
Análisis sonoro del habla en pacientes con Parkinson: Identificación de parámetros y clasificación mediante machine learning

Estefany Eleuteria Batz Cantor



UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Análisis sonoro del habla en pacientes con Parkinson:
Identificación de parámetros y clasificación mediante machine
learning**

Trabajo de graduación presentado por Estefany Eleuteria Batz Cantor
para optar al grado académico de Licenciada en Ingeniería Biomédica

Guatemala,

2024

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Análisis sonoro del habla en pacientes con Parkinson:
Identificación de parámetros y clasificación mediante machine
learning**

Trabajo de graduación presentado por Estefany Eleuteria Batz Cantor
para optar al grado académico de Licenciada en Ingeniería Biomédica

Guatemala,

2024

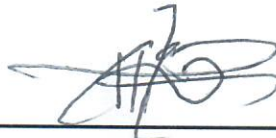
Vo.Bo.:



(f)

Dr.-Ing. José Andrés Leal Ordóñez

Tribunal Examinador:



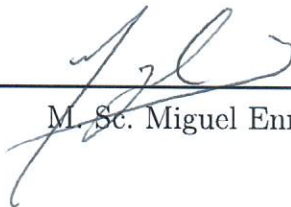
(f)

M. Sc. Carlos Esquit



(f)

Ph.D. Luis Alberto Rivera



(f)

M. Sc. Miguel Enrique Zea

Fecha de aprobación: Guatemala, 14 de Junio de 2024.

La Ingeniería Biomédica destaca por su naturaleza multidisciplinaria, la constante búsqueda de innovación, desarrollo e investigación para abordar desafíos en el campo de la salud y mejorar la calidad de vida. Esto fue lo que me motivo hace cinco años a estudiar la carrera y ahora, al final de mi formación profesional, a realizar el presente trabajo de graduación: *Análisis sonoro del habla en pacientes con Parkinson: Identificación de parámetros y clasificación mediante machine learning*, con el deseo de aportar y dar inicio al desarrollo de investigaciones enfocadas en la enfermedad de Parkinson en el Guatemala.

Quiero expresar mi profundo agradecimiento al Prof. Dr.-Ing. Juan Rafael Orozco Arroyave, quien desde el primer contacto demostró una amabilidad excepcional al ofrecerme su ayuda. También agradezco al Ph.D. Cristian David Ríos Urrego por su asesoría constante y dedicada durante todo el proceso de desarrollo de este trabajo, especialmente por compartir conmigo su invaluable experiencia y conocimiento en el área. Asimismo, deseo expresar mi agradecimiento a M.Sc. (Ing.) José Andrés Leal Ordoñez por su apoyo continuo como asesor a lo largo de la elaboración de este trabajo y por su destacada labor como catedrático durante toda la carrera.

Agradecimiento a Fundación Educación y al comité Local del Programa Nacional de Guatemala por haber visto potencial en mí y brindarme su apoyo para poder culminar mi formación profesional. Agradezco infinitamente a todas las personas que conocí durante estos años, en especial a mis amigos, quienes estuvieron siempre presentes e hicieron de este periodo algo muy divertido y memorable.

Y sobre todo, quiero agradecer a mi familia por su amor, por ser mi mayor fuente de apoyo y motivación durante mi carrera y vida. A mi madre, Ruth Eleuteria Cantor Agustín, quien desde el principio confió en mí y me impulsó a seguir adelante. A mi padre, Adolfo Sebastián Batz Rojas, le agradezco infinitamente por su constante apoyo y compañía. Y a mi hermano, José Adolfo Batz Cantor, por su incondicional respaldo en todo momento. Finalmente, quiero expresar mi gratitud a Dios, quien me ha acompañado en todo momento y ha sido mi fortaleza.

Prefacio	III
Lista de figuras	VIII
Lista de cuadros	IX
Resumen	X
Abstract	XI
1. Introducción	1
2. Antecedentes	2
3. Justificación	4
4. Objetivos	6
5. Alcance	7
6. Marco teórico	8
6.1. Parkinson	8
6.1.1. Definición y epidemiología	8
6.1.2. Causas y patogénesis	9
6.1.3. Métodos de diagnóstico preclínico y clínico	9
6.1.4. Síntomas motores y no motores	12
6.2. Análisis de señales de voz	13
6.2.1. Técnicas utilizadas para el análisis sonoro	14
6.2.2. Parámetros para el análisis acústico de prosodia de la voz	17
6.2.3. Procesamiento de señales sonoras	19
6.3. Aprendizaje automático y algoritmos de clasificación	21
6.3.1. Modelos de aprendizaje supervisado	21
6.3.2. Modelos de aprendizaje no supervisado	24
6.3.3. Modelos de aprendizaje por refuerzo	24

6.3.4.	Aprendizaje profundo - Deep learning	24
6.4.	Evaluación de rendimiento de algoritmos	25
6.4.1.	Modelos de matriz de confusión	25
6.4.2.	Métricas de evaluación	25
6.4.3.	Validación cruzada	26
6.4.4.	Curva ROC	26
7.	Metodología	28
7.1.	Procedimiento de grabación	29
7.2.	Base de datos PC-GITA	32
7.3.	Preprocesamiento de la base de datos	33
7.4.	Adquisición de parámetros.	34
7.4.1.	Implementación de algoritmos supervisados para la clasificación	37
7.4.2.	Modelo de Máquina de Vectores de Soporte	37
7.4.3.	Modelo de regresión logística	37
7.5.	Análisis multimodal de características prosódicas	37
7.5.1.	Evaluación integral de características prosódicas	38
7.5.2.	Análisis detallado de características prosódicas	38
7.5.3.	Evaluación de parámetros de frecuencia en semitonos	38
7.5.4.	Extracción y análisis de cromas	38
7.5.5.	Implementación de nuevas características y cambio dimensional en la frecuencia	39
8.	Resultados	40
8.1.	Evaluación de los métodos de adquisición de audio	40
8.2.	Análisis de tareas diadococinéticas	43
8.2.1.	Análisis de diadococinecia /KA/	43
8.2.2.	Análisis de diadococinecia /PA/	46
8.2.3.	Análisis de diadococinecia /TA/	49
8.2.4.	Análisis de diadococinecia /PAKATA/	52
8.2.5.	Análisis de diadococinecia /PATAKA/	54
8.2.6.	Análisis de diadococinecia /PETAKA/	57
8.2.7.	Resultados de las tareas diadococinecias con la implementación de cromas	59
9.	Discusión	62
10.	Conclusiones	64
11.	Recomendaciones	65
12.	Bibliografía	66
13.	Anexos	73

Lista de figuras

1.	Comparación de la substantia nigra en un cerebro sano y uno con Enfermedad de Parkinson [21]	9
2.	Visualización de la actividad dopaminérgica en el cerebro mediante DATscan en pacientes con un sistema de dopamina saludable (izquierda) y un sistema no saludable (derecha) [28]	10
3.	Identificación de medidas para calcular el Índice de Parkinson de Resonancia Magnética (MPRI) con ecografía transcraneal [30]	11
4.	Cronología de los síntomas clínicos de la Enfermedad de Parkinson separada en: tiempo previo al diagnóstico y posterior al diagnóstico [34]	12
5.	Principales subsistemas de producción del habla y su relación. [38]	14
6.	Forma de onda de una señal de audio (dominio del tiempo) a transformada de Fourier (dominio de la frecuencia) [43]	16
7.	Oscilograma y espectrogramas de las vocales en el idioma español [45]	17
8.	Escala cromática. A) Cromagrama de valores estándar de la escala tonal occidental. B) Señal de audio normalizada. C) Cromagrama obtenido de la señal de audio B. (Elaboración propia)	19
9.	Representación de hiperplano bidimensional [59]	22
10.	Modelo de clasificación K-NN binario [61]	23
11.	Representación de matriz de confusión para la clasificación de pacientes. (Sujetos sanos=0, Pacientes con EP=1).	25
12.	Curva de eficacia de diagnóstico (ROC) y área bajo la curva (AUC). [66]	27
13.	Diagrama esquemático de la metodología implementada.	29
14.	Ventana de Audacity: Configuración de audio	30
15.	Fotografía de sesión individual de grabación en el Laboratorio de Visualización Multimedia de la Universidad del Valle de Guatemala	31
16.	Organización jerárquica del conjunto de tareas diadococinéticas del dataset PC-GITA	33
17.	Proceso de normalización y eliminación. A) Señal de audio original. B) Señal de audio normalizada en una amplitud de [-1,1]. C) Señal sin elementos de frecuencia constante (DC)	34

18.	Análisis temporal de amplitud, frecuencia fundamental, energía y segmentos de duración de la diadicocinesia /PATAKA/. A) Paciente diagnosticado con Parkinson. B) Sujero sano	35
19.	Comparativa de señales de audio adquiridas con diferentes dispositivos. A) Audio adquirido con un micrófono profesional. B) Audio adquirido con el micrófono de una computadora. C) Audio adquirido mediante el micrófono de un celular	41
20.	Comparativa de energía de señales de audio adquiridas con diferentes dispositivos. A) Audio adquirido con un micrófono profesional. B) Audio adquirido con el micrófono de una computadora. C) Audio adquirido mediante el micrófono de un celular	42
21.	Evaluación de los dispositivos de grabación con base en: A) rango dinámico. B) SNR. C) THD y D) Energía	43
22.	Evaluación detallada de características de la tarea: KA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	44
23.	Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de KA	44
24.	Comparativa de medidas de rendimiento de la tareas: KA en diversas experimentaciones.	45
25.	Análisis del rendimiento del modelo de regresión logística para la tarea de KA con parámetros de frecuencia dimensionados en semitonos	45
26.	Principales coeficientes de la tarea KA con parámetros de frecuencia dimensionados en semitonos	46
27.	Evaluación detallada de características de la tarea: PA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	46
28.	Análisis del rendimiento del modelo de regresión logística para la tarea de PA, evaluando características de duración	47
29.	Comparativa de medidas de rendimiento de la tareas: PA en diversas experimentaciones.	47
30.	Principales coeficientes de la tarea PA con parámetros de frecuencia dimensionados en semitonos	48
31.	Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de PA con parámetros de frecuencia dimensionados en semitonos	49
32.	Evaluación detallada de características de la tarea: TA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	49
33.	Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de TA: Evaluando características de frecuencia	50
34.	Comparativa de medidas de rendimiento de la tareas: TA en diversas experimentaciones.	50
35.	Principales coeficientes de la tarea TA con parámetros de frecuencia dimensionados en hercios	51
36.	Análisis del rendimiento del modelo de regresión logística para la tarea de TA con parámetros de frecuencia dimensionados en hercios	51
37.	Evaluación detallada de características de la tarea: PAKATA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	52
38.	Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de PAKATA	52

39.	Comparativa de medidas de rendimiento de la tareas: PAKATA en diversas experimentaciones.	53
40.	Análisis del rendimiento del modelo de regresión logística para la tarea de PATAKA con parámetros de frecuencia dimensionados en hercios	53
41.	Principales coeficientes de la tarea PAKATA con parámetros de frecuencia dimensionados en hercios	54
42.	Evaluación detallada de características de la tarea: PATAKA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	55
43.	Análisis del rendimiento del modelo regresión logística para la tarea de PATAKA	55
44.	Comparativa de medidas de rendimiento de la tareas: PATAKA en diversas experimentaciones.	56
45.	Análisis del rendimiento del modelo de regresión logística para la tarea de PATAKA con parámetros de frecuencia dimensionados en hercios	56
46.	Principales coeficientes de la tarea PAKATA con parámetros de frecuencia dimensionados en hercios	57
47.	Evaluación detallada de características de la tarea: PETAKA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores	57
48.	Comparativa de medidas de rendimiento de la tareas: PETAKA en diversas experimentaciones.	58
49.	Principales coeficientes de la tarea PETAKA con parámetros de frecuencia dimensionados en hercios	59
50.	Gráfico de radar de evaluación de modelos: iniciales considerando características de frecuencia, energía y duración (1) y modelos con la implementación de cromas y cambio dimensional de frecuencia (2) para cada tarea diadocinética	61

Lista de cuadros

1.	Tabla de características prosódicas extraídas con DisVoice	36
2.	Resultados obtenidos	74
3.	Resultados de la primera experimentación con Algoritmo Máquina de Vectores de Soporte (SVM)	75
4.	Resultados experimentación con Algoritmo de Regresión Logística (RL) . . .	76

La Enfermedad de Parkinson es la segunda enfermedad neurodegenerativa más prevalente, afectando al 1% de la población mundial, unos 10 millones de personas. El diagnóstico temprano actualmente no es viable antes de que aparezcan síntomas motores avanzados, indicativos de una pérdida significativa de neuronas dopaminérgicas. Se necesita desarrollar herramientas no invasivas para detectarla de manera temprana, lo que podría permitir tratamientos oportunos y una vida más saludable para los pacientes.

El estudio analizó 600 archivos de audio .wav, divididos entre pacientes sanos y con Parkinson, en seis tareas diadococinesias. Se emplearon máquinas de soporte de vectores y regresión logística para entrenar modelos de aprendizaje supervisado, evaluando diferentes características acústicas y el impacto de incluir cromas. La adición de cromas mejoró significativamente el rendimiento de los modelos. Se compararon diversos dispositivos de grabación, sugiriendo que, en una cabina de grabación, un celular podría ser una alternativa al micrófono profesional. Sin embargo, los datos de audio de los sujetos de prueba no se incorporaron al conjunto de datos de entrenamiento debido a diferencias de edad.

Parkinson's disease is the second most prevalent neurodegenerative disease, affecting 1 % of the world's population, about 10 million people. Early diagnosis is currently not feasible before advanced motor symptoms appear, indicative of a significant loss of dopaminergic neurons. It is necessary to develop non-invasive tools to detect it early, which could allow timely treatments and a healthier life for patients.

The study analyzed 600 .wav audio files, divided between healthy patients and patients with Parkinson's, into six diadochokinesias tasks. Vector support and logistic regression machines were used to train supervised learning models, evaluating different acoustic characteristics and the impact of including chromas. The addition of chromas significantly improved the performance of the models. Various recording devices were compared, suggesting that, in a recording booth, a cell phone could be an alternative to the professional microphone. However, the audio data of the test subjects were not incorporated into the training data set due to age differences.

El desarrollo de herramientas para el diagnóstico preliminar de enfermedades neurodegenerativas es de suma importancia. La identificación temprana de estos trastornos permite atrasar su desarrollo y aumentar la calidad de vida de los pacientes. Una de las enfermedades neurodegenerativas con mayor prevalencia en los últimos años es la enfermedad de Parkinson (EP).

La enfermedad del Parkinson se caracteriza por la manifestación de síntomas motores, que afectan la cotidianidad de los pacientes. Sin embargo, la fase prodromica de la enfermedad inicia 10 años antes de los síntomas motores. Los pacientes desarrollan diversos síntomas en estas primeras etapas: trastornos del sueño, disfunciones autónomas, problemas gastrointestinales, síntomas sensoriales y trastornos en el habla (Disartria). La disartria es uno de los primeros síntomas motores en manifestarse ante la alteración en el control muscular de los órganos del habla. Los pacientes presentan alteraciones en la voz como el cambio en la intensidad vocal, alteraciones en la estabilidad, la manifestación de temblores y fluctuaciones en las cuerdas vocales. En consecuencia el habla se vuelve monótona alterando la acentuación y pronunciación de cada palabra.

Por ello, el estudio y análisis de la voz es de suma importancia. Este tipo de investigaciones permiten reconocer patrones y características que se ven afectadas por la enfermedad de Parkinson de forma accesible, no invasiva y flexible. El presente trabajo evalúa diversas características relacionadas con la prosodia del habla para el entrenamiento de modelos de clasificación en base a Machine Learning. Para ello se utiliza el set de datos PC GITA, elaborado por el Grupo de Investigación en Telecomunicaciones Aplicadas (GITA) de Medellín-Colombia. Se identifica la influencia de las características en el rendimiento de diversos modelos de aprendizaje supervisado.

El Parkinson es una enfermedad neurodegenerativa que se desarrolla rápidamente afectando a las neuronas productoras de dopamina en la sustancia negra del cerebro [1]. En los últimos 25 años la prevalencia de Parkinson ha aumentado, estimando que más de 8.5 millones de personas padecen esta enfermedad. Además de esto, las muertes relacionadas con Parkinson representan un número significativamente mayor que el de las muertes asociadas a otros trastorno neurológicos [2]. Por lo que actualmente existe un gran interés por mejorar el diagnóstico y tratamiento de la enfermedad [3].

El diagnóstico del Parkinson se realiza principalmente mediante una evaluación clínica realizada por un especialista en trastornos del movimiento y exámenes neurológicos detallados. Como apoyo para el diagnóstico clínico se realizan pruebas adicionales para observar diversas características. Las más comunes son pruebas olfativas, evaluaciones de gammagrafía cardíaca con 123-I-metayodobencilguanidina (MIBG), tomografía computarizada por emisión de fotón único y neuroimagen. [4]

En la búsqueda de nuevas tecnologías para el diagnóstico en fase temprana de Parkinson se han evaluado diferentes dominios de la enfermedad, principalmente los referentes a la parte motora y no motora. Para ello se han involucrado diversas disciplinas como el aprendizaje automático y tecnología del habla [5]. Las variaciones en el habla de los pacientes es uno de los parámetros más estudiados en la actualidad, ya que ha demostrado un gran potencial para el diagnóstico temprano de Parkinson. En 2019 se realizó la primera Evaluación Automática del Habla Parkinsoniana (AAPS por sus siglas en inglés). Taller en el que se discutieron los antecedentes en el campo de la evaluación automática del parkinson mediante el habla. Existe un gran interés por explotar este método de diagnóstico pre-clínico debido a su accesibilidad, no-invasividad, flexibilidad y rápido resultado [6].

En 2020 se evaluó el uso de modelos artificiales para estudiar el habla parkinsoniana y generar biomarcadores no invasivos utilizando dos modelos probabilísticos: Modelo de Mezcla Gaussiana - Modelo de Fondo Universal (GMM-UBM) y Análisis Discriminante Lineal Probabilístico (PLDA) [6]. En diversos estudios se han implementado modelos de aprendizaje automático supervisado, como por ejemplo Tsanas et. al., [7] donde se describe

el uso de dos clasificadores estadísticos: bosques aleatorios y máquinas de vector de soporte. Otra de las técnicas utilizadas para la predicción de Parkinson son las redes neuronales de multicapa, perceptrones multicapa y redes neuronales convolucionales (CNN). [8] [9]

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo que se asocia con complicaciones motoras y no motoras del paciente, donde la progresión de estos síntomas afecta directamente la calidad de vida y autonomía de los pacientes.[2] La EP es una de las enfermedades neurodegenerativas con mayor prevalencia a nivel mundial en la actualidad. Entre 1990 y 2015 la prevalencia del Parkinson se duplicó, afectando a 6.2 millones de pacientes y se estima que esta cifra aumente a 14.2 millones para 2040. En 2019 se registraron 329.000 muertes por Parkinson, un aumento del 100 % respecto a las cifras del año 2000. [10] [2]

La EP surge cuando las neuronas en los ganglios basales del cerebro mueren y por consiguiente la producción de dopamina disminuye. Una reducción en este neuromodulador conlleva a problemas en el control motriz del cuerpo. Además de esto, la EP reduce la producción de norepinefrina, lo que afecta directamente al sistema nervioso autónomo, dificultando la capacidad de regular el ritmo cardiaco, respiración, presión arterial y digestión. El motivo por el cual las neuronas mueren durante la EP es aún debatido, por lo que una cura a la enfermedad no existe y el diagnóstico temprano juega un papel importante en la progresión de la enfermedad. [11]

Actualmente el diagnóstico de la EP se realiza por medio de una evaluación clínica por parte de neurólogos con experiencia en trastornos del movimiento. En 2020 la Organización Mundial de la Salud realizó un informe sobre la enfermedad del Parkinson en el que se calculó que existen 0.03 neurólogos por cada 100.000 habitantes en países de bajos ingresos. Estimando que sólo el 23 % de los países de todo el mundo tienen acceso a neurólogos en las áreas rurales [2]. En Guatemala aproximadamente la mitad de la población vive en zonas rurales con un acceso limitado a atención médica.[12] Esto sumado a las limitaciones económicas, lingüísticas y de movilidad, justifica la necesidad de tener alternativas para apoyar el diagnóstico clínico de Parkinson con el objetivo de optimizar los recursos de pacientes y especialistas.

El criterio principal para diagnosticar el parkinsonismo causado por la EP es la presencia de bradicinesia, cuando la actividad motora del paciente disminuye progresivamente, y se

presenta al menos un temblor (movimiento involuntario) en reposo o rigidez [13]. En la actualidad no existe una cura para la EP, sin embargo se tienen diversos tratamientos médicos (e.g., tratamiento con dopamina) y neuroquirúrgicos (e.g., estimulación cerebral profunda) que ayudan a reducir los síntomas clínicos y mejorar la calidad de vida de los pacientes. Por lo que tener acceso a un diagnóstico temprano permite a los pacientes controlar los síntomas y retardar las afecciones motoras de la enfermedad [14]

Estudios recientes han evidenciado la relación entre los cambios de voz respecto a su calidad, volumen y tono debido a la rigidez y bradicinesia causada por el Parkinson. Estos cambios se han estudiado fisiológica y anatómicamente mediante evaluaciones y pruebas diagnósticas de la laringe y vías respiratorias. Siendo un potencial biomarcador para el diagnóstico temprano y progresivo de la enfermedad. [15] [16] [17]. El análisis del reconocimiento del habla utilizando tecnologías de aprendizaje automático (e.g., machine learning, inteligencia artificial) provee un método no invasivo, accesible y con alta versatilidad que puede apoyar al diagnóstico clínico de la enfermedad de Parkinson. Al necesitar únicamente la grabación de una fonación, repetición de sílabas o la lectura de un texto es un método accesible y de bajo costo para un pre-diagnóstico. Este tipo de diagnóstico representa una alternativa eficiente y rápida para la detección temprana del Parkinson, lo que puede acelerar el tratamiento de la enfermedad y aumentar los años de vida sana en los pacientes. Además de esto, puede romper las limitaciones lingüísticas en poblaciones con idiomas nativos como sucede en Guatemala y proveer acceso a diagnóstico remoto en áreas rurales

4.1 Objetivo general

Realizar un análisis sonoro basado en los patrones en el cambio del habla para la clasificación de pacientes con la Enfermedad de Parkinson y pacientes sanos utilizando Machine Learning.

4.2 Objetivos específicos

- Comparar los parámetros de calidad de audio entre distintos dispositivos para la adquisición de audio en un entorno controlado.
- Evaluar el rendimiento de modelos de aprendizaje supervisado, como máquinas de soporte de vectores y regresión logística, en la clasificación de pacientes sanos y con Parkinson utilizando características de prosodia extraídas del dataset PC-GITA.
- Analizar el impacto de diferentes combinaciones de características de prosodia en la capacidad de modelos de aprendizaje supervisado para distinguir entre los grupos de pacientes con enfermedad de Parkinson y sujetos sanos.
- Analizar el impacto de la inclusión de cromas en el conjunto de características en el rendimiento de los modelos de aprendizaje supervisado para la clasificación de pacientes con Parkinson.

El análisis detallado de muestras de audio de pacientes con Parkinson se realiza con el objetivo de contribuir al diagnóstico clínico de la enfermedad de forma temprana para acelerar el tratamiento de la enfermedad y dar al paciente una mejor calidad de vida. Este análisis tiene el potencial de contribuir a la asistencia médica temprana de los pacientes independientemente de su ubicación. En este trabajo se analizan e identifican las características de prosodia (cómo se pronuncian las palabras y las frases, incluyendo el tono de voz, el ritmo y la entonación), que tienen mayor influencia en el reconocimiento y correcta clasificación de pacientes con EP.

La implementación de los algoritmos de clasificación de pacientes con EP considera la base de datos PC GITA, obtenida del Grupo de Investigación en Telecomunicaciones Aplicadas (GI-TA) de la Universidad de Antioquia (Medellín-Colombia); conjunto de 600 grabaciones, clasificadas en seis tareas diadococinéticas (/pataka/, /petaka/, /pakata/, /pa/, /ta/ y /ka/) distribuido equitativamente entre pacientes diagnosticados con Parkinson y sujetos sanos, con 100 grabaciones por tarea. La parametrización de las señales se realiza utilizando el framework DisVoice [18][19] obteniendo parámetros de frecuencia, energía y duración.

Se transforma la dimensional de los parámetros de frecuencia de Hz a semitonos. Se añaden nuevas características que evalúan los perfiles de 12 clase de tonos denominados cromas, para verificar cuál de estos presenta un mejor desempeño, según la tarea analizada, en base a los resultados de métricas estándar de medidas de desempeño para modelos de clasificación supervisados.

Se evalúan diversos dispositivos de grabación: un micrófono profesional, el integrado en una computadora y un celular, mediante la medición del rango dinámico, la relación señal-ruido (SNR) y la distorsión armónica total (THD) al grabar a sujetos sanos en un entorno controlado. En base a los resultados se determina qué medio de grabación es más efectivo para la obtención de audio, para su consideración en futuros trabajos que desarrollen un conjunto de datos propios de Guatemala.

6.1. Parkinson

6.1.1. Definición y epidemiología

La enfermedad del Parkinson (EP) es un trastorno neurodegenerativo crónico y progresivo, actualmente es la segunda enfermedad neurodegenerativa más común, después del alzheimer. 14 de cada 100.000 personas en toda la población padecen de Parkinson, y 160 de cada 100.000 personas mayores a 65 años. Con una frecuencia de 2% en hombres y 1.3% en mujeres. [20]

El Parkinson afecta principalmente la producción de dopamina y en consecuencia origina la pérdida de células nerviosas de la sustancia negra (Figura 1) del cerebro provocando daños neuronales.[21] Esto causa afecciones en la motricidad de los pacientes y contribuye a la manifestación de bradicinesia (ralentización del movimiento o reducción progresiva de su velocidad), rigidez muscular, temblores involuntarios e inestabilidad en la postura. La enfermedad se manifiesta de forma heterogénea y los cambios provocados por la patología se extienden desde el hipotálamo, el tronco encefálico hasta la columna de células intermedio laterales de la médula espinal [22] [23]

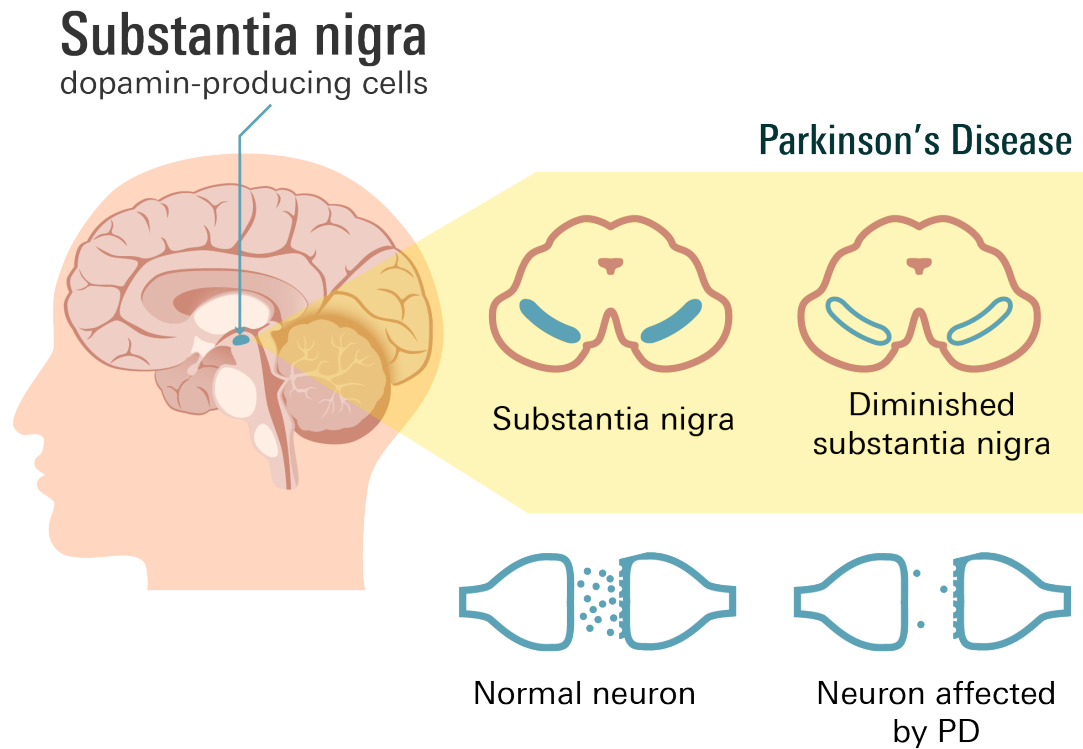


Figura 1: Comparación de la substantia nigra en un cerebro sano y uno con Enfermedad de Parkinson [21]

6.1.2. Causas y patogénesis

La enfermedad de Parkinson es un trastorno heterogéneo. Actualmente no se puede describir la etiología de la enfermedad y se desconoce la causa de la enfermedad, sin embargo se cree que la patogénesis se debe a la combinación de diversos factores genéticos y ambientales como lo son la exposición a pesticidas, herbicidas, metales pesados, el consumo de tabaco y cafeína; historial de lesiones cerebrales y un historial familiar con prevalencia de EP [24] [25].

