



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería de la Salud



INGENIERÍA  
DE LA SALUD

## **TFG del Grado en Ingeniería de la Salud**

### **Análisis de la voz para la detección de la Enfermedad de Parkinson**

Presentado por Maider Murugarren  
Ilundain en Universidad de Burgos

4 de julio de 2023

Tutores: José Francisco Díez Pastor –  
Esther Cubo Delgado







UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería de la Salud



D. José Francisco Díez Pastor, profesor del departamento de Ingeniería Informática, Área de lenguajes y Sistemas informáticos.

Expone:

Que el alumno D. Maider Murugarren Ilundain, con DNI 73139258V, ha realizado el Trabajo final de Grado en Ingeniería de la Salud titulado Análisis de la voz para la detección de la Enfermedad de Parkinson.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 4 de julio de 2023

Vº. Bº. del Tutor:

Vº. Bº. del Tutor:

D. José Francisco Díez Pastor

D. Esther Cubo Delgado





## Resumen

Una detección temprana y precisa de cualquier enfermedad es fundamental para mejorar los resultados clínicos y la calidad de vida de los pacientes. En este caso en concreto, se trabajará con la enfermedad de Parkinson, un trastorno neurodegenerativo crónico que afecta principalmente al sistema motor y en consecuencia a la vida de miles de personas. Comúnmente conocida por uno de sus síntomas característico de un temblor en las manos de los pacientes que lo padecen. En sus comienzos los síntomas pueden ser sutiles y difíciles de distinguir de otros trastornos, lo que lleva a retrasos en el diagnóstico y la implementación de intervenciones adecuadas.

Por ello, para este proyecto de fin de grado se abordará el desafío de hacer uso de audios de voz y técnicas de aprendizaje automático. El objetivo principal es desarrollar un modelo de clasificación automática que pueda identificar la presencia de la enfermedad de Parkinson a partir de características extraídas de los audios de voz, proporcionando así una herramienta de detección efectiva y no invasiva.

Para ello, se ha comenzado con una investigación pertinente en cuanto a la enfermedad, los modelos de aprendizaje automático y las características de la voz. Relacionando todo ello de manera coherente. Recogiendo los audios que se usarán como fuente de información. Haciendo uso de la minería de datos para extraer la información sobre ellos y crear una base con la que entrenar distintos modelos de aprendizaje. Hacer una comparación y conclusión sobre los resultados obtenidos. Y finalmente, diseñar y crear un prototipo capaz de detectar la enfermedad de Parkinson haciendo uso de dicha investigación, con la mayor eficacia posible, partiendo de seis audios de voz del paciente.

## Descriptores

Enfermedad de Parkinson, Trastorno neurodegenerativo, Analisis de voz, Minería de datos, Aprendizaje automático, Clasificador de audios y Herramienta de detección.

## **Abstract**

Early and accurate detection of any disease is crucial to improve clinical outcomes and patients' quality of life. In this specific case, we will be working with Parkinson's disease, a chronic neurodegenerative disorder that primarily affects the motor system and directly impacts the lives of thousands of individuals. It is commonly known for one of its characteristic symptoms, hand tremors, exhibited by affected patients. The initial symptoms of the disease can be subtle and difficult to distinguish from other disorders, leading to delays in diagnosis and the implementation of appropriate interventions.

Therefore, this undergraduate thesis project aims to address the challenge by utilizing voice recordings and machine learning techniques. The main objective is to develop an automated classification model capable of identifying the presence of Parkinson's disease based on extracted features from voice recordings, providing an effective and non-invasive detection tool.

To achieve this, we have embarked on relevant research regarding the disease, machine learning models, and voice characteristics, establishing coherent connections between them. We have collected audio samples to serve as the database and employed data mining techniques to extract information from the recordings, creating a training dataset for various machine learning models. A comparison and conclusion of the obtained results will be made. Finally, we will design and create a prototype capable of detecting Parkinson's disease using the aforementioned research, striving for the highest possible efficacy, utilizing six patient voice recordings as the foundation.

## **Keywords**

Parkinson's Disease, Neurodegenerative Disorder, Voice Analysis, Data Mining, Machine Learning, Audio Classifier, Detection Tool.



---

# Índice general

---

Índice general	iii
Índice de figuras	iv
Índice de tablas	v
Objetivos	1
Conceptos teóricos básicos	3
2.1. Conceptos teóricos básicos . . . . .	3
2.2. Estado del arte y trabajos relacionados. . . . .	14
Metodología	19
3.1. Descripción de los datos. . . . .	19
3.2. Técnicas y herramientas. . . . .	20
Conclusiones	25
4.1. Resumen de resultados . . . . .	25
4.2. Discusion . . . . .	33
Aspectos relevantes del desarrollo del proyecto	41
Lineas de trabajo futuras	43
Bibliografía	45

---

# Índice de figuras

---

2.1. Diagrama Validación Cruzada . . . . .	11
4.1. Ejemplo conjunto de datos extraídos con Spicy . . . . .	26
4.2. Ejemplo conjunto de datos extraídos con Pytorch . . . . .	26
4.3. Comparación para el clasificador árbol de decisiones. . . . .	29
4.4. Comparación para el clasificador árbol de decisiones. . . . .	30
4.5. Comparación para el clasificador árbol de decisiones. . . . .	31
4.6. Comparación para el clasificador árbol de decisiones. . . . .	32
4.7. Comparación para el clasificador árbol de decisiones. . . . .	34
4.8. Comparación para el clasificador KNN. . . . .	35
4.9. Comparación para el clasificador Random Forest. . . . .	35
4.10. Resumen para la comparación de los modelos de clasificación. . .	37
4.11. Captura del prototipo. . . . .	40

---

# Índice de tablas

---

2.1. Matriz de Confusión . . . . .	12
------------------------------------	----



---

# Objetivos

---

Objetivos principales del trabajo realizado serán los siguientes:

1. Recopilación de una base de datos, en este caso audios, de pacientes con Parkinson y pacientes control, respetando los consentimiento y permisos necesarios para usar dicha información en esta investigación.
2. Identificar y hacer uso de las características acústicas en los pacientes con Parkinson, para poder analizarlas y compararlas con personas sin dicha patología. Buscando así un patrón capaz de identificar signos de dicha enfermedad.
3. Hacer de algoritmos de aprendizaje automático para detectar los signos de la enfermedad de Parkinson en la acústica de los pacientes.
4. Hacer una comparación de los diferentes resultados en base a la librería que se haya utilizado para extraer las características de los audios, la librería *Spicy* y la librería *torch* y en consecuencia para entrenar los modelos de aprendizaje automáticos.
5. Evaluar la efectividad de dichos resultados y hacer uso de los modelos de entrenamiento que mejores competencias obtengan.
6. Contribuir a la investigación de nuevas técnicas no invasivas para la detección de la enfermedad de Parkinson.
7. Desarrollar una aplicación web, para el que uso del proyecto resultante. De manera que sea de fácil uso y visual, para cualquier sanitario que quiera detectar signos del Parkinson en la voz de un paciente.



---

# Conceptos teóricos básicos

---

## 2.1. Conceptos teóricos básicos

En este apartado se abordará algunos conceptos teóricos que se han usado como base del proyecto y se mencionaran a lo largo de él. Se comienza mencionando los conceptos más relacionados con la Enfermedad de Parkinson, seguidos de las características del habla y fonación y se terminará con conceptos relacionados con la computación y los métodos que se usan a lo largo de todo el proceso para llegar al objetivo final.

### 2.1.1. Enfermedad neurodegenerativa

Las denominadas enfermedades neurodegenerativas con enfermedades crónicas con tendencia a generar una discapacidad, pudiendo ser tanto física, intelectual y social. Produciendo en consecuencia una dependencia de la persona afectada: su ámbito laboral, económico y social. [56]

Son enfermedades que producen un déficit en el sistema nervioso y en especial en las neuronas, provocando un deterioro progresivo de las funciones cognitivas, motoras y autónomas. Caracterizadas por la degeneración y pérdida de las neuronas. Siendo las enfermedades más conocidas de este grupo: el Alzheimer, la enfermedad de Huntington, la esclerosis lateral amiotrófica o también conocida como la ELA y la enfermedad de Parkinson, cuya patología es el centro de este proyecto. [56]

### 2.1.2. Enfermedad de Parkinson

Se trata de una las enfermedades neurodegenerativas más frecuentes que afecta directamente al movimiento de los seres humanos [59]. Pertenecer al

grupo de trastornos del movimiento, siendo una afección crónica y progresiva [45].

La enfermedad de Parkinson es de causa desconocida, pero se cree que una de las causas principales de la degeneración de las neuronas es por la acumulación anómala de la proteína alfa-sinucleína en las neuronas. De tal manera que acaba desencadenando una serie de procesos celulares que implican la muerte de células nerviosas en el cerebro, como las más afectadas en este caso, las neuronas [55].

Afecta al 1-2% de la población mayor de 60 años [44]. Evoluciona progresivamente con el deterioro temporal de las neuronas, en específico de las dopaminérgicas. Ya que es causada por un defecto en las neuronas que hace que no produzcan la suficiente cantidad de dopamina requerida por nuestro organismo, implicando así una enfermedad neurodegenerativa crónica en los afectados.

Sus manifestaciones físicas más destacables son, el temblor en reposo, la rigidez y afectación en el movimiento, postura o incluso en el andar. Además, se pueden dar incluso en la aparición temprana de la enfermedad problemas de motores <sup>1</sup> o alucinaciones causadas por la medicación. Siendo los síntomas principales el temblor y la bradicinesia <sup>2</sup>, causadas por la alteración de los circuitos motores que conectan los ganglios de la base con la corteza motora [45].

Actualmente, no existe un tratamiento que detenga por completo la progresión de la enfermedad, pero si en cambio tratamientos para ralentizar su progresión y mejorar los síntomas. Como lo son, por ejemplo, la reposición de la dopamina por medio del uso de su precursor, administración de sustancia dopaminérgicas para estimular los receptores o la inhibición de las enzimas que destruyen la dopamina como la catecol- O- metiltransferasa (COMT) con la entacapona, y a la monoamino oxidasa tipo B (MAOB) con la selegilina y la rasagilina [45].

### **2.1.3. Escalas de Valoración: La MDS-UPDRS**

Son las siglas de la escala unificada para la evaluación de la Enfermedad de Parkinson en su nueva versión. Esta nueva versión mantiene el formato anterior de la UPDRS, abordando además las partes que suponían mayor

---

<sup>1</sup>Posibles padecimientos neurológicos complejos que provocan movimientos involuntarios o anormales, o afectan la capacidad de controlar el movimiento.

<sup>2</sup>Lentificación de los movimientos, especialmente de los movimientos voluntarios complejos.



ambigüedad y debilidades. La escala es usada para seguir el avance de la Enfermedad de Parkinson y está compuesta cuatro partes distintas. La primera evalúa las experiencias no motoras de la vida diaria y consta de dos partes, la parte A, compuesta de un determinado número de conducta evaluada por el investigador con la información proporcionada por los pacientes y los cuidadores, y la parte B, es cumplimentada por el paciente con o sin ayuda del cuidador, pero independientemente del investigador. La segunda, las experiencias no motoras de la vida diaria. La tercera, consta de una exploración motora. Y por último, la cuarta, las complicaciones motoras [46].

Además, es de importante destacar que la tercera parte de la MDS-UPDRS también incluye una clasificación por estadios denominada escala de Hoehn y Yahr. Esta escala es de fácil aplicación y relacionada con el deterioro motor y calidad de la vida del paciente. Según la misma existen los siguientes posibles estadios [9]:

- 0.0 – No hay signos de enfermedad.
- 1.0 – Enfermedad exclusivamente unilateral.
- 1.5 – Afectación unilateral y axial.
- 2.0 – Afectación bilateral sin alteración del equilibrio.
- 2.5 – Afectación bilateral leve con recuperación en la prueba de retro-pulsión. (Test del empujón)
- 3.0 – Afectación bilateral leve a moderada; cierta inestabilidad postural, pero físicamente independiente.
- 4.0 – Incapacidad grave; aún capaz de caminar o de permanecer en pie sin ayuda.
- 5.0 – Permanece en una silla de ruedas o encamado si no tiene ayuda.

