



Evaluación automática de la enfermedad del Parkinson usando metodologías de aprendizaje clásico y profundo a partir de tareas de escritura cotidianas

Autor:

Jeferson David Gallo Aristizábal

Trabajo de Grado para optar por el título de:

Ingeniero Electrónico

Asesor:

Prof. Dr.-Ing. Juan Rafael Orozco Arroyave

Co-asesor:

MSc. Daniel Escobar Grisales

Línea de investigación:

Análisis de patrones y reconocimiento de señales

Grupo de investigación GITA

Universidad de Antioquia

Facultad de Ingeniería

Departamento de Ingeniería Electrónica y Telecomunicaciones

Ingeniería Electrónica

Medellín, Antioquia, Colombia

2023

Cita	Gallo-Aristizábal [1]
Referencia Estilo IEEE (2020)	[1] Gallo-Aristizábal, J.D. “Evaluación automática de la enfermedad del Parkinson usando metodologías de aprendizaje clásico y profundo a partir de tareas de escritura cotidianas”, Trabajo de grado, Ingeniería Electrónica, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2023.



Grupo de Investigación Telecomunicaciones Aplicadas (GITA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio César Saldarriaga Molina.

Jefe departamento: Eduard Emiro Rodríguez Ramírez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Agradecimientos

En primer lugar, quiero agradecer a mis padres Omaira Aristizábal y Oscar Gallo por todo el apoyo incondicional que siempre me han brindado y por ser siempre la mayor fuente de motivación para seguir adelante tanto en los procesos académicos como personales. No existen palabras que puedan representar el gran esfuerzo que realizan a diario para brindarnos las herramientas con las cuales podemos enfrentarnos a los retos del día a día. Agradezco a mis hermanos Yerry, Yeison y Jenifer quienes también estuvieron presentes durante todo este camino y me brindaron el mayor apoyo posible.

Durante este largo proceso que involucra ser estudiante de la Universidad de Antioquia he conocido grandes amigos con quienes he podido crecer tanto académica, como personalmente. Quiero expresar un especial agradecimiento hacia Diego López por quien tengo una inmensa admiración y quien durante muchos procesos me ha brindado su gran apoyo. Julio Zapata, a quien conocí iniciando mi carrera y juntos logramos llegar al final de este gran logro. A Laura, Helena, Carlos, Julián y Santiago con quienes entablamos una excelente amistad y pudimos estar juntos mucho tiempo sorteando los retos que nos imponía la universidad y a quienes felicito por ser los excelentes Ingenieros que son hoy. A ellos les deseo siempre la mejor de las suertes y espero sigamos juntos en el camino que nos queda.

Finalmente quiero agradecer a Daniel Escobar y al profesor Rafael Orozco quienes me han brindado todo su conocimiento y apoyo en el grupo de investigación GITA y me han permitido crecer académicamente en el ambito de la investigación. Además, por ser un ejemplo a seguir en mi carrera profesional.

A todos ustedes ¡¡Gracias Totales!!

Índice

1. Resumen	5
2. Abstract	6
3. Introducción	7
4. Objetivos	10
4.1. Objetivo general	10
4.2. Objetivo específicos	10
5. Marco teórico	11
5.1. Métodos de extracción de características clásicos	11
5.1.1. Patrones Binarios Locales (LBP, del inglés Local Binary Patterns)	11
5.1.2. Histograma de Gradientes Orientados (HOG)	13
5.2. Métodos de extracción de características profundos	15
5.2.1. Redes Neuronales Convolucionales (CNN)	15
5.2.2. Redes Neuronales Recurrentes (RNN)	16
5.3. Algoritmos de clasificación	17
5.3.1. Maquinas de Soporte Vectorial (SVM)	17
5.4. Metricas de Desempeño	19
5.4.1. Matriz de confusión	19
5.4.2. Curva ROC	20
6. Base de datos	21
7. Metodología general	22
7.1. Enfoque clásico	23
7.2. Enfoque basado en redes neuronales profundas	23
7.3. Optimización de los parametros y clasificación	26
8. Experimentos y Resultados	26
8.1. Resultados enfoque clásico	26
8.2. Resultados del enfoque basado en aprendizaje profundo	27
8.3. Resultados utilizando una estrategia de fusión	29
9. Conclusiones	30
10. Referencias	32

Índice de figuras

1.	Proceso para el descriptor LBP básico.	12
2.	Patrones uniformes para LBP.	12
3.	Filtros de Roberts, Prewitt y Sobel para el cálculo de los gradientes en las direcciones X y Y	13
4.	Direcciones de los gradientes y su histograma correspondiente.	14
5.	Imagen objetivo, dividida en celdas.	14
6.	Esquema general de una red neuronal recurrente.	17
7.	Transformación del espacio de características mediante la función kernel.	18
8.	SVM de margen blando.	18
9.	Matriz de confusión.	20
10.	Ejemplo ilustrativo de la curva ROC.	21
11.	Metodología general.	22
12.	Arquitectura de la CNN.	24
13.	Secuencias de preentrenamiento para la RNN.	25
14.	Arquitectura RNN.	25
15.	Curva ROC y distribución de los scores para el mejor resultado del enfoque clásico.	27
16.	Curva ROC y distribución para el mejor resultado del enfoque profundo.	29
17.	Curva ROC y distribución para la estrategia de fusión temprana.	30

Índice de tablas

1.	Información de los participantes.	22
2.	Resultados obtenidos para el conjunto de experimentos del enfoque clásico. Valores en términos de $(\mu \pm \sigma)$ y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).	26
3.	Resultados obtenidos para el conjunto de experimentos del enfoque basado en arquitecturas de aprendizaje profundo. Valores en términos de $(\mu \pm \sigma)$ y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).	28
4.	Resultados obtenidos cuando se utiliza una estrategia de fusión temprana con los mejores resultados de cada enfoque y características basadas en señales dinámicas. Valores en términos de $(\mu \pm \sigma)$ y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).	30

1. Resumen

La enfermedad de Parkinson es un desorden neurodegenerativo caracterizado por producir trastornos en el movimiento como temblor, rigidez, entre otros. Estos síntomas afectan diferentes actividades de la vida cotidiana entre ellas la escritura a mano. Generalmente, la escritura de los pacientes con enfermedad de Parkinson presenta diferentes síntomas como la micrografía, la cual está relacionada con la reducción del tamaño de la escritura y la disgrafía, que afecta la coordinación de los movimientos finos que necesitan las personas al escribir. En este trabajo se clasificaron pacientes con la enfermedad de Parkinson y controles saludables a partir de tareas de escritura cotidianas con el objetivo de lograr un enfoque cercano a la evaluación no intrusiva. Para ello se consideró un enfoque clásico donde se implementan diferentes técnicas de procesamiento de imágenes como los histogramas de gradientes orientados y patrones binarios locales. Además, se implementaron arquitecturas de aprendizaje profundo como las redes neuronales convolucionales y las redes neuronales recurrentes. Finalmente, ambos enfoques fueron evaluados y comparados en la tarea de clasificación de pacientes con enfermedad de Parkinson a partir de la escritura de los dígitos del cero al nueve.

Palabras claves: Enfermedad de Parkinson, Histograma de Gradientes Orientados, Patrones Binarios Locales, Redes Neuronales Convolucionales, Redes Neuronales Recurrentes, Escritura a Mano, Procesamiento de Imágenes.

2. Abstract

Parkinson's disease is a neurodegenerative disorder characterized by movement disorders such as tremors and rigidity, among others. These symptoms impact various activities of daily living, including handwriting. Patients with Parkinson's disease commonly exhibit different symptoms in handwriting, such as micrographia, which is associated with a reduction in writing size, and dysgraphia, which affects the fine motor coordination required for writing. This study aims to classify patients with Parkinson's disease and healthy controls based on everyday writing tasks, with the goal of achieving a non-intrusive evaluation approach. A classical approach is adopted, incorporating various image processing techniques such as Histograms of Oriented Gradients and Local Binary Patterns. Additionally, deep learning architectures such as convolutional neural networks and recurrent neural networks were implemented. Finally, both approaches are evaluated and compared in the task of classifying patients with Parkinson's disease from the writing of the digits from zero to nine.

Keywords: Parkinson's disease, Histogram of Oriented Gradients, Local Binary Patterns, Convolutional Neural Networks, Recurrent Neural Networks, Handwriting, Image Processing.

3. Introducción

La enfermedad de Parkinson (PD, Parkinson's Disease), es la segunda enfermedad neurodegenerativa más común alrededor del mundo afectando en una mayor proporción a las personas mayores [1], Esta enfermedad se debe principalmente a la pérdida progresiva de neuronas dopaminérgicas en el sistema nervioso central [2], lo que produce diferentes síntomas motores y no motores. Los trastornos motores suelen aparecer en las etapas intermedias de la PD y se manifiestan principalmente como: rigidez muscular, temblor de reposo, inestabilidad en la postura y bradiquinesia [3], [4]. Dichos síntomas afectan actividades de la vida cotidiana de las personas donde puede estar involucrada la escritura a mano, en la cual se requiere una coordinación motora altamente compleja para el correcto movimiento de los músculos del antebrazo, la muñeca y los dedos [5]. Las alteraciones más evidentes en la escritura de los pacientes con la enfermedad de Parkinson son la micrografía, la cual implica reducción del tamaño de la escritura y la disgrafía, que está relacionada con la coordinación de los movimientos finos que necesitan las personas para escribir [6].

Las deficiencias en la producción escrita son uno de los primeros síntomas que exhiben los pacientes con PD lo que ha incrementado el interés en la comunidad científica por usar la escritura como biomarcador para el monitoreo y evaluación de pacientes con PD [7], [8]. Además, estas señales pueden ser recolectadas de manera remota mediante el uso de computadores y tabletas, lo que la convierte en una metodología de monitoreo de bajo costo [9]. Típicamente, los estudios acerca del análisis de escritura en pacientes con PD han abordado dos metodologías: el procesamiento de señales dinámicas y el procesamiento de imágenes. En el análisis dinámico de escritura se tienen en cuenta señales provenientes de dispositivos electrónicos como tabletas digitalizadoras, lapiceros inteligentes, entre otros. Por otro lado, en el análisis de imágenes se utilizan datos provenientes de la escritura sobre el papel o se trabaja con imágenes reconstruidas a partir de las señales recolectadas de algún dispositivo.

