

UAX

TRABAJO FIN DE MÁSTER

**MUIA
24-25**

UNIVERSIDAD ALFONSO X EL SABIO

Business Tech

Máster Universitario en Inteligencia Artificial



TRABAJO DE FIN DE MÁSTER

**Desarrollo de un modelo de IA para la detección
temprana del Parkinson**

**Unai Roa Cepeda
Valeria Rolle Sónora**

06/2025

RESUMEN

Este proyecto se ha centrado en el desarrollo de un sistema basado en inteligencia artificial para la detección temprana de la enfermedad de Parkinson mediante el análisis de señales de voz. La iniciativa busca aportar una herramienta no invasiva, accesible y eficiente que facilite el diagnóstico precoz de esta enfermedad neurodegenerativa, contribuyendo así a mejorar la calidad de vida de los pacientes y apoyar a los profesionales sanitarios en la toma de decisiones.

Para ello, se ha realizado un estudio exhaustivo del estado del arte, recopilando conocimientos sobre la enfermedad, sus biomarcadores y los métodos tradicionales de diagnóstico. Se han explorado las aplicaciones de la inteligencia artificial en el ámbito médico, con especial atención a modelos aplicados a enfermedades neurodegenerativas. A partir de esta base, se seleccionaron y analizaron conjuntos de datos de voz, realizando un preprocesamiento y un análisis exploratorio para extraer características relevantes.

El desarrollo incluyó el diseño, entrenamiento y evaluación de distintos modelos de machine learning y deep learning, priorizando el equilibrio entre precisión, sensibilidad y robustez. Entre los modelos evaluados, los enfoques tradicionales, como el SVM, demostraron ser especialmente eficaces en este caso, si bien se consideraron técnicas avanzadas para futuras mejoras.

A pesar de las limitaciones en cuanto al tamaño y calidad de los datos disponibles, los resultados obtenidos fueron satisfactorios y comparables con estudios similares, lo que confirma el potencial de esta aproximación. Este proyecto no solo aporta resultados técnicos, sino también aprendizajes importantes sobre la metodología y los retos de investigación en este campo.

En definitiva, el trabajo representa un paso importante hacia el desarrollo de herramientas inteligentes para el diagnóstico precoz de la enfermedad de Parkinson, con amplias posibilidades de mejora y aplicación futura.

Palabras clave

Enfermedad del Parkinson, inteligencia artificial, análisis de voz, aprendizaje automático, detección temprana

ABSTRACT

This project focused on developing an artificial intelligence-based system for the early detection of Parkinson's disease through voice signal analysis. The initiative aims to provide a non-invasive, accessible, and efficient tool to facilitate early diagnosis of this neurodegenerative disease, thereby improving patients' quality of life and supporting healthcare professionals in decision-making.

To achieve this, an extensive review of the state of the art was conducted, gathering knowledge about the disease, its biomarkers, and traditional diagnostic methods. Applications of artificial intelligence in the medical field were explored, with special attention to models applied to neurodegenerative diseases. Based on this foundation, voice datasets were selected and analyzed, performing preprocessing and exploratory analysis to extract relevant features.

The development included designing, training, and evaluating various machine learning and deep learning models, prioritizing a balance between accuracy, sensitivity, and robustness. Among the evaluated models, traditional approaches such as SVM proved particularly effective in this case, although advanced techniques were considered for future improvements.

Despite limitations in the size and quality of available data, the obtained results were satisfactory and comparable to similar studies, confirming the potential of this approach. This project not only delivers technical outcomes but also important insights into methodology and research challenges in this field.

In summary, this work represents a significant step toward developing intelligent tools for the early diagnosis of Parkinson's disease, with broad potential for improvement and future application.

Keywords

Parkinson's disease, artificial intelligence, voice analysis, machine learning, early detection

ÍNDICE

Capítulo 1: Introducción al Trabajo Fin de Máster	9
1. Contexto y justificación del proyecto	9
2. La Inteligencia Artificial en la medicina	10
Capítulo 2: Objetivos del TFM	12
1. Objetivos principales	12
2. Objetivos secundarios	13
Capítulo 3: Estado del arte	14
1. Introducción	14
2. Estudios previos sobre la detección temprana de la EP	15
2.1. Biomarcadores y señales tempranas	15
2.2. Modelos, algoritmos y datos empleados en investigaciones con IA	16
2.3. Proyecto destacado: Análisis del habla asistido por aprendizaje automático para la detección temprana de la EP	18
3. Conclusiones del estado del arte	19
Capítulo 4: Metodología	20
1. Enfoque de la investigación	20
2. Datos empleados	21
2.1. Justificación del uso de datos de voz	21
2.2. Descripción del conjunto de datos NeuroVoz_v3	22
2.3. Limitaciones y consideraciones	23
3. Modelado y técnicas a utilizar	24
3.1. Análisis exploratorio de datos (EDA) y selección de características	24
3.2. Modelos predictivos	24
3.3. Estrategias de validación	25
3.4. Métricas de evaluación	25
4. Herramientas y entornos de desarrollo	26
4.1. Bibliotecas y entornos utilizados	26
4.2. Entornos de desarrollo	27
4.3. Hardware	27

Capítulo 5: Desarrollo del proyecto..... 28

1.	Exploración inicial y modelado con datos originales.....	28
1.1.	Análisis de los metadatos.....	28
1.1.1.	Carga de datos y exploración inicial	29
1.1.2.	Limpieza básica de datos	29
1.1.3.	Análisis exploratorio de datos (EDA).....	30
1.2.	Análisis y preprocesamiento de las características acústicas proporcionadas ..	32
1.2.1.	Carga de datos y exploración inicial	33
1.2.2.	Limpieza básica de datos	33
1.2.3.	Análisis exploratorio de datos (EDA).....	34
1.2.4.	Preprocesamiento de datos post-EDA.....	36
1.3.	División del conjunto de datos	36
1.4.	Modelado y evaluación con las características proporcionadas	37
1.4.1.	Modelo de Regresión Logística	37
1.4.2.	Modelo SVM	38
1.4.3.	Modelo Random Forest	39
1.4.4.	Conclusiones preliminares y limitaciones detectadas	39
2.	Procesamiento y extracción propia de características de los audios.....	40
2.1.	Procesamiento y limpieza de audios.....	40
2.2.	Extracción de características acústicas y espectrales	41
2.3.	Análisis y preprocesamiento de las nuevas características	42
2.3.1.	Carga de datos y exploración inicial	43
2.3.2.	Limpieza básica de datos	43
2.3.3.	Análisis exploratorio de datos (EDA).....	43
2.3.4.	Preprocesamiento de datos post-EDA.....	45
2.3.5.	División del conjunto de datos	45
2.4.	Entrenamiento y evaluación de modelos con las nuevas características.....	46
2.4.1.	Modelo de Deep Learning	46
2.4.2.	Modelo SVM	46
2.4.3.	Modelo Random Forest	47
2.4.4.	Modelo Gradient Boosting	47
3.	Modelado avanzado con Deep Learning	48
3.1.	Arquitectura y diseño del modelo de red neuronal	48
3.2.	Optimización de hiperparámetros con Keras Tuner	48
3.3.	Resultados preliminares y estado actual	49
4.	Comparativa y análisis conjunto de resultados	49
4.1.	Ajuste del umbral de decisión en el modelo SVM.....	51
4.2.	Comparación con estudios previos	51

Capítulo 6: Conclusiones 53

1. Cumplimiento de objetivos	53
2. Limitaciones y desafíos	54
3. Lecciones aprendidas	54
4. Consideraciones finales y reflexiones personales.....	55
 Capítulo 7: Trabajo a futuro	57
 Bibliografía.....	58

Capítulo 1: Introducción al Trabajo Fin de Máster

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo progresivo que afecta el sistema nervioso central, deteriorando de forma gradual la calidad de vida de quienes la padecen. Es la segunda enfermedad neurodegenerativa más frecuente a nivel mundial, después del Alzheimer [1]. Afecta aproximadamente al 1 % de la población mayor de 60 años, alcanzando hasta un 3,5 % entre los 85 y 89 años [1]. Se caracteriza por la presencia de síntomas motores como temblores en reposo, rigidez muscular, bradicinesia (movimientos lentos) y alteraciones posturales, así como síntomas no motores que incluyen trastornos del sueño, deterioro cognitivo y depresión [2]. A pesar de los avances en el tratamiento, actualmente la EP no tiene cura, y las intervenciones médicas se centran en mejorar muchos de los síntomas del movimiento [3]. A medida que la EP avanza, la pérdida progresiva de autonomía repercute significativamente tanto en el bienestar de los pacientes como en el de sus cuidadores y familiares.

1. Contexto y justificación del proyecto

La detección temprana de la EP es un desafío crucial en el ámbito de la neurología. En la actualidad, el diagnóstico se basa en la observación clínica de los síntomas motores, lo que implica que la enfermedad ya ha progresado significativamente en el momento de su detección. Sin embargo, estudios recientes han revelado la existencia de biomarcadores y señales tempranas que podrían facilitar una identificación más precoz de la enfermedad, incluso antes de la aparición de síntomas evidentes [4]. Identificar la EP en sus etapas iniciales permitiría la aplicación de estrategias terapéuticas que podrían retrasar su avance y mejorar la calidad de vida de los pacientes.

En este contexto, la inteligencia artificial (IA) surge como una herramienta poderosa para abordar este problema. La IA ha demostrado ser eficaz en el análisis de grandes volúmenes de datos médicos, facilitando la identificación de patrones complejos que podrían pasar desapercibidos en los métodos tradicionales. La aplicación de algoritmos de aprendizaje automático y aprendizaje profundo en el diagnóstico de enfermedades neurodegenerativas permite analizar datos de diversas fuentes, como imágenes médicas, señales de voz, escritura manual o registros de movimiento, con el objetivo de detectar anomalías sutiles indicativas de la presencia temprana de la EP[5].

2. La Inteligencia Artificial en la medicina

El uso de IA en la medicina ha revolucionado múltiples áreas, mejorando la precisión diagnóstica y optimizando los procesos de toma de decisiones clínicas. Desde la identificación de patrones en imágenes de resonancia magnética hasta el análisis de biomarcadores en fluidos biológicos, los modelos de IA han permitido avanzar en la detección y predicción de diversas patologías [6]. En el caso de la EP, su integración en herramientas de diagnóstico podría reducir el tiempo necesario para identificar la enfermedad y aumentar la eficacia de las intervenciones médicas [5].

A pesar de los avances prometedores, la implementación de la IA en el ámbito médico presenta varios desafíos. La disponibilidad y calidad de los datos utilizados para entrenar los modelos es un aspecto fundamental, ya que la falta de datos representativos o sesgados podría afectar la fiabilidad de los resultados [7]. Asimismo, la interpretabilidad de los modelos sigue siendo un reto importante, ya que en el ámbito clínico es crucial que los profesionales de la salud puedan comprender y confiar en las decisiones tomadas por los algoritmos. Además, la integración de estas herramientas en los flujos de trabajo clínicos existentes requiere la colaboración entre expertos en IA y médicos, garantizando que los modelos sean aplicables y útiles en la práctica diaria.

De este modo, la detección precoz de la EP mediante IA representa una oportunidad significativa para mejorar el diagnóstico y tratamiento de la enfermedad. Aunque existen desafíos por superar, los avances en esta área podrían transformar la manera en que se diagnostican y manejan las enfermedades neurodegenerativas, proporcionando herramientas más precisas y accesibles para los profesionales de la salud y los pacientes.

