

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el CONACYT (Consejo Nacional de Ciencia y Tecnología).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos Autor.

CICESE@ 2022. Todos los derechos reservados

**Centro de Investigación Científica y de Educación  
Superior de Ensenada, Baja California**



---

**Maestría en Ciencias  
en Ciencias de la computación**

---

**Recolección y análisis de voz en adultos mayores para el  
diagnóstico de demencia**

Tesis  
para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Lisset Cabrera Leyva**

Ensenada, Baja California, México  
2022

Tesis defendida por  
**Lisset Cabrera Leyva**

y aprobada por el siguiente Comité

---

Dr. Jesús Favela Vara  
Codirector de tesis

---

Dr. Dagoberto Cruz Sandoval  
Codirector de tesis

**Dra. Ana Isabel Martínez García**

**Dr. Humberto Pérez Espinosa**



---

Dr. Pedro Gilberto López Mariscal  
Coordinador del Posgrado en Ciencias de la Computación

---

**Dr. Pedro Negrete Regagnon**  
Director de Estudios de Posgrado

Resumen de la tesis que presenta **Lisset Cabrera Leyva** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la computación.

### **Recolección y análisis de voz en adultos mayores para el diagnóstico de demencia**

Resumen aprobado por:

---

Dr. Jesús Favela Vara  
Codirector de tesis

---

Dr. Dagoberto Cruz Sandoval  
Codirector de tesis

La demencia es una de las principales causas de dependencia de los adultos mayores. Si bien no tiene cura conocida, su diagnóstico temprano puede ayudar a adultos mayores y familiares a tomar medidas que permitan atenuar sus efectos. En los últimos años se han publicado estudios orientados a diagnosticar esta enfermedad y su trayectoria mediante el uso de biomarcadores digitales. Este trabajo de investigación se centra en el estudio de la voz como biomarcador digital para apoyar el diagnóstico de demencia. Se elaboró un conjunto de datos con registros de voz de 12 adultos mayores con demencia y 20 sanos que asistieron a una clínica geriátrica a consulta. Los participantes desarrollaron tres pruebas clínicas para grabar su voz: “Repetir frases”, “Fluidez verbal semántica” y descripción de la imagen “Robo de galletas”. Las grabaciones se realizaron con el apoyo de una herramienta desarrollada como parte de esta tesis. Un total de 7500 rasgos acústicos aproximadamente se extrajeron de las señales de habla utilizando las herramientas Praat y openSmile. Primero se le aplicaron técnicas de ingeniería de características para reducir el número de rasgos a los más significativos. Posteriormente se crearon modelos utilizando los siguientes algoritmos de clasificación: Análisis discriminante lineal, Árboles de decisión, k-vecinos más cercanos, Máquina de soporte vectorial y Random forest, capaces de discriminar automáticamente, en función de las características acústicas, a los adultos mayores sanos y con demencia. Los modelos se interpretaron a través de técnicas de Inteligencia artificial explicable. Los resultados obtenidos son similares o mejores que los reportados en la literatura. Las técnicas de Inteligencia artificial explicable aportan evidencia de que los modelos están generando decisiones en función de rasgos acústicos asociados a alteraciones de voz de adultos mayores con demencia. La calidad de los datos de voz grabados y el uso de técnicas de reducción de características diferencian este trabajo con respecto a los reportados en la literatura y son las principales causales del buen desempeño obtenido, aunque el número de sujetos analizados es relativamente pequeño. Nuestros resultados aportan evidencia de que las características acústicas pueden ser utilizadas para predecir el diagnóstico de demencia.

**Palabras clave:** biomarcadores digitales, demencia, análisis acústico, aprendizaje automático.

Abstract of the thesis presented by **Lisset Cabrera Leyva** as a partial requirement to obtain the Master of science degree in Computer's science

**Collection and analysis of voice in older adults for the dementia diagnosis**

Abstract approved by:

---

Dr. Jesús Favela Vara  
Codirector de tesis

---

Dr. Dagoberto Cruz Sandoval  
Codirector de tesis

Dementia is one of the principal causes of dependency in older adults. Even though there is no known cure, an early diagnosis can help older adults and their families to take actions to attenuate its effects. In the last few years several studies have been published aimed at developing digital biomarkers for the diagnosis of dementia diagnosis and its trajectory. This work focuses on the use of speech analysis as a digital biomarker to support dementia diagnosis. We created a dataset with voice recordings of 12 older adults diagnosed with dementia and 20 healthy older adults, who assisted to a geriatric clinic for consultation. These participants performed three clinical tests to record their voice: "Repeat phrases", "Semantic verbal fluency", and description of the image "Theft of Cookies image ". The recordings were carried out with a software developed as part of this thesis. A total of approximately 7500 acoustics features were extracted from speech signals using the Praat and openSmile tools. Feature engineering techniques based on ranking fusion were applied to select the most relevant features. Then, classification models were built using the Linear Discriminant Analysis, Decision Tree, k-Nearest Neighbors, Support Vector Machine, and Random Forest techniques, to automatically discriminate, based on the acoustic features, healthy older adults, and those with dementia. The models were interpreted through explainable artificial intelligence techniques. The results obtained are similar or better than those reported in the literature. The explainable artificial intelligence techniques provide evidence that the models are generating decisions based on acoustic features associated with voice alterations in older adults with dementia. The quality of the recorded voice data and the use of feature reduction techniques differentiate this work from those reported in the literature and are the main causes of the good performance obtained, although the number of subjects analyzed is relatively small. Our results provide evidence that acoustic features can be used to predict the dementia diagnosis.

**Keywords: digital biomarkers, dementia, acoustic analysis, machine learning**

## **Dedicatoria**

A mi mamá, mi papá y mi hermana por apoyarme siempre.

A mi pequeño gigante Aaron César por ser el motor impulsor de cada mañana

A mi esposo César Raúl García por ayudarme a cumplir mis sueños sin importar los obstáculos

## Agradecimientos

Al CICESE y Conacyt por el apoyo económico para realizar mis estudios.

A las personas más especiales de mi vida, a mis padres, por todos los consejos, sacrificios, enseñanzas, dedicación y amor que siempre me han dado, por entender mis limitaciones de tiempo a pesar de los tres años que llevamos separados.

A mi esposo por su apoyo incondicional, por siempre animarme y hacer que mis días más grises fueran coloridos.

A mis asesores, Dr. Jesús Favela Vara y Dr. Dagoberto Cruz Sandoval, por su oportuna guía, por conducirme con sabiduría y apoyarme en todo momento. Disfruté mucho trabajar este año junto a ustedes.

A la doctora geriatra Diana Paniagua, por tenernos tanta paciencia, por permitirnos entrar a su consulta y transmitirnos muchos tips que nos facilitaron la recolección de datos con los adultos mayores. A la psicóloga Maricruz Huerta, como siempre te digo, eres una parte importante en este proyecto, sin tu dedicación, ganas de superarte, empeño y disciplina no habiéramos obtenido los archivos de audios con la calidad requerida.

A las profesoras miembros del comité de tesis, Dra. Ana Isabel Martínez y Dr. Humberto Pérez, por sus correcciones y comentarios apropiados para mejorar este trabajo.

A los profesores que aportaron un granito de arena en mi formación, a mis amigos Nadia y Karel por transmitirme sus conocimientos de modelación de proceso y a mis compañeros de CICESE que contribuyeron a estos resultados y siempre estuvieron presente cuando los necesité.

*A todos les estoy eternamente agradecida*

## Tabla de contenido

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatoria .....	iv
Agradecimientos.....	v
Lista de figuras.....	viii
Lista de tablas .....	ix
<b>Capítulo 1. Introducción.....</b>	<b>1</b>
1.1 Planteamiento del problema .....	2
1.2 Objetivos .....	4
1.2.1 Objetivo general.....	4
1.2.2 Objetivos específicos.....	4
1.3 Metodología .....	4
1.4 Organización de la tesis.....	6
<b>Capítulo 2. Biomarcadores digitales de demencia basados en audio .....</b>	<b>8</b>
2.1 Biomarcadores digitales.....	8
2.2 Demencia en el adulto mayor .....	9
2.2.1 Evaluación diagnóstica .....	10
2.3 Detección de demencia analizando la voz .....	12
2.3.1 Metodología para la revisión de la literatura.....	12
2.3.2 Corpus de audios.....	13
2.3.3 Aprendizaje automático para el análisis de la voz .....	16
<b>Capítulo 3. Protocolo, captura y etiquetado del conjunto de datos.....</b>	<b>21</b>
3.1 Protocolo utilizado para la captura de datos .....	21
3.2 Herramienta para la captura de datos .....	24
3.3 Estudio piloto y resultados.....	28
3.4 Descripción del corpus de audio propuesto .....	29
3.4.1 Preprocesamiento de datos .....	31
3.4.2 Etiquetado de registros de audio .....	32
3.4.3 Limitaciones en la calidad del corpus propuesto .....	33

<b>Capítulo 4. Metodología para el análisis de datos .....</b>	<b>35</b>
4.1 Descripción general.....	35
4.2 Rasgos acústicos.....	36
4.3 Construcción del conjunto de datos .....	37
4.3.1 Balanceo de clases.....	39
4.4 Selección de rasgos relevantes .....	39
4.5 Desarrollo de modelos predictivos .....	40
4.6 Interpretación de los resultados .....	41
<b>Capítulo 5. Resultados y discusión .....</b>	<b>43</b>
5.1 Análisis descriptivo del conjunto de rasgos .....	43
5.1.1 Análisis de curtosis y correlación para eliminar rasgos constantes y redundantes.....	43
5.1.2 Análisis visual utilizando Vecinos estocásticos distribuidos (tSNE).....	44
5.1.3 Importancia de características analizando diferentes estadísticos .....	45
5.2 Modelos generados.....	49
5.2.1 Desempeño de los algoritmos de clasificación para datos no balanceados .....	49
5.2.2 Desempeño de los algoritmos de clasificación para datos balanceados .....	51
5.3 Variables representativas identificadas con técnicas de XAI .....	53
5.4 Discusión .....	53
5.4.1 Desempeño de los algoritmos de clasificación .....	53
5.4.2 Interpretación de los modelos a partir de XAI .....	57
<b>Capítulo 6. Conclusiones .....</b>	<b>59</b>
6.1 Aportaciones .....	60
6.2 Limitaciones .....	61
6.3 Trabajo futuro .....	62
<b>Literatura citada .....</b>	<b>63</b>
<b>Anexos .....</b>	<b>70</b>
Anexo A: Carta de consentimiento informado .....	70
Anexo B: Manual de etiquetado .....	73
Anexo C: Reglas generadas por el algoritmo Anchors en las 3 actividades.....	74

## Lista de figuras

Figura	Página
1. Metodología seguida en el estudio.....	5
2. Algoritmos de clasificación utilizados para el diagnóstico de demencia. ....	18
3. Imagen llamada Robo de galletas, utilizada para desarrollar la actividad de “Habla espontánea” (Cummins, 2019). ....	23
4. Modelo conceptual del procedimiento manual de captura de voz. ....	25
5. Modelo conceptual del procedimiento autónomo de captura de voz. ....	25
6. Aplicación desarrollada para la captura de los datos. ....	27
7. Preprocesamiento de datos. (a) Señal inicial y (b) Señal filtrada.....	31
8. Ejemplo del proceso de etiquetado. ....	32
9. Metodología utilizada empleada para el análisis de los datos. ....	35
10. Visualización de rasgos utilizando la técnica t-SNI. (a) actividad “Repetir frases”, (b) actividad “Fluidez verbal semántica” y (c) actividad “Robo de galletas”. ....	44
11. Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Repetir frases”.....	45
12. Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Fluidez verbal semántica”. ....	46
13. Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Robo de galletas”. ....	46
14. Distribución de los rasgos identificados a través de fusión por ranking en la actividad “Repetir frases”. ....	47
15. Distribución de los rasgos identificados a través de fusión por ranking en la actividad “Fluidez verbal.....	48
16. Distribución de los rasgos identificados a través de fusión por ranking en la actividad descripción de la imagen “Robo de galletas”.....	48
17. Variabilidad de los resultados por estrategia de selección de características.....	56
18. Desempeño de los clasificadores por cada actividad. ....	57

## Lista de tablas

Tabla	Página
1. Síntomas de las diferentes etapas de la demencia.....	11
2. Características básicas de los participantes en el estudio Pitt Corpus. ....	14
3. Características básicas de los participantes pertenecientes al conjunto de entrenamiento. .	15
4. Características básicas de los participantes pertenecientes al conjunto de prueba. ....	15
5. Características básicas de los participantes en el Ivanova corpus.....	16
6. Información extraída de los estudios analizados.....	19
7. Descripción de los procesos.....	26
8. Descripción demográfica de los sujetos que participaron en el estudio.....	30
9. Características de los ficheros de audios del corpus propuesto.....	30
10. Limitaciones identificadas en los ficheros de audios.....	33
11. Rasgos acústicos base extraídos en las tres tareas y rasgos acústicos extraídos por tipo de actividad.....	37
12. Características de los conjuntos de datos.....	37
13. Conjuntos resultantes del proceso de ensemble.....	40
14. Parámetros para optimizar en cada algoritmo de clasificación.....	41
15. Análisis de curtosis ( $K > 8$ ).....	43
16. Análisis de correlación ( $\text{corr} > 0.95$ ).....	43
17. Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Repetir frases” para LOSO.....	49
18. Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Fluidez verbal semántica” para LOSO.....	50
19. Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Robo de galletas” para LOSO.....	50
20. Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Repetir frases” para LOSO.....	51
21. Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Fluidez verbal semántica” para LOSO.....	52

22. Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Robo de galletas” para LOSO..... 52
23. Variables representativas por cada actividad. El número ( ) es la cantidad de ocurrencias. ... 54
24. Desempeño de los algoritmos de clasificación del conjunto de datos ADReSS reportado en la literatura y con un flujo de la metodología propuesta en este estudio. .... 55
25. Información acústica que codifican los rasgos relevantes..... 58

## Capítulo 1. Introducción

---

Los biomarcadores digitales se definen como datos fisiológicos y de comportamientos objetivos que se recopilan y miden por medio de dispositivos digitales, como sensores ambientales integrados, portátiles o implantables, para explicar o predecir resultados relacionados con la salud (Kourtis et al., 2019). El cómputo ubicuo hace posible inferir y monitorear actividades y comportamientos humanos de manera continua y a gran escala, lo que representa una de las principales áreas de estudio del cómputo ubicuo aplicado a salud. Esto abre nuevas oportunidades para el uso de biomarcadores digitales, particularmente para la detección y seguimiento de condiciones mentales que suelen inferirse a partir de comportamientos observados. Una de las áreas en las que se ha empezado a explorar el uso de biomarcadores digitales es la demencia (Kourtis et al., 2019)(Cavedoni et al., 2020).

La demencia es una de las principales causas de discapacidad y dependencia entre las personas mayores en el mundo entero. La Organización Mundial de Salud (OMS) reveló que 50 millones de personas padecen demencia y se prevé que esta cifra se triplique en el 2050 (Demencia, 2020). El Alzheimer es el tipo de demencia más frecuente que afecta a las personas mayores, este síndrome es originado por una serie de enfermedades que de algún modo afectan a la función cerebral. En México, el Instituto Nacional de Geriátrica (INGER) estima que aproximadamente el 7.5% de los adultos mayores presentan esta condición (“Boletín INGER,” 2020).

Una posible detección temprana de la demencia puede garantizar una mayor calidad de vida al paciente y a sus cuidadores, y puede reducir el costo de la atención médica. En los últimos años, numerosos estudios han sugerido que la disfunción del lenguaje es uno de los primeros signos de deterioro cognitivo (Petti et al., 2020). Los adultos mayores con demencia producen un habla más lento e interrumpido temporalmente, involucran mucho más tiempo en la fonación y en la producción del habla en general, sus tasas de articulación y elocución son más lentas. Además, incluye más pausas y una duración más irregular de los intervalos silábicos (Ivanova et al., 2022). De aquí el interés por estudiar el análisis de la voz en adultos mayores para proponer posibles biomarcadores digitales de trastornos neurocognitivos. Los registros de voz pueden estudiarse a lo largo del tiempo en las diferentes etapas que se manifiesta la enfermedad. Además, la captura de voz se puede realizar de forma no invasiva con dispositivos de bajo costo.

Hasta hace poco, los estudios relacionados con la voz se analizaban manualmente. Sin embargo, el desarrollo de la tecnología ha permitido automatizar este análisis impulsado por los avances en la rama de inteligencia artificial (IA), específicamente el aprendizaje automático. Con esta automatización se promueve la inclusión de más datos y de un análisis más detallado de los mismos, para revelar patrones que pueden pasar desapercibidos en el análisis manual. A la fecha, se han reportado resultados prometedores con el uso de técnicas de aprendizaje automático en el procesamiento automático de la señal de voz (Petti et al., 2020), permitiendo cuantificar las propiedades de la señal, que son relevantes para la descripción de patologías específicas. Además, ofrece la posibilidad de analizar estadística, matemática y automáticamente las propiedades lingüísticas.

## 1.1 Planteamiento del problema

Si bien la detección de la demencia por medio del análisis de voz ha sido estudiada en años recientes, estos estudios presentan limitaciones importantes que siguen dejando el tema como un problema abierto. En los estudios reportados del área se identificaron elementos interesantes a abordar, tales como: población utilizada, idioma de los corpus de audios empleados y actividades desarrolladas para la recolección de datos.

La población es descrita en función del número de participantes, uso de grupo de control saludable, datos demográficos como edad, sexo y escolaridad. No obstante, se evidencia un desbalance con respecto a los datos demográficos. Más de la mitad de los estudios mostraron diferencias significativas en las edades de los participantes entre los grupos sanos y los que presentaban deterioro cognitivo, se incluyeron más mujeres que hombres en los estudios y en más de la mitad de los datos reportados en relación con la escolaridad, las personas con una condición clínica tendían a tener menos educación que el grupo de control.

La mayoría de los estudios reportados en la literatura utilizan conjuntos de datos en idioma inglés (Fraser et al., 2018)(Calz et al., 2020)(Guinn et al., 2014)(Haider et al., 2019)(Luz et al., 2020), con algunos adicionales en chino mandarín (Chien et al., 2019), sueco (Fraser et al., 2018), francés (Tr et al., 2017) y húngaro (Tóth et al., 2018). Al mejor de nuestro conocimiento, solo se tiene acceso a dos conjuntos de este tipo en español, los corpus AZTIAHO (Ammar and Ayed, 2018) y el reportado en Ivanova (Ivanova et al., 2022) publicados recientemente. AZTIAHO está compuesto por ficheros de videos, quedando su análisis fuera del objetivo de nuestro estudio. Con respecto al corpus de Ivanova, a pesar de que está

formado por ficheros de audios es español, sus datos fueron recolectados en España y los dialectos del español de España con respecto a Hispanoamérica, son diferentes tanto léxica como fonéticamente. La disponibilidad de corpus de audios en idioma español constituye una limitante para realizar estudios de este tipo en México.

Diferentes algoritmos de aprendizaje automático han mostrado resultados prometedores al distinguir los sujetos con deterioro neurocognitivo y sano. Uno de los factores que contribuye a alcanzar un buen acierto en la predicción del diagnóstico de demencia, utilizando estos modelos, es una correcta identificación de características relevantes que proporcionen información útil. En la literatura se evidencia una tendencia al análisis de características extraídas de actividades que promueven el habla (Calz et al., 2020)(Lizarduy et al., 2017)(Warnita et al., 2018). Aunque pudiera estar sesgado por la disponibilidad de conjuntos de datos, con ese tipo de actividad no se pueden analizar características derivadas de la evaluación del lenguaje, ni del déficit cognitivo que presentan los adultos mayores con algún deterioro.

Aunque se ha demostrado un adecuado desempeño de estos algoritmos, se continúa criticando la falta de interpretabilidad en los resultados. Conocer las variables que están tomando en cuenta los modelos para decidir si el adulto mayor presenta demencia, mejora la capacidad de toma de decisiones al mismo tiempo que fortalece la confianza y la transparencia en este tipo de estudio. En la literatura consultada hasta el momento no hay evidencia del uso de técnicas que ayuden a comprender e interpretar las predicciones del diagnóstico.

Esta tesis aborda estas problemáticas a partir de la confección de un corpus de audios en idioma español de México, para entrenar modelos de aprendizaje automático e interpretarlos a través de técnicas de inteligencia artificial explicable, para apoyar el diagnóstico de demencia.

Por lo anterior mencionado, el punto de partida de nuestro estudio es la siguiente **pregunta de investigación**:

*¿De qué manera se puede analizar la voz en adultos mayores, a partir del uso de técnicas de aprendizaje automático, para diferenciar entre adultos mayores sanos y con demencia?*

## 1.2 Objetivos

### 1.2.1 Objetivo general

Desarrollar y evaluar modelos computacionales con el fin de predecir el diagnóstico de demencia en adultos mayores, a partir de análisis de voz, mediante técnicas de aprendizaje automático.

### 1.2.2 Objetivos específicos

Para dar cumplimiento al objetivo general se abordarán los siguientes objetivos específicos:

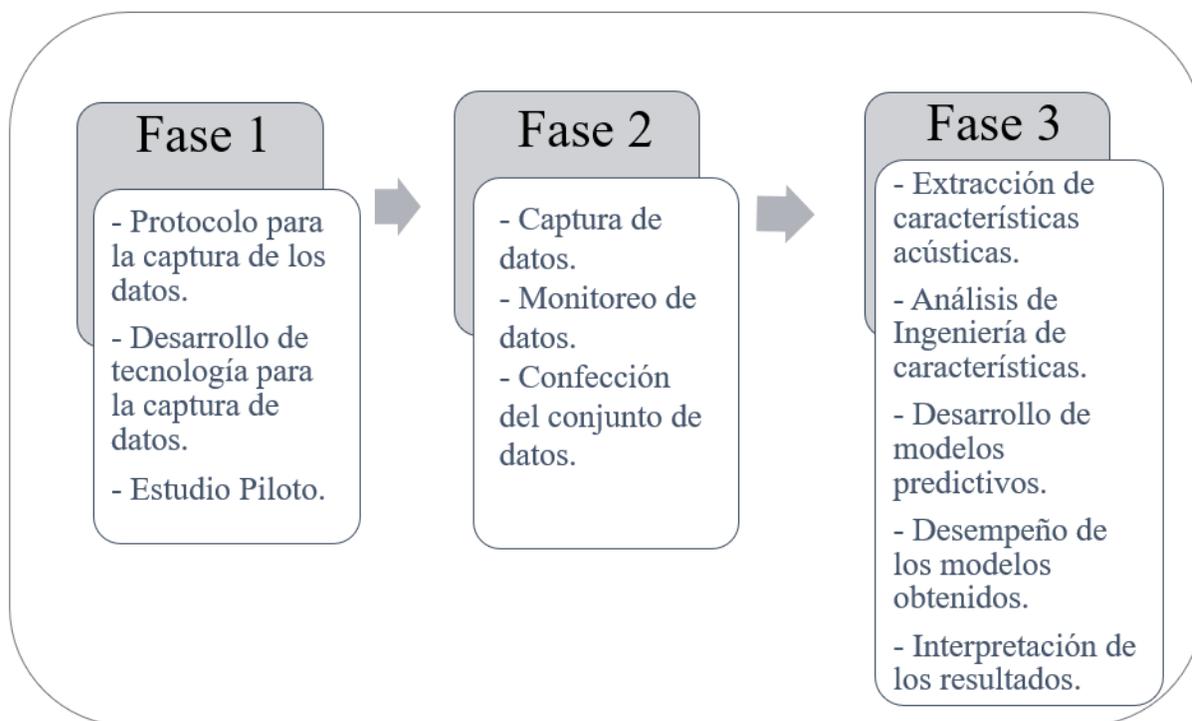
1. Diseñar un protocolo para la recolección de muestras de audio en adultos mayores que asisten a consulta con un médico geriatra.
2. Elaborar un corpus con los registros de audios de adultos mayores y etiquetados con un diagnóstico clínico realizado por un médico geriatra. Hacer públicos los datos anonimizados de los participantes que acepten compartir los registros de su voz.
3. Construir modelos computacionales para la predicción del diagnóstico de demencia utilizando diferentes técnicas de aprendizaje automático.
4. Evaluar el desempeño de los modelos propuestos.

## 1.3 Metodología

Para el desarrollo del presente trabajo de tesis se siguió una metodología en tres fases (ver Figura 1). La primera fase se enfocó en el desarrollo de la metodología para la recolección de datos, la segunda fase se centra en la creación de un conjunto de datos en español de México y en la tercera fase se desarrollan y evalúan modelos predictivos para predecir el diagnóstico de demencia.

El objetivo principal de la *fase 1* fue la confección del protocolo para la captura de datos, teniendo en cuenta el desarrollo de la tecnología para capturarlos y un estudio piloto donde se prueban tanto el protocolo, como la herramienta implementada. Para dar cumplimiento al objetivo de esta fase, se realiza una revisión de la literatura relacionada al perfil de los sujetos que participan en el estudio, actividades comunes que se desarrollan para el diagnóstico de demencia, dispositivos utilizados para la recolección de la información y las metodologías utilizadas.

El objetivo principal de la *fase 2* es la confección de un conjunto de datos en español de México, formado por los archivos de audios resultantes de pedir al adulto mayor desarrollar las actividades predefinidas. El periodo de captura de los datos duró 5 meses, se realiza un constante monitoreo verificando la calidad de estos. Con la finalidad de compartir corpus de audios utilizables, se aplica a cada archivo de audio procesos de limpieza de datos y etiquetados de elementos no analizables. Después de ese tiempo, se comienza a explorar la fase 3, aunque se continuará grabando para seguir enriqueciendo el conjunto de datos.



