

Análisis Fonoaudiológico para la detección de alzheimer con IA

Juan David Guarnizo Gutierrez, *Ingeniería de Sistemas y Computacion*

Resumen—En Colombia, el sistema de salud y la academia enfrentan importantes desafíos en la implementación e investigación de tecnologías que beneficien a la población, especialmente a los adultos mayores que padecen condiciones como el Alzheimer. Esta enfermedad neurodegenerativa afecta significativamente la calidad de vida de los pacientes, generando dependencia y pérdida cognitiva. La Inteligencia Artificial (IA) presenta una oportunidad para mejorar el diagnóstico y el manejo del Alzheimer, permitiendo reducir costos y mejorar la calidad de vida. Este proyecto examina los avances recientes en IA aplicada a la medicina, enfocándose en la implementación de una IA que utiliza reconocimiento de voz para identificar el Alzheimer. Los resultados preliminares muestran una precisión de 95% en entrenamiento y 78% en validación, con un AUC de 0.87 en entrenamiento y 0.89 en validación. Aunque estos resultados son prometedores, indican la posibilidad de sobreajuste, sugiriendo la necesidad de más datos y refinamiento del modelo. Además, se destaca que la calidad de los datos recolectados es crucial para el éxito de los proyectos, ya que aquellos proyectos con datos bien recopilados y procesados muestran un mejor rendimiento en comparación con los que utilizan datos de menor calidad. Este trabajo subraya la importancia de la investigación continua en IA para mejorar el diagnóstico y tratamiento del Alzheimer en Colombia y Latinoamérica.

Abstract— In Colombia, the healthcare system and academia face important challenges in the implementation and research of technologies that benefit the population, especially older adults suffering from conditions such as Alzheimer's disease. This neurodegenerative disease significantly affects the quality of life of patients, generating dependence and cognitive loss. Artificial Intelligence (AI) presents an opportunity to improve the diagnosis and management of Alzheimer's disease, enabling cost reduction and improved quality of life. This project examines recent advances in AI applied to medicine, focusing on the implementation of an AI that uses speech recognition to identify Alzheimer's disease. Preliminary results show an accuracy of 95% in training and 78% in validation, with an AUC of 0.87 in training and 0.89 in validation. Although these results are promising, they indicate the possibility of overfitting, suggesting the need for more data and model refinement. Furthermore, it is highlighted that the quality of the data collected is crucial to the success of the projects, as those projects with well-collected and processed data show better performance compared to those using lower quality data. This work underscores the importance of continued AI research to improve the diagnosis and treatment of Alzheimer's disease in Colombia and Latin America.

Index Terms—Alzheimer, dementia, AI, diagnosis, medicine



1. INTRODUCCIÓN

Que es el Alzheimer?

El alzheimer es una enfermedad de alto riesgo que provoca en el paciente perdida de memoria, depresion, falta de capacidad para administrar los recursos, afasia, agnosia, apraxia y, resumiendo, el deterioro general de las capacidades cognitivas, de lenguaje y sociales lo que genera que las personas con la enfermedad dependan de sus seres queridos en esencialmente todo generando altos niveles de stres en ellas mismas y sus familiares. En el pais tenemos nuevos casos cada 4 minutos, siendo un padecimiento con casos mas recurrentes y entre los factores de riesgo podemos hallar alto o poco consumo de alcohol, consumo de cigarrillo y entre otros, depresion o factores geneticos [1] factores de riesgo presentes en la mayoria de la poblacion de tercera edad.

Entre los retos a los que nos enfrentamos con el Alzheimer, uno de los mas criticos es el diagnostico debido a que los medios

actuales de diagnostico suelen ser invasivos y costosos y a que ya con el diagnostico, el precio economico del tratamiento es muy alto. Generalmente el umbral de diagnostico se encuentra apartir de los 60 años y esto es un dato primordial para entender por que, el uso de la tecnologia y el avance de la medicina es esencial para cuidar y proteger la vida de miles de personas de la 3ra edad y sobretodo en paises de tercer mundo debido a que la reduccion de costos pueden beneficiar a la cantidad de personas que pueden acceder a una medicina justa y eficaz, puesto que como se mencionaba anteriormente, los tratamientos son costosos al involucrar el uso de farmacos, antioxidantes y antipsicoticos, lo cual hace que los costos aumenten conforme progresa la enfermedad y la ausencia de acceso a ellos atacan la calidad de vida y los derechos fundamentales de los pacientes.

Segun un estudio en Estados unidos [2], el solo valor de diagnostico de la enfermedad en el 91 ya presentaba una serie de costos importantes, y el tratamiento de la enfermedad presenta a un mas costos, no obstante, esto sin considerar los aumentos

en inflacion, cuanto antes se logre detectar la enfermedad, sera mucho mas sencillo evitar que la persona se deteriore lo cual mejoraria mucho mas la calidad de vida de la persona y tambien facilitaria la manutencion del paciente al poseer mas control sobre su enfermedad, ademas de brindarle a la persona el beneficio de decidir como va a enfrentarse a la enfermedad cuando aun posee las capacidades mentales para lograr hacerlo.

TABLE 1—Estimated Net Costs of Alzheimer's Disease per Person in 1991 (Midrange Estimates)	
Annual (undiscounted) direct costs	
Diagnosis (first year only)	\$ 1 450
Nursing home	7 570
Long-term mental hospital	392
Paid home care	3 140
Regular physician care	233
Acute care hospitalization	1 202
Other patient direct costs	0
Caregiver medical care	153
Total direct cost (first year only)	14 140
Total direct cost (second and later years)	12 690
Annual (undiscounted) indirect costs of unpaid home care	20 900
Total cost first year, excluding morbidity and mortality	35 040
Total cost second and later years, excluding morbidity and mortality	33 590
Total discounted direct cost*	47 581
Total discounted (direct cost + unpaid caregiver cost)*	123 556
Total discounted (direct cost + unpaid caregiver cost + disability and premature mortality cost)*	173 932

*Assuming a 4% annual discount rate and survival of 3.3 years for men, 4.3 years for women.

