

DETECCIÓN DE LA ENFERMEDAD DE PARKINSON MEDIANTE LA VOZ USANDO
TÉCNICAS DE MACHINE LEARNING

Jouver Fabian Delgado Moreno y Luzdey Sthefanny Duque Mantilla

Director

Carlos Augusto Fajardo Ariza

Doctor en Ingeniería Electrónica

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Ingeniería Electrónica

Bucaramanga

2025

Dedicatoria

Todo lo que he alcanzado hasta el momento es gracias al apoyo constante e imperturbable de mi mamá, a ella le agradezco por poner toda su fe en mí, por cada vez que sus palabras me dieron ganas de continuar aun cuando la situación era la más compleja. Es por esto que a ella le dedico no solo este trabajo de grado, sino todos estos años de formación profesional y todos los logros que conlleva ello.

A mí compañera y amiga Luzdey Duque, a quien le tengo una gran admiración y alguien por quién tengo que agradecerle a Dios por el resto de mis días por cruzarla en mi camino.

Jouver Fabian Delgado Moreno

A Dios por ser siempre mi guía y mi fortaleza. A mi mamá, Cielo Mantilla, mi mayor fuente de amor e inspiración, porque sin su apoyo y sacrificio, este logro no habría sido posible. Este triunfo es tan suyo como mío. A los dos papás que la vida me regaló, Rubén Duque y Edwin Herreño, a mis hermanos, Daniel Duque y Juan Herreño, y a mis abuelos, por ser mi ancla y refugio en esta vida. Finalmente, a Jouver Delgado, gracias por haber sido el mejor compañero en esta travesía universitaria y por brindarme su amistad sincera fuera de ella.

Luzdey Sthefanny Duque Mantilla

Agradecimientos

Al culminar esta etapa tan significativa de mi vida, es imposible no mirar atrás y recordar cada momento, cada desafío y cada persona que, de una u otra forma, hizo parte de este camino. Hoy, con gratitud infinita, quiero dedicar estas palabras a quienes me han acompañado y apoyado incondicionalmente.

En primer lugar, a la persona que ha sido mi mayor inspiración, mi apoyo inquebrantable y quien siempre confió en mí desde el primer día: mi mamá. A ella, dedico este trabajo de grado, aunque sé que jamás será suficiente para retribuir todo su amor, su sacrificio y sus palabras de aliento, esas que, en los momentos más difíciles, me llenaron de fuerza y motivación.

A mis hermanos, Edwin, Sandrith y Kendry Delgado, con quienes he compartido innumerables momentos llenos de alegría y complicidad. Gracias por ser mi refugio y mi alegría, y por regalarme la hermosa noticia de la llegada de mi sobrino Emilio, quien ha iluminado nuestras vidas con su presencia.

A mis ángeles: Laura Sofía Monroy Morales, María José Lizarazo Laguado, Fabián Andrés Amado Núñez y Luzdey Sthefanny Duque Mantilla. Junto a ustedes viví los recuerdos más hermosos de mi etapa universitaria. Su compañía fue mi fortaleza y su amistad, un regalo invaluable. Gracias por ser ese apoyo emocional que todos necesitamos en algún momento.

A mis profesores, quienes con paciencia y vocación nos guiaron en este camino del saber. Gracias por compartir su conocimiento con dedicación y por ser faros de luz en nuestra formación profesional y personal.

Y finalmente, a la Universidad Industrial de Santander, por brindarme la oportunidad de ser parte de esta prestigiosa institución. Aquí encontré retos que me desafiaron constantemente, no solo para convertirme en un mejor profesional, sino en una mejor persona. A todos, gracias.

Jouver Fabian Delgado Moreno

Hoy, al llegar a este momento tan significativo en mi vida, quiero dedicar unas palabras de gratitud a todos los que han sido una parte esencial en este viaje. Primero que todo, quiero agradecer a mis padres, quienes han sido la base de todo lo que soy. A mi madre, por su amor constante y su fortaleza, que ha sido mi refugio en los momentos de incertidumbre. A mi padre, por su ejemplo de sacrificio y dedicación, que me ha enseñado a nunca rendirme y a seguir adelante, sin importar lo difícil que pueda ser el camino.

A mis hermanos, por su apoyo incondicional y por estar siempre a mi lado, inspirándome a seguir mis sueños con el mismo amor que siempre me han brindado. A todos mis familiares y amigos, por sus palabras de aliento, por su fe en mí, y por recordarme en cada paso que todo es posible cuando se cree en uno mismo.

También quiero agradecer profundamente a la Universidad Industrial de Santander por darme la oportunidad de formarme y crecer en un ambiente académico que siempre me desafió y me impulsó a ser mejor. A mis profesores, cuya pasión por la enseñanza y el conocimiento me ha dejado huellas imborrables, gracias por ser guías y por compartir su sabiduría.

Por último, quiero reconocerme a mí misma, por no haberme detenido a pesar de las dificultades, por mantener la determinación cuando todo parecía complejo, y por dar lo mejor de mí en cada paso. Este logro es el resultado de la combinación del esfuerzo propio y del apoyo de todos ustedes, y por eso les agradezco profundamente.

Luzdey Sthefanny Duque Mantilla

Tabla de Contenido

	Pág.
Introducción	14
1. Objetivos	16
1.1 Objetivo General	16
1.2 Objetivos Específicos.....	16
2. Trabajos Relacionados	17
3. Metodología	19
3.1 Datos	19
3.1.1. Conjunto de datos	20
3.1.1.1. Oxford Parkinson's Disease Detection Dataset (Parkinsons).	20
3.1.1.2. Corpus PC-GITA.	20
3.1.2. Características	21
3.1.3. Extracción de Características	22
3.2. Métodos de selección de características	24
3.2.1. CHI ² (χ^2)	25
3.2.2. Matriz de Correlación	25
3.2.3. Clasificador Extra Tree	25
3.2.4. Preprocesamiento de datos.....	26
3.2.5. División de datos.....	26
3.3. Modelos de Machine Learning	26
3.4. Métricas de Evaluación.....	27

3.4.1. Accuracy	27
3.4.2. AUC-ROC.....	27
3.4.3. Precision.....	28
3.4.4. Recall	28
3.4.5. F1-Score.....	28
3.5. Matriz de Confusión	28
4. Resultados.....	30
4.1. Resultados de los mejores modelos por conjuntos de datos	31
5. Análisis de Resultados	35
6. Conclusiones.....	37
Referencias Bibliográficas	38
Apéndices.....	41

Lista de Tablas

	Pág.
Tabla 1 Descripción de las características usadas en el conjunto de datos Parkinsons	21
Tabla 2 Recopilación de los mejores resultados por cada conjunto de datos	32

Lista de Figuras

	Pág.
Figura 1 Diagrama que ilustra el proceso desarrollado para esta investigación	20
Figura 2 Análisis de espectro de la voz y sección para obtener reporte de voz.....	23
Figura 3 Reporte de características de voz extraídas mediante el software Praat	24
Figura 4 Matriz de confusión teórica	29
Figura 5 Esquema de la secuencia de pruebas diseñada para identificar el modelo con mejor desempeño.....	30
Figura 6 Representación de las curvas AUC-ROC de los mejores modelos obtenidos para cada conjunto de datos	31
Figura 7 Matrices de confusión correspondientes a los mejores modelos obtenidos para cada conjunto de datos: (a) Vocal A, (b) Vocal E, (c) Vocal I, (d) Vocal O, (e) Vocal U, (f) Vocales juntas, (g) Parkinsons.....	33

Lista de Apéndices

	pág.
Apéndice A. Repositorio Github DEPAVUM-232	41

Glosario

Espectrograma: representación visual de la energía de una señal en el tiempo y la frecuencia.