**Figura 1.** Metodología seguida en el estudio

El objetivo de la *fase 3* es crear modelos utilizando técnicas de aprendizaje automático para distinguir entre los adultos mayores sanos y aquellos con trastornos neurocognitivos. Para dar cumplimiento al

objetivo de esta fase se realiza una revisión a profundidad de la literatura analizando las características a extraer por cada tipo de actividad, los procedimientos de preprocesado y las herramientas utilizadas, con el objetivo de identificar los rasgos que mejor información ofrecen. Además, se analizan los datos recolectados utilizando diferentes técnicas de aprendizaje automático siguiendo la Razón de Ockham (Lazar, 2010). Las tareas para completar esta fase se enumeran a continuación:

1. Extraer rasgos acústicos utilizando las herramientas Praat (Themistocleous et al., 2018) y openSmile (Haider et al., 2019).
2. Aplicar diferentes técnicas de selección de características para identificar rasgos relevantes.
3. Analizar el conjunto de datos con diferentes algoritmos de ia, teniendo en cuenta las características extraídas de una sola actividad, o posibles combinaciones entre ellas.
4. Analizar los resultados obtenidos en las métricas.
5. Aplicar algoritmos de Inteligencia artificial explicable para la interpretación de los modelos.

## **1.4 Organización de la tesis**

El presente trabajo de tesis se encuentra organizado en 6 capítulos. A continuación, se da una breve descripción de cada uno de ellos:

El capítulo 2 trata sobre el estado del arte relacionado al análisis de la voz como biomarcador digital para el diagnóstico de demencia. Los trabajos relacionados sobre las técnicas de aprendizaje automático para discriminar a través de rasgos acústicos si un adulto mayor tiene demencia. Además, de un análisis detallado de los corpus de audios más utilizados con este propósito.

El capítulo 3 describe el protocolo creado y la herramienta desarrollada para la captura de los datos, también, se analizan los resultados de un estudio piloto que se realiza para probar el funcionamiento de ambos en la población objetivo. Se describe el corpus de audio propuesto, el preprocesamiento de los archivos de audios recabados y sus limitaciones.

En el Capítulo 4 se describe detalladamente la metodología que se utiliza para el análisis de los datos, las técnicas utilizadas para la ingeniería de características, para la generación de los modelos, para el balance de los datos y para la interpretación de los modelos.

En el capítulo 5 se presentan los resultados que se obtuvieron de los análisis descriptivos de los rasgos acústicos calculados, el desempeño de los algoritmos de clasificación y la interpretación de los modelos a partir de técnicas de inteligencia artificial explicable. Además, de una discusión de los resultados.

El capítulo 6 presenta las conclusiones de esta investigación, se especifican las aportaciones, sus limitaciones y posibles trabajos futuros.

## Capítulo 2. Biomarcadores digitales de demencia basados en audio

---

En este capítulo se analiza qué son los biomarcadores digitales y cómo estos se pueden utilizar para el diagnóstico de demencia. Se describe la metodología que se emplea para la revisión de la literatura y los trabajos relacionados sobre las técnicas de aprendizaje automático para discriminar a través de rasgos acústicos si un adulto mayor tiene demencia. Además, se lleva a cabo un análisis detallado sobre los corpus de audios más utilizados.

### 2.1 Biomarcadores digitales

El espacio digital se ha abierto cada vez más en el campo de la salud, ejemplo de ello es el uso de biomarcadores digitales en el estudio de diferentes enfermedades. Los biomarcadores digitales son patrones de respuestas obtenidas por interacción de pacientes con herramientas digitales que se correlacionan fuertemente con una condición de salud (Kourtis et al., 2019). Estos contrastan con datos obtenidos a través de mediciones informadas por el paciente o dispositivos y equipos biomédicos tradicionales. Su principal ventaja está marcada en que los datos objetivos son medidos en el mundo real, incluido el hogar, a diferencia de los métodos tradicionales que suelen tomar medidas solamente durante las visitas al hospital, a pesar de que estas mediciones acostumbran a ser más precisas, no son tan frecuentes ni en entornos naturalistas (Piau et al., 2019).

Un área interesante donde comúnmente se empiezan a proponer los biomarcadores digitales es en la detección de demencia en adultos mayores, sustentado por la prevalencia de la condición, dificultad para diagnosticarla y la importancia de prever la trayectoria de la enfermedad para el manejo de esta. Además, se evidencia en la literatura (Varoquaux and Cheplygina, 2022) que los cambios cognitivos, sensoriales y motores pueden preceder a las manifestaciones clínicas por varios años. En particular, los cambios sensoriales y motores (no cognitivos) pueden ayudar a detectar una enfermedad neurológica o neurodegenerativa 10 o 15 años antes de su diagnóstico efectivo.

## 2.2 Demencia en el adulto mayor

La demencia, denominada recientemente como trastorno neurocognitivo, es un síndrome que implica el deterioro de la capacidad mental, el intelecto, el comportamiento, el lenguaje y la capacidad para realizar actividades de la vida diaria, por ejemplo, vestirse, comer, perder la capacidad de resolver problemas sencillos, sufrir alteraciones de comportamientos, perder la orientación espacio-temporal, sufrir estados de confusión, entre otras (López de Ipiña et al., 2013).

En función de la región del cerebro que esté dañada, podemos hablar de varios tipos de demencia: demencias neurodegenerativas, demencias cerebrovasculares y demencias relacionadas con infecciones. La demencia neurodegenerativa es producida por la muerte de las neuronas, lo que implica una disminución permanente y progresiva de la función mental y física en un periodo de tiempo (Matej et al., 2019). Entre las enfermedades relacionadas a este tipo de demencia encontramos la enfermedad de Alzheimer, Parkinson, Huntington, entre otras. En las demencias cerebrovasculares el daño se produce en los vasos sanguíneos del cerebro, provocando un angostamiento de los vasos sanguíneos que alimentan dicho cerebro (Matej et al., 2019). Finalmente, en las demencias relacionadas con infecciones las células se destruyen por causa de virus, bacterias o por un parásito (Matej et al., 2019).

La enfermedad de Alzheimer es el tipo de demencia más común. Se trata de una patología de la enfermedad que es progresiva y afecta a las neuronas de la corteza cerebral y de zonas próximas. Suele iniciarse con una pérdida de la memoria reciente y de la capacidad de aprendizaje, afectando posteriormente a la memoria remota y ocasionando un progresivo deterioro cognitivo y funcional, que concluye con el fallecimiento del paciente en un plazo de tiempo variable, pero que puede alcanzar hasta los diez años desde el momento del diagnóstico (Ferreira et al., 2020).

La enfermedad presenta tres fases. En la fase asintomática o presintomática puede no presentar síntomas clínicos aparentes, comúnmente se diagnostica a través de la medición de biomarcadores en el plasma sanguíneo (es el caso del biomarcador p-tau181), y en el líquido cefalorraquídeo (p-tau217 y p-tau231) (Ashton et al., 2021). Los primeros síntomas son reconocibles en una fase sintomática previa a la demencia y la fase de demencia es solo la etapa final (Panegyres et al., 2016). Por lo tanto, a pesar de que la enfermedad de Alzheimer es el tipo de demencia más frecuente, un adulto mayor puede ser diagnosticado con Alzheimer y no presentar síntomas de demencia, o más comúnmente, sobre todo en países en desarrollo como México, puede ser diagnosticado aún años después de exhibir síntomas, ya que éstos eran asociados por los familiares como parte del proceso normal de envejecimiento.

Se han propuesto distintas maneras de clasificar las fases de la demencia. Las más representativas son: 1) deterioro subjetivo cognitivo (subjective cognitive impairment, SCI) (Garcia, Haider, et al., 2020), 2) deterioro cognitivo no especificado (unspecified impairment, CI) (Garcia, Ritchie, et al., 2020), y 3) deterioro cognitivo leve (Mild cognitive impairment, MCI) (Tóth et al., 2018) y demencia (Mirzaei et al., 2018). El SCI es una categoría diagnóstica que se encuentra entre la normalidad y el envejecimiento patológico, y es un posible predictor de demencia en personas mayores. En el CI se contabilizan los deterioros no especificados. El MCI es una etapa temprana de pérdida de memoria u otra pérdida de capacidad cognitiva (como lenguaje o percepción visual/espacial) en personas que mantienen la capacidad de realizar de forma independiente la mayoría de las actividades de la vida diaria, generalmente no desemboca en demencia. La demencia implica un estado del paciente en el que sufre una gran pérdida de la memoria, así como graves problemas de conducta, movilidad y problemas comunicativos.

El Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-5) (Sachdev and Mohan, 2016) ha reemplazado el término de *demencia* por *trastornos neurocognitivos mayor* (major neurocognitive disorder, Major NCDs) y *MCI* por *trastorno neurocognitivo leve* (mild neurocognitive disorder, Mild NCDs). Esto se hizo en un esfuerzo por reducir el estigma asociado al término demencia. Sin embargo, muchos médicos y organizaciones continúan utilizando los términos demencia y MCI, ya que sus pacientes están acostumbrados a estos términos. La clasificación que vamos a utilizar en esta tesis es la que utiliza el médico geriatra que colabora en la investigación. En este sentido, la demencia se divide en tres categorías, demencia etapa leve, demencia etapa moderada y demencia etapa severa (According to the Alzheimer's Association the Stages of Alzheimer's., n.d.), en la Tabla 1 se muestran los diferentes síntomas de cada etapa. Puede ser difícil ubicar a una persona con demencia en una etapa específica, ya que las etapas pueden superponerse. Nuestra investigación se centra en el estudio de demencia etapa leve y etapa moderada, como una posible vía para identificar cambios en la población preclínica, ya que los tratamientos actuales son más efectivos en etapas tempranas.

### 2.2.1 Evaluación diagnóstica

Tradicionalmente, el diagnóstico de demencia se basa en un conjunto de actividades que realiza el geriatra para evaluar la memoria y las habilidades para pensar de un adulto mayor. Este conjunto de actividades está compuesto por una serie de pruebas cognitivas para detectar el grado de deterioro cognitivo. Cada una consiste en responder a una serie de preguntas o hacer tareas simples. Están diseñadas para medir funciones mentales como memoria, lenguaje y capacidad de reconocer objetos. Algunos tipos de pruebas

más comunes son: Mini-Mental State Examination (MMSE) (López de Ipiña et al., 2018), Clinical Dementia Rating (CDR) (García, Ritchie, et al., 2020), Clock Drawing Test (Gosztolya et al., 2019), General Deterioration Scale (GDS) (Duong et al., 2005) y Montreal Cognitive Assessment (MoCA) (Calz et al., 2020).

**Tabla 1.** Síntomas de las diferentes etapas de la demencia.

Tipo de demencia	Signos y síntomas	Duración esperada
<b>Etapa leve</b>	Una persona puede funcionar de manera independiente, conducir, trabajar y ser parte de actividades sociales. A pesar de esto, la persona puede sentir que tiene lapsos de memoria, como olvidar palabras familiares o la ubicación de objetos cotidianos.	Duración promedio de 2 años
<b>Etapa moderada</b>	La persona puede confundir las palabras, frustrarse o enojarse y actuar de manera inesperada, como negarse a bañar. Presenta dificultad para expresar sus pensamientos y realizar tareas rutinarias sin ayuda.	Duración promedio 1.5 años
<b>Etapa severa</b>	La persona pierde la capacidad de responder a su entorno, de mantener una conversación y, finalmente, de controlar el movimiento. Puede decir palabras o frases, pero comunicar el dolor se vuelve difícil.	Duración promedio de 2.5 años

Por otro lado, también se realizan un conjunto de actividades con diferentes niveles de complejidad del lenguaje, que permiten examinar diferentes aspectos de la memoria, el procesamiento semántico y las medidas acústicas y fonéticas. Algunas de las tareas que se aplican están enfocadas en desarrollar habla espontánea (López de Ipiña et al., 2018), por ejemplo, se le pide al adulto mayor describir una imagen, recordar una película, un sueño o entablar una conversación. Otras a evaluar la función cognitiva, como las tareas de fluidez verbal semántica, que consiste en pedirle al adulto mayor que mencione los animales o frutas que recuerda en un minuto (Gogtay and Thatte, 2017).

Otra vía prometedora tanto para diagnosticar el deterioro neurocognitivo como para controlar la progresión de este es la imagenología. Para el análisis de esta se utilizan dos tipos de estudios que detectan

alteraciones morfológicas y moleculares específicas de la enfermedad: la resonancia magnética funcional (RMF) (Jagust, 2006) que mide el metabolismo cerebral de la glucosa y el oxígeno, y la neuro imagenología molecular con sus dos variantes: la tomografía por emisión de positrones (PET) (Jagust, 2006) y la tomografía computarizada por emisión de fotones simples (SPET). Estos dos últimos estudios pueden detectar los depósitos amiloides, característicos de la enfermedad de Alzheimer.

Los exámenes neuro imagenológicos han revolucionado el estudio estructural, suelen ser más precisos, pero no están asequibles para todos los pacientes por los altos costos y naturaleza invasiva, impidiendo su uso repetido y frecuente para evaluar a un individuo, específicamente en la etapa presintomática temprana.

El diagnóstico precoz y preciso de trastornos neurocognitivos han ido atrayendo progresivamente más atención en los últimos años. La relativa eficiencia y confiabilidad de varios procedimientos reportados en la literatura han sido probados en diferentes estudios. Sin embargo, estos métodos son costosos, difíciles y lentos de usar. Actualmente el diagnóstico de demencia continúa siendo un problema abierto, en esta dirección se enmarca nuestra investigación, en el análisis de la voz, una técnica relativamente económica y no invasiva.

## **2.3 Detección de demencia analizando la voz**

La evolución de los problemas derivados de la demencia provoca una degeneración paulatina en el lenguaje del adulto mayor. Las limitaciones comienzan a entorpecer los procesos comunicativos de la persona afectada con su entorno, caracterizando dichos intercambios como frustrantes, generadores de ansiedad, de aislamiento y de reacciones anímicas depresivas. El potencial para usar el habla como un biomarcador para el análisis de demencia se basa en varios valores prospectivos, que incluyen: la facilidad con la que se puede grabar y rastrear el habla a lo largo del tiempo, es un método no invasivo y el hecho de que las tecnologías para el análisis del habla han mejorado marcadamente en la última década.

### **2.3.1 Metodología para la revisión de la literatura**

Se realiza una revisión sistemática, en la que se consultaron tres bases de datos: PubMed, IEEExplore y ACM entre los años 2011-2022, con el objetivo de analizar la literatura más reciente sobre el análisis de

voz utilizando técnicas de aprendizaje automático para detección de demencia. La selección de las bases de datos estuvo dada porque Pubmed proporciona acceso a temáticas relacionadas con salud, y las otras dos tienen buena cobertura a temas de computación, procesamiento de audio y áreas afines.

El cribado para la selección de artículos se realizó en dos fases. En la primera fase se examinaron los títulos y resúmenes de los artículos resultantes de las búsquedas en función de los siguientes criterios de inclusión: 1) los participantes debían presentar demencia; 2) evidencia de la presencia de análisis del habla de los participantes; 3) había un elemento de aprendizaje automático, y 4) los artículos completos debían estar disponibles en inglés. Además, se excluyeron los artículos que solo tuvieran el resumen disponible y aquellos relacionados a revisiones, libros y capítulos de libros. En la segunda fase se realizó una lectura profunda a los artículos que no pudieron incluirse o excluirse de manera absoluta basándose únicamente en el título y la información del resumen. Todos los títulos emergentes que se consideraron relevantes se agregaron al proceso de selección.

### 2.3.2 Corpus de audios

Diferentes investigaciones (Calz et al., 2020)(Gosztolya et al., 2016)(Warnita et al., 2018) enfatizan la relevancia de los corpus de audios utilizados para el diagnóstico de demencia. A pesar de que en la actualidad han estado surgiendo conjuntos de datos, aún es un tema abierto, mostrando la necesidad de crear nuevos, preferentemente de acceso público. En la literatura se identificaron varios corpus, tales como: Carolina Conversations Collection (CCC) (Haider et al., 2019), PGA-OREKA (López de Ipiña et al., 2018), MINI-PGA (López de Ipiña et al., 2018). Los dos más utilizados son Pitt Corpus (Haider et al., 2019)(Luz et al., 2020)(Warnita et al., 2018)(Zimmerer et al., 2016) y ADReSS challenge dataset (Garcia, Ritchie, et al., 2020)(Luz et al., 2020) y el publicado por Ivanova (Ivanova et al., 2022) recientemente en idioma español. A continuación, vamos a describir y analizar por relevancia a nuestra investigación los tres últimos.

El **Pitt Corpus** fue publicado en el año 1994 (Becker and Boller, 1994), es el resultado de un tipo de estudio que se focaliza en las evaluaciones neuropsicológicas longitudinales a través de las siguientes tareas: descripción de la imagen llamada “Robo de galletas”, prueba de fluidez verbal semántica, recuerdo de historias, y construcción de oraciones. Contiene la información de 208 participantes con demencia, 104 pertenecientes al grupo de control y 85 con diagnóstico desconocido. En la Tabla 2 se muestran las características demográficas de los mismos.

**Tabla 2.** Características básicas de los participantes en el estudio Pitt Corpus.

	<b>Grupo de control</b> <i>Media (sd)</i>	<b>Demencia</b> <i>Media (sd)</i>
<b>Edad</b>	63.8(8.3)	71.4(8.3)
<b>Educación</b>	14.3(2.9)	12.1(2.9)
<b>Mini mental (MMSE)</b>	29.1 (1.1)	18.4(5.2)
<b>Sexo (M/F)</b>	44/57	59/121

El diagnóstico de cada participante se realizó mediante la recopilación de la información de detección, haciendo que el cuidador principal completara un cuestionario por correo sobre síntomas de demencia e historial médico. Luego se desarrolló una evaluación neuropsiquiátrica exhaustiva que incluía la historia clínica y el examen físico, la historia y el examen neurológicos, la entrevista psiquiátrica semiestructurada y la evaluación neuropsicológica. Además, les aplicaron exámenes de laboratorio. Cada evaluación se completó en aproximadamente tres sesiones, generalmente dentro de un período de 2 semanas. Al finalizar estos estudios, el equipo revisó cada conjunto individual de resultados y establecieron el diagnóstico clínico y neuropatológico final a través de consenso y con ayuda de la autopsia de los pacientes fallecidos (Becker and Boller, 1994).

El **ADReSS challenge dataset** (Luz et al., 2020) es un conjunto de datos, más reciente, pero también muy utilizado. El ADReSS Challenge tiene como objetivo proporcionar a la comunidad, que investiga sobre la demencia, una plataforma para realizar comparaciones metodológicas integrales y analizar diferentes enfoques para la detección de esta. El conjunto de datos ADReSS está compuesto por 78 participantes sanos y 78 con demencia. La información recolectada es a partir de la descripción de la imagen llamada "Robo de galletas". En las Tablas 3 y 4 se muestran las características demográficas de los participantes, divididas en los grupos de entrenamiento y de prueba.

Ambos conjuntos de datos están en inglés y se han utilizado con el objetivo de predecir a través de técnicas de aprendizaje automático, si el adulto mayor presenta demencia o no. Aunque el Pitt corpus no fue creado con esa finalidad, es muy empleado en este tipo de estudio, pudiendo estar sesgado por la disponibilidad de este.

**Tabla 3.** Características básicas de los participantes pertenecientes al conjunto de entrenamiento.

Edad	Demencia			Grupo de control		
	M	F	MMSE (sd)	M	F	MMSE (sd)
[50, 55)	1	0	30.0 (n/a)	1	0	29.0 (n/a)
[55, 60)	5	4	16.3 (4.9)	5	4	29.0 (1.3)
[60, 65)	3	6	18.3 (6.1)	3	6	29.3 (1.3)
[65, 70)	6	10	16.9 (5.8)	6	10	29.1 (0.9)
[70, 75)	6	8	15.8 (4.5)	6	8	29.1 (0.8)
[75, 80)	3	2	17.2 (5.4)	3	2	28.8 (0.4)

**Tabla 4.** Características básicas de los participantes pertenecientes al conjunto de prueba.

Edad	Demencia			Grupo de control		
	M	F	MMSE (sd)	M	F	MMSE (sd)
[50, 55)	1	0	23.0 (n.a)	1	0	28.0 (n.a)
[55, 60)	2	2	18.7 (1.0)	2	2	28.5 (1.2)
[60, 65)	1	3	14.7 (3.7)	1	3	28.7 (0.9)
[65, 70)	3	4	23.2 (4.0)	3	4	29.4 (0.7)
[70, 75)	3	3	17.3 (6.9)	3	3	28.0 (2.4)
[75, 80)	1	1	21.5 (6.3)	1	1	30.0 (0.0)

Considerando que son los dos conjuntos de datos más usados, se realiza un análisis más detallado de ellos.

Encontrando los siguientes resultados:

- En más de la mitad de los archivos se identificó ruido de fondo y en otros se escuchaba lejano.
- Intervención del entrevistador con muletillas como “ok”, “mhm”, y otras expresiones orientadas a motivar al adulto mayor a continuar con la tarea como, por ejemplo: “Qué más está viendo”.
- Intervención de risas, personas y ruidos externos.
- En el caso del Pitt corpus, se evidenció un fuerte desbalance de condición en la cantidad de sujetos que desarrollan la actividad de Fluidez verbal semántica, en el grupo de control solamente participaron 2 sujetos y en el de demencia 333 sujetos.

Se puede constatar que estos conjuntos de datos son inadecuados para obtener un diagnóstico certero, dadas algunas limitaciones en la calidad de los archivos de audios utilizados, así como la introducción de posibles sesgos en la recolección de estos. Esta interpretación se puede evidenciar en el caso del corpus ADReSS, donde el desempeño de los algoritmos de clasificación no sobrepasó el 62,5 % (Luz et al., 2020).

**Ivanova corpus** (Ivanova et al., 2022) fue publicado en el año 2022 en la Universidad de Salamanca. El corpus contiene información de 361 adultos mayores con diferentes perfiles cognitivos (Sanos, MCI, Demencia). Los archivos de audios están formados por la lectura del primer párrafo de “El ingenioso hidalgo don Quijote de la Mancha”, de Miguel de Cervantes. Todos los participantes fueron sometidos a una evaluación neuropsicológica previa que incluyó evaluación de las actividades de la vida diaria, una evaluación psicológica y cognitiva. En la Tabla 5 se muestran las características demográficas de los adultos mayores que participaron en el estudio.

**Tabla 5.** Características básicas de los participantes en el Ivanova corpus.

<b>Grupo</b>	<b>No de sujetos</b>	<b>MMSE (SD)</b>	<b>Sexo</b>	<b>Edad</b>	<b>Escolaridad</b>
<b>Sano</b>	n = 197	28.26 (1.890)	Men = 58 Women = 139	75.5 (7.929)	9.58 (3.848)
<b>MCI</b>	n = 90	23.89 (4.082)	Men = 25 Women = 65	79.49 (9.605)	8.67 (3.938)
<b>Demencia</b>	n = 74	19.97 (5.174)	Men = 30 Women = 44	79.49 (7.921)	8.81 (4.095)

Los archivos de audios que conforman el corpus tienen una excelente calidad auditiva, no se evidencia la presencia de elementos no analizables, como en los casos mencionados arriba.

### 2.3.3 Aprendizaje automático para el análisis de la voz

La afectación del lenguaje es una de las consecuencias iniciales del deterioro neurocognitivo (Petti et al., 2020). Entre las dificultades que proyecta el lenguaje del adulto mayor, se encuentran: afasia, anomia, automatismos, parafasias, circunloquios, estereotipias y la ecolalia, el uso de neologismos y de muletillas o palabras carentes de sentido, alteraciones en la longitud de las emisiones verbales y severas dificultades para seguir el tema de una conversación (Petti et al., 2020).

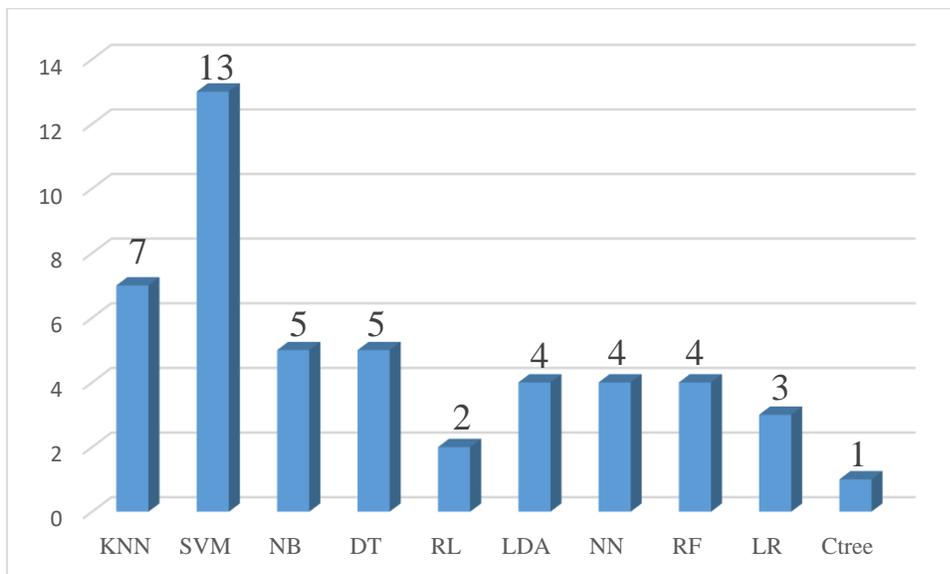
Se han logrado resultados prometedores en la detección de demencia mediante el procesamiento de lenguaje natural (Natural language processing, NLP) (Ammar and Ayed, 2018)(Fang et al., 2017)(Fraser et al., 2018), el procesamiento de señales (Signal Processing, SP) (Purwins et al., 2019) y el aprendizaje automático (Machine Learning, ML) (Calz et al., 2020)(Mirzaei et al., 2018)(Sadeghian et al., 2017). La NLP se ocupa de la comprensión, el aprendizaje y la producción del lenguaje humano utilizando herramientas computacionales. SP explora las señales y la información que transmiten, se ocupa de cómo se pueden transformar, manipular y representar. ML se centra en algoritmos que pueden mejorar automáticamente en función de la experiencia y permite detectar patrones entre los datos recopilados facilitando su análisis.

El aprendizaje automático ha demostrado ser eficaz en el diagnóstico de demencia a través del análisis de la voz (Petti et al., 2020). Garantizar un correcto desempeño de estas técnicas depende desde la captura inicial de los datos hasta los algoritmos que se utilicen para modelar el problema. Nuestra investigación se enfoca en analizar mediante técnicas de ML si un adulto mayor sano y uno con demencia se pueden discriminar automáticamente en función de las características acústicas.