Dentro de las señales dinámicas generalmente se cuenta con información de la presión, ángulos de inclinación y movimientos fuera de la superficie de escritura. Por ejemplo, en [5] los autores evalúan características cinemáticas y temporales tanto para los movimientos en el aire como en la superficie de la tablet cuando los sujetos escriben una frase en lenguaje checo. Exactitudes de hasta un 85,6% fueron obtenidas cuando se combinan las características de los movimientos en el aire y sobre la superficie. En otro estudio similar realizado en [7], y mediante el uso de características no lineales geométricas y cinemáticas, obtuvieron exactitudes superiores al 94% cuando utilizaron la tarea de la espiral de Arquímedes. Algunos trabajos buscan analizar las secuencias temporales asociadas al proceso de escritura mediante arquitecturas de

aprendizaje profundo. En [10] utilizaron un modelo de CNN (del inglés Convolutional Neural Network) seguido de una BIGRU (del inglés Bidirectional Gated Recurrent Unit) para el procesamiento de las señales de pacientes y personas sanas. Exactitudes de 96,3 % y 94,4 % fueron obtenidas para la base de datos PaHaW y NewHandPD, respectivamente. Además, las tareas con mejor desempeño constan de: espirales, figuras geométricas y escritura de frases. En otro estudio relacionado desarrollado en [11] donde se hace uso la base de datos PDMultiMC, los autores implementaron un ensamble de una CNN junto con una BLSTM (del inglés Bidirectional Long Short Term Memory). Los autores reportaron exactitudes de hasta un 83 % usando 7 tareas diferentes dentro de las cuales están: el dibujo de una señal triangular, tren de pulsos, la escritura de algunos días, entre otras tareas.

Generalmente las señales dinámicas brindan más información que las imágenes debido a que estas van capturando datos durante el proceso de escritura. Sin embargo, el procesamiento de las imágenes permite el monitoreo y evaluación de los pacientes sin la necesidad del uso de dispositivos electrónicos, logrando no solo una estrategia de evaluación económica sino también más amigable con el paciente. Por ejemplo, en [12] los autores usan una tarea donde le piden a los sujetos dibujar la espiral de Arquímedes siguiendo una plantilla. Características basadas en la diferencia entre el radio de la plantilla y el trazo del sujeto y otras denominadas temblor relativo medio fueron extraídas para caracterizar los dibujos de los pacientes. En este trabajo los investigadores obtuvieron una exactitud del 78,9 % al utilizar el clasificador Naïve Bayes. En [13] se propone una metodología similar, pero se usan 3 plantillas diferentes: espiral de Arquímedes, un triángulo y un cubo. En este caso los sujetos deben seguir la plantilla con un mouse y se extraen características basadas en la distancia del trazo del sujeto a la plantilla. Exactitudes de hasta el 100 % fueron reportadas cuando se usa una SVM como clasificador. En otro estudio realizado en [14], los autores buscan clasificar pacientes con PD y controles saludables usando diferentes técnicas de procesamiento de imágenes como: Histogramas de gradientes orientados (HOG, del inglés Histogram of Oriented Gradients), ORB (del inglés Oriented FAST and rotated BRIEF), histogramas de gradientes de color (CGH, del inglés Color Gradients Histogram), entre otros. Estas técnicas son evaluadas en una base de datos compuesta por 102 sujetos (51 pacientes), los cuales desarrollaron la tarea de la espiral en una tableta digitalizadora. Finalmente, exactitudes de hasta un 82,6 % fueron obtenidas usando las características basadas en HOG y como clasificador un RF.

Las CNNs también han sido usadas para el análisis de las imágenes extraídas en el proceso de escritura de pacientes con PD. Por ejemplo, los investigadores en [15] utilizan las imágenes construidas de los datos dinámicos recolectados de diferentes tareas de escritura (espirales, círculos, curvas y rotaciones de la muñeca en el aire) de

74 pacientes con PD y 18 controles. Por otro lado, los modelos implementados constan de diferentes enfoques de CNN basados en dos arquitecturas, CNN-ImageNet y CNN-Cifar-10. Finalmente, cuando todas las tareas son combinadas obtuvieron exactitudes superiores al 93,4% usando CNN-ImageNet, la cual resulto ser una de las arquitecturas más estables en cuanto a desempeño. Debido a la poca cantidad de datos presentes en los conjuntos de datos de escritura dedicados a la investigación de PD, algunos estudios han explorado técnicas de aumento de datos y transferencia de aprendizaje, para suplir la gran cantidad de información que normalmente necesitan las arquitecturas basadas en Deep Learning para reconocer diferentes patrones. Por ejemplo, en [16], los autores implementaron algunos modelos basados en CNNs como ResNet50 y AlexNet, entrenados inicialmente con el conjunto de datos ImageNet. Estos modelos son ajustados para clasificar pacientes con PD y controles saludables en 4 conjuntos de datos, PaHaW, HandPD, NewHandPD y Parkinson's Drawing. Los autores reportaron exactitudes del 99,2% ajustando la arquitectura AlexNet mediante estrategias de transferencia de aprendizaje y técnicas de aumento de datos como: rotaciones, cambios en la iluminación y contraste, entre otras. En otro estudio realizado en [17], además de utilizar técnicas de Aumento de Datos los autores proponen un enfoque de múltiples CNN afinadas, teniendo en cuenta dos arquitecturas (VGG y SqueezeNet). Para ello, primero se preentrena la red con la base de datos de ImageNet, luego utilizan las bases de datos de MNIST y UJIPenchars2, para finalmente afinar la arquitectura con la tarea objetivo, en este caso los conjuntos de datos de escritura de pacientes con Parkinson PaHaw y NewHandPD. El mejor desempeño fue obtenido con la tarea de "meandros" de la base de datos NewHandPD con una exactitud del 94,7%. En [8], los autores realizaron una comparación entre características hechas a mano y características basadas en CNN para la identificación de la disgrafía mediante los dibujos de la espiral de Arquímedes y la escritura de frases. Con base en los datos dinámicos se calcularon características cinemáticas, dinámicas, espaciales y temporales. Además, para el enfoque de CNN, se utilizó una red VGG16 preentrenada con la base de datos ImageNet, la cual es usada como extractor de características para las imágenes generadas por los sujetos. El conjunto de datos utilizado consta de 143 pacientes con PD y 151 controles sanos de diferentes ubicaciones geográficas como: Republica Checa, Hungría, Estados Unidos de América y Colombia. Los investigadores reportaron exactitudes de 82% para las características hechas a mano usando como clasificador un algoritmo XGBoost y una exactitud de 77% para las características basadas en la CNN, ambos resultados cuando se utilizó el conjunto de datos de USA.

De acuerdo con la literatura revisada, muchos de los estudios utilizan tareas como el dibujo de figuras geométricas y la espiral de Arquímedes, sin embargo, estas

tareas están específicamente diseñadas para la evaluación de los déficits motores de los pacientes y no son frecuentes en la vida cotidiana de las personas, lo que provoca un sesgo en la evaluación. Por otro lado, el análisis de la enfermedad de Parkinson ha sido poco abordado con tareas de escritura cotidiana como escribir un número telefónico o el número de identificación. En este trabajo se propone un enfoque basado en técnicas clásicas de procesamiento de imágenes y otro utilizando arquitecturas de aprendizaje profundo con el objetivo de analizar la escritura de los pacientes con PD cuando escriben los dígitos del cero al nueve en una tableta. En el enfoque clásico, se implementarán técnicas como HOG y LBP. En el enfoque de aprendizaje profundo, serán consideradas arquitecturas basadas en CNNs y RNNs con el fin de analizar la información estática y dinámica asociada al proceso de escritura de los dígitos. Ambos enfoques serán comparados según el desempeño para clasificar pacientes con PD a partir de una tarea de escritura no intrusiva.

4. Objetivos

4.1. Objetivo general

Diseñar y evaluar una metodología basada en algoritmos de aprendizaje automático para clasificar pacientes con enfermedad de Parkinson a partir de tareas no intrusivas de escritura.

4.2. Objetivo específicos

- Implementar y evaluar una metodología basada en técnicas clásicas de procesamiento de imágenes como LBP o HOG para la caracterización y clasificación de pacientes con enfermedad de Parkinson.
- Diseñar y evaluar una metodología basada en arquitecturas de aprendizaje profundo como las CNN o las RNN para la caracterización y clasificación de pacientes con enfermedad de Parkinson a partir de las imágenes generadas en las tareas no intrusivas de escritura a mano.
- Comparar y seleccionar los enfoques y métodos con mejor rendimiento para la caracterización y clasificación de pacientes con la enfermedad de Parkinson a partir de métricas de desempeño como la exactitud, la precisión, la sensibilidad, entre otras.

5. Marco teórico

5.1. Métodos de extracción de características clásicos

El análisis de la escritura puede proporcionar valiosa información para la clasificación de la enfermedad de Parkinson. En esta sección, exploraremos algunos métodos tradicionales de procesamiento de imágenes mediante los cuales se pueden detectar patrones distintivos en la escritura de pacientes con Parkinson. Estos métodos se enfocan en extraer características de la morfología de las letras y detectar diferentes irregularidades que se puedan presentar en la escritura de los pacientes. Además, a diferencia de aquellas técnicas basadas en aprendizaje profundo, para la implementación de estas técnicas clásicas no es necesario poseer una gran cantidad de datos y generalmente tiene mayor interpretabilidad, lo cual puede brindar un mejor entendimiento de los patrones de escritura asociados a pacientes con la enfermedad de Parkinson.

5.1.1. Patrones Binarios Locales (LBP, del inglés Local Binary Patterns)

LBP básico: es un método de extracción de características de textura el cual ha sido utilizado en tareas de detección facial [18] y verificación de firmas [19]. Para la implementación de esta técnica primero se toma un bloque de 3×3 píxeles de una imagen. El pixel central de este bloque se utiliza como un umbral para realizar comparaciones entre las escalas de grises de este pixel con los píxeles vecinos en el mismo bloque. Así, si el pixel vecino es mayor o igual al pixel central a este se le asigna el valor de 1, de lo contrario se le asigna 0. Luego la matriz que contiene los valores binarios resultantes de la comparación se multiplica por una matriz de pesos donde a cada posición se le asigna una potencia de 2, los cuales se distribuyen como se muestra en la [Figura 1](#). Finalmente se realiza la sumatoria de los valores en la matriz resultante y se le asigna el resultado en la posición de pixel central en la nueva imagen. Posteriormente la celda se desplaza a otra porción de la imagen y se realiza el cálculo nuevamente. En la [Figura 1](#) se observa un ejemplo del procedimiento mencionado [20].