Filtro pasa banda Butterworth: filtro que permite el paso de frecuencias dentro de un rango específico y atenúa las demás.

HNR: evalúa la calidad armónica de la voz, siendo un indicador clave para determinar la claridad vocal y la presencia de ruido.

Jitter DDP: evalúa las irregularidades en la producción vocal.

MDVP F0: representa el tono promedio de la voz y es un indicador importante en el análisis acústico, ya que puede estar alterado en pacientes con Parkinson.

MDVP Fhi: evalúa el rango superior de la capacidad vocal del hablante.

MDVP Flo: ayuda a analizar la estabilidad vocal en los registros más bajos.

MDVP Jitter(%): indica la regularidad del tono entre ciclos vocales consecutivos, siendo útil para detectar alteraciones en la voz.

MDVP Jitter(Abs): mide la variación promedio de los períodos de vibración vocal.

MDVP RAP: Calcula las variaciones en la frecuencia de cada ciclo en relación con los vecinos inmediatos.

MDVP Shimmer(dB): representa la fluctuación en la amplitud de la voz y es relevante para analizar variaciones en la intensidad vocal.

NHR: mide la proporción de ruido presente en la señal vocal con respecto a sus componentes armónicos, indicando claridad o deterioro en la voz.

Resolución: número de bits que representan cada muestra de una señal digital.

Shimmer APQ5: mide las fluctuaciones de amplitud en un marco más amplio de análisis (5 períodos consecutivos).

Shimmer DDA: refleja variaciones en la energía vocal entre períodos consecutivos.

Tasa de muestreo: frecuencia con la que se registran muestras de una señal analógica para convertirla en digital, medida en Hertz (Hz).

Vocales sostenidas: Se refiere a la emisión prolongada de un sonido vocálico, como /a/, /e/, /i/, /o/ o /u/, mantenido de forma continua durante varios segundos.

Resumen

Título: DETECCIÓN DE LA ENFERMEDAD DE PARKINSON MEDIANTE LA VOZ USANDO TÉCNICAS DE MACHINE LEARNING *

Autor(es): Jouver Fabian Delgado Moreno, Luzdey Sthefanny Duque Mantilla **

Director: Carlos Augusto Fajardo Ariza. Dr. Ingeniería Electrónica

Palabras Clave: Aprendizaje Automático, Enfermedad de Parkinson (EP), Clasificadores.

Descripción: La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo causado por la disminución de dopamina, afectando funciones motoras y no motoras. Esta deficiencia altera actividades como la generación del habla, que requiere la coordinación de estructuras como la laringe, cuerdas vocales y lengua. En pacientes con EP, esta coordinación se ve comprometida, provocando cambios significativos en las características del habla, afectando tanto la claridad como la fluidez de la comunicación. El objetivo de esta investigación se centró en la detección de la enfermedad de Parkinson a partir del análisis de características vocales usando dos bases de datos: PC-GITA, donde se extrajeron las características mediante el *software* Praat, y el conjunto de datos Parkinsons, que ya incluía características preprocesadas. Se seleccionaron las diez características más relevantes utilizando los métodos Chi-cuadrado, clasificador *Extra Tree* y Matriz de Correlación. Se evaluaron diversos modelos de aprendizaje automático, incluidos Árbol de decisión o *Decision Tree* (DT), K- Vecinos Más Cercanos o *K-Nearest Neighbors* (KNN), Bosque Aleatorio o *Random Forest* (RF), *Bagging*, *Ada Boosting* y *Gradient Boost*, empleando las métricas de *accuracy*, *precision*, *recall*, *F1-score*, *R2-score* y AUC-ROC. Los resultados obtenidos sugieren que el modelo que demostró el mejor rendimiento fue *Ada Boost*, aplicado al conjunto de datos que combina todas las vocales en PC-GITA, donde alcanzó un AUC-ROC de 0.82 ± 0.04 . Estos resultados indican que el modelo *Ada Boost* proporcionó las estimaciones más precisas y consistentes en nuestro estudio, resaltando su eficacia en la identificación de características asociadas con la enfermedad de Parkinson.

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y Telecomunicaciones. Ingeniería Electrónica. Director: Carlos Augusto Fajardo Ariza. Doctor en Ingeniería Eelcetrónica.

Abstract

Title: DETECTION OF PARKINSON'S DISEASE THROUGH VOICE USING MACHINE LEARNING TECHNIQUES *

Author(s): Jouver Fabian Delgado Moreno, Luzdey Sthefanny Duque Mantilla **

Director: Carlos Augusto Fajardo Ariza. PhD in Electronic Engineering

Key Words: Machine Learning, Parkinson's Disease (PD), Classifiers

Description: Parkinson's disease (PD) is a neurodegenerative disorder caused by a decrease in dopamine, affecting both motor and non-motor functions. This deficiency alters activities such as speech production, which requires the coordination of structures such as the larynx, vocal cords, and tongue. In patients with PD, this coordination is compromised, leading to significant changes in speech characteristics that affect both clarity and communication fluency. The objective of this research was to detect Parkinson's disease through the analysis of vocal features using two databases: PC-GITA, where the features were extracted using Praat software, and the Parkinsons dataset, which already included preprocessed features. The ten most relevant features were selected using the Chi-square method, Extra Tree classifier, and Correlation Matrix. Various machine learning models were evaluated, including Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), Bagging, Ada Boosting, and Gradient Boost, using the metrics of accuracy, precision, recall, F1-score, R2-score, and AUC-ROC. The results suggest that the model demonstrating the best performance was Ada Boost, applied to the dataset that combines all vowels in PC-GITA, where it achieved an AUC-ROC of 0.82 ± 0.04 . These findings indicate that the Ada Boost model provided the most precise and consistent estimates in our study, highlighting its effectiveness in identifying characteristics associated with Parkinson's disease.

* Degree Work

** Faculty of Physical and Mechanical Engineering, School of Electrical, Electronic and Telecommunications Engineering. Electronic Engineering. Director: Carlos Augusto Fajardo Ariza. PhD. In Electronic Engineering.

Introducción

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo progresivo que afecta principalmente al sistema nervioso central y que se ha consolidado como la enfermedad de más rápido crecimiento entre los trastornos neurológicos, según la OMS (World Health Organization, 2022). Este trastorno, que impacta a millones de personas en todo el mundo, afecta predominantemente a la población mayor de 60 años (de Lau, 2006) y se caracteriza por síntomas motores como temblores en reposo, rigidez muscular y bradicinesia, así como síntomas no motores, entre los que destacan el deterioro cognitivo, trastornos del sueño y alteraciones del habla (World Health Organization, 2023).

Las alteraciones del habla, en particular la disartria hipocinética, afectan hasta al 90% de los pacientes diagnosticados con EP (Bahaddad, 2022). Estas alteraciones se manifiestan como cambios significativos en la calidad de la voz, tales como monotonía, debilidad vocal y dificultad para articular palabras, lo que subraya su importancia como un indicador temprano de la enfermedad. La detección y el análisis de estos cambios ofrecen una oportunidad valiosa para el diagnóstico temprano de la EP, lo que permite intervenciones más oportunas y potencialmente más efectivas.

