

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

**Detección Temprana de la Enfermedad de Parkinson Mediante Análisis
Multimodal de Escritura a Mano Utilizando Aprendizaje Profundo.**

Proyecto de Titulación

Previo a la obtención del Título de:

Magíster en Ciencias de la Computación

Presentado por:

Alisson Asunción Constantine Macías

Guayaquil - Ecuador

0

Reconocimientos

Cada página de este trabajo guarda mucho más que resultados: contiene silencios compartidos, apoyo sincero y muchas formas de estar presentes. Es, sobre todo, el reflejo del acompañamiento, la confianza y la generosidad de quienes caminaron conmigo a lo largo de este proceso.

A mi tutor de tesis, Enrique Pelaéz Jarrín, PhD, por su guía honesta y paciente, por su exigencia que me impulsó a dar más, y por sus observaciones lúcidas que enriquecieron cada etapa de este trabajo.

A mi grupo de investigación *AI in Health*, por convertirse en un espacio de aprendizaje colectivo, en donde cada conversación y discusión técnica fueron oportunidades para crecer. Así mismo, agradezco profundamente a los colaboradores del área de neurología de los Hospitales IESS Los Ceibos y Teodoro Maldonado Carbo, cuya disposición y compromiso fueron clave para que esta investigación pudiera llevarse a cabo con responsabilidad y sentido.

Al coordinador de la maestría, Daniel Ochoa Donoso, PhD, por su cercanía, por animarme a mirar más allá del aula y descubrir en la investigación una forma de construir conocimiento con propósito. Gracias por confiar en mis capacidades.

A la ESPOL y a CEDIA, por brindarme el acceso a la infraestructura tecnológica que hizo posible cada experimento. Su apoyo permitió materializar este proyecto con los recursos necesarios.

A mi mamá, por ser ese lugar seguro al que siempre puedo volver. Gracias por su confianza, por su amor que no necesita palabras, y por sostenerme en cada parte de este proceso.

A mi papá, por mostrarme desde niña que la ingeniería también se trata de crear con sentido. Gracias por inspirarme.

A mi hermana, por estar presente a su manera, con sus preguntas constantes, su curiosidad genuina y su manera única de preocuparse por mí.

A quien ha estado a mi lado compartiendo esta etapa —con paciencia, ternura y convicción— gracias por su apoyo constante y por recordarme que los logros compartidos tienen un significado más profundo. Su presencia ha sido abrigo y empuje.

A Vaquita, mi gato, por su compañía nocturna mientras escribía estas páginas, por dormirse cerca del teclado y recordarme, con su serenidad, que el descanso también forma parte del proceso creativo.

Y a todas las personas que en algún momento tendieron una mano, ofrecieron un consejo, compartieron una idea o simplemente me escucharon: gracias. Su presencia fue más importante de lo que imaginan.

A cada persona que dejó huella en este camino.

Evaluadores

Colon Enrique Pelaez Jarrín

Tutor de proyecto

Miguel Andrés Realpe Robalino

Revisor de proyecto

Daniel Erick Ochoa Donoso

Coordinador del programa

Resumen

Este estudio propone un marco metodológico para la detección temprana de la enfermedad de Parkinson (EP) mediante un análisis multimodal basado en tareas de escritura registradas mediante una tableta de dibujo, comparando la eficacia de modelos de aprendizaje profundo (Deep Learning, DL) y no profundo (Machine Learning tradicional, ML) así como la explicabilidad de modelos de ML. Se utilizaron dos conjuntos de datos provenientes de regiones culturalmente diversas: PaHaW (República Checa) y una base de datos propietaria recolectada en Ecuador. Estos conjuntos integran señales motoras tales como presión, azimut, altitud y coordenadas espaciales capturadas durante tareas de dibujo de espirales.

El análisis comparativo reveló que los modelos de aprendizaje automático no profundo, específicamente Extra Trees y Gradient Boosting, obtuvieron precisiones destacables de hasta el 93,27% y 94,20% respectivamente, superando a las redes neuronales convolucionales (CNN) y a las redes neuronales de picos (SNN). Adicionalmente, mediante técnicas de explicabilidad basadas en SHapley Additive exPlanations (SHAP), se identificaron biomarcadores clave para la detección de la enfermedad, destacando principalmente la presión máxima ejercida por el lápiz, la repetición angular (azimut) y bajos valores de altitud.

La investigación demuestra la generalización potencial de los biomarcadores identificados a través de poblaciones con características lingüísticas diferentes. Además, proporciona herramientas interpretables que facilitan el entendimiento clínico de los resultados generados por modelos predictivos complejos. Sin embargo, también se identificaron limitaciones relevantes, tales como los altos requerimientos computacionales y limitaciones en escalabilidad asociadas a las SNN, lo cual abre futuras líneas de investigación hacia la optimización de recursos y mejora en la eficiencia computacional.

Declaración expresa

Yo Alisson Asunción Constantine Macías reconozco que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí/nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al/los autor/es que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 22 de Mayo del 2025

Alisson Asunción Constantine Macías

Índice general

Abstract	iii
Índice de figuras	vii
Índice de cuadros	ix
Lista de abreviaciones	x
1 Introducción	1
2 Trabajos Relacionados	3
2.1 Características neuromotoras en escritura manuscrita	4
2.2 Modelos clásicos de aprendizaje automático	5
2.3 Aprendizaje profundo y multimodalidad	5
2.4 Redes neuronales de picos (SNN) y enfoques bioinspirados	5
2.5 Explicabilidad en contextos clínicos (XAI)	6
2.6 Variabilidad cultural y lingüística	6
2.7 Síntesis del estado del arte	6
3 Metodología	7
3.1 Conjunto de Datos	7
3.1.1 Descripción de las Bases de Datos	7
3.1.2 Selección e Integración de Tareas	9
3.1.3 Preprocesamiento	10
3.1.4 Segmentación de las Señales	11
3.2 Modelos Propuestos	11
3.2.1 Modelos de Aprendizaje No Profundo	12
3.2.2 Modelos de Aprendizaje Profundo	14
3.3 Evaluación	23
3.3.1 Métricas de Desempeño	23
3.3.2 Protocolo Experimental	23
3.4 Pruebas de Significancia	25
3.5 Explicabilidad	25

3.5.1	Técnicas Utilizadas	25
4	Resultados	26
4.1	Diferencias Estadísticas entre Conjuntos de Datos	26
4.2	Aumento de Datos	27
4.3	Enfoque de Aprendizaje Poco Profundo	27
4.3.1	Rendimiento de los Modelos	28
4.3.2	Rendimiento con Conjuntos de Datos Combinados	28
4.3.3	Resultados sobre el Conjunto de Datos Propietario	29
4.3.4	Matrices de Confusión	29
4.3.5	Mejores Hiperparámetros	29
4.4	Enfoque de Aprendizaje Profundo	29
4.4.1	Redes Neuronales Convolucionales	30
4.4.2	Redes Neuronales de Picos	32
4.5	Explicabilidad de los Modelos	36
4.5.1	Características Seleccionadas	38
4.5.2	Impacto de las características específicas de clase	39
4.5.3	Interacciones de características y análisis de dependencia	41
5	Conclusiones	53
Referencias		55
Apéndices		60
Apéndice A: Visualización de la importancia de características		61

Índice de figuras

3.1	Protocolo gráfico utilizado para la evaluación de la escritura en los grupos PD y HC	8
3.2	Plantilla utilizada en la base de datos propietaria para la evaluación de la escritura en los grupos PD y HC . . .	9
3.3	Características Dinámicas que captura el Lápiz	10
3.4	Ejemplo de las imágenes de los grupos PD y HC generadas a partir de las coordenadas x, y , utilizadas como entrada para los modelos de aprendizaje profundo. Las imágenes corresponden a ventanas de 400 puntos con un solapamiento del 75%.	12
3.5	Ejemplo de imágenes del conjunto de prueba después del preprocesamiento	17
3.6	Ejemplo de imágenes después de aplicar el aumento de datos de los grupos PD y HC para la ventana de 400 con overlpangi del 75%.	18
3.7	Ejemplo de codificación en spikes aplicado a señales motoras de escritura. Las dos primeras filas muestran los spikes generados mediante <i>Rate Coding</i> a partir de las componentes seno y coseno del azimut. Las dos filas inferiores representan la codificación mediante <i>Latency Coding</i> para las señales de altitud y presión, donde la aparición temprana de spikes indica mayor intensidad en la señal original.	20
3.8	Secuencia de imágenes generadas a partir de una espiral manuscrita, correspondiente a una ventana de 400 puntos dividida en 100 fotogramas. Cada imagen muestra el estado progresivo del dibujo al acumular 4 nuevos puntos por frame, permitiendo representar la dinámica del trazo en el tiempo. El avance ocurre en sentido antihorario. Los números indican el índice del fotograma dentro de la secuencia.	21
3.9	Esquema de arquitectura Early Fusion . Todas las señales multimodales son concatenadas desde el inicio y procesadas por una única red spiking compuesta por bloques LIF o SLSTM.	22
3.10	Esquema de arquitectura Late Fusion . Cada modalidad es procesada por separado a través de bloques spiking (LIF o SLSTM), y sus salidas se fusionan mediante concatenación antes de la capa final de clasificación.	23
4.1	Matrices de confusión de los distintos modelos de aprendizaje automático evaluados con la base de datos propietaria.	30
4.2	Matrices de confusión de los distintos modelos de aprendizaje profundo evaluados con la base de datos propietaria.	32
4.3	Matrices de confusión de los distintos modelos de SNNs evaluados con la base de datos propietaria.	34
4.4	Dependence plots de las tres características más relevantes para el modelo SVC. El eje x representa el valor real de la característica, mientras que el eje y muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representa valores bajos y rojo valores altos.	42

4.5	Dependence plots de las tres características más relevantes para el modelo de Regresión Logística. El eje <i>x</i> representa el valor real de la característica, mientras que el eje <i>y</i> muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representa valores bajos y rojo valores altos.	43
4.6	Dependence plots de las tres características más relevantes para el modelo Gradient Boosting. El eje <i>x</i> representa el valor real de la característica, mientras que el eje <i>y</i> muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representa valores bajos y rojo valores altos.	44
4.7	Dependence plots de las tres características más importantes del modelo Random Forest. Cada punto representa una observación, el eje <i>x</i> indica el valor real de la característica, el eje <i>y</i> su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos de la característica de interacción, mientras que azul indica valores bajos.	46
4.8	Dependence plots de las tres características más relevantes del modelo KNN. Cada punto representa una observación; el eje <i>x</i> indica el valor de la característica, el eje <i>y</i> su valor SHAP. El color representa la interacción con otra característica.	47
4.9	Dependence plots de las tres características más importantes del modelo Extra Trees. Cada punto representa una observación, el eje <i>x</i> indica el valor real de la característica, el eje <i>y</i> su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos de la característica de interacción, mientras que azul indica valores bajos.	49
4.10	Dependence plots de las tres características más importantes del modelo Naïve Bayes. Cada punto representa una observación; el eje <i>x</i> indica el valor real de la característica, el eje <i>y</i> su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos, mientras que azul representa valores bajos.	50
4.11	Dependence plots de las tres características más importantes del modelo Decision Tree. Cada punto representa una observación; el eje <i>x</i> indica el valor real de la característica, el eje <i>y</i> su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos, mientras que azul representa valores bajos.	52
1	Importancia de características determinada por los modelos SVC, Regresión Logística y Gradient Boosting. . . .	61
2	Importancia de características determinada por los modelos Random Forest, KNN, Extra Trees, Naïve Bayes y Árboles de Decisión.	62

Índice de cuadros

3.1	Información demográfica de pacientes con diagnóstico de la enfermedad de Parkinson (Parkinson Disease, PD) y controles sanos (Health Control, HC) de las bases de datos PaHaW y propietaria.	7
3.2	Valores máximos y mínimos de las variables en las bases de datos PaHaW y propietaria.	10
3.3	Hiperparámetros ajustados para cada modelo	15
3.4	Detalles arquitectónicos de cada modelo CNN, incluyendo el número de capas, transformaciones internas y la estructura de las capas de clasificación. (FC) se refiere a capas totalmente conectadas, (ReLU) a la función de activación Rectified Linear Unit y las capas Dropout se utilizan para prevenir sobreajuste.	16
3.5	Detalles arquitectónicos de cada modelo SNN. Se incluyen las dimensiones de entrada, estructura por capas, tipo de neuronas (LIF o SLSTM), y forma de fusión.	24
4.1	Resultados de las pruebas de Shapiro-Wilk y Mann-Whitney U para las características seleccionadas.	26
4.2	Cantidad de datos antes del proceso de aumento con ventanas deslizantes.	27
4.3	Cantidad de datos después del aumento con ventanas deslizantes.	27
4.4	Métricas de precisión para diferentes tamaños de ventana y solapamientos en distintos modelos de clasificación.	28
4.5	Métricas de rendimiento utilizando el conjunto combinado (100% PaHaW + 30% Propietario).	28
4.6	Resultados de prueba sobre el conjunto de datos propietario.	29
4.7	Mejores hiperparámetros para cada modelo (30% Datos Propietarios + PaHaW).	30
4.8	Métricas de rendimiento de los modelos CNN utilizando el conjunto combinado (100% PaHaW + 30% Propietario)	31
4.9	Métricas de rendimiento sobre el conjunto de datos propietario	31
4.10	Hiperparámetros óptimos seleccionados para cada modelo CNN.	32
4.11	Métricas de rendimiento de los modelos SNNs utilizando el conjunto combinado (100% PaHaW + 30% Propietario).	33
4.12	Métricas de rendimiento sobre el conjunto de datos propietario (70% Propietario).	33
4.13	Mejores hiperparámetros por grupos funcionales para cada arquitectura SNN.	36
4.14	Mejores hiperparámetros por grupos funcionales para cada arquitectura SNN Recurrente.	37
4.15	Frecuencia de aparición de características en el top 3 de importancia SHAP.	40

Lista de abreviaciones

AI Artificial Intelligence.

CNN Convolutional Neural Network.

DL Deep Learning.

EP Enfermedad de Parkinson.

FC Fully Connected.

GNN Graph Neural Network.

GPU Graphics processing unit.

HC Healthy Control.

KNN K-Nearest Neighbors.

LIF Leaky Integrate-and-Fire.

ML Machine Learning.

PD Parkinson Disease.

SD Standard Deviation.

SHAP SHapley Additive exPlanations.

SLSTM Spiking Long Short-Term Memory.

SNN Spiking Neural Network.

SVC Support Vector Classifier.

TSFresh Time Series Feature Extraction Library.

1

Introducción

La enfermedad de Parkinson (EP) es una enfermedad neurodegenerativa progresiva que afecta aproximadamente al 1-2% de la población mayor de 60 años [1]. Su diagnóstico temprano es crucial, ya que permite tratamientos oportunos que pueden ralentizar el deterioro motor y mejorar la calidad de vida de los pacientes. No obstante, los métodos diagnósticos convencionales suelen ser subjetivos y dependen principalmente de evaluaciones clínicas que pueden no detectar cambios sutiles en etapas tempranas [2, 3].

Una alternativa emergente es el análisis digital de la escritura a mano, que aprovecha características motoras como la presión, velocidad, dinámica angular y la fluidez del trazo. La escritura es una tarea neuromotora compleja que requiere control motor fino, planificación cognitiva y retroalimentación sensorial, por lo que pequeñas alteraciones neuromotoras pueden ser detectadas cuantitativamente a través de análisis [4, 5]. Diversos estudios han reportado que pacientes con EP presentan cambios específicos en la escritura, como micrografía, irregularidad en perfiles de presión, y variabilidad en la fluidez del trazo, incluso antes del diagnóstico clínico formal [6, 7].

La creciente disponibilidad y precisión de dispositivos digitales, como tabletas gráficas y bolígrafos inteligentes, ha facilitado considerablemente la captura de biomarcadores a partir de tareas de dibujo y escritura. Particularmente, pruebas como la espiral de Arquímedes han mostrado alta sensibilidad en detectar patrones anormales en pacientes con EP, superando ampliamente las técnicas tradicionales de papel y lápiz [8, 9]. Estos avances tecnológicos han transformado el análisis de dibujo y escritura manuscrita en una herramienta no invasiva, escalable y objetiva para la detección temprana y monitoreo continuo de enfermedades neurodegenerativas.

Sin embargo, a pesar del éxito de estos métodos digitales, aún persisten limitaciones significativas relacionadas con la generalización lingüística. La mayoría de los estudios previos se han basado en datos recolectados en poblaciones homogéneas, lo que podría llevar a modelos poco generalizables cuando se aplican en contextos multilingües [10, 11]. Diferencias educativas y lingüísticas pueden influir notablemente en las características motoras de la escritura, haciendo imperativo el desarrollo de modelos capaces de normalizar y adaptarse a esta variabilidad para lograr una robustez universal.

Este trabajo aborda estas limitaciones mediante un análisis multimodal y multinacional, utilizando datos provenientes de dos conjuntos culturalmente diversos: la base pública PaHaW (República Checa) y un conjunto propietario recolectado en Ecuador. Al integrar técnicas avanzadas de aprendizaje profundo (CNN y SNN) y métodos clásicos de aprendizaje automático, este estudio evalúa críticamente la capacidad de estos enfoques para detectar patrones neuromotores universales en el con-

1 Introducción

texto del diagnóstico temprano de la EP. Asimismo, se aplican métodos de explicabilidad con SHapley Additive exPlanations (SHAP), proporcionando transparencia clínica sobre las decisiones de los modelos y facilitando la interpretación médica de las características identificadas [12, 13].

Además, este estudio investiga explícitamente si modelos de aprendizaje automático poco profundos pueden igualar o superar el rendimiento de las redes neuronales profundas en contextos donde la disponibilidad de datos es limitada, lo cual es común en entornos clínicos reales. También se realiza un análisis estadístico riguroso para determinar si las variaciones en los patrones motores pueden ser significativamente influenciadas por diferencias culturales o lingüísticas entre los conjuntos de datos analizados.

La contribución principal de esta investigación es ofrecer un marco metodológico integral, robusto y explicable para la detección temprana de Parkinson mediante tareas digitales de dibujo, por medio de la espiral de Arquímedes, proporcionando herramientas prácticas para su implementación clínica y facilitando el diagnóstico oportuno en poblaciones diversas.

La estructura del documento es la siguiente: en el capítulo 2 se revisa detalladamente el estado del arte en técnicas digitales de análisis de escritura para la EP. El capítulo *Metodología* describe exhaustivamente los procedimientos experimentales, incluyendo la recolección, integración, preprocesamiento y análisis de los datos, así como el diseño e implementación de los modelos propuestos. En el capítulo *Resultados* se presentan análisis cuantitativos y cualitativos de los hallazgos obtenidos, destacando los biomarcadores neuromotores más relevantes y las capacidades de generalización intercultural de los modelos. Finalmente, el capítulo *Conclusiones y Recomendaciones* resume los aportes fundamentales del estudio y propone futuras líneas de investigación orientadas a mejorar la precisión diagnóstica y la interpretabilidad clínica de las herramientas desarrolladas.

2

Trabajos Relacionados

La escritura a mano constituye una actividad motora compleja, altamente sensible a alteraciones neurológicas. En la enfermedad de Parkinson (EP), las disfunciones en los ganglios basales afectan directamente la motricidad fina, reflejándose en características específicas de la escritura como la micrografía, patrones de presión irregulares, velocidad errática y mayor variabilidad en la fluidez del trazo [6, 5]. Diversos estudios han demostrado que estos cambios pueden presentarse incluso antes de establecerse un diagnóstico clínico formal, consolidando así los biomarcadores digitales extraídos del análisis de escritura como indicadores tempranos sólidos de la enfermedad [7].

Entre los biomarcadores neuromotores más relevantes se destacan la micrografía (reducción anormal del tamaño de letra), la disfluencia asociada a variaciones irregulares en la velocidad y la variabilidad en la presión ejercida durante la escritura. Pacientes con EP generalmente presentan trayectorias del lápiz con menor amplitud, mayor tiempo en el aire (*in-air time*) y mayor cantidad de cambios en la velocidad y aceleración del trazo. Mediciones cinemáticas detalladas, incluyendo duración del trazo, perfiles de presión y variaciones dinámicas en la inclinación del lápiz, han demostrado ser indicadores confiables del deterioro neuromotor asociado con la EP. Estas características permiten no solo la detección temprana, sino también el monitoreo de la progresión de la enfermedad y la efectividad de intervenciones terapéuticas [7].

Las evaluaciones tradicionales de escritura se han empleado durante décadas en la práctica clínica para identificar déficits neuromotores y cognitivos; no obstante, limitaciones como la sensibilidad reducida y la subjetividad inherente han motivado el desarrollo de técnicas digitales avanzadas. Estas técnicas permiten capturar, en tiempo real, características cinemáticas y dinámicas con gran precisión, facilitadas por el uso de tecnologías digitales como tabletas gráficas y bolígrafos inteligentes. Estos dispositivos posibilitan la extracción enriquecida de información espacial, temporal y dinámica, consolidándose como herramientas importantes para evaluar enfermedades neurodegenerativas [14, 15].

Una herramienta ampliamente validada en este contexto es la prueba digital de dibujo de la espiral de Arquímedes. Esta prueba digital ha reportado mejoras significativas en la sensibilidad diagnóstica en comparación con métodos tradicionales de papel y lápiz [16], puesto que permite cuantificar parámetros específicos como el número de picos de velocidad, variaciones en la altitud del lápiz, velocidad del dibujo y amplitud del temblor. La alta resolución espacial y temporal proporcionada por tabletas digitales posibilita detectar variaciones microestructurales, como sutiles cambios en la curvatura y la presión del trazo, comúnmente desapercibidas en inspecciones visuales clínicas. Estudios indican que los pacientes con EP suelen mostrar un temblor unidireccional y asimétrico claramente observable en esta prueba [8, 9].

El proceso de escritura también está influido por factores culturales y características inherentes a los distintos sistemas

2 Trabajos Relacionados

de escritura. Por ejemplo, alfabetos como el árabe, cursivos y sensibles al contexto, implican una ejecución motora continua, mientras que alfabetos latinos presentan caracteres segmentados y discretos. Estas diferencias demandan técnicas adaptativas para la extracción de características espaciotemporales específicas según el idioma y la cultura [11]. Estudios en poblaciones bilingües han evidenciado diferencias significativas en características de escritura entre distintas lenguas y estilos educativos [17, 18, 10]. La normalización adecuada es crítica para corregir estas diferencias culturales y evitar que los modelos diagnósticos atribuyan incorrectamente patrones motores culturales específicos a síntomas patológicos. Técnicas recientes de normalización han demostrado mejorar significativamente el rendimiento de los algoritmos diagnósticos [19].

Una estrategia particularmente efectiva ha sido la integración de datos multimodales. Al combinar señales como presión, azimut, altitud y coordenadas x, y , se logra representar distintos aspectos del deterioro motor. Cascarano et al. [20] demostraron que esta integración mejora la clasificación e incluso permite estimar el grado de severidad de la enfermedad. Estos enfoques multimodales ofrecen una visión más rica del proceso de escritura [21].

Paralelamente, ha surgido una preocupación creciente sobre la interpretabilidad de los modelos utilizados. En entornos clínicos, es fundamental comprender las razones detrás de una decisión automatizada. Por ello, técnicas de explicabilidad como SHAP han sido incorporadas exitosamente. Zemmar et al. [13] identificaron, mediante SHAP, las regiones más influyentes de una espiral manuscrita en la predicción de EP, mientras que Parziale et al. [22] emplearon modelos evolutivos para generar reglas de decisión comprensibles sin comprometer el rendimiento. SHAP permite obtener explicaciones globales y locales, y ha demostrado ser eficaz en aplicaciones médicas, incluyendo el análisis de escritura [12, 23, 24].

Finalmente, las redes neuronales de picos (SNNs) han ganado atención como una alternativa bioinspirada. Estas redes procesan información mediante eventos discretos llamados "spikes", emulando el comportamiento de las neuronas biológicas. Su principal ventaja radica en su capacidad para representar de forma eficiente la información temporal y en su bajo consumo energético, lo que las hace aptas para aplicaciones en hardware neuromórfico [25].

A continuación, se revisan los principales avances y limitaciones existentes en la literatura sobre análisis digital de escritura para el diagnóstico temprano de la EP, organizados en bloques donde se detallan las contribuciones específicas de esta investigación.

