trastorno neurodegenerativo producido cuando las células nerviosas no son capaces de producir suficiente dopamina, produciéndose así alteraciones significativas en la capacidad de

movimiento. Actualmente, es una de las enfermedades más frecuentes entre la población de más de 65 años y, se estima que su número crezca de forma considerable en los próximos años.

Esta enfermedad se caracteriza por la pérdida de neuronas en la sustancia negra, suponiendo así una producción insuficiente de dopamina. Además, estos pacientes suelen presentar acumulaciones de la proteína α -sinucleína en las neuronas que las inhabilita y conduce a la muerte celular.

No obstante, durante el desarrollo de la enfermedad se ven involucrados diferentes mecanismos entre los que destacan el estrés oxidativo¹, déficits en el sistema de neurotransmisores² o la neuroinflamación³. Esto supone la aparición de síntomas de carácter motor y no motor que agravan la enfermedad.

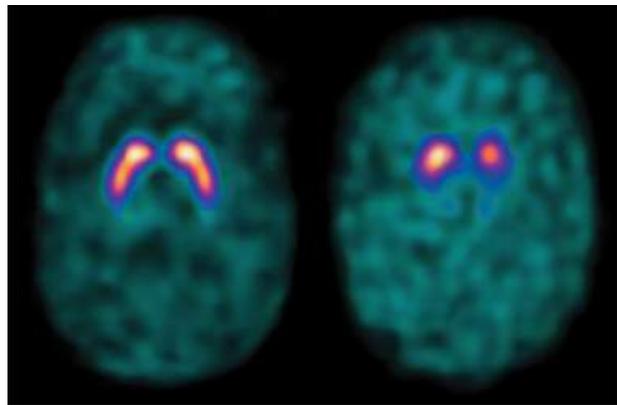


Figura 1.1: Imágenes de la activación dopaminérgica mediante la técnica de imagen DaTSCAN de una persona sana respecto de un paciente PD respectivamente.

1.2.1. Síntomas

A diferencia de otras, la enfermedad de Parkinson no tiene unos síntomas claramente definidos si no que, pueden variar o incluso pasar desapercibidos durante los primeros meses de enfermedad. No obstante, el temblor en reposo suele ser la característica inicial más común en estos pacientes [6].

Los síntomas **no motores** establecidos en la actualidad de la PD son los siguientes:

- Trastornos en el olfato: los pacientes pueden desarrollar anosmia⁴, dificultando la identificación de olores familiares. Los pacientes de PD también pueden padecer hiposmia⁵ o disosmia⁶.
- Alteraciones en la fase REM del sueño: se pueden manifestar en forma de cuadros de sueño fragmentado o insomnio. Además, una vez la enfermedad está avanzada, se suelen sumar alucinaciones y delirios debidos al declive cognitivo.

¹Desequilibrio entre la producción de especies reactivas del oxígeno y la capacidad de un sistema biológico de neutralizar o reparar el daño resultante

²Alteración del material genético

³Respuesta del cerebro a una lesión, infección o enfermedad

⁴Pérdida del sentido del olfato.

⁵Disminución del sentido del olfato.

⁶Percepción distorsionada de los olores.

- Otros síntomas no motores: el estreñimiento o problemas relacionados con la vejiga o el sistema nervioso autónomo.

Paralelamente, se establecen como **síntomas motores**:

- Bradicinesia: trastorno caracterizado por la lentitud de los movimientos de forma incontrolable acompañado con una sensación de rigidez.
- Temblor asimétrico en reposo: hace referencia a sacudidas rítmicas normalmente en las manos, brazos, cabeza, piernas o torso. También puede verse reflejado al tener voz temblorosa o dificultad para escribir o dibujar
- Rigidez: los músculos están contraídos imposibilitando el movimiento de estos. Suele provocar dolor en los pacientes.
- Inestabilidad postural: suele aparecer cuando la enfermedad está más desarrollada y se caracteriza por la dificultad para mantener el equilibrio. Esto puede incapacitar al paciente pues, afecta a acciones como puede ser el andar.
- Otros síntomas motores: la disfagia, alteraciones del lenguaje, o síndrome de piernas inquietas son algunos de otros síntomas que pueden aparecer en pacientes con esta enfermedad.

1.2.2. Factores de riesgo

Los factores de riesgo de la enfermedad de Parkinson se pueden dividir en dos principales categorías en función de si son características modificables o no.

1. Factores de riesgo intrínsecos: aquellos que son incorregibles. Entre ellos se pueden encontrar características como:
 - Edad: la PD puede afectar a toda la población, sin distinción de edad. Sin embargo, dado que este factor supone la pérdida dopaminérgica en la vía nigraestriada y esto es acentuado con la edad. Estudios establecen que las personas suelen contraer la enfermedad aproximadamente a partir de los 60 años [5].
 - Sexo: se presenta con mayor presencia en hombres que en mujeres [5].
 - Factores genéticos: aproximadamente el 15% de los pacientes diagnosticados con la enfermedad tienen antecedentes familiares que la han sufrido [47].
2. Factores de riesgo extrínsecos: se pueden tomar medidas para reducir su impacto en la salud.
 - Estilo de vida: varias investigaciones han demostrado que llevar un estilo de vida saludable puede reducir el riesgo de desarrollar la enfermedad de Parkinson. Esto incluye hacer ejercicio regularmente, mantener un peso saludable, evitar fumar y beber alcohol en exceso, y llevar una dieta saludable y equilibrada [18].
 - Exposición a toxinas: la exposición a ciertas toxinas, como los pesticidas y los herbicidas, se ha relacionado con un mayor riesgo de desarrollar la PD. Reducir la exposición a estas sustancias puede ayudar a reducir el riesgo [27].

- **Trauma cerebral:** las lesiones cerebrales traumáticas pueden aumentar el riesgo de desarrollar la enfermedad de Parkinson. Tomar medidas para prevenir lesiones cerebrales, como usar casco al practicar deportes de contacto (entre otros) puede reducir este riesgo [8]
- **Otros trastornos de salud:** algunos trastornos de salud, como la diabetes y la hipertensión arterial, se han relacionado con un mayor riesgo de desarrollar la enfermedad de Parkinson. Controlar estos trastornos puede ayudar a reducir el riesgo [39].

1.2.3. Diagnóstico

Actualmente, el diagnóstico de la PD se basa en la evaluación clínica de los síntomas motores y no motores, junto al historial clínico del paciente ya que no exista ninguna prueba que pueda diagnosticar esta enfermedad de forma definitiva.

Además de la evaluación clínica, el uso de pruebas de imagen como la tomografía por emisión de positrones (PET) , la resonancia magnética (fMRI) o Tomografía Computarizada de Emisión Monofotónica (SPECT) pueden ayudar a detectar la degeneración dopaminérgica. Para el seguimiento de los pacientes de una forma más sencilla se han desarrollado diferentes escalas que, además, permiten una normalización de la evolución de la enfermedad. Entre estas, se encuentra la escala de Hoehn y Yahr:

Estadio	Características
Estadio 0	No existen signos de la enfermedad
Estadio 1	Enfermedad exclusivamente unilateral
Estadio 1.5	Enfermedad unilateral y axial
Estadio 2	Enfermedad bilateral y axial
Estadio 2.5	Enfermedad bilateral y axial con leve recuperación en el test del empujón
Estadio 3	Enfermedad bilateral y axial moderada con inestabilidad postural
Estadio 4	Incapacidad grave. Capaz de caminar o permanecer de pie sin ayuda
Estadio 5	Dendencia total. Permanece en cama o sillón si no tiene ayuda.

Tabla 1.1: Escala de Hoehn y Yahr

1.2.4. Tratamiento

En la actualidad, no existe cura para la PD. No obstante, se han desarrollado una serie de tratamientos con el fin de aliviar los síntomas iniciales de la enfermedad y retrasar la evolución de la patología. Los principales son los siguientes:

- **Levodopa:** esta sustancia se inserta en el cerebro y se convierte en dopamina produciendo así un retraso significativo de la aparición de los síntomas ya que compensa la falta de dopamina en las células nerviosas [22]. A día de hoy es el tratamiento más utilizado para retrasar el avance de la enfermedad.

- **Bromocriptina y pergolida:** fármacos utilizados para reducir la bradicinesia y rigidez de los pacientes [30]. Su uso es más limitado debido a los efectos secundarios que genera.
- **Selegilina:** se utiliza junto a la levodopa, permitiendo reducir el consumo de esta y alargando su efecto [9].

Es importante destacar que los tratamientos farmacológicos tienen que ser personalizados según cada paciente, pudiéndose adaptar así de la mejor forma posible a sus necesidades. Además, aunque en los últimos años se han desarrollado tratamientos muy prometedores para el futuro, actualmente no siempre son efectivos para todos los pacientes. Es por esto por lo que muchas veces, los pacientes requieren de tratamientos no farmacológicos que complementen el tratamiento.

La terapia física u ocupacional, además de la fisioterapia y la logopedia son algunos tratamientos que aportan mejores resultados al paciente. Recientemente se están incluyendo nuevas relacionadas con el habla y el lenguaje o terapias cognitivas, pudiendo así ayudar a mejorar la función cognitiva, la memoria y la resolución de problemas de los enfermos.

1.2.5. Impacto socioeconómico

La PD es una de las enfermedades neurodegenerativas más frecuentes, afectando no solo a la calidad de vida de las personas que lo padecen, si no también en su entorno social y económico.

El coste de la atención requerida para esta enfermedad es uno de los mayores impactos a nivel económico en la actualidad en relación con las enfermedades neurodegenerativas. Esto se debe a la atención médica que esta enfermedad requiere, además del elevado número de medicamentos y terapias. Varios artículos, como el publicado en 2020 en la revista científica *Movement Disorders* [14] analizó los costes relacionados con esta enfermedad en Estados Unidos debido a los medicamentos, la atención a corto y largo plazo y la improductividad laboral debido a la enfermedad. Se calculó que, durante 2017 la atención médica para personas con PD supuso 52 mil millones de dólares. La mayoría de esta cifra fue destinada a la compra de medicamentos (80 % de los costes directos). El cuidado a largo plazo representó a los costes indirectos en su mayoría.

Además, este estudio señala que la cifra de estos costes tiende a aumentar a medida que pasan los años. Esto se debe principalmente a la relación de la edad con la PD y el aumento de la esperanza de vida en las últimas décadas. Según la Organización Mundial de la Salud (OMS)[32], se estima que la enfermedad de Parkinson se duplicará a nivel mundial en los próximos 20 años, llegando a 12 millones de enfermos en 2040.

En cuanto al propio paciente se refiere, la PD afecta a la capacidad de las personas a la hora de realizar tareas cotidianas y llevar una vida independiente. A medida

que la enfermedad progresa, en la mayoría de casos es necesario la ayuda de cuidadores y/o asistentes, suponiendo un impacto económico para el paciente y la familia.

La calidad de vida también se ve afectada, pudiéndose ver afectadas la movilidad, el estado de ánimo y la capacidad para participar en actividades sociales. Esto puede derivar en el aislamiento social y, en ocasiones, la depresión.

La mortalidad de la enfermedad varía según factores como la edad del paciente, la gravedad de esta y el tratamiento que recibe. Aún así, estudios afirman que los pacientes de PD tienen una tasa de mortalidad 2.3 veces mayor que la población general [31].

1.3. Motivación del proyecto

De todos los síntomas presentes en la enfermedad, el deterioro cognitivo es uno de los que más afectan a los pacientes con PD debido a los problemas de memoria, atención y funciones ejecutivas como la resolución de problemas. Además, las tareas del día a día se ven afectadas notablemente y aumenta el riesgo de accidentes. El 25 % de los pacientes diagnosticados al inicio del estudio padecían un deterioro cognitivo leve (MCI⁷), y esto supone un factor de riesgo importante para el desarrollo de demencia.

Conocer el estado cognitivo de los pacientes PD y entender su progresión a medida que avanza la enfermedad ayuda a obtener un diagnóstico más preciso de esta. Esto puede resultar altamente beneficioso para intentar abordar la enfermedad desde edades tempranas y poder saber como va a evolucionar el paciente.

Es por esto por lo que este estudio tiene como objetivo principal el diseño de modelos predictivos que permitan ayudar a predecir la evolución de la enfermedad del paciente y su estado cognitivo.

1.4. Estructura del documento

A continuación y para facilitar la lectura del documento, se detalla el contenido de cada capítulo.

- En el capítulo 1 se realiza una introducción del estudio a realizar, explicando los conceptos básicos de la neurociencia y la PD.
- En el capítulo 2 se hace un repaso de todos los artículos consultados antes de realizar el estudio, que han servido de ayuda a la hora de desarrollar diferentes algoritmos o ideas.
- En el capítulo 3 se muestran los datos y herramientas utilizados durante el análisis y la contribución de los mismos.
- En el capítulo 4 se presentan los métodos que se han desarrollado para la obtención de los resultados y objetivos del estudio.

⁷ *Mild Cognitive Impairment*

- En el capítulo 5 se exponen los resultados obtenidos, fruto de aplicar los métodos desarrollados en el capítulo anterior.
- En el capítulo 6 se procede al análisis de los resultados anteriores, obteniendo así diferentes conclusiones del proyecto. Además, se proporciona una visión global del estudio, incluyendo nuevas posibles vías de investigación.

Capítulo 2

Estado del arte

En este capítulo se muestran de forma resumida los artículos que han sido utilizados como referencia y han ayudado a desarrollar el estudio.

2.1. Clinical variables and biomarkers in prediction of cognitive impairment [41]

2.1.1. Introducción

El deterioro cognitivo es uno de los signos más evidentes que presentan los pacientes que desarrollan la PD durante un periodo de tiempo elevado (se estima que un 75 % de estos). Además, los estudios realizados hasta la fecha indican que los pacientes con deterioro cognitivo leve que son diagnosticados con Parkinson muy probablemente acaben desarrollando demencia. El déficit dopaminérgico se muestra actualmente como una de las causas principales por las que se desarrolla este deterioro cognitivo.

Antes de este estudio, no se habían hecho investigaciones significativas donde se hubiera planteado la correlación entre biomarcadores y síntomas clínicos del paciente para esta enfermedad. Es por esto que el objetivo principal de este estudio realizado en 2016 fue ampliar información sobre los valores previstos en marcadores clínicos, genéticos y DAT (transportador de dopamina) para el desarrollo del deterioro cognitivo en pacientes diagnosticados con Parkinson. Además se estimó el riesgo de deterioro cognitivo a partir de cada uno de estos valores.

2.1.2. Materiales

Mayor edad, sexo, menor educación o nivel cognitivo son algunos de los factores que se presentaron como posibles causantes del desarrollo del declive cognitivo en el paciente. Como biomarcadores principales para el estudio se encontraron el déficit en el DAT además de los valores α -sinucleína, A β 42, t-tau y p-tau en el CSF (líquido cefalorraquídeo). También se valoraron los valores de APOE¹.

Se procedió al estudio de estos datos clínicos de manera individual y conjunta para poder así establecer una predicción en el desarrollo del deterioro cognitivo re-

¹ApolipoproteínaE

lacionado a la enfermedad de Parkinson. Se hizo especial hincapié en los resultados de valores del CSF y DAT durante los dos años de estudio.

Simultáneamente, las variables clínicas establecidas para el estudio fueron la edad, el sexo del paciente, años de educación, depresión y sentido del olfato (UPSIT). Los datos de pacientes para el análisis se extrajeron del PPMI (Parkinson's Progression Marker Initiative), incluyendo aquellos con un seguimiento de al menos 2 años. En el punto de partida, estos pacientes tenían que tener al menos 30 años y haber tenido un diagnóstico reciente de la detección del Parkinson. De forma paralela, se eligió un grupo control con el mismo reglamento pero sin indicios de esta enfermedad.

2.1.3. Métodos

Para la evaluación del declive cognitivo se utilizó la prueba cognitiva MoCA (Evaluación Cognitiva de Montreal) la cual a partir de 30 preguntas ayuda a evaluar las habilidades cognitivas. Se estudiaron la evolución de los resultados de este ensayo durante 2 años con los pacientes seleccionados. Por otro lado, se procedió a la clasificación de la función cognitiva a partir de pruebas como HVLT-R (Prueba de Aprendizaje Verbal de Hopkins) para la memoria, JLO *Judgement of Line Orientation* para la función visuoespacial, el Test de Modalidades Símbolo-Dígito (SDMT) para el procesamiento de velocidad-atención o la Secuenciación Letra-Número (LNS) para función ejecutiva y memoria de trabajo. La puntuación en dos o más pruebas se relacionó al deterioro cognitivo leve.

Solo se diagnosticó la demencia si además había evidencia de que el deterioro era tal que afectaba a las tareas de la vida diaria.

A partir de los resultados en la prueba MoCA se hicieron análisis lineales univariados y multivariados que permitieron establecer la relación entre los parámetros utilizados. En el caso de los univariados las variables con resultados de $p < 0.2$ se introdujeron en un proceso de eliminación hacia atrás (Backward Stepwise Regression).

Por otro lado, si existía una alta correlación entre las variables el menor valor de p se introducía como una variable independiente. Esto permitió determinar posibles factores de riesgo del deterioro cognitivo o demencia a los 2 años de haberse diagnosticado la enfermedad.

Los análisis multivariados se hicieron variando la variable independiente de estudio. Estos fueron los siguientes casos :

- Edad
- Edad con variables clínicas
- Edad con resultados de imágenes DAT
- Edad con biomarcadores de LCR
- Edad con imágenes clínicas, DAT y biomarcadores del CSF

Tras la obtención de los resultados correspondientes si las variables a estudio no estaban altamente relacionadas ($r > 0.5$) y eran diferentes entre los pacientes con y sin deterioro cognitivo ($p > 0.05$) se incluían como variables independientes.

Finalmente, para poder predecir el riesgo de cada paciente de forma individual se procedió al uso de los resultados de los coeficientes de progresión en los análisis anteriores, obteniendo así la siguiente expresión:

$$\text{riesgodedeteriorocognitivodelpaciente} = \frac{\exp(\text{puntuacionriesgodelpaciente})}{1 + \exp(\text{puntuacionriesgodelpaciente})}$$

Donde la puntuación de riesgo del paciente se obtiene a partir de los coeficientes de progresión.

2.1.4. Resultados

El estudio se llevó a cabo en 390 pacientes del PPMI. El 82 % tuvo un seguimiento en los resultados de las pruebas MoCA durante 2 años. Se estableció que no había una diferencia significativa entre los individuos en los que se produjo un seguimiento y los que no.

En los análisis multivariados la variación de puntuación en MoCA desde el inicio hasta los 2 años de seguimiento se vio relacionada con factores como la edad, el estado de APOE, el ratio de $A\beta$:tau en CSF y el test de olfato UPSIT.

Finalmente, la puntuación a los 2 años de los grupos control se vio referenciada con parámetros como la edad, el sexo o la puntuación MoCA en el punto de partida. Además se pudo concluir que el 16 % de los pacientes recién diagnosticados con Parkinson desarrollaron deterioro cognitivo leve durante los 2 años de estudio, y un 1 % de la muestra fueron diagnosticados con demencia. En el grupo de control solo a 2 de 178 personas (1 %) se les determinó declive cognitivo leve y no hubo ningún diagnóstico con demencia durante la duración del estudio.

Con estos datos se pudo establecer que los 52 pacientes con Parkinson que desarrollaron un deterioro cognitivo o demencia tenían, por lo general, una mayor edad y puntuación en RBDSQ (*REM sleep Behavior Disorder Screening Questionnaire*) desde el punto de partida de la investigación, además de valores menores del ratio de $A\beta$:tau en CSF y del test de olfato UPSIT con respecto a los pacientes sin declive cognitivo.

Por otro lado, cabe destacar los resultados obtenidos del AUC (Área Bajo la Curva) en los modelos en los que se hizo uso de la edad y variables clínicas ($p = 0.03$) y la edad y parámetros CSF ($p = 0.02$) ya que ambos se diferencian notablemente de los resultados obtenidos con el modelo final ($p = 0.62$).

Los demás modelos estudiados no presentaron diferencias significativas. A modo de conclusión se propuso el ejemplo de un paciente recién diagnosticado con Parkinson con una edad de 70 años. Tras el estudio de las variables clínicas y biomarcadores en punto de partida se predijo un riesgo de desarrollo de deterioro cognitivo en los

siguientes 2 años del 13%. Con los resultados obtenidos a lo largo de la investigación, el riesgo de declive cognitivo se disminuiría hasta un 5% si el paciente tuviera 50 años. Además, si sus valores de CSF $A\beta$ disminuyeran notablemente y su puntuación en RBDSQ aumentara este riesgo también aumentaría hasta un 34%.

2.2. Parkinson disease-associated cognitive impairment [4]

2.2.1. Introducción

La PD se caracteriza por la pérdida de neuronas en la sustancia negra, suponiendo así una producción insuficiente de dopamina. Además, estos pacientes suelen presentar acumulaciones de la proteína α -sinucleína en las neuronas que las imposibilita y conduce a la muerte celular.

No obstante, durante el desarrollo de la enfermedad se ven involucrados diferentes mecanismos entre los que destacan el estrés oxidativo, déficits en el sistema de neurotransmisores o la neuroinflamación. Esto supone la aparición de síntomas de carácter motor y no motor que agravan la enfermedad. El deterioro cognitivo es una de las manifestaciones más determinantes entre los diagnosticados, siendo hasta 6 veces más probable en personas con PD respecto a la población sana. Se puede determinar niveles entre los diagnósticos de deterioro cognitivo, dividiéndolo en los siguientes grupos: SCD (Subjective Cognitive Decline), PD-MCI (*Parkinson Disease Mild Cognitive Impairment*) y PDD (*Parkinson Disease Dementia*).

SCD se caracteriza por una leve disminución de la capacidad cognitiva, detectado a partir de diferentes pruebas regularizadas.

Por otro lado, PD-MCI se asocia a personas que presentan un declive cognitivo de forma gradual una vez se les ha diagnosticado PD, suponiendo así problemas en su día a día a la hora de realizar ciertas actividades funcionales dificultosas.

Finalmente, los pacientes diagnosticados con PDD sufren, al menos, déficit en dos de los cuatro dominios cognitivos existentes (atención, memoria, habilidades de visión espacial y habilidades ejecutivas). Además dentro de los pacientes con PDD pueden diferenciarse en tres subgrupos: leve (efecto en tareas diarias de forma leve), moderada y severa (incapacidad de realizar tareas de forma independiente)

Actualmente no se ha conseguido desarrollar un tratamiento que permita retrasar o evitar el declive cognitivo en pacientes con PD, por lo que se requiere un estudio en mayor profundidad de la epidemiología y biomarcadores del declive cognitivo en relación con esta enfermedad neurodegenerativa.

2.2.2. Epidemiología

El diagnóstico del deterioro cognitivo se puede dar antes o durante el desarrollo de la PD, y sus efectos y progresión pueden variar según el paciente. Estudios revelan que los pacientes con PD tienen hasta 6 veces mayor riesgo de desarrollar declive cognitivo en comparación con personas sanas de su misma edad. No obstante, la

distribución y los factores determinantes para el desarrollo del deterioro cognitivo no están del todo claros debido a la ausencia de estudios donde prevalezcan pacientes con PDD o PD-MCI.

- **Demencia:** Los estudios confirman que la conversión a PDD de pacientes 5 años después del diagnóstico es del 17 %, el 46 % a los 10 años y 83 % a los 20 años. Estas cifras son significativamente mayores que la prevalencia de demencia en la población de más de 60 años (5-7 %). Cabe destacar que, aunque se han podido establecer algunos factores de riesgo que pueden explicar la demencia en los pacientes, aún es necesario estudiar los mecanismos que impulsan este declive para llegar al estado de demencia.

Por otro lado, el deterioro cognitivo en PDD es similar al de AD (Alzheimer's Disease), desembocando ambos en la dependencia total de los enfermos.

- **Defecto cognitivo leve:** es una etapa previa a la demencia que se presenta en personas con PD, AD y recientemente también en DLB (Disease in Lewy Bodies). Estudios afirman que en el momento del diagnóstico un 20.2 % de los pacientes padecen MCI. Además, a lo largo del seguimiento de la enfermedad esta cifra aumenta hasta un 40-50 % a los 5 años desde el inicio. En cambio, en la población la prevalencia de MCI en personas de 60 a 90 años se encuentra entre el 16 % y 20 %.

Este estado se define como una transición entre los pacientes de PD sin ningún tipo de deterioro cognitivo (PD-NC) y PDD.

- **Defecto cognitivo subjetivo:** es un término relativamente nuevo desarrollado en el área de la PD y AD. En este grupo se encuentran los pacientes recién diagnosticados con PD que aseguran tener problemas de memoria aunque no las pruebas médicas y test realizados no indiquen lo mismo.

2.2.3. Fisiopatología

Los pacientes con PD tienen en común la pérdida de las neuronas encargadas de producir dopamina en la sustancia negra, además de la acumulación de la proteína α -sinucleína de forma irregular. Actualmente, el declive cognitivo en pacientes con EP se relaciona con los cambios neurodegenerativos que produce esta enfermedad en el cerebro.

Degeneración de los sistemas de neurotransmisión

La pérdida de neuronas dopaminérgicas en el cuerpo estriado ocurre mayoritariamente en los pacientes que presentan PD-MCI. Sin embargo, estos pacientes preservan otros sistemas dopaminérgicos en el cerebro que en los pacientes con PDD desaparecen. Los pacientes con demencia tienen una pérdida de estos sistemas mayoritariamente en las regiones frontal, parietal y temporal. En pacientes sanos no se produce la pérdida de este tipo de neuronas.

Locus coeruleus noradrenérgico y sistemas simpáticos

El Locus Coeruleus se encuentra como principal fuente del neurotransmisor noradrenalina. Esta neurona es la encargada de la detección de señales y modulación de aspectos relacionados con la cognición como la atención o la memoria a largo plazo.

En pacientes con MCI se puede observar una relación significativa entre la reducción de neuromelanina. Además se puede establecer una relación entre las pérdidas del transportador de noradrenalina y el deterioro cognitivo en PD. Es por esto por lo que se establece como un biomarcador potencial para el estudio del declive cognitivo en trastornos neurodegenerativos los niveles de noradrenalina.

Sistemas colinérgicos del prosencéfalo basal

Las neuronas colinérgicas son las encargadas de proporcionar el control del circuito cognitivo, en concreto en el proceso de la memoria y la atención. Estas se encuentran en el prosencéfalo basal (parte inferior frontal del cerebro). En pacientes con PD, se puede relacionar una reducción del volumen de la región colinérgica con el deterioro cognitivo. En pacientes con PD-NC se puede predecir un diagnóstico de declive cognitivo en los siguientes 5 años.

Además, los diagnósticos de PDD están relacionados con una pérdida de fibras colinérgicas superior a la pérdida de este tipo de neuronas, interactuando así con una mayor pérdida de dopamina. Todo esto posibilita de una forma más rápida el declive cognitivo. En diferentes estudios se ha podido observar durante el desarrollo de PD-MCI y progreso hacia PDD un aumento selectivo en la inervación de las neuronas colinérgicas por fibras que contienen galanina como una respuesta a la lesión. Esto es provocado mayoritariamente por el aumento de α -sinucleína en las células.

2.2.4. Neuropatología

La neuropatología es el campo que se encarga del estudio de las enfermedades propias del sistema nervioso, especialmente a nivel tisular. Examina las alteraciones en los tejidos nerviosos y busca comprender los procesos patológicos que subyacen a diversas condiciones neurológicas.

α -sinucleína

Aunque, como se ha descrito antes, las deposiciones anormales de α -sinucleína en las neuronas es una de las causas de la PD, se quiere saber cuando esto puede tener un efecto significativo sobre el deterioro cognitivo del paciente. En la corteza etorrinal, la densidad de α -sinucleína diferencia a los pacientes que derivan en demencia y los que no. Esto supone que las personas con depósitos elevados de esta proteína en el neocortex tengan aproximadamente el doble de disminución cognitiva en un año.

La α -sinucleína interactúa además con el ADN, suponiendo una peor reparación de este en PD. Las mitocondrias también se ven gravemente afectadas, reduciendo así su número en las células.

Diagnóstico, detección y prevención

- **Diagnóstico:** se requiere un diagnóstico preciso del deterioro cognitivo en pacientes con EP para así poder establecer de manera efectiva un asesoramiento y, en la medida de lo posible, un tratamiento.

Este se puede hacer de dos formas. La primera es a partir de una evaluación de la función cognitiva de forma global. La segunda forma incluye una evaluación detallada de carácter neuropsicológico, incluyendo el estudio de factores como la memoria, la visión visuoespacial o el lenguaje (entre otros). El uso de biomarcadores será clave para permitir identificar las causas y, en algunos casos, favorecer la predicción de un posible deterioro cognitivo del paciente.

- **Detección:** la detección de la función cognitiva en pacientes con PD resulta primordial para poder suministrar el tratamiento correcto. Para ello, existen herramientas estandarizadas que permiten la correcta detección. Entre ellas se encuentran la evaluación MoCA, MDRS-2 (*Mattis Dementia Rating Scale Second Edition*) y PD-CRS (*Parkinson's Disease – Cognitive Rating Scale*). Entre todas ellas, el MoCA es el más utilizado en los pacientes con EP, teniendo una efectividad de un 68 % en pacientes con PD-MCI. Este test permite una evaluación neuropsicológica de forma detallada, teniendo en cuenta otros factores previos.

Existen pruebas neuropsicológicas estandarizadas a partir de una norma poblacional. En los resultados de estas tienen en cuenta factores como la edad o el nivel educativo del paciente. Esto supone que la selección de pruebas a realizar se haga en base a la presencia de las normas adecuadas de la población local para obtener así los mejores resultados posibles.

Actualmente, no se cuentan con recursos validos capaces de detectar la presencia de SCD en pacientes con EP, aunque esta puede ser informada por el paciente, los familiares o el mismo médico. Aunque la presencia de SCD no se relaciona con ningún tipo de declive cognitivo en el paciente según las pruebas reguladas, en algunos estudios sugiere que esto puede suponer un mayor riesgo de desarrollar en alguna fase de la enfermedad el deterioro cognitivo. Dado que los pacientes con EP pueden sufrir ansiedad o depresión, es importante tener esto en cuenta especialmente en los sujetos con SCD.

Para el diagnóstico de PD-MCI es necesario diferenciar los efectos del deterioro cognitivo para diferenciarlo de PDD. Esto es posible gracias a métodos como PD-CFRS (*Parkinson's Disease – Cognitive Functional Rating Scale*) o PDAQ (*Penn Parkinson's Daily Activities Questionnaire*). Se pueden diferenciar dos niveles de evaluación de MCI. El primero se basa en una evaluación cognitiva o pruebas neuropsicológicas limitadas, mientras que el nivel dos contempla pruebas neuropsicológicas más exhaustivas con los cinco dominios.

