Phonoaudiological analysis for the detection of Alzheimer's disease with AI (AFDA in Spanish)

Juan David Guarnizo Gutierrez, *Computer Science*

Resumen—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. Introduction

What is Alzheimer's disease?

Alzheimer's is a high-risk disease that causes memory loss, depression, lack of ability to manage resources, aphasia, agnosia, apraxia and, in short, the general deterioration of cognitive, language and social skills, which causes people with the disease to depend on their loved ones in essentially everything, generating high levels of stress in themselves and their families. In the country we have new cases every 4 minutes, being a disease with more recurrent cases and among the risk factors we can find high or low alcohol consumption, cigarette consumption and among others, depression or genetic factors [1] risk factors present in most of the elderly population.

Among the challenges we face with Alzheimer's disease, one of the most critical is the diagnosis because the current means of diagnosis are often invasive and costly, and the cost of treatment is very high. Generally the threshold of diagnosis is found after the age of 60 years and this is essential to understand why the use of technology and the advancement of medicine is necessary to take care and protect the lives of thousands of people of the 3rd age and especially in third world countries because the reduction of costs can benefit the number of people who can access a fair and effective medicine, Since as mentioned above, treatments are expensive because they involve the use of drugs, antioxidants and antipsychotics, which makes costs increase as the disease progresses and the lack of access to them attack the quality of life and the fundamental rights of patients.

According to a study in the United States [2], the mere diagnosis of the disease in 91 already presented a series of important costs, and the treatment of the disease presents more costs, however, without considering the increases in inflation, the earlier the disease is detected, the easier it will be to prevent the person from deteriorating, which would improve the quality of life of the person and also facilitate the patient's maintenance by having more control over the disease, It will be much easier to prevent the person from deteriorating, which would improve the quality of life of the person and also facilitate the maintenance of the patient to have more control over his disease, in addition to giving the person the benefit of deciding how to cope with the disease when he still has the mental abilities to do so.

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[a]	47 581
Total discounted (direct cost + unpaid caregiver cost)[a]	123 556
Total discounted (direct cost + unpaid caregiver cost + disability and premature mortality cost)[a]	173 932

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

Figura 1. Costs of having Alzheimer's

Medicine and technology. AI has made great progress in recent years thanks to advances in hardware and development of abstract models, we have managed to generate models that allow us to recognize more and more diseases in the human body through computer vision and image analysis [3], with voice processing or even multimodal analysis [4], because in reality many diseases are already diagnosed by analyzing this type of information and the body gives many signals in the presence of a disease. In neurodegenerative diseases such as Alzheimer's, the use of these technologies is useful because its affectations can be recognized through the patient's speech and language processing, and language is one of the most important indicators of the social, emotional and cognitive state of a person [14]. This is because the speed of processing, the lexicon and in general the communication helps to recognize pathological patterns since a healthy and functional brain is necessary for an adequate use of language [5]. Likewise, the analysis of the voice and body sounds provide us with several pathological clues in medicine, with the study of this in the past have been detected diseases such as laryngeal cancer, subglottic stenopsis or spasmodic torticollis [6]. To achieve the technological implementation within the diagnosis it is necessary to find models that allow finding the appropriate patterns and biomarkers in the language to distinguish each case of Alzheimer's disease and help to give more accurate diagnoses in all ranges of disease progression, This type of modeling can be done more effectively thanks to the latest developments in deep learning and supervised learning [7], which allow us to study and create models with greater focus and importance, which in contrast to MRI, psychological analysis and other traditional tests is quite inexpensive and minimally invasive once the model is solidified.