Figura 1. Costos de tener Alzheimer

La medicina y la tecnologia

la IA ha logrado avanzar mucho en los ultimos años gracias a los avances en hardware y desarrollo de modelos abstractos, hemos logrado generar modelos que permiten reconocer cada vez mas enfermedades en el cuerpo humano a traves de vision computacional y analisis de imagen [3], con procesamiento de voz o incluso con analisis multimodal [4], esto debido a que en realidad muchas enfermedades ya se diagnostican analizando este tipo de informacion y a que el cuerpo da muchas señales ante la presencia de una enfermedad. En enfermedades neurodegenerativas como el Alzheimer el uso de estas tecnologias es util ya que sus afectaciones se pueden reconocer a traves del habla y procesamiento de lenguaje del paciente y el lenguaje es uno de los indicativos mas importantes del estado social, emocional y cognitivo de una persona [14]. Lo anterior debido a que la velocidad de procesamiento, el lexico y en general la comunicacion ayuda a reconocer patrones patologicos ya que para un uso adecuado del lenguaje es necesario un cerebro sano y funcional[5]. Asi mismo, el analisis de la voz y los sonidos del cuerpo nos brindan varias pistas patologicas en medicina, con el estudio de esto en el pasado se han detectado enfermedades como el cancer de laringe, la estenopsis subglotica o torticollis espasmódica [6]. Para lograr la implementacion tecnologica dentro del diagnostico es necesario hallar modelos que permitan encontrar los patrones adecuados y biomarcadores en el lenguaje para distinguir cada caso de alzheimer y ayudar a dar diagnosticos mas precisos en todos los rangos de avance de la enfermedad, este tipo de modelamiento puede ser realizado mas eficazmente gracias a los ultimos desarrollos de deep learning y supervised learning [7], los cuales permiten estudiar y crear modelos con mayor focalizacion e importancia, lo cual en contra posicion con el MRI, analisis psicologicos y otros examenes tradicionales es bastante economico y poco invasivo una vez solidificado el modelo.

Todo este avance se ha dado gracias a distintos grupos de investigacion y empresas de Inteligencia Artificial y software que han generado herramientas y repositorios como Hugging face, data2vec [8], bases de datos de Audio y videos de pacientes [13] de facil de acceso y gracias a la implementacion de otras herramientas mas tradicionales como modelos de regresion linear y algoritmos de optimizacion para optimizar estos modelos [9]. Todo este avance comunitario nos permitira con el tiempo perfeccionar estos sistemas predictivos y lograr cada vez mas avances dentro de la medicina y la ingenieria aplicada a la misma, no obstante, como todo problema medico y de machine learning es necesario una gran cantidad de datos, por lo cual un mayor indice de investigacion e interes por parte de la comunidad cientifico en estas herramientas puede generar que se obtenga una mayor cantidad de datos y se financien mas proyectos para ayudar al ser humano, entrenar modelos no es sencillo y es necesario datos de alta calidad y tener en consideracion otras enfermedades que pueden pasar como falsos positivos de no ser considerados, todo esto mientras se tiene en claro tambien la importancia de la etica y del factor humano dentro de la creacion de estos modelos para evitar sesgos en las predicciones y tambien asegurar la privacidad de los datos del paciente y entender los deseos del mismo adecuadamente, aplicando modelos de inteligencia artificial junto a humanos para asegurar un tratamiento digno y humano del paciente[10], puesto que una falta del conexion con el sector medico, los doctores y los pacientes, estos modelos pueden avanzar mucho pero nunca ser implementados [11].

2. ANTECEDENTES

El diagnostico en alzheimer ha sido investigado principalmente en las ultimas decadas debido al avance tecnologico, una mayor compresion de la enfermedad y a el envejecimiento prolongado de la mayoria de la poblacion, lo que ha incentivado al desarrollo de mas metodos para ayudar a la poblacion con Alzheimer. Convencionalmente se ha usado el MMSE y el MoCa [4][15] los cuales ayudan a diagnosticar a traves de examenes cognitivos si el paciente posee o no alzheimer, estos examenes nos pueden ayudar a extraer informacion acerca de como y en que circunstancias podemos obtener datos para determinar el diagnostico del paciente, tambien es posible recurrir a escaneos (MRI) pero estos son mucho mas invasivos y costosos para los pacientes, lo cual incentivo a la comunidad medica a empezar a utilizar otros medios para reconocer enfermedades.

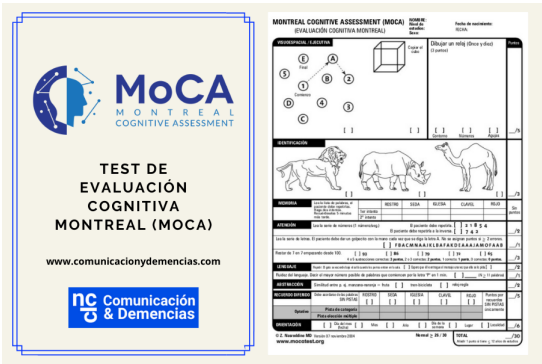


Figura 2. MoCA

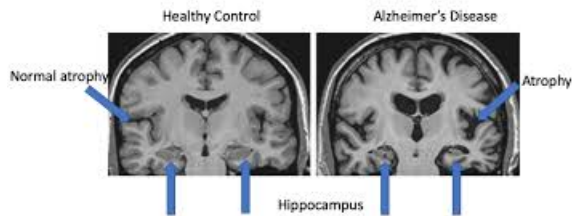


Figura 3. MRI Alzheimer

Otro aspecto de investigación a considerar son los biomarcadores del cuerpo [3][12] los cuales han permitido recolectar mas información acerca de como el cuerpo se comporta, como obtener datos de el y obtener información valiosa de los mismos. Con el desarrollo tecnológico y la mayor capacidad de procesamiento que nace de la era de la información, es muy común en la actualidad implementar modelos estadísticos, como los modelos de probit y los modelos de regresión. Con la acumulación de datos valiosos también se le empezó a dar mas valor a las redes neuronales, en particular a las redes neuronales recurrentes y convolucionales, ya que pueden complementar fácilmente los sistemas de diagnóstico tradicionales y también nos permite extraer y extrapolar mas información que no se había observado antes por los ojos humanos lo cual ayuda a diagnosticar la enfermedad, investigarla y desarrollar fármacos, los cuales han demostrado ser efectivos para la prevención y protección de los pacientes[1] lo cual demuestra el valor de estos últimos avances.

