

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

## PROYECTO DEEPLARNING 2023-1

### ENTREGA 1

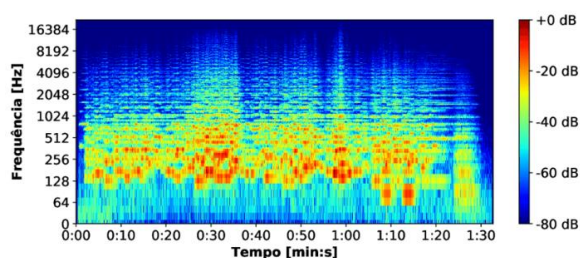
**Estudiante: Diego Alexander López Santander**

#### Contexto

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.

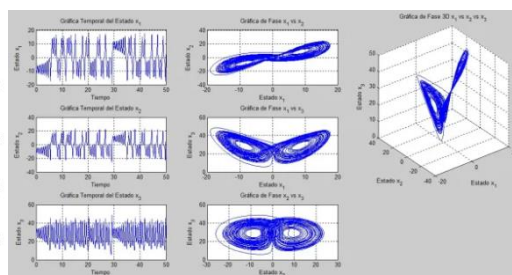
#### Objetivo de Machine Learning

El objetivo 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. Específicamente se planea utilizar dos representaciones gráficas de señales de audio: Espectrogramas y la reconstrucción del espacio de fase usando el teorema de Takens [1]. Posteriormente se usarán redes neuronales convolucionales para entrenar modelos capaces de 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.



Espectrograma

Tomado de [2]



Espacio de fase

Tomado de [3]

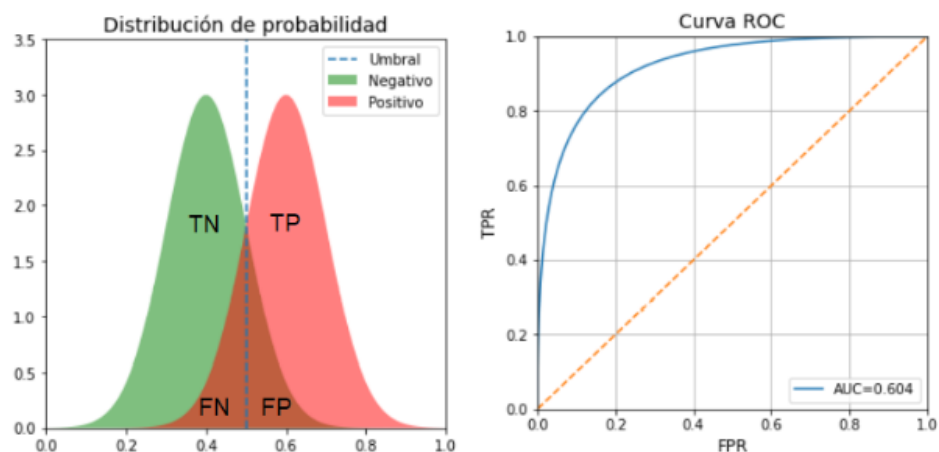
## Dataset

Se usará la base de datos PC-GITA [4] 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 usará la evaluación diadococinética correspondiente a la repetición de sílabas \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.

## Métricas de desempeño

Un esquema de validación cruzada permite evaluar de forma robusta el desempeño de un modelo, independiente de la selección específica de conjuntos de entrenamiento y prueba. Además, permite calcular medidas estadísticas de las métricas de desempeño. Entre las métricas de desempeño que se planea calcular se encuentran: Tasa de acierto, especificidad, sensibilidad, puntaje F1 y la curva ROC.



Curva ROC

## Referencias y resultados previos

En diversas publicaciones se han utilizado redes neuronales convolucionales para modelar la enfermedad de Parkinson a partir de los espectrogramas de las señales de habla obteniendo buenos resultados. En Johri et al [5] los autores reportan una tasa de acierto del

88% con el uso de CNNs sobre espectrogramas. En Er et al [6] se encuentran tasas de acierto de 95% usando una combinación de CNNs y LSTMs.

Por otro lado, no se encuentra bibliografía documentando el uso de la reconstrucción del espacio de fase para la identificación de la enfermedad de Parkinson. En [7] se usan CNNs para clasificar diferentes tipos de fonemas (fricativos, vocales, paradas, nasales) a partir de su reconstrucción en el espacio de fase, obteniendo una tasa de acierto del 60%.

## **Bibliografía**

[1] Noakes, L. (1991). The Takens embedding theorem. *International Journal of Bifurcation and Chaos*, 1(04), 867-872.

[2] [https://www.researchgate.net/figure/Figura-220-Exemplo-de-espectrograma-de-uma-musica\\_fig8\\_339434039](https://www.researchgate.net/figure/Figura-220-Exemplo-de-espectrograma-de-uma-musica_fig8_339434039)

[3] <https://tellesyaretzi.wordpress.com/2014/07/07/el-sistema-de-lorenz-y-sus-comportamientos-dinamicos/>

[4] 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).

[5] Johri, A., & Tripathi, A. (2019, August). Parkinson disease detection using deep neural networks. In *2019 twelfth international conference on contemporary computing (IC3)* (pp. 1-4). IEEE.

[6] Er, M. B., Isik, E., & Isik, I. (2021). Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with variational mode decomposition. *Biomedical Signal Processing and Control*, 70, 103006.

[7] Wesley, R. J., Khan, A. N., & Shahina, A. (2020). Phoneme classification in reconstructed phase space with convolutional neural networks. *Pattern Recognition Letters*, 135, 299-306.