Capítulo 2: Objetivos del TFM

Este proyecto tiene como finalidad desarrollar un modelo de inteligencia artificial capaz de detectar de forma precoz la EP. Para ello, se han establecido una serie de objetivos que guiarán el trabajo desde la fase de investigación inicial hasta la evaluación del modelo. Estos objetivos se dividen en principales y secundarios, en función de su relevancia para el desarrollo del estudio.

1. Objetivos principales

1. Analizar el estado actual del conocimiento sobre la EP, incluyendo sus características clínicas, posibles biomarcadores tempranos y los métodos de diagnóstico tradicionales.
2. Estudiar las aplicaciones actuales de la IA en el ámbito médico, con especial énfasis en los modelos desarrollados para el diagnóstico de enfermedades neurodegenerativas, y evaluar su aplicabilidad en el caso de la EP.
3. Investigar qué tipos de datos podrían ser útiles para la detección precoz de la EP, incluyendo señales de voz, escritura manual, patrones de movimiento o imágenes cerebrales, y seleccionar los más apropiados para el desarrollo del modelo.
4. Localizar y analizar conjuntos de datos existentes, asegurando su calidad y relevancia, y preparar los datos mediante técnicas de preprocesamiento y análisis exploratorio.
5. Diseñar, entrenar y evaluar modelos de IA adecuados, utilizando técnicas como aprendizaje supervisado o profundo, y seleccionar la

arquitectura que ofrezca un mejor equilibrio entre precisión, sensibilidad y generalización.

6. Evaluar el rendimiento del modelo mediante métricas estándar, como precisión, sensibilidad, especificidad, AUC-ROC o F1-score, y comparar los resultados con estudios similares o relacionados.

2. Objetivos secundarios

1. Investigar la posibilidad de desarrollar enfoques más avanzados para mejorar la precisión y robustez del modelo, utilizando por ejemplo datos multimodales, técnicas de aprendizaje profundo u otras metodologías de IA.

Capítulo 3: Estado del arte

La investigación sobre la EP y su detección temprana ha cobrado una creciente relevancia en los últimos años, especialmente con el desarrollo acelerado de las tecnologías digitales y IA. Este capítulo tiene como finalidad realizar un análisis detallado de los estudios más representativos y actuales en esta área, identificando los avances logrados, las técnicas empleadas y las limitaciones presentes. Con ello, se busca establecer una base sólida que justifique el planteamiento del proyecto, así como los aspectos técnicos que se considerarán en los siguientes capítulos.

1. Introducción

La capacidad para identificar las primeras señales de la EP antes de la aparición de síntomas evidentes, como los motores, es clave para permitir intervenciones más efectivas que retrasen la progresión de la enfermedad. Actualmente, el diagnóstico suele realizarse cuando ya se ha perdido entre un 60 % y 80 % de las neuronas dopaminérgicas (células que producen dopamina, un neurotransmisor clave para el control del movimiento), indicando que la enfermedad lleva años avanzando antes de ser detectada [1]. En este contexto, la IA cobra protagonismo permitiendo el análisis eficiente de grandes volúmenes de datos biomédicos y la detección de patrones complejos en estas fases iniciales.

Para conocer el estado actual de este campo, se ha realizado una búsqueda exhaustiva en bases de datos académicas reconocidas como PubMed¹, IEEE Xplore² y Google Scholar³, centrando el análisis en estudios publicados durante los últimos cinco años. Este enfoque permitió identificar las tendencias más recientes, los avances tecnológicos más relevantes, y las principales contribuciones teóricas y prácticas en la detección precoz de la EP.

¹ <https://pubmed.ncbi.nlm.nih.gov/>

² <https://ieeexplore.ieee.org/Xplore/home.jsp>

³ <https://scholar.google.es/>

Los modelos de IA más frecuentes incluyen algoritmos clásicos como Support Vector Machines (SVM) y Random Forest (RF), junto con arquitecturas más avanzadas como redes neuronales profundas. La elección del modelo depende en gran medida del tipo de datos disponibles, pudiendo abarcar desde señales de voz o escritura manual hasta registros de movimiento, imágenes médicas o biomarcadores clínicos. A pesar del progreso logrado, existen desafíos importantes relacionados con la calidad, la representatividad y la heterogeneidad de los datos, así como con la robustez y generalización de los modelos desarrollados.

2. Estudios previos sobre la detección temprana de la EP

Se han revisado investigaciones centradas, por un lado, en la identificación de datos clínicos y biomarcadores con potencial para detectar la EP en sus fases iniciales, y por otro, en el desarrollo y evaluación de modelos de IA aplicados a dicha detección. Estos estudios permiten comprender tanto las bases médicas como las estrategias tecnológicas más prometedoras para el diagnóstico precoz.

2.1. Biomarcadores y señales tempranas

Diversos estudios han identificado señales que pueden anticipar el desarrollo de la enfermedad años antes del diagnóstico clínico, incluyendo síntomas no motores, biomarcadores bioquímicos y técnicas de imagen [8,9].

Entre las señales clínicas más tempranas destacan alteraciones como la pérdida del olfato (presente en aproximadamente el 75 % de los pacientes), el estreñimiento crónico y los trastornos del sueño en fase de movimiento ocular rápido (REM). Éstas pueden presentarse hasta 10 o 20 años antes de los primeros síntomas motores y están asociadas a la acumulación anómala de α -sinucleína (proteína implicada en la formación de cuerpos celulares irregulares) en diferentes regiones del sistema nervioso.

Desde el punto de vista bioquímico, se han identificado diversos biomarcadores en el líquido cefalorraquídeo (fluido que rodea y protege el cerebro y la médula espinal), así como en la sangre o saliva. Algunos de los más estudiados son la propia α -sinucleína, la proteína DJ-1 (relacionada con la protección de las células) y el ácido úrico (un antioxidante natural), cuyos niveles alterados podrían reflejar daños neurodegenerativos en curso. También se han analizado microARNs, pequeñas moléculas presentes en la sangre que podrían facilitar una detección temprana mediante procedimientos más accesibles, como análisis sanguíneos.

En cuanto a las técnicas de imagen, herramientas como la ecografía transcraneal o la resonancia magnética funcional permiten detectar cambios estructurales o funcionales antes del diagnóstico clínico. Una de las más utilizadas es el escáner DAT-SPECT (Dopamine Transporter Single Photon Emission Computed Tomography), que evalúa el estado de las neuronas que producen dopamina, y ha demostrado una alta precisión para identificar la enfermedad en fases iniciales.

Finalmente, también se han identificado ciertos factores genéticos asociados a un mayor riesgo de desarrollar la EP, como mutaciones en los genes *leucine-rich repeat kinase 2* (LRRK2), *glucocerebrosidase* (GBA) y *alpha-synuclein* (SNCA). Aunque su presencia no garantiza el desarrollo de la enfermedad, sí pueden contribuir a su aparición y se están explorando como parte de modelos predictivos más amplios.

2.2. Modelos, algoritmos y datos empleados en investigaciones con IA

Según un estudio reciente [10], el desarrollo de modelos de IA en el ámbito de la medicina se enfrenta a importantes limitaciones, especialmente relacionadas con la calidad y disponibilidad de los datos. Los biomarcadores más específicos, como las proteínas mencionadas para la detección precoz de la EP, requieren procedimientos más complejos y costosos, lo que dificulta la

disponibilidad de grandes bases de datos necesarias para entrenar modelos efectivos capaces de detectar la enfermedad en fases muy tempranas.

Por otro lado, existen alternativas más accesibles, como el análisis de la voz, que puede reflejar cambios sutiles en etapas iniciales de la EP. Del mismo modo ocurre con los trastornos del sueño REM y la pérdida del olfato. Aunque no son las primeras señales en aparecer, suelen manifestarse antes que los síntomas motores.

Para analizar estos datos, se emplean desde algoritmos tradicionales como regresión logística, SVM y árboles de decisión, hasta métodos de aprendizaje profundo que pueden capturar patrones complejos en datos heterogéneos. Las técnicas de balanceo de clases, reducción de dimensionalidad y selección de características son fundamentales para mitigar problemas como el desequilibrio y la limitada cantidad de datos disponibles.

Sin embargo, la gran variabilidad de los datos, el tamaño reducido de las muestras y la dificultad para interpretar modelos complejos siguen siendo desafíos importantes. Para ello, se han desarrollado técnicas de Explainable Artificial Intelligence (XAI), que aportan transparencia y ayudan a entender mejor las decisiones de los modelos y aumentan la confianza de los profesionales de la salud.

A pesar de estas limitaciones, la integración de datos multimodales y el uso de modelos avanzados han mostrado resultados prometedores, alcanzando precisiones de hasta el 96,45 % en la detección temprana de la EP.

2.3. Proyecto destacado: Análisis del habla asistido por aprendizaje automático para la detección temprana de la EP

Un estudio reciente [11] investigó la detección temprana de la EP mediante análisis de voz asistido por aprendizaje automático, utilizando grabaciones capturadas con teléfonos móviles. El conjunto de datos incluía grabaciones de pacientes y controles sanos realizando tareas de lectura y conversación. Para procesar las conversaciones, se aplicó un algoritmo de diarización (técnica que permite separar e identificar quién habla en una grabación de audio) que separó con alta precisión las distintas voces, permitiendo un análisis más preciso en contextos reales.

Se extrajeron características acústicas complementarias, como los coeficientes Mel Frequency Cepstral Coefficients (MFCC) y Gammatone Frequency Cepstral Coefficients (GTCC), que alimentaron modelos de clasificación basados en SVM, KNN y redes neuronales. Estos modelos alcanzaron una precisión de hasta 92.3%, destacando la combinación de ambos indicadores como la más efectiva para detectar pacientes con EP. Además, se comprobó estadísticamente que las características vocales presentan diferencias significativas entre ambos grupos.

Este enfoque innovador destaca por el uso de dispositivos móviles en condiciones reales y la incorporación de técnicas de diarización, lo que facilita su aplicación en la práctica real. Aunque el estudio está limitado por el tamaño de la muestra, abre las puertas al desarrollo de aplicaciones portátiles que puedan ayudar a la detección y seguimiento de la EP en entornos cotidianos.

3. Conclusiones del estado del arte

La revisión realizada muestra como la detección temprana de la EP es un área en constante crecimiento, impulsada por los avances en tecnologías digitales, especialmente la IA. Se han identificado diversas señales clínicas, bioquímicas, genéticas y de imagen que pueden anticipar la enfermedad, muchas de ellas presentes varios años antes de que aparezcan los síntomas motores tradicionales.

En cuanto a la aplicación tecnológica, la IA ha demostrado ser una herramienta muy útil para analizar estos datos complejos y variados, empleando desde modelos clásicos como SVM y regresión logística, hasta redes neuronales profundas y métodos que permiten interpretar mejor sus resultados. A pesar de estos avances, todavía existen retos importantes relacionados con la disponibilidad y calidad de datos disponibles, así como con la capacidad de los modelos para generalizar y ser interpretables.

Es especialmente relevante el uso de las señales accesibles, como el análisis de la voz, que combinado con técnicas avanzadas como la diarización y el aprendizaje automático, abre nuevas posibilidades para la detección precoz en contextos que facilitan su implementación. No obstante, sigue siendo necesario avanzar en la validación clínica y garantizar la escalabilidad de estas soluciones para facilitar su uso práctico.

Estos avances sientan una base sólida para futuras investigaciones orientadas a la mejora de la detección precoz de la EP, contribuyendo así a un diagnóstico más temprano y tratamientos más efectivos.

Capítulo 4: Metodología

Para desarrollar una solución basada en IA capaz de detectar de forma precoz la EP, es fundamental seguir una metodología bien definida que guíe todo el proceso, desde la obtención de los datos hasta la evaluación del modelo. Este capítulo describe el enfoque empleado en el proyecto, detallando los pasos clave para elaborar una propuesta sólida y fundamentada principios técnicos y científicos.