Basado en nuestra revisión de la literatura diferentes algoritmos de ML han mostrado resultados favorables en este sentido, tales como: K-vecinos más cercanos (k-nearest neighbors, KNN) (Nasrolahzadeh et al., 2018) y Máquina de soporte vectorial (Support Vector Machine, SVM) (Syed et al., 2021), Naive Bayes (NB) (Luz, 2017), Árboles de decisión (Decision Tree, DT) (Garcia, Ritchie, et al., 2020), Regresión logística (Logistic regression, RL) (Chinaei et al., 2017), Análisis discriminante lineal (Linear Discriminant Analysis, LDA) (Haider et al., 2019), Redes neuronales (Neural Network, NN) (López de Ipiña et al., 2018), Random forest (RF) (Balagopalan et al., 2021), Regresión lineal (Linear Regression, LR) (Tanaka et al., 2016) y Árboles de inferencia condicional (Conditional Inference Trees, Ctree) (Khodabakhsh et al., 2015). En la Figura 2 se muestra que los algoritmos más populares fueron el SVM y KNN.

La construcción de una representación de características adecuada y el diseño de un clasificador apropiado para estas características a menudo se han tratado como problemas separados en el procesamiento de audio. Un señalamiento de este enfoque es que las características diseñadas pueden no ser óptimas para el objetivo de clasificación en cuestión (Purwins et al., 2019). Por otro lado, investigaciones ya han demostrado que varios trastornos se han diagnosticado satisfactoriamente utilizando diferentes arquitecturas de aprendizaje profundo (Deep Learning) (Gautam and Sharma, 2020). Los métodos basados en aprendizaje profundo se han empleado con éxito en áreas médicas y han logrado un rendimiento a nivel humano en tareas de clasificación. Estas han demostrado ser muy eficaces en la clasificación de imágenes y son prometedoras para el análisis de audio (Chlasta and Wołk, 2021). Sin embargo, explorar

estas técnicas en nuestro estudio no fue posible, por la cantidad de datos que pudimos recabar en el periodo de duración de la tesis.



**Figura 2.** Algoritmos de clasificación utilizados para el diagnóstico de demencia.

La Tabla 6 incluye de forma resumida, las características acústicas analizadas, las técnicas de ML utilizadas y las métricas reportadas por cada estudio. El número de características acústicas empleadas en la mayoría de los estudios fue elevado, y en algunos no se reportaron en su totalidad o detalladamente. Las principales herramientas utilizadas para la extracción de características fueron Praat (Tóth et al., 2018) y OpenSmile (Haider et al., 2019), en tanto que para la tarea de clasificación fueron Weka (Yeung et al., 2021) y Matlab (Garcia, Haider, et al., 2020). Si bien estos estudios han abordado primordialmente el diagnóstico de demencia, el análisis de audio podría también ser usado para caracterizar variantes del problema. Por ejemplo: 1) Analizar el diagnóstico de Demencia/Sano o Demencia/MCI/Sano, 2) Trayectoria de la demencia, considerando que los adultos mayores con MCI tienen una probabilidad de un 10 o 15% de evolucionar a demencia (Ivanova and Meilán, 2022), y 3) Diagnosticar si un adulto mayor está en un periodo de lucidez, por ejemplo, para recibir explicaciones y firmar documentos legales.

En este capítulo se corrobora que la predicción de demencia en adultos mayores aún continúa siendo un problema abierto, permitiendo estudiar la voz como biomarcadores digitales para el diagnóstico de esta condición. El desempeño de los algoritmos de clasificación obtenidos en los trabajos relacionados evidencia la aplicabilidad de estudiar la voz a partir de técnicas de aprendizaje automático, y por lo tanto

sugiere que estos pudieran ser útiles para discriminar a partir del análisis de rasgos acústicos la condición de demencia. Sin embargo, se identificaron hallazgos que resaltan la necesidad de crear corpus de datos en español con la calidad requerida para obtener diagnósticos certeros. Además, al mejor de nuestro conocimiento no se identificaron en la literatura estudios donde se analizan rasgos acústicos en función de la evaluación de diferentes métricas utilizando decisión multicriterio, ni la presencia de optimización de parámetros en los algoritmos de clasificación, ni estudios que involucren interpretación de los modelos obtenidos utilizando técnicas de inteligencia artificial explicable.

**Tabla 6.** Información extraída de los estudios analizados

Estudio	Tamaño AD/MCI	Tamaño control	Recopilación de datos	Rasgos acústicos	Algoritmo clasificación	Medidas
A. Khodabakhsh y otros. 2015 (Khodabakhsh and Demiroglu, 2015)	AD n=27	n=27	OT: preguntas casuales / conversacionales	Actividad de voz, articulación y tasa de habla	SVM, LDA, DT	Acc= 88%
A. Khodabakhsh y otros. 2015 (Khodabakhsh et al., 2015)	AD n=51	n=28	OT: preguntas casuales / conversacionales	Actividad de voz, articulación y tasa de habla	SVM, NB, NN, CTree	Acc = 80%
H. Tanaka y otros. 2016 (Tanaka et al., 2016)	AD n=10	n=10	SS: diálogos multimodales	F0, potencia, velocidad de voz y calidad de voz.	SVM, LR	AUC=94%
F. Martínez y otros. 2017 (Martínez-sáñez et al., 2017)	AD n= 45	N=82	SS: Leer 1er párrafo de Don Quijote de la Mancha	Mediciones de la variabilidad silábica	-	AUC = 87% SEN = 82.2% SPE = 81.7%
S. Al-hameed y otros. 2017 (Al-hameed et al., 2017)	AD n =16 MCI n =6	n=42	SS: descripción de la fotografía "cookie theft"	Frecuencias, jitter, shimmer, etc	KNN, SVM	Acc = 89.2% - 92.4%
H. Chinaei y otros. 2017 (Chinaei et al., 2017)	AD n=167	n=97	SS: descripción de la fotografía "cookie theft"	Proporción de pausa, F0, duración total del habla, etc.	RL	Acc = 82% - 96.1%
P. Vasco y otros. 2017 (Vasco et al., 2017)	MCI n=38	n=187	SVF: nombrar animales	Dominio espectral tiempo, no lineales	KNN, SVM, aprendizaje profundo	-
S. Luz . 2017 (Luz, 2017)	AD n=214	n=184	SS: descripción de la fotografía "cookie theft"	Eventos de vocalización y velocidad del habla	NB	Acc= 68% F1 = 64% - 70%
M. Nasrolahzadeh y otros. 2018 (Nasrolahzadeh et al., 2018)	AD n = 30	n=30	SS: conversaciones sobre historias amistosas, y sus sentimientos	Espectros de orden superior (HOS)	KNN, SVM, NB, DT	Acc = 97.71% Sen = 93.33 Spe = 98.33

V. Vincze y otros. 2018 (Tóth et al., 2018)	MCI n=48	n=36	SS: Recuerdo inmediato y diferido de una película	Velocidad de articulación, duración total pausas, etc	SVM, NB, RF	Acc = 75.0% F1= 78.8% Spe= 80.3% Sen = 81.3% Prec = 79.4%
C. Themistocleous y otros. 2018 (Themistocleous et al., 2018)	MCI n=25	n=30	OT: leer un pasaje corto	Formantes de las vocales, F0, duración de las vocales, género, sexo	Aprendizaje profundo	Acc = 75% Prec = 100% Sen = 72% F1= 72%
U. Martínez-delizarduy y otros. 2018 (López de Ipiña et al., 2018)	Estudio 1: AD n=20 Estudio2: MCI n=38	Estudio 1: n=50 Estudio 2: n=187	SS: Descripción de fotografía. SVF: nombrar animales	MFCC, energía, características no lineales, etc	NN y Aprendizaje profundo	Acc=73%-95%
Z. Guo y otros. 2019 (Guo et al., 2019)	AD n=169	n=99	SS: descripción de la fotografía "cookie theft"	Fonética-fonología	KNN, SVM, DT, RL, RF	Acc = 80.7% - 85.4%
F. Haider y otros. 2019 (Haider et al., 2019)	AD n=82	n=82	SS: descripción de la fotografía "cookie theft"	Derivados del openSMILE	KNN, SVM, DT, LDA, RF	Acc =71.34% - 78.70%
B. Mirheidari y otros. 2019 (Mirheidari et al., 2019)	AD n=19 MCI n=18	n=14	Pruebas de fluidez verbal: SVF: animal PVF: letra "P" SS: Temas personales	Entonación general media, duración y silencio, jitter y shimmer, etc.	LR	Acc = 48% - 62% AUC = 82%
S. De y otros. 2020 (Garcia, Haider, et al., 2020)	Estudio 1: AD n=82 Estudio 2: AD n=30	Estudio 1 n=82 Estudio 2 n=16	SS: descripción de la fotografía "cookie theft"	Derivados del openSMILE	KNN, SVM, DT, LDA	Acc = 70.73% - 84.78%
R. Nagumo y otros. 2020 (Nagumo et al., 2020)	MCI n=1601	n=6343	OT: Leer sentencias cortas	F1 and F2, F0, tempo del habla, etc	LDA	AUC = 61%-77%
Z. S. Syed y otros. 2021 (Syed et al., 2021)	-	-	SS: descripción de la fotografía "cookie theft"	Acento, tempo y entonación	SVM, LR, y Aprendizaje profundo	Acc = 5.37% Sen = 6.30% SPE = 4.44%
A. Yeung y otros. 2021 (Yeung et al., 2021)	AD n=10 MCI n=10	n=10	SS: descripción de la fotografía "cookie theft"	Velocidad del habla, pausas largas, etc	-	-
A. Balagopalan y otros. 2021 (Balagopalan et al., 2021)	AD n=78	n=78	SS: descripción de la fotografía "cookie theft"	Pausas y rellenos, F0, tasa de cruce MFCC	SVM, NB, NN, RF	Acc= 87% Prec = 90% Sen = 83% Spe = 91% F1 = 87%
K. Chlasta y otros. 2021 (Chlasta and Wołk, 2021)	AD n =78	n=78	SS: descripción de la fotografía "cookie theft"	Extractor de rasgos VGGish	KNN, SVM, NN y Aprendizaje profundo	Acc = 64% F1= 69% Sen= 69% Prec = 69%
A. Balagopalan y otros. 2021 (Balagopalan et al., 2021)	AD n =78	n=78	SS: descripción de la fotografía "cookie theft"	Pauses and filler F0 Duration-related Zero-crossing rate MFCC	SVN, NN, RF, NB, BERT	ACC= 83% SEN = 83% SPE = 83% F1 = 83% AUC = 83%

## Capítulo 3. Protocolo, captura y etiquetado del conjunto de datos

---

En este capítulo se describe el protocolo para la captura de datos que se confeccionó y un estudio piloto para verificar que fuera adecuado y hacer las modificaciones pertinentes para proceder con la captura y etiquetado de datos. Este protocolo fue aprobado por el comité de bioética del CICESE (BIOETICA\_HUM\_2021\_06) antes de comenzar el periodo de grabaciones, teniendo en cuenta: las cartas de consentimientos tanto para el paciente como para el familiar, en caso de que el paciente se encuentre imposibilitado de firmar (ver Anexo A), medidas de ética y privacidad correspondientes y autorización de los adultos mayores (o familiar) para hacer sus datos públicos. Además, se detalla en el capítulo la aplicación desarrollada para recabar los archivos de audio que se analizan en esta tesis y la descripción del corpus de audio (de acceso público) que se creó en idioma español de México, formado por los archivos de audios resultantes del desarrollo de las actividades con los adultos mayores y su etiquetado.

### 3.1 Protocolo utilizado para la captura de datos

El protocolo para la captura de los datos se elaboró por un grupo interdisciplinario que incluye a especialistas en ciencia de datos, interacción humano-computadora, geriatría y psicología, llegando a consensos sobre los puntos fundamentales que se han desarrollado, y teniendo en cuenta las recomendaciones reportadas en la literatura. Los criterios que se tienen en cuenta son los siguientes:

**Sitio del estudio:** Consulta privada del médico geriatra en Ensenada, Baja California.

**Duración del estudio:** La colección de las grabaciones se realiza en una primera fase de 5 a 6 meses. El rango de fechas podrá variar ante la incertidumbre provocada por la epidemia por COVID-19 que podría restringir el acceso a la clínica para recabar datos.

**Participantes:** Los sujetos que participan en el estudio son adultos mayores que asisten a la consulta privada del médico geriatra, para una revisión de rutina y, además cumplen con los siguientes criterios de inclusión:

- 1) Adultos mayores con más de 60 años.

2) Adultos mayores cuya lengua nativa sea el español.

Y los siguientes criterios de exclusión:

- 1) Adultos mayores que no presentan capacidad para comunicarse verbalmente
- 2) Adultos mayores que presentan problemas auditivos.
- 3) Adultos mayores que no aceptaran participar.
- 4) Adultos mayores que no hayan firmado la carta de consentimiento, ni sus familiares o cuidadores estén de acuerdo en firmar.

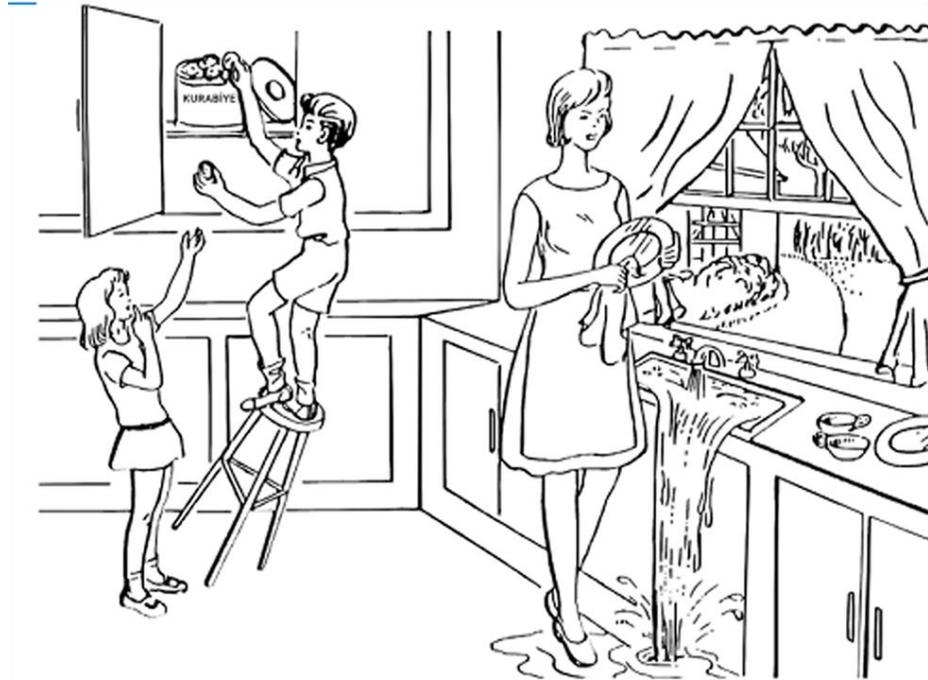
**Pruebas cognitivas que desarrollará el médico geriatra:** El diagnóstico de demencia será determinado por la médico geriatra basado en una batería de pruebas que incluye Mini-Mental State Examination (MMSE) (Garcia, Ritchie, et al., 2020), prueba de Fluidez verbal semántica (semantic verbal fluency test, SVF) (Petti et al., 2020) y prueba de Dibujo del reloj (Clock Drawing Test, CDT) (Garcia, Haider, et al., 2020).

**Actividades para desarrollar:** Durante el estudio se solicita a los participantes realizar las siguientes tareas:

- Habla espontánea. Tiene el objetivo de desencadenar el habla para analizar una variedad de atributos del lenguaje, como procesos de recuperación de palabras, deterioro sintáctico, semántico, acústico y errores de comunicación (Faroqi Shah et al., 2020). Para elicitación de habla espontánea en este estudio se pide al participante describir la imagen llamada "Robo de galletas" (ver Figura 3), utilizando la siguiente instrucción "*Mencione todo lo que ve que está pasando en la siguiente imagen*".
- Fluidez Verbal Semántica (McDonnell et al., 2020). Se utiliza para determinar una evaluación global y rápida del lenguaje, la habilidad de recuperación de información y el funcionamiento ejecutivo. En este caso, la instrucción utilizada es la siguiente: "*Nombre las frutas que recuerde en un minuto*".
- Repetir frases (Nagumo et al., 2020). Es utilizada para observar una respuesta posterior, velocidad del habla y errores en las repeticiones como resultado del déficit cognitivo. Con este propósito, la instrucción utilizada es "*Repita cada una de las siguientes oraciones*" y a continuación se enuncian

un total de 10 oraciones, dando tiempo al participante de repetir cada una de ellas antes de proceder con la siguiente.

Cada actividad tiene una duración aproximada de 60 segundos.



**Figura 3.** Imagen llamada Robo de galletas, utilizada para desarrollar la actividad de “Habla espontánea” (Cummings, 2019).

**Descripción del procedimiento:** A continuación, se describe el protocolo propuesto para la captura de datos con cada uno de los participantes.

1. El asistente de investigador (pasante de medicina o psicología previamente capacitado por el médico geriatra) explica el estudio al participante y familiar (si viene acompañado).
2. Se entrega el formato de consentimiento informado al participante y/o familiar. Se le pide leerlo y firmar si está de acuerdo en participar en el estudio.
3. Se desarrollan las actividades de “Repetir frases”, “Fluidez verbal semántica” y “Robo de galletas” con apoyo de la herramienta de captura desarrollada (ver sección 3.2). Se explica al paciente las instrucciones antes de comenzar cada actividad.

Este procedimiento se realiza antes de la consulta con el médico geriatra que aplica sus propias pruebas (MMSE, Prueba del Reloj y SVF:animales) para realizar un diagnóstico del participante. El estudio es doble ciego: el médico geriatra no conoce el resultado de las actividades realizadas en el estudio de registro de voz y el asistente de investigador no conoce el resultado de las pruebas realizadas por el médico geriatra. Se utiliza un identificador único de paciente (IUP) para cruzar los datos posteriormente para su análisis, manteniendo la privacidad del participante.

**Características de la grabación:** Para estandarizar los archivos de audios recolectados se tomarán en cuenta las siguientes características:

- Todos los participantes usarán cubrebocas mientras sea un requisito para asistir a la clínica.
- Los datos de voz se registrarán con un micrófono cardioide modelo Q2U USB/XLR Dynamic Micrófono Grabación y Podcasting Pack, que será colocado sobre una mesa. El participante estará sentado en una silla aproximadamente a 30 centímetros del micrófono.
- Las grabaciones tendrán una duración de aproximadamente 3-5 minutos, y se almacenarán con una frecuencia de muestreo de 48kHz y cuantificación a 16 bits.
- El médico geriatra a través del IUP proporcionará el diagnóstico de cada individuo, así como datos demográficos como edad y género del participante.

### 3.2 Herramienta para la captura de datos

Se creó una aplicación con el objetivo de capturar los registros de voz de los adultos mayores en el desarrollo de las actividades cognitivas descritas anteriormente. La interacción con la aplicación es a través de dos procedimientos, uno es de forma manual bajo el control de un operador, y el otro es por medio de un procedimiento autónomo que guía al usuario en las distintas actividades y requiere mínima participación de un operador. En ambos procedimientos los adultos mayores desarrollan las mismas actividades. El sistema incluye tres módulos que corresponden a cada una de las actividades: “Repetir frases”, “Fluidez verbal semántica” y “Robo de galletas”. Las mismas suelen ser usadas para elicitación de respuestas de adultos mayores para el diagnóstico de demencia o afasia. Como resultado se obtiene 1

fichero por cada una de las actividades “Robo de Galletas” y “Fluidez verbal semántica” y 10 ficheros de la actividad “Repetir frases”. Las Figuras 4 y 5 muestran la modelación conceptual de los dos procedimientos.

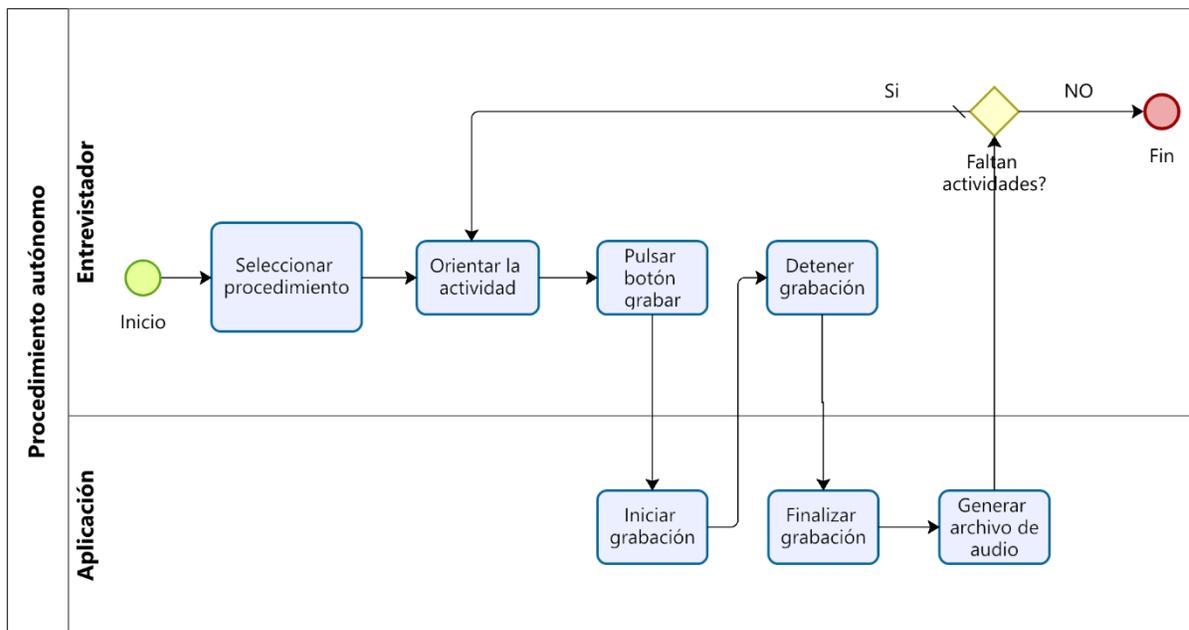


Figura 4. Modelo conceptual del procedimiento manual de captura de voz.

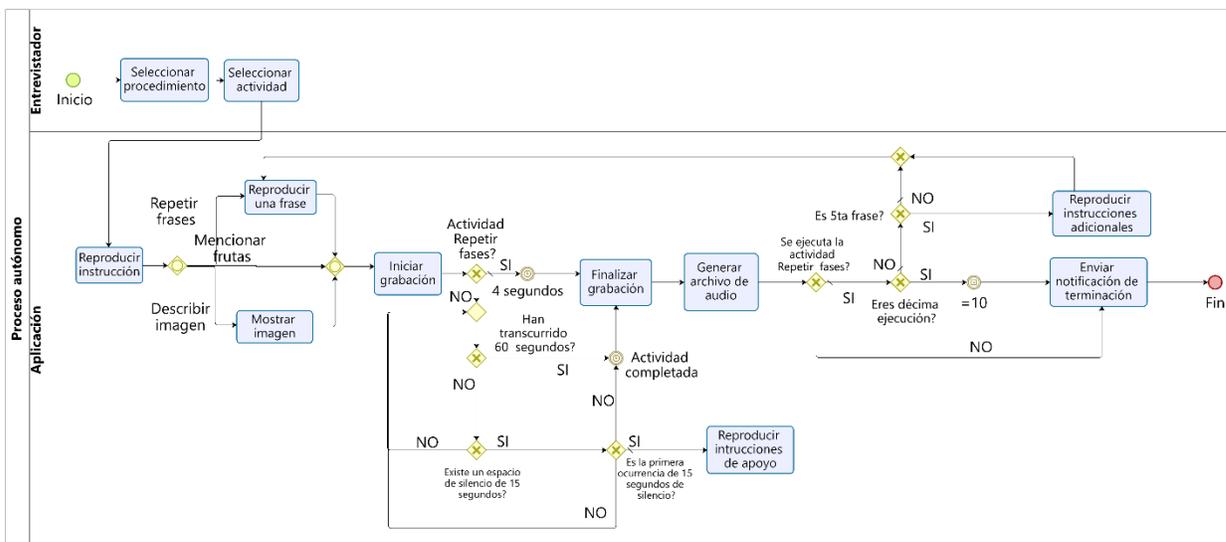


Figura 5. Modelo conceptual del procedimiento autónomo de captura de voz.

En la Figura 5 se puede observar que en la actividad 2 “Robo de galletas” y en la actividad 3 “Mencionar frutas”, si el adulto mayor sostiene 15 segundos de silencio el sistema le hace un recordatorio como el siguiente: “Qué más puede identificar en la imagen?”. Si se detecta un segundo espacio de silencio, se

finaliza la actividad automáticamente con un mensaje indicando que ha terminado. En la actividad 1 “Repetir frases”, después de la quinta oración se le repite la instrucción de la actividad. En la tabla 7 se muestra la descripción detallada de cada una de las actividades y bifurcaciones utilizadas en los modelos conceptuales de los procedimientos mostrados en la Figura 4 y 5.