Cabe destacar que será de relevnate entender y diferenciar los diferentes posibles estadios en esta escala. Ya que las muestras de los audios que se utilizarán para este proyecto tienen como atributo los estadios de cada paciente, que se usarán como referencia para determinar lo avanzada que está la enfermedad de Parkinson en cada caso y poder de esta manera, identificar el nivel de progresión de la enfermedad del resto.

#### 2.1.4. Diagnóstico

Se trata del proceso por el cual se identifica una enfermedad basándonos en los signos y síntomas presentes [6]. En el caso de la Enfermedad de Parkinson, es difícil tener un diagnóstico concluyente en los primeros estadios de la enfermedad. Ya que actualmente, no existe ningún marcador o examen de diagnóstico para esta enfermedad. Pero se hace uso de diferentes signos, así como el temblor en las manos, brazos, piernas o rigidez e las extremidades, acompañada de una lentitud en los movimientos que irá en aumento con el progreso de la enfermedad, además de problemas de equilibrio y coordinación [59][57]. Cabe mencionar otros factores que determinan el diagnóstico como por ejemplo la respuesta a la levodopa, que es un requisito necesario o por ejemplo la detección de síntomas atípicos.

En cuanto a los métodos para obtener el diagnóstico de esta enfermedad, el más común es el diagnóstico por imágenes, mediante tomografías computarizadas o resonancias magnéticas. También cabe mencionar que la mayoría de los diagnósticos, se consideran idiopáticos<sup>3</sup> y sin causa genética conocida. [59].

#### 2.1.5. Biomarcadores

Son moléculas o signos biológicos cuya presencia indican la normalidad de un proceso o por lo contrario la anormalidad causada por una afectación o enfermedad [5]. Para el caso que nos atañe, se tratará como biomarcador la huella cinética acústica, por disfunciones articulatorias, basadas en técnicas de comparación de audios entre un grupo de casos y otro de control.

Se ha encontrado una relación aparente entre los pacientes que padecen la enfermedad de Parkinson y sus dificultades en el habla, debido a la pérdida de coordinación en los músculos que controlan la respiración, la fonación, la articulación y la prosodia. Presentando en consecuencia síntomas en la voz como debilidad en el volumen, monotonía en el tono, cambios de calidad del sonido, velocidad del habla y dificultades de entender el propio mensaje [49].

#### 2.1.6. Las señales de voz y las características físicas

Las señales de voz son producidas por la vibración de las cuerdas vocales por el aire que es expulsado de los pulmones. Para ello, se coordinan distintos movimientos de músculos y la articulación mandibular, lengua, garganta y labios [47].

---

<sup>3</sup>Enfermedad de causa desconocida o mecanismo de aparente origen espontáneo.

Para comenzar se identifican algunas de las características del habla, que se pueden dividir en cinco niveles distintos [52]:

- **Fonatorio:** característica del habla que determinar el nivel de trabajo de la musculatura para emitir sonidos.
- **Resonante:** determina la potencia del tono de la voz vibrando las cuerdas vocales.
- **Articulatorio:** se trata del análisis del habla en relaciones a los fonemas, palabras u oraciones expresadas.
- **Prosódico:** análisis del habla en relacion con el discurso del lenguaje.
- **Lingüístico:** son las alteraciones en el habla continua, manifestándose como repeticiones o sustituciones.

Por otro lado, se definen estas misma características pero expresadas de manera numérica, ya que están relacionadas más con el entorno matemático y en este caso serán las que se tengan en cuenta al analizar los audios de los pacientes computacionalmente [58].

- **Jitter:** medida de variabilidad de los periodos entre los ciclos de vibración de las cuerdas vocales durante la producción de voz. Siendo asi una variable capaz de evaluar la variabilidad, estabilidad y regularidad de la voz. [61].
- **Shimmer:** medida de variabilidad de la amplitud de la señal de audio durante los ciclos de vibración de las cuerdas vocales. Utilizada para evaluar la estabilidad y regularidad de la calidad de la voz.[26].
- **APQ:** proveniente del nombre en inglés Amplitude Perturbation Quotient. Es una medida de variabilidad de la amplitud de la voz, utilizada para medir la perturbación en la voz. [54]
- **PPQ:** proveniente del nombre en inglés Pitch Perturbation Quotient. En la medida de variabilidad en la frecuencia fundamental de la señal la voz.[54]
- **MFCC:** estas son las siglas de Coeficientes Cepstrales en las Frecuencias de Mel en inglés. Es una característica ampliamente utilizada en audios que captura características relacionadas con la forma del espectro de la frecuencia del audio. [15]

- **1º y 2º Derivada de la frecuencia fundamental:** la frecuencia fundamental se puede definir como la frecuencia más baja y dominante en la señal de audio. En este caso, la 1º derivada, representará el cambio de la frecuencia fundamental a lo largo del tiempo. Y la 2º derivada, representa la tasa de cambio de la frecuencia fundamental. [11]
- **Energía logarítmica:** es la cantidad de energía que hay en la señal de audio y es extraída haciendo el cálculo del logaritmo de la suma de los cuadrados de todas las amplitudes de la señal. Refleja la intensidad de la señal del audio. [43]
- **Amplitud:** es la diferencia entre los valores extremos positivos y negativos de una señal. Y en este caso, representa el volumen percibido de la señal del audio. [1]
- **Amplitud a decibelios:** se trata de una transformación que convierte la representación de la amplitud al uso de una señal en una escala logarítmica de decibelios. De manera, que se puede usar para resaltar las diferencias cuando los niveles de sonido son más pequeños de lo habitual. [42]
- **Duración:** intervalo temporal durante el cual la pista de audio persiste sin discontinuidad. [7]
- **Espectrograma:** es la representación visual de la variación de la energía de una señal de audio en función del tiempo y la frecuencia en cada momento. Siendo así la muestra de la intensidad del sonido a lo largo del tiempo del audio. [37]
- **Densidad espectral de potencia (PSD):** es una medida de distribución de la energía en las señales de audio. [8]
- **Loudness:** se trata de una transformación encargada de calcular el nivel de volumen conceptual de una señal de audio en función a la sensibilidad del oído humano. [39]
- **Centroide espectral:** medida indicadora de la gravedad espectral de una señal de audio. Utilizado para evaluar la tonalidad de la señal del audio. [3]

### 2.1.7. Minería de datos y aprendizaje automático

La minería de datos se puede definir como la aplicación de técnicas de la inteligencia artificial sobre grandes cantidades de datos, con el objetivo de descubrir tendencias, patrones, o relaciones ocultas. Dicho en otras palabras, encontrar patrones en los que los datos proporcionan algún tipo de entendimiento o permitan tomar decisiones rápidas y precisas [51].

Para ello hace uso del aprendizaje automático o *machine learning* como técnica. El aprendizaje automático se puede dividir en el aprendizaje supervisado, en los cuales los datos tienen un atributo especial y puede ser un aprendizaje supervisado de tipo clasificación o de tipo regresión. Y por otro lado, se tiene el aprendizaje no supervisado, en el cual, no hay atributo especial como tal y su objetivo es encontrar grupos de ejemplo con características similares para poder proyectarlos, visualizarlos y obtener conclusiones [58].

### 2.1.8. Preprocesamiento de datos

Los datos no pueden ser utilizados de manera directa y que tienen que estar tratados, necesitan una entrada factible para el algoritmo y que tengan las características necesarias para poder obtener la mayor cantidad de información a partir de ellas. Para ello se realizan varios pasos, eliminar el silencio al principio y al final de los audios, excluir el ruido que pueda afectar a la calidad de los audio y por último, normalizar el volumen en todos ellos. Este apartado se explicará con más profundidad a lo largo de la metodología.

### 2.1.9. Extracción de características

La extracción de características es un proceso fundamental para analizar la información y crear un nuevo conjunto de datos con el que poder usar el aprendizaje automático. Tiene como objetivo principal identificar y seleccionar las características de mayor relevancia en cada tipo y cuantificarlos, así mismo explotar al máximo la información que se pueda extraer de la base de datos. En este caso específico se usan los audios y se hará uso de la extracción de características para poder identificar y clasificar las grabaciones en base a distintas variables. Creando así una base de datos, que se usará para entrenar diferentes modelos de aprendizaje automático. [10]

### 2.1.10. Algoritmos de clasificación

Los algoritmos de clasificación son métodos utilizados para etiquetar o categorizar un conjunto de información basándose en las características de cada uno. Partiendo de datos con diferentes atributos o variables, son capaces de aprender de manera automática a partir dichos conjuntos y clasificarlos en diferentes grupos basándose en las características de cada uno. [13]

Estos son algunos de los algoritmos de clasificación que se utilizan en este proyecto:

- **Algoritmo de árbol de decisiones:** se trata de un algoritmo de clasificación que toma decisión de manera recursiva en base al conjunto de reglas y condiciones que se le imponga. Su nombre proviene porque actúa en base a una estructura similar a la de un árbol. Se compone de diversos nodos y en base a la respuesta dada en la pregunta de dicho nodo se desarrollará una rama u otra, que compondrán las diferentes repuestas. Finalmente, llegando a un nodo denominada hoja que contendrá la etiqueta de clasificación final. De esta manera se irán formando y tomando las decisiones hasta clasificar todos los datos. [23]
- **Algoritmo de K vecinos más cercanos (KNN):** es un algoritmo de clasificación basado en la proximidad de los datos según las características de cada uno. Esto es, primero se almacenan todos los datos de manera multidimensional en el espacio y se irán clasificando en base a su cercanía. Y así con todos los datos reordenándose cada vez a la clase más común entre sus vecinos. De esta manera, finalmente irán agrupando y clasificando. Para determinar las etiquetas se utiliza el voto mayoritario, esto es el valor más representativo o el más frecuente entre los datos que componen el grupo. [38]
- **Algoritmo de Random Forest:** : este algoritmo de clasificación combina varios árboles de decisión para hacer las predicciones. Para ello, cada árbol se entrena con un conjunto de muestras aleatorias con un conjunto muestrar de entrenamiento determinados. Posteriormente, para realizar la predicción cada árbol hace su propia predicción mediante la votación o clasificación para determinar el árbol final. Dado que se hace uso de la construcción de diferentes árboles y la combinación de resultados, se reduce el riesgo de sobreajuste mejorando así la precisión y la capacidad de generar un modelo óptimo. [2]

### 2.1.11. Validación cruzada

Se trata de una técnica para concretar y evaluar la capacidad de un modelo de aprendizaje. Cada conjunto de pruebas divide en múltiples subconjuntos llamados “folds”. Con ellos entrenará y evaluará la respuesta del rendimiento del modelo. Principalmente utilizado para detectar el sobre ajuste, esto es los modelos que no logren identificar un patrón eficaz en el caso concreto[36].

La siguiente imagen, muestra de manera visual como se crearan los folds para entrenar y evaluar el aprendizaje.



Figura 2.1: Diagrama Validación Cruzada

Fuente: [36]

### 2.1.12. Métricas de evaluación

Se pueden definir las métricas de evaluación como medidas de cuantificación para determinar y evaluar la calidad de los modelos de aprendizaje en este caso. Son utilizadas para poder evaluar de forma objetiva a través de la información que proporcionan el desempeño, siendo varias métricas las posibles a evaluar. Las diferentes métricas se extraen con los cálculos pertinentes entre los valores de los resultados de la matriz de confusión.

En este caso concreto, estas son las métricas que se han tenido en cuenta para evaluar los modelos: [16]

- **Matriz de confusión:** se trata d una tabla que muestra la combinación de los valores de los casos correctamente clasificados y los que no. Por ello, se compone de cuatro celdas representadas con los valores de las siguientes métricas.
  - **Verdaderos positivos (TP):** el número de instancias positivas clasificadas correctamente.

- **Falsos Positivos (FP)**: el número de instancias positivas incorrectamente clasificadas.
- **Verdaderos negativos (TN)**: el número de instancias negativas correctamente clasificadas.
- **Falsos negativos (FN)**: el número de instancias negativas incorrectamente clasificadas.