6.1.3. Métodos de diagnóstico preclínico y clínico

No existe un examen específico para la detección de EP, el diagnóstico se basa en el historial clínico y evaluaciones físicas de los pacientes realizadas por un psiquiatra especializado en trastornos del movimiento utilizando la Escala Unificada de Calificación de la Enfermedad de Parkinson de la Sociedad de Trastornos del Movimiento (MDS-UPDRS). [22] Por lo que, el diagnóstico es principalmente clínico, usualmente cuando los pacientes reciben el diagnóstico ya han perdido una parte sustancial, entre el 50 y 80%, de los dopaminérgicos neuronales en la sustancia nigra pars compact, (SNpc). Por ello actualmente existen investigaciones que buscan poder realizar un diagnóstico diferencial de la enfermedad. [24]

[26]. En la actualidad la atención se ha dirigido hacia el diagnóstico preciso o prodrómico de la enfermedad. La fase prodrómica de la enfermedad corresponde al período de 10 años antes de la manifestación de síntomas motores en los pacientes. [27] Las metodologías en estudio para desarrollar el diagnóstico precoz de la Enfermedad de Parkinson son: imágenes dopaminérgicas, escáner de transporte de dopamina, ecografía transcraneal y la evaluación de cambios en los niveles de neuroquinas endógenas.

Imágenes preclínicas de diagnóstico

Las imágenes dopaminérgicas tienen un gran potencial como biomarcadores para la enfermedad. La modalidad más estudiada es la tomografía por emisión de positrones fluoro-3,4-dihidroxifenilalanina, una técnica altamente sensible y específica para el diagnóstico de la enfermedad. Sin embargo, la modalidad más usada y comercial es el escáner de transporte de dopamina (DAT scan) que permite identificar cambios en la distribución y densidad de los transportadores de dopamina en el cerebro. como se muestra en la Figura 2 Algunas desventajas de estas técnicas son el costo, disponibilidad de equipos y el uso de radioligandos. [28]

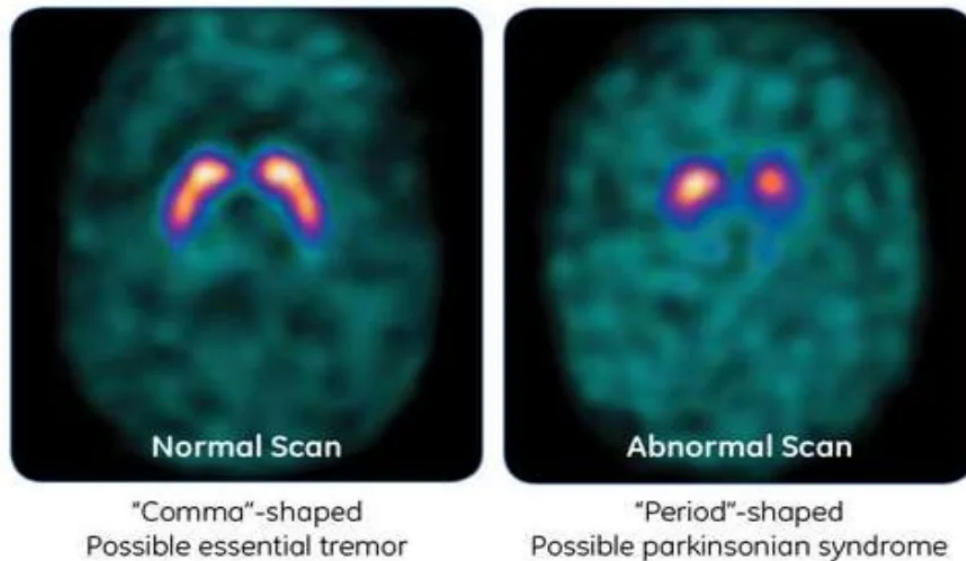


Figura 2: Visualización de la actividad dopaminérgica en el cerebro mediante DATscan en pacientes con un sistema de dopamina saludable (izquierda) y un sistema no saludable (derecha) [28]

Otra potencial herramienta es la ecografía transcraneal, al ser un método no invasivo permite la visualización de la sustancia negra a través de ventanas de hueso temporal. Algunos pacientes no poseen un hueso temporal adecuado para la visualización de las estructuras necesarias, lo que limita el uso de la técnica. [29] Utilizando la resonancia magnética es posible una visualización de diferencias en la sustancia negra lateral y del locus coeruleus para pacientes con EP. Con base en estas imágenes se determina el Índice de Parkinson de Resonancia Magnética (MRPI) que se calcula utilizando el área de la protuberancia (P) y

mesencéfalo (M) en el plano sagital medio y la anchura media de los pedúnculos cerebelosos medios (MCP) y superiores (SCP) [30]

$$MRPI = \left(\frac{P}{M} \right) \cdot \left(\frac{MCP}{SCP} \right) \quad (1)$$

Donde al obtener un índice mayor a 13.55 se sospecha de un trastorno de movimiento como lo es la EP.

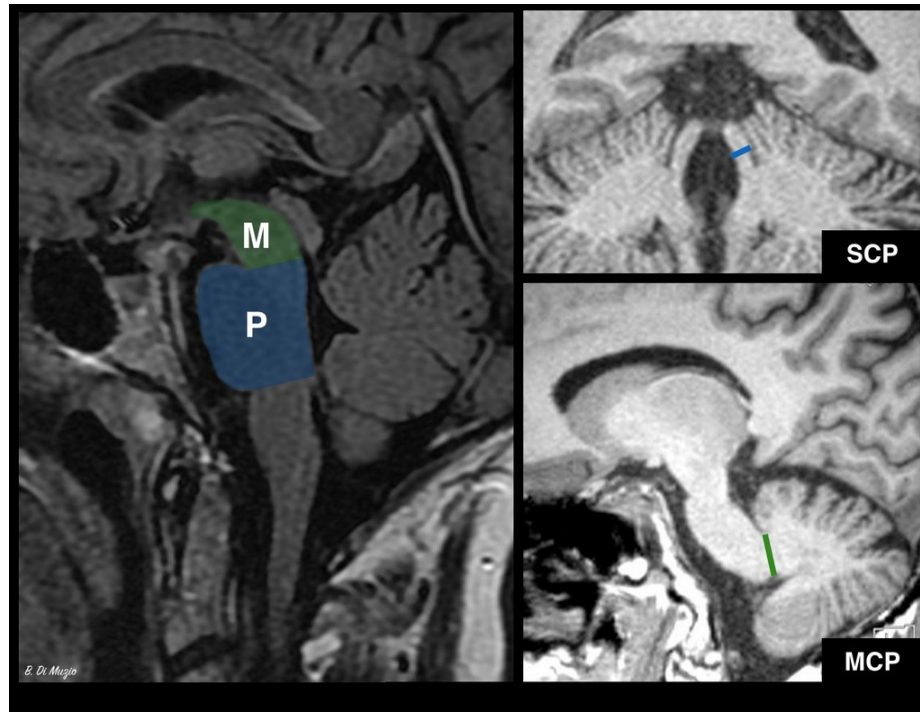


Figura 3: Identificación de medidas para calcular el Índice de Parkinson de Resonancia Magnética (MPRI) con ecografía transcraneal [30]

Reconocimiento de biomarcadores bioquímicos

Los cambios en los niveles de neurotoxinas endógenas, metabolitos, lípidos, citoquinas inflamatorias, microARN y anticuerpos específicos en fluidos corporales, específicamente en el líquido cerebroespinal y la sangre, son indicadores de EP [27]. Los principales biomarcadores en estudio son la alfa-sinucleína, esta proteína se acumula en exceso en las neuronas durante la EP y es el componente principal de los cuerpos de Lewy asociados a la EP [26]; las enzimas lisosomales que poseen un papel importante en el mantenimiento de la homeostasis neuronal; marcadores de la patología de amiloide y tau, y la cadena ligera de neurofilamentos. [31]

6.1.4. Síntomas motores y no motores

Los síntomas de la enfermedad se dividen en dos grupos generales, los síntomas motores y no motores. Más del 60 % de los pacientes sufren de algún síntoma no motor previo a la manifestación de los síntomas motores característicos de la enfermedad [22].

Existen diversas subdivisiones de los síntomas, La Figura 4 muestra el desarrollo de las síntomas de la enfermedad a partir de la etapa de prediagnóstico, 20 años antes del desarrollo de los primeros síntomas motores. Inicialmente el paciente presenta síntomas que incluyen estreñimiento, anosmia, trastorno REM del sueño y depresión; En la etapa temprana de la enfermedad el paciente desarrolla los primeros síntomas motores: temblor, rigidez y bradicinesia. En el inicio de la etapa avanzada de la enfermedad se presentan síntomas autonómicos como náuseas, vómitos, babeo, incontinencia, disfunción eréctil y olfativa; anomalías sensoriales como anosmia, dolor, ageusia y parestesia. En la etapa más avanzada de la enfermedad los pacientes desarrolla síntomas no motores neuro psiquiátricos como ansiedad, apatía, trastorno de control de impulsos, psicosis, alucinaciones, abulia y ataques de pánico; síntomas cognitivos que incluyen la demencia y pérdida de memoria. [32] [33] [26] [34]

En la enfermedad la aparición de síntomas motores es la característica principal del inicio de la etapa temprana para el diagnóstico. Los principales síntomas son la bradicinesia, que define la dificultad y lentitud para iniciar movimientos acompañada con una disminución en la velocidad de respuesta de diversas acciones repetitivas; la rigidez y temblor en reposo con una frecuencia entre los 4 y 6 Hz; la inestabilidad en la postura; los trastornos en la marcha y la disartria hipocinética. [26][32]

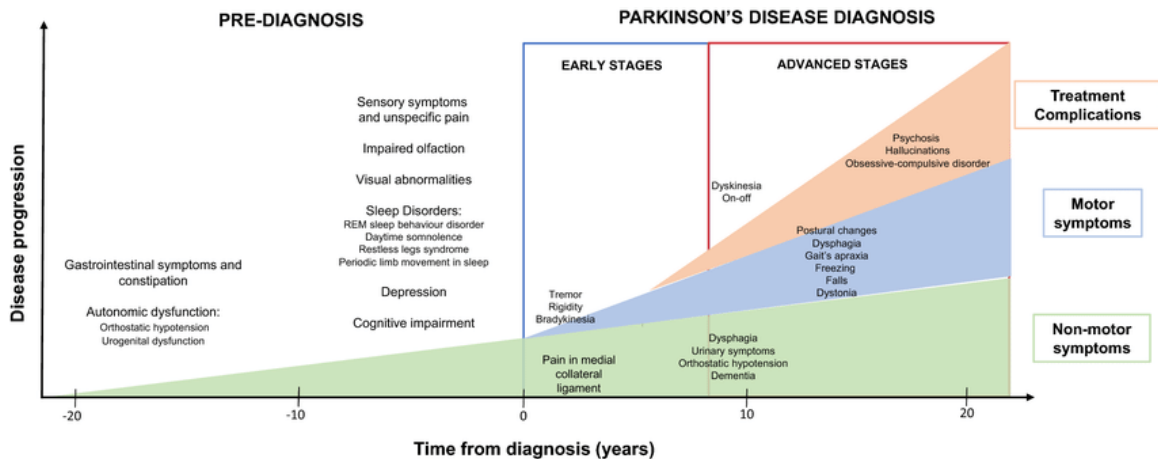


Figura 4: Cronología de los síntomas clínicos de la Enfermedad de Parkinson separada en: tiempo previo al diagnóstico y posterior al diagnóstico [34]

Disartria hipocinética

La disartria es una alteración de origen neurológico en el sistema nervioso que afecta el control, movilidad, velocidad, fuerza, tono y precisión de los movimientos de los órganos musculares encargados del mecanismos del habla. Afecta los procesos de respiración, fona-

ción, resonancia y prosodia. La respiración se vuelve superficial y poco funcional, se tienen alteraciones en la fase inspiratoria y espiratoria, reduciendo el flujo de aire previo a la fonación. En la fonación, no se produce un cierre completo de las cuerdas vocales debido al poco control y flexibilidad laríngea generando una voz débil, apagada, ronca y temblorosa. Por lo que a lo largo del habla la intensidad disminuye y se tiene una pérdida progresiva del volumen de la voz. La articulación se vuelve imprecisa; la coordinación, control y precisión de los movimientos de la mandíbula, labios, lengua y paladar disminuye y se generan temblores en la mandíbula al estar en reposo. Respecto a la prosodia, el habla se caracteriza por ser monótona, con una reducción significativa en la acentuación y entonación. El discurso tiene un ritmo variable de lento a excesivamente rápido al final de la frase [35]. Estos problemas de voz son característicos en el 90 % de las personas con EP. [36] [37].

6.2. Análisis de señales de voz

El proceso del habla se logra a través de la coordinación de 5 subsistemas: respiración, fonación, resonancia, articulación y prosodia. La coordinación precisa de estos subsistemas produce sonidos con un significado interpretable (Figura 5).

El origen de la voz se da con base en la respiración, con la producción de flujo de aire en los pulmones. El aire interactúa con las cuerdas vocales generando fonaciones, que se refiere a las vibraciones que conducen a sonidos propios de la voz. En reposo las cuerdas vocales se separan y permiten el paso del aire a través de la tráquea, al hablar se da la actividad muscular que tensa y relaja las cuerdas vocales. Para producir el sonido, se lleva a cabo un proceso de resonancia en el cual la presión del aire debajo de la laringe aumenta lo suficiente como para que las cuerdas vocales vibren; que en conjunto con la articulación, que implica la posición de los labios, dientes y lengua, se modifica el sonido para generar diferentes resonancias y tonalidades. [15] [38] Finalmente la prosodia completa el proceso del habla, aportando significado y expresividad mediante la acentuación y pronunciación correcta de cada palabra.

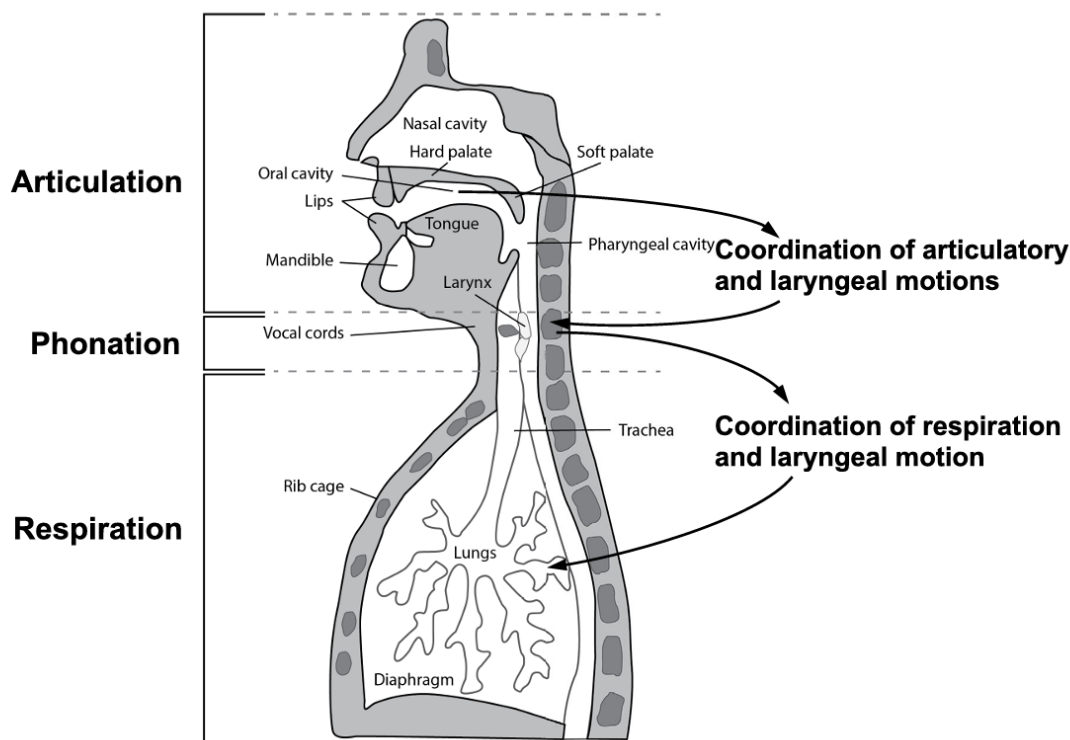


Figura 5: Principales subsistemas de producción del habla y su relación. [38]

La fonética acústica estudia las propiedades físicas de la voz. El objetivo es analizar las señales sonoras que se dan en el habla con características cuantificables como frecuencia, amplitud y duración. [39] La fonética acústica se utiliza en diversos campos, como la fonética, la lingüística, la patología del habla y el procesamiento del habla. Algunas aplicaciones comunes incluyen la detección de trastornos del habla, la evaluación de la calidad vocal, la identificación biométrica por voz y el estudio de las características vocales en distintos grupos de hablantes

En pacientes con EP es recurrente el desarrollo de disartria hipocinética, trastorno de la expresión verbal derivado de la alteración en el control muscular de los órganos del habla, en etapas avanzadas de la enfermedad es notable la alteración en el habla sienta lenta, débil y monótona. Inicialmente se dan cambios en el sistema respiratorio de los pacientes, lo que afecta la intensidad vocal; los cambios en el sistema fonador afectan el ritmo de vibración en las cuerdas vocales, alterando la estabilidad vocal con la presencia de temblores o fluctuaciones; La resonancia vocal de los pacientes disminuye debido a la falta de control muscular. En conjunto con estos cambios la prosodia de los pacientes resulta monótona afectando el ritmo, energía y frecuencia de las entonaciones. [40]

6.2.1. Técnicas utilizadas para el análisis sonoro

Existen diferentes técnicas para el análisis de una señal sonora. El contenido espectral de una señal es tan importante como el curso temporal de la intensidad y energía, así como

el análisis de la evolución del tiempo y frecuencia a lo largo de la señal.

Análisis en el dominio del tiempo y frecuencia

En el dominio del tiempo se observan los cambios de una señal durante determinado periodo de tiempo. Para transformar una señal en el dominio del tiempo a otra en el dominio de la frecuencia se hace uso de la transformada de Fourier (FT) cuyo principio es que todas las funciones periódicas pueden ser representadas mediante una combinación de sinusoides. [41]

$$F(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\omega_T t) + b_n \sin(n\omega_T t)) \quad (2)$$

Para utilizar la transformada de Fourier en una señal discreta se hace uso de la Transformada de Fourier Discreta (DFT). Esta se obtiene mediante la expresión

$$x_k = x(t_k) = \frac{a_0}{2} + \sum_{i=1}^{\frac{N}{2}} \left(a_i \cos\left(\frac{2\pi i t_k}{T}\right) + b_i \sin\left(\frac{2\pi i t_k}{T}\right) \right) \quad (3)$$

donde:

$$a_0 = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} x_k$$

$$a_i = \frac{1}{N} \sum_{i=1}^N x_k \cos\left(\frac{2\pi i t_k}{N}\right)$$

$$b_i = \frac{1}{N} \sum_{i=1}^N x_k \sin\left(\frac{2\pi i t_k}{N}\right)$$

al tener N muestras dicho sistema de ecuaciones se representa en la forma matricial X.

$$X = C \cdot a \implies a = C^{-1} X \quad (4)$$

Siendo X el vector de muestras, a los coeficientes por determinar y C la matriz de armónicos. Al invertir la matriz de armónicos C se obtiene la Transformada Rápida de Fourier, FFT por sus siglas en inglés. Es un algoritmo utilizado para determinar la FT en señales digitales. El análisis espectral se obtiene del uso de la FFT que descompone la señal en sus frecuencias constituyentes. Con lo que es posible identificar frecuencias dominantes y destacar las diferencias de la señal con base en sus características espectrales, que dan información de la energía promedio. [42][43](Figura 6)

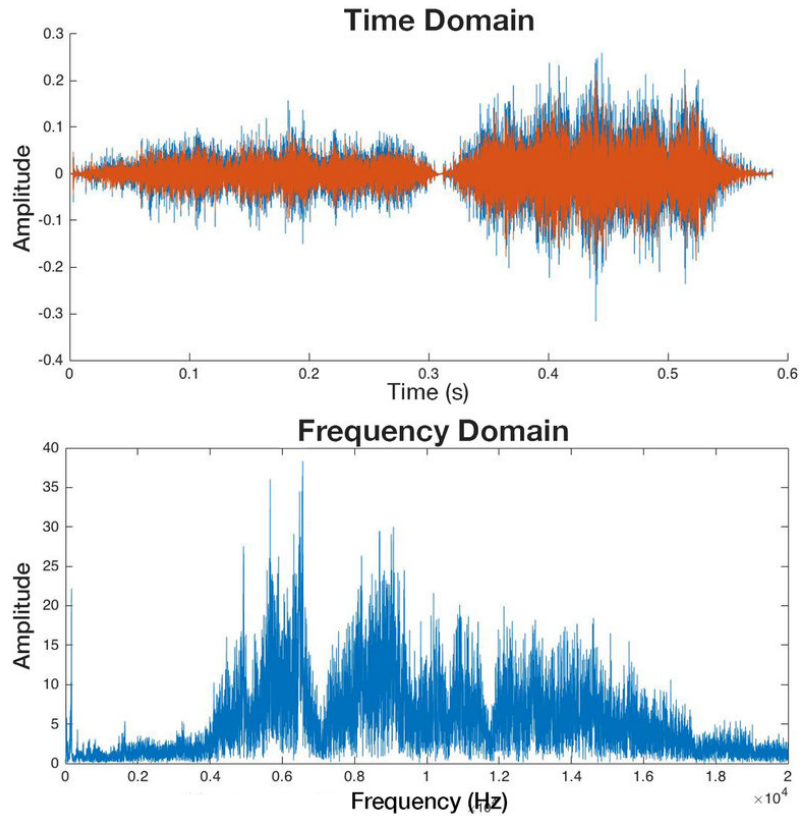


Figura 6: Forma de onda de una señal de audio (dominio del tiempo) a transformada de Fourier (dominio de la frecuencia) [43]

Técnicas de tiempo-frecuencia y tiempo-escala

El análisis de tiempo-frecuencia da un panorama completo respecto a la distribución total de energía de la señal. Para realizar este análisis existen diferentes métodos como la distribución Wigner-Ville, una distribución de tipo no lineal; la transformada de Fourier con ventanas deslizantes; la transformada de Gabor y la transformada wavelet. [42] Métodos como la transformada de Fourier en tiempo corto (STFT), la transformada Wavelet y la distribución de Wigner Ville permiten generar espectrogramas.

Los espectrogramas dan una visualización completa del contenido espectral de la señal a lo largo del tiempo, esta representación es un híbrido entre una gráfica FFT en dominio de la frecuencia y un gráfico en el dominio del tiempo. El espectrograma muestra el tiempo en su eje horizontal, la frecuencia en el eje vertical y la amplitud mediante una escala de grises o un mapa de calor (Figura 7) [44] [45]

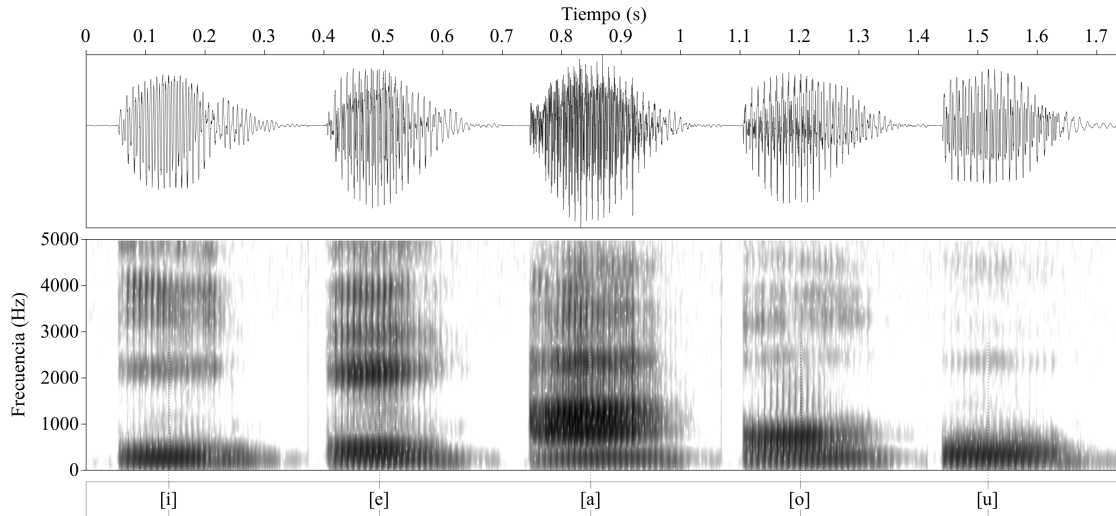


Figura 7: Oscilograma y espectrogramas de las vocales en el idioma español [45]

Evaluaciones de diadococinesia del habla (DDK)

La diadococineasia es la facultad de ejecutar rápida y sucesivamente determinadas tareas con el objetivo de evaluar la coordinación motora. En el habla estas tareas se basan en la producción silábica alternada o secuencial de sonidos de las consonantes oclusivas sordas /p/, /t/, y /k/ en sus combinaciones /pa/, /ta/, /ka/, /pataka/, /pakata/ y /petaka/. [46]

6.2.2. Parámetros para el análisis acústico de prosodia de la voz

El análisis acústico cuantifica y analiza parámetros específicos de la voz. Los parámetros evaluados se basan en el subsistema del habla a analizar. Los parámetros más utilizados son la frecuencia fundamental, fluctuación, brillo (shimmer) y relación armonía-ruido. No obstante, el enfoque de evaluación de prosodia se basa en los parámetros de frecuencia, entonación, estrés de cadencia y la duración asociada a las pausas y segmentos de sonido.

Frecuencia fundamental (F0)

Se define como el número de veces que una onda sonora se repite durante un tiempo determinado (Hz) donde cada unidad equivale a un ciclo por segundo. En el análisis del habla la frecuencia describe la periodicidad de la vibración de las cuerdas vocales, relacionado con los ciclos de tensión y relajación de la glotis. [39] [47]

La evaluación de la frecuencia no es lineal, ya que existen diferencias significativas en el análisis de voces con tonos más altos (como la voz de mujeres y niños) contra voces más graves (la voz de un hombre). Generalmente las medidas de frecuencia de voces femeninas tienen mayor valor, al tener un rango tonal más alto y amplio en comparación a la voz de un hombre. En consecuencia, el análisis de la amplitud del rango de tono de los hablantes

se realiza en escalas logarítmicas como los semitonos y octavas.[48]. La frecuencia en Hz se expresa como una tonalidad musical utilizando la siguiente fórmula:

$$\text{semitonos} = \log_2 \left(\frac{\text{frequency}}{\text{reference}} \right) \times 12 + 57 \quad (5)$$

donde *reference* es el tono de referencia, generalmente 440 Hz equivalentes a A4 en el sistema de notación musical occidental.