El diagnóstico de los pacientes con PDD es importante para poder implementar un tratamiento adecuado. El paciente con PDD se caracteriza por tener una

disminución del dominio cognitivo lo suficientemente grande como para imposibilitarle a la hora de realizar tareas de la vida diaria de manera independiente. Además, es frecuente que presenten apatía, alteraciones del estado de ánimo o alucinaciones (entre otros). Para la detección de demencia se utilizan las mismas herramientas que MCI.

Para entender el funcionamiento y proceso del deterioro cognitivo, es esencial el uso de biomarcadores. Estos ayudan a identificar los pacientes con riesgo de un declive cognitivo temprano. Uno de los marcadores predictivos más utilizados es la atrofia temporo-parietal detectada a partir de la resonancia magnética.

Por otra parte, el hipometabolismo en las regiones frontal medial y parietal utilizando FDG-PET (*Fluorodeoxyglucose-Positron Emission Tomography*) se asocia con una disminución de la función ejecutiva y de la memoria. En cuanto a biomarcadores específicos para patologías asociadas con el deterioro cognitivo cabe destacar los marcadores CSF de la patología AD que pueden predecir el desarrollo cognitivo futuro del paciente. A día de hoy, es complicado establecer ciertos marcadores relacionados con α -sinucleína debido al papel que toma esta proteína en PD.

2.2.5. Tratamientos para el deterioro cognitivo en la PD

Farmacoterapia para la demencia

Hasta la fecha, el único RCT (Ensayo clínico aleatorizado) para pacientes con PDD fue para el inhibidor de la colinesterasa. Esto supone que el único medicamento aprobado por la FDA (*Food and Drug Administration*) sea la rivastigmina oral para el tratamiento de PDD en niveles leves a moderados. Este fármaco supone una mejora en la cognición y algunos síntomas conductuales.

Otros posibles medicamentos como donepezil y la galantamina han sido calificados como posiblemente útiles por el Comité de Medicina Basada en la Evidencia de la Sociedad Internacional de Parkinson y Trastornos del Movimiento debidos a sus efectos, aunque no tienen evidencia comprobada para el tratamiento específico de PDD.

Tratamiento para MCI

A día de hoy no existen tratamientos predeterminados para pacientes con PD-MCI. Dado que este estado de la enfermedad es, en muchos casos, la transición a PDD, los tratamientos suministrados se focalizan en frenar esta evolución intentando modificar los síntomas a largo plazo y los efectos de la enfermedad. Actualmente se están realizando estudios con un receptor nicotínico de acetilcolina α -7 y múltiples tratamientos no farmacológicos.

2.3. Diagnosing PD-MCI by MDS TaskForce Criteria [21]

2.3.1. Introducción

Actualmente, se estima que entre un 20 % y un 50 % de los pacientes con PD presenta este trastorno que, en ocasiones progresa a la demencia. Ante esto la organización MDS ha establecido diagnósticos para la detección de los pacientes PD-MCI, diferenciando dos niveles: el primero hace referencia a una evaluación abreviada mientras que el segundo nivel hace referencia a una evaluación completa que requiere pruebas neuropsicológicas formales con al menos dos pruebas en cada uno de los cinco dominios cognitivos.

Aunque la MDS ha propuesto estas pruebas para confirmar la validez del diagnóstico, reconoció que era necesario una investigación más exhaustiva acerca del número y tipo de pruebas por dominio cognitivo.

A partir de la investigación de datos combinados se está llevando a cabo la validación de los criterios establecidos por la MDS (Sociedad de Trastornos del Movimiento) acerca de los pacientes PD-MCI. De forma paralela, otros estudios están aplicando los criterios establecidos en el nivel II a pacientes PD sin deterioro cognitivo (PD-CU), informando PD-MCI en un 20 % a 62 %.

Es por esto que todavía no existe un consenso sobre la batería óptima de pruebas neuropsicológicas para el diagnóstico de PD-MCI. Anteriormente, se demostró que una batería de 19 pruebas de nivel II proporcionó la mejor clasificación para diagnosticar PD-MCI. Sin embargo, las baterías completas de pruebas pueden tardar y complicadas para los pacientes y administradores de pruebas, por lo que el objetivo actual es optimizar la clasificación diagnóstica con una batería lo más reducida posible.

2.3.2. Dominios cognitivos

Los cinco dominios cognitivos son las categorías utilizadas para describir diferentes tipos de procesos mentales y habilidades de pensamiento que los seres humanos utilizan para procesar la información. Estos cinco dominios son los siguientes:

1. **Dominio cognitivo de conocimiento:** Este dominio se refiere al conocimiento factual y conceptual que las personas tienen sobre el mundo, incluyendo información sobre hechos, conceptos, principios y teorías.
2. **Dominio cognitivo de comprensión:** Este dominio se refiere a la capacidad de las personas para comprender la información, interpretarla y hacer inferencias a partir de ella. Incluye habilidades como la lectura crítica, la interpretación de textos, la comprensión de ideas complejas y la inferencia.
3. **Dominio cognitivo de aplicación:** Este dominio se refiere a la capacidad de aplicar el conocimiento y las habilidades en situaciones prácticas. Incluye habilidades como la resolución de problemas, la toma de decisiones, la planificación y la ejecución de tareas.

4. **Dominio cognitivo de análisis:** Este dominio se refiere a la capacidad de analizar y descomponer la información en sus componentes más básicos, identificar patrones y relaciones, y evaluar argumentos. Incluye habilidades como el análisis de datos, el pensamiento crítico y la evaluación de argumentos.
5. **Dominio cognitivo de síntesis:** Este dominio se refiere a la capacidad de integrar información de diferentes fuentes, crear nuevas ideas y generar soluciones creativas a los problemas. Incluye habilidades como la creatividad, la síntesis de información y la generación de nuevas ideas.

2.3.3. Métodos

En el estudio mencionado en el artículo se incluyeron 76 sujetos con PD sin demencia para formar parte de un estudio prospectivo de marcadores clínicos e de neuroimagen de la afectación cognitiva de la PD. Los pacientes pasaron por evaluaciones clínicas detalladas y neuropsicológicas, incluyendo la MMSE (Mini-Prueba de Estado Mental) y 19 pruebas recomendadas por la MDS que representan cinco dominios cognitivos. Las puntuaciones brutas se transformaron en puntuación z en base a datos normativos. Se calcularon puntuaciones compuestas para cada prueba de memoria y se determinaron las puntuaciones de los dominios cognitivos promediando la puntuación z de las pruebas neuropsicológicas dentro de cada dominio.

Una vez se terminó con la evaluación de los sujetos a estudio, se procedió a determinar una clasificación entre PD-MCI y PD-NC mediante dos métodos. En el primero se tuvo en cuenta los datos clínicos y neuropsicológicos de cada paciente en una conferencia de consenso determinando la clasificación correspondiente a partir de un proceso uniforme y discutido de forma individual.

En cuanto al segundo método se refiere, se aplicaron las pautas de diagnóstico de la MDS, incluyendo la presencia de discapacidad en al menos dos pruebas neuropsicológicas. Esto podría deberse a dos pruebas dañadas en un dominio cognitivo o una prueba dañada den dos dominios cognitivos diferentes. Los pacientes con PD que no cumplen los criterios PD-MCI establecidos por la MDS eran clasificados como PD-NC.

Como se menciona anteriormente, el estudio tiene como objetivo principal determinar el número necesario de pruebas utilizadas para determinar el deterioro cognitivo en pacientes con PD-MCI. Para ello la investigación calculó la probabilidad de determinar este deterioro cognitivo en base al número de pruebas utilizadas por dominio. Se aplicó un modelo de 10 pruebas para clasificar el PD-MCI basándose en los criterios de nivel II de dos pruebas por dominio seleccionados por medio de la clasificación LASSO. Este modelo fue evaluado frente al diagnóstico consensuado, determinando sus valores de sensibilidad, especificidad, predictivo positivo y negativo, y kappa. Además, se investigaron los subtipos de PD-MCI para identificar los que presentan deterioro en un solo dominio o en múltiples dominios.

2.3.4. Resultados

Una vez realizado el estudio al completo se obtuvieron los siguientes resultados: Según el diagnóstico consensuado un 63.2% de los pacientes sometidos a estudio fueron clasificados como PD-MCI.

Teniendo en cuenta el número de pruebas por dominio se demostró que a mayor inclusión de tests, mayor posibilidad existía de detectar una discapacidad en un a prueba. No obstante, este se estabilizó en dos pruebas en los dominios de Atención / Memoria de trabajo y Función Ejecutiva (36.8% y 57.9%, respectivamente), y no aumentó con tres o cuatro pruebas en el dominio. Los otros dominios (Lenguaje, Memoria, Función Visoespacial) demostraron un aumento menor al 5% en la probabilidad máxima de tener al menos una prueba discapacitada cuando se utilizaron más de dos pruebas. Por lo tanto, se persiguió una batería con dos pruebas en cada uno de los dominios en análisis futuros.

En cuanto a la selección de las mejores pruebas, se utilizó el método LASSO para su determinación, clasificándolas según su coeficiente de regresión (a mayor coeficiente mayor rango). Las pruebas con mejor rendimiento fueron las siguientes:

1. TMT-A, SDMT (Atención / Memoria de trabajo)
2. TMT-B, Dibujo del Reloj (Función Ejecutiva)
3. Boston Naming Test, Nombramiento de animales (Lenguaje)
4. FCSRT, Memoria Figural (Memoria)
5. JLO, Intersección de pentágonos (Función Visoespacial).

El uso de este listado de pruebas (dos por dominio) proporcionó una sensibilidad del 81.3% y una precisión del 85.7% para diagnosticar PD-MCI mediante los criterios del nivel II. Los resultados obtenidos demostraron una buena concordancia con el diagnóstico determinado a partir del consenso. En cuanto a la especificidad para clasificar a los sujetos como PD-MCI y PD-NC se determinó un valor de 82.9% en comparación con la clasificación consensuada.

2.3.5. Conclusión

En resumen, el estudio realizado es el primero en examinar los criterios del Task Force nivel II de MDS para la enfermedad de Parkinson con declive cognitivo. Se encontró que una batería neuropsicológica de dos pruebas por dominio es un enfoque eficiente y robusto para diagnosticar PD-MCI. Se seleccionaron diez pruebas que ofrecieron una alta sensibilidad y especificidad en la clasificación de los pacientes como PD-MCI. La batería propuesta incluye pruebas como TMT-A, SDMT, TMT-B, Clock Drawing Test, Boston Naming Test, Animal Naming, FCSRT, Figural Memory, JLO e Intersecting Pentagons. El estudio tiene como fortaleza una cohorte de PD bien definida y una evaluación completa, pero también tiene limitaciones, como una muestra relativamente pequeña y un grupo de alta educación. Se necesitan más estudios para validar los resultados y determinar qué baterías y pruebas son más sensibles a los cambios a lo largo del tiempo y con la intervención.

2.4. Toward a biological definition of Alzheimer's disease [25]

2.4.1. Introducción

El marco de investigación NIA-AA propone una definición biológica de la enfermedad de Alzheimer (AD) en lugar de una definición basada en los efectos clínicos visibles y síntomas. La enfermedad se define a través de marcadores biológicos, como la deposición de b-amiloide, la tau patológica y la neurodegeneración, para identificar a los participantes en la investigación y para caracterizarlos en el Alzheimer continuo.

Este marco de investigación no limita la investigación en la demencia en general, sino que la fortalece al proporcionar una definición biológica de una de las causas de la demencia, la AD. La evaluación neuropatológica es el estándar para definir la AD y los marcadores validados y ampliamente utilizados son sustitutos de los cambios neuropatológicos en la AD.

2.4.2. Enfermedad de Alzheimer

Actualmente, se hace referencia al término "enfermedad de Alzheimer" cuando se habla de una serie de cambios neuropatológicos. Esta enfermedad es definida in vivo por biomarcadores y por examen post-mortem, sin tener en cuenta síntomas clínicos.

Para tener en cuenta este enfoque de la AD se tuvo que realizar una distinción entre los términos *síndrome* y *enfermedad*. Aunque en ciertos casos se podrá ver definido la AD en personas vivas como un síndrome específico, es decir, una demencia multidominio, la realidad es que la demencia no es una enfermedad como tal. Esta es un síndrome compuesto por signos y síntomas que pueden ser causados por diferentes enfermedades (entre las que se puede encontrar AD).

Si se define la AD como un síndrome se encuentran dos problemas: no es sensible ni específico a los cambios neuropatológicos que definen la enfermedad, y no puede identificar a las personas que tienen evidencia biológica de la enfermedad pero aún no manifiestan signos o síntomas. Esto supone que aparezca una la definición de AD por medio de biomarcadores indicativos de cambios neuropatológicos independientes de los síntomas clínicos.

2.4.3. Biomarcadores de la AD

El esquema de investigación AT(N) reconoce tres grupos de biomarcadores basados en la naturaleza del proceso patológico que cada uno mide.

- A: biomarcadores de placas $A\beta$
- T: biomarcadores de tau fibrilar
- (N): biomarcadores de lesión neuronal

El sistema AT(N) fue diseñado con un biomarcador en CSF y un biomarcador de imagen en cada uno de los tres grupos de biomarcadores, permitiendo la caracterización completa de los participantes en la investigación con solo imágenes o solo LCR.

Algunos grupos de investigación pueden preferir una combinación de biomarcadores de imagen y LCR para la caracterización AT(N).

Se acordó que solo los biomarcadores específicos de AD ($A\beta$ y tau patológico) deben considerarse como posibles definiciones de biomarcadores de la enfermedad. Se valoraron diferentes perfiles de biomarcadores. Estudios han demostrado que las personas con biomarcadores anormales de amiloide tienen una progresión más rápida de la atrofia, hipometabolismo y declive cognitivo que las personas sin evidencia de depósito de $A\beta$.

La evidencia de amiloides es la primera señal de cambios neuropatológicos de AD detectables en personas vivas. Por lo tanto, los biomarcadores anormales de amiloides podrían ser una afirmación de la existencia de AD. Sin embargo, tanto $A\beta$ como tau patológico son necesarios para cumplir con los criterios neuropatológicos de AD, lo que sugiere que deben estar presentes ambos biomarcadores para etiquetar a una persona como "enfermedad de Alzheimer". Esto supone que la etiqueta "cambio patológico de Alzheimer" se impondrá si solo hay evidencia de $A\beta$. Ambas definiciones son aplicadas independientemente de los síntomas clínicos.

2.4.4. Estados

Para poder determinar la gravedad de la enfermedad se se crearon dos clasificaciones. La primera se basa en los biomarcadores y las segunda en la disminución de la capacidad cognitiva. Aunque las medidas utilizadas para la definición de la enfermedad en si deben de específicas del AD, las utilizadas para indicar la gravedad no tienen necesariamente que serlo.

Los biomarcadores de $A\beta$ determinan si un individuo está en el continuo de la enfermedad de Alzheimer. Los biomarcadores patológicos de tau determinan si alguien que está en el continuo de la enfermedad de Alzheimer tiene AD (ambos $A\beta$ y tau son necesarios para un diagnóstico neuropatológico de la enfermedad).

Sin embargo, existen biomarcadores de lesión neurodegenerativa-neuronal y los síntomas cognitivos, que no son específicos para la AD, solo se utilizan para clasificar la gravedad, no para definir la presencia del continuo de la enfermedad de Alzheimer.

2.4.5. Perfiles y categorías según los biomarcadores

El sistema de marcadores AT(N) clasifica a los individuos en tres categorías de marcadores:

1. Individuos con marcadores de AD normales
2. Individuos en el continuo de la enfermedad de Alzheimer (subdivididos en cambios patológicos de Alzheimer y enfermedad de Alzheimer)
3. Individuos con un marcador de amiloide normal pero con T o (N) anormales o ambos. Este último perfil de marcadores indica evidencia de uno o más procesos neuropatológicos distintos de la enfermedad de Alzheimer y se ha etiquetado como "patofisiología no sospechada de Alzheimer".

La puntuación AT(N) de los marcadores es independiente de los síntomas clínicos.

La tasa de declive cognitivo es significativamente mayor para aquellos con deterioro cognitivo que tienen anormalidades tanto en un biomarcador de amiloide como en un segundo tipo, en comparación con aquellos que no tienen ninguno o solo una anormalidad. Estos datos establecen que la enfermedad más avanzada definida por biomarcadores previene una mayor probabilidad y un declive cognitivo más rápido. Por lo tanto, existe una base sólida de evidencia que demuestra que las *combinaciones de anormalidades de biomarcadores son útiles para clasificar el continuo de la enfermedad de Alzheimer*.

En ocasiones, para una mayor facilidad a la hora de la investigación, se trata a los biomarcadores como variables continuas. Este enfoque se suele utilizar para el cuidado clínico. Además, cabe destacar los avances en la medicina personalizada que permitirá el tratamiento individualizado de cada paciente a partir de los biomarcadores nombrados anteriormente combinado con los conocimientos de genética.

2.5. Brain Amyloid Contribution to Cognitive Dysfunction in Early PD [17]

2.5.1. Introducción

La enfermedad de Parkinson es una enfermedad neurodegenerativa que afecta a millones de personas en todo el mundo. Uno de los síntomas comunes de la enfermedad de Parkinson es la disfunción cognitiva, que puede incluir problemas de memoria, atención y toma de decisiones. Sin embargo, no está claro por qué algunos pacientes experimentan un declive cognitivo mientras que otros no.

Es por esto que en este estudio el objetivo principal es investigar la asociación entre la acumulación de amiloide en el cerebro, los niveles de amiloide en el CSF y la disfunción cognitiva en pacientes con PD en etapa temprana utilizando el conjunto de datos PPMI (Proyecto de Marcadores de Progresión de la Enfermedad de Parkinson). Los investigadores hipotetizaron que la acumulación de amiloide en el cerebro puede afectar las habilidades cognitivas tempranas en pacientes con PD.

2.5.2. Métodos

Este estudio se basa en el PPMI, que es una iniciativa de investigación global que tiene como objetivo identificar biomarcadores de la PD para mejorar el diagnóstico y el tratamiento. Se excluyeron del análisis a los participantes sanos, los que presentaban síntomas tempranos de la enfermedad y los sujetos sin relación con la cohorte genética, lo que resultó en una muestra final de 48 pacientes con PD sintomática investigados en cinco centros.

En cuanto a las variables demográficas se refiere, se incluyó la edad, años de educación y sexo, mientras que las variables clínicas incluyeron la duración de la enfermedad, la dosis diaria equivalente de levodopa y agonistas de dopamina, la gravedad de la enfermedad medida con la MDS-UPDRS III (Escala de Evaluación

de la Enfermedad de Parkinson de la Sociedad de Trastornos del Movimiento) y el funcionamiento diario utilizando la ADL (Escala de Actividades de la Vida Diaria).

La evolución cognitiva se valoró a través del MoCA, mientras que funciones cognitivas de pacientes con enfermedad de Parkinson se valoraron utilizando diferentes pruebas:

- JLO (Prueba de Juicio de Orientación de Benton)
- SDMT (Test de Símbolos y Dígitos Modales)
- HVLT-R (Test de Aprendizaje Verbal de Hopkins)
- LNS (Test de Secuencias de Letras y Números)
- Prueba fluidez semántica
- Escala de Depresión Geriátrica de 15 ítems

Los pacientes incluidos en el estudio eran independientes en sus actividades diarias y no presentaban demencia.

En una submuestra de los participantes del estudio, compuesta por 34 individuos con enfermedad de Parkinson y positividad de amiloide-beta (PDA+) y 34 individuos con enfermedad de Parkinson y negatividad de amiloide-beta (PDA-), se realizó una punción lumbar para obtener muestras de líquido cefalorraquídeo (CSF). Las muestras de CSF se analizaron para determinar las concentraciones de $A\beta$ 1-42, tau total (t-tau), tau fosforilada (p-tau) y α -sinucleína. También se calcularon las relaciones pTau y T-tau a $A\beta$ 1-42, así como la relación pTau a T-tau.

Por otro lado, en cuanto al uso de imágenes PET para la detección de $A\beta$ en el cerebro se refiere se utilizó un protocolo estandarizado. Se adquirieron las imágenes de acuerdo con un procesamiento a partir del software PMOD, con su posterior normalización a MNI space y su final conversión a Valor de Absorción Estándar (SUV) tomando la corteza cerebelosa como referencia. Se estableció un SUVR (Valor de Absorción Estándar de Ratio Compuesto) para cada sujeto y se consideró positivo si el valor era mayor a 1.43. Los sujetos se clasificaron como PDA+ o PDA- según este límite.

Para el análisis, y posterior estudio de los resultados de la investigación se utilizó una muestra con distribución no normal. Se utilizaron estadísticas descriptivas y no paramétricas (prueba U de Mann-Whitney) para analizar los datos demográficos, clínicos e de imagen. Las variables categóricas se compararon utilizando la prueba exacta de *Fisher*. Para estudiar la asociación entre la cognición y las deposiciones de amiloide se realizaron correlaciones de rango de *Spearman*.

2.5.3. Resultados

Una vez realizadas las pruebas necesarias a los pacientes sometidos a estudio se procedió al análisis y obtención de los siguientes resultados: de los 48 pacientes con PD incluidos en el estudio, 10 (21%) presentaron un SUVR de florbetaben PET superior a 1.43 y fueron clasificados como PDA+ mientras que los 38 restantes se

clasificaron como PDA-.

Además, los pacientes PDA+ presentaron una puntuación significativamente menor en las pruebas de MoCA y SDMT en comparación con los pacientes PDA- ($p=0.01$ y $p=0.03$, respectivamente). También se observaron diferencias significativas entre los grupos en las relaciones CSF p-tau/A β 1-42 ($p < 0.002$) y T-tau/A β 1-42 ($p < 0.001$).

El análisis de la imagen PET reveló un aumento significativo en la captación en varias regiones, especialmente en la corteza, núcleos subcorticales y en la protuberancia.

Se encontró una correlación negativa moderada entre la puntuación de MoCA y la captación de amiloide en regiones corticales y subcorticales entre los grupos PDA+ y PDA-. También se encontró una asociación similar entre la puntuación de SDMT y la captación de florbetaben SUVR en regiones corticales seleccionadas. Esto puede significar que una mayor carga de amiloide en áreas corticales y subcorticales esté asociada con un peor desempeño en tareas de cognición global, atención, escaneo visual y velocidad de procesamiento.

2.5.4. Conclusiones

El objetivo principal del estudio se centró en examinar la presencia y los efectos del depósito de amiloide en el rendimiento cognitivo en pacientes con PD en la etapa temprana sin presencia de demencia. Los resultados indicaron que en torno al 21 % de los pacientes sometidos a estudio tenían una captación cerebral positiva de amiloide [18]florbetaben en varias regiones neocorticales y subcorticales, pero la prevalencia general fue similar a la de las poblaciones de control emparejadas por edad, lo que sugiere que el aumento de la deposición de amiloide no confiere un riesgo específico para el desarrollo de la PD.

Por otro lado, la amiloidosis se asoció con un aumento en la relación p-tau/A β y T-tau/A β y, el estudio sugiere que la patología de PD+AD aumenta con tanto un alto nivel de T-tau/A β como una alta carga de amiloide cerebral.

Desde una perspectiva cognitiva, los pacientes con PD y amiloidosis tuvieron puntuaciones medias más bajas en la escala cognitiva MoCA y en la medida de velocidad de procesamiento de información SDMT. Estos resultados, junto a otros estudios como Schrag [41] promueven la idea de que el deterioro cognitivo en los pacientes se puede ver contribuido por las cargas de amiloide en zonas corticales y subcorticales.

Además, en ciertos pacientes se han podido encontrar patrones estructurales de deposición de esta proteína en ciertas áreas comparables con la enfermedad de Alzheimer (AD). Esto permite pensar que la degeneración de la materia gris frontal puede estar causada, en parte, por la deposición de amiloide en ciertas regiones. Sin embargo, se requieren de más estudios que puedan corroborar esta información.

Tras las evaluaciones con imágenes PET se ha podido determinar que, tanto los pacientes con PD, como los sujetos control, tienen deposiciones de amiloide similares, que aumenta de forma genérica con la edad. Esto sugiere que la deposición de esta proteína en sí no supone un riesgo a la hora de desarrollar PD. De hecho, se ha podido comprobar que las personas con deterioro cognitivo leve y que presentan cambios en las imágenes MRI hacia un proceso neurodegenerativo de la AD tienen riesgo de desarrollar demencia en comparación con los controles sanos negativos, pero, si además existe una deposición adicional de amiloide este riesgo puede ser hasta tres veces mayor. Finalmente, mediante imágenes PET se pudo comprobar la similitud entre $A\beta$ cortical y tau en pacientes PD y controles, corroborando así la importancia de la proteína α -sinucleína en el deterioro cognitivo durante la enfermedad.

La obtención de todos estos resultados sugiere que la coexistencia de las proteínas amiloide y sinucleína contribuyen al desarrollo de la PD, aumentando así el riesgo de deterioro cognitivo. No obstante, se requieren estudios longitudinales para poder confirmar estos datos y así, permitir desarrollar técnicas para frenar el declive cognitivo en pacientes con esta enfermedad.

2.6. Longitudinal survival analysis and two-group comparison for MCI to AD [34]

2.6.1. Introducción

La AD es un trastorno degenerativo cerebral que causa un trastorno en la respuesta cognitiva que puede derivar en demencia a edades tardías. Se espera que, en los próximos años, haya un aumento de casos de AD entre la población debido al aumento de la esperanza de vida. El deterioro cognitivo leve (MCI) se considera la fase intermedia entre un envejecimiento saludable y la demencia y un elemento clave para poder estudiar el desarrollo de la enfermedad. Predecir el diagnóstico de AD resulta clave para poder ayudar a tomar decisiones clínicas sobre el tratamiento o detección de sujetos en riesgo de convertir.

Este estudio ha analizado diversas modalidades de indicadores para estimar la progresión de AD, incluyendo elementos como biomarcadores de neuroimagen y pruebas neuropsicológicas o sanguíneas. Dependiendo de que marcadores son utilizados se obtienen diferentes predicciones de la progresión de MCI a AD (pMCI).

En cuanto a los test neuropsicológicos se refiere, se ha demostrado su eficacia a la hora de predecir esta progresión. Además, los marcadores de neuroimagen proporcionan un respaldo de estos resultados y se complementan entre sí. No obstante, la mayoría de modelos predictivos sobre la conversión de MCI a AD utilizan únicamente datos de base. El uso de un estudio longitudinal permitiría la observación de patrones de cambio a medida que avanza la enfermedad que puede modificar el diagnóstico, haciéndolo más preciso.

Actualmente no hay homogeneidad en los grupos pMCI y sMCI por lo que no se

tiene una estimación concreta de la conversión de MCI a AD. Ante ello, se proponen los modelos de supervivencia, considerando en el grupo clínico los tiempos de conversión y censura. A partir de esto se aplica los modelos de Cox.

En este estudio se han comparado dos enfoques diferentes de modelos predictivos multivariados longitudinales para poder estimar la progresión de MCI a AD. El primero se basa en la comparación de los grupos MCI estable y progresivo, mientras que el segundo se basa en realizar un análisis de supervivencia a partir del modelo de Cox extendido y la regresión logística. Para ambos se utilizó un número reducido de marcadores basados en MRI con medidas cognitivas estándar y una modelización LME.

En el primer enfoque se trató en residuos la trayectoria longitudinal marginal, mientras que en el análisis de supervivencia se utilizó razones de riesgo. Al final del estudio se determinó que los modelos predictivos basados en la supervivencia con datos de diferentes fuentes presentaban mejores resultados de predicción a partir del segundo año de seguimiento que los modelos predictivos de comparación de grupos.

2.6.2. Materiales

Para el estudio se utilizaron los datos de la Iniciativa de ADNI (Neuroimagen de la Enfermedad de Alzheimer) con el fin de comparar dos enfoques diferentes de modelos predictivos multivariados longitudinales para estimar la conversión de la MCI a la AD. En estos datos se incluyeron los relacionados con la resonancia magnética (MRI) longitudinal del cerebro obtenidos al inicio y en intervalos regulares de tiempo, además de datos longitudinales en diferentes meses. Con estos datos se pretendió predecir la conversión de MCI a AD en tres años, incluyendo al inicio pacientes con diagnóstico MCI.

En cada visita, se realizaba un diagnóstico y se establecía si el paciente había convertido a AD (pMCI) o no (sMCI). En cuanto a las características de los pacientes, no se observaron diferencias significativas en edad y sexo entre los grupos clínicos, pero sí hubo diferencias significativas en las puntuaciones de la MMSE.

2.6.3. Métodos

Para predecir la conversión de MCI a AD se utilizó un método determinado por tres etapas.

1. Extracción de datos, donde se obtuvieron las características de la MRI y medidas cognitivas.
2. selección de características a partir del algoritmo mRMR que propuso subconjuntos de marcadores para la previsión.
3. Clasificación longitudinal de los resultados de ambos métodos a partir de la validación cruzada.

Para predecir la conversión de MCI a AD se utilizó un método determinado por tres etapas. La primera de ellas consistió en la extracción de datos, donde se obtuvieron las características de la MRI y medidas cognitivas. En la segunda etapa se procedió a la selección de características a partir del algoritmo mRMR que propuso subconjuntos de marcadores para la predicción. Finalmente, en la tercera se evaluaron dos enfoques de clasificación longitudinal. Los biomarcadores seleccionados para la predicción fueron los volúmenes de regiones cerebrales específicas y el grosor cortical en zonas vulnerables a la AD.

Para determinar si los pacientes tenían demencia se utilizó las características neuropsicológicas (NMs), las cuales valoran diferentes dominios cognitivos como el aprendizaje, el lenguaje o la memoria. Además se valoraron los resultados obtenidos en los siguientes tests: Test de Aprendizaje Verbal Auditivo de Rey (RAVLT), la Escala de Evaluación de la Enfermedad de Alzheimer - Subescala Cognitiva (ADAS-cog), el MMSE, la Evaluación Clínica de Demencia (CDR) y el Cuestionario de Actividades Funcionales (FAQ).

A continuación, se procesaron los datos longitudinales mediante modelos de efectos mixtos. Se optó por etapa combinada de ordenamiento de características y clasificación para identificar los subconjuntos de características óptimas y, posteriormente, se definieron subconjuntos a partir del algoritmo mRMR.

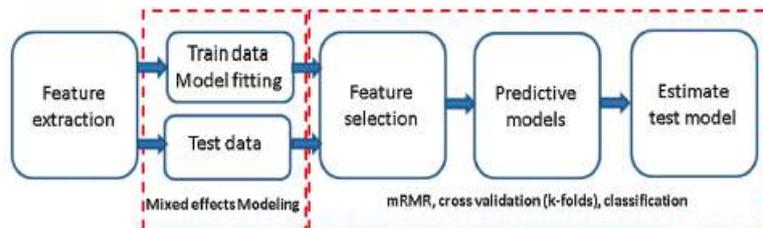


Figura 2.1: Visión general del método propuesto.