All this progress has been made thanks to different research groups, Artificial Intelligence and software companies that have generated tools and repositories such as Hugging face, data2vec [8], patient audio and video databases [13] and implementation of other more traditional tools such as linear regression models and optimization algorithms to optimize these models [9]. All this progress will allow us over time to improve these predictive systems and achieve more and more advances in medicine and engineering applied to it, however, as most medical problem and machine learning is necessary a large amount of data, so a higher rate of research and interest by the scientific community in these tools can generate a greater amount of data is obtained and more projects are funded to help the human being, training models is not easy and requires high quality data and taking into consideration other diseases that may pass as false positives if not considered, all this while also keeping in mind the importance of ethics and the human factor in the creation of these models to avoid biases in the predictions and also ensuring the privacy of patient data and understanding the patient's wishes properly, applying artificial intelligence models together with humans to ensure a dignified and humane treatment of the patient [10], since a lack of connection with the medical sector, doctors and patients, these models may advance a lot but never be implemented [11].

2. Background

Diagnosis in Alzheimer's disease has been investigated mainly in the last few decades due to technological advances, a greater understanding of the disease and the prolonged aging of the majority of the population, which has encouraged the development of more methods to help the Alzheimer's population. Conventionally, the MMSE and MoCa [4][15] have been used to help diagnose through cognitive testing whether or not the patient has Alzheimer's. These tests can help us extract information about how and under what circumstances we can obtain data to determine the patient's diagnosis, it is also possible to resort to scans (MRI) but these are much more invasive and expensive for patients, which encouraged the medical community to start using other means to recognize diseases.

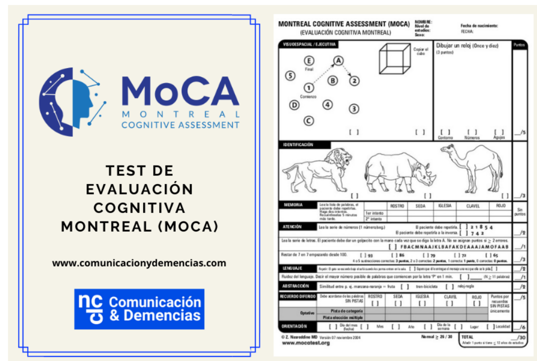


Figura 2. MoCA

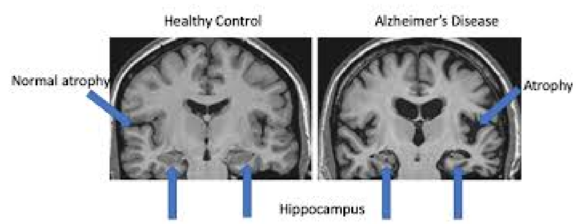


Figura 3. MRI Alzheimer's

Another aspect of research to consider is body biomarkers [3][12] which have made it possible to collect more information about how the body behaves, how to obtain data from it and gain valuable information from it. With the technological development and the increased processing power that comes from the information age, it is very common nowadays to implement statistical models, such as probit models and regression models. With the accumulation of valuable data, neural networks, in particular recurrent and convolutional neural networks, have also become more valuable as they can easily complement traditional diagnostic systems and also allow us to extract and extrapolate more information not previously observed by human eyes which helps in disease diagnosis, research and drug development, which have proven to be effective for prevention and protection of patients[1] which demonstrates the value of these latest advances.

With particular regard to artificial intelligence, transformers and the development of natural language models have been the easiest to incorporate in the field of medicine due to our experience with techniques such as auscultation and big data processing, giving us a lot to work with[7] since multimodal artificial intelligence allows us to process all types of data, images, audio and text, which is ideal for combining different tests that are performed to diagnose Alzheimer's and also to study biological correlations that have gone unnoticed until now, images, audio and text, which is ideal for combining different tests that are used to diagnose

Alzheimer's disease and also to study biological correlations that have gone unnoticed until now.