Una matriz de confusión tiene la siguiente forma:

Tabla 2.1: Matriz de Confusión

	Clase Real	
	Positiva	Negativa
	VP	FN
Clase Predicha	FP	VN

Partiendo de estas mismas métricas se podrá calcular otras de mayor interés como la precisión, la sensibilidad, especificidad o puntaje F1. Siendo todas ellas de gran utilidad a la hora de evaluar el rendimiento de un modelo de clasificación entrenado. Las cuales se explican a continuación. [14]

- **Exactitud (accuracy)**: métrica para evaluar el rendimiento general del modelo, midiendo la predicción global del modelo para determinar correctamente todas las clases. [14]

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- **Precisión**: métrica encargada de medir la proporción entre los casos clasificados correctamente como positivos sobre el total de ejemplos clasificados como positivos, teniendo en cuenta los falsos positivos. [14]

$$Precision = \frac{TP}{TP + FP}$$

- **Sensibilidad**: métrica también llamada Recall, se usa para medir la proporción de resultados positivos sobre los positivos reales, teniendo en cuenta los falsos negativos. [14]

$$Recall = \frac{TP}{TP + FN}$$



- **Especificidad:** métrica para evaluar la proporciones de negativos reales sobre el total de resultados negativos [14].

$$Specificity = \frac{TN}{FP + TN}$$

- **F-1 score:** esta métrica combina el valor de la precisión y de la sensibilidad de los modelos, proporcionando una medida general del rendimiento del modelo [14].

Partiendo de la base que el programa resultante es una herramienta muy específica para la detección de la enfermedad de Parkinson, en este caso den concreto, objetivo principal que es tener un modelo de clasificación entrenado que sea capaz de detectar casos y clasificarlos en positivos o negativos. Posteriormente la decisión de dirige a la elección del modelo de clasificación con mejor rendimiento, en los valores de la sensibilidad, siendo esta de vital importancia a la hora de maximizar la detección de enfermos y por lo contrario minimizar los falsos negativos, esto es, minimizar la probabilidad de clasificar erróneamente a un paciente sin la enfermedad como positivo. En el valor de la precisión, siendo relevante también para minimizar los falsos positivos. Y finalmente en el valor de F-1 score, siendo indicador de la evaluación general del rendimiento a la hora de hacer una correcta clasificación.

### 2.1.13. Conjunto de datos

Un conjunto de datos es una colección de datos en relación con el tema tratado en cada caso. Como en este caso se trabaja en el ámbito del análisis de los datos mediante el aprendizaje automático, se trata de una colección estructurada de datos, utilizada para entrenar, evaluar y validar modelos y algoritmos. Estos información contiene tanto de variables numéricas como textos. Suelen estar ordenados en filas y columnas, representando así en las filas los datos específicos de cada audio y en las columnas cada propiedad o característica de cada caso [41].

### 2.1.14. Interfaz

Se puede definir el concepto de interfaz como el punto de conexión entre la interacción del usuario de forma física y el programa o dispositivo dependiendo del caso. Facilitando así la comunicación entre el usuario y el ordenador. A nivel computacional, se compone del conjunto de funciones,

métodos y protocolos, facilitando la colaboración entre sí. En este caso concreto, la interfaz será la encargada de permitir que el usuario pueda interactuar con los modelos y las herramientas de análisis utilizadas de manera intuitiva y visual. Esta interfaz puede ser diseñada tanto en formato página web como de aplicación móvil [28].

## 2.2. Estado del arte y trabajos relacionados.

Tal y como hemos podido ver en el apartado de los conceptos teóricos la enfermedad de Parkinson afecta al sistema motor de diversas formas. Una de las claras afectaciones se puede ver reflejada en el habla. Dando lugar a alteraciones caracterizadas por un discurso monótono, bajo tono e intensidad, pausas o ritmo del habla alteradas [47]. En por eso que el análisis aústico de la voz puede ofrecer mucha información y actuar como un claro biomarcador para el diagnóstico temprana de la patología. Además de poder indicar tras su detección los resultados de los tratamiento farmacológicos o quirúrgicos del paciente [47].

Estas son algunas de las alteraciones en la voz causadas por la enfermedad de Parkinson:

- **Alteraciones en la fonación:** implicando una dificultad para producir abundantes pliegues vocales, con falta de flexibilidad y control de movimientos laríngeos. Siendo en consecuencia una voz débil y monótona, sin expresión [50].
- **Alteraciones en los articuladores:** implicando poca precisión articulatoria y en consecuencia un habla difusa e imprecisa [50].
- **Alteraciones del ritmo:** implicando una aceleración anormal y progresiva en el habla, amontonando y haciendo pausas imprecisas [50].

Para corroborar lo mencionado anteriormente y la viabilidad e importancia de poder desarrollar un sistema de análisis de la voz para detectar el Parkinson, se hará una investigación de los trabajos publicados hasta ahora y se comentarán los de mayor interés para este caso.

Dado que la investigación es esencial para poder contar con los avances más recientes y relevantes, para este estado del arte se proporcionará una visión general del tema tratado en este caso, basándonos en tres artículos de

interés. Siendo el objetivo principal de este apartado presentar la revisión completa de artículos de interés, por su relevancia y contribución al cambio de la investigación con el uso de audios en la detección de enfermedades neurodegenerativas.

Para este análisis, se hará un resumen del contenido de cada uno de los artículos de interés, abordando los aspectos más relevantes para nuestro caso. Permitiendo contextualizar y sentar las bases sólidas para nuestro propio desarrollo del proyecto.

El artículo titulado “Automatic detection of Parkinson’s disease in running speech spoken in three different languages” de Hesam Sagha y col. [53], se centra en la detección automática de la enfermedad de Parkinson con el análisis del habla en tres idiomas, en español, en alemán y en inglés. Tiene como objetivo principal el estudio de las señales del habla continuada, mediante la segmentación automática de locuciones sonoras.

El estudio involucra a 101 pacientes en total, 61 diagnosticados con la enfermedad de Parkinson y otros 40 como control. Con el procedimiento general de recolección de los audios a analizar, el uso de técnicas de procesamiento de señales y aprendizaje automático se intentó determinar si el cambio de la voz podría ser una herramienta útil para la detección de la enfermedad de Parkinson.

Para ello, se usó el modelado basado en supervectores MFCC-GMM, el análisis de la prosodia, el contenido de ruido, medidas de fonantes y el marco sonoro. Finalmente se usó la clasificación y la validación para determinar si era posible diferenciar los audios de los pacientes respecto a los de los controles.

Como resultado, demostraron que, si era posible diferenciarlos con éxito, gracias a las características específicas de la voz. Dando pie con dicha conclusión a que el análisis del habla es una herramienta de interés para la detección efectiva de manera automática y no invasiva de la enfermedad de Parkinson en diferentes idiomas. Reafirmando de esta manera la hipótesis planteada al principio de este apartado, de que el análisis acústico de la voz puede ofrecer mucha información y actuar como un claro biomarcador para el diagnóstico temprano de dicha patología.

Por otro lado, se tiene el artículo titulado “Parkinson’s Disease Detection from Speech Using Convolutional Neural Networks” de Evaldas Vaiciukynas, Adas Gelzinis, Antanas Verikas y Marija Bacauskiene [48]. Tiene como objetivo principal entrenar un modelo de CNN para clasificar las muestras

según si son procedentes de personas diagnosticadas con la enfermedad de Parkinson o no.

Partiendo de una base de datos de audios usan un modelo de CNN diseñado para procesar características espectrales del habla y aprender a identificar los posibles patrones que se dan con la enfermedad de Parkinson. Dicho modelo se ajustó de forma iterativa para minimizar la pérdida de información y a la vez maximizar la precisión de dicho aprendizaje.

Tras entrenar el modelo se procedió a la evaluación de este utilizando un conjunto de pruebas por separando y calculando las métricas pertinentes como la precisión, la sensibilidad o la especificidad del modelo. Determinando que la precisión de dicho modelo fue de un 98,63

De esta manera se llegó a las siguientes conclusiones, las redes neuronales con una técnica de aprendizaje automático prometedoras para la investigación en la detección de la enfermedad de Parkinson, el uso de las características espectrales puede ser muy efectivas para identificar patrones en el habla, y por último, el modelo de CNN utilizado resulto de gran precisión en la clasificación de muestras de voz entre paciente con enfermedad de Parkinson y pacientes control.

En este artículo se ve que el enfoque de esta investigación y el uso de los modelos de aprendizaje automático es acertado y prometedor en la enfermedad de Parkinson. Y por ello, se usará como base y ejemplo para esta investigación.

Por último, se tiene el artículo titulado “Characterization of speech impairment in Huntington’s disease using language-agnostic acoustic features” de Vitória S Fahed, Emer P Doheny, Carla Collazo, Joanna Krzysztofik, Elliot Mann, Philippa Jones, Laura Mills, Cheney Drew, Anne Rosser, Rebecca Cousins, Grzegorz Witkowski, Esther Cubo, Monica Busse y Madeleine M Lowery.[60]

En este caso, la investigación se basa en La enfermedad de Huntington, una enfermedad genética y neurodegenerativa que afecta el sistema nervioso central. Caracterizada por la degeneración progresiva de las células nerviosas en ciertas áreas del cerebro, particularmente en los ganglios basales y es causada por una mutación genética en el gen HTT (huntingtina).

En este estudio en concreto se examina la caracterización del trastorno del habla en la enfermedad de Huntington, utilizando características acústicas independientes al lenguaje hablado.

Para dicho estudio se recopilan muestras de grabaciones en diferentes etapas del desarrollo de la enfermedad de Huntington, además de las grabaciones de los controles. Se analizaron características como la frecuencia fundamental, la intensidad y la duración de los segmentos, siendo siempre características seleccionadas independientes al lenguaje. De tal manera, que son resultados pudieran ser aplicables en diferentes idiomas después.

Como resultados obtuvieron diferencias significativas en la acústica del habla entre el grupo diagnosticado con la enfermedad de Huntington y el grupo control. Pudiendo identificar diferencias considerables en la mayoría de las características analizadas.

Para todo ello, desarrollaron un modelo de clasificación basado en aprendizaje automático para poder distinguir los grupos. Logrando una alta precisión en la clasificación de los grupos. Indicando de manera efectiva que las características acústicas son indicadores efectivos del trastorno del habla en las enfermedades neurodegenerativas.

En conclusión, el estudio proporciona una caracterización detallada de los trastornos del habla en la EH utilizando características acústicas independientes del lenguaje hablado, demostrando ser indicadores relevantes de los trastornos del habla en la EH y pueden ser utilizadas para el desarrollo de herramientas de detección y seguimiento de la enfermedad. Siendo este un enfoque potencial para ser aplicado en otras enfermedades neurodegenerativas y pudiendo ser aplicado en diferentes idiomas.

Volviendo a reafirmar con este último artículo analizado que el enfoque de la investigación es correcto y que es un campo con especial potencial para la detección no invasiva y eficiente de las distintas enfermedades neurodegenerativas. Y en este proyecto en concreto se centra en la enfermedad de Parkinson.



---

# Metodología

---

## 3.1. Descripción de los datos.

Los datos base para este proyecto han sido unas pistas de audio, grabadas con diversos móviles en formato “.wav”. En total se disponen de 390 pistas de audios, 198 audios provenientes de pacientes control, esto es, de pacientes que no padecen de la enfermedad de Parkinson, y 192 de pacientes diagnosticados con enfermedad de Parkinson en diferentes estadios.

Los audios pueden ser agrupados por su contenido. Y en este caso se pueden diferenciar seis tipos de audios, todos ellos nombrados en base al tipo de contenido de cada uno.

- Los nombrados que terminan en 1, contienen la vocal “A” mantenida de forma constante.
- Los que terminan en 2, contienen la vocal “I” mantenida de forma constante.
- Los que terminan en 3, contienen la vocal “U” mantenida de forma constante.
- Los que terminan en 4, contienen la palabra “campana”.
- Los que terminan en 5, contienen la palabra “gato”.
- Los que terminan en 6, contienen la palabra “petaca”.

A la hora de la recogida de los audios se han nombrado siguiendo un patrón específico, para poder diferenciar cada audio a que paciente pertenecían y el contenido de los audios.

El formato ha sido el siguiente: IDXX\_D-M-A\_Y.wav

- XX: número identificador del paciente.
- D-M-A: fecha de la grabación.
- Y: Identificador del contenido del audio (Número variable entre el 1 y el 6).