En los últimos años, el análisis de voz se ha convertido en una herramienta de diagnóstico relevante debido a su naturaleza no invasiva y accesible. Sin embargo, los métodos tradicionales, como las evaluaciones perceptivas y los análisis acústicos dependientes del tiempo, enfrentan limitaciones al analizar señales de voz complejas, particularmente en pacientes en etapas avanzadas de la enfermedad. Este desafío resalta la necesidad de enfoques más precisos y eficientes para abordar las limitaciones actuales en el diagnóstico de la EP.

En el desarrollo de este proyecto, se presenta una estrategia para abordar estas limitaciones mediante el diseño de un enfoque híbrido que combina métodos de clasificadores de características y técnicas avanzadas de *Machine Learning*. En primer lugar, se emplean tres clasificadores complementarios para la selección de características relevantes: el estadístico chi-cuadrado, un clasificador *Extra Tree* y una matriz de correlación. Estos métodos permiten evaluar a cada característica en función de su relevancia para el diagnóstico de la EP. Esta elección se hizo con el fin de optimizar la selección de características y evitar redundancias o correlaciones no significativas.

1. Objetivos

1.1 Objetivo General

Desarrollar un modelo de *Machine Learning* para la detección del Parkinson a partir del análisis de características vocales.

1.2 Objetivos Específicos

- Preparar la base de datos PC-GITA para ser utilizada en el proceso de entrenamiento de los modelos de *Deep Learning*.
- Evaluar diversos modelos de *Machine Learning*, con el objetivo de identificar aquel que demuestre la mejor capacidad predictiva con base a las métricas de *precision*, *recall* y AUC-ROC.
- Realizar pruebas del modelo seleccionado mediante el uso de técnicas de validación cruzada.

2. Trabajos Relacionados

La investigación previa en el diagnóstico y clasificación de la enfermedad de Parkinson (EP) ha desarrollado metodologías que integran técnicas de selección de características con modelos de aprendizaje automático. Un análisis exhaustivo realizado con 40 sujetos, 20 con EP y 20 sanos, recopiló 26 señales de voz que incluían palabras, números y vocales. Este estudio utilizó el programa de análisis acústico Praat para registrar y extraer características acústicas relevantes de las grabaciones (Sakar, 2010). Otro enfoque combinó lógica difusa con *Principal Component Analysis* (PCA) para seleccionar las características más representativas, integrándolas en un modelo *Support Vector Machine* (SVM) que alcanzó una precisión del 93,47% (Li, 2011). Ambos estudios destacan la eficacia de combinar técnicas de reducción de características con clasificadores para mejorar la precisión diagnóstica.

En un trabajo relacionado, se diseñó un enfoque que utiliza señales de voz para distinguir entre 23 pacientes con EP y 8 sujetos sanos, empleando *Support Vector Machines* (SVM) como modelo de clasificación. Esta estrategia logró una precisión del 91,4%, demostrando el potencial de las características vocales como herramientas diagnósticas (Little, 2008). Además, otro estudio seleccionó 132 características relacionadas con el discurso disfónico mediante métodos de selección como LASSO, Relief y MRMR, reduciéndolas a 10 características óptimas para la clasificación (Tsanas, 2011).

Por último, se exploró la efectividad de los *Variational Autoencoders* (VAE) para extraer características relevantes en datos de voz relacionados con la EP. Este enfoque alcanzó una precisión del 91,6% al combinar estas características con un clasificador basado en *Support Vector Machines* (SVM) *multi-kernel* (Soke, 2021).

Estos estudios resaltan la importancia de metodologías bien estructuradas que integren técnicas de selección de características con algoritmos de clasificación para optimizar el diagnóstico y monitoreo de la EP.

3. Metodología

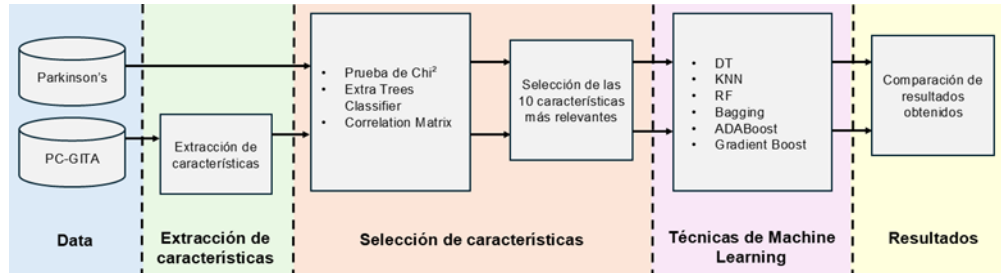
La metodología propuesta se dividió en cuatro etapas principales: extracción de características, clasificación de características, modelado con técnicas de *Machine Learning* y comparación de resultados. Se trabajó con dos bases de datos: PC-GITA (Orozco-Arroyave, 2014), al que se le realizó la extracción de características utilizando el *software* Praat (Boersma, 2025) para analizar señales de voz, y el conjunto de datos Parkinsons que ya contenía características preprocesadas (Little M. , 2025). Posteriormente, se seleccionaron las 10 características más relevantes mediante tres enfoques complementarios: el estadístico chi-cuadrado, el clasificador *Extra Tree* (ET) y una matriz de correlación. Con los conjuntos de datos con las características relevantes, se entrenaron y evaluaron seis clasificadores: *Decision Tree* (DT), *K-Nearest Neighbors* (KNN), *Random Forest* (RF), *Bagging*, *Ada Boosting* y *Gradient Boost*. Finalmente, se realizó una comparación de los resultados obtenidos utilizando métricas de rendimiento estándar como *accuracy*, *precision*, *recall*, *F1-score*, *R2-score* y el área bajo la curva ROC (AUC-ROC), identificando el modelo con mejor rendimiento en la detección de la enfermedad de Parkinson. Para mayor comprensión de la estructura de la metodología, visualizar la Figura 1.

3.1 Datos

En un proyecto de *Machine Learning*, no solo las técnicas desempeñan un papel fundamental al moldear el comportamiento de un algoritmo. Contar con una fuente de datos de calidad, es esencial para garantizar que los modelos predictivos sean precisos y logren métricas de rendimiento satisfactorias. Entre las fuentes de datos más utilizadas en investigaciones relacionadas con la detección de la enfermedad de Parkinson, destacan el *Oxford Parkinson's Disease Detection Dataset* y el corpus PC-GITA.

Figura 1

Diagrama que ilustra el proceso desarrollado para esta investigación



3.1.1. Conjunto de datos

3.1.1.1. Oxford Parkinson's Disease Detection Dataset (Parkinsons).

Es un conjunto de datos con características de 195 audios recolectados y extraídos de 31 pacientes, de los cuales 23 son diagnosticados con Parkinson y 8 son controles sanos. Siendo un conjunto de datos desbalanceado (Little M. , 2025). Este conjunto de datos es ampliamente referenciado en proyectos relacionados; sin embargo, al ser desarrollado con angloparlantes, puede no ser representativo en regiones lingüísticas como la nuestra. Por ello, su principal utilidad radica en servir como referencia al momento de desarrollar conjuntos de datos más adecuados para contextos específicos.