Clasificación longitudinal

Cox propuso un modelo de riesgos proporcionales para el análisis de datos de supervivencia [10]. El modelo de Cox expresa la relación entre una función de riesgo en el tiempo t para un sujeto con un vector $X = \{X_1, X_2, \dots, X_p\}$ independiente del tiempo:

$$h(t, X) = h_0(t) \exp \left(\sum_{k=1}^p \alpha_k X_k \right) \quad (2.1)$$

Donde:

- $h(t, X)$: función de riesgo
- $h_0(t)$: función de riesgo base
- X : vector explicativo
- α : vector de coeficientes de regresión.

Se incluyeron características como la edad, el sexo y los años de educación de los pacientes, además del genotipo APOE, que es el factor de riesgo genético más prevalente en la AD. También se tuvo en cuenta la relación entre el APOE y el tiempo, aunque los modelos de efectos mixtos (LME) con estas variables no demostraron grandes diferencias en la predicción. Los mejores resultados se obtuvieron al incluir la edad, el sexo y los años de educación como covariables en los modelos LME.

Para los modelos predictivos a partir de un análisis de supervivencia, se utilizaron las medidas longitudinales y resultados de los test neuropsicológicos. Se utilizaron la edad, el sexo y la educación como covariables. Se estimó el valor de cada marcador para cada sujeto y momento temporal, y se calculó el tiempo de conversión para aquellos cuya enfermedad había progresado a AD. Se construyó un modelo de Cox extendido para los puntos de tiempo significativos, y se calcularon las razones de riesgo y se convirtieron en probabilidades de conversión mediante regresión logística. Se construyeron cuatro modelos predictivos para el inicio del estudio y 12, 24 y 36 meses de seguimiento.

Selección de características y construcción de modelos predictivos

Se procedió a una validación cruzada anidada para impedir el sobreajuste y obtener estimaciones imparciales del rendimiento del modelo. Esto supuso dos bucles anidados: un bucle interno para seleccionar los subconjuntos de características óptimos y uno externo encargado de estimar el rendimiento del modelo. Se utilizó una validación cruzada de 10 pliegues para ambos bucles.

En el bucle interno, un método de remuestreo buscó los subconjuntos de características que aparecían con más frecuencia utilizando el algoritmo mRMR. Se construyeron modelos predictivos utilizando los datos de entrenamiento con los subconjuntos de características seleccionados. El procedimiento se repitió 100 veces para cada bucle interno. Este enfoque ayudó a evitar la doble inmersión y obtener estimaciones confiables del rendimiento del modelo. A continuación, se seleccionaron las tres combinaciones de marcadores con un mejor desempeño en términos de precisión de clasificación evaluando los modelos correspondientes.

En el bucle externo de la validación cruzada, se construyó un modelo predictivo a partir de los datos de entrenamiento para cada selección de características candidatas y se evaluó su desempeño con los datos de prueba retenidos, que no se utilizaron durante la selección de características, selección de modelos o construcción del modelo final. Para una mejor replicabilidad, se repitió el procedimiento de validación cruzada anidada de 10 pliegues con diferentes particiones de los datos, generando múltiples valores de estimación del desempeño. En total, se evaluaron 30 000 subconjuntos seleccionados para cada dimensión.

Se seleccionaron modelos predictivos con combinaciones de características que aparecían con más frecuencia (número de veces que se evaluó la combinación propuesta de características mediante el procedimiento de validación cruzada), valores del área bajo la curva (AUC) más altos y una sensibilidad y especificidad equilibradas.

2.6.4. Resultados

Se procesaron un total de 1330 escaneos de 321 sujetos utilizando los enfoques propuestos. Para validar la consistencia del proceso de procesamiento de imágenes longitudinales aplicado, se utilizó un control de calidad entre los escaneos y las segmentaciones de cada sujeto. Para cada visita de cada sujeto, se obtuvieron volúmenes subcorticales y mediciones de tomografías computarizadas (CT). La etapa de extracción de características resultó en 40 predictores de MRI basados en estructuras corticales y subcorticales y 11 NMs. Se realizó un análisis univariado de cada marcador y luego se llevó a cabo un análisis multivariado para generar los modelos predictivos propuestos, considerando las 51 características.

Rendimiento de los modelos predictivos de progresión de MCI a AD

Se utilizó el proceso indicado anteriormente para desarrollar el modelo predictivo de la conversión de MCI a AD, seleccionando así los mejores subconjuntos de características. Los datos de entrenamiento se reutilizaron para construir modelos predictivos, y se utilizaron dos enfoques diferentes para la información de los sujetos de prueba. En el enfoque de comparación de dos grupos, se utilizó el residuo de la trayectoria longitudinal en el análisis discriminante lineal, mientras que en los modelos predictivos basados en análisis de supervivencia, las variables exploratorias de los sujetos de prueba se compararon con la población de referencia de esa visita censurada.

Para los modelos de una sola fuente, los marcadores seleccionados fueron similares en ambos enfoques y los modelos que utilizaron el análisis de supervivencia mostraron mejores valores de sensibilidad y especificidad en comparación con el enfoque de comparación de dos grupos. Para los modelos de múltiples fuentes, los marcadores elegidos fueron similares en ambos enfoques y las medidas de volumetría y espesor cortical se centraron en el hipocampo, así como en el lóbulo temporal medial y parietal inferior. En general, los modelos de supervivencia mostraron puntuaciones mejores y más consistentes que los modelos de comparación de dos grupos en el transcurso del estudio longitudinal.

Al utilizar el modelo extendido de Cox, se pudo observar como mejoraron las predicciones de conversión de MCI a AD respecto al enfoque de comparación de dos grupos a partir del segundo año de estudio. Por otro lado, se obtuvieron puntuaciones de referencia comparables con otros enfoques transversales, pero con un vector de características más pequeño e identificable. Además, considerar la trayectoria longitudinal de los pacientes llevó a una mejora en la precisión de la predicción de la conversión a AD en comparación con otros enfoques.

2.6.5. Conclusiones

A diferencia de estudios previos que se enfocaron en un único biomarcador, este estudio desarrolló modelos predictivos combinando un pequeño subconjunto de marcadores basados en imágenes MRI con medidas cognitivas estándar y datos demográficos y clínicos como covariables. Los modelos se construyeron utilizando datos

longitudinales de pacientes con MCI. Se extrajeron datos longitudinales de una amplia gama de características corticales y subcorticales de las imágenes MRI y de pruebas neuropsicológicas. Se preseleccionaron subconjuntos de características de diferentes dimensiones utilizando el algoritmo mRMR, y se utilizó un método de remuestreo para buscar los subconjuntos de características que aparecían con mayor frecuencia.

Posteriormente, se evaluaron los subconjuntos de características propuestos en términos de precisión de clasificación validada cruzada. Se seleccionaron los modelos predictivos con mayor frecuencia de aparición, valores de AUC más altos y una buena sensibilidad y especificidad equilibrada. Este enfoque ayudó a mejorar la precisión en la predicción de la conversión a AD y a identificar un pequeño grupo de marcadores interpretables útiles en la comprensión de la evolución de la enfermedad.

Se realizaron dos enfoques para los modelos predictivos. Ambos seleccionaron un pequeño número de marcadores fácilmente interpretables, y los modelos predictivos incorporaron información demográfica y clínica como covariables. El estudio recomienda utilizar validación cruzada de k-fold para evaluar los modelos predictivos de la progresión de MCI a AD.

El estudio encontró que las características más seleccionadas en los NMs incluían ADAS13, FAQ y RAVLT, mientras que los volúmenes del hipocampo y del palidum y las medidas de TC para varias regiones cerebrales temporoparietales, con una preferencia por el hemisferio izquierdo, eran las características más frecuentemente seleccionadas en los datos de MRI. El estudio también encontró que la combinación de ADAS13 con otros marcadores propuestos, ya sea basados en datos de MRI u otros NMs, mejoró los resultados de predicción.

2.7. Estimating long-term multivariate progression from short-term data [13]

2.7.1. Introducción

Para intentar entender la progresión de las enfermedades neurodegenerativas y desarrollar algún método de detección normalmente se procede al estudio de diferentes cohortes en etapas de la enfermedad en diferentes periodos de tiempo. Para ello, es necesario detectar el "tiempo cero" de la enfermedad para poder ajustar diferentes modelos que permiten estimar la progresión.

La enfermedad de Alzheimer (AD) es un ejemplo de estas, dado que se cree que se desarrolla mucho antes del inicio de los síntomas. Ante esto la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) realizó un seguimiento a pacientes sin deterioro cognitivo (CN), con deterioro cognitivo temprano (EMCI) y tardío (LMCI) y aquellos con posible AD leve. Se tiene estimado que el análisis dure unos 6 años y este se ayuda de imágenes de resonancia magnética (fMRI) y tomografías por emisión de positrones (PET).

Algunos autores ya han propuesto modelos donde queda reflejado la trayectoria

de diferentes biomarcadores causantes de la AD a lo largo de los años antes de la aparición de los primeros síntomas de la demencia [24] [26]. Sin embargo, actualmente las investigaciones se enfocan a estudios a corto plazo.

Con el objetivo de encontrar un modelo capaz de describir la progresión de los biomarcadores de la enfermedad, se aplicó la regresión de auto-modelado (SEMOR). Este enfoque permite ajustar diferentes conjuntos de curvas bajo una forma común en la progresión de AD. Dado que las mediciones de cada sujeto tienen diferente duración y no se tienen disponibilidad de todas ellas, se utilizan efectos aleatorios para minimizar el efecto que las diferencias individuales ante el grupo.

Se aplicó SEMOR a los datos disponibles en ADNI para una modelar la progresión de las medidas cognitivas y funcionales además de los biomarcadores en distintas etapas de AD. Fueron incluidos tanto sujetos cognitivamente normales como con demencia. Se concluyó que SEMOR puede proporcionar información sobre la evolución de la enfermedad y puede servir de ayuda a la hora de seleccionar las características necesarias para conseguir una detección temprana de la enfermedad.

Aunque existen estudios sobre la AD a largo plazo, se ha podido comprobar que estos no proporcionan información diferente a los realizados en un corto periodo de tiempo. Esto implica que los pequeños avances obtenidos con este tipo de análisis pueden ser de ayuda para perfeccionar los algoritmos desarrollados en los estudios a corto plazo.

2.7.2. Supuestos del modelo

Se supone que varios resultados Y_{ij} surgen durante el tiempo t para un individuo i y resultado j :

$$Y_{ij}(t) = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \epsilon_{ij}(t) \quad (2.2)$$

Donde:

- $g_j(t + \gamma_i)$: función continuamente diferenciable y monótona.
- γ_i : desfase temporal.
- α_{0ij} : efectos aleatorios no dependientes del tiempo
- α_{1ij} : efectos aleatorios dependientes del tiempo
- $\epsilon_{ij}(t)$: error residual.

A partir de una estimación de los parámetros conocidos, con un algoritmo de suma de cuadrados residual (RSS) se estiman los parámetros que faltan (g_j , γ_i y α). Finalmente para la obtención de los residuos parciales se utiliza el vector: $(R_{ij}^g(t), R_{ij}^\gamma(t), R_{ij}^\alpha(t))$

2.7.3. Resultados

Debido a las diferentes escalas de los grupos diagnósticos a analizar, primero se requiere una conversión a una escala percentil ponderada de todos ellos para homogeneizar los datos. A continuación se aplica el algoritmo diferentes grupos de ADNI

según diferentes características como la deposición de placa amiloide en el cerebro o al menos un alelo $\epsilon 4$ de apolipoproteína E (APOE).

Se encontró que los sujetos con deposiciones de amiloide, la media de los valores de $A\beta$ en el líquido cefalorraquídeo (CSF) sigue una trayectoria lineal mientras que, en los PET sigue una trayectoria sigmoide. Aun así, son formas más planas que las propuestas por Jack[24], permaneciendo entre el percentil 40 y 80.

2.8. Aportación a la investigación

2.8.1. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study

El artículo de Schrag [41] se centra en la evaluación de la capacidad de diversas variables clínicas y biomarcadores para predecir el deterioro cognitivo en pacientes con enfermedad de Parkinson recién diagnosticada. La idea principal del artículo es investigar y analizar la utilidad de diferentes factores en la predicción de la progresión del deterioro cognitivo en esta población específica.

Proporciona un análisis para identificar qué combinación de características clínicas y biomarcadores podría ser más efectiva para predecir la evolución de la función cognitiva en pacientes con la enfermedad.

2.8.2. Parkinson disease-associated cognitive impairment

Este estudio se centra en la investigación de los factores de riesgo que engloban a la PD [4]. En concreto, como factores se estudian los cambios cognitivos al inicio de la enfermedad, las alucinaciones visuales, la edad y las alteraciones en biomarcadores específicos.

Además, se exploran las fases de progresión desde la PD a estados como el deterioro cognitivo leve en la PD (PD-MCI) y la demencia en la PD (PDD). Por último, se presentan abordajes terapéuticos para el PD-MCI, tanto de naturaleza no farmacológica como farmacológica.

2.8.3. Diagnosing PD-MCI by MDS TaskForce Criteria: How Many and Which Neuropsychological Tests?

La idea principal del artículo [21] es investigar la eficacia y la cantidad óptima de pruebas neuropsicológicas necesarias para diagnosticar el deterioro cognitivo leve en la enfermedad de Parkinson (PD-MCI) utilizando los criterios establecidos por el Grupo de Trabajo de la Sociedad Internacional de Enfermedad de Parkinson y Trastornos del Movimiento (MDS).

El estudio se centra en determinar qué combinación de pruebas neuropsicológicas es más precisa para detectar y diagnosticar el PD-MCI, con el objetivo de mejorar

la precisión diagnóstica y la identificación temprana de problemas cognitivos en pacientes con PD.

2.8.4. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease

En ese estudio, se presenta un marco de investigación con el objetivo de avanzar hacia una definición biológica más precisa de la enfermedad de Alzheimer. El artículo apuesta por un enfoque basado en biomarcadores y características biológicas para diagnosticar y clasificar la enfermedad de Alzheimer, en contraste con la tradicional definición clínica basada en síntomas.

Esta nueva visión de las enfermedades neurodegenerativas es interesante, pues proporciona un nuevo enfoque metodológico y conceptual para la identificación de biomarcadores y características biológicas que puedan ser útiles en la predicción y seguimiento del deterioro cognitivo en la PD.

2.8.5. Brain Amyloid Contribution to Cognitive Dysfunction in Early-Stage Parkinson's Disease: The PPMI Dataset

Fiorenzato [17] establece que depósito de amiloide cerebral está relacionado con la disfunción cognitiva en las etapas iniciales de la enfermedad de Parkinson. Para ello, el estudio utiliza datos del estudio PPMI para examinar cómo la presencia de amiloide en el cerebro puede contribuir al deterioro cognitivo en personas en las primeras etapas de la enfermedad de Parkinson.

Estos hallazgos son muy relevantes a la hora de utilizarlos para calibrar y desarrollar modelos predictivos de la progresión del deterioro cognitivo en pacientes con PD. Los resultados y las conclusiones del estudio pueden ser incorporados en los modelos para mejorar su precisión y capacidad de predecir la evolución de la enfermedad en términos cognitivos.

2.8.6. Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease

En el artículo Platero desarrolla un análisis de supervivencia longitudinal y comparación de dos grupos para predecir la progresión del deterioro cognitivo leve en pacientes con AD. El estudio busca identificar patrones de cambio a lo largo del tiempo y factores predictivos que puedan ayudar en la identificación temprana de aquellos pacientes con mayor probabilidad de desarrollar Alzheimer en comparación con otros con el mismo diagnóstico inicial.

Las técnicas y enfoques estadísticos utilizados en el artículo podrían inspirar el desarrollo de modelos predictivos de progresión del deterioro cognitivo específicamente para pacientes con Parkinson. Estos modelos podrían aprovechar la información longitudinal sobre las características clínicas, biomarcadores y otros factores relevantes, con el objetivo de identificar patrones de progresión y factores de riesgo específicos que influyan en el deterioro cognitivo en esta población.

2.8.7. Estimating long-term multivariate progression from short-term data

Los autores de este artículo [13] proporcionan una guía detallada para la creación de modelos de progresión de la PD utilizando el paquete GRACE desarrollado por el autor a lo largo del tiempo con trayectorias a largo plazo. Esta metodología permite visualizar diversas trayectorias correspondientes a cada biomarcador del estudio, lo que facilita la comprensión de cómo se modela la progresión de la enfermedad. Mediante este estudio, se logra una mejor comprensión de la interpretación de los resultados obtenidos a partir del script implementado en R.

Capítulo 3

Datos empleados, herramientas y materiales

3.1. Introducción

El objetivo de este capítulo es mostrar los datos, herramientas y materiales con los que se ha realizado el estudio. En primer lugar se detalla información sobre los datos escogidos para este, además de la base de datos de donde provienen. A continuación, se listan las herramientas utilizadas para poder llevar a cabo el estudio con una breve explicación.

3.2. Datos empleados

A lo largo de este estudio, se han empleado datos obtenidos a partir de la Iniciativa de Marcadores de Progresión del Parkinson (PPMI). Este es un proyecto de investigación colaborativo que tiene como objetivo identificar y validar biomarcadores de la PD.



Figura 3.1: Logo PPMI.

El PPMI involucra a una red internacional de investigadores y pacientes que colaboran en la recolección de datos clínicos, imágenes cerebrales y muestras biológicas de pacientes con Parkinson y controles sanos [3]. Los datos se utilizan para estudiar la progresión de la enfermedad y para identificar biomarcadores que puedan ayudar en la detección temprana y el seguimiento de la enfermedad.

Para la extracción de datos se utilizan diversos métodos, entre los que se encuentran:

- Imágenes cerebrales: entre ellas se incluyen PET, MRI y fMRI. Estas imágenes pueden estar ponderadas en T1 o T2.
- **Evaluaciones clínicas:** estas miden la gravedad de los síntomas de la enfermedad y el estado cognitivo de los pacientes. Entre ellas se encuentran escalas como UPDRS, la escala de Hoen y Yahr o la evaluación neuropsicológica.
- **Análisis genéticos:** estos sirven para detectar factores de riesgo de la PD y evaluar el impacto de las variaciones genéticas durante la enfermedad.
- **Análisis bioquímicos:** a partir de muestras biológicas se consigue detectar biomarcadores determinantes en la enfermedad. Algunos ejemplos son proteínas como la α -sinucleína o CSF (entre otros).

Los datos del PPMI se obtienen de varios centros de investigación, clínicas y hospitales que participan en el estudio en todo el mundo. Los participantes en el estudio son evaluados a lo largo del tiempo para el seguimiento de la progresión de la PD.

Estos datos incluyen información clínica, neurológica, neuropsicológica, genética y de imagenología, además de muestras biológicas de sangre, CSF y tejido cerebral. La información se recopila y almacena de forma anónima y confidencial, y se utilizan para estudios de investigación destinados a mejorar la comprensión de la PD y desarrollar nuevas terapias y tratamientos.

En este estudio, se han utilizado los datos de 423 pacientes PD:

PD-CU	323
PD-MCI	83
PDD	17
TOTAL	423

Tabla 3.1: Pacientes PD del experimento.

Como se puede observar en la tabla 3.1, de los 423 pacientes de PD 323 no presentan deterioro cognitivo (PD-CU). Además, se encuentran 83 pacientes con deterioro cognitivo leve (PD-MCI) y 17 con demencia (PDD). Adicionalmente, en el análisis se cuenta con 196 sujetos control (HC).

3.2.1. Test neuro-psicológicos utilizados en PPMI

El estudio PPMI incluye la administración de la Evaluación Cognitiva de Montreal (MoCA) y pruebas neuropsicológicas detalladas anualmente a todos los participantes. Las pruebas neuropsicológicas en PPMI evalúan diversas habilidades cognitivas, incluyendo el aprendizaje y la memoria verbal, la fluidez verbal, la velocidad de procesamiento, la memoria de trabajo y la habilidad visoespacial. Se utilizan tres métodos normativos en el estudio: normas de edad publicadas, normas internas de promedio de grupo general y normas basadas en la edad del grupo control saludable.

Para clasificar a los participantes con deterioro cognitivo leve relacionado con la enfermedad de Parkinson (PD-MCI), se siguieron las pautas del MDS Task Force

Nivel I. Se consideró que los participantes tenían MCI si obtenían una puntuación de 1.5 desviaciones estándar o menor que el promedio normativo en al menos 2 de las pruebas neuropsicológicas. La prueba MoCA no se utilizó como criterio para el MCI en este estudio con el fin de comparar las tasas de deterioro basadas en diferentes muestras normativas [45].

- **MoCA¹ (Evaluación Cognitiva de Montreal)**

La evaluación Cognitiva de Montreal (MoCA) [12] es una herramienta de evaluación cognitiva que se utiliza para detectar la presencia de deterioro cognitivo o demencia en los pacientes. Actualmente es una de las pruebas más utilizadas para evaluar la función cognitiva de los pacientes con enfermedades neurodegenerativas.

La MoCA es un examen compuesto por una variedad de tareas y preguntas diseñadas para evaluar aspectos cruciales de la función cognitiva, como la memoria, la atención, la orientación, la habilidad para resolver problemas, el lenguaje, la percepción espacial y la capacidad de abstracción. En general, la prueba MoCA dura cerca de 15 minutos

En cuanto a la puntuación se refiere, se puede obtener un máximo de 30 puntos. Una calificación por debajo de los 26 puntos se considera deterioro cognitivo y, por debajo de 21 demencia. No obstante, esta prueba no es la única que se realiza a los pacientes para determinar un diagnóstico definitivo, si no que se utiliza en conjunto con otros exámenes y pruebas clínicas para conseguir tener la evaluación del sujeto más completa posible.

Actualmente esta prueba es una de las más utilizadas debido a que es breve y no invasiva, además de que no requiere un personal muy especializado para realizarla. Sin embargo, hay que tener en cuenta que los resultados obtenidos a partir de este test pueden estar influenciados por factores no cognitivos como la depresión o el estrés (entre otros).

Es por todo esto, que la prueba MoCA actualmente es un recurso valioso para evaluar la cognición en adultos y puede ayudar a identificar signos tempranos de deterioro cognitivo. No obstante, es esencial combinarla con otros exámenes clínicos e imágenes para obtener una evaluación completa y precisa del estado cognitivo del paciente.

- **HVLT-R² (Test de Aprendizaje Verbal de Hopkins)**

La prueba de Aprendizaje Verbal de Hopkins en su versión revisada es una de las pruebas neuropsicológicas más utilizadas en la actualidad para evaluar los procesos de aprendizaje y memoria verbal en personas adultas [40]. El HVLT-R ha sido utilizado con frecuencia como estándar de oro en estudios que evalúan el funcionamiento cognitivo. La prueba consta de tres ensayos consecutivos en los

¹Montreal Cognitive Assessment

²Hopkins Verbal Learning Test-Revised

cuales la persona debe recordar, en cada ensayo, el mayor número de palabras de una lista de 12 posibles agrupadas en tres categorías semánticas. Posteriormente, se lleva a cabo un ensayo de recuerdo diferido y otro de reconocimiento.

■ **Nombramiento de animales**³

La fluidez animal es una tarea ampliamente utilizada para evaluar a las personas con enfermedad de Alzheimer (AD) y otros trastornos neurológicos. Se argumenta que los mecanismos que impulsan el rendimiento en esta tarea se basan en el lenguaje y las funciones ejecutivas.

La prueba de fluidez animal generalmente la administra un médico psiquiatra o un trabajador social e implica que el paciente nombre tantos animales como sea posible en un período de tiempo de 60 segundos. El clínico registra el número de animales nombrados. Si el paciente nombró 15 o menos animales dentro del marco de tiempo de 60 segundos, esto puede indicar etapas tempranas de demencia o el desarrollo de un deterioro cognitivo

Aunque la prueba del estado mental MMSE⁴ es la más utilizada para evaluar la demencia, a menudo es señalada por la falta de precisión a la hora de detectar un deterioro leve. Por lo contrario, la prueba de fluidez animal puede distinguir de forma precisa las personas que padecen demencia con el mismo nivel de sensibilidad que el MMSE.

La puntuación de esta prueba será el número de animales sin repetir que el paciente sea capaz de nombrar en el tiempo establecido.

■ **SDMT**⁵ (Test de Símbolos y Números)

La SDMT es una prueba de uso común para evaluar la velocidad psicomotora, que mide la velocidad de procesamiento, además de la velocidad motora [35]. En esta prueba se requiere que el paciente sustituya los dígitos por símbolos abstractos utilizando una clave de referencia . El rendimiento también se ve afectado por la atención, el escaneo y el seguimiento visual y la memoria de trabajo. El SDMT tiene excelentes propiedades psicométricas, con alta fiabilidad y validez.

Se ha demostrado que el rendimiento de la prueba es muy sensible a una serie de trastornos neuropsiquiátricos y afecciones neurológicas. El rendimiento en el SDMT es un predictor significativo de la conversión de deterioro cognitivo normal a un deterioro cognitivo leve (MCI). La brevedad y la facilidad de administración del SDMT, junto con su puntuación irrefutable, hacen del SDMT una prueba de elección en muchas baterías neuropsicológicas y estos factores han contribuido a su uso generalizado.

■ **LNS**⁶ (Secuenciación de Letras y Números)

³Animal Naming Test

⁴Mini-Mental State Examination

⁵Symbol Digit Modalities Test

⁶Letter Number Sequencing

La LNS es una prueba que mide las habilidades de memoria a corto plazo de un individuo para poder procesar y volver a secuenciar la información [43]. Esta prueba hace cierto hincapié en el dominio del lenguaje receptivo por parte del sujeto debido a la necesidad de entender la información que se está transmitiendo.

■ **JLO⁷ (Juicio de Orientación Lineal)**

Esta es una prueba estandarizada de habilidades visuoespaciales comúnmente asociadas con el funcionamiento del lóbulo parietal en el hemisferio derecho [7]. La prueba mide la capacidad de una persona para hacer coincidir el ángulo y la orientación de las líneas en el espacio. Se pide a los sujetos que combinen dos líneas en ángulo con un conjunto de 11 líneas que están dispuestas en un semicírculo y separadas 18 grados entre sí.

La prueba consta de cinco ensayos de práctica seguidos de 30 elementos de prueba. Es adecuado para poblaciones adultas y pediátricas. La prueba tiene dos formas, H y J, que presentan las mismas 30 pruebas pero en orden diferente. Las respuestas a los mensajes se pueden señalar o hablar. Una puntuación de 17 o menos se considera un signo de déficit grave.

3.3. Herramientas

3.3.1. MATLAB

MATLAB⁸ es un software de programación y entorno de desarrollo para cálculo numérico y análisis de datos [29]. Esta herramienta permite realizar cálculos numéricos avanzados, gráficos y visualizaciones, así como la creación de interfaces de usuario para aplicaciones interactivas. Es ampliamente utilizado en áreas como la ingeniería, física, matemáticas, finanzas y ciencias de la computación.



Figura 3.2: Logo MATLAB.

Para este estudio, las funciones más utilizadas del software han sido las siguientes:

- Organización de datos en tablas.
- Desarrollo de algoritmos.
- Representación de datos y algoritmos de forma gráfica.
- Cálculo numérico y análisis de datos.

⁷Judgment of Line Orientation

⁸MATrix LABoratory

3.3.2. R

R es un lenguaje de programación y un entorno de desarrollo estadístico utilizado para el análisis de datos, la visualización de datos, la creación de modelos estadísticos y el desarrollo de software estadístico [37]. Este lenguaje se puede utilizar en varios



Figura 3.3: Logo lenguaje R.

entornos de programación. En este estudio se ha hecho uso de la aplicación RStudio para la ejecución de la librería GRACE.

3.3.3. FileZilla

FileZilla es un software libre de transferencia de archivos que permite a los usuarios descargar y subir archivos como documentos, imágenes o scripts desde un servidor a través de protocolos como FTP⁹(Protocolo de Transferencia de Archivos) o SFTP¹⁰ (FTP seguro) [16].



Figura 3.4: Logo FileZilla.

3.3.4. PuTTY

PuTTY¹¹ es un programa que permite a los usuarios acceder a servidores remotos a partir de SSH¹², Tenet y otros protocolos [36]. En concreto, en este experimento se utiliza la conexión remota con una máquina de sistema operativo Linux y su software.



Figura 3.5: Logo PuTTY.

⁹File Transfer Protocol

¹⁰Safe File Transfer Protocol

¹¹Port Unique TeleTYpe

¹²Secure Shell

3.3.5. Xming

Xming es un servidor X Window System para Microsoft Windows que permite ejecutar aplicaciones gráficas de Unix en un sistema Windows [46]. El servidor X permite a los usuarios ejecutar aplicaciones gráficas remotas en un sistema local. Se utiliza junto con PuTTY en la conexión con la máquina para poder visualizar programas que puede ejecutar.



Figura 3.6: Logo Xming.

3.3.6. Tmux

Tmux es una herramienta de línea de comandos que permite a los usuarios de Unix y Linux crear y administrar sesiones de terminal múltiples en una sola ventana. Tmux también permite dividir la ventana de la terminal en paneles para ejecutar varias tareas al mismo tiempo en diferentes paneles. Además, esta herramienta permite la posibilidad de conectarse y desconectarse de sesiones de terminal sin perder el trabajo que se está realizando [23].



Figura 3.7: Logo Tmux.

3.3.7. Magerit-3

Magerit-3 es un supercomputador producto de CesViMa¹³ [2]. Se trata de una infraestructura de computación diseñada para manejar cargas de trabajo científicas que demandan un alto nivel de procesamiento computacional (HPC¹⁴). Este cluster cuenta con 68 nodos, cada uno con las siguientes características:

- Procesador Intel Xeon Gold 6230 (20 cores @ 2.1 GHz y 40 hilos)
- 192 GB de memoria RAM
- Disco de memoria SSD de 480 GB

3.3.8. LaTeX

LaTeX es un sistema de composición de documentos que permite la creación de documentos con alta calidad tipográfica, especialmente enfocado en la escritura de

¹³Centro de Supercomputación y Visualización de Madrid

¹⁴High Performance Computing



Figura 3.8: Logo CesVisMa.

documentos científicos y técnicos que contengan ecuaciones, gráficos, tablas, bibliografías y otros elementos similares [1]. Este sistema está compuesto por multitud de macros de comandos TeX que permiten la correcta ubicación y composición del texto.



Figura 3.9: Logo LaTeX.

Capítulo 4

Métodos

4.1. Introducción

El objetivo de este capítulo es recopilar y documentar todos los métodos utilizados durante el estudio para conseguir procesar los datos de los diferentes sujetos.

Primero, se ha hecho un análisis y clasificación de los sujetos según su capacidad cognitiva y puntuaciones en los diferentes test neuropsicológicos y valores de evaluaciones clínicas y bioquímicas. A continuación se ha procedido a la realización de tablas de contingencia de los mismos datos para obtener los resultados necesarios para las variables categóricas.

Una vez se tienen los resultados de las características de los sujetos, se definen los tiempos de conversión y censura utilizados en el análisis de supervivencia.