Luego de realizar la codificación de la imagen se obtiene un histograma con los valores resultantes. Los valores del histograma corresponden al vector de características con una longitud de 256 (2^8) que son el número total de barras que se pueden obtener con 8 píxeles. Finalmente, en este vector se tiene una representación de las texturas que contiene la imagen.

LBP uniforme: Es una extensión del descriptor original (LBP básico) que se conoce como *LBP patrones uniformes* [21]. La medida de uniformidad de los patrones

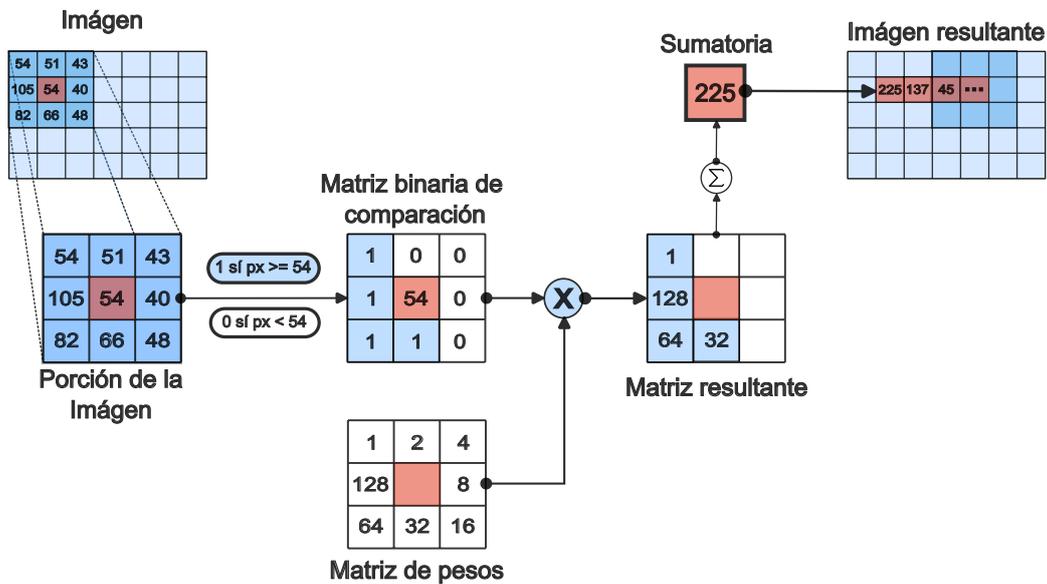


Figura 1. Proceso para el descriptor LBP básico.

es el número de transiciones bit a bit de 0 a 1 y viceversa cuando el patrón de bits se considera circular. De este modo un patrón es llamado uniforme si la secuencia de bits tiene como máximo dos transiciones. Por ejemplo, los patrones 11111111 (0 transiciones), 00001110 (2 transiciones) y 11110011 (2 transiciones) son considerados patrones uniformes. Mientras que los patrones 01101100 (4 transiciones) y 01010110 (6 transiciones) no serían uniformes. En este descriptor existe una categoría separada para cada patrón uniforme y todos los posibles patrones no uniformes se asignan a esta sola categoría. Así, el número de categorías resultantes para el mapeo de los patrones de P bits es: $P(P - 1) + 3$. Por ejemplo, un bloque de 3×3 pixeles donde se tienen 8 bits vecinos, tendrá 59 categorías de salida las cuales formarán el vector de características obtenido mediante este método. Finalmente, los patrones uniformes que se pueden detectar mediante el descriptor LBP incluyen: puntos, fondo, bordes, extremos de bordes, curvas, etc. Algunos ejemplos de estas transiciones uniformes son mostrados en la [Figura 2](#).

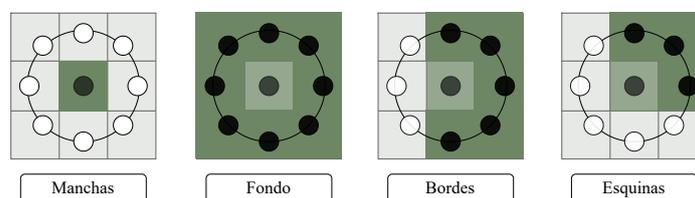


Figura 2. Patrones uniformes para LBP.

5.1.2. Histograma de Gradientes Orientados (HOG)

HOG es conocido como un descriptor basado en el gradiente de la imagen el cual se enfoca en la estructura y la forma de los objetos [22]. En este proceso el gradiente puede ser definido como la dirección del cambio en los niveles de intensidad de una imagen, así, cuando existen grandes variaciones se pueden detectar los bordes de los objetos presentes en dicha imagen.

Para realizar la extracción de características mediante HOG, primero se calculan los gradientes de la imagen en las direcciones x y y . Este cálculo se realiza normalmente mediante la operación de convolución de un filtro sobre toda la imagen. Existen diferentes filtros con los cuales se puede calcular los gradientes, algunos son los filtros de Roberts, los filtros de Prewitt y los filtros de Sobel los cuales se muestran en la Figura 3 [23].

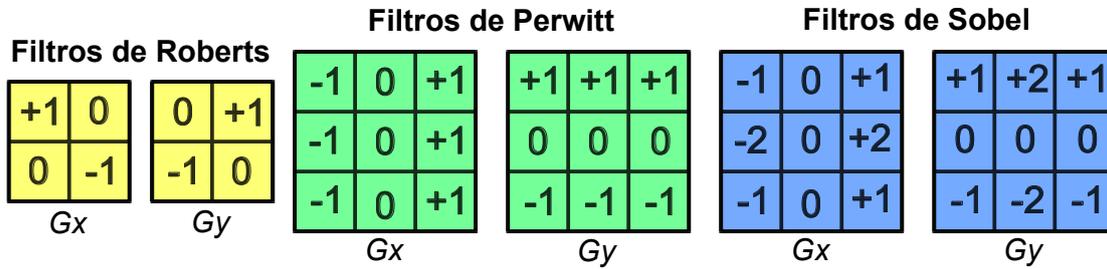


Figura 3. Filtros de Roberts, Prewitt y Sobel para el cálculo de los gradientes en las direcciones X y Y .

Mediante los gradientes obtenidos se debe calcular las matrices de magnitud y fase. Para la primer matriz, cada pixel representa el valor de magnitud del gradiente en esa posicon, el cual se calcula mediante (1). En la matriz de fase cada pixel representa el angulo del gradiente en esa posicon, el cual se calcula mediante (2) [24].

$$Magnitud(\mu) = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$Angulo(\theta) = \arctan\left(\frac{G_y}{G_x}\right) \quad (2)$$

En este descriptor, se suele utilizar un numero definido de direcciones el cual permite discretizar los ángulos de los gradientes y determinan el número de barras en el histograma. Típicamente, se suelen utilizar 8 direcciones diferentes como se observa en la Figura 4, donde cada dirección representa con un rango de 45° . Para construir el histograma de gradientes, se debe tener en cuenta los pixeles en cada una de las direcciones y se suman las magnitudes correspondientes a dichos pixeles. En este proceso

cada histograma (Figura 4) nos da un vector de 8 valores los cuales representan la información de los gradientes en cada una de las direcciones. Luego para el proceso de extracción de características, la imagen objetivo se divide en celdas de tamaño $p \times p$ píxeles, Figura 5, y por cada celda se calcula un histograma. Así, los histogramas de todas las celdas se concatenan para obtener el vector de características.

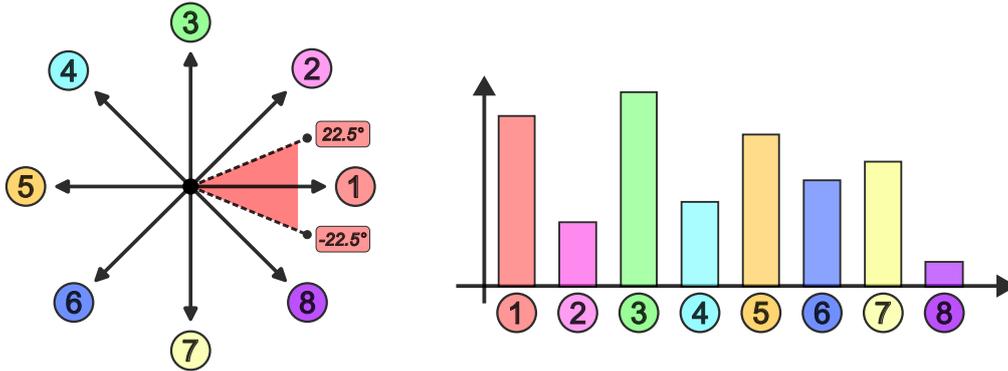


Figura 4. Direcciones de los gradientes y su histograma correspondiente.

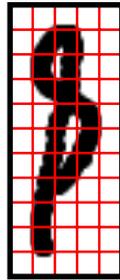


Figura 5. Imagen objetivo, dividida en celdas.

Finalmente, el vector de características se debe normalizar utilizando la norma \mathcal{L}_2 para reducir el efecto de los cambios de contraste entre las imágenes de un mismo objeto. Para realizar la normalización se toman bloques de $c \times c$ celdas y se opera el vector de características mediante (3) donde \mathbf{v} es el vector de características sin normalizar y ϵ una constante pequeña [22].

$$\mathbf{v} \rightarrow \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}} \quad (3)$$

5.2. Métodos de extracción de características profundos

En los últimos años, las técnicas de aprendizaje profundo han revolucionado el campo del procesamiento de imágenes y el análisis de datos secuenciales. Estas técnicas han demostrado ser especialmente efectivas en la detección y clasificación de patrones complejos en diferentes dominios. En el caso del análisis de la enfermedad de Parkinson mediante datos de escritura a mano, se han implementado diferentes arquitecturas como las CNN y las RNN, las CNN son especialmente adecuadas para extraer características espaciales de las imágenes, mientras que las RNN son efectivas para modelar la estructura secuencial de los datos. Así, al utilizar técnicas de aprendizaje profundo, se logra un enfoque más automatizado y preciso para el análisis de la escritura a mano en relación con la enfermedad de Parkinson. Estas redes son capaces de aprender características relevantes directamente de los datos y generar modelos que pueden clasificar y detectar patrones distintivos en las muestras de escritura.

5.2.1. Redes Neuronales Convolucionales (CNN)

Las CNNs son una de las redes más populares y utilizadas del Deep Learning ya que una de sus principales ventajas comparadas con las técnicas clásicas es su capacidad de obtener características relevantes sin supervisión de un experto. Este tipo de redes también han sido ampliamente usadas en aplicaciones como el reconocimiento de objetos y el análisis de imágenes médicas para la detección de enfermedades [25]. Las CNNs son arquitecturas que reciben como entrada diversos tipos de matrices, donde generalmente representan imágenes, y es capaz de capturar diferentes tipos de dependencias espaciales [26].