3.1.1.2. Corpus PC-GITA.

Este corpus fue desarrollado por el Grupo de Investigación en Telecomunicaciones Aplicadas (GITA) de la Universidad de Antioquia, en colaboración con la Clínica Noel de Medellín, Colombia, y se enfoca en hablantes de español. Contiene registros de audio de 50 pacientes diagnosticados con Parkinson y 50 controles sanos, equilibrados en cuanto a género (25 mujeres y 25 hombres por grupo) y edades.

El rango etario y las estadísticas descriptivas de los participantes se desglosan de la siguiente manera:

- Hombres con EP: edades entre 33 y 77 años, con una media de 62.2 ± 11.2 .
- Mujeres con EP: edades entre 44 y 75 años, con una media de 60.1 ± 7.8 .
- Hombres controles sanos: edades entre 31 y 86 años, con una media de 61.2 ± 11.3 .
- Mujeres controles sanos: edades entre 43 y 76 años, con una media de 60.7 ± 7.7 .

Los registros incluyen diferentes tipos de muestras, como monólogos, lecturas de texto, frases, y vocales sostenidas, con una tasa de muestreo de 44.1 kHz y una resolución de 16 bits (Orozco-Arroyave, 2014).

3.1.2. Características

Las características del conjunto de datos *Oxford Parkinson's Disease Detection* (Parkinsons) se utilizan como referencia para la extracción de características en el conjunto de datos PC-GITA. A continuación, en la Tabla 1, se describen las características clave, junto con sus abreviaturas (Wu, 2017), para facilitar su comprensión y aplicación en el análisis de la perturbación de la voz.

Tabla 1

Descripción de las características usadas en el conjunto de datos Parkinsons

	Abreviación	Descripción
Frecuencia fundamental vocal	MDVP:F0	Frecuencia fundamental vocal promedio (en Hz)
	MDVP:Fhi	Frecuencia fundamental vocal máxima (en Hz)
	MDVP:Flo	Frecuencia fundamental vocal mínima (en Hz)
	MDVP:Jitter(%)	Porcentaje de fluctuación de la frecuencia
	MDVP:Jitter(Abs)	Fluctuación absoluta de la frecuencia

Perturbaciones en la frecuencia (Jitter)	MDVP:PPQ	Cociente de perturbación del período de cinco puntos
	MDVP:RAP	Parámetro de perturbación de amplitud relativa
	Jitter:DDP	Diferencia promedio absoluta de las variaciones entre los ciclos de Jitter
Perturbaciones en la amplitud (Shimmer)	MDVP:Shimmer	Shimmer local en unidades originales
	MDVP:Shimmer(dB)	Shimmer local en unidades logarítmicas
	Shimmer:APQ3	Cociente de perturbación de Shimmer de tres puntos
	Shimmer:APQ5	Cociente de perturbación de Shimmer de cinco puntos
	MDVP:APQ11	Cociente de perturbación de amplitud de 11 puntos
	Shimmer:DDA	Diferencia promedio absoluta entre las amplitudes de períodos consecutivos
Relación ruido-armónico y armónico-ruido	NHR	Relación ruido-armónico
	HNR	Relación armónico-ruido

3.1.3. Extracción de Características

Para la extracción de características del corpus PC-GITA, se seleccionaron las vocales sostenidas como fuente de análisis. Esta decisión se fundamenta en estudios previos que indican que las vocales sostenidas ofrecen resultados consistentes en la detección de patrones asociados con la enfermedad de Parkinson (Bang, 2013).

Para el procesamiento de los audios de las vocales sostenidas, se utilizó el *software* libre Praat, desarrollado por P. Boersma y D. Weenink en la Universidad de Ámsterdam (Boersma, 2025).

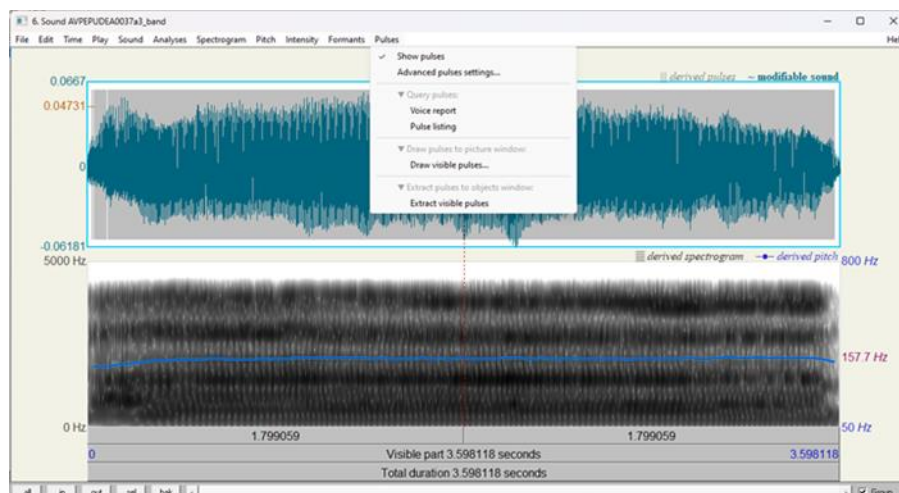
Este programa, ampliamente reconocido en estudios científicos de lingüística, es ideal para realizar análisis detallados de espectrogramas de señales de audio grabadas, facilitando tanto el preprocesamiento como la extracción de características relevantes.

En este proyecto, el preprocesamiento incluyó la aplicación de un filtro pasa banda Butterworth de cuarto orden, con frecuencias de corte en 50 Hz y 4 kHz (Pah, 2022). Este rango asegura que se capturen principalmente las frecuencias asociadas a la voz humana, descartando componentes de ruido.

Después de aplicar el filtro pasa banda, se procede a la visualización del espectro del audio, como se muestra en la Figura 2. Esta representación espectral permite analizar la calidad del preprocesamiento y verificar que las frecuencias no deseadas hayan sido correctamente filtradas.

Figura 2

Análisis de espectro de la voz y sección para obtener reporte de voz

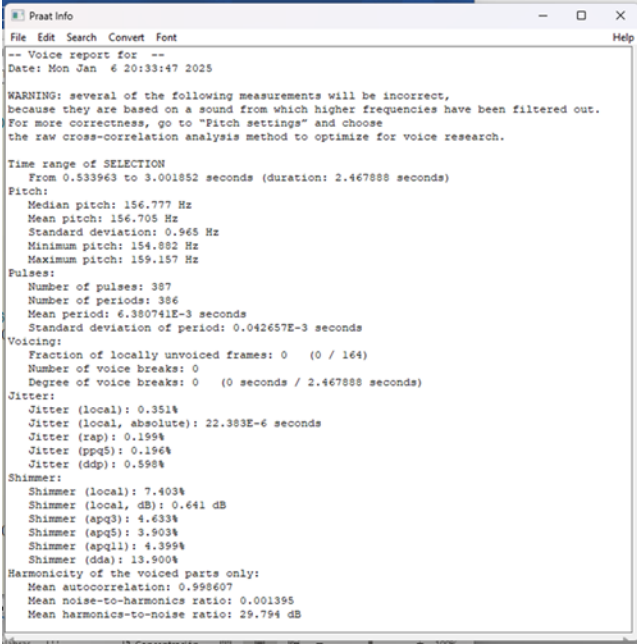


A partir de esta visualización, se genera un reporte de voz que contiene las características necesarias para la construcción del conjunto de datos, como se observa en

la Figura 3. Estas características extraídas de las vocales sostenidas constituyen la base de datos que se utilizará en el entrenamiento y evaluación de los modelos predictivos.