Además de los audios también se recopiló información relevante para el desarrollo del proyecto, así como la edad del paciente, el sexo del paciente y el estadio del desarrollo de la enfermedad de Parkinson en la escala *MDS – UPDRS*. En el caso de los pacientes control, este último fue rellenado con el valor de 0, indicativo de no haber signos de la enfermedad en la escala *MDS – UPDRS*.

Por otro lado, tras haber extraídos las características fonéticas de los audios, partiendo de dichos 390 audios, se ha obtenido diferentes conjuntos de datos. Cada conjunto está compuesto por características extraídas a través de un algoritmo que usa una librería distinta y que contiene el nombre de cada archivo de audio seguido de 10 columnas con los valores de diferentes características extraídas.

Finalmente se ha hecho un Join <sup>4</sup> con la información de los pacientes y con las características obtenidas en cada caso, logrando de esta manera un conjunto de datos con mucha información para poder utilizarlo al entrenar los modelos de clasificación automática.

El formato y la estructura de este nuevo conjunto de datos creados serán explicados en el anexo D.1 Descripción formal de los datos.

## 3.2. Técnicas y herramientas.

En este apartado se habla de las diferentes técnicas y herramientas que se han utilizado a lo largo del desarrollo del proyecto. Para ello se habla de las librerías y modelos utilizados agrupándolos por objetivos y secciones en las que hayan sido utilizadas en cada caso.

Se comienza con el único framework utilizado en el desarrollo de este proyecto:

---

<sup>4</sup>Función para combinar varibales de dos o más conjuntos basándose en un campo común entre ellas.



- **PyTorch**: es un conjunto de herramienta con unas funciones pre-determinadas para hacer más fácil el desarrollo de programas. En este caso es de código abierto y está desarrollado por Facebook's AI Research Lab, con el objetivo principal de facilitar la implementación de algoritmos. Y se basa en estructuras de datos similares a matrices multidimensionales, siendo su característica principal el enfoque llamado modo imperativo, para escribir código de manera dinámica y flexible [22].

Se sigue nombrando de las bibliotecas utilizadas para implementar los códigos de aprendizaje automático:

- **“os”**: la biblioteca “os” tiene funciones para interactuar con el sistema operativo, como en este caso para hacer uso de rutas de archivo o creación de archivos [18].
- **“numpy”**: es una biblioteca utilizada para la manipulación científica computacional en Python. Proporcionando funciones matemáticas, operadores algebraicos o trabajar con estructuras multidimensionales [17].
- **“pandas”**: biblioteca comúnmente usada para el análisis y manipulación de datos, mediante estructuras eficientes y herramientas para la ordenación de estos [19].
- **“scikit-learn”**: biblioteca de aprendizaje automáticos, que dispone de diversos algoritmos y herramientas para trabajar con la minería de datos y el análisis de datos [24].

Además, también hemos trabajado diferentes clases de esta misma librería:

- **“sklearn.tree.DecisionTreeClassifier”**: clase que representa el clasificador de árbol de decisiones [32].
- **“sklearn.ensemble.RandomForestClassifier”**: clase que representa el clasificador de random forest [29].
- **“sklearn.neighbors.KNeighborsClassifier”**: clase que representa el clasificador de K-vecinos más cercanos [31].
- **“sklearn.model\_selection.train\_test\_split”**: función utilizada para dividir el conjunto de datos en grupos de entrenamiento y prueba [30].

Por otro lado, se explican las herramienta y librerías utilizadas tanto en la fase de preprocesado de datos como en el procesamiento de señales de los audios:

- **“pydub.AudioSegment”**: es una biblioteca de pyhub con funciones como para cargar, manipular y exportar audios en diversos formatos [21].
- **“shutil”**: la biblioteca proporciona funciones de manipulación de archivos y directorios en Python [27].
- **“csv”**: biblioteca con funciones como para leer y escribir de manera sencilla en archivos con formato *“.csv”* [4].

También se tiene un grupo de librerías que se han utilizado en este caso como herramientas de extracción de características de los audios de voz:

- **“librosa”**: es una biblioteca de Python utiliza para el análisis de señales de audio, proporcionando una amplia variedad de funcionalidades de carga, procesamiento, analizadores y visualizadores [12].
- **“spicy”**: biblioteca que proporciona funcionalidades avanzadas del área de las matemáticas, ciencia e ingeniería y en este caso utilizada para el análisis de audios de voz mediante diferentes subbibliotecas [25]:
  - **“scipy.io.wavfile”**: biblioteca que permite leer y escribir archivos de audio en formato WAV.
  - **“scipy.stats”**: proporciona funciones de estadística y distribución de probabilidades como herramientas de cálculos de medidas estadísticas.
  - **“scipy.signal”**: proporciona funciones para el procesamiento de señales, como por ejemplo filtros digitales, transformada de Fourier, análisis espectrales...
- **“torchaudio”**: biblioteca de Pytorch con funcionalidades para el procesamiento de señales de audio [34].

Además, se plantean también diversas bibliotecas utilizadas en el software del análisis estadístico, que en este caso han sido las utilizadas para evaluar el rendimiento de los modelos entrenados:

- “**matplotlib**”: es una biblioteca para visualizar los datos con herramientas como gráficos, diagramas, histogramas, gráficos de barras... facilitando así la interpretación de estos [35].
  - **sklearn.metrics.accuracy\_score**: función utilizada para calcular la precisión de un clasificador.
  - **sklearn.metrics.recall\_score**: función utilizada para calcular la recuperación de un clasificador.
  - **sklearn.metrics.precision\_score**: función utilizada para calcular la precisión de un clasificador.
  - **sklearn.metrics.f1\_score**: función utilizada para calcular la medida que combina precisión y recuperación de un clasificador.

Se usará la biblioteca de **Streamlit**, utilizada en este caso para realizar la interfaz y la demo del programa final. Se trata de una herramienta de código abierto que permite crear aplicaciones web con Python. Siendo una manera útil de crear prototipos de manera sencilla e intuitiva. Ya que permite una fácil integración con otras bibliotecas y código realizado mediante Python [33].

Por último, se expone el entorno de lenguaje utilizado en este proyecto y el lenguaje de programación utilizado.

**Jupyter Notebook** será el entorno de programación utilizado para el desarrollo de este proyecto. Se trata de una aplicación web de código abierto y en este caso se usará de manera local. Hace uso de archivos notebook para crear y compartir el código creado en ellos en diferentes lenguajes de programación. Es un entorno intuitivo y fácil para desarrollar y visualizar el código a tiempo real.

Se basa en archivos notebooks compuestos por diferentes celdas, donde puede contener tanto código ejecutable como texto escrito en formato *markdown*. Permitiendo así a los programadores mezclar tanto la documentación escrita con el código como los resultados de salida de dichas celdas a tiempo real [20].

En cambio, **Python** será el lenguaje con el que se desarrollará toda la implementación del código. Se trata de un lenguaje de programación interpretado, que fue creado el 1990 por Guido van Rossum. Se trata de uno de los lenguajes de programación más popular de la actualidad, destacando sobre el resto su sintaxis sencilla y intuitiva, además de las amplias áreas de aplicación, como son, por ejemplo, las ciencias de datos, inteligencia artificial o desarrollo web [40].



---

# Conclusiones

---

## 4.1. Resumen de resultados

En este apartado se dará un resumen de los resultados obtenidos a lo largo del desarrollo de todo el proyecto, así como las métricas de evaluación obtenidas de los entrenamientos de diferentes modelos de manera objetiva. Tiene como objetivo general informar de los resultados cuantitativos y del análisis mediante las métricas pertinentes.

En primer lugar se explica la descripción de los conjuntos de datos creados, su estructura y su contenido.

Se han creado en total dos conjuntos de datos distintos partiendo del análisis de todos los audios, para ello se han extraído diferentes características para componer cada conjunto. Además, para realizar el proceso de extracción de características se ha hecho uso de dos librerías distintas para cada caso.

El primer conjunto de características ha sido obtenido mediante la librería ‘*Spicy*’ mayoritariamente y se han extraído las siguientes características, 1º derivada de la frecuencia fundamental, 2º derivada de la frecuencia fundamental, Jitter, Shimmer, APQ, PPQ, energía logarítmica, amplitud media del audio y la duración del audio.

Todas ellas, se han almacenado en archivos de formato “.csv”, con una estructura por columnas por características, identificando cada fila con el nombre de cada archivo de audio, número de id del paciente y el tipo de audio al que se corresponde. Después, se ha hecho un *join* (con el punto de unión el id del paciente en ambos archivos), con otro archivo “.csv”, unificando de esta manera tanto los datos de las extracciones de las características como los datos clínicos relevantes de cada paciente. Obtenido de esta manera

finalmente un conjunto de datos por cada tipo de audio con el que se trabajará.

Ejemplo del primer conjunto de datos creado, al que denominaremos de aquí en adelante como conjunto de características Spicy:

Genero	Edad	ID	Estadio	archivo	tipo	ff1	ff2
Mujer		78 ID151		0 ID151_16-2-2 1.0		3,83E-02	-1,41E-03
Mujer		56 ID152		0 ID152_17-2-2 1.0		2,90E-02	1,87E-02
Hombre		56 ID153		0 ID153_17-2-2 1.0		-2,25E+16	-1,50E+16
Mujer		66 ID154		0 ID154_17-2-2 1.0		-3,49E-01	-2,47E-01
Mujer		58 ID155		0 ID155_17-2-2 1.0		1,54E-01	1,08E-01
Mujer		58 ID156		0 ID156_17-2-2 1.0		3,77E-02	2,51E-02
Hombre		78 ID157		0 ID157_17-2-2 1.0		1,48E-02	9,85E-03
Hombre		67 ID158		0 ID158_17-2-2 1.0		-5,20E-03	-3,93E-03
Mujer		92 ID159		0 ID159_17-2-2 1.0		-4,70E-02	-3,13E-02
Mujer		50 ID160		0 ID160_19-02- 1.0		-1,92E-02	-2,09E-03
jitter	shimmer	log_energy	apq	ppq	Amplitud media	duracion	
-1,40E+15	1,82E-01	1,94E+16	2,27E+12	2,89E+15	2,57E+16	3,17E+00	
-8,44E+15	1,54E-01	2,03E+16	3,45E+16	4,65E+13	2,63E+16	1,57E+04	
-4,44E+16	1,55E-01	1,98E+15	1,05E+16	1,36E+16	2,22E+15	7,88E+03	
-2,84E+15	1,22E-01	1,90E+16	1,59E+13	2,15E+13	3,78E+16	4,26E+03	
-1,13E+16	1,14E-01	1,83E+16	3,37E+15	4,37E+15	1,86E+16	5,05E+03	
-3,03E+16	1,24E-01	2,06E+15	1,43E+13	1,92E+13	2,71E+16	1,05E+04	
-6,52E+15	8,91E-02	2,08E+15	3,16E+13	4,02E+16	3,45E+16	1,49E+04	
-2,77E+15	1,04E-01	2,04E+16	5,53E+12	7,12E+12	1,92E+16	7,04E+03	
-6,75E+15	1,23E-01	2,05E+16	1,39E+16	1,79E+16	2,22E+15	8,71E+03	
-3,13E+14	1,72E-01	1,54E+16	1,36E+16	4,03E+15	6,63E+15	1,41E+04	

Figura 4.1: Ejemplo conjunto de datos extraidos con Spicy

Para el segundo conjunto de características, se ha utilizado la librería ‘*Torch*’ extrayendo las siguientes, MFCC, la amplitud en decibelios, media del espectrograma, densidad espectral de potencia (PSD), loudness y el centroide espectral. Haciendo con estas extracciones el mismo procedimiento que en el caso del primer conjunto. Obteniendo finalmente otros seis conjuntos diferentes de cada tipo de audio con estructura similar a los conjuntos anteriores.