2.1 Características neuromotoras en escritura manuscrita

Numerosos estudios han identificado características neuromotoras de la escritura como biomarcadores tempranos de la EP. Por ejemplo, Tai et al. (2021) alcanzaron una sensibilidad del 88% al analizar cambios en la presión y velocidad [5], mientras que Di Febbo et al. (2023) reportaron una precisión del 91% empleando mediciones cinemáticas detalladas (duración de trazo, perfiles de presión, dinámica angular) [7]. En la prueba de espiral digital, Danna et al. (2019) y Kamble et al. (2021) consiguieron sensibilidades cercanas al 90% y 92% respectivamente, gracias a la cuantificación de picos de velocidad y variaciones en la altitud del lápiz [8, 9].

Limitaciones. Estos trabajos suelen basarse en poblaciones culturalmente homogéneas, lo que puede sesgar los biomarcadores al no considerar variaciones lingüísticas y educativas.

Contribución de este estudio. Se integran dos bases de datos de regiones diversas (PaHaW de República Checa y un dataset

2 Trabajos Relacionados

propio de Ecuador) para evaluar la robustez intercultural de estas características neuromotoras y validar su universalidad.

2.2 Modelos clásicos de aprendizaje automático

Modelos como SVM, Random Forest y regresión logística han sido eficazmente aplicados al análisis de escritura. Ali et al. (2019) y Poon et al. (2019) obtuvieron precisiones del 85–92% utilizando atributos diseñados manualmente (velocidad, aceleración, presión) [26, 27]. En particular, Poon et al. alcanzaron un 85% de exactitud al emplear picos de aceleración e irregularidades en la presión como entradas [27].

Limitaciones. La dependencia de características artesanales limita la escalabilidad y adaptación a nuevas poblaciones, además de requerir expertos en diseño de features.

Contribución de este estudio. Se compara el rendimiento de estos modelos con el de arquitecturas profundas en escenarios de datos limitados y heterogéneos, determinando si pueden igualar la eficacia de los enfoques más complejos.

2.3 Aprendizaje profundo y multimodalidad

Las CNN y arquitecturas híbridas CNN–LSTM han demostrado una elevada capacidad de extracción automática de características. Kamran et al. (2021) reportaron sensibilidades del 95% con una CNN–LSTM en tareas de escritura [28], y Ali et al. (2023) alcanzaron un 81% de precisión combinando características estáticas y dinámicas extraídas por CNNs [29]. Por su parte, Cascarano et al. (2019) y Ríos-Urrego et al. (2020) obtuvieron precisiones de hasta el 97% al integrar señales de presión, azimut y altitud [20, 21].

Limitaciones. La opacidad de las redes profundas dificulta su aceptación clínica y escasa interpretabilidad de las decisiones.

Contribución de este estudio. Se incorporan técnicas de explicabilidad (SHAP) para generar interpretaciones globales y locales de las predicciones, facilitando la validación médica de los biomarcadores extraídos.

2.4 Redes neuronales de picos (SNN) y enfoques bioinspirados

Las Spiking Neural Networks (SNN) ofrecen eficiencia energética y representación temporal precisa. El modelo SEFRON de Das et al. (2024) alcanzó una precisión media del 91% (hasta 100% en ciertos subconjuntos) en detección de EP a partir de señales de voz codificadas en spikes [30]. AbouHassan et al. (2023) demostraron su aplicabilidad multimodal con resultados comparables a DNNs tradicionales [31].

Limitaciones. El entrenamiento de SNNs requiere ajustes complejos de hiperparámetros y codificaciones de spike, lo que dificulta su adopción práctica.

Contribución de este estudio. Se evalúa de forma sistemática el desempeño comparativo de SNNs y CNNs en series tempo-

2 Trabajos Relacionados

rales de escritura, explorando estrategias de codificación (rate vs. latency coding) y optimización de entrenamiento.

2.5 Explicabilidad en contextos clínicos (XAI)

La interpretabilidad de modelos es esencial para la integración en entornos médicos. SHAP ha emergido como técnica de elección por su consistencia global y local. Zemmar et al. (2024) usaron SHAP para identificar regiones de la espiral más influyentes, y Ladbury et al. (2022) validaron su aplicabilidad en diagnósticos clínicos con un aumento del 15% en la confianza del médico [13, 12].

Limitaciones. Aunque SHAP facilita la interpretación, su implementación en modelos de muy alta dimensión puede resultar costosa en tiempo de cómputo.

Contribución de este estudio. Se aplica SHAP tanto sobre modelos de árbol como profundos, generando visualizaciones (summary y dependence plots) que correlacionan clínicamente cada biomarcador con la probabilidad de EP.

2.6 Variabilidad cultural y lingüística

Estudios comparativos de escritura árabe y latina han mostrado diferencias de hasta 15% en métricas de fluidez y continuidad, afectando la precisión diagnóstica cuando no se normaliza adecuadamente [11, 19, 10]. Estas variaciones provienen de diferencias en la segmentación de trazos y en la pedagogía de enseñanza de la escritura.

Limitaciones. La mayoría de modelos no incorpora estrategias de normalización cultural, exponiéndolos a errores sistemáticos en poblaciones diversas.

Contribución de este estudio. Se prueba un protocolo de normalización basado en ajustes de escala (min–max) y compensaciones por estilo de escritura, evaluando su impacto en la robustez de los modelos.

2.7 Síntesis del estado del arte

A pesar de los avances en el análisis digital de la escritura para la enfermedad de Parkinson (EP), persisten desafíos en la generalización intercultural, la interpretabilidad clínica y la eficiencia con datos limitados. Aunque los biomarcadores neuromotores en la escritura muestran gran potencial para el diagnóstico temprano no invasivo de la EP, existen retos significativos, especialmente relacionados con la diversidad cultural y lingüística. La expansión de conjuntos de datos multilingües y culturalmente diversos es fundamental para desarrollar modelos diagnósticos robustos y generalizables, capaces de captar plenamente la variabilidad natural de la escritura en contextos demográficos y lingüísticos diversos [32]. Esta investigación aporta un enfoque integral que combina modelos multimodales, técnicas de explicabilidad y normalización cultural, creando un marco robusto y clínicamente interpretable para la detección temprana de Parkinson en entornos multiculturales, fortaleciendo así la aplicabilidad clínica de los modelos desarrollados.

3

Metodología

Este capítulo describe el enfoque metodológico adoptado en el estudio, abarcando desde la recolección y normalización de los datos hasta la evaluación e interpretación de los resultados. Se detallan los procesos de integración y preprocesamiento aplicados a los conjuntos de datos multiculturales, así como la segmentación temporal de las señales. Posteriormente, se exploran distintos enfoques de clasificación: modelos de *machine learning* tradicionales, redes profundas (*deep learning*) y arquitecturas bioinspiradas basadas en redes neuronales de picos (SNN). Finalmente, se presentan las estrategias de evaluación y las técnicas de explicabilidad utilizadas para interpretar los modelos.

3.1 Conjunto de Datos

3.1.1 Descripción de las Bases de Datos

Dos conjuntos de datos recopilados de manera independiente se emplearon, los cuales se estandarizaron y consolidaron en una base unificada para su análisis integral. El conjunto de datos está compuesto por variables digitalizadas mediante dispositivos electrónicos, que registran las dinámicas cinemáticas en la tarea de dibujo de espirales. Los detalles demográficos de ambos conjuntos de datos se resumen en la tabla 3.1.

Cuadro 3.1: Información demográfica de pacientes con diagnóstico de la enfermedad de Parkinson (Parkinson Disease, PD) y controles sanos (Health Control, HC) de las bases de datos PaHaW y propietaria.

	PaHaW PD (n=37)	PaHaW HC (n=38)	Propietaria PD (n=9)	Propietaria HC (n=7)
	Media ± SD	Media ± SD	Media ± SD	Media ± SD
Edad, Años	65.83 ± 11.64	69.3 ± 5.74	64.94 ± 13.54	61.56 ± 5.88
Duración de la enfermedad, años	8.38 ± 4.80	-	9.79 ± 8.29	-
Dosis diaria equivalente de Levodopa (mg)	1432.19 ± 704.78	-	447.77 ± 283.92	-
Genero (Masculino/Femenino)	19/18	20/18	1/8	1/6
Lateralidad (Diestro/Zurdo)	37/0	38/0	8/1	7/0

Valores son media ± desviación estandar (SD) excepto para genero y lateralidad.

3 Metodología

Base de Datos PaHaW La base de datos "Parkinson's Disease Handwriting" contiene los trazos de 37 pacientes diagnosticados con PD -19 hombres y 18 mujeres- y 38 personas de control sanos HC -20 hombres y 18 mujeres-, equiparados por edad y género. Los datos fueron recolectados mediante colaboración entre el Centro de Trastornos del Movimiento del primer departamento de Neurología de la Universidad Masaryk y el Hospital Universitario St. Anne's en Brno, República Checa.

Para el registro de los datos se utilizó una tableta Wacom Intuos 4M con una frecuencia de muestreo de 150 Hz. Con el fin de mantener condiciones naturales de escritura durante la digitalización, se empleó un bolígrafo de tinta convencional.

Las tareas de trazado se ilustran en la Figura 3.1, que muestra el protocolo estandarizado utilizado para la evaluación comparativa entre los grupos PD y HC.



Figura 3.1: Protocolo gráfico utilizado para la evaluación de la escritura en los grupos PD y HC.

Base de Datos Propietaria La base de datos propietaria fue recolectada por el equipo de investigación *AI in Health* de la Universidad Escuela Superior Politécnica del Litoral (ESPOL) en Guayaquil, Ecuador. El conjunto de datos consiste en trazados de 9 pacientes diagnosticados con enfermedad de Parkinson y 7 sujetos de control, obtenidos en el Hospital IESS Los Ceibos. El estudio contó con la aprobación del Comité de Ética de la Universidad Técnica de Manabí (UTM).

Para la captura de parámetros digitales se empleó la tableta Wacom Pro PTH-660, equipada con un sensor de presión de 8192 niveles, resolución espacial de 5080 lpi y un lápiz digital Wacom Pro Pen 2.

Se solicitó a cada participante realizar tareas de escritura mediante una plantilla predefinida como se muestra en la Figura 3.2. En este conjunto de datos, la tarea 5 consiste en replicar el patrón de la tarea 4, trazando sobre la línea y comenzando desde el símbolo "+". La tarea 7 requiere replicar el patrón de la tarea 6, comenzando y terminando en el símbolo "+". Por último, la tarea 8 solicita a los participantes dibujar un reloj con las manecillas marcando 10 minutos antes de las 11.

Para ambas bases de datos, las pruebas fueron ejecutadas a velocidad natural sin restricciones en repeticiones de sílabas, palabras o altura de trazos. La captura de datos inició cuando el lápiz óptico estableció contacto con la tableta y finalizó al completarse la actividad.

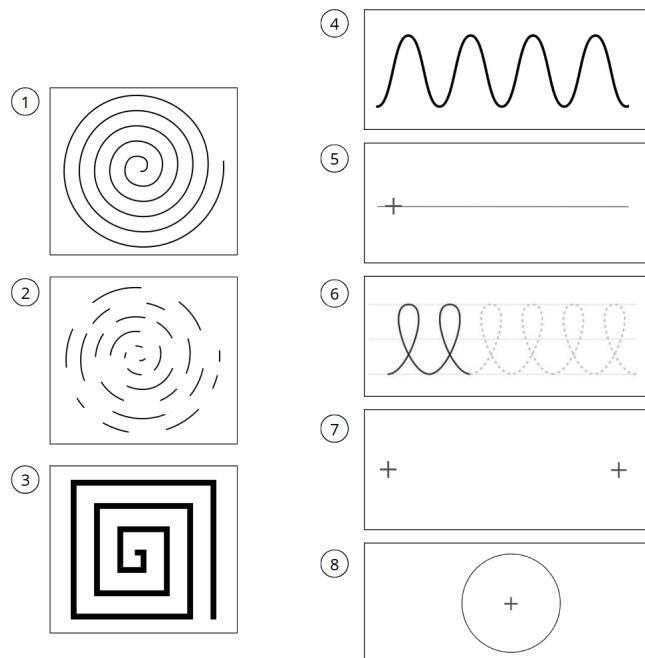


Figura 3.2: Plantilla utilizada en la base de datos propietaria para la evaluación de la escritura en los grupos PD y HC.

Por cada tarea, se capturaron los siguientes parámetros en forma de secuencias/señales temporales:

- Coordenada x : posición horizontal del bolígrafo.
- Coordenada y : posición vertical del bolígrafo.
- Marca de tiempo: registro temporal de cada punto capturado.
- Estado del botón: indicador binario, donde 0 representa el bolígrafo levantado (movimiento en el aire) y 1 el bolígrafo sobre la superficie (movimiento sobre la superficie).
- Presión perpendicular: mide la fuerza aplicada por el bolígrafo, brindando información sobre el control motor y la fuerza de agarre.
- Azimut: captura los cambios direccionales en la orientación del bolígrafo con respecto al eje horizontal de la tableta.
- Altitud: representa la inclinación del bolígrafo con respecto al eje vertical durante las tareas de escritura.

3.1.2 Selección e Integración de Tareas

La tarea #1 (Espirales de Arquímedes), común en ambos conjuntos de experimentos 3.1.1 (PaHaW) y 3.1.1 (Propietaria), ha sido seleccionada para el análisis y desarrollo de este estudio por su relevancia en la evaluación del control motor. Las características consideradas incluyen: **[Coordenada X, Coordenada Y, Altitud, Azimut, Presión]** como se ilustra en la Figura 3.3. Estas variables permiten analizar la dinámica del movimiento de escritura y diferenciar los patrones motores entre pacientes con Parkinson y controles sanos.

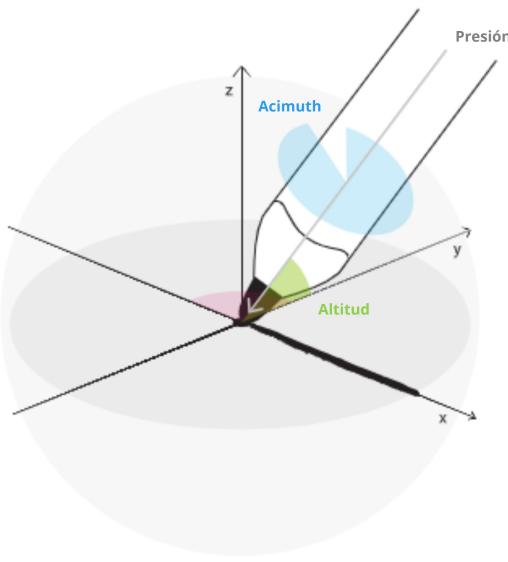


Figura 3.3: Características Dinámicas que captura el Lápiz

3.1.3 Preprocesamiento

Dado que las tabletas utilizadas en este estudio generan valores en rangos numéricos marcadamente diferentes, se adoptó un esquema de normalización en dos etapas con el fin de garantizar la comparabilidad entre dispositivos y evitar que los modelos aprendan patrones específicos del hardware.

1. Reescalado específico por dispositivo (etapa de calibración). Cada variable del conjunto de datos propietario fue linealmente transformada para corresponder al rango físico utilizado por la tableta PaHaW, siguiendo la fórmula:

$$\tilde{x} = x \cdot (x_{\max, \text{PaHaW}} - x_{\min, \text{PaHaW}}) + x_{\min, \text{PaHaW}},$$

donde x representa un valor crudo de la tableta propietaria y $x_{\min, \text{PaHaW}}$, $x_{\max, \text{PaHaW}}$ son los límites mínimo y máximo físicos observados en el dispositivo PaHaW. Este paso garantiza que, por ejemplo, un valor de presión de 0.5 corresponda al mismo nivel de fuerza en ambas tabletas, eliminando así los sesgos inducidos por el hardware. En la Tabla 3.2 se presentan los valores mínimos y máximos de cada variable utilizados en la etapa de calibración por dispositivo.

Cuadro 3.2: Valores máximos y mínimos de las variables en las bases de datos PaHaW y propietaria.

Variable	PaHaW		Propietaria	
	Mínimo	Máximo	Mínimo	Máximo
Coordenada X (xcoord)	-9079.0	9079.0	0.102235	0.389202
Coordenada Y (ycoord)	-9079.0	9079.0	-0.014404	0.234685
Azimut (azimuthY)	0.0	3589.0	-1.94	3.14
Altitud (altitudeY)	300.0	892.0	0.39	1.44
Presión (pressureY)	0.0	2048.0	0.0	1.00

3 Metodología

2. Normalización min–max global. Una vez combinados ambos conjuntos de datos (PaHaW y propietario ya reescalado), se aplicó una normalización min–max por variable:

$$x' = \frac{\tilde{x} - \tilde{x}_{\min}}{\tilde{x}_{\max} - \tilde{x}_{\min}},$$

donde \tilde{x}_{\min} y \tilde{x}_{\max} son los valores mínimos y máximos observados globalmente en la variable correspondiente. Este segundo paso asegura que todas las variables estén contenidas en el intervalo $[0, 1]$ al momento del entrenamiento, favoreciendo la estabilidad numérica de los modelos y evitando que características con mayor escala dominen el aprendizaje.

Este proceso doble de calibración y normalización mejora significativamente la comparabilidad de las muestras, previene el sesgo por dispositivo y garantiza que los valores numéricos reflejen significados físicos consistentes entre cohortes.

3.1.4 Segmentación de las Señales

El aumento de datos se llevó a cabo con el objetivo de ampliar y equilibrar las muestras de señales disponibles, mejorando así la robustez y la capacidad de generalización del modelo frente a la variabilidad inherente a los datos. Este proceso fue crucial debido al tamaño limitado de las bases de datos utilizadas.

Para implementar el aumento de datos, se emplearon ventanas deslizantes con solapamiento (*overlap*), lo que permitió segmentar las señales medidas en intervalos de diferentes tamaños. La longitud promedio de cada señal medida fue de 2301.40 puntos para la clase HC y 2709.05 puntos para la clase PD. Las configuraciones de ventanas utilizadas fueron las siguientes:

- 100 puntos (50% overlap)
- 200 puntos (25%, 50% y 75% overlap)
- 250 puntos (50% overlap)
- 400 puntos (12.5%, 25%, 50% y 75% overlap)
- 500 puntos (50% overlap)

Cada ventana generada representa un subconjunto de la señal original, permitiendo capturar información localizada y específica del movimiento de escritura. Para garantizar que cada ventana tuviera una longitud uniforme, se realizó una imputación de datos basada en la media de la ventana correspondiente. Esto asegura la consistencia en las secuencias generadas y evita la pérdida de información en los bordes de las señales segmentadas.

3.2 Modelos Propuestos

Para la detección temprana de la enfermedad de Parkinson mediante la tarea de clasificación supervisada, se seleccionaron diversos modelos de aprendizaje automático y aprendizaje profundo, considerando sus distintas fortalezas y capacidades de generalización. Este trabajo se enmarca dentro de un enfoque netamente *multimodal*, ya que, aunque todas las señales provienen de una misma fuente cada una representa aspectos distintos del comportamiento motor del sujeto: las coordenadas trazan el movimiento, la presión indica la fuerza aplicada, mientras que los ángulos capturan la orientación del dispositivo de entrada. Esta riqueza semántica justifica un tratamiento multimodal, con el objetivo de explotar la complementariedad entre las distintas modalidades.

3 Metodología

Los experimentos se llevaron a cabo en un sistema de computación de alto rendimiento (HPC) que cuenta con cuatro GPU NVIDIA TESLA A100, cada una equipada con 40 GB de memoria VRAM. Estos recursos computacionales fueron proporcionados por la *Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA)*. Para la implementación, se emplearon PyTorch 2.0.1¹ y Python 3.8².

El conjunto de datos fue procesado mediante dos representaciones complementarias:

1. **Imágenes generadas a partir de las coordenadas x, y :** Esta representación permite capturar patrones espaciales y morfológicos del trazo mediante imágenes bidimensionales, las cuales fueron posteriormente analizadas utilizando modelos de aprendizaje profundo basados en redes neuronales convolucionales. Para generar estas imágenes, se empleó la librería Matplotlib de Python, configurando cada imagen con un tamaño estándar de 224x224 píxeles, fondo blanco, líneas negras con grosor 3, y formato de almacenamiento JPG. Un ejemplo ilustrativo se presenta en la Figura 3.4.
2. **Serie temporal de presión, azimuth y altitud:** Esta representación permite capturar las propiedades dinámicas y temporales finas asociadas al movimiento del trazo, como la variación de la presión aplicada y los cambios en los ángulos del lápiz durante la escritura.

La integración de estas dos representaciones complementarias, junto con sus respectivos modelos, permite realizar un análisis integral del movimiento desde múltiples perspectivas, incrementando así la capacidad de detección temprana de patrones asociados a la enfermedad de Parkinson.

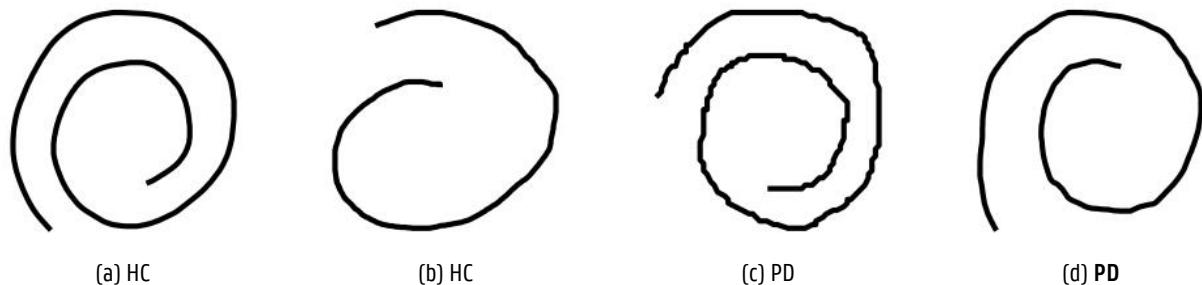


Figura 3.4: Ejemplo de las imágenes de los grupos PD y HC generadas a partir de las coordenadas x, y , utilizadas como entrada para los modelos de aprendizaje profundo. Las imágenes corresponden a ventanas de 400 puntos con un solapamiento del 75%.

3.2.1 Modelos de Aprendizaje No Profundo

Utilizando el conjunto de datos integrado, se entrenaron los siguientes modelos de aprendizaje no profundo para procesar las señales temporales de presión, azimuth y altitud. La elección de cada uno de los modelos se describe a continuación:

- **Regresión Logística** La regresión logística es un modelo lineal ampliamente utilizado en clasificación binaria por su simplicidad y robustez. Su principal ventaja radica en su interpretación directa, ya que estima probabilidades

¹<https://pytorch.org/get-started/previous-versions/#v201>

²<https://www.python.org/downloads/release/python-380/>

3 Metodología

basadas en una función sigmoide. En este caso, se ajustaron el parámetro de regularización C y el solucionador (*solver*) *liblinear*, diseñado específicamente para problemas de alta dimensionalidad [33].