En lo que respecta particularmente a la Inteligencia artificial, los transformers y el desarrollo de los modelos con lenguaje natural, son los que mas facilidades han tenido para incorporarse en el campo de la medicina debido a nuestra experiencia con técnicas como la auscultación y con el procesamiento de big data dándonos mucho con lo que trabajar[7] ya que la inteligencia artificial multimodal permite procesar todo tipo de datos, tanto imágenes, como audio y texto, lo cual es ideal para combinar diferentes test que se realizan para diagnosticar el Alzheimer y así también poder estudiar correlaciones biológicas que han pasado desapercibidas hasta ahora

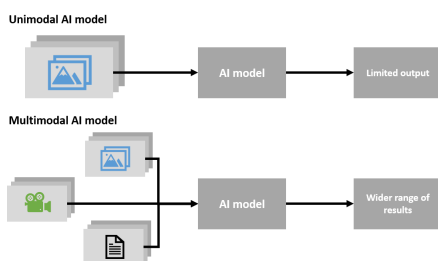


Figura 4. IA multimodal

Para observar mas a profundidad a continuación estudiaremos brevemente 5 casos y las herramientas que utilizaron para lograr realizar sus respectivas investigaciones

1) Use of Deep Neural Networks to Predict Obesity With Short Audio Recordings: Development and Usability Study El objetivo del estudio fue poder predecir si una persona poseía o no obesidad con la ayuda de un audio, el estudio se postulo de la siguiente manera:

- **Consideraciones éticas:** Se presento el estudio al comité ético de la universidad de Shanghai y se obtuvo en consentimiento de cada uno de los participantes
- **Recolección de Datos:** Se recolectaron los datos con un encuesta web donde se le solicitaban a los participantes información y se les solicitaba enviar audios, se obtuvieron un total de 696 participantes, entre los cuales 500 eran hombres con una edad promedio de 24 años y aproximadamente 400 participantes no poseían obesidad
- **Procesamiento de los datos:** Los audios fueron segmentados y transformados en espectrogramas, se porcionaron los datos en 592 espectrogramas para el entrenamiento y 105 para la validación de los datos, finalmente se utilizo un 5-fold cross-validation durante el entrenamiento
- **Modelo:** Se adapto el Framework YOLO especializado en procesamiento de imagen, utilizando batch normalization, learning rate optimization y early stopping para optimizar y mejorar el rendimiento de la IA, en comparación también se utilizo una red neuronal convolucional entrenada con datos recolectados con MFCC (un método para procesamiento de audio que extrae información acústica y armónica importante)
- **Conclusiones:** El modelo logro una precisión del 70 % a nivel global, no obstante obtuvo una mayor precisión para detectar patrones de no obesidad, por lo cual el modelo necesita explorar mas alternativas de recolección de datos, probar con otras técnicas de procesamiento y estudiar mas a profundidad el tema para lograr aplicar con mayor eficacia las redes neuronales al estudio de la obesidad y la prevención y diagnóstico de la misma

2) Technology COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings Durante la pandemia del Covid y la necesidad de diagnosticar rapidamente a las personas ante la gravedad de la situación, se realizo un modelo buscando explorar las posibilidades de diagnosticar covid con información auditiva recolectada a través de audios

- **Consideraciones éticas:** El estudio fue realizado considerando los parametros éticos de la IEEE y en específico la EMB
- **Recolección de Datos:** Se creo un sitio web donde se recolecto información y muestras acerca de personas con Covid y sin covid con un total de 5230 muestras, con la mitad siendo pacientes positivos para covid y la otra mitad negativos, teniendo un set balanceado en todos los parametros
- **Procesamiento de los datos:** Para el procesamiento de los datos se dividió el audio en 6 segmentos y se utilizo el MFCC para lograr obtener información valiosa acerca de los datos, para posteriormente procesar aun mas los datos con una máscara de Poisson, se utilizo el 80% de los datos para el entrenamiento y el 20% para la validación del modelo
- **Modelo:** El modelo se baso en un esquema de procesamiento por partes donde se utilizaron redes neuronales ResNet50, se utilizaron 4 redes, 3 para procesar y caracterizar diferentes características del audio (la salud de las cuerdas vocales, el tracto respiratorio y los sentimientos) pasar los resultados de esas redes neuronales a una capa de pooling sobre las cuales finalmente, pasaran sus resultados a una última shallow network densa que finalmente daría

el resultado, se utilizaron datos tambien de Librispeech para pre entrenar las redes para procesamiento de audio y posteriormente se terminaron de entrenar con los datos recolectados

- **Conclusiones:** El modelo fue un exito ya que alcanzo una precision del 98.5 %, evitando los falsos positivos, lo cual puede ayudar a diagnosticar la enfermedad facilmente y sin un costo elevado, demostrando el potencial de la inteligencia artificial para el diagnostico de enfermedades

3) On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis

El diagnostico de alzheimer es muy costoso e invasivo y un diagnostico temprano es importante para el cuidado de los pacientes por lo cual se busca explorar el uso de la IA para abaratar los costos del diagnostico de la enfermedad