**Tabla 7** Descripción de los procesos

<b>Actividad</b>	<b>Descripción</b>
Seleccionar procedimiento	Se selecciona el procedimiento que se va a utilizar para el desarrollo de las actividades, son dos opciones: a) Procedimiento manual y 2) Procedimiento autónomo (Entrevistador).
Orientar la actividad	Se enuncia la orientación de las actividades (Entrevistador).
Pulsar botón grabar	Se presiona manualmente el botón para comenzar la grabación (Entrevistador).
Iniciar grabación	Se graba la respuesta del adulto mayor (Aplicación).
Detener grabación	Se presiona manualmente el botón para detener la grabación (Entrevistador).
Finalizar grabación	Se detiene la grabación de la respuesta del adulto mayor (Aplicación).
Generar archivo de audio	Se genera un archivo de audio con la respuesta grabada (Aplicación).
Seleccionar actividad	Se selecciona la actividad que se va a desarrollar. Son tres actividades: 1) “Repetir frases”, 2) “Fluidez verbal semántica” y 3) “Robo de galletas” (Entrevistador).
Reproducir instrucción	Se reproduce la instrucción de la actividad (Aplicación).
Reproducir una frase	Se reproduce cada una de las 10 frases de la actividad “Repetir frases” (Aplicación).
Mostrar imagen	Se muestra la imagen a describir en la actividad “Robo de galletas” (Aplicación).
Reproducir instrucciones de apoyo	Se reproducen instrucciones para incentivar a que el adulto mayor continúe hablando en el caso de las actividades “fluidez verbal semántica” y “Robo de galletas” (Aplicación).
Reproducir instrucciones adicionales	Se reproducen instrucciones para recordarle al adulto mayor en que consiste la actividad en el caso de “Repetir frases” (Aplicación).
Enviar notificación de terminación	Se le indica al adulto mayor a través de la reproducción de una sentencia que ha terminado la actividad (Aplicación).
<b>Bifurcación</b>	<b>Descripción</b>
¿Faltan actividades?	Si faltan actividades por desarrollar se regresa a la actividad <i>Orientar Actividad</i> y si NO faltan actividades termina el proceso
¿Actividad Repetir frases?	Si es la actividad de “Repetir frases” se espera 4 segundos y se dirige a la actividad <i>Finalizar grabación</i> y si NO continúa el flujo de otra actividad.
¿Han transcurrido 60 segundos?	Si han transcurrido 60 segundos se completa la actividad y se dirige a la actividad <i>Finalizar grabación</i> y si NO se continúa grabando.
¿Existe un espacio de silencio de 15 segundos?	Si el sistema identifica un espacio de 15 segundos de silencio revisa que sea la primera vez y si NO continúa grabando.
¿Es la primera ocurrencia de 15 segundos de silencio?	Si es la primera ocurrencia se dirige a la actividad <i>Reproducir instrucciones de apoyo</i> y si NO da la actividad como completada y se dirige a <i>Finalizar grabación</i> .
¿Se ejecuta la actividad Repetir frases?	Si se está desarrollando la actividad de “Repetir frases” verificar número de ejecución y si NO continúa el flujo de otra actividad.
¿Eres décima ejecución?	Si es la décima ejecución dirigirse a la actividad <i>Enviar notificación de terminación</i> y si NO verificar número de ejecución.

La tecnología que se empleó para la captura de datos es una aplicación de escritorio desarrollada en Python. Las principales bibliotecas que se utilizaron fueron: *pyaudio* y *wave* para todo lo relacionado con ficheros de audios, *PIL* para el trabajo con imágenes, *threading* para ejecutar subprocessos múltiples, y *tkinter* para la construcción de las interfaces gráficas de usuario. En la Figura 6 se muestran algunas de las principales pantallas de la aplicación desarrollada.



**Figura 6.** Aplicación desarrollada para la captura de los datos.

En la figura 6(a) se muestra la interfaz gráfica para indicar cuando inicia o termina el estudio con el adulto mayor. Además, se puede seleccionar el procedimiento que se va a aplicar (manual o autónomo). En la 6(b) la interfaz gráfica donde se selecciona la tarea cognitiva que se va a aplicar en el procedimiento autónomo. En la 6(c) la interfaz gráfica donde se graban las respuestas de los adultos mayores, por cada actividad, en el procedimiento manual. Finalmente, en la 6(d) la imagen que se describe en la actividad de “Habla espontánea”.

Los archivos de audios de cada participante se almacenan en carpetas diferentes. Por cada actividad se genera un fichero de audio en formato .wav que tendrá una duración máxima de 60 segundos para las actividades de “Fluidez verbal semántica” y descripción de la imagen llamada “Robo de galletas”, mientras

que en el caso de la actividad “Repetir frases” la duración será de 4 segundos por cada oración, que es el tiempo máximo estimado como necesario para repetir las oraciones.

### 3.3 Estudio piloto y resultados

Se realizó un estudio piloto con el propósito de probar la herramienta desarrollada y analizar el comportamiento de la población objetivo. Este estudio se llevó a cabo en la consulta del médico geriatra y participaron 13 adultos mayores, 6 utilizaron el procedimiento manual y 7 el procedimiento autónomo.

La realización del piloto resultó fundamental para identificar potenciales contratiempos en el estudio que eran difíciles de prever. Algunas de las observaciones que generaron cambios en el sistema o el protocolo fueron realizadas directamente por la geriatra mientras que otras emergieron de la observación del comportamiento de los participantes.

De este estudio se obtuvieron los siguientes resultados:

- Se identificaron oportunidades de mejora en la aplicación para mejorar su funcionalidad. Por ejemplo, se decidió incorporar un cronómetro y manejar espacios de silencio.
- Cambios en algunas instrucciones de las actividades que se aplican. Por ejemplo, inicialmente se preguntaba “Mencione las frutas que usted conozca”, dando la oportunidad que el adulto mayor pierda tiempo de la prueba preguntando “El aguacate es una fruta”, actualmente se pregunta “Mencione las frutas que usted sepa”.
- En el caso de que el adulto mayor no entienda la actividad, o se quede callado al inicio, ejemplificar para lograr su entendimiento. Por ejemplo, “En la imagen se ve un niño encima de un banco queriendo agarrar las galletas...”.
- Mencionarle al adulto mayor un conjunto de instrucciones antes de comenzar a realizar las actividades, por ejemplo, en el caso de la tarea de “Fluidez verbal semántica” se le especifica que “la prueba involucra tiempo, mientras más rápido usted la haga es mejor”. En el caso de la tarea de “Repetir frases”, se le menciona que “es importante que ponga atención porque no se le va a

repetir la oración”. Un elemento relevante para resaltar en esta población es que las instrucciones deben de ser cortas y precisas, para evitar distracciones.

- Se incorporó lenguaje corporal, por ejemplo, en el caso del procedimiento autónomo, explicarle al adulto mayor señalando para la tableta que va a interactuar con ese dispositivo. Tocarles el hombro para llamar su atención.
- Uso de reforzadores positivos, por ejemplo, “lo está haciendo muy bien”.
- Dirigirse al adulto mayor modulando la voz con tonos graves.
- Reconocer cuándo suspender la actividad porque el adulto mayor no quiere colaborar, o se encuentra ansioso.

Como se puede observar la mayoría de las sugerencias corresponden a cambios sutiles en la forma de dirigirse a los participantes. El piloto también permitió a la asistente familiarizarse con la tarea y el perfil de los participantes.

Los archivos de audios resultantes fueron analizados teniendo en cuenta el ruido externo, intervenciones del familiar, dificultad del adulto mayor en el desarrollo de las actividades, entre otras, con la finalidad de asegurar la calidad de estos para un mayor análisis y exploración. Finalizado el estudio piloto se iniciaron las grabaciones para la confección del corpus de audio, teniendo en cuenta los elementos aprendidos y modificados del mismo.

### **3.4 Descripción del corpus de audio propuesto**

Se creó un conjunto de datos con la participación de 32 adultos mayores, 12 sujetos con demencia y 20 sujetos sanos. La descripción demográfica de los mismos es mostrada en la Tabla 8. En la misma se evidencia que los grupos no están demográficamente balanceados ni por *Edad* ni por *Sexo*, lo cual constituye una limitación para el conjunto de datos, ya que los años de educación influyen en pruebas diagnósticas como MMSE (Pinto et al., 2019) y hay más prevalencia de demencia en mujeres (Beam et al., 2018).

**Tabla 8.** Descripción demográfica de los sujetos que participaron en el estudio.

	<b>Sujetos con Demencia (sd)</b>	<b>Sujetos Sano (sd)</b>	<b>P value</b>
<b>Edad (Años)</b>	79.7 (7.2)	76.0 (7.8)	< 0.05
<b>Educación (Años)</b>	8.7 (4.8)	8.9 (5.1)	> 0.05
<b>Sexo (F/M)</b>	10/2	12/8	> 0.05

*El valor de p se calculó para las variables Edad y Educación mediante la prueba t-Student, para la variable Sexo mediante la prueba chi-cuadrado.*

El conjunto de datos propuesto está formado por 378 muestras de audios, 140 pertenecientes a sujetos con demencia y 238 a sujetos sanos. Todas las muestras fueron recolectadas utilizando el procedimiento autónomo. En la Tabla 9 se muestra la distribución de ficheros de audios y duración promedio de estos, por las tres tareas desarrolladas en el estudio.

**Tabla 9.** Características de los ficheros de audios del corpus propuesto.

<b>Tipo de tarea</b>	<b>Cantidad de archivos de audios</b>		<b>Promedio en segundos (sd)</b>	
	Demencia	Sano	Demencia	Sano
<b>“Robo de galletas”</b>	12	19	43.14(12.0)	48.24(17.1)
<b>Fluidez verbal semántica</b>	12	20	45.42(12.4)	53.23(8.5)
<b>Repetir frases</b>	116(12*10)	199(20*10)	2.85(0.9)	2.69(0.8)

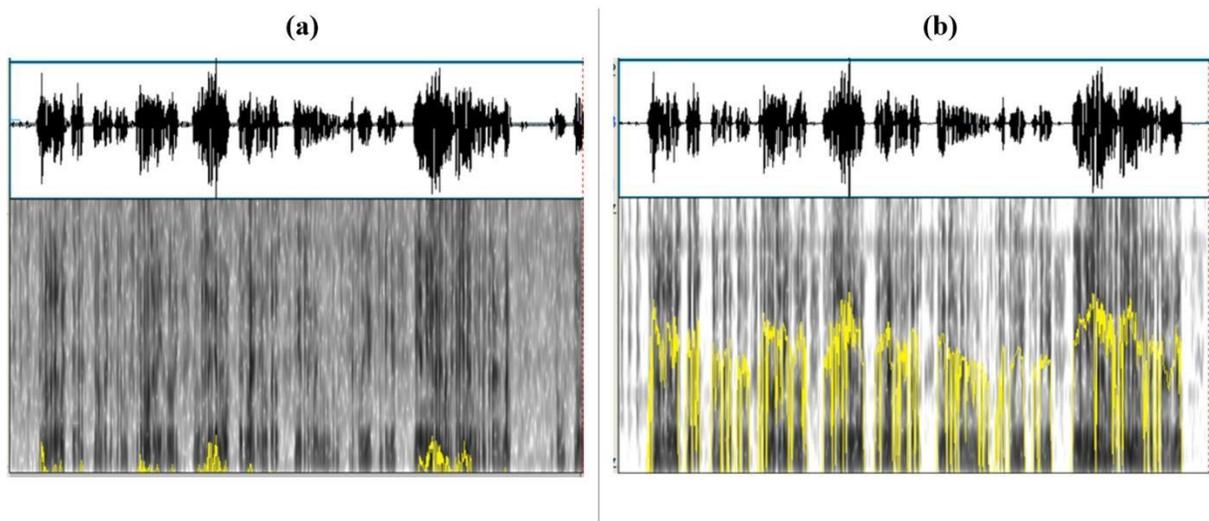
Los archivos de audios faltantes en las tareas de “Repetir frases” (grupo demencia 4 archivos y grupo sano 1 archivos) y “Robo de galletas” (1 archivos del grupo sano) está causado porque los adultos mayores no desarrollaron las actividades en el tiempo establecido.

Por otro lado, se evidencia un comportamiento esperado en las tareas de “Robo de galletas” y “Fluidez verbal semántica” en cuanto a la duración de los audios, porque los adultos mayores sanos tienden a hablar todo el tiempo que dura la actividad, y en la tarea de “Repetir frases” los adultos mayores con demencia tienden a demorar más a la hora de repetir las frases. Las grabaciones se mejoraron acústicamente con diferentes técnicas de limpieza de datos, que se describen en la siguiente sección.

### 3.4.1 Preprocesamiento de datos

En el análisis de esta tesis se incluyeron 378 archivos de audios. Para garantizar la calidad de los datos recabados se aplicaron las siguientes técnicas de limpieza.

1. Se comprobó que la intensidad del sonido de cada muestra se encontrara entre los 30 y 80 decibelios (dB)
2. Se eliminaron de forma manual espacios de silencio al inicio y al final de cada archivo de audio.
3. Se normaliza cada señal de audio escalando su amplitud al máximo nivel disponible en los ficheros .wav, para controlar la variación causada por las condiciones de grabación, como la ubicación del micrófono.
4. Se filtran las grabaciones en *paso alto* a 80 Hz para eliminar el ruido de baja frecuencia.
5. Se elimina el ruido de fondo mediante el algoritmo de Sustracción Espectral (SE) (Fukane and Sahare, 2011).



**Figura 7.** Preprocesamiento de datos. (a) Señal inicial y (b) Señal filtrada.

En la Figura 7 se muestra un fragmento de la descripción de la imagen llamada “Robo de galletas”, de un adulto mayor que participó en el estudio. En la misma podemos observar la señal inicial (ver Figura 7(a)) y



teléfono, aplausos, risas, murmullo, intervención familiar, ruido externo del ambiente, entre otros. También, se encuentra la familia de fáticos, que indican cuando el adulto mayor aún necesita la ayuda del asistente, por ejemplo, cuando mencionan “*mhm*”. En el Anexo B se muestra el manual que se creó para este propósito, en el que se describen detalladamente las etiquetas que se utilizaron. La Figura 8 muestra un ejemplo de etiquetado de la tarea de fluidez verbal semántica de un adulto mayor que participa en el estudio.

El proceso de etiquetado sólo se aplicó sobre las actividades de “Fluidez verbal semántica” y “Robo de galletas”. En el primer caso, se etiquetó toda la señal de audio para identificar elementos semánticos interesantes de analizar en estudios posteriores y en el segundo solo cuando se presenciaron elementos no analizables para indicarle a investigadores que utilicen nuestros datos, donde se encuentran estos fragmentos.

### 3.4.3 Limitaciones en la calidad del corpus propuesto

La calidad de los ficheros de audios recolectados influye en el diagnóstico certero de la demencia. A pesar de que, en nuestro estudio a través de los protocolos creados tratamos de tener en cuenta elementos que pudieran atentar con la calidad de nuestras muestras, se evidenciaron algunos que eran difíciles de prever. En la siguiente Tabla 10 se muestra un reporte de estos. Con el monitoreo constante de los datos, estas limitaciones fueron identificadas de forma oportuna y tomadas en cuenta para siguientes grabaciones. Además, se etiquetaron estos ficheros utilizando el procedimiento descrito en la sección 3.4.2 Etiquetado de registros de audio, para que sea tomado en cuenta por otros grupos de investigación.

**Tabla 10.** Limitaciones identificadas en los ficheros de audios.

	<b>Nombre del fichero</b>	<b>Señalamiento</b>
Adulto mayor 1	Adulto1_fluidez	Sonido teléfono de la doctora.
Adulto mayor 3	Adulto3_robogalletas	Sonido teléfono de la doctora.
Adulto mayor 8	Adulto8_fluidez	Intervención familiar
	Adulto8_robogalletas	Sonido celular del hijo (respondió la llamada y habló), caída de objetos, ruido de bolsa de nylon.
Adulto mayor 27	Adulto27_fluidez	Intervención familiar
	Adulto27_oracion1	Intervención familiar

En este capítulo se describe el protocolo elaborado y los resultados obtenidos en el estudio piloto, para analizar su comportamiento en la población objetivo. Se describe el conjunto de datos creado en cuanto a características demográficas y a muestras de audios obtenidas por cada actividad. También se detallan las técnicas de preprocesamiento de datos que se utilizan para mejorar la calidad de los archivos, el proceso de etiquetado que se realiza y las limitaciones con respecto a la calidad de estos.

Además, se presentó una herramienta para capturar los registros de voz de los adultos mayores en el desarrollo de las actividades cognitivas. Está formado por dos procedimientos fundamentales, el procedimiento manual y el autónomo. El sistema incluye tres módulos que corresponden a cada una de las actividades: “Repetir frases”, “Prueba de fluidez verbal semántica” y “Robo de galletas”. Es una aplicación de escritorio desarrollada en Python.

## Capítulo 4. Metodología para el análisis de datos

En este capítulo se presenta la metodología utilizada para el análisis de los datos recabados en el corpus de voz de adultos mayores descrito en el capítulo anterior.

### 4.1 Descripción general

El análisis de los datos juega un rol fundamental en las investigaciones relacionadas a la predicción del diagnóstico de demencia. En nuestro estudio, este comienza a partir de los archivos de audios resultantes del procedimiento de captura. A cada archivo de audio se le aplican técnicas de limpieza de datos para mejorar la calidad de estos (descrito es sección 3.4.1 Preprocesamiento de datos), luego se extraen los rasgos acústicos que los describen y se analizan a través de técnicas de aprendizaje automático, como se muestra en la Figura 9.

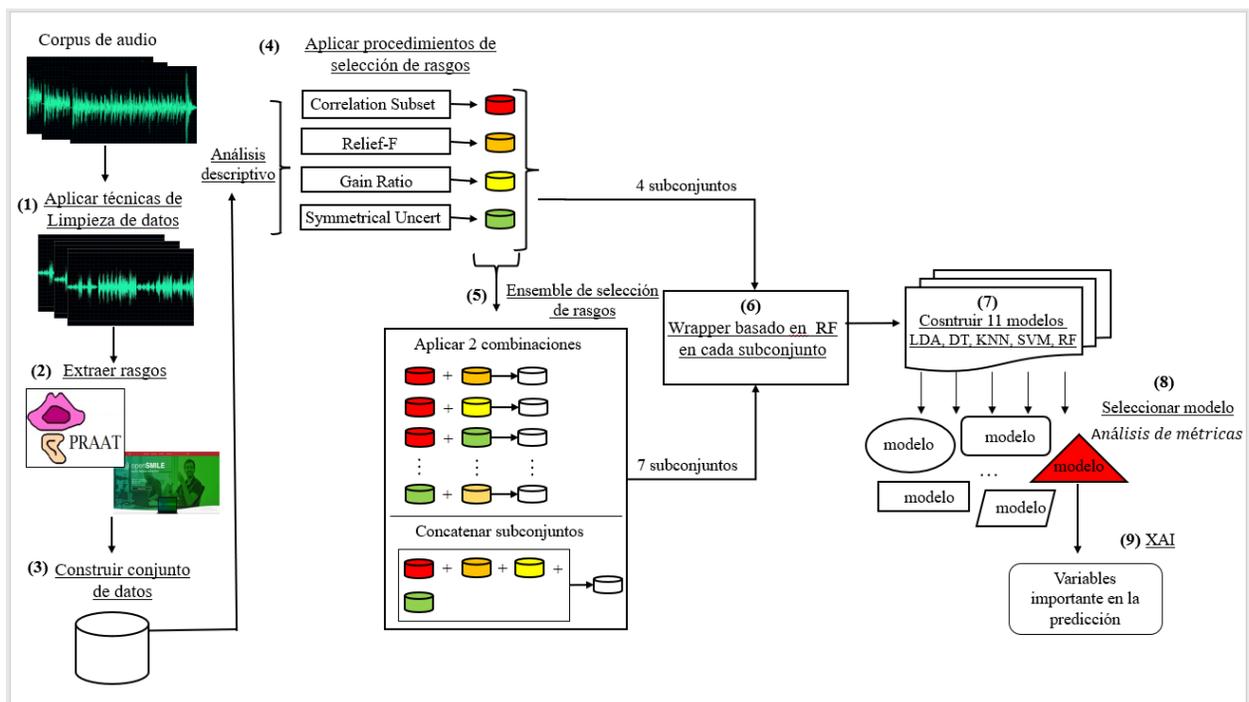


Figura 9. Metodología utilizada empleada para el análisis de los datos.

Esta metodología se aplica a cada una de las tareas desarrolladas: “Repetir frases”, “Fluidez verbal semántica” y “Robo de galletas”. Cada tarea se analiza de forma independiente. En las siguientes secciones se explica detalladamente cada paso de la metodología.

## 4.2 Rasgos acústicos

La extracción de rasgos es un aspecto fundamental de cualquier proceso de análisis de la voz, define qué características serán tenidas en cuenta para describir los ficheros de audios. Se realizó una revisión de la literatura para identificar el estado del arte de los rasgos acústicos que usualmente son utilizados en este tipo de estudio. Por un lado, se identificó un conjunto de rasgos bases, donde las características a tener en cuenta están relacionadas con el proceso de percepción humana del habla, por ejemplo, la frecuencia fundamental, los formantes, frecuencia de Mel, energía, intensidad, entre otras. Por otro lado, se reconoció otro grupo que estaba más enfocado a capturar las características que mejor describieran las actividades que se estaban analizando. Por ejemplo, en el caso de la actividad de “Fluidez verbal semántica” los *segmentos de silencio* no son relevantes para interpretarla, sin embargo, el rasgo *Distancias (tiempo) de la 5-10-15 posición de las palabras detectadas* codifica información relevante al análisis intrínseco de la misma.

En este estudio (Figura 9, paso (2)) se utilizan dos herramientas para la extracción de rasgos, el openSmile y el Praat. Estas herramientas son de libre acceso y se usan para realizar análisis acústico del habla. Este procedimiento se realiza de forma automática, utilizando Python y las librerías praat-parselmouth y Opensmile 2.4.1.

Para el propósito de esta tesis, consideramos un conjunto de rasgos acústicos base, que se calcularon para cada archivo resultante de cada una de las 3 actividades. Además, se generó un conjunto adicional por cada actividad individual. El número de rasgos resultantes es elevado, por lo tanto, en este documento se detallan en la Tabla 11 sólo los más característicos. Adicionalmente, para cada grupo se calcula un conjunto de rasgos apoyados en estadísticas básicas como promedio, varianza, desviación estándar, máximo y mínimo. En total utilizando los dos programas, por cada archivo de audio referente a las actividades de “Repetir frases” se extrajeron 7450 rasgos, “Fluidez verbal semántica” 7476 rasgos y “Robo de galletas” 7502 rasgos.

**Tabla 11.** Rasgos acústicos base extraídos en las tres tareas y rasgos acústicos extraídos por tipo de actividad.

	Repetir Frases	Fluidez verbal semántica	Robo de galletas
<b>Rasgos acústicos base</b>	Rasgos de calidad vocal: Regularidad de la frecuencia (Jitter) Irregularidad en la amplitud (Shimmer) Relación Ruido-Armónicos (NHR)		
	Rasgos de dominio Spectral: Formantes F1, F2, F3 Mel Frequency Cepstral Coefficients (MFCC)		
	Rasgos prosódicos: Percepción de la frecuencia de un sonido (Pitch/F0)		
<b>Rasgos acústicos por tipo de actividad.</b>	Duración relativa	<ul style="list-style-type: none"> <li>- Distancias (tiempo) de la 5-10-15 posición de las palabras detectadas.</li> <li>- Número de respuestas por cuartil de tiempo.</li> <li>- Número de silencios.</li> <li>- Tiempo total de silencio.</li> <li>-Tiempo promedio entre silencios.</li> <li>- Longitud promedio de las pausas de silencio por cuartil de tiempo.</li> </ul>	<ul style="list-style-type: none"> <li>-Longitud del segmento de voz (seg).</li> <li>- Duración del segmento de silencio (seg).</li> <li>- Longitud del segmento periódico (seg).</li> <li>- Longitud del segmento aperiódico (seg).</li> <li>- Cantidad de sílabas.</li> <li>- Velocidad del habla.</li> <li>- Duración media de la sílaba (ASD).</li> <li>- Pausas por sílaba.</li> <li>- Tasa de pausas.</li> </ul>

### 4.3 Construcción del conjunto de datos

Las características de los conjuntos de datos construidos se detallan en la siguiente Tabla 12.

**Tabla 12.** Características de los conjuntos de datos.

Conjunto de datos	Cantidad de instancias	Cantidad de rasgos por instancia
dataset_repetirFrases	316	7450
dataset_fuidez	32	7476
dataset_roboGalletas	31	7502

En cada conjunto de datos se realiza un estudio descriptivo para analizar el comportamiento del conjunto de rasgos. El mismo se detalla a continuación:

1. **Análisis de Kurtosis para eliminar valores constantes.** La curtosis (Gogtay & Thatte, 2017) es una métrica que se utiliza para identificar los valores concentrados alrededor de la media. En este estudio se eliminaron los rasgos que mostraran una curtosis por encima de 8 ( $k > 8$ ).

2. **Análisis de correlación para eliminar valores iguales o muy correlacionados** (Gogtay and Thatte, 2017). Cómo se están utilizando dos herramientas para calcular los rasgos acústicos, se podría procesar rasgos que estuvieran codificando la misma información con diferentes nombres o información muy cercana. Por lo tanto, se decidió eliminar los que presentan una correlación mayor a 0.95.
3. **Análisis visual utilizando la técnica de Vecinos estocásticos distribuidos (tSNE)** (Ozgode Yigin and Saygili, 2022). t-SNE es un algoritmo de reducción de dimensionalidad no lineal, que se utiliza para explorar datos de alta dimensión. Mapea datos multidimensionales a dos o más dimensiones adecuadas para la observación humana. Específicamente, se utiliza en este estudio para analizar visualmente el comportamiento de las variables con respecto a la clase.
4. Análisis de los rasgos acústicos, teniendo en cuenta la importancia de estos en función de diferentes estadísticos.
  - GainRatio (Karegowda et al., 2010): Es la razón de ganancia de información, con esta métrica seleccionamos los rasgos que son capaces de identificar pequeños cambios en archivos de audios distintos.
  - ChiSquared (Patel and Upadhyay, 2012): Examina las diferencias entre los rasgos de la misma clase. Entre más dependencia entre el rasgo y la clase mejor valor de ChiSquared.
  - ReliefF (Urbanowicz et al., 2018): Asume que un rasgo es fuertemente relevante si este permite distinguir fácilmente entre dos instancias de clases diferentes.

Para llegar a un consenso de cuáles pueden ser las mejores variables identificadas a través de los 3 estadísticos antes mencionados, se realiza un análisis basado en decisión multicriterio mediante la técnica de fusión por ranking (Hashemi Amin et al., 2022). Las 10 mejores variables seleccionadas según el ranking promedio obtenido para este problema específico se analizan visualmente a través de un diagrama de cajas.

