

Universidad de las Américas Puebla

Escuela de Ingeniería

Departamento de Computación, Electrónica y Mecatrónica.



Sometido al departamento de Computación, Electrónica y Mecatrónica en cumplimiento parcial de los requerimientos para la obtención del grado de Doctor en Sistemas Inteligentes.

“Detección Automática de la enfermedad de Parkinson utilizando Análisis de Voz y análisis basado en vóxeles de Resonancia Magnética Estructural”.

Gabriel Solana Lavallo

Director de tesis:

Dr. Roberto Rosas Romero

mayo 2023

Abstract

Parkinson's Disease (PD) is a chronic neurodegenerative disorder that affects over six million people worldwide. It is considered one of the most challenging neurological diseases to diagnose, resulting in wrong medication and treatments that affect the patient's quality of life. Because of the recent availability of large databases, the advances in technology, signal processing, and machine learning, researchers have been working to find new biomarkers and assist the diagnosis of neurodegenerative disease with more accurate and affordable detection systems. In this thesis, we present two methodologies for analyzing voice recordings and structural Magnetic Resonance Imaging (sMRI) for the detection of PD. Voice recordings present the opportunity for detecting the disease in a non-invasive and affordable manner up to ten years before a clinical diagnosis is made. In this work, the use of various classifiers and a reduced subset of features from the voice recordings obtained a 95.9% accuracy in the detection of PD in female patients and 94.36% when detecting PD in male patients. We also report and analyze the features that are selected in the best subset for each gender. The medical images used in this thesis provide an accurate representation of the structure of the tissues in the brain, which we can use to detect PD. The results obtained in this work with 1.5 and 3 Tesla sMRI include 99.01% accuracy in the male patient's dataset and 96.97% accuracy in the female patient's dataset. The regions of interest in the brain used for the detection of PD are shown in the results. In both voice recording and medical images experiments, we provide information to avoid the "black box" of the binary classification algorithms and facilitate the clinical interpretation of the results, such as specific affectations on the voice, statistical information, and the features from the regions of the brain that are used by the classification algorithms for each specific gender.

Resumen

La Enfermedad de Parkinson (EP) es un trastorno crónico neurodegenerativo que afecta a más de seis millones de personas, y se considera una de las enfermedades neurológicas más difíciles de diagnosticar, lo que puede resultar en tratamientos erróneos que afectan a la calidad de vida de los pacientes. Este problema ha llamado la atención de la comunidad académica debido a la reciente disponibilidad de bases de datos más amplias, así como los avances en tecnología, procesamiento de señales y aprendizaje automático, que pueden utilizarse para encontrar nuevos biomarcadores y ayudar al diagnóstico de las enfermedades neurodegenerativas con sistemas de detección más precisos y asequibles. En esta tesis presentamos dos metodologías, la primera para el análisis de grabaciones de voz y la segunda para el análisis de Imágenes de Resonancia Magnética Estructural (IRMe) para la detección de la EP. Las grabaciones de voz son capaces de detectar la enfermedad de forma no invasiva y asequible hasta diez años antes de que se realice un diagnóstico clínico. En este trabajo, el uso de varios clasificadores y un grupo reducido de características de las grabaciones de voz, obtuvieron una exactitud del 95.9% en la detección de EP en pacientes de sexo femenino y del 94.36% al detectar la EP en pacientes de sexo masculino. También reportamos el análisis de las características que se seleccionaron en el mejor subconjunto para cada sexo. En esta tesis utilizamos IRMe, que proporcionan una representación precisa de la estructura de los tejidos en el cerebro. Los resultados obtenidos en este trabajo con IRMe de 1.5 y 3 Tesla incluyen una exactitud de 99.01% en el conjunto de datos de pacientes masculinos y de 96.97% en el conjunto de datos de pacientes femeninos. En los resultados se muestran las regiones de interés del cerebro utilizadas para la detección de la EP. Tanto en los experimentos de grabación de voz como en los de imágenes médicas, proporcionamos información para evitar el problema de los resultados binarios de “caja negra” y facilitar la interpretación clínica de los resultados, como las afectaciones específicas de la voz, información estadística y las características de las regiones del cerebro que utilizan los algoritmos de clasificación para cada sexo específico.

Agradecimientos

Quiero agradecer a mi asesor el Dr. Roberto Rosas Romero, por dirigir este trabajo de tesis y mi formación como investigador. Siempre escuchó mis ideas y me apoyó en cada uno de los pasos que dí a lo largo del doctorado, supo guiarme y motivarme a hacer el mejor trabajo que fuera posible con comentarios de gran nivel académico y humano, sin su apoyo este trabajo nunca hubiera sido posible.

A mis sinodales el Dr. Juan Horacio Espinoza Rodríguez, el Dr. Oleg Starostenko Basarab, el Dr. Edgar Guevara Codina, el Dr. Jorge Martínez Carballido y el Dr. Jorge Rodríguez Asomoza por sus acertados y valiosos comentarios para mejorar el nivel académico de esta tesis.

A los profesores que me transmitieron su experiencia y conocimiento lo largo de este proceso de formación, especialmente al Dr. Juan Carlos Galán Hernández quien me brindó consejo desde el inicio de mis estudios de doctorado.

Al Dr. Pascal Tyrrell por aceptarme en su laboratorio para realizar una estancia de investigación y permitirme ampliar mis horizontes.

A mis compañeros y amigos que recorrieron este camino conmigo, aprendimos y crecimos juntos a lo largo de estos 5 años.

Quiero agradecer especialmente a mis padres, mis hermanas y mi novia que siempre creyeron en mi capacidad para superar cualquier reto, me brindaron su apoyo incondicional y estuvieron junto a mí durante todo este proceso. Sin ellos no habría llegado hasta aquí.

Finalmente agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por otorgarme la beca nacional para posgrados, así como a la Universidad de las Américas Puebla (UDLAP) por la beca investigación para realizar mis estudios de doctorado.

Índice

Abstract	2
Agradecimientos	3
Lista de Figuras	12
Lista de Tablas	22
Lista de Acrónimos	23
Capítulo 1 Introducción	26
1.1 Motivación	26
1.2 Trabajo relacionado	28
1.2.1 Detección de la enfermedad de Parkinson basada en grabaciones de voz	28
1.2.2 Imágenes médicas	30
1.3 Hipótesis	33
1.4 Dirección de la tesis	33
1.4.1 Objetivo general	35
1.4.2 Objetivos específicos	35
1.5 Organización del documento	35
Capítulo 2 La enfermedad de Parkinson	37
2.1 Etapas patológicas y síntomas cardinales de la enfermedad de Parkinson	37
2.1.1 Esquemas de las etapas	37
2.1.2 Características cardinales de la enfermedad de Parkinson en etapas tempranas	40
2.1.3 Características cardinales de la enfermedad de Parkinson en etapas intermedias	40
2.2 Diagnóstico de la enfermedad de Parkinson	40
2.2.1 Diagnóstico con estudios epidemiológicos	40
2.2.2 Diagnóstico diferencial	41
2.2.3 Escalas de clasificación	41
2.3 Biomarcadores	41
2.4 Desórdenes de la voz, habla y deglución	43
2.4.1 Problemas de deglución	44
2.5 Neuropatología	45

2.5.1	Histopatología	45
2.5.2	Cambios estructurales en el cerebro	45
2.5.3	Ganglios basales	45
2.5.4	Rol de los ganglios basales en el control motor	47
2.5.5	Dopamina	47
2.5.6	Sistema Nigrostriatal y síntomas de la Enfermedad de Parkinson	48
2.6	Discusión	48
Capítulo 3 Metodología y antecedentes teóricos		50
3.1	Descripción de las bases de datos	51
3.1.1	Grabaciones de voz provenientes de la facultad de Medicina de la Universidad de Estambul	51
3.1.2	Base de datos de imágenes de Resonancia Magnética Estructural del Parkinson's Progression Marker Initiative	51
3.2	Preprocesamiento de los datos	51
3.2.1	Preprocesamiento de Imágenes de Resonancia Magnética Estructural	53
3.3	Morfometría Basada en Vóxeles	56
3.4	Extracción de características	57
3.4.1	Características base de las señales de grabaciones de voz	58
3.4.2	Coefficientes Cepstrales de la Frecuencia de Mel	59
3.4.3	Características basadas en la Transformación de Ondículas Discreta	60
3.4.4	Transformada de Ondículas con Factor-Q Ajustable	61
3.4.5	Características de textura de Harlick	62
3.5	Selección de Características	65
3.5.1	Análisis de Componentes Principales	66
3.5.2	Selección del Grupo de Características Basado en Wrappers	67
3.6	Clasificación	68
3.6.1	Perceptrón Multicapa	68
3.6.2	K-vecino más cercano	69
3.6.3	Bosques Aleatorios	70
3.6.4	Máquina de Soporte Vectorial	71
3.7	Métricas de rendimiento	72
3.8	Metodologías propuestas	72
3.8.1	Metodología propuesta para la detección de la enfermedad de Parkinson en grabaciones de voz	73

3.8.2	Metodología propuesta para la detección de la enfermedad de Parkinson en Imágenes de Resonancia Magnética Estructural	74
Capítulo 4	Resultados	77
4.1	Grabaciones de voz	77
4.1.1	Base de datos con participantes de sexo femenino y masculino	78
4.1.2	Base de datos de participantes de sexo femenino	82
4.1.3	Base de datos con participantes de sexo masculino	85
4.1.4	Discusión	89
4.2	Imágenes de Resonancia Magnética Estructural	93
4.2.1	Resultados del proceso de Morfometría basada en Vóxeles	93
4.2.2	Selección de características de las Regiones de Interés	95
4.2.3	Resultados de la clasificación	102
4.2.4	Discusión	105
Capítulo 5	Conclusión y Trabajo Futuro	111
5.1	Contribuciones	112
5.2	Limitaciones de la investigación	113
5.3	Trabajo a Futuro	114
Bibliografía		115
Apéndice		122

Lista de Figuras

1	Morfología interna del cerebro presentada desde un corte axial. Adaptada de una IRMe de la base de datos del iniciativa de Marcadores de Progresión del Parkinson (PPMI) . . .	38
2	Morfología interna del cerebro presentada desde un corte coronal. Adaptada de una IRMe de la base de datos del PPMI	39
3	Morfología interna del cerebro presentada desde un corte sagital. Adaptada de una IRMe de la base de datos del PPMI	39
4	Representación esquemática del sistema motor extrapiramidal en un corte transversal. .	46
5	Representación general de la metodología para la detección de la EP utilizando señales biomédicas y algoritmos de aprendizaje automático.	50
6	La imagen en la esquina superior izquierda muestra la vista coronal del cerebro, la esquina superior derecha muestra la vista sagital y la imagen en la esquina inferior izquierda muestra la vista axial.	53
7	Origen seleccionado manualmente en el software de SPM12 en una IRMe de un paciente con la EP.	54
8	Resultados de las etapas de preprocesamiento de las IRMe de los pacientes.	57
9	Metodología propuesta para detectar la EP en grabaciones de voz. Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC.	73
10	Metodología propuesta para la detección de la EP utilizando IRMe.	75
11	Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos mixta. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC	81
12	Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos de participantes de sexo femenino. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC	84
13	Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos de participantes de sexo masculino. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC	88

14	<p>Mapa de colores de la vista coronal de los resultados de Morfometría basada en vóxeles (VBM) aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	94
15	<p>Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos.(para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	95
16	<p>Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos.(para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	96
17	<p>Mapa de colores de la vista coronal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	97

18	<p>Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	97
19	<p>Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	98
20	<p>Mapa de colores de la vista coronal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	99
21	<p>Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.</p>	100

22	Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.	101
23	Las dos Regiones de Interés (RDI) con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo femenino. En la figura se pueden ver las RDI 6 (mostrada en rojo) y 14 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.	102
24	Las cuatro RDI con las características más seleccionadas en las IRMe de 3 Teslas de participantes de sexo femenino. En la figura se pueden ver las RDI 1 (mostrada en rojo) y 14 (mostrada en verde) en la primera fila. Las RDI 18 (azul) y 19 (amarillo) se localizan en la segunda y tercera fila respectivamente. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.	103
25	Las dos RDI con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo masculino. En la figura se pueden ver las RDI 4 (mostrada en rojo) y 6 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.	103
26	Las tres RDI con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo masculino. En la figura se pueden ver las RDI 2 (mostrada en rojo) y 14 (mostrada en amarillo) en la primera fila. En la segunda fila, se observa la región 7 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.	104
27	Portada del artículo “Analysis of voice as an assisting tool for detection of Parkinson’s disease and its subsequent clinical interpretation”. (Gabriel Solana-Lavalle, 2020). CC-BY-NC	134

28	Portada del artículo “Analysis of voice as an assisting tool for detection of Parkinson’s disease and its subsequent clinical interpretation”. (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC	135
29	Portada del artículo “Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson’s disease”. (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC	136
30	Capítulo de libro: Technology, Science and Culture: A Global Vision, Volume III. (Solana-Lavalle & Rosas-Romero, n.d.). CC-BY-NC.	137

Lista de Tablas

1	Trabajos recientes sobre detección de la EP con grabaciones de voz.	31
2	Tabla comparativa entre los diferentes trabajos de detección de la EP basada en imágenes médicas.	34
3	Biomarcadores Clínicos de la EP	42
4	Las técnicas de imagen más utilizadas como biomarcadores para la EP. (Factor and Weiner 2017, Regnault et al. 2019)	43
5	Concentración de Dopamina (DA) en pacientes con la EP y sujetos sanos es comparada en las diferentes regiones de interés (Factor and Weiner, 2007).	48
6	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 384 observaciones provenientes de pacientes con la EP y sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el Selección del Grupo de Características Basado en Wrappers (SGCBW) para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	78
7	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 384 observaciones provenientes de 64 pacientes con la EP y 64 sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el SGCBW para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	79
8	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 756 observaciones provenientes de 188 pacientes con la EP y 64 sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el SGCBW para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	80
9	Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de ambos sexos.	81

10	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 246 observaciones provenientes de pacientes con la EP y controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	82
11	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 246 observaciones provenientes de 41 pacientes con la EP y 41 controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	83
12	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 366 observaciones provenientes de pacientes con la EP y controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	83
13	Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de sexo femenino.	85
14	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 138 grabaciones provenientes de pacientes con la EP y sujetos control de sexo masculino fueron utilizadas para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	86
15	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 138 observaciones provenientes de 23 pacientes con la EP y 23 controles de sexo masculino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	87

16	Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 390 grabaciones provenientes de pacientes con la EP y sujetos control de sexo masculino fueron utilizadas para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.	87
17	Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de sexo masculino.	89
18	Número de características seleccionadas de cada grupo (características base, Coeficientes Cepstrales de la Frecuencia de Mel (CCFM), Transformada de Ondículas (TO) y Transformada de Ondículas con Factor-Q Ajustable (TOFQA)) ejecutando el algoritmo SGCBW con cada clasificador en el conjunto de datos equilibrado.	90
19	Número de características seleccionadas de cada grupo (características base, CCFM, TO y TOFQA) ejecutando el algoritmo SGCBW con cada clasificador en el conjunto de datos no balanceada.	90
20	Comparación de los resultados de exactitud de la clasificación de pacientes con la EP con otras bases de datos de grabaciones de voz con características similares a las utilizadas en este trabajo	91
21	Número de características extraídas en cada uno de los grupos, y el número de características seleccionados después de aplicar Análisis de Componentes Principales (ACP) y correr el algoritmo de SGCBW para cada clasificador.	104
22	Resultados de la detección de la EP en participantes de sexo masculino y femenino con siete clasificadores diferentes y IRMe de 1.5 Teslas.	105
23	Características de las IRMe de 1.5T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick	106

24	Características de las IRMe de 1.5T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick	107
25	Características de las IRMe de 3T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick	108
26	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente una grabación y se entrenó cada algoritmo con el grupo de características óptimo de una Máquina de Soporte Vectorial (MSV).	122
27	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un Bosques Aleatorios (BA).	122
28	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un Perceptrón Multicapa (PM)	123

29	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un K-Vecino mas Cercano (KNN)	123
30	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	123
31	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	124
32	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	124
33	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	124

34	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	125
35	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	125
36	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	125
37	Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	126
38	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	126
39	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	126

40	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	127
41	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	127
42	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	127
43	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	128
44	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	128
45	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	128

46	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	129
47	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	129
48	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	129
49	Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	130
50	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	130
51	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	130

- 52 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA 131
- 53 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN 131
- 54 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV 131
- 55 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM 132
- 56 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA 132
- 57 Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN 132

58	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV	133
59	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM	133
60	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA	133
61	Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN	133

Lista de Acrónimos

ABC Área Bajo de la Curva

ACP Análisis de Componentes Principales

AM Aprendizaje de Máquina

AP Aprendizaje Profundo

BA Bosques Aleatorios

BI Bayes Ingenuous

CB Características Base

CCFM Coeficientes Cepstrales de la Frecuencia de Mel

CCM Coeficiente de Correlación de Mathew

CL Cuerpos de Lewis

CP Componentes Principales

DA Dopamina

DaTscan Tomografía Computarizada de Emisión de Fotón único para detectar Dopamina

DaT-SPECT Tomografía Computarizada de Emisión de Fotón único para detectar Dopamina

DFA Análisis de fluctuación de tendencia

DICOM Digital Imaging and Communications in Medicine

DT-MRI Resonancia magnética con tensor de difusión

EP Enfermedad de Parkinson

FWHM Kernel gaussiano con anchura máxima a la mitad

GPe Globo Pálido Pars Externa

GPI Globo Pálido Pars Interna

IA Inteligencia Artificial

IRM Imágenes de Resonancia Magnética

IRMe Imágenes de Resonancia Magnética Estructural

IRMf Imágenes de Resonancia Magnética Funcional

ITDRM Imágenes con Tensores de Difusión en Resonancia Magnética

KNN K-Vecino mas Cercano

LC Locus Ceruleus

MCCNG Matriz de Co-Ocurrencia de Niveles de Gris

MDS-UPDRS Movement Disorder Society-Unified Parkinson's Disease Rating Scale

ML Machine Learning

MLCP Memoria de Largo y Corto Plazo

MOG Mezcla de gaussianas

MSV Máquina de Soporte Vectorial

MSV-FBR Máquina de Soporte Vectorial con Función de Base Radial

NM Neuromelanina

NM-IRM Imágenes de Resonancia Magnética sensible a la Neuromelanina

PET Tomografía de Emisión de Positrones

PM Perceptrón Multicapa

PPE Entropía del periodo de tono

PPMI iniciativa de Marcadores de Progresión del Parkinson

PPQ5 Cociente de Perturbacion en un Periodo de Cinco Puntos en el Tiempo

RB Red Bayesiana

RDI Regiones de Interés

RNA Red Neuronal Artificial

RNC Red Neuronal Convolutacional

RNP Red Neuronal Profunda

RPDE Periodo de recurrencia Densidad Entropía

SGCBW Selección del Grupo de Características Basado en Wrappers

SN Substantia Nigra

SNc Substantia Nigra Pars Compacta

SNr Substantia Nigra pars Reticulata

SPECT Trazador de yodo-123 fluoropropilo combinado con tomografía computarizada por emisión monofotónica

STN Núcleo Subtalámico

TDO Transformada Discreta de Ondículas

TDRF Transformada Discreta Rápida de Fourier

TDF Transformada Discreta de Fourier

TO Transformada de Ondículas

TOFQA Transformada de Ondículas con Factor-Q Ajustable

UPDRS Escala unificada de valoración de la enfermedad de Parkinson

VBM Morfometría basada en vóxeles

sMRI structural Magnetic Resonance Imaging

PD Parkinson's Disease

Capítulo 1 Introducción

En este capítulo, la *Enfermedad de Parkinson (EP)*, sus características *cardinales*, los síntomas en etapas tempranas así como los retos para detectar la enfermedad a tiempo son presentadas. Después, se presenta la revisión del trabajo relacionado, en la cual se incluye el estado del arte para técnicas de detección de la EP las cuales se combina el procesamiento de señales y el Aprendizaje de Máquina (AM). Ya con el contexto proporcionado por el trabajo relacionado a la EP, se formula la hipótesis de la tesis. El objetivo de la tesis, así como las contribuciones tanto en el área de medicina, como en el área de la ingeniería son mencionados. Finalmente, la organización del resto del documento se muestra al final del capítulo.

1.1 Motivación

La EP es un desorden crónico degenerativo que fue estudiado por primera vez por el médico James Parkinson, quien describió los síntomas y progresión de la enfermedad en el “*Ensayo de la parálisis agitante*” en 1817, y fue nombrada en honor a él por William Rutherford Sanders en 1865 (Factor & Weiner, 2007)(Goedert & Compston, 2018). La EP es un desorden del sistema central nervioso, el cual es el resultado de la pérdida de células dopaminérgicas en diferentes regiones del cerebro, incluyendo la región de la Substantia Nigra (SN). Las células de ésta región son las encargadas de producir la *dopamina*, un mensajero químico el cual es responsable de transmitir señales al cerebro que permiten la coordinación de movimientos entre otras funciones. Algunos de los resultados de la pérdida de dopamina en el cerebro incluyen desórdenes cognitivos, afectaciones motoras, trastornos del estado de ánimo, entre otros (Michael J. Fox Foundation, 2020).

La EP es el desorden neurodegenerativo más común después de la enfermedad de Alzheimer, y es la enfermedad neurodegenerativa más común en personas en el rango de entre 50 a 60 años de edad (Qureshi, Baez, & Reitz, 2020). Estudios muestran que la mayor parte de los casos de la EP se presentan en pacientes que tienen más de 50 años de edad, y únicamente de un 4 a 10 por ciento de los pacientes tienen un diagnóstico antes de los 40 años de edad. En el año 2015, seis millones de personas fueron diagnosticadas con la EP, y se espera que para el año 2040 el número de casos sea más del doble; esto debido a el incremento en la esperanza de vida de la población (DESA, 2015).

Los *síntomas cardinales* de la EP incluyen síntomas motores primarios como el temblor en estado de reposo, donde los miembros como los brazos o piernas se mueven sin control. Otro síntoma car-

dinal es la lentitud al iniciar un movimiento (dyskinesia), lentitud en los movimientos (bradkynesia), problemas de balance y rigidez. Los síntomas motores secundarios incluyen trastornos de la marcha, pequeñas desviaciones en la escritura (micrographia) entre otras (Moustafa et al., 2016). No todos los pacientes presentan todos los síntomas, y cuando se llegan a presentar las afectaciones motoras primarias, generalmente es durante una etapa avanzada de la enfermedad, lo que hace que el diagnóstico de la EP sea muy complicado. Por otro lado, los síntomas no motrices incluyen el deterioro cognitivo, trastornos del estado de ánimo, dificultades para dormir, pérdida del sentido del olfato, problemas de habla y deglución, deterioro de habilidades de escritura, babeo y presión arterial baja al estar de pie. Estos síntomas pueden ayudar a la detección y seguimiento de la progresión de la enfermedad en etapas tempranas, antes de que aparezcan síntomas cardinales que pueden ser más evidentes (Michael J. Fox Foundation, 2020).

El diagnóstico de la EP es una de los más complicados en el área de la neurología, y frecuentemente resulta en la administración de tratamientos inadecuados para los pacientes, principalmente en los casos en los cuales el paciente experimenta temblores aislados o afectaciones poco evidentes. Debido a que la EP progresa lentamente, en la mayoría de los casos los síntomas más evidentes pueden tardar años en aparecer, lo que significa que el paciente puede haber vivido con la enfermedad durante varios años sin saberlo ni tomar los medicamentos adecuados (Heim Beatrice, 2017). Diagnosticar correctamente la EP es una tarea compleja, ya que algunos de los pacientes pueden cumplir con las características de la enfermedad sin realmente padecerla. Los parkinsonismos (un grupo de desórdenes neurológicos que causan afectaciones motoras parecidas a las que se observan en pacientes con la EP) pueden llegar a causar confusión en el diagnóstico clínico de la enfermedad. Esto es corroborado con un estudio del año 2010 (Esiri, 2010), en el cual se realizó una autopsia al cerebro de pacientes que habían sido clínicamente diagnosticados con la EP. En este estudio se encontró que el 35% de los pacientes fueron incorrectamente diagnosticados. Sin embargo, los estudios patológicos y biopsias son posibles en un número muy reducido de pacientes (A schrag, 2002; Esiri, 2010). Estas limitantes en la detección de la enfermedad y su correcto diagnóstico puede costar tiempo y dinero muy valioso para los pacientes, así como reducir su calidad de vida.

Grupos de investigación en el área de medicina ha tenido un interés cada vez mayor en encontrar biomarcadores que puedan indicar la presencia de la EP y al mismo tiempo proveer información relevante acerca de las causas, progresión y tratamiento. Síntomas prodrómicos como las afectaciones de voz son características de las primeras etapas de la EP cuando el cerebro está comenzando a tener afectaciones. Grabaciones de voz por medio de dispositivos móviles o computadoras, son herramientas que pueden ayudar a detectar la enfermedad con un bajo costo económico y sin tener que encontrarse físicamente en una clínica u hospital. De la misma manera, imágenes médicas como las IRMe y las

Imágenes de Resonancia Magnética Funcional (IRMf) pueden confirmar el diagnóstico de la EP. La detección en etapas tempranas permite iniciar los tratamientos necesarios para mejorar la calidad de vida de los pacientes.

Una de las mayores preocupaciones de la comunidad médica para aceptar algoritmos de AM e Inteligencia Artificial (IA), es la falta de transparencia en la toma de decisiones. Algunos modelos como el Aprendizaje Profundo, tienen miles de parámetros, lo que hace su comportamiento y predicciones extremadamente difíciles de explicar; ésta es una de las razones por las cuales los llaman “*cajas negras*” (Wang et al., 2020; Wongvibulsin et al., 2020; Josephson & Wiebe, 2020). A pesar de que los algoritmos de aprendizaje profundo tienen una alta exactitud en las predicciones, en la mayoría de los casos los resultados son binarios, lo que significa que únicamente nos indican si el paciente tiene la enfermedad o no. Los médicos requieren más información para complementar los diagnósticos y para rastrear la progresión y severidad de las enfermedades. En este trabajo de tesis nos enfocamos en detectar la EP con una exactitud similar o superior a la del estado del arte, utilizando el procesamiento de señales y algoritmos de AM. También tenemos el objetivo de proveer información relevante en las señales de voz y las imágenes de resonancia magnética para detectar la enfermedad y ayudar a la interpretación médica de los resultados.

1.2 Trabajo relacionado

Los avances en la tecnología han hecho posible la aplicación de algoritmos de AM e IA para la detección y rastreo de la progresión de enfermedades neurodegenerativas para ayudar el diagnóstico clínico realizado por los médicos. La detección de enfermedades neurodegenerativas como el Alzheimer y la EP ha sido de interés para los investigadores tanto en los campos de medicina, como en los de bio-informática, debido a la complejidad del problema y el impacto positivo que puede generar en la calidad de vida de los pacientes. El incremento del número de bases de datos para el entrenamiento y la validación de los algoritmos de IA ha contribuido al desarrollo de modelos alternativos para la detección de estas enfermedades. Algunos de los trabajos más relevantes para la detección de la EP se presentan en esta sección.

1.2.1 Detección de la enfermedad de Parkinson basada en grabaciones de voz

Las afectaciones en la voz son presentadas por hasta el 90% de los pacientes dentro de las primeras dos etapas del modelo de la progresión de la EP presentada por Braak et al. (Braak et al., 2004), siendo estos síntomas unos de los primeros en indicar que se tiene la enfermedad. Debido a que los pacientes pueden fácilmente grabar sus voces, ha habido un incremento en el interés por utilizar las grabaciones de voz para la detección temprana de la EP. Algunas deficiencias del habla presentadas por los pacientes de

Parkinson incluyen la *dysphonia* (uso defectuoso de la voz), *hypophonia* (reducción en el volumen de la voz), *monotone* (reducción en el rango de frecuencias), and *dysarthria* (dificultad con la articulación de sonidos o sílabas, causado por debilidad muscular) (Sakar et al., 2013). Los trabajos más relevantes en esta área son presentados en esta sección. Los trabajos recientes en la detección de la EP basado en el análisis de grabaciones de voz y sus bases de datos se muestran en la Tabla 1.

La mayor base de datos pública de grabaciones de voz provenientes de pacientes con la EP e individuos sanos, fue introducida por primera vez en el año 2019 por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019). En su trabajo, Sakar et al. extrajeron cuatro grupos de características provenientes de las grabaciones de voz, incluyendo las *Características Base*, los Coeficientes Cepstrales de Frecuencia de Mel CCFM, los coeficientes de la Transformada de ondículas TO y coeficientes de la Transformada de Ondículas con Factor-Q Ajustable TOFQA, para generar un total de 754 características por cada grabación. De acuerdo con los autores, el grupo de características de TOFQA fue usado por primera vez en este trabajo para la detección de la EP. La principal contribución de su trabajo incluye la comparación entre los grupos de características que extrajeron, mostrando la relevancia de los CCFM y los coeficientes TOFQA para la detección de la EP en grabaciones de voz. La clasificación fue realizada con un grupo de 50 características, seleccionadas utilizando el algoritmo de *mínima redundancia máxima relevancia* y una Máquina de Soporte Vectorial con Función de Base Radial (MSV-FBR), que fue el clasificador que obtuvo los mejores resultados, con alrededor de un 86% de exactitud.

Los algoritmos de Aprendizaje Profundo (AP) son herramientas poderosas, las cuales han ganado popularidad debido a su buen desempeño en problemas de clasificación. En el trabajo de (Rizvi et al., 2020), se utilizó una Red Neuronal Profunda (RNP) y una red de Memoria de Largo y Corto Plazo (MLCP) para la predicción de EP utilizando grabaciones de voz. Este modelo fue propuesto como una alternativa a la naturaleza secuencial de las redes pre-alimentadas, ya que no han tenido buenos resultados en problemas como el reconocimiento de voz. El uso de redes MLCP resuelve el problema de “memorizar” la información previa para calcular la salida de la red neuronal. Para entrenar y probar el modelo, características base fueron extraídas de las muestras de voz provenientes de 20 individuos sanos y 20 pacientes con la EP. Los parámetros de la Red Neuronal Artificial (RNA) y de la red MLCP fueron escogidos heurísticamente, obteniendo una exactitud del 97.96 %, con una desviación estándar de menos de $\pm 1\%$. Los resultados de éste modelo son de los mejores reportados para la detección de la EP con grabaciones de voz, sin embargo, es posible que debido al tamaño de la base de datos utilizada, exista un sobre ajuste de los datos.

En el estudio conducido por (Quan, Ren, & Luo, 2021), un modelo de red bidireccional MLCP es propuesto en conjunto con características estáticas y dinámicas de señales de voz para la detección

de la EP. En éste trabajo, se encontró que la articulación entre las transiciones y la tendencia de la curva de la frecuencia fundamental, varía en significado entre los pacientes con la EP y los controles sanos, resultando en una característica importante para la clasificación. El modelo propuesto recibe series de tiempo dinámicas de articulaciones como características de entrada. La base de datos contiene 268 grabaciones de voz de la vocal /a/ sostenida durante unos segundos, proveniente de 35 sujetos (20 con la EP, y 15 controles sanos). Las grabaciones fueron utilizadas para el entrenamiento y la prueba del modelo. Las muestras de voz fueron adquiridas con un celular inteligente con un muestreo de 96 kHz. Este dato es importante ya que prueba que aparatos disponibles al público pueden ser utilizados para obtener grabaciones de voz con suficiente calidad para la detección de la EP. Utilizando las características dinámicas de las muestras de voz, la red bidireccional MLCP, obtuvo una exactitud del 84.29 %, mientras que una Red Convolutiva obtuvo una exactitud del 83.52 %, lo que demuestra que el modelo propuesto obtiene mejores resultados.

El trabajo de (D. Braga, 2019) prueba que es posible detectar la EP en etapas tempranas en grabaciones de voz de habla libre con condiciones no controladas de ruido ambiental. En este trabajo se utilizaron tres bases de datos diferentes, las cuales tienen condiciones de sonido ambiental no controladas y fueron grabadas con diferentes micrófonos con una velocidad de muestreo de 44.1 kHz. Características base como el Jitter, Shimmer, y el tono, entre otras, fueron utilizadas para el proceso de clasificación. Tres clasificadores diferentes fueron probados y la validación cruzada *deja-uno-fuera* fue utilizada. Los resultados muestran que el algoritmo de clasificación de BA, obtiene la mayor exactitud de entre los tres clasificadores, pero se menciona que esto puede deberse al sobre ajuste a la base de datos. Los algoritmos de MSV-FBR y RNA obtuvieron una exactitud de un 91-92% y generalizaron de mejor manera en las tres bases de datos. También se muestra que los resultados de la RNA tienen una mayor varianza cuando existen valores atípicos. Los resultados obtenidos para la detección de la EP en condiciones no controladas fueron relevantes, ya que abren la posibilidad de utilizar la tele-diagnóstico con diferentes dispositivos, permitiendo al paciente recibir un diagnóstico aun cuando no se encuentra en un hospital.

1.2.2 Imágenes médicas

Cuando las células dopaminérgicas en la región de la *substantia nigra* en el cerebro comienzan a morir, la falta del neurotransmisor llamado *dopamina* resulta en la aparición de los síntomas de la EP. Por esta razón, las imágenes médicas han jugado un rol importante a lo largo de los años para entender y detectar la EP. Algunas de las técnicas de imagenología más utilizadas para detectar la EP son IRMe, IRMf, Imágenes con Tensores de Difusión en Resonancia Magnética (ITDRM), e imágenes de Tomografía Computarizada de Emisión de Fotón único para detectar Dopamina (DaTscan). DaTscan es una técnica

Tabla 1: Trabajos recientes sobre detección de la EP con grabaciones de voz.

Autor, Año	Base de datos	Resultados
(Guruler, 2017)	195 mediciones de voz provenientes de 8 sujetos sanos y 23 con la EP	Exactitud = 0.99, Sensibilidad = 1, Especificidad = 0.99, Valor-F = 0.99
(B. E. Sakar, 2017)	42 pacientes con la EP y 8 sujetos control	Exactitud = 0.96, Coeficiente de Correlación de Mathew (CCM) = 0.77
(Sakar et al. 2019)	756 grabaciones de voz provenientes de 64 individuos sanos y 188 pacientes con la EP	Exactitud = 0.86, CCM = 0.59
(D. Braga, 2019)	22 hablantes con la EP y 30 hablantes sanos	Exactitud = 0.99 con un clasificador de BA
(Rizvi et al., 2020)	195 muestras de voz provenientes de 8 personas sanas y 23 pacientes con la EP	Exactitud = 0.97
(Quan et al., 2021)	268 muestras de voz obtenidas con un celular inteligente provenientes de 15 personas sanas y 40 pacientes con la EP	Exactitud = 0.8429
(Vital et al., 2021)	1200 muestras de voz provenientes de 51 personas sanas y 62 pacientes con la EP	Exactitud = 1
(Pramanik et al., 2021)	756 grabaciones de voz provenientes de 64 personas sanas y 188 pacientes con la EP	Exactitud = 0.95

que permite el monitoreo de los niveles de dopamina en diferentes áreas del cerebro utilizando inyecciones de Iodine-123 y marcadores de fluoro propileno combinadas con tomografías de emisiones computarizadas de un único fotón. Esta técnica es la que naturalmente se elegiría para detectar la EP, sin embargo, debido a su elevado costo y a su poca disponibilidad en algunos países, se están buscando alternativas a esta técnica. IRMe y IRMf son algunas de las alternativas más comunes y económicamente viables que permiten observar la estructura del cerebro, y el flujo de sangre a través del tiempo respectivamente. La Tabla 2 muestra trabajos recientes para detectar la EP por medio de IRMe y IRMf.