- **Support Vector Classifier (SVC)** El clasificador SVC es un modelo ampliamente utilizado en tareas de clasificación debido a su capacidad para encontrar un hiperplano óptimo que maximice la separación entre clases en el espacio de características. Su principal ventaja radica en su efectividad para trabajar con datos de alta dimensionalidad y manejar problemas linealmente no separables mediante el uso de núcleos (*kernels*) como el lineal, el radial base (RBF) o el polinómico. En este modelo, se ajustaron parámetros clave como C , que controla la regularización, el tipo de *kernel* y el coeficiente γ , que define la influencia de los puntos de datos en la decisión del modelo [34].
- **Decision Trees** Los árboles de decisión dividen iterativamente el espacio de características maximizando la ganancia de información en cada división. Este modelo es fácil de interpretar y comprender, siendo ideal para identificar patrones en datos complejos. Se ajustaron parámetros como la profundidad máxima del árbol (*max_depth*), el criterio para medir la calidad de las divisiones (*criterion*) y el número mínimo de muestras necesarias para dividir un nodo (*min_samples_split*) [35].
- **Random Forest** El modelo Random Forest es un algoritmo basado en ensamblajes que combina múltiples árboles de decisión para mejorar la precisión y reducir el riesgo de sobreajuste. Su ventaja principal es su capacidad para manejar conjuntos de datos con ruido y variables no lineales, proporcionando predicciones robustas y generalizables. En este estudio, se optimizaron parámetros como el número de árboles (*n_estimators*), la profundidad máxima de los árboles (*max_depth*) y el número mínimo de muestras requeridas para dividir un nodo (*min_samples_split*) [36].
- **Naive Bayes** El clasificador Naive Bayes Gaussiano es eficiente y simple, ya que asume una distribución normal para cada característica. Su ventaja principal es su capacidad para trabajar bien con conjuntos de datos pequeños y de alta dimensionalidad. En este caso, se ajustó el parámetro de suavizado (*var_smoothing*) para mejorar la estabilidad de las predicciones [37].
- **Grading Boosting** Gradient Boosting es un modelo de ensamblaje que combina árboles de decisión de manera secuencial, corrigiendo los errores cometidos por modelos previos. Su ventaja principal es su alta precisión en problemas complejos, ya que optimiza una función de pérdida utilizando gradientes. Se ajustaron parámetros como la tasa de aprendizaje (*learning rate*), la profundidad de los árboles (*max_depth*) y el número de estimadores (*n_estimators*) [38].
- **Extra Trees** Extra Trees (Extremely Randomized Trees) es una variación de Random Forest que introduce una mayor aleatoriedad en el proceso de división de los árboles, lo que mejora su robustez frente a datos ruidosos. Su ventaja principal radica en su eficiencia computacional y en su capacidad para manejar grandes volúmenes de datos. Se ajustaron parámetros como el número de árboles (*n_estimators*) y la profundidad máxima (*max_depth*) [39].
- **K-Nearest Neighbors** El algoritmo KNN clasifica nuevas instancias en función de la distancia a sus vecinos más cercanos en el espacio de características. Es especialmente útil en escenarios donde los datos tienen distribuciones no lineales. Su principal ventaja es su simplicidad y ausencia de supuestos sobre la distribución de los datos. En este estudio, se ajustaron parámetros como el número de vecinos k , la métrica de distancia (*euclidean* o *manhattan*) y el peso asignado a los vecinos (*uniform* o *distance*) [40].

3 Metodología

Ingeniería de Características

A partir de cada segmento extraído de las ventanas deslizantes, se generaron características adicionales de series de datos temporales utilizando la librería TS-Fresh v0.21.0³[41]. TSFresh es una herramienta especializada en la extracción automatizada de características a partir de series temporales. Se optó por esta implementación debido a su capacidad para ofrecer explicabilidad de las características en tareas de regresión y clasificación, aprovechando los métodos que esta provee (*FC-ComprehensiveParameters*).

Las características generadas se pueden agrupar en categorías como:

- **Estadísticas básicas:** media, desviación estándar, mínimo, máximo, y percentiles.
- **Transformaciones basadas en Fourier:** coeficientes de la transformada de Fourier y análisis de frecuencias.
- **Medidas de correlación:** características relacionadas con correlaciones entre segmentos de la señal.
- **Autocorrelaciones y pruebas estadísticas:** coeficientes de autocorrelación y resultados de pruebas como la prueba de Augmented Dickey-Fuller.
- **Características derivadas de la dinámica de los datos:** información sobre la tendencia, curtosis y asimetría.

De las características extraídas, se seleccionaron aquellas que fueran ortogonales entre sí, es decir, con baja redundancia, y que tuvieran una correlación significativa con la salida. Para esto, se analizaron tanto las correlaciones lineales como no lineales, utilizando un umbral de correlación del 70%.

Posteriormente, se utilizó la función *SelectKBest* [42] de la librería scikit-learn v1.6.1⁴ para seleccionar las 10 mejores características que representan los datos considerando un umbral de correlación mayor a 40%. La función *SelectKBest* evalúa cada característica individualmente en función de su relación con la variable objetivo, utilizando pruebas estadísticas como el estadístico chi-cuadrado o ANOVA según el tipo de datos procesados.

Entrenamiento

Todos los modelos fueron optimizados utilizando una búsqueda en rejilla (*grid search*) para ajustar sus hiperparámetros. Para cada modelo, se seleccionaron hiperparámetros específicos en función de su impacto en el rendimiento, como se muestra en la Tabla 3.3.

Estos hiperparámetros fueron elegidos debido a su influencia significativa en el rendimiento y la capacidad de generalización de sus respectivos modelos.

Adicionalmente, se aplicó validación cruzada con $k = 5$ particiones utilizando *StratifiedKFold* de scikit-learn para todos los modelos, con el fin de asegurar la robustez de los resultados y mejorar su capacidad de generalización.

3.2.2 Modelos de Aprendizaje Profundo

En este estudio se consideraron dos enfoques principales de aprendizaje profundo, cada uno orientado al análisis de diferentes representaciones de los datos recolectados durante la tarea de dibujo de espirales.

³<https://tsfresh.readthedocs.io/en/v0.21.0/>

⁴<https://scikit-learn.org/stable/index.html>

3 Metodología

Cuadro 3.3: Hiperparámetros ajustados para cada modelo

Modelo	Hiperparámetros
Support Vector Classifier (SVC)	Regularization (C), Kernel Type, Polynomial Degree , Kernel Coefficient
Random Forest	Number of Trees, Maximum Depth, Minimum Samples Split , Minimum Samples Leaf
Logistic Regression	Regularization Strength , Penalty Type , Solver
Gradient Boosting	Learning Rate, Maximum Depth, Number of Estimators
K-Nearest Neighbors (KNN)	Number of Neighbors, Distance Metric, Weighting
Decision Tree	Criterion, Maximum Depth, Minimum Samples Split
Naive Bayes	Variance Smoothing Factor
Extra Trees	Number of Trees, Maximum Depth, Minimum Samples Split

El primer enfoque se basa en el uso de Redes Neuronales Convolucionales (*Convolutional Neural Networks*, CNN) para procesar imágenes generadas a partir de las coordenadas x e y , capturando así patrones espaciales y morfológicos del trazo. Este tipo de arquitectura ha demostrado ser eficaz en tareas de clasificación de imágenes biomédicas, permitiendo extraer características relevantes asociadas a la regularidad, simetría y fluidez del dibujo.

El segundo enfoque corresponde al uso de Redes Neuronales de Picos (*Spiking Neural Networks*, SNN), una clase de modelos bioinspirados que simulan de manera más realista la dinámica de disparo de las neuronas biológicas. Estas redes fueron utilizadas para la clasificación tanto de las series temporales (presión, azimut y altitud) como de las representaciones visuales derivadas de las coordenadas espaciales. Las SNN presentan ventajas en el modelado de señales temporales complejas y pueden ofrecer un marco más interpretativo al estar alineadas con principios neurofisiológicos.

Ambos enfoques permiten comparar el desempeño y la interpretabilidad de arquitecturas profundas tradicionales y bioinspiradas, en el contexto de un análisis multimodal para la detección temprana de la enfermedad de Parkinson.

Redes Neuronales Convolucionales (CNN)

Utilizando el conjunto de datos integrado por las imágenes generadas a partir de las coordenadas x , y , se entrenaron cinco arquitecturas de Redes Neuronales Convolucionales (CNN) para la tarea de clasificación entre pacientes con PD y HC. En todos los modelos, la capa de salida fue modificada para incluir un perceptrón multicapa (*Multilayer Perceptron*, MLP) totalmente conectada, adaptado específicamente para clasificación binaria.

Las arquitecturas seleccionadas fueron elegidas en función de su rendimiento comprobado tanto en estudios previos del grupo de investigación como en la literatura científica reciente [43, 44, 45], donde han demostrado una alta precisión y eficiencia en tareas de clasificación similares, especialmente en el contexto de imágenes biomédicas. Cada modelo fue optimizado específicamente para las imágenes de escritura a mano utilizadas en este estudio. La Tabla 3.4 presenta un resumen de las especificaciones arquitectónicas, incluyendo el número de capas, las transformaciones aplicadas y las capas finales de clasificación.

A continuación, se describen brevemente las arquitecturas utilizadas:

3 Metodología

- **ResNet50:** Red residual profunda con 50 capas que resuelve el problema del desvanecimiento del gradiente mediante bloques residuales, permitiendo el entrenamiento eficiente de redes muy profundas. Fue seleccionada por su capacidad para extraer características ricas y por su amplio uso exitoso en aplicaciones médicas [46, 47].
- **EfficientNet-B0:** Esta arquitectura destaca por su escalabilidad eficiente en profundidad, anchura y resolución, lo que le permite alcanzar alta precisión con un número reducido de parámetros. Fue seleccionada por su balance entre eficiencia computacional y rendimiento [48].
- **DenseNet-121:** Conecta cada capa con todas las capas posteriores de forma directa, lo que mejora el flujo del gradiente, reduce la redundancia y fomenta la reutilización de características. Fue elegida por su capacidad de mitigar el desvanecimiento del gradiente y su eficacia comprobada en tareas de clasificación médica [49, 50].
- **MobileNetV2:** Diseñada para dispositivos con recursos limitados, emplea convoluciones separables en profundidad y bloques residuales tipo bottleneck. Fue seleccionada por su eficiencia y aplicabilidad en contextos donde se requiere bajo consumo computacional [51].
- **Ensamble de Modelos:** Se construyó un modelo de ensamble combinando las predicciones de ResNet50, EfficientNet-B0, DenseNet-121 y MobileNetV2. El ensamble se realizó mediante la técnica de *bagging* (Bootstrap Aggregating), entrenando cada modelo individualmente sobre subconjuntos aleatorios del conjunto de entrenamiento generados con reemplazo. Esta estrategia permite reducir el sobreajuste y la varianza del modelo, mejorando su capacidad de generalización. Las predicciones finales se obtuvieron promediando las salidas de cada modelo individual, lo cual resultó en un clasificador más robusto y preciso [52].

Cuadro 3.4: Detalles arquitectónicos de cada modelo CNN, incluyendo el número de capas, transformaciones internas y la estructura de las capas de clasificación. (FC) se refiere a capas totalmente conectadas, (ReLU) a la función de activación Rectified Linear Unit y las capas Dropout se utilizan para prevenir sobreajuste.

Modelo	Arquitectura
ResNet50	50 capas: 49 convolucionales, 1 totalmente conectada. Conexiones residuales. Salida: FC(2048) → FC(512) → ReLU → Dropout → FC(1)
EfficientNet-B0	16 capas: bloques Conv-BN-ReLU, convoluciones separables en profundidad. Salida: FC(1280) → Dropout → FC(1)
DenseNet-121	121 capas: bloques densos, convoluciones 1x1 y 3x3. Salida: FC(1024) → FC(512) → ReLU → Dropout → FC(1)
MobileNetV2	53 capas: convoluciones separables en profundidad, bloques residuales tipo bottleneck. Salida: FC(1280) → Dropout → FC(1)
Ensamble	Combinación de ResNet50, EfficientNet-B0, DenseNet-121 y MobileNetV2. Predicciones promedio entre los modelos individuales.

3 Metodología

Preprocesamiento

Todas las imágenes fueron normalizadas utilizando los valores de media y desviación estándar de [0.0287, 0.0285, 0.0285] y [0.0071, 0.0070, 0.0070] respectivamente, ejemplos de estas se pueden observar en la Figura 3.5. Estos valores corresponden a la distribución de intensidades de píxeles en las imágenes del dataset, y fueron adoptados para alinear los datos con los requisitos de entrada de las arquitecturas convolucionales utilizadas. La normalización es un paso esencial en el entrenamiento eficiente de modelos de aprendizaje profundo, ya que permite estabilizar el proceso de aprendizaje al centrar los valores de píxel alrededor de cero y escalarlos dentro de un rango consistente. En concreto, durante el preprocesamiento se resta la media a cada canal de color y se divide por su respectiva desviación estándar, evitando problemas como la explosión o desaparición del gradiente durante la retropropagación.

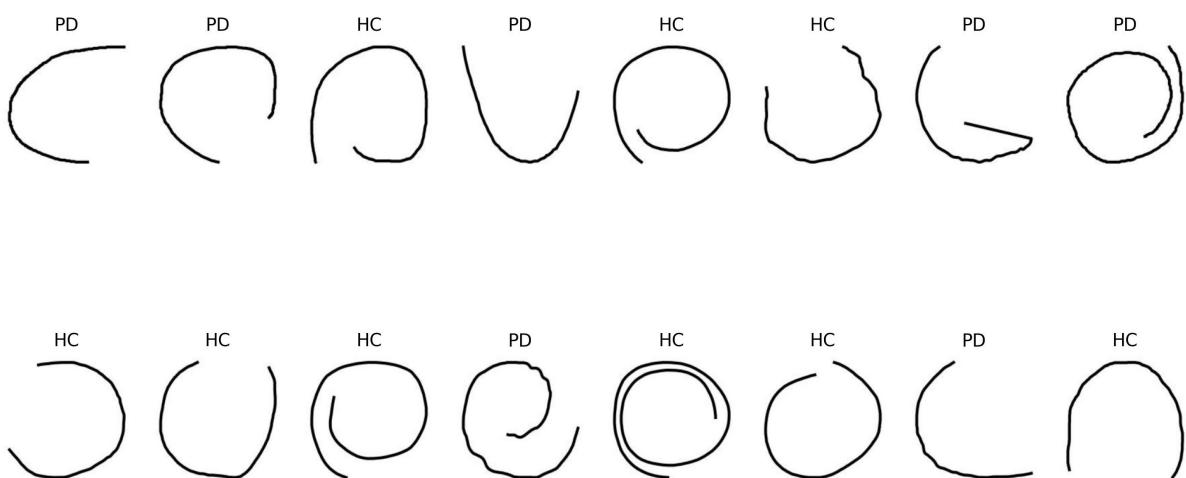


Figura 3.5: Ejemplo de imágenes del conjunto de prueba después del preprocesamiento

Aumento de Datos

Con el objetivo de aumentar la diversidad del conjunto de entrenamiento y mejorar la capacidad de generalización de los modelos, se aplicaron diversas técnicas de aumento de datos exclusivamente sobre las imágenes de entrenamiento. Estas transformaciones fueron clave para mitigar el sobreajuste, ya que proporcionaron múltiples variaciones del mismo trazo sin alterar la información esencial de los patrones de escritura de los sujetos. Las transformaciones se diseñaron para preservar la estructura del trazo mientras se introducían variaciones visuales que simulan diferentes condiciones de adquisición.

Las transformaciones se agruparon en dos categorías: (1) aquellas aplicadas a todas las imágenes, y (2) transformaciones aleatorias, en las cuales se seleccionó una de cada grupo con una probabilidad determinada. Tras el aumento, el conjunto de datos se incrementó a un total de 1200 imágenes, permitiendo una mayor variabilidad durante el entrenamiento.

Las transformaciones aplicadas a todas las imágenes fueron:

- **Rotación aleatoria:** Gira la imagen aleatoriamente entre ángulos de 90° y -90°.
- **Flip:** Aplica volteos aleatorios horizontales y verticales.

3 Metodología

- **Transposición:** Transpone la imagen aleatoriamente.
- **Ruido gaussiano:** Agrega ruido gaussiano con una probabilidad del 20%, simulando ruido en la adquisición de imágenes.

Las siguientes transformaciones se aplicaron seleccionando aleatoriamente una opción dentro de cada grupo, con una probabilidad del 30%:

- **Desenfoque:** Aplica una de las siguientes transformaciones: *MotionBlur*, *MedianBlur* o *Blur*, simulando variaciones en el enfoque.
- **Distorsiones geométricas:** Aplica una de las siguientes: *Optical Distortion*, *Grid Distortion* o *Piecewise Affine*, introduciendo deformaciones geométricas.
- **Mejoras de imagen:** Aplica una de las siguientes: *CLAHE*, *Sharpen*, *Emboss* o *Random Brightness Contrast*, mejorando el contraste o la nitidez para simular condiciones de iluminación variadas.

Estas transformaciones se aplicaron con una probabilidad del 20–30%, logrando un balance entre variabilidad e integridad del dato. El aumento de datos se realizó de manera dinámica durante el entrenamiento (*on-the-fly*), lo cual evitó la necesidad de almacenamiento adicional y proporcionó una mayor variabilidad en tiempo real.

La Figura 3.6 presenta ejemplos de imágenes aumentadas utilizadas durante el entrenamiento lo que evidencia la diversidad introducida en el entrenamiento y la consistencia mantenida en la evaluación.

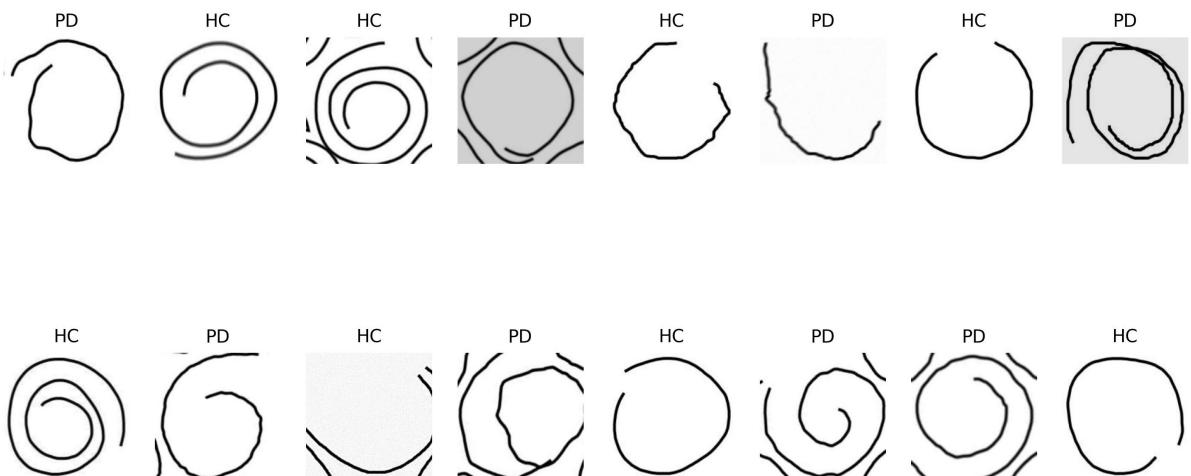


Figura 3.6: Ejemplo de imágenes después de aplicar el aumento de datos de los grupos PD y HC para la ventana de 400 con overlaping del 75%.

Entrenamiento

Todos los modelos fueron optimizados mediante una búsqueda en rejilla (*grid search*) para ajustar hiperparámetros clave: el valor de *dropout*, la tasa de aprendizaje (*learning rate*) y el *weight decay*. Estos parámetros fueron seleccionados por su

3 Metodología

influencia directa en la capacidad del modelo para aprender de manera eficiente sin sobreajustarse.

Además, se empleó validación cruzada estratificada con $k = 5$ particiones utilizando la clase *StratifiedKFold* de la librería *scikit-learn*, lo que permitió garantizar la robustez del entrenamiento y una mejor capacidad de generalización. Esta técnica asegura que cada subconjunto preserve la proporción de clases original, lo cual es especialmente importante en conjuntos de datos con clases desbalanceadas o limitadas.

Spiking Neural Networks

Las Redes Neuronales de Picos (*Spiking Neural Networks*, SNNs) son un tipo de red neuronal inspirada en el funcionamiento del cerebro humano. A diferencia de las redes neuronales tradicionales, donde la información se transmite de manera continua, en las SNNs las neuronas se comunican enviando señales discretas llamadas *spikes* (picos), que ocurren en momentos específicos del tiempo.

Estas redes se basan en un modelo llamado *Leaky Integrate-and-Fire* (LIF), donde cada neurona acumula señales de entrada a lo largo del tiempo. Cuando la señal acumulada supera un cierto umbral, la neurona genera un pico y transmite esa señal a otras neuronas. Si el umbral no se alcanza, parte de la señal se pierde con el tiempo a través de un "canal de fuga", simulando cómo las neuronas reales dejan de responder si no reciben suficiente estimulación.

Este tipo de codificación permite que las SNNs procesen información no solo por la cantidad de señal, sino también por el momento en que ocurre. Gracias a esto, las SNNs son especialmente útiles para detectar patrones en señales temporales, como las que se producen al escribir a mano, ya que pueden capturar mejor la dinámica del movimiento y los cambios sutiles en el tiempo.

Preprocesamiento de Datos

Para entrenar la red neuronal de picos (SNN), se procesaron tanto las series de tiempo como las imágenes generadas a partir de los datos de escritura. En esta sección se detalla el preprocesamiento aplicado a las señales temporales de presión, azimut y altitud.

Series de Tiempo Las Spiking Neural Networks requieren que los datos de entrada se representen en forma de secuencias de spikes (picos). Para lograr esto, se aplicaron dos estrategias de codificación:

- **Rate Coding:** La información se codifica mediante la frecuencia de disparo de una neurona dentro de un intervalo de tiempo. Cuanto mayor es la magnitud del estímulo, más frecuente será la emisión de spikes. En la implementación, cada valor de la señal se normaliza al rango $[0, 1]$ y se asocia a una distribución de Bernoulli con esa probabilidad. Durante un número fijo de pasos temporales (*num_steps*), se realiza un muestreo en cada paso; si el resultado es un "éxito", se genera un spike, en caso contrario no.
- **Temporal Coding (Latency Coding):** En esta estrategia, la información se representa a través del momento en que ocurre el primer spike dentro de la ventana temporal. Un valor alto provoca un spike temprano, mientras que un valor bajo lo retrasa. Esta codificación permite capturar relaciones temporales de manera más precisa.

3 Metodología

Las tres variables de entrada fueron transformadas en secuencias de spikes utilizando las siguientes configuraciones:

- **Altitud y presión:** Codificadas mediante *Latency Coding*, dado que su dinámica temporal puede reflejar cambios sutiles en el control motor.
- **Azimut:** Debido a su naturaleza cíclica ($0\text{--}360^\circ$), se descompuso en sus componentes seno y coseno para preservar la continuidad angular. Luego, ambas componentes fueron codificadas utilizando *Rate Coding*.

La Figura 3.7 muestra un ejemplo visual de las señales de altitud, azimut y presión codificadas en formato spike, correspondiente a una ventana de 400 puntos con un solapamiento del 75%, utilizando 50 pasos temporales, un umbral (*threshold*) de 0.5 y una constante de tiempo (τ) de 10.

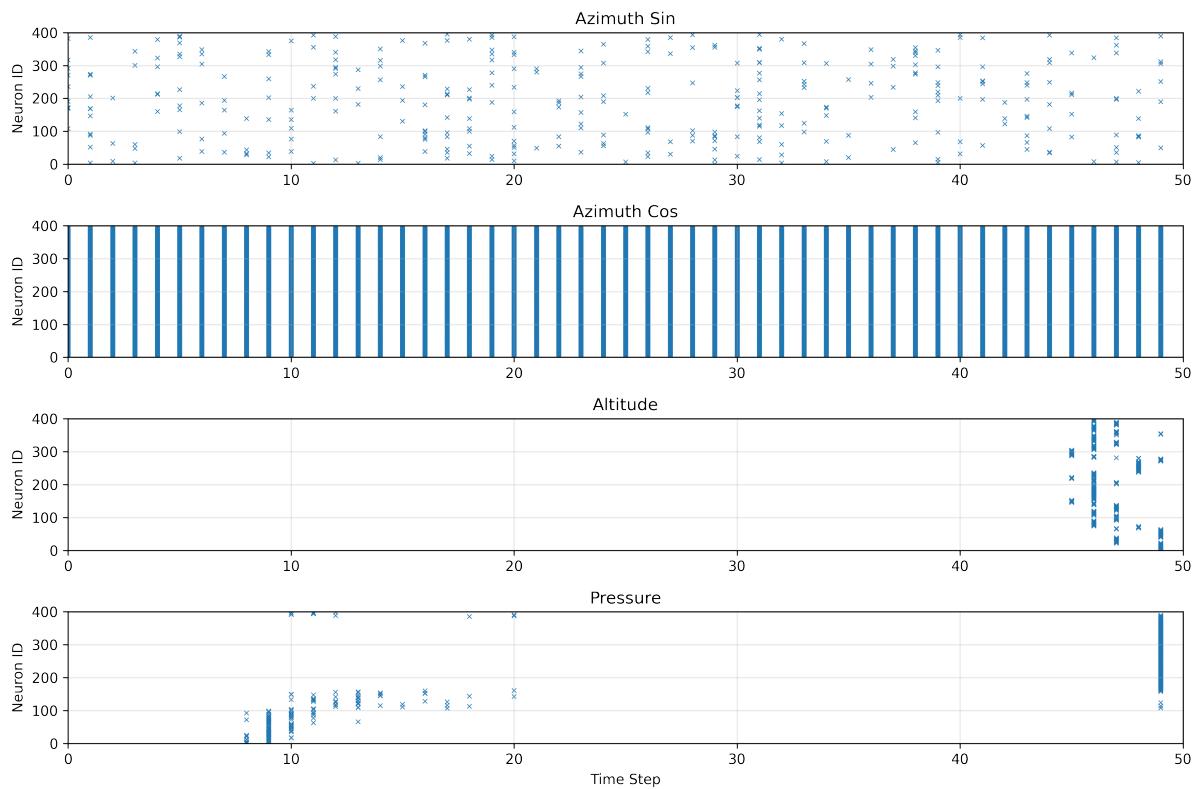


Figura 3.7: Ejemplo de codificación en spikes aplicado a señales motoras de escritura. Las dos primeras filas muestran los spikes generados mediante *Rate Coding* a partir de las componentes seno y coseno del azimut. Las dos filas inferiores representan la codificación mediante *Latency Coding* para las señales de altitud y presión, donde la aparición temprana de spikes indica mayor intensidad en la señal original.