Figura 3

Reporte de características de voz extraídas mediante el software Praat



```
Praat info
File Edit Search Convert Font
-- Voice report for --
Date: Mon Jan 6 20:33:47 2025

WARNING: several of the following measurements will be incorrect,
because they are based on a sound from which higher frequencies have been filtered out.
For more correctness, go to "Pitch settings" and choose
the raw cross-correlation analysis method to optimize for voice research.

Time range of SELECTION
From 0.533963 to 3.001852 seconds (duration: 2.467888 seconds)
Pitch:
Median pitch: 156.777 Hz
Mean pitch: 156.705 Hz
Standard deviation: 0.965 Hz
Minimum pitch: 154.882 Hz
Maximum pitch: 159.157 Hz
Pulses:
Number of pulses: 387
Number of periods: 386
Mean period: 6.380741E-3 seconds
Standard deviation of period: 0.042657E-3 seconds
Voicing:
Fraction of locally unvoiced frames: 0 (0 / 164)
Number of voice breaks: 0
Degree of voice breaks: 0 (0 seconds / 2.467888 seconds)
Jitter:
Jitter (local): 0.351%
Jitter (local, absolute): 22.383E-6 seconds
Jitter (zap): 0.199%
Jitter (ppq5): 0.196%
Jitter (ddp): 0.598%
Shimmer:
Shimmer (local): 7.403%
Shimmer (local, dB): 0.641 dB
Shimmer (apq3): 4.633%
Shimmer (apq5): 3.903%
Shimmer (apq11): 4.399%
Shimmer (dda): 13.900%
Harmonicity of the voiced parts only:
Mean autocorrelation: 0.998607
Mean noise-to-harmonics ratio: 0.001395
Mean harmonics-to-noise ratio: 29.794 dB
```

Los datos obtenidos se almacenan en archivos CSV para su posterior procesamiento. Por cada vocal sostenida se generan 300 datos, lo que da un total de 1500 datos considerando las cinco vocales. En total, se trabaja con 7 conjuntos de datos, siendo 5 conjuntos por cada vocal, un conjunto para todas las vocales juntas y el conjunto de datos de Parkinsons.

3.2. Métodos de selección de características

Se decidió no entrenar los modelos con todas las características extraídas, sino centrarse en aquellas consideradas más relevantes (Yadav, 2023). Estas características extraídas corresponden a las utilizadas en el conjunto de datos Parkinsons (Little M. M., 2007). Para su selección se emplearon los métodos Chi-Cuadrado (χ^2), Matriz de correlación y *Extra Trees* (ET), los cuales se describen a continuación.

3.2.1. *CHI² (χ^2)*

Este método evalúa la dependencia entre las características y la variable objetivo mediante el cálculo del estadístico χ^2 (Pearson, 1900). Se utiliza la siguiente fórmula:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (1)$$

Donde O representa las frecuencias observadas y E las esperadas. Las características con mayores valores χ^2 son seleccionadas como las más relevantes, ya que tienen una relación más significativa con la variable objetivo.

3.2.2. *Matriz de Correlación*

La correlación es una medida estadística que evalúa la fuerza y dirección de la relación entre las características del conjunto de datos y la variable objetivo, siendo esta relación positiva (ambas aumentan) o negativa (una aumenta mientras la otra disminuye). Una alta correlación sugiere una fuerte asociación con la variable objetivo, mientras que valores cercanos a cero indican una relación débil o inexistente. A través de un mapa de calor, se visualizan fácilmente estas relaciones, permitiendo identificar las características más relevantes para la variable objetivo.

3.2.3. *Clasificador Extra Tree*

Para identificar las características más relevantes, se utiliza el modelo *Extra Tree* (ET), que asigna puntuaciones a cada característica basándose en su importancia (Sturges, 1926). Cuanto mayor sea la puntuación, más relevante será la característica para la variable objetivo. En este caso, se emplea el clasificador ET para seleccionar las diez características principales del conjunto de datos, facilitando la extracción de las más destacadas.

3.2.4. Preprocesamiento de datos

Para garantizar la compatibilidad y uniformidad en las características analizadas, se aplicaron pasos comunes de preprocesamiento antes de utilizar los métodos de selección de características. En primer lugar, las variables continuas se transformaron en categorías discretas para la prueba de Chi-cuadrado (χ^2), utilizando partición en intervalos equitativos basada en percentiles determinados por el Criterio de Sturges (Sturges, 1926):

$$q = \log_2(n) + 1 \quad (2)$$

Posteriormente, para todos los métodos (Chi-cuadrado, matriz de correlación y *Extra Trees Classifier*), las características fueron normalizadas mediante escalado Min-Max, ajustando sus valores a un rango entre 0 y 1. Este proceso garantiza rangos homogéneos entre las características, reduce posibles sesgos y facilita un análisis estadístico coherente.

3.2.5. División de datos

Utilizando los clasificadores para calcular las características más prometedoras, se seleccionan las diez mejores características de cada clasificador y se crearon tres subconjuntos de datos por cada uno de los siete conjuntos originales. A partir de estos subconjuntos, las muestras se dividen en un conjunto de entrenamiento y un conjunto de prueba. El 20% de las muestras se destina para pruebas, mientras que el 80% restante se utiliza para el entrenamiento del modelo.

3.3. Modelos de Machine Learning

En este proyecto se emplearon los siguientes modelos de aprendizaje automático: *Decision Tree* (DT), *K-Nearest Neighbors* (KNN), *Random Forest* (RF), *Bagging*, *AdaBoost* y *Gradient Boosting*.

El *Decision Tree*, introducido por Quinlan, organiza los datos en una estructura jerárquica basada en decisiones, evaluando características en nodos internos hasta alcanzar etiquetas en los

nodos hoja (Quinlan, 1986). Por otro lado, el *K-Nearest Neighbors* (KNN) clasifica nuevas instancias en función de la proximidad a sus k vecinos más cercanos, siendo un método no paramétrico, pero con alta demanda computacional (Altman, 1992).

Entre las técnicas de ensamblado, *Random Forest* mejora la precisión al combinar múltiples árboles de decisión generados con subconjuntos aleatorios del conjunto de datos y características (Anisha, 2022). De manera similar, *Bagging* utiliza muestreo con reemplazo para entrenar modelos independientes y reducir la varianza (Tiwari, 2016). *AdaBoost* ajusta iterativamente los pesos de las muestras para centrarse en las mal clasificadas, permitiendo la creación de un modelo más robusto (Freund, 1997). Finalmente, *Gradient Boosting* emplea el descenso por gradiente para minimizar errores y refinar el modelo en cada iteración (Li X. L., 2018).

3.4. Métricas de Evaluación

Con el objetivo de poder determinar qué modelo se ajusta más a los datos, se hace necesario usar métricas de rendimiento que comprueben los resultados.