Se abordarán aspectos como el enfoque general de la investigación, las características de los datos empleados y las técnicas de modelado, así como las métricas de evaluación consideradas. Además, se describirán las herramientas y entornos utilizados para el desarrollo, con el fin de ofrecer una visión completa y estructurada del proceso llevado a cabo.

1. Enfoque de la investigación

Este estudio adopta con un enfoque exploratorio y experimental, orientado al desarrollo de un modelo predictivo capaz de identificar de forma temprana la EP a partir de datos biomédicos estructurados. Por un lado, esto permite examinar en profundidad las características y calidad de los datos disponibles, identificar patrones relevantes y comprender la naturaleza del problema desde una perspectiva técnica. Al mismo tiempo, facilita la aplicación y evaluación de diversas técnicas de aprendizaje automático y profundo.

Este estudio se plantea con un enfoque exploratorio y experimental. El carácter exploratorio responde al interés por investigar nuevas formas de abordar el diagnóstico precoz de la EP mediante tecnologías avanzadas como la Inteligencia Artificial (IA). El trabajo busca indagar cómo estas tecnologías

pueden mejorar los métodos diagnósticos actuales, explorando nuevas fuentes de datos, como las señales de voz y las imágenes cerebrales.

La dimensión experimental del estudio se reflejará en la futura implementación y análisis de distintos modelos de IA para evaluar su rendimiento en tareas de detección temprana, utilizando datos reales obtenidos de diversas fuentes. En particular, se prestará especial atención al análisis de los datos de voz, que, como se discutió previamente, son un indicador prometedor de las alteraciones neurológicas precoces. Además, la integración de imágenes cerebrales para detectar patrones sutiles en el cerebro de los pacientes permitirá mejorar la precisión de los diagnósticos.

El enfoque de investigación es cuantitativo, orientado a la clasificación de pacientes mediante la identificación de patrones complejos presentes en los datos. Este enfoque permitirá evaluar objetivamente el desempeño de los modelos en relación con los diagnósticos clínicos disponibles.

2. Datos empleados

Para el desarrollo de este proyecto se ha seleccionado un conjunto de datos biomédicos enfocado en el análisis de la voz de pacientes con EP y sanos. Comprender bien las características de los datos es fundamental, ya que influye directamente en cómo se plantean las metodologías y en la fiabilidad de los resultados.

2.1. Justificación del uso de datos de voz

Aunque existen biomarcadores más tempranos y específicos para la detección de la EP, como los relacionados con alteraciones moleculares, el análisis de la voz ha ganado relevancia en los últimos años como una alternativa viable para el diagnóstico precoz. Su principal ventaja consiste en permitir acceder a señales potencialmente indicativas de la enfermedad de forma no invasiva

y económica. Diversos estudios han demostrado que las alteraciones vocales pueden manifestarse en fases iniciales, incluso antes de que se detecten síntomas motores evidentes [11], lo que convierte a la voz en una herramienta prometedora para el desarrollo de modelos predictivos tempranos.

2.2. Descripción del conjunto de datos NeuroVoz_v3

El conjunto de datos **NeuroVoz v3** es una base de datos pública diseñada para el estudio del habla hispana en pacientes con EP. Fue desarrollado por la Universidad Politécnica de Madrid en colaboración con diversos hospitales de Madrid [12].

Contiene grabaciones de voz de 112 participantes, 58 individuos sanos y 54 pacientes con la EP, todos bajo la medicación habitual para asegurar condiciones homogéneas. El corpus contiene 2.977 archivos de audio, con un promedio aproximado de 27 grabaciones por participante.

El protocolo de grabación se diseñó para capturar diferentes aspectos del habla afectada por la EP, incluyendo las siguientes tareas principales:

- **Fonación sostenida:** consiste en pronunciar las vocales /a/, /e/, /i/, /o/ y /u/ en un tono cómodo, repitiendo cada una tres veces, con una duración media de alrededor de 4 segundos.
- **Escuchar y repetir:** los participantes escucharon y repitieron 16 frases predefinidas, método que reduce la carga cognitiva y favorece una articulación más natural.
- **Prueba diadocinéctica (DDK):** implica la repetición rápida de la secuencia /pa-ta-ka/ durante al menos 5 segundos para evaluar la articulación, midiendo la rapidez y coordinación de los movimientos del habla.

- **Monólogos espontáneos:** los sujetos describieron libremente una imagen predeterminada, con una duración media de 31 segundos para pacientes y 46 segundos para controles, que permite analizar patrones de habla natural y fluidez.

Cada grabación está asociada a metadatos clínicos y demográficos que enriquecen el análisis, así como la edad, sexo, diagnóstico, duración de la enfermedad y estado de medicación. También se registraron indicadores binarios para la presencia de síntomas específicos (como temblores vocales o disfonía), junto con evaluaciones clínicas según las escalas Unified Parkinson's Disease Rating Scale (UPDRS) y Hoehn-Yahr (H-Y).

Por otra parte, las grabaciones cuentan con transcripciones manuales detalladas para detectar errores de articulación y variaciones lingüísticas. Además, se extrajeron parámetros acústicos que miden perturbaciones en frecuencia y amplitud, niveles de ruido y características del temblor vocal, aportando información precisa sobre alteraciones en la voz relacionadas con la enfermedad.

2.3. Limitaciones y consideraciones

El conjunto de datos NeuroVoz v3 incluye grabaciones de audio de 112 participantes, lo que puede limitar la generalización de los. La calidad de las grabaciones, aunque es buena, podría mejorarse utilizando un entorno acústico más controlado para reducir ruidos de fondo y mejorar la precisión del análisis.

Al estar centrado exclusivamente en el habla en español, su aplicabilidad a otros idiomas y contextos culturales es limitada. Además, la mayoría de los pacientes se encontraban bajo medicación estable durante la grabación, lo que podría dificultar identificar ciertas características.

3. Modelado y técnicas a utilizar

El modelado es una etapa clave que abarca desde la exploración y selección de las características más relevantes hasta la implementación de los modelos predictivos. Además, se aplican técnicas sólidas de validación para garantizar resultados fiables, junto con ciertas métricas para evaluar el desempeño y la efectividad de las soluciones desarrolladas.

3.1. Análisis exploratorio de datos (EDA) y selección de características

Antes de la construcción de los modelos, se realizará un EDA para comprender la distribución, calidad y posibles anomalías o inconsistencias en las características disponibles. Se examinarán estadísticas descriptivas y se visualizarán los datos para identificar patrones y valores atípicos. En cuanto a la selección de características, se emplearán técnicas estadísticas, destacando el análisis de la matriz de correlación. También se aplicarán métodos basados en algoritmos como la selección por importancia en árboles de decisión o técnicas de reducción dimensional, con el fin de identificar las variables más relevantes para la detección de la enfermedad. Este paso es crucial para mejorar la interpretabilidad y el rendimiento de los modelos.

3.2. Modelos predictivos

Se explorarán diferentes enfoques para la clasificación de pacientes con y sin EP a partir de las características de voz disponibles. Entre los modelos a considerar se encuentran algoritmos tradicionales de aprendizaje automático, como la regresión logística, SVM y Random Forest, que han mostrado buen desempeño en problemas similares. También se evaluarán modelos basados en redes neuronales, que pueden capturar relaciones no lineales y patrones complejos en los datos. La elección final del modelo o combinación de modelos

dependerá de los resultados obtenidos durante las fases de validación y optimización.

3.3. Estrategias de validación

Para asegurar la generalización y robustez de los modelos, se implementarán estrategias de validación rigurosas. Se utilizarán técnicas como la validación cruzada, que permite evaluar el rendimiento del modelo en diferentes particiones del conjunto de datos, minimizando el riesgo de sobreajuste. Además, se considerarán métodos de balanceo de clases en caso de existir un desbalance significativo entre pacientes y sujetos sanos, para evitar sesgos en la predicción. La correcta definición de los conjuntos de entrenamiento, validación y prueba será fundamental para evaluar el desempeño real de los modelos.

3.4. Métricas de evaluación

La efectividad de los modelos evaluará mediante métricas adecuadas para problemas de clasificación binaria. Entre ellas, se incluirán la exactitud (accuracy), precisión, sensibilidad (recall), especificidad y la puntuación F1 (F1-Score), que equilibra precisión y sensibilidad. También se analizarán las curvas Receiver Operating Characteristic (ROC) y el área bajo la curva (AUC) para medir la capacidad discriminativa del modelo. Asimismo, se evaluará la matriz de confusión para obtener una visión detallada de los errores de clasificación, incluyendo falsos positivos y falsos negativos.

Dado el contexto del problema se pondrá un especial énfasis en el recall, al tratarse de un aspecto clave para la detección temprana de la enfermedad y la reducción del riesgo de pasar por alto casos positivos. Estas métricas ofrecerán una visión completa del rendimiento, ayudando a identificar tanto puntos fuertes como posibles aspectos a mejorar en los modelos.

4. Herramientas y entornos de desarrollo

El desarrollo de los modelos se realizará en Python, uno de los lenguajes más comunes para IA y ciencia de datos, con especial énfasis en bibliotecas y herramientas que se adaptan bien a los datos de voz e imágenes cerebrales.

4.1. Bibliotecas y entornos utilizados

En esta sección se describen las bibliotecas esenciales para el desarrollo de los modelos, que facilitarán tanto el manejo de datos como la construcción de modelos de IA:

- **scikit-learn**: Para implementar algoritmos clásicos como SVM, Random Forest y KNN, que serán utilizados en el análisis inicial de los datos y en la comparación con modelos más avanzados de deep learning.
- **TensorFlow y Keras**: Estas herramientas serán fundamentales para la construcción y entrenamiento de las Redes Neuronales Convolucionales (CNN) para las imágenes cerebrales y Redes Neuronales Recurrentes (RNN) con LSTM para los datos de voz.
- **pandas y NumPy**: Para la manipulación y análisis de los datos, como la conversión de las señales de voz en espectrogramas y la manipulación de imágenes cerebrales.
- **matplotlib y seaborn**: Para la visualización de los resultados y el análisis exploratorio de los datos, tanto de las señales de voz como de las imágenes cerebrales.

4.2. Entornos de desarrollo

El entorno de desarrollo juega un papel crucial en la eficiencia del trabajo, especialmente cuando se trabaja con grandes volúmenes de datos. Los siguientes entornos proporcionan la infraestructura adecuada para el desarrollo y la ejecución de los modelos:

- Google Colab: Se utilizará como entorno principal debido a su accesibilidad y soporte para GPU, lo cual es esencial para entrenar modelos complejos de deep learning, especialmente cuando se trabajan con grandes volúmenes de datos como los espectrogramas de voz y las imágenes cerebrales.
- Jupyter Notebooks: Se usará para documentar el proceso de desarrollo de manera detallada, facilitando la reproducibilidad de los experimentos.

4.3. Hardware

Se utilizarán recursos en la nube, como la GPU NVIDIA Tesla T4 de Google Colab, para entrenar modelos complejos de deep learning que requieren una mayor capacidad de procesamiento. Además, se contará con un equipo local equipado con un procesador Ryzen 7 5800X y una tarjeta gráfica RTX 5070, destinado a tareas de preprocesamiento de datos y pruebas preliminares.

Capítulo 5: Desarrollo del proyecto

La aplicación de técnicas de IA para identificar de manera precoz patologías complejas requiere un proceso detallado del desarrollo. Este proceso abarca desde la preparación y el análisis de los datos hasta la extracción de las características clave y la construcción de los modelos predictivos. Todo ello se acompaña de un trabajo cuidadoso de limpieza exploración y preprocesamiento de los datos, así como de la optimización y comparación de los modelos para obtener los mejores resultados posibles. Se trata de seleccionar la solución más adecuada para el problema planteado.