Estructura del segundo conjunto de datos, al que a partir de ahora llamaremos conjunto de características torch:

Genero	Edad	ID	Estadio	archivo	tipo	amplitude_to_loudness	mfcc	psd	spectral_cent	spectrogram
Mujer		60 ID150		0 ID150_16-2-2 2.0		-3.997.069	-22.243.849	-16.495.598	34.336.107	26.542.778 0.64969814
Mujer		78 ID151		0 ID151_16-2-2 2.0		-351.483	-14.331.858	-14.778.366	15.522.028	21.258.718 3.061.758
Mujer		56 ID152		0 ID152_17-2-2 2.0		-36.855.324	-11.617.362	-14.761.337	24.702.545	19.516.483 48.062.606
Hombre		56 ID153		0 ID153_17-2-2 2.0		-46.072.514	-13.019.314	-15.618.957	1.924.682	13.798.083 37.850.149
Mujer		66 ID154		0 ID154_17-2-2 2.0		-3.876.945	-12.851.757	-142.731.285	22.731.693	1.229.181 43.928.537
Mujer		58 ID155		0 ID155_17-2-2 2.0		-38.814.064	-15.865.065	-15.989.177	10.339.837	16.816.471 20.375.626

Figura 4.2: Ejemplo conjunto de datos extraidos con Pytorch

Con los conjuntos de datos creados, se han utilizado tres modelos de calificación diferentes para entrenarlos. Y para entrenar estos modelos, se ha

optado por seguir el proceso de validación cruzada con cada subconjunto de datos (recordar que los subconjuntos estarán compuestos con la información extraída de 65 audios, todos del mismo tipo).

A continuación, se hablará de los diferentes resultados de rendimiento obtenidos para el entrenamiento de cada modelo, para cada tipo de audio, con el conjunto concreto de características.

Tras entrenar los diferentes modelos se han usado distintas métricas del rendimiento. Para poder evaluar y cuantificar de esta forma la eficacia de cada de ellos en cada tipo de audios.

Los valores de estas métricas se han almacenado en un archivo con formato “.csv”, para que se fácil para comparar los valores de las métricas según el conjunto utilizado para cada modelo y tipo de audio. En total se han tenido en cuenta y se han calculado las siguientes 10 métricas: exactitud, precisión, sensibilidad, puntaje de F1, valores de TN, FP, FN, TP, FPR y TPR.

Las tablas con todos los valores en crudo se encuentran en el apartado G.2 Detalle de resultados de los anexos. En el resumen de resultados de exponen únicamente los valores de las métricas más relevantes en el proyecto. Esto es, los valores de la exactitud, precisión, sensibilidad y puntaje de F1. Y para una mejor visualización de los datos y comparación entre los valores según el conjunto y tipo de audio analizado. Se ha hecho una síntesis de dichos resultados en forma de gráfico de radar. Ya que son especialmente útiles a la hora de comparar múltiples conjuntos de datos. Visualizando así de forma clara el comportamiento de cada uno de los modelos según el subconjunto de datos utilizado para el entrenamiento de cada uno.

Concretando, todos los resultados en cuatro gráficas de radar o también conocidas como gráficas de araña, donde se representan en cada una de ellas una de las métricas más relevantes para la evaluación del modelo en una gráfica independiente, con todos los tipos de modelos como variables.

Estas gráficas se han representado de la siguiente manera. Cada gráfica corresponde a una de las cuatro métricas tenidas en cuenta. En total se componen de 18 ejes distintos. Agrupados por cada 6, correspondientes a cada tipo de clasificador utilizado. El centro de la circunferencia corresponde a un valor de 0 en la métrica y en los extremos corresponde al 1, esto es al valor máximo posible. La representación con la línea azul son los datos obtenidos al entrenar los modelos con el subconjunto de valores extraídos mediante la librería Spicy. Por el contrario, la línea naranja a los modelos con las características extraídas con la librería torch.

El gráfico así adopta una forma más clara y concisa para poder analizar los resultados, también se han usado las siguientes abreviaturas para nombrar a cada tipo de modelo en los ejes.

- **Árbol X**: Clasificador de árbol de decisiones, con un número detrás correspondiente al tipo de audio para el que ha sido entrenado.
- **KNN X**: Clasificador de vecinos más cercanos, con un número detrás correspondiente al tipo de audio para el que ha sido entrenado.
- **RF X**: Clasificador Random Forest, con un número detrás correspondiente al tipo de audio para el que ha sido entrenado.

El tipo de audio será representado con un número, que se referirá al audio el contenido correspondiente en cada caso:

- **Tipo 1**: Audio con la vocal A.
- **Tipo 2**: Audio con la vocal I.
- **Tipo 3**: Audio con la vocal U.
- **Tipo 4**: Audio con la palabra campana.
- **Tipo 5**: Audio con la palabra gato.
- **Tipo 6**: Audio con la palabra petaca.



Se comienza con el gráfico que representa los valores de la métrica exactitud. Esta métrica nos indica el rendimiento general de cada modelo, evaluando la predicción global para clasificar correctamente las clases asignadas. Por ende, podemos ver que de forma uniforme hay un claro predominio de la exactitud de los modelos entrenados con el subconjunto spicy, sobre los entrenados con el subconjunto de datos torch, a excepción del modelo de clasificación árbol de decisiones de tipo de audio 3,. Siendo destacable también que todos los valores predominantes rondan sobre la cuarta línea del radar, esto es, en un valor aproximado y mantenido uniformemente de 0,8.

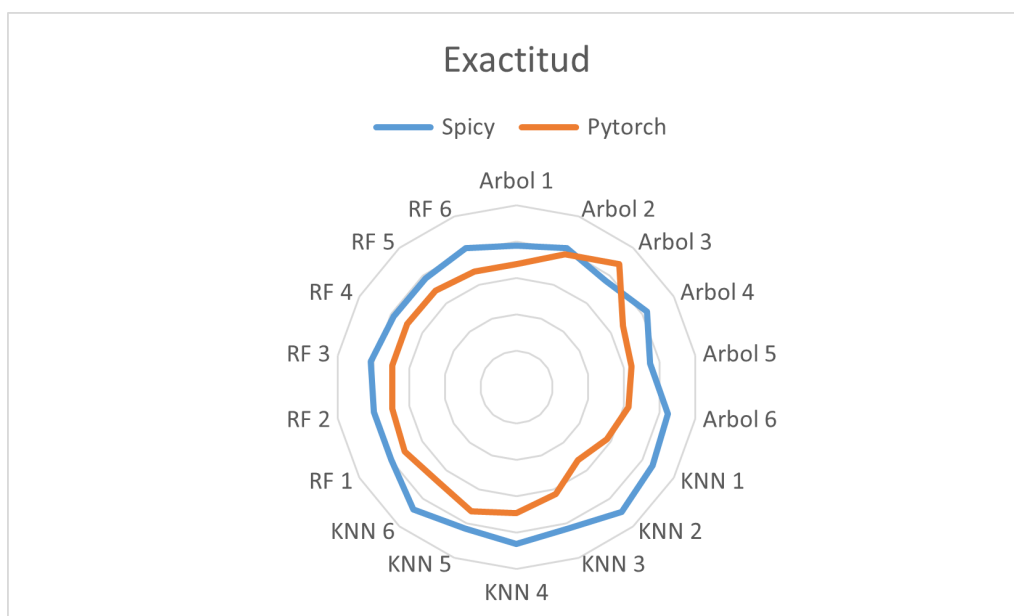


Figura 4.3: Comparación para el clasificador árbol de decisiones.

Fuente: Elaboración propia

Después se tiene la representación de los valores de precisión de los modelos. Usada para medir la proporción de resultados reales positivos clasificados sobre todos los clasificados como positivos.

En este otro caso, podemos ver como tenemos dos exposiciones sobre la norma, el caso del modelo de clasificación árbol de decisiones de tipo audio 2 y 3. Aunque en este caso los valores de la precisión son más aproximados entre ambos subconjuntos, siguen predominando los modelos entrenados con el subconjunto de datos Spicy. También podemos apreciar que en este caso los valores de la precisión son algo menos estables que el caso de la métrica de exactitud.

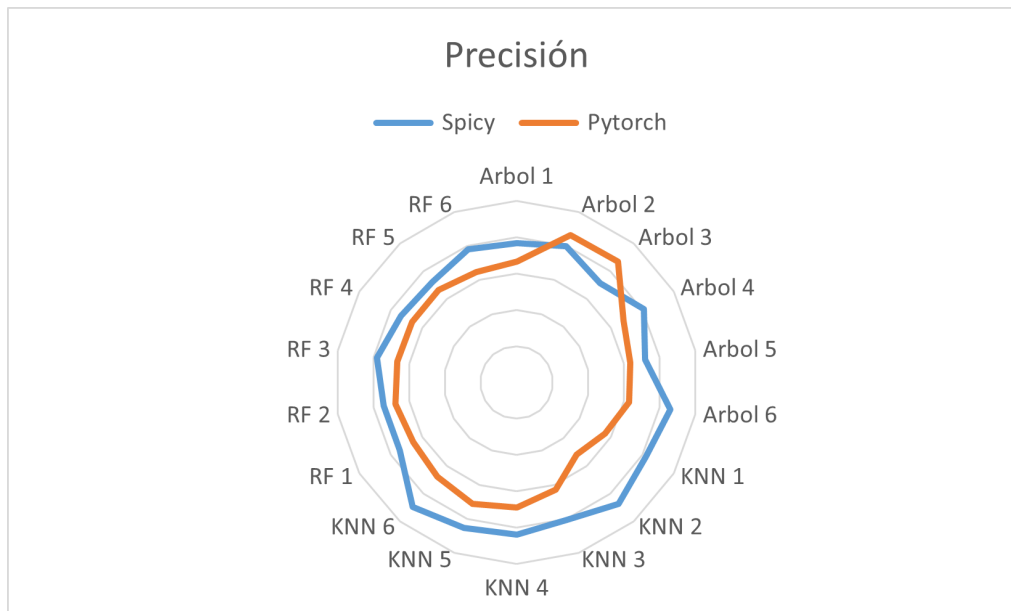


Figura 4.4: Comparación para el clasificador árbol de decisiones.

Fuente: Elaboración propia

Por otro lado, se tiene el gráfico correspondiente a los valores de la métrica sensibilidad, para cada modelo. La métrica proporciona la evaluación de los positivos clasificados correctamente sobre el total de predicciones positivas, teniendo en cuenta los falsos negativos.

En el que predomina en todo momento el rendimiento de los modelos entrenados con el subconjunto de datos Spicy, a excepción del modelo de clasificación árbol de decisiones de tipo audio 3. Cabe destacar lo poco uniformes que son los valores de los modelos con el subconjunto torch, en comparación con los modelos al ser entrenados con el subconjunto de datos Spicy. Siendo estos uniformes entre ellos, manteniendo en todo momento unos valores por encima de la cuarta línea del radar, esto es, valores mayores al 0.8 en todo momento. Siendo en consecuencia la métrica con mejor rendimiento sobre las cuatro comentadas.

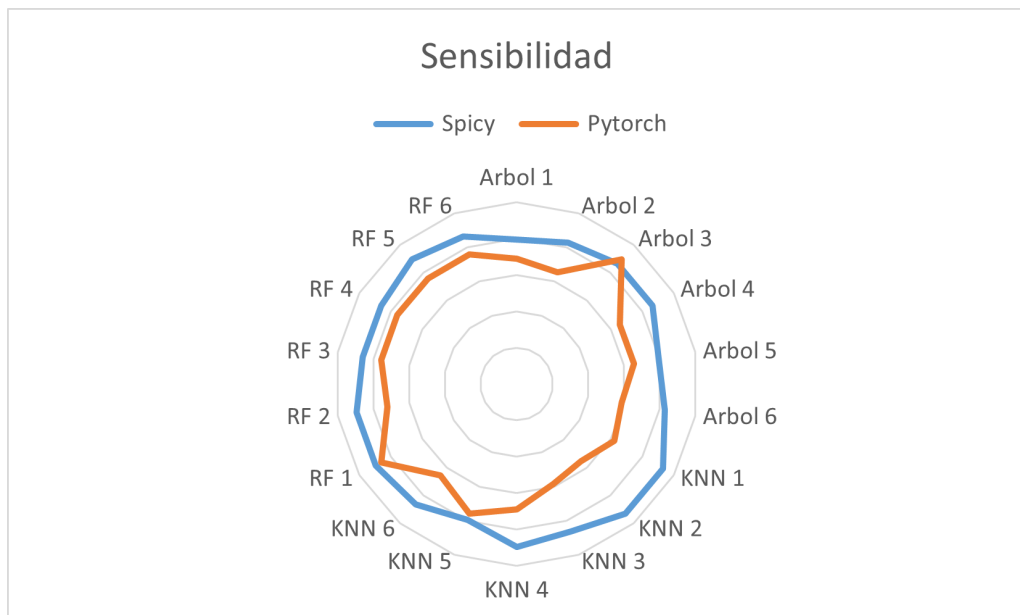


Figura 4.5: Comparación para el clasificador árbol de decisiones.