Energía

La energía hace referencia a la intensidad que genera una señal en un periodo de tiempo específico. Esta medida se relaciona con la amplitud de la señal de voz. La medida de energía para cuantificar la intensidad de la señal es decibelios (dB), una escala logarítmica que compara la potencia de la señal con un referente específico. [49]

$$\text{dB} = 10 \log_{10} \left(\frac{I}{I_0} \right) \quad (6)$$

Donde:

I es la intensidad del sonido en watts por metro cuadrado.

I_0 es la intensidad de referencia, que es típicamente 10^{-12} watts por metro cuadrado, que es el umbral de audición humano.

Duración

La duración en el contexto del análisis del habla, evalúa el tiempo que transcurre entre eventos sonoros como: segmentos de habla, silencios y pausas en la señal; la velocidad y ritmo del habla. Los segmentos sonoros son todas aquellas muestras donde se producen sonidos vocales, las pausas son las interrupciones intencionadas en el habla y los silencios son aquellos periodos de tiempo en los que no se producen sonidos vocales. La velocidad del habla se define como la duración total de una oración, medida en segmentos por segundo.[49]

Escala cromática

Los cromas son una representación de la distribución de energía espectral en una señal de audio. La escala cromática es el vectores de características de 12 elementos que representan la energía de cada clase de tono musical, descompuesto en cromas correspondientes a las octavas utilizadas en la música occidental. $[C, C\#, D, D\#, E, \dots, B]$ [50]

La comprensión la escala cromática se visualiza representando los doce elementos de la energía espectral en un cromagrama (Figura 8, elaboración propia). En esta representación, compartimento corresponde a uno de los doce tonos de la escala cromática, ordenados de

forma ascendente según su frecuencia fundamental. Al utilizar esta representación visual, se puede apreciar cómo la energía espectral se distribuye a lo largo de los doce tonos.

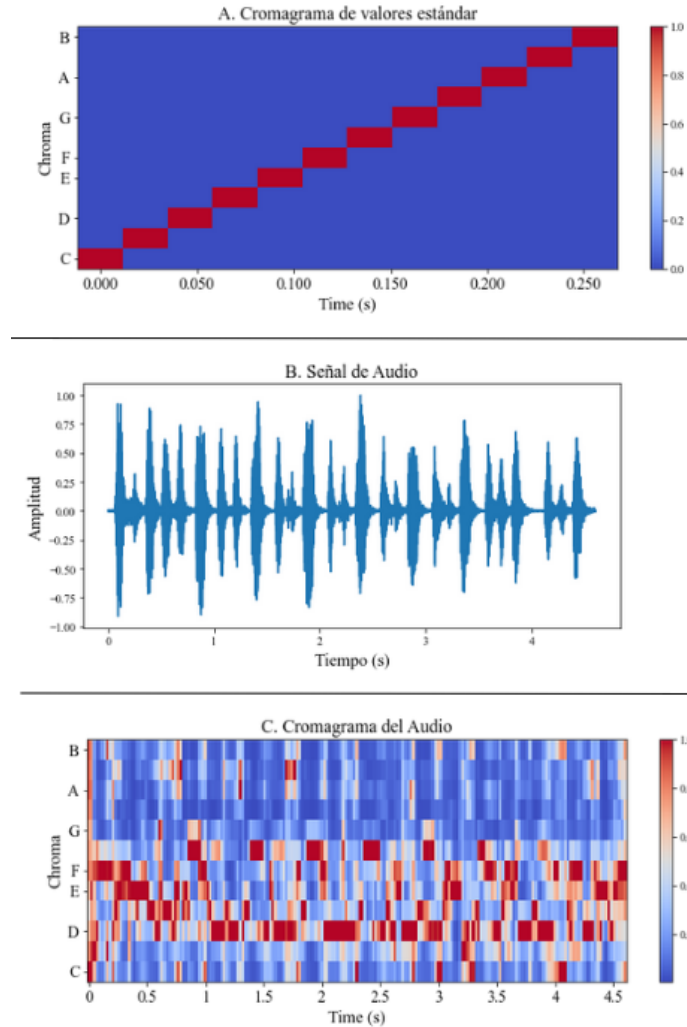


Figura 8: Escala cromática. A) Cromagrama de valores estándar de la escala tonal occidental. B) Señal de audio normalizada. C) Cromagrama obtenido de la señal de audio B. (Elaboración propia)

6.2.3. Procesamiento de señales sonoras

Las ondas sonoras se convierten en señales eléctricas haciendo uso de micrófonos. La señal obtenida del micrófono es una señal analógica cuya amplitud varía de manera continua en el tiempo. Para convertir la señal analógica en una señal digital es necesario realizar un proceso de digitalización, que implica el muestreo, cuantificación y codificación de la señal. [51] Para la digitalización de una señal se toman muestras discretas en intervalos regulares de la señal analógica. Esta etapa de muestreo determina la fidelidad y calidad de la muestra digital del audio. La frecuencia de muestreo es el número de muestras tomadas por segundo, la elección de la frecuencia de muestreo es crucial para evitar la pérdida de información [52]. Con base en el teorema de Nyquist-Shannon la frecuencia de muestreo debe ser al menos el

doble de la frecuencia más alta en la señal analógica para evitar fenómenos como aliasing.

$$f_{\text{Nyquist}} = f_s = 2 \cdot f_{\text{max}} \quad (7)$$

Para la cuantificación se determina cuánta información se captura en cada muestra mediante el bit depth, número de bits, utilizado en cada muestra. La amplitud de cada muestra de la señal original se traduce a valores discretos enteros expresados en bits. Esto tiene un impacto en el rango dinámico de la señal digitalizada, errores en la cuantificación pueden dar lugar a distorsión en la señal. [51]

Normalización

La normalización ajusta el nivel de amplitud de la señal de audio para que esta se encuentre en el rango determinado, asegurando una amplitud relativa constante entre todas las muestras, independientemente de la tarea. Esto se realiza tomando el valor máximo absoluto de la señal original y multiplicando por el valor máximo deseado:

$$x_{\text{norm}} = \frac{x}{\text{máx}(|x|)} \times \text{max_value} \quad (8)$$

donde:

X_{norm} es el valor normalizado de la muestra

x es el valor original de la muestra

$\text{máx}(|x|)$ es el valor máximo absoluto encontrado en la señal original

La normalización evita la saturación en las muestras y asegura que las señales analizadas tengan homogeneidad en los límites. [53]

Eliminación de desplazamiento de corriente continua de una grabación (DC Offset)

El desplazamiento de corriente continua (DC) es el desplazamiento de la amplitud media desde cero. Esto representa una potencial pérdida de volumen de audio o distorsión en la muestra. En la mayoría de casos este desplazamiento surge por la presencia de un voltaje fijo en la señal analógica que se convierte en valores digitales [54]. La eliminación se realiza sustrayendo el valor promedio de la señal de cada muestra a lo largo de la señal analizada.

$$x_{\text{DCremoved}} = x - \text{media}(x) \quad (9)$$

Este proceso garantiza que la señal a evaluar está centrada alrededor de cero, lo que es útil para mantener un rango dinámico y reducir los sesgos en los análisis posteriores.

6.3. Aprendizaje automático y algoritmos de clasificación

El aprendizaje automático, mejor conocido como machine learning, es una rama de la inteligencia artificial, en la cual mediante el desarrollo de algoritmos y modelos estadísticos las máquinas aprenden sin ser programadas explícitamente para ello. La base de estos sistemas es la identificación de patrones y su capacidad predictiva [55]. Según los datos proporcionados y resultados esperados existen diferentes modelos de machine learning.

6.3.1. Modelos de aprendizaje supervisado

Los modelos de aprendizaje supervisado utilizan datos etiquetados que especifican la entrada y salida del algoritmo. Una vez entrenado el modelo con los datos iniciales se utiliza el modelo para realizar predicciones con base en nuevas entradas.[55] Este modelos se suelen utilizar para problemas de clasificación y regresión.

La clasificación es el proceso de categorización de elementos. Existen métodos de clasificación simples como la clasificación binaria donde las muestras pueden ser de clase positiva o negativa; y modelos de clasificación múltiple con más de dos clases.[56] En estos modelos el valor de salida es discreto.

La regresión, a diferencia de la clasificación, es el proceso de estimación de valores basados en entradas actuales, por lo que tiene una salida continua. Existen dos tipos de regresión, la univariada donde se estima un valor de salida y la regresión multivariada donde se tiene más de un valor de salida. [57]

Regresión logística

Este algoritmo se utiliza para la solución de problemas de clasificación binaria, cuyas variables resultantes son dicotómicas. Se consideran la observación de entrada x , que representa el vector de características $[x_1, x_2, \dots, x_n]$ es clasificada en y igual a 1 o 0 con base en la probabilidad $P(y = 1|x)$. Para ello se implementa la función sigmoide $\sigma(z)$, también conocida como función logística, representada como:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (10)$$

al aplicar la función sigmoide a cálculo de probabilidad $P_{y=1|x}$ se obtiene la expresión:

$$P(Y = 1|X) = \sigma(w \cdot x + b) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_N X_N)}} \quad (11)$$

que describe la relación entre las características y su probabilidad de pertenencia a la clase positiva ($Y = 1$). La clasificación de los parámetros se decide mediante el límite de decisión (decision boundary), donde aquellos valores con una probabilidad mayor a 0.5 se clasifican como 1 [58]

$$\text{decision}(x) = \begin{cases} 1 & \text{if } P(y = 1|x) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Máquina de soporte de vectores (SVM)

El objetivo de este algoritmo es determinar un límite de decisión, denominado hiperplano (Figura 9), óptimo para la solución de problemas de clasificación en un espacio N-dimensional. El algoritmo elige los vectores de soporte que son puntos de datos cercanos al hiperplano para definir los límites del mismo. Los clasificadores de vectores de soporte se basan en los hiperplanos de clase:

$$\langle w, x \rangle + b = 0 \quad (13)$$

siendo

$$w, x \in \mathbb{R}^n \quad y \quad b \in \mathbb{R}$$

donde el vector de coeficientes adimensionales w es normal al hiperplano de la superficie; b , es el valor de desplazamiento desde el origen (sesgo) y x son los valores del conjunto de datos . [59]

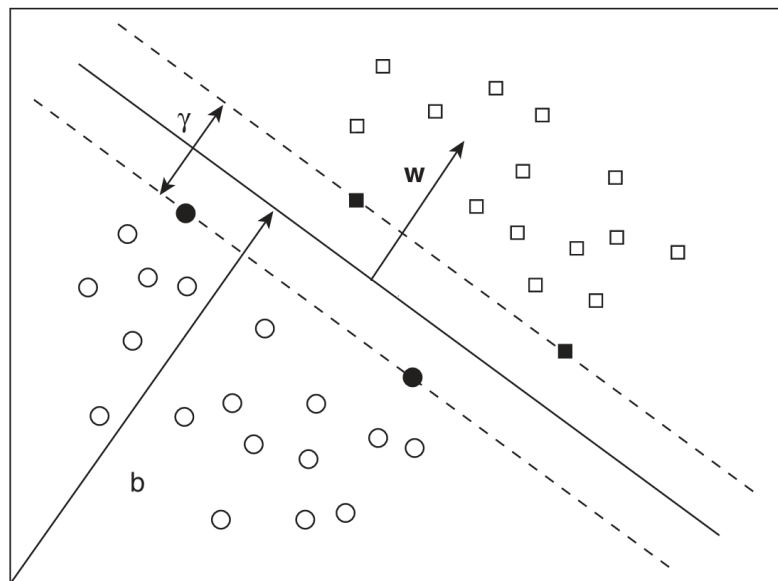


Figura 9: Representación de hiperplano bidimensional [59]

Para obtener los valores w y b Se utilizan los multiplicadores de Lagrange. Estos maximizan el margen entre las clases para cumplir con las restricciones de los vectores de soporte. El conjunto de entrenamiento $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$, donde x_i es la muestra y y_i la etiqueta de la clase, puede expresar w como la combinación lineal de los vectores de soporte por los multiplicadores de Lagrange óptimos i en los n vectores de soporte.

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (14)$$

Sin embargo, no es necesario expresar explícitamente a w en la función lineal discriminante:

$$f(x) = \sum_{i=1}^b \alpha_i y_i x_i^T + b \quad (15)$$

Utilizar una clasificación lineal en situaciones del mundo real resulta ineficiente. Las funciones de kernel adecuadas permiten implementar SVM para el reconocimiento de patrones no lineales operando en el espacio original n , esto se conoce como Kernel Trick. El Kernel se usa en la función lineal discriminante para expresar la función de decisión: [58]

$$f(x) = \sum_{i=1}^b \alpha_i y_i K(x_i, x_j) + b \quad (16)$$

K-Nearest Neighbor (KNN)

El método de Clasificación K-Nearest Neighbor (KNN) es un método de clasificación basado en que los patrones más cercanos a un patrón objetivo x' , el cual se desea etiquetar, brindan información útil sobre la etiqueta (Figura 10). Por lo que x' está etiquetado con la clase de la mayoría de los K-Nearest en el espacio. K representa el número de vecinos a considerar para la predicción. [60]

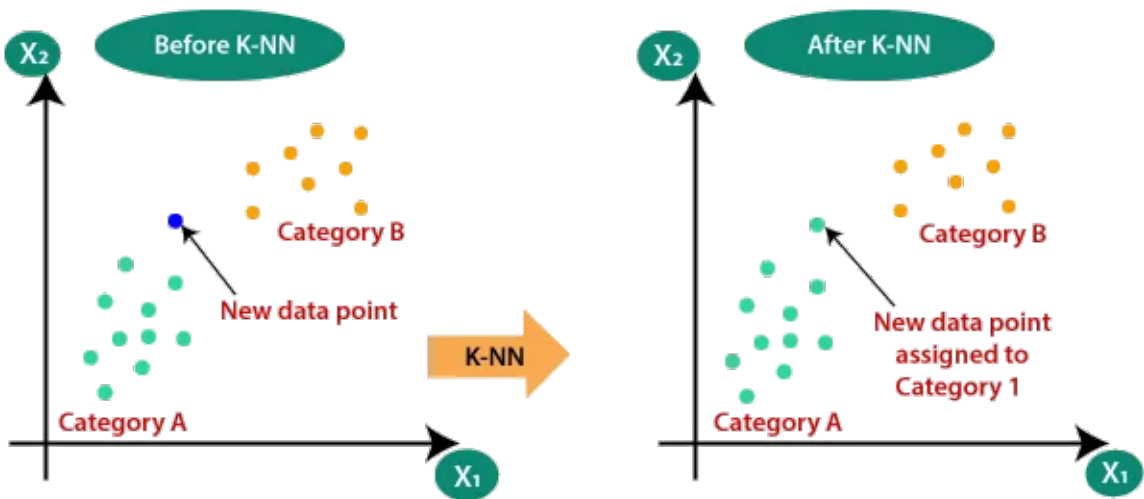


Figura 10: Modelo de clasificación K-NN binario [61]

Una de las métricas principales en este algoritmo es la distancia para determinar los puntos más cercanos al objetivo x' . Para ello se hace uso de la distancia de Minkowski:

$$d(x, X_i) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{\frac{1}{p}} \quad (17)$$

cuyos casos especiales son la distancia euclidiana con $p = 1$, que determina la distancia cartesiana entre dos puntos en un plano y se utiliza en datos cuya interpretación es geométrica. Con $p = 2$ Se obtiene la ecuación de la distancia a Manhattan, utilizada cuando la distancia total recorrida del objeto es de interés. [62].El rendimiento de este algoritmo se ve directamente afectado por la elección del parámetro k y la métrica de la distancia.

6.3.2. Modelos de aprendizaje no supervisado

Los datos utilizados en los modelos no supervisados no se encuentran etiquetados previamente. Los datos de entrenamiento constan únicamente de variables de entrada definidas y el modelo encuentra información estructural de interés para el usuario. Estos modelos son utilizados en problemas de agrupación y asociación.

Los problemas de agrupación, también llamados clustering, constan en la segmentación de un grupo en subgrupos basado en similitudes entre los elementos. Para este proceso es necesario determinar una métrica de similitud para la agrupación. En la asociación se determinan reglas que describen el comportamiento de los datos. [57]

Entre estos modelos existen los modelos de aprendizaje semi-supervisado. Que se basan en una porción de datos etiquetados, pero la mayoría de la información son datos sin etiquetar. Situándose entre el aprendizaje supervisado y el no supervisado.

6.3.3. Modelos de aprendizaje por refuerzo

Este modelo aprende mediante prueba y error, utilizando un sistema de retroalimentación. Se mantienen en el continuo aprendizaje de su entorno de forma interactiva, lo que permite a las máquinas y agentes determinar el comportamiento ideal dentro de un contexto específico, con el objetivo de maximizar el rendimiento. [63]

6.3.4. Aprendizaje profundo - Deep learning

El continuo desarrollo de algoritmos y la capacidad computacional de nuevas tecnologías dan lugar al aprendizaje profundo, comúnmente llamado deep learning, una rama del machine learning donde se hacen uso de algoritmos que se asemejan a las conexiones neuronales biológicas del cerebro humano [63]. En la mayoría de algoritmos de deep learning se hace uso de arquitecturas de redes neuronales.

6.4. Evaluación de rendimiento de algoritmos

6.4.1. Modelos de matriz de confusión

Esta herramienta da una evaluación tabular de los resultados de clasificación de los modelos binarios. Representa la calidad del modelo con base en la relación entre las predicciones del modelo y la clasificación real del conjunto de datos. Muestran el número de clasificaciones correctas (verdadero positivo TP, verdadero negativo TN) y clasificaciones incorrectas (falso positivo FP, falso negativo FN) (Figura 11, elaboración propia) [58]

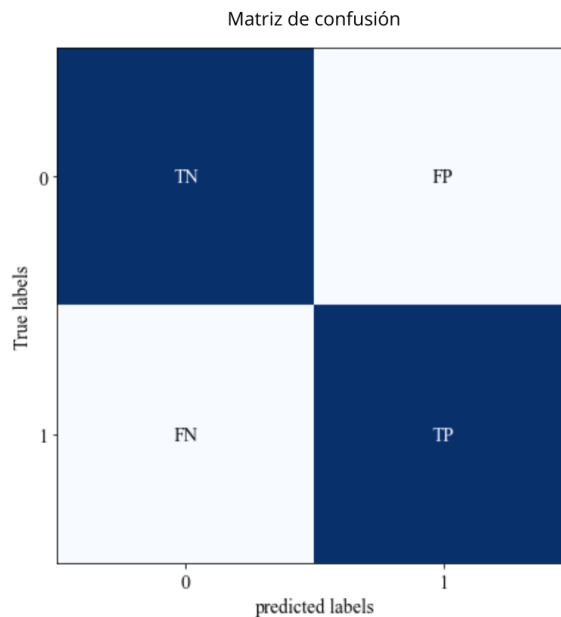


Figura 11: Representación de matriz de confusión para la clasificación de pacientes. (Sujetos sanos=0, Pacientes con EP=1).

Fuente: Elaboración propia

6.4.2. Métricas de evaluación

Para evaluar la eficiencia de los diversos algoritmos se hace uso de medidas estadísticas basadas en las tasas de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. La precisión, mide la proporción de valores positivos correctamente clasificados entre todos los valores positivos que el modelo predijo. [58]

$$\text{Precisión} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (18)$$

La sensibilidad, también llamada recall, es la tasa de verdaderos positivos correctamente identificados en por el modelo.

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \times 100 \quad (19)$$

En contraparte, la especificidad mide la proporción de verdaderos negativos correctamente identificados por el modelo.

$$\text{Especificidad} = \frac{TN}{TN + FP} \times 100 \quad (20)$$

6.4.3. Validación cruzada

La validación evalúa y compara el aprendizaje automático del modelo utilizando dos conjuntos de datos: prueba y entrenamiento. Se utilizan diversos pares de conjuntos a lo largo del entrenamiento del algoritmo para obtener una mejor estimación del rendimiento del modelo. Un enfoque específico de la validación cruzada es K-fold cross-validation, el conjunto de datos se divide en múltiples subconjuntos k, de igual tamaño, y es entrenado k veces utilizando k-1 pliegues. Utilizando el pliegue restante como conjunto de prueba. Es importante es necesario garantizar que cada pliegue represente correctamente todos los datos del conjunto general [64]

6.4.4. Curva ROC

La curva de eficacia diagnóstica (Receiver Operating Characteristic, ROC por sus siglas en inglés) (Figura 12), es método estadístico que evalúa la exactitud de clasificación de pruebas dicotómicas. La curva ROC determina el punto de corte en el que se alcanza la sensibilidad y especificidad más alta del modelo. Se basa en los cálculos de sensibilidad y especificidad para todos los posibles puntos de corte. La sensibilidad se sitúa en el eje de ordenadas y el complementario de la especificidad en el eje de abscisas. La curva se traza mediante la unión de los pares de valores de sensibilidad y especificadas correspondientes al punto de corte. El área bajo la curva, AUC, mide la capacidad discriminadora del modelo. [65]. Con base en el AUC se puede estimar la capacidad discriminadora del test. Un valor cercano a 1 demuestra una alta capacidad discriminadora, entre más cercano sea el valor a 0,5 esta capacidad disminuye.

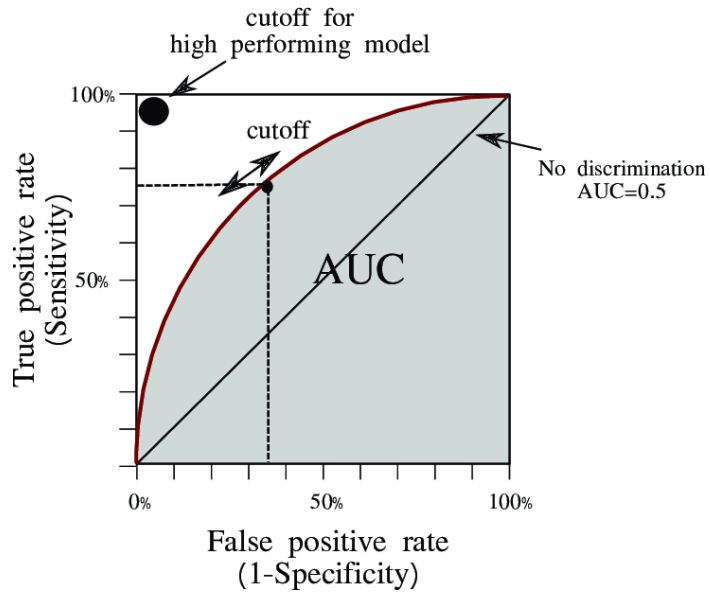


Figura 12: Curva de eficacia de diagnóstico (ROC) y área bajo la curva (AUC). [66]

La metodología descrita en este capítulo detalla el proceso de adquisición de audios para la evaluación de métodos de grabación; el análisis y clasificación de muestras basado en características de prosodia. Se define el dataset utilizado, el diseño, la implementación y la evaluación de los algoritmos de machine learning. A su vez, se describe el proceso de extracción de características para diferentes tareas diadococinéticas.

Los audios originales se clasifican en dos grupos correspondientes al estado del sujeto: paciente con enfermedad de Parkinson (EP) y sujetos de control (HC). Para ello se emplean dos modelos de aprendizaje de máquinas supervisado: Máquina de soporte de vectores (SVM) y regresión logística (RL). El rendimiento en ambos modelos es evaluado y se calculan las medidas de desempeño correspondientes. La Figura 13. Detalla los pasos clave de la metodología implementada.

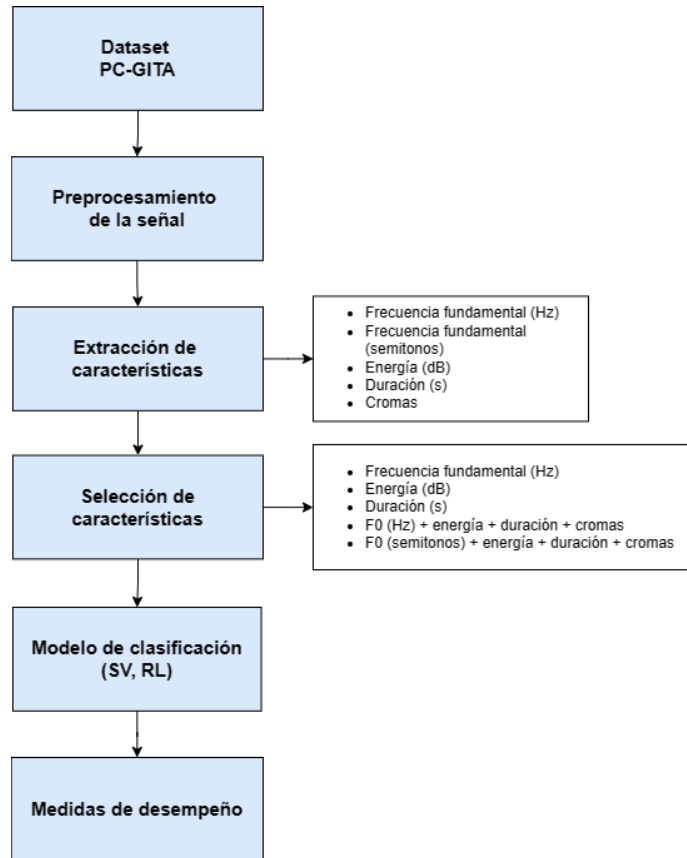


Figura 13: Diagrama esquemático de la metodología implementada.

7.1. Procedimiento de grabación

Las grabaciones de audio son realizadas utilizando tres micrófonos diferentes: un micrófono profesional de condensador de 3 cápsulas modelo uber-mic (M-Audio., China) en configuración cardioide; el micrófono incorporado de una computadora portátil HP laptop 15-dw3xxx (Hewlett-Packard Company, California, Estados Unidos) y el micrófono integrado en un dispositivo móvil gama media Samsung A23 (Samsung Electronics., Corea del Sur). Todas las grabaciones se desarrollan en un entorno controlado, el Laboratorio de Visualización Multimedia de la Biblioteca Amparo Codina de Campollo, ubicada en las instalaciones de la Universidad del Valle de Guatemala.

Cada micrófono se coloca a una distancia aproximada de 10 centímetros de los participantes, con una alineación constante para todas las grabaciones. Para la grabación de audio que utiliza el micrófono profesional y el micrófono incorporado en una computadora se emplea el software de edición de audio Audacity versión 3.4.2 (Muse Group, Estados Unidos). Las sesiones de grabación se realizan con la configuración estándar de la Figura 14 con una frecuencia de muestreo de 44100 Hz y una resolución de 16 bits.