1. Exploración inicial y modelado con datos originales

Para establecer una primera aproximación al problema, se ha trabajado con los datos originales de NeuroVoz_v3, que incluyen tanto las características previamente extraídas de los audios de los pacientes como los metadatos asociados a cada persona.

1.1. Análisis de los metadatos

Los metadatos proporcionan información demográfica, clínica y técnica relacionada con los audios, fundamental para contextualizar el problema y enriquecer el análisis. A continuación, se describen las variables disponibles en los archivos `metadata_hc.csv` y `metadata_pd.csv`:

Variable	Descripción
ID	Identificador único del paciente o muestra.
Group	Grupo de origen del paciente: HC (sano) o PD (con Parkinson).
Date	Fecha de grabación del audio o evaluación clínica.
Sex	Sexo del paciente: 1 (hombre) o 0 (mujer).
Age	Edad del paciente en años.
Diagnosis	Información textual sobre el diagnóstico (puede estar vacía o redundante).
Vocal tremor	Presencia de temblor vocal. Valor binario (1 / 0).
Cephalic tremor	Presencia de temblor cefálico (cabeza). Valor binario (1 / 0).
Mandibular tremor	Presencia de temblor mandibular (mandíbula). Valor binario (1 / 0).
Sialorrhoea	Presencia de babeo excesivo. Valor binario (1 / 0).
Dysphagia	Presencia de dificultad al tragar. Valor binario (1 / 0).
Hypophonic voice	Voz hipofónica (débil o de baja intensidad).
Time Disease (years)	Tiempo desde el diagnóstico de la enfermedad de Parkinson (en años).
UPDRS scale	Puntuación en la escala UPDRS (Unified Parkinson's Disease Rating Scale).
H-Y Stadium	Estadio en la escala de Hoehn y Yahr (gravedad del Parkinson).
Date Evaluation Scales	Fecha de evaluación clínica o cuestionarios.
Medication	Medicación que toma el paciente.
Medication status	Estado del paciente respecto a la medicación (ON / OFF , etc.).
Occupation	Profesión del paciente.
Fiber/VocalFolds	Observación sobre las cuerdas vocales. (Normal / Not performed / Texto libre)
Observations	Notas clínicas adicionales.
Doctor	Nombre del médico o evaluador.
Audio	Nombre del archivo de audio asociado.

Tabla 1: Variables de los metadatos de NeuroVoz_v3

1.1.1. Carga de datos y exploración inicial

Se han cargado los metadatos de ambos archivos, unificándolos en un único DataFrame para facilitar su tratamiento y análisis conjunto. Para verificar que la estructura era correcta y que la carga de datos se realizó adecuadamente, se han mostrado las primeras filas del DataFrame.

1.1.2. Limpieza básica de datos

Para esta limpieza básica se han estandarizado los nombres de las columnas eliminando espacios en blanco para evitar inconsistencias. Además, se ha realizado una revisión de valores nulos y duplicados, junto con el cálculo de estadísticas descriptivas básicas de las variables.

1.1.3. Análisis exploratorio de datos (EDA)

Es fundamental entender la distribución y características de las variables presentes en los metadatos, especialmente aquellas categóricas y demográficas relevantes.

En primer lugar, se ha examinado la distribución de variables categóricas como el grupo (Control / Parkinson), diagnóstico, estado y tipo de medicación, así como el sexo de los participantes, detectando alguna inconsistencia leve que fueron corregidas. Para complementar este análisis, se han generado diversas visualizaciones que permiten observar la distribución y relación entre estas variables.

Entre las gráficas más relevantes destacan la distribución del grupo, que muestra un equilibrio razonable entre participantes sanos y con EP.

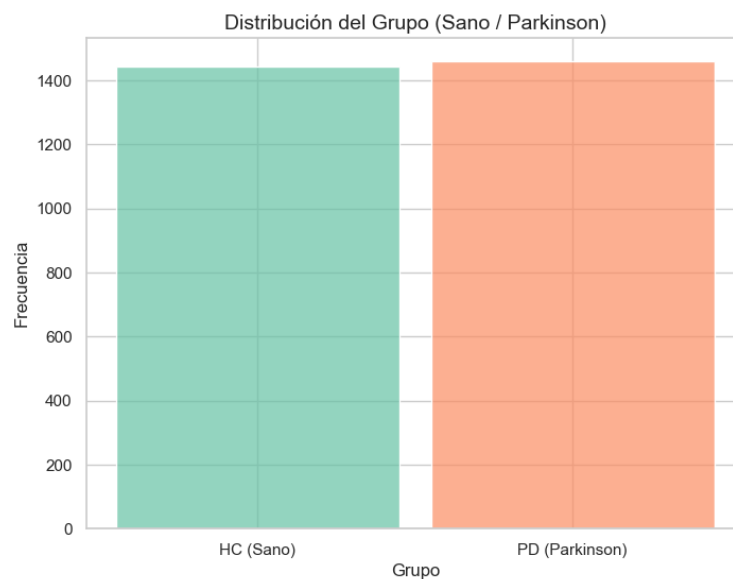


Ilustración 1: Distribución de Grupo

En la distribución por sexo, se observa una mayoría de hombres con un porcentaje menor de mujeres y algunos casos sin especificar. La relación entre

ambas variables muestra proporciones similares en ambos grupos, sin sesgos demográficos importantes.

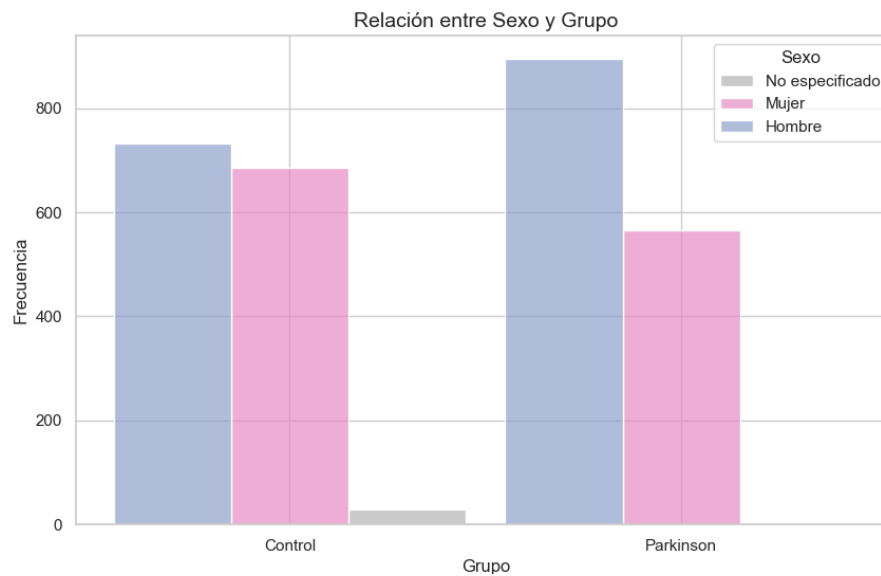


Ilustración 2: Relación entre Edad y Grupo

Por otro lado, se ha analizado la distribución de la edad, que presenta una media cercana a los 68 años y donde se observa que la mayoría de los participantes supera los 50 años, siendo escasos los casos por debajo de esa edad.

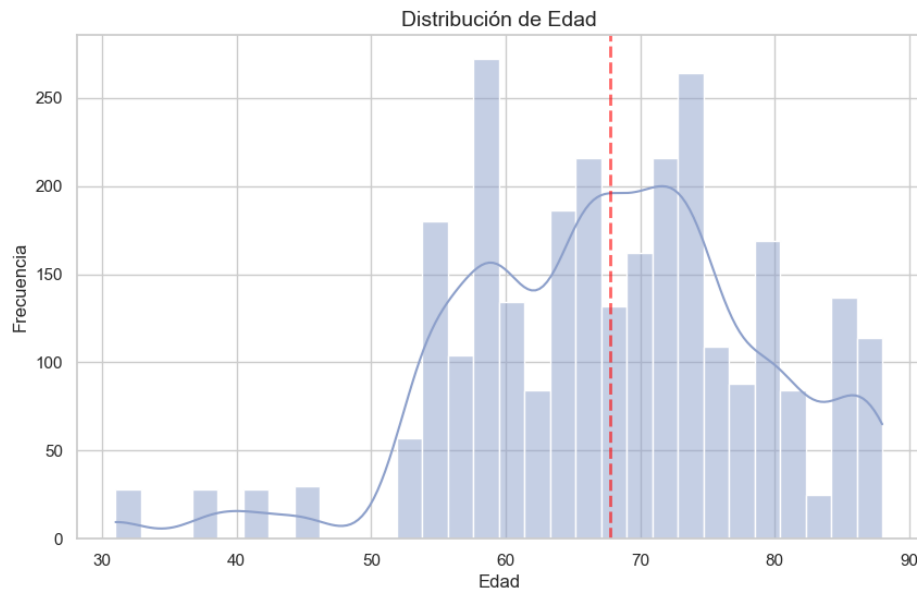


Ilustración 3: Distribución de Edad

También se ha explorado la relación entre la edad y el grupo, y otras características como la variabilidad en el diagnóstico textual y el tiempo desde el diagnóstico, aportando un contexto clínico relevante para el estudio.

1.2. Análisis y preprocesamiento de las características acústicas proporcionadas

Para continuar con el análisis, se ha trabajado con las características acústicas proporcionadas en el conjunto de datos NeuroVoz_v3, previamente extraídas de los audios de los pacientes. Estas servirán como base para el desarrollo y evaluación de los modelos de clasificación en las siguientes fases del estudio. A continuación, se describen las variables disponibles en el archivo `audio_features.csv`:

Variable	Descripción
AudioPath	Ruta al archivo de audio correspondiente.
Jitter	Variación absoluta del período fundamental, indicador de inestabilidad en la frecuencia (irregularidad).
rJitter	Jitter relativo: variación del período fundamental en porcentaje.
RAP	Jitter relativo promedio absoluto: medida de variación entre ciclos consecutivos.
rPPQ	Pitch perturbation quotient: variación de frecuencia a corto plazo.
rSPPQ	Smoothed pitch perturbation quotient: versión suavizada del PPQ.
ShimmerDb	Variación de amplitud (en dB) entre ciclos consecutivos (amplitud perturbation).
rShimmer	Shimmer relativo: porcentaje de variación de amplitud entre ciclos.
APQ	Amplitude perturbation quotient: variación promedio de amplitud.
sAPQ	Smoothed amplitude perturbation quotient: versión suavizada del APQ.
FTRI	Frequency tremor intensity index: mide la intensidad del temblor en frecuencia.
ATRI	Amplitude tremor intensity index: mide la intensidad del temblor en amplitud.
FFTR	Fundamental frequency tremor frequency: frecuencia del temblor en el tono fundamental.
ATRF	Amplitude tremor frequency: frecuencia del temblor en amplitud.
NNE	Noise-to-noise energy ratio: proporción de energía del ruido respecto al tono.
HNR	Harmonics-to-noise ratio: relación entre la energía de los armónicos y el ruido.
CHNR	Cepstral harmonics-to-noise ratio: variante cepstral del HNR.
GNE	Glottal-to-noise excitation ratio: medida de la relación señal-ruido glotal.
CPP	Cepstral peak prominence: prominencia del pico cepstral (relacionado con la calidad vocal).

Tabla 2: Características acústicas de NeuroVoz_v3

1.2.1. Carga de datos y exploración inicial

Se han cargado las características proporcionadas, guardándolas en un DataFrame para su análisis. Para verificar que la estructura de los datos es la esperada y que la carga se ha realizado correctamente, se han mostrado las primeras filas del conjunto, permitiendo una comprobación inicial de las variables y su formato.