También se proponen diferentes modelos predictivos para la progresión de los sujetos y se analizan para intentar obtener el que de mejores prestaciones.

Todos los algoritmos desarrollados se encuentran en el usuario *egomez* de la dirección IP: **138.100.100.132**

4.2. Análisis de los sujetos al inicio del seguimiento

Inicialmente, se procede al análisis de los pacientes sometidos a estudio para poder tener una visión global del estado cognitivo de estos y los resultados en diferentes tests.

El script utilizado para el análisis de los sujetos al inicio se encuentra en la dirección: *egomez/matlab/summaryPPMI_PD_features/summaryPPMI_PD_features.m*

Este algoritmo permite obtener de una forma visual los valores de las variables clínicas y demográficas de las variables no categóricas de los pacientes sometidos a estudio.

Algoritmo 1 Características clínicas y demográficas iniciales de los pacientes sometidos al estudio

Entrada: features, grupos del estudio, tipo de características y datos de los pacientes

Salida: Medidas de los grupos PD-CU, PD-MCI y PDD.

- 1: Configuración de los tipos de grupos que hay.
 - 2: **para** para $i=1$:número de grupos **hacer**
 - 3: Máscara para usar solo los datos del grupo i
 - 4: Número de pacientes de la máscara con datos al principio del estudio
 - 5: Cálculo número de mujeres en el grupo i
 - 6: Cálculo de pacientes con valor de $A\beta$ distinto de cero en el grupo i
 - 7: Mostrar número de mujeres, media y valores máximos y mínimos del grupo i
 - 8: Mostrar media de años de educación y desviación típica del grupo i
 - 9: Mostrar pacientes con datos de $A\beta$ en el grupo i
 - 10: **fin para**
 - 11: **para** $i=1$:número de características a estudiar **hacer**
 - 12: Mostrar la característica i
 - 13: **para** $j=1$:número de grupos **hacer**
 - 14: Mostrar número de pacientes, valor medio, desviación típica y valores máximos y mínimos de la variable i del grupo j
 - 15: **fin para**
 - 16: Cálculo del pvalor de la característica i
 - 17: Multicomparación entre los grupos de la característica i
 - 18: **fin para**
-

A partir de este algoritmo, se comparan los resultados de los grupos de sujetos control y pacientes PD y seguidamente los correspondientes a los sujetos PD-CU, PD-MCI y PDD.

Para hacer el análisis de las variables categóricas y realizar las tablas de contingencia necesarias se utilizó el código ubicado en:

egomez/matlab/crossTab_PPMI/crossTab_PPMI_PD.m

Algoritmo 2 Análisis de las variables categóricas del estudio

Entrada: features, grupos del estudio, tipo de características y datos de los pacientes (variables categóricas)

Salida: Distribución de la frecuencia de las variables categóricas de los grupos PD-CU, PD-MCI y PDD.

- 1: Configuración de los tipos de grupos que hay.
 - 2: Configuración criterios cognitivos
 - 3: **si** Tipo de datos es correspondiente a sujetos PD de PPMISep20 **entonces**
 - 4: Establece características a utilizar
 - 5: **fin si**
 - 6: Número de cohorte
 - 7: Obtención de datos a partir de la función *getDataPPMI-MCI.v1* introduciendo las características y cohorte
 - 8: Obtención de variables PPMI feature_names
 - 9: **si** varCogCriteria= SITE **entonces**
 - 10: Asigna a la variable PPMI.MCI el valor de PPMI.cogstate_end
 - 11: **fin si**
 - 12: Categoriza a los pacientes en tres grupos en función de sus valores PPMI.MCI
 - 13: Máscara para utilizar los datos del año 0.
 - 14: Variable categórica de género (0 hombre, 1 mujer)
 - 15: Análisis tabla de contingencia datos género de los pacientes
 - 16: Máscara para excluir pacientes sin valores en APOE4
 - 17: Variable categórica APOE4
 - 18: Análisis tabla de contingencia datos APOE4 de los pacientes
-

4.3. Perfiles AT en sujetos PD y tiempos de conversión y censura

4.3.1. Perfiles AT en sujetos PD

Una vez se han analizado diferentes características que podrían actuar como factores de riesgo en la PD y, por consiguiente, ser predictores de la enfermedad, se propone establecer una relación entre el deterioro cognitivo que estos pacientes poseen y los constructos de la AD tal como se muestra en el artículo de Jack et al. [25].

Para ello, se establecen una serie de pautas: un sujeto es A+ cuando sus valores de $A\beta < 683,45$ pg/ml y T+ cuando la pTau > 23 pg/ml.

A partir del script ubicado en:

egomez/matlab/demo_histConvertTime/egomez/matlab/demo_histConvertTime.m

se pueden clasificar a los sujetos del estudio según los diferentes perfiles AT.

Algoritmo 3 Perfiles AT de los pacientes PD-novo**Entrada:****Salida:** Perfiles AT de los diferentes grupos cognitivos de los sujetos PD-novo

- 1: Configuración de los tipos de características, población y criterio MCI.
- 2: Datos de población PD y características seleccionadas
- 3: Mascara para datos del baseline y con valores de abeta y ptau no nulos
- 4: Aplicar máscara a PPMI
- 5: Máscara para MCI=1
- 6: Máscara para MCI=2
- 7: Máscara para MCI=3
- 8: Aplicar máscaras a máscara del baseline PPMI
- 9: **para** cada máscara **hacer**
- 10: **para** cada perfil AT **hacer**
- 11: Filtrar datos con esas condiciones
- 12: Calcular porcentaje del grupo respecto el total
- 13: **fin para**
- 14: Mostrar resultados del grupo en el baseline
- 15: **fin para**
- 16: Representación valores $A\beta$ y pTau en baseline
- 17: Diagrama de barras según estado cognitivo y perfil AT
- 18: **para** cada máscara **hacer**
- 19: **para** cada perfil AT **hacer**
- 20: Guardar mínimo valor de $A\beta$
- 21: Guardar máximo valor de pTau
- 22: Filtrar datos con esas condiciones
- 23: Calcular porcentaje del grupo respecto el total
- 24: **fin para**
- 25: Mostrar resultados de los perfiles del grupo
- 26: **fin para**
- 27: Representación valores $A\beta$ y pTau en el último año de estudio
- 28: Diagrama de barras según estado cognitivo y perfil AT en el último año de estudio

4.3.2. Tiempos de conversión y censura

Se realiza un análisis de supervivencia entre los sujetos del estudio con el fin de valorar si los pacientes sufren un deterioro cognitivo a lo largo de los años del experimento. Se diferencian tres grupos:

- PD-CU: sujetos no convierten a MCI a lo largo del estudio.
- PD-MCI: la última evaluación cognitiva del sujeto es MCI.
- PDD: la última evaluación cognitiva del sujeto es demencia.

Además, cabe destacar que en los tiempos de conversión, se establece como primera visita cuando el paciente tiene una evaluación cognitiva de MCI o demencia. Sin embargo, en cuanto a los tiempos de censura se refiere, se establece como la última visita de los pacientes. El script correspondiente de los tiempos de conversión y censura se puede encontrar en:

egomez/matlab/demo_histConvertTime/hist_PD-CU-MCI-dementia.m

Algoritmo 4 Tiempos de conversión y censura

Entrada:**Salida:** Histogramas de los tiempos de conversión y censura

- 1: Configuración de los tipos de características, población y criterio MCI.
 - 2: Datos de población PD y características seleccionadas
 - 3: Calculo tiempo conversión
 - 4: Mascara para datos del baseline
 - 5: Aplicar máscara a datos PPMI
 - 6: Máscara para MCI=1
 - 7: Máscara para MCI=0
 - 8: Organizar converTime en función del año de conversión
 - 9: Contar sujetos en cada grupo
 - 10: **para** i=1:número de grupos **hacer**
 - 11: Máscara grupo i
 - 12: Contar sujetos con esa máscara
 - 13: **fin para**
 - 14: Histograma con MCI=1 según año de conversión
 - 15: Histograma con MCI=1 según año de censura
 - 16: Datos PPMI con MCI=1
 - 17: Organizar cogstate_end en función de demencia o MCI
 - 18: Máscara para grupo cogstate_end MCI
 - 19: Aplicar mascara MCI=1 al grupo cogstate_end MCI
 - 20: Máscara para grupo cogstate_end demencia
 - 21: Aplicar mascara MCI=1 al grupo cogstate_end demencia
 - 22: Histograma de la población según el año de conversión
-

4.4. Análisis estadístico

4.4.1. Modelos predictivos

Para poder definir un vector de características con el cual se pueda determinar la conversión de los pacientes PD a PD-MCI o PDD es necesario desarrollar modelos predictivos.

Para establecer la trayectoria de la enfermedad se realiza un modelo de progresión a partir de los métodos mencionados en Platero [34].

Procedimiento

Dado que se quiere realizar un seguimiento de una muestra de varios individuos en un periodo de tiempo y recoger datos de múltiples variables de estos, se emplea un estudio longitudinal multivariante. Los resultados permiten analizar cómo cambian las variables estudiadas en relación con el tiempo y la relación que tienen entre sí.

Se analizan datos al inicio del estudio (bl), a los 12 meses ($m12$) y a los 24 meses ($m24$). Las características a analizar incluyen resultados de diferentes test neuropsicológicos como: MCI_{test} , $UPDRS1$, $SITE$, marcadores biológicos determinantes en la PD (DAT, NM, CSF) o habilidades motoras. Se utilizan los métodos de Cox y Cox extendido para analizar el tiempo que tarda en ocurrir la conversión en los pacientes.

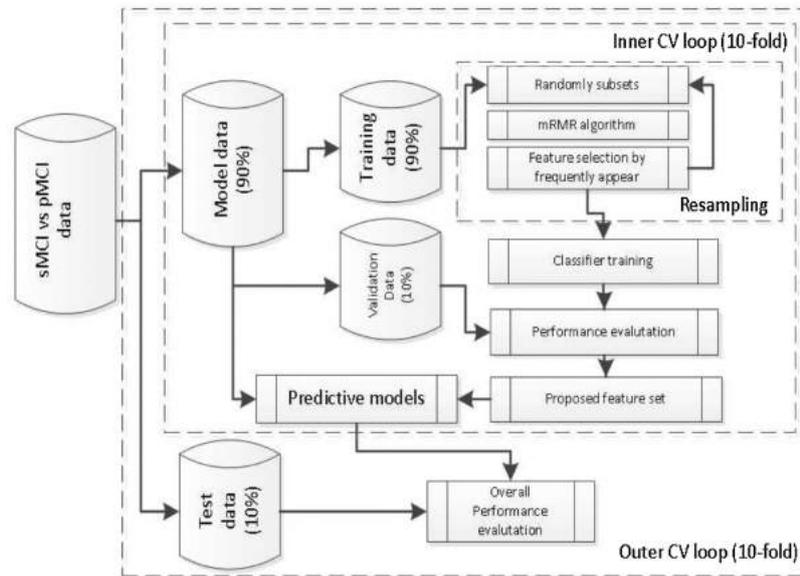


Figura 4.1: Método de validación cruzada para el desarrollo de posibles modelos predictivos.

Como se muestra en la figura, para obtener los modelos predictivos se realizan dos bucles anidados obteniendo la validación cruzada k-fold. Esta técnica permite evaluar el rendimiento del modelo dividiendo los datos en entrenamiento (*data training*) y prueba (*test set*). Para ello los datos se dividen en k subconjuntos y el modelo se entrena k veces, utilizando cada vez un subconjunto k como test set y el resto como data training. El rendimiento del modelo se obtiene a partir de la media de los k resultados.

En el bucle interno se dividen los datos en dos grupos: data training (90%) y test set (10%). Con los datos de entrenamiento se aplica el criterio de mínima-redundancia-máxima-relevancia (mRMR) y se forman los modelos con las características obtenidas con mayor frecuencia.

En el bucle externo se realiza de nuevo la división de los datos para evaluar los resultados obtenidos en el bucle interno.

Este procedimiento se realiza ejecutando el script ubicado en:

`egomez/matlab/demoPredictiveModels/PPMI_SA_230324/builtPredictiveModels_PDtoPDMCI_v1.m`

Una vez se han creado los diferentes modelos, se obtienen archivos .mat que recogen los resultados de entrenarlos y evaluarlos.

Algoritmo 5 builtPredictiveModels_PDtoPDMCI_v1.m**Entrada:****Salida:** Vectores de características con mejores prestaciones

- 1: Configuración características para modelo
- 2: Cargar matriz con datos seleccionados
- 3: **si** extCox_LME o LME **entonces**
- 4: Obtener matriz con características
- 5: **si no, si** LME **entonces**
- 6: Mascara año de visita 0
- 7: Obtener matriz con características
- 8: **fin si**
- 9: **para** i=1:totales bucle externo **hacer**
- 10: **para** i=1:totales bucle interno **hacer**
- 11: Dividir población en dos lotes para obtener el mejor conjunto
- 12: Entrenar modelo
- 13: **fin para**
- 14: Calcular rendimiento del modelo entrenado con el lote de entrenamiento
- 15: **fin para**
- 16: Guardar vectores con mejor rendimiento

Para seleccionar el mejor resultado, se utiliza el script que se encuentra en:

egomez/matlab/demoPredictiveModels/PPMI_SA_230324/analysis_withSavedData_Mux_PPMI.m

Es necesario introducir los siguientes parámetros de entrada:

- **process_data**: valor booleano que determina si se tienen que procesar los datos (1) o se van a cargar datos ya obtenidos anteriormente (0).
- **n_Approach**: indica el modelo que se va a utilizar.
 - Regresión de Cox.
 - Modelo exclusivo LME.
 - Regresión de Cox extendido.
- **n_Cov**: indica la covariable que se va a utilizar en el modelo.
 - Edad.
 - Edad, sexo y años de educación.
 - Correlación entre covariables.
 - Edad, UPSIT, REM y GDS.
 - Sin la edad.
- **str_visit**: especifica la fecha de los datos con los que se van a obtener los modelos.
 - Al inicio del estudio: *'bl'*
 - A los 12 meses: *'m12'*
 - A los 24 meses: *'m24'*
- **_cohort**: especifica la cohorte utilizada en el estudio. En este análisis se utilizarán todos los sujetos.

- **iter**: número de iteraciones utilizadas para los modelos.
- **n_Population**: población a analizar (*'PPMISep20'*)
- **n_cogCriteria**: establece el criterio cognitivo del análisis.
 - MCI test
 - SITE

A partir de este script se obtienen los mejores modelos, mostrando los valores de sensibilidad, especificidad, exactitud, característica operativa del receptor y frecuencia.

Algoritmo 6 `analysis_withSavedData_Mux_PPMI.m`

Entrada: Modelos predictivos del algoritmo 5

Salida: Mejores vectores de características

- 1: Configuración características para modelo
 - 2: **si** no existe documento con mejores vectores **entonces**
 - 3: Obtener rendimientos de los vectores
 - 4: Guardar resultados en archivo `.mat`
 - 5: **si no**
 - 6: Cargar fichero
 - 7: **fin si**
 - 8: Imprimir mejores vectores y sus características.
-

Con el código de *`analysis_predictiveModels_table_PDtoMCI.m`* se obtiene una tabla con todos los modelos predictivos obtenidos y sus valores.

4.4.2. Modelos de progresión con GRACE

Una vez se han obtenido los vectores de características con mejores prestaciones, se procede a estudiar la trayectoria temporal de la progresión de la enfermedad. Esto supone realizar un análisis que permita mostrar como avanza y se desarrolla la enfermedad a lo largo del tiempo. En este proceso se valora la evolución de los síntomas, la velocidad a la que empeoran y como afecta la PD al cuerpo.

Estos modelos son importantes ya que no solo permiten comprender de una manera más efectiva como se desarrolla la enfermedad, si no que también permiten diseñar tratamientos más personalizados según el paciente.

En este estudio, se ha utilizado el método GRACE que está formado por diferentes funciones en R que posibilitan la construcción de modelos predictivos a partir del análisis de los datos longitudinales de los pacientes PD. Como resultado, se obtienen las curvas de crecimiento multivariantes en largos periodos de tiempo. Los scripts se encuentran disponibles en:

<https://bitbucket.org/mdonohue/grace/src/master/>

Para el análisis de las trayectorias a corto plazo se emplea un tiempo relativo γ a la evolución de la enfermedad. Se supone que varios resultados Y_{ij} surgen durante

el tiempo t para un individuo i y resultado j :

$$Y_{ij}(t) = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \epsilon_{ij}(t) \quad (4.1)$$

Donde:

- $g_j(t + \gamma_i)$: función continuamente diferenciable y monótona.
- γ_i : desfase temporal.
- α_{0ij} : efectos aleatorios no dependientes del tiempo
- α_{1ij} : efectos aleatorios dependientes del tiempo
- $\epsilon_{ij}(t)$: error residual.

A partir de una estimación de los parámetros conocidos, con un algoritmo de suma de cuadrados residual (RSS) se estiman los parámetros que faltan (g_j , γ_i y α). Finalmente para la obtención de los residuos parciales se utiliza el vector: $(R_{ij}^g(t), R_{ij}^\gamma(t), R_{ij}^\alpha(t))$

Algoritmo 7 GRACE

Entrada: Residuos parciales

Salida: Estimación de los parámetros g_j , γ_i y α .

- 1: Inicialización de $\gamma_i = 0$
 - 2: Establecer $\alpha_{0ij} = 0$ y $\alpha_{1ij} = 0$
 - 3: **mientras** $R_{ij}^g(t)$ y $R_{ij}^\alpha(t)$ no converjan **hacer**
 - 4: Estimar g_j con el suavizado del residuo $R_{ij}^g(t)$
 - 5: Estimar α_{0ij} , α_{1ij} con el modelo linear mixto del residuo $R_{ij}^\alpha(t)$
 - 6: Comprobación de la convergencia de $R_{ij}^g(t)$ y $R_{ij}^\alpha(t)$
 - 7: **fin mientras**
 - 8: **mientras** $R_{ij}^g(t)$, $R_{ij}^\gamma(t)$, $R_{ij}^\alpha(t)$ no converjan **hacer**
 - 9: **mientras** $(R_{ij}^g(t), R_{ij}^\gamma(t))$ no converjan **hacer**
 - 10: Estimar g_j con el suavizado del residuo $R_{ij}^g(t)$
 - 11: Estimar α_{0ij} , α_{1ij} con el modelo linear mixto del residuo $R_{ij}^\alpha(t)$
 - 12: Comprobar la convergencia de $R_{ij}^g(t)$ y $R_{ij}^\alpha(t)$
 - 13: **fin mientras**
 - 14: Estimación de g_j
 - 15: Establecer $\epsilon_{ij}(t) = 0$ y $\alpha = 0$
 - 16: Estimación de $R_{ij}^\gamma(t)$ para todos los instantes
 - 17: Comprobación de la convergencia de $R_{ij}^g(t)$, $R_{ij}^\gamma(t)$, $R_{ij}^\alpha(t)$
 - 18: **fin mientras**
-

Para obtener los modelos de progresión con GRACE, es necesario trabajar con MATLAB y R. Es por ello por lo que se siguen las siguientes pautas:

1. Generación de tabla EXCEL con los datos necesarios para GRACE

A partir del código ubicado en:

egomez/matlab/dem_Grace_PPMI/tableGracePPMI_2.m

se genera y guarda la tabla correspondiente con los siguientes datos:

- **t**: tiempo en el que se ha medido el marcador a analizar.
- **Y**: vector de marcadores expresados en percentiles.

- **raw**: marcadores sin cambios.
- **Outcome**: número de identificador del marcador.
- **id**: número de identificador del sujeto.
- **Convert**: indica la conversión o no a MCI del sujeto.
- **RID**: identificador del paciente.
- **ConverTime**: tiempos de conversión o censura.

2. Se emplea el algoritmo GRACE con RStudio

A partir de la tabla anterior, se ejecuta el código:

```
egomez/matlab/dem_Grace_PPMI/grace_PPMI.R
```

con RStudio tal y como se detalla en el artículo de Donohue et al. [13].

Como resultado de ejecutar este código se obtiene una tabla EXCEL con los parámetros de la ecuación 4.1 ya calculados.

Algoritmo 8 grace_PPMI.R

Entrada: tabla de datos *dd* generados con *tableGracePPMI_2.m*

Salida:

- 1: Instalar librerías necesarias
 - 2: Cargar librerías
 - 3: Establecer directorio de trabajo donde se encuentran los datos
 - 4: Imprimir el directorio de trabajo actual
 - 5: Lectura de archivo Excel "*dd.xlsx*" ubicado en el directorio y se guarda en el objeto "*dd*"
 - 6: Análisis "*grace.simulation.fits*" utilizando la función "*grace*"
 - 7: Se crea un nuevo objeto llamado "*dd1*" mediante la combinación de los resultados
 - 8: Se muestra un resumen estadístico de "*dd1*" utilizando la función "*summary*"
 - 9: Se guarda el objeto "*dd1*" en un archivo Excel
-

3. Trazado de trayectorias con MATLAB

Finalmente, con el script que se encuentra en:

```
/home/egomez/matlab/demo_Grace_PPMI/plot_gracePPMI_2.m
```

se obtienen las trayectorias a corto y largo plazo de la población sometida a estudio. Además, se calculan los siguientes parámetros:

- MAE¹ (Error Medio Absoluto) de los residuos de estimación de trayectorias para cada marcador. Representa la diferencia absoluta entre los valores pronosticados y los valores observados.
- Rendimiento del modelo según el grupo cognitivo en estudio.
- Correlación entre el tiempo conocido y el calculado por el algoritmo de GRACE.

¹Mean Absolute Error

Algoritmo 9 plot_gracePPMI.2.m**Entrada:** tabla de datos *dd1* generados con *grace_PPMI.R***Salida:** Figuras con las trayectorias de los marcadores seleccionados en los modelos progresivos

```

1: Define las variables population y cognitiveCriteria en función de los argumentos de entrada.
2: Cargar la tabla grace a partir de "tableGracePPMI.2.m"
3: si dataset seleccionado es PPMISep20 entonces
4:   si el criterio seleccionado es SITE entonces
5:     si n_trial coincide entonces
6:       Carga datos de archivo .xls en variable dd1
7:       Establece lista de nombres de características en la variable "feature_names"
8:     fin si
9:   si no, si el criterio seleccionado es MCI entonces
10:    Carga datos de archivo .xls en variable dd1
11:    Establece lista de nombres de características en la variable "feature_names"
12:  fin si
13: fin si
14: Establecer a 0 el número de visitas ("num_max_visits")
15: para i=1:longitud del vector outcomes hacer
16:   Número de visitas que cumplen la condición dd1.outcome=outcomes(i)
17:   si número de visitas  $\geq$  num_max_visits entonces
18:     num_max_visits toma el valor del número de visitas del bucle (num_visits)
19:   fin si
20: fin para
21: Inicializar las variables "MAE", "t_app", "yhat_app" y "grace_t_ghat" como matrices NaN de
    dimensiones adecuadas
22: para i=1:longitud del vector outcomes hacer
23:   Máscara ("mask_outcome") para filtrar los datos relevantes en dd1 según el valor del outcome
    actual.
24:   Subconjunto dd_outcome de los datos en dd1 que cumplen la condición de la máscara.
25:   Vector "id" con los valores únicos de la columna "id" en dd_outcome
26:   Inicialización del índice idx_first en 1
27:   para i=1:longitud del vector id hacer
28:     Encontrar el índice ("index") donde los valores en la columna "id" de dd_outcome coinciden
    con el valor actual de id.
29:   fin para
30: fin para

```

4.5. Modelos de progresión de forma masiva con Magerit

Una vez se ha encontrado el mejor vector de características a partir de Grace, se procede a la combinación masiva de todas las características para conseguir así hallar el vector con mayor rendimiento de todos. Para ello, se utiliza el algoritmo desarrollado en *multi_grace_PPMI.m*, que se puede encontrar en :

/home/u895/u895715/demo_multi_grace_PPMI.230512/multi_grace_PPMI.m

Para la optimización de recursos y reducir el tiempo de ejecución, se crea un archivo .sh para lanzar el script en Magerit. Para ello es necesario la instalación de GRACE en Magerit. Esto se hace siguiendo los siguientes pasos:

1. Se inicia sesión en Magerit con el usuario correspondiente por SSH.
2. Carga de modulo de R (*module load R*) y ejecución de R.

3. Se procede a la instalación de las librerías necesarias para ejecutar posteriormente el análisis masivo. Estas son:

```
install.packages("devtools")
install.packages("mvtnorm")
install.packages(readxl")
```

Esto solo será necesario hacerlo una vez, para que nuestro módulo contenga los paquetes necesarios a utilizar. A continuación, se carga la librería "devtools", necesaria para la instalación de grace en R (*library(devtools)*)

4. Se procede a la instalación de GRACE ejecutando las siguientes líneas de código:

```
install.bitbucket("mdonohue/grace")
install.bitbucket("mdonohue/grace",lib=-/R/libs")
install.packages("writexl",lib=-/R/libs")
install.packages("fda", lib=-/R/libs")
.libPaths( c( .libPaths(), -/R/libs/ ) )
```

La instalación de las librerías solo será necesario una vez, a partir de entonces con cargarlas será suficiente.

5. Para comprobar la correcta instalación de estas, se cargan las librerías:

```
library(grace)
library(fda)
library(mvtnorm)
library(readxl)
library(writexl)
```

A partir de aquí ya se podrá lanzar el código de R correspondiente que hace uso de las librerías anteriores.

Se lanza el archivo .sh ubicado en:

```
/home/u895/u895715/demo_multi_grace_PPMI_230512/multi_Grace_PD_Magerit.sh
```

El cual cargará los módulos de MATLAB y R necesarios para ejecutar el script *multi_grace_PPMI.m*. Este código realiza un proceso de análisis de datos a partir del algoritmo GRACE para determinar el mejor con mejor rendimiento con un análisis masivo. Se divide en tres fases:

1. Creación de las tablas GRACE a partir de MATLAB: cada tabla hace referencia a diferentes subconjuntos de características. Estas contienen información sobre la relación entre las características y la variable objetivo.
2. Procesado de las tablas GRACE con R y generación de las tablas dd1: a partir del script *multi_Grace.R* se obtienen los archivos dd1 los cuales contienen información adicional sobre la relación entre las características y la variable de respuesta.

3. Resultado de las tablas dd1 con MATLAB: se procede al análisis de las tablas generadas en la segunda fase, obteniendo así los resultados de todos los vectores analizados.

Algoritmo 10 Análisis masivo a partir de multi_grace.PPMI.m

Entrada:

Salida: Vectores de características con mejor rendimiento

- 1: Configuración de la población, tipo de características y criterio
 - 2: Inclusión o no de la edad como covariable para el análisis
 - 3: Definición de la combinación de las características
 - 4: Definición de la longitud de los vectores
 - 5: Habilidad de las fases del algoritmo
 - 6: Selección población PD con criterio MCI
 - 7: Ajuste de los tiempos de pacientes para que el tiempo cero sea el punto medio de la duración del estudio.
 - 8: Incluir la edad como covariable si esta se ha seleccionado
 - 9: Combinación de vectores
 - 10: Transformación a percentiles
 - 11: Generación y almacenamiento de los vectores a analizar en *features_subset_table*
 - 12: **si** phase=1 **entonces**
 - 13: **para** 1 hasta tamaño de *features_subset_table* **hacer**
 - 14: Creación de la tabla GRACE del vector
 - 15: **fin para**
 - 16: **fin si**
 - 17: **si** phase=2 **entonces**
 - 18: **para** 1 hasta tamaño de *features_subset_table* **hacer**
 - 19: Creación de tabla dd1 del vector de características a partir del script *multi_Grace.R*
 - 20: **fin para**
 - 21: **fin si**
 - 22: **si** phase=3 **entonces**
 - 23: **para** 1 hasta tamaño de *features_subset_table* **hacer**
 - 24: **si** existe el archivo dd1 **entonces**
 - 25: Se carga la tabla grace desde grace.mat.
 - 26: Se lee la tabla dd1 desde dd1.xlsx.
 - 27: Se obtienen datos del vector actual
 - 28: Se calculan correlaciones entre los resultados de las tablas grace y dd1 para el vector.
 - 29: Los valores de correlación se almacenan en la tabla *features_subset_table*
 - 30: **fin si**
 - 31: **si** no existe el archivo dd1 **entonces**
 - 32: Mensaje de error en la consola
 - 33: **fin si**
 - 34: **fin para**
 - 35: **fin si**
-

Algoritmo 11 Creación de tablas dd1 en la fase 2 del algoritmo: *multi_Grace.R*

Entrada: Archivo dd en formato .xlsx

Salida: Tabla dd1 del vector

- 1: Carga de las librerías necesarias
 - 2: Se lee el archivo de entrada y se guarda en dd
 - 3: Se aplica el algoritmo de GRACE a dd
 - 4: Se guardan los resultados en la tabla dd1
 - 5: Se escribe la tabla dd1 en formato .xlsx
-

Una vez se ha obtenido el vector que proporciona unas mejores prestaciones, se

procede a su análisis. Al igual que con el análisis de los vectores propuestos por el análisis de Cox, se obtienen las matrices de correlación tras la ejecución de:

```
/home/egomez/matlab/demo_Grace_PPMI/tableGracePPMI_2.m
```

introduciendo el nuevo vector de características. A continuación, se muestran tanto las trayectorias a corto y largo plazo como la recta de regresión del modelo con el script ubicado en:

```
/home/egomez/matlab/demo_Grace_PPMI/plot_gracePPMI_2.m
```

4.5.1. Bootstrap

Finalmente, se procede a estudiar la influencia de la patología amiloide en el desarrollo de la enfermedad de Parkinson, debido a la existencia de resultados en otras enfermedades neurodegenerativas como el Alzheimer (AD).

Para ello se hace uso de la técnica bootstrap o remuestreo bootstrap, el cual es un método estadístico utilizado para estimar la distribución de un estadístico y proporcionar intervalos de confianza. A partir de la tabla dd del vector obtenida anteriormente, se procede al cálculo de las estimaciones de los resultados de las características del vector con el script *bootstrap_PPMI_ATN.R*, diferenciando si se tiene patología amiloide (A+) o no (A-).

Cabe destacar que es necesario ejecutar el script por separado para los perfiles A+ y A-.