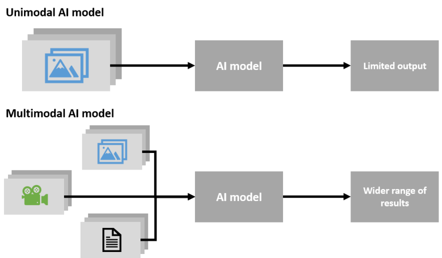


Figura 4. multimodal AI

For a more in-depth look at 5 cases and the tools they used to carry out their respective investigations, we will briefly review them below

1) Use of Deep Neural Networks to Predict Obesity With Short Audio Recordings: Development and Usability Study The objective of the study was to predict whether or not a person was obese with the help of audio, the study was postulated as follows:

- **Ethical Considerations:** The study was submitted to the ethics committee of Shanghai University and consent was obtained from each of the participants.
- **Data collection.** Data were collected with a web survey where participants were asked for information and asked to submit audios, a total of 696 participants were obtained, among which 500 were male with an average age of 24 years old and approximately 400 participants did not have obesity.
- **Data processing:** The audios were segmented and transformed into spectrograms, the data were portioned into 592 spectrograms for training and 105 for data validation, finally a 5-fold cross-validation was used during training.
- **Model:** We adapted the YOLO Framework specialized in image processing, using batch normalization, learning rate optimization and early stopping to optimize and improve the AI performance, in comparison we also used a convolutional neural network trained with data collected with MFCC (a method for audio processing that extracts important acoustic and harmonic information).
- **Conclusions:** The model achieved an accuracy of 70 % globally, however it obtained a higher accuracy to detect patterns of non-obesity, so the model needs to explore more alternatives for data collection, try other processing techniques and study more in depth the subject to achieve more effectively apply neural networks to the study of obesity and the prevention and diagnosis of obesity.

2) Technology COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings During the Covid pandemic and the need to diagnose people quickly given the severity of the situation, a model was developed to explore the possibilities of diagnosing Covid with auditory information collected through audios.

- **Ethical considerations.** The study was conducted considering the ethical parameters of the IEEE and specifically the EMB.
- **Data Collection:** A website was created where information and samples were collected about people with and without covid with a total of 5230 samples, with half being covid positive patients and the other half covid negative, having a balanced set in all parameters.
- **Data processing:** For data processing, the audio was divided into 6 segments and the MFCC was used to obtain valuable information about the data, and then the data was further processed with a Poisson mask, 80 % of the data was used for training and 20 % for model validation.
- **Model:** The model was based on a piecewise processing scheme where ResNet50 neural networks were used, 4 networks were used, 3 to process and characterize different characteristics of the audio (vocal cord health, respiratory tract and sentiments) passing the results of those neural networks to a pooling layer on which they will finally pass their results to a last dense shallow network that will finally give the result, The results will be passed to a final dense shallow network that will finally give the result, Librispeech data was also used to pre-train the networks for audio processing and then finished training with the collected data.
- **Conclusions.** The model was a success as it achieved an accuracy of 98.5 %, avoiding false positives, which can help diagnose the disease easily and without high cost, demonstrating the potential of artificial intelligence for disease diagnosis.

3) On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis The diagnosis of Alzheimer's is very expensive and invasive and early diagnosis is important for patient care, so we are looking to explore the use of AI to lower the cost of diagnosing the disease.

- **Ethical considerations:**Participants were asked for their consent and also complied with the data protection and ethical measures required by each entity that provided assistance or information for the development of the project.
- **Data Collection:** A total of 70 video samples were used from 8 to 12 hours interacting with patients in a friendly conversation in a friendly and familiar environment, all this considering the way of acting and talking of people with Alzheimer's disease, 50 of them were healthy and 20 had Alzheimer's disease.
- **Data processing:** We collected the audio of the videos and removed all the non-analyzable events, leaving as usable 50 % and 80 % of the data obtained, the first from Alzheimer's patients and the second from healthy patients, the audios were divided into segments of 60 seconds ending with a total of 600 audio segments, ASSA was used to obtain auditory information from the audios, we also studied the possibility of studying the emotional factors and the Higuchi fractal dimension.The diagnosis of Alzheimer's disease is very costly and invasive and early diagnosis is important for the care of patients, so we sought to explore the use of the model.
- **Model:** A multilayer Perceptron with a hidden each of 100 neurons and 1000 training steps together with Cross validation was used to optimize the model, the model was fed with features extracted from the data using emotional

information, harmonic ratios and in general the essential acoustic features.