1.2.2. Limpieza básica de datos

Para la limpieza inicial de las características acústicas, se ha seguido un procedimiento similar al de los metadatos, estandarizando los nombres de las columnas y examinando las dimensiones del conjunto de datos, así como los tipos de las variables y la presencia de valores nulos o duplicados. También se han calculado las estadísticas descriptivas básicas para identificar posibles anomalías.

Cabe destacar que el conjunto de características acústicas contiene 1035 filas, significativamente menor que los casi 3000 audios disponibles, lo que indica que no se dispone de características para todos los audios y que el archivo CSV no estaba completo.

1.2.3. Análisis exploratorio de datos (EDA)

Se ha realizado un análisis exploratorio para comprender la distribución de las características acústicas. En primer lugar, se ha extraído la clase (HC / PD) a partir de la ruta del archivo de audio, creando una nueva columna para el grupo en el DataFrame. Se ha comprobado que esta información es coherente con la de los metadatos, manteniéndose un reparto prácticamente equilibrado entre ambos grupos.

Posteriormente, se han generado visualizaciones de la distribución de las variables numéricas mediante histogramas y diagramas de caja, para identificar posibles valores atípicos y evaluar la dispersión de los datos. En general, muchas de las distribuciones presentan cierto sesgo, acumulando la mayoría de los datos en uno de los extremos. Además, los diagramas de caja presentan numerosos valores atípicos.

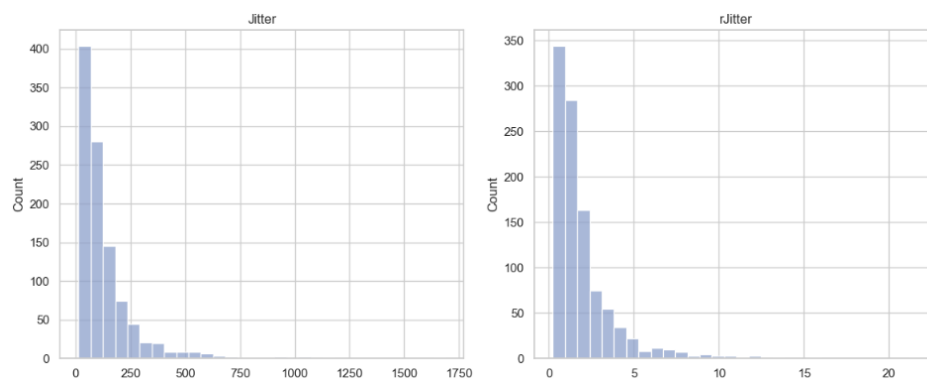


Ilustración 4: Histogramas de Jitter y rJitter (datos originales)

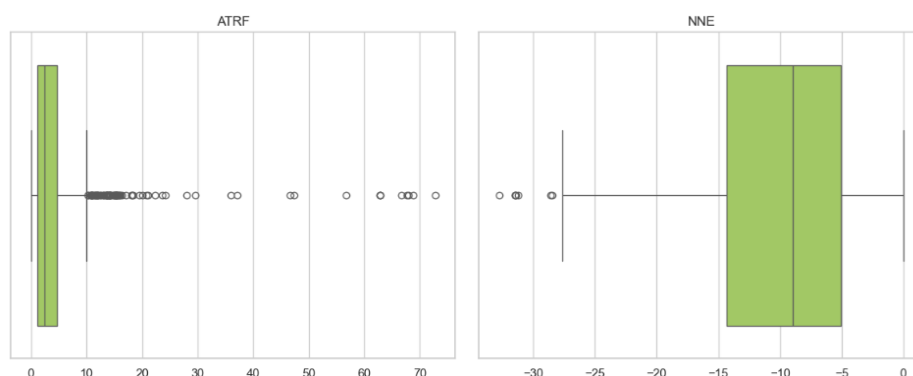


Ilustración 5: Diagramas de caja de ATRF y NNE (datos originales)

Esto podría deberse a la propia variabilidad natural de las características acústicas en la voz de los pacientes, donde algunas métricas reflejan desviaciones asociadas a la intensidad o a particularidades propias del EP.

Finalmente, se ha calculado la matriz de correlación para analizar las relaciones entre las distintas características. Esto permite identificar variables muy correlacionadas entre sí, como es el caso de rJitter y RAP, que muestran una correlación muy alta, o ShimmerDb, rShimmer y APQ, que también muestran cierta correlación. Esto podría indicar redundancia entre variables, que se podría reducir en fases posteriores.

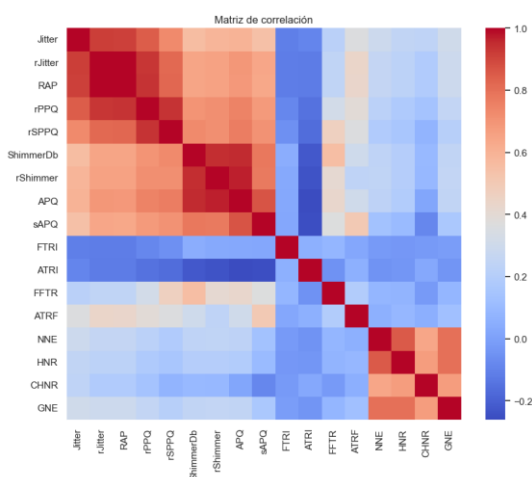


Ilustración 6: Matriz de correlación de las características acústicas de NeuroVoz_v3

1.2.4. Preprocesamiento de datos post-EDA

Tras el análisis exploratorio, se ha realizado un preprocesamiento de las características para optimizar su uso en los modelos posteriores. En primer lugar, se ha establecido un umbral de correlación del 0.95, eliminando aquellas variables cuya correlación superaba dicho valor. Como resultado, se eliminaron las variables RAP y APQ, reduciendo el conjunto inicial de 20 a 18 características.

Respecto a la variable objetivo, se ha codificado el grupo Group en formato binario, asignando 0 a los sujetos sanos (HC) y 1 a los pacientes con EP (PD).

Finalmente, se ha eliminado la columna AudioPath del DataFrame, puesto que resultaba irrelevante para el modelado. Además, la variable CPP también fue descartada, debido a la gran cantidad de datos incompletos. Esta característica estaba compuesta por un array de valores, donde en muchos casos mostraba solo algunos valores acompañados de puntos suspensivos. Esto dificultaba su interpretación y uso directo en los modelos.

1.3. División del conjunto de datos

Se han separado las variables predictoras (X) de la variable objetivo (y). Posteriormente, se ha dividido el conjunto de datos en subconjuntos de entrenamiento y prueba, asignando un 80 % para entrenamiento y un 20 % para evaluación. La división se realiza de forma estratificada para mantener la proporción original de clases en ambos subconjuntos, asegurando así una representación equilibrada.

1.4. Modelado y evaluación con las características proporcionadas

Como enfoque inicial, se han desarrollado tres modelos de clasificación entrenados con los datos acústicos de NeuroVoz_3, con el objetivo de evaluar el rendimiento de cada uno y comparar su capacidad para diferenciar entre sujetos sanos y pacientes con EP. Esto permite obtener una primera impresión sobre la efectividad de los modelos con los datos disponibles.

Para todos los modelos se ha realizado una búsqueda de los hiperparámetros óptimos mediante validación cruzada. La evaluación de cada modelo incluye métricas como precisión, recall, F1-Score y AUC-ROC, además de la visualización de la matriz de confusión y la curva ROC para analizar su desempeño de forma detallada.

1.4.1. Modelo de Regresión Logística

La regresión logística es un modelo lineal sencillo y eficiente, que suele funcionar bien con conjuntos de datos de tamaño moderado y con características no demasiado complejas. Su interpretabilidad y rapidez de entrenamiento lo hacen adecuado como referencia inicial.

Este primer modelo se ha construido utilizando una cadena de pasos (pipeline) que incluye imputación de valores faltantes mediante la media, escalado estándar de las características y el propio clasificador de regresión logística.

Los resultados obtenidos han mostrado una precisión global del 62%, con un rendimiento relativamente equilibrado entre ambas clases. Sin embargo, estas métricas no son especialmente buenas y reflejan que el modelo está limitado a la hora de distinguir con claridad entre sujetos sanos y pacientes con EP. La matriz de confusión confirma una tasa moderada de aciertos, pero también un número considerable de errores en ambas clases. El área bajo la curva ROC (AUC-ROC) fue de 0.652, indicando una capacidad reducida de discriminación.

1.4.2. Modelo SVM

El clasificador SVM es un modelo que busca un hiperplano óptimo para separar las clases en el espacio de características, pudiendo manejar tanto separaciones lineales como no lineales mediante el uso de diferentes kernels. Su capacidad para modelar relaciones complejas y su robustez frente a datos con alta dimensionalidad lo hacen adecuado para problemas como el presente.

Este modelo se ha construido mediante un pipeline que incluye imputación de valores faltantes con la media, escalado estándar y el propio clasificador SVM. Se ha llevado a cabo una búsqueda de hiperparámetros por validación cruzada, explorando distintos valores de regularización (C), tipos de kernel (lineal y RBF) y modos de cálculo del parámetro gamma.

Los resultados mostraron una mejora respecto a la regresión logística, con una precisión global del 69%, y un mejor equilibrio en las métricas de precisión y recall, especialmente para la clase de pacientes con Parkinson. La matriz de confusión también ha mostrado algunas mejoras, aunque aún persisten errores en ambas clases. El AUC-ROC ha alcanzado el valor de 0.727, indicando una capacidad discriminativa más elevada, aunque aún con cierto margen de mejora.

1.4.3. Modelo Random Forest

El clasificador Random Forest es un modelo basado en la combinación de múltiples árboles de decisión, que funcionan en conjunto para mejorar la precisión y evitar el sobreajuste. Su capacidad para manejar datos no lineales, captar interacciones complejas entre variables y su robustez frente al ruido lo hacen especialmente útil.

Este modelo se ha construido mediante un pipeline que incluye imputación de valores faltantes con la media, escalado estándar y el clasificador Random Forest. Se realizó una búsqueda de hiperparámetros mediante validación cruzada, explorando diferentes números de árboles, profundidad máxima, y parámetros que controlan la división y tamaño mínimo de las hojas, así como la selección de características para la división.

Los resultados obtenidos han mostrado una precisión global del 63%, con un rendimiento algo mejor para la clase de pacientes con Parkinson. Sin embargo, el modelo ha presentado una tasa relativamente alta de falsos negativos y falsos positivos, como refleja la matriz de confusión. El AUC-ROC ha sido de 0.716, indicando una capacidad discriminativa aceptable, pero que no supera al SVM.

1.4.4. Conclusiones preliminares y limitaciones detectadas

Los modelos desarrollados ofrecen una primera aproximación para clasificar entre sujetos sanos y pacientes con EP a partir de las características acústicas de NeuroVoz_3. De los tres, el clasificador SVM ha mostrado el mejor rendimiento general. Sin embargo, ninguno alcanza una precisión o capacidad discriminativa lo suficientemente alta como para considerarse fiable.

Las principales limitaciones detectadas están relacionadas con la cantidad y calidad de los datos, así como con la presencia de características que contienen valores incompletos o inconsistentes. Aunque se ha empleado validación cruzada para ajustar los hiperparámetros y evitar el sobreajuste, los resultados indican que es necesario explorar métodos más avanzados, incorporar nuevas variables o utilizar técnicas más sofisticadas de extracción de características.