El trabajo de (Yasaka et al., 2021), presenta un análisis del posible uso de algoritmos de AP para detectar la EP e identificar desórdenes en los circuitos neuronales en los pacientes. Para este estudio, 115 pacientes con la EP y 115 participantes sanos se ofrecieron como voluntarios para tomarse imágenes de IRMe IRMf y ITDRM con un sensor con un campo magnético de 3 Teslas. Uno de los retos en el AP es la falta de transparencia en cuanto al funcionamiento del algoritmo. Para atacar este problema, los autores utilizaron un *Mapa de activaciones de clase ponderada por gradientes (Grad-CAM)* que hace posible la visualización de las regiones de las imágenes que está utilizando el modelo de AP con mayor frecuencia. Otra de las ventajas del modelos, es que permite la visualización de los circuitos neuronales con alteraciones en los pacientes con la EP. El rendimiento que se obtuvo fue del 0.895 Área Bajo de la Curva (ABC) y una Exactitud del 81%. Se encontró que la Red Neuronal Convolutiva (RNC) se concentró en la conexión entre el Ganglio Basal y el Cerebelo.

Las RNC han estado creciendo en popularidad y son especialmente buenas para los problemas de clasificación de imágenes. Estas consisten en una RNA con capas convolucionales que pueden extraer automáticamente características de las imágenes. La arquitectura de la RNC “AlexNet”, la cual consiste en cinco capas convolucionales y tres capas completamente conectadas en conjunto con la técnica de transferencia de aprendizaje, es utilizada por (Sivaranjini & Sujatha, 2019) para ayudar a los médicos a tener una mejor exactitud en el diagnóstico de la EP. El objetivo del trabajo es que algunas capas extrajeran características de las Imágenes de Resonancia Magnética (IRM), y que las capas especializadas pudieran clasificar a los pacientes con la EP. En este trabajo, la base de datos del PPMI fue utilizada para realizar los experimentos. La base de datos contiene IRM provenientes de 182 individuos, 100 con la EP y 82 sujetos sanos, incluyendo pacientes del sexo femenino y masculino. Los resultados muestran una exactitud del 88.9% en cuanto a la habilidad de diferenciar entre participantes sanos y pacientes con la EP.

En el trabajo de (Lorio Sara, 2019), la Tomografía Computarizada de Emisión de Fotón único para detectar Dopamina (DaT-SPECT), ITDRM y IRMe fueron utilizadas para observar los neurotransmisores y microestructuras de los pacientes con la EP. La combinación de estas imágenes médicas permitieron la predicción de la progresión y severidad de los síntomas en pacientes *novo* con la EP.

En este estudio participaron 105 pacientes con la EP y 105 sujetos sanos, con y sin tratamiento de L-Dopa. Para identificar las diferencias entre ambos grupos, los autores realizaron un estudio basado en la covarianza entre vóxeles en MatLab. Encontraron que las regiones cortical y subcortical sufrieron alteraciones en pacientes novo con la EP, así como afectaciones en el striatum y regiones medias del cerebro, las cuales se ven generalmente afectadas por la enfermedad. El Movement Disorder Society- Unified Parkinson's Disease Rating Scale (MDS-UPDRS), es una escala estándar, utilizada para evaluar la severidad de los síntomas percibidos por los pacientes con la EP. En este estudio, los resultados de la prueba de MDS-UPDRS estuvieron altamente relacionados con alteraciones en el putamen, cortex-prefrontal y la ínsula. Sin embargo, el MDS-UPDRS es contestado por pacientes que están bajo la influencia del tratamiento de L-Dopa, y esto debe ser tomado en cuenta.

La *Neuromelanina (NM)* es un pigmento que se encuentra en las células dopaminérgicas, y en las neuronas *locus coeruleus noradrenergic*. Este pigmento se pierde cuando las neuronas en la sustantia nigra comienzan a morir, lo que es un indicador de que el paciente pueda tener la EP. La NM está relacionada con la concentración de metales en el cerebro como el hierro, cobre y zinc. En el trabajo de (Sulzer et al., 2018), se propone que utilizar altos campos magnéticos de 7 Teslas en las IRM, puede resultar en imágenes en las cuales se puede detectar la NM, debido a las propiedades paramagnéticas de esta sustancia. Los autores proponen que las Imágenes de Resonancia Magnética sensible a la Neuromelanina (NM-IRM) pueden ser un biomarcador para la detección de la EP, y que es viable para realizar estudios longitudinales, debido a que la acumulación o decrecimiento de la NM en el cerebro se puede observar. Sin embargo, la concentración de NM debido al envejecimiento normal, la administración de L-Dopa para reducir los síntomas en los pacientes con la EP, entre otras, puede interferir con las NM-IRM y es un problema que se debe de solucionar.

1.3 Hipótesis

Basada en la revisión de literatura, se formula la siguiente hipótesis:

“La enfermedad de Parkinson puede detectarse utilizando señales biomédicas adquiridas de forma no invasiva, como grabaciones de voz e imágenes de resonancia magnética estructural, con gran precisión seleccionando subconjuntos de características con diferentes clasificadores y obteniendo más del 95% de exactitud y proporcionando datos para interpretación clínica”

1.4 Dirección de la tesis

Es importante encontrar alternativas para la detección de la EP que le den al médico suficiente información para poder diagnosticar e interpretar los resultados de los algoritmos de aprendizaje de máquina, y así poder dar tratamientos adecuados a los pacientes. Las señales que estamos utilizando para la de-

Tabla 2: Tabla comparativa entre los diferentes trabajos de detección de la EP basada en imágenes médicas.

Autor y año	Base de datos	Rendimiento
(Yasaka et al., 2021)	115 pacientes con la EP y 115 sujetos sanos	Exactitud del 81%
(Lei et al., 2018)	base de datos del PPMI con IRM	Exactitud del 86.48%
(Sivaranjini & Sujatha, 2020)	base de datos del PPMI con IRM	Exactitud del 88.9%
(Esmailzadeh et al., 2018)	base de datos del PPMI con IRM	Exactitud del 100%
(Lorio Sara, 2019)	28 pacientes con la EP y 28 sujetos sanos	Exactitud del 92.7%
(Shah et al., 2018)	base de datos del PPMI con IRM	Exactitud del 93%
(Shinde et al., 2019)	45 pacientes con la EP y 35 sujetos sanos	Exactitud del 85%
(Amoroso et al., 2018)	base de datos del PPMI con IRM	Exactitud del 93%, sensibilidad del 93%, especificidad del 92%

tección de la EP son las grabaciones de voz e imágenes médicas, incluyendo IRMe y IRMf. Estas señales fueron escogidas debido a que se pueden detectar alteraciones provocadas por la EP en etapas tempranas y son económicamente viables comparadas con otras técnicas como las imágenes de DaTscan. En esta tesis, los estudios son realizados de manera independiente para los pacientes femeninos y los pacientes masculinos, tanto en el análisis de las señales de voz y en las imágenes médicas. La voz y el cerebro de los pacientes muestran diferencias significativas entre los dos sexos; extrayendo las características más relevantes para cada sexo en específico, los algoritmos pueden incrementar su eficiencia.

1.4.1 Objetivo general

- Detectar la EP por medio de grabaciones de voz y de IRMe durante las primeras 3 etapas de Braak (Bryce & Sprague, 2012) con algoritmos de AM con más de un 95 % de exactitud y proporcionar información para la interpretación clínica de los resultados.

1.4.2 Objetivos específicos

- Preprocesar las IRMe de 1.5 y 3 Teslas, realizar VBM y encontrar las RDI resultantes para detectar la EP.
- Extraer características de primer y segundo orden estadístico de las RDI encontradas en las IRMe.
- Realizar selección de características de primer y segundo orden estadístico de las RDI encontradas en las IRMe.
- Entrenar algoritmos de AM con los subconjuntos de características seleccionados para detectar la EP en grabaciones de voz y en IRMe en sujetos de sexo femenino y sexo masculino.
- Obtener las métricas de rendimiento de clasificación para detectar la enfermedad con validación cruzada de 10 veces para las grabaciones de voz y las IRMe.
- Encontrar las características que resultan más útiles para la detección de la EP para cada género.
- Interpretar el tipo de características que fueron seleccionadas para la clasificación y visualizar en las IRMe las regiones del cerebro que fueron afectadas.
- Analizar las RDI y características que se comparan entre géneros, así como las diferencias del campo magnético.

1.5 Organización del documento

El resto del documento de tesis está organizado de la siguiente manera:

Los conceptos biológicos de la EP se revisan en el **Capítulo 2**. Se explican los síntomas cardinales, el modelo de Braak (Braak et al., 2004) para definir las etapas de la EP, y la neuropatología de la enfermedad. El objetivo de este capítulo es entender la enfermedad, su origen y su progresión. En el **Capítulo 3** se explica el marco teórico del procesamiento de señales, la extracción de características, la selección de características y los algoritmos de clasificación. Se muestra la metodología propuesta para detectar la EP con grabaciones de voz y IRMe. En el **Capítulo 4**, se muestran los resultados de la selección de características y la clasificación para la detección de la EP en grabaciones de voz, y IRMe. Los resultados obtenidos para cada sexo y clasificador se discuten al final de cada sección. El **Capítulo 5** muestra la conclusión de la tesis, las limitaciones, y el trabajo a futuro.

Capítulo 2 La enfermedad de Parkinson

En este capítulo nos enfocamos en las características más importantes de la EP y las consideraciones neuropatológicas para el trabajo. Primero, describimos el modelo de Braak et al. (Braak et al., 2004), el cual propone un modelo de diferentes etapas y los síntomas asociados a la enfermedad. Este modelo describe las etapas y los síntomas que pueden ser útiles para entender la progresión de la EP. Después, se muestran los diagnósticos epidemiológicos y diferenciales de la EP. Los síntomas cardinales y biomarcadores son fundamentales para desarrollar nuevas herramientas de diagnóstico y tratamientos para la enfermedad. Tanto los síntomas cardinales como los biomarcadores son discutidos en este capítulo. Desórdenes relacionados con la voz, el habla y problemas para deglutir son de especial interés, ya que son de los primeros síntomas en aparecer y pueden ayudar a la detección de la EP y comenzar a tratar a los pacientes con medicamentos adecuados y a tiempo (D. Braga, 2019). Finalmente, se discute la neuroanatomía y neuropatología del cerebro de los pacientes con la EP.

2.1 Etapas patológicas y síntomas cardinales de la enfermedad de Parkinson

Para entender y diagnosticar la EP, es importante identificar las diferentes etapas de la enfermedad, así como las características de cada una de éstas etapas. En esta sección vamos a presentar uno de los modelos que mejor describe las diferentes etapas, y que es utilizado en este trabajo, así como las características cardinales de la EP.

2.1.1 Esquemas de las etapas

Braak et al. (Braak et al., 2004) propone un modelo de la progresión de la EP, como una sinucleinopatía, la cual es una enfermedad neurodegenerativa caracterizada por la acumulación anormal de agregados de la proteína alfa-sinucleína en las neuronas, fibras nerviosas, y células gliales con 6 etapas neuropatológicas.

Las afectaciones que ocurren durante las primeras dos etapas de la enfermedad son caracterizadas por los cambios en los nervios autónomos periféricos y en los ganglios, el núcleo olfativo y el núcleo motor dorsal, del nervio vago. En las etapas tres y cuatro, la substantia nigra y las porciones basales del cerebro medio y el cerebro anterior son involucradas. La pars compacta de la substantia nigra, donde las neuronas dopaminérgicas del camino nigrostriatal son localizadas, es involucrado en estas

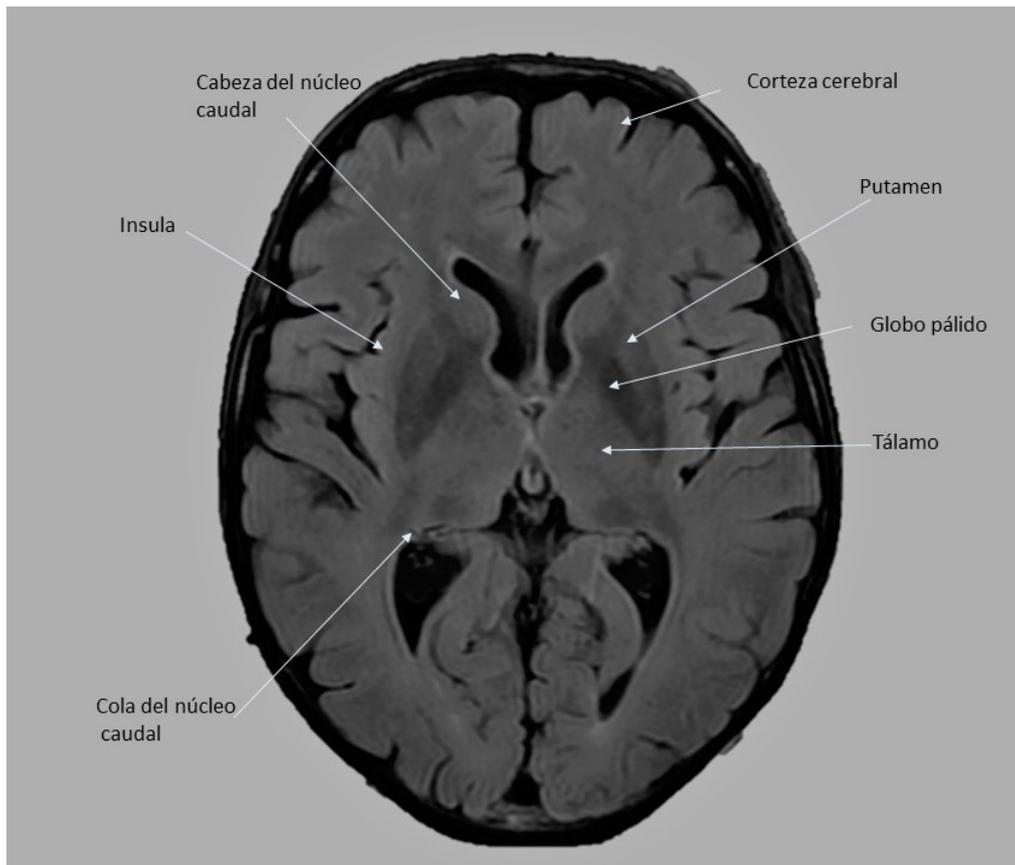


Figura 1: Morfología interna del cerebro presentada desde un corte axial. Adaptada de una IRMe de la base de datos del PPMI

etapas intermedias. Otras regiones del cerebro que muestran cambios patológicos incluyen el núcleo colinérgico tegmental pedunculopontino, el núcleo rafe (que se puede observar en los cortes sagitales de las IRMe debajo de la sustancia nigra), el núcleo colinérgico magnocelular del cerebro anterior basal y el núcleo tuberomamaria hipotalámica. También, los pacientes comienzan a mostrar síntomas motores durante estas etapas.

Durante las etapas cinco y seis, el proceso degenerativo ya se extendió por el cerebro y envuelve una región amplia del neocortex (región superficial del cerebro encima del sistema límbico). Las patologías previamente descritas en otras regiones se vuelven mucho más pronunciadas y patologías de cuerpos de inclusión están ahora presentes en el área prefrontal, así como en los córticos asociados, e incluso en áreas neocorticales primarias (Braak et al., 2004) (Factor & Weiner, 2007).

Regiones relevantes del cuerpo estriado, como la sustancia nigra y el putamen, son observables desde los cortes axiales, coronales y sagitales en las Figuras 1, 2 y 3 respectivamente.

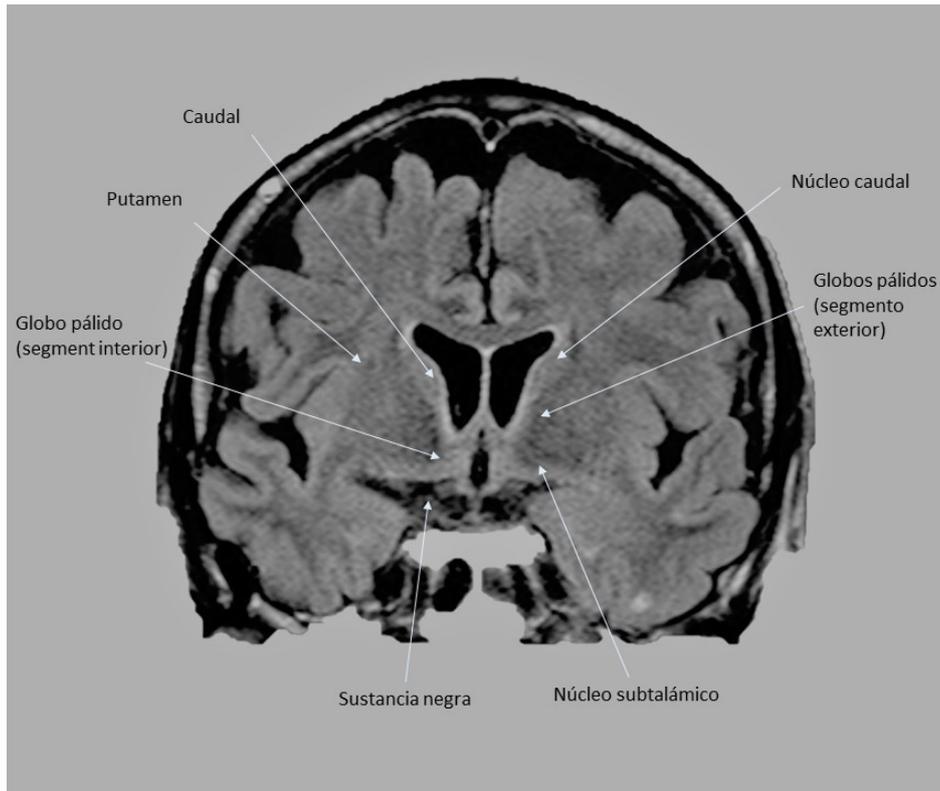


Figura 2: Morfología interna del cerebro presentada desde un corte coronal. Adaptada de una IRMe de la base de datos del PPMI

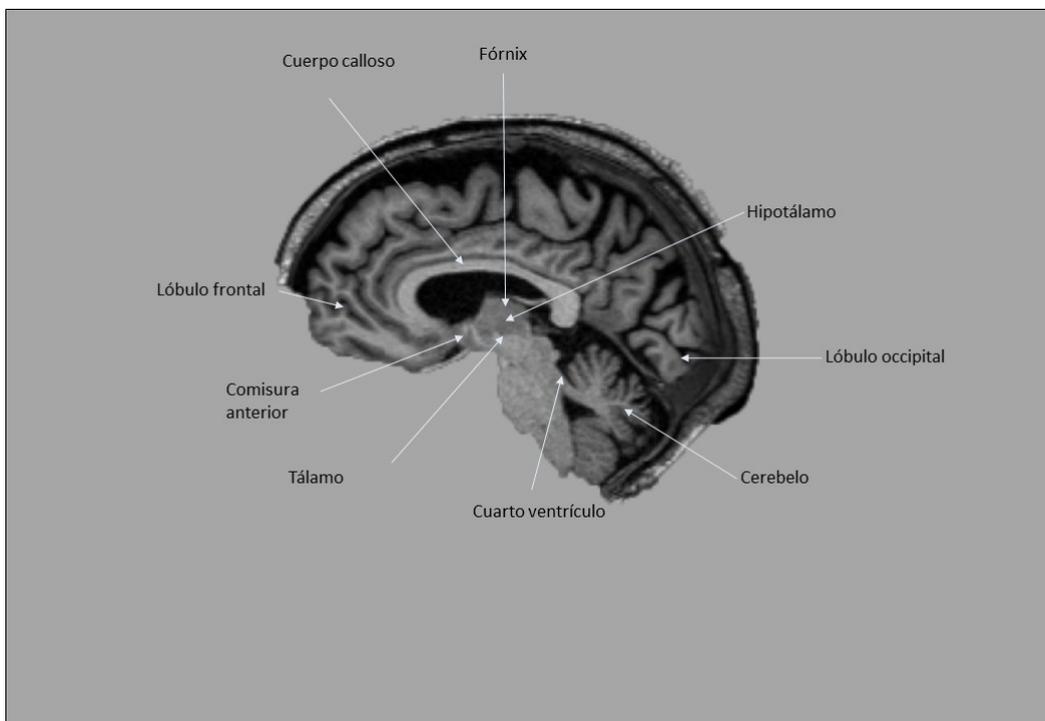


Figura 3: Morfología interna del cerebro presentada desde un corte sagital. Adaptada de una IRMe de la base de datos del PPMI

2.1.2 Características cardinales de la enfermedad de Parkinson en etapas tempranas

Síntomas tempranos de la EP pueden incluir fatiga, reducción de energía, rigidez en las articulaciones, calambres musculares o afectaciones sensoriales vagas. Una escritura lenta con letras pequeñas (micrographia), ronquera o una voz suave (hypophonia), afectaciones en el sueño, el síntoma de la pierna inquieta, estreñimiento y sentido del olfato disminuido (hyposmia), son síntomas comunes de las etapas tempranas de la enfermedad (antes de que comiencen a aparecer síntomas motores). Es importante notar que desórdenes del sueño REM idiopático y disfunciones del olfato pueden estar correlacionados con la presencia de disfunción nigroestriatal (Stiasny-Kolster et al., 2004). Todas estas características son útiles para identificar la EP antes de que el paciente muestre síntomas evidentes, lo que ayuda al diagnóstico de la enfermedad en etapas tempranas.

2.1.3 Características cardinales de la enfermedad de Parkinson en etapas intermedias

Los síntomas de la EP son muchos, sin embargo, los síntomas más importantes para el diagnóstico clínico se llaman síntomas *cardinales*. Estos síntomas generalmente aparecen en las etapas intermedias de la enfermedad e incluyen movimientos involuntarios de las extremidades cuando la extremidad está en reposo (Disquinesia), lentitud de movimiento (Bradiquinesia), inestabilidad postural, problemas de balance y rigidez.

2.2 Diagnóstico de la enfermedad de Parkinson

El diagnóstico de la EP es considerada como uno de los más complicados en el área de neurología. Debido a que la enfermedad progresa lentamente, en la mayoría de los casos los síntomas pueden tardar años en que sean evidentes, lo que significa que en una gran cantidad de casos, cuando el paciente es correctamente diagnosticado, ya ha vivido con la enfermedad durante varios años (Heim Beatrice, 2017).

2.2.1 Diagnóstico con estudios epidemiológicos

Los estudios epidemiológicos requieren largos periodos de tiempo para dar resultados, al contrario de los estudios de laboratorio, los cuales se completan en el transcurso de horas o días. Sin embargo, este tipo de estudios epidemiológicos son conocidos por dar resultados sesgados. La epidemiología de la EP puede ser dividida en las siguientes categorías: descriptiva, analítica y experimental. El diagnóstico epidemiológico descriptivo busca detectar la incidencia, mortalidad y prevalencia en los pacientes. Los estudios analíticos identifican aquellos factores que influyen la frecuencia de aparición de la enfermedad, tendencias, y la significancia de estos. Los estudios epidemiológicos experimentales se enfocan principalmente en los tratamientos, lo que requiere que se hagan consideraciones clínicas cuidadosas y

sean documentados adecuadamente. Estos estudios tratan de identificar las relaciones entre los diferentes eventos y variables que están involucrados en la EP para tratar de determinar su causa. Para realizar el diagnóstico, los médicos buscan la presencia de síntomas cardinales; si dos o más síntomas cardinales están presentes, el paciente es diagnosticado con la EP (Factor & Weiner, 2007).

2.2.2 Diagnóstico diferencial

El diagnóstico diferencial se realiza cuando el médico intenta diferenciar entre dos enfermedades que presentan síntomas y características similares, como en el caso en el que se busca diferenciar entre la EP y *los parkinsonismos*. El diagnóstico diferencial de los parkinsonismos es extensivo; incluye el envejecimiento normal, temblores, parkinsonismos inducidos por medicamentos o drogas, parkinsonismos vasculares, presión normal de hidrocefalia, entre otros. Por otro lado, el diagnóstico clínico de la EP requiere el historial clínico del paciente, examinación neurológica, documentación de mejora en el desarrollo de los problemas motores durante el tratamiento dopaminérgico. Las técnicas de neuroimagenología también se usan para el diagnóstico, sin embargo, la interpretación es realizada por un médico, lo que puede llevar a errores humanos, por lo que no se consideran biomarcadores inequívocos de la progresión de la enfermedad (Factor & Weiner, 2007).

2.2.3 Escalas de clasificación

La escala del Escala unificada de valoración de la enfermedad de Parkinson (UPDRS) fue desarrollada para diagnosticar y clasificar la severidad de la EP, así como monitorear su progreso. Esta escala ha sido ampliamente utilizada por los médicos para ajustar y medir la efectividad de los medicamentos como la L-Dopa (Savoiaro, 2003).

2.3 Biomarcadores

Un biomarcador es definido como “Una característica que es objetivamente medida y evaluada como un indicador de procesos biológicos normales, procesos patogénicos, o respuestas farmacológicas a una intervención terapéutica” por el *Biomarkers definitions working group*. Los biomarcadores son utilizados para entender las fisiopatologías (desórdenes fisiopatológicos) de la enfermedad, respuestas terapéuticas, y para ayudar con el diagnóstico y monitoreo de la enfermedad. Conforme se vayan descubriendo nuevos biomarcadores, y sean validados, se va a incrementar aceleradamente la investigación y el correcto diagnóstico de la EP. Sin embargo, los biomarcadores actuales no están completamente validados y no son óptimos (Foundation, 2021) (Factor & Weiner, 2007).

Tabla 3: Biomarcadores Clínicos de la EP

Prueba	Descripción	Beneficios	Limitaciones
Pruebas motoras	Evaluación del rendimiento en pruebas motoras para cuantificar las afectaciones	Fácil de realizar y cuantificar la respuesta al tratamiento	Pruebas y datos limitados para apoyar el uso clínico. Añade poca información para entender la fisiopatología
Cambios farmacológicos	Examinar la respuesta a la levodopa o agonistas dopaminérgicos para diferenciar la EP de otros síndromes parkinsonianos	Se encuentran disponibles y se realizan rutinariamente. Tienen buena sensibilidad (~ 70%) y especificidad (~ 70%)	Únicamente proveen información para diagnósticos primarios.
Olfato	Evalúa el sentido del olfato, el cual se deteriora a lo largo de la EP	Sencillo y barato de realizar. Se está estudiando para diferenciar la EP de otros parkinsonismos	Aplicaciones limitadas más allá del diagnóstico. (Factor and Weiner 2007)

Los biomarcadores que están siendo estudiados para la EP incluyen:

- **Pruebas clínicas:** Es el método de diagnóstico más utilizado para la EP. En la Tabla 3, tres pruebas son descritas. Estas pruebas son consideradas biomarcadores de la EP e incluyen pruebas motrices, retos farmacológicos y pruebas de olfato.
- **Estudios de imagenología:** Estos biomarcadores incluyen diferentes modalidades de imágenes que proveen información acerca de los procesos neuroanatómicos y fisiopatológicos asociados a la enfermedad. En la Tabla 4, las técnicas utilizadas que están siendo consideradas como biomarcadores son presentadas (Saeed et al., 2017).
- **Pruebas de Sangre y de fluido Cerebroespinal:** se cree que estos métodos pueden incrementar el entendimiento de la EP. Sin embargo, estos biomarcadores tienen poco impacto como herramientas para diagnosticar o que provean métricas terapéuticas. Se necesita más investigación en este campo.
- **Estudios genéticos:** Son indicadores de los procesos patogénicos, y pueden ser considerados biomarcadores de la EP. Aun cuando los factores genéticos no juegan un rol claro en la EP, algunos genes han sido asociados con la enfermedad. Existe un creciente interés en esta área de investigación.

Tabla 4: Las técnicas de imagen más utilizadas como biomarcadores para la EP. (Factor and Weiner 2017, Regnault et al. 2019)

Modalidad de la imagen	Biomarcador
IRMe (T1,T2) Resonancia magnética con tensor de difusión (DT-MRI) Espectroscopia de resonancia magnética de protones	Análisis morfométrico volumétrico basado en vóxeles. Lesión del tracto de la sustancia blanca e integridad microestructural. Cuantificación de metabolitos cerebrales que contienen protones.
Trazador de yodo-123 fluoropropilo combinado con tomografía computarizada por emisión monofotónica (SPECT) Tomografía por emisión de positrones Sonografía transcraneal	Evaluación de la integridad nigroestriatal y de la perfusión cerebral. Aumento de las funciones nigroestriatales, metabolismo de la glucosa, amiloide, imagen molecular de tau y neuroinflamación. Medir la ecogenicidad de la sustancia negra y del núcleo lentiforme.
DaT-SPECT con 123I ioflupano	Examinar la pérdida de dopamina y la progresión de la enfermedad. La interpretación de las imágenes no es sencilla

2.4 Desórdenes de la voz, habla y deglución

Los desórdenes de la voz se manifiestan en más del 90% de los pacientes de la EP durante la primera etapa de Braak. Estas afectaciones de la voz han sido reconocidas como uno de los primeros síntomas prodrómicos de la EP. Algunas de estas afectaciones incluyen la disfonía (uso defectuoso de la voz), hipofonía (volumen reducido), monótono (reducción del rango de tono), disartria (dificultad con la articulación de sonidos y sílabas) e imprecisiones en la articulación hipoquinética, son algunas de las principales características de la voz de los pacientes con EP (B. E. Sakar, M. E. Isenkul, C. O. Sakar, et al., 2013).

Los desórdenes de voz y del habla en la EP han sido atribuidos a las características físicas de la enfermedad, especialmente a la rigidez muscular. Sin embargo, los mecanismos que se encuentran detrás de la fisiopatología específica aún no son claros, y los diferentes estudios apuntan a la activación muscular y los déficits en el área de la Gaita sensoriomotora (alteración de la inhibición del sistema nervioso central), señales internas, amplitud de las señales de voz y la autorregulación de los esfuerzos y producciones vocales.

La naturaleza de las anomalías laríngeas ha sido investigada con técnicas de electromiografías (grabaciones de las señales eléctricas en los músculos y las células motoras). Estos resultados indican la reducción de conexiones neuronales con los músculos de la laringe, o actividad anormal y elevada en la laringe. Estos resultados muestran incongruencias con la teoría que sugiere que la debilidad vocal,

inhabilidad para deglutir y la hipofonía en la EP están relacionadas con la rigidez (Factor & Weiner, 2007).

La producción de consonantes y los problemas de fluidez en el habla pueden estar relacionados al control de las articulaciones. Estudios cinemáticos y electromiográficos de movimientos orofaciales durante el habla de pacientes con la EP pueden indicar una reducción en el tamaño (cerca de la mitad de lo normal), y en la velocidad máxima de los movimientos de la mandíbula, aumento de los niveles de actividad neuromuscular en reposo y de fondo, así como pérdida de reciprocidad entre los grupos musculares agonistas y antagonistas (músculos relacionados con la inhibición del movimiento) en los músculos periorales.

Rosen et al. (Rosen, Kent, & Duffy, 2005) encontró que los pacientes con la EP no tienen diferencias significativas en la caída de la intensidad del habla, comparado con el habla de sujetos sanos en la prolongación de las vocales. Sin embargo, la intensidad de las vocales de los pacientes con la EP si cae más rápidamente que la de los sujetos sanos.

Se ha demostrado que la cirugía de estimulación profunda del cerebro, la cual mitiga algunos de los síntomas de la EP a través de la estimulación específica de algunas regiones del cerebro, puede mejorar la movilidad de las cuerdas vocales. Estos hallazgos se encuentran relacionados con el área del cerebro que se está estimulando, la cual en este caso es el núcleo subtalámico, el núcleo intermedio ventral del tálamo y el globo pálido interno (Arocho-Quinones, Hammer, Bock, & Pahapill, 2017).

2.4.1 Problemas de deglución

Cerca del 90% de los pacientes con la EP desarrollan problemas de deglución (dysphagia). Estos desórdenes suelen aparecer dentro de los primeros síntomas de la enfermedad. Los problemas de deglución incluyen dificultades con la movilidad lingual, la formación del bolo, y la iniciación de la deglución, así como retrasos en las respuestas de la faringe y reducción en la contracción de la faringe. Sin embargo, estudios neuro farmacológicos y neuroquirúrgicos han resultado efectivos para controlar los problemas del habla y la deglución. Fisiológicamente, la red de deglución es influenciada por regiones altas del cerebro, incluyendo la cortical y subcortical. Estudios de IRMf de la deglución de pacientes con la enfermedad y participantes sanos, demuestran que las regiones de activación incluyen el núcleo ambiguo, el córtex sensorimotor, el córtex premotor, el córtex inferior frontal, el cortex parietal posterior, el córtex insular, el giro temporal superior, el cerebelo, el putamen, el globo pálido, el tálamo, el giro cingulado anterior, el gyrus temporal superior y la SN. Muchas de estas regiones del cerebro también son activadas durante el proceso de habla normal (Factor & Weiner, 2007).

2.5 Neuropatología

Los pacientes con la EP sufren diferentes cambios en la estructura, funcionamiento y activación de las regiones del cerebro. En esta sección, algunos de los cambios más relevantes son mencionados, así como la función de la dopamina y la importancia de este neurotransmisor en la degeneración del cerebro de los pacientes con la EP.

2.5.1 Histopatología

La histopatología (estudio de las enfermedades de los tejidos) de la EP se caracteriza clásicamente por la pérdida de neuronas dopaminérgicas en la región de la SN en el cerebro. La mayoría de las afectaciones que producen parkinsonismos tienen en común la deficiencia dopaminérgica en el área nigroestriada del cerebro, aún si se tienen diferentes patologías moleculares. Las neuronas que suelen degenerarse en pacientes con la EP, incluyen los Cuerpos de Lewis (CL), los cuales son la principal característica histológica de la EP, y pueden estar asociados a un rango de síndromes clínicos como la demencia, la psicosis, la disautonomía (condición médica que afecta la autonomía del sistema nervioso). La distribución de los CL en el sistema nervioso central y periférico está mucho más esparcido que únicamente la substantia nigra (Factor & Weiner, 2007).

