IDENTIFICACIÓN DE LA ENFERMEDAD DE PARKINSON USANDO CNNS SOBRE DOS REPRESENTACIONES GRÁFICAS DE SEÑALES DE HABLA: ESPECTROGRAMAS Y ESPACIO DE FASE

PROYECTO DEEPLEARNING 2023-1

ENTREGA 2

Estudiante: Diego Alexander López Santander

Contextualización

La enfermedad de Parkinson es un trastorno neurológico progresivo que afecta al movimiento, el equilibrio y la coordinación. Está causada por la degeneración de las neuronas cerebrales productoras de dopamina. Además de los síntomas motores, la enfermedad de Parkinson también puede afectar al habla y la comunicación. Por este motivo, la predicción automática mediante técnicas de Machine Learning de trastornos neurodegenerativos como la enfermedad de Parkinson a partir de señales de habla es un campo de investigación en rápido crecimiento. Con este tipo de sistemas es posible apoyar a los profesionales de la salud en el diagnóstico de este tipo de enfermedades y brindar herramientas accesibles para el tratamiento temprano de las mismas

El objetivo del proyecto es realizar una clasificación bi-clase entre individuos que padecen la enfermedad de Parkinson e individuos de control sanos a partir de señales de habla. Para ello usaron 2 representaciones de las señales de habla: Espectrogramas y la reconstrucción del espacio de fase usando el teorema de Takens [1], obteniendo un atractor. Posteriormente se entrenan redes neuronales convolucionales para clasificar pacientes con la enfermedad de Parkinson e individuos de control sanos. Con las dos representaciones de las señales se busca analizar que tan bien pueden caracterizar el fenómeno de interés (Enfermedad de Parkinson) y de ser posible, complementar la información contenida por cada representación usando técnicas de fusión.

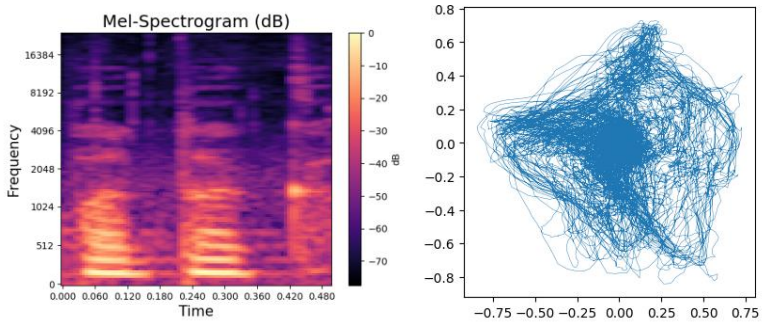


Figura 1. Representación de audio como espectrograma y como atractor.

Datos

Se usa la base de datos PC-GITA [2] la cual consta de grabaciones del habla de 50 pacientes con enfermedad de Parkinson y 50 sujetos de control sanos en formato WAV (waveform audio format), emparejados por edad y sexo. Todos los participantes son hablantes nativos de español y las grabaciones se recogieron siguiendo un protocolo que tiene en cuenta tanto los requisitos técnicos como varias recomendaciones dadas por expertos en lingüística, foniatría y neurología. Este corpus incluye tareas como la fonación sostenida de vocales, la evaluación diadococinética, 45 palabras, 10 frases, un texto de lectura y un monólogo. Las señales de audio se encuentran muestreadas con una frecuencia de 48 KHz. En el proyecto solo se usa la evaluación diadococinética correspondiente a la repetición de silabas \PaTaKa. Por cada individuo se cuenta con dos grabaciones de dicha tarea, acumulando un promedio de 15 segundos de grabación por sujeto. En total, las grabaciones ocupan un tamaño en disco de 143 Mb.

ESTRUCTURA DE NOTEBOOOKS ENTREGADOS

Se entregan 3 notebooks numerados:

- 01_Exploración_de_datos.ipynb
- 02_Arquitecturas_individuales_CNN.ipynb
- 03_Modelo_Joint_Fusion.ipynb

El primer notebook “01_Exploración_de_datos.ipynb” contiene una exploración de los datos, mostrando distintas representaciones gráficas y principalmente calculando, a partir de las señales de audio, las representaciones como espectrogramas y atractores en el espacio de fase para cada individuo de la base de datos.