- **Conclusions:** The model was a success as it achieved an accuracy of 95 %, probably due to the deep processing and analysis of the auditory characteristics to be studied, the model is an example of how specialized knowledge and interdisciplinary teams can achieve and optimize the proper use of Artificial Intelligence in medicine.

4) Artificial Intelligence-Enabled End-To-End Detection and Assessment of Alzheimer's Disease Using Voice Artificial Intelligence presents many aids that can complement the diagnosis of diseases, the objective of this project was to study the use of pre-trained networks to diagnose Alzheimer's disease.

- **Ethical considerations:** We used ethical considerations from Drexel University in Philadelphia, USA and used data requested from the ADReSSo database.
- **Data Collection:** Using the ADReSSo data, a total of 247 audio samples were collected, of which slightly more than half had Alzheimer's disease and the rest were healthy people.
- **Data processing:** 70 % of the data were used during training and the remaining 30 % were used for validation, Data2Vec was used to process the audios and the extra information contained in the database such as the MMSE value to generate vectors with which to train the model, it was also used to compare wav2vec2 models using only the auditory data.
- **Model:** The Hugging face library for neural networks was used to process the data together with Cross validation.
- **Conclusions:**An accuracy of 73 % was achieved using data2vec and 72 % using wav2vec2, to diagnose the disease it seems that a larger amount of data and preprocessing of the data is necessary, however the model shows promising results.

5) A deep learning approach to dysphagia-aspiration detecting algorithm through pre- and post-swallowing voice changes The costs of diagnosing dysphagia are very high and uncomfortable for most patients, so we are looking for diagnostic methods that are easier for patients and can reach more people, so we are looking to implement a model that uses audio and process it quickly to obtain diagnostic results.

- **Ethical considerations:** Everything was done under the ethical considerations and requested by Bundang Hospital, Seoul National University and asked for the consent of the patients.
- **Data Collection:** Data were collected from 198 patients in the hospital between October 2021 and February 2023 by collecting auditory information before food consumption and after food consumption and different audio collection devices were used to avoid particular biases of the collection medium.
- **Data processing:**The audios were cleaned by removing noise, split into 2-second audios and processed using Mel's frequencies, and various audio procedures including reduced time Fourier transformation, and combinations of the before and after food consumption fragments were made to generate approximately 8000 samples.
- **Model:** Pretrained neural networks were used with PyTorch, the network was fed with spectrograms of Mel, 3 models were made, one for female patients, one for male patients and the other mixed.
- **Conclusions:** The purpose of this network was to use it as a support for the diagnosis of the disease and not as a replacement, which helps us to reduce the costs of diagnosis in the institutions, the model was relatively successful with an accuracy of 81 %.

3. Methodology

To generate a good model it is necessary to have an adequate data collection and analysis procedure, so it is important to understand and study the scope of our tools. We need to observe what information we have and what is possible to obtain, in the implementations of the last decades we have mainly used closed databases with diagnostic data and audio tests[6] so we will review the availability of this information and work with the valuable information collected within the framework of the possibilities, since the information is our greatest ally to achieve our task, since without good data it is not possible to achieve good models[10].

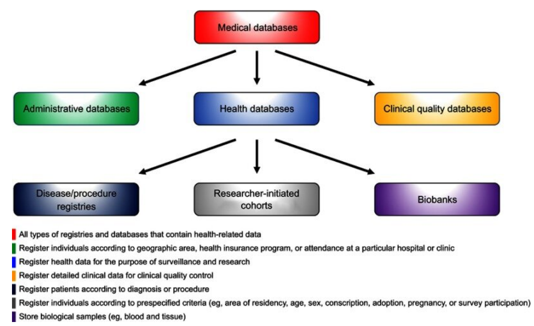


Figura 5. Medical databases

For this we will study and look for the necessary libraries, the data where to train the model and the respective analysis will be made, the respective study of the data will be carried out.