Capa convolucional: es la capa principal de la CNN en la cual se aplica la operación de convolución entre un kernel y la imagen de entrada, obteniendo así lo que se denomina el mapa de características. Los valores del kernel son parámetros que la red aprende durante el proceso de entrenamiento y dependiendo de estos valores se obtendrán diferentes mapas de características al aplicar la convolución. A medida que se agregan capas convolucionales a la red, la imagen inicial se va comprimiendo espacialmente mientras la cantidad de mapas de características va aumentando, así, se van detectando patrones más complejos y extrayendo representaciones de más alto nivel [26].

Stride y Padding: el parámetro que controla el deslizamiento del kernel a través de la imagen se denomina stride, este indica cuantos pixeles se avanza en cada iteración. Por otro lado, el padding determina la cantidad de pixeles con valor igual a cero que se agregarán alrededor de la imagen de entrada para controlar así el tamaño del

mapa de características a la salida [26].

Capa de pooling: la capa de pooling generalmente se aplica luego de una capa convolucional con la idea de reducir el volumen espacial del mapa de características obtenido. Esta operación también ayuda a disminuir la complejidad computacional de la red mediante la reducción de dimensionalidad de los datos, además es útil para extraer características dominantes sin alterar el proceso de entrenamiento. Esta capa cuenta con dos parámetros, el kernel y el stride. En este caso el kernel determina la cantidad de píxeles que serán reducidos a un solo valor mediante la operación de pooling [26]. Existen diferentes tipos de pooling, entre los más destacados están el max pooling y el average pooling. El max pooling retorna el valor máximo de la porción de la imagen que está cubierta por el kernel, mientras que el average pooling retorna el promedio de dichos valores.

Capa totalmente conectada (capa densa): Esta capa corresponde a la etapa de clasificación de la CNN. Además, las capas densas son una forma de aprender combinaciones no lineales de las características de alto nivel representadas por la salida de la capa convolucional. Para ello, el último mapa de características es transformado en un vector concatenando las filas o columnas de la matriz; a este proceso se le conoce como aplanado o flatten. Comúnmente a la salida de la capa densa existen tantas neuronas como clases a predecir [26].

5.2.2. Redes Neuronales Recurrentes (RNN)

Las RNN son una clase de redes neuronales que son capaces de modelar datos secuenciales. Estas redes han sido ampliamente utilizadas para el procesamiento de series temporales como: texto, habla, datos financieros, audio, video, entre otros [27]. En estas redes las salidas de cada neurona dependen del estado actual y de la salida del estado anterior, además, los instantes de tiempo son definidos por la cantidad de elementos dentro de la secuencia de datos.

Cada bloque (instante de tiempo) de la red cuenta con una entrada (x_t), una salida (y_t), el estado oculto anterior (h_{t-1}) y el estado oculto actual (h_t). Estos estados ocultos son los que le transportan y guardan la información entre diferentes instantes de tiempo. En la [Figura 6](#) se tiene un esquema de una RNN donde se observa el flujo de la información entre diferentes instantes de tiempo [28]. La salida de la red recurrente (y_t) y el estado oculto (h_t) están dados por:

$$h_t = f_1(W_{hh} \cdot h_{t-1} + W_{hx} \cdot x_t + b_h) \quad (4)$$

$$y_t = f_2(W_{yh} \cdot h_t + b_y) \quad (5)$$

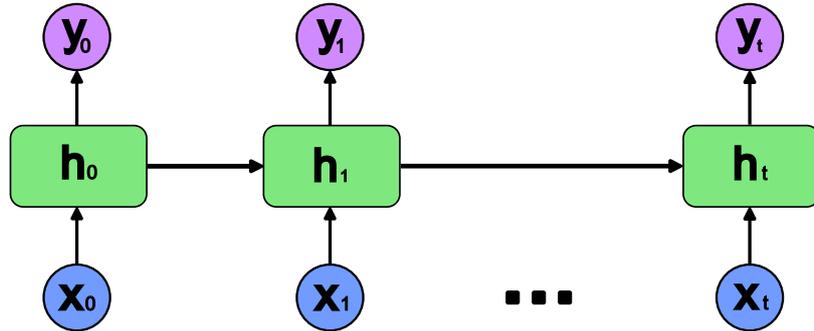


Figura 6. Esquema general de una red neuronal recurrente.

Donde W_{hh} , W_{hx} , W_{yh} son los pesos y b_h y b_y los sesgos de la red, los cuales se aprenden durante el proceso de entrenamiento. Las funciones de activación de la red están dadas por f_1 y f_2 , donde generalmente se suelen utilizar funciones \tanh y softmax , respectivamente.

5.3. Algoritmos de clasificación

5.3.1. Maquinas de Soporte Vectorial (SVM)

Las SVM son un algoritmo de Machine Learning las cuales han sido desarrolladas como una herramienta para la clasificación y regresión en diferentes tareas complejas, como el análisis de texto, análisis de escritura a mano, el reconocimiento de rostros y objetos, entre muchos otros [29]. Para la clasificación las SVM construyen un hiperplano de separación que se utiliza como un umbral de decisión para establecer a que clase pertenece determinado dato. Los puntos de cada clase más cercanos al hiperplano se denominan vectores de soporte, a su vez, la distancia entre los vectores de soporte de ambas clases determina el margen de separación [30]. De acuerdo con el algoritmo de SVM, se busca que el margen de separación maximice la distancia entre las clases. Por otro lado, los conjuntos de datos normalmente no son linealmente separables, por lo tanto, se hace uso de las funciones kernel las cuales permiten transformar el espacio de características inicial en uno de mayor dimensión donde las clases sean linealmente separables y se maximice la separación entre las clases [30]. Un ejemplo de esta transformación se muestra en la [Figura 7](#).

Además de las transformaciones, cuando los conjuntos de datos no son linealmente separables se suele definir un margen blando el cual permite una pequeña cantidad de errores. En la [Figura 8](#), se observa una representación de un hiperplano de separación con un margen blando. Los datos que se encuentren dentro de este margen serán penalizados, sin embargo, esto permitirá que la SVM sea tolerante al ruido e incremente la capacidad de generalización en el sistema [30].

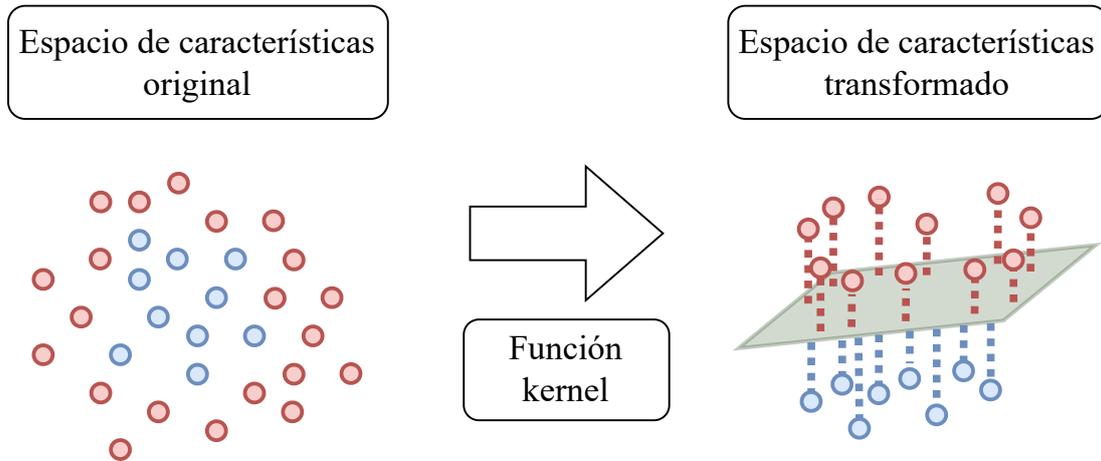


Figura 7. Transformación del espacio de características mediante la función kernel.

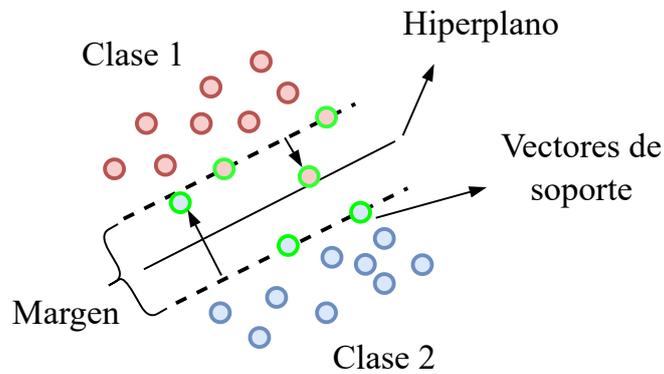


Figura 8. SVM de margen blando.

Además, existen dos parámetros que controlan el comportamiento del margen. El parámetro (C) determina el balance entre la dimensión del margen y los puntos penalizados, así cuanto más grande sea el parámetro (C) el margen de decisión estará más ajustado a los datos de entrenamiento. Por otro lado, el parámetro (γ), el cual se utiliza para las funciones kernel radial, polinomial y sigmoideal, determina la influencia de los datos en el proceso de entrenamiento. De este modo para valores altos de (γ) el límite de decisión se ajustará a los datos de entrenamiento, de lo contrario, para valores pequeños de (γ) el hiperplano tiende a ser más lineal [30].

5.4. Métricas de Desempeño

Para medir el desempeño de un sistema es necesario cuantificar la capacidad de discriminación entre las diferentes clases para saber que tan aproximadas son sus predicciones a los datos reales. En problemas de clasificación existen métricas importantes como la matriz de confusión y la curva ROC.

5.4.1. 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 [31]. 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 2×2 donde en cada celda se pueden definir los siguientes términos [31], [32]:

- *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.

La [Figura 9](#) muestra una matriz de confusión bi-clase con la identificación de cada celda basado en los términos previamente definidos.

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.

$$acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

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

$$S = \frac{TP}{TP + FN} \quad (7)$$

		Clase real	
		positiva	negativa
Clase predicha	positiva	TP	FP
	negativa	FN	TN

Figura 9. Matriz de confusión.

