

# Entrega 2: Evaluación de la Enfermedad de Parkinson Mediante Datos de Escritura a Mano Fundamentos de Deep Learning

Jeferson Gallo

jeferson.gallo@udea.edu.co

Estudiante de Maestría en Ingeniería de Telecomunicaciones  
Facultad de Ingeniería, Universidad de Antioquia



**UNIVERSIDAD  
DE ANTIOQUIA**

1 8 0 3

9 de junio de 2024

## Contexto

La enfermedad de Parkinson (EP) es una enfermedad neurodegenerativa que afecta el sistema nervioso central mediante la pérdida progresiva de neuronas dopaminérgicas [1]. La dopamina es un neurotransmisor encargado del funcionamiento de las vías neuronales asociadas con la motivación, el estado de ánimo y el control motor [3]. Por esta razón, la EP puede causar síntomas motores como la bradicinesia (lentitud en los movimientos voluntarios), la rigidez muscular y los temblores en reposo. Estos síntomas afectan significativamente diferentes actividades de los pacientes como lo es la escritura a mano, ya que los procesos de escritura necesitan una alta coordinación de los músculos y tendones que conectan el antebrazo, la muñeca y los dedos. Adicionalmente, las principales anomalías que se evidencia en la escritura son la micrografía (reducción anormal en el tamaño de la escritura) y la disgrafía (déficits en la producción grafomotora debido a una combinación de varios signos cardinales de la enfermedad de Parkinson) [2].

## Objetivo de Machine Learning

En este proyecto se abordará una clasificación bi-clase entre pacientes con la enfermedad de Parkinson versus personas sanas (HC) mediante señales de escritura a mano, utilizando diferentes tareas como dibujos, escritos y una tarea compleja. Las imágenes correspondientes a cada tarea se procesan mediante una red convolucional VGG-16, la cual fue entrenada con la base de datos ImageNet. El objetivo principal se basa en determinar cuál de las diferentes tareas de escritura me permite un mejor modelamiento de la enfermedad de parkinson.

## Datos

Los datos fueron recopilados utilizando una tableta Wacom Cintiq 13 HD con retroalimentación visual para los participantes y con una frecuencia de muestreo de 180 Hz. Esta tableta proporciona seis señales diferentes provenientes del proceso de escritura: la posición  $x$ , la posición  $y$ , la distancia desde la superficie hasta el lápiz (la posición  $z$ ), la presión del lápiz, el ángulo azimut y la inclinación del lápiz (la altitud). Para la creación de las imágenes se utilizan las señales  $(x, y)$  las cuales corresponden a los trazos realizados sobre la superficie de la tablet. Un ejemplo de las imágenes resultantes para algunas tareas se encuentran en la Figura 1.

Para este trabajo se tienen un total de 7 tareas dentro de las cuales se encuentran algunos escritos como el alfabeto, frase libre y nombre; tareas de dibujo como una línea curva, espiral y una casa; y finalmente se tiene una tarea compleja llamada rey. En la Tabla 1 se encuentran la cantidad de participantes por cada una de las tareas.

Los datos son de uso académico exclusivamente y para solicitar acceso se debe escribir correo al coordinador del grupo GITA [rafael.orozco@udea.edu.co](mailto:rafael.orozco@udea.edu.co).

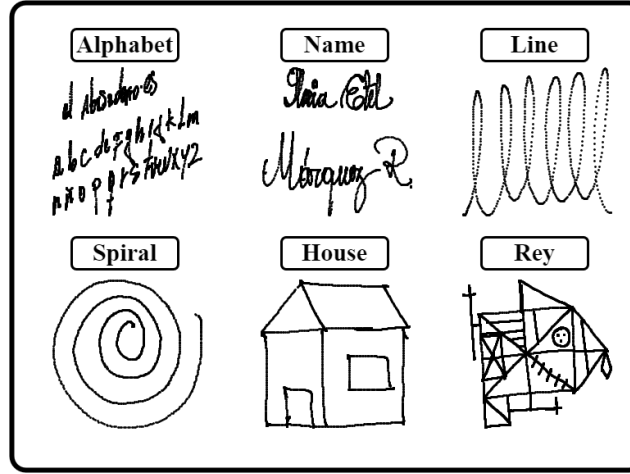


Figura 1: Imágenes de la base de datos.