- **Consideraciones eticas:** Se solicito a los participantes el consentimiento y tambien se cumplio con las medidas de proteccion de datos y eticas exigidas por cada entidad que brindo ayuda o informacion para el desarrollo del proyecto
- **Recoleccion de Datos:** Se utilizaron un total de 70 muestras de videos de 8 a 12 horas interactuando con pacientes en una conversacion amistosa en un ambiente amistoso y familiar, todo esto considerando la forma de actuar y de hablar de las personas con alzheimer, 50 de ellos eran sanos y 20 tenian alzheimer
- **Procesamiento de los datos:** Se recolecto el audio de los videos y se removieron todos los eventos no analizables, dejando como utilizable el 50% y 80% de los datos obtenidos, los primeros de los pacientes con Alzheimer y los segundos los pacientes sanos, los audios se dividieron en segmentos de 60 segundos terminando con un total de 600 segmentos de audio, se utilizo ASSA para obtener informacion auditiva de los audios, se estudio tambien la posibilidad de estudiar los factores emocionales y la dimension fractal Higuchi
- **Modelo:** Se utilizo un Perceptron multicapa con una capa oculta de 100 neuronas y 1000 pasos de entrenamiento junto a Cross validation para lograr optimizar el modelo, el modelo fue alimentado con características extraídas de los datos utilizando la informacion emocional, los ratios armonicos y en general las características esenciales acusticas
- **Conclusiones:** El modelo fue un exito ya que se logro una precision del 95 %, seguramente debido al debido procesamiento y analisis profundo de las características auditivas a estudiar, el modelo es una muestra de como el conocimiento especializado y los equipos interdisciplinarios pueden lograr y optimizar el adecuado uso de la Inteligencia Artificial en la medicina

4) Artificial Intelligence-Enabled End-To-End Detection and Assessment of Alzheimer's Disease Using Voice

La Inteligencia Artificial presenta muchas ayudas que pueden complementar el diagnostico de enfermedades, el objetivo de este proyecto fue estudiar el uso de las redes pre entrenadas para diagnosticar alzheimer

- **Consideraciones eticas:** Se usaron las consideraciones eticas de la Universidad de Drexel en Filadelfia, USA y se utilizaron datos solicitados de la base de datos de ADReSSo

- **Recoleccion de Datos:** Utilizando los Datos de ADReSSo se recolectaron un total de 247 muestras de audio de las cuales un poco mas de la mitad tenian alzheimer y el resto eran personas sanas
- **Procesamiento de los datos:** Se utilizaron el 70 % de los datos durante el entrenamiento y el 30 % restante se utilizo la validacion, se utilizo Data2Vec para procesar los audios y la informacion extra que contenia la base de datos como el valor de el MMSE para generar vectores con los cuales entrenar el modelo, tambien se utilizo para comparar modelos wav2vec2 utilizando solamente los datos auditivos
- **Modelo:** Se utilizo la libreria de Hugging face para redes neuronales para procesar los datos junto a Cross validation
- **Conclusiones:** Se logro una precision del 73 % utilizando data2vec y un 72 % utilizando wav2vec2, para lograr diagnosticar la enfermedad parece ser que es necesario una mayor cantidad de datos y preprocesamiento de los datos, no obstante el modelo presenta resultados prometedores

5) A deep learning approach to dysphagia-aspiration detecting algorithm through pre- and post-swallowing voice changes

Los costos de diagnosticar Disfagia son muy elevados e incomodos para la mayoria de pacientes por lo cual se busca lograr metodos de diagnostico que sean mas faciles para los pacientes y pueda alcanzar a mas personas por lo cual se busca implementar un modelo que utilice el audio y lo procese rapidamente para obtener los resultados de diagnostico

- **Consideraciones eticas:** Se realizo todo bajo las consideraciones eticas y solicitadas del Hospital Bundang de la Universidad Nacional de Seoul y pidiendo el consentimiento de los pacientes
- **Recoleccion de Datos:** Se recolectaron los datos de 198 pacientes en el hospital entre octubre de 2021 y febrero de 2023 recolectando informacion auditiva antes de consumir alimentos y despues de consumirlos y se utilizaron diferentes dispositivos de recoleccion de audio para evitar sesgos particulares del medio de recoleccion
- **Procesamiento de los datos:** Los audios se limpiaron quitando el ruido, se dividieron en audios de 2 segundos y se procesaron usando las frecuencias de Mel, y varios procedimientos de audio entre los cuales estaba la transformacion de Fourier de tiempo reducido, se hicieron combinaciones de los fragmentos de antes y despues de consumir alimentos para generar aproximadamente 8000 muestras
- **Modelo:** Se utilizaron redes neuronales preentrenadas con PyTorch, la red fue alimentada con espectrogramas de Mel, se realizaron 3 modelos uno para pacientes femeninos, uno para pacientes masculinos y el otro mixto
- **Conclusiones:** El proposito de esta red fue utilizarla como apoyo al diagnostico de la enfermedad y no como reemplazo, lo cual nos ayuda a reducir los costos de diagnostico en las instituciones, el modelo fue relativamente exitoso con una precision del 81 %

3. METODOLOGIA

Para generar un buen modelo es necesario tener un adecuado procedimiento de recoleccion y analisis de los datos por lo cual es importante comprender y estudiar los alcances de nuestras

herramientas. Necesitamos observar que informacion tenemos y cual es posible obtener, en las implementaciones de las ultimas decadas principalmente se han usado bases de datos cerradas con datos diagnosticos y pruebas de audio[6] por lo cual se revisara la disponibilidad de esta informacion y se trabaja con la informacion valiosa recolectada dentro del marco de las posibilidades, ya que la informacion es nuestro mayor aliado para lograr nuestra tarea, ya que sin buenos datos no es posible lograr buenos modelos[10]

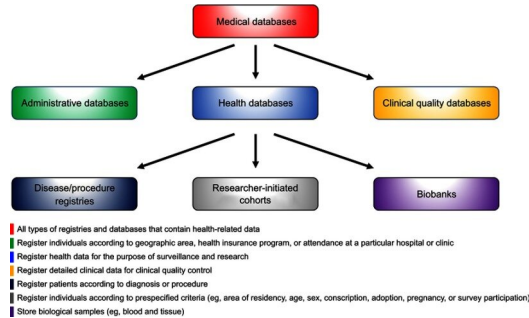


Figura 5. Bases de datos medicas

Para esto se estudiara y buscaran las librerias necesarias, los datos donde entrenar el modelo y se realizaran los respectivos analisis, se realizara el respectivo estudio de los datos