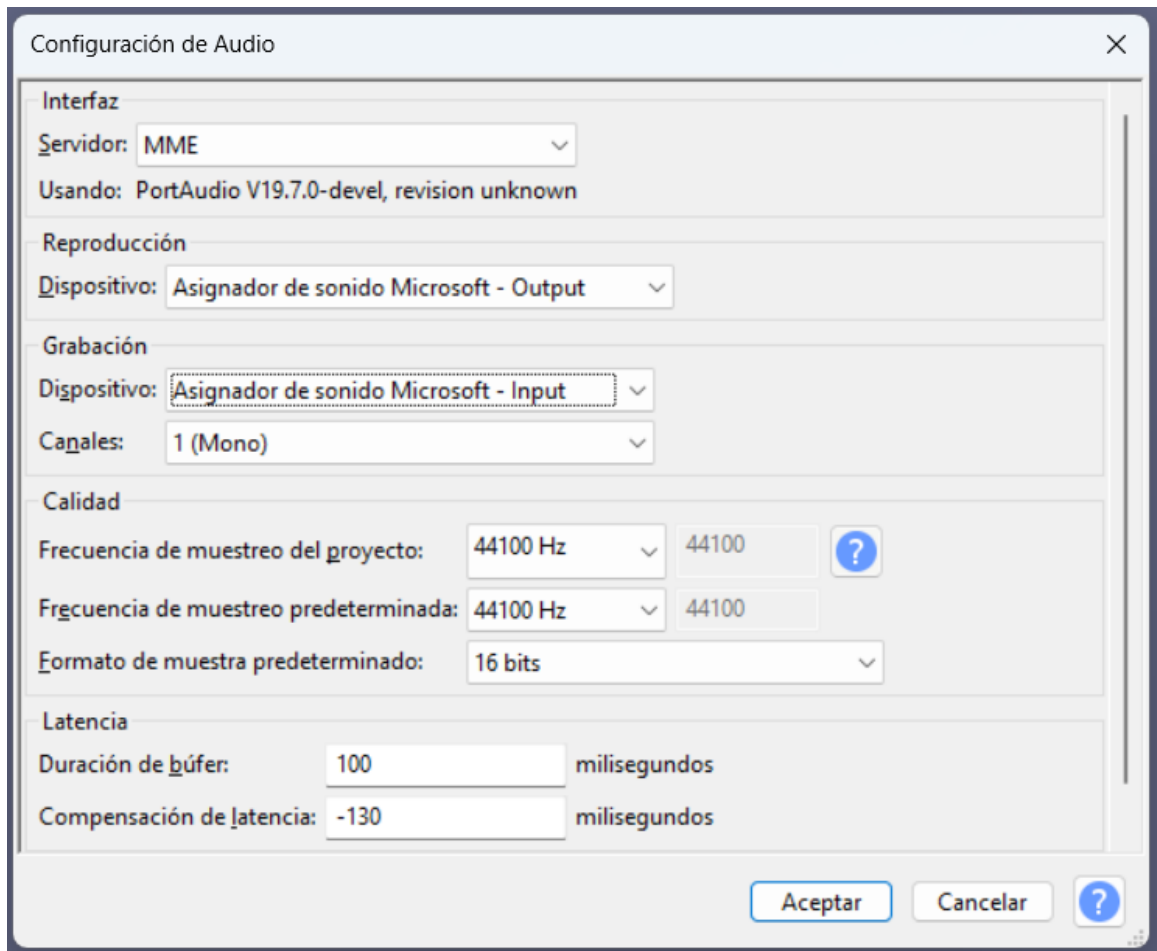


Figura 14: Ventana de Audacity: Configuración de audio

Cada participante es grabado de forma individual mientras llevaba a cabo un conjunto de 11 tareas específicas repetidas dos veces. Estas tareas consisten en fonaciones prolongadas de las cinco vocales del idioma español, así como la repetición de secuencias diadococinéticas, incluyendo los sonidos /pa/, /ta/, /ka/, /pataka/, /petaka/ y /pakata/.

Se organiza una sesión individual para la grabación de cada participante en el laboratorio de Visualización Multimedia (Figura 15). En a primera prueba se utiliza el micrófono profesional y en una segunda prueba en el participante es grabado simultáneamente con el micrófono incorporado en la computadora y el micrófono integrado en el dispositivo móvil, permitiendo una comparación directa entre los distintos dispositivos de grabación.

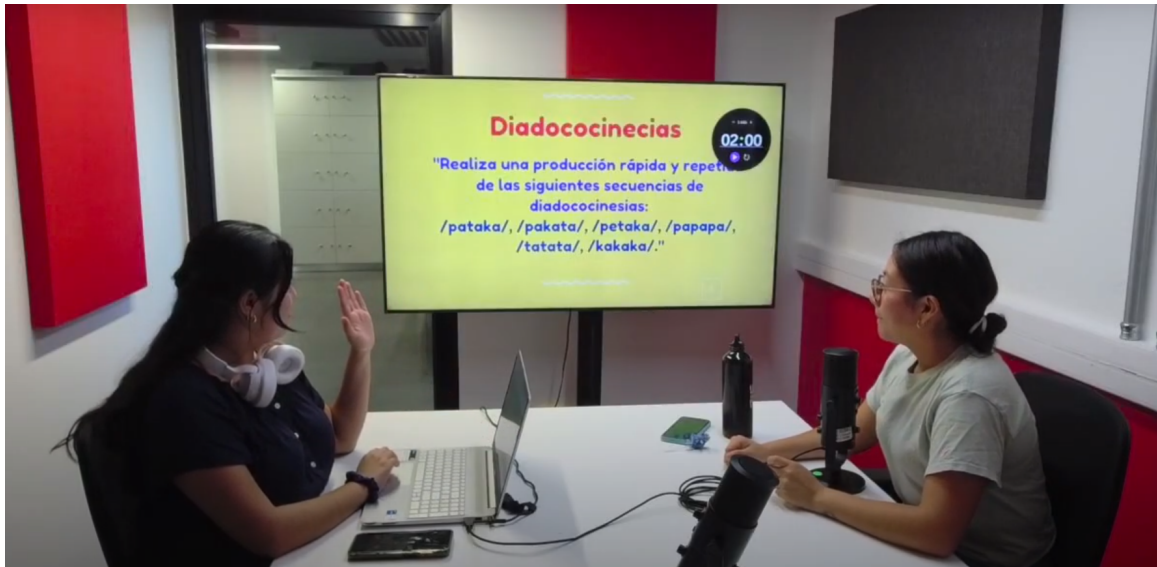


Figura 15: Fotografía de sesión individual de grabación en el Laboratorio de Visualización Multimedia de la Universidad del Valle de Guatemala

Tras la adquisición de las grabaciones, se realiza la separación de los fragmentos de interés utilizando nuevamente el software Audacity. Cada fragmento recortado corresponde a una tarea específica realizada por el participante. Los fragmentos se organizan y guardan en carpetas específicas de acuerdo al tipo de micrófono utilizado en la grabación. Cada archivo de audio es etiquetado con un identificador único que representa el número asignado al participante (S1); el tipo de micrófono utilizado donde: M corresponde al dispositivo móvil, C al micrófono integrado en la computadora y P al micrófono profesional; la tarea realizada y el número de repetición de la tarea. Por ejemplo, el archivo S1M_pataka_1.wav corresponde al participante S1, grabado con un dispositivo móvil, realizando la primera repetición de la tarea “pataka”. Es importante destacar que en el caso de los audios grabados con el dispositivo móvil, se exportan a una computadora para su posterior análisis. Durante este proceso, los archivos de audio, que originalmente tenían la extensión .m4a, se convierten a archivos con extensión .wav para facilitar su manejo y análisis en Audacity. Tras la organización y etiquetado de los fragmentos de audio se realiza un preprocesamiento a las señales para garantizar la homogeneidad en las muestras. Esto se realiza en Visual Studio Code (Microsoft, Estados Unidos) con lenguaje de programación Python, utilizando el módulo librosa para la manipulación de los archivos de audio. Todos los fragmentos de audio se normalizan para estar en una escala entre -1 y 1. Con el objetivo de asegurar que el nivel de volumen es consistente entre todas las grabaciones. Esto implica ajustar el nivel de volumen de cada muestra para que alcance un nivel máximo predefinido sin distorsionar la señal original. A su vez, se elimina cualquier componente de corriente continua de la señal, para evitar distorsiones en las muestras de audio. Para tener homogeneidad en las muestras se evalúan los primeros 4 segundos de cada muestra a partir del primer sonido detectado. Con un total de 66 muestras de audio por participante y 4 participantes se obtienen 264 muestras.

Una vez completados estos procesamientos, se procede a evaluar las métricas de calidad de audio para cada muestra, incluyendo el rango dinámico, la relación señal a ruido (SNR) y la distorsión armónica total (THD).

Finalmente, se compararon y analizaron los resultados obtenidos para cada métrica con el objetivo de determinar y comparar la calidad relativa de las grabaciones realizadas con los diferentes micrófonos utilizados en el estudio.

7.2. Base de datos PC-GITA

Se utiliza en dataset denominado PC-GITA proporcionado por Rafael Orozco et al. [67], que incluye grabaciones de 50 personas con EP y 50 personas sanas, de los cuales 25 son hombres y 25 son mujeres en cada grupo. Todos los participantes son hispanohablantes nativos de Colombia. Las grabaciones fueron realizadas a 44100 Hz con 16 bits de resolución, con un micrófono profesional (Shure, SM 63L). Los pacientes con EP fueron diagnosticados por neurólogos con base a las escalas UPDRS (Unified Parkinson’s Disease Rating Scale) y H&Y (Hoehn and Yahr scale); Se asegura que los pacientes sanos no presentaron ningún síntoma asociado a EP u otra enfermedad neurológica.

El protocolo de grabación utilizado en el dataset considera diversas tareas para el análisis de diferentes aspectos de la voz en los pacientes, agrupadas en tres aspectos principales: fonación, articulación y prosodia, con un total de 21 tareas. En este trabajo se utiliza el conjunto de evaluaciones diadococinesias orales a través de repeticiones monosilábicas y palabras conformadas por las mismas (/pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, /ka/.), con un total de 6 tareas, cada una con 50 grabaciones de pacientes diagnosticados y 50 grabaciones de sujetos sanos, con una duración entre 1 y 5 segundos.

Todos los procesos se llevan a cabo en el entorno de programación Visual Studio Code con lenguaje de programación Python. La base de datos conformada por un total de 600 archivos tipo .wav es cargada utilizando la biblioteca Pandas para la manipulación posterior de los datos. Para el análisis y extracción de las características se utiliza el framework DisVoice [18][19].

El dataset se encuentra organizado en una estructura jerárquica de carpetas (Figura 16). La carpeta principal denominada “Diadococinéticas”, almacena todas las subcarpetas que contienen los datos de las diferentes tareas a evaluar. Dentro del directorio principal se encuentran seis subcarpetas, correspondientes a las tareas específicas: /pa-ta-ka/ ,/pa-ka-ta/ ,/pe-ta-ka/ ,/pa/ ,/ta/ y /ka/. Adicionalmente estas carpetas específicas tienen una clasificación interna para la diferenciación de las muestras de pacientes diagnósticas (PD) y las controles sanos (HC). Dentro de las subcarpetas condicionales se encuentran los archivos de audio correspondientes, cada archivo con un identificador único asociado con el número de muestra y tipo de tarea.

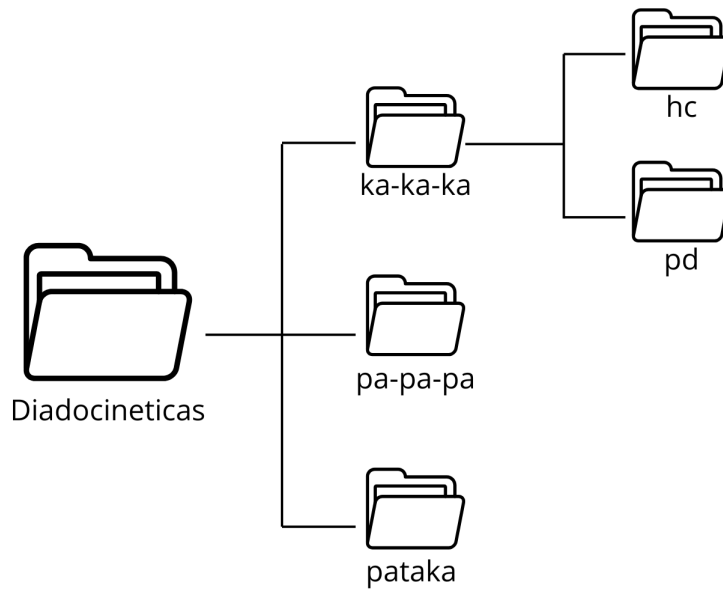


Figura 16: Organización jerárquica del conjunto de tareas diadococinéticas del dataset PC-GITA

7.3. Preprocesamiento de la base de datos

Para garantizar la uniformidad y precisión de la información se realiza un proceso de normalización y eliminación de señal DC para todos los audios del dataset. La normalización se realiza tomando el valor máximo absoluto de la señal original y multiplicando por el valor máximo deseado, en este caso 1. La eliminación de la señal DC se realizó utilizando la ecuación 9

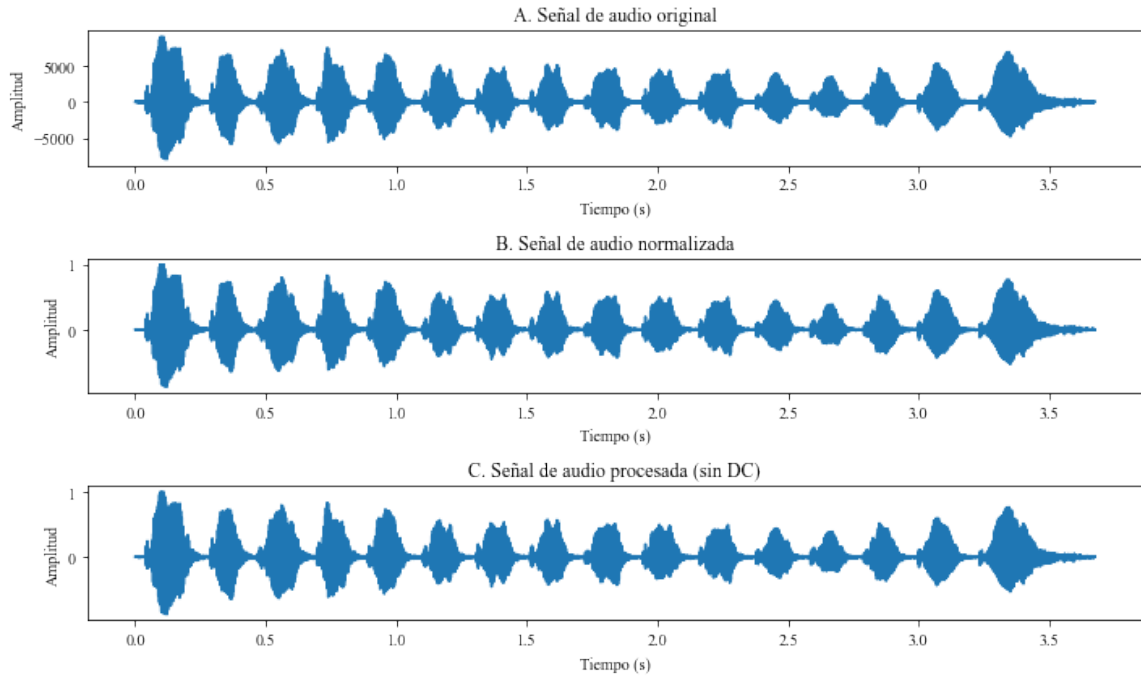


Figura 17: Proceso de normalización y eliminación. A) Señal de audio original. B) Señal de audio normalizada en una amplitud de $[-1,1]$. C) Señal sin elementos de frecuencia constante (DC)

7.4. Adquisición de parámetros.

Para la adquisición de parámetros se emplea la herramienta DisVoice [18][19]. Esta herramienta utiliza algoritmos específicos para extraer parámetros divididos en características fonológicas, glotales, de fonación, de articulación y de prosodia. Usando la clase Prosody del modulo disvoice.prosody se obtiene un total de 103 características de prosodia, las cuales están divididas en tres grupos principales: frecuencia fundamental, energía y duración (Figura 18). Específicamente con 30 características corresponden a frecuencia, 48 a energía y 25 a la duración de la muestra. En conjunto estos parámetros (frecuencia fundamental, duración y energía) son cruciales para analizar las características prosódicas del habla y pueden ser de la enfermedad de Parkinson.

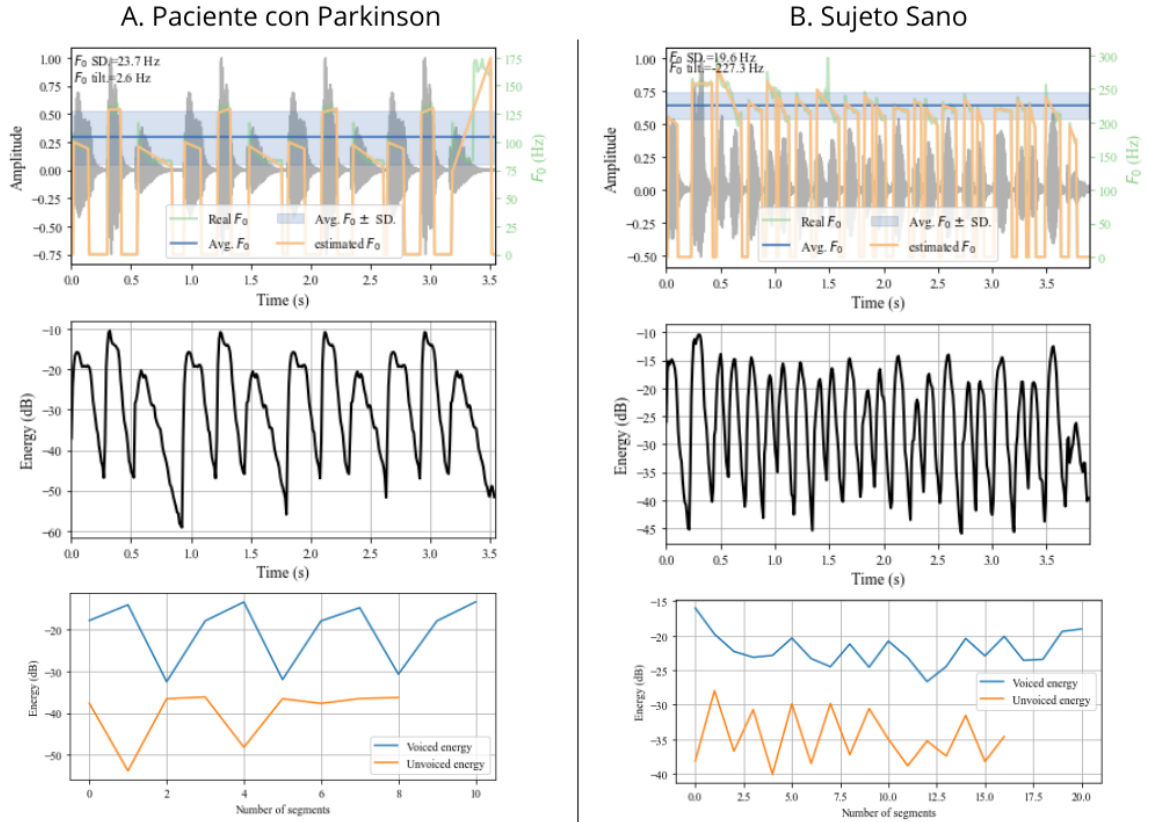


Figura 18: Análisis temporal de amplitud, frecuencia fundamental, energía y segmentos de duración de la diadicocinesia /PATAKA/. A) Paciente diagnosticado con Parkinson. B) Sujero sano

La Tabla 1 describe las características por grupo acompañas por su identificación utilizada por DisVoice. Cada característica se compone de medidas descriptivas: promedio (avg), desviación estándar (std), máximo (max), mínimo (min), asimetría (skw) y curtosis (kurtosis). Las características marcadas con un * incluyen solo promedio, desviación estándar, asimetría y curtosis. Aquellas marcadas con ** no tienen medidas descriptivas asociadas. Las característica **relación de duración**, marcada con ***, se compone de relaciones proporcionales de los segmentos: $PVU = \text{Pausa} / (\text{voz} + \text{Sordo})$, $PU = \text{Pausa} / \text{Sordo}$, $UVU = \text{Sordo} / (\text{Voz} + \text{Sordo})$, $VVS = \text{Voz} / (\text{Voz} + \text{Sordo})$, $VP = \text{Voz} / \text{Pausa}$ y $UP = \text{Sordo} / \text{Pausa}$.

Las 103 características se nombran con base al identificador del parámetro evaluado y la medida estadística descriptiva asociada. Por ejemplo, avgtiltEvoiced corresponde al promedio (avg) de la Inclinación de la estimación lineal del contorno de energía para segmentos sonoros (tiltEvoiced).

Para facilitar los análisis posteriores el formato de salida de las características extraídas se estable como “dataframe”. Obteniendo una estructura bidimensional de una fila por cada muestra de audio y 104 columnas incluyendo el identificador del audio analizado. En función de la carpeta de origen de clasificación (hc o pd) se añade una columna qué indica la clase de la muestra. Se asignaron valores binarios para representar las clases: los controles sanos se etiquetan como 0, mientras qué los pacientes diagnosticados con EP se etiquetan como 1.

Cuadro 1: Tabla de características prosódicas extraídas con DisVoice

Grupo	Característica
Frecuencia	Contorno F0 (F0)
	Inclinación de la estimación lineal de F0 para cada segmento sonoro (F0tilt)
	MSE de la estimación lineal de F0 para cada segmento sonoro (F0mse)
	F0 en el primer segmento sonoro (1F0)
	F0 en el último segmento sonoro (lastF0)
Energía	Contorno de energía para segmentos sonoros* (Evoiced)
	Inclinación de una estimación lineal del contorno de energía para segmentos sonoros* (tiltEvoiced)
	MSE de la estimación lineal del contorno de energía para el segmento sonoro* (mseEvoiced)
	Energía del primer segmento sonoro (1Evoiced)
	Energía del último segmento sonoro (lastEvoiced)
	Contorno de energía para segmentos sordos* (Eunvoiced)
	Inclinación de una estimación lineal del contorno de energía para segmentos sordos* (tiltEunvoiced)
	MSE de la estimación lineal del contorno de energía para segmentos sordos* (mseEunvoiced)
	Energía del primer segmento sordo (1Eunvoiced)
Energía del último segmento sordo lastEunvoiced	
Duración	Velocidad de voz** (Vrate)
	Duración de la voz (durvoiced)
	Duración de segmentos sordos (durunvoiced)
	Duración de pausas (durpause)
	Relaciones de duración*** (PVU, PU, UVU, VVU, VP, UP)

Para la división y análisis de las señales de audio en segmentos temporales el tamaño de ventana y del paso son 0.02 y 0.01 segundos respectivamente, estos valores se relacionan con la muestra al multiplicarse por la frecuencia de muestreo de la señal de audio analizada. El tamaño de ventana define la duración de cada segmento, mientras que el tamaño de paso controla la superposición entre los segmentos. Al igual que estos parámetros se determina la longitud mínima de pausa de la muestra con un valor de 0.14 segundos, que se multiplica con la frecuencia de muestreo fs. Dichos valores están definidos en la inicialización de atributos de la clase prosody del modulo disvoice.prosody. Por lo que la cantidad de ventanas depende de la duración de la muestra evaluada.

Todos las muestras de audio son de tipo monofónicas, utilizando una única señal de audio para reproducir el sonido. En caso de tener una muestra con múltiples canales, por ejemplo una señal estereofónica, se promedian los valores de los canales para obtener una señal monoaural. Se crean seis Data Frames correspondientes a cada tarea diadococinética, para el análisis específico y detallado de las características extraídas en relación a cada una de las tareas. Cada uno de estos DataFrames se guarda como un archivo separado en formato .txt, para su uso posterior en los métodos de clasificación.

7.4.1. Implementación de algoritmos supervisados para la clasificación

Para analizar cada tarea diadococinesia se implementan dos modelos de machine learning: Support Vector Machine y Regresión logística. La división de los datos de entrenamiento y prueba se realiza mediante validación cruzada (k-folds validation), donde k se establece en 4. Con un total de 100 muestras en cada dataset, se generan 4 pliegues de manera equitativa, lo que representa 25 muestras en cada pliegue. Durante el proceso de validación cruzada se emplean 3 pliegues como conjunto de entrenamiento, 75 % de las muestras, y el pliegue restante como conjunto de prueba, el 25 % restante de las muestras. Se agrupan los resultados de cada conjunto de prueba evaluar el rendimiento del modelo, empleando un total de 100 muestras correspondientes a las 25 muestras de prueba de cada pliegue. En ambos modelos se emplea un enfoque de búsqueda de hiperparámetros, utilizando `param_grid` para optimizar los modelos. Cada combinación de hiperparámetros se evalúa durante la validación cruzada. La estrategia para evaluar el desempeño del modelo con validación cruzada en el conjunto de prueba se establece con base a la puntuación de precisión.

7.4.2. Modelo de Máquina de Vectores de Soporte

Para la implementación del modelo se utiliza la clase SVC del módulo `sklearn.svm` en python. Se evalúan diferentes combinaciones de hiper parámetros, incluyendo el tipo kernel, el parámetro de regularización C y el coeficiente gamma. El propósito de la implementación de hiperparámetros es encontrar la mejor configuración para proporcionar el mejor rendimiento en términos de precisión para la clasificación del modelo. Se analizan tres claves principales: 'kernel' para evaluar dos tipos de kernel: lineal y radial basis function (RBF); la clave C, correspondiente a la regularización de modelo con valores numéricos: 0.1, 1, 10, 100 y 1000; Por último, la clave gamma correspondiente con valores: 0.000, 0.001, 1, 10 y 100.

7.4.3. Modelo de regresión logística

El modelo de regresión logística se implementa utilizando la clase `LogisticRegression` del módulo `sklearn.linear_model` en python. Se realiza la selección de hiperparámetro para el modelo, incluyendo el tipo de penalización L_1 y L_2 , los valores de regularización C en un rango de 0.01, 0.1, 1, 10 y 100. Y el solucionador a utilizar, en este caso únicamente se evalúa "liblinear" por la naturaleza de la clasificación binaria.

En este modelo se evalúan los pesos asignados a cada característica y su influencia, se extraen los coeficientes de regresión logística utilizando `grid.best_estimator_.coef_` y se almacenan para su análisis posterior.

7.5. Análisis multimodal de características prosódicas

Para la clasificación y análisis del dataset PC-GITA se realizan diversos experimentos. Cada uno de estos considera los modelos de clasificación: maquina de soporte de vectores

(SVM) y regresión logística. Además, se realiza de forma individual para cada tarea diadococinética.

7.5.1. Evaluación integral de características prosódicas

Esta experimentación aplica los modelos de clasificación a cada tarea diadococinética, considerando las 103 características de forma conjunta. Para cada prueba, se obtiene la matriz de confusión correspondiente, la curva ROC y se calculan medidas de desempeño: exactitud, sensibilidad que determina la precisión del modelo para identificar verdaderos positivos: paciente con Parkinson, y especificidad que describe la habilidad del modelo para clasificar verdaderos negativos: Sujetos sanos.

7.5.2. Análisis detallado de características prosódicas

Se profundiza el análisis de los conjuntos de parámetros: frecuencia fundamental, energía y duración. En cada experimentación se evalúa el desempeño de los modelos clasificación y se utilizan los coeficientes del modelo de regresión logística para determinar cuales son las características que tienen mayor peso.

7.5.3. Evaluación de parámetros de frecuencia en semitonos

Se modifica la dimensionalidad de las 30 características de frecuencia, de Hz a semitonos. El objetivo es evaluar si el cambio de dimensionalidad elimina el sesgo de género presente en la frecuencia en Hz. Se realizan las pruebas individuales de este conjunto de características y se compara su desempeño con los obtenidos en la segunda experimentación para el conjunto de frecuencia.

7.5.4. Extracción y análisis de cromas

En esta experimentación se introducen nuevos parámetros relacionados con los 13 cromas de la escala tonal occidental. Se utiliza `ShortTermFeatures.feature_extraction` de `PyAudioAnalysis` [68] para la extracción de las los cromas. Se mantiene uniformidad respecto al tamaño de ventanas y paso para la división de las señales de audio, por lo que se implementan los mismos valores definidos por `DisVoice` (tamaño de ventana=0.02, paso = 0.01)

Por cada croma se obtiene un conjunto de estadísticas descriptivas (promedio, desviación estandar, mínimo, máximo, asimetría y curtosis), siendo un total de 78 nuevas características. Las características se identifican por el número de croma evaluado (`chr_n`) y la medida estadística descriptiva asociada, al igual que la nomenclatura utilizada en `DisVoice`. Por ejemplo, `chr1avg` corresponde al promedio del primer croma. Se evalúa el desempeño al analizar aisladamente estos nuevos parámetros para cada tarea.

7.5.5. Implementación de nuevas características y cambio dimensional en la frecuencia

En esta última experimentación se introduce el nuevo conjunto de 78 parámetros cromáticos a cada tarea. Con un total de 181 características se evalúa cada tarea con base en el desempeño de los modelos de clasificación. La nueva información considerando la dimensionalidad original de la frecuencia y el cambio a semitonos.