### 4.3.1 Balanceo de clases

El desbalance de datos es un problema común en el análisis de datos que afecta la calidad de un modelo y su capacidad de predicción (Cui et al., 2019). En los conjuntos de datos que se analizan en esta tesis, la clase *Demencia* tiene una representación menor que la clase *Sano* (clase mayoritaria). Para mejorar los resultados de la clasificación sin alterar las características del proceso de los algoritmos de aprendizaje que se utilizan, se aplican técnicas de balanceo de clases como método de preprocesamiento de datos en el proceso de entrenamiento. Se generan muestras sintéticas de la clase minoritaria utilizando el algoritmo SMOTE (Ganganwar, 2012).

## 4.4 Selección de rasgos relevantes

El procedimiento para la selección de rasgos relevantes se realiza en dos etapas, en una primera etapa se seleccionan rasgos utilizando algoritmos filtros (Figura 9, paso (4)). Los algoritmos filtros se basan en las características de los rasgos para seleccionar el subconjunto que contenga aquellos que son relevantes (la relevancia se establece al aplicar diferentes criterios) para la tarea de clasificación (Billsus et al., 1998). En la segunda etapa se aplica un algoritmo wrapper basado en el desempeño del clasificador Random Forest (Figura 9, paso (6)). Los algoritmos wrapper utilizan un algoritmo de búsqueda para explorar a través del espacio de características posibles y evaluar cada subconjunto corriendo un modelo en el subconjunto (Lerman et al., 2003).

Algoritmos filtros empleados:

- Correlation subset (Karegowda et al., 2010): El algoritmo selecciona los rasgos altamente correlacionados con la clase, pero con poca o ninguna correlación con otros atributos.
- ReliefF (Urbanowicz et al., 2018): Descrito en la sección 4.3 Construcción del conjunto de datos.
- Gain Ratio (Karegowda et al., 2010): Descrito en la sección 4.3 Construcción del conjunto de datos.
- Symmetrical Uncert (Dimić et al., 2017): Es un algoritmo basado en ranking, que evalúa el valor de un rasgo midiendo la incertidumbre simétrica con respecto a la clase.

Los algoritmos ReliefF, Gain Ratio, Symmetrical Uncert son basados en ranking, para el proceso de selección, se tomaron en cuenta la misma cantidad de rasgos que seleccionó el algoritmo Correlation subset.

Después de obtener estos 4 subconjuntos de datos, se aplica ensemble de selección de rasgos (Figura 9, paso (5)). El ensemble consiste por un lado en construir las posibles combinaciones con 2 subconjuntos y, por otro lado, concatenarlos todos. A los 7 subconjuntos formados se le eliminan los datos duplicados. Finalmente, a los 11 subconjuntos resultantes mostrados en la Tabla 13 se le aplica el algoritmo wrapper basado en Random Forest.

**Tabla 13.** Conjuntos resultantes del proceso de ensemble.

<b>Subconjuntos iniciales</b>	CFS
	Relief
	SymmetricalUncert
	GainRatio
<b>Combinaciones con 2 subconjuntos</b>	CFS_Relief
	CFS_SymmetricalUncert
	CFS_GainRatio
	Relief_SymmetricalUncert
	Relief_GainRatio
	SymmetricalUncert_GainRatio
<b>subconjuntos concatenados</b>	All

## 4.5 Desarrollo de modelos predictivos

Los modelos de clasificación binaria se crean para distinguir entre el habla de adultos mayores con demencia y sanos. En este estudio, se utilizan algunos de los más referenciados en la literatura y los que mejores desempeños han mostrado. Se construyen 11 modelos por cada algoritmo de clasificación (Figura 9, paso (7)). Los algoritmos de clasificación que se utilizan son LDA (Haider et al., 2019), DT (Guo et al., 2019), KNN (Nasrolahzadeh et al., 2018), SVM (Syed et al., 2021) y RF (Balagopalan et al., 2021). A cada algoritmo se le optimiza un conjunto de parámetros para encontrar la mejor selección de valores y generar el mejor modelo posible (descritos en la Tabla 14).

Se utiliza el procedimiento de validación cruzada dejando uno fuera (Leave-one-subject-out, LOSO), porque acorde a la cantidad de datos que se tiene (en el caso del conjunto derivado de la actividad “Repetir frases” 316 muestras y en el caso de las actividades “Fluidez verbal semántica” y “Robo de Galletas” 32 y

31 respectivamente) es la técnica de validación más adecuada para evaluar el poder predictivo de los clasificadores. El mejor modelo se selecciona (Figura 9, paso (8)) a través de la métrica de precisión. Los métodos de clasificación se implementan en Python, con la librería *sklearn*.

**Tabla 14.** Parámetros para optimizar en cada algoritmo de clasificación.

Algoritmos	Criterios	Parámetro que se optimizan
<b>DT</b>	Indica cuál es la función que se utilizará para dividir los datos	criterion = ['gini', 'entropy']
	Máxima profundidad a la cual debe de llegar el árbol.	max_depth = [2,4,6,8,10,12,14,16,18,20]
<b>KNN</b>	Número de vecinos cercanos	n_neighbors = [1,3,5,7,11,13]
<b>SVM</b>	Coeficiente de penalización	C = [0.01, 1, 10, 0.001, 0.01, 0.1]
	Función matemática utilizada en la transformación	kernel = ['linear', 'rbf']
<b>RF</b>	Número de árboles que va a tener el bosque aleatorio.	n_estimators = [50,100, 200,300,500,1000]
	Indica cuál es la función que se utilizará para dividir los datos	criterion = ['gini','entropy']
	Profundidad máxima del árbol	max_depth = [5, 10,15,20, 25,30]

## 4.6 Interpretación de los resultados

Una parte integral de la adopción y aceptación de los sistemas de IA en entornos del mundo real es la capacidad de las personas para dar sentido y evaluar dichos sistemas, un área creciente de esfuerzos de desarrollo y diseño conocida como XAI (IA explicable) (Barredo Arrieta et al., 2020). Los algoritmos de IA se basan en el propio conocimiento que contienen los datos para realizar predicciones. Los algoritmos de XAI ayudan a detectar si nuestros modelos están sesgados. Además, para personas ajenas al campo de la IA puede ser complejo comprender cómo funcionan los modelos de predicción y se preguntan por qué toman ciertas decisiones o cómo afecta cada variable en la decisión final. Con ese propósito, se aplica XAI a los modelos predictivos generados en esta investigación.

Los modelos predictivos en este estudio se explican a través de métodos de interpretación local, con el objetivo de justificar las predicciones individuales. Específicamente, se utiliza para identificar las variables importantes que están influyendo en el modelo, el Scoped Rules (Anchors) (Duell, 2021). Las explicaciones resultantes de este algoritmo se expresan como reglas SI-ENTONCES, que describen qué rasgos acústicos

causaron el resultado del modelo para una instancia específica. Su razón de ser consiste en encontrar reglas que se apliquen preferentemente a una gran parte del espacio de entrada del modelo.

Para identificar los rasgos acústicos relevante en el modelo, primero se seleccionan las reglas que presentan una cobertura (*proporción de la población que satisface los predicados*) mayor al 13% para la actividad de “Repetir frases” y mayor a 18% para “Fluidez verbal semántica” y “Robo de galletas”<sup>1</sup>, de las reglas seleccionadas se toman las variables que más se repiten. El siguiente paso consiste en hacer una interpretación de las variables respecto a características del habla. Algunos parámetros pueden tener una interpretación más simple, tales como número de silencios, y otros más abstractos, pero se buscará asociarlos a cambios del habla que han sido reportados como ocasionados por la demencia.

En este capítulo se describe la metodología utilizada para el análisis de los datos. En la misma se detallan las técnicas utilizadas para la extracción de rasgos acústicos, selección de características, balanceo de datos, generación de modelos, incluyendo los parámetros de optimización e interpretación de los modelos a través de técnicas de XAI. El siguiente capítulo presenta los resultados de aplicar la metodología descrita con el conjunto de datos descrito en el capítulo anterior.

---

<sup>1</sup> En la actividad “Repetir frases” la cobertura fluctuó entre 13% y 16%, en la de “Fluidez verbal semántica” entre 18% y 23% y en el “Robo de galletas” entre 18% y 26%.

## Capítulo 5. Resultados y discusión

---

En este capítulo se presentan los resultados del análisis descriptivo realizado a los conjuntos de datos obtenidas de cada una de las tres actividades desarrolladas, el desempeño de los algoritmos de clasificación sobre cada subconjunto resultante del proceso de selección de características y las variables representativas para explicar los modelos obtenidos, descritos en el capítulo anterior. Además, una discusión de los resultados.

### 5.1 Análisis descriptivo del conjunto de rasgos

#### 5.1.1 Análisis de curtosis y correlación para eliminar rasgos constantes y redundantes.

En las Tablas 15 y 16 se presentan los resultados de aplicar el análisis de curtosis y de correlación a los conjuntos de datos de las 3 actividades. En la actividad “Repetir frases” el análisis de curtosis elimina 1152 rasgos, en tanto que el de correlación elimina 1440 adicionales, para dejar un total de 4858 rasgos. En la de “Fluidez verbal semántica” el análisis de curtosis elimina 473 rasgos, entretanto el de correlación elimina 2428 adicionales, para dejar un total de 4575 rasgos. Finalmente, en la actividad “Robo de galletas” el análisis de curtosis elimina 1448 rasgos, mientras que el de correlación elimina 2074 adicionales, para dejar un total de 3980 rasgos.

**Tabla 15.** Análisis de curtosis ( $K > 8$ )

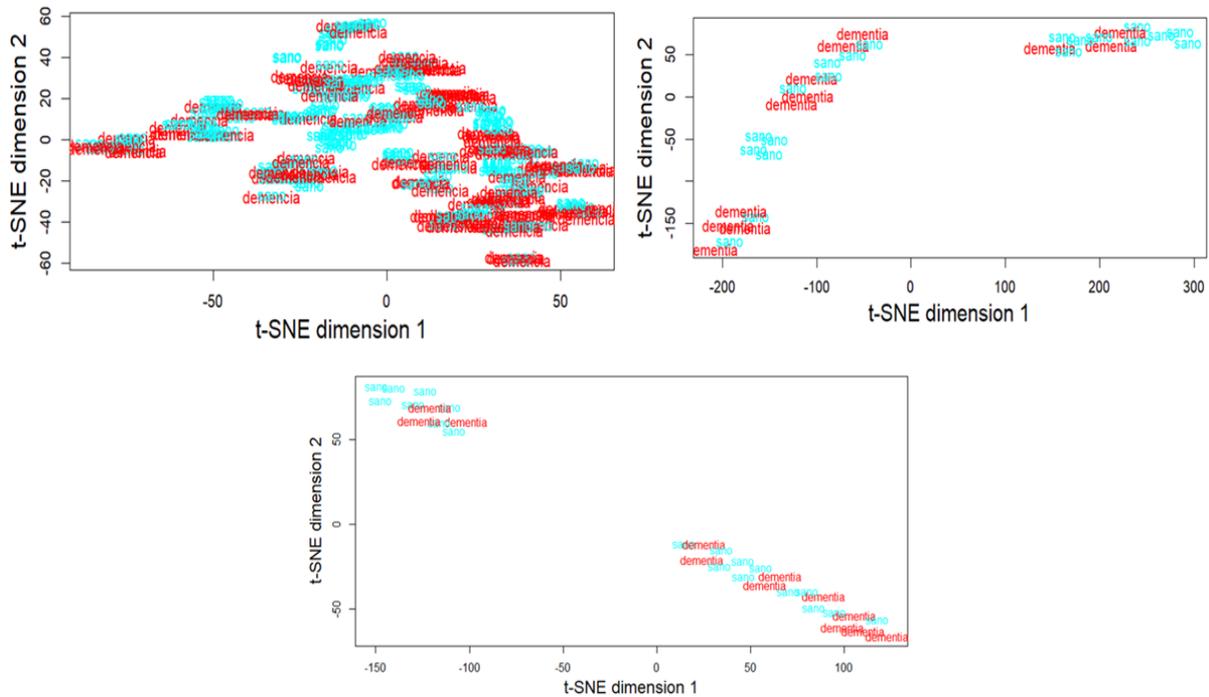
	Cantidad de rasgos iniciales	Cantidad de rasgos eliminados	Cantidad de rasgos finales
“Repetir frases”	7450	1152	6298
“Fluidez verbal semántica”	7476	473	7003
“Robo de galletas”	7502	1448	6054

**Tabla 16.** Análisis de correlación ( $\text{corr} > 0.95$ )

	Cantidad de rasgos iniciales	Cantidad de rasgos eliminados	Cantidad de rasgos finales
“Repetir frases”	6298	1440	4858
“Fluidez verbal semántica”	7003	2428	4575
“Robo de galletas”	6054	2074	3980

### 5.1.2 Análisis visual utilizando Vecinos estocásticos distribuidos (tSNE)

Para visualizar la separabilidad de las clases a partir de los sujetos, se observa en la Figura 10 los rasgos acústicos resultantes del análisis de curtosis y correlación (descrito en la sección anterior) utilizando la técnica tSNE, por cada una de las actividades.



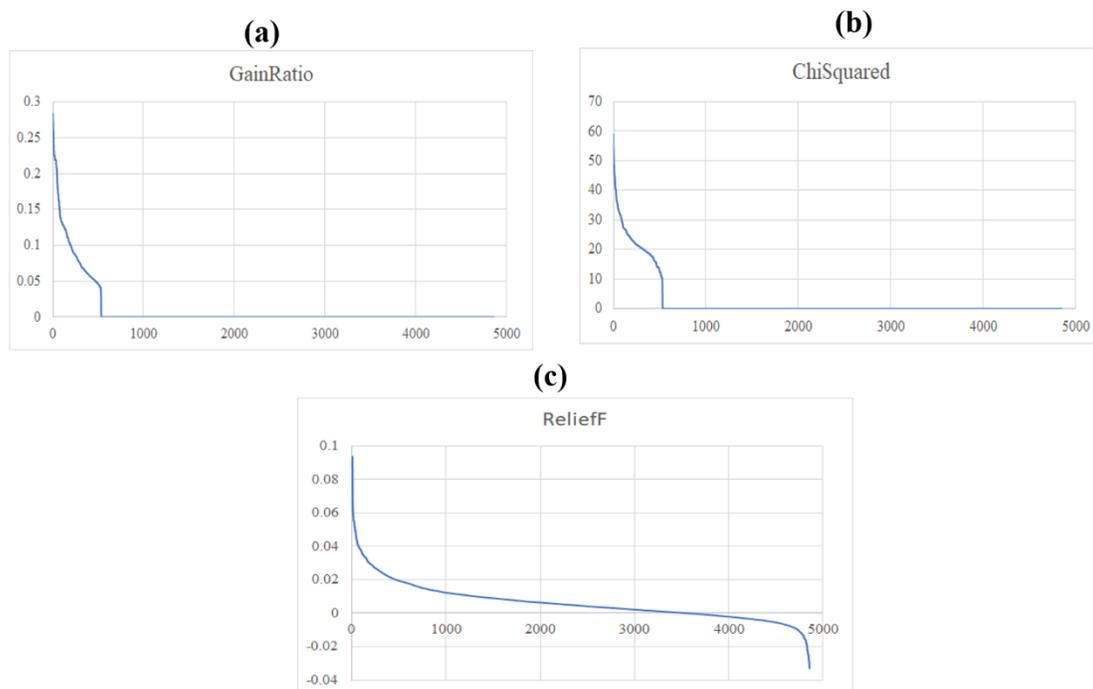
**Figura 10.** Visualización de rasgos utilizando la técnica t-SNE. (a) actividad “Repetir frases”, (b) actividad “Fluidez verbal semántica” y (c) actividad “Robo de galletas”.

La Figura 10 muestra que en las tres actividades las clases están mezcladas, los 4858, 4575 y 3980 rasgos correspondientes a las actividades “Repetir frases”, “Fluidez verbal semántica” y “Robo de galletas” respectivamente, por sí solos no son capaces de identificar cambios cuando codifican la información de diferentes clases.

### 5.1.3 Importancia de características analizando diferentes estadísticos

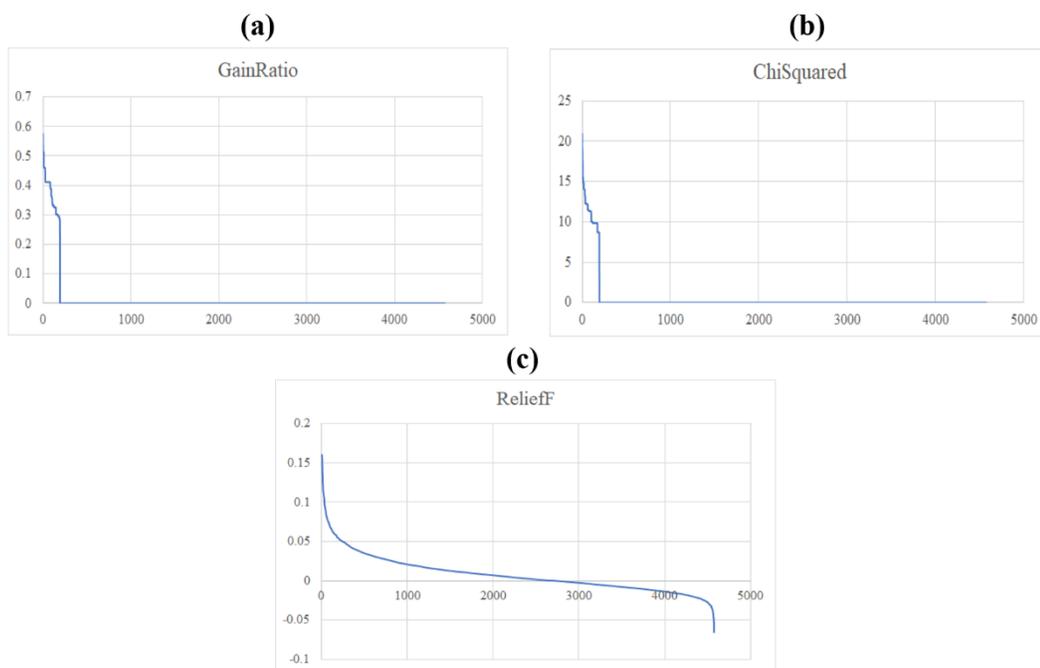
A partir del comportamiento obtenido de los rasgos en las gráficas tSNI (Figura 10), se realiza un análisis detallado de los mismos, para observar a través de diferentes métricas (descritas en sección 4.3 Construcción del conjunto de datos) los rasgos más significativos.

Las Figuras desde 11 hasta 13 muestran el comportamiento de los diferentes estadísticos por cada una de las actividades.

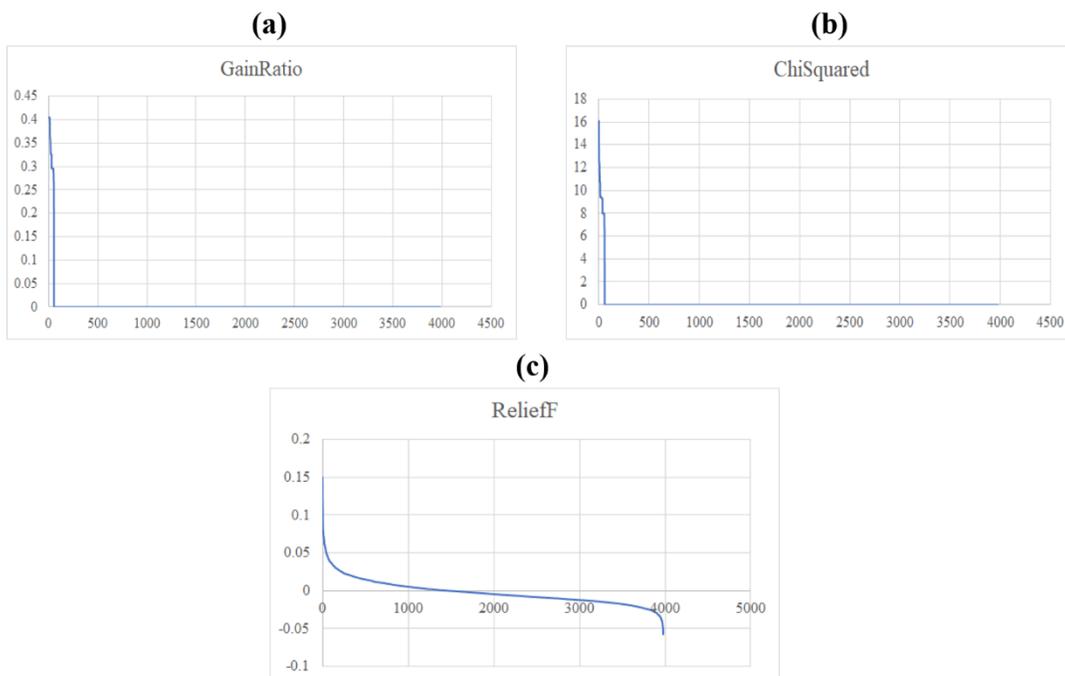


**Figura 11.** Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Repetir frases”.

En las Figuras desde la 11 hasta la 13 se identifican visualmente que un gran número de variables no están aportando información útil según las métricas analizadas. Las métricas GainRatio y Chi-cuadrado siguiendo sus criterios, solo identifican en la actividad “Repetir frases” 533 rasgos importantes, en “Fluidez verbal semántica” 195 y en la actividad “Robo de galletas” 57.



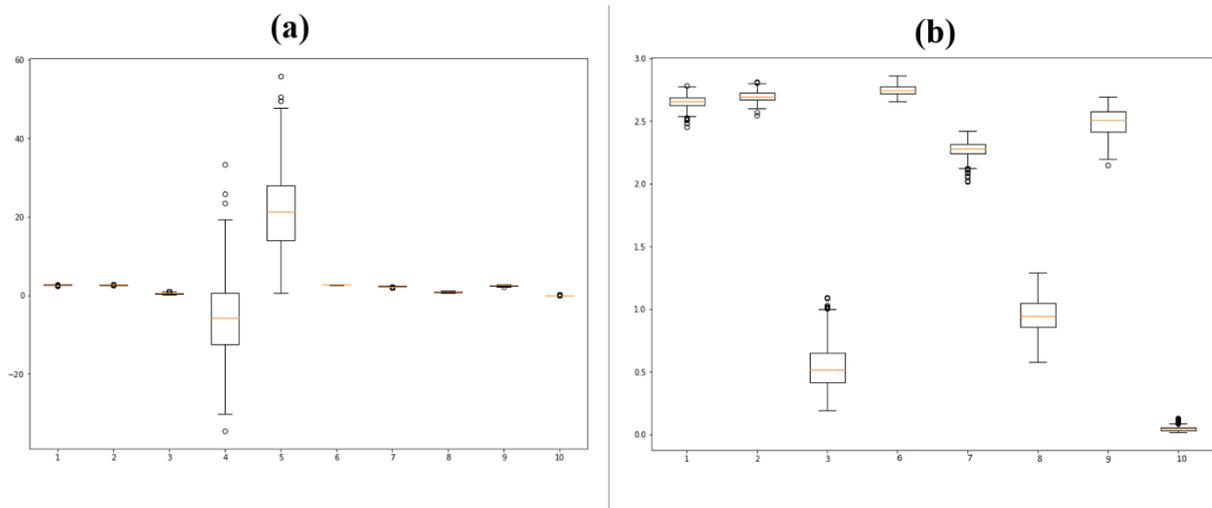
**Figura 12.** Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Fluidez verbal semántica”.



**Figura 13.** Visualización de los rasgos a través de la métrica GainRatio (a), Chi-cuadrado (b), ReliefF (c) en la actividad “Robo de galletas”.

La calidad de una variable no se puede definir por un único criterio, depende de qué criterios se utilizan para medirla, no hay un mejor criterio para confirmar quien es la mejor variable o quien es la variable más adecuada para un problema determinado. En este estudio se analizan 3 métricas *GainRatio*, *Chi-cuadrado* y *ReliefF*, cada una mide de forma diferente la importancia de las variables con respecto a la clase que se está modelando. Para llegar a un consenso de cuáles pueden ser las mejores variables, se aplica una técnica de ensemble de características. En nuestro estudio, se utilizó la técnica de fusión por ranking (Barredo Arrieta et al., 2020). El ranking promedio que se obtiene, lo que indica son las mejores variables según el promedio obtenido a partir de diferentes criterios para este problema específico.

Las Figuras desde 14 hasta la 16 muestran una distribución visual de las 10 primeras variables (más alto ranking promedio) identificadas en el proceso de fusión por ranking, en cada una de las actividades.



**Figura 14.** Distribución de los rasgos identificados a través de fusión por ranking en la actividad "Repetir frases".

En la Figura 14(a) de las 10 variables más importantes la 4 y la 5 presentan una alta variación en cuanto a su orden de importancia de acuerdo con las diferentes métricas, un mayor rango de valores reflejados por los valores atípicos, por consiguiente, una distribución diferente con respecto a las otras variables. Si se eliminan estos dos rasgos como se muestra en la Figura 14(b), no se evidencia una gran dispersión en los valores de las variables y tienden a estar normalmente distribuidos. Comportamiento similar se evidencia en las Figuras 15 y 16.