En este contexto:

- **Pasos temporales (*time steps*):** representan las unidades discretas de tiempo en las que se simula la propagación de los spikes. A mayor número de pasos, mayor resolución temporal tiene la codificación.
- **Umbral (*threshold*):** valor mínimo que debe alcanzar la señal para que se genere un spike. Un valor de 0.5 implica

3 Metodología

que solo los estímulos de media o alta intensidad desencadenan actividad neuronal.

- **Constante de tiempo (τ):** parámetro que regula la rapidez con la que decrece la probabilidad de disparo en la codificación por latencia. Valores bajos generan spikes más tempranos para estímulos intensos, mientras que valores más altos permiten una codificación más dispersa en el tiempo.

Imágenes Para el procesamiento de las imágenes en el contexto de las redes neuronales de picos (SNNs), se emplearon dos estrategias distintas de representación visual de las coordenadas x, y . A diferencia de las señales temporales, en las imágenes no es necesario realizar una codificación explícita en spikes, ya que la red puede inferir directamente la información a partir de los valores de intensidad de los píxeles durante el entrenamiento. Es decir, la codificación de la información está implícita en la variación espacial del trazo.

Las estrategias utilizadas fueron las siguientes:

- **Secuencia completa:** Se generaron imágenes a partir de la totalidad de la secuencia de coordenadas x, y , trazando el dibujo completo con líneas negras sobre fondo blanco. Esta representación estática proporciona información morfológica global del trazo. Esta técnica fue también utilizada en el entrenamiento de las CNN, como se describe en la sección anterior. Ejemplos de estas imágenes pueden observarse en la Figura 3.4.
- **Secuencia por fotogramas (frames):** En esta modalidad, el trazo se construye de forma progresiva, añadiendo cuatro nuevos puntos por cada fotograma, lo que genera una secuencia visual que simula el proceso temporal del dibujo. De esta forma, se obtiene una representación visual dinámica de la escritura, tratándola como una serie de tiempo de imágenes. Esta estrategia permite a la red capturar la evolución temporal del trazo a través de la progresión espacial de los píxeles. Un ejemplo de esta representación puede verse en la Figura 3.8.

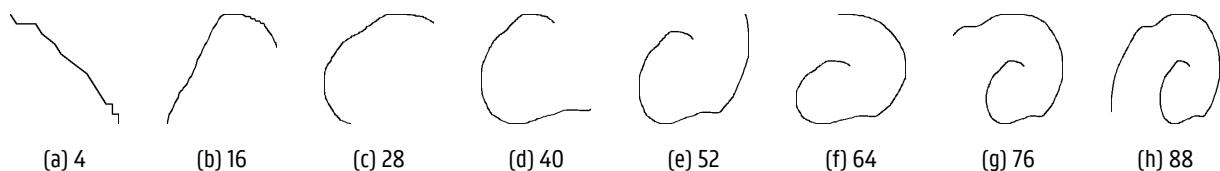


Figura 3.8: Secuencia de imágenes generadas a partir de una espiral manuscrita, correspondiente a una ventana de 400 puntos dividida en 100 fotogramas. Cada imagen muestra el estado progresivo del dibujo al acumular 4 nuevos puntos por frame, permitiendo representar la dinámica del trazo en el tiempo. El avance ocurre en sentido antihorario. Los números indican el índice del fotograma dentro de la secuencia.

Arquitecturas propuestas

En este estudio se emplearon distintas arquitecturas de SNNs ajustadas específicamente para procesar información multimodal derivada de tareas de escritura a mano. Estas arquitecturas fueron seleccionadas con base en su respaldo en la literatura y su capacidad para capturar tanto patrones espaciales como dinámicas temporales complejas, características inherentes a las señales biomotoras involucradas en la detección temprana de la enfermedad de Parkinson.

3 Metodología

Se exploraron seis configuraciones principales de modelos SNN, diferenciadas por el tipo de entrada (series de tiempo o imágenes), el tipo de fusión multimodal (temprana o tardía), y el uso de bloques spiking simples (LIF) o recurrentes (SLSTM):

- **Early Fusion (LIF y SLSTM):** Todas las señales multimodales (presión, azimut y altitud) son concatenadas desde el inicio (Figura 3.9) y procesadas por una única red. Esta variante es útil cuando existe una fuerte correlación entre las señales. El uso de neuronas LIF permite entrenamientos más rápidos, mientras que los bloques SLSTM agregan memoria temporal, mejorando la modelación de relaciones de largo plazo.
- **Late Fusion (LIF y SLSTM):** Las señales se procesan por separado a través de bloques independientes de capas spiking como se muestra en la Figura 3.10. Esto permite explotar la especificidad de cada modalidad. Finalmente, las salidas son fusionadas mediante concatenación antes de la clasificación. La inclusión de SLSTM en cada rama mejora la capacidad para extraer dependencias temporales particulares en cada señal, algo clave en variables como presión o altitud.
- **Convolutional SNN:** Utilizada para procesar imágenes completas generadas a partir del trazo x, y del paciente. Emplea capas convolucionales seguidas de neuronas spiking y una capa totalmente conectada. La entrada se replica a lo largo de varios pasos temporales para adaptarse al paradigma spiking. Esta arquitectura permite capturar patrones morfológicos relevantes en los trazos.
- **Convolutional + SLSTM SNN:** Procesa secuencias de imágenes construidas a partir de fotogramas del trazo (frames de coordenadas). Combina convoluciones espaciales con bloques SConv2dLSTM, permitiendo modelar simultáneamente tanto la variación intra-imagen como la evolución temporal entre imágenes. Esta arquitectura resulta especialmente adecuada para representar el proceso de escritura como una serie de micro-movimientos.
- **Ensamble:** Dada la naturaleza complementaria de las arquitecturas anteriores, se implementó un modelo de ensamble que combina las predicciones de los modelos Late Fusion SLSTM, Early Fusion SLSTM y Convolutional+SLSTM SNN. Este enfoque busca aumentar la robustez del sistema y mejorar la capacidad de generalización al mitigar los sesgos de cada arquitectura individual.

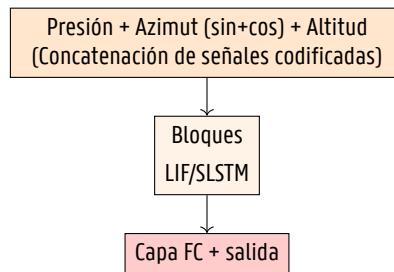


Figura 3.9: Esquema de arquitectura **Early Fusion**. Todas las señales multimodales son concatenadas desde el inicio y procesadas por una única red spiking compuesta por bloques LIF o SLSTM.

La exploración sistemática de estas configuraciones permite comparar sus capacidades de representación y discriminación bajo un enfoque netamente multimodal. En la tabla 3.5 se muestra en detalle la arquitectura de cada una de las redes utilizadas.

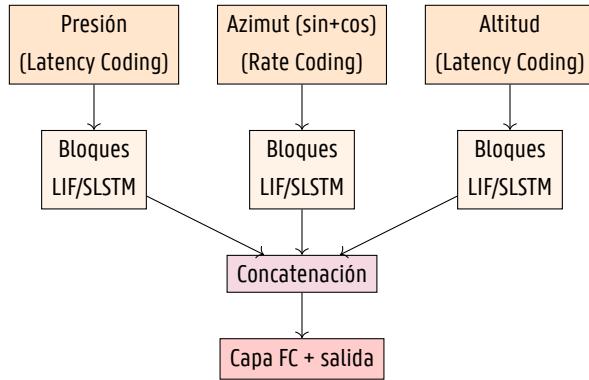


Figura 3.10: Esquema de arquitectura **Late Fusion**. Cada modalidad es procesada por separado a través de bloques spiking (LIF o SLSTM), y sus salidas se fusionan mediante concatenación antes de la capa final de clasificación.

Aumento de Datos

El aumento de datos aplicado en los modelos spiking sobre imágenes se realizó utilizando las mismas configuraciones descritas previamente en la sección de las CNNs. Tanto las imágenes completas generadas a partir de las coordenadas x, y como las secuencias de fotogramas (frames) de las espirales fueron procesadas con técnicas de aumento de datos consistentes.

Para asegurar la coherencia temporal y preservar la integridad de los trazos, se aplicó la misma configuración de transformaciones a todas las imágenes pertenecientes a una misma secuencia de espiral. Sin embargo, cada secuencia recibió un conjunto de transformaciones distinto, con el objetivo de incrementar la diversidad del conjunto de entrenamiento sin comprometer la continuidad interna de los frames.

3.3 Evaluación

3.3.1 Métricas de Desempeño

El desempeño de los modelos fue evaluado utilizando métricas ampliamente aceptadas en tareas de clasificación binaria: exactitud (*accuracy*), precisión, sensibilidad (*recall*) y puntaje F1. Estas métricas permiten obtener una visión integral sobre la capacidad de los modelos para distinguir entre sujetos sanos (HC) y pacientes con enfermedad de Parkinson (PD) [53, 54]. En particular, el puntaje F1 resulta especialmente útil en escenarios con posibles desbalances entre clases, al combinar en una sola medida la precisión y la sensibilidad del modelo.

3.3.2 Protocolo Experimental

El protocolo de evaluación se desarrolló en varias fases, garantizando tanto la comparabilidad como la capacidad de generalización de los modelos:

- Inicialmente, el conjunto de datos *PaHaW* fue dividido en un 80% para entrenamiento y validación, y un 20% para prueba.

3 Metodología

Cuadro 3.5: Detalles arquitectónicos de cada modelo SNN. Se incluyen las dimensiones de entrada, estructura por capas, tipo de neuronas (LIF o SLSTM), y forma de fusión.

Modelo	Arquitectura detallada
Late Fusion (LIF)	3 ramas paralelas: <ul style="list-style-type: none"> Presión y Altitud: entrada de 400. Azimut (sin + cos): entrada de 800. Cada rama: FC → LIF → Dropout → FC → LIF → Dropout → FC → LIF. Fusión por concatenación: FC(3×128) → LIF → FC(1).
Early Fusion (LIF)	Entrada fusionada de 1600 (azimut sin/cos, presión, altitud). Arquitectura MLP secuencial: FC(1600) → LIF → Dropout → FC → LIF → Dropout → FC → LIF → FC(1).
Late Fusion SLSTM	3 ramas paralelas con proyección inicial: <ul style="list-style-type: none"> Presión y Altitud: Proyección FC(400 → 200). Azimut (sin/cos): FC(800 → 400). Cada rama: SLSTM(2 bloques). Fusión por concatenación → SLSTM combinado → FC(1).
Early Fusion SLSTM	Entrada fusionada (1600). Procesamiento por: SLSTM → Dropout → SLSTM → Promedio temporal → FC(1). Ideal para capturar dinámicas globales de las señales codificadas.
Convolutional SNN	Imágenes RGB (1 canal) de tamaño 224×224 replicadas durante 50 pasos temporales. Procesamiento: Conv2D(3 bloques) + LIF (stride=2, sin pooling) → Flatten → FC → LIF → Dropout → FC(1).
Convolutional + SLSTM SNN	Secuencia de imágenes frame por frame (cada una representa el trazo progresivo). 2 capas SConv2dLSTM + Dropout + Pooling adaptativo (8×8) → FC(1). Promedio temporal aplicado antes de la clasificación final.
Ensamble	Fusión de predicciones de: Late Fusion (SLSTM), Early Fusion (SLSTM), y Convolutional SNN.

- Para la etapa preliminar de selección de ventanas deslizantes, se utilizó un subconjunto de 200 muestras por clase (HC y PD) a fin de acelerar el proceso de extracción de características mediante la herramienta TSFRESH.
- Una vez identificada la configuración óptima de ventana (longitud de 400 y solapamiento del 75%), se utilizó el conjunto completo de datos *PaHaW* para entrenamiento.
- Para mejorar la capacidad de generalización del modelo, se incorporó aleatoriamente el 30% del conjunto de datos propietario tanto en el entrenamiento como en la prueba.
- Finalmente, el 70% restante del conjunto de datos propietario fue reservado exclusivamente para pruebas sobre datos no vistos.

3 Metodología

Durante todo el proceso, se empleó validación cruzada estratificada (*Stratified K-Fold*) para asegurar una distribución balanceada de las clases en cada partición y evitar sesgos en la evaluación.

Comparación de enfoques: Todos los modelos de aprendizaje automático (ML) y aprendizaje profundo (DL y SNN) fueron evaluados bajo los mismos conjuntos de prueba, lo cual permitió realizar comparaciones directas entre enfoques tradicionales y arquitecturas más complejas de manera justa y controlada.

3.4 Pruebas de Significancia

Dado que este estudio busca desarrollar modelos de detección temprana de la enfermedad de Parkinson que sean robustos y generalizables, independientemente del origen geográfico o cultural de los sujetos, se planteó como paso inicial la evaluación estadística de las diferencias entre los conjuntos de datos empleados. En particular, se analizó si existían diferencias significativas entre el conjunto *PaHaW* (República Checa) y el conjunto propietario (Ecuador), considerando que posteriormente serán integrados en un único dataset para el entrenamiento y validación de los modelos.

Esta comparación es especialmente relevante debido a que los datos provienen de poblaciones que utilizan alfabetos distintos, y cuyas dinámicas motoras y hábitos de escritura pueden diferir notablemente por factores lingüísticos, educativos o culturales. El objetivo final es construir un modelo *agnóstico*, capaz de generalizar entre poblaciones sin depender del contexto específico de escritura.

Para ello, se realizaron pruebas de significancia sobre tres variables de gran relevancia biomotriz: la presión perpendicular, el ángulo de azimut y el ángulo de altitud del lápiz. Estas características fueron seleccionadas debido a que, si bien el patrón de dibujo (una espiral) es el mismo para ambos grupos, dichas señales reflejan aspectos individuales del control motor que pueden variar entre poblaciones, incluso bajo tareas gráficas idénticas.

En primer lugar, se aplicó la prueba de Shapiro-Wilk para evaluar la normalidad de cada variable. Dado que todas las variables presentaron valores $p < 0.05$, se rechazó la hipótesis de normalidad y se procedió con la prueba no paramétrica de Mann-Whitney U para comparar las distribuciones entre los dos conjuntos de datos.

3.5 Explicabilidad

3.5.1 Técnicas Utilizadas

La interpretabilidad de los modelos desarrollados en este estudio fue una prioridad fundamental, especialmente considerando su relevancia en el uso clínico.

Para ello, en los modelos clásicos de aprendizaje automático basados en características extraídas (como presión máxima, altitud promedio, etc.), se utilizó **SHAP (Shapley Additive Explanations)** [55]. Esta técnica se basa en teoría de juegos y permite cuantificar la contribución individual de cada característica en una predicción específica. SHAP asigna un valor de Shapley a cada variable de entrada, facilitando así la interpretación de decisiones del modelo, la identificación de las variables más relevantes y la detección de posibles sesgos o relaciones inesperadas.

4

Resultados

En este capítulo se presentan los resultados obtenidos a partir de los diferentes experimentos realizados. Se incluyen análisis estadísticos para evaluar la compatibilidad entre los conjuntos de datos utilizados, así como los desempeños alcanzados por los distintos modelos de clasificación. También se exploran las salidas de las técnicas de explicabilidad del mejor grupo de modelos para una mejor interpretación de los resultados.

4.1 Diferencias Estadísticas entre Conjuntos de Datos

La Tabla 4.1 resume los resultados obtenidos al aplicar la prueba de normalidad de Shapiro-Wilk y la prueba no paramétrica de Mann-Whitney U. Como se observa, todas las variables presentaron distribuciones no normales, lo cual justificó el uso de la prueba de Mann-Whitney U para comparar los grupos.

Cuadro 4.1: Resultados de las pruebas de Shapiro-Wilk y Mann-Whitney U para las características seleccionadas.

Característica	Conjunto de Datos	Shapiro-Wilk (p)	Mann-Whitney U (p)
Altitud (Media)	Propietario	0.9733 (3.81×10^{-9})	0.4378
	PaHaW	0.9814 (3.27×10^{-15})	
Azimut (Media)	Propietario	0.9056 (4.66×10^{-19})	4.29×10^{-24}
	PaHaW	0.6943 (1.42×10^{-50})	
Presión (Media)	Propietario	0.8737 (6.55×10^{-22})	0.0293
	PaHaW	0.9915 (3.33×10^{-9})	

Los resultados muestran que existen diferencias significativas en dos de las tres variables analizadas: la **presión perpendicular** ($p = 0.0293$) y el **azimut del lápiz** ($p = 4.29 \times 10^{-24}$). No se observaron diferencias significativas en la **altitud del lápiz** ($p = 0.4378$), lo que sugiere cierta estabilidad de este parámetro entre poblaciones.

Estas diferencias confirman que los patrones de escritura varían entre los dos grupos, probablemente influenciados por factores culturales, biomecánicos o de estilo de escritura. Esta observación justifica el uso de técnicas de normalización y el diseño de modelos multimodales, robustos a este tipo de variabilidad intergrupal. Así, se busca garantizar que el sistema de

4 Resultados

detección propuesto no dependa del origen del participante, sino de las señales biomotoras asociadas al Parkinson.

4.2 Aumento de Datos

La Tabla 4.2 presenta la cantidad inicial de muestras disponibles por grupo (sanos y pacientes con EP) antes de aplicar el aumento de datos con ventanas deslizantes.

Cuadro 4.2: Cantidad de datos antes del proceso de aumento con ventanas deslizantes.

PaHaW		Propietaria	
HC	PD	HC	PD
32	32	9	12

La Tabla 4.3 resume la cantidad total de muestras obtenidas después del aumento. Cabe destacar que las diferencias en el número de muestras por sujeto se deben a la duración variable que cada participante requería para completar la tarea de dibujo, lo cual afectaba el número total de puntos capturados por la tableta digitalizadora.

Cuadro 4.3: Cantidad de datos después del aumento con ventanas deslizantes.

Ventana	Paso	PaHaW		Propietaria	
		HC	PD	HC	PD
100	50	1653	2156	306	435
200	50	1534	2092	288	411
200	100	1151	1071	151	217
200	150	563	726	105	149
250	125	650	851	120	169
400	50	1461	1964	252	363
400	100	754	1007	133	193
400	200	403	525	133	193
400	300	284	368	53	77
500	250	316	418	58	81

4.3 Enfoque de Aprendizaje Poco Profundo

Esta sección presenta los resultados experimentales obtenidos al aplicar técnicas de aprendizaje no profundo para la clasificación de la EP a partir del análisis de escritura. Se incluyen los resultados de rendimiento de los modelos, el proceso de optimización de hiperparámetros, las evaluaciones sobre ambos conjuntos de datos utilizados, y los hallazgos derivados de la aplicación de técnicas de inteligencia artificial explicable.

4 Resultados

4.3.1 Rendimiento de los Modelos

La Tabla 4.4 resume las métricas de precisión alcanzadas por ocho modelos de clasificación, evaluados con distintas configuraciones de ventanas deslizantes y tamaños de paso (overlap). Estos resultados permiten observar el impacto que tienen la longitud de la ventana y el grado de solapamiento en el rendimiento de los modelos.

Cuadro 4.4: Métricas de precisión para diferentes tamaños de ventana y solapamientos en distintos modelos de clasificación.

Tamaño de ventana	Paso	SVC	Random Forest	Regresión Logística	Gradient Boosting	KNN	Árbol de Decisión	Naive Bayes	Extra Trees
100	50	0.7333	0.8333	0.7500	0.8000	0.7000	0.7333	0.8333	0.8000
200	50	0.8500	0.8667	0.8667	0.8333	0.9167	0.8833	0.8500	0.8833
200	100	0.8167	0.8667	0.7667	0.9167	0.8833	0.8500	0.7500	0.8833
200	150	0.8500	0.8167	0.8500	0.8500	0.8500	0.8167	0.8167	0.8333
250	125	0.9000	0.9000	0.9000	0.9167	0.9000	0.8000	0.8667	0.9167
400	50	0.9833	0.9667	0.9333	0.9167	0.9833	0.9500	0.9000	0.9833
400	100	0.9667	0.9333	0.8833	0.9333	0.9500	0.8167	0.8500	0.9667
400	200	0.7500	0.7667	0.7833	0.7667	0.7500	0.6333	0.8000	0.7833
400	300	0.7500	0.7500	0.7667	0.7167	0.7167	0.7333	0.7500	0.7333
500	250	0.7667	0.7000	0.7500	0.7500	0.7833	0.6833	0.7833	0.8167

Los resultados muestran que el mejor desempeño se alcanzó utilizando una ventana deslizante de 400 puntos y un paso de 50 puntos, configuración con la cual todos los modelos superaron el 90% de precisión. Esta configuración fue, por tanto, seleccionada para los experimentos posteriores con modelos de aprendizaje profundo.

4.3.2 Rendimiento con Conjuntos de Datos Combinados

Una vez identificada la configuración óptima de ventana y solapamiento, los modelos fueron reentrenados utilizando un conjunto de datos combinado que incluyó el 100% del conjunto PaHaW y el 30% del conjunto de datos propietario. Los resultados obtenidos se resumen en la Tabla 4.5.

Cuadro 4.5: Métricas de rendimiento utilizando el conjunto combinado (100% PaHaW + 30% Propietario).

Modelo	Acc	Precision	Recall	F1-Score
SVC	0.8826	0.8826	0.8826	0.8826
Random Forest	0.9155	0.9157	0.9155	0.9154
Logistic Regression	0.6737	0.6731	0.6737	0.6728
Gradient Boosting	0.9554	0.9555	0.9554	0.9554
KNN	0.9413	0.9413	0.9413	0.9413
Decision Tree	0.8732	0.8733	0.8732	0.8731
Naive Bayes	0.6761	0.6778	0.6761	0.6720
Extra Trees	0.9507	0.9508	0.9507	0.9507

Los modelos Gradient Boosting, KNN y Extra Trees obtuvieron los mejores resultados de forma consistente, destacándose Gradient Boosting como el modelo con mayor precisión (95.54%).

4 Resultados

4.3.3 Resultados sobre el Conjunto de Datos Propietario

La prueba final se realizó utilizando el 70% restante del conjunto de datos propietario, el cual contenía 177 muestras de sujetos sanos (HC) y 255 muestras de pacientes con Parkinson (PD), tras aplicar la técnica de segmentación por ventanas (window slicing). Los resultados obtenidos se resumen en la Tabla 4.6.

Cuadro 4.6: Resultados de prueba sobre el conjunto de datos propietario.

Modelo	Accuracy	Precision	Recall	F1-Score
SVC	0.9072	0.9084	0.9072	0.9075
Random Forest	0.9026	0.9038	0.9026	0.9029
Logistic Regression	0.6961	0.6922	0.6961	0.6924
Gradient Boosting	0.9327	0.9328	0.9327	0.9325
KNN	0.9258	0.9265	0.9258	0.9259
Decision Tree	0.7819	0.7815	0.7819	0.7817
Naive Bayes	0.5545	0.5236	0.5545	0.5221
Extra Trees	0.9420	0.9420	0.9420	0.9420

El modelo Extra Trees obtuvo el mejor desempeño en esta etapa de prueba, alcanzando una precisión del 94.20%, seguido por Gradient Boosting con 93.27% y KNN con 92.58%. Estos resultados confirman la capacidad de generalización de los modelos entrenados, incluso al enfrentarse a datos no vistos provenientes de otra población.