2.5.2 Cambios estructurales en el cerebro

Patología Macroscópica: Las afectaciones en las regiones externas del cerebro de los pacientes con la EP es imperceptible, con ligera atrofia frontal en algunos casos. No existe una atrofia significativa en el tronco encefálico que pueda llegar a ayudar a los médicos a conducir un diagnóstico diferencial clínico de los desórdenes parkinsonianos. Sin embargo, agrandamiento ventricular leve, particularmente en el cuerno frontal ha sido reportado. En algunos casos, coloración parecida al óxido en la región del globo pálido puede ser notada. Secciones del tronco encefálico pueden revelar pérdida del pigmento normalmente oscuro de la región de la SN y los Locus Ceruleus (LC). La pérdida de este pigmento está relacionada con la pérdida de neuronas.

Patologías Microscópicas: Además de las regiones de la SN y el LC, la pérdida de neuronas es aparente en el núcleo basal de Meynert y en el núcleo motor dorsal del vago. En todas estas regiones, la pérdida neuronal es acompañada de CL así como procesos neuríticos inmunorreactivos de α -sinucleína (Factor & Weiner, 2007).

2.5.3 Ganglios basales

Las características motoras de la EP, son el resultado de las funciones de los ganglios basales que han sido alteradas. Los ganglios basales incluyen el núcleo caudate y el putamen (que en conjunto

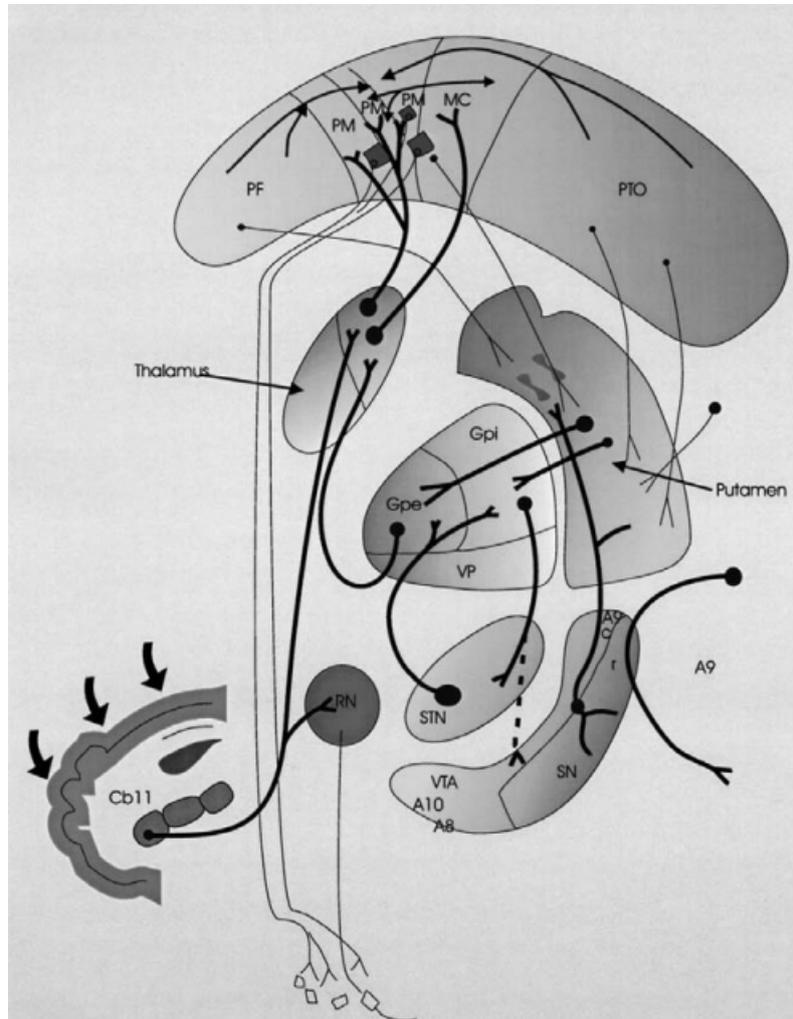


Figura 4: Representación esquemática del sistema motor extrapiramidal en un corte transversal. Las líneas gruesas muestran los caminos entre el Pálido-tálamo-cortical y el cerebelo-tálamo-cortical. También en líneas gruesas se muestran las entradas al globo pálido desde el cuerpo estriado, el putamen, y el ciclo pálido-subtálamo-palidal. MC, corteza motora; PM, corteza premotora; PF, corteza prefrontal; PTO, asociación parieto-temporo-occipital con la corteza; GP, globos pálidos; SN, sustancia nigra; AQ, células dopaminérgicas que comprenden la sustancia nigra compacta; Núcleo Subtalámico (STN); RN núcleo rojo; VP pálido ventral; VTA, área tegmental ventral; Cb11, cerebelo.

Figura obtenida de Factor & Weiner 2007.

conforman el cuerpo estriado), el Globo Pálido Pars Interna (GPi), el Globo Pálido Pars Externa (GPe) el STN, y los Substantia Nigra pars Reticulata (SNr) y Substantia Nigra Pars Compacta (SNc). Otras estructuras que se integran en los circuitos de los ganglios basales y juegan un rol importante en la función extrapiramidal incluyen la corteza premotora, la corteza primaria y el tálamo. Los principales aferentes a los ganglios basales que surgen de los grupos neuronales extrínsecos, así como las vías desde los ganglios basales a los grupos neuronales externos se muestran en la Figura 4.

2.5.4 Rol de los ganglios basales en el control motor

En términos funcionales, los movimientos voluntarios se inician a nivel cortical del circuito motor con salidas simultáneas hacia el tronco encefálico y la médula espinal, así como hacia múltiples puntos subcorticales, incluyendo el tálamo, putamen y STN.

Se cree que los ganglios basales pueden desempeñar un puesto en especificar la amplitud y velocidad de los movimientos. Para lograr escalamiento y terminación de los movimientos, la salida del cuerpo estriado primero inhibiría poblaciones neuronales específicas en las neuronas GPi/SNr a través de entradas sobre vías indirectas, lo que conduce a una inhibición del movimiento en curso. Otra función motora que se piensa está relacionada con los ganglios basales, incluye la autoiniciación de movimientos, el aprendizaje de movimientos y secuencias de movimientos. Los estudios Tomografía de Emisión de Positrones (PET) han mostrado un alto nivel de correlación entre la velocidad de los movimientos y la activación de los ganglios basales, los cuales se pierden con la EP.

2.5.5 Dopamina

La DA, norepinefrina y epinefrina, son catecolaminas (neurotransmisores monoaminas) sintetizadas de los mismos precursores, el aminoácido L-tirosina y su producto hidroxilado L-Dopa (el cual es usado como tratamiento para los pacientes con la EP). La DA es un neurotransmisor del sistema nervioso central que juega un papel importante en la regulación del comportamiento y funciones motoras; así mismo, juega un papel importante en la regulación del comportamiento y funciones motoras, y juega un papel central en muchas enfermedades neuropsiquiátricas. También es importante para el sistema cardiovascular, los riñones, la modulación de la presión arterial, el balance de sodio y funciones renales independientes del sistema neuro-dopaminérgico.

El sistema dopaminérgico está formado por tres circuitos neuronales principales: el nigroestriatal, el mesolímbico/mesocortical, y el tubular-infundibular. La DA está involucrada con el sistema nervioso central, control del estado de ánimo, movimiento, cognición, motivación y el sistema de recompensas. En el sistema periférico del sistema nervioso, la DA controla la función de los riñones, el flujo de sangre renal, y funciones cardiovasculares (Kudo & Fujii, 2012).

Durante las primeras etapas de la EP, el agotamiento de la DA sucede principalmente en la región sensorimotora en el cuerpo estriado (la región postcomissural del putamen), lo que sugiere que el circuito motor está involucrado en la fisiopatología subyacente a las anomalías motoras. La reducción en la activación de receptores de DA en la región estriada, puede resultar en un aumento de cambios en la actividad en los ganglios basales. Muchos de los cambios fisiopatológicos en los ganglios basales, que resultan de la pérdida de DA en la región estriada, se interpretan como el resultado de la reducción de activaciones en los receptores de DA en la región estriada.

Se considera que la pérdida de DA en la región estriada debe de haberse reducido en un 70%-80% antes de que los síntomas motores de la EP se vuelvan aparentes. En la Tabla 5, una comparación entre la concentración de DA y las diferentes regiones del cerebro de participantes control y pacientes con la EP es presentada (Factor & Weiner, 2007).

Tabla 5: Concentración de DA en pacientes con la EP y sujetos sanos es comparada en las diferentes regiones de interés (Factor and Weiner, 2007).

Región	Control	pacientes con la EP
Putamen	5.06 ± 0.39	0.14 ± 0.13
Núcleo Caudado	4.06 ± 0.47	0.2 ± 0.19
Núcleo accumbens	3.79 ± 0.82	1.61 ± 0.28
Circunvolución paraolfatoria	0.35 ± 0.09	< 0.03
Hipotálamo lateral	0.51 ± 0.08	< 0.03
Substantia nigra	0.46 ± *	0.07 ± *

Resultados expresados en *ng/mg*

2.5.6 Sistema Nigrostriatal y síntomas de la Enfermedad de Parkinson

La deficiencia de DA en la región estriada juega un papel fundamental en el desarrollo de los síntomas motores de la EP. Sin embargo, no está claro si estos síntomas son el resultado de la falta de dopamina únicamente, o si otro neurotransmisor está involucrado también. Se ha encontrado una alta correlación entre la severidad de la aquinesia (el deterioro global de la iniciación del movimiento), reducción de dopamina en la región estriada y el grado de pérdida neuronal en la SN. La aquinesia es más severa en el lado contrario al cual se presenta la mayor pérdida neuronal de la SN.

2.6 Discusión

La búsqueda de biomarcadores en la EP es de gran interés para desarrollar nuevos tratamientos y herramientas de diagnóstico. Imágenes médicas como la IRMe y IRMf, son dos candidatas para encontrar

alternativas al DaTscan. Sin embargo, es necesario contar con más investigación e interpretación de los resultados obtenidos para poder ser aceptadas como biomarcadores. Grabaciones de voz y los problemas para deglutir, como se describió anteriormente, involucran diferentes regiones del cerebro. Considerando que estos problemas comienzan durante las primeras dos etapas de la enfermedad, el análisis de estas regiones presenta una oportunidad para detectar la enfermedad de manera proactiva.

Capítulo 3 Metodología y antecedentes teóricos

Las señales biomédicas como las IRMe y las grabaciones de voz requieren múltiples pasos de preprocesamiento antes de poder ser utilizadas para detectar, o dar información acerca de la progresión de la EP. Los pasos de procesamiento de las señales se pueden resumir como (1) preprocesamiento de las señales, (2) extracción de características, (3) clasificación y (4) interpretación clínica de los resultados, como es mostrado en la Figura 5.

El preprocesamiento de las señales generalmente involucra limpiar y preparar los datos para poder extraer características útiles, sin la presencia de ruido. La normalización de la señal, el filtrado de ruido, la segmentación, entre otras técnicas, pueden incrementar dramáticamente la exactitud con la que se realizan predicciones (Akçay & Oğuz, 2020) (Gargouri et al., 2018). La extracción de características por otro lado, depende específicamente del tipo de señal que está siendo analizada. Sin embargo, características estadísticas de frecuencia, espaciales y de multi-resolución, son comúnmente utilizadas. La selección de características ayudar a reducir la complejidad del modelo y evitar problemas como la alta correlación entre las características que puede llegar a reducir el rendimiento de los clasificadores. Los algoritmos de clasificación son entrenados y probados para cada problema en específico, y los que obtienen los mejores resultados con métricas como la exactitud, la sensibilidad, y la precisión entre otras, son los que se seleccionan. Finalmente, la interpretación clínica de los resultados es un paso importante para evitar resultados binarios que resulten poco claros para los médicos, y no permitan justificar el diagnóstico o el inicio de los tratamientos.

Cada uno de estos pasos se explican a detalle tanto para grabaciones de voz como para las señales de IRMe en el resto de este capítulo.

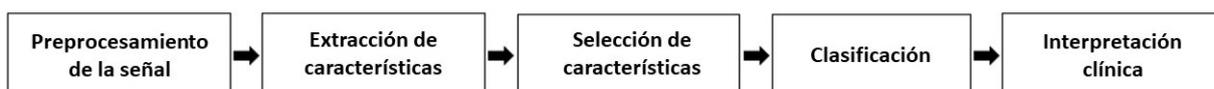


Figura 5: Representación general de la metodología para la detección de la EP utilizando señales biomédicas y algoritmos de aprendizaje automático.

3.1 Descripción de las bases de datos

3.1.1 Grabaciones de voz provenientes de la facultad de Medicina de la Universidad de Estambul

La base de datos que se utiliza en este trabajo para detectar la EP por medio de grabaciones de voz es proporcionada por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019). Esta base de datos fue publicada por primera vez en 2019, y contiene grabaciones de 188 pacientes con la EP, incluyendo 107 pacientes masculinos y 81 pacientes femeninos con un rango de edad de entre 33 y 87 años, y 64 sujetos sanos, incluyendo 23 sujetos masculinos y 41 sujetos femeninos de entre 41 y 82 años de edad. Los datos se obtuvieron en el departamento de Neurología en la *Facultad de medicina Cerrahpasa, en la Universidad de Estambul*. Las grabaciones se obtuvieron con micrófonos que grabaron con una frecuencia de muestreo de 44.1 kHz, mientras el paciente pronunciaba tres iteraciones de la vocal sostenida /a/. Esta vocal fue escogida debido a que los cambios como las articulaciones y la reducción de espacio en la vocal de los pacientes con la EP pueden ser identificadas con la pronunciación de la vocal (Bang, Min, Sohn, & Cho, 2013). Todos los participantes del estudio firmaron una carta de consentimiento informado voluntariamente. Esta base de datos contiene un total de 756 grabaciones en total, 546 provenientes de pacientes con la EP y 192 de sujetos control.

3.1.2 Base de datos de imágenes de Resonancia Magnética Estructural del Parkinson's Progression Marker Initiative

Las IRMe de pacientes con la EP utilizadas en este trabajo se obtuvieron de la base de datos del PPMI. El PPMI es un estudio observacional que se encuentra disponible desde el año 2010, y el cual busca definir y medir de mejor manera la EP. El PPMI contiene datos longitudinales de 14,000 pacientes y sujetos control, que provienen de más de 11 países y 33 clínicas. El PPMI está patrocinado por la fundación de Michael J. Fox (Foundation, 2021). Las IRMe de 226 pacientes masculinos y 104 pacientes femeninos con la EP, así como 86 sujetos control de sexo masculino y 64 sujetos control de sexo femenino fueron obtenidas de la PPMI para este trabajo. La edad de los participantes en el estudio se varía en el rango de los 30 a los 89 años. Las IRMe utilizadas son de peso-T1 y fueron generadas con un campo magnético de 1.5 o 3 Teslas, dependiendo del escáner utilizado. El tiempo de escaneo fue de 20 a 30 minutos, y los resultados fueron imágenes en una secuencia tridimensional compuesta de tres vistas (sagital, axial y coronal), con cortes de un 1.5mm de grueso.

3.2 Preprocesamiento de los datos

Las señales biomédicas son comúnmente utilizadas para detectar enfermedades, monitorear la salud, y comparar cambios longitudinales en los pacientes. Sin embargo, estas señales pueden contener ruido

causado por artefactos fisiológicos, ruido electrónico proveniente de los instrumentos, entre otras causas. Cuando se trabaja con estas señales, es importante considerar la variabilidad de los instrumentos que se utilizan para obtener las muestras, la posición de los pacientes, las alteraciones en los sistemas, movimientos involuntarios de los sujetos, y otros factores que pueden causar ruido en la señal durante la medición y examinación. Por estas razones, el preprocesamiento de las señales juega un papel muy importante para obtener buenos resultados en los estudios subsecuentes (Chahid, 2020).

En el caso de las grabaciones de voz, algunas de las técnicas de preprocesamiento más comunes incluyen el *enmarcado* (división de la señal en muestras con el mismo tamaño), *ventanas* (para reducir el efecto de las desbordes durante los análisis de frecuencia), *activación por medio de la detección de voz* (identificar los periodos de silencio, de habla, y pausas del habla), *normalización* (reducir la variabilidad sin afectar la habilidad de generalizar de las características extraídas), *reducción de ruido* (reducir la presencia de valores atípicos), entre otros (Akçay & Oğuz, 2020). Sin embargo, las bases de datos de grabaciones de voz de (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019), no mencionan cuales fueron las etapas de preprocesamiento, aparte de dividir las grabaciones en pequeñas ventanas de 25 milisegundos utilizando la *ventana de Hamming* para procesar las señales como si fueran estacionarias, resultando en 9,702,000 muestras por grabación.

Una ventana es una función matemática la cual regresa cero a todos los valores que se encuentran fuera de un intervalo. Esta función puede ser multiplicada por una señal, y de esta manera se obtienen únicamente los valores deseados dentro del segmento de interés (Podder, Khan, Khan, & Rahman, 2014). La función más simple es una ventana rectangular, sin embargo, este tipo de función puede resultar en ruido de alta frecuencia en los bordes. Para evitar este problema, se utiliza una secuencia de ventanas de Hamming, dada por la ecuación 1, la función no causal de la ventana dada por la ecuación 2, y el espectro de la ventana mostrado en la ecuación 3.

$$W(n) = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right) \quad \text{for} \quad -\frac{N-1}{2} \leq n \leq \frac{N-1}{2} \quad (1)$$

Donde N es el número de muestras, $\alpha = 0.54$ y $\beta = 1 - \alpha = 0.46$.

$$W_H = W_R(n) \left[0.54 + 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \right] \dots \quad (2)$$

Donde $W_R(n)$ es la función rectangular de la ventana.

$$W_H(e^{j\omega\tau}) = 0.54 \frac{\sin\left(\frac{\omega N\tau}{2}\right)}{\sin\left(\frac{\omega\tau}{2}\right)} + 0.46 \frac{\sin\left(\frac{\omega N\tau}{2} - \frac{N\pi}{N-1}\right)}{\sin\left(\frac{\omega\tau}{2} - \frac{\pi}{N-1}\right)} + 0.46 \frac{\sin\left(\frac{\omega N\tau}{2} + \frac{N\pi}{N-1}\right)}{\sin\left(\frac{\omega\tau}{2} + \frac{\pi}{N-1}\right)} \dots \quad (3)$$

En el caso de las IRMe, es muy importante preparar los datos antes de comparar los grupos control y los grupos con pacientes con la EP. La variabilidad de los escáneres con los que se toman las

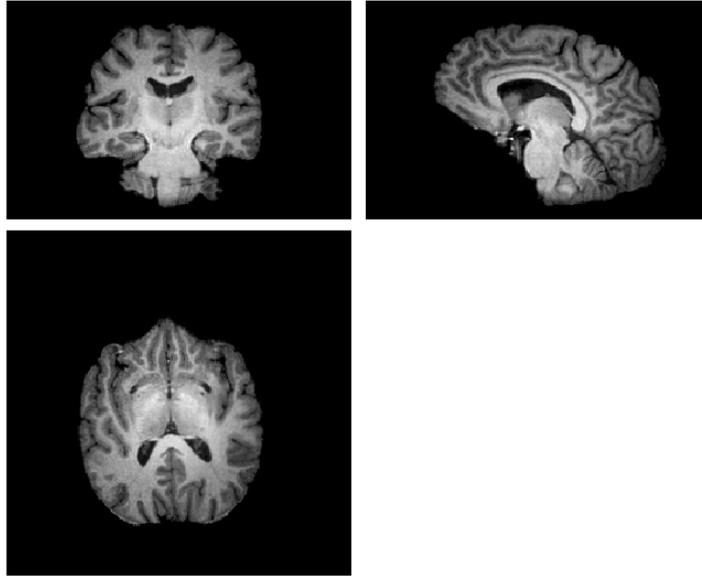


Figura 6: La imagen en la esquina superior izquierda muestra la vista coronal del cerebro, la esquina superior derecha muestra la vista sagital y la imagen en la esquina inferior izquierda muestra la vista axial.

imágenes, el campo magnético, la posición del paciente, el ruido de las imágenes, normalización y plantilla estándar, entre otros problemas, deben de ser tomados en cuenta para obtener resultados válidos. Estos pasos de preprocesamiento para las IRMe se describen en la siguiente subsección.

3.2.1 Preprocesamiento de Imágenes de Resonancia Magnética Estructural

Todos los pasos descritos en esta subsección fueron realizados con el software de libre uso SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>), en el lenguaje de programación de MatLab (R2020b; MathWorks, Natick, MA, USA), y la caja de herramientas de CAT12, desarrollada por Christian Gaser y Robert Dahnke (Hospital de la Universidad de Jena, departamento de psiquiatría y neurología).

(1) Formato de los archivos: Las imágenes para cada sujeto de la base de datos fueron proporcionadas por el PPMI en múltiples archivos con el formato Digital Imaging and Communications in Medicine (DICOM), el cual es utilizado para almacenar imágenes médicas como *Ultrasonidos*, *IRM*, Y *Tomografías computarizadas*. Para trabajar con las imágenes como una reconstrucción en 3-dimensiones de los cortes sagitales, coronales y axiales, las imágenes fueron convertidas al formato *Nifti*. Las imágenes sin procesar en las diferentes vistas son mostradas en la Figura 6. En este formato, los datos volumétricos y los metadatos pueden ser accedados desde un único archivo *.NII*.

(2) Selección del origen: Una vez que las imágenes se encuentran en el formato correcto, se selecciona la posición de origen de todas las imágenes. La convención es seleccionar la región de la *comisura anterior* (anterior commissure) como el origen. El origen de cada archivo *Nifti* debe ser seleccionado

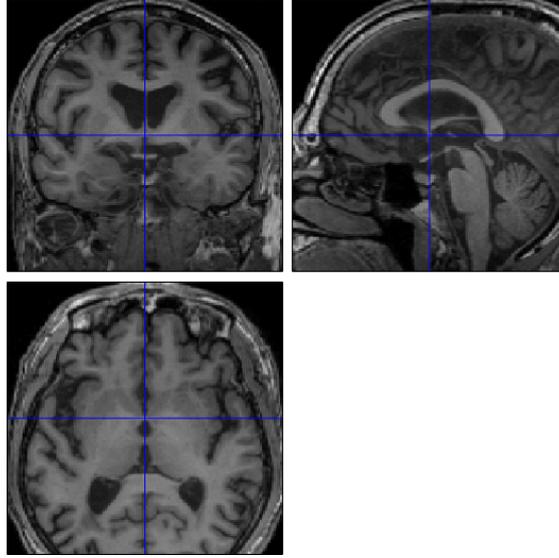


Figura 7: Orígen seleccionado manualmente en el software de SPM12 en una IRMe de un paciente con la EP.

manualmente con el software de SPM12. La comisura anterior en cada una de las tres vistas se muestra en la Figura 7.

(3) Alineamiento de la imagen: La orientación de la cabeza de los pacientes dentro del escáner de Resonancia Magnética no siempre es exactamente igual. Todas las imágenes deben ser alineadas con una plantilla común para poder comparar grupos grandes de pacientes. Para lograr esto, se realiza un proceso de deformación de la imagen. Este proceso es una alteración de la forma de la IRMe reposicionando los vóxeles en la imagen, de acuerdo a $s_i = T(q_i, p)$, donde la posición de un vóxel está dada por q_i y se le asigna un nuevo valor s_i , dependiendo del conjunto de parámetros de transformación p . Para el caso de una *transformación afín*, la cual es una transformación lineal seguida de un proceso de traslación, y el cual preserva colinealidad entre los puntos, p corresponde a los ángulos de la rotación de los objetos y los escalares son parte del proceso de escalamiento (aumento o disminución) y compensaciones para el desplazamiento espacial del objeto. La transformación afín tiene la forma de la ecuación 4.

$$\begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} \quad (4)$$

El alineamiento de las imágenes consiste en estimar los parámetros p de transformación que optimizan el proceso para igualar el conjunto de vóxeles transformados s_i a otro conjunto de referencia de vóxeles r_i donde ambos conjuntos son extraídos de un objeto observado I_1 dada la transformación

$T(I_1, p)$ y el objeto de referencia I_2 , respectivamente. El objeto de referencia utilizado para ese trabajo es una versión modificada del atlas probabilístico de tejidos de la ICBM, el cual se deriva de 452 imágenes T1, las cuales son alineadas, se corrige la falta de homogeneidad y son segmentadas en materia blanca, materia gris y fluido cerebroespinal (Mazziotta et al., 1995).

Para realizar el alineamiento de las imágenes, una transformación afín debe ser aplicada iterativamente al objeto I_1 , lo cual resulta en la minimización de la *función de costo* definida como la *distancia euclídeana* entre los vóxeles de los diferentes objetos de acuerdo a $C(p) = \sum_{i=1}^N [T(q_i, p) - r_i]^2$, donde N es el número de vóxeles en el objeto. El método del *gradiente descendente* o método de *Levenberg-Marquardt* para calcular Δp en $p^k + \Delta p$ puede ser utilizado para minimizar la función de costo $C(p)$ durante la transformación afín. La transformación afín utilizada en el software CAT12, incluye regularización, lo que hace el método más robusto penalizando estiramientos o encogimiento excesivo de las imágenes (Gaser & Dahnke, 2016).

Además de alinear las imágenes con la plantilla del ICBM, las intensidades de las imágenes resultantes deben de ser normalizadas. La normalización de las imágenes se realiza para mitigar las variaciones intra e inter escaneado entre las IRMe de diferentes pacientes. Algunos de los métodos utilizados para normalizar las imágenes en escala de grises incluyen un máximo para todas las imágenes, y valores dinámicos limitados por $\mu \pm 3\sigma$ (Collewet, Strzelecki, & Mariette, 2004).

(4) Segmentación de IRMe: En las IRMe, se consideran tres tejidos: materia gris, materia blanca y fluido cerebro espinal. La segmentación de IRMe consiste en clasificar los píxeles de cada uno de estos tejidos. Esto se realiza calculando las probabilidades que tiene cada píxel de pertenecer a uno de los tejidos. El modelo de mezcla Gaussiana logra este objetivo construyendo un mapa de probabilidades de 3 dimensiones, modelando la función de costo como una mezcla de k funciones Gaussianas (Bandhyopadhyay & Paul, 2012). El número k se elige igual al número de tejidos, el cual para este trabajo es de tres. La probabilidad de un pixel con intensidad x_i dado que pertenece a k , es dada por $P(x_i|k, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}}$, donde la función Gaussiana para cada tejido está caracterizada por μ_k (valor promedio) y σ_k (varianza). La probabilidad a priori de que un vóxel pertenezca a un tejido, dada la proporción de vóxeles que pertenecen a la Gaussiana está dada por: $P(k|\gamma_k) = \gamma_k$, donde γ es una de la serie de pesos, donde $\sum_1^k \gamma_k = 1$. Un vóxel con intensidad x_i pertenece al tejido k_{th} con la probabilidad conjunta más grande, dada por: $P(x_i, k|\mu_k, \sigma_k, \gamma_k) = P(x_i|k, \mu_k, \sigma_k)P(k|\gamma_k)$. Para obtener la probabilidad de un vóxel de valor x_i , dados los parámetros μ, σ and γ , el conjunto del tejido k es añadido $P(x_i|\mu, \sigma, \gamma) = \sum_1^k P(x_i, k|\mu_k, \sigma_k, \gamma_k)$. Para aprender los parámetros de Mezcla de gaussianas (MOG), se asume que todos los elementos son independientes, y que el dataset $x_1, x_2, x_3, \dots, x_I$ es derivado de acuerdo a la ecuación 5.

$$P(x_1, \dots, x_I | \mu, \sigma, \gamma) = \prod_1^I P(x_i | \mu, \sigma, \gamma) P(k | \gamma_k) = \prod_1^I \left[\sum_1^k \frac{\gamma_k}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \right] \quad (5)$$

La función que se va a maximizar se obtiene tomando el logaritmo de la función de costo mostrada en la ecuación 5. Esto resulta en una función monótonica, mostrada en la ecuación 6.

$$\varepsilon = - \sum_1^I \log \left[\sum_1^K \frac{\gamma_k}{\sqrt{2\pi}\sigma_k} \exp^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \right] \quad (6)$$

En el software CAT12, un método de segmentación unificada es propuesto por (Ashburner & Friston, 2005). Este método combina el MOG e incorpora una variación suave de intensidad y registro no lineal con mapas probabilísticos de tejidos. Esto es realizado combinando el registro con las plantillas estándar y la clasificación del tejido (*A Computational Anatomy Toolbox for SPM*, n.d.).

(5) Suavizado El último paso del preprocesamiento es el suavizado de la imagen. Se suavizan las IRMe y poder realizar el análisis estadístico el cual puede incrementar los niveles de señal-ruido y la sensibilidad de las señales de tamaños y formas específicas. Otra razón para realizar un suavizado de las imágenes, es filtrar los resultados de los artefactos que se puedan generar durante las etapas de registro y segmentación. Sin embargo, el suavizado puede generar líneas poco definidas en los bordes (Mikl et al., 2008). Las imágenes 3D se convolucionan con una Kernel gaussiano con anchura máxima a la mitad (FWHM) del kernel Gaussiano definido en milímetros. Para entender el FWHM, necesitamos considerar la distribución normal de la ecuación 7.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] \quad (7)$$

donde σ es la desviación estándar y x_0 es el valor esperado. La relación entre σ y la FWHM se muestra en la ecuación 8.

$$FWHM = 2\sqrt{2\ln 2}\sigma \approx 2.355\sigma \quad (8)$$

En la Figura 8, se muestra el resultado del alineamiento, segmentación y suavizado. En la imagen de la izquierda, podemos ver la imagen original una vez que fue convertida a un formato NIfTI. La segunda imagen muestra la alineación y segmentación de las IRMe, utilizando únicamente los tejidos de la materia gris. Finalmente, la imagen es suavizada como se puede ver en la imagen de la derecha.

3.3 Morfometría Basada en Vóxeles

La morfometría basada en vóxeles o VBM por sus siglas en inglés, es una herramienta estadística que permite encontrar diferencias en la sustancia gris de las IRMe realizando un comparación vóxel a vóxel

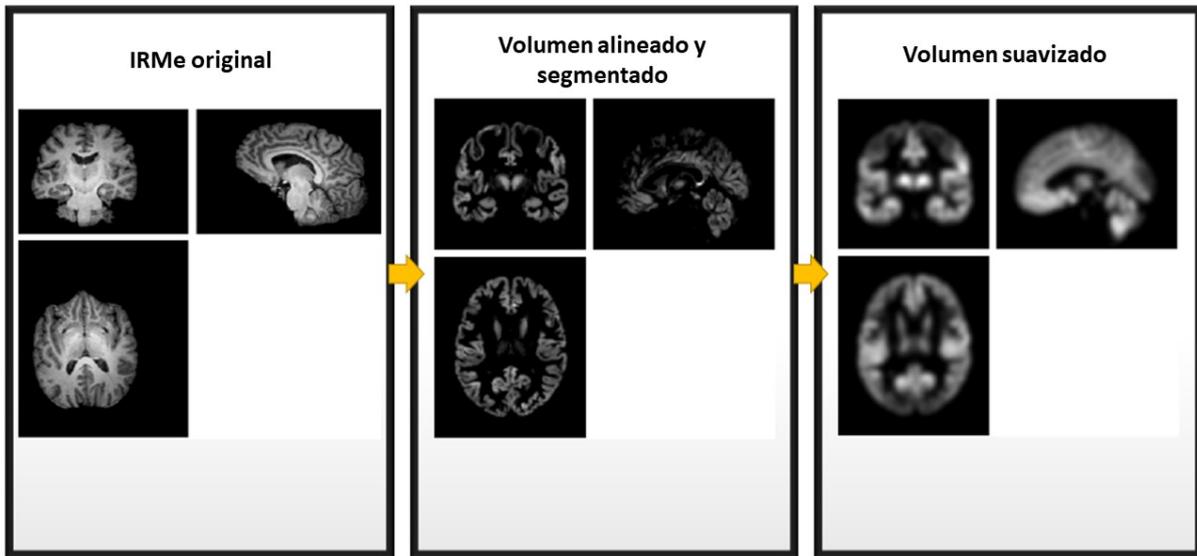


Figura 8: Resultados de las etapas de preprocesamiento de las IRMe de los pacientes.

entre dos grupos de imágenes. El proceso involucra el preprocesamiento de las imágenes, previamente descrito en este capítulo, que incluye la normalización, segmentación y suavizado de las IRMe. Una vez que las imágenes se encuentran en la misma escala y espacio (lo cual se logra generalmente utilizando una plantilla), los vóxeles son comparados con pruebas estadísticas paramétricas utilizando modelos lineales estándar. El objetivo es encontrar las regiones de interés con diferencias significativas entre ambos grupos. Para probar la hipótesis de similitud entre vóxeles, se realiza la prueba de student para cada vóxel. Esta prueba estadística asume que se cuenta con una distribución normal de los datos, y se utiliza la teoría de campos aleatorios para corregir el número de falsos positivos. El resultado del VBM es un *mapa-t* el cual contiene los valores t para cada vóxel, y muestra cuáles son los vóxeles que contienen diferencias relevantes entre los grupos (Ashburner & Friston, 2007).

3.4 Extracción de características

Cuando trabajamos con señales biomédicas, la señal original puede contener ruido, lo que dificulta que el clasificador pueda aprender y reconocer patrones en los datos ruidosos. Por esta razón, es importante extraer información relevante de la señal original para reducir la dimensionalidad de los datos y proveer características ya procesadas para entrenar el algoritmo de aprendizaje automático. Entre mejor información se le da al clasificador, tendrá un mejor rendimiento. Algunas de las características más comunes que se extraen de señales biomédicas incluyen información estadística, características de frecuencia, relaciones espaciales y análisis multiresolución de las señales.

Es importante considerar la dimensionalidad de la señal para seleccionar la característica que se va a extraer. Para este trabajo, se consideran dos tipos diferentes de señales. Con las grabaciones de voz, las señales son unidimensionales, mientras que con las IRMe, se necesita procesar datos tridimensionales. Las características que se extraen de las diferentes señales utilizadas en este trabajo, son descritas en esta sección.

3.4.1 Características base de las señales de grabaciones de voz

Las Características Base (CB), son algunas de las características más utilizadas para los estudios del análisis de la EP. Algunas de estas características incluyen el Jitter, Shimmer, parámetros de los armónicos, Periodo de recurrencia Densidad Entropía (RPDE), parámetros de las frecuencias fundamentales, Análisis de fluctuación de tendencia (DFA) y Entropía del periodo de tono (PPE), entre otros. El Jitter y el Shimmer son medidas que tienen que ver con las variaciones ciclo a ciclo de la frecuencia fundamental y la amplitud respectivamente. Es común utilizar estas medidas para encontrar variaciones entre los diferentes hablantes, pero también puede utilizarse para detectar patologías y defectos de la voz, y en este caso, para detectar la EP (Farrús, Hernando, & Ejarque, 2007). Algunas de las medidas del Jitter incluyen el Cociente de Perturbación en un Periodo de Cinco Puntos en el Tiempo (PPQ5) relativo y el absoluto (Teixeira, Oliveira, & Lopes, 2013). Algunos de los parámetros más relevantes para este trabajo se describen a continuación.