3.4.1. Accuracy

Mide el porcentaje de instancias correctamente clasificadas entre todas las instancias del conjunto de datos. Es representada por la siguiente fórmula:

$$Accuracy = \frac{T_P + T_N}{N} \quad (3)$$

3.4.2. AUC-ROC

Evalúa la capacidad discriminativa del modelo a partir de la curva *Receiver Operating Characteristic* (ROC). Representa el área bajo esta curva (AUC), que muestra la relación entre la tasa de verdaderos positivos (*True Positive Rate*) y la tasa de falsos positivos (*False Positive Rate*) en distintos umbrales de clasificación.

3.4.3. Precision

La precisión indica la proporción de instancias clasificadas como positivas que realmente son positivas, en relación con todas las instancias que el modelo identificó como positivas. Se calcula utilizando la siguiente fórmula:

$$\mathbf{Precision} = \frac{T_P}{T_P + F_P} \quad (4)$$

3.4.4. Recall

Mide la proporción de predicciones positivas correctas en relación con el total de instancias positivas presentes en el conjunto de datos. La siguiente fórmula se utiliza para calcularlo:

$$\mathbf{Recall} = \frac{T_P}{T_P + F_n} \quad (5)$$

3.4.5. F1-Score

El *F1-Score* equilibra *Precision* y *Recall*, considerando tanto falsos positivos como falsos negativos. Se calcula como la media armónica entre ambos. Se calcula mediante la fórmula (6):

$$\mathbf{F1\ Score} = \frac{2}{\left(\frac{1}{Precision} + \frac{1}{Recall}\right)} \quad (6)$$

3.5. Matriz de Confusión

Es una métrica utilizada en la evaluación de modelos de clasificación, ya que ofrece una representación visual del desempeño del modelo al comparar sus predicciones con las etiquetas reales del conjunto de datos. Esta matriz organiza la información en cuatro categorías fundamentales: Verdaderos Positivos o *True Positives* (TP), Verdaderos Negativos o *True Negatives* (TN), Falsos Positivos o *False Positives* (FP) y Falsos Negativos o *False Negatives* (FN), así como se ve en la Figura 4, permitiendo analizar detalladamente los aciertos y errores del modelo.

Figura 4

Matriz de confusión teórica

A 2x2 confusion matrix diagram. The vertical axis is labeled 'Valores de Predicción' and the horizontal axis is labeled 'Valores Reales'. The four quadrants are: Top-Left (green): Verdaderos positivos (TP); Top-Right (orange): Falsos Positivos (FP); Bottom-Left (orange): Falsos Negativos (FN); Bottom-Right (green): Verdaderos Negativos (TN).

Valores de Predicción	Verdaderos positivos (TP)	Falsos Positivos (FP)
	Falsos Negativos (FN)	Verdaderos Negativos (TN)
	Valores Reales	

4. Resultados

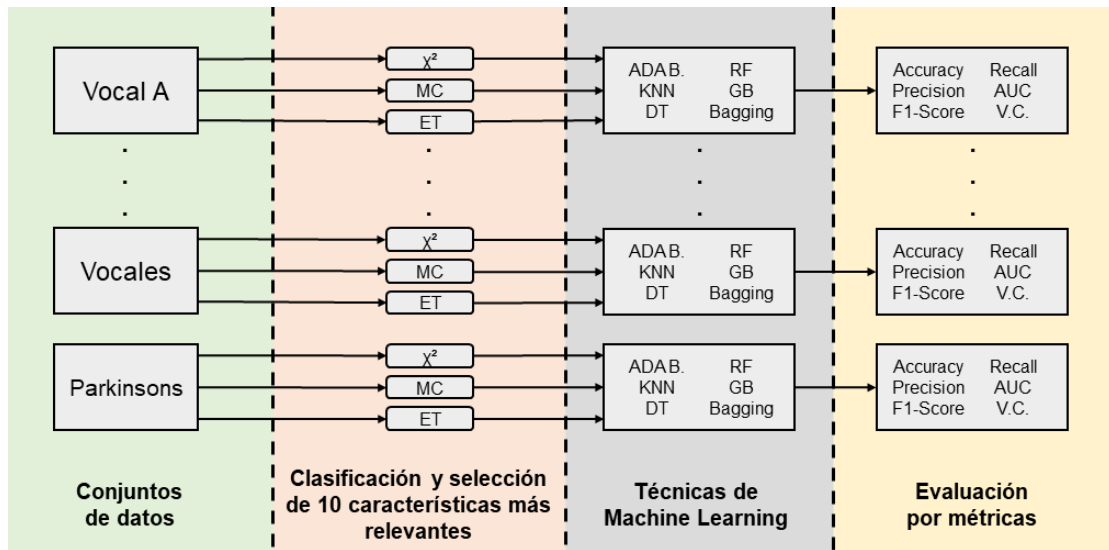
Para identificar el modelo con el mejor desempeño, se diseñó un sistema de pruebas estructurado que comenzó con la preparación de siete conjuntos de datos originales: Vocal A, Vocal E, Vocal I, Vocal O, Vocal U, todas las Vocales juntas y Parkinsons. A partir de estos conjuntos, se generaron tres subconjuntos por cada uno, seleccionando las 10 características más relevantes mediante los métodos de clasificación: Chi-Cuadrado (χ^2), Matriz de Correlación (MC) y *Extra Tree* (ET). Estos subconjuntos permitieron centrar el entrenamiento únicamente en las características más prometedoras, optimizando así la eficacia de los modelos (Yadav, 2023).

Posteriormente, los modelos de aprendizaje automático (*ADA Boost*, *Random Forest*, *Gradient Boosting*, *Bagging*, entre otros) se entrenaron utilizando los subconjuntos generados. Para garantizar la confiabilidad de los resultados, se aplicaron técnicas de validación cruzada con $K=10$ y se evaluaron los modelos con la métrica AUC-ROC. Este enfoque permitió analizar el impacto tanto de los métodos de selección de características como de los diferentes modelos sobre cada conjunto de datos.

La Figura 5 resume el flujo del proceso, mostrando cómo se crean los subconjuntos a partir de los conjuntos de datos originales, para que sean estos los que entrenen modelos de aprendizaje automático y se evalúen con las métricas de rendimiento.

Figura 5

Esquema de la secuencia de pruebas diseñada para identificar el modelo con mejor desempeño

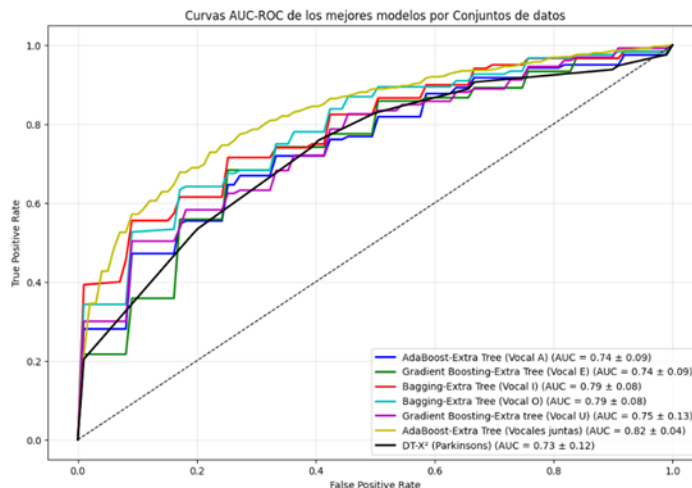


4.1. Resultados de los mejores modelos por conjuntos de datos

En la Figura 6 se presentan las curvas AUC-ROC correspondientes a los mejores modelos obtenidos para cada conjunto de datos. Estas curvas ilustran cómo varían las tasas de verdaderos positivos (*True Positive Rate*) frente a las tasas de falsos positivos (*False Positive Rate*) para cada modelo, proporcionando una visión integral de su capacidad discriminativa.