4.3.4 Matrices de Confusión

La Figura 4.3 presenta las matrices de confusión correspondientes a todos los modelos. Estas matrices permiten analizar con mayor detalle el comportamiento de clasificación de cada modelo, mostrando la distribución de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Los altos valores en la diagonal principal reflejan una alta precisión en la clasificación de ambas clases.

4.3.5 Mejores Hiperparámetros

La Tabla 4.7 presenta los mejores hiperparámetros encontrados para cada modelo después del proceso de ajuste, utilizando el conjunto de datos combinado. Estos hiperparámetros fueron seleccionados para maximizar el rendimiento y garantizar predicciones confiables.

4.4 Enfoque de Aprendizaje Profundo

A continuación, se presentan los resultados de los experimentos realizados con redes neuronales convolucionales (CNNs) y redes neuronales de picos (SNNs), entrenadas con el conjunto de datos combinado (100% del conjunto PaHaW y 30% del conjunto propietario). Posteriormente, se evaluaron los modelos utilizando el 70% restante del conjunto propietario, correspondiente exclusivamente a participantes de Ecuador.

4 Resultados

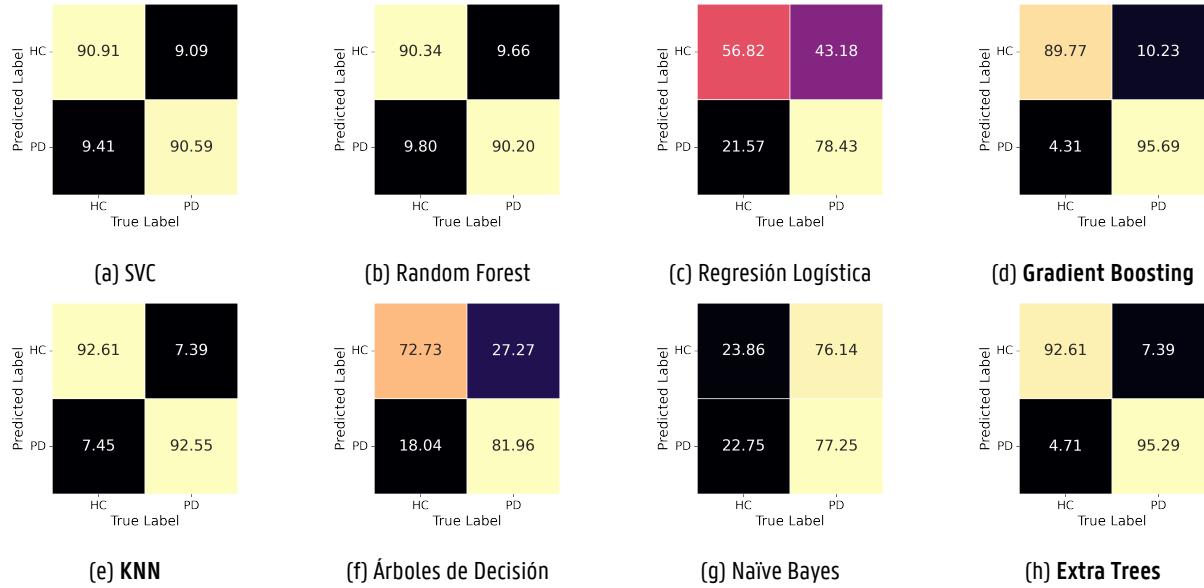


Figura 4.1: Matrices de confusión de los distintos modelos de aprendizaje automático evaluados con la base de datos propietaria.

Cuadro 4.7: Mejores hiperparámetros para cada modelo (30% Datos Propietarios + PaHaW).

Modelo	Mejores Hiperparámetros
SVC	$C: 100$, Degree: 2, Gamma: Auto, Kernel: RBF
Random Forest	Max Depth: None, Min Samples Leaf: 1, Min Samples Split: 2, Number of Estimators: 100
Logistic Regression	Regularization (C): 0.01, Penalty: L2, Solver: Liblinear
Gradient Boosting	Learning Rate: 0.2, Max Depth: 7, Number of Estimators: 200
KNN	Distance Metric: Manhattan, Number of Neighbors (k): 3, Weights: Distance
Decision Tree	Criterion: Gini, Max Depth: 15, Min Samples Split: 2
Naïve Bayes	Variance Smoothing: 1×10^{-9}
Extra Trees	Max Depth: None, Min Samples Split: 2, Number of Estimators: 200

4.4.1 Redes Neuronales Convolucionales

La Tabla 4.8 presenta las métricas de rendimiento para los cuatro modelos de CNN entrenados (ResNet50, EfficientNet-B0, DenseNet121 y MobileNetV2), así como para el modelo *ensemble* construido a partir de ellos.

En esta evaluación inicial, el modelo **EfficientNet-B0** se destacó por su capacidad de balancear precisión y recall, lo cual sugiere que puede detectar casos de Parkinson con menor número de falsos negativos. Sin embargo, el modelo **ensemble**, aunque logra una mayor precisión (0.8573), presenta un leve descenso en el recall, lo que indica una mayor tasa de falsos negativos. Este comportamiento sugiere que la combinación ponderada de modelos mejora la confiabilidad general pero puede ser más conservadora al predecir la clase minoritaria. ResNet50 mostró el peor rendimiento con un recall extremadamente bajo, lo cual indica un sesgo significativo hacia la clase sana.

4 Resultados

Cuadro 4.8: Métricas de rendimiento de los modelos CNN utilizando el conjunto combinado (100% PaHaW + 30% Propietario)

Modelo	Acc	Precisión	Recall	F1-Score
ResNet50	0.6140	0.8187	0.1927	0.3180
EfficientNet	0.7968	0.8410	0.6825	0.7512
DenseNet	0.6838	0.7974	0.4125	0.5460
MobileNet	0.7624	0.7896	0.6582	0.7145
Ensemble	0.7968	0.8573	0.6637	0.7451

La Tabla 4.9 muestra los resultados sobre el conjunto de datos propietario donde los modelos EfficientNet y el ensemble siguen mostrando un desempeño competitivo.

Cuadro 4.9: Métricas de rendimiento sobre el conjunto de datos propietario

Modelo	Acc	Precisión	Recall	F1-Score
ResNet50	0.5960	0.7949	0.1890	0.3054
EfficientNet	0.7736	0.8195	0.6646	0.7340
DenseNet	0.6648	0.7765	0.4024	0.5301
MobileNet	0.7421	0.7721	0.6402	0.7000
Ensemble	0.7736	0.8346	0.6463	0.7285

Matrices de Confusión

La Figura 4.2 muestra las matrices de confusión para los cinco modelos CNN, revelando patrones consistentes con las métricas previamente discutidas. En particular:

- El modelo **ResNet50** presenta un comportamiento claramente sesgado hacia la clase HC, con una baja capacidad de detección de casos PD.
- **EfficientNet** y **MobileNetV2** logran un mayor balance entre clases, aunque aún se evidencia una tendencia a clasificar erróneamente algunos sujetos con Parkinson como sanos.
- **DenseNet121** muestra un desempeño intermedio con un número importante de falsos negativos.
- El **ensemble** refuerza la robustez al combinar fortalezas de los modelos base, aunque aún existe una brecha en la sensibilidad para la clase PD.

En términos biomecánicos, los resultados obtenidos con las redes neuronales convolucionales sugieren que las representaciones visuales derivadas de la escritura capturan patrones motores relevantes para la clasificación de la enfermedad de Parkinson. Modelos como EfficientNet y MobileNet lograron detectar con mayor eficacia a los sujetos con deterioro motor.

A pesar de que algunos modelos presentaron un sesgo hacia la clase sana, los resultados indican que el enfoque basado en imágenes tiene potencial clínico, especialmente cuando se combina con técnicas de ensamblado que integran fortalezas

4 Resultados

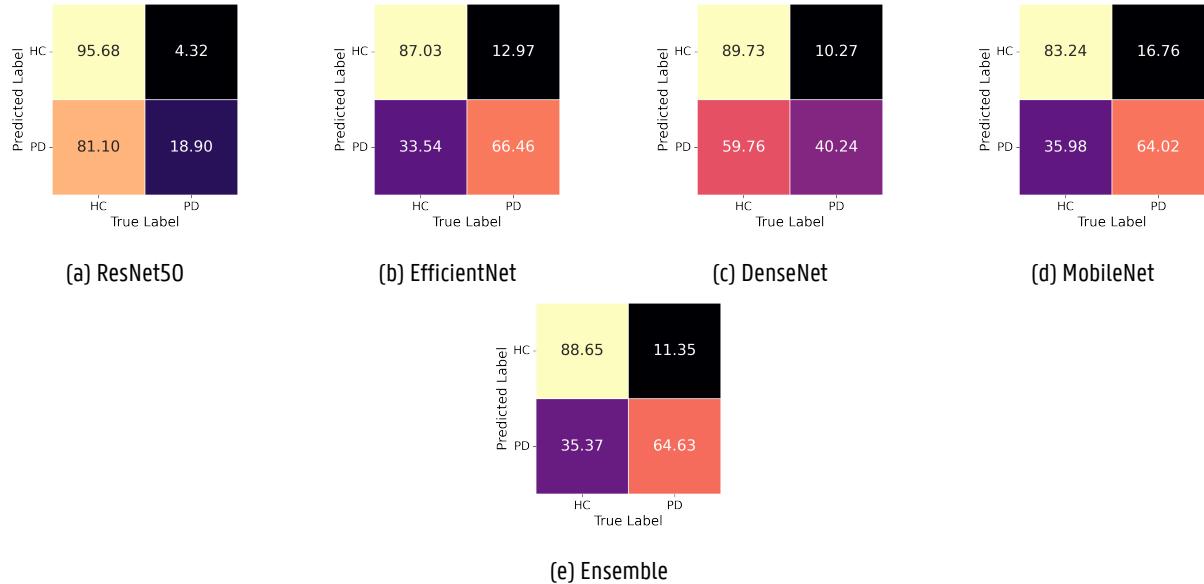


Figura 4.2: Matrices de confusión de los distintos modelos de aprendizaje profundo evaluados con la base de datos propietaria.

individuales de múltiples arquitecturas. Esta capacidad de discriminación basada en trazos escritos posicione a las CNNs como una herramienta prometedora para el desarrollo de biomarcadores digitales no invasivos en contextos de cribado temprano o monitoreo remoto de síntomas motores.

Mejores Hiperparámetros

La Tabla 4.10 resume los mejores hiperparámetros encontrados para cada arquitectura, utilizando búsqueda en rejilla y validación cruzada estratificada. Cabe destacar que todos los modelos fueron entrenados con la función de pérdida **BCE with logits loss**, que permitió una mayor estabilidad numérica al tratarse de una tarea de clasificación binaria con predicción basada en logits.

Cuadro 4.10: Hiperparámetros óptimos seleccionados para cada modelo CNN.

Modelo	Dropout Rate	Learning Rate	Weight Decay
ResNet50	0.3	0.001	0.001
EfficientNet-BO	0.3	0.001	0.0001
DenseNet121	0.5	0.001	0.001
MobileNetV2	0.5	0.001	0.0001

4.4.2 Redes Neuronales de Picos

La Tabla 4.11 resume el rendimiento de los modelos SNN evaluados sobre el conjunto de validación (100% PaHaW + 30% Propietario), incluyendo un ensemble basado en las tres mejores arquitecturas: Late Fusion SLSTM, Early Fusion SLSTM y

4 Resultados

Convolutional + SLSTM SNN.

Cuadro 4.11: Métricas de rendimiento de los modelos SNNs utilizando el conjunto combinado (100% PaHaW + 30% Propietario).

Modelo	Acc	Precisión	Recall	F1-Score
Late Fusion (LIF)	0.7357	0.71823	0.7182	0.7182
Early Fusion (LIF)	0.7049	0.6894	0.6768	0.6831
Late Fusion SLSTM	0.7564	0.7257	0.7744	0.7493
Early Fusion SLSTM	0.7116	0.7211	0.8346	0.7737
Convolutional SNN	0.6524	0.6145	0.6037	0.6055
Convolutional + SLSTM SNN	0.5337	0.5017	0.8122	0.6203
Ensamble	0.6931	0.6468	0.7631	0.7001

Por su parte, la Tabla 4.12 muestra el rendimiento sobre el conjunto de prueba (70% del dataset propietario).

Cuadro 4.12: Métricas de rendimiento sobre el conjunto de datos propietario (70% Propietario).

Modelo	Acc	Precisión	Recall	F1-Score
Late Fusion (LIF)	0.7163	0.7019	0.6890	0.6954
Early Fusion (LIF)	0.6877	0.6687	0.6646	0.6667
Late Fusion SLSTM	0.7507	0.7278	0.7500	0.7387
Early Fusion SLSTM	0.7070	0.7162	0.8346	0.7709
Convolutional SNN	0.6493	0.6256	0.6218	0.6237
Convolutional + SLSTM SNN	0.4721	0.5398	0.7205	0.6172
Ensamble	0.6775	0.6323	0.7459	0.6844

Los resultados sobre el conjunto de validación muestran que el modelo **Early Fusion SLSTM** alcanza el mejor desempeño general, con un *F1-Score* de 0.7737 y un *recall* de 0.8346. Esto sugiere que la fusión temprana de señales spiking, en combinación con una arquitectura recurrente, permite capturar eficazmente los patrones temporales asociados a la enfermedad de Parkinson.

Le sigue el modelo **Late Fusion SLSTM**, que también presenta un rendimiento destacado (*F1-Score*: 0.7493), manteniendo un buen equilibrio entre precisión y sensibilidad. Por otro lado, los modelos **LIF sin recurrencia** obtienen resultados aceptables, aunque inferiores, evidenciando las ventajas de incorporar mecanismos de memoria temporal.

El modelo **Convolutional + SLSTM SNN**, a pesar de haber sido entrenado solo durante tres épocas debido a sus altas exigencias computacionales, logra un notable *recall* (0.8122), aunque con baja precisión (0.5017). Esto indica una fuerte tendencia a identificar correctamente los casos positivos, aunque a costa de un mayor número de falsos positivos. Aun así, demuestra el potencial de combinar información espacial y temporal proveniente de imágenes derivadas de las señales originales.

En el conjunto de prueba, se observa una tendencia similar. El modelo **Early Fusion SLSTM** mantiene el mejor desempeño (*F1-Score*: 0.7709, *Recall*: 0.8346), lo que demuestra su capacidad de generalización. El modelo **Convolutional + SLSTM SNN**

4 Resultados

incrementa levemente su precisión (0.5398), manteniendo una alta sensibilidad, lo que refuerza su utilidad en escenarios donde los falsos negativos son críticos.

En general, los modelos basados en **SLSTM** destacan por su capacidad para modelar secuencias complejas, mientras que las variantes convolucionales abren nuevas posibilidades en el uso de representaciones visuales. Estos resultados reflejan el alto potencial de las redes neuronales de picos en aplicaciones biomédicas, particularmente cuando se integran componentes recurrentes o convolucionales.

Matrices de Confusión

La Figura 4.3 muestra las matrices de confusión de los modelos SNN sobre el conjunto de prueba. El modelo **Early Fusion SLSTM** no solo alcanza el mayor *recall*, sino también una clara capacidad de identificar correctamente sujetos con Parkinson. En contraste, el modelo **CNN + SLSTM**, a pesar de su baja precisión, logra una notable sensibilidad, lo cual puede ser útil en tareas donde los falsos negativos deben minimizarse.

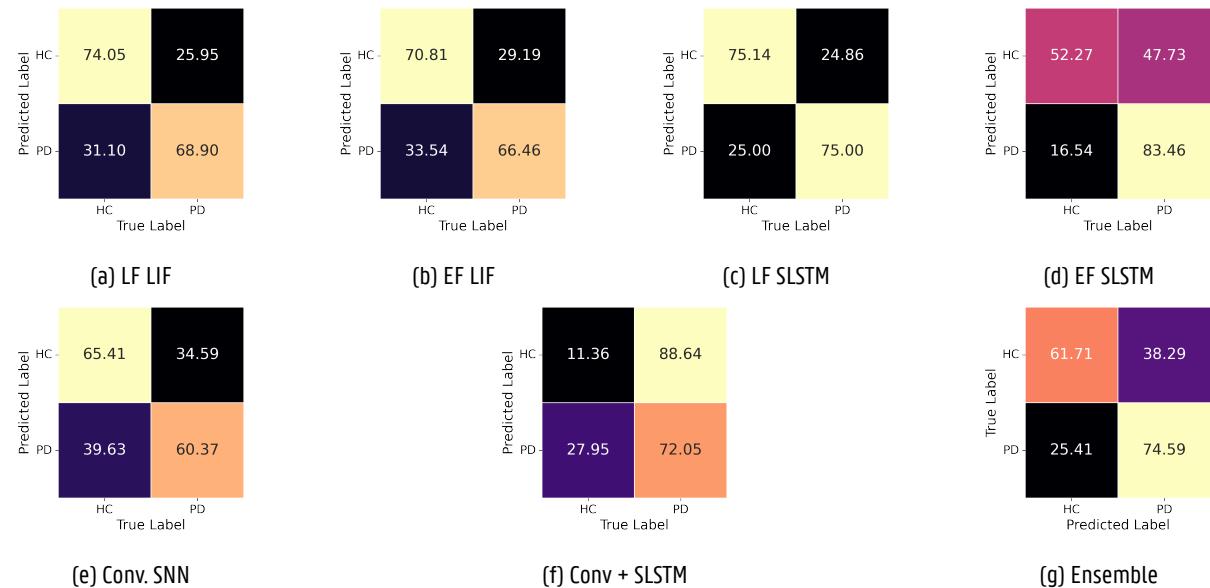


Figura 4.3: Matrices de confusión de los distintos modelos de SNNs evaluados con la base de datos propietaria.

Los resultados obtenidos con las SNNs reflejan el gran potencial de esta tecnología para capturar patrones temporales complejos en tareas multimodales, como el análisis de escritura para la detección de la EP. Entre las distintas variantes exploradas, los modelos basados en **SLSTM** destacaron por su mayor capacidad para modelar secuencias temporales prolongadas y dinámicas neuronales complejas.

Uno de los hallazgos más interesantes se obtuvo con el modelo **Convolutional + SLSTM**, que, aunque presentó un rendimiento general más bajo debido a su alto costo computacional, permitió explorar la representación temporal de secuencias de imágenes derivadas de las señales originales. Este modelo fue entrenado durante **3 épocas**, sin aplicar técnicas de reducción de resolución, lo que refleja tanto sus elevadas exigencias como su prometedor potencial.

El número de épocas se fijó tras aplicar un criterio de parada temprana, ya que experimentos preliminares mostraron que

4 Resultados

la pérdida de validación se estabilizaba antes de la cuarta época. Además, dado que cada época en una red de tipo SNN con 50 pasos temporales equivale a unas 10 épocas en una red convencional, entrenar más allá de este punto incrementaba considerablemente el costo computacional sin aportar mejoras significativas en el rendimiento, y aumentaba el riesgo de sobreajuste debido al tamaño limitado del conjunto de datos. La arquitectura de este modelo permite aprovechar simultáneamente la información espacial y temporal, lo que lo convierte en una herramienta valiosa para futuros desarrollos, especialmente en contextos con mayor capacidad de cómputo.

En general, si bien los modelos tradicionales como LIF mostraron un rendimiento competitivo, el uso de estructuras recurrentes y convolucionales dentro del marco de las SNNs permite modelar de manera más realista el procesamiento cerebral. Esto se refleja en el mejor desempeño de los modelos SLSTM frente a las arquitecturas puramente feedforward.

Cabe destacar que el entrenamiento de estos modelos fue computacionalmente intensivo, requiriendo el uso de técnicas de optimización avanzadas así como múltiples rondas de ajuste de hiperparámetros por grupo funcional. Esta limitación impidió explorar entrenamientos más prolongados, especialmente en arquitecturas híbridas (CNN+SNN), pero no compromete su validez conceptual ni su aplicabilidad futura en sistemas con recursos especializados.

Mejores Hiperparámetros

A continuación, se presentan los criterios que guiaron la definición de los rangos de búsqueda y la metodología utilizada para el ajuste de hiperparámetros, con el objetivo de garantizar configuraciones óptimas, reproducibles y acordes a las características de cada arquitectura evaluada.

- **Búsqueda jerárquica mixta (coarse – fine).** Se combinó *random search* (50 iteraciones) para explorar rangos amplios para cada hiperparámetro, por ejemplo, $\text{LR} \in [10^{-5}, 10^{-2}]$, $\text{hidden size} \in [64, 1\,600]$, $\tau \in [5, 25]$ —con una fase posterior de *grid search* fina ($\pm 20\%$ alrededor de los mejores valores) guiada por la métrica *balanced accuracy*.
- **Validación estratificada $k=5$ + parada temprana.** Cada configuración se evaluó con $k = 5$ particiones estratificadas y *early stopping* (pacienza = 8 épocas) para evitar sobreajuste y reducir tiempo de cómputo.
- **Optimización y regularización.** AdamW se prefirió sobre Adam clásico por su desacoplamiento explícito del *weight decay*, crucial para SNNs con muchos parámetros latentes [56]; los valores de *weight decay* se barrieron log-uniformemente en $[10^{-6}, 10^{-3}]$.
- **Parámetros de disparo (τ , *threshold*, *beta*).** Los intervalos se fijaron a partir de estudios previos sobre estabilidad de gradiente en neuronas LIF/SLSTM y se ajustaron en función de la longitud de secuencia (*num steps*); τ altos favorecen la integración temporal en secuencias largas, mientras que *threshold/beta* controlan la esparsidad de los spikes y, por ende, el consumo de memoria [57, 58, 59].
- **Funciones de gradiente diferenciable.** Se probaron aproximaciones STE (*Straight-Through Estimator*) [60], *atan*, *sigmoid*, *fast_sigmoid* y *triangular*, la selección final se basó en la combinación que maximizó la estabilidad del entrenamiento (varianza del gradiente < 0.15) y la puntuación de validación.
- **Dropout y tamaños de batch.** Los coeficientes de *dropout* se ajustaron de forma inversa al tamaño de *batch* para

4 Resultados

mantener constante la tasa efectiva de regularización; en arquitecturas "Early Fusion" el *batch size* se redujo (8–16) para acomodar secuencias más largas, compensándolo con mayores valores de *dropout*.

Este procedimiento permitió identificar configuraciones que equilibran rendimiento, estabilidad numérica y coste computacional; los hiperparámetros resultantes se sintetizan en las Tablas 4.13 y 4.14.

Cuadro 4.13: Mejores hiperparámetros por grupos funcionales para cada arquitectura SNN.

Modelo	Hiperparámetros Agrupados
Late Fusion LIF	<p>Optimización: AdamW, LR: 2.03×10^{-4}, Weight Decay: 3.78×10^{-4}</p> <p>Arquitectura: Hidden Size: 716, Num Steps: 66, Batch Size: 64</p> <p>Spiking: Tau: 19, Threshold: 0.821, Learn Threshold: False, Learn Beta: False</p> <p>Codificación Neuronal: Gradientes por variable: Altitud (atan), Azimut (sigmoid), Presión (fast_sigmoid)</p> <p>Dropout: Altitud [0.81, 0.77, 0.75], Azimut [0.26, 0.51, 0.71], Presión [0.53, 0.78, 0.10]</p> <p>Betas: Altitud [0.98, 0.61, 0.64], Azimut [0.82, 0.28, 0.33], Presión [0.77, 0.96, 0.68]</p> <p>Umbrales por Variable: Altitud [0.007, 0.056, 0.033], Azimut [0.476, 0.025, 0.255], Presión [0.274, 0.026, 0.005]</p>
Early Fusion LIF	<p>Optimización: AdamW, LR: 4.68×10^{-4}, Weight Decay: 6.46×10^{-5}</p> <p>Arquitectura: Hidden Size: 1200, Num Steps: 182, Batch Size: 8</p> <p>Spiking: Tau: 7, Threshold: 0.959, Learn Beta: True, Learn Threshold: False</p> <p>Dropout: [0.27, 0.20, 0.06]</p> <p>Betas: [0.77, 0.23, 0.08]</p> <p>Gradientes: [STE, atan, atan]</p> <p>Umbrales: [0.0071, 0.7506, 0.2049]</p>
CNN + SNN	<p>Optimización: AdamW, LR: 3.76×10^{-4}, Weight Decay: 2.66×10^{-6}</p> <p>Arquitectura: Hidden Size: 143, Num Steps: 22, Batch Size: 32</p> <p>Convoluciones: Conv1: [Canales=8, Kernel=4], Conv2: [64, 6], Conv3: [40, 4]</p> <p>Spiking: Thresholds: [0.0033, 0.0264, 0.0173, 0.0172]</p> <p>Betas: [0.91, 0.62, 0.59, 0.50], Learn Beta: True (except Beta4)</p> <p>Gradientes: [fast_sigmoid, triangular, triangular, fast_sigmoid]</p> <p>Dropout: 0.049</p>

4.5 Explicabilidad de los Modelos

La explicabilidad en modelos de aprendizaje automático es fundamental para interpretar los patrones detectados por los algoritmos, especialmente en aplicaciones clínicas donde las decisiones deben ser comprensibles y confiables para profesionales de la salud.