8.1. Evaluación de los métodos de adquisición de audio

Se evalúan las diferencias entre los métodos de adquisición de sonido, utilizando 3 micrófonos diferentes. En la Figura 19 se visualiza la forma de onda de la tarea /pakata/ realizada por el mismo sujeto en los 3 dispositivos utilizados, durante 2 segundos. La forma de onda de la primera señal, grabada con un micrófono profesional, tiene un rango de amplitud de $[-0.7, 0.5]$, la señal es más uniforme y constante con una amplitud estable a lo largo del tiempo; En contraste, la segunda señal, adquirida mediante el micrófono de una computadora, tiene una variación constante de amplitud entre los $[-0.16, 0.20]$. destacan picos y valles a lo largo de muestra. La última forma de onda, correspondiente al audio adquirido mediante el micrófono integrado en un celular, muestra variaciones de amplitud menos pronunciadas que la señal anterior, en un rango de $[-0.6, 0.8]$, aunque en comparación con la primera señal es menos uniforme.

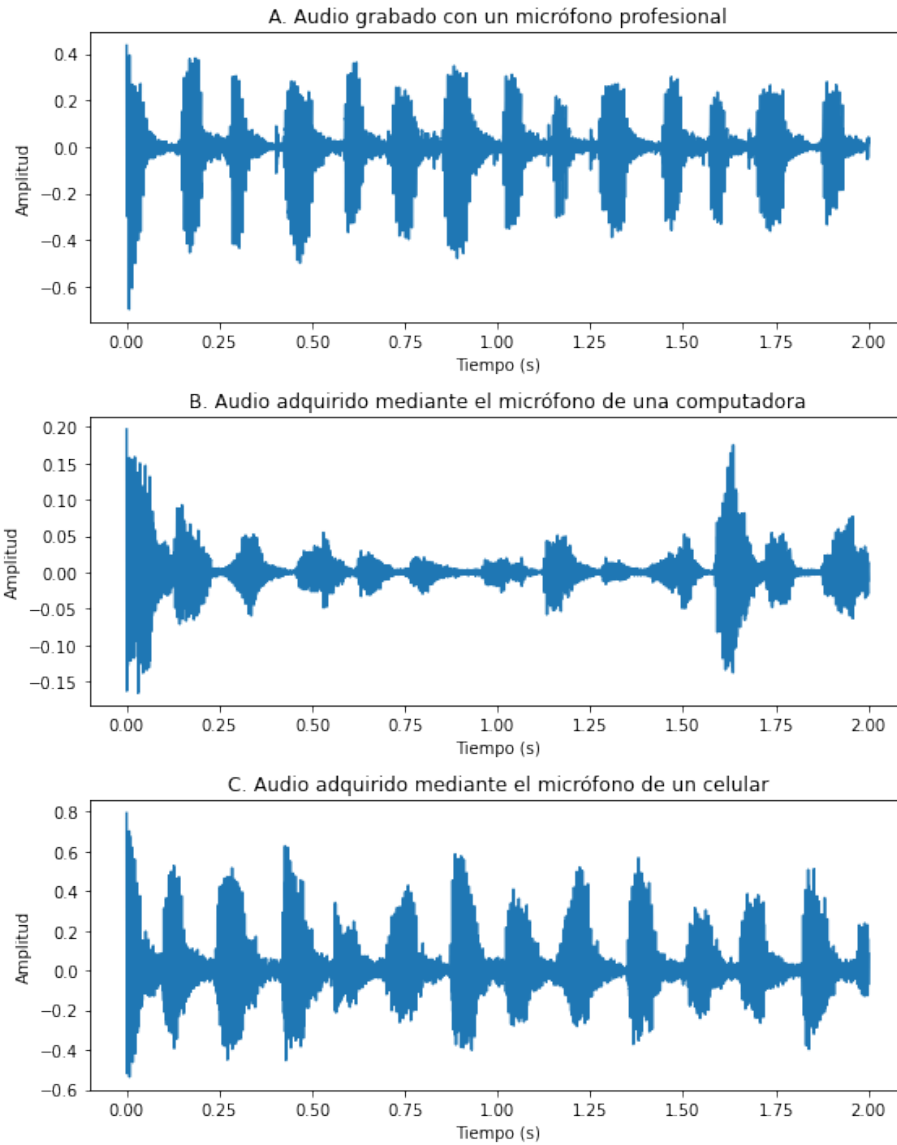


Figura 19: Comparativa de señales de audio adquiridas con diferentes dispositivos. A) Audio adquirido con un micrófono profesional. B) Audio adquirido con el micrófono de una computadora. C) Audio adquirido mediante el micrófono de un celular

Las grabaciones correspondientes al micrófono de una computadora y el micrófono de un celular son realizadas en simultáneo. La señal adquirida mediante el micrófono integrado en una computadora tiene una disminución de amplitud, esto se visualiza mejor en la Figura 20 donde se observan diferencias significativas entre la energía de las señales de audio capturada. Las figuras 20.A y 20.C tienen un comportamiento de energía similar. Se destaca que el comportamiento de las formas de onda fueron semejantes a lo largo de las muestras evaluadas.

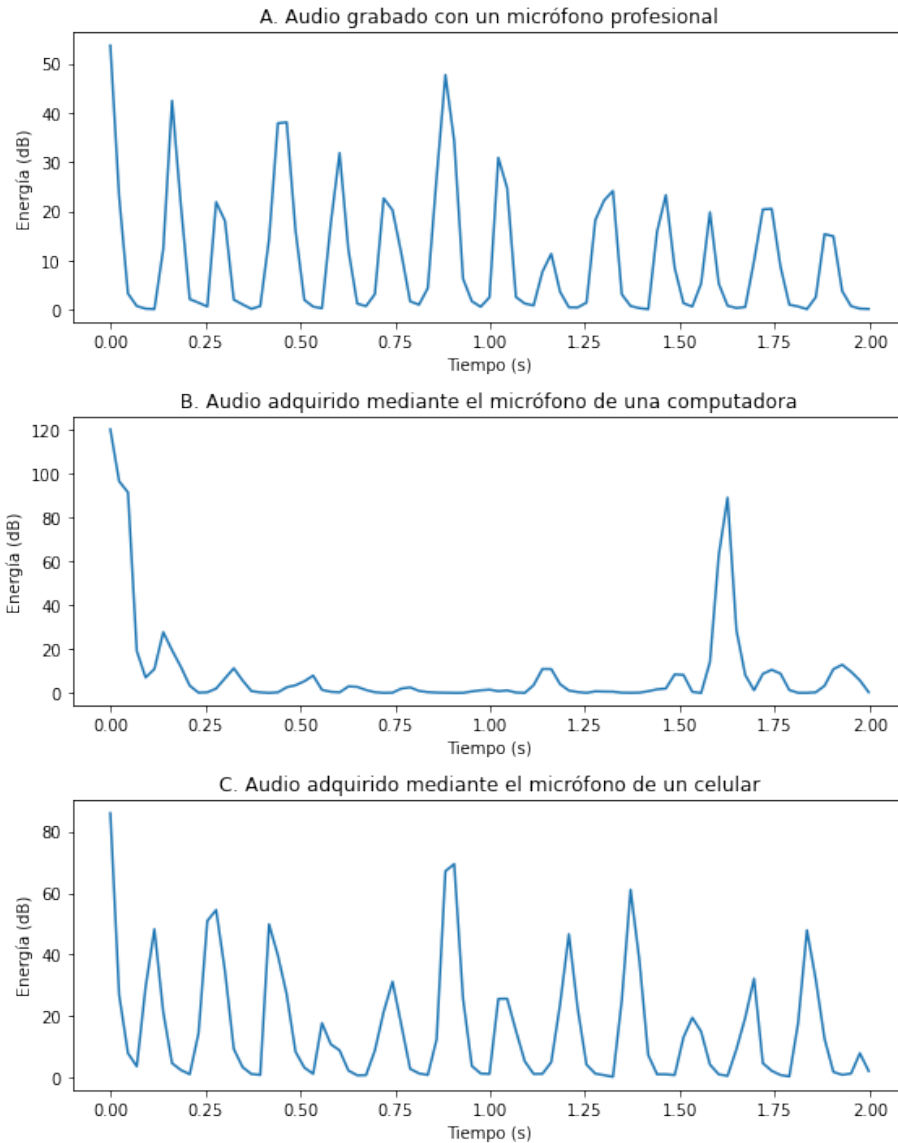


Figura 20: Comparativa de energía de señales de audio adquiridas con diferentes dispositivos. A) Audio adquirido con un micrófono profesional. B) Audio adquirido con el micrófono de una computadora. C) Audio adquirido mediante el micrófono de un celular

En la Figura 21 se evalúan las diferencias de los distintos dispositivos de adquisición con base a el rango dinámico, la relación señal a ruido (SNR), la distorsión armónica total (THD) y la energía promedio de todos los audios adquiridos separados por dispositivo. Referente al rango dinámico se observa que el micrófono profesional tiene el valor más bajo (1.6851) y el micrófono de la computadora tiene el valor más alto (1.8579). La relación señal-ruido de las tres señales es bastante baja, cercana a 0, por lo que se reportan los resultados en notación $1e-7$. Pese a que el micrófono profesional tiene una relación SNR negativa, esta es poco significativa al considerar la escala. En los tres casos el ruido de fondo no es significativo en comparación a la señal de interés, considerando que todas las grabaciones se llevaron a cabo en un entorno controlado, caracterizado por condiciones estables y predecibles.

La distorsión armónica de todas las experimentaciones es 7.071 lo que sugiere que la distorsión no es un factor determinante al evaluar la calidad de adquisición en un entorno controlado. Por su parte la evaluación de energía muestra las diferencias más significativas entre los métodos de grabación. Se evalúa la energía promedio de todos los audios. Los valores más altos se presentan al utilizar una computadora (0.0035), en relación a lo reportado en los rango dinámico y la Figura 20, representa una distorsión a las muestras resultando en una grabación inexacta.

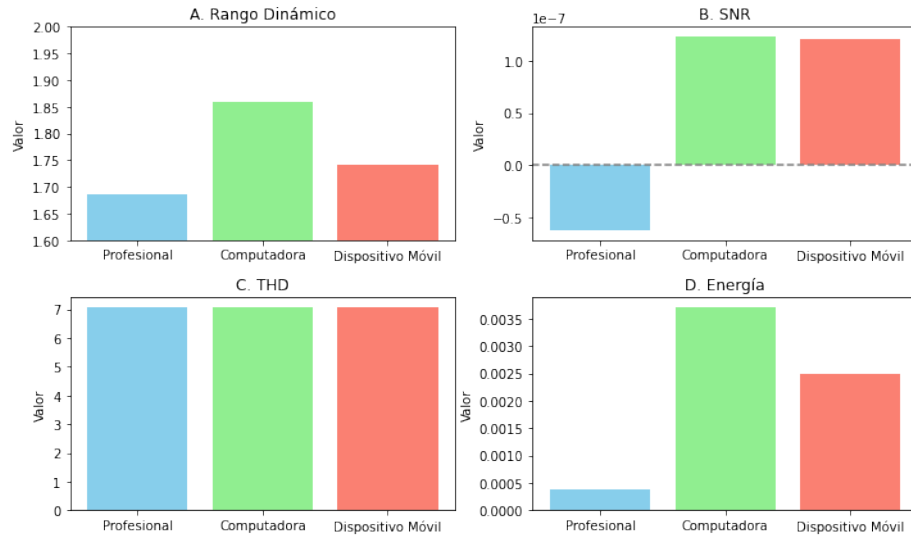


Figura 21: Evaluación de los dispositivos de grabación con base en: A) rango dinámico. B) SNR. C) THD y D) Energía

8.2. Análisis de tareas diadococinéticas

8.2.1. Análisis de diadococinecia /KA/

La evaluación integral de las características prosódicas de la tarea /KA/, Figura 22, muestra un mejor rendimiento al evaluar todas las características proporcionadas por Dis-Voice, utilizando un modelo de Máquina de soporte de vectores, ya que en dicho modelo refleja uniformidad en la clasificación. El rendimiento de este modelo se describe detalladamente en la Figura 23: la curva ROC tiene un área bajo la curva de 0.78, una precisión de 0.70, sensibilidad de 0.69 y especificidad de 0.70.

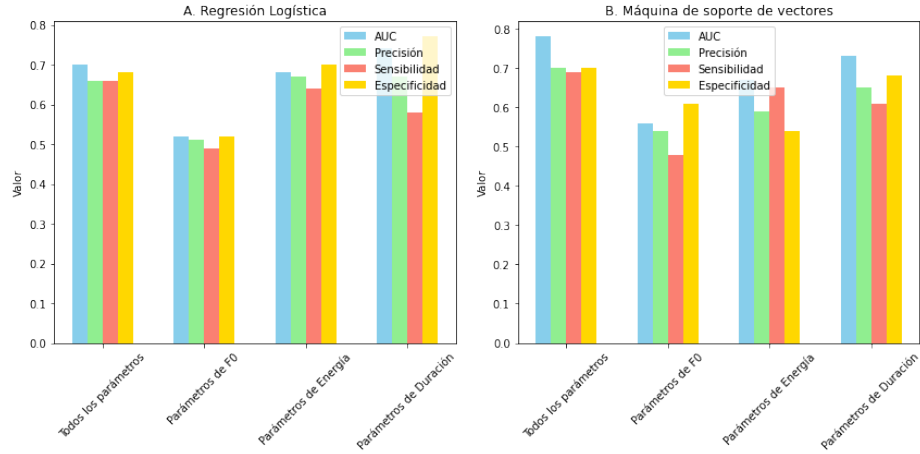


Figura 22: Evaluación detallada de características de la tarea: KA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

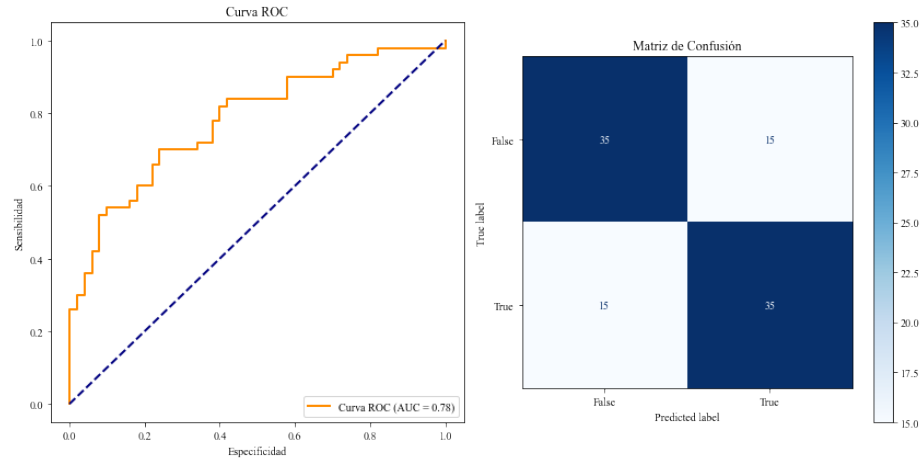


Figura 23: Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de KA

Con el objetivo de mejorar el rendimiento de dicho modelo se añadieron los parámetros descritos en sección 7.5.4. y se evaluaron las características de frecuencia dimensionadas en semitonos, en lugar de Hz como se realizó para las evaluaciones de la Figura 22. En la Figura 24 se comparan los resultados de area bajo la curva ROC, precisión, sensibilidad y especificidad de diversas experimentaciones. La experimentación 1 corresponde a la misma realizada en la la Figura 22 "Todos los parámetros", y las experimentaciones 5 a los resultados de implementar las nuevas características. Se añaden los mejores dos resultados de las cuatro pruebas realizadas en la experimentación 5.

En los cada modelo de clasificación se obtuvo un resultado diferente con base a la dimensional de frecuencia utilizada. El modelo de regresión logística funcionó mejor al utilizar semitonos. En contraste, el modelo de máquina de soporte de vectores funcionó mejor al utilizar la dimensionalidad original. Entre ambas experimentaciones se tiene un mejor rendimiento, detallado en la Figura 25, al implementar regresión logística con frecuencia en

semitonos

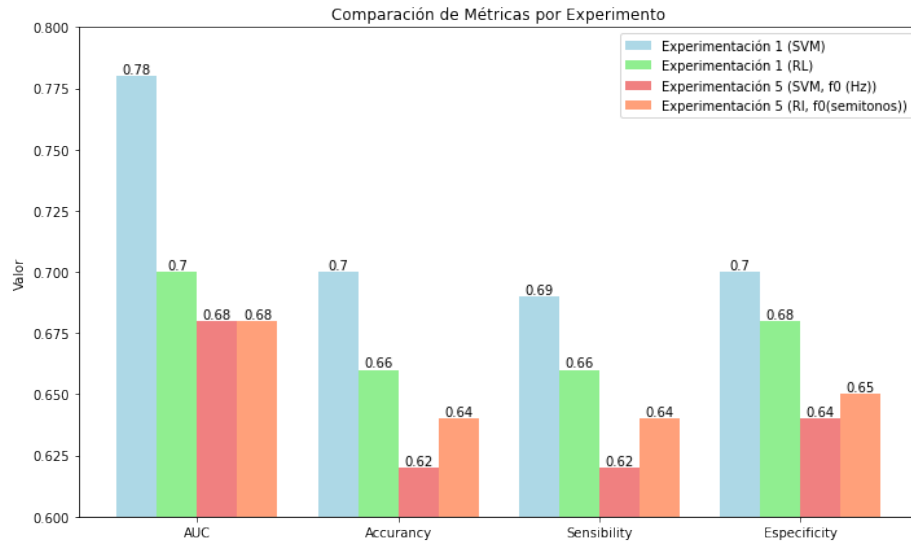


Figura 24: Comparativa de medidas de rendimiento de la tareas: KA en diversas experimentaciones.

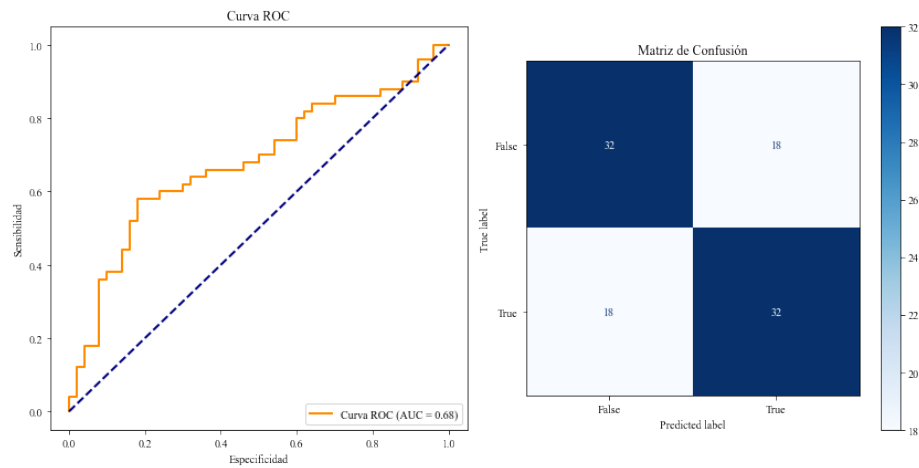


Figura 25: Análisis del rendimiento del modelo de regresión logística para la tarea de KA con parámetros de frecuencia dimensionados en semitonos

Se definen los principales coeficientes, con base en el modelo de regresión logística, para la evaluación de la tarea /KA/. Estos se muestran en la Figura 26, Los 3 principales coeficientes corresponden a las características: el valor máximo de la energía del primer segmento sordo (max1Eunvoiced), la kurtosis del cuarto croma (chr4Kurtosis) y el mínimo de la duración de los segmentos sordos (mindurunvoiced).

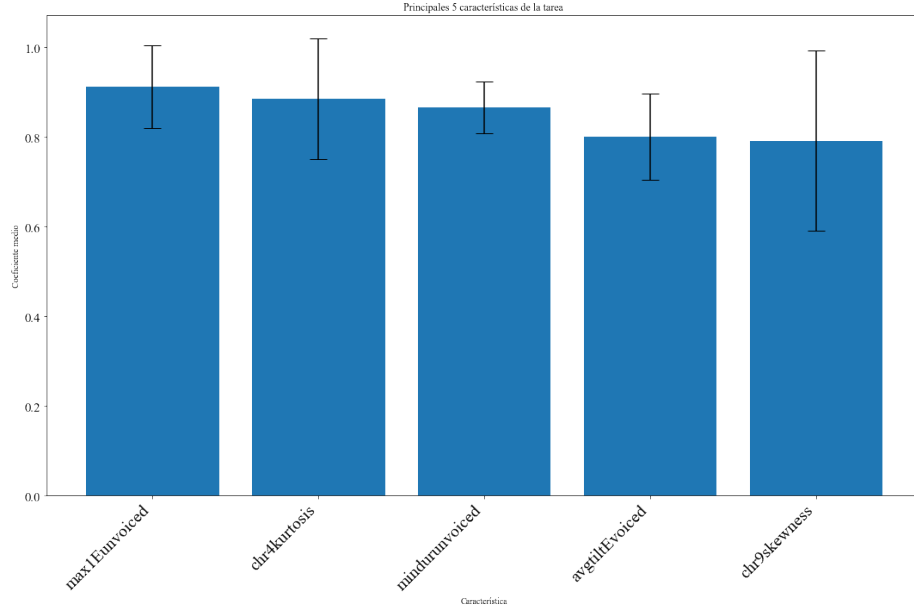


Figura 26: Principales coeficientes de la tarea KA con parámetros de frecuencia dimensionados en semitonos

8.2.2. Análisis de diadococinecia /PA/

Al analizar la tarea /PA/ en las principales agrupaciones de frecuencia, energía y duración (Figura 27). se observa un rendimiento deficiente al evaluar únicamente los parámetros de frecuencia, independiente del modelo de clasificación implementado.

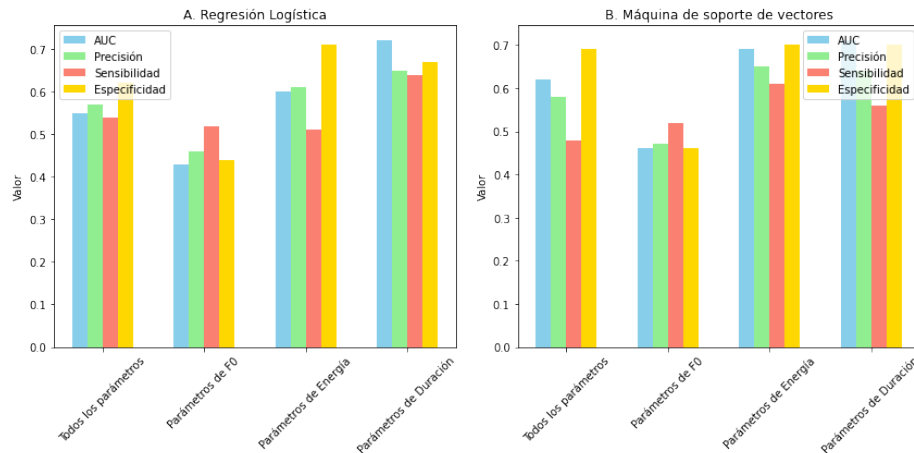


Figura 27: Evaluación detallada de características de la tarea: PA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

En esta tarea el análisis enfocado únicamente en la duración tiene mejores resultados, como se observa en la Figura 28. Al evaluar exclusivamente estas características con un modelo de regresión logística se tiene un área bajo la curva de 0.72, la matriz de confusión

asociada demuestra una homogeneidad para la clasificación con una precisión de 0.65, una sensibilidad de 0.64 y especificidad de 0.67.

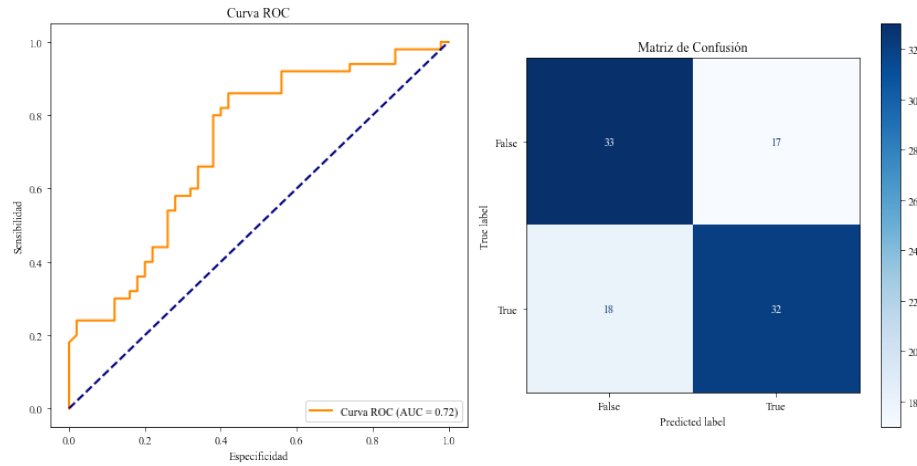


Figura 28: Análisis del rendimiento del modelo de regresión logística para la tarea de PA, evaluando características de duración

Ante el cambio dimensional en las características de frecuencia y la introducción del conjunto de cromas se observa una mejora en el modelo. La Figura 29 compara las principales medidas de rendimiento en las diversas experimentaciones. Se observa un mejor rendimiento al añadir los cromas y específicamente en la experimentación 5 (SVM, f0(semitonos)) al utilizar las medidas de frecuencia en semitonos. En esta experimentación aumenta el área bajo la curva; disminuye la especificidad del modelo en comparación a la experimentación 1, que corresponde al análisis de los parámetros de frecuencia, energía y duración en conjunto con un modelo SVM; pero se tiene una mejor distribución en las medidas de precisión y sensibilidad, el modelo mejora considerablemente su capacidad para predecir verdaderos positivos que corresponde a la clasificación de pacientes con Parkinson.

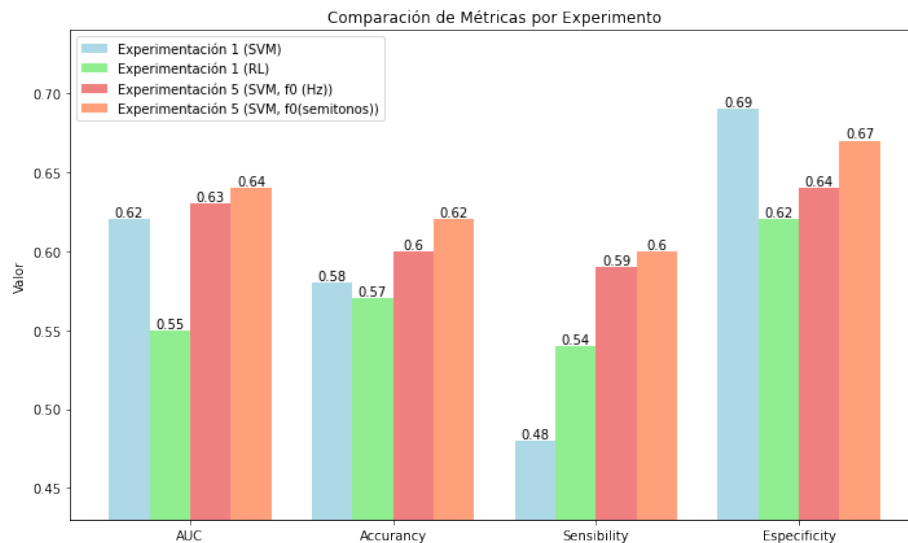


Figura 29: Comparativa de medidas de rendimiento de la tareas: PA en diversas experimentaciones.

En esta tarea se observa la influencia de la implementación de los cromas. En la Figura 30 se muestra como tres de los cinco principales coeficientes corresponden a características asociadas a este grupo: la curtosis del cromata 8, el promedio del cromata 13 y la asimetría del cromata 11.