Figura 6

Representación de las curvas AUC-ROC de los mejores modelos obtenidos para cada conjunto de datos



Para evaluar el desempeño de los modelos de aprendizaje automático en cada conjunto de datos, se recopilaron los mejores resultados obtenidos. Este análisis permite identificar las combinaciones más efectivas de modelos y métodos de selección de características para cada conjunto de datos evaluado.

La Tabla 2 muestra los mejores resultados por conjuntos de datos. Los conjuntos de datos correspondientes a las vocales “A”, “E”, “I”, “O” y “U” alcanzaron los mejores resultados utilizando el método de selección *Extra Tree* (ET). En particular, para las vocales “A” y “E”, los modelos *AdaBoost* y *Gradient Boosting* lograron un AUC de 0.74 ± 0.09 , respectivamente. Para las vocales “I” y “O”, el modelo *Bagging* obtuvo un AUC de 0.79 ± 0.08 en ambos casos, mientras que para la vocal “U”, el mejor desempeño se observó con el modelo *Gradient Boosting*, alcanzando un AUC de 0.75 ± 0.13 .

Por otro lado, el conjunto de datos que combina todas las vocales juntas presentó el mejor resultado general, con un AUC de 0.82 ± 0.04 , utilizando el método ET y el modelo *AdaBoost*, lo que evidencia la efectividad de integrar características de múltiples vocales para optimizar el desempeño del modelo. Finalmente, para el conjunto de datos Parkinsons, el método Chi-Cuadrado (χ^2) y el modelo *Decision Tree* (DT) alcanzaron un AUC de 0.73 ± 0.12 . Estos resultados reflejan cómo el método de selección de características y el modelo de aprendizaje automático afectan el desempeño en cada conjunto de datos, evidenciando que la selección con *Extra Tree* (ET) y los modelos *AdaBoost*, *Gradient Boosting* y *Bagging* son particularmente efectivos en escenarios específicos.

Tabla 2

Recopilación de los mejores resultados por cada conjunto de datos

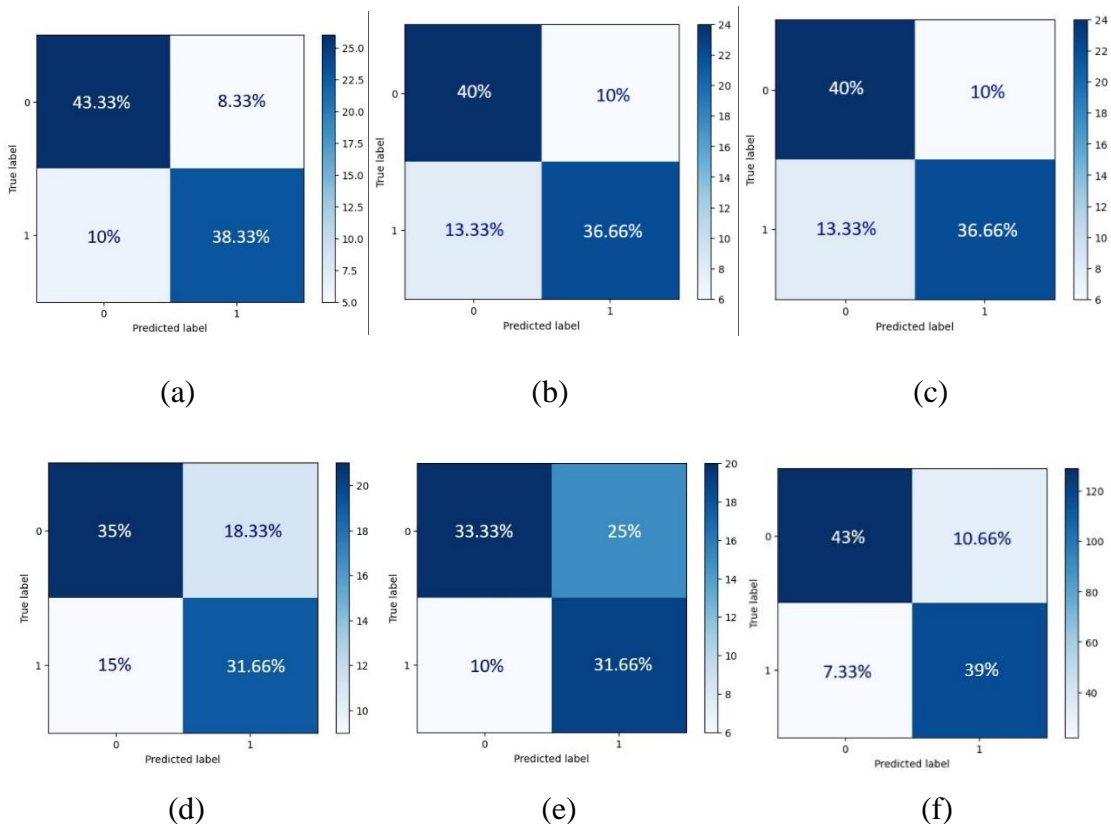
Datos	AUC	Clasificador	Modelo
Vocal A	0.74 ± 0.09	ET	AdaBoost
Vocal E	0.74 ± 0.09	ET	Gradient Boosting
Vocal I	0.79 ± 0.08	ET	Bagging
Vocal O	0.79 ± 0.08	ET	Bagging
Vocal U	0.75 ± 0.13	ET	Gradient Boosting
Vocales juntas	0.82 ± 0.04	ET	AdaBoost
Parkinsons	0.73 ± 0.12	X ²	DT

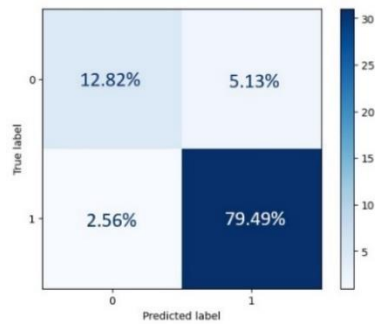
Las matrices de confusión de la Figura 7, muestran el desempeño de los mejores modelos seleccionados para cada conjunto de datos, evidenciando un comportamiento variable según el tipo de vocal o conjunto analizado.

Figura 7

Matrices de confusión correspondientes a los mejores modelos obtenidos para cada conjunto de datos: (a) Vocal A, (b) Vocal E, (c) Vocal I, (d) Vocal O, (e) Vocal U, (f) Vocales juntas, (g)

Parkinsons





(g)

En las vocales individuales, los modelos *AdaBoost*, *Gradient Boosting* y *Bagging* lograron resultados consistentes, aunque con variaciones en la proporción de verdaderos y falsos negativos. *Bagging* mostró un buen equilibrio en las vocales “I” y “O”, mientras que *Gradient Boosting* presentó más falsos negativos en las vocales “E” y “U”. El conjunto que combina todas las vocales obtuvo el mejor desempeño con *AdaBoost*, logrando una alta precisión y reduciendo significativamente los falsos negativos. Por otro lado, el conjunto *Parkinsons*, modelado con *Decision Tree* y Chi-Cuadrado (χ^2), mostró un desempeño aceptable, pero con mayores desafíos debido a su desbalance y complejidad.