Jitter (absolute): También conocido como *jitta*, representa la diferencia absoluta promedio entre dos periodos consecutivos, dividida por el periodo promedio como se muestra en la ecuación 9.

$$jitt(absolute) = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (9)$$

Jitter(relative): También conocido como *jitt*, esto representa la diferencia absoluta promedio entre los periodos consecutivos, divididos por un periodo promedio como se muestra en la ecuación 10.

$$jitt(relative) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (10)$$

Jitter (PPQ5): Se refiere al cociente de perturbación sobre un periodo de cinco puntos. Se calcula la diferencia del promedio absoluto entre un periodo y el promedio de los cuatro más cercanos, divididos por el periodo promedio, como se muestra en la ecuación 11.

$$jitt(ppq5) = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |T_i - (\frac{1}{5} \sum_{n=i-2}^{i+2} T_n)|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (11)$$

donde T_i es el periodo extraído en el ciclo $i4$ y N es el número de ciclos que se está analizando.

Otra característica importante para la detección de la EP en grabaciones de voz, es el DFA. Esta característica detecta ruido turbulento en la señal de voz, e incrementa el ruido nasal (Little, McSharry,

Roberts, Costello, & Moroz, 2007). Este se calcula sobre una serie de tiempo $x(n); n = 1, \dots, N$, y requiere diferentes pasos. Primero, la versión acumulada de la serie de tiempo de media cero, se calcula con la ecuación 12.

$$y(m) = \sum_{n=1}^m [x(n) - x] \quad (12)$$

después $y(m)$ es segmentado en k ventanas de tamaño $s = \frac{N}{K}$. En cada ventana, un polinomio (usualmente lineal) y_{LS} se ajusta a $y(m)$ utilizando una regresión lineal de *Mínimos-Cuadrados*. La fluctuación es el promedio del residuo al cuadrado $F(s)$ mostrado en la ecuación 13.

$$F(s) = \sqrt{\frac{1}{N} \sum_{m=1}^N [y(m) - Y_{LS}(m)]^2} \quad (13)$$

La función $F(s)$ es evaluada en diferentes ventanas. La relación $F(s) \propto s^H$, donde H es el exponente de Hurst, busca la auto-similaridad en la serie de tiempo. Un doble logaritmo (“log-log”) $F(s)$ contra s puede ser estimado utilizando una regresión lineal de mínimos-cuadrados (Bryce & Sprague, 2012). Esta pendiente es utilizada para calcular H , donde:

$$\begin{cases} H > 0.5 & x(n) \text{ está positivamente correlacionado} \\ H = 0.5 & x(n) \text{ no está correlacionado} \\ H < 0.5 & x(n) \text{ está anti-correlacionado} \end{cases} \quad (14)$$

3.4.2 Coeficientes Cepstrales de la Frecuencia de Mel

Los CCFM son usados ampliamente en el reconocimiento y análisis de habla y voz. Son una herramienta que responde de manera similar al sistema auditivo humano, el cual no es lineal en las diferentes frecuencias, y no está diseñado para reconocer frecuencias similares. Esta característica se vuelve más pronunciada conforme más alta es la frecuencia (Yujin, Peihua, & Qun, 2010). Para lograr esto, la señal de audio es filtrada con una colección triangular de filtros llamado el *Banco de filtros de Mel* (Sigurdsson, Petersen, & Lehn-Schiøler, 2006). Este banco de filtros procesa de manera diferente las frecuencias bajas de la señal, con las frecuencias altas. Para obtener sus coeficientes, el primer paso es dividir la señal discreta en N -segmentos $x(n)w(n); n = 1, \dots, N$, donde $w(n)$ es una ventana de Hamming. Después, se calcula el Transformada Discreta Rápida de Fourier (TDRF) para cada una de las ventanas con la ecuación 15.

$$X(k) = \sum_{n=0}^{N-1} w(n) x(n) \exp\left(\frac{-j2\pi kn}{N}\right) \quad (15)$$

Donde k corresponde a la frecuencia $f(k) = \frac{kf_s}{N}$, dado que f_s es la frecuencia de muestreo.

Se calcula la energía de cada banco, y el ancho del filtro se obtiene con la ecuación 16 la cual es la relación idiográfica entre la frecuencia de Mel y la frecuencia estándar en Hz.

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (16)$$

Donde la f se mide en Hz, y el logaritmo se utiliza debido a que el sistema auditivo no funciona de manera lineal, si no que por lo contrario, requiere mayor energía en cuanto más alta es la frecuencia para percibirla.

Los CCFM se calculan utilizando la transformada discreta de coseno, mostrada en la ecuación 17.

$$c_{mel} = \sum_{m=1}^{M-1} Y(m) \cos \left(\frac{\pi}{M} \left(m - \frac{1}{2} \right) n \right); n = 1, 2, \dots, M. \quad (17)$$

Donde $Y(m)$ es dada por el logaritmo de la suma ponderada de los coeficientes de Transformada Discreta de Fourier (TDF) utilizando la ecuación 18.

$$Y(m) = \ln \sum_{n=0}^{N-1} |X(k)| H(k, m); m = 1, \dots, M. \quad (18)$$

Las características dinámicas pueden ser extraídas de la señal relacionando los coeficientes CCFM. Esto se logra seleccionando derivadas de tiempo, llamadas características delta (Hossan, Memon, & Gregory, 2010). Las características deltas se calculan con la ecuación 19.

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (C_{i+\theta} - C_{i-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2} \quad (19)$$

Donde d_t es el coeficiente delta en el instante de tiempo t , $c_{t-\theta}$ y $c_{t+\theta}$ y Θ es el tamaño de la ventana delta.

Una consideración para los CCFM, es que son afectados por el ruido de fondo, por lo que estas características funcionan mejor cuando las grabaciones se obtienen en ambientes controlados.

3.4.3 Características basadas en la Transformación de Ondículas Discreta

La transformada de ondículas TO, son funciones matemáticas basadas en un grupo de funciones base y transformaciones que contienen información de una señal. Una de las ideas principales detrás de la TO, es que, para señales representadas en función del tiempo, esta herramienta provee buena resolución tanto de tiempo como de frecuencia, contrario a la transformada de Fourier, la cual únicamente proporciona buena resolución en frecuencia. Esto significa que la mayor parte de la energía de la señal está concentrada en un pequeño grupo de coeficientes. La TO es generada por medio de escalamiento y traslación de una señal llamada la *Ondícula Madre*, mostrada en la ecuación 20 y 21 respectivamente (Burrus, 1997).

$$\psi_{j,k}(n) = \psi(2^j t - k) \quad j, k \in Z \quad (20)$$

$$\phi_{j,k}(n) = \phi(2^j t - k) \quad j, k \in Z \quad (21)$$

Donde Z es el grupo de todos los enteros, y el factor $2^{j/2}$ mantiene una norma constante, independiente de la escala j .

En este trabajo, la función de escalamiento Haar ϕ y la ondícula ψ son usadas para descomponer la señal. La señal puede ser representada como una combinación lineal de la función de escalamiento recorrida en el tiempo $\phi_{j,k}(n)$ y la función de la ondícula mostrada en la ecuación 22.

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t - k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j t - k) \quad (22)$$

La Transformada Discreta de Ondículas (TDO) de la función $f(t)$ es un conjunto de coeficientes de expansión bidimensional: los coeficientes de escalamiento c_k , y los coeficientes de la ondícula $d_{j,k}$. Estos coeficientes se pueden calcular utilizando el producto punto mostrado en la ecuación 23.

$$f(t) = \sum_{j,k} \langle \psi_{j,k}(t), f(t) \rangle \psi_{j,k}(t) \quad (23)$$

El filtrado recursivo y la reducción de muestreo de la señal se calculan para encontrar los coeficientes de expansión utilizando un banco de filtros. El banco de filtros tiene una estructura de árbol, en la cual la señal es dividida en dos partes iguales, una con el espectro de alta frecuencia de la señal d_{j-1} , y otro con el espectro de las frecuencias bajas c_{j-1} . Después, la rama con el espectro de frecuencias bajas se subdivide en dos ramas de la misma señal, una con las frecuencias altas d_{j-2} , otro con las frecuencias bajas c_{j-2} . A cada una de estas etapas se le llama *nivel de descomposición*, y el proceso continúa hasta que la señal no puede ser dividida más. Los coeficientes pueden ser utilizados para describir, analizar, filtrar, y aproximar la señal de tiempo. De estos coeficientes se extraen características de las grabaciones voz.

3.4.4 Transformada de Ondículas con Factor-Q Ajustable

La TOFQA está basada en la TDO, con la diferencia de que el factor-Q del banco de filtros puede ser ajustado de acuerdo al comportamiento de oscilación de la señal que se está analizando, lo que resulta en el cambio de la tasa de muestreo de la señal. Si el comportamiento oscilatorio de la señal es alto, entonces un factor Q alto es más adecuado. Por el contrario, si la señal no presenta comportamientos oscilatorios, un factor Q bajo es preferido (Selesnick, 2011). La TOFQA se calcula con un filtrado iterativo de la señal con dos canales del banco de filtros, en la parte de las frecuencias bajas de la

señal. Un filtro de escalamiento pasa-bajos con parámetro α mantiene el contenido de frecuencias bajas de la señal y está dado por la ecuación 24, mientras que un escalamiento pasa-altas, filtrado con el parámetro β , conserva el contenido de las frecuencias altas de la señal. El parámetro de escalamiento debe satisfacer $0 < \alpha < 1$ y $0 < \beta \leq 1$ para evitar la redundancia en TOFQA, y puede ser calculado con la ecuación 25.

$$Y(\omega) = X(\alpha\omega), \quad |\omega| \leq \pi \quad (24)$$

$$Y(\omega) = \begin{cases} X(\beta\omega + (1-\beta)\pi), & \text{for } 0 < \omega < \pi \\ X(\beta\omega - (1-\beta)\pi), & \text{for } -\pi < \omega < 0 \end{cases} \quad (25)$$

La frecuencia central en el nivel j está dado por el promedio de ω_1 y ω_2 , como se muestra en la ecuación 26.

$$\omega_c = \frac{1}{2}(\omega_1 + \omega_2) = \alpha^j \frac{2-\beta}{2\alpha} \pi \quad (26)$$

Donde $\omega_1 = (1-\beta)\alpha^{j-1}\pi$ y $\omega_2 = \alpha^{j-1}\pi$.

El factor Q es la relación entre la frecuencia central sobre el ancho de banda, y puede ser expresado en términos de β como se muestra en la ecuación 27.

$$Q = \frac{\omega_c}{BW} = \frac{2-\beta}{\beta} \quad (27)$$

Donde BW es el ancho de banda de la respuesta en frecuencia, dado por $BW = \frac{1}{2}(\omega_2 - \omega_1) = \frac{1}{2}\beta\alpha^{j-1}\pi$.

El factor Q es independiente del nivel j en el cual el análisis se está realizando y debe de ser $Q \leq 1$. La tasa de sobre muestreo, o de redundancia, está dada por $r = \frac{\beta}{1-\alpha}$. Modificando los valores de α y de β , la frecuencia de descomposición y de redundancia puede ser ajustada, contrario que con la TO, donde el factor Q permanece constante.

3.4.5 Características de textura de Harlick

El análisis de textura está basado en estadística de alto orden, donde al contrario que con la estadística de primer orden, se considera la relación espacial entre múltiples píxeles al mismo tiempo. Algunos de los métodos que consideran estas relaciones entre la intensidad de los píxeles incluyen las ondículas, los filtros de Gabor, y las características de textura de Haralick (Löfstedt, Brynolfsson, Asklund, Nyholm, & Garpebring, 2019). Las características de textura fueron propuestas por primera vez en 1973 por Robert Haralick. Estas han sido utilizadas ampliamente para el análisis de microfotografías, fotografía

aérea, e imágenes satelitales, entre otras (Wibmer et al., 2015). En este trabajo, las características de Haralick son utilizadas para extraer características de imágenes médicas.

Las imágenes médicas como las IRMe consisten de un gran número de píxeles con diferente nivel de gris. Las características de Haralick se pueden calcular en el dominio espacial y consisten en 14 características de textura. Estas características contienen información sobre la homogeneidad, el contraste, número de límites, complejidad, estructura lineal de las imágenes, entre otras. Para extraer esta información, las relaciones espaciales promedio entre las diferentes intensidades de los píxeles en las imágenes son consideradas como texturas.

El primer paso para extraer las características, es cuantizar la imagen I de tamaño $M \times K$ y rango $[a, b]$, a un rango $[1, N]$, donde N es el número deseado de niveles de gris. El segundo paso es obtener la Matriz de Co-Ocurrencia de Niveles de Gris (MCCNG) de la región de interés, o de la imagen completa, dependiendo de la aplicación. La MCCNG es una matriz de tamaño $N \times N$ en la que el número de ocurrencias de un nivel de gris de los píxeles es representado. La distancia entre los vecinos es definida, dado el desplazamiento del vector $\delta = (d_x, d_y)$, 270° o 315° . Los valores $X(i, j)$ de la MCCNG son calculados como se muestra en la ecuación 28.

$$X(i, j) = \sum_{m=1}^M \sum_{k=1}^K \begin{cases} 1, & \text{if } I(m, k) = i \text{ and } I(m + d_x, k + d_y) = j, \\ 0, & \text{otherwise} \end{cases} \quad (28)$$

Es importante considerar que para calcular los valores de la MCCNG, el número de operaciones requeridas es directamente proporcional al número de células de resolución n presentes en la imagen, por lo que es costoso el calcular áreas extensas de la imagen.

El tercer paso consiste en construir la MCCNG normalizada. Esta matriz está compuesta de las probabilidades de aparición de cada uno de los niveles de gris, y se calcula con la ecuación 29.

$$P = \frac{X}{\sum_{i=1}^N \sum_{j=1}^N X(i, j)} \quad (29)$$

Cuando la MCCNG ya se ha calculado, es posible obtener 14 características de textura. Algunas de estas mediciones se relacionan con características específicas de textura en la imagen, como la homogeneidad, el contraste, o la presencia de estructuras organizadas en la imagen. Otras, caracterizan la compleja naturaleza de las transiciones de tonos de gris conforme ocurren en la imagen. Las características de Haralick, así como algunas características derivadas de éstas se muestran a continuación.

1. Segundo Momento Angular:

$$f_1 = \sum_i \sum_j (p(i, j))^2 \quad (30)$$

2. Contraste:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \right\}, |i - j| = n \quad (31)$$

3. Correlación:

$$f_3 = \frac{\sum_i^N \sum_j^N (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (32)$$

Donde $\mu_x, \mu_y, \sigma_x,$ y σ_y son el promedio y la desviación estandar de p_x y p_y .

4. Suma de cuadrados: Varianza

$$f_4 = \sum_i^N \sum_j^N (i - \mu)^2 p(i, j) \quad (33)$$

5. Momento de diferencia inversa:

$$f_5 = \sum_i^N \sum_j^N \frac{1}{1 + (i - j)^2} p(i, j) \quad (34)$$

6. Suma de Promedios:

$$f_6 = \sum_{i=2}^{2N} i p_{x+y}(i) \quad (35)$$

7. Suma de Varianzas:

$$f_7 = \sum_{i=2}^{2N} (i - f_8)^2 p_{x+y}(i) \quad (36)$$

8. Suma de Entropía:

$$f_8 = - \sum_{i=2}^{2N} p_{x+y}(i) \log(p_{x+y}(i)) \quad (37)$$

9. Entropía:

$$f_9 = - \sum_i^N \sum_j^N p(i, j) \log(p(i, j,)) \quad (38)$$

10. Disimilitud:

$$\sum_{i=1}^N \sum_j^N j = 1^N |i - j| * p(i, j) \quad (39)$$

11. Prominencia de Grupos

$$\sum_{i=1}^N \sum_{j=1}^N (i + j - 2\mu)^3 p(i, j) \quad (40)$$

12. Sombra de Grupos

$$\sum_{i=1}^N \sum_{j=1}^N (i + j - 2\mu)^4 p(i, j) \quad (41)$$

13. Diferencia de Varianza:

$$f_{10} = \text{varianza of } p_{x-y} \quad (42)$$

14. Diferencia de Entropía:

$$- \sum_{i=0}^{N-1} p_{x-y}(i) \log \{ p_{x-y}(i) \} \quad (43)$$

15. Probabilidad Máxima:

$$\max_{ij} p(i, j) \quad (44)$$

16. Medidas de Correlación de la Información (12 y 13):

$$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (45)$$

17.

$$f_{13} = (1 - \exp[-2.0(HXY2 - HXY)])^{\frac{1}{2}} \quad (46)$$

$$HXY = - \sum_i^N \sum_j^N p(i, j) \log(p(i, j)) \quad (47)$$

Donde: HX y HY son la entropía de p_x y p_y , y

$$HXY1 = - \sum_i^N \sum_j^N p(i, j) \log(p_x(i)p_y(j)) \quad (48)$$

$$HXY2 = - \sum_i^N \sum_j^N p_x(i)p_y(j) \log(p_x(i)p_y(j)) \quad (49)$$

18. Coeficiente de Máxima Correlación:

$$f_{14} = (\text{Segundo mayor eigenvalor de } Q)^{\frac{1}{2}} \quad (50)$$

Donde

$$Q(i, j) = \sum_k^N \frac{p(i, k)p(j, k)}{p_x(i)p_y(k)}$$

La media y el rango de cada una de estas 14 medidas, promediadas sobre los cuatro valores (ángulos $0^\circ, 45^\circ, 90^\circ$ y 135°), comprenden el conjunto de 28 características que pueden ser utilizadas como entradas para un clasificador. Sin embargo, algunas de estas características pueden estar fuertemente correlacionadas entre ellas, por lo si se aplica la selección de características, se puede obtener la mejor combinación de las 28 (Haralick, Shanmugam, & Dinstein, 1973).

3.5 Selección de Características

La selección de características es un proceso de selección del mejor grupo de características para una aplicación en específico, o un clasificador. La aplicación de la selección de características puede tener un gran impacto en la exactitud y rendimiento de los algoritmos de clasificación. algunos de los objetivos de la selección de características son: reducir la complejidad del modelo, eliminar las características irrelevantes, ruidosas o las características altamente correlacionadas para incrementar el rendimiento

del clasificador. Existen tres formas principales de atacar el problema de la selección de características: selección basada en *Filtros*, basada en *Wrappers* o basada en *enfoques incorporados*. La selección basada en filtros tiene la ventaja de ser independiente al algoritmo de clasificación que se va a utilizar, y ocupar menos recursos computacionales. Por otro lado, la selección de características basada en *Wrappers*, encuentra el mejor grupo de características para un clasificador en específico, a cambio de un costo computacional mayor. La selección de características basada en el tercer enfoque es una mezcla de filtros y wrappers (Sayed, Hassanien, & Azar, 2019). Algunos de los modelos propuestos para una selección de características óptima, incluye el uso de AP (Raj et al., 2020), algoritmos evolutivos, algoritmos bio-inspirados (Sayed et al., 2019), algoritmos basados en *Wrappers* (Kohavi, John, et al., 1997) o algoritmos basados en filtros (Alirezanejad, Enayatifar, Motameni, & Nematzadeh, 2020).

En este trabajo, selección de características basadas en filtros y en *Wrappers* se utilizar para la selección óptima de características que fueron extraídas tanto de imágenes como de grabaciones de voz. En esta sección, la teoría detrás de los algoritmos de ACP y del SGCBW es explicada.

3.5.1 Análisis de Componentes Principales

El ACP es uno de los análisis más utilizados para la reducción de datos de alta-dimensionalidad, los cuales están altamente correlacionados. El objetivo de el ACP es reducir la dimensionalidad de los datos, proyectando los datos a un nuevo subespacio de menor dimensionalidad. Cada una de estas proyecciones, llamada Componentes Principales (CP), explica un porcentaje de la varianza de los datos, con un mínimo de pérdida de información, y manteniendo los datos linealmente no correlacionados. El primer CP está compuesto por la proyección de los datos originales a un nuevo eje que contiene la mayor varianza. El segundo CP es proyectado en un segundo eje con la mayor varianza que es ortogonal al CP anterior, y el proceso continúa de manera iterativa. El máximo número de componentes posible es el número de dimensiones de los datos originales (Kherif & Latypova, 2020).

Para calcular el ACP, el primer paso es extraer la media de cada una de las características de los datos originales para obtener una base de datos con media zero. El siguiente paso es calcular la matriz de covarianza de los datos. En el caso de que los datos tengan más de dos dimensiones, el número de matrices de covarianza que necesitan ser calculadas es de $\frac{n!}{(n-2)! \times 2}$. Después, los eigenvectores y eigenvalores $(\lambda_1, \lambda_2, \dots, \lambda_n)$ de las matrices de covarianza son calculados. Los eigenvectores representan los ejes principales, y los eigenvalores asociados a cada eigenvector, representan el nivel de varianza que explican los datos proyectados en el primer eje. Estos valores son llamados puntuaciones (y_1, y_2, \dots, y_n) . Debido a que los eigenvectores con el mayor eigenvalor pertenecen al primer CP, los eigenvalores pueden ser ordenados de mayor a menor y de esta manera ordenar la varianza explicada por el CP. Usualmente, los CP que explican la menor cantidad de varianza pueden ser ignorados, y de esta manera

reducir la dimensionalidad de los datos.

En este trabajo, el ACP es utilizado para seleccionar las características x con la mayor contribución para la clasificación de las grabaciones de voz. Las características con los resultados más significativos son definidos como los que contribuyen de mayor manera. La ecuación 51 da el porcentaje de la varianza y la puntuación y_i (Solana-Lavalle & Rosas-Romero, 2021a).

$$\text{porcentaje explicado de varianza}_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} \quad (51)$$

La carga $\phi_{j,i}$ representa el peso de la contribución de la característica x_j al CP i . Debido a que el CP es el resultado de la combinación lineal normalizada de las características, las cargas deben satisfacer la ecuación 52.

$$\sum_{j=1}^n \phi_j^2 = 1 \quad (52)$$

La carga $\phi_{i,j}$ corresponde a la contribución de cada una de las características, que son dadas por la ecuación 53.

$$\text{contribución}_j = \frac{\phi_{i,j} \lambda_i}{\sum_{k=1}^n \lambda_k} \quad (53)$$

Las características con la mayor contribución de los CP que explican la mayor varianza, son las que utilizamos en el proceso de selección de características.

3.5.2 Selección del Grupo de Características Basado en Wrappers

El SGCBW (wrapper feature subset selection en inglés) es un algoritmo de selección de características que encuentra el grupo de características que maximiza la eficiencia del algoritmo de clasificación. Esto se logra buscando el grupo de características óptimo para un algoritmo de clasificación en específico, mientras que se descartan el resto de las características, lo que reduce la complejidad del modelo. La SGCBW utiliza una estrategia de búsqueda que reduce el número de iteraciones necesarias para encontrar un grupo con buen rendimiento (Kohavi et al., 1997).

El SGCBW requiere un espacio de búsqueda, un estado inicial, una condición de término y un motor de búsqueda. El espacio de búsqueda consta de los posibles subconjuntos de características donde el tamaño del espacio de búsqueda con n características es de $O(2^n)$. El estado inicial del estado de búsqueda consiste en un vector sin ninguna característica seleccionada, el estado final de búsqueda es el que contiene las características que proporcionan el mejor rendimiento al algoritmo de clasificación, de acuerdo a la métrica deseada, y que para este trabajo es la exactitud. El motor de búsqueda que se utiliza en este trabajo es el *primero el mejor* (best first en inglés).

El algoritmo de búsqueda primero el mejor está basado en árboles, donde la idea es expandir el mejor nodo generado hasta ese punto. El mejor nodo se define como el nodo que obtiene la mayor exactitud hasta ese punto. La condición de paro está basada en una búsqueda obsoleta, lo que significa que cuando el nodo subsecuente no consigue mejorar en un número k de expansiones, la búsqueda se da por terminada.

Para comenzar la búsqueda, un algoritmo de *selección paso a paso hacia adelante* (forward step wise selection en inglés) es utilizado, en el cual el conjunto de características contiene únicamente una característica y el número de características va incrementando con el tiempo. La otra forma de abordar el problema es a través de *selección paso a paso hacia atrás* (backward stepwise selection en inglés), el cual comienza con un vector que contiene todas las características, y en cada iteración, va removiendo las características que no ayudan a mejorar la exactitud del modelo. Sin embargo, este tipo de búsqueda no se utiliza en este trabajo debido a su alta complejidad inicial.

Teóricamente, la búsqueda de primero el mejor termina cuando se alcanza el objetivo. En la realidad, una búsqueda más minuciosa podría reducir la precisión del subconjunto final debido al equilibrio entre sesgo y varianza. La heurística para la búsqueda reduce el tiempo de ejecución necesario para encontrar un grupo de características óptimo. Mientras un algoritmo de búsqueda que busca todas las posibles combinaciones necesita buscar 2^n combinaciones de grupos, una búsqueda de primero el mejor reduce esta búsqueda a $O(n * \text{Log}n)$, donde n es el número de grupos posibles (Kohavi et al., 1997).

3.6 Clasificación

3.6.1 Perceptrón Multicapa

Las RNA son uno de los modelos de aprendizaje más populares en el área del aprendizaje automático para problemas supervisados y no supervisados. Este modelo obtiene su nombre del modelado de los sistemas biológicos de neuronas que se utilizan para procesar la información. El PM es una red neuronal alimentada hacia adelante que consiste en múltiples unidades de procesamiento llamadas *perceptrones* (Rosenblatt, 1958), las cuales son entrenadas con el algoritmo de propagación hacia atrás (Back Propagation en inglés) para aprender las salidas deseadas. El modelo básico de redes neuronales es una extensión del modelo de regresión lineal. Es un algoritmo que permite que los parámetros se ajusten en la etapa de entrenamiento, y toman la forma de la ecuación 54.

$$y(x, w) = f\left(\sum_{j=1}^M w_j \phi_j(x)\right) \quad (54)$$

Donde f es la función de activación que puede ser modificada para ser lineal para los problemas de regresión, o no lineal para los de clasificación. La diferencia entre el algoritmo de *regresión lineal* y las

RNA, reside en la función base $\phi(x)$ que para el caso de las redes neuronales depende de parámetros ajustables, junto con los pesos w_j . Para obtener $\phi(x)$, necesitamos las cantidades a_j , llamadas *activaciones*, y las cuales se muestran en la ecuación 55

$$a_j = \sum_{i=1}^D w_{ji}^{(l)} x_i + w_{j0}^{(l)} \quad (55)$$

Donde x_i son las características de entrada, $j = 1, \dots, M$ y (1) indica los parámetros y la capa de la red respectivamente. w_{ji} son los pesos que son ajustados, y w_{j0} es el peso del bias.

El valor de a es procesado por funciones de activaciones diferenciables y no-lineales como la *función sigmoide*, representada por $h(\cdot)$ la cual se muestra en la ecuación 56.

$$z_j = h(a_j) \quad (56)$$

El valor de z_j se llama un valor unitario oculto. El valor unitario oculto se combina linealmente para dar la salida de la unidad activaciones como se muestra en la ecuación 57.

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)} \quad (57)$$

Donde $k = 1, \dots, k$, y k es el número de salidas de la PM.

La función de activación de la capa final de la red neuronal está determinada por las salidas deseadas (de regresión o clasificación) y dan el valor de y_k .

Para entrenar el PM, una de las técnicas más populares es el algoritmo de *propagación hacia atrás* basada en el gradiente descendiente y propuesta por Rumelhart et. al (Rumelhart, Hinton, & Williams, 1986). El objetivo del algoritmo es ajustar los pesos de las redes para minimizar el error en cada iteración. El algoritmo consta de dos pasos: Primero, se obtiene la derivada del error con respecto a cada uno de los pesos. El segundo paso consiste en ajustar los pesos de acuerdo con el error. La propagación hacia atrás es una técnica eficiente para evaluar el gradiente del error en las capas del PM (Bishop & Nasrabadi, 2006).

3.6.2 K-vecino más cercano

El algoritmo supervisado de clasificación de KNN, fué desarrollado para ser usado cuando no existe mucha información acerca de la distribución de los datos. Los fundamentos del algoritmo fueron el trabajo de Fix and Hodges (Fix & Hodges, 1989) en el que la regla del K vecino más cercano fue introducida. Este algoritmo consiste en calcular la distancia euclídeana entre una muestra desconocida, y un grupo de muestras etiquetadas que constituyen el grupo de entrenamiento.

La distancia euclídeana entre las muestras de entrada x_i con p número de características y la muestra x_j son definidas en la ecuación 58.

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{iN} - x_{jN})^2} \quad (58)$$

Donde N es el número total de muestras.

Para encontrar la clase de la nueva observación, el KNN depende del conjunto de entrenamiento, el cual es etiquetado con valores verdaderos del grupo de pertenencia de cada muestra. El parámetro k se refiere al número de vecinos cercanos que van a ser utilizados para predecir una nueva clase. Es importante notar que k necesita ser un número impar para evitar tener el mismo número de vecinos de diferentes clases. La clase para cada nueva muestra se obtiene con la ecuación 59.

$$d(m_i, x) = \min_j(d(m_j, x)) \quad (59)$$

Donde m_i es el vecino más cercano a la muestra x , y se calcula para los k vecinos más cercanos.

La clase que fue seleccionada con más frecuencia por los vecinos es asignada a la nueva muestra (Peterson, 2009).

3.6.3 Bosques Aleatorios

Los BA son un algoritmo de aprendizaje automático que fueron desarrollados como una mejora al algoritmo de árboles de decisión en bolsas (bagged decision tree en inglés), con un incremento en la exactitud de la clasificación y la regresión, y al mismo tiempo decorrelacionando las características de los árboles de decisión, lo que resulta en una reducción de la variaciones en los resultados (James, Witten, Hastie, & Tibshirani, 2013). Los BA consisten en varios árboles individuales que otorgan su voto en cuanto a la clase a la que pertenece la nueva muestra. Los árboles reciben muestras aleatorias para su construcción y eligen características aleatorias para dividir los nodos. Estas características ayudan a reducir el sobreajuste e incrementar su resistencia al ruido. La clase con el mayor número de votos es la que se elige para la nueva muestra (Livingston, 2005).

Para entrenar el algoritmo de BA, un conjunto de muestras S_n se elige de manera aleatoria del conjunto de entrenamiento con tamaño N para construir un árbol de decisión T_b de tamaño n_b . Este proceso es conocido como *bagging*. Cada uno de los árboles individuales es dividido en dos nodos hijos seleccionando $k = \sqrt{n_b}$ variables de manera aleatoria, y seleccionando el conjunto con la mayor cantidad de ganancia de información. El proceso se repite hasta que no se tienen más ganancias en la información del nuevo conjunto, o un nodo de tamaño mínimo es alcanzado. El nodo hoja es utilizado para generar la probabilidad $P(c|l)$, y emitir un voto. La clasificación de la nueva muestra x se muestra

en la ecuación 60 (Mishina, Murata, Yamauchi, Yamashita, & Fujiyoshi, 2015) (Hastie, Tibshirani, Friedman, & Friedman, 2009).

$$P(c|x) = \frac{1}{T} \sum_{t=1}^T P_t(c|x) \quad (60)$$

Donde c es la clase, y T es el árbol.

3.6.4 Máquina de Soporte Vectorial

La MSV es un algoritmo de aprendizaje automático supervisado basado en el *clasificador de margen máximo* (maximal margin classifier en inglés) desarrollado entre 1992 y 1997. Se basa en un proceso de optimización para encontrar un *hiperplano* óptimo: $aX + bY = C$, que separa a una máxima distancia dos clases linealmente separables. El problema de optimización para encontrar un hiperplano se muestra en la ecuación 61.

$$\text{Min } \phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w, w), \quad (61)$$

Tal que

$$y_i(w \cdot x_i + b) \geq 1$$

Para lograr la separación lineal de los patrones no-linealmente separables, la MSV proyecta los datos a nuevo espacio de mayor dimensión hasta que es linealmente separable. La transformación se logra con una función *kernel* $\phi(x)$ donde x es un vector con las características extraídas de las observaciones. Existen diferentes kernels que pueden ser utilizados incluyendo el kernel *lineal*, el kernel *polinomial* y el kernel de *función de base radial Gaussiana*, (los kernels de BA están dados por la ecuaciones 62, 63 y 64 respectivamente.)

$$k(x, y) = x^T y + c \quad (62)$$

$$k(x, y) = (\alpha x^T y + c)^d \quad (63)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (64)$$

Donde c es una constante, α representa la pendiente, d es el grado polinomial del kernel y σ es la desviación estándar.

Otro parámetro que se puede modificar es C, el cual está relacionado con el error de los puntos que fueron clasificados incorrectamente en el hiperplano. Conforme el valor de C crece, el algoritmo permite un error más grande en la clasificación.

3.7 Métricas de rendimiento

Las métricas de evaluación utilizadas para evaluar el rendimiento del algoritmo en este trabajo son *Exactitud*, *Precisión* o valor predictivo positivo, *Sensibilidad*, *especificidad* y el *Coficiente de correlación de Mathew* (CCM). Estas métricas fueron elegidas ya que permiten elegir el algoritmo que mejor detecta la EP, así como su gran uso en la literatura, lo que nos da la oportunidad de comparar los resultados obtenidos en este trabajo. La elección del CCM se tomó para contar con una métrica que no sea sensible al desbalance de las clases, y así comparar los resultados de los diferentes datasets de manera adecuada. Si un paciente con la EP es correctamente clasificado, contará como un Positivo Verdadero (PV); de lo contrario será contado como un Falso Negativo (FN). Por otro lado, si un persona sana se es clasificada en el grupo de pacientes con la EP, será contado como un Falso Positivo (FP), de lo contrario, será contado como un Verdadero Negativo (VN). Las métricas de evaluación se pueden calcular con las ecuaciones 65, 66, 67, 68 y 69 respectivamente.

$$Exactitud = \frac{PV + VN}{PV + VN + FP + FN} \quad (65)$$

$$Precisión = \frac{PV}{PV + FP} \quad (66)$$

$$Sensibilidad = \frac{PV}{PV + FN} \quad (67)$$

$$Especificidad = \frac{VN}{VN + FP} \quad (68)$$

$$CCM = \frac{(PV * VN) - (FP * FN)}{\sqrt{(PV + FP)(PV + FN)(VN + FP)(VN + FN)}} \quad (69)$$

3.8 Metodologías propuestas

Una vez descritos todos los antecedentes teóricos para preprocesar las señales biomédicas, extraer y utilizar los algoritmos de selección de características, en esta sección se describen las metodologías propuestas para detectar la EP en grabaciones de voz y IRMe, respectivamente.