Fuente: Elaboración propia

En cuarto lugar, estamos ante la representación de los valores de la métrica de puntaje F1. La cual, combina el valor de la precisión con el valor de sensibilidad de cada modelo, proporcionando una medida general del rendimiento de los mismos.

Donde se puede apreciar una situación similar de comportamiento de la métrica de exactitud. Volviendo a tener el mismo patrón de predominancia de los modelos entrenados con el subconjunto Spicy, a excepción del caso árbol 3.

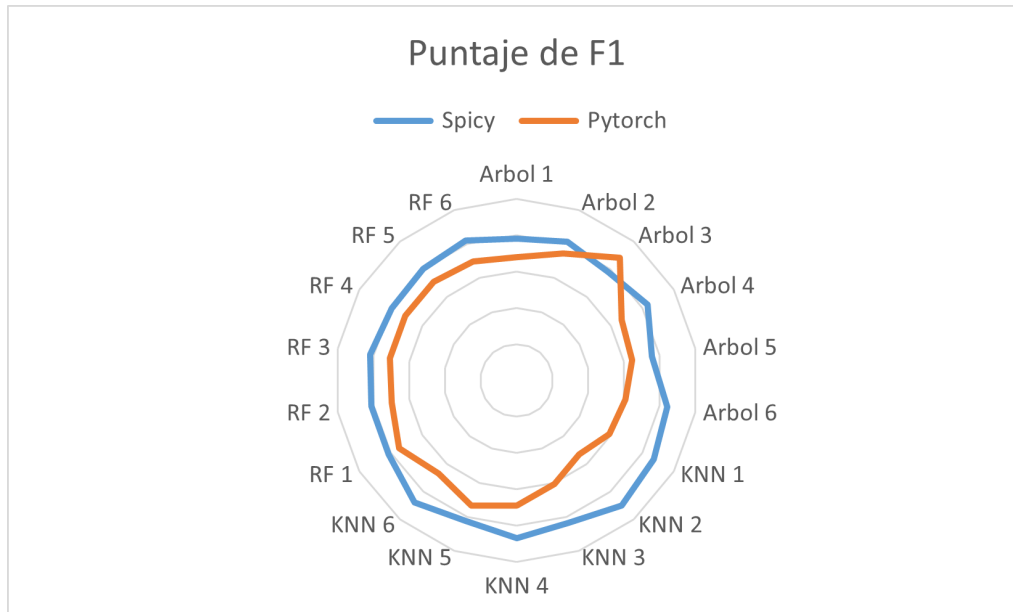


Figura 4.6: Comparación para el clasificador árbol de decisiones.

Fuente: Elaboración propia

Gracias a la comparación de los valores de estas métricas para cada tipo de audio con cada conjunto y modelo. Se tiene la capacidad de decidir qué conjunto de datos es el más adecuado para entrenar los modelos según el tipo de audio con el que se trabaja.

Además, tras tener el conjunto más optimo a utilizar para cada tipo de audio, también se hará una comparación de la eficacia y rendimiento entre los diferentes modelos de clasificación. Relacionando los valores necesarios incluidos en crudo en el anexo G.1 Cuaderno de trabajo.

Todo ello, tiene como objetivo principal, identificar y seleccionar tanto el mejor conjunto como modelo de clasificación para poder desarrollar un prototipo. Todos los modelos ya entrenados se han ido almacenando en los directorios pertinentes en cada caso, con un nombre identificativo y en un archivo de formato *“.jolib”*, proporcionando la opción de poder cargar el modelo entrenado en otro notebook cuando sea preciso. Todos ellos se encuentran en los directorios correspondientes junto con los valores de las métricas en cada caso guardadas en archivos *“.csv”*.

## 4.2. Discusion

En este apartado, se procederá a discutir y concluir sobre los resultados mencionados en el apartado anterior. Tomando diversas decisiones sobre la composición del prototipo creado para el proyecto.

Se comienza con la representación de los resultados en tres gráficos diferentes. Cada uno de ellos, representa los valores de rendimiento de cada audio con cada unidad de conjunto de datos, para cada tipo de modelo entrenado. De tal manera, que sea más fácil compararlos todos conjuntamente de manera visual e intuitiva.

Los gráficos seguirán la siguiente estructura. En el eje X serán representados cada audio y para cada uno de ellos cada tipo de conjunto utilizado. Y en el eje Y, se representaran los valores de las cuatro métricas mencionadas anteriormente. Cada métrica será correspondida con un color distintivo. De manera que sea más visuales de ver los resultados y poderlos comparar entre los distintos subconjuntos de datos utilizados. Además de resaltar en cada gráfica los mejores casos para cada modelo de clasificación, con un rectángulo de color rojo identificativo.

En este primer gráfico se ven los resultados para el modelo de clasificación Árbol de decisiones. Donde se determinan de una manera clara que el modelo entrenado con mejores resultados será el entrenado con el subconjunto de datos Spicy, a excepción del tipo de audio 3, donde se obtiene un mejor resultado al entrenar el modelo con el subconjunto de datos torch.

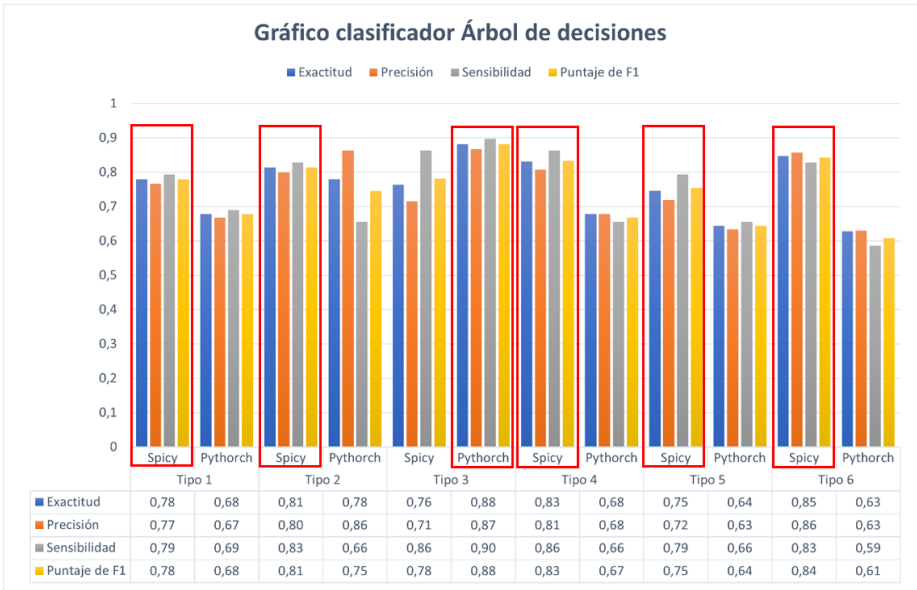


Figura 4.7: Comparación para el clasificador árbol de decisiones.

Fuente: Elaboración propia

En segundo lugar, se visualiza el gráfico que representa los resultados para el modelo de clasificación de KNN. En esta representación se observa unos mejores resultados utilizando el subconjunto de datos extraídos con la librería spicy. Siendo relevante en todos los tipos de audios y con una tendencia significativa.

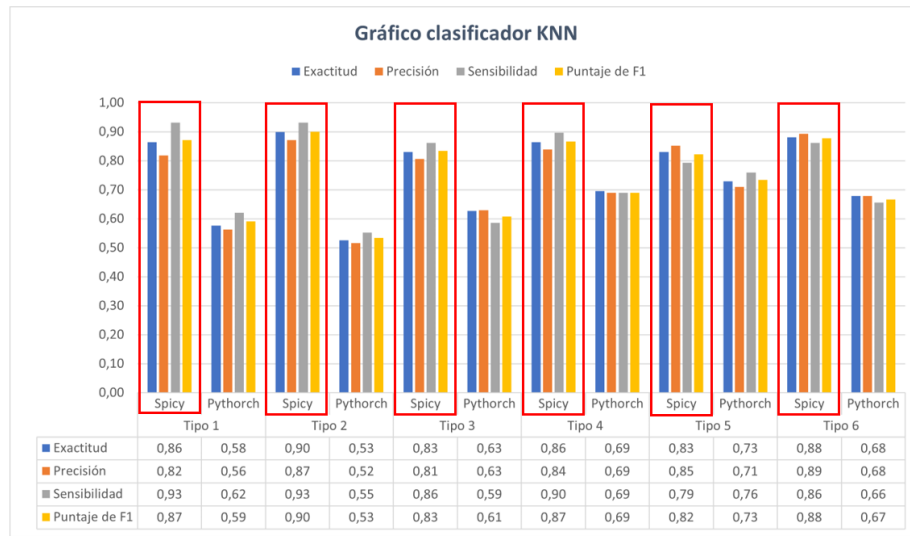


Figura 4.8: Comparación para el clasificador KNN.

Fuente: Elaboración propia

En la tercera gráfica se visualiza la comparación del modelo de clasificación Random Forest. En la cual se observa que a pesar de que la diferencia de rendimiento del modelo no sea tan relevante como en el caso anterior, también predomina para todos los tipos de audios el modelo entrenado con el subconjunto de datos extraídos con la librería spicity.

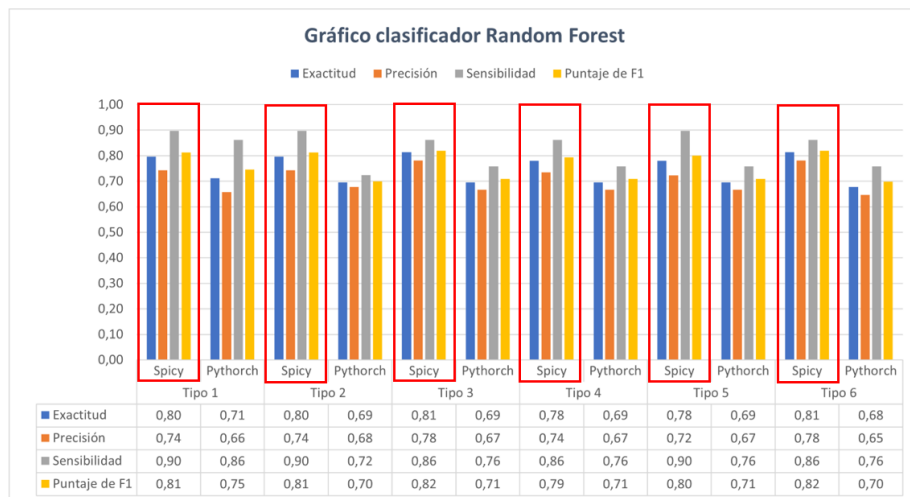


Figura 4.9: Comparación para el clasificador Random Forest.

Fuente: Elaboración propia

Tras analizar el caso de cada modelo de clasificación para cada tipo de audio, se puede identificar un patrón claro, el cual se repite en todos los casos a excepción de uno. Siendo que los modelos entrenados con el subconjunto compuesto por las características como, 1º derivada de la frecuencia fundamental, 2º derivada de la frecuencia fundamental, Jitter, Shimmer, APQ, PPQ, energía logarítmica, amplitud media del audio y la duración del audio, extraídas mediante la librería *spicy*, obtienen unos valores mejores a los del otro subconjunto en la mayoría todas las métricas valoradas.

Esto podría deberse a varios factores fundamentales, que logran que los clasificadores puedan ser mejor entrenados y den resultados mucho más precisos, sensibles y con el menor margen de error posible.

1. El factor de la cantidad de características que se han tenido en cuenta en cada subconjunto, teniendo más características en el subconjunto de datos *spicy*.
2. El factor de la relevancia de las características del subconjunto *Spicy*, son más adecuadas y significativas a la hora de analizar un audio de voz y su comportamiento.
3. El factor de la especificidad respecto a la enfermedad de Parkinson, teniendo en cuenta que las características extraídas para el subconjunto *spicy*, representan mejor los cambios en la voz como en la fonatoria del habla, en la resonante de la potencia del tono, en la articulación de los fonemas o palabras o incluso las alteraciones en el ritmo del habla.