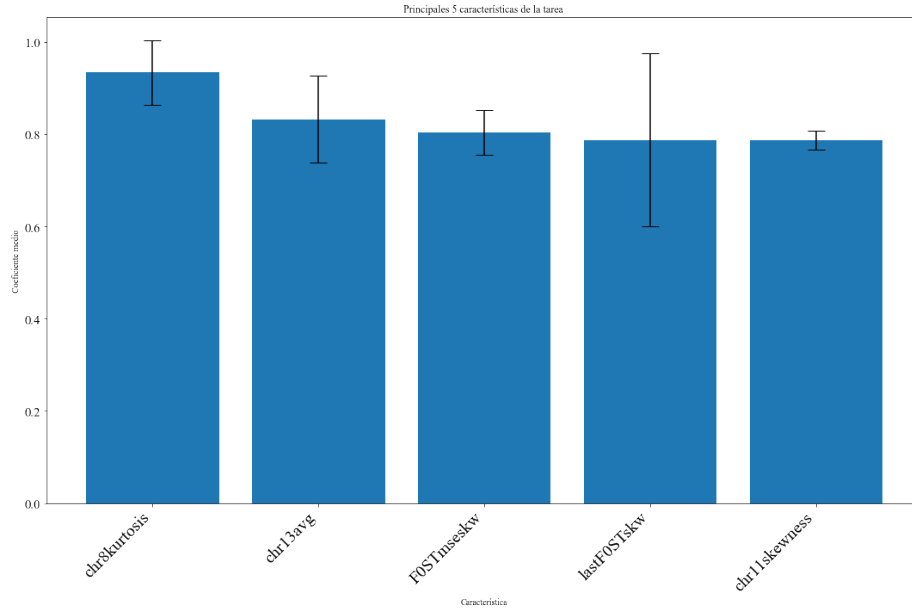


Figura 30: Principales coeficientes de la tarea PA con parámetros de frecuencia dimensionados en semitonos

Se detalla el comportamiento del modelo en la Figura 31. El AUC es igual al obtenido en el mejor modelo sin implementación de cromas (0.64), la precisión es de 0.62, la sensibilidad de 0.60 y la especificidad de 0.67. El modelo mejora su relación entre la clasificación de verdaderos positivos y verdaderos negativos respecto al modelo que no implementa cromas. Ya que como se muestra en la Figura 27 al evaluar todos los parámetros hay baja sensibilidad. Sin embargo, el mejor modelo sigue siendo el que evalúa únicamente los parámetros de duración.

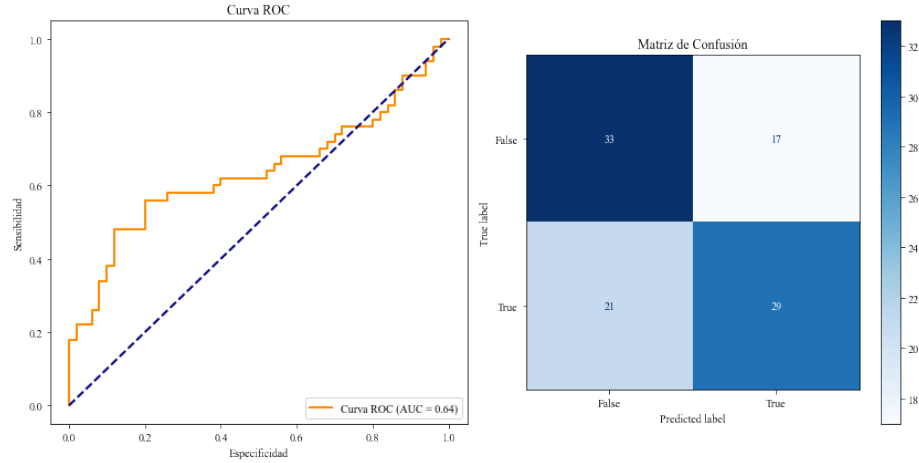


Figura 31: Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de PA con parámetros de frecuencia dimensionados en semitonos

8.2.3. Análisis de diadococinecia /TA/

La tarea TA se caracteriza por la clasificación de verdaderos negativos. En la Figura 32 se observa la tendencia en esta métrica para todas las evaluaciones. Sin embargo, ambos modelos son propensos a clasificar erróneamente los casos positivos por su baja sensibilidad. En esta tarea la evaluación aislada de los parámetros de frecuencia tiene mejores resultados. Ya que las medidas de precisión, sensibilidad y especificidad están equilibradas evitando el sesgo a la clasificación de positivos o negativos.

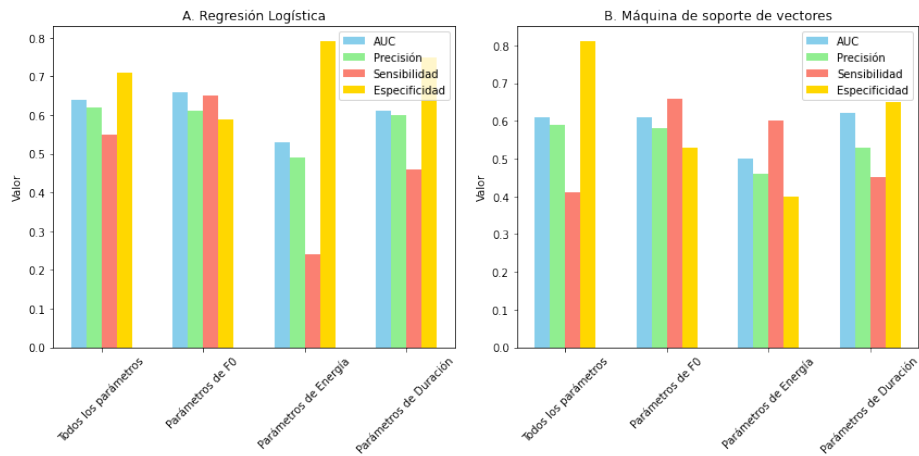


Figura 32: Evaluación detallada de características de la tarea: TA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

Entre ambos modelos de clasificación se tienen mejores resultados utilizando regresión logística. La Figura 33 detalla la matriz de confusión asociada al modelo, en este caso se tiene mayor tendencia a la clasificación correcta de verdaderos positivos (Pacientes con Parkinson) con una sensibilidad de 0.65. Sin embargo, la clasificación de verdaderos negativos disminuye

con un valor de especificidad de 0.59, en general el modelo tiene una precisión de 0.61.

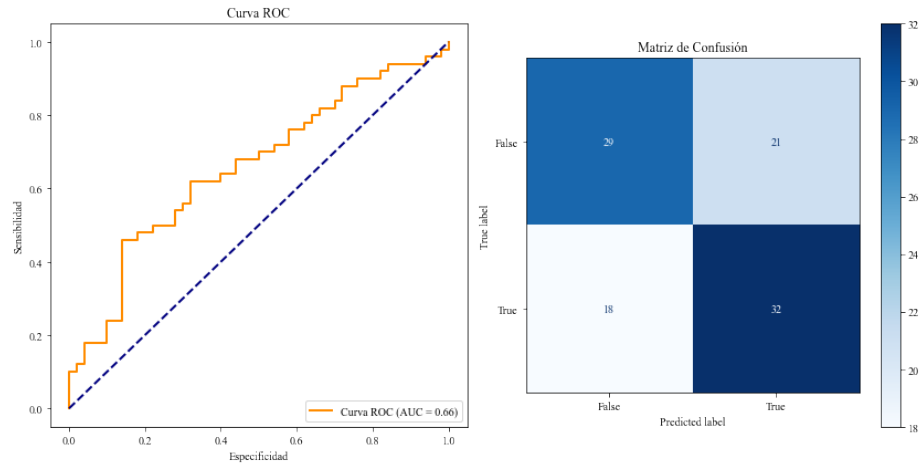


Figura 33: Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de TA: Evaluando características de frecuencia

con base en lo observado las características de frecuencia influyen significativamente en la evaluación de esta tarea. Por lo que implementar el cambio dimensional de frecuencia a semitonos no representa una mejora en el modelo. Como se observa en la Figura 34, al utilizar una frecuencia en semitonos se obtienen un menor rendimiento en comparación al uso de frecuencia en Hz.

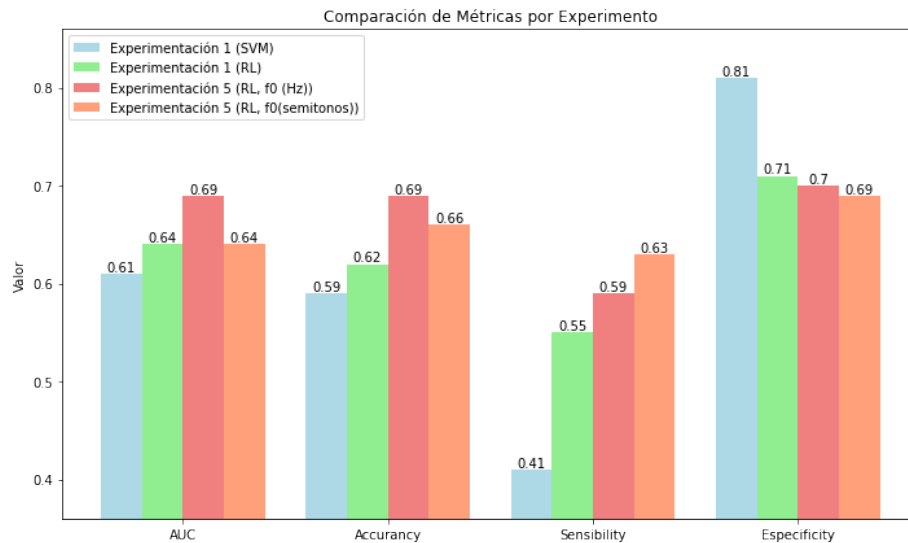


Figura 34: Comparativa de medidas de rendimiento de la tareas: TA en diversas experimentaciones.

El modelo mejora al implementar los cromas como se observa en la experimentación 5 de la Figura 34 con un modelo de regresión logística y en la Figura 35, donde los principales coeficientes corresponden a parámetros de frecuencia: el promedio y la asimetría de la inclinación de la estimación lineal de frecuencia para cada segmento sonoro; y parámetros relacionados con los cromas: promedio del croma 11 y asimetría del croma 4.

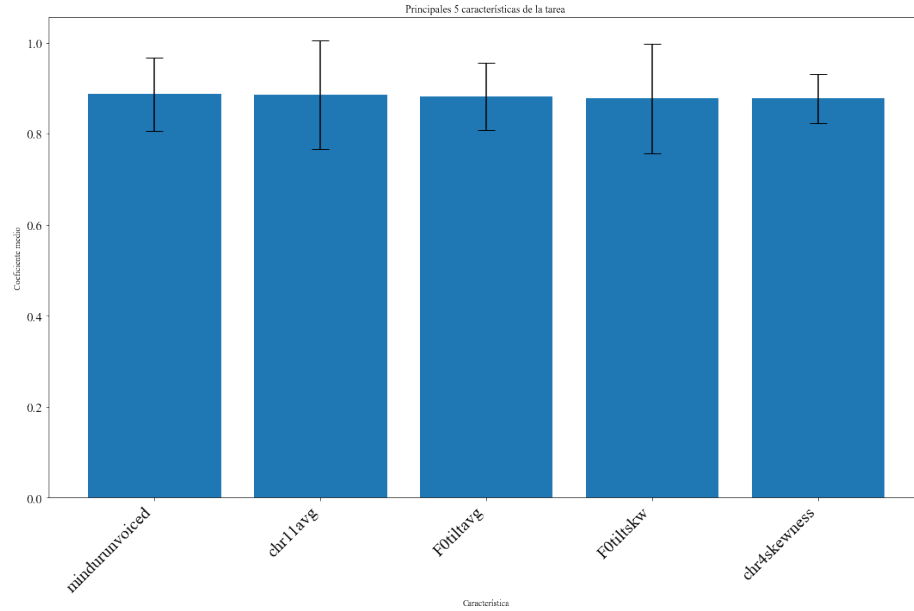


Figura 35: Principales coeficientes de la tarea TA con parámetros de frecuencia dimensionados en hercios

En consecuencia al implementar los cromas y realizar un análisis con las medidas de frecuencia en hercios se tiene un mejor rendimiento. El modelo obtenido de esta configuración de parámetros se observa en la Figura 36 con una tendencia por la clasificación correcta de los valores verdaderos positivos como se observa en la matriz de confusión. Con una sensibilidad de 0.59, una especificidad de 0.69 y un una precisión general de 0.69 el AUC es de 0.69

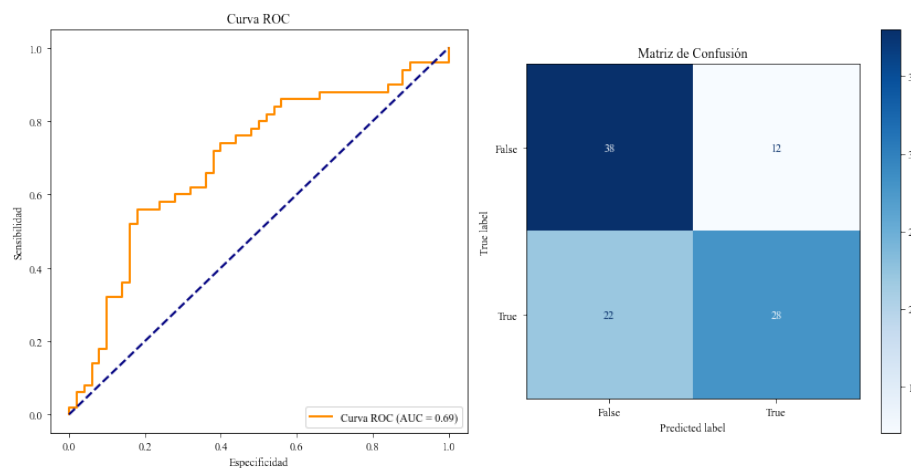


Figura 36: Análisis del rendimiento del modelo de regresión logística para la tarea de TA con parámetros de frecuencia dimensionados en hercios

8.2.4. Análisis de diadococinecia /PAKATA/

La tarea /PAKATA/ se caracteriza por la influencia de los parámetros de energía. En la Figura 37 se observa mayor estabilidad entre las medidas de rendimiento para ambos modelos al evaluar únicamente los parámetros de energía. Por el contrario, los parámetros de frecuencia y duración no tienen buenos resultados independientemente al modelo de clasificación las evaluaciones aisladas de estas características tienen una precisión menor a 0.50, lo que contribuye al bajo rendimiento al utilizar todos los parámetros.

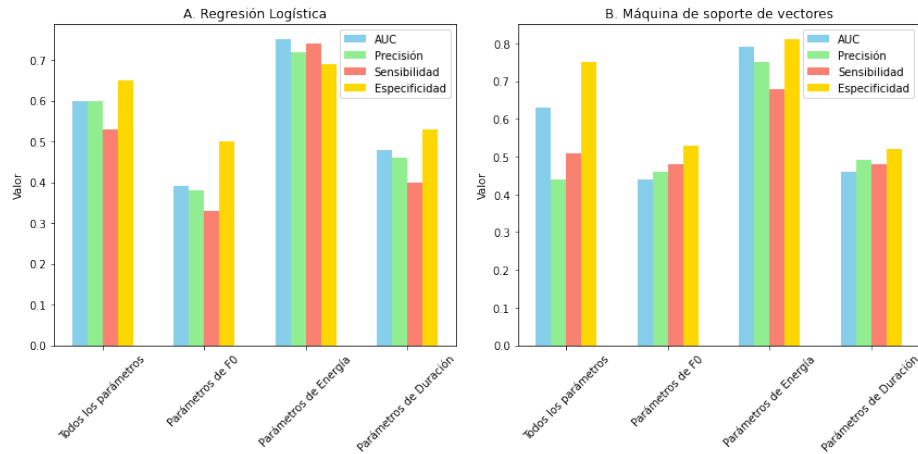


Figura 37: Evaluación detallada de características de la tarea: PAKATA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

Al evaluar las características de energía exclusivamente, (43), se tiene una precisión de 0.79, sensibilidad de 0.68 y especificidad de 0.81. El modelo clasifica mejor los valores verdaderos negativos. Cabe destacar que este es el mejor modelo obtenido entre todas las tareas y experimentaciones realizadas.

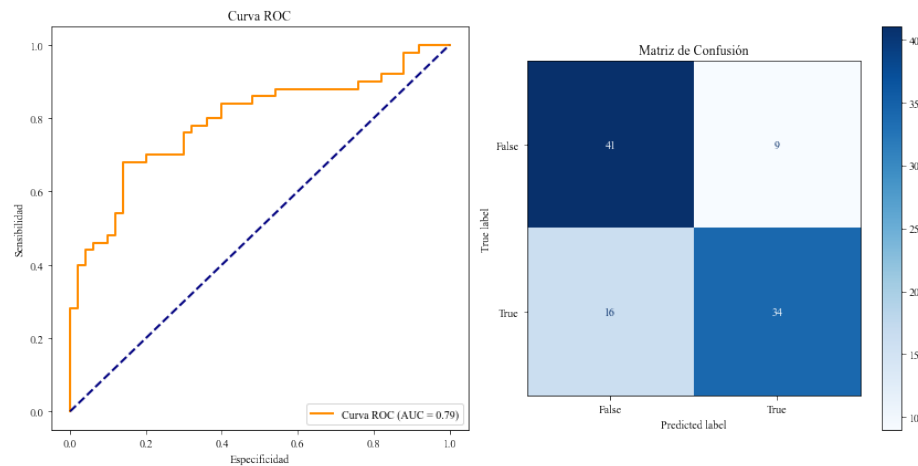


Figura 38: Análisis del rendimiento del modelo de máquina de soporte de vectores para la tarea de PAKATA

Al realizar el cambio dimensional de frecuencia se observa una mejora en la relación de clasificación de los verdaderos positivos y negativos (Figura 39). No obstante, las características de cromas son las que tienen mayor influencia, como se observa en las barras correspondientes a la experimentación 5 RL, $f_0(\text{Hz})$, la relación entre la precisión, sensibilidad y especificidad mejora significativamente. En la Figura 40 se detalla el rendimiento del modelo. Con un valor AUC con 0.74, precisión de 0.68, sensibilidad de 0.73 y especificidad de 0.65. Los resultados mejoran respecto a los resultados de la Figura 37 al evaluar todos los parámetros.

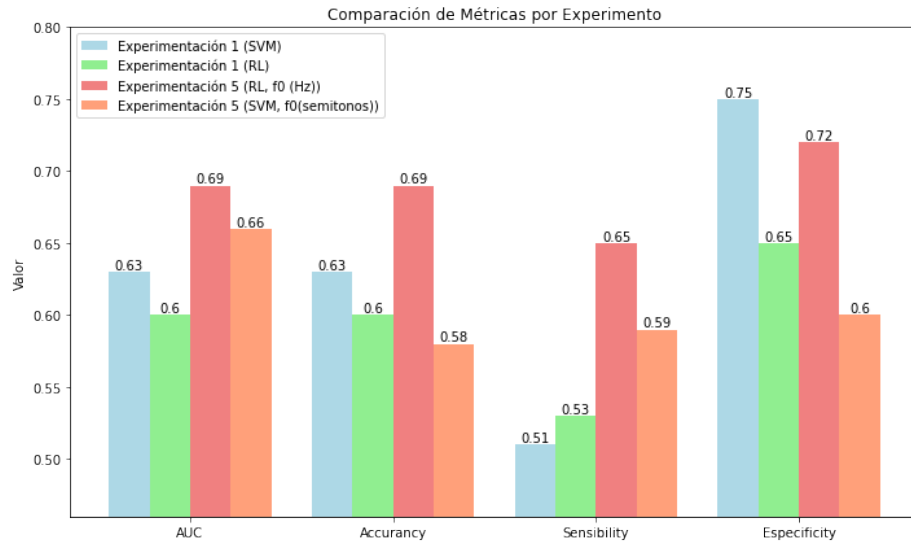


Figura 39: Comparativa de medidas de rendimiento de la tareas: PAKATA en diversas experimentaciones.

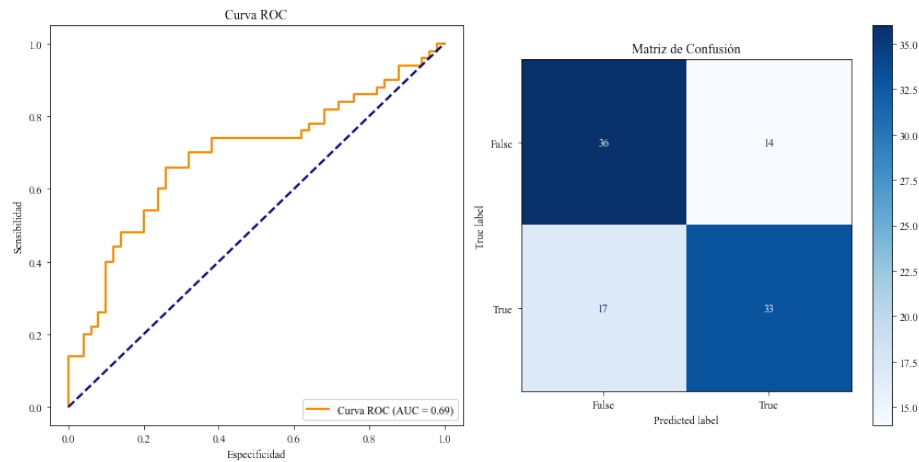


Figura 40: Análisis del rendimiento del modelo de regresión logística para la tarea de PATAKA con parámetros de frecuencia dimensionados en hercios

La influencia de los cromas se comprueba al evaluar los parámetros asociados a los principales coeficientes del modelo de regresión logística (Figura 41). Dos de los principales

parámetros son la curtosis del cromograma 12 y 7. Conjuntamente, se comprueba el peso de los parámetros de energía al tener la asimetría y curtosis del contorno de energía para los segmentos sonoros (Evoiced)

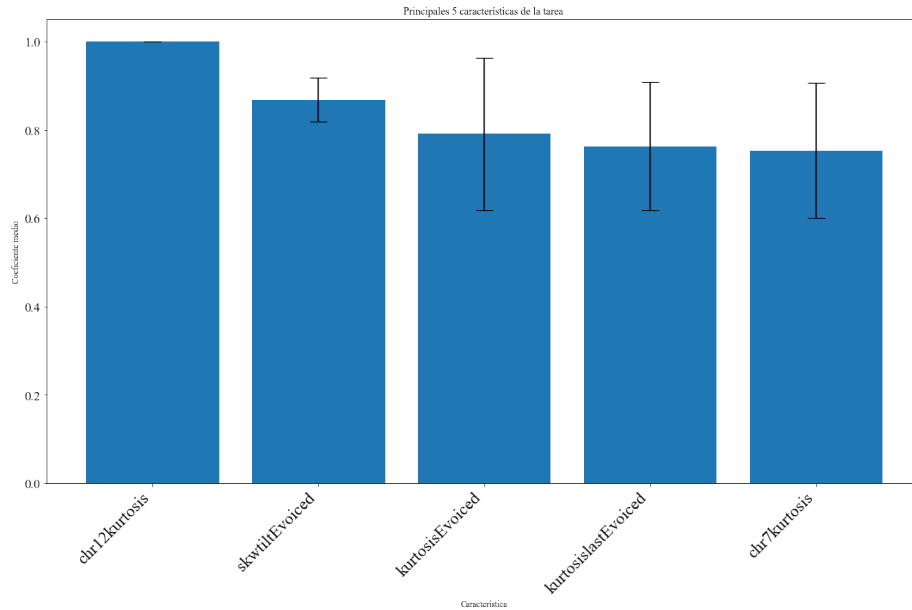


Figura 41: Principales coeficientes de la tarea PAKATA con parámetros de frecuencia dimensionados en hercios

El añadir los cromas mejora la evaluación conjunta de todos los parámetros. Sin embargo, el rendimiento de esta evaluación no supera el rendimiento del modelo al analizar exclusivamente los parámetros de energía.

8.2.5. Análisis de diadococinecia /PATAKA/

En esta tarea el mejor rendimiento lo tienen las pruebas evaluando todas las características y únicamente los parámetros de duración. Sin embargo, el rendimiento de estos sigue siendo bajo con una precisión de 0.66 y 0.69 respectivamente. Como se observa en la Figura 42 la evaluación con mayor precisión es la que considera los parámetros de duración utilizando regresión logística, sin embargo este modelo tiene una desproporción entre la sensibilidad y especificidad. Lo que se refleja en una mejor capacidad para la clasificación de verdaderos negativos.

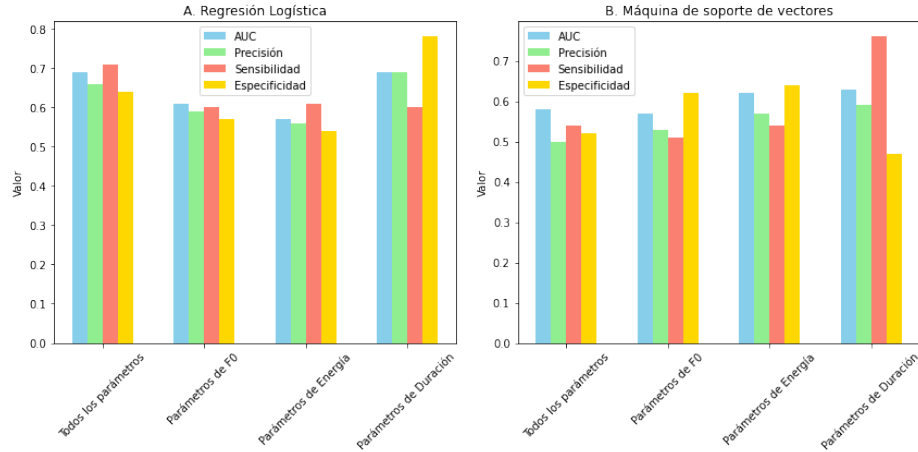


Figura 42: Evaluación detallada de características de la tarea: PATAKA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

La evaluación de todos los parámetros, bajo el mismo modelo de regresión logística, tiene mejor relación entre la sensibilidad (0.71) y especificidad (0.64). La Figura 43 describe este análisis. La matriz de confusión muestra la mejor relación para la predicción de verdaderos positivos (35) y verdaderos negativos (31), el área bajo la curva ROC de 0.69 y la precisión del modelo de 0.66, más baja que la precisión al evaluar únicamente la duración, pero la relación entre las medidas de rendimiento es mejor.

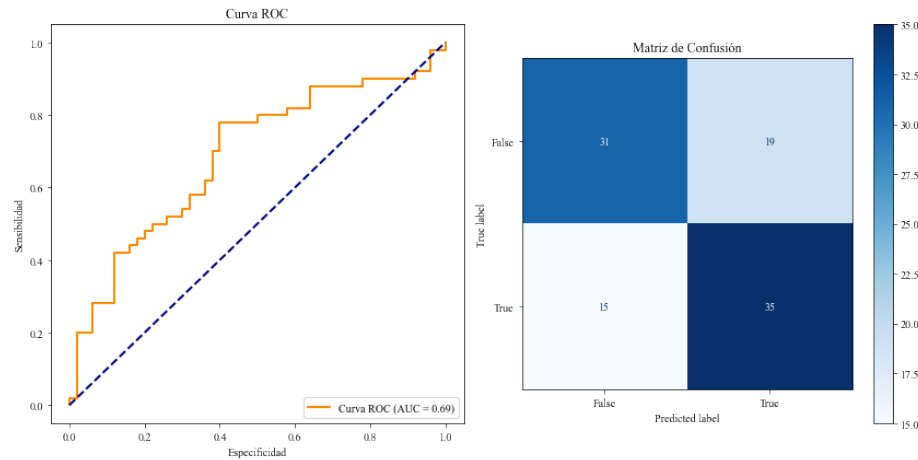


Figura 43: Análisis del rendimiento del modelo regresión logística para la tarea de PATAKA

Al añadir el conjunto de cromas a la evaluación se observan mejores resultados y una mejor relación entre las medidas de rendimiento. (Figura 44) En este caso la precisión aumenta a 0.68, sensibilidad a 0.73 y la especificidad 0.65. En general, los mejores resultados se obtienen al utilizar el modelo de regresión logística. En esta tarea el cambio de frecuencia de Hercios a semitonos no tiene un impacto tan significativo, por lo que se considera que los mejores resultados se obtuvieron manteniendo la dimensional de frecuencia en hercios.