En definitiva, estos resultados representan un buen punto de partida que aporta información útil, pero requieren mejoras y estudios adicionales para avanzar hacia un sistema de diagnóstico robusto y fiable.

2. Procesamiento y extracción propia de características de los audios

Dado que los resultados obtenidos con las características iniciales han mostrado ciertas limitaciones, se ha decidido aprovechar el acceso a los audios originales para realizar un procesamiento propio y más exhaustivo. El objetivo es limpiar y preparar adecuadamente las grabaciones, eliminando ruidos o señales no deseadas que puedan afectar la calidad de la señal. Esto permitirá extraer tanto las características iniciales como otras adicionales, de forma más completas y fiables. De este modo, se espera mejorar la calidad de los datos y, en consecuencia, el rendimiento de los modelos de clasificación.

2.1. Procesamiento y limpieza de audios

Se ha desarrollado el script `audio_processing.py` para limpiar y preparar de forma exhaustiva los audios originales. En este proceso, los archivos se cargan y se convierten a una sola pista de sonido, conocida como mono, con una frecuencia de muestreo uniforme de 16 kHz. Esto asegura que todas las grabaciones tengan el mismo formato y calidad, facilitando su comparación y análisis posterior. Además, se normaliza el volumen de los audios a

aproximadamente -20 decibelios relativos a Full Scale (dBFS), lo que ayuda a mantener una intensidad similar en todas las señales.

Además, el script detecta y elimina los silencios largos que puedan aparecer al principio o al final de las grabaciones, aplicando umbrales específicos para evitar incluir partes sin información relevante. Esta limpieza ayuda a reducir ruidos y pausas innecesarias, mejorando la calidad del audio para la extracción de características en las etapas siguientes.

2.2. Extracción de características acústicas y espectrales

Para la extracción de características acústicas y espectrales de los audios, se ha desarrollado el script *audio_features_extracting.py*. Este script automatiza la obtención de una amplia variedad de parámetros relevantes para el análisis vocal, combinando métodos basados en dos librerías especializadas, Parselmouth y librosa.

Primero, se extraen características acústicas básicas utilizando Parselmouth, que incluyen medidas del tono fundamental (pitch), así como parámetros de variabilidad en la frecuencia y amplitud, como jitter y shimmer, respectivamente. Además, se calcula la relación entre armónicos y ruido (HNR), que indica la calidad vocal y el nivel de ruido presente. Para complementar, se implementa una función que calcula la prominencia del pico cepstral (CPP), un índice relacionado con la claridad del tono.

Por otra parte, usando la librería librosa, se extraen los coeficientes cepstrales en las frecuencias de Mel (MFCCs), que capturan propiedades espectrales del sonido muy útiles para tareas de reconocimiento y análisis acústico. Se calculan tanto la media como la desviación estándar de estos coeficientes para reflejar la variabilidad temporal.

El script procesa automáticamente todos los archivos de audio en un directorio específico, extrae estas características y las combina en un único diccionario por archivo. Finalmente, los resultados se guardan en un fichero CSV para facilitar su análisis y uso en etapas posteriores.

2.3. Análisis y preprocesamiento de las nuevas características

Ahora que los audios han sido procesados y se han extraído las características acústicas y espectrales de forma completa y controlada, se ha realizado un análisis similar al aplicado anteriormente. Esto implica cargar los datos, explorar su estructura y calidad, limpiar valores inconsistentes o erróneos, y llevar a cabo un análisis exploratorio para detectar patrones o posibles problemas. Finalmente, se han aplicado las técnicas de preprocesamiento necesarias para preparar estas nuevas características para el posterior entrenamiento de los modelos. A continuación, se describen las nuevas variables disponibles en el archivo `audio_features_new.csv`:

Variable	Descripción
<code>duration_s</code>	Duración total del audio en segundos.
<code>mean_F0_Hz</code>	Frecuencia fundamental media (pitch) en Hertzios, indicador del tono base de la voz.
<code>stdev_F0_Hz</code>	Desviación estándar de la frecuencia fundamental, mide la variabilidad tonal.
<code>jitter_local</code>	Variación absoluta local del período fundamental, indicador de inestabilidad en la frecuencia.
<code>rJitter_rap</code>	Jitter relativo promedio absoluto (RAP), variación relativa de frecuencia entre ciclos consecutivos.
<code>rJitter_ppq5</code>	Jitter relativo Pitch Perturbation Quotient a corto plazo (PPQ5).
<code>rJitter_ddp</code>	Jitter relativo DDP (derivado de la diferencia de períodos).
<code>shimmer_local</code>	Variación local absoluta de la amplitud entre ciclos consecutivos.
<code>rShimmer_local_dB</code>	Shimmer relativo medido en decibelios (variación porcentual de amplitud).
<code>APQ3_Shimmer</code>	Amplitude Perturbation Quotient calculado en ventanas de 3 ciclos, mide variabilidad de amplitud.
<code>APQ5_Shimmer</code>	Amplitude Perturbation Quotient en ventanas de 5 ciclos.
<code>sAPQ_Shimmer_dda</code>	Smoothed Amplitude Perturbation Quotient (dda), versión suavizada del APQ.
<code>mean_HNR_dB</code>	Relación armónicos-ruido media en decibelios, indica calidad vocal y nivel de ruido.
<code>CPP</code>	Cepstral Peak Prominence, prominencia del pico cepstral, relacionada con la claridad del tono.
<code>mfcc_1-13_mean</code>	Media de los 13 coeficientes cepstrales en escala Mel, que representan características espectrales del audio.
<code>mfcc_1-13_std</code>	Desviación estándar de los 13 coeficientes cepstrales en escala Mel, reflejando su variabilidad temporal.
<code>AudioPath</code>	Ruta al archivo de audio correspondiente.

Tabla 3: Descripción de las características acústicas y espectrales extraídas

2.3.1. Carga de datos y exploración inicial

Se han cargado las características extraídas y almacenadas en el nuevo archivo CSV, guardándolas en un DataFrame para su posterior análisis. Para verificar que la estructura de los datos es la esperada y que la carga se realizó correctamente, se han mostrado las primeras filas del conjunto, lo que permite una comprobación inicial de las variables y su formato.

2.3.2. Limpieza básica de datos

Para asegurar la calidad y coherencia del conjunto de datos, se han realizado varias operaciones básicas de limpieza. Aunque los nombres de las columnas ya fueron definidos correctamente en el script de extracción se han eliminado posibles espacios en blanco al inicio y al final de estos nombres como una medida preventiva, para evitar posibles errores derivados del formato al exportar o cargar el archivo CSV.

Por otro lado, se han inspeccionado las dimensiones del DataFrame para confirmar el tamaño del conjunto de datos, y se han revisado los tipos de datos junto con la presencia de valores nulos en cada columna para detectar posibles inconsistencias.

Finalmente, se han identificado y contabilizado posibles filas duplicadas para valorar su eliminación o revisión, aunque no se han encontrado ninguna.

2.3.3. Análisis exploratorio de datos (EDA)

Siguiendo un procedimiento similar al EDA de los datos acústicos originales, en primer lugar, se ha extraído una variable adicional llamada Group a partir del nombre del archivo de audio, que permite clasificar las muestras en dos grupos principales: HC (sanos) y PD (Parkinson). Además, se ha verificado la

distribución de estas etiquetas mostrando sus frecuencias y se ha comprobado la ausencia de valores nulos en esta nueva columna.

Nuevamente, se han analizado visualmente las características numéricas mediante histogramas y diagramas de caja (boxplots). Los histogramas muestran que, en comparación con los datos originales proporcionados en el conjunto NeuroVoz_v3, las nuevas características presentan distribuciones mucho menos sesgadas, aunque persisten algunos sesgos en ciertos parámetros.

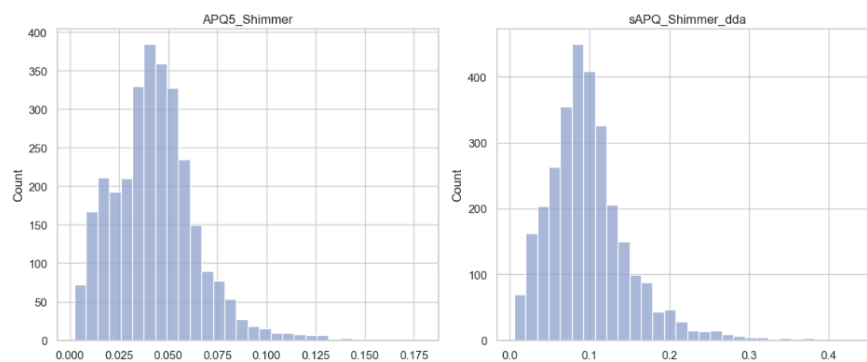


Ilustración 7: Histogramas de APQ5_Shimmer y sAPQ_Shimmer_dda (datos extraídos)

Por otra parte, los diagramas de caja siguen mostrando múltiples valores atípicos o anomalías fuera del rango esperado. Sin embargo, tal y como se justifica previamente, estas anomalías pueden deberse a la variabilidad natural en la señal vocal y no necesariamente a errores o datos corruptos.

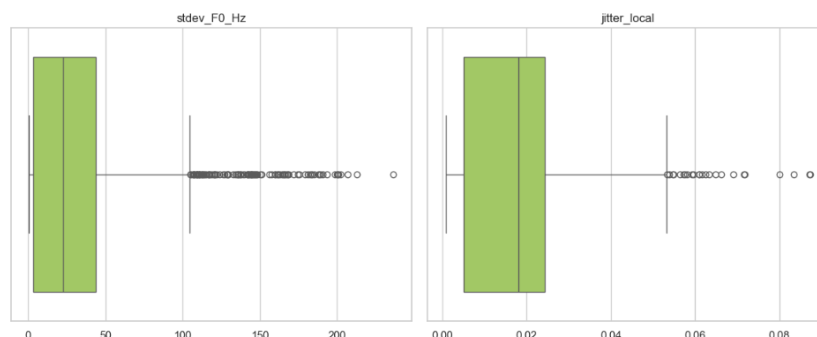


Ilustración 8: Diagramas de caja de stdev_F0_Hz y jitter_local (datos extraídos)

Finalmente, se calcula y visualiza la matriz de correlación para evaluar las relaciones entre las variables numéricas, aportando información útil para futuras selecciones o transformaciones de características.

2.3.4. Preprocesamiento de datos post-EDA

Se ha calculado la matriz de correlación absoluta entre las variables numéricas para identificar relaciones muy altas que pudieran indicar redundancia en las características. Se ha definido un umbral de correlación mayor a 0.95 para detectar variables altamente correlacionadas.

Como resultado, se han identificado cinco variables que superan este umbral y que se consideran redundantes: rJitter_rap, rJitter_ppq5, rJitter_ddp, rShimmer_local_dB y sAPQ_Shimmer_dda. Estas características han sido eliminadas para evitar multicolinealidad y mejorar la eficiencia del modelo.

Tras esta selección, el conjunto de datos reducido mantiene la mayoría de las variables, pasando de 42 a 37 columnas.

Finalmente, se ha preparado el conjunto de datos para el modelado, eliminando columnas irrelevantes como la ruta del archivo (AudioPath) y convirtiendo la variable objetivo Group en un valor binario, donde HC (sano) se codifica como 0 y PD (Parkinson) como 1.

2.3.5. División del conjunto de datos

Se han separado las variables predictoras (X) de la variable objetivo (y). Posteriormente, se ha dividido el conjunto de datos en subconjuntos de entrenamiento y prueba, asignando un 80 % para entrenamiento y un 20 % para evaluación. La división se realiza de forma estratificada para mantener la proporción original de clases en ambos subconjuntos, asegurando así una representación equilibrada.