La segunda parte de este apartado tendrá como objetivo determinar cual es el modelo que mejor se comporta a la hora de poder detectar la enfermedad de Parkinson con la mayor eficacia y menor rango de error posible. Para ello, se han sintetizado los resultados comentados en el apartado anterior, pero esta vez ya únicamente haciendo uso de los modelos con un subconjunto en específico, el subconjunto que mejores resultados haya obtenido para cada tipo de audio y para cada modelo, según la conclusión dada en la primera parte de este mismo apartado.

Con la idea de que sea más visual poder comparar los datos de los tres modelos a la vez, se ha realizado un gráfico representando los valores de las métricas para cada caso.



Se trata de un gráfico de barras que sigue la siguiente estructura. Un eje Y con los posibles valores de las cuatro métricas más relevantes para evaluar el rendimiento de un modelo de clasificación. Y un eje X, donde se representan los tres modelos utilizados dentro de cada tipo de audio. Además de tener asignada a cada métrica a un color distintivo para diferenciarlos con mayor facilidad.

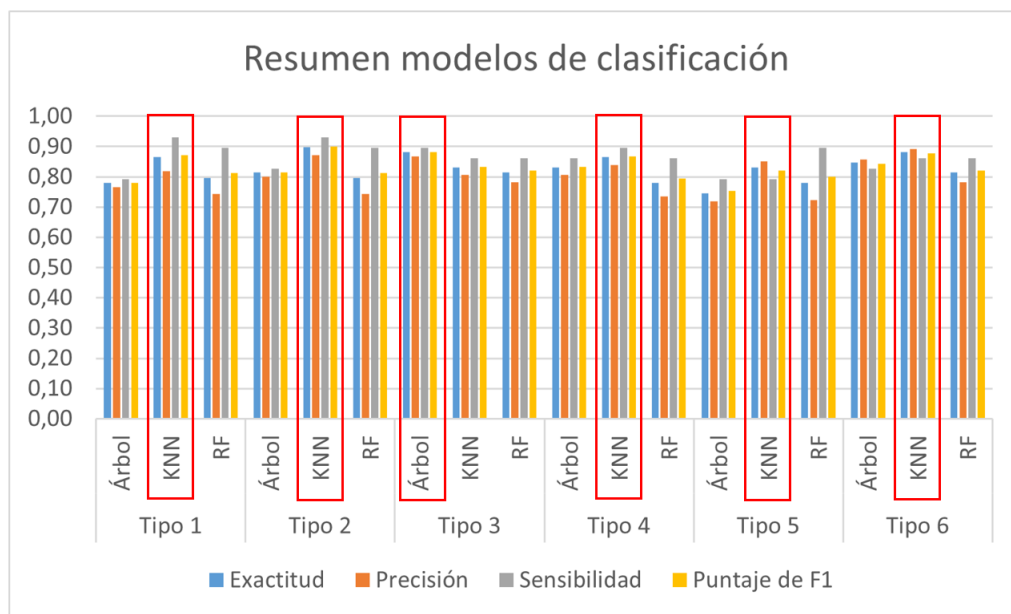


Figura 4.10: Resumen para la comparación de los modelos de clasificación.

Fuente: Elaboración propia

En base a la representación anterior. Se analiza el rendimiento de cada modelo para cada tipo de audio.

En primer lugar, se tienen los valores de rendimiento con los audios de tipo 1 (vocal A), donde se puede objetivizar una pequeña diferencia entre los valores del modelo KNN sobre los otros dos. Predominando en los valores de las cuatro métricas sobre los otros dos modelos. Y con un valor superior en la sensibilidad del modelo KNN. Indicando así que el modelo de clasificación entrenado con el algoritmo de KNN para los audios de tipo 1(vocal A), tiene un acierto casi sin errores a la hora de clasificar los positivos dentro de los positivos reales.

En segundo lugar, se tiene la comparación para los audios de tipo 2 (vocal I). En este caso, también se puede destacar un predominio claro del rendimiento del modelo de KNN respecto a los demás. En este caso dentro de los valores de este mismo modelo, se visualiza que las cuatro métricas son valor aproximado al 1 lo cual indica, que el modelo elegido es idóneo.

En tercer lugar, se tiene la comparación de los modelos haciendo uso de los audios de tipo 3 (vocal U). Siendo los tres modelos bastante similares en todas las métricas. Pero siendo el modelo de árbol de decisiones en este caso el de mayor rendimiento por poseer los valores más próximos al 1 entre los tres.

En cuarto lugar, se puede observar la comparación de los modelos con los audios de tipo 4 (palabra campana). En este caso pese a estar muy parejas las métricas tanto del modelo de árbol de decisiones como las del modelo de KNN. Tiene una pequeña superioridad el clasificador KNN. Siendo la consecuencia el óptimo para clasificar los audios de tipo 4.

En penúltimo lugar está la comparación para los audios de tipo 5 (palabra gato). Donde a pesar de que el valor destacablemente alto lo tiene la métrica de sensibilidad para el modelo de clasificación Random Forest. En conjunto obtiene un mejor rendimiento el modelo KNN. Por lo que se concluye que será el óptimo para clasificar correctamente los audios de este tipo.

Por último, está la representación que compara el comportamiento de los modelos para los audios de tipo 6 (palabra petaca). Para este caso, los tres modelos obtienen un alto valor en sus métricas. Por lo que cualquiera de los tres podría ser utilizado. Aunque, el modelo de clasificación KNN será el elegido, puesto que es el que mayor valores de rendimiento obtiene en las cuatro características.

A excepción del caso de los audios de tipo 3. Se destaca una clara dominancia de rendimiento del modelo de clasificación KNN sobre el resto para todos los tipos de audios. Para determinar o detectar si el paciente padece de Parkinson o no. A excepción del caso de los audios de tipo 3 (vocal U), donde anteriormente se ha indicado el modelo de elección sería el entrenado con el algoritmo de árbol de decisiones.

Una de las causas posibles puede ser que el modelo de clasificación KNN se basa en la similitud de las muestras y no en las etiquetas de la enfermedad durante el entrenamiento. Ya que utiliza la cercanía de los datos para realizar las predicciones. En este caso, la estructura de los datos de las características de los audios son complejas y variadas. Implicando así que el modelo KNN haya podido aprovechar la similitud entre las muestras para afianzar sus

resultados de la manera más precisa posible. Siendo esto una ventaja sobre los otros dos modelos que no son capaces de aprovechar esta característica.

Finalmente, tras la valoración de todos los subconjuntos y modelos posibles para cada tipo de audio. Se ha llegado a la conclusión de cuáles serán las mejores combinaciones para crear un prototipo de clasificador automático para la enfermedad de Parkinson. Siendo los siguientes modelos entrenados los que obtienen la mejor eficacia entre los comparados y analizados. Estos han sido los cargados en el programa para crear el prototipo.

Los resultados de las selección han sido los siguientes:

- **Tipo 1 (vocal A):** modelo de clasificación KNN, entrenado con el subconjunto de características spicy.
- **Tipo 2 (vocal I):** modelo de clasificación KNN, entrenado con el subconjunto de características spicy.
- **Tipo 3 (vocal U):** modelo de clasificación árbol de decisiones, entrenado con el subconjunto de características torch.
- **Tipo 4 (palabra campana):** modelo de clasificación KNN, entrenado con el subconjunto de características spicy.
- **Tipo 5 (palabra gato):** modelo de clasificación KNN, entrenado con el subconjunto de características spicy.
- **Tipo 6 (palabra petaca):** modelo de clasificación KNN, entrenado con el subconjunto de características spicy.

Para cerrar el apartado de conclusiones, se hablará del prototipo o demo creada.

El uso del programa podrá hacerse a través de una interfaz sencilla e intuitiva de usar. La cual, se ejecuta a través de un archivo de formato “.py”, directamente en la terminal del ordenador desde donde se podrá utilizar. Para su creación se ha hecho uso de la librería de Python Streamlit.

Tal y como se ve en la captura de la estructura del prototipo. Por un lado, tiene en el lado izquierdo una entrada donde se introducirá directamente el la ruta del directorio que contiene los seis tipos de audios necesarios para hacer una predicción. Además, para determinar el paciente a estudiar se introduciras una serie de datos personarles, nombre, 1º apellido, 2º apellido, edad y género.



Figura 4.11: Captura del prototipo.

Fuente: Elaboración propia

Tras introducir todos los datos de manera adecuada, se debe pulsar el botón de realizar predicción. Y devolvera los resultados. Comenzando por los datos personales del paciente, los resultados para cada tipo de audio y una conclusión final de cuantos de los audios serán positivos en proporción a los seis audios, con una representación visual del mismo en forma de gráfico de tarta. Se verá el funcionamiento general con un ejemplo, en los anexos.

Cuando se pulsa el botón de predecir se ejecuta el programa completo. Este está compuesto por diferentes fases por las cuales pasará el audio antes de hacer la predicción para cada uno de ellos. Empezando por un preprocesado, para normalizar el volumen y eliminar el ruido de fondo de estos. A continuación el programa extrae las características para cada uno de ellos, añadiéndoles los datos de identificación personal. Posteriormente, tras hacer la carga de los modelos entrenados, hará uso de cada uno de ellos, segun el tipo de audio correspondiente y obtendrá el resultado. Devolviendo de esta manera, el grupo al que pertenece cada audio. Determinando como resultado el porcentaje de audios positivos al Parkison respecto del total y representándolo con una gráfica de tarta.

---

## Aspectos relevantes del desarrollo del proyecto

---

Para poder desarrollar este proyecto, fue fundamental la primera fase. En esta se recogieron los audios tanto de personas con enfermedad de Parkinson, como de personas sin diagnosticar siendo estos pacientes control. Dichas personas han sido participes en este estudio de manera altruista.

Estos audios han sido recogidos en el HUBU<sup>5</sup>, gracias al equipo de neurología de Dra Esther Cubo Delgado, siendo así recogidos los audios entre los profesionales de su equipo y la persona que presenta este proyecto. Los audios de los pacientes diagnosticados con Parkinson fueron recogidos de pacientes ya diagnosticados de dicha enfermedad por la doctora. En cambio, los audios recogidos de las personas no diagnosticados, pacientes control, tienen una procedencia más diversa.

Para la grabación de todos los audios se ha seleccionado el procedimiento con 7 pasos:

1. Informar al paciente en concreto de la investigación en la que va a participar, así como, los objetivos de esta, los aspectos legales que le puedan influir y el procedimiento que se seguirá para la recogida de los audios.
2. Administración de todo lo indicado de forma verbal en físico, con toda la información explicada de forma precisa.
3. Firma del paciente del consentimiento informado para poder participar como voluntario de manera altruista en la investigación.

---

<sup>5</sup>Hospital Universitario de Burgos

4. Obtención de información personal del paciente para uso exclusivo de la investigación: nombre completo, genero, edad.
5. Grabar los audios: para este paso, como se grababan 6 audios con contenido distinto con cada paciente, se les explica que tenían que decir para cada grabación y cual sería tanto la señal para el inicio de la prueba como el termino de la misma. Si es necesario se ha recordado antes de cada grabación, lo que se tiene que decir y el procedimiento.
6. Tras la grabación de los audios se agradece a cada voluntario su participación y el empleo de su tiempo en dicha investigación.
7. Recopilación de los audios, con la nomenclatura definida anteriormente y descrita en el apartado 3.1 de la memoria. Para después almacenar los audios de forma ordenada en una carpeta a disposición de los investigadores.
8. Recogida bajo custodia en el HUBU de todos los consentimientos de las personas partícipes en dicho proyecto.

---

## Lineas de trabajo futuras

---

A medida que la detección automática de la enfermedad de Parkinson mediante audios de voz continúa mostrando resultados prometedores, es crucial explorar nuevas líneas de investigación para mejorar aún más la precisión y la efectividad de los modelos de clasificación.

Pudiéndose identificar así diversas líneas futuras para la mejora de este y otros prototipos en la detección temprana y un diagnóstico preciso de la enfermedad de Parkinson.