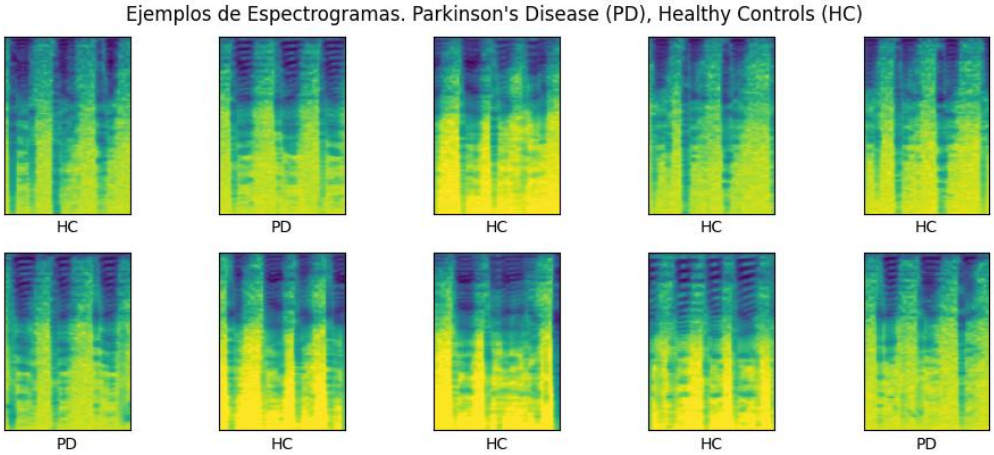


Figura 2. Espectrogramas etiquetados por segmento

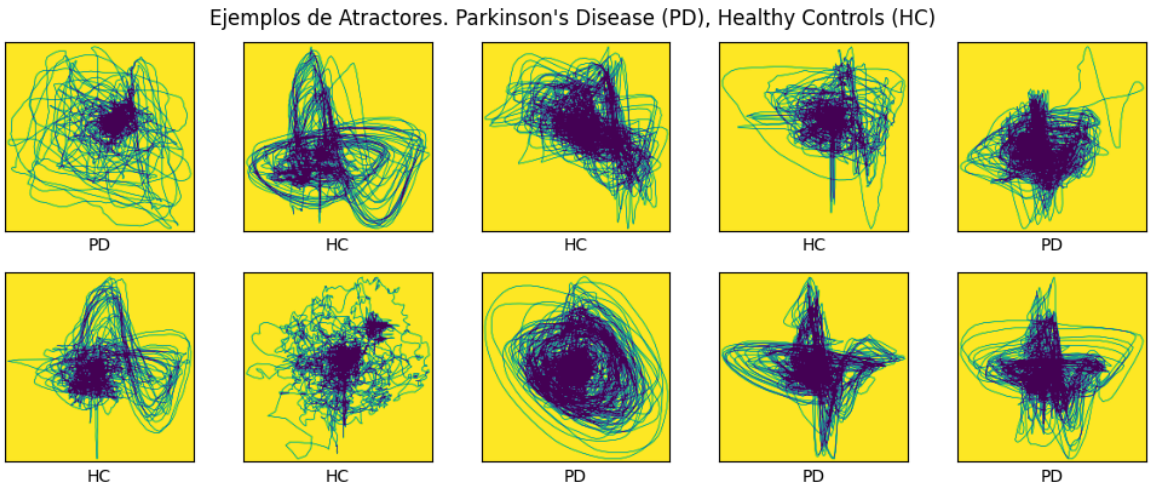
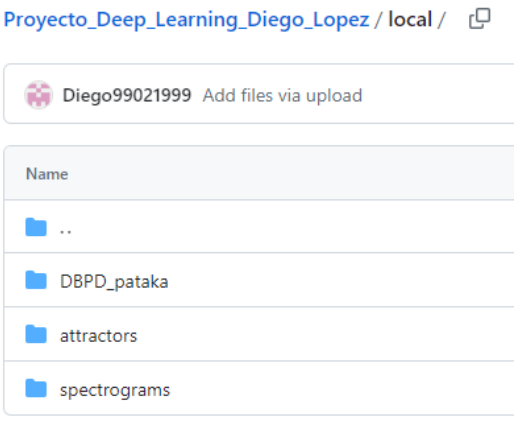


Figura 3. Atractores etiquetados por segmento

El segundo notebook "02_Arquitecturas_individuales_CNN.ipynb " contiene los experimentos realizados entrenando modelos de redes neuronales convolucionales individualmente, usando como entrada solamente las representaciones de atractores o de espectrogramas respectivamente.

El tercer notebook "03_Modelo_Joint_Fusion.ipynb" finalmente muestra la experimentación realizada usando la técnica de Joint Fusion para generar un modelo que combina la información de ambos tipos de entradas (Espectrogramas y atractores).



En la carpeta local se encuentran las carpetas: "DBPD_pataka", "attractors" y "spectrograms", las cuales contienen la base de datos, y sus representaciones como atractores espectrogramas, respectivamente.