En este estudio, las técnicas de explicabilidad se aplicaron exclusivamente a los modelos de aprendizaje no profundo (ML clásico), debido a su rendimiento superior, su interpretación directa y la madurez de las herramientas disponibles. Esta

4 Resultados

decisión se sustenta en dos consideraciones técnicas principales:

- **Rendimiento y estabilidad.** Los modelos clásicos (como *Extra Trees*, *Gradient Boosting* y *SVC*) obtuvieron las mejores métricas de precisión ($\approx 94\%$) durante la validación cruzada, superando a las arquitecturas profundas (CNN) y bioinspiradas (SNN). Esta estabilidad los posiciona como los candidatos más adecuados para su interpretación clínica.
- **Interpretabilidad clínica y adecuación de herramientas XAI.** Los modelos clásicos utilizados en este estudio operan sobre un conjunto reducido de variables con significado fisiológico claro (como presión máxima, entropía de altitud, cuantil bajo de azimut), lo que permite aplicar técnicas como SHAP para generar explicaciones directamente vinculadas con biomarcadores interpretables para expertos médicos [61, 62]. En contraste, aunque existen métodos de explicabilidad para modelos profundos—como Grad-CAM, DeepSHAP [63] e Integrated Gradients [64] para CNN, o métricas basadas en gradientes para SNN [65]—estos suelen producir salidas complejas y de difícil interpretación clínica.

Por estas razones, enfocar la explicabilidad en modelos de aprendizaje clásico maximiza la utilidad clínica y la robustez técnica del presente estudio. A continuación, se detallan los resultados obtenidos.

Cuadro 4.14: Mejores hiperparámetros por grupos funcionales para cada arquitectura SNN Recurrente.

Late Fusion SLSTM	Optimización: AdamW, LR: 5.78×10^{-5} , Weight Decay: 1.96×10^{-5} Arquitectura: Hidden Size: 472, Num Steps: 49, Batch Size: 4 Spiking: Tau: 12, Threshold: 0.493, Learn Beta: False, Learn Threshold: False Dropout: Altitud [0.15, 0.42], Azimut [0.24, 0.40], Presión [0.28, 0.26] Gradientes: Altitud (fast_sigmoid), Azimut (triangular), Presión (sigmoid) Umbrales por Variable: Altitud [0.0089, 0.0368], Azimut [0.0059, 0.0395], Presión [0.224, 0.0049], Combinado: 0.149
Early Fusion SLSTM	Optimización: AdamW, LR: 2.02×10^{-4} , Weight Decay: 3.10×10^{-4} Arquitectura: Hidden Size: 519, Num Steps: 79, Batch Size: 16 Spiking: Tau: 10, Threshold: 0.535, Learn Threshold1/2: False Dropout: [0.286, -, -] Gradientes: [sigmoid, sigmoid] Umbrales: [0.0478, 0.0279]
CNN + SLSTM	Optimización: AdamW, LR: 3.15×10^{-3} , Weight Decay: 3.33×10^{-4} Arquitectura: Batch Size: 1, Canales: [64, 64] Spiking: Threshold: 0.0818, Inhibición: True, Reset: subtract Dropout: 0.143 Otros: Pos Weight: 1.149, Cuantización del estado: False

4 Resultados

4.5.1 Características Seleccionadas

Las diez características más relevantes seleccionadas mediante el método *SelectKBest* se presentan a continuación, ordenadas según su correlación con la variable objetivo. La numeración del 0 al 9 indica su importancia relativa, siendo 0 la característica con mayor correlación.

0. Altitude Permutation Entropy (Dimension 7, $\tau = 1$)

Entropía por Permutación de la Altitud (Dimensión 7, $\tau = 1$). Esta característica mide la complejidad de la señal de altitud utilizando entropía por permutación, una técnica utilizada comúnmente para evaluar el grado de aleatoriedad en series temporales. Valores más altos indican mayor complejidad o irregularidad en los movimientos verticales del lápiz [66].

1. Pressure Change Quantiles Aggregated Mean ($isabs=True$, $qh=0.6$, $ql=0.0$)

Promedio Agregado de Cambios Absolutos en Presión (cuantil 0.0 a 0.6). Representa el promedio de los cambios absolutos en la presión dentro del rango definido por los cuantiles inferiores y superiores ($ql = 0.0$, $qh = 0.6$). Esta medida permite capturar fluctuaciones locales en la señal de presión [67].

2. Azimuth Sum of Reoccurring Data Points

Suma de Puntos Recurrentes en el Azimut. Calcula la suma de los valores de azimut que se repiten en la señal, lo cual puede reflejar patrones angulares repetitivos durante el trazado. Este tipo de repetición es relevante en el análisis de escritura, especialmente en el contexto de trastornos motores [68].

3. Pressure Change Quantiles Aggregated Mean ($isabs=True$, $qh=0.4$, $ql=0.2$)

Promedio Agregado de Cambios Absolutos en Presión (cuantil 0.2 a 0.4). Similar a la característica anterior, pero enfocada en un rango de cuantiles más estrecho ($ql = 0.2$, $qh = 0.4$), lo que proporciona un análisis más fino de las variaciones de presión [67].

4. Altitude Change Quantiles Aggregated Variance ($isabs=True$, $qh=1.0$, $ql=0.2$)

Varianza Agregada de Cambios Absolutos en Altitud (cuantil 0.2 a 1.0). Mide la variabilidad de los cambios absolutos en la altitud dentro del rango de cuantiles especificado. Este tipo de característica basada en la varianza permite capturar dinámicas de la señal en términos de dispersión [69].

5. Azimuth Number of Continuous Wavelet Transform Peaks ($n=1$)

Número de Picos mediante Transformada Wavelet Continua en Azimut ($n=1$). Detecta la cantidad de picos en la señal de azimut utilizando la transformada wavelet continua. Las características basadas en wavelets son útiles para detectar cambios abruptos o patrones periódicos en señales temporales [70].

6. Altitude FFT Aggregated Kurtosis

Curtosis Agregada del FFT de la Altitud. Calcula la curtosis de la señal de altitud después de aplicar la Transformada Rápida de Fourier (FFT). La curtosis evalúa la prominencia de los picos en la distribución de la señal en el dominio de la frecuencia [69].

7. Pressure Maximum

Presión Máxima. Valor máximo registrado de la presión ejercida durante el proceso de escritura. Este tipo de carac-

4 Resultados

terística es relevante para analizar la intensidad del trazo y la fuerza ejercida por el usuario [68].

8. Altitude Sample Entropy

Entropía de Muestra en la Altitud. Evalúa la complejidad de la señal de altitud mediante la entropía de muestra. Valores elevados indican mayor irregularidad o imprevisibilidad en la señal, lo cual puede estar relacionado con un control motor deteriorado [71].

9. Altitude Quantile (q=0.2)

Cuantil del 20% de Altitud. Representa el valor por debajo del cual se encuentra el 20% más bajo de los valores de altitud. Las características basadas en cuantiles permiten obtener información sobre los rangos extremos o atípicos de la señal [67].

4.5.2 Impacto de las características específicas de clase

Las Figuras 1 y 2 del Aprendice A, muestran la importancia de las 10 características seleccionadas por *SelectKBest* para distintos clasificadores, diferenciando el impacto medio por clase. Las barras azules representan la clase 0 (HC) y las barras rojas la clase 1 (PD). Cuanto mayor sea el valor absoluto del SHAP, mayor es la contribución de la característica a la predicción.

Existen dos tipos de visualización en las gráficas generadas:

- Las **gráficas tipo beeswarm** (gráficos de puntos) permiten visualizar el efecto individual de cada observación. Los puntos representan ejemplos individuales, donde el color indica el valor de la característica (de bajo a alto, de azul a rojo). La dispersión horizontal muestra el rango del impacto SHAP: puntos alejados del centro tienen mayor influencia en la predicción, ya sea positiva o negativa.
- Las **gráficas de barras apiladas** muestran el valor promedio absoluto de SHAP por característica y clase. El color azul representa la clase 0 (sujetos sanos) y el rojo la clase 1 (pacientes con Parkinson). Cuanto más larga es la barra, mayor es el impacto de esa característica en el resultado del modelo.

Para el modelo **SVC** (**Figura 1a**): las características 1 (cambios en presión en quantiles altos), 2 (suma de puntos recurrentes en azimut), 7 (presión máxima) y 9 (cuantil bajo de altitud) presentan el mayor impacto en las predicciones. Estas variables están relacionadas con la fuerza ejercida en el trazo y la repetitividad angular, aspectos clave para detectar rigidez o temblores en la escritura.

Por su parte, en el modelo de **Regresión Logística** (**Figura 1b**): se observa un patrón similar al de SVC, destacando nuevamente las variables 1, 2, 7 y 9 como determinantes. Dado que la regresión logística es más lineal, estas características son probablemente las más separables entre clases.

Mientras que en **Gradient Boosting** (**Figura 1c**): la característica 9 (cuantil bajo de altitud) domina el ranking, seguida por la presión máxima (7) y los cambios de presión (1). Este modelo muestra una alta sensibilidad a fluctuaciones finas en el trazo, lo cual es coherente con su naturaleza basada en árboles optimizados.

En el modelo de **Random Forest** (**Figura 2a**): se destaca la característica 2 (repeticiones en azimut), lo que indica que el patrón angular durante la escritura tiene un peso significativo. También sobresale la presión máxima (7), corroborando su

4 Resultados

importancia biomecánica.

De igual manera, en el modelo **KNN (Figura 2b)**: las características más importantes fueron 2, 7 y 9. Como este modelo se basa en la proximidad en el espacio de características, estas variables probablemente muestran buena discriminación espacial entre clases.

Extra Trees (Figura 2c): coincide con Random Forest, situando la característica 2 en primer lugar, seguida por 9 y 1. El modelo refuerza la idea de que el patrón angular y la elevación del lápiz contienen información diagnóstica relevante.

En **Naive Bayes (Figura 2d)**: aunque suposiciones de independencia pueden afectar la interpretación, destaca la variable 2, junto con 7 y 1. El impacto más disperso sugiere que este modelo aprovecha múltiples variables con moderado efecto, sin depender fuertemente de una sola.

Finalmente en el modelo de **Árboles de Decisión (Figura 2e)**: se observa un énfasis en las características 1 y 7, aunque con menor consistencia general. Su comportamiento más local puede explicar la variabilidad en la importancia de las variables según los nodos.

A partir del análisis de las ocho gráficas, se contabilizó cuántas veces cada característica apareció en el top 3 de importancia SHAP. La Tabla 4.15 resume esta información:

Cuadro 4.15: Frecuencia de aparición de características en el top 3 de importancia SHAP.

Característica	Frecuencia en Top 3 (de 8 modelos)
2: Azimuth sum of reoccurring data points	7
7: Pressure maximum	6
9: Altitude quantile (q=0.2)	6
1: Pressure change quantiles (qh=0.6, ql=0.0)	4
0: Altitude permutation entropy (d=7, t=1)	3

Las variables más influyentes en el diagnóstico de Parkinson en tareas de escritura fueron:

- **Repetitividad angular (Feature 2)**: Presente en la mayoría de los modelos, lo que destaca la utilidad del análisis de trayectorias angulares.
- **Presión máxima (Feature 7)**: Indicador clave de la intensidad del trazo, asociada con el control motor fino.
- **Altitud en percentil bajo (Feature 9)**: Refleja sutiles variaciones en la altura del lápiz, relacionadas con temblores o pérdida de precisión.

Estos hallazgos refuerzan la hipótesis de que los aspectos biomecánicos de la escritura, como presión, angulación y elevación, contienen señales discriminativas útiles para el diagnóstico temprano y no invasivo de la enfermedad de Parkinson.

4 Resultados

4.5.3 Interacciones de características y análisis de dependencia

Para comprender con mayor profundidad la interacción entre características y su influencia directa en las predicciones de los modelos, se analizaron los *dependence plots* de las tres características más influyentes: la **Feature 2**, **Feature 7**, y **Feature 9**. Estos gráficos representan en el eje *x* el valor real de la característica, mientras que en el eje *y* se visualiza el valor SHAP, es decir, la contribución individual de dicha característica en la predicción. El gradiente de color indica la interacción con una segunda característica, lo cual permite observar relaciones no lineales entre múltiples variables.

En los modelos **Random Forest**, **Extra Trees**, **KNN**, **Naïve Bayes**, **Árboles de Decisión** se grafican por separado las clases HC y PD, lo cual permite analizar el comportamiento específico de cada característica por clase. Esta separación es posible porque el algoritmo de interpretación SHAP calcula contribuciones diferenciadas por clase en modelos no lineales o no probabilísticos. En cambio, en los modelos **SVC**, **Regresión Logística** y **Gradient Boosting** utiliza un único gráfico por característica, donde se combina el efecto de ambas clases. Esta diferencia se debe al tipo de implementación del clasificador y cómo se distribuyen los valores SHAP cuando se hace una predicción en términos de margen de decisión en lugar de probabilidad por clase.

Support Vector Classifier

Feature 2 – Suma de puntos recurrentes del azimut. La Figura 4.4a muestra una tendencia decreciente en los valores SHAP a medida que aumenta el valor de la característica 2. Esto sugiere que mayores repeticiones angulares reducen la probabilidad de clasificar un sujeto como enfermo de EP. El color representa la interacción con la característica 8 (Entropía de muestra en altitud), y se observa que valores más altos de entropía (en rojo) tienden a intensificar esta relación negativa.

Feature 9 – Cuantil del 20% de altitud. En la Figura 4.4b, los valores SHAP disminuyen a medida que disminuye el valor de la característica 9. Esto indica que valores bajos en el cuantil del 20% de altitud favorecen las predicciones hacia la clase 1 (PD), en línea con los patrones observados en modelos previos. La interacción, en este caso, es nuevamente con la presión máxima (Feature 7), cuya intensidad modula el impacto de la altitud baja sobre la predicción.

Feature 7 – Presión máxima. La Figura 4.4c presenta una relación positiva entre el valor de la presión máxima y el valor SHAP, implicando que a mayor presión, aumenta la probabilidad de ser clasificado como paciente con EP. Además, el gradiente de color (Feature 4: Varianza de cambios en altitud) muestra que esta interacción puede amplificar la influencia de la presión, reforzando la hipótesis de que trazos más intensos y variables verticalmente son característicos de sujetos con deterioro motor.

Regresión Logística

En el modelo de **Regresión Logística**, los *dependence plots* muestran relaciones completamente lineales entre cada característica y su contribución SHAP, lo cual es coherente con la naturaleza del modelo. Al ser un clasificador lineal, la relación entre el valor de entrada y su impacto sobre la predicción es directa y proporcional.

4 Resultados

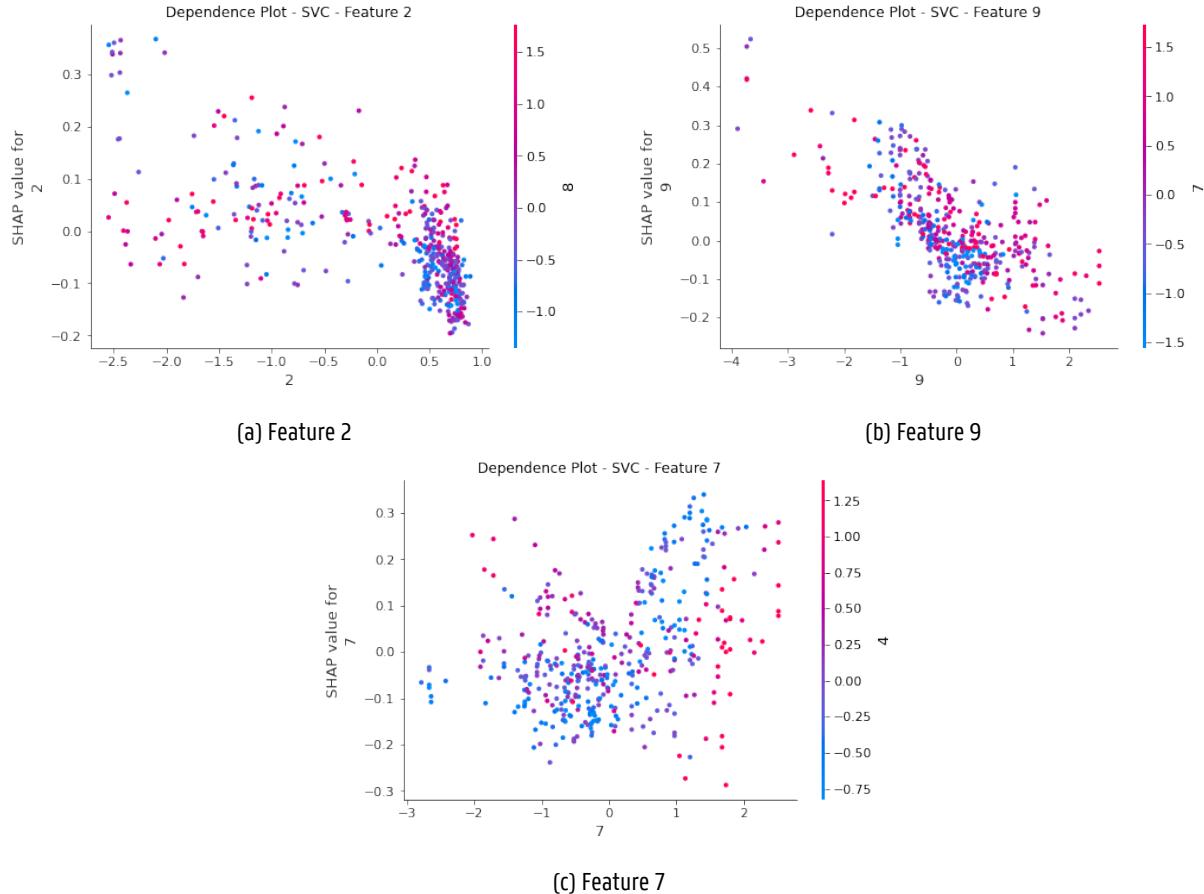


Figura 4.4: Dependence plots de las tres características más relevantes para el modelo SVC. El eje x representa el valor real de la característica, mientras que el eje y muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representa valores bajos y rojo valores altos.

Feature 2 – Suma de puntos recurrentes del azimut. La Figura 4.5a evidencia una relación lineal negativa. A medida que aumenta la repetición de patrones en el azimut, el valor SHAP disminuye, lo que indica que esta característica reduce la probabilidad de pertenecer a la clase PD. Esto sugiere que una mayor regularidad angular está asociada a sujetos sanos.

Feature 9 – Cuantil del 20% de altitud. En la Figura 4.5b, se observa una clara relación negativa: valores bajos de esta característica (elevación mínima del lápiz) incrementan el valor SHAP y, por tanto, empujan la predicción hacia la clase PD. Esta relación sugiere que pacientes con Parkinson tienden a mantener el lápiz más cerca de la superficie durante la escritura.

Feature 7 – Presión máxima. Contrariamente a las dos características anteriores, la Figura 4.5c muestra una relación positiva. A mayor presión, el valor SHAP también se incrementa, indicando una mayor probabilidad de clasificación como PD. Esto respalda la hipótesis de que los pacientes con Parkinson ejercen más presión durante la escritura.

En todos los casos, el gradiente de color refleja la interacción con una segunda característica. Sin embargo, dado que el modelo es lineal, estas interacciones tienen un impacto mínimo en la variación de los valores SHAP, lo que se traduce en una

4 Resultados

coloración relativamente homogénea a lo largo de las rectas.

Gradient Boosting

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.6a, se observa una relación inversa entre los valores de la característica 2 y su valor SHAP: a medida que la recurrencia angular disminuye (valores negativos en el eje x), el valor SHAP aumenta, indicando un mayor empuje hacia la predicción de la clase PD. Esto sugiere que los patrones angulares repetitivos son más frecuentes en sujetos sanos. La interacción con la Feature 7 (presión máxima), representada por el gradiente de color, muestra que niveles altos de presión (tonos rosados) intensifican el impacto de la recurrencia angular en la predicción.

Feature 9 – Cuantil del 20% de altitud. Como se muestra en la Figura 4.6b, los valores bajos de la característica 9 están asociados con valores SHAP altos, lo que indica una mayor contribución hacia la clase PD. Esto refuerza la idea de que una menor elevación del lápiz (cuantiles bajos de altitud) es un rasgo característico en pacientes con Parkinson. La interacción

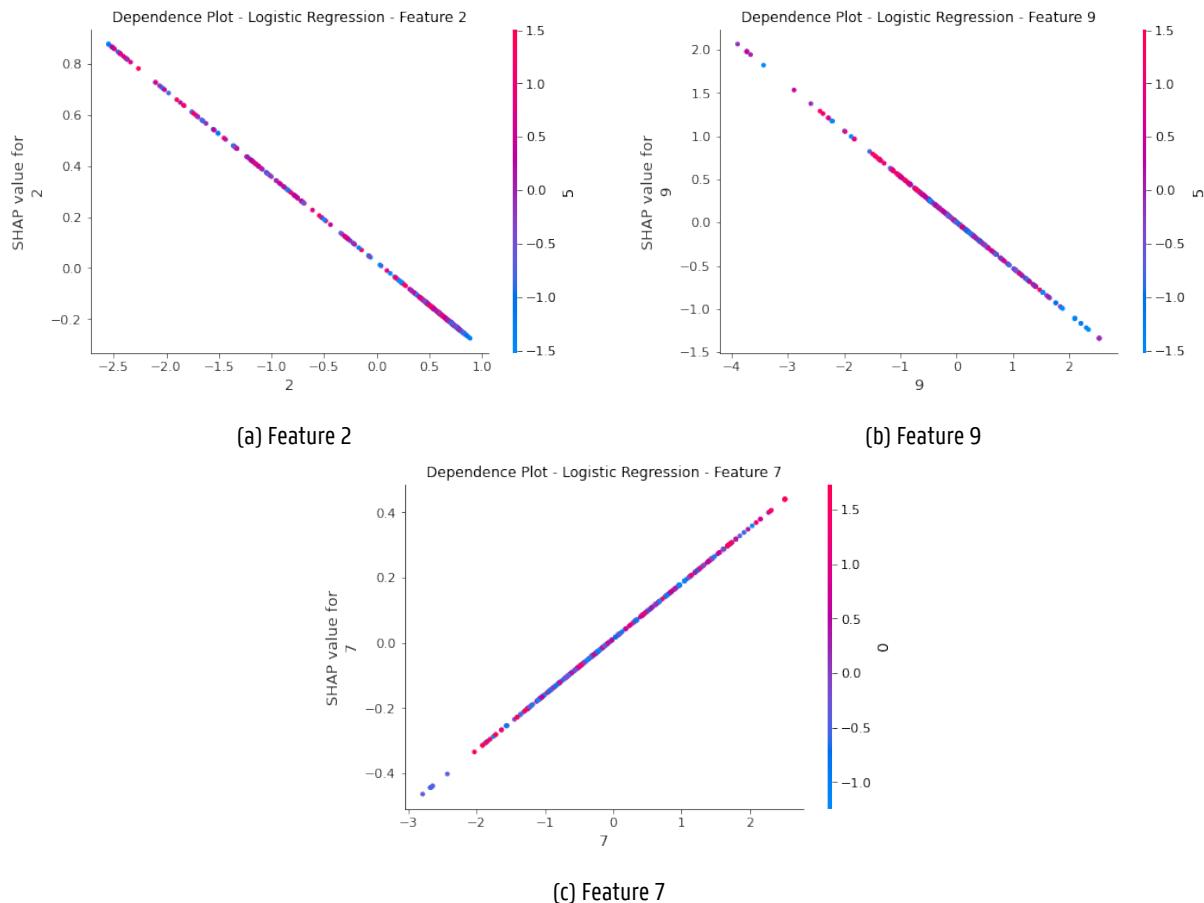


Figura 4.5: Dependence plots de las tres características más relevantes para el modelo de Regresión Logística. El eje x representa el valor real de la característica, mientras que el eje y muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representa valores bajos y rojo valores altos.