4. ESTRUCTURA DEL PROYECTO

Enfoque y Tecnicas

Este proyecto tiene un enfasis en el estudio del alzheimer, la recoleccion de datos y la estructuracion de un prototipo del modelo, implementando inteligencia artificial para lograr obtener un modelo que con un audio pueda determinar si alguien posee o no alzheimer **Herramientas** Para el proyecto se va a hacer uso de las siguientes Herramientas y Tecnologias:

Github: Donde se Hosteara el acceso al codigo y el Articulo <https://github.com/HanamDavid/AFDA.git>

Google Colab: Para lograr manipular el codigo y procesar los datos y generar el modelo

https://drive.google.com/drive/folders/1NKC_4UG0Hdzff4SLvLF7ZKIxrg4uJIW?usp=drive_link

Librerias:

- Tensorflow: Para implementar los algoritmos de machine learning necesario y poder clasificar los datos correctamente y generar procesos utiles
- Sklearn: Para poder dividir los datos del entrenamiento
- Librosa: Pra procesar la informacion de los audios y poder normalizar los datos lo cual nos permitira optimizar y mejorar el rendimiento de la red neuronal y en general con la limpieza de datos
- Matplotlib: Para Visualizar la informacion y asi poder tener una forma de entender los datos con mayor facilidad
- Pandas y Numpy: Para la manipulacion de las Series y los Datos una vez abstraídos de las muestras de Audio

Se utilizo la base de datos **DementiaNet** la cual incluye una serie de datos medicos de pacientes con demencia y pacientes sanos recolectados especificamente para ML <https://github.com/shreyasgite/dementianet>

#	name	dementia type	birth	death	first symptom URL's after symptoms	5 years	5 < 10 years	10 < 15 years
1	Alce Burnes	Alzheimer	1910	1985	1975	https://www.youtube.com/watch?v=3j5p8p8C8e4		
2	Alcen Hernandez	Alzheimer	1926	2017	2012	https://youtu.be/Ca5G2k-mkYk		
3	Alan Ramsey	Dementia	1928	2020	2015	https://www.youtube.com/watch?v=C2K4K-Au4dQ		
4	Alan Bums	Alzheimer	1935	2011		https://www.youtube.com/watch?v=3j5p8p8C8e4		
5	Andrew Sachs	Dementia	1930	2018	2012	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4
6	Arnette McClinton	Dementia	1922	2018	2013	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4
7	Arney Fine	Dementia	1923	2010	2009	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4	https://youtu.be/3j5p8p8C8e4
8	B B King	Dementia	1925	2015	2010	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
9	Benny Carter	Alzheimer	1929	2017	2012	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
10	Bert Brackley	Alzheimer	1921	2014	2009	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
11	Bill Buckner	Alzheimer	1949	2019		https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
12	Bobbi Weisbach	Alzheimer	1944	2014	2013	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
13	Cathy Kasem	Alzheimer	1932	2014	2007	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
14	Charles Boncomp	Alzheimer	1921	2010		https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
15	Charles K Khan	Alzheimer	1933	2018	2004	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
16	Charmian Carr	Dementia	1942	2016	2006	https://youtu.be/3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
17	Don Hagan	Alzheimer	1934	2019	2013	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
18	David Maltch	Alzheimer	1945	present	2015	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
19	David Prosser	Dementia	1935	2020	2013	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
20	Deena Moore	Alzheimer	1945	2011	2013	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
21	Don Lane	Alzheimer	1933	2009	2003	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4
22	Donald Sterling	Alzheimer	1934	present	2014	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4	https://www.youtube.com/watch?v=3j5p8p8C8e4

Figura 6. Datos originales

Esta base de datos contiene un total de 359 muestras de audios recolectados de internet de personas famosas entre las cuales 131 tiene demencia, los audios poseen distintas duraciones y la hay personas que llevan poco tiempo diagnosticadas como pacientes con mas de decada y media con la enfermedad, los datos fueron descargados y procesados en local para luego ser subidos a google drive

Procesamiento de Datos

Una vez con los datos descargados, debido a que la base de datos tiene muestras de personas con otras enfermedades, se eliminaron mas de la mitad de los datos de las muestras de pacientes enfermos y terminamos con un aproximado de 54 pacientes con alzheimer, posteriormente para compensar la perdida de datos en el set de pacientes con alzheimer y permitir un procesamiento de datos mas rapido, los audios de alzheimer se recortaron en 4 segmentos de 4 segundos cada uno, y los de los pacientes saludables en 2 segmentos de 4 segundos y se limpio el csv para guardar solo los datos relevantes, terminando con un total de 202 muestras de pacientes sanos y 202 de pacientes con alzheimer

```

[ ] #1 alzheimer
#0 no alzheimer
#veremos nuestro csv que contiene el path de cada archivo junto a si posee o no alzheimer, tenemos 202 muestras con alzheimer y 202 sin alzheimer
data=pd.read_csv('content/drive/My Drive/AFDA/data/alz_data.csv')
data.head()
data["alzheimer"].value_counts()

```

	count
alzheimer	
1	202
0	202

dtype: int64

Figura 7. Datos Limpios

Vampiro_5_segment3.wav
 Vampiro_5_segment2.wav
 Vampiro_5_segment1.wav
 Vampiro_0_segment4.wav
 Vampiro_0_segment3.wav
 Vampiro_0_segment2.wav
 Vampiro_0_segment1.wav
 TrevorPeacock_5_segment4.wav
 TrevorPeacock_5_segment2.wav
 TrevorPeacock_5_segment1.wav
 TonyParkes_10_segment4.wav
 TonyParkes_10_segment3.wav
 TonyParkes_10_segment2.wav

Figura 8. Segmentos de Datos

Posteriormente ya con las carpetas y el csv, se intercambiaron de posicion cada una de las muestras para facilitar el entrenamiento al aleatorizar el orden de las muestras. Se utilizo librosa para extraer las caracteristicas esenciales de cada uno de los audios con MFCC ya que MFCC permite extraer datos vitales para obtener informacion auditiva y armonica de manera mas facil y suele ser usada en procesamiento de audio