4. Project Structure

Focus and Techniques

This project has an emphasis on the study of Alzheimer's disease, the collection of data and the structuring of a prototype model, implementing artificial intelligence to obtain a model that with an audio can determine whether or not someone has Alzheimer's disease. **Tools** The following tools and technologies will be used for the project:

Github: Where the access to the code and the article https://github.com/HanamDavid/AFDA.git will be hosted.
Google Colab: To manipulate the code and process the data and generate the model.

https://drive.google.com/drive/folders/1NKC_4UG0HdzffI4SLvLF7ZKIxrg4uJlW?usp=drive_link
Tech:

- Tensorflow: To implement the necessary machine learning algorithms and to be able to classify the data correctly and generate useful processes.
- Sklearn: To be able to split the training data.
- Librosa: To process the information of the audios and to be able to normalize the data which will allow us to optimize

and improve the performance of the neural network and in general with the data cleaning.

- Matplotlib: To visualize the information and thus be able to have a way to understand the data more easily.
- Pandas and Numpy: For the manipulation of the Series and the Data once abstracted from the Audio samples.

We used the database **DementiaNet** which includes a series of medical data of dementia patients and healthy patients collected specifically for ML https://github.com/shreyasgite/dementianet

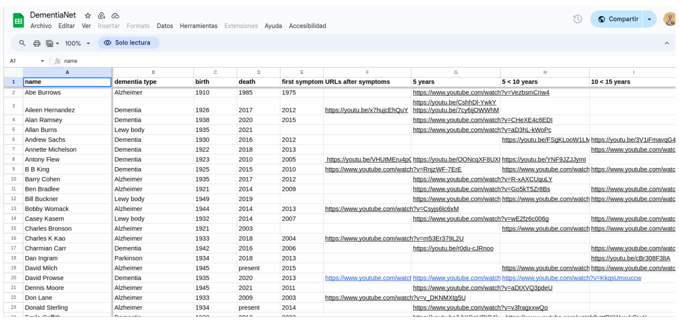


Figura 6. original data

This database contains a total of 359 audio samples collected from the internet of famous people among which 131 have dementia, the audios have different durations and there are people who have been diagnosed recently as patients with more than a decade and a half with the disease, the data were downloaded and processed locally and then uploaded to google drive.

Data Processing

Once the data was downloaded, because the database has samples from people with other diseases, more than half of the data from the sick patient samples were removed and we ended up with approximately 54 alzheimer's patients, later to compensate for the loss of data in the alzheimer's patient set and to allow for faster data processing, the alzheimer's audios were trimmed into 4 segments of 4 seconds each, and those of the healthy patients into 2 segments of 4 seconds and the csv was cleaned to keep only the relevant data, ending up with a total of 202 samples of healthy patients and 202 of alzheimer's patients.

.

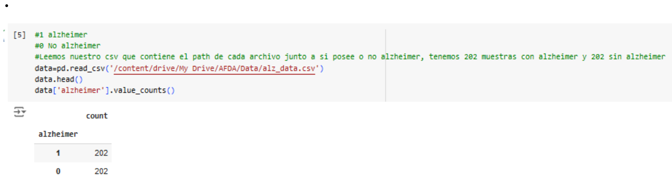


Figura 7. Clean Data

.