DESCRIPCIÓN DE LA SOLUCIÓN

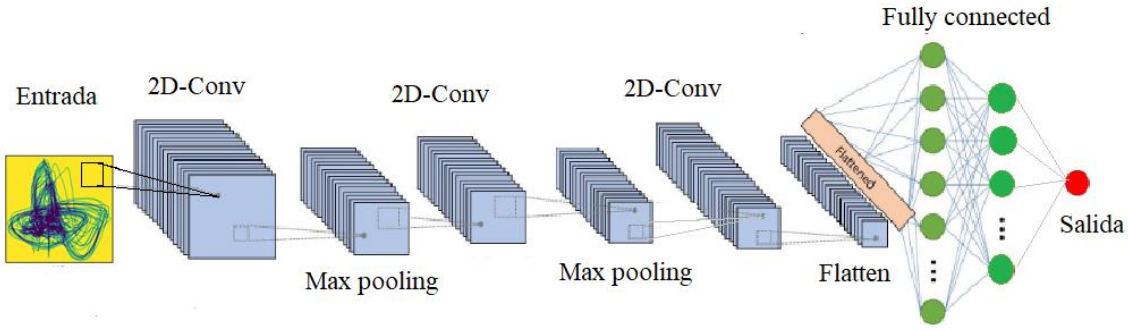


Figura 4. Modelo CNN

En la figura anterior se visualiza una representación gráfica del modelo de redes neuronales convolucionales implementado individualmente para cada representación de la señal. El modelo consiste en 3 capas convolucionales de dos dimensiones, cada una seguida por una respectiva capa de max-pooling y una de batch-normalization. El tamaño de la entrada de cada modelo depende de la representación: las dimensiones de los espectrogramas son de (128, 91, 1) mientras que las dimensiones de los atractores son de (182, 182, 1). El resultado de las capas convolucionales se 'aplana' y posteriormente se usan capas completamente conectadas hasta obtener una única salida (usando una neurona con activación sigmoide).

Model: "sequential_10"

Layer (type)	Output Shape	Param #
conv2d_31 (Conv2D)	(None, 126, 89, 16)	160
max_pooling2d_22 (MaxPooling2D)	(None, 42, 29, 16)	0
batch_normalization_20 (BatchNormalization)	(None, 42, 29, 16)	64
conv2d_32 (Conv2D)	(None, 40, 27, 16)	2320
max_pooling2d_23 (MaxPooling2D)	(None, 20, 13, 16)	0
batch_normalization_21 (BatchNormalization)	(None, 20, 13, 16)	64
conv2d_33 (Conv2D)	(None, 18, 11, 32)	4640
flatten_10 (Flatten)	(None, 6336)	0
dense_25 (Dense)	(None, 64)	405568
dense_26 (Dense)	(None, 1)	65

Total params: 412,881
Trainable params: 412,817
Non-trainable params: 64

Model: "sequential_11"

Layer (type)	Output Shape	Param #
conv2d_34 (Conv2D)	(None, 180, 180, 16)	160
max_pooling2d_24 (MaxPooling2D)	(None, 60, 60, 16)	0
batch_normalization_22 (BatchNormalization)	(None, 60, 60, 16)	64
conv2d_35 (Conv2D)	(None, 58, 58, 16)	2320
max_pooling2d_25 (MaxPooling2D)	(None, 29, 29, 16)	0
batch_normalization_23 (BatchNormalization)	(None, 29, 29, 16)	64
conv2d_36 (Conv2D)	(None, 27, 27, 32)	4640
flatten_11 (Flatten)	(None, 23328)	0
dense_27 (Dense)	(None, 64)	1493056
dense_28 (Dense)	(None, 1)	65

Total params: 1,500,369
Trainable params: 1,500,305
Non-trainable params: 64

Tabla 1. Summary completo de cada modelo. (Espectrogramas, Atractores)

Fusión de ambos modelos

El modelo de fusión se obtiene removiendo la capa de salida de los modelos anteriores, concatenando la salida de las 64 neuronas de la capa completamente conectada de cada modelo y llevando la salida de las 128 neuronas resultantes a una única neurona para obtener la clasificación binaria.

En la siguiente imagen se visualiza la fusión de ambos modelos. Cada recuadro azul corresponde a uno de los modelos vistos anteriormente, mientras que el recuadro rojo representa la etapa de fusión.