4 Resultados

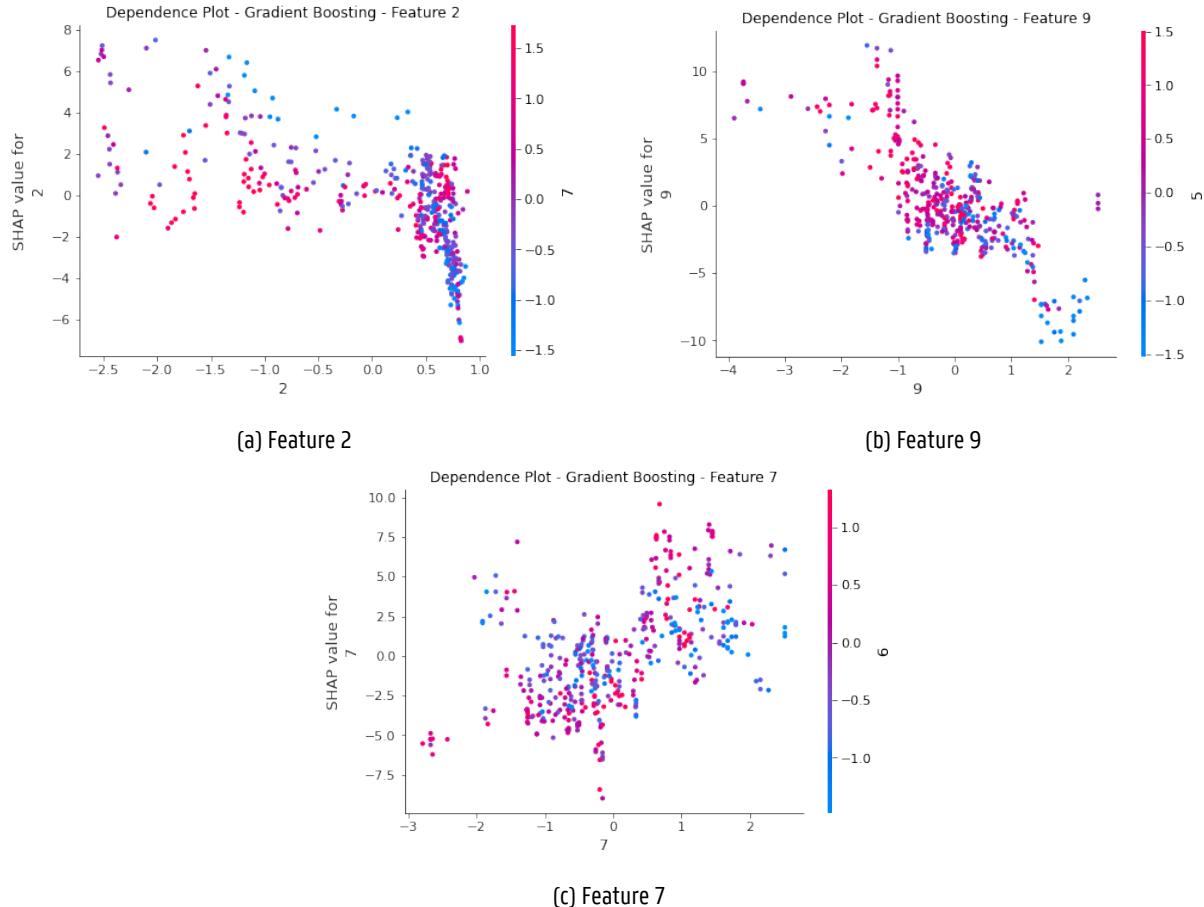


Figura 4.6: Dependence plots de las tres características más relevantes para el modelo Gradient Boosting. El eje x representa el valor real de la característica, mientras que el eje y muestra su valor SHAP. El color indica la interacción con una segunda característica: azul representan valores bajos y rojo valores altos.

con la Feature 5 (número de picos en el azimut) se visualiza en los colores: tonos rojizos (valores altos de picos) amplifican el efecto de esta característica en la predicción.

Feature 7 – Presión máxima. En la Figura 4.6c, se evidencia una correlación positiva entre la presión máxima y su contribución SHAP: a mayor presión, mayor impacto hacia la predicción de la clase PD. Este hallazgo es consistente con estudios que asocian trazos más intensos con deterioro motor. La interacción con la Feature 6 (curtosis del FFT de altitud) es visible en el gradiente de color, donde valores altos (rosados) aumentan el efecto de la presión sobre la decisión del modelo.

Random Forest

El modelo **Random Forest** muestra patrones similares a los observados en Extra Trees, destacando nuevamente la relevancia de las características Feature 2 (repetitividad angular), Feature 7 (presión máxima) y Feature 9 (cuantil de altitud). Este modelo, al ser basado en árboles y tener capacidad de modelar relaciones no lineales, permite observar contribuciones diferenciadas por clase.

4 Resultados

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.11a, para la clase HC, se observa una tendencia creciente en el valor SHAP conforme aumenta la repetitividad angular, indicando que valores altos de esta característica favorecen la clasificación como sujeto sano. En la Figura 4.11b, correspondiente a la clase PD, se evidencia un patrón opuesto: los valores altos en la característica 2 tienden a disminuir su contribución a favor de la clase 1. El color refleja la interacción con la Feature 7 (presión máxima), lo cual sugiere que la fuerza del trazo puede moderar el efecto de la repetición angular.

Feature 9 – Cuantil del 20% de altitud. En la Figura 4.11c, el cuantil bajo de altitud muestra una relación positiva con los valores SHAP para la clase HC, mientras que en la Figura 4.11d, la relación es inversa para pacientes con Parkinson. Este comportamiento sugiere que una altitud mínima más baja está más asociada con la clase PD, posiblemente debido a movimientos verticales reducidos o inestables en la escritura de pacientes.

Feature 7 – Presión máxima. Como se observa en las Figuras 4.11e y 4.11f, la presión máxima tiene un impacto opuesto según la clase. Para sujetos sanos, valores más altos de presión tienden a disminuir el SHAP, mientras que en pacientes con Parkinson, una presión elevada refuerza la predicción hacia la clase PD. Esto puede reflejar el esfuerzo compensatorio o rigidez muscular durante la escritura. Nuevamente, el color representa la interacción con otras variables, particularmente Feature 6 (kurtoisis de la FFT de altitud), que puede amplificar estos efectos.

KNN

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.8a, correspondiente a sujetos sanos, se observa una ligera tendencia creciente: a mayor valor de la característica 2, el valor SHAP tiende a aumentar, lo que sugiere una mayor probabilidad de ser clasificado como clase HC. Por el contrario, en pacientes con Parkinson (Figura 4.8b), se evidencia una tendencia decreciente: valores más altos del azimut recurrente reducen el valor SHAP, es decir, tienen una influencia negativa hacia la clase 1. En ambos casos, el color representa la interacción con la característica 8 (entropía de muestra de altitud), y se observa que altos niveles de irregularidad vertical pueden modular el efecto del azimut en la clasificación.

Feature 7 – Presión máxima. En la Figura 4.8e, para la clase HC, no se identifica una relación clara entre la presión máxima y el valor SHAP, lo que indica una influencia neutra o débil de esta característica en sujetos sanos. Sin embargo, en la Figura 4.8f, correspondiente a sujetos con Parkinson, se evidencia una fuerte relación positiva: a mayor presión ejercida, mayor es el valor SHAP, incrementando la probabilidad de pertenecer a la clase PD. La interacción con la característica 6 (kurtoisis FFT de altitud) sugiere que señales con mayor prominencia o agudeza en el dominio de la frecuencia podrían intensificar el impacto de la presión máxima.

Feature 9 – Cuantil del 20% de altitud. En la Figura 4.8c (clase HC), se aprecia una relación lineal creciente entre el valor del cuantil y el SHAP, lo que indica que una mayor elevación mínima favorece la clasificación como sano. En cambio, para la clase PD (Figura 4.8d), la relación es inversa: valores bajos en el cuantil de altitud aumentan el SHAP, fortaleciendo la predicción hacia Parkinson. El color, que representa la interacción con la característica 8 (entropía de muestra en altitud), sugiere que sujetos con baja altitud y menor complejidad vertical son más propensos a ser clasificados como pacientes con deterioro motor.

4 Resultados

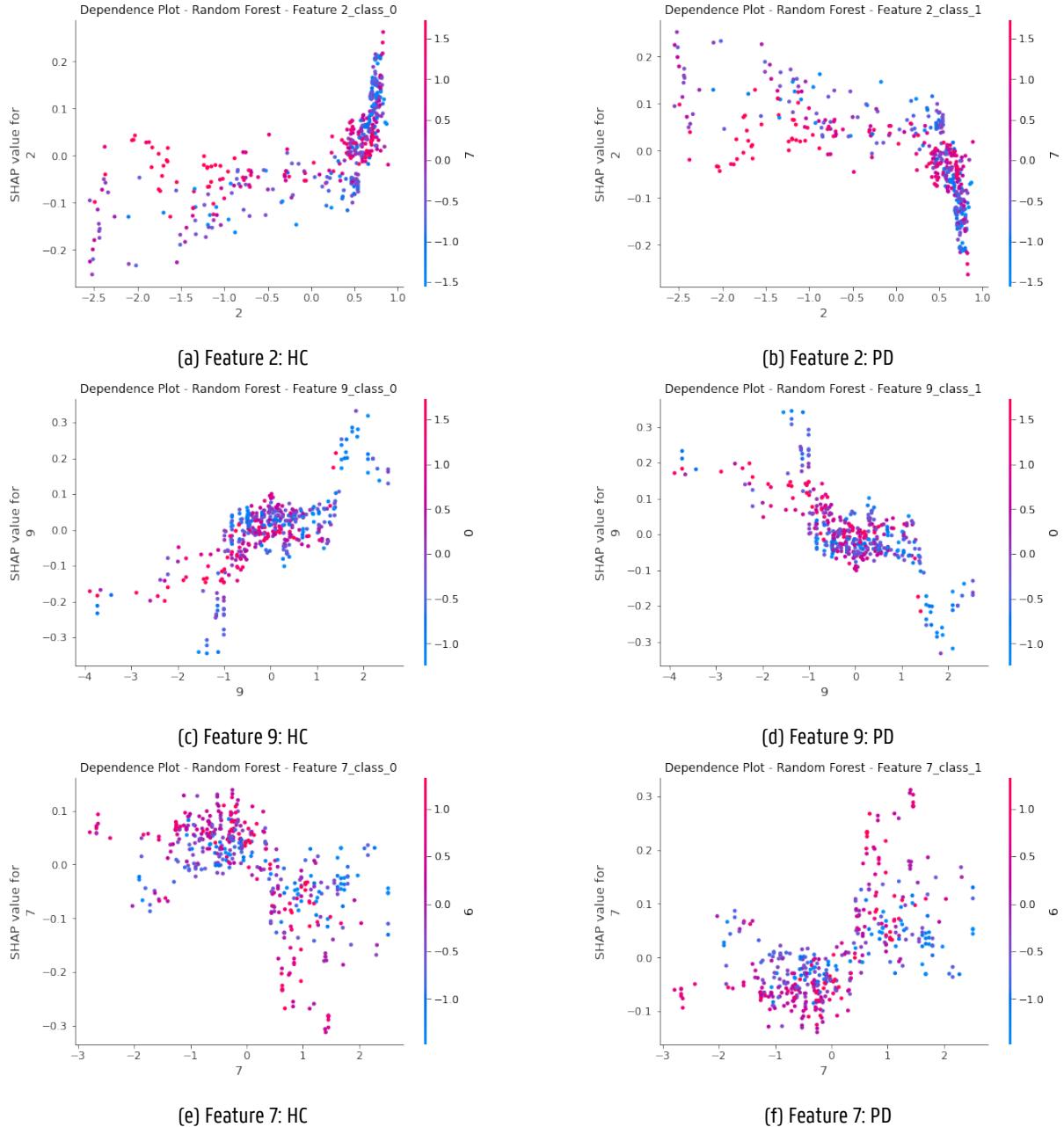


Figura 4.7: Dependence plots de las tres características más importantes del modelo Random Forest. Cada punto representa una observación, el eje x indica el valor real de la característica, el eje y su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos de la característica de interacción, mientras que azul indica valores bajos.

Extra Trees

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.11a (Clase HC), se observa una correlación positiva: a medida que el valor de la característica 2 aumenta, el valor SHAP también lo hace, lo que indica una mayor contribución

4 Resultados

hacia la predicción de clase sana. En contraste, la Figura 4.11b muestra que para pacientes con Parkinson, a mayor recurrencia angular (feature 2), el valor SHAP disminuye, lo que implica una menor probabilidad de ser clasificado como clase 1. El color (Feature 7) actúa como variable de interacción, mostrando que niveles altos de presión pueden acentuar o reducir el impacto del azimut.

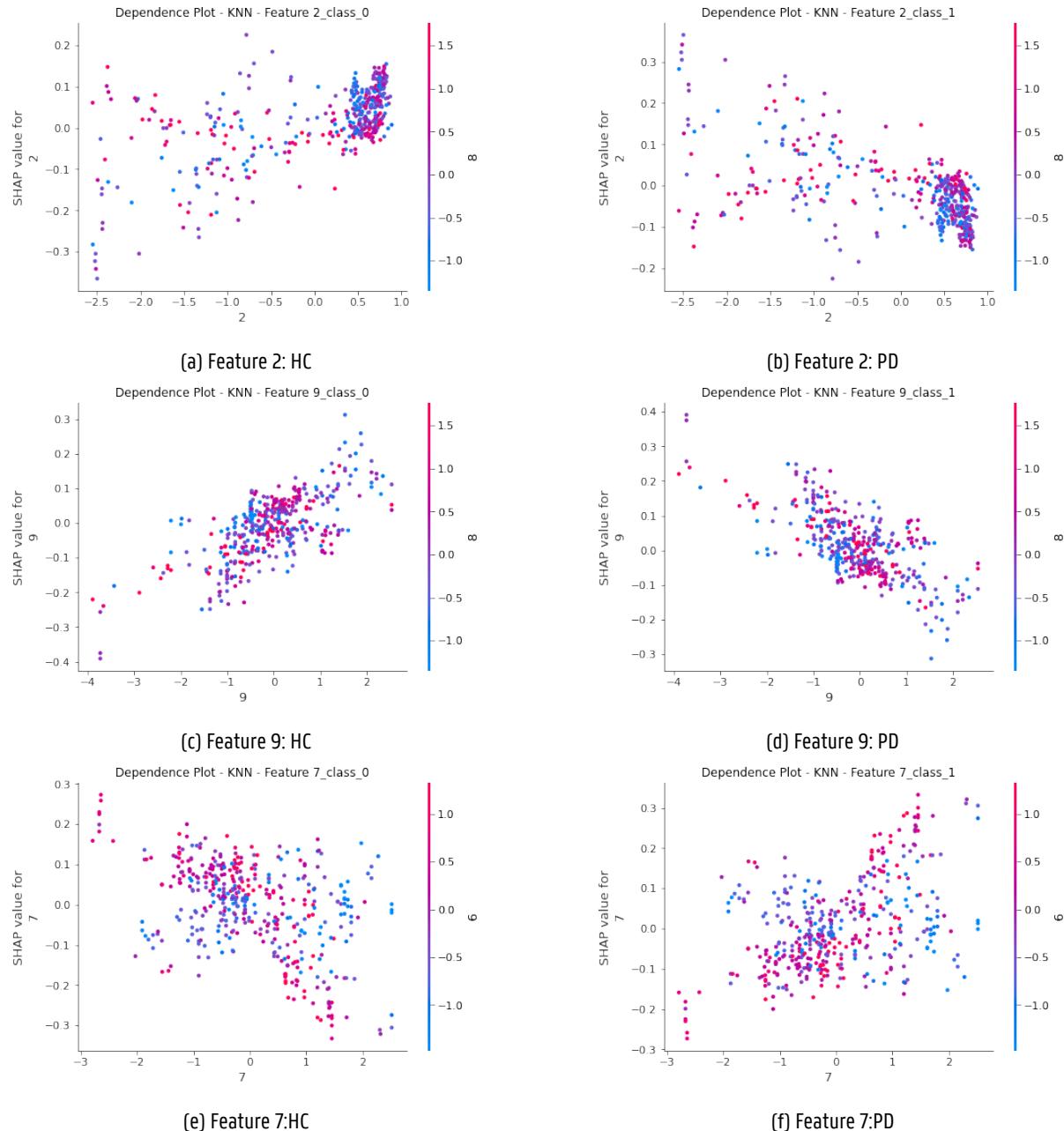


Figura 4.8: Dependence plots de las tres características más relevantes del modelo KNN. Cada punto representa una observación; el eje x indica el valor de la característica, el eje y su valor SHAP. El color representa la interacción con otra característica.

4 Resultados

Feature 9 – Cuantil del 20% de altitud. En sujetos sanos (Figura 4.11c), se evidencia que a mayor altitud baja (valores más negativos), el valor SHAP decrece, favoreciendo la predicción de la clase HC. Para la clase PD (Figura 4.11d), la relación es inversa: menores valores en el cuantil de altitud aumentan el SHAP hacia la clase Parkinson, confirmando que movimientos verticales con valores bajos son más característicos en pacientes con la enfermedad. La variable de interacción es nuevamente la presión máxima (Feature 7), evidenciando patrones complejos entre intensidad del trazo y variabilidad vertical.

Feature 7 – Presión máxima. En la Figura 4.11e, para sujetos sanos, se observa una relación negativa: a mayor presión, menor valor SHAP, lo cual reduce la probabilidad de clasificar como clase HC. Por otro lado, la Figura 4.11f muestra que en pacientes con Parkinson, la presión alta aumenta significativamente el valor SHAP, reforzando su influencia hacia la clase PD. Este comportamiento sugiere que una presión intensa es un marcador distintivo en sujetos con deterioro motor. En ambos casos, la variable de interacción (Feature 6: curtosis FFT de altitud) parece modular el impacto de la presión.

En conjunto, los dependence plots del modelo Extra Trees confirman que las variables relacionadas con patrones angulares repetitivos (Feature 2), elevación baja del lápiz (Feature 9) y fuerza de presión (Feature 7) son indicadores diferenciadores robustos entre sujetos sanos y pacientes con Parkinson. Las variaciones de color en los puntos muestran cómo estas características interactúan entre sí, revelando relaciones no lineales que serían difíciles de detectar con análisis tradicionales.

Naïve Bayes

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.11a (Clase HC), se observa una relación creciente: valores más altos en la característica 2 se asocian con mayores valores SHAP, lo que implica una mayor contribución hacia la predicción de sujetos sanos. En contraste, la Figura 4.11b muestra que para la clase PD, el patrón es inverso: valores altos de la característica disminuyen significativamente la contribución SHAP, lo que favorece la clasificación hacia la clase con enfermedad. Esta dicotomía refuerza la relevancia de los patrones angulares repetitivos del azimut en la diferenciación entre clases.

Feature 9 – Cuantil del 20% de altitud. Para los sujetos sanos (Figura 4.11c), a medida que el cuantil de altitud aumenta, también lo hace el valor SHAP, fortaleciendo la predicción hacia HC. En cambio, en la Figura 4.11d, se observa un patrón decreciente para pacientes con Parkinson, indicando que valores bajos de altitud están fuertemente asociados a la enfermedad. Este comportamiento sugiere que la elevación mínima del lápiz durante el trazo es un rasgo característico de deterioro motor.

Feature 7 – Presión máxima. En la Figura 4.11e (HC), se evidencia una relación negativa: mayor presión reduce el valor SHAP, disminuyendo la probabilidad de predecir clase sana. En cambio, la Figura 4.11f muestra una relación lineal positiva entre la presión y la contribución hacia PD. Este hallazgo refuerza el patrón identificado en otros modelos, donde una presión elevada es un signo distintivo en sujetos con alteraciones motoras. La interacción con la Feature 1 (cuartiles de cambio de presión) modula este efecto, como se evidencia en la gradiente de color.

4 Resultados

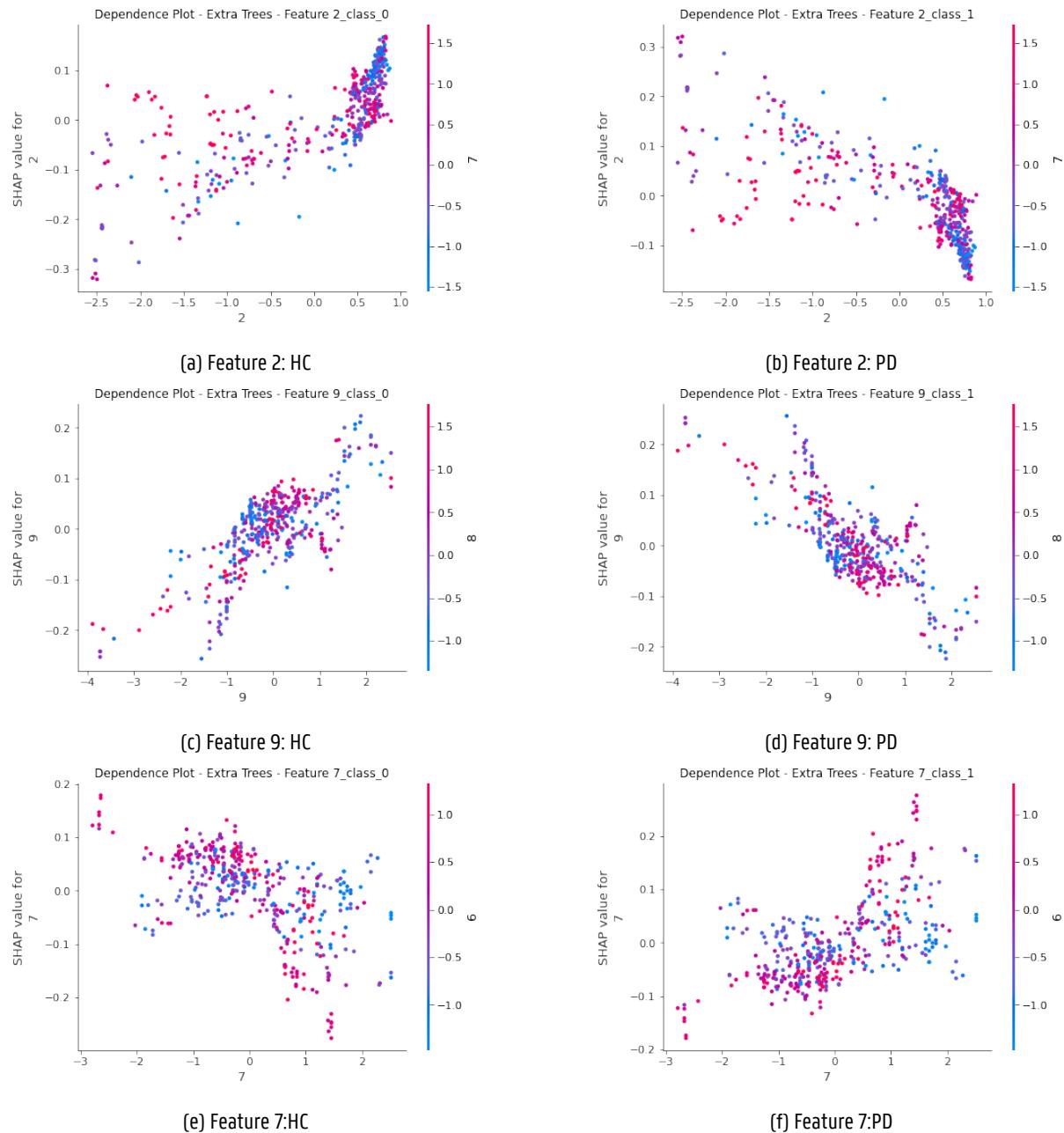


Figura 4.9: Dependence plots de las tres características más importantes del modelo Extra Trees. Cada punto representa una observación, el eje x indica el valor real de la característica, el eje y su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos de la característica de interacción, mientras que azul indica valores bajos.

Decision Tree

El modelo **Decision Tree** presenta relaciones claramente diferenciadas para las clases HC y PD en las tres características más influyentes.