Se puede encontrar el script en:

```
/home/egomez/matlab/demo_Grace_PPMI/bootstrap_PPMI_ATN.R
```

Algoritmo 12 Bootstrap con *bootstrap_PPMI_ATN.R*

Entrada:

Salida: Trayectorias perfiles A+ o A-

- 1: Carga de las librerías necesarias
 - 2: Establecer ruta donde se encuentra el archivo dd
 - 3: Se lee el archivo dd
 - 4: Dividir perfiles dd entre A+ y A-
 - 5: Seleccionar solo los perfiles A+ o A-
 - 6: **para** 1 hasta 100 **hacer**
 - 7: Ajuste de modelos grace para una muestra aleatoria
 - 8: Guardar archivo tipo .rdata
 - 9: **fin para**
 - 10: **para** 1 hasta 100 **hacer**
 - 11: Escribe los archivos .rdata en formato .xlsx
 - 12: **fin para**
 - 13: Muestra resultados
-

Una vez se ha realizado el análisis para ambos perfiles (con y sin patología amiloide), a partir del código desarrollado en el archivo *boot_grace_PPMI_Apos.m* se puede observar las estimaciones de los sujetos A+ y A- simultáneamente en las diferentes características del vector seleccionado. Este script se puede encontrar en:

```
/home/egomez/matlab/demo_Grace_PPMI/boot_grace_PPMI_Apos.m
```

Algoritmo 13 boot_grace_PPMI_Apos.m

Entrada:**Salida:** Figura con la estimación de las trayectorias de los sujetos A+ y A-

- 1: Configuración de la población, vector de características y nombre de estas
 - 2: Establece número de repeticiones del bootstrap (n_trials)
 - 3: **para** 1 hasta n_trials **hacer**
 - 4: Lee archivos dd1_Aneg y dd1_Apos y los incluye en los vectores data_long_Aneg y data_long_Apos respectivamente
 - 5: **fin para**
 - 6: Dibuja las trayectorias de los perfiles A+ y A-
-

Capítulo 5

Resultados

5.1. Introducción

El objetivo de este capítulo es mostrar los resultados obtenidos con los métodos desarrollados anteriormente.

Para ello, se comienza con el análisis de los datos recogidos de las características clínicas y demográficas de los sujetos sometidos a estudio. A continuación, se procede al estudio de los tiempos de conversión y censura de los pacientes según los diferentes estados cognitivos de la PD.

Una vez se tienen estos resultados, se presentan los modelos predictivos propuestos, con sus características correspondientes.

Por último, a través de enfoques amplios de combinación de características, se logra identificar el vector que exhibe el mejor rendimiento, lo que permite discernir modelos con mejores prestaciones.

5.2. Análisis de los sujetos al inicio del seguimiento

5.2.1. Análisis de sujetos HC y PD al inicio del seguimiento

En un primer lugar, se analizan características demográficas y clínicas además de puntuaciones de diferentes test neuropsicológicos entre sujetos control (HC) y sujetos PD pudiéndose observar diferencias significativas en la mayoría de características demográficas y clínicas estudiadas ($p_{\text{valor}} < 0,001$). Además, en cuanto a los resultados de los test cognitivos, JLO, LNS INMEDIAT o ADL son los únicos cuyos valores son estables independientemente de si el sujeto es PD o HC. Si se analizan los marcadores CSF o imágenes DAT se encuentran menos diferencias entre ambos grupos.

Características demográficas y clínicas			
Sujetos	PD (N=423)	HC (N=196)	pvalor
Mujeres	146(34.5%)	70 (35.7%)	
APOE4	101(26.3%)	47 (26.4%)	
Edad	61.67(9.71)	60.82 (11.23)	
Educación	15.56(2.97) (5.00 26.00)	16.04 (2.89) (8.00 24.00)	
GDS	2.32(2.22) (0.00 14.00)	1.29 (2.10) (0.00 15.00)	< 0,001
UPSIT	2.32(2.22) (0.00 14.00)	1.29 (2.10) (0.00 15.00)	< 0,001
REM	4.12 (2.69) (0.00 12.00)	2.83 (2.26) (0.00 11.00)	< 0,001
UPDRS3	20.89(8.85) (4.00 51.00)	1.21 (2.19) (0.00 13.00)	< 0,001
Resultados cognitivos			
MOCA	27.13(2.32) (17.00 30.00)	1.21 (28.23) (26.00 30.00)	< 0,001
LNS	10.59(2.66) (2.00 20.00)	10.87 (2.57) (2.00 20.00)	0.3381
SDMT	41.18(9.73) (7.00 82.00)	46.77 (10.53) (20.00 83.00)	< 0,001
JLO	12.77(2.13) (5.00 15.00)	13.12 (1.98) (4.00 15.00)	0.0495
INMEDIAT	24.44(4.98) (9.00 36.00)	26.05 (4.50) (15.00 35.00)	0.002
HVLTRDLY	8.36(2.52) (0.00 12.00)	9.29 (2.32) (2.00 12.00)	< 0,001
SFT	48.67(11.63) (20.00 103.00)	1.80 (11.20) (22.00 80.00)	< 0,001
ADL	93.14(5.89) (70.00 100.00)	100.00(0.00) (100.00 100.00)	0.2118
SCOPA	9.51(6.18) (0.00 39.00)	5.83 (3.69) (0.00 20.00)	< 0,001
Marcadores CSF e imágenes DAT			
ABETA	909.55(410.73) (283.80 3707.00)	1019.37 (499.46) (239.10 3297.00)	0.029
ASYN	1506.70(666.59) (432.40 5256.90)	1695.19 (747.42) (488.60 4683.10)	0.0015
TAU	169.54(56.80) (80.93 467.00)	191.64 (79.26) (81.96 580.80)	0.0027
PTAU	14.87(5.25) (8.01 40.13)	17.52 (8.35) (8.08 73.61)	< 0,001
A β /TAU	5.56(1.60) (1.19 10.38)	5.58 (1.67) (0.71 9.80)	0.8336
MEANPUT	0.83(0.30) (0.24 2.17)	2.15 (0.56) (0.64 3.89)	< 0,001
MEANCAU	2.00(0.56) (0.39 3.71)	2.98 (0.63) (1.32 5.20)	< 0,001

Tabla 5.1: Características demográficas y clínicas de los pacientes con PD y HC al inicio del seguimiento.

5.2.2. Análisis de sujetos PD-CU PD-MCI y PDD al inicio del seguimiento

Características demográficas y clínicas					
Sujetos	PD-CU (N=323)	PD-MCI (N=83)	PD-Dementia (N=17)	pvalue	Post-Hoc
Age	60.13 (9.82) (33.50,82.25)	66.10 (7.37) (48.03,84.89)	69.45 (7.45) (54.82,82.98)	< 0.001	a,b
Women	120 (37.2%)	20 (24.1%)	6 (35.3%)	0.083	-
Education years	15.70 (2.87)(8.00 26.00)	14.99 (2.82) (8.00 21.00)	15.59 (4.91) (5.00 21.00)	0.0869	-
ABETA	314	80	16	-	-
GDS	2.17 (2.40) (0.00 14.00)	2.86 (2.61) (0.00 13.00)	2.71 (2.11) (0.00 8.00)	0.058	-
UPSIT	23.41 (7.87) (5.00 39.00)	19.54 (8.60) (1.00 40.00)	16.06 (7.39) (5.00 31.00)	< 0.001	a,b
REM	3.82 (2.54) (0.00 12.00)	5.06 (2.90) (0.00 12.00)	5.24 (3.11) (0.00 12.00)	< 0.001	a
Resultados cognitivos					
MOCA	27.33 (2.17) (19.00 30.00)	26.58 (2.70) (17.00 30.00)	25.94 (2.38) (21.00 29.00)	0.003	-
LNS	11.04 (2.50) (2.00 20.00)	9.29 (2.58) (4.00 17.00)	8.29 (2.78) (3.00 13.00)	< 0.001	a,b
SDMT	42.94 (8.75) (17.00 82.00)	36.69 (10.18) (7.00 58.00)	29.82 (10.85) (15.00 50.00)	< 0.001	a,b
JLO	12.98 (1.96) (5.00 15.00)	12.19 (2.44) (6.00 15.00)	11.53 (2.67) (6.00 15.00)	< 0.001	a,b
INMEDIATE	25.26 (4.74) (9.00 36.00)	22.13 (4.90) (11.00 34.00)	20.24 (4.28) (14.00 28.00)	< 0.001	a,b
HVLTDRDLY	8.75 (2.35) (0.00 12.00)	7.13 (2.75) (0.00 12.00)	7.00 (2.12) (3.00 11.00)	< 0.001	a,b
SFT	50.60 (11.54) (24.00 103.00)	43.48 (9.39) (20.00 70.00)	37.35 (9.37) (25.00 54.00)	< 0.001	a,b
SCOPA	8.74 (5.70) (0.00 35.00)	11.79 (6.96) (3.00 39.00)	13.06 (7.20) (4.00 32.00)	< 0.001	a,b
ADL	93.25 (6.02) (70.00 100.00)	93.01 (5.29) (80.00 100.00)	91.76 (6.11) (80.00 100.00)	0.583	-
Marcadores CSF y de imágenes DAT					
MEANCAU	2.07 (0.54) (0.39 3.71)	1.80 (0.54) (0.62 3.27)	1.59 (0.50) (0.74 2.52)	< 0.001	a,b
ABETA	942.40 (413.12) (315.90 3707.00)	844.08 (395.26) (238.80 2572.00)	592.27 (251.10) (308.80 1300.00)	0.001	b
ASYN	1519.17 (680.58) (432.40 5256.90)	1470.97 (620.62) (472.00 3443.70)	1443.49 (642.42) (539.40 2780.50)	0.783	-
TAU	167.17 (54.61) (80.93 467.00)	172.53 (59.27) (83.93 345.30)	202.68 (78.07) (97.17 327.40)	0.053	-
PTAU	14.52 (4.99) (8.02 40.13)	15.51 (5.40) (8.36 32.80)	18.59 (7.68) (8.01 30.84)	0.007	b
PTAU /A β	0.02 (0.01) (0.01 0.07)	0.02 (0.01) (0.01 0.08)	0.03 (0.02) (0.01 0.07)	< 0.001	a,b
UPDRSTOTON	31.35 (12.94) (7.00 72.00)	35.35 (13.12) (7.00 68.00)	36.94 (14.37) (18.00 70.00)	0.016	-
UPDRS1	5.21 (3.90) (0.00 24.00)	6.59 (4.42) (0.00 18.00)	7.53 (4.17) (2.00 18.00)	0.003	-
UPDRS2	5.63 (4.13) (0.00 22.00)	6.54 (4.24) (1.00 20.00)	8.12 (4.27) (2.00 15.00)	0.260	-
UPDRS3ON	20.51 (8.73) (5.00 47.00)	22.29 (9.40) (4.00 51.00)	21.29 (8.18) (10.00 39.00)	0.260	-

Tabla 5.2: Características demográficas y clínicas de los pacientes con PD-CU, PD-MCI y PDD al inicio del seguimiento.

A continuación, se procede al análisis entre los sujetos PD, según su estado cognitivo. Se determinan 3 grupos:

- PD-CU: sujetos con PD cognitivamente normal.
- PD-MCI: sujeto PD con declive cognitivo.
- PDD: sujeto PD con demencia.

Para conseguir un análisis completo que englobe los tres grupos se ha procedido a un análisis a partir de la herramienta anova1. Esto permite ver de forma más visual las diferencias que pueden existir entre las características y establecer factores de riesgo adecuados.

- a: diferencia entre PD-CU y PD-MCI.
- b: diferencia entre PD-CU y PDD.
- c: diferencia entre PD-MCI y PDD.

Tras el análisis completo de esta tabla se pueden establecer como algunos de los factores de riesgo más influyentes en el declive cognitivo en la enfermedad del Parkinson la edad, junto a la disminución de puntuación en los test neuropsicológicos llevados a cabo y la medida de pTau/A β .

5.3. Perfiles AT y tiempos de conversión en pacientes PD-novo

En este estudio se ha realizado un análisis transversal con sujetos control (HC) y pacientes con la enfermedad de Parkinson (PD).

5.3.1. Perfiles AT

Entre los individuos que padecen PD se puede hacer una clasificación entre los que presentan la patología amiloide (PD-A+) y los que no (PD-A-). Además, según el deterioro cognitivo que presenten se ha procedido a una clasificación entre cognitivamente normales (PD-CU), con deterioro cognitivo leve (PD-MCI) y con demencia (PDD).

En cuanto al perfil AT se refiere, este se analiza con el fin de intentar encontrar una relación entre el declive cognitivo de los sujetos y los valores de las proteínas $A\beta$ y pTau. Se establecen así las diferentes categorías:

- A representa los biomarcadores de las placas $A\beta$
- T representa los biomarcadores que determinan los valores de pTau
- N representa los biomarcadores de lesiones neuronales o neurodegeneración.

Como se menciona anteriormente, se ha establecido el siguiente criterio a lo largo del estudio para el análisis de los datos: un sujeto es A+ cuando sus valores de $A\beta$ < 683,45 pg/ml y T+ cuando la pTau > 23pg/ml [25]. En la población sometida a estudio, al inicio se obtuvieron los siguientes resultados.

En los pacientes PD-CU un 23.4 % presentaba patología amiloide (A+), que engloba tanto a los perfiles A+T+ como a los perfiles A+T-. Estos resultados son similares en sujetos con PD-MCI con un 26.5 %. Sin embargo, existen diferencias significativas entre estos sujetos y los diagnosticados con PDD, aumentando la patología amiloide hasta un 66.7 %.

Incluyendo en el análisis la patología pTau, se puede observar como en el caso de los pacientes con PD-CU, prácticamente no existen casos en el que ambas patologías coexistan (0.7 %). No obstante, se puede observar como en pacientes con PDD un 26.7 % son perfiles A+T+, manteniéndose a lo largo del estudio. Finalmente, cabe destacar como, al principio del estudio, los perfiles A-T- son predominantes en pacientes PD-CU (70.8 %), disminuyendo hasta 67.6 % en PD-MCI y 12.3 % en pacientes PDD.

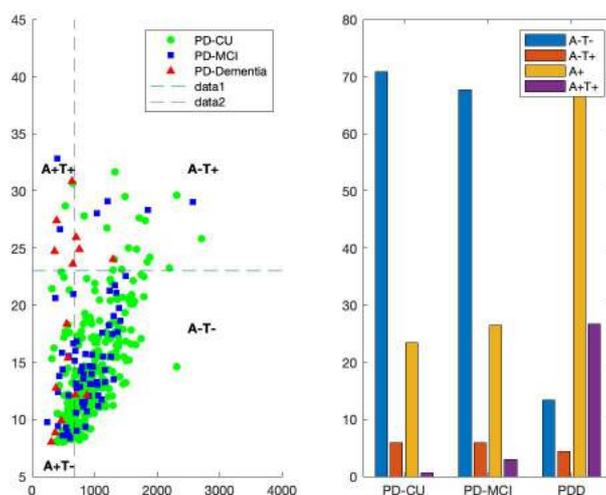


Figura 5.1: Perfiles AT al principio del estudio de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.

A lo largo de los cinco años se realizó un estudio longitudinal que permite observar la progresión de estos sujetos. Para la obtención de estos datos se seleccionaron los valores mínimos de $A\beta$ y máximos de pTau de cada paciente y aplicando el mismo criterio que en el estudio anterior, se obtuvieron los siguientes resultados.

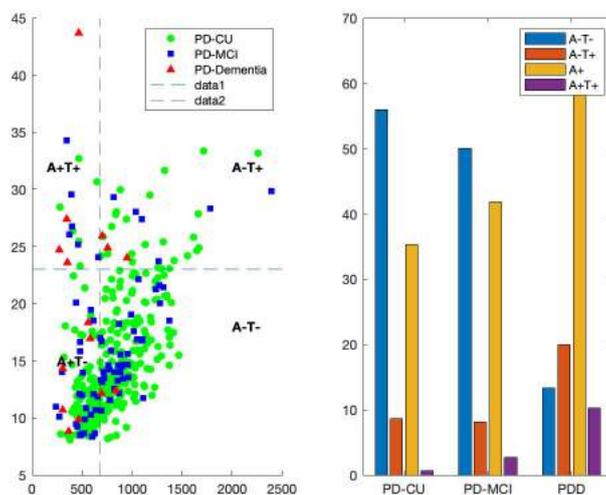


Figura 5.2: Perfiles AT tras el análisis longitudinal de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.

5.3.2. Tiempos de conversión

Los histogramas mostrados a continuación muestran de forma gráfica la conversión de los sujetos PD-CU sometidos a estudio a PD-MCI o PDD, además de los

tiempos de censura en los primeros cinco años de estudio.

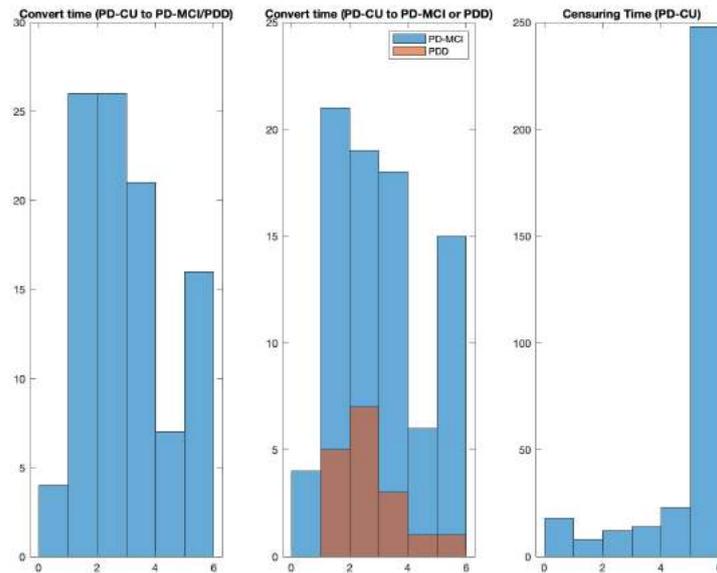


Figura 5.3: Histograma con los tiempos de conversión de los sujetos PD-novo y su evolución en los 5 primeros años de estudio: a) Conversión de sujetos PD-CU a PD-MCI o PDD, b) Distinción entre sujetos convertidos a PD-MCI y PDD c) Tiempo de censura de los sujetos PD-CU.

En la primera gráfica se muestra los tiempos de conversión de los sujetos PD-CU al principio del estudio a PD-MCI o PDD (sin distinción) durante los 5 primeros años de estudio. Se puede observar como a lo largo de este periodo aproximadamente un 24 % de los sujetos progresan de PD-CU a PD-MCI o PDD ($n=100/423$). Además, cabe destacar que la mayoría de los pacientes que progresan a PDD en el estudio aparecen en el segundo año ($n=7/423$) y a lo largo de los cinco años de estudio aproximadamente un 4 % convierte a demencia. Por otro lado, se puede observar como en cuanto a los casos de PD-MCI se refiere (predominantes a lo largo del experimento), aproximadamente un 20 % progresan a este estado ($n=83/423$).

Si se comparan estos resultados con los obtenidos con los sujetos control (HC), se pueden observar grandes diferencias. En este último grupo de pacientes, menos de un 3 % ($n=5/176$) convierte a MCI y no se detecta ningún diagnóstico de demencia en los 5 años de estudio.

5.4. Análisis estadístico

5.4.1. Modelos predictivos

Para comenzar, se realizan modelos predictivos utilizando una única variable aplicando un análisis de Cox. En cuanto al tiempo se refiere, se realiza el análisis con la edad y sin la edad como covariable.

Se obtienen los siguientes resultados:

Modelos univariantes con Age en el baseline					
Marcador	SEN	SPE	ACC	AUC	FREQ
SFT	71.76 % (71.42 % 72.10 %)	59.77 % (59.57 % 59.97 %)	62.56 % (62.39 % 62.72 %)	72.92 % (72.73 % 73.12 %)	7333
LNS	71.48 % (71.15 % 71.82 %)	60.04 % (59.84 % 60.23 %)	62.70 % (62.53 % 62.86 %)	62.70 % (62.53 % 62.86 %)	7185
SDMT	70.96 % (70.55 % 71.36 %)	60.02 % (59.79 % 60.26 %)	62.62 % (62.42 % 62.81 %)	72.71 % (72.47 % 72.94 %)	5140
pTauAB	71.73 % (71.30 % 72.15 %)	60.27 % (60.01 % 60.52 %)	62.98 % (62.76 % 63.19 %)	73.29 % (73.04 % 73.54 %)	4468
INMEDIATA	70.24 % (69.69 % 70.79 %)	60.95 % (60.62 % 61.27 %)	63.06 % (62.79 % 63.33 %)	72.58 % (72.26 % 72.90 %)	2780
HVLTRDLY	71.79 % (71.01 % 72.58 %)	60.00 % (59.54 % 60.47 %)	62.84 % (62.45 % 63.24 %)	72.84 % (72.37 % 73.30 %)	1291
MEANCAU	74.96 % (74.16 % 75.77 %)	60.58 % (60.09 % 61.07 %)	63.94 % (63.53 % 64.34 %)	75.00 % (74.50 % 75.50 %)	1167
ABETA	73.26 % (71.53 % 75.00 %)	60.12 % (58.97 % 61.28 %)	63.32 % (62.37 % 64.27 %)	74.79 % (73.73 % 75.86 %)	266
REM	71.47 % (69.47 % 73.47 %)	60.48 % (59.26 % 61.69 %)	63.30 % (62.37 % 64.23 %)	73.37 % (72.10 % 74.65 %)	156
SCOPA	73.17 % (70.68 % 75.65 %)	60.59 % (59.08 % 62.10 %)	63.71 % (62.41 % 65.01 %)	74.70 % (73.22 % 76.19 %)	139

Tabla 5.3: Modelos univariantes con **Age** como covariable

Modelos univariantes sin Age en el baseline					
Marcador	SEN	SPE	ACC	AUC	FREQ
LNS	58.73 % (58.29 % 59.17 %)	66.42 % (66.16 % 66.67 %)	64.57 % (64.39 % 64.74 %)	68.89 % (68.67 % 69.11 %)	7330
SFT	57.75 % (57.29 % 58.22 %)	66.51 % (66.24 % 66.78 %)	64.43 % (64.25 % 64.61 %)	68.59 % (68.36 % 68.82 %)	6927
SDMT	58.39 % (57.88 % 58.90 %)	66.69 % (66.40 % 66.99 %)	64.69 % (64.49 % 64.89 %)	69.15 % (68.89 % 69.41 %)	5590
pTauAB	59.06 % (58.50 % 59.62 %)	66.84 % (66.52 % 67.16 %)	64.97 % (64.76 % 65.19 %)	69.73 % (69.46 % 70.00 %)	4654
INMEDIATA	56.22 % (55.50 % 56.94 %)	66.69 % (66.26 % 67.11 %)	64.17 % (63.89 % 64.46 %)	67.95 % (67.59 % 68.30 %)	2777
HVLTRDLY	58.70 % (57.65 % 59.75 %)	66.36 % (65.72 % 67.00 %)	64.57 % (64.13 % 65.01 %)	69.57 % (69.05 % 70.10 %)	1226
MEANCAU	61.00 % (59.82 % 62.17 %)	66.40 % (65.64 % 67.15 %)	65.16 % (64.64 % 65.68 %)	70.66 % (70.03 % 71.29 %)	948
SCOPA	62.70 % (59.43 % 65.96 %)	67.63 % (65.90 % 69.36 %)	66.46 % (65.26 % 67.66 %)	73.27 % (71.54 % 75.01 %)	132
ABETA	60.40 % (56.46 % 64.34 %)	68.15 % (66.14 % 70.15 %)	66.63 % (65.10 % 68.17 %)	70.75 % (68.53 % 72.96 %)	104
REM	56.76 % (51.68 % 61.84 %)	67.44 % (64.62 % 70.26 %)	64.73 % (62.59 % 66.86 %)	70.23 % (67.55 % 72.91 %)	52

Tabla 5.4: Modelos univariantes sin **Age** como covariable

Tras el estudio longitudinal multivariante se obtienen los siguientes resultados:

Time	Marcador	SEN	SPE	ACC	AUC	FREQ
Baseline	LNS, SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, pTauAB	74.59 % (73.81 % 75.38 %)	72.37 % (71.93 % 72.81 %)	72.84 % (72.49 % 73.20 %)	81.68 % (81.30 % 82.06 %)	1422
Baseline	LNS, SDMT, HVLTRDLY, SFT, SCOPA, MEANCAU, pTauAB	74.85 % (74.18 % 75.53 %)	72.50 % (72.11 % 72.90 %)	73.06 % (72.74 % 73.38 %)	81.26 % (80.89 % 81.62 %)	1776
Baseline	LNS, SDMT, INMEDIATA, SFT, MEANCAU, ABETA, pTauAB	73.39 % (72.54 % 74.25 %)	72.76 % (72.31 % 73.21 %)	72.84 % (72.47 % 73.22 %)	81.40 % (80.97 % 81.82 %)	1227
m12	REM, LNS, SDMT, INMEDIATA, MEANCAU, pTauAB	68.13 % (67.31 % 68.94 %)	68.98 % (68.57 % 69.39 %)	68.68 % (68.32 % 69.04 %)	77.58 % (77.15 % 78.02 %)	2212
m12	REM, LNS, SDMT, INMEDIATA, MEANCAU, pTauAB	67.72 % (66.71 % 68.74 %)	68.59 % (68.07 % 69.11 %)	68.21 % (67.76 % 68.66 %)	77.40 % (76.85 % 77.95 %)	1401
m12	REM, LNS, SDMT, INMEDIATA, MEANCAU, pTauAB	68.90 % (67.62 % 70.17 %)	68.55 % (67.90 % 69.20 %)	68.48 % (67.90 % 69.06 %)	78.18 % (77.52 % 78.85 %)	896
m24	REM, LNS, SDMT, INMEDIATA, MEANCAU, pTauAB	73.22 % (72.23 % 74.22 %)	72.38 % (71.87 % 72.89 %)	72.41 % (71.97 % 72.85 %)	80.92 % (80.00 % 81.04 %)	1457
m24	LNS, SDMT, HVLTRDLY, SFT, SCOPA, MEANCAU, pTauAB	73.51 % (72.45 % 74.57 %)	71.41 % (70.86 % 71.96 %)	71.66 % (71.19 % 72.13 %)	80.21 % (79.64 % 80.78 %)	1222
m24	REM, LNS, SDMT, HVLTRDLY, MEANCAU, pTauAB	75.02 % (73.93 % 76.11 %)	71.41 % (70.82 % 72.00 %)	72.13 % (71.63 % 72.63 %)	80.88 % (80.30 % 81.47 %)	1140

Tabla 5.5: Modelos multivariantes con **Age** como covariable

Time	Marcador	SEN	SPE	ACC	AUC	FREQ
Baseline	LNS, SDMT, SFT, SCOPA, MEANCAU, pTauAB	71.77 % (70.85 % 72.68 %)	75.20 % (74.73 % 75.68 %)	74.31 % (73.92 % 74.71 %)	81.21 % (80.71 % 81.71 %)	1039
Baseline	LNS, SDMT, HVLTRDLY, SFT, SCOPA, pTauAB	71.67 % (70.67 % 72.67 %)	75.41 % (74.92 % 75.90 %)	74.53 % (74.11 % 74.95 %)	81.30 % (80.78 % 81.83 %)	955
Baseline	LNS, SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, pTauAB	72.57 % (71.81 % 73.34 %)	75.27 % (74.88 % 75.65 %)	74.62 % (74.29 % 74.95 %)	81.56 % (81.15 % 81.96 %)	1438
m12	LNS, SDMT, SCOPA, pTauAB	62.11 % (61.09 % 63.14 %)	75.15 % (74.63 % 75.67 %)	71.76 % (71.31 % 72.21 %)	75.09 % (74.46 % 75.72 %)	1367
m12	LNS, SDMT, MEANCAU, pTauAB	62.58 % (61.50 % 63.66 %)	75.33 % (74.77 % 75.89 %)	72.14 % (71.66 % 72.62 %)	75.43 % (74.76 % 76.10 %)	1231
m12	REM, LNS, SDMT, HVLTRDLY, MEANCAU, pTauAB	63.30 % (62.12 % 64.48 %)	73.79 % (73.18 % 74.40 %)	71.22 % (70.69 % 71.75 %)	74.76 % (74.01 % 75.50 %)	1052
m24	REM, LNS, SDMT, pTauAB	65.65 % (64.87 % 66.44 %)	76.66 % (76.27 % 77.05 %)	74.15 % (73.81 % 74.49 %)	78.81 % (78.37 % 79.25 %)	2424
m24	LNS, SDMT, MEANCAU, pTauAB	65.77 % (64.82 % 66.71 %)	76.75 % (76.29 % 77.20 %)	74.33 % (73.93 % 74.73 %)	78.71 % (78.19 % 79.22 %)	1801
m24	LNS, SDMT, SCOPA, pTauAB	65.61 % (64.53 % 66.70 %)	76.32 % (75.78 % 76.86 %)	73.90 % (73.43 % 74.38 %)	78.80 % (78.21 % 79.39 %)	1278

Tabla 5.6: Modelos multivariantes sin **Age** como covariable

A continuación, se procede a analizar los modelos predictivos compuestos por **varias características**. Para ello, se analizan los resultados en tres periodos de tiempo: al inicio del estudio, a los 12 meses y a los 24 meses. Se obtienen resultados con y sin la edad como covariable para comparar los resultados.

Una vez se hayan obtenido todos los resultados de los distintos modelos univariantes y multivariantes, se procede a seleccionar uno de ellos en función del mayor

rendimiento que presente. A continuación se presentan los resultados:

Si se considera la edad como covariable, se propone el siguiente vector:

LNS, SDMT, HVLTRDLY, MEANCAU, pTAUAB, Age

Si no se considera la edad como covariable se propone el vector de características:

LNS, SDMT, MEANCAU, pTAUAB

5.4.2. Modelos de progresión con GRACE

Una vez se ha propuesto un vector de características, se procede a analizar los resultados de los modelos de progresión correspondientes, que permiten estudiar el patrón o tendencia de los datos y la evolución a lo largo del tiempo.

Modelos de progresión a partir de los modelos predictivos multivariantes

Si analizamos los modelos de progresión con la edad como covariable, se hace uso del vector de características:

[LNS, SDMT, HVLTRDLY, MEANCAU, pTAUAB, Age]

Donde se encuentran los test neuropsicológicos:

- LNS¹
- SDMT²
- HVLTRDLY³

Además de las características clínicas:

- MEANCAU
- pTAUAB

y la edad.

Se obtienen las siguientes matrices de correlación para obtener la relación lineal entre el conjunto de datos:

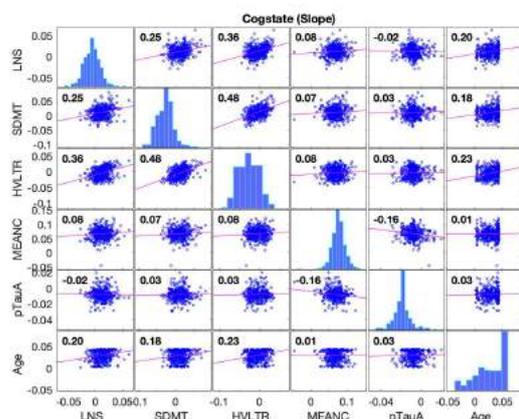


Figura 5.4: Matriz de correlación de la pendiente según el criterio SITE con Age.