2.4. Entrenamiento y evaluación de modelos con las nuevas características

Se han empleado los modelos previamente definidos y desarrollados para entrenar y evaluar el desempeño con las nuevas características acústicas y espectrales extraídas. La estructura y configuración de los modelos permanece sin cambios, pero la incorporación de estas variables ha permitido obtener resultados significativamente mejorados. Además, se ha añadido un nuevo modelo basado en Gradient Boosting (XGBoost), que se ha entrenado y evaluado junto con los demás para completar el análisis comparativo.

2.4.1. Modelo de Deep Learning

El modelo ha logrado un AUC ROC de 0.809 en el conjunto de test, mostrando un buen equilibrio entre precisión y recall para ambas clases, con una precisión global del 73%. La matriz de confusión refleja que, de 302 muestras sanas, 230 se clasifican correctamente y 72 se clasifican erróneamente como Parkinson. Por otro lado, de 294 muestras de Parkinson, 207 se detectan correctamente y 87 se clasifican como sanas. La curva ROC confirma una buena capacidad de discriminación del modelo.

2.4.2. Modelo SVM

El modelo SVM ha conseguido un AUC ROC de 0.942 en el conjunto de test, lo que indica una excelente capacidad de discriminación entre las clases. La precisión global es del 86%, con un buen equilibrio entre precisión y recall para ambos grupos. Según la matriz de confusión, de las 302 muestras sanas, 271 se clasifican correctamente y 31 se etiquetan erróneamente como Parkinson. En cuanto a las 294 muestras de Parkinson, 239 se detectan correctamente y 55 se clasifican como sanas. Los valores de f1-score para ambas clases reflejan un desempeño sólido y equilibrado del modelo.

2.4.3. Modelo Random Forest

El modelo Random Forest ha alcanzado un AUC ROC de 0.918 en el conjunto de test, mostrando un buen rendimiento general. La precisión global es del 83%, con un equilibrio notable entre precisión y recall en ambas clases. La matriz de confusión indica que de las 302 muestras sanas, 251 se clasifican correctamente, mientras que 51 se etiquetan erróneamente como Parkinson. Por otro lado, de las 294 muestras con Parkinson, 243 se detectan correctamente y 51 se clasifican como sanas. Los valores del f1-score reflejan un desempeño consistente y equilibrado del modelo.

2.4.4. Modelo Gradient Boosting

El modelo Gradient Boosting, implementado con XGBoost, es un clasificador basado en árboles que mejora iterativamente los errores de predicción. Se construyó un pipeline con imputación de la media, escalado estándar y búsqueda de hiperparámetros mediante validación cruzada para optimizar parámetros como la profundidad, tasa de aprendizaje y número de estimadores.

Los mejores parámetros obtenidos fueron 500 estimadores, profundidad máxima 7 y tasa de aprendizaje 0.1. El modelo alcanzó una precisión del 85% en test, con un buen balance entre precisión y recall en ambas clases.

La matriz de confusión mostró un buen desempeño con pocos errores, y la curva ROC alcanzó un AUC de 0.928, destacando su capacidad discriminativa superior a otros modelos evaluados.

3. Modelado avanzado con Deep Learning

Los resultados obtenidos con los modelos iniciales entrenados con los nuevos datos extraídos han sido más que aceptables, aunque quizás se puedan mejorar con otros enfoques más avanzados como en el caso de las redes neuronales profundas.

El uso de redes neuronales facilita la captura de interacciones y patrones más sutiles en las características extraídas, algo especialmente útil en problemas biomédicos como el presente, donde las señales pueden presentar relaciones no triviales. El objetivo es comprobar si mediante este tipo de modelado es posible obtener un modelo más robusto y preciso.

3.1. Arquitectura y diseño del modelo de red neuronal

El modelo consiste en una red multicapa secuencial compuesta por entre 1 y 3 capas ocultas densas con función de activación *ReLU*. El número de neuronas por capa varía entre 64 y 512 y se complementa con capas de *dropout* con tasas entre 0.2 y 0.5 para reducir el riesgo de sobreajuste. La salida es una única neurona con activación sigmoide adecuada para clasificación binaria. La red se entrena con el optimizador Adam.

3.2. Optimización de hiperparámetros con Keras Tuner

Se ha empleado Keras Tuner mediante búsqueda aleatoria para ajustar el número de capas, unidades por capa, tasas de dropout y la tasa de aprendizaje del optimizador. La optimización se ha realizado mediante validación cruzada estratificada en 5 particiones, maximizando el AUC en validación.

Este enfoque no solo permite encontrar la mejor combinación de hiperparámetros, sino que asegura que dicha combinación ofrece un rendimiento consistente y estable en distintas particiones de los datos, al seleccionar los parámetros en función del AUC medio obtenido en los diferentes folds. Una vez identificada la mejor configuración, el modelo final se ha reentrenado usando el conjunto de entrenamiento completo con validación interna.

3.3. Resultados preliminares y estado actual

El modelo de red neuronal profunda, ajustado mediante validación cruzada y búsqueda de hiperparámetros con Keras Tuner, logró un rendimiento destacado en el conjunto de validación, alcanzando un AUC cercano a 0.9081. Tras el reentrenamiento final con los mejores hiperparámetros encontrados, el modelo mostró una capacidad robusta para discriminar entre las clases en el conjunto de test, consiguiendo un AUC ROC de 0.910.

Además, se obtuvo un buen equilibrio entre precisión y recall en ambas clases, con una exactitud global del 82 %. Esto indica que el uso de redes neuronales permitió capturar relaciones no lineales y complejas en los datos que, en este caso, se tradujeron en un modelo competitivo y estable.

4. Comparativa y análisis conjunto de resultados

Para la comparación de los resultados obtenidos de todos los modelos con los nuevos datos extraídos, se ha realizado una representación gráfica un diagrama de barras que resume las métricas clave de clasificación (accuracy, precision, recall y F1 score) para cada modelo, y una superposición de las curvas ROC que permite evaluar y contrastar la capacidad discriminativa global de cada clasificador.

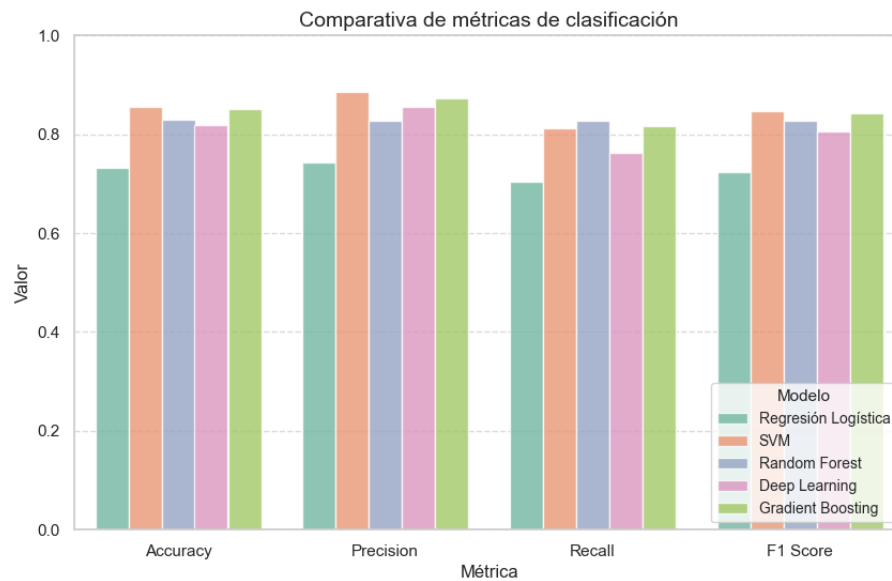


Ilustración 9: Comparativa de métricas de clasificación para los distintos

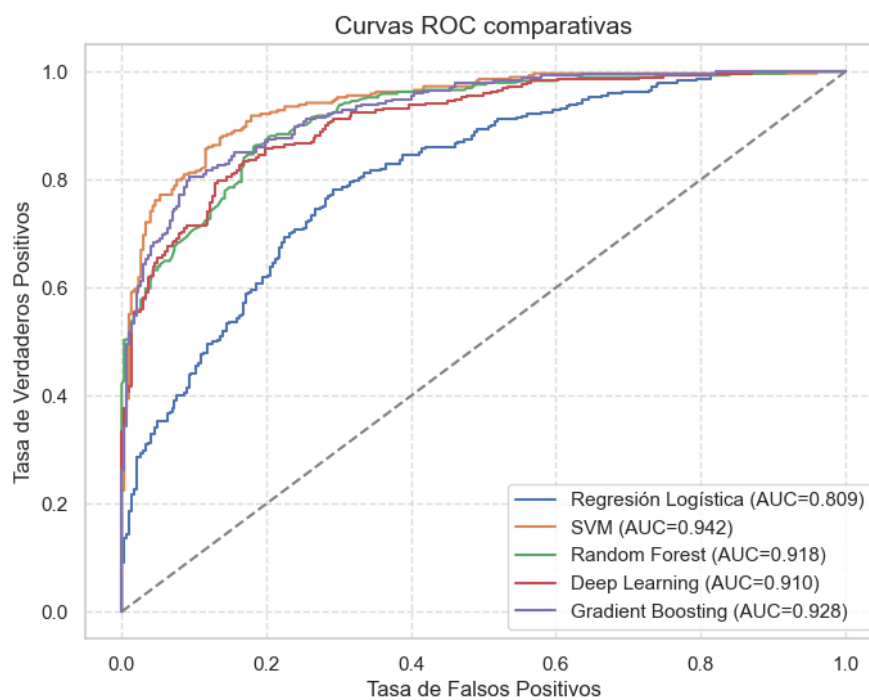


Ilustración 10: Curvas ROC comparativas de los modelos evaluados

Aunque todos los modelos muestran un rendimiento competitivo, el clasificador SVM destaca por un equilibrio más consistente entre las

métricas, manteniendo buenos valores tanto en precisión como en recall. Esto sugiere que el SVM ofrece una clasificación robusta y balanceada entre la detección correcta de casos positivos y negativos, lo que es clave en aplicaciones donde ambos tipos de error tienen consecuencias relevantes.

4.1. Ajuste del umbral de decisión en el modelo SVM

En problemas de clasificación binaria, el umbral estándar para asignar una clase suele ser 0.5 en la probabilidad predicha. Sin embargo, este valor puede no ser óptimo dependiendo de la aplicación y de qué tipo de error se quiera priorizar minimizar. En el contexto de detección de la EP, resulta fundamental maximizar la capacidad del modelo para identificar correctamente a los pacientes afectados (alta sensibilidad o recall), ya que un falso negativo podría retrasar un diagnóstico crucial para el tratamiento.

Por ello, se ha evaluado el comportamiento del modelo SVM (el mejor según la comparativa previa) ajustando el umbral de decisión en distintos valores, desde 0.7 hasta 0.2. Este ajuste permite observar cómo varían las métricas de precisión, recall y F1-score para encontrar un punto de compromiso adecuado que priorice la detección correcta de casos positivos sin sacrificar excesivamente la precisión.

Dado que la prioridad es maximizar el recall sin perder demasiado en precisión, se ha seleccionado el umbral con valor 0.3, que presenta un recall alto, 0.92, junto con un F1-score máximo, 0.87, y una precisión razonable, 0.83, lo que indica un buen equilibrio para la detección temprana de Parkinson en este contexto.

4.2. Comparación con estudios previos

En este proyecto se ha alcanzado un recall del 92.3% y una precisión del 83% en la detección temprana de la EP a partir del análisis de características

acústicas. Estos resultados son comparables a los reportados por Di Cesare et al. (2024), quienes lograron una precisión del 91.5% y un recall del 93.1% utilizando coeficientes MFCC y GTCC junto con modelos SVM, KNN y redes neuronales.