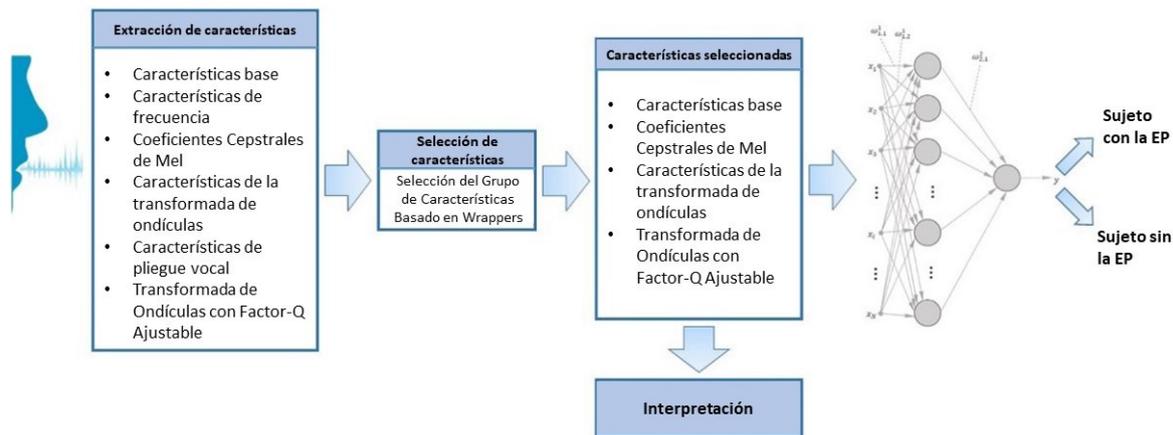


Figura 9: Metodología propuesta para detectar la EP en grabaciones de voz. Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC.

3.8.1 Metodología propuesta para la detección de la enfermedad de Parkinson en grabaciones de voz

La metodología propuesta para detectar la EP en grabaciones de voz comienza con el proceso de selección de características, ya que en el conjunto de datos que utilizamos en este trabajo, la obtención de las señales biomédicas y la extracción de características fueron realizados anteriormente por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019).

Un aspecto importante en la metodología que se propone, es el equilibrio entre la explicabilidad de los resultados y la precisión de la clasificación. Esto se logra con tres procesos diferentes: el uso del método de reducción de características utilizando SGCBW, el uso de diferentes algoritmos de aprendizaje automático incluyendo RNA, KNN, BA, y MSV, y el análisis de los subconjuntos de características seleccionadas por el algoritmo SGCBW para cada clasificador específico con ACP.

En el primer paso, utilizamos SGCBW para encontrar las características más relevantes para cada algoritmo de aprendizaje de máquina, lo que aumenta la precisión de la clasificación y reduce la complejidad del modelo. Dado que estamos entrenando cuatro algoritmos diferentes, se obtienen cuatro subconjuntos diferentes de características para analizar.

El segundo paso de la metodología propuesta es la clasificación de las muestras y el cálculo de las métricas para medir el rendimiento de cada algoritmo. Éstas se validan utilizando la validación cruzada de diez veces, que divide aleatoriamente el conjunto de datos en un conjunto de entrenamiento y un conjunto de pruebas dividido en diez partes para entrenar y realizar la clasificación. La media de los resultados de la validación cruzada de diez veces revela qué algoritmo es el mejor para clasificar estas señales y que al mismo tiempo es capaz de generalizar en diferentes conjuntos de datos. En el Anexo a, se muestran las tablas 1 a la 12, las cuales contienen los resultados de la clasificación en bases de

datos en las que se incluye únicamente una grabación por paciente y una grabación por sujeto control para los casos de poblaciones mixtas, de sexo masculino y de sexo femenino. Las métricas muestran un rendimiento menor, lo que se puede deber al número reducido de grabaciones, sin embargo se evita la posible fuga de datos en la validación.

El tercer paso consiste en analizar cada característica en los subconjuntos generados con el SGCBW. Para ello se utiliza el algoritmo ACP como sistema de clasificación. El ACP da como resultado una combinación lineal de las distintas características y sus CP, como se explica en los antecedentes teóricos. Estos CP están clasificados de acuerdo a la varianza que explica cada una de las características. Las cuatro características más importantes son las que tienen la mayor contribución a los 10 CP que mayor varianza explican. Estas características son las que mayor influencia ejercen de acuerdo a la ecuación 70.

$$Influencia\ de\ la\ característica_j = \sum_i^{10} (contribución_{j,i} * porcentaje\ explicado\ de\ varianza_i) \quad (70)$$

Donde la *contribución_{j,i}* de la característica *j* al CP *i* está dada por la ecuación 53 y el porcentaje explicado de varianza del CP *i* está dado por la ecuación 51.

La metodología propuesta para la detección de la EP con grabaciones de voz se muestra en la Figura 9.

3.8.2 Metodología propuesta para la detección de la enfermedad de Parkinson en Imágenes de Resonancia Magnética Estructural

La metodología propuesta para detectar la EP en IRMe consta de los siguientes pasos. En primer lugar, los volúmenes de IRMe se transforman de un formato de archivo *.DICOM* al formato de archivo *.NIFTI*. El segundo paso es el preprocesado de las imágenes, que incluye la alineación de la imagen, la segmentación en materia blanca, materia gris y fluido cerebroespinal, y por último, el suavizado de la imagen, todo ello explicado al principio del capítulo. En el tercer paso, se realiza el VBM utilizando el software SPM y CAT12 en MATLAB con un valor *p* de 0.001. La elección del valor *p* para el VBM en la mayoría de los casos se realiza arbitrariamente, y en gran parte de los estudios se utiliza el valor de *p* = 0.005. Sin embargo, en este trabajo se decidió reducir el valor ya que el número de vóxeles aislados que se mostraban como relevantes era muy alto. Con un valor de *p* = 0.001 se tiene un margen de error de un falso positivo en 1000 vóxeles, los cuales se consideran independientes uno de otro. Debido a que la corrección de tipo Bonferroni implicaría el utilizar una *p* = .000000471, lo cual resulta en regiones muy pequeñas, decidimos utilizar la técnica de clústeres significativos para reducir el número de falsos positivos, la cual busca que los vóxeles seleccionados necesitan estar conectados con otros vóxeles con significancia estadística, por lo que se elimina la suposición de que los vóxeles son independientes (Voormolen et al., 2010) (Kurth, Gaser, & Luders, 2015). Del mapa-t resultante, consideramos una RDI

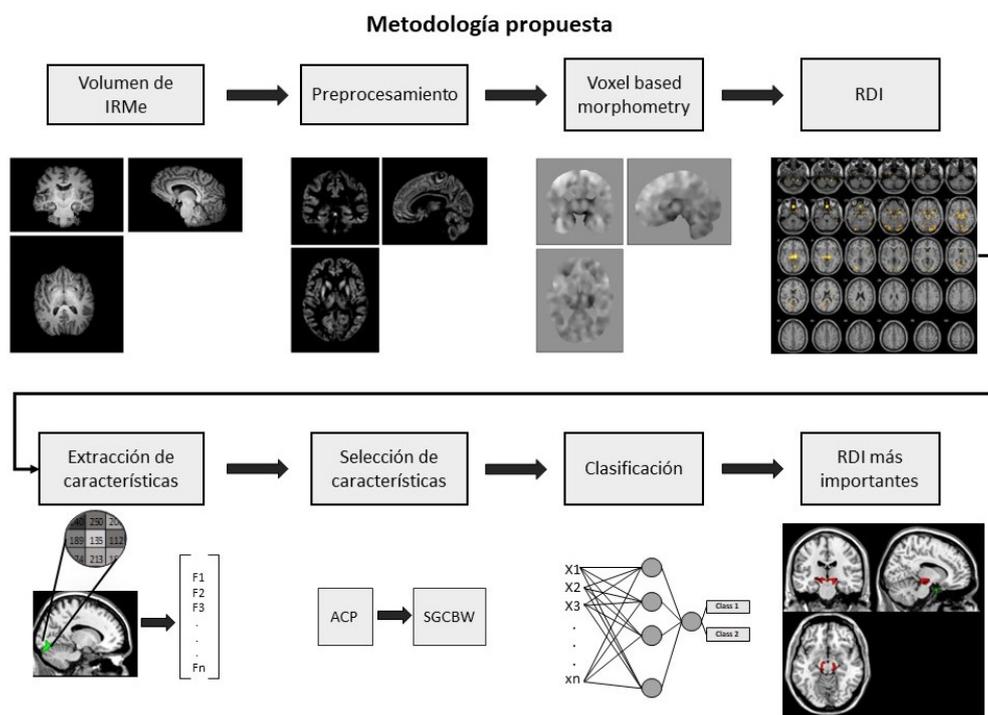


Figura 10: Metodología propuesta para la detección de la EP utilizando IRMe.

las regiones que contienen 50 o más vóxeles conectados con el valor t requerido. Es necesario comentar que el utilizar esta técnica puede resultar en RDI las cuales no están perfectamente delineadas debido a los posibles vóxeles falsos positivos.

Una vez identificados las RDI, el siguiente paso consiste en extraer características estadísticas de primer y segundo orden de los grupos de vóxeles. Las características basadas en histogramas y las características de Haralick (Haralick et al., 1973) se extraen de los RDI. Para reducir la complejidad del modelo y aumentar el rendimiento de los clasificadores de AM, utilizamos el ACP y SGCBW. En primer lugar, el ACP reduce el gran conjunto de características a CP, de donde seleccionamos las características que más contribuyen al CP. Una vez que reducimos el número original de características a la mitad, utilizamos SGCBW para encontrar el subconjunto de características que maximiza el rendimiento para cada clasificador. La función de costo para SGCBW es la exactitud obtenida en la validación cruzada de 5 grupos con el algoritmo de AM. Una vez obtenido el subconjunto óptimo de características de cada algoritmo, se realiza una validación cruzada de diez partes con los clasificadores para evaluar su rendimiento. Los clasificadores utilizados en este trabajo incluyen el logístico, BA, Bayes Ingenuous (BI), Red Bayesiana (RB), KNN, RNA y MSV.

Finalmente, las RDI con las características que contribuyen en mayor grado para la clasificación son reportadas para facilitar la interpretación clínica de los resultados y proveer información de interés para los médicos.

La Figura 10 muestra la metodología propuesta en este trabajo para detectar la EP con IRMe, y muestra las RDI más importantes.

Capítulo 4 Resultados

Este capítulo muestra los resultados de los diferentes enfoques para la detección de la EP. Debido a que se utilizaron dos modalidades de señales biomédicas para detectar la EP, este capítulo se divide en subsecciones que incluyen los resultados y la discusión de la clasificación de las grabaciones de voz y la clasificación de la EP utilizando IRMe.

4.1 Grabaciones de voz

La base de datos de (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019) contiene 756 grabaciones de voz y 754 características de cada una de las grabaciones. Estas grabaciones fueron extraídas de participantes de ambos sexos. El gran número de grabaciones nos permitió realizar experimentos y seleccionar las mejores características para cada uno de los sexos de manera independiente. Uno de los experimentos fue la comparación entre bases de datos con el mismo número de observaciones para las clases (balanceado) y otra con números diferentes de muestras en cada clase (desbalanceado). Estas diferencias resultaron en seis bases de datos diferentes, las cuales consisten en dos bases de datos con participantes de sexo masculino y femenino con bases de datos balanceadas y desbalanceadas, dos bases de datos con participantes de sexo femenino (balanceado y desbalanceado) y dos bases de datos con participantes de sexo masculino (balanceado y desbalanceado).

Para cada base de datos se utilizó la selección de características para incrementar el rendimiento de los clasificadores y obtener información acerca de las características que pueden ayudar a la evaluación e interpretación clínica de los resultados. Primero, el algoritmo de SGCBW fue utilizado para encontrar el grupo de características que mejor se adapta a cada clasificador. Después, el ACP nos permitió encontrar las características que contribuyeron de mayor manera a la clasificación. Estas características corresponden a los mayores eigenvalores de la transformación. El criterio para determinar la contribución de cada una de las características del conjunto se muestra en la ecuación 53 en el capítulo 3. Los algoritmos de clasificación utilizados en este trabajo para comparar y seleccionar los conjuntos de características más relevantes son KNN, MSV-FBR, PM y BA.

4.1.1 Base de datos con participantes de sexo femenino y masculino

En esta sección presentamos los resultados que corresponden a la base de datos que contiene las observaciones de participantes de ambos sexos. Para cada base de datos, cuatro conjuntos de características son seleccionados con SGCBW y ACP, uno por cada algoritmo de clasificación. Estos conjuntos son utilizados para entrenar a los clasificadores y obtener las métricas de rendimiento para compararlos. Las métricas incluyen la exactitud, sensibilidad, especificidad y precisión.

La base de datos balanceada es construida con un total de 192 grabaciones de cada clase (pacientes y controles). El número de grabaciones que se utiliza está definido por el número de participantes control, debido a que es la clase con el menor número de observaciones. Las 192 grabaciones de pacientes con la EP fueron seleccionadas de manera aleatoria de la base de datos original, y el resto de las observaciones fue descartado. En la Tabla 6, se muestran los resultados de los clasificadores con la base de datos balanceada. La tabla se divide en cuatro secciones, las cuales muestran los resultados de diferentes grupos de características obtenidos con SGCBW. Para cada uno de los grupos de características, probamos cada uno de los clasificadores. Los mejores resultados se muestran resaltados en negritas.

Tabla 6: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 384 observaciones provenientes de pacientes con la EP y sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el SGCBW para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.87	0.849	0.891	0.886	0.740	PM	0.875	0.901	0.849	0.856	0.751
BA	0.854	0.828	0.880	0.874	0.709	BA	0.932	0.927	0.937	0.937	0.865
KNN	0.885	0.859	0.911	0.907	0.771	KNN	0.917	0.906	0.927	0.925	0.833
MSV	0.888	0.865	0.911	0.907	0.776	MSV	0.823	0.885	0.760	0.787	0.650

Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.865	0.896	0.833	0.843	0.730	PM	0.849	0.865	0.833	0.838	0.698
BA	0.862	0.875	0.849	0.853	0.724	BA	0.909	0.906	0.911	0.911	0.818
KNN	0.799	0.828	0.771	0.783	0.599	KNN	0.950	0.958	0.943	0.944	0.901
MSV	0.825	0.812	0.838	0.834	0.651	MSV	0.838	0.844	0.833	0.835	0.677

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

Se generó una base de datos en la que se asegura que la evaluación de los resultados se realiza con el mismo número de grabaciones de voz de cada participante. Se incluyeron las 3 grabaciones provenientes de 46 pacientes con la EP, de los cuales 23 son de sexo masculino y 23 son de sexo

femenino, y 46 sujetos control de los cuales 23 son de sexo masculino y 23 de sexo femenino. Los resultados se muestran en la tabla 7.

Tabla 7: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 384 observaciones provenientes de 64 pacientes con la EP y 64 sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el SGCBW para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.710	0.734	0.690	0.659	0.422	PM	0.895	0.904	0.886	0.884	0.790
BA	0.848	0.880	0.82	0.804	0.698	BA	0.924	0.933	0.915	0.913	0.848
KNN	0.789	0.833	0.756	0.725	0.584	KNN	0.920	0.914	0.926	0.927	0.841
MSV	0.855	0.882	0.831	0.819	0.712	MSV	0.492	0.494	0.491	0.594	-0.014
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.931	0.934	0.928	0.927	0.862	PM	0.873	0.860	0.887	0.891	0.747
BA	0.898	0.898	0.898	0.898	0.797	BA	0.873	0.865	0.881	0.884	0.746
KNN	0.924	0.953	0.898	0.891	0.849	KNN	0.927	0.954	0.904	0.898	0.856
MSV	0.493	0.494	0.491	0.594	-0.014	MSV	0.493	0.493	0.491	0.594	-0.014

La base de datos desbalanceada es construida con un total de 756 grabaciones de 188 pacientes con la EP y 64 controles. En la Tabla 8, se muestran los resultados de las clasificaciones con una base de datos desbalanceada. La tabla se divide en cuatro secciones, las cuales muestran los resultados de los diferentes grupos de características que se obtuvieron con el SGCBW. Para cada uno de los grupos de características, se prueba cada clasificador. El mejor resultado es mostrado en negritas.

Tabla 8: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 756 observaciones provenientes de 188 pacientes con la EP y 64 sujetos control fue utilizada. La base de datos contiene sujetos del sexo masculino y femenino. Cada grupo de características se obtuvo con el SGCBW para cada uno de los clasificadores. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.921	0.959	0.807	0.935	0.786	PM	0.899	0.947	0.760	0.921	0.728
BA	0.902	0.977	0.682	0.900	0.730	BA	0.922	0.984	0.74	0.917	0.788
KNN	0.903	0.966	0.719	0.91	0.735	KNN	0.910	0.968	0.74	0.916	0.754
MSV	0.947	0.964	0.929	0.972	0.868	MSV	0.874	0.977	0.573	0.870	0.648

Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.866	0.936	0.661	0.890	0.632	PM	0.856	0.952	0.573	0.867	0.593
BA	0.848	0.943	0.568	0.865	0.571	BA	0.848	0.931	0.604	0.873	0.577
KNN	0.820	0.917	0.536	0.853	0.494	KNN	0.856	0.938	0.615	0.877	0.599
MSV	0.831	0.956	0.463	0.84	0.510	MSV	0.836	0.97	0.443	0.836	0.526

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

En la Figura 11, el diagrama de cajas muestra las cuatro características que contribuyen en mayor cantidad a la varianza de la base de datos mixta. El criterio para seleccionar estas características se muestra en la ecuación 53. En los diagramas de cajas, se muestra la variabilidad entre los valores de las características de la EP y los sujetos control. Las cajas azules corresponden a los valores de las características de los pacientes con la EP, y las cajas naranjas corresponden a los sujetos control. El valor promedio se representa con una cruz en la caja, la media se muestra con una línea a la mitad de la caja. El gráfico está dividido en cuartiles. El primer cuartil consiste en valores de entre la parte inferior de la caja y la media. El valor medio es llamado el segundo cuartil. El tercer cuartil incluye los valores que van de la media a la parte superior de la caja. Los puntos inferiores y superiores representan los valores mínimos y máximos respectivamente. En la Tabla 9 se muestra el valor exacto del promedio (μ), desviación estándar (σ), Q1, Q2 y Q3 para los pacientes con la EP y sujetos control.

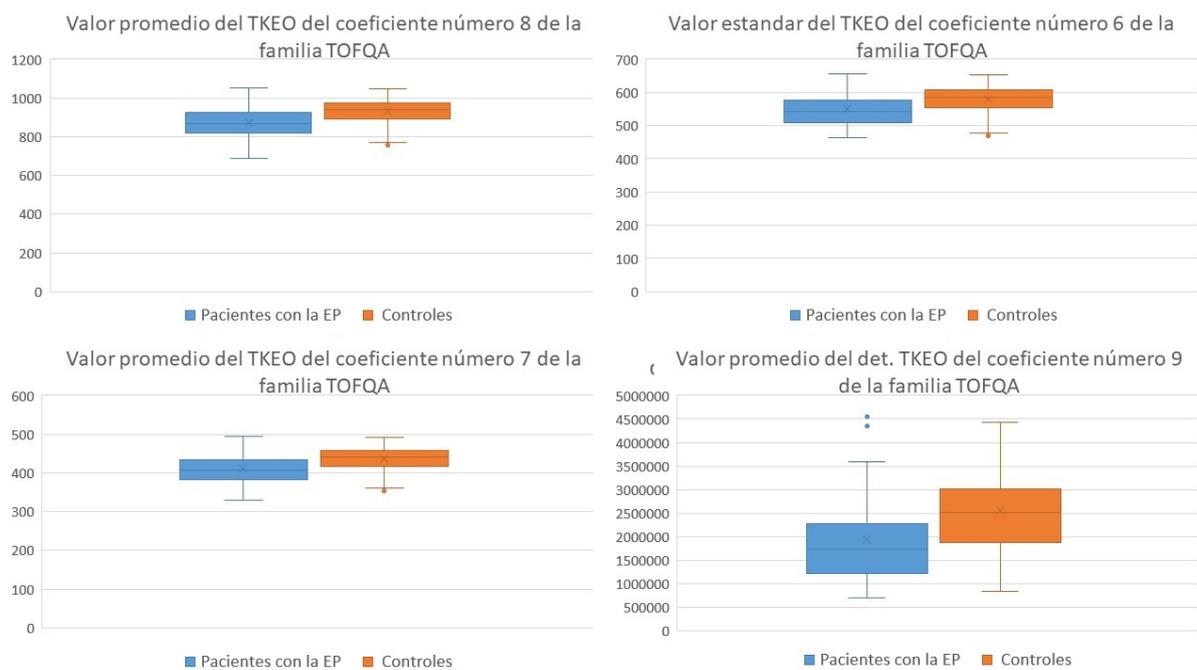


Figura 11: Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos mixta. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

Tabla 9: Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de ambos sexos.

Pacientes Femeninos y Masculinos con la EP					
Característica	μ	σ	Q1	Q2	Q3
LT TKEO mean value for the eight TOFQA coefficient	880.8585	84.8514	828.47	885.33	945.75
LT TKEO standard value for the sixth TOFQA coefficient	553.8791	53.18925	516.98	554.52	590.9053
LT TKEO mean value for the seventh TOFQA coefficient	416.53	40.001	389.33	414.96	445.35
det TKEO mean value for the ninth TOFQA coefficient	2074499	1066726	133×10^4	187×10^4	256×10^4
Controles Femeninos y Masculinos					
Feature	μ	σ	Q1	Q2	Q3
LT TKEO mean value for the eight TOFQA coefficient	924.423	106.8708	882.73	961.97	1014.82
LT TKEO standard value for the sixth TOFQA coefficient	589.0155	67.11649	550.007	600.9408	633.26
LT TKEO mean value for the seventh TOFQA coefficient	443.31	50.3	413.8	450.88	476.69
det TKEO mean value for the ninth TOFQA coefficient	2975794	1801452	181×10^4	2785×10^3	3745×10^3

4.1.2 Base de datos de participantes de sexo femenino

La base de datos contiene grabaciones de 81 pacientes femeninos y 41 sujetos control. Esto resulta en 366 grabaciones las cuales fueron analizadas para encontrar las características más relevantes para la clasificación de participantes del sexo femenino. Los resultados de cada clasificador correspondientes a la base de datos balanceada con participantes de sexo femenino se muestran en la Tablas 10 y 11 y los resultados de la base de datos desbalanceada se muestran en la Tabla 12. Los mejores resultados son resaltados en negritas. La base de datos balanceada consiste en 123 grabaciones de cada una de las clases. La base de datos desbalanceada consiste en grabaciones de 81 pacientes con la EP femeninos y 81 controles femeninos, con un total de 366 grabaciones. Los diferentes grupos de características fueron seleccionados utilizando SGCBW con cada uno de los clasificadores para cada una de las bases de datos.

Tabla 10: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 246 observaciones provenientes de pacientes con la EP y controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV					Selección del grupo de características con el algoritmo de BA						
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.723	0.829	0.724	0.750	0.555	PM	0.756	0.756	0.756	0.756	0.512
BA	0.752	0.789	0.715	0.735	0.505	BA	0.886	0.886	0.886	0.886	0.772
KNN	0.74	0.733	0.748	0.744	0.480	KNN	0.825	0.805	0.845	0.839	0.651
MSV	0.817	0.886	0.748	0.779	0.640	MSV	0.760	0.805	0.715	0.739	0.522
Selección del grupo de características con una PM					Selección del grupo de características con el algoritmo de KNN						
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.890	0.911	0.87	0.875	0.781	PM	0.764	0.805	0.724	0.744	0.530
BA	0.833	0.837	0.829	0.831	0.667	BA	0.760	0.780	0.74	0.750	0.521
KNN	0.793	0.813	0.772	0.781	0.586	KNN	0.797	0.829	0.764	0.779	0.594
MSV	0.785	0.74	0.829	0.812	0.571	MSV	0.691	0.862	0.520	0.642	0.406

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

Para asegurar que la evaluación de los resultados se realiza con el mismo número de grabaciones de voz de cada sujeto, se generó una base de datos con las 3 grabaciones de cada uno de los participantes provenientes de 41 pacientes de sexo femenino con la EP, y 41 sujetos control. Los resultados se muestran en la tabla 11.

Tabla 11: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 246 observaciones provenientes de 41 pacientes con la EP y 41 controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.825	0.845	0.808	0.797	0.645	PM	0.833	0.836	0.831	0.830	0.667
BA	0.853	0.865	0.842	0.837	0.706	BA	0.902	0.930	0.878	0.870	0.806
KNN	0.821	0.816	0.826	0.829	0.642	KNN	0.869	0.853	0.889	0.894	0.741
MSV	0.882	0.867	0.898	0.902	0.765	MSV	0.492	0.49	0.493	0.398	-0.016
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.927	0.941	0.913	0.910	0.854	PM	0.764	0.782	0.748	0.731	0.529
BA	0.866	0.895	0.841	0.829	0.734	BA	0.845	0.876	0.819	0.805	0.693
KNN	0.833	0.873	0.801	0.780	0.670	KNN	0.906	0.924	0.890	0.886	0.813
MSV	0.488	0.485	0.489	0.390	-0.024	MSV	0.5	0.5	0.5	0.390	0.008

Tabla 12: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 366 observaciones provenientes de pacientes con la EP y controles de sexo femenino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.891	0.946	0.780	0.895	0.751	PM	0.877	0.897	0.837	0.916	0.727
BA	0.866	0.951	0.699	0.862	0.693	BA	0.907	0.963	0.797	0.903	0.789
KNN	0.861	0.942	0.699	0.861	0.680	KNN	0.855	0.885	0.797	0.896	0.677
MSV	0.896	0.971	0.748	0.884	0.765	MSV	0.839	0.930	0.658	0.843	0.628
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.888	0.938	0.789	0.898	0.745	PM	0.891	0.942	0.789	0.898	0.751
BA	0.863	0.946	0.699	0.861	0.687	BA	0.88	0.942	0.756	0.884	0.725
KNN	0.863	0.909	0.772	0.888	0.691	KNN	0.959	0.983	0.911	0.956	0.907
MSV	0.817	0.951	0.553	0.808	0.575	MSV	0.869	0.975	0.658	0.849	0.703

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

En la Figura 12, se muestran los diagramas de bloques de las cuatro características que contribuyeron en mayor cantidad a la varianza de la base de datos de participantes de sexo femenino. Los diagramas de cajas, muestran la variabilidad entre los valores de las características de los pacientes con la EP y los

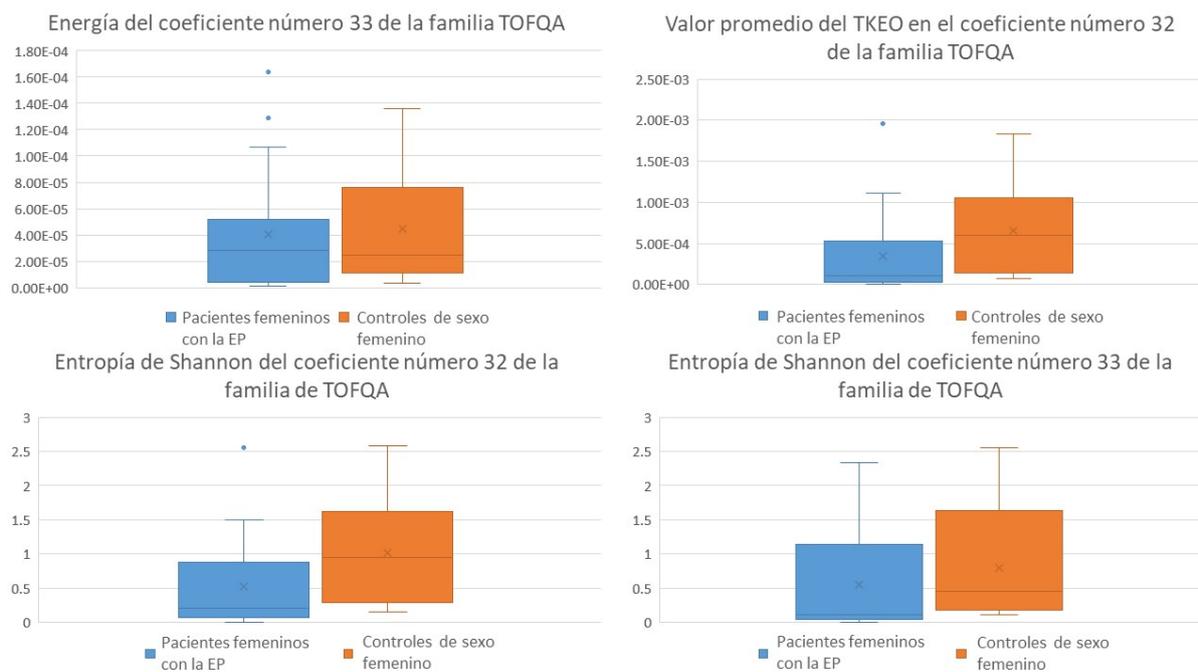


Figura 12: Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos de participantes de sexo femenino. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

sujetos control. Las cajas azules corresponden a los valores de las características de los pacientes con la EP, y las cajas naranjas corresponden a los sujetos sanos. En la tabla 13 se muestra el valor exacto del promedio (μ), desviación estándar (σ), Q1, Q2 y Q3 para los pacientes con la EP y sujetos control.

Tabla 13: Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de sexo femenino.

Pacientes de sexo femenino con la EP					
Característica	μ	σ	Q1	Q2	Q3
Energía del coeficiente 33 de TOFQA	3.73×10^{-5}	3.806×10^{-5}	0.00000544	0.0000218	0.0000566
Valor promedio de TKEO del coeficiente 32 de TOFQA	3.43×10^{-4}	0.0005086	0.00003248	0.00120245	0.0004611
Entropía de Shannon del coeficiente 32 de TOFQA	0.4786	0.5855	0.07624	0.2053	0.7244
Entropía de Shannon del coeficiente 33 de TOFQA	0.4432	0.6244	0.0465	0.1366	0.6183
Sujetos control de sexo femenino					
Feature	μ	σ	Q1	Q2	Q3
Energía del coeficiente 33 de TOFQA	3.61×10^{-5}	4×10^{-5}	7.11×10^{-6}	0.000023	4.29×10^{-5}
Valor promedio de TKEO del coeficiente 32 de TOFQA	5.16×10^{-4}	0.000505	0.0000121	0.00028	0.000795
Entropía de Shannon del coeficiente 32 de TOFQA	0.8426	0.7395	0.2589	0.5474	1.2676
Entropía de Shannon del coeficiente 33 de TOFQA	0.824	0.7489	0.1802	0.5902	1.12966

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

4.1.3 Base de datos con participantes de sexo masculino

El tercer grupo de experimentos fue llevado a cabo con una base de datos que contiene grabaciones de 107 pacientes con la EP de sexo masculino y 23 controles. Esto resultó en 390 grabaciones las cuales se analizaron para encontrar las características más relevantes para la clasificación de participantes de sexo masculino.

Los resultados de cada uno de los clasificadores correspondientes a la base de datos balanceada con grabaciones de sujetos de sexo masculino se muestran en las Tablas 14 y 15, y los resultados de la base de datos desbalanceada se muestran en la Tabla 16. Los mejores resultados son resaltados en negritas. Las bases de datos balanceadas contienen 69 grabaciones de cada clase. La base de datos desbalanceada contiene 107 sujetos de sexo masculino y 23 sujetos control, con un total de 390 grabaciones. Los diferentes grupos de características fueron seleccionados utilizando SGCBW con cada uno de los clasificadores para cada una de las bases de datos.

Tabla 14: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 138 grabaciones provenientes de pacientes con la EP y sujetos control de sexo masculino fueron utilizadas para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.804	0.826	0.783	0.798	0.610	PM	0.855	0.841	0.870	0.866	0.710
BA	0.833	0.812	0.855	0.848	0.667	BA	0.927	0.884	0.971	0.968	0.858
KNN	0.797	0.812	0.783	0.789	0.594	KNN	0.848	0.754	0.942	0.929	0.708
MSV	0.870	0.841	0.899	0.892	0.740	MSV	0.819	0.826	0.812	0.814	0.638
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.906	0.913	0.899	0.900	0.812	PM	0.761	0.783	0.739	0.750	0.522
BA	0.819	0.783	0.855	0.844	0.639	BA	0.848	0.826	0.870	0.864	0.696
KNN	0.848	0.855	0.841	0.843	0.696	KNN	0.920	0.870	0.971	0.968	0.845
MSV	0.775	0.710	0.841	0.817	0.555	MSV	0.754	0.652	0.855	0.818	0.518

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

Se generó una base de datos que incluye el mismo número de sujetos masculinos y las 3 grabaciones pertenecientes a cada participante. Esta base de datos contiene las grabaciones de 23 pacientes de sexo masculino con la EP, y 23 sujetos control para tener un total de 138 grabaciones. Los resultados de la clasificación con cada uno de los algoritmos se muestran en la tabla 15.

Tabla 15: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 138 observaciones provenientes de 23 pacientes con la EP y 23 controles de sexo masculino fue utilizada para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.811	0.787	0.841	0.855	0.625	PM	0.898	0.910	0.887	0.884	0.797
BA	0.840	0.840	0.840	0.840	0.681	BA	0.963	0.984	0.944	0.942	0.928
KNN	0.855	0.845	0.866	0.869	0.710	KNN	0.898	0.898	0.898	0.898	0.797
MSV	0.848	0.863	0.833	0.826	0.696	MSV	0.492	0.487	0.494	0.289	-0.016
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.964	0.957	0.970	0.971	0.927	PM	0.920	0.926	0.914	0.913	0.841
BA	0.935	0.917	0.954	0.956	0.8703	BA	0.862	0.847	0.878	0.884	0.725
KNN	0.898	0.910	0.887	0.884	0.797	KNN	0.978	0.985	0.971	0.971	0.957
MSV	0.507	0.533	0.504	0.116	0.023	MSV	0.528	0.516	0.611	0.898	0.086

Tabla 16: Resultados de clasificación de pacientes con la EP y sujetos control con 4 algoritmos diferentes. La base de datos con 390 grabaciones provenientes de pacientes con la EP y sujetos control de sexo masculino fueron utilizadas para entrenar los algoritmos. Cuatro grupos de características fueron seleccionados con el algoritmo de SGCBW para cada clasificador. Cada uno de los grupos de características fue probado con los cuatro clasificadores.