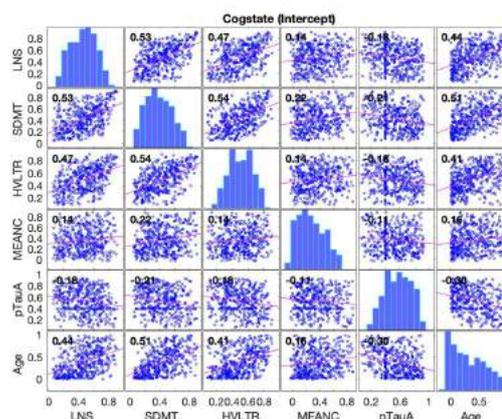


Figura 5.5: Matriz de correlación del intercepto según el criterio SITE con Age.

Donde la mayor correlación de la pendiente y el intercepto se encuentra entre HVLTRDLY y SDMT.

A continuación, se procede a analizar las trayectorias de las características intrínsecas del vector previamente seleccionado. Estas trayectorias son esenciales para comprender cómo evolucionan y cambian las variables en estudio a lo largo del tiempo. Al examinar estas trayectorias, se permite la identificación de patrones, tendencias y posibles variaciones en el comportamiento de las características del vector seleccionado.

¹Letter-Number Sequencing

²Symbol Digit Modalities Test

³Hopkins Verbal Learning Test Revised Delayed Recall

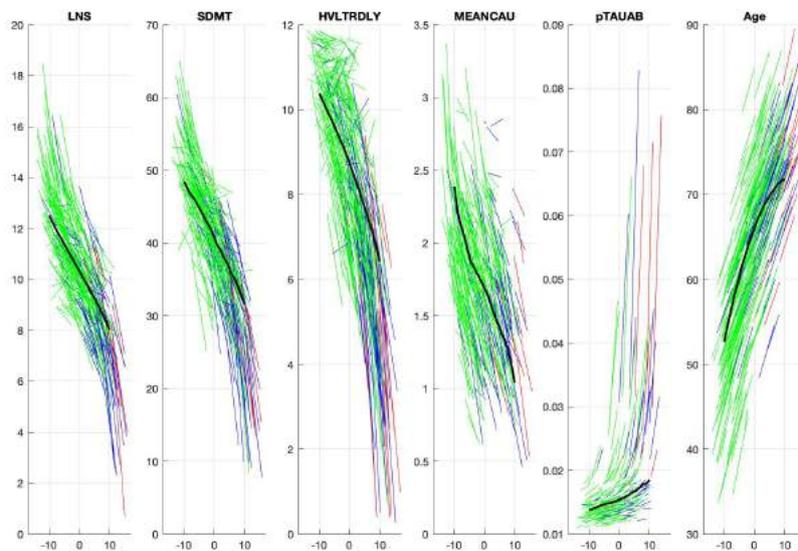


Figura 5.6: Trayectorias de los biomarcadores del vector con **Age**. Trayectorias a corto plazo en verde para sujetos PD-CU, azul para PD-MCI y rojo para PDD. Trayectoria a largo plazo representada en negro.

En la Figura 5.6 se pueden observar las diferentes trayectorias a corto plazo de los sujetos representadas a lo largo del tiempo. El valor 0 del eje x indica el momento en el que el paciente convierte. Se muestra en negro la trayectoria que sigue a largo plazo cada característica, obtenida en función de las trayectorias anteriores de cada sujeto.

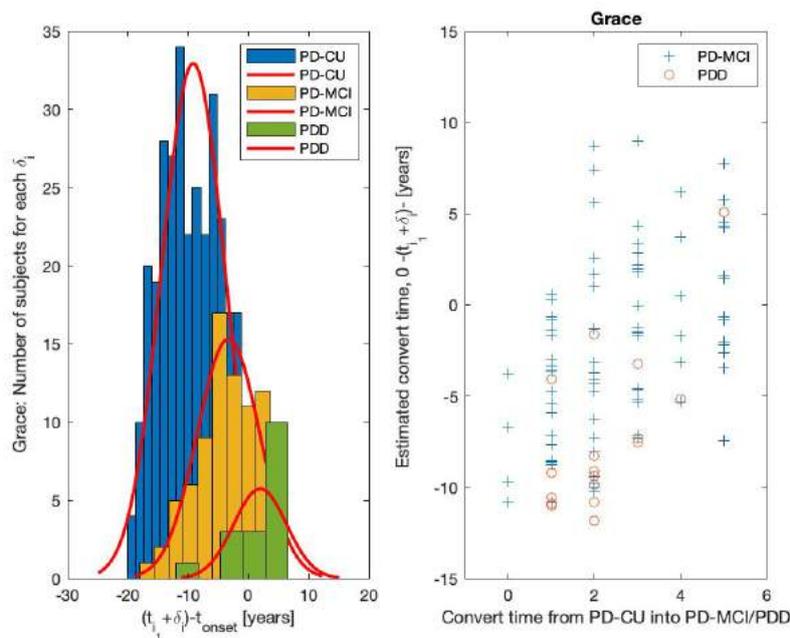


Figura 5.7: Distribución de los sujetos PD según el tiempo de conversión a partir del vector con **Age** como covariable

De una forma más visual, se puede observar en la Figura 5.7 el número de pacientes PD-CU, PD-MCI y PDD del experimento y sus tiempos de conversión.

De manera análoga se realiza este análisis con el vector de características seleccionado sin la edad como covariable:

[LNS, SDMT, MEANCAU, pTAUAB]

Donde se encuentran los test neuropsicológicos:

- LNS
- SDMT

y las características clínicas:

- MEANCAU
- pTAUAB

Se obtienen las siguientes matrices de correlación a partir de este vector:

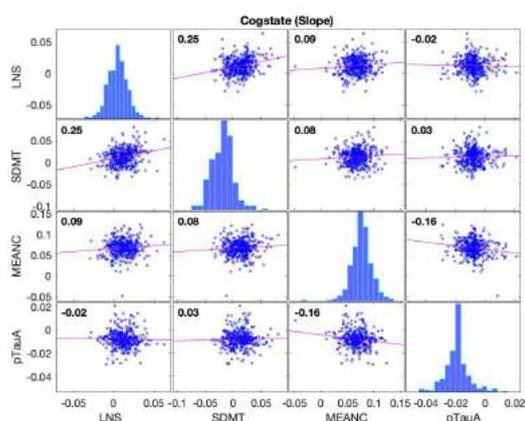


Figura 5.8: Matriz de correlación de la pendiente según el criterio SITE sin Age.

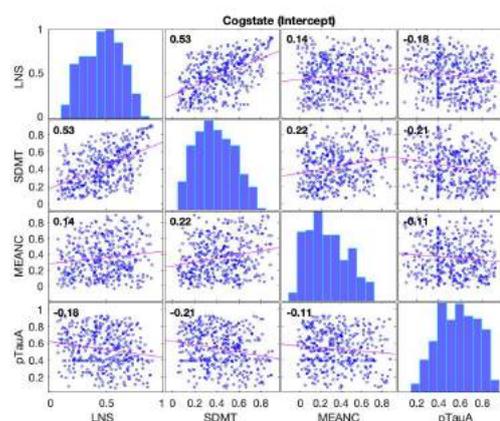


Figura 5.9: Matriz de correlación del intercepto según el criterio SITE sin Age.

Si se procede al análisis de estas matrices, se puede observar como la correlación más alta se encuentra en la matriz de intercepto entre los tests LNS y SDMT (0,53). Por otro lado, la correlación más negativa está definida por las variables pTAUAB y SDMT (-0,23).

A continuación, en el estudio de las trayectorias del modelo progresivo se obtienen los siguientes resultados:

Al igual que en el vector con la edad como covariable, la Figura 5.10 representa las trayectorias según los biomarcadores seleccionados para los sujetos sometidos a estudio. Además, en la Figura 5.11 se puede observar el número de pacientes con su tiempo de conversión correspondiente.

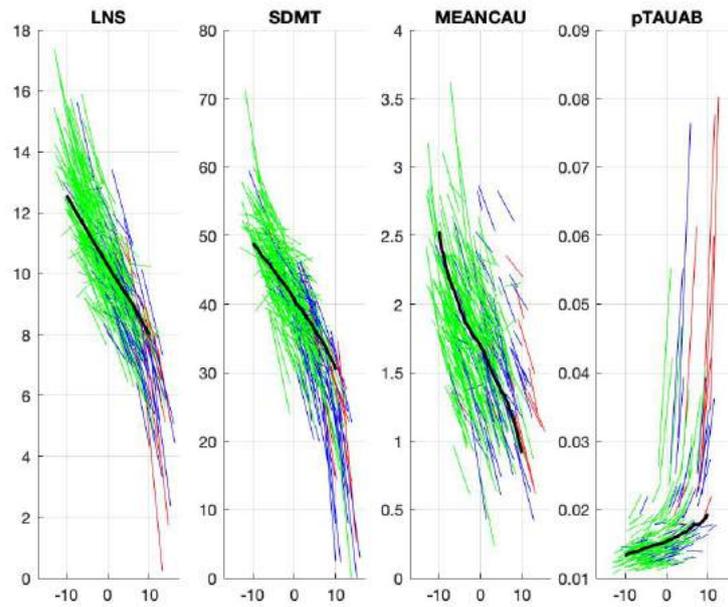


Figura 5.10: Trayectorias de los biomarcadores del vector sin **Age**. Trayectorias a corto plazo en verde para sujetos PD-CU, azul para PD-MCI y rojo para PDD. Trayectoria a largo plazo representada en negro.

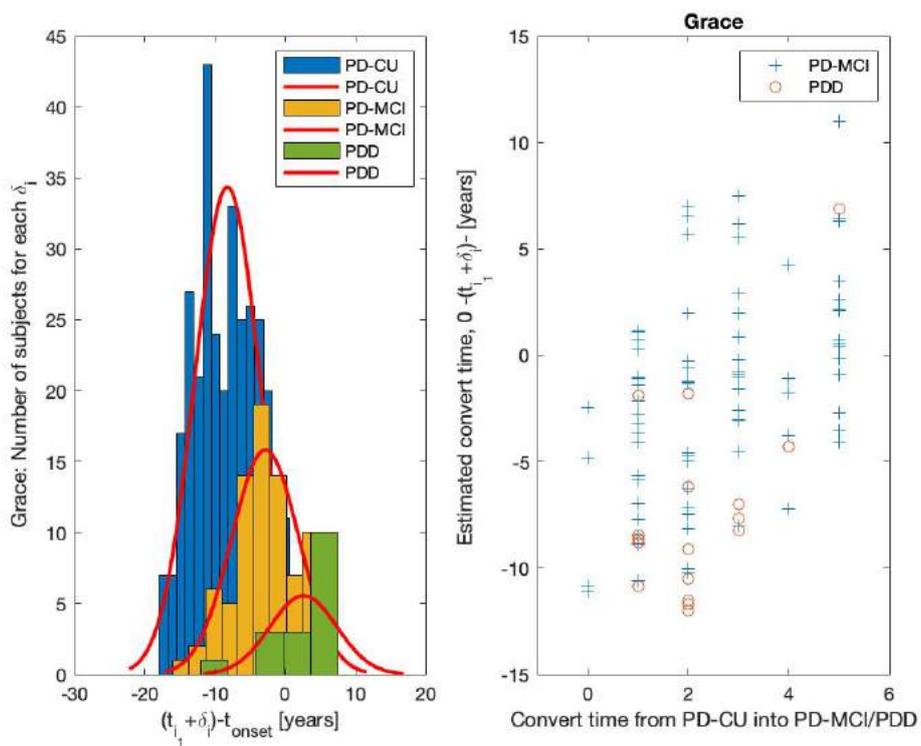


Figura 5.11: Distribución de los sujetos PD según el tiempo de conversión a partir del vector sin **Age** como covariable

5.5. Modelos de progresión de forma masiva con Magerit

Una vez se han ejecutado el código *multi_grace_PPMI.m* se obtienen las combinaciones de características con mejor rendimiento.

En este estudio se ha hecho una evaluación de los mejores vectores con y sin la edad, para comparar así cuales ofrecen mejores resultados.

Si se contempla la edad como covariable, se obtienen estos resultados:

Vector	Time onset	t	R	Scores1	Scores2	Scores3
SDMT HVLTRDLY SFT MEANCAU A β Age	4.4	1.78x-6.44	0.54	75.23 %	66.00 %	58.00 %
LNS SDMT HVLTRDLY SFT MEANCAU A β Age	5.8	1.92x-7.23	0.54	81.73 %	65.00 %	64.00 %
LNS HVLTRDLY SFT MEANCAU A β Age	5	1.73x-6.23	0.53	79.26 %	65.00 %	64.00 %

Tabla 5.7: Características de los vectores seleccionados, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.

En el momento que se produce el deterioro cognitivo (time onset) de cada vector, los biomarcadores de estos toman los siguientes valores:

Vector	LNS	SDMT	HVLTRDLY	SFT	MEANCAU	A β
SDMT HVLTRDLY SFT MEANCAU A β Age	-	36.98	7.77	43.66	1.44	762.16
LNS SDMT HVLTRDLY SFT MEANCAU A β Age	9.25	35.40	7.44	43.03	1.38	877.86
LNS HVLTRDLY SFT MEANCAU A β Age	9.26	-	7.56	43.67	1.41	747.52

Tabla 5.8: Valores de diferentes características según el tiempo onset del vector.

Vector	MOCA	Age
SDMT HVLTRDLY SFT MEANCAU A β Age	25.74	69.16
LNS SDMT HVLTRDLY SFT MEANCAU A β Age	25.08	70.32
LNS HVLTRDLY SFT MEANCAU A β Age	25.35	69.61

Tabla 5.9: Valores del test MoCA y la edad según el tiempo onset del vector.

Una vez se han obtenido estos resultados, se evalúa el tiempo (en años) que dura la evolución del declive cognitivo a demencia a partir de la puntuación del test neuropsicológico MoCA. Un valor menor de 25 indica que se ha iniciado el deterioro cognitivo (PD-MCI) mientras que un valor menor de 21 indicaría la progresión a demencia (PDD).

Vector	$t_{MOCA=21}-t_{MOCA=25}$	$A\beta_{683}-t_{MOCA=25}$
SDMT HVLTRDLY SFT MEANCAU $A\beta$ Age	7.3	0.5
LNS SDMT HVLTRDLY SFT MEANCAU $A\beta$ Age	7.7	1.9
LNS HVLTRDLY SFT MEANCAU $A\beta$ Age	7.2	0.7

Tabla 5.10: Tiempo entre el inicio del deterioro cognitivo y demencia según las puntuaciones del test MoCA y el biomarcador $A\beta$ para los diferentes vectores seleccionados con la edad.

Se procede a un análisis análogo sin la edad como covariable para analizar los diferentes resultados, obteniéndose así:

Vector	Time onset	t	R	Scores1	Scores2	Scores3
SDMT HVLTRDLY SFT MEANCAU pTau $A\beta$	6	$1.94x-7.11$	0.55	82.35 %	62.00 %	61.00 %
LNS HVLTRDLY SFT MEANCAU pTau $A\beta$	5.8	$2x-7.4$	0.55	80.80 %	64.00 %	62.00 %

Tabla 5.11: Características de los vectores seleccionados, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.

Los biomarcadores de estos vectores en el time onset tienen los siguientes valores:

Vector	LNS	SDMT	HVLTRDLY	SFT	MEANCAU	$A\beta$
SDMT HVLTRDLY SFT MEANCAU pTau $A\beta$	-	34.97	7.24	42.56	1.36	727.53
LNS HVLTRDLY SFT MEANCAU pTau $A\beta$	9.20	-	7.27	42.87	1.41	735.42

Tabla 5.12: Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.

Vector	MOCA
SDMT HVLTRDLY SFT MEANCAU pTau $A\beta$	24.95
LNS HVLTRDLY SFT MEANCAU pTau $A\beta$	25.11

Tabla 5.13: Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.

El tiempo (en años) que transcurre entre el inicio del deterioro cognitivo y la demencia de cada vector es:

Vector	$t_{MOCA=21}-t_{MOCA=25}$	$A\beta_{683}-t_{MOCA=25}$
SDMT HVLTRDLY SFT MEANCAU pTau $A\beta$	7.5	1.8
LNS HVLTRDLY SFT MEANCAU pTau $A\beta$	7.9	1.8

Tabla 5.14: Tiempo entre el inicio del deterioro cognitivo y demencia según las puntuaciones del test MoCA y el biomarcador $A\beta$ para los diferentes vectores seleccionados sin la edad.

Se selecciona como vector más óptimo:

$$SDMT\ HVLTRDLY\ SFT\ MEANCAU\ pTauA\beta$$

El cual presenta las siguientes trayectorias a corto y largo plazo con un time onset de 4.2 años y un coeficiente de Pearson de 0.55:

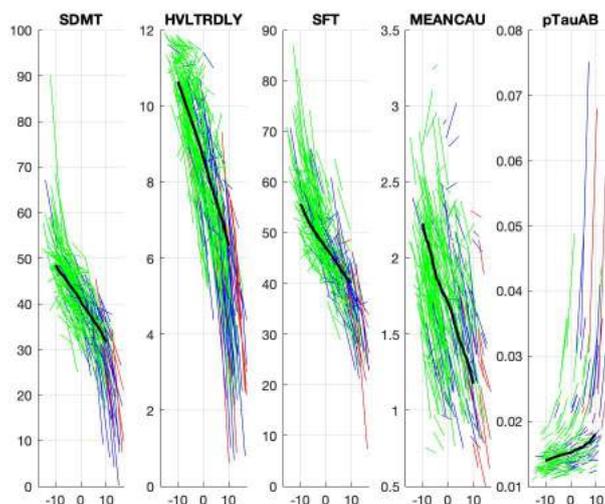


Figura 5.12: Trayectorias a corto y largo plazo de las características del vector con mayor rendimiento.

5.5.1. Bootstrap

Al realizar el análisis del mejor vector con la técnica bootstrap se obtienen las siguientes trayectorias en los test neuropsicológicos y biomarcadores:

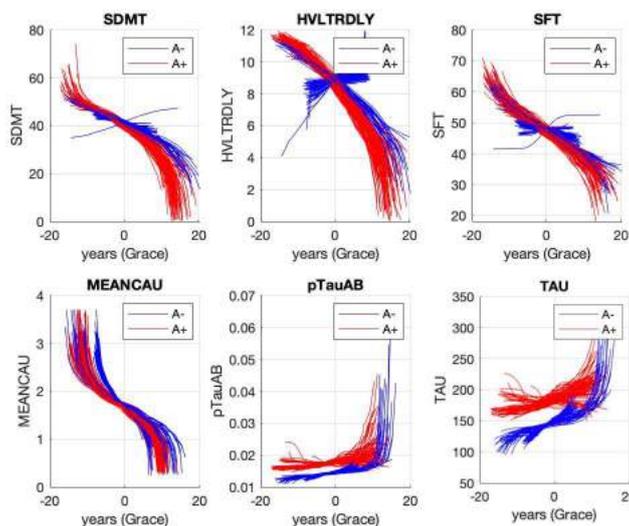


Figura 5.13: Distribución de los sujetos PD según el tiempo de conversión a partir del vector con **Age** como covariable

Donde los perfiles A+ se representan en color rojo y los perfiles A- en azul.

Capítulo 6

Discusiones

6.1. Introducción

En este capítulo, se procede al análisis e interpretación de los resultados obtenidos a lo largo de la investigación de una forma más precisa. Esto permitirá obtener conclusiones sólidas acerca de lo investigado y proporcionar una interpretación coherente y fundamentada en los resultados, respaldada en todo momento por los artículos previos.

Este capítulo finalizará con una conclusión global del proyecto y se propondrán nuevas líneas de investigación que permitan avanzar en lo estudiado hasta ahora.

6.2. Análisis de los sujetos al inicio del seguimiento

6.2.1. Análisis de los sujetos HC y PD

En el análisis entre los sujetos HC y PD, los resultados de la Tabla 5.1 muestran las características de ambos grupos sometidos a estudio con sus valores al inicio. Como se puede observar, el pvalor de la mayoría de características expuestas en la tabla es menor a 0.001, lo que implica que existen diferencias significativas entre los resultados obtenidos por los pacientes PD y HC.

No obstante, en ciertas pruebas neuropsicológicas como LNS, JLO o ADL se pueden observar valores muy similares entre ambos grupos (se puede ver referenciado en el pvalor). En cuanto a los marcadores relacionados con las imágenes DAT y CSF cabe destacar la disminución en las características $A\beta$, ASYN, Tau y P τ en el grupo PD respecto del grupo HC.

6.2.2. Análisis

Si el análisis se centra en el estado cognitivo de los pacientes PD (Tabla 5.2), se puede observar como en cuanto a características demográficas se refiere, los hombres padecen mayoritariamente la enfermedad en los tres grupos, aunque no hay diferencias altamente significativas respecto de las mujeres. Sin embargo, se puede observar como la edad supone un factor de riesgo para el desarrollo del declive

cognitivo en los pacientes con PD, aumentando desde los 60.1 años de media en las personas con PD-CU hasta los 66.1 años en PD-MCI y 69.45 años en pacientes PDD.

Si se analizan los resultados obtenidos en las pruebas cognitivas se puede observar como en MoCA las diferencias entre los tres grupos son mínimas (27.33, 26.58 y 25.94) para los grupos PD-CU, PD-MCI y PDD respectivamente. En cuanto a los marcadores en el líquido cefalorraquídeo (CSF) e imágenes DAT cabe destacar el dato de la $A\beta$ que disminuye notablemente entre los sujetos sin deterioro cognitivo y aquellos con demencia pasando de una media de 942.4 a 592.27 respectivamente. Otros cambios menos significativos, pero también existentes en estos marcadores se pueden encontrar en los valores de Tau o pTau, ambos aumentando a medida que avanza el deterioro cognitivo en los sujetos.

Para la correcta multicomparación entre los tres grupos se ha procedido a un análisis a partir de la herramienta *anova1*. Esto permite ver de forma más visual las diferencias que pueden existir entre las características y establecer factores de riesgo adecuados.

- a: diferencia entre PD-CU y PD-MCI.
- b: diferencia entre PD-CU y PDD.
- c: diferencia entre PD-MCI y PDD.

A continuación, se puede observar de una forma visual las discrepancias entre los diferentes grupos en los distintos marcadores. Se puede observar como en las pruebas cognitivas realizadas a los diferentes sujetos como la secuenciación de números y letras (LNS), la prueba de modalidades de dígitos de símbolo (SDMT) o el juicio de orientación lineal (JLO) las diferencias entre los tres grupos de pacientes son significativas. Por ejemplo, en el test LNS la puntuación disminuye de una media de 11.04 en pacientes con PD-CU a 9.29 en aquellos con PD-MCI. Si comparamos los valores entre los grupos PD-CU y PDD observamos como los valores disminuyen desde 11.04 hasta 8.29 respectivamente.

Estas diferencias también son notables en ciertos factores determinados por los marcadores DAT y CSF como la medida del caudado (MEANCAU) donde los grupos PD-CU, PD-MCI y PDD presentan valores medios de 2.07, 1.80 y 1.59 respectivamente.

Tras el análisis completo de esta tabla se pueden establecer como algunos de los factores de riesgo más influyentes en el declive cognitivo en la enfermedad del Parkinson la edad, junto a la disminución de puntuación en los test neuropsicológicos llevados a cabo y la medida de pTau/ $A\beta$.

6.3. Perfiles AT y tiempos de conversión y censura

Si se comparan los resultados al inicio del estudio y tras el análisis longitudinal se observan cambios significantes sobretodo en la patología amiloide en sujetos PD-CU y PD-MCI. Sin embargo, los perfiles AT de sujetos PDD no sufren variaciones a lo largo del experimento.

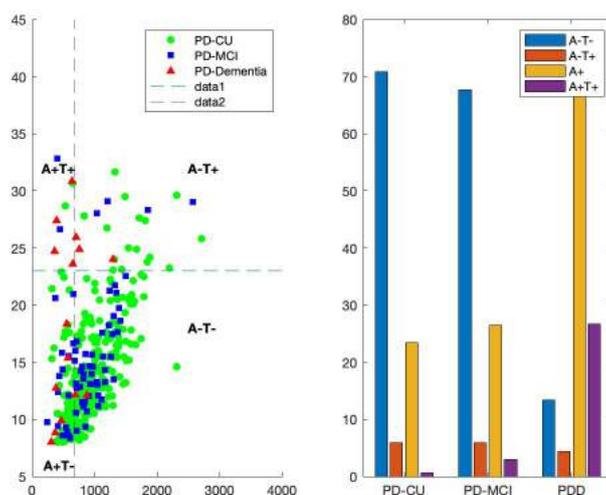


Figura 6.1: Perfiles AT al principio del estudio de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.

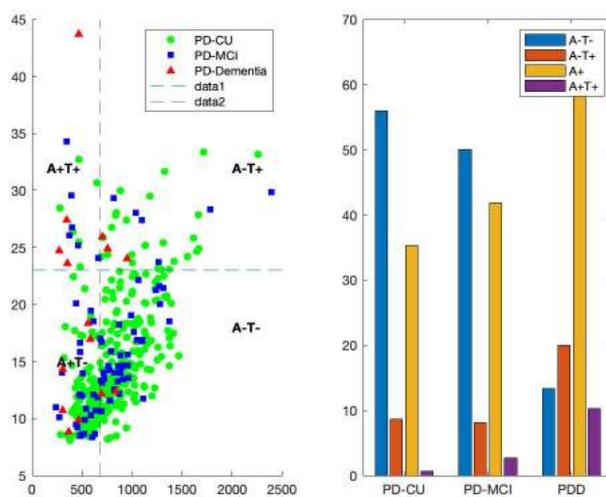


Figura 6.2: Perfiles AT tras el análisis longitudinal de los sujetos PD-CU, PD-MCI y PDD: a) Diagrama de dispersión entre $A\beta$ y pTau para los diferentes grupos, b) Porcentaje de perfiles AT según el estado cognitivo.

Se puede observar como, al final del estudio, la patología amiloide aumenta obteniéndose así: un 36.7 % en sujetos PD-CU y 41.63 % en PD-MCI (frente 23.4 % y 26.5 % al inicio del estudio respectivamente). Sin embargo en los sujetos PDD se mantiene constante (66.7 %). No se observan otros cambios importantes.

Tras el análisis de estos datos se pueden recoger los resultados obtenidos en la siguiente tabla:

Clinical	Time	A+	A+T+	A-T+	A-T-
PD-CU (n=291)	Baseline	23.37 %	0.69 %	5.84 %	70.79 %
	Long	35.33 %	2.00 %	8.67 %	56.00 %
PD-MCI (n=68)	Baseline	26.47 %	2.94 %	5.88 %	67.65 %
	Long	41.89 %	8.11 %	8.11 %	50.00 %
PDD (n=15)	Baseline	66.67 %	26.67 %	20.00 %	13.33 %
	Long	66.67 %	26.67 %	20.00 %	13.33 %

Tabla 6.1: Porcentajes de los perfiles AT en el estudio al inicio del estudio y tras el análisis longitudinal.

En cuanto a los tiempos de conversión se refiere, se puede observar en la Figura 5.3 como aproximadamente el 24 % (n=100/423) de los sujetos experimenta una progresión a PD-MCI o PDD, siendo más notoria en el segundo año, donde la mayoría de los casos de conversión a PDD ocurren (n=7/423). A lo largo de los 5 años, cerca del 4 % desarrolla demencia. En contraste, alrededor del 20 % de los casos de PD-MCI (que predominan durante el estudio) evolucionan hacia este estado.

Comparados con el grupo de control HC, los resultados revelan diferencias significativas. Menos del 3 % de los pacientes en el grupo de control (n=5/176) transitan a MCI, y no se identifica demencia en 5 años.

6.4. Análisis estadístico

6.4.1. Modelos predictivos

Una vez se han ejecutado los scripts necesarios con los que se obtienen los diferentes modelos, se estudian los resultados obtenidos.

A la hora de analizar un modelo predictivo, es importante considerar el equilibrio que existen entre todas las medidas a analizar (SEN¹,SPE²,ACC³,AUC⁴ y frecuencia).

Modelos predictivos univariantes

Tanto si se considera la edad como covariable en la tabla (Tabla 5.3) como si no (Tabla 5.4), se obtienen modelos con una ACC similar, en torno al 63 % y 65 % respectivamente. Esto indica que hay una alta proporción de casos clasificados correctamente por el modelo (tanto positivos como negativos).

Si se analiza el rendimiento del modelo a partir de factores como la SEN y SPE, se puede observar como existen diferencias en torno al 10 % entre sus valores en los modelos mostrados en la Tabla 5.3. Para poder considerar un rendimiento óptimo en un modelo predictivo es necesario un equilibrio entre estas dos métricas, pues representan el rendimiento en la detección de casos positivos (SEN) y negativos (SPE). En este caso, dado que la sensibilidad del modelo es ligeramente mayor, se

¹Sensibilidad.

²Especificidad.

³Exactitud.

⁴Área Bajo la Curva.

clasificarán muchos casos como positivos, lo que puede aumentar los falsos positivos.

Si se examinan los resultados de la Tabla 5.4, se observa como la diferencia entre estas dos variables es aproximadamente de un 8%. En este caso, los valores de SPE son ligeramente mayores, lo que puede significar que se pasen por alto algunos casos positivos.

Finalmente, si se evalúa la capacidad global del modelo para distinguir entre las clases a partir de la AUC en ambos modelos con y sin edad, se puede concluir que en términos generales, los modelos en los que se incluye esta covariable muestran resultados ligeramente mejores (en torno al 73%) en comparación con los modelos que no (en torno al 69%).

Modelos predictivos multivariantes

A continuación, se procede al análisis de modelos multivariantes en diferentes periodos de tiempo, para poder comparar y elegir los resultados más óptimos.

Al inicio del estudio se utiliza un análisis de Cox para obtener las propiedades necesarias del modelo. Se puede observar como en las tablas 5.5 y 5.6 la ACC es muy elevada en los 3 modelos seleccionados con y sin la covariable, obteniendo unos valores máximos de 73,06% y 74,62% respectivamente. Si se compara con los modelos univariantes, se puede observar como existe una mejora significativa de los casos detectados correctamente con las características propuestas, aumentando en torno al 10%.

A la hora de estudiar el rendimiento en el *baseline*, se consigue un equilibrio entre los valores de SEN y SPE en ambos modelos. Si se contempla la edad como covariable (Tabla 5.5) se obtienen valores de SEN y SPE como: 74,59% y 72,37%, 74,85% y 72,50% o 73,39% y 72,84% respectivamente. Si no se incluye la edad (Tabla 5.6) se consiguen resultados como: 71,77% y 75,20%, 71,67% y 75,41% o 72,57% y 75,27%. Estos resultados demuestran un rendimiento mayor a los modelos univariantes con resultados como: 71,76% y 59,77% o 58,73% y 66,42%.

En cuanto a la distinción de clases se refiere, se obtiene una mejora en torno al 10% en modelos con y sin la edad multivariantes respecto a los univariantes, obteniendo valores en torno a 81,40% (con la edad) y 81,30% (sin ella).