Mientras que su estudio contó con un conjunto de datos mayor y aplicó técnicas de diarización para separar las voces en grabaciones de conversaciones captadas con dispositivos móviles, el presente proyecto se basó en aproximadamente 3.000 audios individuales. Para esta cantidad “relativamente reducida” de datos, los resultados obtenidos son bastante competitivos.

Esta comparación destaca la eficacia del enfoque empleado y sugiere que, a pesar de las limitaciones en cantidad de datos y diferencias metodológicas, el análisis de voz sigue siendo una herramienta prometedora para la detección precoz de la EP.

Capítulo 6: Conclusiones

En este capítulo se presentan los resultados alcanzados durante el desarrollo del proyecto, destacando el cumplimiento de los objetivos planteados inicialmente, así como las lecciones aprendidas y las oportunidades de mejora identificadas a partir del análisis realizado. Este análisis permite valorar el trabajo realizado tanto desde un punto de vista técnico como personal.

1. Cumplimiento de objetivos

A lo largo del desarrollo de este proyecto, se ha logrado cumplir con los objetivos establecidos inicialmente, tanto principales como secundarios. Se ha llevado a cabo un análisis detallado del conocimiento actual sobre la enfermedad de la EP, incluyendo sus características clínicas y posibles biomarcadores tempranos. Se han estudiado las aplicaciones de la IA en el ámbito médico, con especial atención a los modelos para enfermedades neurodegenerativas, y se evaluó su aplicabilidad en el diagnóstico precoz de la EP.

Asimismo, se han investigado diferentes tipos de datos relevantes para la detección temprana, seleccionando los más adecuados y realizando un preprocesamiento y análisis exploratorio exhaustivo. Se han diseñado, entrenado y evaluado varios modelos de IA, desde técnicas tradicionales hasta enfoques de aprendizaje profundo, optimizando el equilibrio entre precisión, sensibilidad y generalización.

Finalmente, se ha evaluado rigurosamente el rendimiento de los modelos mediante métricas estándar y se han comparado los resultados con trabajos similares. En cuanto al objetivo secundario, se ha abordado el desarrollo de enfoques avanzados para mejorar la precisión y robustez del modelo,

utilizando técnicas de aprendizaje profundo, cumpliendo así con la propuesta inicial de forma ampliada.

2. Limitaciones y desafíos

Uno de los principales desafíos al inicio del proyecto fue que el CSV con las características acústicas de los audios estaba incompleto y contenía datos inconsistentes, lo que dificultó el análisis inicial. Para solucionar esto, fue necesario desarrollar scripts propios para procesar manualmente los audios y extraer de nuevo las características relevantes.

Como limitación destacable, el conjunto de datos contaba con alrededor de 3,000 audios, una cantidad moderada que, aunque suficiente para obtener resultados interesantes, limita la capacidad de generalización y robustez del modelo. Estas dificultades técnicas y de datos marcaron el desarrollo, subrayando la importancia de mejorar la calidad y cantidad de información en futuras investigaciones.

3. Lecciones aprendidas

A lo largo del desarrollo de este proyecto, he podido comprobar lo crucial que es disponer de datos de calidad y bien organizados para que los modelos de IA funcionen correctamente. El tener que procesar y extraer manualmente las características de los audios me hizo ver lo importante que es dedicar tiempo a la preparación y limpieza de los datos antes de cualquier análisis.

También aprendí que no siempre el uso de técnicas más avanzadas garantiza mejores resultados. En este caso, los modelos tradicionales como el SVM ofrecieron un desempeño más sólido que algunos enfoques de aprendizaje profundo, lo que demuestra la importancia de evaluar y seleccionar cuidadosamente la metodología según el contexto y los datos disponibles.

Además, este trabajo me ha permitido mejorar mis habilidades técnicas en el manejo de datos, programación para extracción de características y evaluación de modelos, pero también me ha dado una visión más clara sobre las oportunidades y limitaciones actuales de la IA para el diagnóstico temprano de enfermedades neurodegenerativas.

Por último, esta experiencia me ha enseñado la importancia de ser paciente y constante, manteniendo la flexibilidad ante los desafíos técnicos y metodológicos que aparecen durante un proyecto de investigación.

4. Consideraciones finales y reflexiones personales

Este proyecto me ha permitido consolidar y poner en práctica conocimientos técnicos y metodológicos adquiridos durante el Máster, especialmente en el manejo de datos, diseño de modelos y evaluación rigurosa. Los resultados obtenidos, aunque condicionados por el tamaño limitado del conjunto de datos, son alentadores y muestran el potencial de las técnicas aplicadas para avanzar en la detección temprana de la EP.

A nivel personal y profesional, esta experiencia me ha enseñado a ser constante y a adaptarme frente a los retos, así como a valorar la importancia de la paciencia y la flexibilidad en un trabajo de investigación. Además, he aprendido que, en muchos casos, no siempre las técnicas más avanzadas garantizan mejores resultados, y que es fundamental elegir y ajustar los métodos adecuados según el problema y los recursos disponibles.

También he reflexionado sobre la importancia de comunicar los hallazgos de forma clara y accesible, lo cual considero una competencia clave para mi desarrollo, especialmente si en el futuro desempeño un rol docente o de divulgación.

Quiero agradecer a todas las personas que me han apoyado durante este proceso, por su guía, ayuda y ánimo constante. Espero que este trabajo sirva como una pequeña contribución en el ámbito de la IA aplicada a la salud, y que ayude a inspirar futuras investigaciones que mejoren la calidad de vida de quienes padecen esta enfermedad.

Capítulo 7: Trabajo a futuro

Aunque los resultados obtenidos en este proyecto son prometedores, existen varias vías para continuar y profundizar en esta línea de investigación. En primer lugar, sería recomendable ampliar el conjunto de datos, tanto en número de muestras como en diversidad, para mejorar la robustez y generalización de los modelos desarrollados. Contar con datos multimodales, como combinaciones de señales de voz, escritura y movimientos, también podría aportar información complementaria valiosa para la detección precoz de la EP.

En cuanto a los modelos, aunque las técnicas tradicionales han mostrado un buen desempeño en este caso, explorar y optimizar enfoques avanzados de aprendizaje profundo o modelos híbridos podría ayudar a mejorar la precisión y sensibilidad, especialmente con conjuntos de datos más grandes y variados. Asimismo, implementar métodos de interpretabilidad ayudaría a entender mejor las decisiones de los modelos y facilitar su aplicación clínica.

Por otro lado, la integración de sistemas en entornos reales, como aplicaciones móviles o plataformas portátiles, permitiría validar los modelos en condiciones prácticas y fomentar su adopción en la detección temprana de la enfermedad.

Finalmente, continuar desarrollando habilidades de análisis, modelado y comunicación científica será clave para avanzar en futuros proyectos, especialmente aquellos con impacto en la salud y el bienestar de las personas.

Bibliografía

- [1] P. L. Rodríguez García, “Diagnóstico y tratamiento médico de la enfermedad de Parkinson”, Revista Cubana de Neurología y Neurocirugía, vol. 10 (1), 2020. Disponible en: <https://revneuro.sld.cu/index.php/neu/article/view/285>
- [2] Organización Mundial de la Salud, “Enfermedad de Parkinson”, abr. 2022. Disponible en: <https://www.who.int/es/news-room/factsheets/detail/parkinson-disease>
- [3] National Institute of Neurological Disorders and Stroke, “Parkinson’s Disease”, NINDS, U.S. Department of Health and Human Services. Disponible en: <https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease>
- [4] Hospital Clínic Barcelona, “¿Se puede detectar el Parkinson años antes de la aparición de los primeros síntomas?”, 3 de enero de 2024. Disponible en: <https://www.clinicbarcelona.org/noticias/se-puede-detectar-el-parkinson-anos-antes-de-la-aparicion-de-los-primeros-sintomas>
- [5] W. Wang, J. Lee, F. Harrou and Y. Sun, "Early Detection of Parkinson’s Disease Using Deep Learning and Machine Learning", vol. 8, pp. 147635-147646, 2020, doi: 10.1109/ACCESS.2020.3016062. Disponible en: <https://ieeexplore.ieee.org/abstract/document/9165732>
- [6] V. Kaul, S. Enslin, and S. A. Gross, “History of artificial intelligence in medicine”, vol. 388, no. 1, p. 111840, 2020. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0016510720344667>

- [7] A. Govindu and S. Palwe, “Early detection of Parkinson's disease using machine learning”, vol. 218, pp. 284–291, 2023. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1877050923000078>
- [8] S. Das and H. Ramteke, “A Comprehensive Review of the Role of Biomarkers in Early Diagnosis of Parkinson's Disease”, vol. 16, no. 2, p. e54337, Feb. 2024, doi: 10.7759/cureus.54337. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10945043/>
- [9] T. Li and W. Le, “Biomarkers for Parkinson's Disease: How Good Are They?”, *Neuroscience Bulletin*, vol. 36, no. 2, pp. 183–194, Feb. 2020, doi: 10.1007/s12264-019-00433-1. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6977795/>
- [10] B. Ndlovu, K. Maguraushe, and O. Mabikwa, “Machine Learning and Explainable AI for Parkinson's Disease Prediction: A Systematic Review”, *Indonesian Journal of Computer Science*, vol. 14, no. 2, Apr. 2025, doi: 10.33022/ijcs.v14i2.4837. Disponible en: <http://www.ijcs.net/ijcs/index.php/ijcs/article/view/4837/1016>
- [11] M. G. Di Cesare, D. Perpetuini, D. Cardone, and A. Merla, “Machine Learning-Assisted Speech Analysis for Early Detection of Parkinson's Disease: A Study on Speaker Diarization and Classification Techniques”, vol. 24, no. 5, p. 1499, Feb. 2024, doi: 10.3390/s24051499. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10934449/>
- [12] J. Mendes-Laureano, J. A. Gómez-García, A. Guerrero-López, et al., “NeuroVoz: a Castilian Spanish corpus of parkinsonian speech”, vol. 11, p. 1367, 2024, doi: 10.1038/s41597-024-04186-z. Disponible en: <https://www.nature.com/articles/s41597-024-04186-z>



Anexos

Anexo A: Código fuente y scripts del proyecto

El código completo desarrollado durante este Trabajo Fin de Máster está disponible en el repositorio público de GitHub: <https://github.com/Unai3105/TFM-Unai-Roa>

En dicho repositorio se incluyen los siguientes archivos principales, que permiten reproducir y analizar los procesos realizados:

- **audio_processing.py**: Script dedicado al preprocesamiento general de los audios, incluyendo limpieza, normalización y preparación de los datos para su posterior análisis.
- **audio_features_extracting.py**: Script encargado de la extracción de características acústicas específicas a partir de las grabaciones de audio. Incluye funciones para procesar las señales y generar los vectores de características utilizados en el modelado.
- **neurovoz_v3_metadatos.ipynb**: Notebook dedicado al análisis y exploración de los metadatos asociados a las grabaciones, como información demográfica y clínica, que sirven para complementar el estudio y mejorar la comprensión del conjunto de datos.
- **neurovoz_v3_audios.ipynb**: Notebook en formato Jupyter donde se lleva a cabo la exploración inicial, análisis y visualización de las características extraídas de los audios, así como algunas pruebas preliminares de modelado.
- **neurovoz_v3_audios.new.ipynb**: Versión actualizada del notebook anterior, con mejoras en el preprocesamiento y análisis de los datos de audio, incluyendo ajustes en las técnicas de extracción y preparación para el modelado avanzado.

Estos recursos permiten replicar los experimentos realizados y facilitan la extensión y mejora futura del proyecto.