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

$$E = \frac{TN}{TN + FP} \quad (8)$$

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

$$P = \frac{TP}{TP + FP} \quad (9)$$

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

$$F1 - score = 2 \times \frac{P \times S}{P + S} \quad (10)$$

5.4.2. Curva ROC

La curva ROC (Receiver Operating Characteristic) es una representación gráfica que nos permite evaluar el desempeño de un modelo [33]. Esta curva se construye graficando la tasa de verdaderos positivos (S), versus la tasa de falsos positivos ($1 - E$), a medida en que se desplaza el umbral de decisión. Un ejemplo de la curva ROC se observa en la [Figura 10](#). Por otro lado, existe una métrica que se puede obtener mediante la curva ROC la cual se denomina, area bajo la curva ROC (AUC). Esta

es una medida cuantitativa que indica la capacidad del del modelo para discriminar entre dos clases. Además, esta métrica tiene un rango de $[0.5,1]$ siendo 1 el mejor desempeño posible donde el sistema logra clasificar todos los datos correctamente y 0,5 el mínimo valor el cual indica que el clasificador está prediciendo las calses aleatoriamente. De la [Figura 10](#), en la parte de abajo de la curva ROC se observa el valor AUC correspondiente.

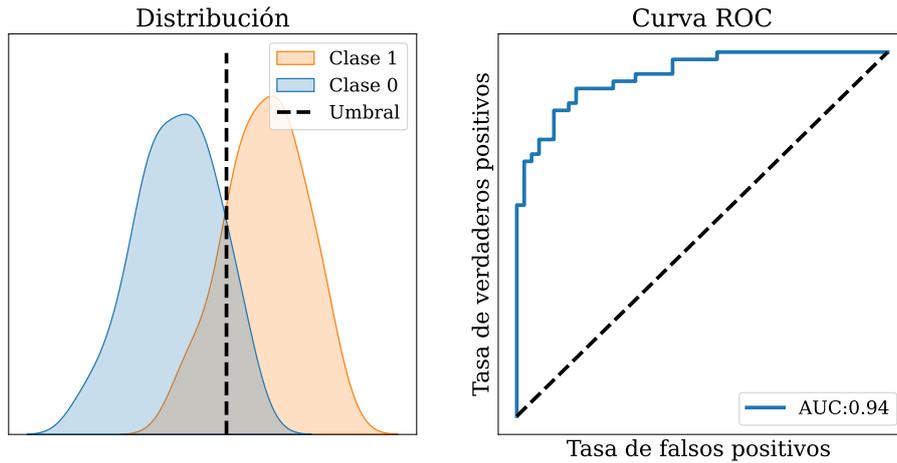


Figura 10. Ejemplo ilustrativo de la curva ROC.

6. Base de datos

Los datos para este estudio fueron recolectados mediante una tableta digitalizadora *Wacom Cintiq 13 HD* con una frecuencia de captura de $180Hz$. La tablet permite la captura de 6 señales diferentes: las posiciones (x, y) , la altura del lápiz a la superficie de la tableta eje (z) , la presión del lápiz, el ángulo azimut, y la inclinación del lápiz (ángulo altitud). Para este estudio se tuvieron en cuenta 110 sujetos en total de los cuales 55 son pacientes con PD y para 47 de ellos se cuenta con la evaluación neurológica de acuerdo con la escala *MDS-UPDRS-III* [34]. La información general de los participantes se encuentra resumida en la [Tabla 1](#).

Tabla 1. Información de los participantes.

	Pacientes con PD		Controles sanos	
	Hombres	Mujeres	Hombres	Mujeres
Número de sujetos	26	29	33	22
Edad ($\mu \pm \sigma$)	68.8 \pm 10.3	62.2 \pm 11.6	67.0 \pm 10.4	58.7 \pm 10.5
Rango de edad	50 - 90	29 - 84	49 - 85	43 - 83
MDS-UPDRS-III ($\mu \pm \sigma$)	38.0 \pm 16.7	32.3 \pm 15.7		
Rango de MDS-UPDRS-III	16 - 82	14 - 77		

7. Metodología general

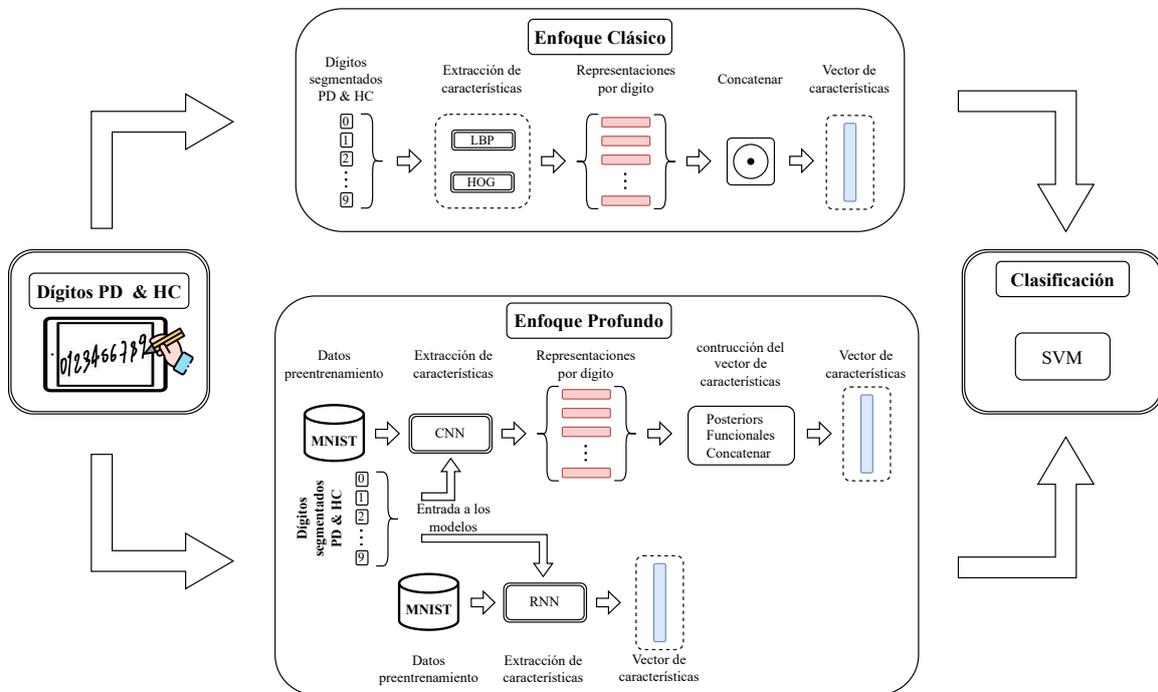


Figura 11. Metodología general.

La metodología general abordada en este estudio consiste en 2 enfoques para la caracterización de las imágenes de los pacientes con PD y controles sanos. En primer lugar, se tiene el enfoque clásico en el cual se implementarán dos técnicas diferentes para la extracción de características. El descriptor LBP que permite identificar texturas en la imagen y el descriptor HOG que permite reconocer los bordes de los objetos presentes en una imagen. En segundo lugar, para el enfoque profundo se implementan

dos arquitecturas para la caracterización. Una CNN, donde la red se preentrena para reconocer dígitos de la base de datos MNIST y de la cual se obtendrán representaciones vectoriales de los dígitos de los sujetos. Además, se tiene la arquitectura basada en RNN en el cual la red se preentrena para detectar secuencias de los números del 0 al 9 de MNIST y posteriormente se ajusta para procesar las secuencias de dígitos de los participantes y obtener representaciones para estos. Finalmente se emplean el algoritmo SVM para realizar la clasificación utilizando los vectores de características obtenidos de los enfoques propuestos. Un resumen de la metodología abordada en este trabajo se encuentra en la [Figura 11](#).

7.1. Enfoque clásico

Para este enfoque se tienen en cuenta dos descriptores que serán utilizados para la extracción de características, LBP y HOG. En primer Lugar, para LBP se implementaron dos métodos diferentes, LBP básico y LBP uniforme los cuales fueron descritos en la sección 5.1.1. En el caso de LBP básico, cada dígito segmentado tendrá una representación vectorial de 256 valores correspondientes al histograma de salida. Por otro lado, en LBP uniforme se tendrán representaciones con una dimensión de 59 por cada dígito. Además, en ambos métodos de LBP se utilizó un bloque de 3×3 que corresponde a un radio igual a 1 con 8 bits vecinos para la comparación con el pixel central.

Para el segundo descriptor que corresponde a HOG, el cual se introdujo en la sección 5.1.2, se utilizaron celdas de 4×4 y 7×7 pixeles con bloques de 1×1 celdas y 9 direcciones para la construcción de los histogramas. Cuando se utilizan celdas de 4×4 pixeles se tienen representaciones con una dimensión de 144 por dígito, mientras que con una celda de 7×7 se tendrán representaciones de 441 valores por dígito.

7.2. Enfoque basado en redes neuronales profundas

Para el enfoque profundo se implementaron dos arquitecturas para obtener representaciones de los dígitos de los participantes, una CNN y una RNN. En primer lugar, para la CNN se toma una arquitectura LeNet utilizada para la tarea de clasificación de los dígitos del 0 al 9 de la base de datos MNIST, el criterio de selección consiste en la arquitectura más simple con un buen desempeño en la tarea propuesta, con base en los modelos probados en [35]. La arquitectura seleccionada cuenta con 3 capas convolucionales con 8, 16 y 32 filtros respectivamente, además, cada capa tiene un kernel de 5×5 y una función de activación *ReLU*. Cada capa convolucional se encuentra conectada a una capa maxpooling que cuenta con un kernel de 2×2 y un stride igual a 2 que reduce el mapa de características a la mitad de su dimensión inicial. Posteriormente el mapa de características de la última capa convolucional se

aplana y se conecta a una capa densa (completamente conectada), la cual tiene una capa intermedia de 64 neuronas con una función de activación *ReLU* y una capa de salida de 10 neuronas con una función de activación *Softmax*, esta entregará un vector de 10 posiciones donde cada posición corresponde a la probabilidad de que la imagen de entrada sea un dígito del cero al nueve. Un esquema de esta arquitectura es mostrado en la [Figura 12](#). Finalmente, esta arquitectura es pre-entrenada con la base de datos MNIST, lo que permite extraer representaciones de los dígitos de los pacientes y controles.