Para analizar las características del modelo a los 12 y 24 meses, se procede a aplicar un análisis de Cox extendido.

Cuando ha transcurrido un año desde el estudio, se pueden observar como todos los valores del modelo muestran una peor valoración, descendiendo en torno a un 5% en todas las características. En la Tabla 5.5 se muestra como el mejor modelo tiene las siguientes características:

SEN:68,13% SPE:68,98% ACC:68,68% AUC:77,58% FREQ:2212,

que son similares al modelo más óptimo sin la edad:

SEN: 62,11 % SPE: 75,15 % ACC: 71,76 % AUC: 75,09 % FREQ: 1367.

Sin embargo, al estudiar los resultados del modelo a los 24 meses, se consigue una mejora significativa de todos los parámetros del modelo, no solo respecto a los 12 meses, si no también respecto al inicio.

Cabe destacar la diferencia de resultados (hasta ahora no relevante) entre los modelos con y sin la edad. Se puede observar como se obtienen resultados más óptimos en la SEN del modelo cuando se contempla la covariable, no solo porque es más elevada (73,22 % respecto 65,65 %), si no porque hay un mayor equilibrio entre las medidas de SEN y SPE, resultando en un mayor rendimiento. También se encuentran mejoras a la hora de evaluar la calidad del modelo mediante la AUC, obteniendo valores más elevados en los modelos con la edad (80,80 % respecto 78,80 %). El análisis en este periodo de tiempo proporciona los mejores resultados en términos generales.

Una vez se han analizado todos los modelos obtenidos en los distintos periodos de tiempo, teniendo en cuenta diferentes características y covariables, se procede a proponer un vector de características que consiga las mejores propiedades del modelo.

Si se considera la edad como covariable, se propone el siguiente vector:

LNS, SDMT, HVLTRDLY, MEANCAU, pTAUAB, Age

Si no se considera la edad como covariable se propone el vector de características:

LNS, SDMT, MEANCAU, pTAUAB

6.4.2. Modelos de progresión con GRACE

Edad como covariable

Tras el estudio del vector: [LNS, SDMT, HVLTRDLY, MEANCAU, pTAUAB, Age] con GRACE, se obtienen los siguientes resultados:

Como se puede observar en las figuras 5.4 y 5.5, el coeficiente de relación más elevado es 0,54, entre las características HVLTRDLY y SDMT, lo cual implica que estas tienden a aumentar o disminuir juntas. La correlación más negativa más alta es de -0,30 entre la pTAUAB y la edad. Esto implica que mientras una tiende a aumentar, la otra disminuye.

La naturaleza de esta correlación sugiere una relación significativa entre los cambios en los valores de HVLTRDLY y SDMT, lo que podría indicar cierta coherencia en la manera en que estas dos variables se desarrollan a lo largo del tiempo.

Sin edad como covariable

Se hace uso del vector [LNS, SDMT, MEANCAU, pTAUAB] con GRACE, obteniendo un coeficiente de correlación entre las variables implicadas de 0,53 entre SDMT y LNS.

6.5. Modelos de progresión de forma masiva con Magerit

Una vez se han obtenido los mejores vectores con y sin la edad como covariable, se puede observar como los vectores con mayor rendimiento tienen un coeficiente de Pearson en torno al 0.50-0.55 tanto con la edad como sin ella.

En cuanto a los valores obtenidos de las características pertenecientes a los vectores en los time onset correspondientes, se pueden observar valores dentro de los rangos estimados, sin encontrarse diferencias significativas con y sin la edad como covariable.

Sin embargo, se pueden observar pequeñas mejoras en aquellos en los que la edad no está presente, obteniendo como mayor coeficiente de Pearson 0.55 en el vector:

$$SDMT\ HVLTRDLY\ SFT\ MEANCAU\ pTauA\beta$$

el cual no contiene la edad como covariable. Los resultados obtenidos de este vector son:

Vector	Time onset	t	R	Scores1	Scores2	Scores3
SDMT HVLTRDLY SFT MEANCAU pTauA β	6	1.98x-7.19	0.55	82.35 %	62.00 %	61.00 %

Tabla 6.2: Características del mejor vector, incluyendo el time onset, la recta de regresión (t), el coeficiente de Pearson (R) y puntuaciones.

Donde se obtiene un time onset de 6 años, una recta de regresión: $y=1.94x-7.11$ y un coeficiente de Pearson de 0.55.

En cuanto a la recta de regresión se refiere, se puede observar como esta no presenta una forma similar a la ideal: $y=x+\theta$ lo cual indica que, no existe una relación lineal en su totalidad entre el vector de características seleccionado y la variable respuesta.

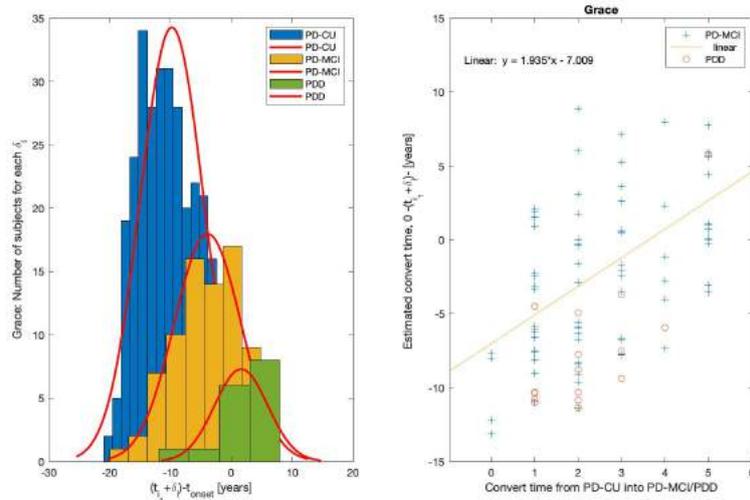


Figura 6.3: Distribución de los sujetos PD según el tiempo de conversión y recta de regresión del vector

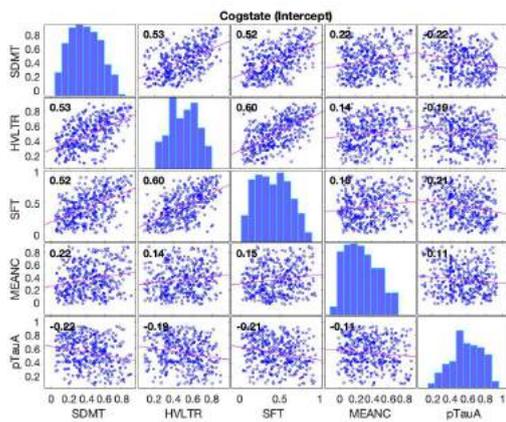


Figura 6.4: Matriz de correlación del intercepto según el criterio SITE con Age.

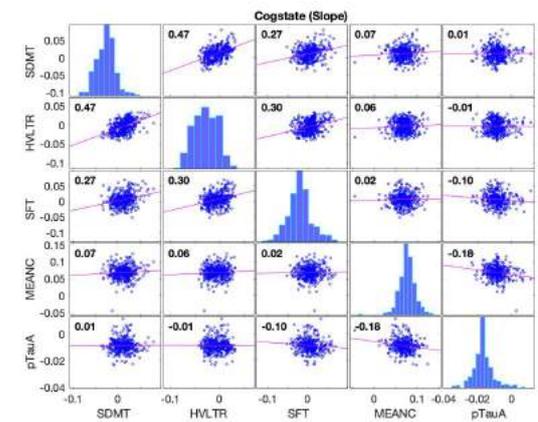


Figura 6.5: Matriz de correlación de la pendiente según el criterio SITE con Age.

Por otro lado, un coeficiente de Pearson de 0.55 en el modelo indica que existe una correlación positiva entre las variables, implicando que a medida que aumenta una la otra también lo hace. Sin embargo, aunque ambas aumentan o disminuyen en conjunto, al tener un valor de 0.55 su correlación es moderada.

Los valores de las características implicadas en el vector en el time onset (6 años) son los siguientes:

Vector	SDMT	HVLTRDLY	SFT	MEANCAU	$A\beta$
SDMT HVLTRDLY SFT MEANCAU pTauA β	34.97	7.24	42.56	1.36	727.53

Tabla 6.3: Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.

Vector	MOCA
SDMT HVLTRDLY SFT MEANCAU pTauA β	24.95

Tabla 6.4: Etapa del deterioro cognitivo según el vector de características seleccionado en el tiempo de onset.

Para poder determinar el estado cognitivo del paciente, se han establecido una serie de valores cut-off para diferentes test neuropsicológicos. Algunos de los implicados en el vector de características seleccionado son:

- HVLTR: una puntuación típica para personas sin deterioro cognitivo suele ser de 10 o más en las primeras pruebas de aprendizaje y de 5 o más en la prueba de retención. En varios estudios, un valor de corte para el diagnóstico de demencia podría ser una puntuación de 7 o menos en las primeras pruebas de aprendizaje y una puntuación de 2 o menos en la prueba de retención [38].
- SDMT: en esta prueba, se considera que una puntuación típica para personas sin deterioro cognitivo suele estar en el rango de 50 a 60 o más, dependiendo de la población. Para determinar que un paciente sufre MCI se considera una puntuación inferior a 45 o 40, mientras que un valor menor a 10 puede significar demencia [15].
- MoCA: en términos generales, se considera que una puntuación total de MoCA por debajo de 25 puntos puede indicar declive cognitivo del paciente, mientras que valores del test inferiores a 21 indican riesgo de demencia [19].

En el año 6 (time onset), se obtiene un valor de la prueba HVLTR de 7.24. Según los valores cut-off mencionados previamente, al obtener un valor de 7.24, se afirmaría que el paciente estaría en un estado de deterioro cognitivo. En cuanto al test MoCA se refiere, este cuenta con un valor de 24.95 lo que indica PD-MCI. Adicionalmente, se observan valores en torno a 35 en el test SDMT lo cual está en el rango especificado para pacientes con deterioro cognitivo (valores menores a 45).

Siguiendo con el análisis de las variables en los diferentes tests, se puede determinar el tiempo transcurrido entre el inicio del declive cognitivo ($\text{MoCA} < 25$) y demencia ($\text{MoCA} < 21$), obteniendo valores en torno a los 7 años en todos los vectores analizados con y sin la edad.

Concretamente, en el vector con mejores prestaciones se obtiene un tiempo de 7.2 años entre estos dos estados cognitivos.

6.5.1. Bootstrap

Finalmente, al obtener el análisis bootstrap para los perfiles A+ y A-, se puede observar de forma visual como las trayectorias tanto de los test neuropsicológicos como de los biomarcadores no varían en relación con la patología amiloide. Esto permite concluir que en la PD la presencia o no de esta proteína no influye al desarrollo

del deterioro cognitivo del paciente.

Sin embargo, esto no sucede de la misma manera en todas las enfermedades neurodegenerativas, como el AD.

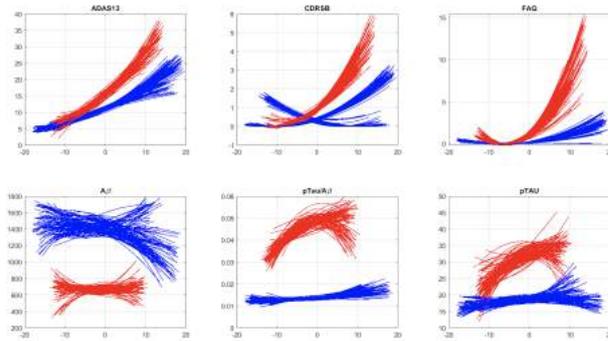


Figura 6.6: Trayectorias de test neuropsicológicos y biomarcadores tras análisis bootstrap para perfiles A+ y A- en AD

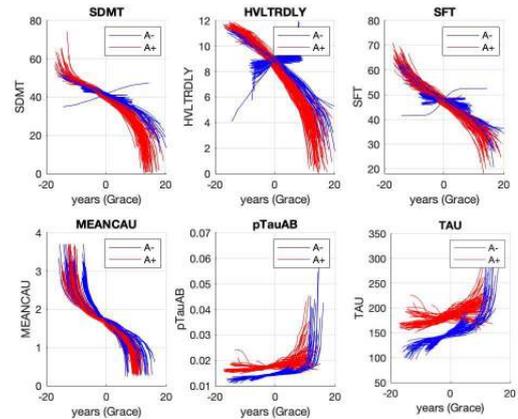


Figura 6.7: Trayectorias de test neuropsicológicos y biomarcadores tras análisis bootstrap para perfiles A+ y A- en PD

Como se puede observar, en la Figura 6.6 las trayectorias de los biomarcadores muestran grandes diferencias entre ambos perfiles a lo largo de su trayectoria, mientras que en los análisis de PD (Figura 6.7) estas son prácticamente iguales.

6.6. Conclusiones

Tras la obtención de los diferentes resultados y su posterior análisis, se presentan las siguientes conclusiones del estudio.

En referencia a los análisis realizados al inicio, se encontraron diferencias significativas entre los pacientes HC y PD en la mayoría de test neuropsicológicos, excepto LNS, INMEDIAT y ADL. Esto también se puede ver reflejado en biomarcadores como ABETA, ASYN o pTAU donde el pvalor es mayor que 0.001. Por otro lado, en el análisis de sujetos PD diferenciados por su estado cognitivo, se obtienen diferencias en la mayoría de los test y biomarcadores entre los grupos *PD-CU PD-MCI* y *PD-CU PDD*. En cuanto a las características clínicas se refiere, también se encuentran diferencias significativas en la edad. Esto permite deducir que para analizar el estado cognitivo de un paciente PD, sería interesante tener en cuenta los resultados de estos marcadores y test, pues su resultados se ven altamente relacionados con el estado cognitivo del sujeto.

En cuanto a los perfiles AT se refiere, se puede concluir que los sujetos A+ suponen un 35 % de los sujetos PD-CU , 41 % PD-MCI y 66.67 % en PDD al final del estudio, siendo el perfil dominante entre los pacientes sometidos a estudio.

Además, tras el análisis estadístico para obtener diferentes modelos predictivos a partir de diferentes métodos se obtiene como vector con mejor rendimiento: *SDMT*

HVLTRDLY SFT MEANCAU pTauA β con un coeficiente de Pearson de $R=0.55$ tras las diferentes combinaciones obtenidas con las diferentes características. Este vector tiene un tiempo de onset de 6 años. Si se analizan los valores de los tests neuropsicológicos y biomarcadores correspondientes en este tiempo y se comparan con los resultados clínicos y otros estudios y artículos publicados, se obtienen resultados similares dentro de los valores estipulados.

Aun así, los valores de correlación no son excesivamente altos. Esto se puede ver reflejado en la recta de regresión del modelo, que no se asemeja a la forma ideal $y=x+\theta$. Esto evidencia que el modelo desarrollado es mejorable, y da cabida a la implementación de nuevos algoritmos o tipos de análisis para intentar conseguir una correlación más cercana a 1.

Finalmente, en cuanto al análisis bootstrap se trata, se puede concluir que la patología amilode no afecta al deterioro cognitivo en los pacientes con PD, a comparación de otras enfermedades neurodegenerativas como AD.

6.7. Desarrollos futuros

Esta investigación se ha podido llevar a cabo gracias a los resultados de de los diferentes artículos mencionados anteriormente y el trabajo de Jaime Palomino [33]. Se ha continuado esta línea de investigación, desarrollando así nuevos modelos predictivos, incluyendo el análisis de un mayor número de variables.

Se abren varias líneas de investigación tras el desarrollo de este proyecto que permitirán un mayor desarrollo de enfoques más precisos y personalizados para predecir y comprender el declive cognitivo en pacientes con Parkinson. Algunas de estas son:

- **Aumento de datos longitudinales:** la recopilación de datos longitudinales en una cohorte más amplia permitiría comprender mejor la evolución del declive cognitivo a lo largo del tiempo en sujetos con PD. El seguimiento a largo plazo brindaría una visión más completa de los patrones del deterioro cognitivo y permitiría identificar marcadores tempranos de progresión.
- **Integración de nuevas herramientas:** el uso de datos de neuroimagen, como resonancia magnética estructural y funcional, permitiría una comprensión más profunda de los cambios cerebrales asociados con el declive cognitivo. Estos datos podrían utilizarse para mejorar la capacidad de predicción y para identificar patrones específicos de daño cerebral.
- **Integración de los datos:** el desarrollo de modelos con diferentes tipos de datos como información clínica, biomarcadores y datos neuroimagen (entre otros) puede llevar a obtener modelos más completos y precisos.
- **Desarrollo de nuevos algoritmos de aprendizaje automático:** como se puede ver a lo largo del estudio, los algoritmos desarrollados no alcanzan un valor de correlación igual a la unidad. El desarrollo de algoritmos o sistemas como redes neuronales profundas y modelos de atención, para capturar relaciones complejas en los datos pueden ser capaces de detectar patrones sutiles

y no lineales en los datos, lo que podría mejorar la capacidad de predicción y progresión del declive cognitivo.

Apéndice A

Análisis estadístico aplicado a las ciencias de la salud.

A.1. Introducción

El análisis estadístico aplicado a las ciencias de la salud es una herramienta fundamental para la investigación en el campo de la salud. Los métodos estadísticos en este área permiten el análisis y la síntesis de datos, además de la identificación de patrones. Adicionalmente, gracias a estas herramientas es posible evaluar la eficiencia de las intervenciones y obtener conclusiones basadas en la evidencia.

Este tipo de análisis es necesario en todas las etapas de la investigación en la salud, desde el diseño y la planificación del estudio hasta la recopilación, análisis y presentación de los datos obtenidos. Entre los métodos más utilizados se pueden encontrar el análisis de la varianza, la regresión lineal y logística, la prueba de hipótesis, el análisis de supervivencia o el análisis de correlación.

A.2. Conceptos básicos

A.2.1. Inferencia estadística

Se define la inferencia estadística como el conjunto de métodos que, a partir de la información de una muestra estadística, permiten inducir el comportamiento de una población con un riesgo de error medible en términos de probabilidad.

Existen dos métodos paramétricos por los cuales se puede desarrollar la inferencia estadística: métodos de estimación de parámetros y métodos de contraste de hipótesis. Ambos se basan principalmente en el conocimiento teórico de la distribución de probabilidad de una muestra de tipo estadística que se utiliza como estimador de un parámetro.

Métodos de estimación de parámetros

Este método consiste en hacer una asignación de un valor específico a los parámetros que caracterizan la distribución de probabilidad de la población. Al estimar un

parámetro de la población, se comete un error de estimación que se puede definir como la diferencia entre la estimación y el verdadero valor del parámetro. Para poder valorar la precisión asociada a este parámetro elegido se parte de dicha estimación y se construye un intervalo de confianza.

Un intervalo de confianza se construye a partir de valores numéricos, entre los que se encuentra el verdadero valor del parámetro entre la probabilidad del intervalo. Esta probabilidad se denomina grado de confianza del intervalo, y la amplitud de éste constituye una medida del grado de precisión con el que se estima el parámetro.

Métodos de contraste de hipótesis

Este segundo método tiene como objetivo principal comprobar si un determinado supuesto que hace referencia a un parámetro poblacional o a parámetros de varias poblaciones es compatible con la evidencia empírica que hay en la muestra. Para ello se hacen uso de parámetros denominados hipótesis paramétricas. Estas hacen referencia a los supuestos que se establecen respecto a los diferentes parámetros de la población. Para toda hipótesis paramétrica se requiere un criterio de decisión, en el que depende de factores como la naturaleza de la población, la ley de probabilidad del estimador de ese parámetro y el nivel de control que se quiere aplicar.

En todo contraste es necesario tener en cuenta dos tipos de hipótesis: la hipótesis nula y la hipótesis alternativa.

- **Hipótesis nula (H_0)**

Esta recoge el supuesto de que el parámetro encargado de describir las características del estudio a realizar tome cierto valor (\mathbf{X}_0) y soporta la carga de la prueba. Se pueden fijar diferentes valores de nivel de significación de contraste (α). Este nivel es el encargado de definir la probabilidad de cometer el error de rechazar la hipótesis nula siendo ésta cierta.

- **Hipótesis alternativa (H_1)** Esta hipótesis establece que el valor que describe las características del estudio (\mathbf{X}) es diferente al determinado por la hipótesis nula (H_0). Es decir, intenta demostrar la falsedad de la hipótesis nula realizada en un estudio. Existen dos tipos de hipótesis alternativas: si la hipótesis alternativa se formula simplemente como 'la hipótesis nula no es cierta', el contraste es bilateral; por el contrario cuando se indica el sentido de la diferencia, el contraste es unilateral.

A.2.2. Intervalos de confianza

Se define intervalo de confianza al rango de valores que se utilizan para estimar un parámetro desconocido en una población. Este se basa en una muestra aleatoria de la población y se calcula a partir de técnicas estadísticas. Para la construcción de este intervalo son necesarios los siguientes parámetros.

- **Tamaño muestral:** número de observaciones en la muestra. Cuanto mayor sea la muestra, más precisos serán los resultados y por consiguiente, menor será el intervalo.

- Desviación estándar: esta medida representa la variabilidad entre los datos de la población.
- Nivel de confianza: probabilidad de que entre el intervalo elegido se encuentre el verdadero valor del parámetro.
- Estimador de la media: esta medida es utilizada para obtener una estimación aproximada de el valor de la media del intervalo.

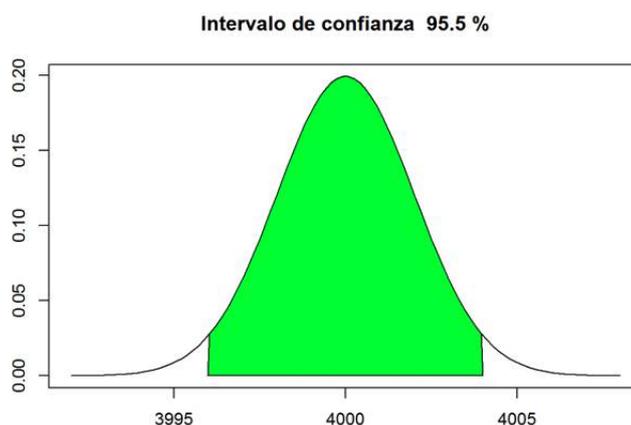


Figura A.1: Representación de un intervalo de confianza de 95.5 %

A.2.3. Estadísticos y p-valor

Los estadísticos son medidas numéricas que se utilizan para resumir y describir un conjunto de datos. Existen dos tipos:

- Medidas de tendencia central: se utilizan para describir la ubicación central de un conjunto de datos y se calculan utilizando diferentes métodos. Las medidas más comunes son la media aritmética, la mediana y la moda. La media aritmética se calcula sumando todos los valores en un conjunto de datos y dividiéndolos por el número total de valores. La mediana es el valor central en un conjunto de datos ordenado de menor a mayor, y la moda es el valor que ocurre con más frecuencia en un conjunto de datos.
- Medidas de dispersión: Las medidas de dispersión se utilizan para describir la variabilidad o la dispersión de los datos. Las medidas más comunes son el rango, la desviación estándar y la varianza. El rango es la diferencia entre el valor más alto y el más bajo en un conjunto de datos. La desviación estándar y la varianza miden la variabilidad de los datos con respecto a la media.

Los estadísticos también se utilizan en la inferencia estadística para realizar pruebas de hipótesis y para estimar parámetros de población desconocidos a partir de una muestra de datos. Estos incluyen el error estándar, el intervalo de confianza y el p-valor.

Al realizar un contraste de hipótesis se obtiene como resultado un estadístico. Este hace referencia a un parámetro empírico cuyo valor sigue una distribución de probabilidad determinada y proviene de datos de la muestra. A partir de esto se

puede definir el concepto de ***p-valor*** como la probabilidad de que el estadístico de la prueba tome un valor igual o superior al de la muestra suponiendo que la hipótesis nula (***H₀***) es cierta.

Esto supone que, si el *p-valor* es menor o igual que el nivel de significación establecido (α), se rechazará la hipótesis nula ***H₀***. Cuando el *p-valor* tome valores próximos a cero la hipótesis nula se rechazará para cualquier nivel de significación establecido.

A.3. Calidad del experimento

Existen indicadores encargados de evaluar el un procedimiento estadístico. Entre ellos se encuentran:

- **Sensibilidad (SEN)**: se define como la proporción de personas con la enfermedad que tienen un resultado positivo en la prueba diagnóstica. Este parámetro determina la capacidad que tiene la prueba de detectar de forma correcta los casos positivos. Su valor se obtiene a partir de:

$$SEN(\%) = \frac{VP^1}{VP + FN^2} \quad (\text{A.1})$$

- **Especificidad (SPE)**: esta característica hace referencia a la proporción de personas que tienen un resultado negativo en la prueba realizada y no padecen la enfermedad. Esto implica que la especificidad determina la capacidad de excluir de forma correcta a los casos negativos. Su valor se obtiene a partir de:

$$SPE(\%) = \frac{VN^3}{VN + FP^4} \quad (\text{A.2})$$

- **Exactitud (ACC)**: capacidad de una medida estadística para estar cerca del valor verdadero de la población que se está estudiando. Este indicador se puede medir a partir del error de medición o el error estándar.

$$ACC(\%) = \frac{VP + VN}{VP + VN + FP + FN} \quad (\text{A.3})$$

El error de medición se define como la diferencia entre el valor verdadero de la población y la medida estadística a partir de la muestra. Por otro lado, el error estándar hace referencia a la variabilidad de la muestra y se utiliza para estimar el error de medición.

- **Área bajo la curva (AUC)**: esta medida estadística se utiliza para cuantificar el número de datos que se encuentran dentro de un rango. Representa el área entre una curva y el eje horizontal de la gráfica. Se utiliza usualmente a la hora de analizar la distribución de un conjunto de datos. Para hallar su valor se puede usar métodos numéricos o fórmulas matemáticas.

Las curvas ROC (*Receiver Operating Characteristic*) son una herramienta utilizada para evaluar el rendimiento de un modelo de clasificación binaria. Las curvas ROC son una representación gráfica de la tasa de verdaderos positivos (SEN) en

función de la tasa de falsos positivos ($1 - \text{SPE}$) para diferentes umbrales de clasificación, situando a la especificidad en el eje horizontal y a la sensibilidad en el vertical.

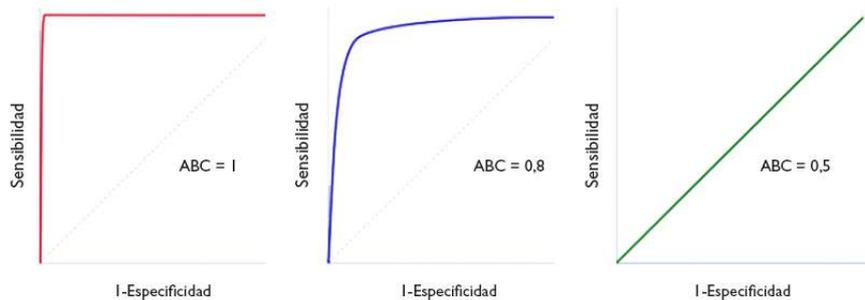


Figura A.2: Curvas ROC de calidades perfecta, buena y mala respectivamente

A.4. Análisis de la varianza

El Análisis de la Varianza, también conocido como ANOVA (ANálisis Of VARiance) es una herramienta utilizada para el estudio de efecto de uno o varios factores sobre una variable continua. Esta técnica se emplea cuando se pretende comparar las medias de dos o más grupos.

Este tipo de análisis es utilizado en el diseño de experimentos y comparación de hipótesis. Estos experimentos buscan examinar cómo una variable alterada influye en la variabilidad de una respuesta. Sin embargo, el análisis puede estar limitado por varios factores, como la naturaleza del experimento o las condiciones de diseño establecidas. Algunos de los términos destacados a la hora de realizar este análisis son:

- Unidad experimental: objeto o magnitud responsable de tomar las medidas
- Factor: variable que puede ser modificada por el autor del experimento (no depende de otras)
- Nivel: variable que indica la intensidad de un factor
- Tratamiento: combinación de los mismos niveles de un factor.
- Respuesta: variable del experimento medida.

El análisis ANOVA se basa en una distribución de Fischer, por la cual se evalúa la correlación dentro de grupos y la variabilidad que existe entre estos. Es indispensable que entre cada población puesta a estudio, se obtengan una misma distribución normal con un mismo valor de la varianza (σ^2).

ANOVA se puede representar gráficamente mediante un diagrama de cajas y bigotes, también conocido como *boxplot*. Este tipo de gráfico muestra la distribución de los datos para cada grupo y facilita la comparación entre ellos. En un boxplot se diferencian las siguientes partes:

- Caja: representa el rango intercuartílico (IQR), que contiene el 50% de los datos
- Línea central: representa la mediana
- Extremos de la caja: representan el primer y tercer cuartil

Los bigotes se extienden desde la caja hasta el valor más extremo que no se considera un valor atípico (outlier). Los valores atípicos se representan como puntos fuera de los bigotes.

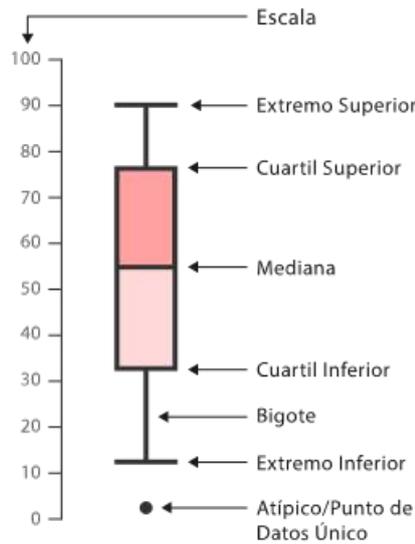


Figura A.3: Visión general del método propuesto.

Comparación de un único factor

En este análisis se pretende estudiar un único factor y variable independiente. Se requiere que la variable dependiente esté distribuida de manera normal en cada grupo y que la variabilidad dentro de los grupos sea similar.

En la comparación de un único factor ANOVA, se utiliza la prueba F para determinar si hay una diferencia significativa en los promedios entre los grupos. La hipótesis nula (H_0) es que no hay diferencias entre los grupos y la hipótesis alternativa (H_a) es que hay al menos una diferencia significativa.

Para realizar la comparación, se calcula la suma de cuadrados entre grupos (SSG) y la suma de cuadrados dentro de los grupos (SSE). La SSG mide la variación entre los promedios de los grupos, mientras que la SSE mide la variación dentro de los grupos.

Se calcula el valor de la prueba F mediante la división de la varianza entre grupos (MSG) por la varianza dentro de los grupos (MSE):

$$F = MSG/MSE \tag{A.4}$$

Si el valor de la prueba F es mayor que el valor crítico, se rechaza la hipótesis nula y se concluye que hay al menos una diferencia significativa entre los grupos. En este caso, se puede realizar un análisis post-hoc para determinar qué grupos son significativamente diferentes entre sí. Si el valor de la prueba F es menor que el valor crítico, se acepta la hipótesis nula y se concluye que no hay evidencia suficiente para afirmar que hay diferencias significativas entre los grupos.