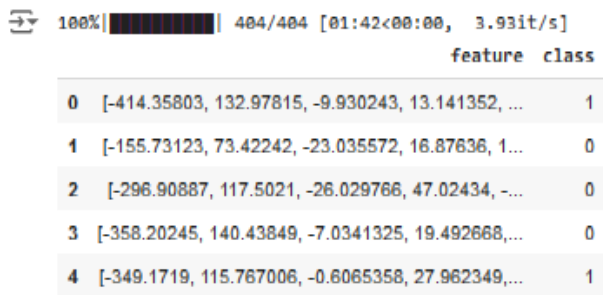


Figura 9. Datos de MFCC

y se guardaron los datos en un vector para el entrenamiento finalmente dividiendo los sets de entrenamiento y de validacion con sklearn, se utilizo el 70% de los datos para el entrenamiento de la red

Modelo

Para el modelo se uso Tensorflow se realizaron varios modelos pero tras la evaluacion de redes normales, recurrentes y convolucionales, se determino elegir la convolucional dada la capacidad de estas redes para explorar patrones dificilmente observables en los datos y una capa recurrente para el procesamiento de la ultima capa. El modelo fue el siguiente:

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.layers import Conv2D, Conv1D, MaxPooling1D, BatchNormalization, Bidirectional, LSTM, Dense, Dropout, Flatten, Add, Input
from tensorflow.keras.regularizers import l2
from tensorflow.keras.metrics import Precision, Recall, AUC
from sklearn.metrics import confusion_matrix

model = Sequential()

#Primera capa convolucional
model.add(Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(48, 1)))
model.add(MaxPooling1D(pool_size=2))
model.add(BatchNormalization())

#Segunda capa convolucional
model.add(Conv1D(filters=128, kernel_size=3, activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(BatchNormalization())

#Tercera capa convolucional
model.add(Conv1D(filters=256, kernel_size=3, activation='relu', kernel_regularizer=l2(0.001)))
model.add(MaxPooling1D(pool_size=2))
model.add(BatchNormalization())

#Cuarta capa convolucional
model.add(Conv1D(filters=512, kernel_size=3, activation='relu', kernel_regularizer=l2(0.001)))
model.add(MaxPooling1D(pool_size=2))
model.add(BatchNormalization())

#Capa LSTM
model.add(LSTM(100, return_sequences=True))
model.add(Dropout(0.5))

model.add(Flatten())
model.add(Dense(num_labels, activation='sigmoid'))
model.summary()

# Métricas
model.compile(
    loss='categorical_crossentropy',
    optimizer=Adam(),
    metrics=['accuracy', Precision(), Recall(), AUC()])
```

Figura 10. Configuración del modelo

Layer (type)	Output Shape	Param #
conv1d_14 (Conv1D)	(None, 58, 64)	256
max_pooling1d_12 (MaxPooling1D)	(None, 29, 64)	0
batch_normalization_12 (BatchNormalization)	(None, 29, 64)	256
conv1d_15 (Conv1D)	(None, 27, 128)	24,704
max_pooling1d_13 (MaxPooling1D)	(None, 13, 128)	0
batch_normalization_13 (BatchNormalization)	(None, 13, 128)	512
conv1d_16 (Conv1D)	(None, 11, 256)	98,560
max_pooling1d_14 (MaxPooling1D)	(None, 5, 256)	0
batch_normalization_14 (BatchNormalization)	(None, 5, 256)	1,024
conv1d_17 (Conv1D)	(None, 3, 512)	393,728
max_pooling1d_15 (MaxPooling1D)	(None, 1, 512)	0
batch_normalization_15 (BatchNormalization)	(None, 1, 512)	2,048
lstm_5 (LSTM)	(None, 1, 100)	245,200
dropout_9 (Dropout)	(None, 1, 100)	0
flatten_7 (Flatten)	(None, 100)	0
dense_9 (Dense)	(None, 2)	202

Total params: 766,490 (2.92 MB)
Trainable params: 764,570 (2.92 MB)
Non-trainable params: 1,920 (7.50 KB)

Figura 11. Resumen del Modelo

```
from tensorflow.keras.metrics import Precision, Recall, AUC
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.callbacks import ReduceLROnPlateau, EarlyStopping
from datetime import datetime

# funciones de Callback
num_epochs = 200
num_batch_size = 16

checkpointer = ModelCheckpoint(filepath='saved_models/audio_classification.keras',
                               verbose=1, save_best_only=True)

early_stopping = EarlyStopping(monitor='val_loss', patience=40, restore_best_weights=True)

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=20, min_lr=0.0001)

# Entrenando el modelo
start = datetime.now()

history = model.fit(
    X_train, y_train,
    batch_size=num_batch_size,
    epochs=num_epochs,
    validation_data=(X_test, y_test),
    callbacks=[checkpointer, early_stopping, reduce_lr],
    verbose=1
)