5. Análisis de Resultados

Si bien las pruebas realizadas con el conjunto de datos de PARKINSONS mostraron métricas aceptables, una revisión más detallada de la matriz de confusión revela una importante limitación. El desbalance en los datos del conjunto provoca que el modelo tenga dificultades para reconocer adecuadamente a los pacientes con controles sanos. Esto genera un sesgo hacia la clasificación de pacientes con enfermedad de Parkinson (EP), lo que disminuye la confiabilidad del modelo para su aplicación en escenarios reales. Este problema pone de manifiesto la necesidad de abordar el desbalance de clases, ya sea mediante técnicas de preprocesamiento, como sobremuestreo o submuestreo, o ajustando los criterios de entrenamiento, para mejorar la capacidad del modelo en la detección de ambas clases de manera equitativa.

En todos los modelos evaluados, el clasificador *Extra Tree* destacó por obtener, en la mayoría de las pruebas realizadas, las métricas más altas en términos de rendimiento. Este comportamiento sugiere que el enfoque de *Extra Tree* para la selección de características es especialmente eficaz para identificar aquellas que tienen un mayor impacto en la clasificación de los datos. Su desempeño consistente puede atribuirse a su capacidad para manejar grandes volúmenes de datos, reducir el sobreajuste y captar relaciones no lineales entre las características. Por lo tanto, se posiciona como la opción más robusta y confiable en comparación con otros clasificadores utilizados en este proyecto, convirtiéndose en una herramienta clave para garantizar la precisión y la eficiencia de los modelos predictivos desarrollados.

Los modelos que demostraron un mejor rendimiento general fueron *AdaBoost* y *Bagging*, destacándose por obtener las mejores métricas en múltiples conjuntos de datos. *AdaBoost* sobresalió en el análisis de las vocales "A" y en el conjunto que combina todas las vocales,

consolidándose como una opción eficaz para la clasificación cuando se utilizan características combinadas. Por su parte, Bagging mostró un desempeño destacado en las vocales "I" y "O", evidenciando su capacidad para manejar características individuales. Estos conjuntos lograron obtener las métricas más altas y consistentes dentro del corpus PC-GITA, lo que los posiciona como las opciones más sólidas para el entrenamiento y evaluación de modelos predictivos en este proyecto.

6. Conclusiones

La selección de características relevantes utilizando métodos como *Extra Tree* permitió optimizar los modelos de aprendizaje automático al centrarse en las variables más importantes, mejorando así el rendimiento de los algoritmos.

Entre los modelos de aprendizaje automático evaluados, *Ada Boost* mostró el mejor rendimiento con un AUC-ROC de 0.82 ± 0.04 , especialmente cuando se aplicó al conjunto de datos de PC-GITA que combina todas las vocales. Esto resalta la capacidad de *Ada Boost* para manejar datos complejos y producir estimaciones consistentes en la detección de la enfermedad de Parkinson.

Al enfocarse en las alteraciones vocales como indicadores iniciales, este proyecto tiene el potencial de reducir significativamente los diagnósticos tardíos, mejorando la calidad de vida de los pacientes mediante diagnósticos más tempranos.

El desarrollo de herramientas basadas en inteligencia artificial para la detección de Parkinson tiene un impacto potencial en la reducción de costos asociados al diagnóstico, mejorando al mismo tiempo la accesibilidad y velocidad en la atención médica.

Los resultados obtenidos en este estudio abren la puerta a la implementación de modelos basados en aprendizaje automático en el ámbito clínico, ofreciendo un enfoque no invasivo y eficiente para apoyar el diagnóstico de la enfermedad de Parkinson. Asimismo, se sugiere explorar la combinación de características adicionales y técnicas más avanzadas para mejorar aún más la precisión de los modelos.

Referencias Bibliográficas

- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 175-185.
- Anisha, C. D. (2022). Tuned Homogenous Ensemble Regressor Model for Early Diagnosis of Parkinson Disorder Based on Voice Features Modality. *Journal of Artificial Intelligence and Capsule Networks*, 188-199.
- Bahaddad, A. R. (2022). Metaheuristics with deep learning-enabled Parkinson's disease diagnosis and classification model. *Journal of Healthcare Engineering*.
- Bang, Y. I. (2013). Acoustic characteristics of vowel sounds in patients with Parkinson disease. *NeuroRehabilitation*, 649-654.
- Boersma, P. &. (12 de 01 de 2025). *Praat: Doing phonetics by computer*. Obtenido de Praat: <https://www.fon.hum.uva.nl/praat/>
- de Lau, L. &. (2006). Epidemiology of Parkinson's disease. *The Lancet Neurology*, 525-535.
- Freund, Y. &. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 119-139.
- Li, D. L. (2011). A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artificial intelligence in medicine*, 45-52.
- Li, X. L. (2018). Human age prediction based on DNA methylation using a gradient boosting regressor. *Genes*, 424.
- Little, M. (12 de 01 de 2025). *Parkinsons [Dataset]*. Obtenido de UCI Machine Learning Repository: <https://archive.ics.uci.edu/dataset/174/parkinsons>

- Little, M. M. (2007). Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMed Eng.*
- Little, M. M. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nature Precedings.*
- Orozco-Arroyave, J. R.-L.-B.-R. (2014). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. *In Lrec*, 342-347.
- Pah, N. M. (2022). Phonemes based detection of parkinson's disease for telehealth applications. *Sci Rep.*
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.*
- Quinlan, J. (1986). Induction of Decision Trees. *Machine Learning*, 81-106.
- Sakar, C. &. (2010). Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of medical systems*, 591-599.
- Soke, F. G.-G. (2021). Task-oriented circuit training combined with aerobic training improves motor performance and balance in people with Parkinson's Disease. *Acta Neurologica Belgica*, 535-543.
- Sturges, H. A. (1926). The choice of a class interval. *Journal of the american statistical association*, 3-42.
- Tiwari, A. K. (2016). Machine learning based approaches for prediction of Parkinson's disease. *Machine Learning Application*, 33-39.

- Tsanas, A. L. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the royal society interface*, 842-855.
- World Health Organization. (2022). Parkinson disease: a public health approach. *World Health Organization*, 1-2.
- World Health Organization. (09 de 08 de 2023). *Enfermedad de Parkinson*. Obtenido de World Health Organization: <https://www.who.int/es/news-room/fact-sheets/detail/parkinsondisease>
- Wu, Y. C. (2017). Dysphonic voice pattern analysis of patients in Parkinson's disease using minimum interclass probability risk feature selection and bagging ensemble learning methods. *Computational and Mathematical Methods in Medicine*.
- Yadav, S. S. (2023). Artificial intelligence model for parkinson disease detection using machine learning algorithms. *Biomedical Materials & Devices*.

Apéndices

Apéndice A. Repositorio Github DEPAVUM-232

En el siguiente enlace se puede consultar gratuitamente el código utilizado para el desarrollo del Proyecto, publicado en Github.

<https://github.com/Luzdeyduque/DEPAVUM-232>

Ese repositorio dispone de un pequeño apartado para su comprensión y uso.