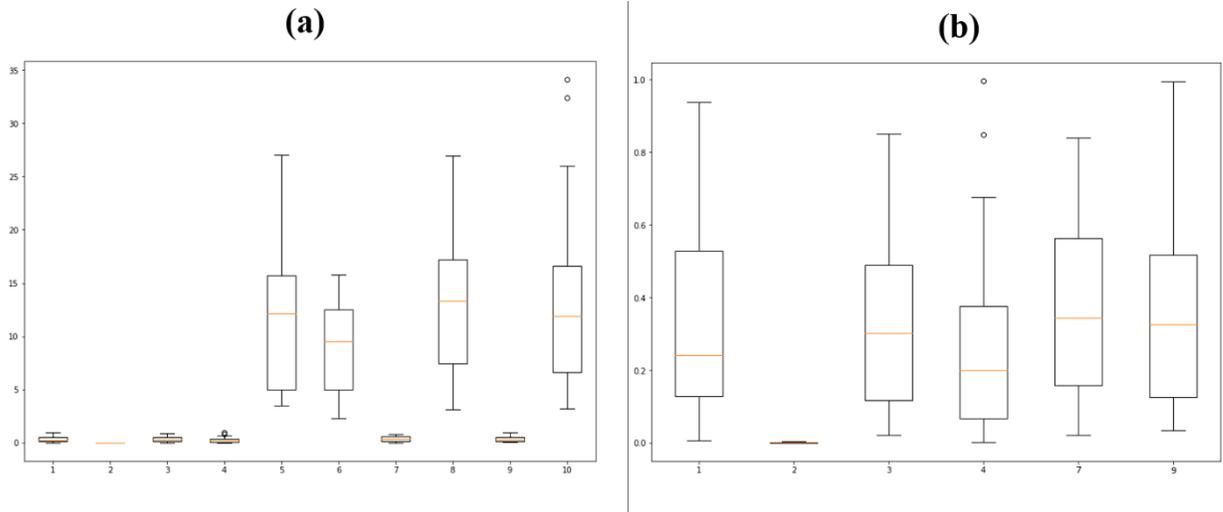


Figura 15. Distribución de los rasgos identificados a través de fusión por ranking en la actividad "Fluidez verbal"

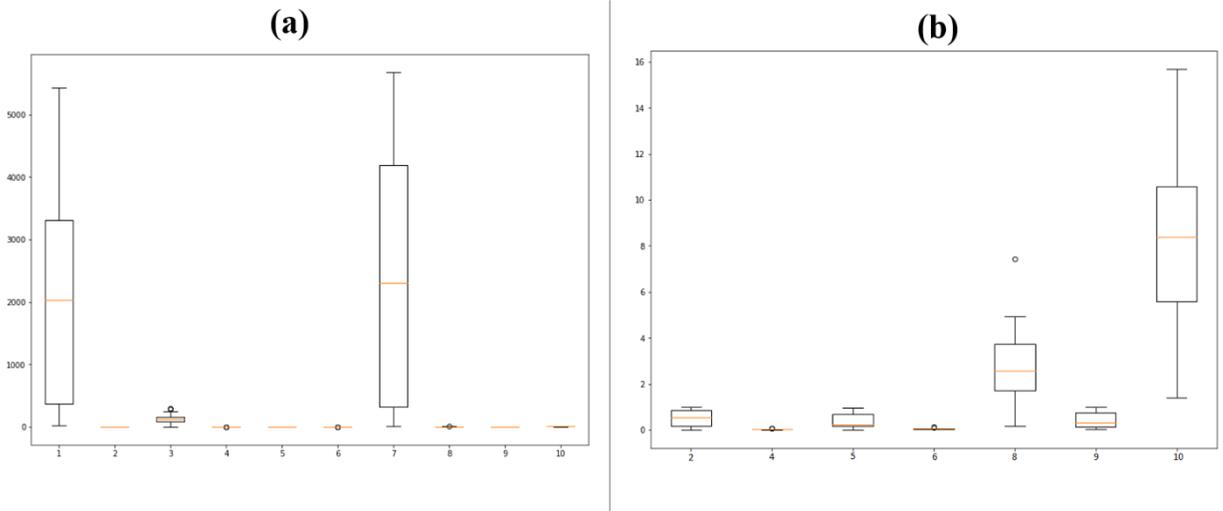


Figura 16. Distribución de los rasgos identificados a través de fusión por ranking en la actividad descripción de la imagen "Robo de galletas".

El comportamiento mostrado en el análisis de los datos no es concluyente con respecto a los resultados esperados, sólo a priori se puede identificar que para algunos algoritmos de clasificación el proceso de predicción será más complejo.

## 5.2 Modelos generados

### 5.2.1 Desempeño de los algoritmos de clasificación para datos no balanceados

En la Tabla 17 se muestra la precisión de la clasificación de los 11 subconjuntos originales (no balanceados) de la actividad “Repetir frases”. El clasificador que mejor desempeño mostró para la mayoría de los conjuntos (6 subconjuntos de 11) es el RF, con una precisión del 92% para el conjunto Relief, seguido de SVM, que también obtuvo una precisión del 92 % para el conjunto All. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

**Tabla 17.** Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Repetir frases” para LOSO.

	LDA	DT	KNN	SVM	RF
<b>CFS(34)</b>	0.825	0.848	0.737	0.873	<b>0.889</b>
<b>Relief(38)</b>	0.838	0.825	0.625	0.870	<b>0.921</b>
<b>SymmetricalUncert(34)</b>	0.790	<b>0.879</b>	0.613	0.854	0.873
<b>GainRatio(36)</b>	0.819	0.762	0.717	0.870	<b>0.876</b>
<b>CFS_Relief(59)</b>	0.816	0.790	0.657	<b>0.898</b>	0.895
<b>CFS_SymmetricalUncert(55)</b>	0.835	0.813	0.679	0.886	<b>0.892</b>
<b>CFS_GainRatio(51)</b>	0.825	0.822	0.676	0.876	<b>0.876</b>
<b>Relief_SymmetricalUncert(62)</b>	0.835	0.825	0.679	0.876	<b>0.914</b>
<b>Relief_GainRatio(57)</b>	0.838	0.778	0.673	<b>0.886</b>	0.882
<b>SymmetricalUncert_GainRatio(51)</b>	0.794	0.768	0.667	<b>0.876</b>	0.873
<b>All(88)</b>	0.806	0.794	0.683	<b>0.917</b>	0.911

En la Tabla 18 se muestra la precisión de la clasificación de los 11 subconjuntos originales (no balanceados) de la actividad “Fluidez verbal semántica”. El clasificador que mejor desempeño mostró para la mayoría de los conjuntos (10 subconjuntos de 11) es el RF. Los mejores desempeños se obtuvieron con los conjuntos *CFS*, *SymmetricalUncert* y *CFS\_GainRatio* con una precisión del 97% para cada uno. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

En la Tabla 19 se muestra la precisión de la clasificación de los 11 subconjuntos originales (no balanceados) de la actividad “Robo de galletas”. El clasificador que mejor desempeño mostró para la mayoría de los conjuntos (8 subconjuntos de 11) es el RF. No obstante, se puede observar que la mejor precisión la obtuvo el LDA en el conjunto *Relief\_GainRatio* con un 100%, seguido de *Relief* con un 94%. Por tanto, aunque el

RF proporciona mejor desempeño en la mayor parte de los subconjuntos, el LDA también exhibe un desempeño prometedor, siendo de hecho más estable en todos los subconjuntos que los otros clasificadores. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

**Tabla 18.** Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Fluidez verbal semántica” para LOSO.

	LDA	DT	KNN	SVM	RF
<b>CFS(19)</b>	0.781	0.688	0.781	0.813	<b>0.969</b>
<b>Relief(18)</b>	0.719	0.812	0.813	<b>0.906</b>	0.844
<b>SymmetricalUncert(21)</b>	0.844	0.688	0.813	0.875	<b>0.969</b>
<b>GainRatio(18)</b>	0.875	0.781	0.625	0.625	<b>0.938</b>
<b>CFS_Relief(36)</b>	0.781	0.781	0.813	0.875	<b>0.906</b>
<b>CFS_SymmetricalUncert(31)</b>	0.688	0.75	0.781	0.875	<b>0.906</b>
<b>CFS_GainRatio(32)</b>	0.719	0.781	0.781	0.906	<b>0.969</b>
<b>Relief_SymmetricalUncert(33)</b>	0.813	0.813	0.813	0.875	<b>0.938</b>
<b>Relief_GainRatio(33)</b>	0.625	0.813	0.813	0.906	<b>0.938</b>
<b>SymmetricalUncert_GainRatio(27)</b>	0.594	0.75	0.781	0.813	<b>0.906</b>
<b>All(43)</b>	0.719	0.781	0.781	0.781	<b>0.906</b>

**Tabla 19.** Desempeño de la clasificación de los conjuntos (no balanceados) resultantes de la actividad “Robo de galletas” para LOSO.

	LDA	DT	KNN	SVM	RF
<b>CFS(10)</b>	<b>0.839</b>	0.774	0.677	0.806	<b>0.839</b>
<b>Relief(14)</b>	<b>0.935</b>	0.516	0.839	0.774	0.871
<b>SymmetricalUncert(13)</b>	<b>0.903</b>	0.613	0.677	0.839	0.871
<b>GainRatio(14)</b>	0.774	0.645	0.806	0.806	<b>0.903</b>
<b>CFS_Relief(21)</b>	0.645	0.710	0.516	0.710	<b>0.871</b>
<b>CFS_SymmetricalUncert(16)</b>	0.742	0.710	<b>0.806</b>	0.903	<b>0.806</b>
<b>CFS_GainRatio(16)</b>	0.710	0.548	0.742	0.710	<b>0.839</b>
<b>Relief_SymmetricalUncert(22)</b>	0.774	0.516	0.839	0.871	<b>0.903</b>
<b>Relief_GainRatio(23)</b>	<b>1.0</b>	0.677	0.839	0.839	0.839
<b>SymmetricalUncert_GainRatio(15)</b>	0.710	0.581	0.677	<b>0.839</b>	<b>0.839</b>
<b>All(29)</b>	0.548	0.581	0.742	<b>0.903</b>	<b>0.903</b>

Estos resultados muestran que el RF proporciona el mejor desempeño para la mayoría de los conjuntos (no balanceados) resultantes de las 3 actividades desarrolladas, con una precisión por encima del 80%.

### 5.2.2 Desempeño de los algoritmos de clasificación para datos balanceados

En la Tabla 20 se muestra la precisión de la clasificación de los 11 subconjuntos balanceados de la actividad “Repetir frases”. El clasificador que mejor desempeño mostró en todos los conjuntos es el RF. Los mejores desempeños se obtienen con los conjuntos *Relief* y *Relief\_GainRatio* con un 97% cada uno y *Relief\_SymmetricalUncert* con 96%. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

**Tabla 20.** Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Repetir frases” para LOSO.

	LDA	DT	KNN	SVM	RF
<b>CFS(44)</b>	0.854	0.864	0.676	0.897	<b>0.945</b>
<b>Relief(46)</b>	0.837	0.892	0.889	0.877	<b>0.970</b>
<b>SymmetricalUncert(55)</b>	0.849	0.889	0.621	0.884	<b>0.950</b>
<b>GainRatio(36)</b>	0.817	0.887	0.681	0.857	<b>0.945</b>
<b>CFS_Relief(74)</b>	0.862	0.864	0.894	0.932	<b>0.957</b>
<b>CFS_SymmetricalUncert(64)</b>	0.887	0.902	0.852	0.927	<b>0.957</b>
<b>CFS_GainRatio(63)</b>	0.872	0.867	0.676	0.915	<b>0.950</b>
<b>Relief_SymmetricalUncert(74)</b>	0.867	0.889	0.802	0.899	<b>0.960</b>
<b>Relief_GainRatio(77)</b>	0.839	0.915	0.729	0.844	<b>0.970</b>
<b>SymmetricalUncert_GainRatio(73)</b>	0.864	0.915	0.653	0.907	<b>0.945</b>
<b>All(115)</b>	0.867	0.879	0.817	0.945	<b>0.950</b>

En la Tabla 21 se muestra la precisión de la clasificación de los 11 subconjuntos balanceados de la actividad “Fluidez verbal semántica”. El clasificador que mejor desempeño mostró en todos los conjuntos es el RF con 100% en el *CFS* y 98% con los conjuntos *CFS\_Relief*, *Relief\_GainRatio* y *SymmetricalUncert\_GainRatio*. Cabe señalar que, aunque el LDA no mostró un comportamiento estable en todos los conjuntos, con el *CFS* obtuvo un 98% de precisión. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

En la Tabla 22 se muestra la precisión de la clasificación de los 11 subconjuntos balanceados de la actividad “Robo de galletas”. El clasificador que mejor desempeño mostró en todos los conjuntos es el RF con una precisión por encima del 95% en la mayoría de los conjuntos (10 conjuntos de 11). La mejor precisión la obtuvo con *GainRatio* con 100%. El valor en () indica la cantidad de rasgos con la que se entrenaron los modelos correspondientes.

**Tabla 21.** Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Fluidez verbal semántica” para LOSO.

	LDA	DT	KNN	SVM	RF
CFS(28)	<b>0.975</b>	0.85	0.925	0.85	<b>1.0</b>
Relief(29)	0.775	0.8	0.8	0.9	<b>0.95</b>
SymmetricalUncert(27)	0.775	0.825	0.775	0.85	<b>0.95</b>
GainRatio(25)	0.875	0.8	0.825	0.925	<b>0.95</b>
CFS_Relief(43)	0.825	0.825	0.85	0.875	<b>0.975</b>
CFS_SymmetricalUncert(43)	0.8	0.85	0.75	0.875	<b>0.925</b>
CFS_GainRatio(40)	0.7	0.775	0.825	0.9	<b>0.925</b>
Relief_SymmetricalUncert(46)	0.8	0.825	0.8	0.875	<b>0.925</b>
Relief_GainRatio(43)	0.75	0.875	0.8	0.875	<b>0.975</b>
SymmetricalUncert_GainRatio(34)	0.675	0.875	0.95	0.825	<b>0.975</b>
All(51)	0.675	0.675	0.875	0.825	<b>0.9</b>

**Tabla 22.** Desempeño de la clasificación de los conjuntos balanceados resultantes de la actividad “Robo de galletas” para LOSO.

	LDA	DT	KNN	SVM	RF
CFS(26)	0.711	0.842	0.868	0.737	<b>0.947</b>
Relief(24)	0.763	0.632	0.789	0.763	<b>0.895</b>
SymmetricalUncert(23)	0.895	0.816	0.789	0.842	<b>0.974</b>
GainRatio(23)	0.842	0.684	0.842	0.868	<b>1.0</b>
CFS_Relief(41)	0.737	0.711	0.711	0.868	<b>0.947</b>
CFS_SymmetricalUncert(31)	0.737	0.711	0.658	0.842	<b>0.947</b>
CFS_GainRatio(33)	0.868	0.737	0.737	0.789	<b>0.947</b>
Relief_SymmetricalUncert(41)	0.947	0.737	0.763	0.789	<b>0.947</b>
Relief_GainRatio(44)	0.842	0.763	0.737	0.789	<b>0.974</b>
SymmetricalUncert_GainRatio(25)	0.947	0.684	0.842	0.842	<b>0.974</b>
All(50)	0.737	0.658	0.737	0.842	<b>0.974</b>

Estos resultados ponen en evidencia que el RF proporciona el mejor desempeño para todos los conjuntos resultantes de las 3 actividades desarrolladas, con una precisión por encima del 90%, mejorando el desempeño del resto de los algoritmos analizados y de los obtenidos con los conjuntos sin balancear.

### 5.3 Variables representativas identificadas con técnicas de XAI

Los resultados de esta sección se presentan en función de las reglas generadas por el algoritmo  *Anchors*  con una cobertura por encima del 13% para la actividad “Repetir frases” y por encima del 18% para las actividades de “Fluidez verbal semántica” y “Robo de galletas”. Las reglas generadas se muestran en el Anexo C.

Los rasgos relevantes que se identificaron a partir de las reglas generadas se muestran en la siguiente Tabla 23, están ordenados por la cantidad de veces que aparecieron en los predicados. En la actividad “Repetir frases” de 74 rasgos se identificaron 19 relevantes, en la de “fluidez verbal semántica” de 28 rasgos 25 fueron elegidos y en la en la actividad “Robo de galletas” de 25 rasgos quedaron 13.

El rasgo  *lspFreq\_sma[7]\_min*  (marcado de color rojo en la tabla 23) es el único rasgo que se repite en las actividades de “Repetir frases” y “Robo de galletas”. El resto de los rasgos no mostraron coincidencias en ninguna de las actividades.

## 5.4 Discusión

### 5.4.1 Desempeño de los algoritmos de clasificación

Los modelos construidos lograron desempeños similares (o superiores) a los estudios de análisis acústicos reportados en la literatura (mostrado en la Tabla 6), para discriminar a través de la voz si un adulto mayor tiene demencia. No obstante, los resultados positivos deben tomarse con cautela, dado que el número de datos analizados es relativamente bajo, y por lo mismo se utilizó el procedimiento de validación LOSO que puede sobreestimar los resultados (Farrahi et al., 2019).

Para comparar los resultados con otros estudios, se retoma el conjunto de datos ADRESS (descrito en la sección 2.3.2 Corpus de audios), siguiendo la metodología de esta investigación para el análisis de los datos (mostrada en el Capítulo 4). En la Tabla 24 muestra los resultados reportados por el procedimiento utilizado en (Luz et al., 2020) y los obtenidos con el flujo correspondiente al algoritmo de selección de rasgos Correlation Subset de la metodología propuesta. Los valores entre () indican la cantidad de rasgos con la que se entrenaron los modelos.

**Tabla 23.** Variables representativas por cada actividad. El número () es la cantidad de ocurrencias.

Repetir frases	Fluidez verbal semántica	Robo de galletas
mfcc_sma[10]_linregc2.1 (4) jitterDDP_sma_rqmean (3) lspFreq_sma[6]_amean (3) lspFreq_sma[2]_quartile1 (2) mfcc_sma[4]_rqmean (2) slopeV500-1500_sma3nz_amean (2) audSpec_Rfilt_sma[0]_percentile 1.0 (2) jitterLocal_sma_percentile99 (1) lspFreq_sma[6]_quartile3 (1) lspFreq_sma[6]_quartile1 (1) <b>lspFreq_sma[7]_min (1)</b> lspFreq_sma[5]_quartile3 (1) pcm_fftMag_spectralFlux_sma_de_meanSegLen (1) F0env_sma_linregc2 (1) audspec_lengthL1norm_sma_peakMeanAbs (1) mfcc_sma[10]_lpc0 (1) mfcc_sma[4]_percentile1.0 (1) audSpec_Rfilt_sma[10]_iqr1-2 (1) voiceProb_sma_max (1)	pcm_fftMag_fband1000-4000_sma_de_maxPos(4) shimmerLocal_sma_de_qregc1(4) jitterLocal_sma_upleveltime90 (4) audSpec_Rfilt_sma[9]_maxSegLen (3) audSpec_Rfilt_sma_de[20]_skewness (3) mfcc_sma[9]_centroid (3) pcm_fftMag_fband250-650_sma_de_upleveltime50(2) pcm_fftMag_fband1000-4000_sma_de_minPos (2) audSpec_Rfilt_sma_de[9]_skewness (2) audspec_lengthL1norm_sma_de_peakRangeAbs(2) audSpec_Rfilt_sma[19]_maxPos (2) alphaRatioV_sma3nz_amean(2) mfcc_sma[2]_minPos(2) jitterLocal_sma_percentile1.0 (2) lspFreq_sma_de[7]_minPos(2) pcm_fftMag_psySharpness_sma_de_maxPos (1) pcm_fftMag_spectralRollOff90.0_sma_de_minSegLen (1) pcm_zcr_sma_minPos.1(1) pcm_fftMag_fband1000-4000_sma_de_peakDistStddev(1) audSpec_Rfilt_sma[24]_minPos(1) audSpec_Rfilt_sma[14]_qregc1 (1) audSpec_Rfilt_sma_de[22]_meanPeakDist (1) audSpec_Rfilt_sma[13]_upleveltime90 (1) audSpec_Rfilt_sma_de[13]_upleveltime75(1) answerQ1(1) lspFreq_sma_de[7]_quartile3(1)	<b>lspFreq_sma[7]_min (4)</b> mfcc_sma_de[2]_lpc3 (4) jitterLocal_sma_rqmean (4) jitterLocal_sma_amean (4) lspFreq_sma[4]_maxPos (2) audSpec_Rfilt_sma[17]_maxPos (2) F0semitoneFrom27.5Hz_sma3nz_meanRisingSlope (2) mfcc_sma_de[4]_stddevFallingSlope (1) mfcc_sma[5]_rqmean (1) mfcc_sma_de[14]_lpc4 (1) jitterLocal_sma_range (1) pcm_RMSenergy_sma_de_maxSegLen (1) pcm_fftMag_spectralRollOff50.0_sma_minSegLen (1)

Analizando la Tabla 24 se muestra que, utilizando la metodología de este estudio, se mejoró el desempeño en todos los algoritmos analizados, además, a un menor costo computacional, porque al utilizar un menor número de rasgos se reduce el tiempo de entrenamiento, el algoritmo toma las decisiones más rápido y mejora su simplicidad. El bajo desempeño del conjunto de datos ADDRESS puede atribuirse a la baja calidad de las grabaciones como se mencionó en la sección 2.3.2 Corpus de audios. Cabe destacar que no obstante estas limitaciones, el conjunto de datos ADDRESS sigue siendo muy utilizado, lo que hace pensar que el

conjunto de datos como el creado en este trabajo pueden resultar una contribución importante para avanzar esta línea de investigación.

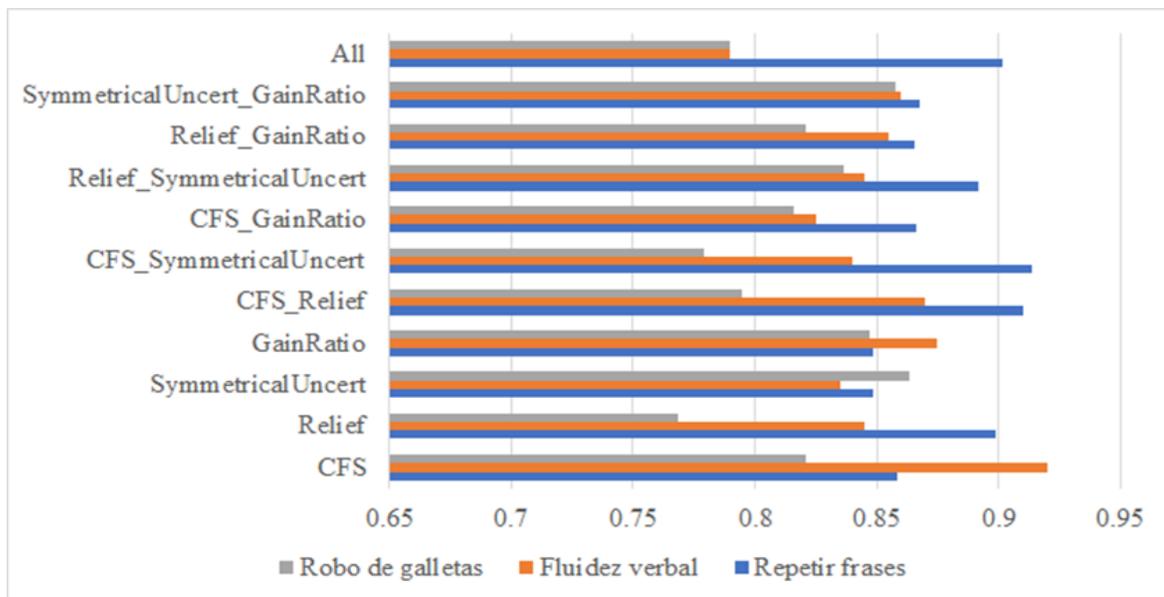
**Tabla 24.** Desempeño de los algoritmos de clasificación del conjunto de datos ADRess reportado en la literatura y con un flujo de la metodología propuesta en este estudio.

		LDA	DT	KNN	SVM	RF
<b>emobase</b>	ADReSS (988)	0.500	0.519	0.398	0.491	0.472
	<i>CFS(15)</i>	<i>0.645</i>	<i>0.591</i>	<i>0.562</i>	<i>0.611</i>	<i>0.675</i>
<b>eGeMAPS</b>	ADReSS (88)	0.482	0.500	0.380	0.333	0.482
	<i>CFS(10)</i>	<i>0.612</i>	<i>0.558</i>	<i>0.581</i>	<i>0.606</i>	<i>0.628</i>

Ninguno de los estudios analizados en el trabajo previo reporta el uso de técnicas de optimización de parámetros. Al mejor de nuestro conocimiento, ninguno de los trabajos reportados en el área utiliza esta técnica. Este constituye junto con la calidad de los archivos de audios recolectados y el preprocesamiento de estos, un elemento distintivo que contribuyó a obtener los buenos resultados que se reportan en este estudio, dada la poca separabilidad que los rasgos mostraron entre las clases en los análisis de tSNI (Figura 10).

La mayoría de los algoritmos de clasificación utilizados mostraron mejor desempeño después de aplicar la técnica de balanceo de datos propuesta. Por lo general el desbalance de datos afecta a los algoritmos en su proceso de generalización, perjudicando a la clase minoritaria (en nuestro caso la de individuos con demencia). En algunos clasificadores fue más evidente la mejoría en su desempeño que en otros. SVM, por ejemplo, no mejoró su desempeño después de balanceados los datos. Sin embargo, el RF si lo mejoró en todos los conjuntos. Este comportamiento se debe a que RF, siendo un método ensemble, funciona a través de consensos y toma las decisiones a partir de muchos árboles de clasificación individuales.

La combinación de los subconjuntos principales: *CFS*, *Relief*, *SymmetricalUncert* y *GainRatio* no mejoraron su desempeño en todos los conjuntos de datos. En el conjunto derivado de la actividad “Repetir frases” mejoró en todas las combinaciones excepto en los subconjuntos *Relief\_GainRatio* y *Relief\_SymmetricalUncert*. En el caso de las actividades de “Fluidez verbal semántica” y descripción de la imagen “Robo de galletas” no mejoraron en ningún subconjunto. Por tanto, se concluye que la combinación de varios selectores no conlleva a obtener mejores resultados. En la gráfica 17 se muestra la variabilidad de los resultados por estrategia de selección de características.