Selección del grupo de características con una MSV						Selección del grupo de características con el algoritmo de BA					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.869	0.953	0.478	0.895	0.501	PM	0.905	0.959	0.652	0.928	0.656
BA	0.879	0.981	0.406	0.885	0.524	BA	0.946	0.997	0.710	0.941	0.807
KNN	0.854	0.963	0.348	0.873	0.410	KNN	0.933	0.969	0.768	0.951	0.764
MSV	0.869	1.00	0.261	0.863	0.474	MSV	0.833	1.00	0.058	0.832	0.219
Selección del grupo de características con una PM						Selección del grupo de características con el algoritmo de KNN					
Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC	Algoritmo	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
PM	0.859	0.959	0.391	0.880	0.441	PM	0.874	0.935	0.594	0.915	0.552
BA	0.879	0.984	0.391	0.883	0.522	BA	0.910	0.978	0.594	0.918	0.665
KNN	0.828	0.931	0.348	0.869	0.330	KNN	0.944	0.966	0.841	0.966	0.806
MSV	0.846	0.997	0.145	0.844	0.327	MSV	0.846	1.00	0.130	0.842	0.331

De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

En la Figura 13, se muestran los diagramas de bloques de las cuatro características que contribuyeron en mayor cantidad a la varianza de la base de datos de participantes de sexo masculino. Los diagramas

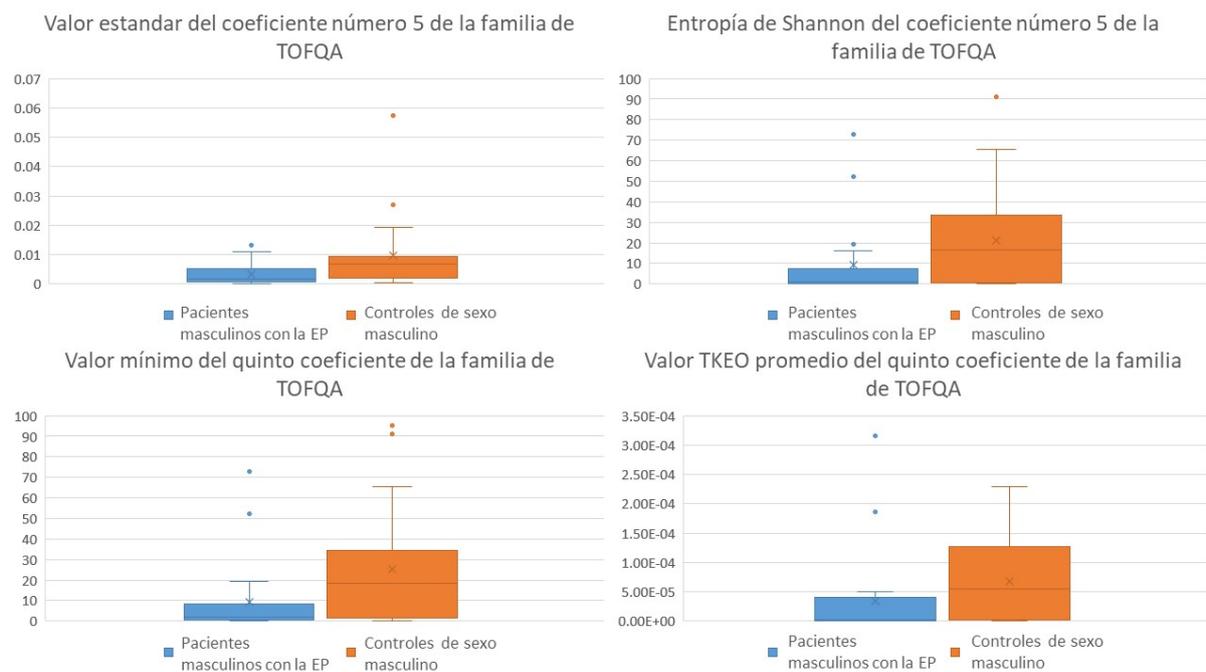


Figura 13: Diagramas de cajas de las cuatro características con la mayor contribución a la detección de la EP en la base de datos de participantes de sexo masculino. De (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

de cajas, muestran la variabilidad entre los valores de las características de los pacientes con la EP y los sujetos control. Las cajas azules corresponden a los valores de las características de los pacientes con la EP, y las cajas naranjas corresponden a los sujetos sanos. En la Tabla 17 se muestra el valor exacto del promedio (μ), desviación estándar (σ), Q1, Q2 y Q3 para los pacientes con la EP y sujetos control.

Tabla 17: Rango de los valores de las características de los pacientes con la EP y sujetos sanos en una población con participantes de sexo masculino.

Pacientes con la EP de sexo masculino					
Característica	μ	σ	Q1	Q2	Q3
Valor estándar de el quinto coeficiente de TOFQA	0.0027651	0.0030759	0.0006299	0.0014699	0.0036567
Entropía de Shannon de el quinto coeficiente de TOFQA	8.201105	15.56329	0.32416	1.2458	6.8209
Valor mínimo del quinto coeficiente de TOFQA	8.07123	15.33394	0.32416	1.5932	6.8209
Valor promedio del valor de TKEO del quinto coeficiente de TOFQA	2.79 $\times 10^{-5}$	5.825 $\times 10^{-5}$	7.115 $\times 10^{-7}$	3.055 $\times 10^{-6}$	2.328 $\times 10^{-5}$
Sujetos control de sexo masculino					
Característica	μ	σ	Q1	Q2	Q3
Valor estándar del quinto coeficiente de TOFQA	0.008074	0.010691	0.002099	0.005479	0.008775
Entropía de Shannon de el quinto coeficiente de TOFQA	17.79456	21.1533	1.8698	11.2268	26.5685
Valor mínimo del quinto coeficiente de TOFQA	20.27424	24.97224	1.8698	11.3183	28.2205
Valor promedio del valor de TKEO del quinto coeficiente de TOFQA	6.13 $\times 10^{-5}$	8.09 $\times 10^{-5}$	4.22 $\times 10^{-6}$	2.81 $\times 10^{-5}$	8.72 $\times 10^{-10}$

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

En las tablas 18 y 19, se muestran el número de características de cada familia que fueron elegidas para cada uno de los subconjuntos resultantes del algoritmo de SGCBW cuando se entrenaron con bases de datos balanceadas y no balanceadas respectivamente.

4.1.4 Discusión

En esta sección, mostramos los resultados de la detección de la EP con la metodología que se muestra en la Figura 9. Analizamos el rendimiento de la metodología propuesta obteniendo diferentes conjuntos de datos separando las grabaciones por sexo y mismo número de controles y pacientes con la EP, para determinar las características más relevantes. Después, reportamos el rendimiento promedio obtenido de la validación cruzada de diez veces para reducir los sesgos en los resultados.

Los experimentos se llevaron a cabo utilizando el conjunto de datos proporcionado por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019), que contiene 756 grabaciones de 189 pacientes con la EP y 64 controles sanos, y la cual se considera una de las bases de datos

Tabla 18: Número de características seleccionadas de cada grupo (características base, CCFM, TO y TOFQA) ejecutando el algoritmo SGCBW con cada clasificador en el conjunto de datos equilibrado.

	Hombres-Mujeres				Mujeres				Hombres			
	KNN	PM	MSV	BA	KNN	PM	MSV	BA	KNN	PM	MSV	BA
Características base	1	1	2	0	0	2	0	0	1	0	0	1
CCFM	3	3	2	2	0	1	1	3	1	0	1	1
TO	2	0	0	5	0	2	0	1	1	1	2	2
TOFQA	8	5	10	12	6	5	6	7	7	9	5	7
Total	14	9	14	19	6	10	7	11	10	10	8	11

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

Tabla 19: Número de características seleccionadas de cada grupo (características base, CCFM, TO y TOFQA) ejecutando el algoritmo SGCBW con cada clasificador en el conjunto de datos no balanceada.

	Mujeres				Hombres			
	KNN	PM	MSV	BA	KNN	PM	MSV	BA
Características base	1	3	1	2	1	0	1	1
CCFM	0	2	6	2	2	3	2	3
TO	2	0	3	2	2	1	1	2
TOFQA	9	6	10	11	6	8	8	10
Total	12	11	20	17	11	12	12	16

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC

más grandes disponibles públicamente. Sin embargo, el conjunto de datos está desequilibrado entre el número de controles y pacientes con la EP tanto en participantes femeninos como masculinos.

La Tabla 8 muestra los resultados de la clasificación utilizando todas las grabaciones de un conjunto de datos desequilibrado. El mejor rendimiento lo obtiene el clasificador MSV-FBR, con una exactitud de 0.947 utilizando sólo 18 características, lo que muestra una mejora significativa en la exactitud respecto a los 0.86 reportados por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019) con un modelo más complejo que consistía en un subconjunto con 50 características y una MSV-FBR. Los resultados de la clasificación dependen del conjunto de datos utilizado para entrenar los clasificadores, si el conjunto de datos está equilibrado o desequilibrado en cuanto al número de pacientes con la EP frente a controles, la Tabla 8 muestra que la MSV-FBR disminuye su rendimiento, y el KNN obtiene mejores resultados. Se observa que el algoritmo de KNN resulta ser el clasificador con mejor rendimiento cuando se manejan conjuntos de datos desequilibrados.

En otros experimentos, se realizaron pruebas separadas para poblaciones masculinas y femeninas.

Tabla 20: Comparación de los resultados de exactitud de la clasificación de pacientes con la EP con otras bases de datos de grabaciones de voz con características similares a las utilizadas en este trabajo

Autor	Base de datos	No. Características	Tipo de características	Exactitud	Algoritmo
(Guruler, 2017)	Controles 8 EP 23 Grabaciones 195	22	Base	0.99	RNA
(B. E. Sakar, 2017)	Controles 8 EP 42 Grabaciones 5923	16	Base	0.964	MSV
(Sakar, 2019)	Controles 64 EP 188 Grabaciones 756	50	Base, CCFM, TO, TOFQA	0.87	MSV
(Rizvi et al., 2020)	Controles 20 EP 20 Grabaciones 40	26	Base	0.99	MLCP
(Pramanik et al., 2021)	Controles 64 EP 188 Grabaciones 756	756	Base, CCFM, TO, TOFQA	0.941	BA
(Quan et al., 2021)	Controles 15 EP 30 Grabaciones 268	53-569	Transición de articulación y Frecuencia Fundamental	0.8-0.84	MLCP
(Vital et al., 2021)	Controles 51 EP 62 Grabaciones 1200	29	Base, edad, sexo, índice de masa corporal	1	RNA probabilística
Base de datos desbalanceada mixta	Controles 64 EP 188 Grabaciones 756	20	Base, CCFM, TO, TOFQA	0.947	MSV
Base de datos balanceada mixta	Controles 64 EP 64 Grabaciones 384	14	Base, CCFM, TO, TOFQA	0.95	KNN
Base de datos balanceada sexo femenino	Controles 41 EP 41 Grabaciones 246	11	Base, CCFM, TO, TOFQA	0.92	PM
Base de datos balanceada sexo masculino	Controles 23 EP 23 Grabaciones 138	14	Base, CCFM, TO, TOFQA	0.978	KNN

Para cada población se utilizaron conjuntos equilibrados y desequilibrados. Los mejores resultados de clasificación para cada población, masculina y femenina, se obtuvieron con conjuntos de datos desequilibrados. Se observó que la sensibilidad aumenta al utilizar conjuntos de datos desequilibrados. Sin embargo, la especificidad de los algoritmos disminuyó significativamente en comparación con los conjuntos de datos equilibrados. Esto es indicativo de un sesgo hacia el diagnóstico de pacientes con la EP, lo cual no es deseable porque el paciente podría empezar a ser tratado con la medicación incorrecta.

Los clasificadores funcionan mejor con un conjunto equilibrado con participantes masculinos que en un conjunto con grabaciones de participantes de sexo femenino, lo que podría explicarse por el menor número de grabaciones en el conjunto de datos femeninos. En los conjuntos de datos no equilibrados, los algoritmos muestran una mayor exactitud en el conjunto de datos femenino. Sin embargo, como se ha comentado anteriormente, el resto de las métricas son inconsistentes y sugieren que el clasificador está sesgado hacia clasificar las grabaciones como procedentes de pacientes con la EP.

Tras el proceso de reducción de características, encontramos que el número óptimo de características para la detección de la EP oscila entre 6 y 20, como se muestra en las Tablas 18 y 19, con una reducción de más del 97% respecto al conjunto original de 754 características. Con la metodología propuesta, observamos que se seleccionan características diferentes para cada sexo. En concreto, a la hora de detectar la EP en participantes de sexo femenino, las características de CCFM son más importantes que las de TO, y en participantes de sexo masculino ocurre lo contrario. Un hallazgo de esta tesis que confirma el trabajo de (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019), es que la familia de características más útil para detectar la EP en cualquier conjunto de datos, ya sea equilibrado o desequilibrado, con participantes femeninos o masculinos, son las características de TOFQA mostrando un 68% y 67% de predominio respectivamente.

Para proporcionar al médico una cantidad mayor de información sobre la enfermedad y evitar los resultados binarios de los clasificadores, analizamos la contribución de las características seleccionadas para el conjunto de datos sólo de hombres y las características seleccionadas para el conjunto de datos sólo de mujeres. Utilizamos el ACP para encontrar las cuatro características que más contribuían a la detección de la EP, y comparamos las características utilizadas para participantes masculinos y femeninos. Las Figuras 11, 12 y 13, muestran las características que son más útiles para discriminar entre pacientes con la EP y sujetos control. En todos los casos, las características pertenecen al grupo de TOFQA, y en 10 de los 12 casos, los primeros cuartiles de ambos grupos no se solapan. Los dos casos en los que los primeros cuartiles se solapan pertenecen al grupo de participantes mujeres.

Aunque todas las características pertenecen a la familia TOFQA, las características de baja frecuencia son las más representativas en los participantes varones, y las características de alta frecuencia se seleccionan para detectar la EP en las participantes mujeres, como se observa en las figuras mencionadas

anteriormente. Esto es esperable debido a las diferencias entre la voz de los participantes femeninos y masculinos y sugiere que el entrenamiento de los clasificadores con características específicas para cada sexo puede tener un impacto positivo en los resultados.

Finalmente, en la tabla 20 se hace una comparación de los resultados obtenidos en este trabajo de tesis con los trabajos mencionados en la revisión de literatura. Se muestran el número de características utilizados en cada trabajo, las cuales van de 16 a 756 por cada grabación. En la mayoría de los casos se utilizan características pertenecientes a la familia de características base. Es importante destacar el número de pacientes y sujetos control que participaron en los estudios, los cuales en la mayoría de los casos son menores a 60, y sin embargo se llegan a analizar hasta 1200 grabaciones provenientes de estos pacientes. La poca cantidad de sujetos en los estudios y el gran número de grabaciones puede generar que se tenga un sesgo a la hora de entrenar los algoritmos de aprendizaje automático y resultar en la memorización de las características de los pacientes. Así mismo es importante destacar que los mejores resultados reportados se obtuvieron con RNA y MLCP. Los resultados de este trabajo se muestran al final de la tabla y se destaca el número reducido de características seleccionadas, así como el número balanceado de participantes y grabaciones analizadas.

4.2 Imágenes de Resonancia Magnética Estructural

En este trabajo se utilizaron IRMe de la base de datos del PPMI, la cual contiene imágenes tomadas con sensores de 1.5 y 3 Teslas. Estas imágenes fueron tomadas de pacientes prodrómicos, pacientes con la EP en etapas intermedias y avanzadas, y sujetos control de ambos sexos. El tamaño de la base de datos nos permiten realizar estudios de manera separada para cada uno de los sexos y campos magnéticos de las IRMe. Antes de aplicar VBM a estas imágenes, fueron normalizadas, segmentadas, alineadas y emparejadas con la plantilla estándar como parte del proceso de preprocesamiento que se explica en el capítulo 3. En esta sección, presentamos los resultados del proceso de VBM. Después, mostramos las regiones de interés que se obtuvieron con la metodología propuesta por la selección de características. Finalmente, se muestra el rendimiento de cada uno de los algoritmos de clasificación aplicado a cada uno de los sexos e intensidad de campo magnético utilizando las RDI encontradas.

4.2.1 Resultados del proceso de Morfometría basada en Vóxeles

La VBM es una técnica utilizada para detectar las diferencias en la concentración de *materia gris* entre dos grupos de IRMe realizando un análisis de cada uno de los vóxeles (Ashburner & Friston, 2000). Los detalles de esta técnica son explicados en el Capítulo 3. Realizamos VBM utilizando herramientas de MatLab, específicamente CAT12 y SPM12 (Gaser & Dahnke, 2016). Generamos plantillas que contienen los valores promedios de los vóxeles de cada grupo, y después aplicamos VBM para encontrar las

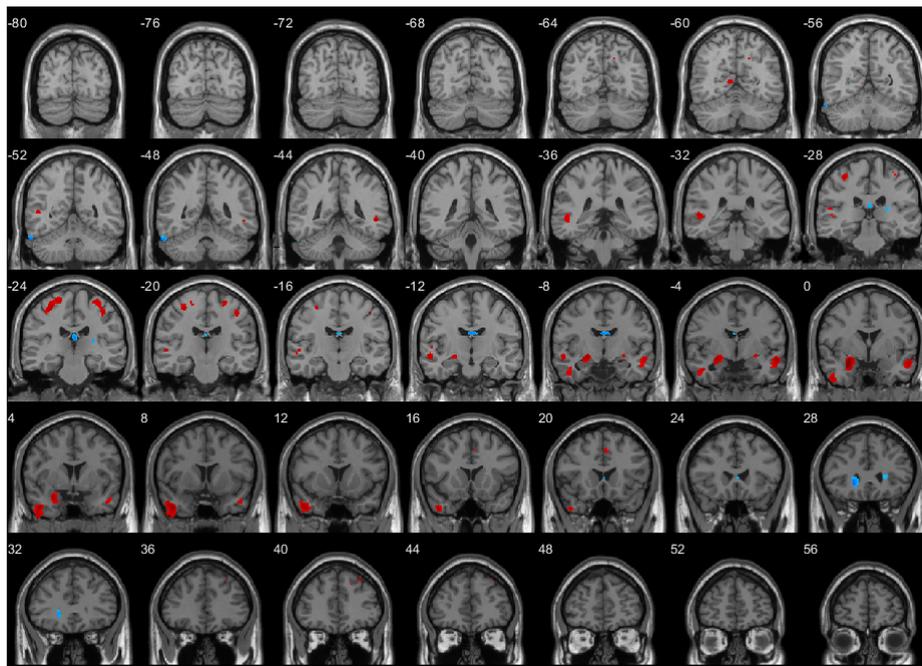


Figura 14: Mapa de colores de la vista coronal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

RDI. Nosotros seleccionamos un umbral de mínimo 50 vóxeles consecutivos con diferencias relevantes entre los dos grupos para ser considerada una RDI.

En las Figuras 20 (coronal), 15 (sagital) y 16 (transversal), las RDI encontradas en pacientes prodrómicos masculinos son mostradas. El número de participantes en la base de datos es de 33 pacientes prodrómicos con la EP y 95 controles. No incluimos el VBM de pacientes prodrómicos femeninos con la EP debido al número reducido de datos. Los vóxeles que se muestran en rojo, representan un valor t positivo, y los vóxeles mostrados en azul representan los valores t negativos. En este trabajo detectamos las RDI durante una etapa temprana en los cerebros de los participantes y podemos observar como se modifica su estructura.

En las Figuras 17 (coronal), 18 (sagital) y 19 (transversal), las RDI encontradas en los pacientes con la EP de sexo femenino son mostradas. En este estudio comparamos las IRMe de 104 pacientes con la EP y 64 sujetos control. Los resultados de VBM muestran que únicamente se encontraron valores t negativos, los cuales se muestran en amarillo. En las Figuras 20 (coronal), 21 (sagital) y 22 (transversal), las RDI más relevantes encontradas en pacientes de sexo masculino son mostradas. El número de pacientes con la EP y sujetos control son 226 y 86 respectivamente. Los vóxeles mostrados en rojo,

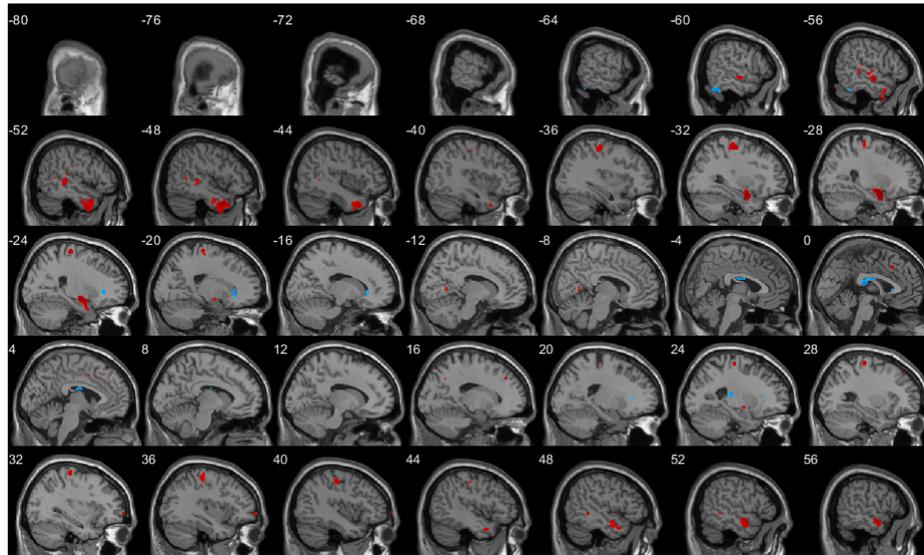


Figura 15: Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos.(para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

representan los valores t positivos, y los vóxeles mostrados en azul representan los vóxeles t negativos.

4.2.2 Selección de características de las Regiones de Interés

Para definir los grupos más importantes de RDI para cada sexo, las RDI con el mayor número de características seleccionadas fueron utilizadas. El proceso de selección de características incluye SGCBW y ACP como se explica en el Capítulo 3. Se encontró que algunas de las RDI seleccionadas por el proceso de VBM, no fueron elegidas en los grupos de características para la clasificación. Las RDI, por lo tanto, se consideran como las regiones que tuvieron una mayor contribución para la clasificación.

Las Figuras 23 y 24, muestran las RDI que se seleccionaron en las IRMe tomadas con campos magnéticos de 1.5 y 3 Teslas, respectivamente, de participantes femeninos. Las RDI que se seleccionaron de las IRMe de 1.5 Teslas, se incluyen las regiones del tálamo (RDI 6 mostrada en rojo), y el tronco encefálico (RDI 7 mostrada en verde). Las RDI de las IRMe de 3 Teslas incluyen el giro temporal inferior (RDI 1 y 18 mostradas en rojo y azul respectivamente), el córtex visual (RDI 14 mostrada en verde), y región del córtex (RDI 19 mostrada en amarillo).

Las Figuras 25 y 26, muestran las RDI que se seleccionaron en las IRMe tomadas con campos magnéticos de 1.5 y 3 Teslas, respectivamente de participantes masculinos. Las RDI que se seleccionaron en las IRMe de 1.5 Teslas, se incluyen las regiones del tálamo (RDI 4 mostrada en rojo), y el

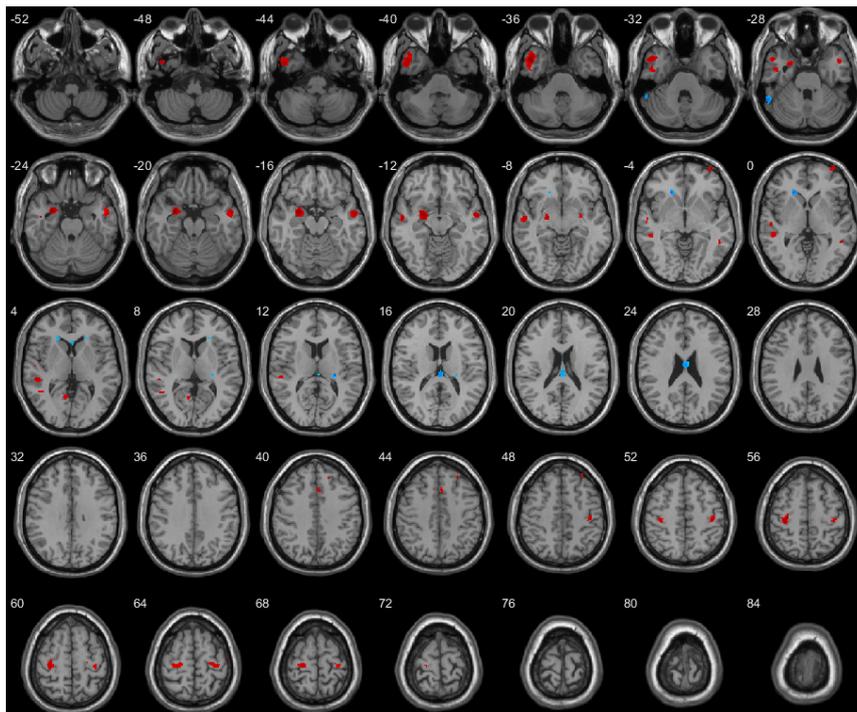


Figura 16: Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 33 pacientes prodrómicos con la EP, y el otro grupo contiene 95 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos y los vóxeles de color azul corresponden a los valores t negativos.(para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

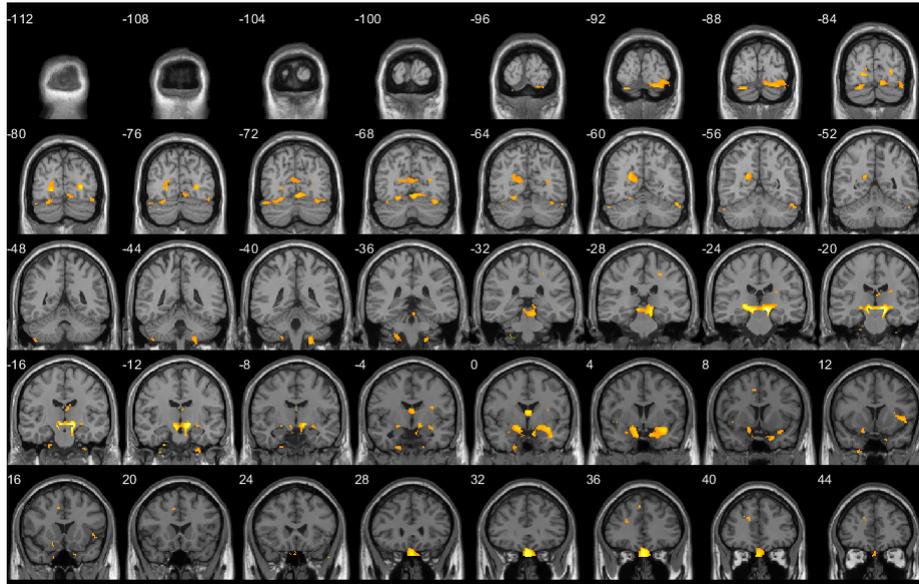


Figura 17: Mapa de colores de la vista coronal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

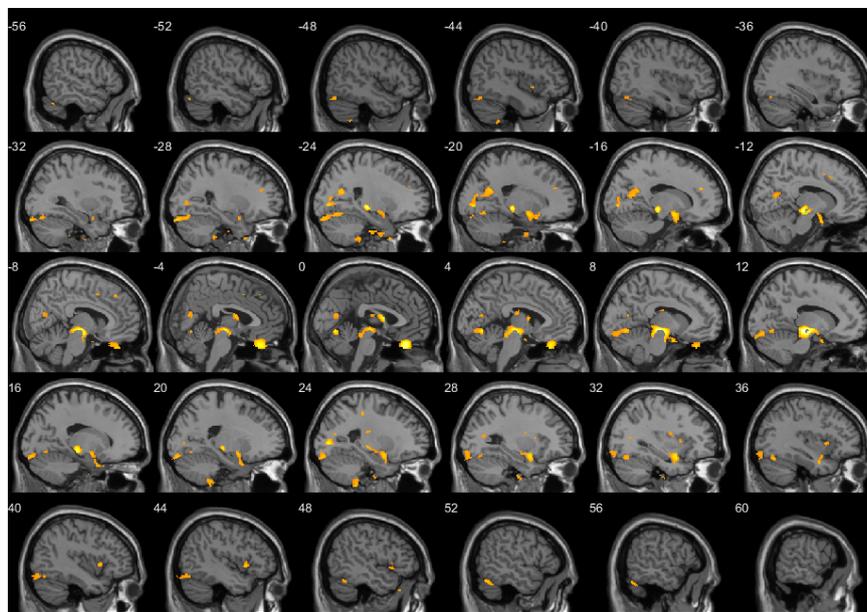


Figura 18: Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

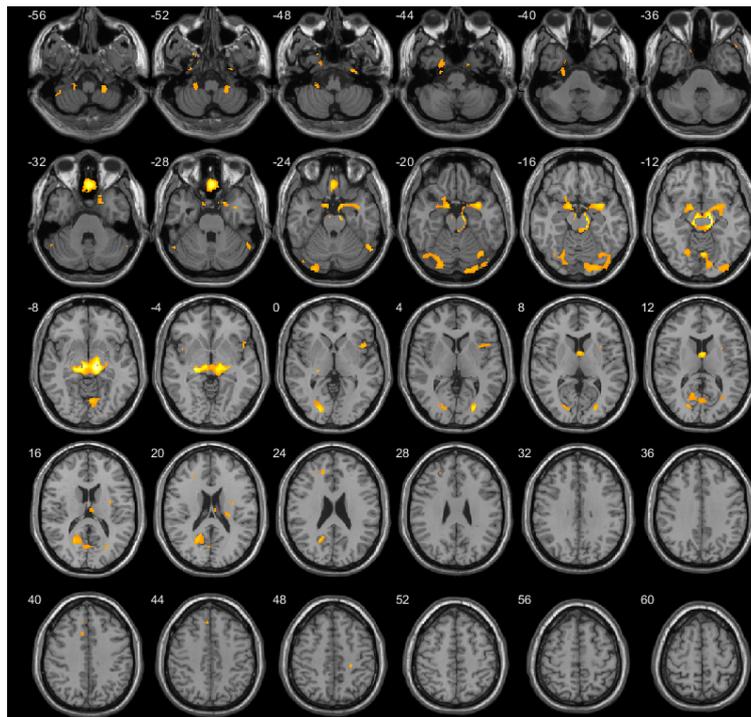


Figura 19: Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 104 pacientes femeninos con la EP, y el otro grupo contiene 64 controles femeninos. El valor p fue seleccionado como 0.001. Los vóxeles de color amarillo corresponden a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

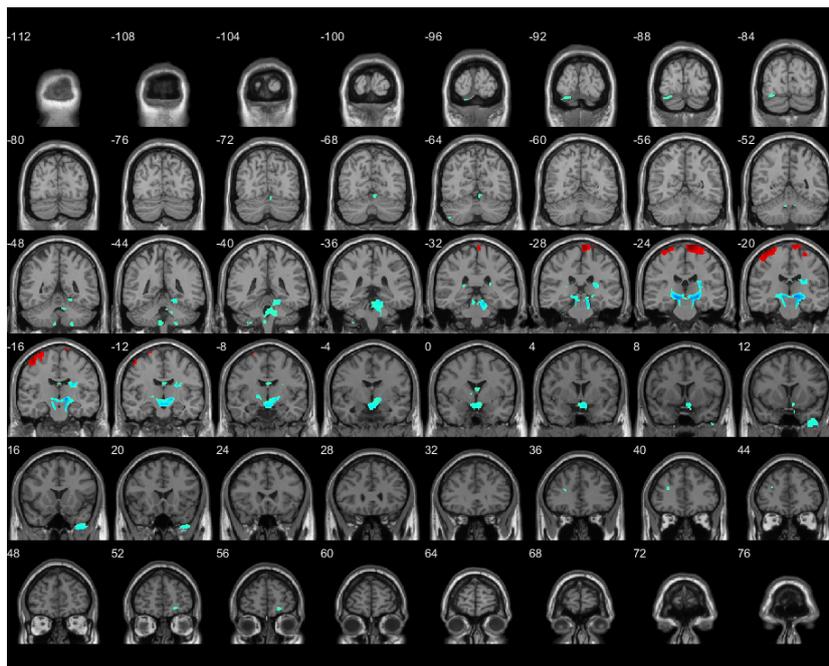


Figura 20: Mapa de colores de la vista coronal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

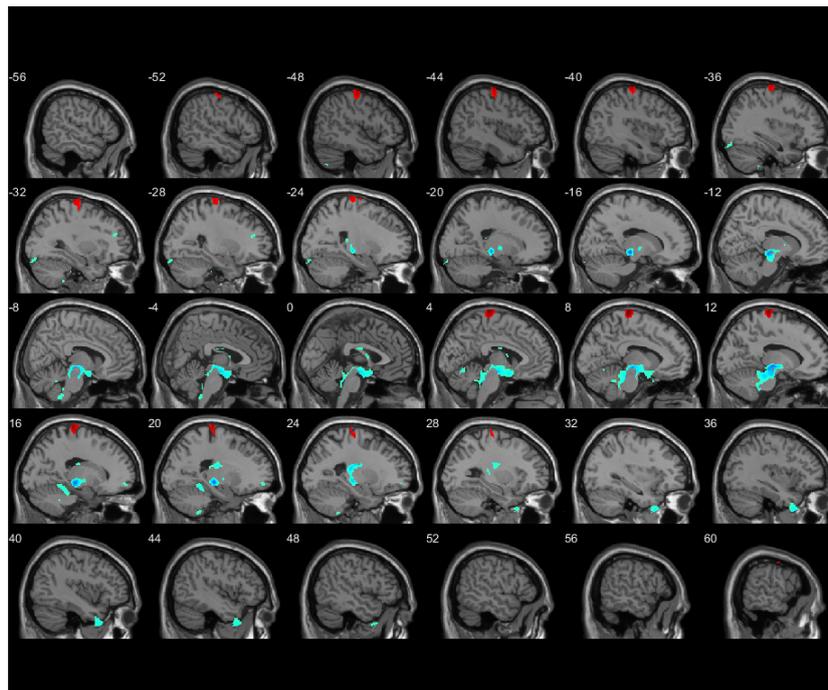


Figura 21: Mapa de colores de la vista sagital de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

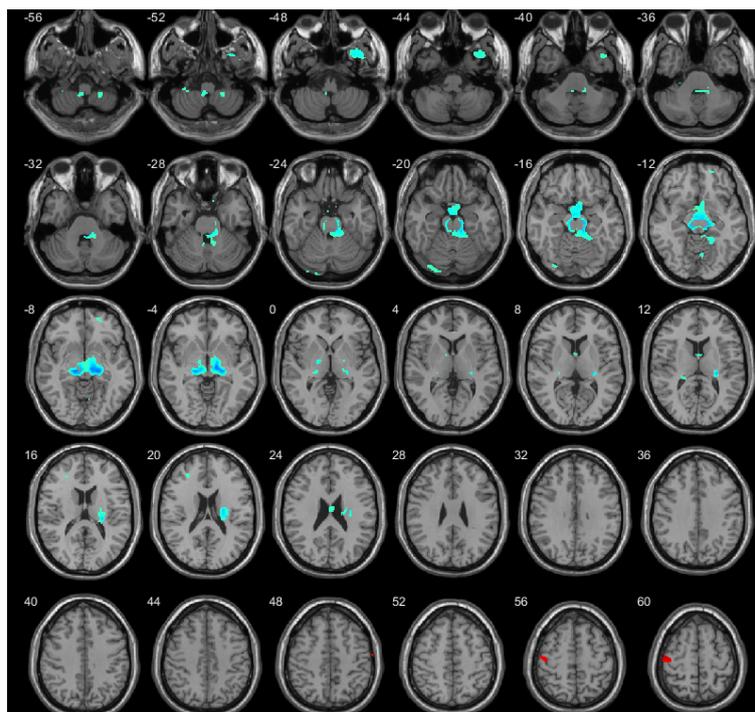


Figura 22: Mapa de colores de la vista transversal de los resultados de VBM aplicados con pruebas t para dos muestras entre los grupos. Uno está conformado por 226 pacientes masculinos con la EP, y el otro grupo contiene 86 controles masculinos. El valor p fue seleccionado como 0.001. Los vóxeles de color rojo corresponden a los valores t positivos, y los vóxeles de color azul a los valores t negativos. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

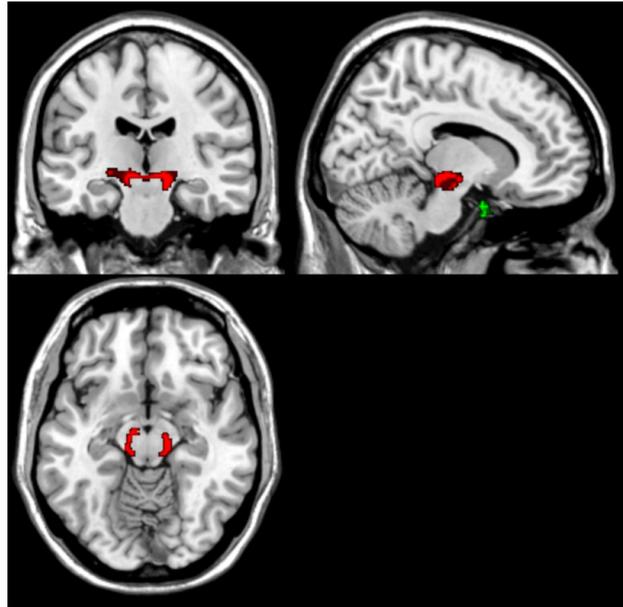


Figura 23: Las dos RDI con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo femenino. En la figura se pueden ver las RDI 6 (mostrada en rojo) y 14 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

ventrículo derecho e izquierdo (RDI mostrada en verde). Las RDI de las IRMe de 3 Teslas incluyen la unión de los dos hemisferios del córtex superior (RDI 2 mostrada en rojo), la parte inferior del cerebelo (RDI 7 mostrada en verde), y región del córtex (RDI 14 mostrada en amarillo).