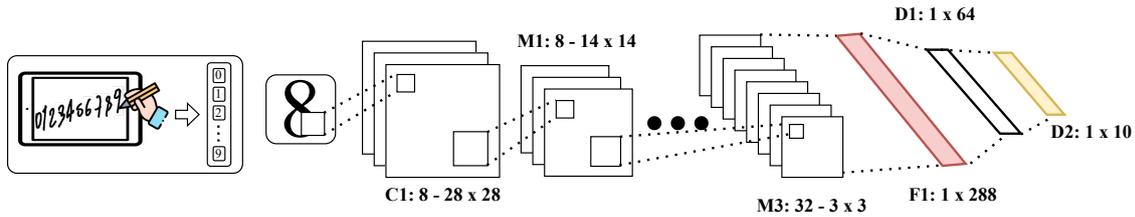


Figura 12. Arquitectura de la CNN.

Para el segundo método de extracción de características del enfoque profundo se utilizó una arquitectura RNN. Con esto se quiere modelar la escritura de los dígitos ya que estos fueron escritos secuencialmente del 0 al 9 por los participantes. del mismo modo como en la CNN, la RNN requiere ser preentrenada para caracterizar los dígitos de los sujetos y para esto se utilizó nuevamente la base de datos MNIST. En este caso se agrupan dígitos al azar del 0 al 9 para simular las secuencias escritas por los pacientes y controles de nuestra base de datos, la mitad de las secuencias se modifican degenerando las imágenes para crear secuencias correctas y secuencias modificadas. Un ejemplo de las secuencias se encuentra en la [Figura 13](#).

Antes de ingresar a la RNN las imágenes deben ser preprocesadas, para ello se utiliza el modelo de CNN preentrenado del método anterior de donde se obtienen representaciones vectoriales (embeddigs) de los dígitos en cada secuencia, estas representaciones se basan en la salida aplanada de la tercera capa convolucional de la red (bloque rojo de la [Figura 12](#)). Luego, los embeddings de cada dígito de la secuencia son usados como entrada de la RNN, la cual se pre-entrena para reconocer secuencias buenas y malas. El siguiente paso es obtener las representaciones para los dígitos de los pacientes y controles. Para esto primero se realiza el mismo preprocesamiento que se utilizó para las secuencias de dígitos creadas con MNIST, obteniendo por cada participante el arreglo de embeddings de la CNN para sus secuencias de dígitos. Luego estos arreglos son procesados por la RNN donde se toma el estado oculto de salida como una representación vectorial (embedding) que modela la escritura de los dígitos

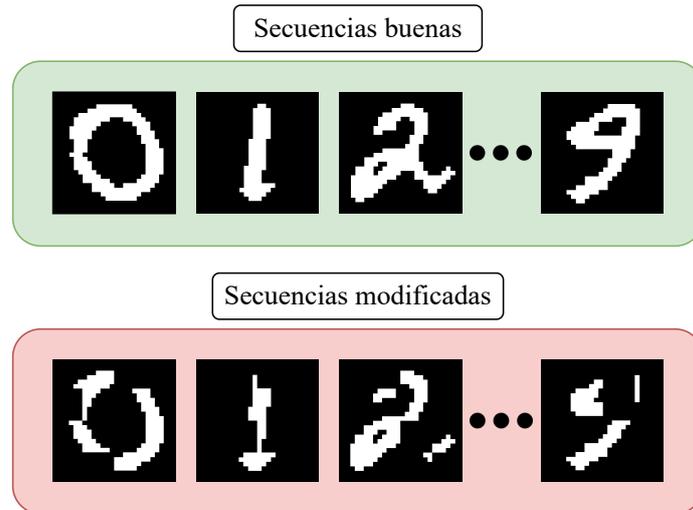


Figura 13. Secuencias de preentrenamiento para la RNN.

de los participantes secuencialmente. Aquí se utilizó una arquitectura que consta de una capa recurrente con 64 unidades y a una capa densa con dos neuronas en la capa de salida. Una representación gráfica de la arquitectura y el proceso realizado se encuentra en la [Figura 14](#).

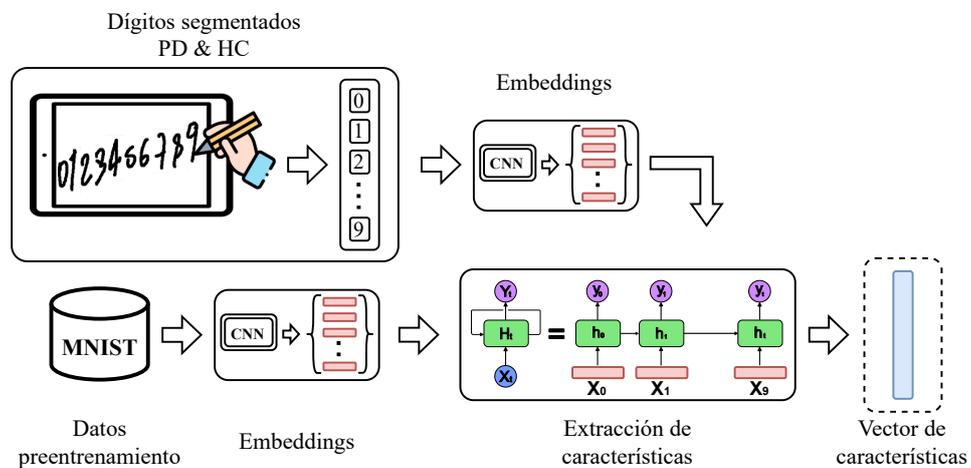


Figura 14. Arquitectura RNN.

7.3. Optimización de los parámetros y clasificación

Para todos los experimentos realizados en este trabajo se utilizó una estrategia de validación cruzada de 5 folds los cuales fueron predefinidos y estratificados para las dos clases, pacientes con PD y controles sanos. Además, se utilizó una SVM con kernel lineal para la clasificación donde el parámetro de costo C fue optimizado mediante una malla de búsqueda. El conjunto de valores del parámetro C se encontraba en potencias de 10 entre 0.0001 y 1000 ($C \in \{0,0001, 0,001, \dots, 1000\}$). Adicionalmente se utilizaron diferentes kernels para la SVM como el kernel gaussiano y el sigmoidal.

8. Experimentos y Resultados

En este trabajo se consideraron 3 tipos de experimentos: (1) las características obtenidas mediante los métodos clásicos de LBP y HOG, (2) las características extraídas de los modelos basados en aprendizaje profundo, CNN y RNN. (3) basados en las características con mejor desempeño de cada enfoque, se realiza un experimento agregando características calculadas a partir de señales dinámicas usando una estrategia de fusión temprana (del inglés, early fusion strategy). Esto con el fin de observar si existe información complementaria entre las características obtenidas de las imágenes y las características basadas en los datos dinámicos de escritura.

8.1. Resultados enfoque clásico

Para los experimentos correspondientes al enfoque clásico se tienen las características basadas en los descriptores LBP y HGO como se menciona en la sección 7.1. Para formar el vector de características por sujeto se concatenan las representaciones extraídas por cada dígito y luego se procede a realizar la validación cruzada con cada conjunto de características obteniendo finalmente 4 experimentos: (1) LBP básico y (2) LBP uniforme, (3) HOG-(4 × 4) y (4) HOG-(7 × 7). los resultados para los experimentos mencionados se encuentran en la [Tabla 2](#).

Tabla 2. Resultados obtenidos para el conjunto de experimentos del enfoque clásico. Valores en términos de ($\mu \pm \sigma$) y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).

Experimento	Exactitud(%)	Especificidad(%)	Sensibilidad(%)	F1-Score	AUC	\hat{C}
LBP-Básico	65.5 ± 8.4	60.0 ± 21.2	70.9 ± 10.6	0.67 ± 0.07	0.65	1×10^{-3} (5/5)
LBP-Uniforme	61.8 ± 10.6	56.4 ± 17.6	67.3 ± 12.3	0.64 ± 0.10	0.67	1×10^{-3} (4/5)
HOG-(4 × 4)	72.7 ± 11.1	72.7 ± 15.2	72.7 ± 16.3	0.72 ± 0.11	0.80	1×10^{-3} (5/5)
HOG-(7 × 7)	67.3 ± 12.3	70.9 ± 17.6	63.5 ± 17.2	0.66 ± 0.14	0.75	1×10^{-3} (3/5)

De manera general, los resultados reportados en la [Tabla 2](#) muestran que las

características basadas en el descriptor HOG logran mejores resultados que aquellas obtenidas con en el descriptor LBP. Esto puede deberse ya que mediante HOG se detectan los bordes de los objetos presentes en la imagen, en este caso los dígitos escritos por los participantes, y así se pueden identificar irregularidades en la escritura causadas por la enfermedad del Parkinson. Ahora en cuanto a LBP, este es un método que permite detectar diferentes texturas en las imágenes mediante las variaciones en la intensidad de los píxeles, y dado que se está trabajando con dígitos cuyas imágenes se encuentran en escalas de grises pueden no existir diferencias significativas entre pacientes y personas sanas. Finalmente, las características de HOG calculadas con una celda de 4×4 píxeles obtuvieron el mejor resultado para este enfoque. Esto se puede dar ya que la celda de 4×4 , al tratarse de porciones más pequeñas en la imagen, brinda una mejor resolución capturando los trazos del dígitos. En la [Figura 15](#) se observa la curva ROC donde se evidencia gráficamente el incremento en el AUC respecto a la exactitud y las distribuciones de los scores de la SVM para este experimento, donde se refleja el balance entre la especificidad y la sensibilidad.

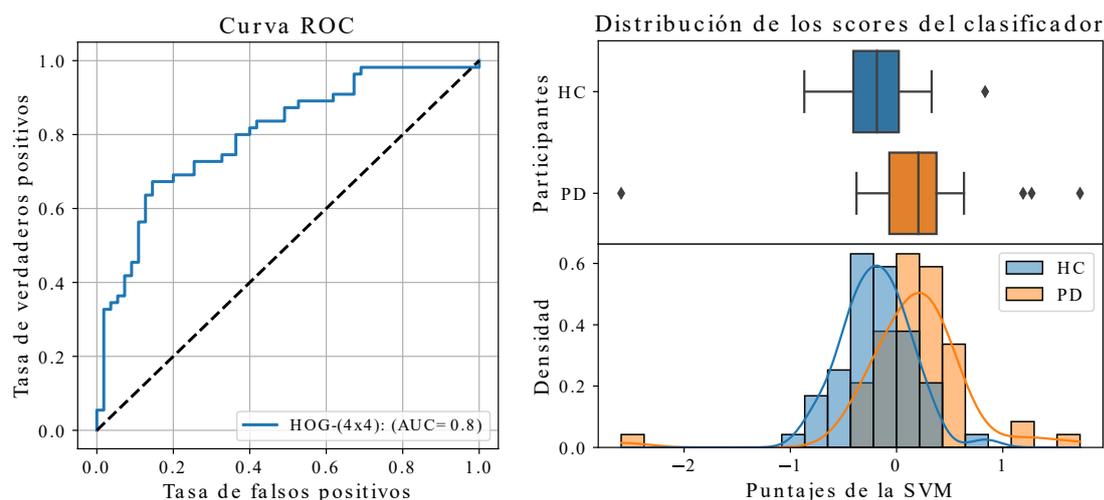


Figura 15. Curva ROC y distribución de los scores para el mejor resultado del enfoque clásico.