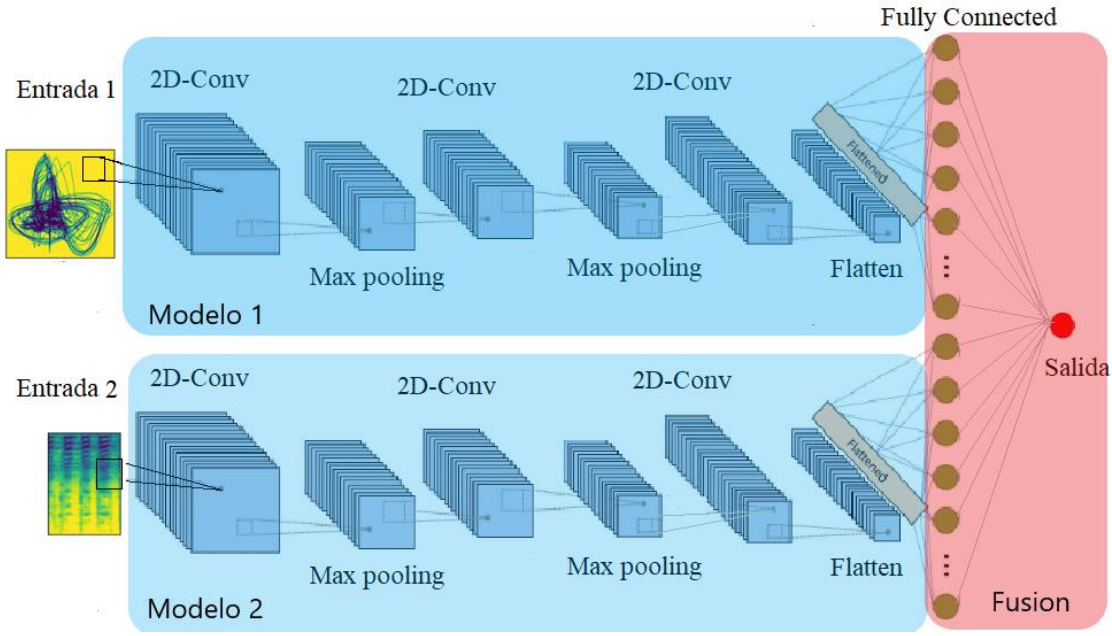


Figura 5. Fusión de ambos modelos

La siguiente tabla muestra el summary completo entregado por Tensorflow, del modelo obtenido mediante la unión de 2 modelos.

```
Model: "model_2"
_________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #     Connected to
=================================================================================================
 input_3 (InputLayer)        [(None, 182, 182, 1       0           []
                             )]

 input_4 (InputLayer)        [(None, 128, 91, 1)       0           []
                             ]

 model_1 (Functional)        (None, 64)                1500304     ['input_3[0][0]']

 model (Functional)          (None, 64)                412816      ['input_4[0][0]']

 concatenate (Concatenate)   (None, 128)               0           ['model_1[0][0]',
                                                                   'model[0][0]']

 dense_2 (Dense)             (None, 1)                 129         ['concatenate[0][0]']

=================================================================================================
Total params: 1,913,249
Trainable params: 1,913,121
Non-trainable params: 128
_________________________________________________________________________________________________
```

Tabla 2. Summary de fusión de modelos

RESULTADOS

Para obtener resultados más robustos e independientes de la elección particular de conjuntos de prueba y entrenamiento, se utiliza una estrategia de validación cruzada con 5 folds. De esta forma, el modelo se entrena en 5 ocasiones con particiones de datos distintas, de forma que cada uno de los datos sea usado tanto como dato de prueba, como dato de entrenamiento en alguna de las iteraciones.

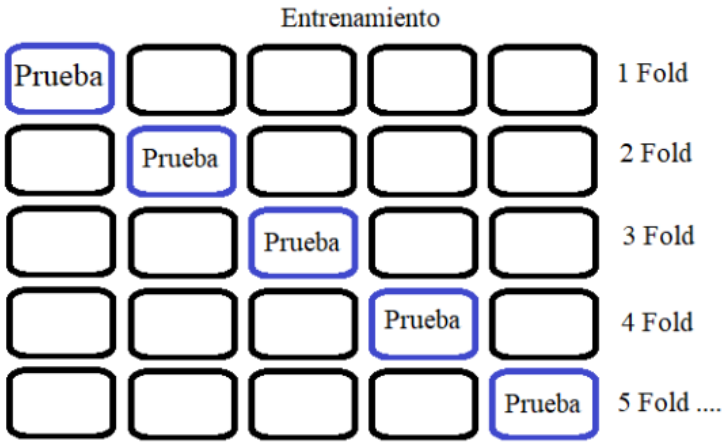


Figura 6. Esquema de validación cruzada

Tras finalizar el entrenamiento en cada Fold, se calcula el Accuracy de las predicciones con el conjunto de prueba. El resultado promedio se muestra a continuación.