# Calculando el tiempo de entrenamiento
duration = datetime.now() - start
print("Training completed in time: ", duration)
```

Figura 12. Configuración Entrenamiento

En el modelo se utilizo

- Conv1D: Redes convolucionales unidimensionales, los cuales se suelen usar para procesar datos secuenciales, se escogieron debido a la naturaleza de nuestro problema que requiere buscar características especiales dentro de la voz de los pacientes
- LSTM: Capas de Long Short-Term Memory, que permiten entender la información secuencial de los datos. Después de extraer las características fundamentales del audio con MFCC y las capas convolucionales, la LSTM proporciona información adicional al capturar dependencias a largo plazo y mitigar el desvanecimiento del gradiente.
- BatchNormalization: Esta técnica mejora la eficacia y rapidez del procesamiento entre capas al normalizar las

activaciones, actuando también como una forma de regularización que puede mejorar la generalización del modelo.

- Pooling: El cual reduce la cantidad de información de la red que reciben las neuronas, aislando los valores más relevantes y reduciendo la dimensionalidad de los datos
- ModelCheckpoint: Una técnica que durante el entrenamiento nos permite guardar la información de los mejores modelos durante el entrenamiento
- Early Stopping: Una técnica que nos permite parar el entrenamiento anticipadamente si no se obtienen mejoras durante un tiempo, lo cual optimiza el rendimiento del entrenamiento y ayuda al cómputo del modelo
- ReduceLROnPlateau: Esta técnica ajusta la tasa de aprendizaje cuando no se observan mejoras durante un período específico, lo que permite que el modelo se adapte mejor a la superficie de pérdida y mejore la convergencia.

En nuestro modelo se combinan capas convolucionales y LSTM para capturar tanto características locales como patrones temporales en datos de audio y buscamos una buena capacidad de parámetros para poder explorar incluso patrones difíciles de percibir, junto a un análisis de las tecnologías y técnicas usadas en los proyectos anteriormente observados

5. DISCUSION DE RESULTADOS

Los resultados fueron los siguientes:

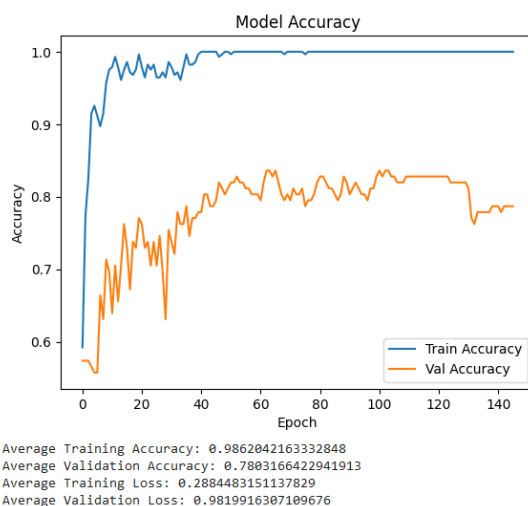


Figura 13. Precision del Modelo 1

y la siguiente matriz de confusión:

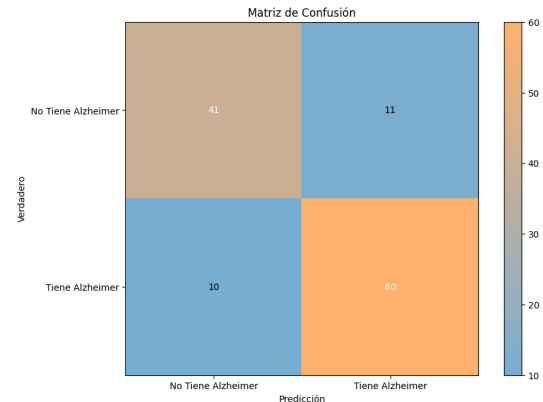


Figura 14. Matriz

Los Datos obtenidos fueron :

- 95 % precision de training: lo cual indica que las características de los datos observados durante el entrenamiento fueron aprendidas correctamente
- 78 % precision de validacion: Mas baja que durante el entrenamiento lo cual podria indicar overfitting, no obstante fue un resultado bastante bueno y puede mejorarse con posteriores configuraciones del modelo y una mayor cantidad de datos
- 0.28 Loss de training: La cual indica que las predicciones del modelo podrian ser buenas, no obstante al compararse con la de validacion, podemos observar que hay overfitting
- 0.98 Loss de validacion
- 0.87 AUC de training: La cual indica que hay una buena capacidad de discriminar entre clases en el entrenamiento
- 0.89 AUC de validacion La cual indica que hay una buena capacidad de discriminar entre clases en la validacion y por lo tanto en general en el modelo

Si comparamos con los proyectos antes estudiados, podemos observar que en general se obtuvo un muy buen resultado dada la poca cantidad de datos, lo cual demuestra la capacidad de las redes convolucionales y de MFCC para discriminar enfermedades, se obtuvo una precision similar a la del proyecto 1 y 4, no obstante en lo que respecta a otros proyectos aun hay mucho que mejorar para poder hacer mejores modelos.

En particular es importante resaltar la calidad de la información recolectada, ya que los proyectos con mayor éxito poseían una gran calidad de datos recolectados especialmente para el proyecto, mientras que en el caso de AFDA los audios fueron recolectados de internet y la cantidad de ruido y de información auditiva innecesaria puede perjudicar la discriminación de la enfermedad. En el proyecto se combinaron varias de las técnicas que se usaron en los proyectos anteriormente mencionados buscando llegar a las mejores herramientas para solucionar el problema, los resultados son prometedores, no obstante es importante estudiar con mayor detenimiento las particularidades fisiológicas y neurológicas del Alzheimer y cómo afecta la voz, para poder darle un mejor procesamiento a los datos antes de alimentarlos en la red y también para hacer un mejor uso de los hiperparámetros, obteniendo mayores niveles de precisión y evitar también el overfitting

6. CONCLUSIONES

El aumento de la capacidad medica para intervenir las enfermedades cambia, se estructura y fomenta buenas practicas medicas dentro de la sociedad. El desarrollo tecnologico es esencial para proteger y cuidar a la poblacion, principalmente a la de alto riesgo, en el caso del alzheimer, cuanto mayor sea la implementacion e investigacion de IA en medicina mayor facilidad abra para mejorar la vida de los pacientes. Esto por supuesto tiene su costo, una gran cantidad de investigacion, un deseo de la sociedad por adoptarlo y una gran cantidad de datos para proporcionarle la informacion adecuada a los modelos. Con el tiempo, la IA toma cada vez mas espacio en nuestra sociedad y al menos en la medicina da buenos prospectos.