8.2. Resultados del enfoque basado en aprendizaje profundo

Para el enfoque de aprendizaje profundo se proponen 3 experimentos basados en CNNs y 2 experimentos basados en RNN. Los experimentos basados en CNN son: (1) CNN-posteriors, los cuales son las 10 predicciones a la salida de la red (bloque amarillo

de la [Figura 12](#)) hechas para cada dígito, luego para formar el vector de características se concatenan las 10 representaciones por participante. (2) CNN-CatEmbb, estas características se basan en las representaciones o embeddings obtenidos de la salida aplanada de la tercera capa convolucional (bloque rojo de la [Figura 12](#)). Este vector cuenta con una dimensión de 288 y aquí las representaciones de cada dígito se concatenan formando un vector de características de 2880 valores por participante. (3) CNN-FuncEmbb, estas características también se basan en los embeddings del experimento anterior pero aquí las representaciones por cada dígito se apilan verticalmente y se extraen 6 funcionales por cada columna (media, desviación estándar, curtosis, Skewness, máximo y mínimo) formando finalmente un vector de 1728 características por sujeto. En segundo lugar, se tienen las caracterizaciones provenientes del modelo RNN. (4) RNN-congelada, donde solamente se extraen las representaciones por cada sujeto de la RNN preentrenada y (5) RNN-afinada, en donde se afina la capa recurrente de la arquitectura RNN con el conjunto de entrenamiento de la base de datos de Parkinson. Los resultados de los experimentos mencionados se muestran en la [Tabla 3](#).

Tabla 3. Resultados obtenidos para el conjunto de experimentos del enfoque basado en arquitecturas de aprendizaje profundo. Valores en términos de $(\mu \pm \sigma)$ y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).

Experimento	Exactitud(%)	Especificidad(%)	Sensibilidad(%)	F1-Score	AUC	\hat{C}
CNN-Posteriors	51.8 \pm 7.4	52.7 \pm 6.8	50.9 \pm 14.8	0.51 \pm 0.11	0.52	1×10^{-2} (3/5)
CNN-CatEmbb	71.8 \pm 4.5	74.5 \pm 3.6	69.1 \pm 9.3	0.71 \pm 0.06	0.75	1×10^{-3} (4/5)
CNN-FuncEmbb	55.5 \pm 12.3	60.0 \pm 13.6	50.9 \pm 17.8	0.52 \pm 0.16	0.60	1×10^{-3} (4/5)
RNN-Congelada	60.9 \pm 6.2	69.1 \pm 9.3	52.7 \pm 21.0	0.55 \pm 0.13	0.61	1×10^{-1} (4/5)
RNN-Afinada	63.6 \pm 7.6	65.6 \pm 6.8	61.8 \pm 15.6	0.62 \pm 0.11	0.68	1×10^{-2} (5/5)

Los resultados mostrados en la [Tabla 3](#) indican que las características basadas en los embeddings extraídos con la CNN (CNN-CatEmbb) muestran el mejor desempeño en la tarea de clasificación de pacientes con PD y controles saludables. Las probabilidades de la CNN (CNN-Posteriors), si bien tienen más interpretabilidad ya que son las predicciones realizadas por el modelo las cuales significan que tanto éste logró identificar el dígito, muestran el desempeño más bajo respecto a las otras caracterizaciones consideradas. Por otro lado, con las características CNN-embedd-funcionales se está perdiendo información útil para la caracterización al calcular los estadísticos de los embeddings en la transformación realizada. De este modo, existen ciertas ventajas en el conjunto de características CNN-embedd-concatenados y es el contenido de información presente en la red luego de que la imagen de entrada fue procesada por las capas convolucionales, así, a la hora de concatenar las representaciones, no se está perdiendo información significativa de los participantes. La [Figura 16](#) muestra una representación gráfica de los resultados donde se corroboran los resultados

obtenidos por este experimento. Ahora respecto al desempeño a las representaciones obtenidas mediante la arquitectura RNN, hubo una ligera mejora cuando se afinó la arquitectura con la base de datos de los pacientes, sin embargo, el desempeño aún está por debajo del mejor resultado obtenido con las representaciones provenientes de la CNN. Por otra parte, el modelo RNN permite analizar el proceso de la escritura de los participantes conjuntamente, sin embargo, la degeneración de las imágenes que corresponden a las secuencias modificadas no modela correctamente las anomalías que presentan los pacientes con Parkinson. En lugar de eso simplemente se enmascaran partes de la imagen para volver los dígitos poco reconocibles. La CNN no cuenta con ese problema de preentrenamiento ya que esta solo modela el reconocimiento de los dígitos uno a uno y luego del procesamiento de la imagen por las capas convolucionales estas condensan la información en un embedding que resulta ser útil para la caracterización de la tarea propuesta.

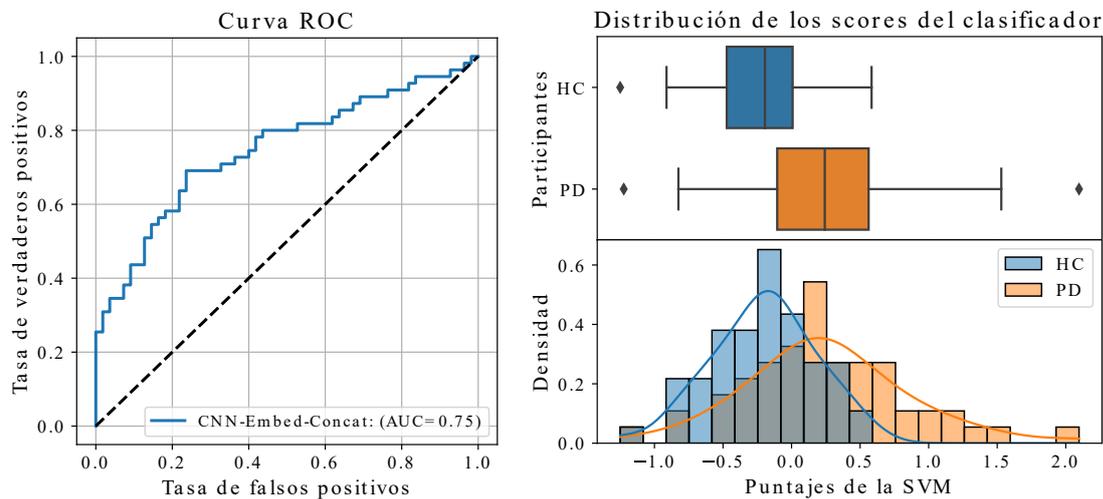


Figura 16. Curva ROC y distribución para el mejor resultado del enfoque profundo.

8.3. Resultados utilizando una estrategia de fusión

En estos experimentos se utilizan las mejores características de cada enfoque y mediante una técnica de fusión temprana se agrega a cada vector un conjunto de características calculadas a partir de las señales dinámicas. Estas características son extraídas de 4 señales provenientes de la tableta digitalizadora (ángulos azimut y altitud, presión y distancia de la tableta a la punta del lápiz). Adicionalmente para cada señal se calculan la primer y segunda derivada, así, para cada subconjunto de señales (señal cruda y sus derivadas) se calculan 9 estadísticos los cuales corresponden

a: media, desviación estándar, curtosis, skewness, máximo, mínimo y los percentiles 25th, 50th y 75th. Además, únicamente cada señal cruda se divide en 10 secciones iguales y se calcula la media de cada sección. Finalmente, se tendrán 37 medidas por cada subconjunto de señales y un vector de características dinámicas de 148 valores en total. De este modo en la [Tabla 4](#) se evidencian los resultados de los experimentos al combinar los dos tipos de características.

Tabla 4. Resultados obtenidos cuando se utiliza una estrategia de fusión temprana con los mejores resultados de cada enfoque y características basadas en señales dinámicas. Valores en términos de $(\mu \pm \sigma)$ y respecto al parámetro \hat{C} se tiene (valor (moda/folds)).

Experimento	Exactitud(%)	Especificidad(%)	Sensibilidad(%)	F1-Score	AUC	\hat{C}
HOG (4 × 4) + Dinámicas	74.5 ± 9.4	74.5 ± 12.1	74.5 ± 15.6	0.74 ± 0.10	0.82	0.001 (5/5)
CNN Concatenados + Dinámicas	74.5 ± 3.6	76.4 ± 7.6	72.7 ± 12.9	0.74 ± 0.06	0.78	0.001 (4/5)

La [Tabla 4](#) indica una mejora en los modelos del enfoque clásico y profundo cuando se considera la información dinámica. Aunque ambos enfoques presentan resultados similares, los resultados de las representaciones basadas en CNN cuentan con menor desviación estándar. Sin embargo, en la [Figura 17](#) el AUC de las representaciones de HOG es levemente mayor que para la CNN.

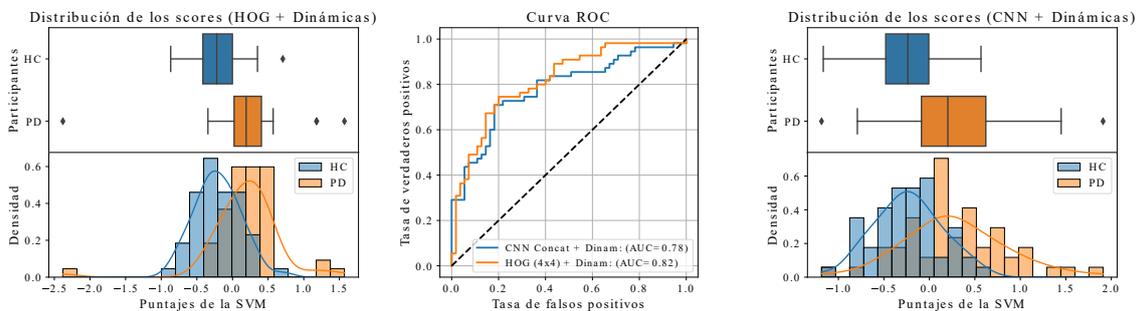


Figura 17. Curva ROC y distribución para la estrategia de fusión temprana.