Tabla 1: Cantidad de participantes entre pacientes (PD), y controles (HC), en cada una de las tareas.

Tarea	# Pacientes (PD)	# Controles (HC)
alphabet	71	55
line1	60	42
freewriting	73	54
house	75	55
rey	69	54
name	74	55
spiral	75	55

## Métricas de desempeño

En este proyecto se utilizarán métricas como la matriz de confusión, como también algunas métricas que se desprenden de esta.

### Matriz de confusión

Es una matriz cuadrada donde el número de filas y columnas depende de la cantidad de clases diferentes que tenga el problema. Esta matriz organiza las predicciones correctas e incorrectas evidenciando el rendimiento del sistema para la predicción de cada clase. En problemas de clasificación bi-clase, la matriz de confusión es  $2 \times 2$  donde en cada celda se pueden definir los siguientes términos:

- *Verdaderos positivos (TP, true positive)*: es la cantidad de datos de la clase positiva que el sistema clasificó correctamente.
- *Falsos positivos (FP, false positive)*: son la cantidad de datos de la clase positiva que el sistema clasificó erróneamente como pertenecientes a la clase negativa.
- *Verdaderos negativos (TN, true negative)*: es el número de datos de la clase negativa clasificados correctamente por el sistema.

- *Falsos negativos (FN, false negative)*: es el número de datos de la clase negativa que el sistema clasificó erróneamente como pertenecientes a la clase positiva.

Además, otras métricas pueden ser estimadas de acuerdo con la información de la matriz de confusión:

- *Exactitud (acc, accuracy)*: se define como la cantidad de aciertos en ambas clases sobre el total de datos.

$$acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

- *Sensibilidad (S, sensitivity)*: es la capacidad del sistema para clasificar la clase positiva.

$$S = \frac{TP}{TP + FN} \quad (2)$$

- *Especificidad (E, specificity)*: es la capacidad del sistema para clasificar la clase negativa.

$$E = \frac{TN}{TN + FP} \quad (3)$$

- *Precisión (P, precision)*: mide la proporción de muestras positivas predichas correctamente entre todas las muestras que se predicen como positivas.

$$P = \frac{TP}{TP + FP} \quad (4)$$

- *F1-score*: es la media armónica entre la precisión y la sensibilidad, la cual permite obtener una medición balanceada del desempeño del modelo.

$$F1 - score = 2 \times \frac{P \times S}{P + S} \quad (5)$$

## Modelo de Deep Learning

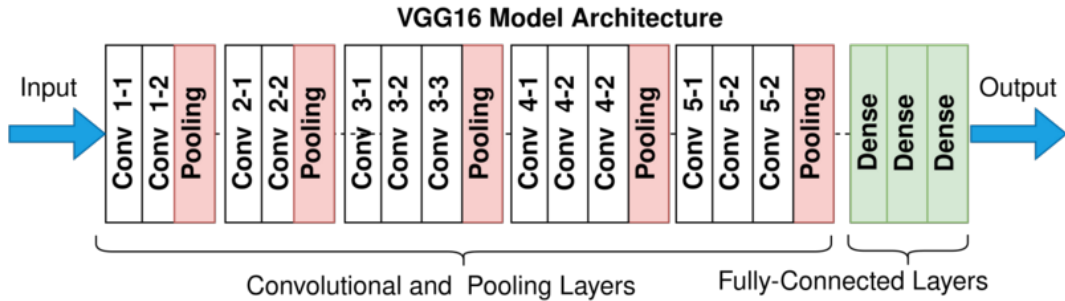


Figura 2: Arquitectura VGG16.