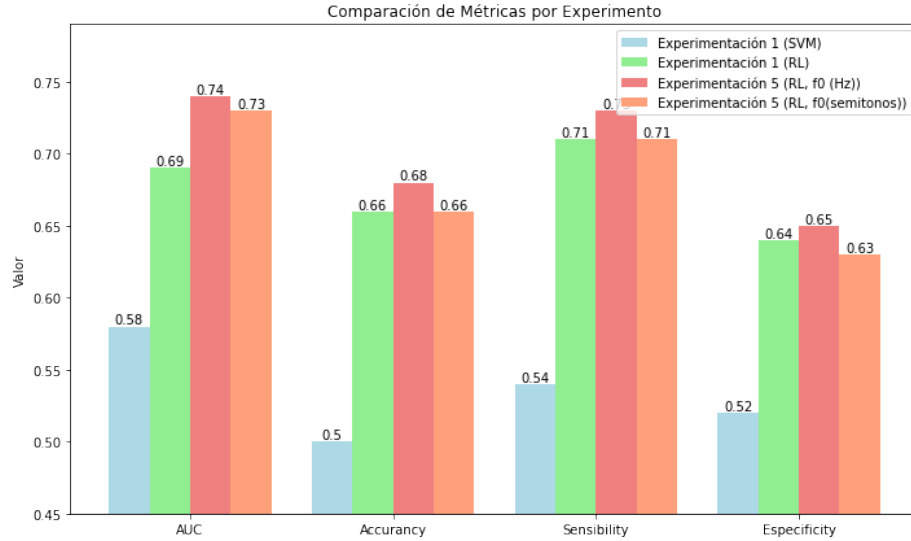


Figura 44: Comparativa de medidas de rendimiento de la tareas: PATAKA en diversas experimentaciones.

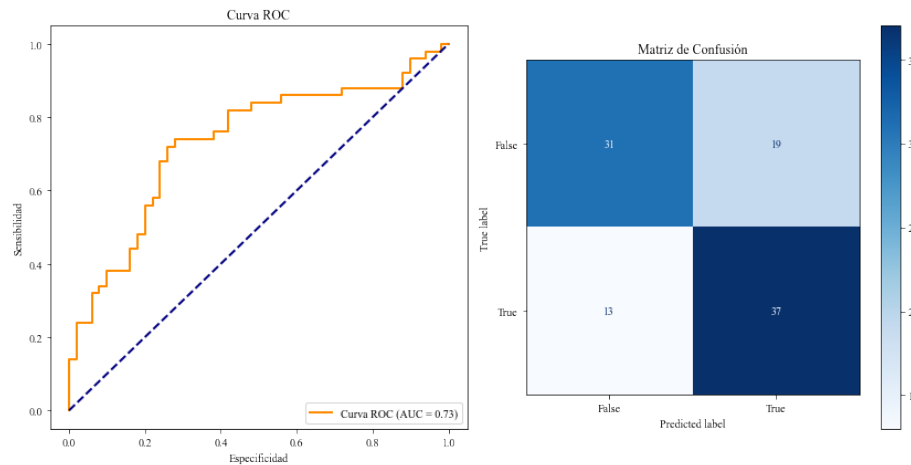


Figura 45: Análisis del rendimiento del modelo de regresión logística para la tarea de PATAKA con parámetros de frecuencia dimensionados en hercios

En la Figura 46 se observan los principales parámetros asociados a los coeficientes del modelo de regresión logística que evalúa todas las métricas (frecuencia en hercios, energía, duración y cromas). Se destaca la influencia de la kurtosis del croma 12 y croma 7. Al igual que la kurtosis de los parámetros de energía: Contorno de energía para segmentos sonoros(kurtosisEvoiced) y Energía del último segmento sonoro (kurtosislastEvoiced).

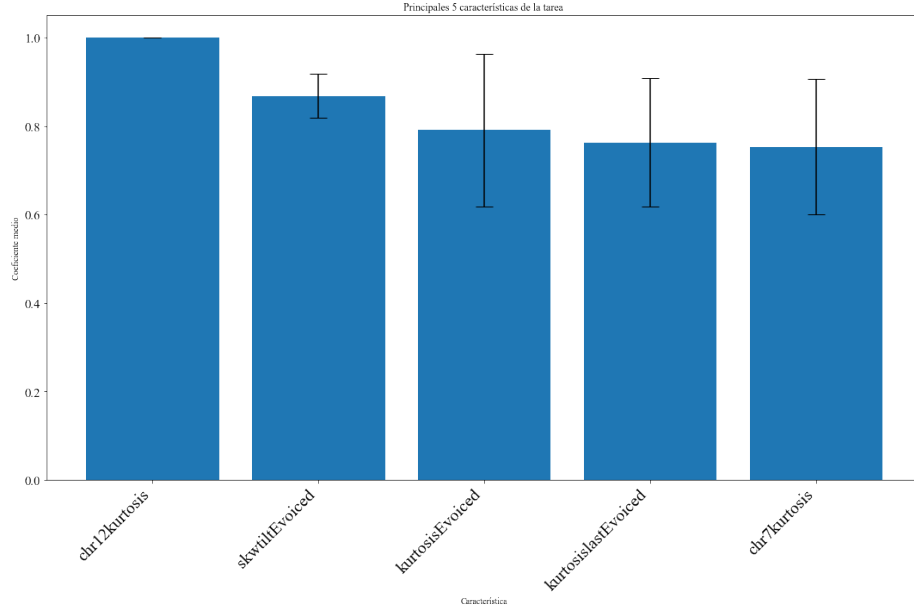


Figura 46: Principales coeficientes de la tarea PAKATA con parámetros de frecuencia dimensionados en hercios

8.2.6. Análisis de diadococinecia /PETAKA/

La Figura 47 muestra un rendimiento similar para todas las evaluaciones de la tarea /PETAKA/. La precisión y AUC en su mayoría son menores a 0.60. Únicamente las experimentaciones relacionadas al análisis aislado de energía tiene resultados de precisión de 0.65 y 0.61 para los modelos RL y SVM respectivamente. Sin embargo, el bajo rendimiento de frecuencia y duración influyen al analizar todos los parámetros en conjunto.

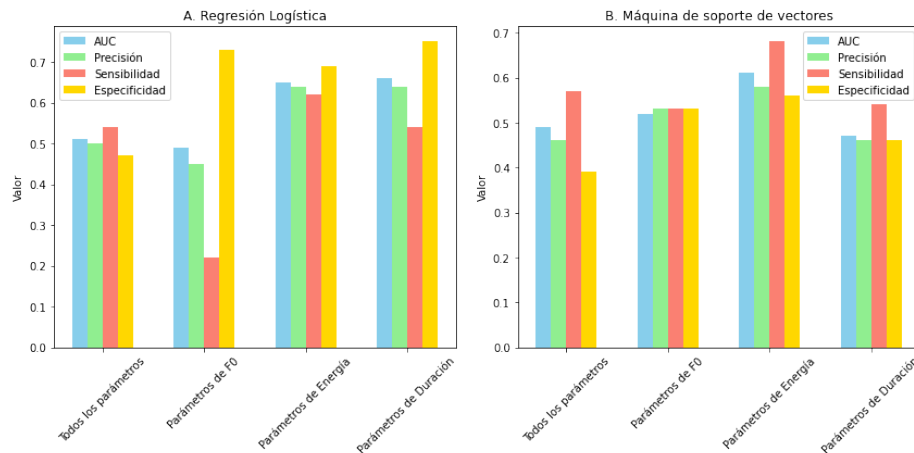


Figura 47: Evaluación detallada de características de la tarea: PETAKA. A) Modelo de regresión logística. B) Modelo de Máquina de Soporte de Vectores

En la Figura 49 se observa el impacto de incorporar los cromas al conjunto de caracterís-

tas. Esta tarea tiene un comportamiento similar a otras analizadas, como lo es /PAKATA/, donde el cambio dimensional de la frecuencia no tiene un impacto considerable, pero si la adición de los cromas, lo que se refleja en las barras correspondientes a la experimentación 5 (RL, F0 (Hz)).

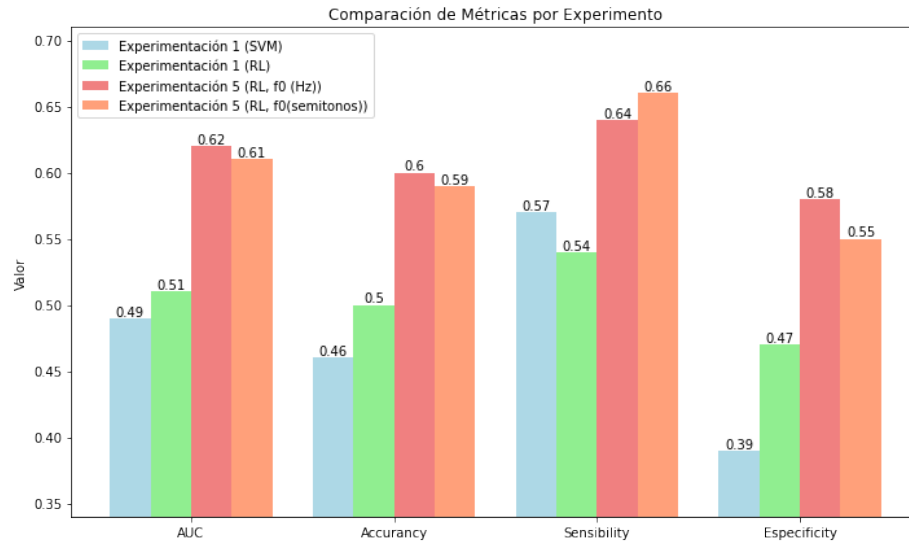


Figura 48: Comparativa de medidas de rendimiento de la tareas: PETAKEA en diversas experimentaciones.

La Figura 49 comprueba la influencia de las características asociadas a la energía. Las medidas de máximo, promedio y desviación estándar de la energía del primer segmento sonoro son tres de las principales características, acompañadas de la kurtosis del croma 4 y el máximo del croma 11.

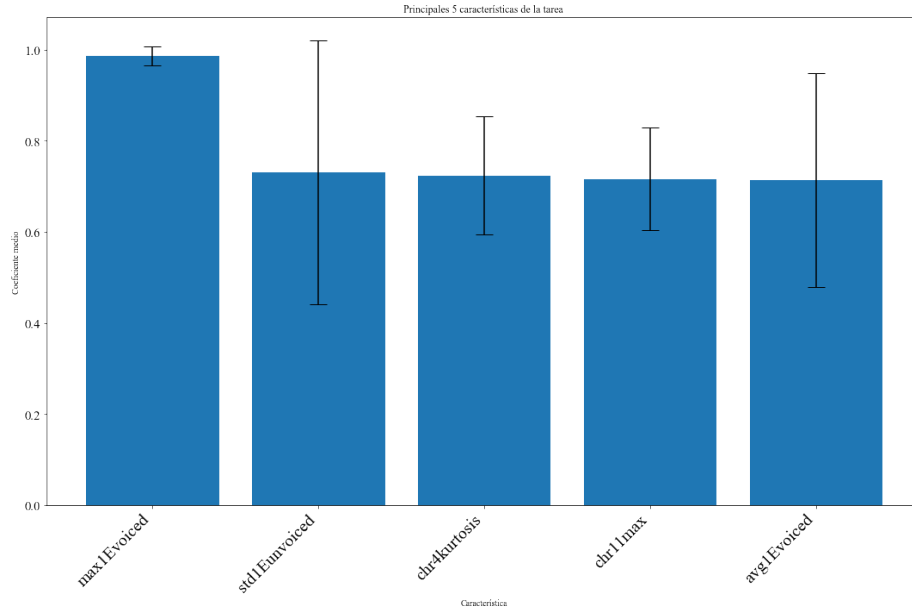


Figura 49: Principales coeficientes de la tarea PETAKA con parámetros de frecuencia dimensionados en hercios

Si embargo, la evaluación conjunta de los parámetros de frecuencia en hercios, energía, duración y cromas tiene menor rendimiento al obtenido al analizar individualmente los parámetros de energía para la tarea /PETAKA/.

8.2.7. Resultados de las tareas diadococinecias con la implementación de cromas

La Figura 50 muestra las diferencias entre el mejor modelo de cada tarea diadococinetica al:

1. evaluar todos los parámetros extraídos de DisVoice en conjunto
2. al evaluar el cambio dimensional de frecuencia a semitonos y añadir las características asociadas a los cromas

Se observa el comportamiento de las medidas de rendimiento de cada modelo con base al área bajo la curva ROC, la precisión, sensibilidad y especificidad. En la tarea Ka se destaca una alta medida de AUC (0.78) sobre el resto de las evaluaciones en la primera experimentación utilizando SVM. Esto se relaciona con la homogeneidad de las otras métricas asociadas (Precisión, especificidad y sensibilidad) todas con un valor 0.70. Al implementar los cromas el rendimiento de esta diadococinecia disminuye. La tarea PA tiene una mejora considerable respecto a su sensibilidad de la evaluación 1 (0.48) a la evaluación 2 (0.60), esto se relaciona con la capacidad de predicción correcta del modelo para verdaderos positivos (pacientes con Parkinson); resultando en un aumento de AUC (0.64) y precisión (0.62). Sin embargo, disminuye su valor para clasificación de verdaderos negativos (Pacientes sanos)

con una especificidad de 0.69 a 0.67. La Tarea TA también tiene una mejora en su clasificación de verdaderos positivos, pasando de 0.41 de sensibilidad a 0.63. Se reduce el valor de especificidad del modelo, pero el AUC y precisión mejoran aumentando ambas a 0.69.

Todas las medidas de rendimiento de la tarea PATAKA aumentan al añadir las características cromáticas a un AUC de 0.74, precisión de 0.68, sensibilidad de 0.73 (el valor más alto entre todos los modelos) y especificidad de 0.65. En la primera evaluación, al considerar todas las características, el rendimiento del modelo fue muy bajo con todas las medidas de rendimiento por debajo de 0.58. La tarea PAKATA aumento su sensibilidad de 0.51 a 0.65. Al igual que las otras tareas, pese la disminución de especificidad, la precisión y AUC del modelo mejoran de 0.63 a 0.69. PETAKA destaca por sus bajas medidas de rendimiento. Diferente a la mayoría de tareas, la especificidad en la primera evaluación es menor a la sensibilidad del modelo, por lo que inicialmente es mejor en el reconocimiento de valores verdaderos positivos (pacientes con Parkinson) sobre los verdaderos negativos (pacientes sanos). Al añadir los cromas mejoran todas las medidas de desempeño: AUC pasa de 0.49 a 0.62, precisión 0.46 a 0.6, la sensibilidad de 0.57 a 0.64 y la especificidad del modelo pasa de 0.39 a 0.58.

Para todos los modelos el añadir los cromas mejora la capacidad de identificar a pacientes con Parkinson (verdaderos positivos). Siendo más significativo el cambio en las diadococinecias: PA, TA, PATAKA y PAKATA. Esta mejora impacta en la precisión y AUC de todos los modelos al reducir la diferencia marcada entre los valores de sensibilidad y especificidad. Los modelos TA y PAKATA tienen un mejor desempeño en el reconocimiento de pacientes sanos (verdaderos negativos). Al incluir los cromas en la tarea PAKATA se tiene un mejor desempeño por su relación equitativa en la clasificación de pacientes con Parkinson y sujetos sanos.

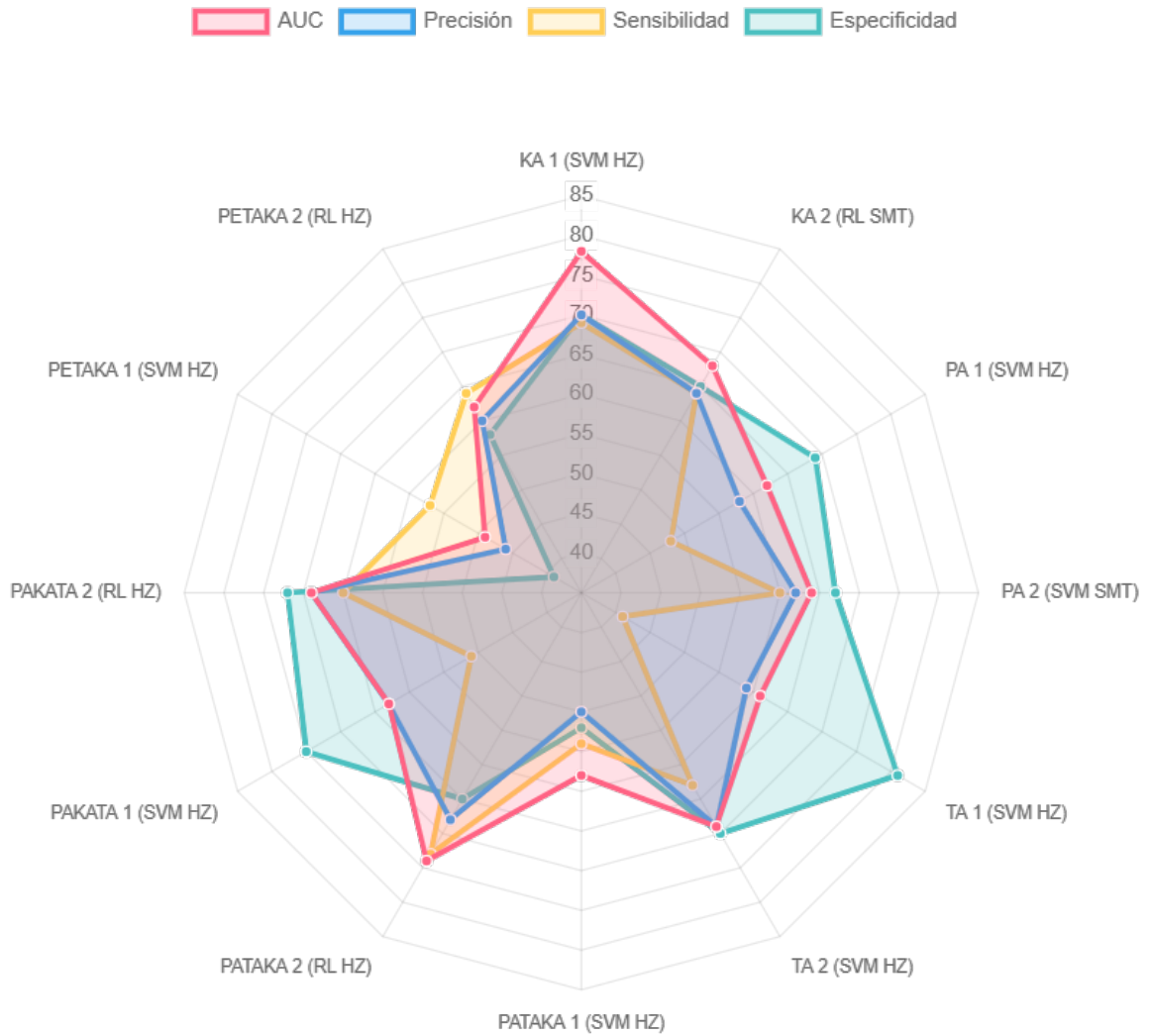


Figura 50: Gráfico de radar de evaluación de modelos: iniciales considerando características de frecuencia, energía y duración (1) y modelos con la implementación de cromas y cambio dimensional de frecuencia (2) para cada tarea diadocinética

En este trabajo se realizó la identificación y análisis de diversos parámetros del habla relacionados con la disartria hipocinética y su afección en la prosodia de pacientes con la Enfermedad de Parkinson. En conjunto con la evaluación de rendimiento de distintos dispositivos de grabación en un entorno controlado.

La evaluación de tres dispositivos de adquisición de audio: Un micrófono profesional, el micrófono integrado en una computadora portátil y el micrófono integrado en un dispositivo móvil de gama media se realiza con el objetivo de determinar la influencia del dispositivo de adquisición en la fidelidad de los audios, con base en las métricas de calidad de audio. Pese a que las grabaciones realizadas con el dispositivo móvil y la computadora fueron realizadas en simultaneo, la señal adquirida mediante el micrófono integrado en una computadora tiene una disminución de amplitud, lo que representa pérdida de información a lo largo de la grabación. En diversos estudios se ha comprobado el bajo rendimiento de los micrófonos integrados en las computadoras. [69] [70]. Estos dispositivos tienden medidas de frecuencia, jitter y shimmer mayores a las de un dispositivo de grabación de disco duro de alta calidad, lo que refleja una saturación en el sonido original [71].

En el análisis de las señales obtenidas con el micrófono integrado en la computadora, las muestras tienen un rango dinámico mayor a los otros dos dispositivos evaluados, debido a las variaciones abruptas en la amplitud de la señal al inicio y final de las tareas. Estos picos y valles son críticos para la evaluación precisa de las muestras y su extracción de características. En pacientes con enfermedad de Parkinson, los cambios bruscos en el habla son comunes. Por lo regular los pacientes inician el discurso de forma lenta y al final tienen un aumento excesivo del ritmo e intensidad. Por lo que esta saturación en los audios al inicio y final es contraproducente en la extracción de características.

Por su parte los dispositivos móviles se han adaptado para su uso multidisciplinario de forma eficiente. Independientemente a la gama de los dispositivos en su mayoría tienen

integrados un mínimo de dos micrófonos, con la finalidad de reducir el ruido y tener archivos de audio más limpios. En los estudios mencionados se observan una similitud entre los dispositivos profesionales y los integrados en dispositivos móviles. Como se observa en la Figura 21 un dispositivo móvil es una buena alternativa al uso de dispositivo profesionales, en términos de accesibilidad y costos, en un entorno controlado como una cabina de grabación.

El análisis de la prosodia en señales de audio se fundamenta en las características de frecuencia fundamental, energía y duración. Dichas características varían entre los Sujetos Sanos y pacientes con Parkinson en diferente medida, dependiendo de la tarea diadococinetica evaluada. Entre las tareas que presentan mayores dificultades para los pacientes con Parkinson, con disartria hipocinética, es la pronunciación de $/KA/$. Esta tarea requiere una coordinación precisa, ya que implica una oclusión completa del tracto vocal al inicio de la sílaba ("K"), seguida de una rápida liberación de aire ("A") [72], combinando varios subsistemas en un corto periodo de tiempo. Debido a esta evidente complejidad y diferenciación entre pacientes con Parkinson y sujetos sanos, la tarea $/KA/$ presenta los mejores resultados en la evaluación de todas las características juntas, sin el uso de cromas. El modelo de clasificación de maquina de soporte de vectores tiene un equilibrio en la clasificación de verdaderos positivos (pacientes con Parkinson) y verdaderos negativos (Pacientes sanos) con un rendimiento, respecto a su precisión, de 0.70.

El resto de tareas diadococineticas evaluadas ($/PA/$, $/TA/$, $/PAKATA/$ y $/PETAKA/$) mejora significativamente su relación de clasificación de verdaderos positivos y verdaderos negativos al implementar las características cromáticas, ya que todas estas tienen a ser mejores para la clasificación de sujetos sanos (verdaderos negativos). Diferente a la tarea $/PATAKA/$, que tiene una relación consistente entre el reconocimiento de verdaderos positivos (Pacientes con Parkinson) y verdaderos negativos (Sujetos sanos) en todos los conjuntos de características evaluados: energía, duración, frecuencia fundamental y los tres grupos juntos. Al añadir el conjunto de cromas es la tarea que mejores resultados tiene en su relación de sensibilidad, especificidad y precisión. Con la particularidad para clasificar mejor a los pacientes con Parkinson entre todos los modelos evaluados.

En diferentes estudios relacionados al análisis del habla enfocado a las emociones, la detección de ideas suicidas [73], suplantación de voz [74] y enfermedades relacionadas con el sistema respiratorio, como el COVID-19 [75], todos ellos enfocados en la evaluación de prosodia, se ha visto la influencia de las características de cromas. Estos resultados destacan el impacto positivo de incorporar características de cromas a las tareas diadococinéticas, lo cual mejora significativamente el rendimiento del modelo de regresión logística. Esta mejora se atribuye a la influencia de los cromas en los coeficientes del modelo, lo que sugiere que estas características aportan información relevante y distintiva que fortalece la capacidad predicativa del modelo. La conversión dimensional de las características de frecuencia de Hz a semitonos no mostró una mejora significativa en los modelos de clasificación. En su mayoría, las experimentaciones arrojaron mejores resultados al mantener la dimensionalidad en Hz y añadiendo los cromas, debido al contenido espectral distintivo que poseen los cromas.

- El micrófono integrado de un celular puede reproducir los resultados de un micrófono profesional al grabar en un entorno controlado, como una cabina de grabación. Destacando que un celular es más barato y accesible en diferentes lugares, lo que lo convierte en una excelente opción para la adquisición de muestras, eliminando así la limitación de acceso a dispositivos profesionales que implican un costo elevado.
- La regresión logística en aprendizaje supervisado proporciona coeficientes interpretables, lo que permite comprender cómo cada característica de energía, frecuencia, duración y cromas contribuye a la clasificación de pacientes con Parkinson frente a sujetos sanos. La identificación de las características más relevantes en el modelo, conduce a una mejor comprensión de la enfermedad y a una mejora en el diagnóstico diferencial.
- Las características de energía, duración y frecuencia fundamental proporciona información valiosa para la evaluación diadococinética, con diferentes niveles de relevancia. Sin embargo, se observa que la evaluación conjunta de todas las características tiende a lograr un mejor equilibrio en la clasificación de pacientes con Parkinson frente a sujetos sanos. Aunque algunas características individualmente pueden conducir a mejores resultados en la clasificación de sujetos sanos, La evaluación integral de todas las características es fundamental para alcanzar una clasificación óptima de pacientes con Parkinson.
- La inclusión de cromas en el análisis de prosodia de pacientes con Parkinson presenta un impacto significativo en la sensibilidad de todas las tareas diadococinéticas. Reflejado en la capacidad mejorada del modelo para clasificar pacientes con Parkinson. Los cromas, al capturar contenido espectral específico, ofrecen información crucial sobre las características vocales asociadas con la disartria hipocinética.

- Considerando que el micrófono de un celular puede ofrecer resultados comparables a los de un micrófono profesional en un entorno controlado como una cabina de grabación, se recomienda evaluar su desempeño en condiciones menos controladas. Probar su funcionalidad en ambientes más accesibles, donde la presencia de ruido ambiental es inevitable pero se puede minimizar. Por ejemplo, realizar pruebas en entornos con la menor cantidad de ruido posible fuera de una cabina de grabación podría proporcionar una perspectiva más realista de su utilidad y precisión en situaciones cotidianas.
- Se sugiere explorar la viabilidad de analizar todas las tareas diadococinéticas en conjunto utilizando un algoritmo supervisado de regresión logística, al agregar parámetros de cromas a los parámetros originales proporcionados por DisVoice. es importante buscar formas de implementarla para simplificar el proceso y reducir la necesidad de separar manualmente las tareas y características.
- Se recomienda continuar investigando el estado del arte en otras investigaciones relacionadas con la evaluación de prosodia, enfocadas en diferentes condiciones o patologías distintas al Parkinson. Ya que al igual que los cromas, la implementación de otras características puede influir en la mejora de modelos de clasificación, enriquecer el análisis y entender los efectos de la disartria hipocinética en la prosodia para un diagnóstico temprano de la enfermedad de Parkinson.
- Se recomienda evaluar este sistema considerando la etapa de la enfermedad en la que se encuentra cada paciente. Es importante realizar evaluaciones longitudinales del rendimiento de los modelos para determinar si es posible detectar cambios en la voz en las etapas tempranas de la enfermedad.