9. Conclusiones

En este estudio se abordó el problema de la clasificación de pacientes con la enfermedad de Parkinson y personas sanas utilizando una tarea que se considera frecuente como lo es la escritura de números, la cual resulta ser más amigable con los pacientes que otras tareas que involucran trazos más complejos. Además, se implementaron dos

enfoques para la caracterización de las imágenes. Un enfoque clásico donde se utilizaron dos descriptores LBP y HOG, como también un enfoque basado en arquitecturas de aprendizaje profundo donde se empleó una CNN y una RNN. Respecto al enfoque clásico las representaciones obtenidas con el método de HOG con celdas de 4×4 y 9 orientaciones alcanzaron una exactitud del 72,7%. Mientras que, para el enfoque profundo, las representaciones obtenidas de los embeddings provenientes de la salida aplanada de la tercera capa convolucional lograron una exactitud del 71,8% siendo estos dos los mejores resultados en el enfoque clásico y profundo, respectivamente. Adicionalmente con estos dos conjuntos de características se realizó una estrategia de fusión temprana agregando información dinámica proveniente de otras señales capturadas por la tablet. En este experimento adicional se incrementó el desempeño de ambas representaciones en al menos 2%. Por otro lado, ya que el descriptor HOG detecta los bordes de los objetos presentes en una imagen, le es posible identificar anomalías en la escritura de los dígitos de los pacientes, superando así la caracterización mediante el descriptor LBP. Ahora en cuanto al enfoque profundo, cuando se extraen las representaciones de las predicciones hechas por la red o los funcionales calculados a los embeddings intermedios, se está perdiendo información útil que pueda diferenciar entre pacientes y controles. Además, respecto a las representaciones basadas en la arquitectura RNN, se evidenció la necesidad de determinar una tarea de preentrenamiento para la red que le permita adquirir un conocimiento significativo a la hora de modelar la información proveniente de los pacientes con la enfermedad de Parkinson.

Aunque los resultados obtenidos no están cerca de otros trabajos que analizan pacientes con PD mediante señales de escritura, se evidenció que los enfoques implementados en este trabajo configuran una metodología conveniente para la clasificación de pacientes y controles a partir de una tarea no intrusiva, lo cual la hace candidata natural a ser incluida en procesos de evaluación no invasivo y longitudinal de pacientes con PD. Además, es importante notar que muchos de estos trabajos del estado del arte utilizan tareas que están específicamente diseñadas para la evaluación de la patología como lo es el dibujo de la espiral de Arquímedes o el dibujo de patrones repetitivos. Este tipo de tareas que son más frecuentes en la vida cotidiana de las personas como la escritura de números telefónicos o la escritura de la firma nos permiten realizar una evaluación no intrusiva a los pacientes donde se analicen primordialmente las afecciones motoras y se evite el sesgo cognitivo que puede implicar la ejecución de tareas más complejas. Finalmente, como trabajo futuro se desea continuar con la evaluación de pacientes con PD mediante tareas no intrusivas donde además se incluya el análisis de tareas de escritura offline, las cuales son más naturales y eliminan posibles barreras tecnológicas en los pacientes. También se buscarán diferentes formas de preentrenar arquitecturas recurrentes de manera que enriquezcan el modelamiento

secuencial, buscando analizar de manera más natural fenómenos como la micrografía en la escritura de pacientes con PD.

10. Referencias

Referencias

- [1] T. Pringsheim, N. Jette, A. Frolkis y T. D. Steeves, “The prevalence of Parkinson’s disease: a systematic review and meta-analysis,” *Movement disorders*, vol. 29, n.º 13, págs. 1583-1590, 2014.
- [2] O. Hornykiewicz, “Biochemical aspects of Parkinson’s disease,” *Neurology*, vol. 51, n.º 2 Suppl 2, S2-S9, 1998.
- [3] S. Sveinbjörnsdóttir, “The clinical symptoms of Parkinson’s disease,” *Journal of neurochemistry*, vol. 139, págs. 318-324, 2016.
- [4] A. Samii, J. G. Nutt y B. R. Ransom, “Parkinson’s disease,” *The Lancet*, vol. 363, n.º 9423, págs. 1783-1793, 2004.
- [5] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal y M. Faundez-Zanuy, “Analysis of in-air movement in handwriting: A novel marker for Parkinson’s disease,” *Computer methods and programs in biomedicine*, vol. 117, n.º 3, págs. 405-411, 2014.
- [6] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet y S. Pinto, “From micrographia to Parkinson’s disease dysgraphia,” *Movement Disorders*, vol. 29, n.º 12, págs. 1467-1475, 2014.
- [7] C. D. Rios-Urrego, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera y J. R. Orozco-Arroyave, “Analysis and evaluation of handwriting in patients with Parkinson’s disease using kinematic, geometrical, and non-linear features,” *Computer methods and programs in biomedicine*, vol. 173, págs. 43-52, 2019.
- [8] Z. Galaz, P. Drotar, J. Mekyska et al., “Comparison of CNN-Learned vs. Handcrafted Features for Detection of Parkinson’s Disease Dysgraphia in a Multilingual Dataset,” *Frontiers in Neuroinformatics*, pág. 35, 2022.
- [9] M. Thomas, A. Lenka y P. Kumar Pal, “Handwriting analysis in Parkinson’s disease: current status and future directions,” *Movement disorders clinical practice*, vol. 4, n.º 6, págs. 806-818, 2017.
- [10] M. Diaz, M. Moetesum, I. Siddiqi y G. Vessio, “Sequence-based dynamic handwriting analysis for Parkinson’s disease detection with one-dimensional convolutions and BiGRUs,” *Expert Systems with Applications*, vol. 168, pág. 114 405, 2021.

- [11] C. Taleb, M. Khachab, C. Mokbel y L. Likforman-Sulem, “Visual representation of online handwriting time series for deep learning Parkinson’s disease detection,” en *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, IEEE, vol. 6, 2019, págs. 25-30.
- [12] C. R. Pereira, D. R. Pereira, F. A. Da Silva et al., “A step towards the automated diagnosis of parkinson’s disease: Analyzing handwriting movements,” en *2015 IEEE 28th international symposium on computer-based medical systems*, IEEE, 2015, págs. 171-176.
- [13] L. S. Bernardo, A. Quezada, R. Munoz et al., “Handwritten pattern recognition for early Parkinson’s disease diagnosis,” *Pattern recognition letters*, vol. 125, págs. 78-84, 2019.
- [14] D. Arraziqi, T. A. Sardjono, H. Miawarni y M. H. Purnomo, “Detection of Parkinson’s Disease at The Level of Motor Experiences of Daily Living Using Spiral Handwriting,” en *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, IEEE, 2020, págs. 39-46.
- [15] C. R. Pereira, D. R. Pereira, G. H. Rosa et al., “Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson’s disease identification,” *Artificial intelligence in medicine*, vol. 87, págs. 67-77, 2018.
- [16] I. Kamran, S. Naz, I. Razzak y M. Imran, “Handwriting dynamics assessment using deep neural network for early identification of Parkinson’s disease,” *Future Generation Computer Systems*, vol. 117, págs. 234-244, 2021.
- [17] M. Gazda, M. Hireš y P. Drotár, “Multiple-fine-tuned convolutional neural networks for Parkinson’s disease diagnosis from offline handwriting,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, n.º 1, págs. 78-89, 2021.
- [18] M. Pietikäinen, *Local Binary Patterns - Scholarpedia — scholarpedia.org*, http://www.scholarpedia.org/article/Local_Binary_Patterns, 2010.
- [19] T. Jadhav, “Handwritten signature verification using local binary pattern features and KNN,” *Int. Res. J. Eng. Technol.(IRJET)*, vol. 6, n.º 4, págs. 579-586, 2019.
- [20] E. Prakasa, “Texture feature extraction by using local binary pattern,” *INKOM Journal*, vol. 9, n.º 2, págs. 45-48, 2016.
- [21] M. Pietikäinen, A. Hadid, G. Zhao y T. Ahonen, *Computer vision using local binary patterns*. Springer Science & Business Media, 2011, vol. 40.
- [22] N. Dalal y B. Triggs, “Histograms of oriented gradients for human detection,” en *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, Ieee, vol. 1, 2005, págs. 886-893.
- [23] S. KAUSHAL, *Beginner’s Guide To Image Gradient — analyticsvidhya.com*, <https://onx.la/101b0>, 2022.

- [24] M. Tyagi, *HOG(Histogram of Oriented Gradients)* — *towardsdatascience.com*, <https://onx.la/1a890>, 2021.
- [25] L. Alzubaidi, J. Zhang, A. J. Humaidi et al., “Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions,” *Journal of big Data*, vol. 8, págs. 1-74, 2021.
- [26] S. Saha, *A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way* — *towardsdatascience.com*, <https://onx.la/d7096>, 2018.
- [27] N. Donges, *A Guide to Recurrent Neural Networks: Understanding RNN and LSTM Networks* — *builtin.com*, <https://builtin.com/data-science/recurrent-neural-networks-and-lstm>, 2022.
- [28] D. Kalita, *A Brief Overview of Recurrent Neural Networks (RNN)* — *analyticsvidhya.com*, <https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/>, 2022.
- [29] H. Bhavsar y M. H. Panchal, “A review on support vector machine for data classification,” *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, n.º 10, págs. 185-189, 2012.
- [30] R. Pupale, *Support Vector Machines(SVM) — An Overview* — *towardsdatascience.com*, <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>, 2018.
- [31] A. Mitrani, *Evaluating Categorical Models* — *towardsdatascience.com*, <https://towardsdatascience.com/evaluating-categorical-models-e667e17987fd>, 2019.
- [32] S. Spichak, *Sensitivity, Specificity and Meaningful Classifiers* — *towardsdatascience.com*, <https://towardsdatascience.com/sensitivity-specificity-and-meaningful-classifiers-8326738ec5c2>, 2020.
- [33] D. Steen, *Understanding the ROC Curve and AUC* — *towardsdatascience.com*, <https://towardsdatascience.com/understanding-the-roc-curve-and-auc-dd4f9a192ecb>, 2020.
- [34] C. G. Goetz, B. C. Tilley, S. R. Shaftman et al., “Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results,” *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, n.º 15, págs. 2129-2170, 2008.
- [35] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh y B. Yoon, “Improved handwritten digit recognition using convolutional neural networks (CNN),” *Sensors*, vol. 20, n.º 12, pág. 3344, 2020.