Para este trabajo se implementó la red neuronal convolucional VGG16. Esta es una arquitectura desarrollada por el Visual Geometry Group (VGG) de la Universidad de Oxford, que se destacó en la competición ImageNet Large Scale Visual Recognition Challenge (ILSVRC) de 2014. Se caracteriza por su simplicidad y efectividad, utilizando principalmente pequeñas capas convolucionales de  $3 \times 3$  píxeles y capas de max-pooling de  $2 \times 2$  píxeles, organizadas en 5 bloques de capas convolucionales. Esta arquitectura ha demostrado ser extremadamente eficiente en tareas de reconocimiento de imágenes, estableciendo un nuevo estándar en precisión y eficiencia en el campo del aprendizaje profundo [4]. El modelo implementado en este trabajo 2 fue pre-entrenado con la base de datos ImageNet y se le agregan dos capas densas para realizar la clasificación entre pacientes y controles.

## Experimentos

### Exploración de tareas

Para este primer experimento se desea conocer cuál de las tareas utilizadas modela de una mejor manera la enfermedad del parkinson en la tarea de clasificación. En la Tabla 2 se encuentran los resultados obtenidos para las tareas cuando no se afinan ningún bloque convolucional y solamente se afinan las capas densas encargadas de realizar la clasificación. En la Tabla 3 se encuentran los resultados del experimento cuando se afinan todos los bloques convolucionales y las capas densas de la red. En ambos casos la tarea de escritura del nombre obtuvo exactitudes del 84% del modelo sin afinar y 73.08% cuando se afinaron todos los bloques convolucionales.

Tabla 2: Resultados de la exploración de tareas utilizando el modelo VGG16 sin afinar.

Task	Accuracy	Specificity	Sensitivity	F1_Score
<b>name</b>	84.62	100.00	75.00	0.86
<b>spiral</b>	76.92	86.67	63.64	0.70
<b>rey</b>	76.00	63.64	85.71	0.80
<b>alphabet</b>	65.38	54.55	73.33	0.71
<b>house</b>	57.69	41.67	71.43	0.65
<b>line1</b>	42.86	37.50	46.15	0.50
<b>freewriting</b>	23.08	23.08	23.08	0.23

Tabla 3: Resultados de la exploración de tareas afinando todos los bloques convolucionales del modelo VGG16.

Task	Accuracy	Specificity	Sensitivity	F1_Score
<b>name</b>	73.08	70.00	75.00	0.77
<b>spiral</b>	73.08	66.67	81.82	0.72
<b>alphabet</b>	65.38	72.73	60.00	0.67
<b>rey</b>	64.00	54.55	71.43	0.69
<b>house</b>	57.69	58.33	57.14	0.59
<b>freewriting</b>	53.85	53.85	53.85	0.54
<b>line1</b>	47.62	37.50	53.85	0.56

## Exploración de la tare "Name"

Debido a que en el anterior experimento la tarea name obtuvo el mejor desempeño en ambos casos, además se observó una variación entre el uso del modelo sin afinar y afinando todas las capas. En este experimento se exploran diferentes grados de afinación de los bloques convolucionales. En la Tabla 4 se observan los resultados al variar la cantidad de bloques convolucionales afinados. El mejor resultado finalmente fue afinando solamente la capa densa que corresponde a la etapa de clasificación. Esto puede estar ocurriendo debido a la poca cantidad de datos, ya que no se alcanzan a afinar correctamente todas las capas convolucionales para modelar el fenómeno del parkinson.

Tabla 4: Resultados de la exploración de tareas afinando todos los bloques convolucionales del modelo VGG16.

Block	Total parameters	Trainable parameters	Accuracy	Specificity	Sensitivity	F1_Score
dense	14,780,610	65,922	87.50	100.00	75.00	0.86
block5	14,780,610	7,145,346	73.08	100.00	56.25	0.72
block4	14,780,610	13,045,122	53.85	60.00	50.00	0.57
block3	14,780,610	14,520,450	65.38	100.00	43.75	0.61
block2	14,780,610	14,741,890	76.92	100.00	62.50	0.77
block1	14,780,610	14,780,610	73.08	70.00	75.00	0.77

## Análisis de resultados

En la Figura 3 se encuentra el proceso de entrenamiento del mejor experimento el cual corresponde a la tarea "name" cuando solo se afina la capa densa. Se configuró un total de 2000 épocas con una paciencia para el early stopping de 100. En total fueron necesarias 1400 épocas para afinar el modelo. Adicionalmente, en la Figura 4 se encuentra la matriz obtenida para el conjunto de test en este experimento. En este caso, el 100 % de los controles fué clasificado correctamente mientras que el 75 % de los pacientes pudieron ser identificados por el modelo.