-
- [1] P. Foundation, *What is Parkinson's?* Accessed April 21, 2023. dirección: <https://www.parkinson.org/understanding-parkinsons/what-is-parkinsons>.
 - [2] W. H. Organization, *Parkinson disease: a public health approach: technical brief*. World Health Organization, 2022, iv, 21 p.
 - [3] N. I. of Health, *Parkinson's Disease: Challenges, Progress, and Promise*, Accessed April 23, 2023. dirección: <https://www.ninds.nih.gov/current-research/focus-disorders/focus-parkinsons-disease-research/parkinsons-disease-challenges-progress-and-promise>.
 - [4] S. Lehericy, G. Bensimon y M. Vidailhet, "Parkinsonian Syndromes," en *Brain Mapping*, A. W. Toga, ed., Waltham: Academic Press, 2015, págs. 769-785, ISBN: 978-0-12-397316-0. DOI: <https://doi.org/10.1016/B978-0-12-397025-1.00088-9>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780123970251000889>.
 - [5] Á. Sánchez-Ferro, M. Elshehabi, C. Godinho et al., "New methods for the assessment of Parkinson's disease (2005 to 2015): A systematic review," *Movement Disorders*, vol. 31, n.º 9, págs. 1283-1292, 2016. DOI: <https://doi.org/10.1002/mds.26723>. eprint: <https://movementdisorders.onlinelibrary.wiley.com/doi/pdf/10.1002/mds.26723>. dirección: <https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.26723>.
 - [6] "Automatic Assessment of Parkinsonian Speech," vol. 1295, J. I. Godino-Llorente, ed., 2020. DOI: 10.1007/978-3-030-65654-6. dirección: <http://link.springer.com/10.1007/978-3-030-65654-6>.
 - [7] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman y L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, págs. 1264-1271, 5 mayo de 2012, ISSN: 00189294. DOI: 10.1109/TBME.2012.2183367.
 - [8] M. Hireš, M. Gazda, P. Drotár, N. D. Pah, M. A. Motin y D. K. Kumar, "Convolutional neural network ensemble for Parkinson's disease detection from voice recordings," *Computers in Biology and Medicine*, vol. 141, pág. 105 021, feb. de 2022, ISSN: 0010-4825. DOI: 10.1016/J.COMPBIOMED.2021.105021.

- [9] W. Liu, J. Liu, T. Peng et al., "Prediction of Parkinson's disease based on artificial neural networks using speech datasets," *Journal of Ambient Intelligence and Humanized Computing*, págs. 1-14, abr. de 2022, ISSN: 18685145. DOI: 10.1007/S12652-022-03825-W/METRICS. dirección: <https://link.springer.com/article/10.1007/s12652-022-03825-w>.
- [10] A. J. Noyce, R. N. Rees, A. P. Acharya y A. Schrag, "An early diagnosis is not the same as a timely diagnosis of Parkinson's disease," *F1000Research* 2018 7:1106, vol. 7, pág. 1106, jul. de 2018, ISSN: 1759796X. DOI: 10.12688/f1000research.14528.1. dirección: <https://f1000research.com/articles/7-1106>.
- [11] W. Poewe, K. Seppi, C. M. Tanner et al., "Parkinson disease," *Nature reviews. Disease primers*, vol. 3, págs. 1-21, mar. de 2017, ISSN: 2056-676X. DOI: 10.1038/NRDP.2017.13. dirección: <https://pubmed.ncbi.nlm.nih.gov/28332488/>.
- [12] A. Adcock y J. E. Frey, "Can telemedicine address neurologic health disparities in rural Guatemala: a health promotor educational intervention study," *Health Promotion International*, vol. 37, 1 feb. de 2022, ISSN: 0957-4824. DOI: 10.1093/HEAPRO/DAAB072. dirección: <https://dx.doi.org/10.1093/heapro/daab072>.
- [13] R. B. Postuma, D. Berg, M. Stern et al., "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, págs. 1591-1601, 12 oct. de 2015, ISSN: 1531-8257. DOI: 10.1002/MDS.26424. dirección: <https://onlinelibrary.wiley.com/doi/full/10.1002/mds.26424>
<https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.26424>
<https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mds.26424>.
- [14] L. Marsh y T. M. Dawson, "Treatment of early Parkinson's disease," *BMJ*, vol. 321, págs. 1-2, 7252 jul. de 2000, ISSN: 0959-8138. DOI: 10.1136/BMJ.321.7252.1. dirección: <https://www.bmj.com/content/321/7252/1>
<https://www.bmj.com/content/321/7252/1.abstract>.
- [15] A. Ma, K. K. Lau y D. Thyagarajan, "Voice changes in Parkinson's disease: What are they telling us?" *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia*, vol. 72, págs. 1-7, feb. de 2020, ISSN: 1532-2653. DOI: 10.1016/J.JOCN.2019.12.029. dirección: <https://pubmed.ncbi.nlm.nih.gov/31952969/>.
- [16] R. Sheibani, E. Nikookar y S. Alavi, "An ensemble method for diagnosis of Parkinson's disease based on voice measurements," *Journal of Medical Signals Sensors*, vol. 9, pág. 221, 4 oct. de 2019, ISSN: 2228-7477. DOI: 10.4103/JMSS.JMSS_57_18. dirección: <https://www.jmssjournal.net/article.asp?issn=2228-7477;year=2019;volume=9;issue=4;spage=221;epage=226;aulast=Sheibani>
<https://www.jmssjournal.net/article.asp?issn=2228-7477;year=2019;volume=9;issue=4;spage=221;epage=226;aulast=Sheibani;type=0>.
- [17] R. M. Sadek, S. A. Mohammed, A. Rahman et al., "Parkinson's Disease Prediction Using Artificial Neural Network," *International Journal of Academic Health and Medical Research*, vol. 3, págs. 1-8, 1 2019. dirección: www.ijeais.org/ijahmr.
- [18] N. Dehak, P. Dumouchel y P. Kenny, "Modeling Prosodic Features With Joint Factor Analysis for Speaker Verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, págs. 2095-2103, oct. de 2007. DOI: 10.1109/TASL.2007.902758.

- [19] J. Vásquez-Correa, J. Orozco-Arroyave, T. Bocklet y E. Nöth, “Towards an automatic evaluation of the dysarthria level of patients with Parkinson’s disease,” *Journal of Communication Disorders*, vol. 76, págs. 21-36, nov. de 2018, ISSN: 0021-9924. DOI: 10.1016/j.jcomdis.2018.08.002. dirección: <http://dx.doi.org/10.1016/j.jcomdis.2018.08.002>.
- [20] A. Ascherio y M. A. Schwarzschild, “The epidemiology of Parkinson’s disease: risk factors and prevention,” *The Lancet Neurology*, vol. 15, n.º 12, págs. 1257-1272, nov. de 2016, ISSN: 1474-4422. DOI: 10.1016/S1474-4422(16)30230-7. dirección: [http://dx.doi.org/10.1016/S1474-4422\(16\)30230-7](http://dx.doi.org/10.1016/S1474-4422(16)30230-7).
- [21] M. Amin, *Marley Drug*, sep. de 2023. dirección: <https://www.marleydrug.com/blog/parkinsons-disease>.
- [22] M. L. Giroux, “Parkinson disease: Managing a complex, progressive disease at all stages,” *Cleveland Clinic Journal of Medicine*, vol. 74, n.º 5, págs. 313-328, 2007, ISSN: 0891-1150. eprint: <https://www.ccjm.org/content/74/5/313.full.pdf>. dirección: <https://www.ccjm.org/content/74/5/313>.
- [23] R. Martínez-Fernández., C. Gasca-Salas C., Á. Sánchez-Ferro y J. Ángel Obeso, “ACTUALIZACIÓN EN LA ENFERMEDAD DE PARKINSON,” *Revista Médica Clínica Las Condes*, vol. 27, n.º 3, págs. 363-379, mayo de 2016, ISSN: 0716-8640. DOI: 10.1016/j.rmclc.2016.06.010. dirección: <http://dx.doi.org/10.1016/j.rmclc.2016.06.010>.
- [24] R. Balestrino y A. Schapira, “Parkinson disease,” *European Journal of Neurology*, vol. 27, n.º 1, págs. 27-42, nov. de 2019, ISSN: 1468-1331. DOI: 10.1111/ene.14108. dirección: <http://dx.doi.org/10.1111/ene.14108>.
- [25] D. K. Simon, C. M. Tanner y P. Brundin, “Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology,” *Clinics in Geriatric Medicine*, vol. 36, n.º 1, págs. 1-12, feb. de 2020, ISSN: 0749-0690. DOI: 10.1016/j.cger.2019.08.002. dirección: <http://dx.doi.org/10.1016/j.cger.2019.08.002>.
- [26] J. S. Saavedra Moreno, P. A. Millán y O. F. Buriticá Henao, “Introducción, epidemiología y diagnóstico de la enfermedad de Parkinson,” *Acta Neurológica Colombiana*, vol. 35, n.º 3 Supl 1, págs. 2-10, sep. de 2019, ISSN: 0120-8748. DOI: 10.22379/24224022244. dirección: <http://dx.doi.org/10.22379/24224022244>.
- [27] M. Ugrumov, “Development of early diagnosis of Parkinson’s disease: Illusion or reality?” *CNS Neuroscience amp; Therapeutics*, vol. 26, n.º 10, págs. 997-1009, jun. de 2020, ISSN: 1755-5949. DOI: 10.1111/cns.13429. dirección: <http://dx.doi.org/10.1111/cns.13429>.
- [28] R. Gilbert, *What is a DaTscan and should I get one? | APDA*, ene. de 2019. dirección: <https://www.apdaparkinson.org/article/what-is-a-datscan-and-should-i-get-one/>.
- [29] S. Mantri, J. F. Morley y A. D. Siderowf, “The importance of preclinical diagnostics in Parkinson disease,” *Parkinsonism amp; Related Disorders*, vol. 64, págs. 20-28, jul. de 2019, ISSN: 1353-8020. DOI: 10.1016/j.parkreldis.2018.09.011. dirección: <http://dx.doi.org/10.1016/j.parkreldis.2018.09.011>.
- [30] D. Smith y M. Czarniecki, *Magnetic resonance parkinsonism index*, mar. de 2014. DOI: 10.53347/rid-28173. dirección: <http://dx.doi.org/10.53347/rid-28173>.

- [31] L. Parnetti, L. Gaetani, P. Eusebi et al., “CSF and blood biomarkers for Parkinson’s disease,” *The Lancet Neurology*, vol. 18, n.º 6, págs. 573-586, jun. de 2019, ISSN: 1474-4422. DOI: 10.1016/s1474-4422(19)30024-9. dirección: [http://dx.doi.org/10.1016/s1474-4422\(19\)30024-9](http://dx.doi.org/10.1016/s1474-4422(19)30024-9).
- [32] T. A. Zesiewicz, “Parkinson Disease,” *CONTINUUM: Lifelong Learning in Neurology*, vol. 25, n.º 4, págs. 896-918, ago. de 2019, ISSN: 1080-2371. DOI: 10.1212/con.0000000000000764. dirección: <http://dx.doi.org/10.1212/con.0000000000000764>.
- [33] M. J. Armstrong y M. S. Okun, “Diagnosis and Treatment of Parkinson Disease: A Review,” *JAMA*, vol. 323, n.º 6, pág. 548, feb. de 2020, ISSN: 0098-7484. DOI: 10.1001/jama.2019.22360. dirección: <http://dx.doi.org/10.1001/jama.2019.22360>.
- [34] I. Castilla-Cortázar, G. A. Aguirre, G. Femat-Roldán, I. Martín-Estal y L. Espinosa, “Is insulin-like growth factor-1 involved in Parkinson’s disease development?” *Journal of Translational Medicine*, vol. 18, n.º 1, feb. de 2020, ISSN: 1479-5876. DOI: 10.1186/s12967-020-02223-0. dirección: <http://dx.doi.org/10.1186/s12967-020-02223-0>.
- [35] S. Luis Parrón, “Tratamiento de voz Lee Silverman en la disartria hipocinética: revisión de la efectividad del tratamiento en los enfermos de Parkinson,” *Revista de Logopedia, Foniatría y Audiología*, vol. 37, n.º 3, págs. 130-144, 2017, ISSN: 0214-4603. DOI: <https://doi.org/10.1016/j.rlf.2017.01.003>. dirección: <https://www.sciencedirect.com/science/article/pii/S0214460317300207>.
- [36] V. P. Vandana, J. K. Darshini, V. H. Vikram, K. Nitish, P. K. Pal e Y. Ravi, “Speech Characteristics of Patients with Parkinson’s Disease—Does Dopaminergic Medications Have a Role?” *Journal of Neurosciences in Rural Practice*, vol. 12, págs. 673-679, sep. de 2021, ISSN: 0976-3147. DOI: 10.1055/s-0041-1735249. dirección: <http://dx.doi.org/10.1055/s-0041-1735249>.
- [37] E. Nikookar, R. Sheibani y S. Alavi, “An ensemble method for diagnosis of Parkinson’s disease based on voice measurements,” *Journal of Medical Signals amp; Sensors*, vol. 9, n.º 4, pág. 221, 2019, ISSN: 2228-7477. DOI: 10.4103/jmss.jmss_57_18. dirección: http://dx.doi.org/10.4103/jmss.JMSS_57_18.
- [38] T. Talkar, J. Williamson, D. Hannon et al., “Assessment of Speech and Fine Motor Coordination in Children With Autism Spectrum Disorder,” *IEEE Access*, vol. PP, págs. 1-1, jul. de 2020. DOI: 10.1109/ACCESS.2020.3007348.
- [39] *Acoustic Phonetics*, [Accessed 23-01-2024]. dirección: <https://www.sheffield.ac.uk/linguistics/home/all-about-linguistics/about-website/branches-linguistics/phonetics/what-do-phoneticians-study/acoustic#:~:text=Acoustic%20phonetics%20is%20the%20study,through%20looking%20at%20a%20waveform>.
- [40] S. Fernández-García, C. G. Dumitrache y J. A. González-López, “Acoustic analysis of the voice in patients with Parkinson’s disease and hypokinetic dysarthria,” *Revista de Logopedia, Foniatría y Audiología*, vol. 41, n.º 3, págs. 142-150, jul. de 2021, ISSN: 0214-4603. DOI: 10.1016/j.rlf.2020.04.002. dirección: <http://dx.doi.org/10.1016/j.rlf.2020.04.002>.
- [41] M. Rodríguez Súnico, *ANÁLISIS MODAL OPERACIONAL: TEORÍA Y PRÁCTICA*, <https://biblus.us.es/bibing/proyectos/abreproy/3828>, [Accessed 09-05-2024].

- [42] R. Kronland-Martinet, P. Guillemin y S. Ystad, “From Sound Modeling to Analysis-Synthesis of Sounds,” 2001. dirección: <https://api.semanticscholar.org/CorpusID:195809834>.
- [43] D. Moore, H. Tennent, N. Martelaro y W. Ju, “Making Noise Intentional: A Study of Servo Sound Perception,” mar. de 2017, págs. 12-21. DOI: 10.1145/2909824.3020238.
- [44] G. Martínez Mascorro Guillermo Arturo Aguilar Torres, “Reconocimiento de voz basado en MFCC, SBC y Espectrogramas,” Español, *Ingenius. Revista de Ciencia y Tecnología*, 2013, ISSN: 1390-650X. dirección: <https://www.redalyc.org/articulo.oa?id=505554816003>.
- [45] J. Llisterri, *Características acústicas de los elementos segmentales del español - Vocales*, [Accessed 05-02-2024]. dirección: https://joaquimllisterri.cat/phonetics/fon_anal_acus/castella_analisi/cast_vocales_acust.html.
- [46] H. J. R. Navia, N. A. G. Quintero y P. L. N. Criollo, “DIADOCOCINESIA DEL HABL A EN ESTUDIANTES UNIVERSITARIOS,” *REVISTA CIENTÍFICA SIGNOS FÓNICOS*, vol. 2, n.º 1, abr. de 2016, ISSN: 2422-1716. DOI: 10.24054/01204211.v1.n1.2016.1786. dirección: <http://dx.doi.org/10.24054/01204211.v1.n1.2016.1786>.
- [47] J. P. Teixeira, C. Oliveira y C. Lopes, “Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters,” *Procedia Technology*, vol. 9, págs. 1112-1122, 2013, ISSN: 2212-0173. DOI: 10.1016/j.protcy.2013.12.124. dirección: <http://dx.doi.org/10.1016/J.PROTCY.2013.12.124>.
- [48] *The Cambridge Handbook of Phonetics*. Cambridge University Press, nov. de 2021, ISBN: 9781108448598. DOI: 10.1017/9781108644198. dirección: <http://dx.doi.org/10.1017/9781108644198>.
- [49] *Springer Handbook of Acoustics*. Springer New York, 2014, ISBN: 9781493907557. DOI: 10.1007/978-1-4939-0755-7. dirección: <http://dx.doi.org/10.1007/978-1-4939-0755-7>.
- [50] D. J. Hirst y C. de Looze, “Measuring Speech. Fundamental frequency and pitch,” en *Cambridge Handbook of Phonetics*, 1, R.-A. Knight y J. Setter, eds., Cambridge University Press, nov. de 2021, págs. 336-361. DOI: 10.1017/9781108644198. dirección: <https://hal.science/hal-03596403>.
- [51] *5.1.2.1 Two Steps: Sampling and Quantization*, <https://digitalsoundandmusic.com/5-1-2-digitization/>, [Accessed 24-01-2024].
- [52] N. Tits, K. El Haddad y T. Dutoit, “The Theory behind Controllable Expressive Speech Synthesis: A Cross-Disciplinary Approach,” en *Human 4.0 - From Biology to Cybernetic*. IntechOpen, ene. de 2021. DOI: 10.5772/intechopen.89849. dirección: <http://dx.doi.org/10.5772/INTECHOPEN.89849>.
- [53] Audacity, *Loudness Normalization | Audacity Support*, mar. de 2023. dirección: <https://support.audacityteam.org/audio-editing/loudness-normalization>.
- [54] Audacity, *DC offset - Audacity Manual*, 2024. dirección: https://manual.audacityteam.org/man/dc_offset.html#:~:text=A%20sound%20that%20has%20DC,maximum%20possible%20level%20without%20clipping..

- [55] C. Janiesch, P. Zschech y K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, n.º 3, págs. 685-695, sep. de 2021. DOI: 10.1007/s12525-021-00475-. dirección: https://ideas.repec.org/a/spr/elmark/v31y2021i3d10.1007_s12525-021-00475-2.html.
- [56] T. Jo, *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer International Publishing, 2021, ISBN: 9783030659004. DOI: 10.1007/978-3-030-65900-4. dirección: <http://dx.doi.org/10.1007/978-3-030-65900-4>.
- [57] J. Fumo, *Types of Machine Learning Algorithms You Should Know*. dirección: <https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>.
- [58] A. Ul Haq, J. Li, M. H. Memon et al., “Comparative Analysis of the Classification Performance of Machine Learning Classifiers and Deep Neural Network Classifier for Prediction of Parkinson Disease,” en *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, IEEE, dic. de 2018. DOI: 10.1109/iccwamtip.2018.8632613. dirección: <http://dx.doi.org/10.1109/ICCWAMTIP.2018.8632613>.
- [59] A. Mammone, M. Turchi y N. Cristianini, “Support vector machines,” *WIREs Computational Statistics*, vol. 1, n.º 3, págs. 283-289, nov. de 2009, ISSN: 1939-0068. DOI: 10.1002/wics.49. dirección: <http://dx.doi.org/10.1002/wics.49>.
- [60] O. Kramer, “K-Nearest Neighbors,” en *Intelligent Systems Reference Library*. Springer Berlin Heidelberg, 2013, págs. 13-23, ISBN: 9783642386527. DOI: 10.1007/978-3-642-38652-7_2. dirección: http://dx.doi.org/10.1007/978-3-642-38652-7_2.
- [61] P. Castagno, *k-nearest neighbors (KNN)*, [Accessed 05-02-2024]. dirección: <https://patriziacastagnod.medium.com/k-nearest-neighbors-knn-9491f6d684ae>.
- [62] S. Gupta, *Understanding and using k-Nearest Neighbours aka kNN for classification of digits*. dirección: <https://towardsdatascience.com/understanding-and-using-k-nearest-neighbours-aka-knn-for-classification-of-digits-a55e00cc746f>.
- [63] I. H. Sarker, “Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions,” *SN Computer Science*, vol. 2, n.º 6, ago. de 2021, ISSN: 2661-8907. DOI: 10.1007/s42979-021-00815-1. dirección: <http://dx.doi.org/10.1007/s42979-021-00815-1>.
- [64] K. J. Danjuma, “Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients,” 2015. DOI: 10.48550/ARXIV.1504.04646. dirección: <https://arxiv.org/abs/1504.04646>.
- [65] J. Martínez Pérez y P. Pérez Martín, “La curva ROC,” *Medicina de Familia. SEMERGEN*, vol. 49, n.º 1, pág. 101 821, ene. de 2023, ISSN: 1138-3593. DOI: 10.1016/j.semerg.2022.101821. dirección: <http://dx.doi.org/10.1016/j.semerg.2022.101821>.
- [66] M. M. Rodríguez-Hernández, R. E. Pruneda y J. M. Rodríguez-Díaz, “Statistical Analysis of the Evolutive Effects of Language Development in the Resolution of Mathematical Problems in Primary School Education,” *Mathematics*, vol. 9, n.º 10, pág. 1081, mayo de 2021, ISSN: 2227-7390. DOI: 10.3390/math9101081. dirección: <http://dx.doi.org/10.3390/math9101081>.

- [67] J. R. Orozco, J. D. Arias-Londoño, J. Vargas-Bonilla, M. González-Rátiva y E. Noeth, “New Spanish speech corpus database for the analysis of people suffering from Parkinsons disease,” *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)*, págs. 342-347, ene. de 2014.
- [68] T. Giannakopoulos, “pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis,” *PloS one*, vol. 10, n.º 12, 2015.
- [69] M.-J. Marsano-Cornejo y Á. Roco-Videla, “Comparison of the Acoustic Parameters Obtained With Different Smartphones and a Professional Microphone,” *Acta Otorrinolaringologica (English Edition)*, vol. 73, n.º 1, págs. 51-55, 2022, ISSN: 2173-5735. DOI: <https://doi.org/10.1016/j.otoeng.2020.08.009>. dirección: <https://www.sciencedirect.com/science/article/pii/S2173573522000199>.
- [70] J. Green, *Cell Phone vs. Microphone: Judging Emotion in the Voice*, 2016. dirección: https://digitalcommons.lib.uconn.edu/gs_theses/868/.
- [71] A. P. Vogel, K. M. Rosen, A. T. Morgan y S. Reilly, “Comparability of Modern Recording Devices for Speech Analysis: Smartphone, Landline, Laptop, and Hard Disc Recorder,” *Folia Phoniatrica et Logopaedica*, vol. 66, n.º 6, págs. 244-250, 2014, ISSN: 1421-9972. DOI: 10.1159/000368227. dirección: <http://dx.doi.org/10.1159/000368227>.
- [72] A. Ganeshan, *La lingüística hispánica: Una introducción (Capítulo 1 y 2 disponibles, otros capítulos en desarrollo)*, 2019. dirección: <https://ohioopen.library.ohio.edu/opentextbooks/3/>.
- [73] A. Belouali, S. Gupta, V. Sourirajan et al., “Acoustic and language analysis of speech for suicidal ideation among US veterans,” *BioData Mining*, vol. 14, n.º 1, feb. de 2021, ISSN: 1756-0381. DOI: 10.1186/s13040-021-00245-y. dirección: <http://dx.doi.org/10.1186/s13040-021-00245-y>.
- [74] J. Boyd, M. Fahim y O. Olukoya, “Voice spoofing detection for multiclass attack classification using deep learning,” *Machine Learning with Applications*, vol. 14, pág. 100 503, 2023, ISSN: 2666-8270. DOI: <https://doi.org/10.1016/j.mlwa.2023.100503>. dirección: <https://www.sciencedirect.com/science/article/pii/S2666827023000567>.
- [75] G. Celik, “CovidCoughNet: A new method based on convolutional neural networks and deep feature extraction using pitch-shifting data augmentation for covid-19 detection from cough, breath, and voice signals,” *Computers in Biology and Medicine*, vol. 163, pág. 107 153, sep. de 2023, ISSN: 0010-4825. DOI: 10.1016/j.combiomed.2023.107153. dirección: <http://dx.doi.org/10.1016/j.combiomed.2023.107153>.

Listing 13.1: Ejemplo de extracción de parámetros

```
from disvoice.prosody import Prosody
prosody = Prosody()

file_audio = "... \ ka-ka-ka \ hc \ AVPEPUDEAC0001_ka.wav"
features=prosody.extract_features_file(file_audio, \
static=True, plots=True, fmt="dataframe")
```

Listing 13.2: Ejemplo de extracción de parámetros

```
from pyAudioAnalysis import ShortTermFeatures

sonido = "... \ ka-ka-ka \ hc \ AVPEPUDEAC0001_ka.wav"

Fs, x = read(sonido)
window = 0.02*Fs
step = 0.01*Fs

chr_features, names= ShortTermFeatures.feature_extraction(x, \
Fs, window, step)
```

Listing 13.3: Hiperparametros utilizados en modelo SVM

```
param_grid_sv = { 'kernel': [ 'linear', 'rbf' ],
                  'C': [0.1, 1, 10, 100, 1000],
                  'gamma': [0.0001, 0.001, 0.1, 1, 10, 100] }
```

Listing 13.4: Hiperparametros utilizados en modelo SVM

```

param_grid_lg = { 'penalty': [ 'l1', 'l2' ],
                  'C': [0.01, 0.1, 1, 10, 100],
                  'solver': [ 'liblinear' ] }

```

Micrófono utilizado	Rango dinámico	SNR	THD	Energía
Profesional	1.6851	-6.20e-08	7.071	0.000390
Computadora	1.8579	1.23e-07	7.071	0.003710
Dispositivo móvil	1.7402	1.20e-07	7.071	0.002501

Cuadro 2: Resultados obtenidos

Tarea	Características	AUC	Precisión	Sensibilidad	Especificidad
KA	Todos los parámetros	0,78	70 %	69 %	70 %
	F0	0,56	54 %	48 %	61 %
	Energía	0,67	59 %	65 %	54 %
	Duración	0,73	65 %	61 %	68 %
PA	Todos los parámetros	0,62	58 %	48 %	69 %
	F0	0,46	47 %	52 %	46 %
	Energía	0,69	65 %	61 %	70 %
	Duración	0,71	64 %	56 %	70 %
TA	Todos los parámetros	0,61	59 %	41 %	81 %
	F0	0,61	58 %	66 %	53 %
	Energía	0,50	46 %	60 %	40 %
	Duración	0,62	53 %	45 %	65 %
PAKATA	Todos los parámetros	0,63	63 %	51 %	75 %
	F0	0,44	46 %	48 %	53 %
	Energía	0,79	75 %	68 %	81 %
	Duración	0,46	49 %	48 %	52 %
PATAKA	Todos los parámetros	0,58	50 %	54 %	52 %
	F0	0,57	53 %	51 %	62 %
	Energía	0,62	57 %	54 %	64 %
	Duración	0,63	59 %	76 %	47 %
PETAKA	Todos los parámetros	0,49	46 %	57 %	39 %
	F0	0,52	53 %	53 %	53 %
	Energía	0,61	58 %	68 %	56 %
	Duración	0,47	46 %	54 %	46 %

Cuadro 3: Resultados de la primera experimentación con Algoritmo Máquina de Vectores de Soporte (SVM)

Tarea	Características	AUC	Precisión	Sensibilidad	Especificidad
KA	Todos los parámetros	0,70	66 %	66 %	68 %
	F0	0,52	51 %	49 %	52 %
	Energía	0,68	67 %	64 %	70 %
	Duración	0,74	67 %	58 %	77 %
PA	Todos los parámetros	0,55	57 %	54 %	62 %
	F0	0,43	46 %	52 %	44 %
	Energía	0,60	61 %	51 %	71 %
	Duración	0,72	65 %	64 %	67 %
TA	Todos los parámetros	0,64	62 %	55 %	71 %
	F0	0,66	61 %	65 %	59 %
	Energía	0,53	49 %	24 %	79 %
	Duración	0,61	60 %	46 %	75 %
PAKATA	Todos los parámetros	0,60	60 %	53 %	65 %
	F0	0,39	38 %	33 %	50 %
	Energía	0,75	72 %	74 %	69 %
	Duración	0,48	46 %	40 %	53 %
PATAKA	Todos los parámetros	0,69	66 %	71 %	64 %
	F0	0,61	59 %	60 %	57 %
	Energía	0,57	56 %	61 %	54 %
	Duración	0,69	69 %	60 %	78 %
PETAKA	Todos los parámetros	0,51	50 %	54 %	47 %
	F0	0,49	45 %	22 %	73 %
	Energía	0,65	64 %	62 %	69 %
	Duración	0,66	64 %	54 %	75 %

Cuadro 4: Resultados experimentación con Algoritmo de Regresión Logística (RL)