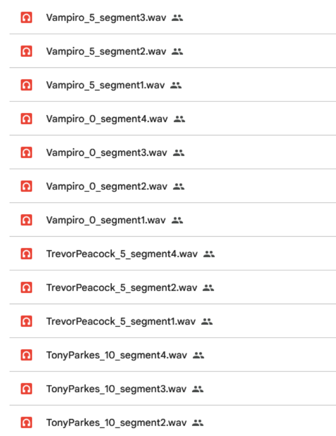
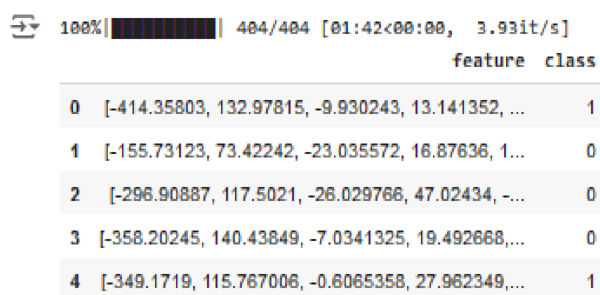


Figura 8. Data Segments

Next, with the folders and the csv, each of the samples were swapped to facilitate the training by randomizing the order of the samples. Booksa was used to extract the essential characteristics of each of the audios with MFCC since MFCC allows to extract vital data to obtain auditory and harmonic information in an easier way and it is usually used in audio processing.



100%| | 404/404 [01:42<00:00, 3.93it/s]

	feature	class
0	[-414.35803, 132.97815, -9.930243, 13.141352, ...	1
1	[-155.73123, 73.42242, -23.035572, 16.87636, 1...	0
2	[-296.90887, 117.5021, -26.029766, 47.02434, -...	0
3	[-358.20245, 140.43849, -7.0341325, 19.492668,...	0
4	[-349.1719, 115.767006, -0.6065358, 27.962349,...	1

Figura 9. MFCC Data

and stored the data in a vector for training, finally splitting the training and validation sets with sklearn, the 70th percentile of the data was used for training the network.

Model

For the model Tensorflow was used, several models were made but after the evaluation of normal, recurrent and convolutional networks, it was determined to choose the convolutional one given the capacity of these networks to explore patterns hardly observable in the data and a recurrent layer for the processing of the last layer. The model was as follows:

```python
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=(60, 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. Model configuration

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. Resume

```python
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. Training configuration

In the model we used

- Conv1D: One-dimensional convolutional networks, which are often used to process sequential data, were chosen due to the nature of our problem which requires searching for special features within the patients' voice
- LSTM: Long Short-Term Memory layers, which allow understanding the sequential information of the data. After extracting fundamental audio features with MFCC and convolutional layers, LSTM provides additional information by capturing long-term dependencies and mitigating gradient fading.
- BatchNormalization: This technique improves the efficiency and speed of cross-layer processing by normalizing activations, also acting as a form of regularization that can improve model generalization.
- Pooling: Which reduces the amount of network information that neurons receive, isolating the most relevant values and reducing the dimensionality of the data
- ModelCheckpoint: A technique that during training allows us to save the information of the best models during training
- Early Stopping: A technique that allows us to stop training early if no improvements are obtained for a period of time, which optimizes training performance and helps the model compute
- ReduceLROnPlateau: This technique adjusts the learning rate when no improvement is observed for a specific period, allowing the model to better fit the loss surface and improve convergence.

In our model, convolutional and LSTM layers are combined to capture both local characteristics and temporal patterns in audio data and we look for a good parameter capacity to be able to explore even patterns that are difficult to perceive, along with an analysis of the technologies and techniques used in the projects. previously observed

5. Discussion of Results

The results were the following:

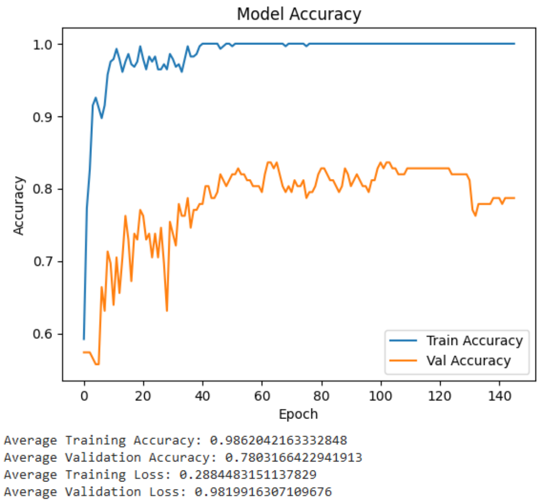


Figura 13. Model 1

and the following confusion matrix:

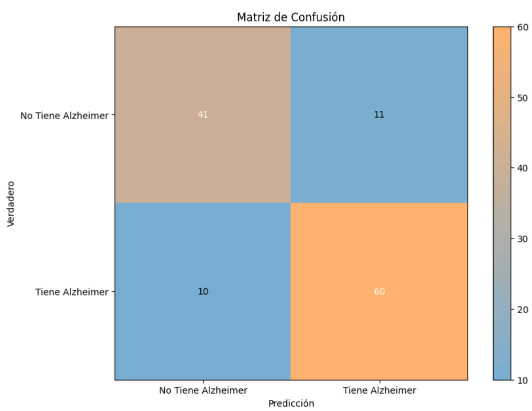


Figura 14. Matrix

The data obtained were:

- 95 % training precision: which indicates that the characteristics of the data observed during training were learned correctly
- 78 % validation precision: Lower than during training which could indicate overfitting, however it was a fairly good result and can be improved with subsequent model configurations and a greater amount of data
- 0.28 Training loss: Which indicates that the model predictions could be good, however when checked with the validation prediction, we can see that there is overfitting
- 0.98 Validation loss
- 0.87 AUC of training: Which indicates that there is a good ability to discriminate between classes in the training
- 0.89 Validation AUC Which indicates that there is a good ability to discriminate between classes in the validation and therefore in general in the model

If we compare with the projects previously studied, we can see that in general a very good result was obtained given the small amount of data, which demonstrates the capacity of convolutional and MFCC networks to discriminate diseases, a precision similar to that of the project 1 and 4, however with regard to other projects there is still a lot to improve to be able to make better models.

In particular, it is important to highlight the quality of the information collected, since the most successful projects had a high quality of data collected especially for the project, while in the case of AFDA the audios were collected from the Internet and the amount of noise and of unnecessary auditory information can impair disease discrimination. In the project, several of the techniques that were used in the aforementioned projects were combined, seeking to arrive at the best tools to solve the problem. The results are promising. However, it is important to study in greater detail the physiological and neurological particularities of Alzheimer's and how affects the voice, to be able to give better processing to the data before feeding it into the network and also to make better use of hyperparameters, obtaining higher levels of precision and also avoiding overfitting

6. Conclusions

The increase in medical capacity to intervene in diseases changes, structures and promotes good medical practices within society. Technological development is essential to protect and care for the population, mainly those at high risk, in the case of Alzheimer's, the greater the implementation and research of AI in medicine, the greater the ease it will open to improve the lives of patients. This of course has its cost, a large amount of research, a desire of society to adopt it and a large amount of data to provide adequate information to the models. Over time, AI takes up more and more space in our society and at least in medicine it gives good prospects.

Of course, it is necessary to first prioritize the moral and ethical obtaining of data and a good implementation of IoT systems to be able to collect this data, since without these we can have models with insufficient capacity to detect the disease or detect it poorly, which can be The lives of many patients are at risk, without a rigid and moral structure that supports artificial intelligence in medicine, it is not possible to ensure that the models are successful and generate the desired impact, but on the contrary, they generate false positives, distrust of the technology and in the worst case the death of many. Finally, it is important to detail that advances in Artificial Intelligence shine above all when they support traditional research and analysis systems since these systems allow data to be extrapolated and complement treatments only when the circumstances where it is applicable are adequately understood, which is why it is It is necessary that all progress be reviewed with interdisciplinary teams in order to exploit the use of technology in the scientific community and in this case in the medical community.

The projects observed and their implementation allowed us to observe that potential, it is important to continue supporting and exploring the use of Artificial Intelligence in the environment and thus be able to improve the quality of life of thousands of people throughout the world.

7. Bibliographic references

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. Laguarta, 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. Laguarta, 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_3$34421419.,Fig5