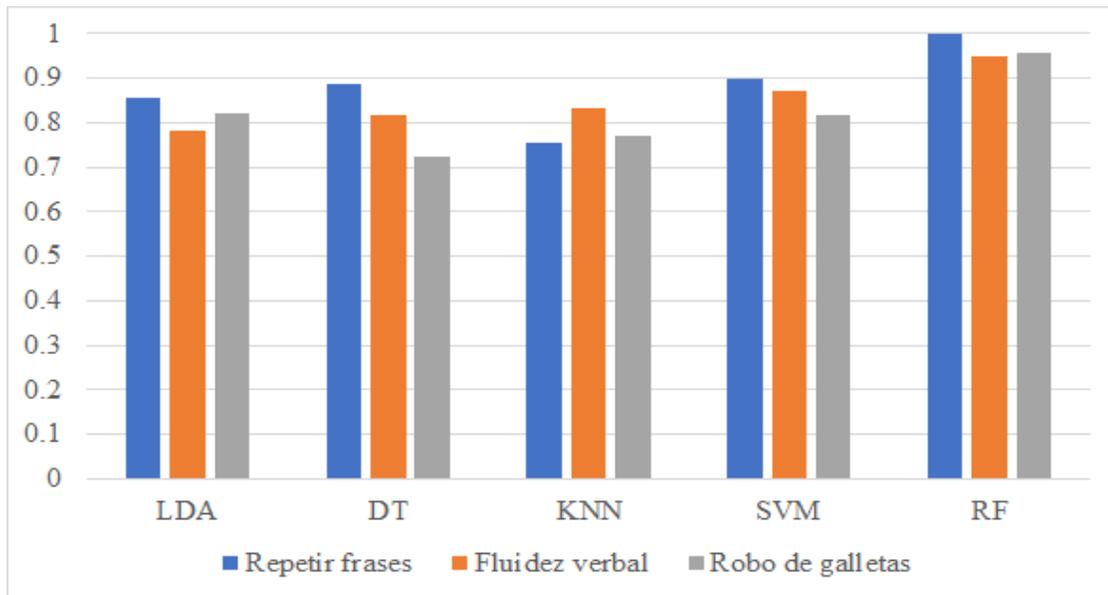


**Figura 17.** Variabilidad de los resultados por estrategia de selección de características.

También vale la pena señalar en la Tabla 19 que, aunque la exactitud de las predicciones de los subconjuntos principales está entre 70 % y 100%, la precisión del LDA, cuando se unieron las características de todos los selectores cae a 54%. Esto puede ser ocasionado, al menos parcialmente, a que como se evidencia en la Tabla 19, la combinación *CFS\_Relief* mostró bajo sinergismos para clasificar, por tanto, cuando se combinan todos los subconjuntos tiene un comportamiento similar.

Realizando un análisis por técnica de clasificación éstas mostraron un comportamiento diferente por cada una de las tareas analizadas. El LDA, DT, SVM y RF mostraron mejor desempeño en la actividad de “Repetir frases” y el KNN en la actividad de “Fluidez verbal semántica”. La Figura 18 muestra el desempeño de los clasificadores por cada actividad.

Una limitante importante para resaltar en este sentido es el proceso de validación que se utilizó en la experimentación, condicionado por la relativamente baja cantidad de datos. LOSO es un procedimiento optimista y puede conllevar a conclusiones erróneas (Farrahi et al., 2019).



**Figura 18.** Desempeño de los clasificadores por cada actividad.

#### 5.4.2 Interpretación de los modelos a partir de XAI

En este estudio, se aplica el algoritmo *Anchors* para identificar variables importantes que puedan explicar la decisión que está generando el modelo, en este caso, identificar si un adulto mayor presenta demencia. Las reglas derivadas de los conjuntos de datos analizados presentaron poco porcentaje de cubrimiento, constituyendo una limitante para generalizar los resultados. Los rasgos relevantes, los que más repeticiones mostraron en los predicados, resultantes del análisis de las 3 actividades codificaron la siguiente información acústica mostrada en la Tabla 25

Los resultados en este estudio de XAI aportan evidencia de que los modelos están generando decisiones en función de rasgos acústicos asociados a alteraciones de voz de adultos mayores con demencia. Por un lado, de las 8 variables que más ocurrencia mostraron (en este caso, las variables que tuvieron 4 repeticiones en la Tabla 23), 6 se encontraron dentro de las primeras 5 posiciones en el análisis de fusión por ranking presentado en la sección 5.1 (Análisis descriptivo del conjunto de rasgos). Por otro lado, estudios recientes se han enfocado en demostrar una asociación entre diferentes características acústicas y la demencia. Cómo se reportó en (Ivanova and Meilán, 2022) en los adultos mayores con demencia se manifiesta una disminución del *shimmer*. En otro estudio (Lin et al., 2020), los rasgos como *jitter*, *pcm* y *mfcc* se asociaron significativamente con esta patología después de la corrección de Bonferroni ( $P <$

0,05/48 = 0,001), de estos el *jitter* y *pcm* se asociaron positivamente y *mfcc* negativamente con el riesgo de demencia.

**Tabla 25.** Información acústica que codifican los rasgos relevantes.

Rasgos	Información acústica que codifican
<i>lspFreq</i>	Pares de líneas espectrales, modelación matemática del tracto vocal.
<i>mfcc</i>	Coefficientes Cepstrales en las Frecuencias de Mel, aproximación del comportamiento del Sistema auditivo humano.
<i>pcm</i>	Representa digitalmente señales analógicas muestreadas.
<i>jitter</i>	Cambio indeseado y abrupto en la propiedad de la señal.
Shimmer	Perturbación en la amplitud de la frecuencia fundamental entre cada ciclo vocal (permite determinar el grado de disfonía de una voz).

El estudio realizado por Lin (Lin et al., 2020), con 157 sujetos con demencia y en inglés, encontró que además de los parámetros *jitter*, *mfcc*, *pcm*, los rasgos *audspec*, *voicingFinalUnclipped*, y *audspecRasta* son buenos discriminadores. La coincidencia en la importancia de este parámetro podría ser un indicio de que son dependientes del idioma. Pero en general no se esperaría que un modelo entrenado en un idioma funcionara bien para sujetos que hablan otro idioma. Un modelo independiente del idioma sería un objetivo interesante para explorar a mediano plazo, y para el que una base de datos federada con criterios homogéneos resultaría muy útil.

Comparando los rasgos relevantes entre cada actividad (mostrado en la Tabla 23), solamente el rasgo *lspFreq\_sma[7]\_min*, coincidió en las actividades de “Repetir frases” y “Robo de galletas”. La poca presencia de solapamiento en los rasgos relevantes de las 3 actividades, puede ser una señal de que las actividades pueden ser complementarias y combinarse en un método de aprendizaje por ensemble para trabajo futuro.

Nuestros resultados confirmaron que las características acústicas pueden ser utilizadas para predecir el diagnóstico de demencia. Además, este tipo de estudio tiene el potencial de convertirse en una herramienta eficiente para evaluar el riesgo de demencia futura a partir de las grabaciones de voz antes de que aparezcan los síntomas cognitivos.

## Capítulo 6. Conclusiones

---

La demencia es una de las patologías que más afecta a los adultos mayores. Actualmente existen varias alternativas para diagnosticarla. Una de las más utilizadas es la voz como biomarcador digital, dado que es una técnica relativamente económica y no invasiva. A pesar de que se han reportado varias investigaciones en este sentido, aún continúa siendo un problema abierto, sobre todo en países para los que no se cuenta con bases de datos, como sucede con países de habla hispana. En este trabajo de tesis, se analiza la voz a partir de técnicas de aprendizaje automático para apoyar el diagnóstico de demencia en adultos mayores.

Se presentó una herramienta para capturar los registros de voz de los adultos mayores en el desarrollo de las actividades cognitivas. El sistema desarrollado está formado por dos procedimientos fundamentales, el procedimiento manual y el autónomo. Además, incluye tres módulos que corresponden a cada una de las actividades: “Repetir frases”, “Prueba de fluidez verbal semántica”, y “Robo de galletas”. Es una aplicación de escritorio desarrollada en Python.

El procedimiento de captura de datos se realizó a través de la interacción de 32 adultos mayores con el procedimiento autónomo en el desarrollo de las actividades mencionadas anteriormente, obteniendo de esta interacción 378 archivos de audios (315 correspondientes a la actividad “Repetir frases”, 32 a la actividad de “Fluidez verbal” y 31 a la actividad “Robo de galletas”) en un intervalo de 5 meses. A estos archivos se le aplicaron técnicas de preprocesamiento de datos para mejorar su calidad.

A cada archivo de audio preprocesado se le extrajeron características acústicas utilizando los programas PRATT y openSmile, estas fueron analizadas a través de técnicas de selección de atributos para identificar los rasgos que estaban aportando información útil. Se crearon los modelos utilizando diferentes algoritmos de clasificación, se analizaron las tareas de forma independiente, y el desempeño de los clasificadores con los conjuntos no balanceados y balanceados. Finalmente, se analizaron los modelos a través de técnicas de XAI, identificando los rasgos relevantes e interpretándose con respecto a características del habla.

Actualmente continuamos capturando datos, con 59 participantes incluidos en el estudio hasta la fecha. El corpus de audio creado será publicado en la base de datos compartida *DementiaBank* de interacciones multimedia para el estudio de la comunicación en la demencia.

Atendiendo a los resultados obtenidos se llegó a las siguientes **conclusiones**:

Los mejores desempeños fueron obtenidos por el algoritmo RF con más de un 90% de precisión para los 11 conjuntos de atributos seleccionados. Todos los algoritmos, excepto el SVM, mostraron mejor desempeño después de balanceados los datos.

El ensamble de características, producto de juntar los atributos obtenidos al aplicar las siguientes técnicas de selección de rasgos *Correlation Subset*, *Relief-F*, *Gain Ratio* y *Symmetrical Uncertainty* (Figura 9, paso (5)) no mejoraron el desempeño de los algoritmos de clasificación.

Los resultados en este estudio de XAI aportan evidencia de que los modelos están generando decisiones en función de rasgos acústicos asociados a alteraciones de voz de adultos mayores con demencia. Se identificaron los rasgos *Shimmer*, *jitter*, *pcm* y *mfcc* como potenciales discriminadores para identificar dicha enfermedad.

## 6.1 Aportaciones

Las aportaciones de este trabajo de tesis son las siguientes:

Un corpus de audios en idioma español, para facilitar este tipo de estudios en países de habla hispana, formado por los archivos de audios resultantes del desarrollo de las actividades de “Repetir frases”, “Fluidez verbal semántica” y “Robo de galletas”. El corpus está formado por la información de 12 adultos mayores con demencia y 20 sanos. Por cada adulto mayor se recabaron 12 archivos de audios. A pesar de que el mismo se capturó evitando los problemas de otros conjuntos de datos como ADReSS (descrito en la sección 2.3.2 Corpus de audios), los errores presentados pueden ser manejados con ayuda del etiquetado realizado a dichos archivos. Además, se cuenta con la autorización de los participantes para compartir los datos con otros grupos para propósitos de investigación.

Una herramienta computacional para la recolección de muestras de audios, que permite capturar los registros de voz resultantes del desarrollo de las actividades “Repetir frases”, “Fluidez verbal semántica” y descripción de la imagen “Robo de galletas”, a partir de los procedimientos manual y autónomo.

Modelos computacionales utilizando algoritmos de clasificación, tales como: LDA, DT, KNN, SVM y RF. Además, se obtuvieron modelos adicionales con las mismas técnicas, pero con los conjuntos balanceados con datos sintéticos obtenidos a través del algoritmo Smote. Para cada modelo se optimizaron los

parámetros correspondientes y se utilizó el procedimiento de validación LOSO. Se interpretaron los modelos obtenidos a partir de técnicas de Inteligencia artificial explicable.

Se participó en el Congreso de Generación de Soluciones Industriales Basadas en Inteligencia Artificial (SIBIA 2021). Se obtuvo el 2do lugar en el Taller de Desarrollo de Ideas de Negocio Basadas en Inteligencia Artificial (INEBIA).

Un artículo aceptado en *IEEE Mexican International Conference on Computer Science (ENC)*:

Leyva, L. C., Vara, J. F., Sandoval, D. C., Santos, D. L. P., & Jauregui, M. H. (2022, August). Creating a Spanish Speech Corpus to Develop Digital Dementia Biomarkers Using Machine Learning. In 2022 IEEE Mexican International Conference on Computer Science (ENC) (pp. 1-5).

## 6.2 Limitaciones

El estudio realizado presenta algunas limitaciones que limitan la generalización de los resultados:

La cantidad de datos recolectados. La recolección de datos en su etapa inicial estuvo muy subordinada primeramente a la aprobación del Comité de bioética del CICESE y del hospital donde se encuentra la clínica del médico geriatra y durante los 5 meses de recolección al semáforo condicionado por la pandemia de Covid. A pesar de que el número de datos recabados no está muy distante a los que se utilizan en este tipo de estudios, para lograr la confiabilidad de los modelos y la capacidad de generalización de las técnicas de aprendizaje automático se necesita una mayor cantidad, permitiéndonos incluso, explorar otras técnicas como Aprendizaje profundo, que han mostrado buen comportamiento en este tipo de estudios.

Procedimiento de validación utilizado (LOSO). En este estudio por la cantidad de datos recolectados solamente se experimentó utilizando la estrategia de validación LOSO, esta estrategia a pesar de ser muy utilizada en este tipo de estudios es un procedimiento optimista y puede conllevar a conclusiones erróneas.

### 6.3 Trabajo futuro

El trabajo realizado en esta tesis abre interesantes líneas de investigación entre las que se destacan las siguientes:

Ampliar corpus de audios. Actualmente continuamos capturando datos y a la fecha se tienen 59 participantes incluidos en el estudio. Generar modelos que incluyan más sujetos y experimentar con otros procedimientos de validación, se vuelve más tentativo para mejorar el aprendizaje de los modelos y para generalizar los resultados.

Publicar el corpus de audio creado en el Dementia Bank. A pesar de que es una tendencia el aumento de estudios de este tipo en países de habla hispana, actualmente, se evidencia una carencia de disponibilidad de corpus de audios en español. Compartir el corpus de audio creado en esta tesis cuando se tengan más de 80 o 100 participantes, sería de gran ayuda a la comunidad que investiga en esta área.

Realizar un estudio longitudinal con datos de varias sesiones por sujetos, para analizar el posible deterioro de la voz a lo largo del tiempo. Este análisis puede permitir determinar si un individuo con deterioro cognitivo leve desarrollará o no demencia (Davis et al., 2018).

Crear modelos que incluyan 2 o más actividades y comparar su precisión con los modelos desarrollados.

Utilizar la metodología propuesta en este estudio con otros conjuntos de datos, por ejemplo, el de Ivanova corpus (Ivanova et al., 2022) que es en español. Además, utilizar otros en corpus en inglés para explorar la influencia del lenguaje en los modelos.

Crear modelos de aprendizaje automático basados en la combinación de rasgos acústicos y lingüísticos, y comparar la precisión con los obtenidos en este trabajo de tesis.

El diagnóstico de demencia utilizando rasgos de voz es un campo de investigación incipiente, pero con gran potencial. Este trabajo aporta a esta línea de conocimiento con importantes aplicaciones sociales y económicas.

## Literatura citada

---

- Alzheimer's Association. Consultado el 16 de agosto de 2022 de <https://www.alz.org/alzheimers-dementia/stages>
- Al-hameed, S., Benaissa, M., and Christensen, H. 2017. Detecting and Predicting Alzheimer ' s Disease Severity in Longitudinal Acoustic Data. In *Proceedings of the International Conference on Bioinformatics Research and Applications*, pp. 57–61. doi: 10.1145/3175587.3175589
- Ammar, R. Ben and Ayed, Y. Ben. 2018. Speech processing for early Alzheimer Disease diagnosis : Machine learning based approach. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, pp. 1–8. doi: 10.1109/AICCSA.2018.8612831
- Ashton, N. J., Pascoal, T. A., Karikari, T. K., Benedet, A. L., Lantero-Rodriguez, J., Brinkmalm, G., Snellman, A., Schöll, M., Troakes, C., Hye, A., Gauthier, S., Vanmechelen, E., Zetterberg, H., Rosa-Neto, P., and Blennow, K. 2021. Plasma p-tau231: a new biomarker for incipient Alzheimer's disease pathology. *Acta Neuropathologica*. Springer Berlin Heidelberg, 141(5), pp. 709–724. doi: 10.1007/s00401-021-02275-6
- Balagopalan, A., Eyre, B., Robin, J., and Rudzicz, F. 2021. Comparing Pre-trained and Feature-Based Models for Prediction of Alzheimer ' s Disease Based on Speech. *Frontiers in Aging Neuroscience*, 13, pp. 635945. doi: 10.3389/fnagi.2021.635945
- Ballester, A. and Santana, C. 1993. Transcription Conventions udes for the Corpus of Spoken Contemporary Spanish. *Literary and Linguistic Computing*, 8(4), pp. 283–292. doi: 10.1093/llic/8.4.283
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. Elsevier B.V., 58(December 2019), pp. 82–115. doi: 10.1016/j.inffus.2019.12.012
- Beam, C. R., Kaneshiro, C., Jang, J. Y., Reynolds, C. A., Pedersen, N. L., and Gatz, M. 2018. Differences between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease. *Journal of Alzheimer's Disease*, 64(4), pp. 1077–1083. doi: 10.3233/JAD-180141
- Becker, J. and Boller, F. 1994. The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6), pp. 585–594. doi: doi:10.21415/T5CS3J
- Billsus, D., Billsus, D., Pazzani, M. J., and Pazzani, M. J. 1998. Learning collaborative information filters. In *Proceedings of the Fifteenth International Conference on Machine Learning*, 54, pp. 47. Recuperado de <http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-005.pdf>
- Boletín INGER. 2020. Estigmas Hacia Demencia Frenan Atención y Diagnóstico Oportuno a Personas Que Desarrollan Alzheimer, pp. 9–13. Recuperado de [https://issuu.com/institucionacionaldegeriatria/docs/boletin\\_sept-2020](https://issuu.com/institucionacionaldegeriatria/docs/boletin_sept-2020)
- Calz, L., Gagliardi, G., Rossini, R., and Tamburini, F. 2020. Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech & Language*, 65, pp. 101113. doi: 10.1016/j.csl.2020.101113

- Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P., and Riva, G. 2020. Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in Human Neuroscience*, 14(July), pp. 1–12. doi: 10.3389/fnhum.2020.00245
- Chien, Y., Hong, S., Cheah, W., Yao, L., and Chang, Y. 2019. An Automatic Assessment System for Alzheimer's Disease Based on Speech Using Feature Sequence Generator and Recurrent Neural Network. *Scientific Reports*, 9(1), pp. 1–10. doi: 10.1038/s41598-019-56020-x
- Chinaei, H., Currie, L. C., Danks, A., and Lin, H. 2017. Identifying and Avoiding Confusion in Dialogue with People with Alzheimer Disease. *Computational Linguistics*, 43(2), pp. 377–406. doi: 10.1162/COLI
- Chlasta, K. and Wołk, K. 2021. Towards Computer-Based Automated Screening of Dementia Through Spontaneous Speech. *Frontiers in Psychology*, 11, pp. 623237. doi: 10.3389/fpsyg.2020.623237
- Cui, Y., Jia, M., Lin, T. Y., Song, Y., and Belongie, S. 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 9260–9269. doi: 10.1109/CVPR.2019.00949
- Cummings, L. 2019. Describing the Cookie Theft picture Sources of breakdown in Alzheimer's dementia. *Pragmatics and Society*, 10(2), pp. 153–176. doi: <https://doi.org/10.1075/ps.17011.cum>
- Davis, M., O'Connell, T., Johnson, S., Cline, S., Merikle, E., Martenyi, F., and Simpson, K. 2018. Estimating Alzheimer's Disease Progression Rates from Normal Cognition Through Mild Cognitive Impairment and Stages of Dementia. *Current Alzheimer Research*, 15(8), pp. 777–788. doi: 10.2174/1567205015666180119092427
- Demencia. 2020. Organización Mundial de Salud. Recuperado de <https://www.who.int/es/news-room/fact-sheets/detail/dementia>
- Dimić, G., Rančić, D., Milentijević, I., Spalević, P., and Plečić, K. 2017. Comparative Study: Feature Selection Methods in the Blended Learning Environment. *Facta Universitatis, Series: Automatic Control and Robotics*, 16(2), pp. 95. doi: 10.22190/fuacr1702095d
- Duell, J. A. 2021. A Comparative Approach to Explainable Artificial Intelligence Methods in Application to High-Dimensional Electronic Health Records: Examining the Usability of XAI. *ArXiv Preprint ArXiv*, 2103.04951. doi: 10.48550/arXiv.2103.04951
- Duong, A., Giroux, F., Tardif, A., and Ska, B. 2005. The heterogeneity of picture-supported narratives in Alzheimer's disease. *Brain and Language*, 93(2), pp. 173–184. doi: 10.1016/j.bandl.2004.10.007
- Fang, C., Janwattanapong, P., Martin, H., Cabrerizo, M., and Barreto, A. 2017. Computerized Neuropsychological Assessment in Mild Cognitive Impairment Based on Natural Language Processing-oriented Feature Extraction. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 543–546. doi: 10.1109/BIBM.2017.8217706
- Faroqi Shah, Y., Treanor, A., Ratner, N. B., Ficek, B., Webster, K., and Tsapkini, K. 2020. Using narratives in differential diagnosis of neurodegenerative syndromes. *Journal of Communication Disorders*. Elsevier Inc., 85, pp. 105994. doi: 10.1016/j.jcomdis.2020.105994

- Farrahi, V., Niemelä, M., Kangas, M., Korpelainen, R., and Jämsä, T. 2019. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait and Posture*. Elsevier, 68(November 2018), pp. 285–299. doi: 10.1016/j.gaitpost.2018.12.003
- Ferreira, D., Nordberg, A., and Westman, E. 2020. Biological subtypes of Alzheimer disease A systematic review and meta-analysis. *Neurology*, 94(10), pp. 436–448. doi: 10.1212/WNL.0000000000009058
- Fraser, K. C., Fors, K. L., and Kokkinakis, D. 2018. Multilingual word embeddings for the assessment of narrative speech in mild cognitive impairment. *Computer Speech & Language*. Elsevier Ltd, 53, pp. 121–139. doi: 10.1016/j.csl.2018.07.005
- Fukane, A. and Sahare, S. 2011. Different approaches of spectral subtraction method for enhancing the speech signal in noisy environments. *International Journal of Scientific & Engineering ...*, 2(5), pp. 1–6. Recuperado de [http://www.ijser.org/researchpaper/Different\\_Approaches\\_of\\_Spectral\\_Subtraction\\_method\\_for\\_Enhancing\\_the\\_Speech\\_Signal\\_in\\_Noisy\\_Environments.pdf](http://www.ijser.org/researchpaper/Different_Approaches_of_Spectral_Subtraction_method_for_Enhancing_the_Speech_Signal_in_Noisy_Environments.pdf)
- Ganganwar, V. 2012. An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, 2(4), pp. 42–47. Recuperado de [http://www.ijetae.com/files/Volume2Issue4/IJETAE\\_0412\\_07.pdf](http://www.ijetae.com/files/Volume2Issue4/IJETAE_0412_07.pdf)
- Garcia, S. de la F., Haider, F., and Luz, S. 2020. Cross-corpus Feature Learning between Spontaneous Monologue and Dialogue for Automatic Classification of Alzheimer ' s Dementia Speech. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 5851–5855. doi: 10.1109/EMBC44109.2020.9176305
- Garcia, S. de la F., Ritchie, C. W., and Luz, S. 2020. Artificial Intelligence, Speech , and Language Processing Approaches to Monitoring Alzheimer ' s Disease : A Systematic Review. *Journal of Alzheimer's Disease*, 78, pp. 1547–1574. doi: 10.3233/JAD-200888
- Gautam, R. and Sharma, M. 2020. Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis. *Journal of Medical Systems*. *Journal of Medical Systems*, 44(2), pp. 1–24. doi: 10.1007/s10916-019-1519-7
- Gogtay, N. J. and Thatte, U. M. 2017. Principles of correlation analysis. *Journal of Association of Physicians of India*, 65(MARCH), pp. 78–81. Recuperado de [https://www.kem.edu/wp-content/uploads/2012/06/9-Principles\\_of\\_correlation-1.pdf](https://www.kem.edu/wp-content/uploads/2012/06/9-Principles_of_correlation-1.pdf)
- Gosztolya, G., Toth, L., Grosz, T., Vincze, V., Hoffmann, I., Szatloczki, G., Pakaski, M., and Janos Kalman. 2016. *Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection*. pp. 107–111. doi: 10.21437/Interspeech.2016-384
- Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., and Hoffman, I. 2019. Identifying Mild Cognitive Impairment and Mild Alzheimer's Disease Based On Spontaneous Speech Using ASR and Linguistic Features. *Computer Speech & Language*, 53, pp. 181–197. doi: 10.1016/j.csl.2018.07.007
- Guinn, C., Singer, B., and Habash, A. 2014. A Comparison of Syntax, Semantics , and Pragmatics in Spoken Language among Residents with Alzheimer ' s Disease in Managed-Care Facilities. In IEEE Symposium on Computational Intelligence in Healthcare and E-Health (CICARE), pp. 98–103. doi: 10.1109/CICARE.2014.7007840