Se destacan cinco vertientes de trabajo.

1. Primeramente se podría hacer una fusión de datos multimodales, esto es, además de partir de los datos extraídos de los audios de voz, integrar también otro tipo de datos y pruebas para mejorar la precisión de diagnóstico. Como por ejemplo, grabación para visualizar datos de afectación física.
2. En segunda línea podría ser mejorar las características de las grabaciones en las audiciones, para evaluar nuevas características y entrenar a los modelos con ellas, consiguiendo así que sean más precisos. Con ello, se intensifican los entrenamientos de los modelos.
3. Otra línea de actuación puede ser el aumento de estudio de campo. Para que estos resultados se puedan generalizar en una población diana determinada previamente.
4. Como cuarta línea de trabajo, se puede considerar opción de hacer algo similar con modelos de aprendizaje profundo como por ejemplo redes neuronales, mejorando la capacidad de detección y clasificación

de la enfermedad. Ya que estos son capaces de capturar patrones más complejos y aprender representaciones más significativas de los audios de voz.

5. Como última línea se podría trabajar sobre las posibles mejoras en la implementación del prototipo. Como por ejemplo, conseguir una interfaz a la cual se pudiera acceder online, además de poder subir los audios de entrada desde la misma nube. De esta manera, se podría adjuntar o incorporar como prueba diagnóstica en la historia clínica del paciente.



---

## Bibliografía

---

- [1] 1.3.2 características básicas de una señal periódica (amplitud, frecuencia, fase). [http://cidecame.uaeh.edu.mx/lcc/mapa/PROYECTO/libro27/132\\_caractersticas\\_bsicas\\_de\\_una\\_seal\\_peridica\\_amplitud\\_frecuencia\\_fase.html](http://cidecame.uaeh.edu.mx/lcc/mapa/PROYECTO/libro27/132_caractersticas_bsicas_de_una_seal_peridica_amplitud_frecuencia_fase.html).
- [2] Aprendizaje supervisado: Random forest classification. <https://aprendeia.com/aprendizaje-supervisado-random-forest-classification/>.
- [3] Centroide espectral cálculo y uso alternativo. [https://hmong.es/wiki/Spectral\\_centroid](https://hmong.es/wiki/Spectral_centroid).
- [4] csv — csv file reading and writing — python 3.11.3 documentation. <https://docs.python.org/3/library/csv.html>.
- [5] Definición de biomarcador - diccionario de cáncer del nci - nci. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/biomarcador>. (Accessed on 02/09/2023).
- [6] Definición de diagnóstico - diccionario de cáncer del nci - nci. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/diagnostico>. (Accessed on 02/09/2023).
- [7] Diferencia entre calcular el espectro de potencia y la densidad del espectro de potencia de una señal - ni. <https://knowledge.ni.com/KnowledgeArticleDetails?id=kA00Z0000015A5tSAE&l=es-MX>. (Accessed on 05/23/2023).

- [8] Diferencia entre calcular el espectro de potencia y la densidad del espectro de potencia de una señal - ni. <https://knowledge.ni.com/KnowledgeArticleDetails?id=kA00Z0000015A5tSAE&l=es-MX>.
- [9] Escala de hoehn y yahr | consultadeneurologia. <https://consultadeneurologia.com/enfermedades-degenerativas/enfermedad-de-parkinson/escala-de-hoehn-y-yahr/>. (Accessed on 02/09/2023).
- [10] Extracción de características | interactive chaos. <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/extraccion-de-caracteristicas>.
- [11] Frecuencia fundamental. <https://www.studio-22.com/blog/enciclopedia/frecuencia-fundamental>.
- [12] librosa — librosa 0.10.1dev documentation. <https://librosa.org/doc/main/>.
- [13] Machine learning: Algoritmos de clasificación y regresión - the black box lab. <https://theblackboxlab.com/2022/05/06/machine-learning-diferencias-entre-algoritmos-clasificacion-regresion/>.
- [14] Machine learning: Seleccin metricas de clasificacion - sitio-bigdata.com. <https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/>.
- [15] Mfcc : definición de mfcc y sinónimos de mfcc (español). <https://diccionario.sensagent.com/MFCC/es-es/>.
- [16] Métricas de evaluación de modelos en el aprendizaje automático. <https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico>.
- [17] Numpy documentation. <https://numpy.org/doc/>.
- [18] os — miscellaneous operating system interfaces — python 3.11.3 documentation. <https://docs.python.org/3/library/os.html>.
- [19] pandas documentation — pandas 2.0.1 documentation. <https://pandas.pydata.org/docs/>.
- [20] Project jupyter | home. <https://jupyter.org/>.

- [21] pydub/api.markdown at master · jiaaro/pydub. <https://github.com/jiaaro/pydub/blob/master/API.markdown#audiosegment>.
- [22] Pytorch documentation — pytorch 2.0 documentation. <https://pytorch.org/docs/stable/index.html>.
- [23] Qué es y para que se utiliza un algoritmo de árbol de decisiones | tokio. <https://www.tokioschool.com/noticias/algoritmo-arbol-decision/>.
- [24] scikit-learn: machine learning in python — scikit-learn 1.2.2 documentation. <https://scikit-learn.org/stable/>.
- [25] Scipy api — scipy v1.10.1 manual. <https://docs.scipy.org/doc/scipy/reference/>.
- [26] shimmer | sottovoce. <https://sottovoce.hypotheses.org/tag/shimmer>. (Accessed on 02/09/2023).
- [27] shutil — high-level file operations — python 3.11.3 documentation. <https://docs.python.org/3/library/shutil.html>.
- [28] Significado de interfaz (qué es, concepto y definición) - significados. <https://www.significados.com/interfaz/>.
- [29] sklearn.ensemble.randomforestclassifier — scikit-learn 1.2.2 documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [30] sklearn.model\_selection.train\_test\_split — scikit-learn 1.2.2 documentation. [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html).
- [31] sklearn.neighbors.kneighborsclassifier — scikit-learn 1.2.2 documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.
- [32] sklearn.tree.decisiontreeclassifier — scikit-learn 1.2.2 documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>.
- [33] Streamlit documentation. <https://docs.streamlit.io/>. (Accessed on 06/24/2023).
- [34] Torchaudio documentation — torchaudio 2.0.1 documentation. <https://pytorch.org/audio/stable/index.html>.

- [35] Users guide — matplotlib 3.7.1 documentation. <https://matplotlib.org/stable/users/index.html>.
- [36] Validación cruzada - amazon machine learning. [https://docs.aws.amazon.com/es\\_es/machine-learning/latest/dg/cross-validation.html](https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/cross-validation.html).
- [37] ¿qué es el espectrograma? usos en el análisis musical | unir. <https://ecuador.unir.net/actualidad-unir/espectrograma/>.
- [38] ¿qué es knn? | ibm. <https://www.ibm.com/mx-es/topics/knn>.
- [39] ¿qué es loudness en el audio y por qué es importante? | tvyvideo. <https://www.tvyvideo.com/2023011620348/articulos/tecnologia/que-es-loudness-en-el-audio-y-por-que-es-importante.html>.
- [40] ¿qué es python? - explicación del lenguaje python - aws. <https://aws.amazon.com/es/what-is/python/>.
- [41] ¿qué es un conjunto de datos? definición, casos de uso, ventajas y ejemplos. <https://brightdata.es/blog/datos-web/what-is-a-dataset>.
- [42] ¿sabes qué es un decibelio? | audiothechno. <https://www.audiothechno.es/que-es-un-decibelio/>.
- [43] ...la escala de richter es logarítmica y mide la energía liberada en un terremoto? – revista e-medida. <https://www.e-medida.es/numero-6/la-escala-de-richter-es-logaritmica-y-mide-la-energia-liberada-en-un-terremoto/>.
- [44] Enfermedad de parkinson | mal de parkinson | medlineplus en español, 2003.
- [45] Enrique Chávez-León, Martha Patricia Ontiveros-Urbe, and José Damián Carrillo-Ruiz. La enfermedad de parkinson: neurología para psiquiatras. *Salud mental*, 36(4):315–324, 2013.
- [46] Bruno Dubois Anette Schrag Matthew B. Stern Anthony E. Lang Peter A. LeWitt Stanley Fahn Joseph Jankovic C. Warren Olanow Pablo Martinez-Martin Andrew Lees Olivier Rascol Bob van Hilten Glenn T. Stebbins Robert Holloway David Nyenhuis Cristina Sampaio Richard Dodel Jaime Kulisevsky-Barbara Tilley Sue Leurgans Jean Teresi Christopher G. Goetz, Werner Poewe.

- [47] Jorge Eduardo Correa Pillajo. Selección de características de la voz de personas con parkinson, usando algoritmos genéticos. B.S. thesis, Quito, 2019., 2019.
- [48] William T. Dauer and Serge Przedborski. Parkinson's disease: mechanisms and models. *Neuron*, 39(6):889–909, 2003.
- [49] Giovanni Dimauro, Vincenzo Di Nicola, Vitoantonio Bevilacqua, Danilo Caivano, and Francesco Girardi. Assessment of speech intelligibility in parkinson's disease using a speech-to-text system. *IEEE Access*, 5:22199–22208, 2017.
- [50] Belén García-Botija Aldana. Diagnóstico de la enfermedad de parkinson usando deep learning y grabaciones de voz mediante teléfono móvil. 2019.
- [51] Mark A. Hall Christopher J. Pal I. H. (Ian H.), Eibe. Frank and Mark A. *Data mining: practical machine learning tools and techniques (Fourth edition.)*. Elsevier, 2017.
- [52] Formal analysis Funding acquisition Methodology Software Writing – original draft Writing – review editing 1 S. Shattuck-Hufnagel Conceptualization Formal analysis-Supervision Writing – review editing 1 J. Y. Choi Formal analysis Supervision Writing – review editing 1 L. Moro-Velázquez Software Visualization Writing – review editing 2 J. I. Godino-Llorente, Conceptualization and Software Writing – review editing 2 J. A. Gómez-García, Data curation. Towards the identification of idiopathic parkinson's disease from the speech. new articulatory kinetic biomarkers - pmc. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5730127/>, 14 December 2017. (Accessed on 02/09/2023).
- [53] J D Arias-Londoño J F Vargas-Bonilla K Daqrouq S Skodda J Rusz E Nöth J R Orozco-Arroyave, F Hönig. Automatic detection of parkinson's disease in running speech spoken in three different languages. page 481–500, 2016.
- [54] T Yoshida Y Hirade H Kasuya Y Kikuchi M Hirano, S Hibi. Acoustic analysis of pathological voice. some results of clinical application - pubmed. <https://pubmed.ncbi.nlm.nih.gov/3400446/>.
- [55] Connie Marras and Susan Fox. Parkinson's disease: Diagnosis and clinical management. *BMJ : British Medical Journal*, 357:j1951, 2018.

- [56] SERVICIOS SOCIALES E IGUALDAD MINISTERIO DE SANIDAD. Estrategia en enfermedades neurodegenerativas del sistema nacional de salud. [https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/pdf/Est\\_Neurodegenerativas\\_APROBADA\\_C\\_INTERTERRITORIAL.pdf](https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/pdf/Est_Neurodegenerativas_APROBADA_C_INTERTERRITORIAL.pdf). (Accessed on 06/24/2023).
- [57] Carlos Morales Passos-Pereira. Análisis del efecto del parkinson en el temblor de la voz: frecuencia fundamental. 2016.
- [58] Adrián Arnaiz Rodríguez. Aplicación de técnicas de minería de datos para la detección de la enfermedad del parkinson. Master's thesis, 2021.
- [59] John Tapia-Núñez and Pedro Chaná. Diagnosis of parkinson's disease. *Revista de neurologia*, 38:61–7, 01 2004.
- [60] Carla Collazo Joanna Krzysztofik Elliot Mann Philippa Jones Laura Mills Cheney Drew Anne Rosser Rebecca Cousins Grzegorz Witkowski Esther Cubo Monica Busse Madeleine M Lowery Vitória S Fahed, Emer P Doheny. Characterization of speech impairment in huntington's disease using language-agnostic acoustic features. 2019.
- [61] Wikipedia. Jitter — wikipedia, la enciclopedia libre, 2022. [Internet; descargado 9-febrero-2023].