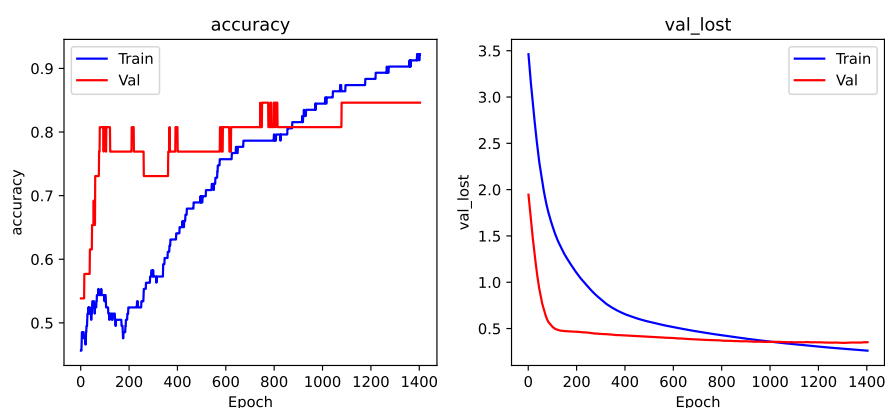


Figura 3: Curvas de entrenamiento en la tarea "name" afinando las capas densas.

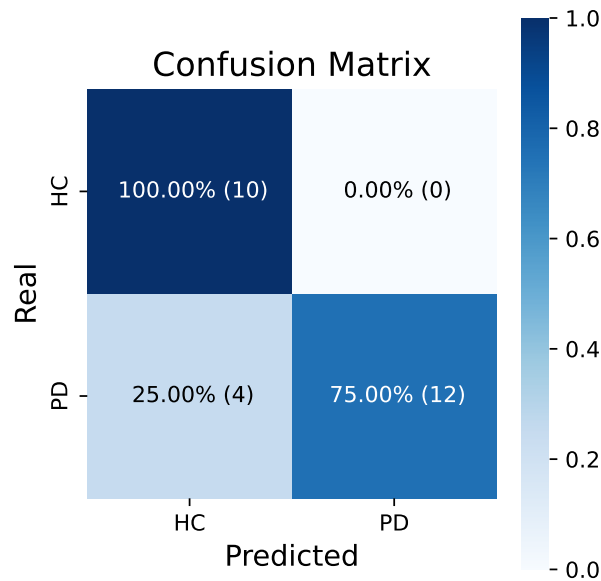


Figura 4: Matiz de confusión para el experimento del nombre cuando se afinan las capas densas.

## Disribución de los notebooks

En el notebook **1\_Data\_Exploration** se encuentra una breve exploración de la base de datos, donde se realiza el conteo de participantes por tarea y la visualización de algunas tareas. En el notebook **2\_CNN\_Model.ipynb** se encuentra la implementación del pipeline de entrenamiento. Respecto al script **3\_Task\_Exploration\_VGG16.py** se realiza el primer experimento donde se exploran las diferentes tareas. Finalmente, en el script **4\_Name\_CNN\_Models.py** se realizan los experimentos aplicando diferentes niveles de afinación para la mejor tarea obtenida en el experimento anterior.

## Referencias

- [1] Hornykiewicz, O.: Biochemical aspects of parkinson's disease. *Neurology* **51**(2-suppl.2), S2–S9 (1998)
- [2] Letanneux, A., Danna, J., Velay, J.L., Viallet, F., Pinto, S.: From micrographia to parkinson's disease dysgraphia. *Movement Disorders* **29**(12), 1467–1475 (2014)
- [3] Meder, D., Herz, D.M., Rowe, J.B., Lehericy, S., Siebner, H.R.: The role of dopamine in the brain-lessons learned from parkinson's disease. *Neuroimage* **190**, 79–93 (2019)
- [4] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)