4.2.3 Resultados de la clasificación

Para reducir la complejidad de los modelos, la selección de características se realiza antes de entrenar el algoritmo de clasificación. Este es un paso necesario para encontrar el mejor grupo de características específico para cada uno de los algoritmos de clasificación, así como para reducir la dimensionalidad del modelo. En la Tabla 21, el número original de características extraídas de las RDI en las IRMe de ambos sexos son presentadas. Originalmente, el número de características extraídas de las IRMe de participantes del sexo masculino eran de 2660 y en el caso de las participantes de sexo femenino, ascendía a 3000 características por paciente. Este número se reduce utilizando ACP seguido de SGCBW. El número final de características depende del algoritmo de clasificación y varía entre 3 y 30 características.

En las Tablas 22 y 23, se muestran los resultados de los clasificadores con diferentes grupos de características de ambos sexos y los diferentes campos magnéticos (1.5 Teslas y 3 Teslas respectivamente). Los clasificadores que se utilizaron incluyen la regresión logística, BA, RB, KNN, PM y MSV.

En las tablas 24 y 25 se muestran las características que fueron seleccionadas por los algoritmos

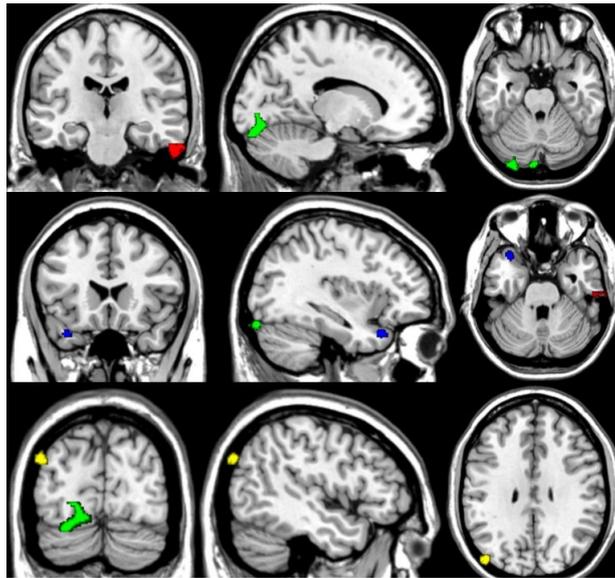


Figura 24: Las cuatro RDI con las características más seleccionadas en las IRMe de 3 Teslas de participantes de sexo femenino. En la figura se pueden ver las RDI 1 (mostrada en rojo) y 14 (mostrada en verde) en la primera fila. Las RDI 18 (azul) y 19 (amarillo) se localizan en la segunda y tercera fila respectivamente. (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

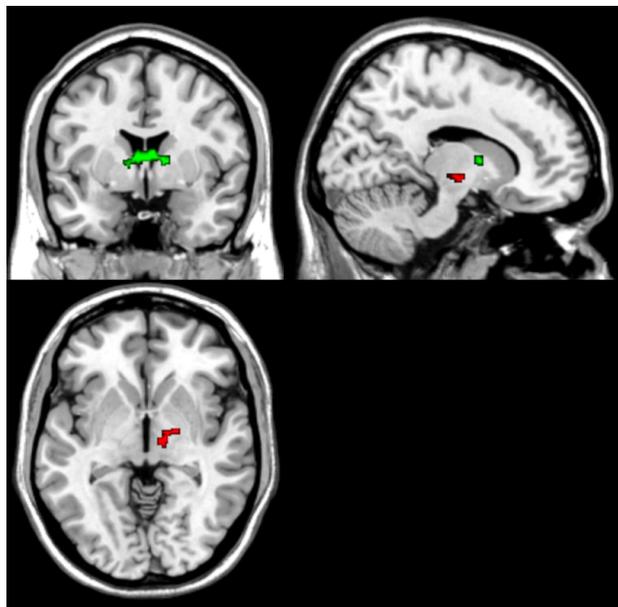


Figura 25: Las dos RDI con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo masculino. En la figura se pueden ver las RDI 4 (mostrada en rojo) y 6 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

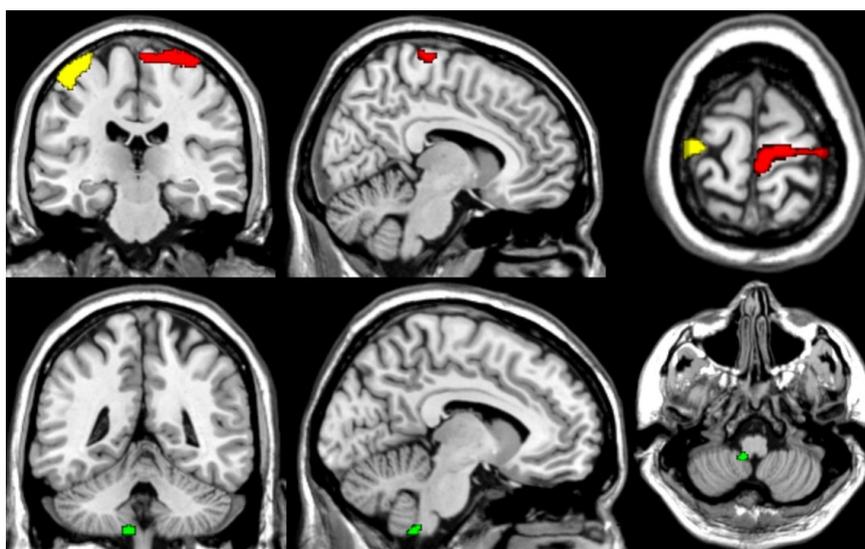


Figura 26: Las tres RDI con las características más seleccionadas en las IRMe de 1.5 Teslas de participantes de sexo masculino. En la figura se pueden ver las RDI 2 (mostrada en rojo) y 14 (mostrada en amarillo) en la primera fila. En la segunda fila, se observa la región 7 (mostrada en verde). (para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de la tesis). De (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC.

Tabla 21: Número de características extraídas en cada uno de los grupos, y el número de características seleccionados después de aplicar ACP y correr el algoritmo de SGCBW para cada clasificador.

Número de características	Clasificador	Hombres		Mujeres	
		1.5 T	3 T	1.5 T	3 T
Extraídas		2660	3192	3192	6384
Seleccionadas por el ACP		1582	1911	1916	2180
Seleccionado por SGCBW	Logístico	19	30	7	16
	BA	3	4	7	6
	BI	21	24	9	10
	RB	8	6	6	4
	KNN	9	8	9	2
	PM	6	3	9	6
	MSV	3	6	6	2

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC

Tabla 22: Resultados de la detección de la EP en participantes de sexo masculino y femenino con siete clasificadores diferentes y IRMe de 1.5 Teslas.

Clasificador	Población	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
Logística	Masculino	0.960	0.974	0.913	0.974	0.887
	Femenino	0.968	0.972	0.961	0.972	0.934
BA	Masculino	0.921	0.974	0.739	0.927	0.766
	Femenino	0.871	0.972	0.731	0.833	0.742
BI	Masculino	0.990	0.987	1	1	0.972
	Femenino	0.871	0.889	0.846	0.889	0.735
RB	Masculino	0.931	0.949	0.87	0.961	0.806
	Femenino	0.887	1	0.731	0.837	0.782
KNN	Masculino	0.931	0.974	0.783	0.938	0.7965
	Femenino	0.903	0.944	0.846	0.895	0.801
PM	Masculino	0.980	0.987	0.956	0.987	0.941
	Femenino	0.935	0.917	0.961	0.971	0.871
MSV	Masculino	0.900	0.895	0.933	0.987	0.702
MSV	Femenino	0.855	0.917	0.769	0.846	0.701

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC

de clasificación para imágenes de 1.5T y 3T respectivamente. También se muestra el porcentaje que cada una de las características aporta al total de las características seleccionadas en los subconjuntos obtenidos con el SGCBW. Finalmente el porcentaje de características que son derivadas de los histogramas que representan el análisis estadístico de primer orden y el número de características que provienen de las características de Haralick, las cuales representan un análisis estadístico de segundo orden.

4.2.4 Discusión

En esta sección del capítulo, mostramos los resultados de la metodología propuesta mostrada en la Figura 10, para detectar la EP en IRMe. En este trabajo, aprovechamos la gran cantidad de datos disponibles en el conjunto de datos de la base de datos de PPMI, que contiene suficientes imágenes para realizar estudios separados para participantes femeninos y masculinos, lo que suele ser una limitación en estudios anteriores que no analizaron estas diferencias debido al reducido número de imágenes en los conjuntos de datos. También realizamos el proceso de VBM con las imágenes disponibles de pacientes prodrómicos para observar los RDI en las primeras fases de la enfermedad. El número de pacientes y controles se reorganizó para lograr un equilibrio y evitar sesgos en el entrenamiento hacia una de las clases.

Tabla 23: Características de las IRMe de 1.5T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick

Clasificador	Población	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
Logística	Masculino	0.916	0.928	0.889	0.947	0.808
	Femenino	0.933	0.951	0.895	0.951	0.845
BA	Masculino	0.844	0.922	0.681	0.86	0.632
	Femenino	0.79	0.914	0.526	0.804	0.489
BI	Masculino	0.84	0.843	0.833	0.915	0.652
	Femenino	0.8403	0.901	0.710	0.869	0.625
RB	Masculino	0.747	0.837	0.555	0.8	0.652
	Femenino	0.706	0.889	0.316	0.735	0.250
KNN	Masculino	0.813	0.863	0.708	0.863	0.571
	Femenino	0.739	0.852	0.5	0.784	0.374
PM	Masculino	0.862	0.935	0.708	0.872	0.674
	Femenino	0.857	0.981	0.658	0.856	0.695
MSV	Masculino	0.956	0.993	0.875	0.944	0.897
MSV	Femenino	0.781	0.757	1	1	0.488

Adaptado de (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC

Tabla 24: Características de las IRMe de 1.5T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick

IRMe 1.5T Sexo Masculino		IRMe 1.5T Sexo Femenino	
Características seleccionadas		Características seleccionadas	
Segundo Momento Angular	16.66%	Autocorrelación	16.66%
Disimilitud	13.88%	Suma de Promedios	16.66%
Entropía	13.88%	Segundo Momento Angular	12.5%
Suma de Promedios	11.11%	Diferencia de Varianzas	8.33%
Medidas de Correlación de la Información	8.33	Prominencia de Grupos	8.33%
Suma de Cuadrados	8.33%	Entropía	8.33%
Promedio del Histograma	5.55%	Diferencia de Entropías	4.16%
Energía	5.55%	Suma de Entropías	4.16%
Prominencia de Grupos	5.55%	Momento Central del Histograma	4.16%
Momento Central Del Histograma	2.77%	Disimilitud	4.16%
Desviación Estándar del Histograma	2.77%	Medida de Correlación de la información	4.16%
Correlación	2.77%	Energía	4.16%
Autocorrelación	2.77%	Momento de Diferencia Inversa	4.16%
Momentos de Diferencia Inversa	2.77%		
Sombra de Grupos	2.77%		
Características de Histogramas:	10.53%		4.17%
Características de Haralick:	89.47%		95.83%

Para detectar la EP, realizamos experimentos de acuerdo a la metodología mostrada en la Figura 10. Tras el preprocesamiento de las imágenes, los resultados del VBM muestran que hay varias RDI que contienen diferencias significativas entre los grupos de participantes control y pacientes con la EP. Las Figuras 14, 15 y 16 muestran los resultados del VBM en pacientes prodrómicos masculinos comparándolos con el grupo control. Las RDI se localizan en diferentes regiones del cerebro, incluyendo la corteza premotora y el giro temporal, que están relacionadas con el sistema motor extrapiramidal mostrado en la Figura 4. Se trata de un hallazgo interesante porque sugiere que el agotamiento de la dopamina afecta a distintas regiones del cerebro alejadas de la SN, incluso en las fases iniciales, y estas regiones están relacionadas con la articulación y fonación, así como la planeación del habla.

Las Figuras 17, 18 y 19 muestran las RDI que se encontraron tras ejecutar el algoritmo de VBM. En estas figuras, observamos que las regiones contienen diferencias entre los dos grupos. Las RDI en estas imágenes se localizan en el lóbulo occipital, el tálamo, los ganglios basales, una pequeña parte

Tabla 25: Características de las IRMe de 3T más seleccionadas por los algoritmos de clasificación con los mejores resultados en bases de datos con pacientes de sexo masculino y en bases de datos de sexo femenino. Se muestra el porcentaje de características seleccionadas en los subconjuntos, aunque pueden pertenecer a diferentes RDI. Al final de la tabla se muestra el porcentaje de características que pertenecen a histogramas y el porcentaje que pertenecen a las características de Haralick

IRMe 3T Sexo Masculino		IRMe 3T Sexo Femenino	
Características seleccionadas		Características seleccionadas	
Momento de Diferencia Inversa	20%	Medida de correlación de la información	34.14%
Medidas de Correlación de la Información	11.42%	Contraste	12.19%
Entropía	11.42%	Disimilitud	9.75%
Autocorrelación	8.57%	Probabilidad Máxima	7.31%
Suma de Varianzas	8.57%	Suma de Promedios	7.31%
Momento Central del Histograma	5.71%	Segundo Mometo Angular	7.31%
Probabilidad Máxima	5.71%	Suma de Varianzas	7.31%
Disimilitud	5.71%	Homogeneidad	4.87%
Contraste	5.71%	Diferencia de Entropía	2.43%
Suma de Cuadrados	2.85%	Suma de Cuadrados	2.43%
Suma de Varianzas	2.85%	Entropía	2.43%
Diferencia de Entropía	2.85%	Energía	2.43%
Características de Histogramas:	5.74%		0%
Características de Haralick:	94.28%		100%

del cerebelo y el lóbulo frontal. Por otro lado, el grupo que contiene pacientes varones con la EP y controles sanos, muestra que las RDI que contienen diferencias significativas incluyen los ganglios basales, el tronco encefálico, el cuarto ventrículo, el ventrículo lateral, el cerebelo, el lóbulo frontal, el lóbulo temporal y el córtex, como puede observarse en las figuras 20, 21 y 22.

Entre sexos, existen importantes diferencias en las RDI mostradas en las Figuras anteriores. Podemos observar que las regiones afectadas están más próximas a la SN en las participantes femeninas, mientras que en los participantes masculinos, las regiones están más diseminadas e incluso alcanzan el córtex. Esto ocurre también en el grupo prodrómico de pacientes varones. Otro hallazgo es que las RDI en el grupo de participantes varones son mayores. Esto puede estar relacionado con el hecho de que el número de participantes varones diagnosticados con la EP es casi el doble que el de mujeres.

Tras ejecutar los algoritmos de ACP y SGCBW, comprobamos que el número de RDI utilizadas para la detección de la EP con aprendizaje automático se reduce en comparación con los precedentes del VBM. En la Figura 23, podemos ver las RDI seleccionadas en las imágenes de las participantes tomadas con un imán de 1.5 Tesla. Las regiones seleccionadas incluyen las regiones del tálamo y

del tronco cerebral. En la Figura, podemos ver que las RDI se concentran en la región central del cerebro cerca del SN. Lo mismo ocurre en las imágenes de 1.5 Tesla tomadas a participantes varones, como se muestra en la Figura 25, donde la región alrededor del putamen y el tálamo son relevantes. Estas regiones están relacionadas con la generación de movimientos involuntarios e instintivos, y cuya deficiencia se asocia a los síntomas de la EP.

En las Figuras 24 y 26 se muestran las RDI seleccionados por el ACP y SGCBW en las imágenes tomadas con un imán de 3 Tesla (mayor resolución) en participantes femeninos y masculinos respectivamente. Las regiones seleccionadas en el grupo femenino de participantes se localizan en el cerebelo, una pequeña región de la corteza y las circunvoluciones temporales inferior y superior, que se han asociado con la conciencia espacial y la estimulación visual. En el caso del grupo de participantes varones, las regiones seleccionadas incluyen la corteza premotora, el lóbulo parietal superior y el lóbulo cerebeloso.

Todas estas regiones muestran un aumento o una reducción del grosor de los tejidos del cerebro en comparación con el grupo control. Dado que la mayoría de las regiones pertenecen al sistema motor extrapiramidal, puede estar relacionado con la depleción de dopamina originada en el SN. Es importante estudiar estas regiones del cerebro, aunque estén alejadas de la región del estriado, ya que puede mejorar la detección de la EP y podría aportar nuevos conocimientos sobre el desarrollo de las primeras fases de la enfermedad.

Utilizamos múltiples algoritmos de aprendizaje automático y validación cruzada de 10 veces para probar el rendimiento en la clasificación de pacientes con la EP, como se muestra en las Tablas 22 y 23. El mejor algoritmo para detectar la enfermedad en las imágenes de 1.5 Tesla del grupo de participantes varones fue el BI, con una exactitud de 0.990, seguido de la RNA, con 0.987. En el caso de las participantes de sexo femeninos, los mejores resultados se obtuvieron con el clasificador *Logístico*, con una exactitud de 0.967, seguido de la RNA con una exactitud de 0.935. En este caso, el clasificador obtuvo resultados mucho mejores en la detección de la EP en el grupo de pacientes varones en comparación con el grupo de mujeres. Esto puede estar relacionado con el mayor número de imágenes para el entrenamiento y las RDI seleccionados en los mejores subconjuntos de características, que contienen regiones más cercanas en comparación con las regiones seleccionadas para el grupo de participantes masculinos.

En el caso de la detección de la EP con imágenes tomadas con imanes de 3 Tesla para mayor resolución, probamos los mismos clasificadores para ambos grupos. La figura 23 muestra los resultados para ambos sexos. En el caso del grupo de participantes masculinos, el algoritmo que obtuvo el mejor rendimiento fue la MSV con una exactitud de 0.956, seguido del clasificador logístico con una exactitud de 0.916. En el caso del grupo de participantes femeninos, el clasificador logístico fue el algoritmo con el mejor rendimiento, con una exactitud de 0.933, seguido de la RNA con una exactitud de 0.857. Estos

resultados muestran una disminución del rendimiento en comparación con los obtenidos con imágenes de 1.5 Tesla. Consideramos que la disminución en la exactitud podría deberse al menor tamaño del conjunto de datos.

Por último, descubrimos que los subconjuntos óptimos de características seleccionados por el algoritmo de SGCBW consisten principalmente en características estadísticas de segundo orden, lo cual se puede observar en las tablas 24 y 25, así como se observa que las características de primer orden son significativamente más importantes para la clasificación en participantes de sexo masculino tanto para IRMe de 1.5T como de 3T. Esto es de esperar porque estas características pueden identificar patrones que muestran la relación entre la intensidad de nivel de gris de los píxeles con sus vecinos en múltiples direcciones. Dado que las IRMe contienen información en 3 dimensiones, las características estadísticas de segundo orden (características Haralick) pueden identificar patrones más complejos. Así mismo en las tablas 24 y 25 se puede observar que las características de Medidas de Correlación de la Información juegan un papel importante en las IRMe de 3T en ambos sexos, Mientras que en las de 1.5T los subconjuntos para ambos sexos contienen características diferentes.

Capítulo 5 Conclusión y Trabajo Futuro

La EP es una enfermedad neurodegenerativa que resulta muy complicada de diagnosticar correctamente y la cual afecta al 1% de la población mayor de 60 años. En esta tesis, presentamos dos metodologías, una para detectar la EP en grabaciones de voz y la segunda para detectar la EP en IRMe. Estos dos enfoques pueden ayudar a los médicos a diagnosticar la enfermedad utilizando herramientas no invasivas y ampliamente disponibles con alta exactitud y bajo costo para el paciente. Las metodologías propuestas en esta tesis tienen como objetivo lograr una alta exactitud de detección y proporcionar información acerca del análisis llevado a cabo para lograr una mejor interpretabilidad de los resultados, en comparación con los algoritmos de “caja negra”.

Los estudios neurológicos han mostrado diferencias entre los cerebros de participantes de sexo femenino y sexo masculino. En esta tesis, dividimos los experimentos por sexo, aprovechando el tamaño de las bases de datos proporcionadas por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019) en el caso de las grabaciones de voz y el PPMI para IRMe. Esto nos permite comprender mejor qué características de la voz de los pacientes contienen información relevante para la detección de la EP. En el caso del análisis de la IRMe, se identifican las características y regiones del cerebro que pueden explicar los cambios estructurales en comparación con el grupo de control. La decisión de dividir los conjuntos de datos por sexo permite encontrar los modelos que funcionan mejor para cada grupo.

Las métricas de rendimiento de la metodología propuesta para detectar la EP con grabaciones de voz y con sMRI se obtienen con validación cruzada de 10 veces. Los mejores resultados de precisión para la clasificación del conjunto de datos con ambos sexos es de 0.9505 en un conjunto de datos equilibrado, 0.959 en un conjunto de datos desequilibrado con participantes femeninos únicamente, y 0.9436 en un conjunto de datos equilibrado con participantes masculinos únicamente, todos ellos con el clasificador KNN. En el caso de la clasificación de IRMe, encontramos que se alcanza una precisión de 0.9677 con un clasificador logístico en un conjunto de datos con participantes femeninos únicamente, y una precisión de 0.99 con un clasificador BI en un conjunto de datos con participantes masculinos únicamente.

A partir del proceso de reducción de características y de la selección de los mejores subconjuntos para entrenar los algoritmos de clasificación, encontramos las características más relevantes para

detectar la EP en la voz de los pacientes. Las características seleccionadas a partir del análisis de las grabaciones de voz muestran que las frecuencias más altas son más útiles para detectar la enfermedad en las pacientes femeninas, y las frecuencias más bajas son más útiles para los pacientes masculinos, lo cual es un resultado esperado debido a las diferencias en el rango medio de frecuencias de la voz de cada sexo. Encontramos que las características procedentes del grupo de CCFM son más relevantes para el grupo de pacientes femeninos, y el grupo de características TDO es más relevante para el grupo de pacientes masculinos. Sin embargo, para ambos grupos, los rasgos más importantes provienen del grupo TOFQA, lo cual fue reportado previamente por (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019), y se confirma en este trabajo. En el caso de la IRMe, encontramos los RDI para pacientes varones prodrómicos, pacientes varones y pacientes mujeres tanto en IRMe de 1.5 Teslas como en IRMe de 3 Teslas. Estas regiones varían en cada grupo y resolución de imagen, lo que es importante a la hora de diagnosticar la enfermedad con estas herramientas. En el caso de las pacientes femeninas, las RDI más importantes incluyen el tálamo, el tronco encefálico, el cerebelo, una pequeña región de la corteza, así como la circunvolución temporal inferior y superior. En el caso de los pacientes varones, las RDI incluyen la región alrededor del putamen, el tálamo, la corteza pre-motora, el lóbulo parietal superior y el lóbulo cerebeloso. Se trata de hallazgos relevantes porque las RDI seleccionadas están contenidas en el sistema motor extrapiramidal mostrado en la Figura 4, que está relacionado con las vías de la dopamina en el cerebro, y por tanto con la EP.

Llegamos a la conclusión de que es posible ayudar a la detección de la EP con alta precisión mediante el uso de algoritmos de aprendizaje automático y señales biomédicas como grabaciones de voz e IRMe. Los resultados sugieren que tanto para las grabaciones de voz como para la IRMe, el mejor subconjunto de características para la clasificación depende del sexo del paciente. El entrenamiento de los algoritmos para cada sexo específico puede aumentar el rendimiento y la interpretabilidad de los resultados. Estas dos señales biomédicas, en conjunto, proporcionan una alternativa a técnicas invasivas y más costosas, como el DaTscan. Las metodologías propuestas en este trabajo pueden encontrar el mejor subconjunto de características para la tarea de clasificación y dar a los algoritmos un mayor nivel de interpretabilidad en comparación con los resultados binarios producidos por los algoritmos de aprendizaje automático.

5.1 Contribuciones

Las principales aportaciones de esta tesis son las dos metodologías para detectar la EP con grabaciones de voz y señales de IRMe. Estas metodologías obtienen subconjuntos reducidos de características de las señales originales adaptadas a clasificadores específicos de aprendizaje automático, reduciendo así la complejidad de los modelos y aumentando la interpretabilidad de los resultados.

Otra contribución de esta tesis son los resultados específicos para cada sexo de la detección de la EP en los mayores conjuntos de datos disponibles públicamente para grabaciones de voz e IRMe. En la literatura, pocos trabajos abordan la detección de la EP de forma específica para cada sexo. En este trabajo, aprovechamos el tamaño de los conjuntos de datos y mostramos resultados no sólo de los conjuntos de datos específicos de cada sexo, sino también de los conjuntos de datos equilibrados y desequilibrados en el caso de las grabaciones de voz.

Obtuvimos los resultados de mayor exactitud en la detección de la EP en grabaciones de voz utilizando el conjunto de datos de (C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al., 2019) en el momento de publicación del artículo (Gabriel Solana-Lavalle, 2020). En el caso de las IRMe obtenemos resultados comparables al estado del arte en el momento de la publicación del artículo (Solana-Lavalle & Rosas-Romero, 2021b) y mostramos el rendimiento de clasificación de ambos sexos.

Por último, reportamos las familias de características que más contribuyen a la clasificación de las grabaciones de voz en función del sexo de los participantes, y descubrimos que las características CCFM contribuyen sobre todo a la clasificación de los pacientes femeninos, mientras que las características TDO contribuyen en mayor medida a la clasificación de los pacientes masculinos. En el caso de las IRMe, mostramos que las RDI que contienen patrones diferenciables en el cerebro de los pacientes los cuales son diferentes para ambos sexos y resoluciones de las IRMe. Estos hallazgos pueden utilizarse para la interpretación clínica del enfoque basado en datos que proponemos.

5.2 Limitaciones de la investigación

Es importante señalar que las metodologías propuestas utilizan el algoritmo SGCBW, el cual puede tomar mucho tiempo para encontrar un subconjunto de características óptimo. Esto se mitiga reduciendo la dimensionalidad con el algoritmo de ACP en la metodología IRMe. Sin embargo, en el caso de la metodología para la clasificación de grabaciones de voz, el principal algoritmo de reducción de características es el SGCBW.

No se puede garantizar que las RDI que se muestran en las IRMe estén perfectamente delineadas, ya que pueden existir vóxeles que resultaron relevantes en el análisis de VBM, pero que en realidad son falsos positivos que se encuentran en estas regiones.

En la validación de 10 veces de la clasificación de grabaciones de voz puede existir la fuga de datos en los resultados debido a que se tienen múltiples grabaciones de voz de cada paciente con la EP y sujetos control, por lo que se pueden tener grabaciones provenientes de un mismo paciente tanto en los grupos de entrenamiento y los grupos de prueba. Para mostrar resultados en los cuales se ataca este problema, se añadieron 36 tablas en el Apéndice *a* en las cuales se muestran los resultados de la clasificación en bases de datos que únicamente tienen una grabación por cada paciente y cada sujeto

control para cada grupo (sexo masculino, sexo femenino, base de datos mixta).

Los resultados mostrados en esta tesis están limitados por el tamaño de los conjuntos de datos disponibles públicamente, y pueden contener algún sesgo debido a la falta de diversidad de los participantes. Por esta razón, no podemos asegurar la generalización de los algoritmos en poblaciones diferentes a los pacientes que participaron en el estudio. Este problema es una limitación común en el área de la investigación médica y debe tenerse en cuenta antes de implementar las metodologías propuestas en aplicaciones con participantes reales.

La interpretación médica y biológica de los resultados de esta tesis no procede de expertos médicos. Un especialista en la materia debería confirmarlos y aportar su punto de vista.

5.3 Trabajo a Futuro

Los resultados de esta tesis pueden ser validados probando los algoritmos entrenados con diferentes conjuntos de datos y poblaciones. Asimismo, las metodologías de esta tesis podrían adaptarse para estudios longitudinales aprovechando las imágenes de pacientes disponibles en los conjuntos de datos proporcionados por el PPMI. Mediante el seguimiento de la progresión de la enfermedad y la detección del estado actual del paciente, las IRMe podrían convertirse en una herramienta más útil para los médicos.

Las futuras investigaciones sobre la detección de la EP y la búsqueda de nuevos biomarcadores podrían beneficiarse del análisis de las IRMf y la DT-MRI, ya que estas técnicas de imagen proporcionan información sobre la actividad, activación y conectividad del cerebro. Si estas técnicas demuestran su utilidad fuera del ámbito académico, especialmente las IRMf, podrían considerarse nuevos biomarcadores de la enfermedad y aportar información relevante para su tratamiento y seguimiento de su progresión.

Por último, el *aprendizaje profundo* puede ser una herramienta poderosa para detectar la EP con una exactitud mayor a las mostradas en este trabajo, sacrificando cierta interpretabilidad de los resultados. El área de la Inteligencia Artificial explicable podría producir nuevos algoritmos y metodologías que podrían implementar RNA que pueden resolver el problema de la detección de la EP con alta precisión al mismo tiempo que proporcionan información relevante para una mejor comprensión de la enfermedad.

Bibliografía

- Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication, 116*, 56–76.
- Alirezanejad, M., Enayatifar, R., Motameni, H., & Nematzadeh, H. (2020). Heuristic filter feature selection methods for medical datasets. *Genomics, 112*(2), 1173–1181.
- Amoroso, N., La Rocca, M., Monaco, A., Bellotti, R., & Tangaro, S. (2018). Complex networks reveal early MRI markers of Parkinson's disease. *Medical image analysis, 48*, 12-24. doi: 10.1016/j.media.2018.05.004
- Arocho-Quinones, E. V., Hammer, M. J., Bock, J. M., & Pahapill, P. A. (2017). Effects of deep brain stimulation on vocal fold immobility in parkinson's disease. *Surgical neurology international, 8*.
- A schrag, N. Q., Y Ben-Shlomo. (2002). How valid is the clinical diagnosis of parkinson's disease in the community? *Journal of Neurology Neurosurgery and Psychiatry, 73*, 529-534. (doi: 10.1136/jnnp.73.5.529)
- Ashburner, J., & Friston, K. (2007). Voxel-based morphometry. *Statistical parametric mapping: the analysis of functional brain images, 92–98*.
- Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—the methods. *Neuroimage, 11*(6), 805–821.
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26*(3), 839–851.
- B. E. Sakar, M. E. Isenkul, C. O. Sakar, et al. (2013). Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform., 19*(4), 828-834. doi: 10.1109/JBHI.2013.2245674
- Bandhyopadhyay, D. S. K., & Paul, T. U. (2012). Segmentation of brain mri image—a review. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(3).
- Bang, Y.-I., Min, K., Sohn, Y. H., & Cho, S.-R. (2013). Acoustic characteristics of vowel sounds in patients with parkinson disease. *NeuroRehabilitation, 32*(3), 649–654.
- B. E. Sakar, C. O. S., G. Serbes. (2017). Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease. *PLOS ONE, 1-18*. doi: 10.1371/journal.pone.0182428
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4) (No. 4). Springer.
- Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., & Del Tredici, K. (2004). Stages in the development of parkinson's disease-related pathology. *Cell and tissue research, 318*(1), 121–134.
- Bryce, R., & Sprague, K. (2012). Revisiting detrended fluctuation analysis. *Scientific reports, 2*(1),

1–6.