Accuracy para Espectrogramas	Accuracy para Atractores
0.61 ± 0.0715	0.547 ± 0.0303

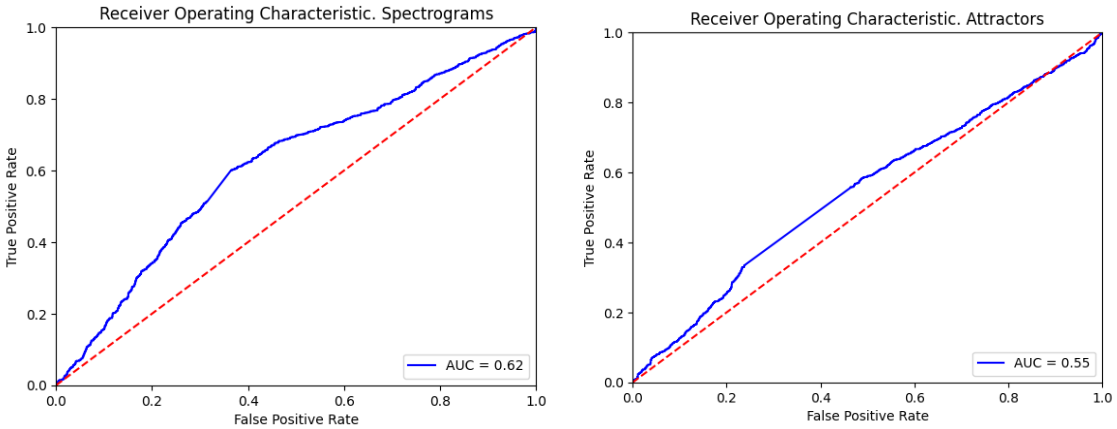


Figura 7. Curva Roc para ambos modelos (Espectrogramas y Atractores)

De las Curvas ROC se puede observar que el modelo de espectrogramas presenta un desempeño mejor el que modelo basado en atractores, pues presenta una mayor área bajo la curva.

Accuracy. Modelo de Fusión
0.612 ± 0.0277

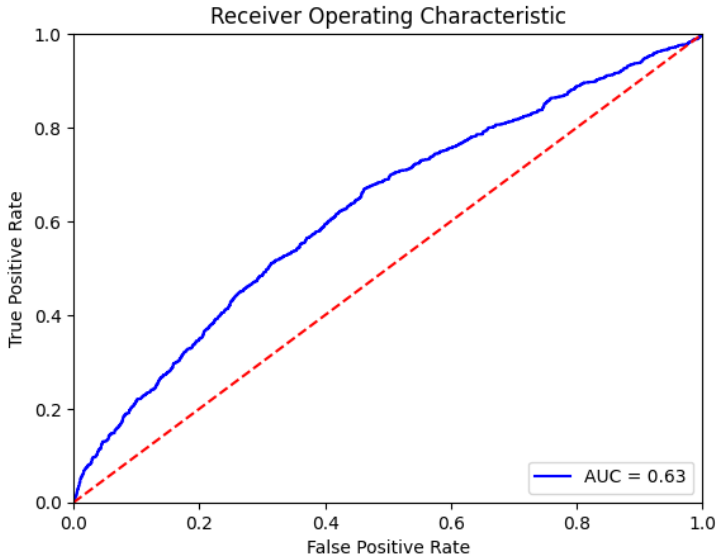


Figura 8. Curva Roc para fusión de modelos

La curva ROC de la fusión de modelos es muy similar a la de los modelos individuales. Se observa que en la fusión de modelos el desempeño es ligeramente mejor comparando las áreas bajo la curva ROC, al igual que también se presenta una mejora en la desviación estándar del accuracy comparado con los modelos individuales.

CONCLUSIONES

Se implementaron modelos de clasificación binaria basados en redes neuronales convolucionales, usando como entradas dos tipos de representaciones de las señales temporales: Como espectrogramas o como atractores.
A partir de ambos modelos se implementó una estrategia de fusión para combinar la información proveniente de ambas fuentes, obteniendo un muy ligero mejoramiento en el desempeño
Tensorflow permite una gran flexibilidad para la definición de modelos de redes neuronales.

REFERENCIAS

[1] Noakes, L. (1991). The Takens embedding theorem. International Journal of Bifurcation and Chaos, 1(04), 867-872.
[2] Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Gonzalez-Rátiva, M. C., & Nöth, E. (2014, May). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In LREC (pp. 342-347).