- Guo, Z., Ling, Z., and Li, Y. 2019. Detecting Alzheimer's Disease from Continuous Speech Using Language Models. *Journal of Alzheimer's Disease*, 70(4), pp. 1163–1174. doi: 10.3233/JAD-190452
- Haider, F., Fuente, S. D. La, and Luz, S. 2019. An Assessment of Paralinguistic Acoustic Features for Alzheimer's Dementia Recognition in Spontaneous Speech. *IEEE Journal of Selected Topics in Signal Processing*. IEEE, 14(2), pp. 272–281. doi: 10.1109/JSTSP.2019.2955022
- Hashemi Amin, Dowlatshahi, M., and Nezamabadi-pour, H. 2022. Ensemble of feature selection algorithms: a multi-criteria decision-making approach. *International Journal of Machine Learning and Cybernetics*, 13(1), pp. 49–69. doi: 0.1007/s13042-021-01347-z
- Ivanova, O. and Meilán, J. J. G. 2022. Speech Analysis in Preclinical Identification of Alzheimer's Disease. *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 13258 LNCS, pp. 363–368. doi: 10.1007/978-3-031-06242-1\_36
- Ivanova, O., Meilán, J. J. G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T. E., and González, N. C. 2022. Discriminating speech traits of Alzheimer's disease assessed through a corpus of reading task for Spanish language. *Computer Speech and Language*, 73(March 2021), pp. 101341. doi: 10.1016/j.csl.2021.101341
- Jagust, W. 2006. Positron emission tomography and magnetic resonance imaging in the diagnosis and prediction of dementia. *Alzheimer's and Dementia*, 2(1), pp. 36–42. doi: 10.1016/j.jalz.2005.11.002
- Karegowda, A. G., Manjunath, A. S., Ratio, G., and Evaluation, C. F. 2010. Comparative study of Attribute Selection Using Gain Ratio. *International Journal of Information Technology and Knowledge and Knowledge Management*, 2(2), pp. 271–277. Recuperado de [https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf%0Ahttp://csjournals.com/IJITKM/PDF 3-1/19.pdf](https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf%0Ahttp://csjournals.com/IJITKM/PDF%203-1/19.pdf)
- Khodabakhsh, A. and Demiroglu, C. 2015. Analysis of Speech-Based Measures for Detecting and Monitoring Alzheimer's Disease. In *Data Mining in Clinical Medicine* (Vol. 1246). doi: 10.1007/978-1-4939-1985-7
- Khodabakhsh, A., Yesil, F., Guner, E., and Demiroglu, C. 2015. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP Journal on Audio, Speech, and Music Processing*, 1, pp. 1–15. doi: 10.1186/s13636-015-0052-y
- Kourtis, L. C., Regele, O. B., Wright, J. M., and Jones, G. B. 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digital Medicine*. Springer US, 2(1), pp. 1–9. doi: 10.1038/s41746-019-0084-2
- Lazar, N. 2010. *Ockham's Razor*. *April*, pp. 243–246. doi: 10.1002/wics.75
- Lerman, K., Minton, S. N., and Knoblock, C. A. 2003. Wrapper maintenance: A machine learning approach. *Journal of Artificial Intelligence Research*, 18, pp. 149–181. doi: 10.1613/jair.1145
- Lin, H., Karjadi, C., Ang, T. F. A., Prajakta, J., McManus, C., Alhanai, T. W., Glass, J., and Au, R. 2020. Identification of digital voice biomarkers for cognitive health. *Exploration of Medicine*, 1(6), pp. 406–417. doi: 10.37349/emed.2020.00028

- Lizarduy, U. M. De, Salomón, P. C., Vilda, P. G., Torres, M. E., and Ipiña, K. L. de. 2017. ALZUMERIC : A decision support system for diagnosis and monitoring of cognitive impairment. *Loquens*, 4(1), pp. e037. doi: 10.3989/loquens.2017.037
- López de Ipiña, K., Alonso, J.-B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Martinez-Lage, P., Ecay-Torres, M., and 1, U. M. de L. 2013. On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. *Sensors*, 13(5), pp. 6730–6745. doi: 10.3390/s130506730
- López de Ipiña, K., Martinez de Lizarduy, U., Calvo, P. M., Mekyska, J., Beitia, B., Barroso, N., Estanga, A., Tainta, M., and Ecay-Torres, M. 2018. Advances on Automatic Speech Analysis for Early Detection of Alzheimer Disease: A Non-linear Multi-task Approach. *Current Alzheimer Research*, 15(2), pp. 139–148. doi: 10.2174/1567205014666171120143800
- Luz, S. 2017. Longitudinal Monitoring and Detection of Alzheimer ' s Type Dementia from Spontaneous Speech Data. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems Longitudinal, pp. 45–46. doi: 10.1109/CBMS.2017.41
- Luz, S., Haider, F., Fuente, S. De, Fromm, D., and Macwhinney, B. 2020. Alzheimer ' s Dementia Recognition through Spontaneous Speech : The ADReSS Challenge. *ArXiv Preprint ArXiv:2004.06833*. doi: 10.48550/arXiv.2004.06833
- Martínez-sáhez, F., Meilán, J. J. G., Vera-ferrandiz, J. A., Carro, J., Pujante-valverde, I. M., Ivanova, O., Carcavilla, N., Meilán, J. J. G., Vera-ferrandiz, J. A., Carro, J., Pujante-valverde, I. M., Ivanova, O., and Carcavilla, N. 2017. Speech rhythm alterations in Spanish-speaking individuals with Alzheimer ' s disease. *Aging, Neuropsychology, and Cognition*. Routledge, 24(4), pp. 418–434. doi: 10.1080/13825585.2016.1220487
- Matej, R., Tesar, A., and Rusina, R. 2019. Alzheimer ' s disease and other neurodegenerative dementias in comorbidity : A clinical and neuropathological overview. *Clinical Biochemistry*. Elsevier, 73(March), pp. 26–31. doi: 10.1016/j.clinbiochem.2019.08.005
- McDonnell, M., Dill, L., Panos, S., Amano, S., Brown, W., Giurgius, S., Small, G., and Miller, K. 2020. Verbal fluency as a screening tool for mild cognitive impairment. *International Psychogeriatrics*, 32(9), pp. 1055–1062. doi: 10.1017/S1041610219000644
- Mirheidari, B., Blackburn, D., Malley, R. O., Walker, T., and Christensen, H. 2019. Computational cognitive assessment: Investigating the use of an intelligent virtual agent for the detection of early signs of dementia. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2732–2736. doi: 10.1109/ICASSP.2019.8682423
- Mirzaei, S., Yacoubi, M. El, Garcia-salicetti, S., Boudy, J., Kahindo, C., Cristancho-lacroix, V., Kerhervé, H., and Rigaud, A. 2018. Two-Stage Feature Selection of Voice Parameters for Early Alzheimer ' s Disease Prediction. *IRBM*. Elsevier Masson SAS, 39(6), pp. 430–435. doi: 10.1016/j.irbm.2018.10.016
- Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T., Sumi, S., Kurita, S., Nakakubo, S., Lee, S., and Doi, T. 2020. Automatic Detection of Cognitive Impairments through Acoustic Analysis of Speech. *Current Alzheimer Research*, 17(1), pp. 60–68. doi: 10.2174/1567205017666200213094513

- Nasrolahzadeh, M., Mohammadpoory, Z., and Haddadnia, J. 2018. Higher-order spectral analysis of spontaneous speech signals in Alzheimer ' s disease. *Cognitive Neurodynamics*. Springer Netherlands, 12(6), pp. 583–596. doi: 10.1007/s11571-018-9499-8
- Ozgode Yigin, B. and Saygili, G. 2022. Confidence Estimation for t-SNE Embeddings Using Random Forest. *International Journal of Machine Learning and Cybernetics*. Springer Berlin Heidelberg, pp. 1–12. doi: 10.1007/s13042-022-01635-2
- Panegyres, P. K., Berry, R., and Burchell, J. 2016. Early dementia screening. *Diagnostics*, 6(1), pp. 6. doi: 10.3390/diagnostics6010006
- Patel, N. and Upadhyay, S. 2012. Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *International Journal of Computer Applications*, 60(12), pp. 20–25. doi: 10.5120/9744-4304
- Petti, U., Baker, S., and Korhonen, A. 2020. Review A systematic literature review of automatic Alzheimer ' s disease detection from speech and language. *Journal Ofthe American Medical Informatics Association*, 27(11), pp. 1784–1797. doi: 10.1093/jamia/ocaa174
- Piau, A., Wild, K., Mattek, N., and Kaye, J. 2019. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: Systematic review. *Journal of Medical Internet Research*, 21(8), pp. 1–14. doi: 10.2196/12785
- Pinto, T. C. C., Machado, L., Bulgacov, T. M., Rodrigues-Júnior, A. L., Costa, M. L. G., Ximenes, R. C. C., and Sougey, E. B. 2019. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? *Int Psychogeriatr*, 31(4), pp. 491–504. doi: 10.1017/S1041610218001370
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. Y., and Sainath, T. 2019. Deep Learning for Audio Signal Processing. *IEEE Journal on Selected Topics in Signal Processing*, 13(2), pp. 206–219. doi: 10.1109/JSTSP.2019.2908700
- Sachdev, P. S. and Mohan, A. 2016. DSM-5 and mental disorders in older individuals: an overview. *Harv Rev Psychiatry*, 23(5), pp. 320–328. doi: 10.1097/HRP.000000000000090.DSM-5
- Sadeghian, R., Schaffer, J. D., and Zahorian, S. A. 2017. Speech Processing Approach for Diagnosing Dementia in an Early Stage. In *INTERSPEECH 2017*, pp. 2705–2709. doi: 10.21437/Interspeech.2017-1712
- Syed, Z. S., Shehram, M., Syed, S., Lech, M., and Pirogova, E. 2021. Automated Recognition of Alzheimer ' s Dementia using Bag-of-Deep-Features and Model Ensembling. *IEEE Access*. IEEE, 9, pp. 88377–88390. doi: 10.1109/ACCESS.2021.3090321
- Tanaka, H., Kudo, T., Nakamura, S., Adachi, H., and Ukita, N. 2016. Automatic Detection of Very Early Stage of Dementia through Multimodal Interaction with Computer Avatars. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 261–265. doi: doi.org/10.1145/2993148.2993193

- Themistocleous, C., Eckerström, M., and Kokkinakis, D. 2018. Identification of Mild Cognitive Impairment From Speech in Swedish Using Deep Sequential Neural Networks. *Frontiers in Neurology*, 9, pp. 975. doi: 10.3389/fneur.2018.00975
- Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákási, M., and Kálmán, J. 2018. A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Current Alzheimer Research*, 15(2), pp. 130–138. doi: 10.2174/1567205014666171121114930
- Tr, J., Linz, N., Alexandra, K., and Robert, P. 2017. Automated Speech-based Screening for Alzheimer ' s Disease in a Care Service Scenario. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp. 292–297. doi: doi.org/10.1145/3154862.3154915
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., and Moore, J. H. 2018. Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*. Elsevier, 85(June), pp. 189–203. doi: 10.1016/j.jbi.2018.07.014
- Varoquaux, G. and Cheplygina, V. 2022. Machine learning for medical imaging: methodological failures and recommendations for the future. *Npj Digital Medicine*. Springer US, 5(1), pp. 1–8. doi: 10.1038/s41746-022-00592-y
- Vasco, P., Herriko, E., and Faundez-zanuy, M. 2017. Analysis of disfluencies for automatic detection of Mild Cognitive Impairment : a deep learning approach. In *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*, pp. 1–4. doi: 10.1109/IWOBI.2017.7985526
- Warnita, T., Inoue, N., and Shinoda, K. 2018. Detecting Alzheimer's Disease Using Gated Convolutional Neural Network from Audio Data. *ArXiv Preprint ArXiv*, 1803, pp. 11344. doi: 10.48550/arXiv.1803.11344
- Yeung, A., Iaboni, A., Rochon, E., Lavoie, M., Santiago, C., Yancheva, M., Novikova, J., Xu, M., Robin, J., Kaufman, L. D., and Mostafa, F. 2021. Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech- language changes in mild cognitive impairment and Alzheimer ' s dementia. *Alzheimer's Research & Therapy*. *Alzheimer's Research & Therapy*, 13(1), pp. 1–10. doi: 10.1186/s13195-021-00848-x
- Zimmerer, V. C., Wibrow, M., and Varley, R. A. 2016. Formulaic Language in People with Probable Alzheimer ' s Disease : A Frequency-Based Approach. *Journal of Alzheimer's Disease*, 53, pp. 1145–1160. doi: 10.3233/JAD-160099

## Anexos

---

### Anexo A: Carta de consentimiento informado

#### CARTA DE CONSENTIMIENTO COMO SUJETO DE INVESTIGACIÓN

Ensenada Baja California, México, a \_\_\_\_\_ de \_\_\_\_\_ del 2022

Se le solicita su consentimiento para participar en el estudio de investigación. La participación en este estudio es completamente voluntaria. Por favor lea la siguiente información y siéntase libre de preguntar cualquier cosa que no entienda antes de decidir si desea participar.

El **propósito** del estudio es evaluar la memoria en adultos mayores a partir de análisis de voz usando herramientas computacionales.

#### Requerimientos

Usted es elegible para participar en este estudio si cumple con los siguientes requisitos:

1. Persona mayor de 60 años.
2. La lengua nativa es el español.

Y, no es elegible si:

1. No presenta capacidad para comunicarse verbalmente.

#### PROCEDIMIENTOS

El estudio se realizará antes de entrar a la consulta con la médica geriatra. Las sesiones durarán aproximadamente 15 minutos. El asistente de investigador (pasante de medicina o psicología) explicará el estudio al participante y familiar (si viene acompañado). Se desarrollarán actividades como: repetir frases cortas, pruebas de fluidez y descripción de una fotografía, explicándole al adulto mayor las instrucciones antes de comenzar cada actividad. Los datos de las sesiones de intervención serán obtenidos por medio de grabaciones de audio. Además de datos demográficos (e.g. edad, sexo) de los participantes obtenidos vía medios escritos.

En visitas posteriores a la consulta médica geriátrica se le volverá a invitar a participar en otra sesión para grabar datos adicionales siguiendo el mismo procedimiento, siempre y cuando hayan pasado al menos tres meses desde la última sesión.

## **RIESGOS E INCOMODIDADES**

Durante el estudio no se realizará ninguna intervención médica.

Se seguirán los protocolos de prevención asociados a la contingencia COVID-19 establecidos en la clínica, incluyendo mantener sana distancia, y darle al participante gel antibacterial y cubrebocas. La información de cada participante será confidencial. La base de datos de registros de audio no incluirá información personal del participante que permita identificarlo.

## **BENEFICIOS**

Con este estudio se busca beneficiar a los adultos mayores, al poder identificar riesgos de deterioro cognitivo en etapa preclínica. Además, al proporcionar un conjunto de datos resultante del desarrollo de las actividades con los adultos mayores, se facilitará el estudio de esta enfermedad en México, utilizando procedimientos no invasivos.

## **TERMINACIÓN VOLUNTARIA DEL ESTUDIO**

Usted es libre de dejar el estudio en cualquier momento. **Si usted decide dejar el estudio deberá avisar al equipo de investigación inmediatamente.**

Datos de los investigadores principales: Dr. Jesús Favela Vara, celular: 646 116 6429

Ing. Lisset Cabrera Leyva, celular: 553944505

## **CONFIDENCIALIDAD**

La anonimización de los datos es imprescindible dentro del proyecto, para cuidar la identidad del participante. De igual forma, el resguardo de los datos será restringida exclusivamente para los miembros del grupo de investigación y almacenados en los medios digitales que aseguren este acceso exclusivo. Queda estrictamente prohibido la transferencia, reproducción y/o publicación de los datos por motivos personales no alineados a los objetivos del proyecto de investigación.

## **PERMISO PARA COMPARTIR DATOS CON LA AUDIENCIA EN ESTA ÁREA**

Pensando en los beneficios a la comunidad científica, el equipo de investigación contempla en un futuro compartir los datos capturados durante el estudio con la comunidad de investigación en esta área. Los datos que se publicarían son: registros de audio, edad, género y resultado de diagnóstico neurocognitivo. Por favor indique a continuación si da su permiso para compartirlos:  Sí  No  Sus iniciales.

**ACUERDO DE PARTICIPACIÓN VOLUNTARIA**

Usted no debería firmar este documento a menos que lo haya leído. La participación en este estudio es voluntaria. Usted puede suspender su participación en cualquier momento sin sanciones ni pérdida de beneficios a los que tendría derecho. Su firma indica que usted ha leído la información en este documento de consentimiento y ha tenido la oportunidad de hacer cualquier pregunta que tenga sobre el estudio.

*Estoy de acuerdo en participar en el estudio.*

---

**Firma y nombre del participante**

---

**Firma y nombre del familiar  
(si procede)**

---

**Fecha**

---

**Firma y nombre del investigador**

---

**Fecha**

## Anexo B: Manual de etiquetado

### *Etiquetas:*

- **palabra\_cortada**
- **silencio**
- **encuestador:** Se escucha la voz del encuestador, Doctora, pasante de psicología o personal que esté dirigiendo la actividad.
- **respuesta\_correcta:** El adulto mayor menciona una fruta
- **respuesta\_incorrecta:** El adulto mayor menciona otro dominio, por ejemplo, mencionar animales, medios de transportes, flores, etc etc.

Los fáticos se etiquetan cuando se evidencia que el adulto mayor aún necesita la ayuda del encuestador. Se dividen en las siguientes etiquetas

- **fatico\_afirmación:** ex) cuando el adulto mayor menciona “tantas frutas que hay y no recuerdo ninguna”
- **fatico\_negación:** ex) cuando el adulto mayor menciona “no me acuerdo de ninguna fruta”
- **fatico\_duda:** ex) cuando el adulto mayor menciona “mmm”, “eh” o elementos que demuestren dudas
- **fatico\_interrogacion:** ex) cuando el adulto mayor menciona “la sandía es una fruta verdad”

Los ruidos se clasifican en las siguientes categorías:

- **ruido\_externo:** es el ruido que se siente fuera del lugar donde se está desarrollando la actividad, por ejemplo, ruidos de construcción, de ambulancias, de carros de policías, ladridos de perros, etc.
- **ruido\_timbreTelefono**
- **ruido\_risas**
- **ruido\_aplausos**
- **ruido\_murmullo**
- **ruido\_ordenarSilencio:** ex) “Shsss”
- **ruido\_intervenciónFamiliar:** cuando el familiar del adulto mayor interviene en el desarrollo de la actividad.

## Anexo C: Reglas generadas por el algoritmo Anchors en las 3 actividades

### Actividad “Repetir frases”

#### *Clase demencia*

Anchor: lspFreq\_sma[7]\_quartile1 > 2.63 AND jitterDDP\_sma\_rqmean > 0.05 AND  
 lspFreq\_sma[6]\_quartile3 > 2.38  
 Anchor: jitterLocal\_sma\_percentile99.0 > 0.17 AND jitterDDP\_sma\_rqmean > 0.05 AND  
 lspFreq\_sma[6]\_quartile3 > 2.34  
 AND slopeV500-1500\_sma3nz\_amean <= -0.01

#### *Clase sano*

Anchor: F0env\_sma\_linregc2 <= 258.10 AND lspFreq\_sma[6]\_amean <= 2.25 AND  
 lspFreq\_sma[2]\_quartile1 <= 0.86  
 Anchor: audspec\_lengthL1norm\_sma\_peakMeanAbs > 1.88 AND lspFreq\_sma[6]\_amean <= 2.25 AND mfcc\_sma[4]\_rqmean > 12.45  
 Anchor: jitterDDP\_sma\_rqmean <= 0.05 AND mfcc\_sma[10]\_linregc2.1 > -0.48 AND  
 audSpec\_Rfilt\_sma[0]\_percentile1.0 <= 0.04  
 Anchor: jitterDDP\_sma\_rqmean <= 0.05 AND mfcc\_sma[10]\_linregc2.1 > -0.48 AND  
 mfcc\_sma[10]\_lpc0 <= -1.48  
 Anchor: slopeV500-1500\_sma3nz\_amean > -0.02 AND lspFreq\_sma[6]\_amean <= 2.25 AND  
 lspFreq\_sma[2]\_quartile1 <= 0.86  
 Anchor: voiceProb\_sma\_max <= 0.95 AND mfcc\_sma[10]\_linregc2.1 > -0.48 AND  
 audSpec\_Rfilt\_sma[0]\_percentile1.0 <= 0.04  
 Anchor: audSpec\_Rfilt\_sma[10]\_iqr1-2 > 0.20 AND mfcc\_sma[10]\_linregc2.1 > -0.48 AND  
 mfcc\_sma[4]\_percentile1.0 <= -19.17

### Actividad “Fluidez verbal semántica”

#### *Clase demencia*

Anchor: pcm\_fftMag\_psySharpness\_sma\_de\_maxPos <= 0.56 AND  
 audSpec\_Rfilt\_sma[9]\_maxSegLen <= 5.24  
 Anchor: answerQ1 <= 6.00 AND shimmerLocal\_sma\_de\_qregc1 <= -0.00  
 Anchor: pcm\_fftMag\_fband1000-4000\_sma\_de\_maxPos > 0.21 AND  
 audSpec\_Rfilt\_sma\_de[20]\_skewness <= -0.06  
 Anchor: mfcc\_sma[9]\_centroid > 0.48 AND shimmerLocal\_sma\_de\_qregc1 <= -0.00  
 Anchor: audSpec\_Rfilt\_sma\_de[20]\_skewness <= 0.28 AND audSpec\_Rfilt\_sma[19]\_maxPos > 0.60  
 Anchor: audSpec\_Rfilt\_sma\_de[9]\_skewness <= 0.19 AND shimmerLocal\_sma\_de\_qregc1 <= -0.00

Anchor: audSpec\_Rfilt\_sma[24]\_minPos > 0.32 AND audSpec\_Rfilt\_sma\_de[9]\_skewness <= 0.19

Anchor: lspFreq\_sma\_de[7]\_minPos > 992.75 AND audSpec\_Rfilt\_sma\_de[20]\_skewness <= -0.06

Anchor: pcm\_fftMag\_fband1000-4000\_sma\_de\_peakDistStddev <= 2.59 AND audSpec\_Rfilt\_sma\_de[9]\_skewness <= 0.19

Anchor: audSpec\_Rfilt\_sma[9]\_maxSegLen <= 13.60 AND audSpec\_Rfilt\_sma\_de[20]\_skewness <= -0.06

Anchor: audSpec\_Rfilt\_sma[19]\_maxPos > 0.16 AND shimmerLocal\_sma\_de\_qregc1 <= -0.00

Anchor: audSpec\_Rfilt\_sma\_de[13]\_uplevertime75 > 0.00 AND shimmerLocal\_sma\_de\_qregc1 <= -0.00

### *Clase Sano*

Anchor: audSpec\_Rfilt\_sma[14]\_qregc1 > 0.00 AND jitterLocal\_sma\_uplevertime90 <= 0.00

Anchor: alphaRatioV\_sma3nz\_amean > -15.77 AND jitterLocal\_sma\_percentile1.0 <= 0.00

Anchor: mfcc\_sma[2]\_minPos <= 0.39 AND jitterLocal\_sma\_uplevertime90 <= 0.00

Anchor: alphaRatioV\_sma3nz\_amean > -15.77 AND

pcm\_fftMag\_spectralRollOff90.0\_sma\_de\_minSegLen > 0.61

Anchor: lspFreq\_sma\_de[7]\_quartile3 > 0.01 AND pcm\_fftMag\_fband250-650\_sma\_de\_uplevertime50 > 0.30

Anchor: pcm\_zcr\_sma\_minPos.1 > 1804.50 AND pcm\_fftMag\_fband250-650\_sma\_de\_uplevertime50 > 0.30

Anchor: jitterLocal\_sma\_uplevertime90 <= 0.00 AND jitterLocal\_sma\_percentile1.0 <= 0.00

Anchor: audspec\_lengthL1norm\_sma\_de\_peakRangeAbs > 0.42 AND pcm\_fftMag\_fband1000-4000\_sma\_de\_minPos <= 0.25

Anchor: pcm\_fftMag\_fband1000-4000\_sma\_de\_maxPos <= 0.36 AND

audspec\_lengthL1norm\_sma\_de\_peakRangeAbs > 0.58

Anchor: pcm\_fftMag\_fband1000-4000\_sma\_de\_maxPos <= 0.56 AND mfcc\_sma[9]\_centroid <= 0.48

Anchor: audSpec\_Rfilt\_sma[9]\_maxSegLen > 5.24 AND jitterLocal\_sma\_uplevertime90 <= 0.00

Anchor: audSpec\_Rfilt\_sma\_de[22]\_meanPeakDist <= 0.91 AND pcm\_fftMag\_fband1000-4000\_sma\_de\_minPos <= 0.25

Anchor: pcm\_fftMag\_fband1000-4000\_sma\_de\_maxPos <= 0.36 AND

audSpec\_Rfilt\_sma[13]\_uplevertime90 <= 0.00

Anchor: mfcc\_sma[2]\_minPos <= 0.39 AND mfcc\_sma[9]\_centroid <= 0.48

Anchor: audSpec\_Rfilt\_sma[19]\_maxPos <= 0.36 AND lspFreq\_sma\_de[7]\_minPos <= 992.75

### **Actividad “Robo de galletas”**

#### *Clase dementia*

Anchor: lspFreq\_sma[7]\_min > 2.33 AND audSpec\_Rfilt\_sma[17]\_maxPos <= 0.14 AND mfcc\_sma\_de[4]\_stddevFallingSlope <= 50.51

*Clase sano*

Anchor: lspFreq\_sma[4]\_maxPos > 1758.00 AND jitterLocal\_sma\_rqmean <= 0.04

Anchor: jitterLocal\_sma\_range <= 0.55 AND jitterLocal\_sma\_rqmean <= 0.04

Anchor: audSpec\_Rfilt\_sma[17]\_maxPos > 0.14 AND jitterLocal\_sma\_rqmean <= 0.04

Anchor: pcm\_RMSEnergy\_sma\_de\_maxSegLen > 8.34 AND mfcc\_sma\_de[2]\_lpc3 > -0.59

Anchor: lspFreq\_sma[4]\_maxPos > 1758.00 AND mfcc\_sma\_de[2]\_lpc3 > -0.59

Anchor: mfcc\_sma\_de[2]\_lpc3 > -0.61 AND lspFreq\_sma[7]\_min <= 2.33

Anchor: lspFreq\_sma[7]\_min <= 2.48 AND pcm\_fftMag\_spectralRollOff50.0\_sma\_minSegLen > 0.57

Anchor: F0semitoneFrom27.5Hz\_sma3nz\_meanRisingSlope <= 178.77 AND jitterLocal\_sma\_amean <= 0.02

Anchor: jitterLocal\_sma\_amean <= 0.03 AND lspFreq\_sma[7]\_min <= 2.33

Anchor: F0semitoneFrom27.5Hz\_sma3nz\_meanRisingSlope <= 128.16 AND mfcc\_sma\_de[2]\_lpc3 > -0.59

Anchor: jitterLocal\_sma\_amean <= 0.03 AND jitterLocal\_sma\_rqmean <= 0.04

Anchor: mfcc\_sma[5]\_rqmean <= 13.39 AND mfcc\_sma\_de[14]\_lpc4 <= 0.33 AND jitterLocal\_sma\_amean > 0.03