- Burrus, C. S. (1997). Introduction to wavelets and wavelet transforms: a primer. *Englewood Cliffs*.
- C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, et al. (2019). A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform. *Applied Soft Computing*, 74, 255-263. doi: 10.1016/j.asoc.2018.10.022
- Chahid, A. (2020). *Pre-processing and feature extraction methods for smart biomedical signal monitoring: Algorithms and applications* (Unpublished doctoral dissertation).
- Collewet, G., Strzelecki, M., & Mariette, F. (2004). Influence of mri acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance imaging*, 22(1), 81–91.
- A computational anatomy toolbox for spm*. (n.d.).
- D. Braga, L. C. R. A., A. M. Madureira. (2019). Automatic detection of parkinson's disease based on acoustic analysis of speech. *Engineering Applications of Artificial Intelligence*, 77, 148-158. doi: 10.1016/j.engappai.2018.09.018
- DESA, U. (2015). United nations department of economic and social affairs, population division. world population prospects: The 2015 revision, key findings and advance tables. In *Technical report*. Working Paper No. ESA/P/WP. 241.
- Esiri, M. M. (2010). Pro: Can neuropathology really confirm the exact diagnosis? *Alzheimer's research & therapy*, 2(2), 10.
- Esmailzadeh, S., Yang, Y., & Adeli, E. (2018). End-to-end Parkinson disease diagnosis using brain MR-images by 3D-CNN. *Computer Vision and Pattern Recognition*, 1-7.
- Factor, S. A., & Weiner, W. (2007). *Parkinson's Disease: Diagnosis & Clinical Management*. Demos Medical Publishing.
- Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. In *Eighth annual conference of the international speech communication association*.
- Fix, E., & Hodges, J. L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 57(3), 238–247.
- Foundation, M. J. F. (2021). *Parkinson's progression marker initiative*. <https://www.ppmi-info.org/about-ppmi/about-pd/>. (Accessed: 2021-03-06)
- Gabriel Solana-Lavalle, R. R.-R., Juan-Carlos Galan-Hernandez. (2020, jan). Automatic parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybernetics and Biomedical Engineering*, 40(1), 505–516. doi: 10.1016/j.bbe.2020.01.003

- Gargouri, F., Kallel, F., Delphine, S., Ben Hamida, A., Lehericy, S., & Valabregue, R. (2018). The influence of preprocessing steps on graph theory measures derived from resting state fmri. *Frontiers in computational neuroscience*, 12, 8.
- Gaser, C., & Dahnke, R. (2016). Cat-a computational anatomy toolbox for the analysis of structural mri data. *Hbm*, 2016, 336–348.
- Goedert, M., & Compston, A. (2018). Parkinson's disease - The story of an eponym. *Nature Reviews Neurology*, 14(1), 57–63. Retrieved from <http://dx.doi.org/10.1038/nrneuro1.2017.165> doi: 10.1038/nrneuro1.2017.165
- Guruler, H. (2017). A novel diagnosis system for parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural. Comput. & Applic.*, 28, 1657-1666. doi: 10.1007/s00521-015-2142-2
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*(6), 610–621.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
- Heim Beatrice, D. M. R. S. K., Krismer Florian. (2017). Magnetic resonance imaging for the diagnosis of parkinson's disease. *Journal of neural transmission*, 124(8), 915–964.
- Hossan, M. A., Memon, S., & Gregory, M. A. (2010). A novel approach for mfcc feature extraction. In *2010 4th international conference on signal processing and communication systems* (pp. 1–5).
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- Josephson, C. B., & Wiebe, S. (2020). Precision medicine: Academic dreaming or clinical reality? *Epilepsia*.
- Kherif, F., & Latypova, A. (2020). Principal component analysis. In *Machine learning* (pp. 209–225). Elsevier.
- Kohavi, R., John, G. H., et al. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273–324.
- Kudo, E., & Fujii, Y. (2012). *Dopamine: Functions, regulation and health effects*. Nova Science Publishers.
- Kurth, F., Gaser, C., & Luders, E. (2015). A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (spm). *Nature protocols*, 10(2), 293–304.
- Lei, H., Zhao, Y., Wen, Y., Luo, Q., Cai, Y., Liu, G., & Lei, B. (2018). Sparse feature learning for multi-class Parkinson's disease classification. *Technology and Health Care : Official Journal of the European Society for Engineering and Medicine*, 26(1), 193-203. doi: 10.3233/thc-174548

- Little, M., McSharry, P., Roberts, S., Costello, D., & Moroz, I. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nature Precedings*, 1–1.
- Livingston, F. (2005). Implementation of breiman’s random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper*, 1–13.
- Lorio Sara, B. A. D. B. D. J., Sambataro Fabio. (2019). The Combination of DAT-SPECT, Structural and Diffusion MRI Predicts Clinical Progression in Parkinson’s Disease. *Frontiers in Aging Neuroscience*, 11, 57. Retrieved from <https://www.frontiersin.org/article/10.3389/fnagi.2019.00057> doi: 10.3389/fnagi.2019.00057
- Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T., & Garpebring, A. (2019). Gray-level invariant Haralick texture features. *PloS one*, 14(2), e0212110.
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., Lancaster, J., et al. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*, 2(2), 89–101.
- Michael J. Fox Foundation. (2020). *Parkinson’s disease causes*. Retrieved from: <https://www.michaeljfox.org/understanding-parkinsons/living-with-pd.html>.
- Mikl, M., Mareček, R., Hlušík, P., Pavlicová, M., Drastich, A., Chlebus, P., . . . Krupa, P. (2008). Effects of spatial smoothing on fmri group inferences. *Magnetic resonance imaging*, 26(4), 490–503.
- Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted random forest. *IEICE TRANSACTIONS on Information and Systems*, 98(9), 1630–1636.
- Moustafa, A. A., Chakravarthy, S., Phillips, J. R., Gupta, A., Keri, S., Polner, B., . . . Jahanshahi, M. (2016). Motor symptoms in parkinson’s disease: A unified framework. *Neuroscience & Biobehavioral Reviews*, 68, 727–740.
- Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883.
- Podder, P., Khan, T. Z., Khan, M. H., & Rahman, M. M. (2014). Comparative performance analysis of hamming, hanning and blackman window. *International Journal of Computer Applications*, 96(18).
- Pramanik, M., Pradhan, R., Nandy, P., Bhoi, A. K., & Barsocchi, P. (2021). Machine learning methods with decision forests for parkinson’s detection. *Applied Sciences*, 11(2), 581.
- Quan, C., Ren, K., & Luo, Z. (2021). A Deep Learning Based Method for Parkinson’s Disease Detection Using Dynamic Features of Speech. *IEEE Access*, 9, 10239–10252. doi: 10.1109/ACCESS.2021.3051432
- Qureshi, Y. H., Baez, P., & Reitz, C. (2020). Endosomal Trafficking in Alzheimer’s Disease, Parkinson’s Disease, and Neuronal Ceroid Lipofuscinosis. *Molecular and Cellular Biology*, 40(19). doi: 10.1128/mcb.00262-20
- Raj, R. J. S., Shobana, S. J., Pustokhina, I. V., Pustokhin, D. A., Gupta, D., & Shankar, K. (2020). Op-

- timal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access*, 8, 58006–58017.
- Rizvi, D. R., Nissar, I., Masood, S., Ahmed, M., & Ahmad, F. (2020). An lstm based deep learning model for voice-based detection of parkinson's disease. *International Journal of Advanced Science and Technology*, 29(5), 8.
- Rosen, K. M., Kent, R. D., & Duffy, J. R. (2005). Task-based profile of vocal intensity decline in parkinson's disease. *Folia Phoniatrica et Logopaedica*, 57(1), 28–37.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533–536.
- Saeed, U., Compagnone, J., Aviv, R. I., Strafella, A. P., Black, S. E., Lang, A. E., & Masellis, M. (2017). Imaging biomarkers in parkinson's disease and parkinsonian syndromes: current and emerging concepts. *Translational neurodegeneration*, 6(1), 8.
- Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... Kursun, O. (2013). Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828-834.
- Savoiaro, M. (2003). Differential diagnosis of parkinson's disease and atypical parkinsonian disorders by magnetic resonance imaging. *Neurological Sciences*, 24(1), s35–s37.
- Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. *Neural computing and applications*, 31(1), 171–188.
- Selesnick, I. W. (2011). Wavelet transform with tunable q-factor. *IEEE transactions on signal processing*, 59(8), 3560–3575.
- Shah, P. M., Zeb, A., Shafi, U., Zaidi, S. F. A., & Shah, M. A. (2018). Detection of Parkinson disease in brain MRI using Convolutional Neural Network. In *2018 24th international conference on automation and computing (icac)* (p. 1-6).
- Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingalthalika, M. (2019). Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *Neuroimage: Clinical*, 22, 101748. doi: 10.1016/j.nicl.2019.101748
- Sigurdsson, S., Petersen, K. B., & Lehn-Schiøler, T. (2006). Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music. In *Ismir* (pp. 286–289).
- Sivaranjini, S., & Sujatha, C. (2019). Deep learning based diagnosis of parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, 1–13.
- Sivaranjini, S., & Sujatha, C. M. (2020). Deep learning based diagnosis of Parkinson's disease using

- convolutional neural network. *Multimedia Tools and Applications*, 79, 15467–15479. doi: 10.1007/s11042-019-7469-8
- Solana-Lavalle, G., & Rosas-Romero, R. (n.d.). Analysis of voice and magnetic resonance images to assist diagnosis of parkinson’s disease with machine learning. *Technology, Science and Culture: A Global Vision, Volume III*, 121.
- Solana-Lavalle, G., & Rosas-Romero, R. (2021a). Analysis of voice as an assisting tool for detection of parkinson’s disease and its subsequent clinical interpretation. *Biomedical Signal Processing and Control*, 66, 102415. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1746809421000124> doi: <https://doi.org/10.1016/j.bspc.2021.102415>
- Solana-Lavalle, G., & Rosas-Romero, R. (2021b). Classification of ppmi mri scans with voxel-based morphometry and machine learning to assist in the diagnosis of parkinson’s disease. *Computer Methods and Programs in Biomedicine*, 198, 105793. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0169260720316266> doi: <https://doi.org/10.1016/j.cmpb.2020.105793>
- Stiasny-Kolster, K., Doerr, Y., Möller, J., Höffken, H., Behr, T., Oertel, W., & Mayer, G. (2004). Combination of ‘idiopathic’rem sleep behaviour disorder and olfactory dysfunction as possible indicator for α -synucleinopathy demonstrated by dopamine transporter fp-cit-spect. *Brain*, 128(1), 126–137.
- Sulzer, D., Cassidy, C., Horga, G., Kang, U. J., Fahn, S., Casella, L., ... others (2018). Neuromelanin detection by magnetic resonance imaging (mri) and its promise as a biomarker for parkinson’s disease. *NPJ Parkinson’s disease*, 4(1), 1–13.
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology*, 9, 1112–1122.
- Vital, T. P. R., Nayak, J., Naik, B., & Jayaram, D. (2021). Probabilistic neural network-based model for identification of parkinson’s disease by using voice profile and personal data. *Arabian Journal for Science and Engineering*, 1–25.
- Voormolen, E. H., Wei, C., Chow, E. W., Bassett, A. S., Mikulis, D. J., & Crawley, A. P. (2010). Voxel-based morphometry and automated lobar volumetry: the trade-off between spatial scale and statistical correction. *Neuroimage*, 49(1), 587–596.
- Wang, F., Kaushal, R., & Khullar, D. (2020). *Should health care demand interpretable artificial intelligence or accept “black box” medicine?* American College of Physicians.
- Wibmer, A., Hricak, H., Gondo, T., Matsumoto, K., Veeraraghavan, H., Fehr, D., ... others (2015). Haralick texture analysis of prostate mri: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different gleason scores. *European radi-*

ology, 25(10), 2840–2850.

- Wongvibulsin, S., Wu, K. C., & Zeger, S. L. (2020). Improving clinical translation of machine learning approaches through clinician-tailored visual displays of black box algorithms: Development and validation. *JMIR medical informatics*, 8(6), e15791.
- Yasaka, K., Kamagata, K., Ogawa, T., Hatano, T., Takeshige-Amano, H., Ogaki, K., . . . others (2021). Parkinson's disease: deep learning with a parameter-weighted structural connectome matrix for diagnosis and neural circuit disorder investigation. *Neuroradiology*, 1–12.
- Yujin, Y., Peihua, Z., & Qun, Z. (2010). Research of speaker recognition based on combination of lpcc and mfcc. In *2010 ieee international conference on intelligent computing and intelligent systems* (Vol. 3, pp. 765–767).

Apéndice a

De la tabla 25 a la tabla 61 se muestran los resultados de la clasificación de pacientes con la EP y sujetos control en bases de datos que contienen únicamente una grabación de voz por cada participante. Las bases de datos que se crearon incluyen 3 bases de datos con ambos sexos, 3 bases de datos con participantes únicamente de sexo masculino y 3 bases de datos únicamente con participantes de sexo femenino. Para cada base de datos se encuentra el subconjunto de características óptimo para cada algoritmo de clasificación el cual se resalta en negritas. Con ese subconjunto de características se obtienen las métricas para los demás clasificadores, por lo que se obtienen 4 tablas por cada base de datos.

Tabla 26: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente una grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV.

Base de datos mixta 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.760	0.723	0.937	0.982	0.513
PM	0.489	0.569	0.296	0.660	-0.125
BA	0.521	0.593	0.357	0.678	-0.046
KNN	0.500	0.578	0.321	0.660	-0.094

Tabla 27: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA.

Base de datos mixta 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.641	0.653	0.588	0.875	0.192
PM	0.576	0.597	0.2	0.928	-0.093
BA	0.728	0.738	0.703	0.857	0.412
KNN	0.445	0.539	0.241	0.607	-0.208

Tabla 28: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos mixta 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.554	0.590	0.222	0.875	-0.114
PM	0.728	0.718	0.761	0.910	0.412
BA	0.652	0.676	0.583	0.821	0.233
KNN	0.554	0.605	0.380	0.767	-0.011

Tabla 29: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos mixta 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.586	0.600	0	0.964	-0.119
PM	0.576	0.600	0.285	0.910	-0.062
BA	0.586	0.625	0.45	0.803	0.063
KNN	0.771	0.777	0.758	0.875	0.510

Tabla 30: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos mixta 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.706	0.704	0.714	0.892	0.359
PM	0.586	0.602	0.250	0.946	-0.061
BA	0.641	0.676	0.555	0.785	0.216
KNN	0.543	0.612	0.400	0.678	0.0123

Tabla 31: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos mixta 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.630	0.622	1	1	0.185
PM	0.521	0.588	0.333	0.714	-0.070
BA	0.793	0.776	0.84	0.928	0.561
KNN	0.630	0.648	0.555	0.857	0.165

Tabla 32: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos mixta 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.641	0.629	1	1	0.228
PM	0.75	0.761	0.724	0.857	0.462
BA	0.641	0.666	0.565	0.821	0.205
KNN	0.6957	0.705	0.667	0.857	0.335

Tabla 33: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos mixta 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.608	0.608	0	1	0
PM	0.576	0.607	0.384	0.857	-0.005
BA	0.717	0.788	0.625	0.732	0.419
KNN	0.804	0.8275	0.764	0.857	0.585

Tabla 34: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos mixta 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.728	0.706	0.823	0.946	0.421
PM	0.489	0.567	0.28	0.678	-0.139
BA	0.510	0.590	0.354	0.642	-0.053
KNN	0.586	0.65	0.468	0.696	0.115

Tabla 35: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos mixta 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.510	0.566	0	0.839	-0.264
PM	0.586	0.673	0.475	0.625	0.150
BA	0.782	0.810	0.735	0.839	0.539
KNN	0.673	0.703	0.607	0.803	0.292

Tabla 36: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos mixta 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.608	0.608	0	1	0
PM	0.782	0.764	0.833	0.928	0.538
BA	0.641	0.661	0.571	0.839	0.200
KNN	0.630	0.677	0.533	0.75	0.202

Tabla 37: Resultados de clasificación con una base de datos mixta y sujetos control con 4 algoritmos diferentes. La base de datos cuenta con 92 grabaciones provenientes de 23 pacientes de sexo femenino, 23 pacientes de sexo masculino, 23 sujetos control de sexo femenino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos mixta 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.608	0.613	0.5	0.964	0.047
PM	0.608	0.608	0	1	0
BA	0.586	0.636	0.461	0.75	0.090
KNN	0.760	0.765	0.75	0.875	0.486

Tabla 38: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos de pacientes de sexo femenino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.780	0.725	0.870	0.902	0.578
PM	0.621	0.656	0.6	0.512	0.25
BA	0.548	0.555	0.543	0.487	0.098
KNN	0.548	0.552	0.545	0.512	0.097

Tabla 39: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos de pacientes de sexo femenino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.5	0.5	0.5	0.634	0.578
PM	0.682	0.659	0.714	0.756	0.25
BA	0.512	0.511	0.513	0.560	0.098
KNN	0.621	0.608	0.638	0.682	0.097

Tabla 40: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos de pacientes de sexo femenino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.487	0.48	0.491	0.292	-0.026
PM	0.524	0.526	0.522	0.487	0.048
BA	0.756	0.733	0.783	0.804	0.514
KNN	0.621	0.613	0.631	0.658	0.272

Tabla 41: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos de pacientes de sexo femenino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.487	0.491	0.48	0.682	-0.026
PM	0.634	0.666	0.612	0.536	0.221
BA	0.573	0.575	0.571	0.560	0.514
KNN	0.756	0.733	0.783	0.804	0.146

Tabla 42: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos de pacientes de sexo femenino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.731	0.787	0.693	0.634	0.472
PM	0.670	0.684	0.659	0.634	0.341
BA	0.646	0.636	0.657	0.682	0.293
KNN	0.756	0.862	0.698	0.609	0.535

Tabla 43: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos de pacientes de sexo femenino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.487	0.491	0.48	0.682	-0.026
PM	0.756	0.783	0.733	0.707	0.463
BA	0.609	0.604	0.615	0.634	0.219
KNN	0.756	0.756	0.756	0.756	0.512

Tabla 44: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos de pacientes de sexo femenino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.487	0.480	0.491	0.292	-0.026
PM	0.536	0.533	0.540	0.585	0.048
BA	0.756	0.733	0.783	0.804	0.314
KNN	0.621	0.613	0.631	0.658	0.244

Tabla 45: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos de pacientes de sexo femenino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.475	0.483	0.45	0.731	-0.064
PM	0.707	0.697	0.717	0.731	0.415
BA	0.548	0.55	0.547	0.536	0.097
KNN	0.780	0.828	0.744	0.707	0.567

Tabla 46: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos de pacientes de sexo femenino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.792	0.833	0.760	0.731	0.589
PM	0.512	0.511	0.513	0.560	0.070
BA	0.621	0.631	0.613	0.585	0.244
KNN	0.390	0.4	0.378	0.439	-0.22

Tabla 47: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos de pacientes de sexo femenino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.475	0.483	0.454	0.707	-0.055
PM	0.731	0.771	0.702	0.658	0.391
BA	0.548	0.55	0.547	0.536	0.097
KNN	0.573	0.571	0.575	0.585	0.121

Tabla 48: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos de pacientes de sexo femenino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.512	0.545	0.507	0.146	0.035
PM	0.512	0.511	0.512	0.536	-0.097
BA	0.780	0.767	0.794	0.804	0.561
KNN	0.512	0.512	0.511	0.487	0.024

Tabla 49: Resultados de clasificación con una base de datos de participantes de sexo femenino con 4 algoritmos diferentes. La base de datos cuenta con 82 grabaciones provenientes de 41 pacientes de sexo femenino y 41 sujetos control de sexo femenino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos de pacientes de sexo femenino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.609	0.615	0.604	0.585	0.219
PM	0.634	0.600	0.703	0.804	0.285
BA	0.573	0.583	0.565	0.512	0.147
KNN	0.719	0.704	0.736	0.756	0.44

Tabla 50: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos con sujetos de sexo masculino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.847	0.833	0.863	0.869	0.609
PM	0.717	0.692	0.750	0.782	0.438
BA	0.608	0.608	0.608	0.608	0.217
KNN	0.673	0.642	0.722	0.782	0.356

Tabla 51: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos con sujetos de sexo masculino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.434	0.421	0.444	0.347	-0.174
PM	0.673	0.666	0.681	0.695	0.348
BA	0.652	0.64	0.666	0.695	0.305
KNN	0.760	0.750	0.772	0.782	0.522

Tabla 52: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos con sujetos de sexo masculino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.478	0.478	0.478	0.478	-0.043
PM	0.566	0.565	0.565	0.565	0.13
BA	0.804	0.769	0.850	0.869	0.613
KNN	0.630	0.650	0.615	0.565	0.263

Tabla 53: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la primera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos con sujetos de sexo masculino 1	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.543	0.522	1	1	-0.139
PM	0.543	0.535	0.555	0.652	0.089
BA	0.695	0.695	0.695	0.695	0.391
KNN	0.826	0.857	0.8	0.782	0.654

Tabla 54: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos con sujetos de sexo masculino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.804	0.937	0.733	0.652	0.639
PM	0.543	0.538	0.550	0.608	0.087
BA	0.652	0.684	0.629	0.565	0.309
KNN	0.608	0.586	0.647	0.739	0.225

Tabla 55: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos con sujetos de sexo masculino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.391	0.368	0.407	0.304	-0.22
PM	0.826	0.826	0.826	0.826	0.652
BA	0.695	0.680	0.714	0.739	0.392
KNN	0.695	0.655	0.764	0.826	0.405

Tabla 56: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos con sujetos de sexo masculino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.586	0.558	0.666	0.826	0.198
PM	0.652	0.684	0.629	0.565	0.309
BA	0.869	0.869	0.869	0.869	0.869
KNN	0.608	0.608	0.608	0.608	0.217

Tabla 57: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la segunda grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos con sujetos de sexo masculino 2	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.521	0.538	0.515	0.304	0.044
PM	0.695	0.714	0.680	0.652	0.304
BA	0.565	0.560	0.571	0.608	0.131
KNN	0.847	0.833	0.863	0.869	0.696

Tabla 58: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de una MSV

Base de datos con sujetos de sexo masculino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.826	0.857	0.800	0.782	0.654
PM	0.521	0.523	0.52	0.478	0.07
BA	0.586	0.576	0.6	0.652	0.175
KNN	0.565	0.56	0.571	0.608	0.13

Tabla 59: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un PM

Base de datos con sujetos de sexo masculino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.456	0.464	0.444	0.565	-0.089
PM	0.891	0.875	0.909	0.913	0.783
BA	0.673	0.700	0.653	0.608	0.350
KNN	0.695	0.645	0.800	0.869	0.417

Tabla 60: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un BA

Base de datos con sujetos de sexo masculino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.673	0.611	0.900	0.956	0.356
PM	0.869	0.814	0.947	0.956	0.750
BA	0.847	0.833	0.863	0.869	0.693
KNN	0.717	0.666	0.812	0.869	0.456

Tabla 61: Resultados de clasificación con una base de datos de participantes de sexo masculino con 4 algoritmos diferentes. La base de datos cuenta con 46 grabaciones provenientes de 23 pacientes de sexo masculino y 23 sujetos control de sexo masculino. De cada participante se utilizó únicamente la tercera grabación y se entrenó cada algoritmo con el grupo de características óptimo de un KNN

Base de datos con sujetos de sexo masculino 3	Exactitud	Sensibilidad	Especificidad	Precisión	MCC
MSV	0.521	0.533	0.516	0.347	0.046
PM	0.652	0.629	0.684	0.739	0.309
BA	0.586	0.576	0.600	0.652	0.175
KNN	0.826	0.826	0.826	0.826	0.652

Apéndice b



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe



Original Research Article

Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features



Gabriel Solana-Lavalle, Juan-Carlos Galán-Hernández, Roberto Rosas-Romero*

Universidad de las Américas-Puebla, Mexico

ARTICLE INFO

Article history:

Received 8 October 2019

Received in revised form

3 December 2019

Accepted 1 January 2020

Available online 05 February 2020

Keywords:

Parkinson disease

Dysphonia features

Feature subset selection

Support vector machine

ABSTRACT

Recent research on Parkinson disease (PD) detection has shown that vocal disorders are linked to symptoms in 90% of the PD patients at early stages. Thus, there is an interest in applying vocal features to the computer-assisted diagnosis and remote monitoring of patients with PD at early stages. The contribution of this research is an increase of accuracy and a reduction of the number of selected vocal features in PD detection while using the newest and largest public dataset available. Whereas the number of features in this public dataset is 754, the number of selected features for classification ranges from 8 to 20 after using *Wrappers feature subset selection*. Four classifiers (*k nearest neighbor*, *multi-layer perceptron*, *support vector machine* and *random forest*) are applied to vocal-based PD detection. The proposed approach shows an accuracy of 94.7%, sensitivity of 98.4%, specificity of 92.68% and precision of 97.22%. The best resulting accuracy is obtained by using a *support vector machine* and it is higher than the one, which was reported on the first work to use the same dataset. In addition, the corresponding computational complexity is further reduced by selecting no more than 20 features.

© 2020 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

1. Introduction

Parkinson disease (PD) is a chronic and degenerative neurological disease that affects 1% of people over age 60 and it is the most common neurological disorder after Alzheimer. PD is the result of losing cells in different regions of the brain, mainly the *substantia nigra*. The *substantia nigra* produces *dopamine*, a carrier of the signals for movement coordination. The loss of

dopamine causes neurons to fire without control so that these patients lose the ability to control limb movement [1].

The cardinal symptoms of PD patients are *motor symptoms* such as a resting tremor where limbs move without control (*diskinesya*), slowness of movement (*bradykinesia*), postural instability (balance problems) and rigidity [1]. Since motor symptoms are revealed at an advanced disease stage, PD is difficult to diagnose at an early stage. Although PD is not cured, the quality of life in PD patients is improved when they are

* Corresponding author at: Universidad de las Américas-Puebla, Mexico.

E-mail address: roberto.rosas@udlap.mx (R. Rosas-Romero).

<https://doi.org/10.1016/j.bbe.2020.01.003>

0208-5216/© 2020 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

Figura 27: Portada del artículo “Analysis of voice as an assisting tool for detection of Parkinson’s disease and its subsequent clinical interpretation”. (Gabriel Solana-Lavalle, 2020). CC-BY-NC



Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation

Gabriel Solana-Lavalle, Roberto Rosas-Romero *

Department of Electrical and Computer Engineering, Universidad de las Américas Puebla, Ex Hacienda Sta. Catarina Mártir S/N, San Andrés Cholula, Puebla, C.P. 72810, Mexico

ARTICLE INFO

Keywords:

Parkinson's disease detection
Voice analysis
Feature subset selection
Assisting tool for diagnosis

ABSTRACT

In this work, a voice-based analysis is conducted with the contribution of providing physicians with a decision tool along with framing information to help them see functional differences and understand why the detection method suspects PD. The voice-based detection method consists in applying *feature subset selection* and four different classifiers to voice recordings from five datasets (gender-based, balanced and unbalanced) derived from the largest public dataset for voice-based PD detection. One of the contributions is an improvement over previous works on voice-based PD detection over the same dataset, in terms of performance and complexity. The detection performance is characterized by 95.9% of accuracy, 98.35% of sensitivity, 91.06% of specificity, and 95.6% of precision in women; and 94.36% of accuracy, 100% of sensitivity, 97.1% of specificity, and 96.83% of precision in men. The number of features, fed to classifiers, ranges from 6 to 20. This work shows that different factors are associated with PD detection according to gender: high-frequency voice content is the most significant functional information to assist PD detection in women, while low-frequency content assists PD detection in men better. It is shown that a comparison of the variability of the most important features between patients with PD and controls can be used as contextual information by a physician to have a better interpretation of the classification.

1. Introduction

Parkinson's disease (PD) is the second most common chronic neurodegenerative disease after Alzheimer's disease. It affects 1% of the population over age 60, and it is estimated that the number of persons with Parkinson will double from 2005 to 2030 [1,2]. The diagnosis of PD is one of the most challenging tasks in neurology since PD is mostly diagnosed at an advanced stage [3].

According to Braak et al. [4], there are six neuropathologic stages during the evolution of the disease. Olfactory and vocal disorders appear at the first two stages. During the third and fourth stages, motor symptoms become more apparent, which facilitates the diagnosis of PD by an expert. Large areas of the brain are severely affected during the fifth and sixth stages. Some of the vocal impairments, in PD patients, are defective use of voice (dysphonia) [5], reduced volume (hypophonia), reduced pitch range (monotone) and difficulty with the articulation of sounds (dysarthria) [6].

Although PD is not cured, the quality of life of PD patients might be improved if prompt treatment is received as a consequence of an early diagnosis [7]. Vocal disorders are prodromal symptoms that manifest in

90% of the PD patients at early stages [6]. Thus, analysis of vocal features has been used for PD detection at early stages [3,7–11]. Furthermore, the vocal analysis method is inexpensive and easy to do early on.

The advantages of applying pattern recognition on voice recordings, as an assisting tool for PD diagnosis, are: (1) vocal disorders start in the first stages of the disease [12], (2) voice is easily recorded, (3) there is a wide diversity of signal processing techniques for extraction of voice features [13,14], (4) the classification of voice features with machine learning algorithms provides assistance during the early diagnosis of the disease [14,13], and (5) PD detection at early stages and the corresponding health cares might improve the quality of life of the patients [15].

According to the reported works on voice-based PD detection, the most commonly used group of vocal features has been the *baseline features* that include *jitter*, *shimmer*, *fundamental frequency parameters*, *harmonicity parameters*, *recurrence period density entropy* (RPDE), *detrended fluctuation analysis* (DFA) and *pitch period entropy* (PPE) [13]. Another comprehensive analysis of features, for PD detection, is based on 132 dysphonia features from sustained vowels [14]. According to the reported works on voice-based PD detection, the best results were

* Corresponding author.

E-mail addresses: gabriel.solanale@udlap.mx (G. Solana-Lavalle), roberto.rosas@udlap.mx (R. Rosas-Romero).

<https://doi.org/10.1016/j.bspc.2021.102415>

Received 9 October 2020; Received in revised form 26 November 2020; Accepted 10 January 2021

1746-8094/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Figura 28: Portada del artículo “Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation”. (Solana-Lavalle & Rosas-Romero, 2021a). CC-BY-NC



Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease



Gabriel Solana-Lavalle, Roberto Rosas-Romero*

Department of Electrical and Computer Engineering, Universidad de las Américas-Puebla Santa Catarina Mártir, San Andrés Cholula, Puebla, 72810, México

ARTICLE INFO

Article history:

Received 21 August 2020
Accepted 6 October 2020

Keywords:

Parkinson's disease
Magnetic resonance imaging
Voxel-based morphometry
Machine learning
Feature selection

ABSTRACT

Background and objectives: Qualitative and quantitative analyses of *Magnetic Resonance Imaging* (MRI) scans are carried out to study and understand *Parkinson's Disease*, the second most common neurodegenerative disorder in people at their 60's. Some quantitative analyses are based on the application of *voxel-based morphometry* (VBM) on *magnetic resonance* images to determine the *regions of interest*, within gray matter, where there is a loss of the nerve cells that generate dopamine. This loss of dopamine is indicative of Parkinson's disease. The purpose of this research is the introduction of a new method to classify the 3-D magnetic resonance scans of an individual, as an assisting tool for diagnosis of Parkinson's disease by using the largest MRI dataset (*Parkinson's Progression Markers Initiative*) from a population of patients with Parkinson's disease and control individuals. A contribution is that separate studies are conducted for men and women since gender plays a significant role within Neurobiology, which is demonstrated by the fact that men are more prone to Parkinson's disease than women are.

Methods: Previous to classification, VBM is conducted on magnetic resonance images to detect the regions where features are extracted by using first- and second-order statistics methods. Furthermore, the number of features is considerably reduced by using *feature selection techniques*. Seven classifiers are used and we are conducting separate experiments for men and women.

Results: The best detection performance achieved in men is 99.01% of accuracy, 99.35% of sensitivity, 100% of specificity, and 100% of precision. The best detection performance achieved in women is 96.97% of accuracy, 100% of sensitivity, 96.15% of specificity, and 97.22% of precision. During classification of magnetic resonance images, the corresponding computational complexity is reduced since few features are selected.

Conclusions: The proposed method provides high performance as an assisting tool in the diagnosis of Parkinson's disease, by conducting separate experiments in men and women. While previous works have focused their analysis to the striatum region of the brain (the largest nuclear complex of the basal ganglia), the proposed approach is based on analysis over the whole brain by looking for decreases of tissue thickness, with the consequence of finding other regions of interest such as the cortex.

© 2020 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder for people at their 60's. The characteristic symptoms are tremor at rest, bradykinesia, and rigidity. The cause of PD is the loss of dopaminergic neurons in the substantia nigra of the brain [1].

Magnetic Resonance Imaging (MRI) of the brain has been used to assist the diagnosis of *Parkinson's disease* (PD) with an increase in the detection accuracy over the last years [2,3]. The brain structure, over different types of tissue, has been visually and qualitatively interpreted by physicians on MR images; however, an automated quantitative analysis is also recommended since it improves the interpretation of the changes in brain structure [2–6].

To conduct this research, we used MRI images from the *Parkinson's Progression Markers Initiative* (PPMI), sponsored by the *Michael J. Fox Foundation*. One result, of this initiative, is one of the most

* Corresponding author.

E-mail address: roberto.rosas@udlap.mx (R. Rosas-Romero).

<https://doi.org/10.1016/j.cmpb.2020.105793>

0169-2607/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Figura 29: Portada del artículo “Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease”. (Solana-Lavalle & Rosas-Romero, 2021b). CC-BY-NC

Chapter

Analysis of Voice and Magnetic Resonance Images to Assist Diagnosis of Parkinson's Disease with Machine Learning

Gabriel Solana-Lavalle and Roberto Rosas-Romero

Abstract

Parkinson's disease (PD) is a chronic neurodegenerative disease that affects 1% of the population and whose diagnosis is considered one of the most challenges in the area of neurology. The goal of our work is to assist physicians with the correct diagnosis and early detection of PD. This chapter provides a review of previous work on PD detection under two perspectives, voice analysis and Magnetic Resonance Imaging analysis, by comparing our work with those from other authors. For the case of voice-based PD detection, accuracy reaches 95.9% in female patients and 94.36% in male patients on the largest available dataset. Another contribution in this area is the analysis of voice features to assist the clinical interpretation of the binary result of voice-based detection. For the case of structural Magnetic Resonance Imaging (sMRI)-based PD detection, detection accuracy reaches 96.97% for female patients and 99.01% for male patients using the Parkinson's Progression Marker Initiative dataset. We provide a discussion about the finding of new regions of interest to assist in the detection of PD on sMRI. There is also a comparison between voice-based and MRI-based PD detection methods. Finally, a perspective on future work for PD detection is discussed.

Keywords: Parkinson's disease, machine learning, biomedical engineering, magnetic resonance imaging, voice analysis, diagnostic tool

1. Introduction

Parkinson's disease (PD) is a chronic neurodegenerative disease that affects over six million people worldwide. Because PD is most common in people over the age of 50, the number of PD patients is expected to double by 2040 due to the increase in life expectancy [1]. The loss of dopaminergic cells in the substantia nigra region of the brain reduces the amount of dopamine in PD patients, causing dyscontrol in several areas of the brain. Some of the main symptoms of PD are motor symptoms, such as tremors, rigidity, and slow movement (bradykinesia). These symptoms, however, become apparent at an intermediate-advanced stage of the disease when the patient may have had the disease for over ten years [2].

The diagnosis of PD is considered one of the most challenging in the area of neurology. The autopsy of the brain of PD patients has shown that 35% of the