Comparación de múltiples medidas

Tras la realización del análisis ANOVA se procede al procesamiento de los datos obtenidos. En el caso de que se confirmen diferencias significativas entre las medidas se procederá al estudio de estas a partir de técnicas de contrastes para comparaciones múltiples.

El objetivo principal es saber si dos grupos difieren de forma significativa entre sí. Una vez esto se cumple, mediante este análisis se pretende conocer los tratamientos que producen mayores diferencias. Para ello se establece el modelo de efectos fijos:

$$y_{ij} = \mu_i + u_{ij} \quad (\text{A.5})$$

Para el estudio cualitativo del comportamiento se efectúa una comparación de forma gráfica de las medidas:

- **Comparaciones gráficas de medidas** Se considera un modelo equilibrado con n observaciones por nivel $(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_I)$ las medias muestrales de los tratamientos. Con la hipótesis del modelo establecido cualquier promedio \bar{y}_i del tratamiento se distribuirá a partir de su media μ_i con una desviación estándar $\frac{\sigma}{\sqrt{2}}$.
Se pueden distinguir los siguientes casos:

1. Efectos de los tratamientos son iguales entre sí: estos son una muestra aleatoria de la misma distribución.
2. Efectos de los tratamientos no son iguales: no tienen una muestra aleatoria con la misma distribución.

- **Comparaciones basadas en la distribución t** Para las comparaciones múltiples las pruebas estadísticas más utilizadas se basan en la distribución t de Student. Esto permite la comparación por parejas de los efectos de I tratamientos. Esto supone que:

$$H_0 : \mu_i = \mu_j \quad (\text{A.6})$$

$$H_1 : \mu_i \neq \mu_j \quad (\text{A.7})$$

La técnica más conocida para el estudio de estas comparaciones múltiples es el procedimiento LSD (Least Significant Difference).

Contraste de la Mínima Diferencia Significativa

Este procedimiento fue sugerido por Fisher en 1935. Dicho procedimiento consiste en una prueba de hipótesis por parejas basada en la distribución t.

Método de Bonferroni

En este procedimiento se fija un nivel de significación (α) que se reparte entre cada una de las comparaciones consideradas y se utiliza la desigualdad de Bonferroni.

$$Pr\left(\bigcup_{m=1}^M A_m\right) \leq \sum_{m=1}^M Pr(A_m) \quad (\text{A.8})$$

Se establece la estimación por intervalos para las $M = \frac{I}{2}$ comparaciones posibles, cada una al nivel de significación $\alpha^* = \frac{\alpha}{M}$. Esto supone la creación de M intervalos de confianza que contienen a cada una de las posibles diferencias $\mu_i - \mu_j$ con una probabilidad de $1 - \alpha^*$. Si C_m se establece como el intervalo m -enésimo se obtiene:

$$Pr[\mu_{1m} - \mu_{2m} \in C_m = 1 - \alpha^*] \quad (\text{A.9})$$

$$m = 1, 2, \dots, M \quad (\text{A.10})$$

donde μ_{1m} y μ_{2m} hacen referencia a la primera y segunda media de la correspondiente comparación, aquí se supone que $1 \leq 1m < 2m \leq I$.

Si se aplica la desigualdad de Bonferroni se obtiene lo siguiente:

$$Pr\left(\bigcap_{m=1}^M C_m\right) = 1 - Pr\left(\bigcup_{m=1}^M \overline{C_m}\right) \geq 1 - \sum_{m=1}^M Pr(\overline{C_m}) = 1 - \sum_{m=1}^M \alpha^* \quad (\text{A.11})$$

Si se quiere garantizar un nivel de significación α para el conjunto de las M comparaciones por parejas, o un nivel de confianza $1 - \alpha$ basta con tomar:

$$\alpha^* = \frac{\alpha}{M} \quad (\text{A.12})$$

Multicompare en Matlab

La función $c = \text{multicompare}(stats)$ en Matlab devuelve una matriz c de los resultados de comparación por pares de una prueba de comparación múltiple utilizando la información contenida en la estructura de estadísticas. Además, realiza un gráfico donde se muestran las estimaciones y los intervalos de comparación.

La media de cada grupo está representada por un símbolo y el intervalo está representado por una línea que se extiende desde el símbolo. Las medias de dos grupos son significativamente diferentes si sus intervalos son disjuntos; no son significativamente diferentes si sus intervalos se superponen. Permite la interacción con el ratón para seleccionar cualquier grupo y resaltar todos los grupos que son significativamente diferentes, en caso de haberlos.

Tablas de contingencia

Estas son una herramienta fundamental para permitir el análisis y estudio de la influencia de una variable de tipo categórica respecto a otras. Para permitir este tipo de comparación se forman estas tablas, formadas a partir de filas horizontales que recogen la información de una variable y verticales que recogen los datos de otra.

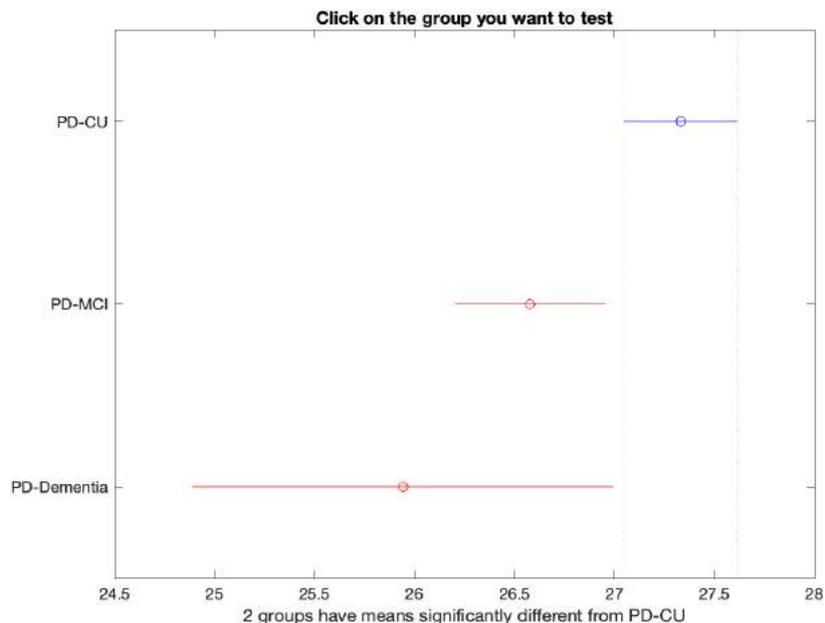


Figura A.4: Gráfico interactivo donde se muestran los intervalos de comparación.

Las celdas que componen estas filas y columnas son las encargadas de proporcionar la información necesaria para permitir la comparación de las variables a estudiar. A partir de los datos obtenidos gracias a la tabla se permite un fácil estudio de las

DATA	PD-CU	PD-MCI	PDD	TOTAL
FEMALE	120	20	6	146
MALE	203	63	11	277
TOTAL	323	83	17	423

Figura A.5: Tabla de contingencia para el estudio del sexo de los pacientes sometidos a estudio.

diferentes variables. En Matlab el estudio de las variables categóricas se realiza a partir de la función *crosstab_analysis.m* la cual permite obtener de una forma rápida los datos de la tabla correspondiente. Una vez los datos de la tabla están especificados es posible hacer la multicomparación de los datos obtenidos.

A.5. Análisis de supervivencia

El análisis de supervivencia es una técnica estadística utilizada para analizar el tiempo que transcurre hasta que ocurre un evento. Es comúnmente utilizado en estudios de ciencias de la salud, ingeniería, economía cuando se requiere analizar la duración de un evento.

Este tipo de análisis se enfoca en el tiempo de supervivencia⁵. Además, considera aquellos individuos que no han experimentado el evento de interés al momento del análisis, los cuales se consideran censurados.

⁵Tiempo hasta que se produce el evento de interés.

Entre las técnicas más comunes utilizadas en el análisis de supervivencia se encuentran el modelo de Kaplan-Meier para la estimación de la función de supervivencia y la regresión de Cox para analizar la influencia de variables predictivas sobre el tiempo de supervivencia.

A.5.1. Método de Cox

El método de Cox o método de riesgos proporcionales de Cox es un modelo estadístico que se encarga de analizar la relación que existe entre la supervivencia y uno o varios predictores.

Para ello, establece una tasa de riesgo en un tiempo determinado (probabilidad que ocurra el evento a estudiar) y asume que esta es proporcional a la tasa de riesgo para los demás individuos en ese instante de tiempo. Esto tiene el nombre de hipótesis de riesgos proporcionales.

Este modelo es utilizado en diferentes estudios o ensayos clínicos, con el objetivo de evaluar el efecto de una intervención en el tiempo hasta que ocurre el evento de interés. Para obtener un modelo adecuado es necesario ajustar el efecto de otras variables (covariables) que pueden estar relacionadas con el evento a estudio. La edad, variables demográficas o diferentes biomarcadores son algunas de las covariables que se pueden tener en cuenta en función del evento [44].

Para ello, se recoge esta información en el vector de datos $z' = (z_1, z_2, \dots, z_n)$. Si se suponen las covariables independientes del tiempo se puede definir el modelo de regresión de Cox como:

$$\lambda(t/z) = \exp\{\beta'z\}\lambda_0(t) = \exp\{\beta_1z_1 + \beta_2z_2 + \beta_3z_3 + \dots + \beta_nz_n\}\lambda_0(t) \quad (\text{A.13})$$

Donde las covariantes tienen un efecto multiplicativo, β son los parámetros de regresión y λ la función de riesgo basal. Por otro lado, la función $\exp\{\beta'z\}$ indica el riesgo relativo de un individuo con perfil z con respecto a otro con $z = 0$ [11]. En el modelo original, se asume una supervivencia lineal en el efecto de las covariantes. En caso de plantearse la no linealidad, se ha demostrado que la hipótesis planteada conduce a una estimación sesgada, reduciendo así la potencia de los test estadísticos.

A.5.2. Modelo de regresión de Cox extendido

Cuando se incluyen variables predictivas dependientes del tiempo a la hora de evaluar la hipótesis de riesgos proporcionales, el modelo se extiende permitiendo así la interacción de las covariantes con las funciones del tiempo [28]. Si se analizan las variables de forma individual se obtiene:

$$h(t, z) = h_0(t)\exp(\beta z + \delta(zg(t))) \quad (\text{A.14})$$

Donde $zg(t)$ hace referencia a la interacción de las covariantes (z) con funciones del tiempo $g(t)$.

Si se analizan simultáneamente varios predictores se obtiene la siguiente expresión:

$$h(t, z) = h_0(t) \exp \left(\sum_{j=1}^p \beta_j z_j + \sum_{j=1}^p \beta_j z_j \right) \quad (\text{A.15})$$

A.5.3. Clasificación logística

Se define la clasificación logística al método estadístico que se utiliza para modelar la probabilidad de una variable categórica binaria (como puede ser éxito o fracaso) según una o varias variables predictoras seleccionadas. Para este modelo se utiliza normalmente el análisis de datos de clasificación [20].

Este método tiene como fundamento principal una función logística, la cual devuelve un 1 si el dominio tiende a infinito positivo o 0 si tiende a infinito negativo. Esta función tiene la forma:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (\text{A.16})$$

A partir de esta función, el modelo permite predecir la probabilidad de que ocurra el evento a estudio a partir de las variables predictoras [42]. Estas pueden ser de dos tipos: continuas o categóricas. El modelo permite el uso de ambas y la interacción entre ellas.

Además, para el ajuste del modelo se utiliza la estimación de los coeficientes de las variables predictoras en la cual se hace uso de la función de verosimilitud, la cual permite encontrar los coeficientes de las variables predictoras que maximizan la probabilidad de obtener los datos observados.

Cuando se consigue un ajuste óptimo del modelo, este se puede utilizar para predecir la probabilidad de que ocurra un evento según el valor de las variables predictoras. Si la probabilidad es mayor que la del umbral establecido, se clasifica como éxito y, si es inferior al umbral, como fracaso.

A.5.4. Variables normalizadas

Las variables normalizadas o z-scores, se definen como una medida de la desviación de un valor particular de una distribución de probabilidad en relación a la media y la desviación estándar de esa distribución. Estas variables son el resultado de la resta entre la media y la distribución entre la desviación estándar:

$$z = \frac{x - \mu}{\sigma} \quad (\text{A.17})$$

Donde x es el valor individual de la variable, μ es la media de la distribución y σ hace referencia a la desviación estándar de la distribución.

La distribución normal z o distribución normal estándar, se define como una distribución de probabilidad normal que se caracteriza por tener una media de valor 0 y una desviación estándar de 1, denotándose como $N(0,1)$. Las variables normalizadas

se distribuyen siguiendo este tipo de distribución.

El uso de los z-scores permiten una fácil análisis y comparación con otros datos medidos en diferentes escalas. Además, permiten estimar la probabilidad de que un valor determinado se produzca en una distribución normal y pueden evaluar la relación que existen entre dos variables en un conjunto amplio de datos.

Apéndice B

Uso de Magerit

B.1. Introducción

Magerit-3 es un sistema de cómputo científico destinado a la ejecución de cargas de trabajo científico que requieren cómputo intensivo. El servicio ofrece la posibilidad de realizar tareas informáticas intensivas en cuanto a capacidad de cómputo, como análisis o simulaciones, a través de lotes de trabajo.

El acceso al servicio se organiza mediante actividades en las que participan miembros de un mismo proyecto. Si hay varios proyectos que lo requieren, es posible tener varias actividades activas simultáneamente.

B.2. Acceso

Para poder acceder a los servicios que proporciona el sistema, es necesario enviar una solicitud para obtener el usuario y clave. Una vez se tienen, se puede utilizar Magerit-3 de forma remota, a partir de SSH, accediendo a la dirección IP:

magerit.cesvisma.upm.es

B.3. Ejecución de tareas

Magerit-3 se utiliza a través de trabajos por lotes utilizando SLURM¹ como gestor y planificador de recursos. Para ejecutar un trabajo, se especifican las características necesarias y el sistema se encarga de reservar los recursos requeridos y llevar a cabo las tareas.

El objetivo principal de SLURM es optimizar la utilización de los recursos disponibles y asegurar la equidad en la asignación de recursos entre los usuarios. El sistema de colas de SLURM permite priorizar y ordenar los trabajos en función de la prioridad o límite de recursos (entre otros).

¹Simple Linux Utility for Resource Management

B.4. Mandatos SLURM

Para controlar los trabajos en ejecución a partir de SLURM, se utilizan diferentes mandatos, entre los que se encuentran:

- **sbatch**: envía un trabajo al sistema devolviendo su identificador.
- **squeue**: muestra la lista los trabajos encolados en el sistema (aquellos que están ejecutando o esperando para ejecutar).
- **scancel**: cancela un trabajo encolado. Si el trabajo ya estaba ejecutando se abortará la ejecución en ese momento.
- **sstat** : Muestra estadísticas de uso de un trabajo que esté ejecutando en ese momento.
- **sacct**: Muestra información de la ejecución de un trabajo ya finalizado.

B.5. Definición de trabajos

Cuando se envía un trabajo a SLURM, se especifican diversos parámetros, como los recursos necesarios (como número de CPU, memoria, tiempo de ejecución, etc.), los archivos de entrada y salida, las dependencias, las variables de entorno y otros detalles relevantes. Estos parámetros permiten al sistema de gestión de recursos de SLURM asignar y programar los recursos necesarios para la ejecución del trabajo de manera eficiente.

Para que el gestor detecte de forma correcta los parámetros introducidos, es necesario que estos empiecen por la palabra *#SBATCH*. Los utilizados en este proyecto son:

- **-job-name**: nombre del trabajo. Es meramente informativo y sirve para facilitar la identificación.
- **-partition**: partición a utilizar. Se puede diferenciar entre:
 - Standar: 600 cores en 160 horas máximo.
 - Debug: 40 cores en 1 hora máxima.
- **-nodes**: número de nodos a utilizar.
- **-ntasks**: número de tareas a ejecutar (paralelismo distribuido).
- **-ntasks-per-node**: número de tareas asignadas a cada nodo físico.
- **-time**: duración máxima de ejecución del trabajo.
- **-mem-per-cpu**: RAM solicitada por CPU en MB.
- **-mem**: RAM solicitada por nodo en MB.
- **-output**: fichero para almacenar la salida estandar del job.
- **-error**: fichero para almacenar la salida de errores del job.
- **-mail-user**: dirección de correo para el envío de notificaciones de eventos.

- **-mail-type**: enviar correos electrónicos al usuario en todos los eventos de trabajo.
- **-chdir**: directorio de trabajo.

Ejemplo de carga de parámetros en la ejecución de un trabajo:

```
#!/bin/bash
##----- Start job description -----
#SBATCH --partition=standard
#SBATCH --job-name=my_job
#SBATCH --ntasks=1
#SBATCH --mem-per-cpu=1G
#SBATCH --time=12:00:00
#SBATCH --mail-type=ALL
#SBATCH --mail-user=user@example.com
##----- End job description -----
```

B.6. Directorio SCRATCH

Este directorio está destinado a ser utilizado por los usuarios para realizar operaciones de entrada y salida (E/S) rápidas durante la ejecución de sus trabajos.

Ejemplo de directorio scratch en la ejecución de un trabajo:

```
# Definición del directorio SCRATCH
mkdir -p SCRATCH/SLURM_JOB_ID
{execution script}
# Cleanup local work directory
rm -rf SCRATCH/SLURM_JOB_ID
```

B.7. Redirección de salidas y notificaciones

Las aplicaciones suelen escribir datos o resultados por la consola. Al procesarse mediante un sistema de proceso por lotes, estas salidas se perderían al no existir una consola en el momento de ejecución. Para evitarlo SLURM dirige esas salidas a ficheros en disco dónde se pueden consultar.

Además, se puede configurar el envío automático de un correo cuando se produzcan modificaciones en el estado de un trabajo. Para ello hay que añadir las opciones `-mail-user` indicando el correo al que se deben enviar las notificaciones y `-mail-type` con los eventos de interés separados por comas:

- BEGIN
- END
- FAIL

- REQUEUE
- STAGE_OUT

Apéndice C

Diccionario de siglas y acrónimos

ACC Exactitud

AD Alzheimer's Disease

ADAS- cog Escala de Evaluación de la Enfermedad de Alzheimer- Subescala Cognitiva (Alzheimer's Disease Assessment Scale - Cognitive Subscale)

ADN Ácido Desoxirribonucleico

APOE apolipoproteína E

ANOVA Análisis de la Varianza (Analysis Of VAriance)

CSF Líquido Cerebroespinal (Cerebrospinal Fluid)

CDR Evaluación Clínica de Demencia (Clinical Dementia Rating)

CT Tomografías Computarizadas (Computed Tomography)

DNI Documento Nacional de Identidad

DAT Transportador de dopamina (Dopamine Transporter)

FDG-PET Tomografía por Emisión de Positrones con Fluorodesoxiglucosa (Fluorodeoxyglucose-Positron Emission Tomography)

fMRI Resonancia Magnética Funcional (Functional Magnetic Resonance Imaging)

GRACE GRowth models via Alternating Conditional Expectation

HC Sujeto sano control (Healthy Control)

HVLT-R Prueba de Aprendizaje Verbal de Hopkins (Hopkins Verbal Learning Test-Revised)

JLO Evaluación de la Orientación de Líneas (Judgment of Line Orientation)

LNS Secuenciación Letra-Número (Letter-Number Sequencing)

MDS Sociedad de Trastornos del Movimiento

MDRS-2 Escala de Disfunción de Rivermead-2 (Mattis Dementia Rating Scale Second Edition)

- MEANCAU** Absorción media en caudado
- MMSE** Mini-Prueba de Estado Mental
- MoCA** Evaluación Cognitiva de Montreal (Montreal Cognitive Assessment)
- MSG** Message (Mensaje)
- NaN** Dato no numérico (Not a Number)
- NIA-AA** Instituto Nacional sobre el Envejecimiento y la Asociación de Alzheimer (National Institute on Aging and Alzheimer's Association)
- NM** Medidas neuropsicológicas (Neuropsychological Measurements)
- OMS** Organización Mundial de la Salud
- PD** Parkinson's Disease
- PD-CFRS** Escala de Clasificación Funcional Cognitiva de la Enfermedad de Parkinson (Parkinson's Disease – Cognitive Functional Rating Scale)
- PD-CRS** Escala de Evaluación Cognitiva de la Enfermedad de Parkinson (Parkinson's Disease – Cognitive Rating Scale)
- PD-CU** Diagnóstico de Parkinson sin deterioro cognitivo
- PDA+** Diagnóstico de Parkinson con depósitos de $A\beta$
- PDA-** Diagnóstico de Parkinson sin depósitos de $A\beta$
- PD-MCI** Diagnóstico de Parkinson con deterioro cognitivo leve
- PET** Tomografía por Emisión de Positrones (Positron Emission Tomography)
- PPMI** Parkinson's Progression Markers Initiative
- RAVLT** Test de Aprendizaje Verbal Auditivo de Rey (Rey Auditory Verbal Learning Test)
- RCT** Ensayo clínico aleatorizado (Randomized Controlled Trial)
- REM** Movimiento Rápido del Ojo (Rapid Eye Movements)
- SCD** Declive Cognitivo Subjetivo (Subjective Cognitive Decline)
- SEN** Sensibilidad
- SPECT** Tomografía por Emisión de Fotón Único (Single Photon Emission Computed Tomography)
- SPE** Especialidad
- UPSIT** Prueba de Identificación de Olores de la Universidad de Pensilvania (University of Pennsylvania Smell Identification Test)
- fMRI** Resonancia Magnética funcional (Functional Magnetic Resonance Imaging)

Bibliografía

- [1] The latex project.
- [2] Magerit: High performance computing services.
- [3] PPMI - Parkinson's Progression Markers Initiative, 2022.
- [4] Dag Aarsland, Lucia Batzu, Glenda M Halliday, Gert J Geurtsen, Clive Ballard, K Ray Chaudhuri, and Daniel Weintraub. Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers*, 7(1):47, 2021.
- [5] Daniela Berg and Ronald B. Postuma. Parkinson's disease: etiology, pathogenesis and risk factors. *Nature Reviews Neurology*, 10(8):415–424, 2014.
- [6] K Berganzo, B Tijero, A González-Eizaguirre, J Somme, E Lezcano, I Gabilondo, M Fernandez, JJ Zarranz, and JC Gómez-Esteban. Síntomas no motores y motores en la enfermedad de parkinson y su relación con la calidad de vida y los distintos subgrupos clínicos. *Neurología*, 31(9):585–591, 2016.
- [7] L Calvo, M Casals-Coll, G Sánchez-Benavides, M Quintana, RM Manero, T Rognoni, R Palomo, F Aranciva, F Tamayo, and J Peña-Casanova. Estudios normativos españoles en población adulta joven (proyecto neuronorma jóvenes): normas para las pruebas visual object and space perception battery y judgment of line orientation. *Neurología*, 28(3):153–159, 2013.
- [8] J Campdelacreu. Parkinson's disease and alzheimer disease: environmental risk factors. *Neurología (English Edition)*, 29(9):541–549, 2014.
- [9] Paul Chrisp, George J Mammen, and Eugene M Sorkin. Selegiline: a review of its pharmacology, symptomatic benefits and protective potential in parkinson's disease. *Drugs & Aging*, 1:228–248, 1991.
- [10] David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.
- [11] Marta Cuadras and Laia Palau. El modelo de cox de riesgos proporcionales: descripciones. *Medicina Clínica*, 158(4):166–172, 2022.
- [12] JC Dalrymple-Alford, MR MacAskill, CT Nakas, L Livingston, C Graham, GP Crucian, TR Melzer, J Kirwan, R Keenan, S Wells, et al. The moca: well-suited screen for cognitive impairment in parkinson disease. *Neurology*, 75(19):1717–1725, 2010.
- [13] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Raman, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr,

- Michael W Weiner, Jean-François Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410, 2014.
- [14] E. Ray Dorsey, Todd Sherer, Michael S Okun, and Bastiaan R Bloem. The economic burden of parkinson's disease in the united states: A systematic review. *Movement Disorders*, 35(8):1392–1404, 2020.
- [15] DA García Estévez. Aproximación al estudio del deterioro cognitivo en la esclerosis múltiple recurrente remitente. un estudio en práctica clínica habitual empleando el symbol digit modalities test y la resonancia magnética convencional. *Neurology Perspectives*, 1(1):27–32, 2021.
- [16] FileZilla Project. Filezilla - the free ftp solution, 2023. Sitio web oficial de FileZilla Project.
- [17] Eleonora Fiorenzato, Roberta Biundo, Diego Cecchin, Anna Chiara Frigo, Jinhee Kim, Luca Weis, Antonio P Strafella, and Angelo Antonini. Brain amyloid contribution to cognitive dysfunction in early-stage parkinson's disease: the ppmi dataset. *Journal of Alzheimer's Disease*, 66(1):229–237, 2018.
- [18] Laura Fratiglioni and Hui-Xin Wang. Smoking and parkinson's and alzheimer's disease: review of the epidemiological studies. *Behavioural brain research*, 113(1-2):117–120, 2000.
- [19] M Gaete, S Jorquera, S Bello-Lepe, YM Mendoza, M Vz, MF Alonso-Sanchez, and J Lira. Resultados estandarizados del montreal cognitive assessment (moca) para cribado neurocognitivo en población chilena. *Neurolog*, 2020.
- [20] Ángel García-Hernández, José Ramón González-Cristóbal, and Lourdes Ureña-López. Detección de noticias falsas: aproximación mediante aprendizaje automático. *Procesamiento del Lenguaje Natural*, 61:57–64, 2018.
- [21] Jennifer G Goldman, Samantha Holden, Bichun Ouyang, Bryan Bernard, Christopher G Goetz, and Glenn T Stebbins. Diagnosing pd-mci by mds task force criteria: How many and which neuropsychological tests? *Movement Disorders*, 30(3):402–406, 2015.
- [22] Gabriel Henrique Hawthorne, Marcelo Picinin Bernuci, Mariza Bortolanza, Victor Tumas, Ana Carolina Issy, and Elaine Del-Bel. Nanomedicine to overcome current parkinson's treatment liabilities: A systematic review. *Neurotoxicity research*, 30:715–729, 2016.
- [23] Hostinger. Usar tmux cheat sheet - tutorial de hostinger, 2023. Tutorial en línea de Hostinger sobre el uso de tmux.
- [24] Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

- [25] Clifford R Jack Jr, David A Bennett, Kaj Blennow, Maria C Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M Holtzman, William Jagust, Frank Jessen, Jason Karlawish, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562, 2018.
- [26] Clifford R Jack Jr, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207, 2013.
- [27] Freya Kamel and Samuel M. Goldman. Occupational exposure to pesticides, and Parkinson's disease: a systematic review and meta-analysis of cohort studies. *Occupational and Environmental Medicine*, 71(7):492–500, 2014.
- [28] DY Lin. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247, 1994.
- [29] MathWorks. Matlab - mathworks, 2023. Plataforma en la de MATLAB.
- [30] AE Mehta and G Tolis. Pharmacology of bromocriptine in health and disease. *Drugs*, 17:313–325, 1979.
- [31] Xiao Meng, Jia-Hao Deng, Xue-Feng Wang, Yu-Jun Liu, Xiao-Qi Huang, Yu-Hang Li, Li Chen, Si-Yan Chen, Yong-Ming Guo, and An-Mu Xie. Mortality in Parkinson's disease: a systematic review and meta-analysis. *Neuroepidemiology*, 54(3):193–202, 2020.
- [32] World Health Organization. Parkinson's disease increasing in many countries, six million people worldwide affected. <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>, 2018. Accessed on April 22, 2023.
- [33] Jaime Palomino Vaquero. Conexión entre el declive cognitivo en sujetos con la enfermedad de Parkinson novo y los constructos biológicos de la enfermedad de Alzheimer, 2022.
- [34] Carlos Platero and M Carmen Tobar. Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease. *Journal of neuroscience methods*, 341:108698, 2020.
- [35] PsicoRevista. Sdmt. test de símbolos y dígitos, 2018.
- [36] PuTTY. Putty: a free ssh and telnet client, 2023. Sitio web oficial de PuTTY.
- [37] R Development Core Team. Comprehensive R archive network (cran), 2023. Sitio web oficial del Comprehensive R Archive Network.
- [38] Diego Rivera, N Cadavid-Ruiz, NC Gutiez-Hernez, JA Calderhangualarlos Jos los Reyes Aragnnd JC Arango-Lasprilla. Test de aprendizaje verbal de Hopkins-Revisado (HVLT-R). datos normativos basados en regresiones múltiples para población colombiana. *Rev. Iberoam. Neuropsicol*, 2:69–81, 2019.
- [39] V. Rozani, N. Giladi, and T. Gurevich. Traumatic brain injury and risk of Parkinson disease. *Neurology*, 90(10):e905–e912, 2018.

- [40] O. Sáez Atxukarro, R. del Pino, J. Peña, D.J. Schretlen, N. Ibarretxe Bilbao, and N. Ojeda. Test de aprendizaje verbal de hopkins revisado: normalización y estandarización de la prueba en población española. *Neurología*, 2020.
- [41] Anette Schrag, Uzma Faisal Siddiqui, Zacharias Anastasiou, Daniel Weintraub, and Jonathan M Schott. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed parkinson’s disease: a cohort study. *The Lancet Neurology*, 16(1):66–75, 2017.
- [42] Barbara G Tabachnick and Linda S Fidell. *Using Multivariate Statistics*. Pearson, 7th edition, 2019.
- [43] F Tamayo, M Casals-Coll, G Sánchez-Benavides, M Quintana, RM Manero, T Rognoni, L Calvo, R Palomo, F Aranciva, and J Peña-Casanova. Estudios normativos españoles en población adulta joven (proyecto neuronorma jóvenes): normas para las pruebas span verbal, span visuoespacial, letter-number sequencing, trail making test y symbol digit modalities test. *Neurología*, 27(6):319–329, 2012.
- [44] Paola Velasco lvarez. Aplicación de técnicas de aprendizaje automático para la predicción del parkinson. Trabajo de fin de grado, Universidad de Sevilla, 2020.
- [45] Kathryn A Wyman-Chick, Phillip K Martin, Daniel Weintraub, Scott A Sperling, Lauren O Erickson, Carol A Manning, and Matthew J Barrett. Selection of normative group affects rates of mild cognitive impairment in parkinson’s disease. *Movement Disorders*, 33(5):839–843, 2018.
- [46] Xming. Xming: the x window system server for windows, 2023. Sitio web oficial de Xming.
- [47] Hui Yu, Dongdong Wang, Xiangjun Meng, Huifang Yang, Junwei Hao, Xinglong Yang, and Lei Wang. Risk factors for parkinson’s disease: a meta-analysis. *J Neural Transm (Vienna)*, 125(3):443–454, 2018.