Por supuesto es necesario primero priorizar la obtencion moral y etica de datos y de una buena implementacion de sistemas de IoT para poder recolectar estos datos, ya que sin estos podemos tener modelos con la insuficiente capacidad para detectar la enfermedad o detectarla mal lo cual puede poder en riesgo la vida de muchos pacientes, sin una estructura rigida y moral que sostenga la inteligencia artificial en la medicina no es posible lograr que los modelos sean exitosos y que generen el impacto buscado si no que por el contrario generen falsos positivos, desconfianza de la tecnologia y en el peor de los casos la muerte de muchos.

Por ultimo es importante detallar que los avances en Inteligencia artificial brillan sobretodo cuando apoyan a los sistemas tradicionales de investigacion y de analisis ya que estos sistemas permiten extrapolar datos y complementar los tratamientos solamente cuando se entiende adecuadamente las circunstancias donde es aplicable, por lo cual es necesario que todo avance este siendo revisado con equipos interdisciplinarios para asi poder explotar el uso de la tecnologia en la comunidad cientifica y en esto caso en la comunidad medica.

Los proyectos observados y la implementacion de este nos permitieron observar ese potencial, es importante, seguir apoyando y explorando el uso de Inteligencia Artificial en el medio y asi poder mejorar la calidad de vida de miles de personas a lo largo del mundo

7. REFERENCIAS BIBLIOGRAFICAS

REFERENCIAS

- [1] "Alzheimer's disease: early diagnosis and treatment," HKMJ, Jan. 17 2015. <https://www.hkmj.org/abstracts/v18n3/228.htm>
- [2] R. L. Ernst and J. W. Hay, "The US economic and social costs of Alzheimer's disease revisited.," American Journal of Public Health, vol. 84, no. 8, pp. 1261–1264, Aug. 1994, doi: 10.2105/ajph.84.8.1261.
- [3] K. Qian, B. Hu, Y. Yamamoto, and B. W. Schuller, "The Voice of the Body: Why AI Should Listen to It and an archive," Cyborg and Bionic Systems, vol. 4, Jan. 2023, doi: 10.34133/cbsystems.0005.
- [4] F. Agbavor and H. Liang, "Artificial Intelligence-Enabled End-To-End detection and assessment of Alzheimer's disease using Voice," Brain Sciences, vol. 13, no. 1, p. 28, Dec. 2022, doi: 10.3390/brainsci13010028.
- [5] H. Lin et al., "Identification of digital voice biomarkers for cognitive health," Exploration of Medicine, vol. 1, no. 6, pp. 406–417, Dec. 2020, doi: 10.37349/emed.2020.00028.
- [6] E. Evangelista et al., "Current Practices in voice data collection and Limitations to Voice AI Research: A national survey," The Laryngoscope, Dec. 2023, doi: 10.1002/lary.31052.
- [7] M. Chen and M. Décary, "Artificial intelligence in healthcare: An essential guide for health leaders," Healthcare Management Forum, vol. 33, no. 1, pp. 10–18, Sep. 2019, doi: 10.1177/0840470419873123.
- [8] A. Baevski, W.-N. Hsu, Q. Xu, A. Babu, J. Gu, and M. Auli, "data2vec: A General Framework for Self-supervised Learning in Speech, Vision and Language," PMLR, Jun. 28, 2022. <https://proceedings.mlr.press/v162/baevski22a.html>
- [9] D. Berrar, "Cross-Validation," in Elsevier eBooks, 2019, pp. 542–545. doi: 10.1016/b978-0-12-809633-8.20349-x.
- [10] Womersley K, Fulford KB, Peile E, Koralus P, Handa A. Hearing the patient's voice in AI-enhanced healthcare. BMJ. 2023 Nov 21;383:2758. doi: 10.1136/bmj.p2758. PMID: 37989502.
- [11] M. Alzubaidi et al., "The role of neural network for the detection of Parkinson's Disease: A scoping review," Healthcare, vol. 9, no. 6, p. 740, Jun. 2021, doi: 10.3390/healthcare9060740.
- [12] J. Laguarda, F. Hueto, P. Rajasekaran, S. E. Sarma, and B. Subirana, "Longitudinal Speech biomarkers for automated Alzheimer's detection," Research Square (Research Square), Aug. 2020, doi: 10.21203/rs.3.rs-56078/v1.
- [13] J. Laguarda, F. Hueto, and B. Subirana, "COVID-19 Artificial intelligence diagnosis using only cough recordings," IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 275–281, Jan. 2020, doi: 10.1109/ojemb.2020.3026928.
- [14] K. L. De Ipiña et al., "On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis," Sensors (Basel), vol. 13, no. 5, pp. 6730–6745, May 2013, doi: 10.3390/s130506730.
- [15] J. Huang, P. Guo, S. Zhang, M. Ji, and R. An, "Use of Deep Neural Networks to Predict Obesity with Short Audio Recordings: A Pilot Study (Preprint)," JMIR AI, Nov. 2023, doi: 10.2196/54885.
- [16] J.-M. Kim, M.-S. Kim, S.-Y. Choi, K. Lee, and J. S. Ryu, "A deep learning approach to dysphagia-aspiration detecting algorithm through pre- and post-swallowing voice changes," Frontiers in Bioengineering and Biotechnology, vol. 12, Aug. 2024, doi: 10.3389/fbioe.2024.1433087.
- [17] UCSF, "AI and Alzheimer's Research," YouTube. May 10, 2022. [Online]. Available: https://www.youtube.com/watch?v=zkvK_iER5tI
- [18] R. L. Ernst and J. W. Hay, "The US economic and social costs of Alzheimer's disease revisited.," American Journal of Public Health, vol. 84, no. 8, pp. 1261–1264, Aug. 1994, doi: 10.2105/ajph.84.8.1261., Fig 1
- [19] *Comunicación y Demencias*. [Online]. Available: <https://comunicacionydemencias.com/test-moca-demencia/>, Fig 2
- [20] *Dementia: Understand Together*. [Online]. Available: <https://dementia.ie/lessons/mri-scan-of-brain-alzheimers-disease/>, Fig 3
- [21] *Compass RAU IAS*. [Online]. Available: <https://compass.rauias.com/current-affairs/multimodal-artificial-intelligence/>, Fig 4
- [22] *ResearchGate*. [Online]. Available: https://www.researchgate.net/figure/Classification-of-medical-databases_fig4_34421419, Fig5