4 Resultados

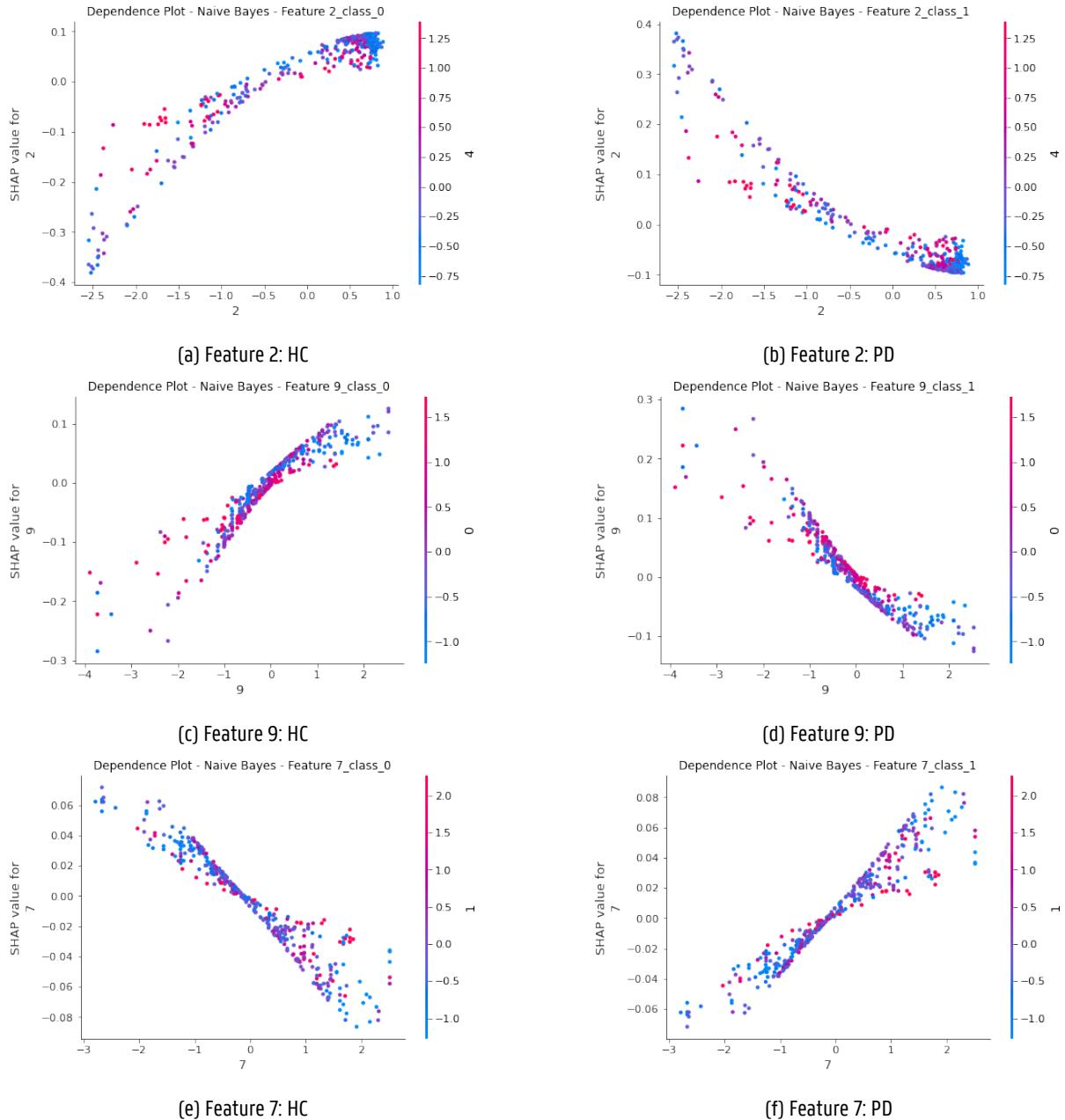


Figura 4.10: Dependence plots de las tres características más importantes del modelo Naïve Bayes. Cada punto representa una observación; el eje x indica el valor real de la característica, el eje y su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos, mientras que azul representa valores bajos.

Feature 2 – Suma de puntos recurrentes del azimut. En la Figura 4.11a, correspondiente a sujetos sanos, se observa una ligera relación creciente: valores más altos de recurrencia angular tienden a generar valores SHAP más elevados. Para la clase PD (Figura 4.11b), el patrón es inverso: a mayor recurrencia, menor contribución SHAP. Este patrón refuerza la idea de que una señal más repetitiva en el ángulo de azimut es característica de la escritura sana.

4 Resultados

Feature 9 – Cuantil del 20% de altitud. En individuos sanos (Figura 4.11c), se identifica una correlación positiva entre el valor del cuantil bajo de altitud y el SHAP, mientras que en pacientes con Parkinson (Figura 4.11d), esta relación se invierte. Los valores bajos en la elevación del lápiz son más característicos de los sujetos con deterioro motor, y esto se ve reflejado en la disminución de SHAP hacia clase PD.

Feature 7 – Presión máxima. El comportamiento de esta característica en la Figura 4.11e para HC es difuso, sin una tendencia clara. Sin embargo, en la Figura 4.11f se muestra que valores altos de presión tienden a incrementar la contribución SHAP hacia la clase PD. Este hallazgo coincide con los observados en otros modelos: una presión elevada es un marcador importante para la predicción de enfermedad.

Los resultados obtenidos a partir de los modelos evaluados —tanto los de mejor rendimiento (Extra Trees, Gradient Boosting, KNN) como los modelos adicionales (SVC, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest)— revelan una consistencia notable en los patrones detectados sobre las tres características más influyentes. Esta convergencia entre modelos fortalece la validez de los hallazgos desde una perspectiva computacional y clínica.

- **Feature 2** (Suma de puntos recurrentes del azimut): En la mayoría de modelos, se observó un patrón dual según la clase. Valores más altos tienden a asociarse con sujetos sanos (HC), mientras que valores más bajos o menos variados son más frecuentes en pacientes con Parkinson (PD). Esto puede reflejar una mayor rigidez o pérdida de fluidez en la orientación angular del lápiz, relacionada con bradicinesia o rigidez motora.
- **Feature 7** (Presión máxima): Esta característica mostró una clara separación entre clases. Los modelos coinciden en que los pacientes con PD tienden a ejercer mayor presión durante la escritura, lo que puede atribuirse al esfuerzo compensatorio, temblores o alteración del control motor fino. Esta variable se destaca como uno de los indicadores más sólidos y consistentes del deterioro motor.
- **Feature 9** (Cuantil bajo de altitud): Los resultados indican que los pacientes con Parkinson tienden a mantener el lápiz más cerca de la superficie, especialmente en los tramos más bajos del trazo (percentil 20%). Esto podría estar relacionado con dificultades para separar el lápiz del papel debido a temblores de reposo, rigidez o lentitud motora. En modelos basados en árboles, esta característica mostró una clara inversión del impacto SHAP entre clases.

En conjunto, los *dependence plots* proporcionan evidencia visual y cuantitativa de que los modelos de aprendizaje automático no solo logran clasificar adecuadamente, sino que lo hacen apoyándose en patrones biomecánicos interpretables y clínicamente relevantes. La orientación angular, la fuerza de presión y la elevación del lápiz emergen como biomarcadores digitales robustos para diferenciar sujetos sanos de pacientes con Parkinson, destacando el potencial de estas señales en aplicaciones de diagnóstico asistido por IA.

4 Resultados

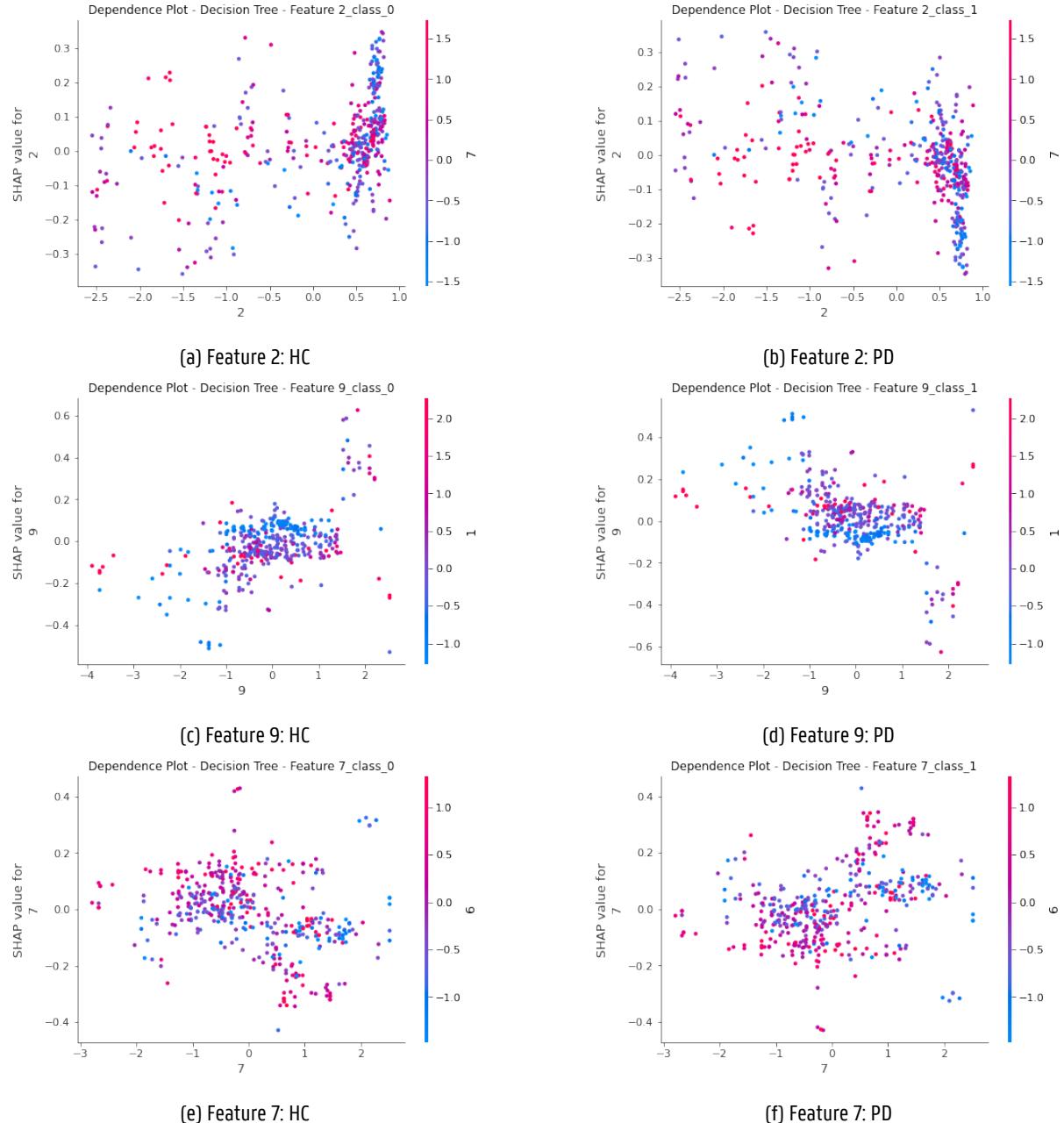


Figura 4.11: Dependence plots de las tres características más importantes del modelo Decision Tree. Cada punto representa una observación; el eje x indica el valor real de la característica, el eje y su contribución SHAP. El color representa la interacción con una segunda característica: mayor intensidad (rojo) sugiere valores altos, mientras que azul representa valores bajos.

5

Conclusiones

Este estudio confirma el potencial de las dinámicas de escritura digital como biomarcadores para evaluar las disfunciones motoras en la EP. Se identificaron la presión, la altitud y el azimut como características clave para distinguir a pacientes con EP de sujetos sanos (HC), aportando información detallada sobre los patrones motores característicos de la enfermedad. Estos hallazgos sugieren que el análisis digital de la escritura podría integrarse en protocolos clínicos como un método no invasivo para monitoreo y diagnóstico temprano del Parkinson.

Con únicamente 85 registros de dos cohortes culturalmente distintas—PaHaW (República Checa) y un conjunto propio de Ecuador—se alcanzaron precisiones de **94,2 %** con *Gradient Boosting* y **93,3 %** con *Extra Trees*. Estos resultados superaron, en la mayoría de las métricas, a las arquitecturas profundas (CNN) y a las redes neuronales de picos (SNN). La combinación de señales cinemáticas relativamente directas (presión, azimut, altitud) y modelos poco profundos mostró que la complejidad algorítmica no siempre es necesaria para capturar rasgos motores distintivos de la EP. No obstante, los modelos profundos conservan un papel importante cuando se disponga de volúmenes de datos mayores o de tareas con mayor variabilidad motora.

Las **CNN** evidenciaron su capacidad para extraer patrones espaciales y morfológicos del trazo cuando las coordenadas *x,y* se transforman en imágenes, si bien no superaron de forma consistente a los modelos tradicionales debido al tamaño muestral limitado. De forma análoga, las **SNN** mostraron ventajas en la codificación temporal, pero requirieron mayor ajuste de hiperparámetros y más datos para explotar completamente su potencial bioinspirado.

El análisis de explicabilidad basado en SHAP corroboró la relevancia clínica de los biomarcadores identificados. La presión máxima, la variabilidad angular del azimut y el cuantil bajo de altitud figuraron entre las variables más influyentes en todos los clasificadores de mejor desempeño, alineándose con la fisiopatología conocida del Parkinson y aumentando la confianza de los especialistas en la salida del modelo.

Finalmente, la fusión de datos de dos contextos lingüísticos distintos demostró que los biomarcadores subyacentes trascienden particularidades de caligrafía; aun así, sólo puede hablarse de *indicios* de robustez intercultural, dado el tamaño limitado de la muestra y la ausencia de otros idiomas o sistemas de escritura.

5 Conclusiones

Limitaciones

El trabajo presenta cuatro restricciones principales: (i) un tamaño muestral relativamente pequeño y con leve desbalance de clases, (ii) la concentración en una única tarea gráfica (espiral de Arquímedes) sin seguimiento longitudinal y (iii) la composición demográfica desigual de las muestras, con posibles sesgos introducidos por diferencias educativas, etarias o culturales entre las cohortes de Ecuador y la República Checa, que podrían haber influido en los patrones de escritura observados.

Trabajos futuros

Se propone, en primer lugar, ampliar el estudio a un conjunto multilingüe y longitudinal que permita validar la generalización intercultural y estimar progresión clínica. En segundo lugar, se explorará la integración de biomarcadores complementarios (voz, marcha, oculomotricidad) para crear un sistema multimodal holístico. En tercer lugar, la comparación sistemática de técnicas de explicabilidad (LIME, Grad-CAM, Integrated Gradients, DeepSHAP, gradient-based spike metrics) enriquecerá la transparencia del modelo. Por último, se prevé la implementación de SNN optimizadas en hardware neuromórfico de baja potencia, así como la extensión del enfoque a otras patologías neurológicas (temblor esencial, distonía, Alzheimer).

Referencias

- [1] G. . P. D. Collaborators, "Global, regional, and national burden of parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet Neurology*, vol. 19, no. 11, pp. 939–953, 2020.
- [2] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review," *Applied Sciences*, vol. 9, no. 21, p. 4666, Nov. 2019. [Online]. Available: <http://dx.doi.org/10.3390/app9214666>
- [3] C. Rodriguez and M. Perez, "Handwriting as a diagnostic and monitoring tool for neurological disorders," *Journal of Neuromotor Research*, vol. 12, no. 3, pp. 123–135, 2020.
- [4] D. Impedovo and G. Pirlo, "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective," *IEEE Reviews in Biomedical Engineering*, vol. 12, p. 209–220, 2019.
- [5] H. Tai and J. Lee, "Micrographia as a diagnostic tool for early-stage parkinson's disease," *Neurological Sciences*, vol. 42, pp. 3215–3224, 2021.
- [6] M. P. Caligiuri and L. A. Mohammed, *The Neuroscience of Handwriting*. CRC Press, Feb. 2012. [Online]. Available: <http://dx.doi.org/10.1201/b11703>
- [7] A. Di Febbo, "Design and control of a novel hybrid soft-rigid manipulator for minimally invasive surgery," PhD thesis, Politecnico di Milano, Milano, Italy, July 2023, available at https://www.politesi.polimi.it/bitstream/10589/176386/1/PhdThesis_DiFebbo.pdf.
- [8] J. Danna, J.-L. Velay, A. Eusebio, L. Véron-Delor, T. Witjas, J.-P. Azulay, and S. Pinto, "Digitalized spiral drawing in parkinson's disease: A tool for evaluating beyond the written trace," *Human Movement Science*, vol. 65, p. 80–88, Jun. 2019. [Online]. Available: <http://dx.doi.org/10.1016/j.humov.2018.08.003>
- [9] M. Kamble, P. Shrivastava, and M. Jain, "Digitized spiral drawing classification for parkinson's disease diagnosis," *Measurement: Sensors*, vol. 16, p. 100047, Aug. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.measen.2021.100047>
- [10] T. D'Alessandro, "Handwriting analysis for the development of a system to support the diagnosis of neurodegenerative diseases," Tesi di dottorato, Università degli Studi di Cassino e del Lazio Meridionale, 2024, acceso el 12 de mayo de 2025. [Online]. Available: <https://tesidottorato.depositolegale.it/handle/20.500.14242/168538>
- [11] A. Ammour, I. Aouraghe, G. Khaissidi, M. Mrabti, G. Aboulem, and F. Belahsen, "Online arabic and french handwriting of parkinson's disease: The impact of segmentation techniques on the classification results," *Biomedical Signal Processing and Control*, vol. 66, p. 102429, Apr. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.bspc.2021.102429>
- [12] C. Ladbury, R. Zarinshenas, H. Semwal, A. Tam, N. Vaidehi, A. S. Rodin, A. Liu, S. Glaser, R. Salgia, and A. Amini, "Utilization of model-agnostic explainable artificial intelligence frameworks in oncology: a narrative review," *Translational Cancer Research*, vol. 11, no. 10, p. 3853–3868, Oct. 2022. [Online]. Available: <http://dx.doi.org/10.21037/tcr-22-1626>
- [13] A. Zemmar, A. Bennour, T. Mekhaznia, F. Ghabban, and M. Al-Sarem, "Unveiling parkinson's: Handwriting symptoms with explainable and interpretable cnn model," *International Journal of Pattern Recognition and Artificial Intelligence*, 2024.

5 Referencias

- [14] D. Impedovo, G. Pirlo, and G. Vessio, "Dynamic handwriting analysis for supporting earlier parkinson's disease diagnosis," *Information*, vol. 9, no. 10, p. 247, 2018.
- [15] P. Drotar, J. Mekyska, and Z. Smekal, "Evaluación de la cinemática y presión de la escritura para el diagnóstico diferencial de la ep," *Computer Methods and Programs in Biomedicine*, pp. 405–411, 2016.
- [16] P. Zham, S. P. Arjunan, S. Raghav, and D. K. Kumar, "Efficacy of guided spiral drawing in the classification of parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1648–1652, 2017.
- [17] R. Karlsdottir and E. Stefansson, "Development of handwriting in the early school years: A longitudinal study," *Perceptual and Motor Skills*, vol. 94, no. 3, pp. 795–806, 2002.
- [18] V. W. Berninger and T. L. Richards, "Handwriting and spelling instruction for students with learning disabilities: A review of the literature," *Learning Disabilities Research Practice*, vol. 30, no. 1, pp. 44–57, 2015.
- [19] A. El Idrissi, "Analyse des performances des réseaux de capteurs sans fil dans les environnements urbains," Thèse de doctorat, Université Mohammed V de Rabat, 2022, consulté le 12 mai 2025. [Online]. Available: <https://toubkal.imist.ma/handle/123456789/16383>
- [20] G. D. Cascarano, C. Loconsole, A. Brunetti, A. Lattarulo, D. Buongiorno, G. Losavio, E. Di Sciascio, and V. Bevilacqua, "Biometric handwriting analysis to support parkinson's disease assessment and grading," *BMC Medical Informatics and Decision Making*, vol. 19, no. Suppl 9, p. 252, 2019.
- [21] C. D. Rios-Urrego, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera, and J. R. Orozco-Arroyave, "Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features," *arXiv preprint arXiv:2002.05411*, 2020. [Online]. Available: <https://arxiv.org/abs/2002.05411>
- [22] A. Parziale, R. Senatore, A. Della Cioppa, and A. Marcelli, "Cartesian genetic programming for diagnosis of parkinson disease through handwriting analysis: Performance vs. interpretability issues," *Artificial Intelligence in Medicine*, vol. 111, p. 101984, 2021.
- [23] Y. Zhang, Z. Zeng, M. S. Mirian, K. Yen, K. W. Park, M. Doo, J. Ji, Z. Shen, and M. J. McKeown, "Investigating the efficacy and importance of mobile-based assessments for parkinson's disease: uncovering the potential of novel digital tests," *Scientific Reports*, vol. 14, no. 1, Mar. 2024. [Online]. Available: <http://dx.doi.org/10.1038/s41598-024-55077-7>
- [24] D. D. Febbo, "Novel ehealth methodologies to support the detection of age related decline," Tesi di dottorato, Politecnico di Milano, 2021, accedido el 12 de mayo de 2025. [Online]. Available: <https://www.politesi.polimi.it/handle/10589/176386>
- [25] J. K. Eshraghian, M. Ward, E. O. Neftci, X. Wang, G. Lenz, G. Dwivedi, M. Bennamoun, D. S. Jeong, and W. D. Lu, "Training spiking neural networks using lessons from deep learning," *Proceedings of the IEEE*, vol. 111, no. 9, p. 1016–1054, Sep. 2023. [Online]. Available: <http://dx.doi.org/10.1109/jproc.2023.3308088>
- [26] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y.-D. Liu, "Reliable parkinson's disease detection by analyzing handwritten drawings: Construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116 480–116 489, 2019.

5 Referencias

- [27] C. Poon, N. Gorji, M. Latt, K. Tsoi, B. Choi, C. Loy, and S. Poon, "Derivation and analysis of dynamic handwriting features as clinical markers of parkinson's disease," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, ser. HICSS. Hawaii International Conference on System Sciences, 2019. [Online]. Available: <http://dx.doi.org/10.24251/hicss.2019.450>
- [28] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of parkinson's disease," *Future Generation Computer Systems*, vol. 117, pp. 234–244, 2021.
- [29] S. Ali, A. Hashmi, A. Hamza, U. Hayat, and H. Younis, "Dynamic and static handwriting assessment in parkinson's disease: A synergistic approach with c-bi-gru and vgg19," *Journal of Computing Theories and Applications*, vol. 1, no. 2, p. 151–162, Dec. 2023.
- [30] P. Das, S. Nanda, G. Panda, S. Dash, A. Ksibi, S. Alsenan, W. Bouchelligua, and S. Mallik, "A robust parkinson's disease detection model based on time-varying synaptic efficacy function in spiking neural network," *BMC Neurology*, vol. 24, no. 1, p. 492, 2024.
- [31] I. AbouHassan, N. K. Kasabov, V. Jagtap, and P. Kulkarni, "Spiking neural networks for predictive and explainable modelling of multimodal streaming data with a case study on financial time series and online news," *Scientific Reports*, vol. 13, p. 18367, 2023.
- [32] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders—a case study on parkinson's disease," *Pattern Recognition Letters*, vol. 121, p. 19–27, Apr. 2019. [Online]. Available: <http://dx.doi.org/10.1016/j.patrec.2018.04.008>
- [33] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Wiley, 2013.
- [34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [35] J. R. Quinlan, "Improved use of continuous attributes in c4.5," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996.
- [36] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [37] I. Rish, "An empirical study of the naive bayes classifier," *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, pp. 41–46, 2001.
- [38] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [39] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, pp. 3–42, 2006.
- [40] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [41] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh) - a python package," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

5 Referencias

- [42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, *Scikit-learn: Machine Learning in Python*. MIT Press, 2011, vol. 12. [Online]. Available: <http://jmlr.org/papers/v12/pedregosa11a.html>
- [43] J. Luna, E. Peláez, F. R. Loayza, A. Saravia, I. Obeso, and V. Asanza, "Anatomical patterns recognition of impulse control disorders of parkinsonian patients using deep learning of mri structural images," in *2022 IEEE ANDESCON*, 2022, pp. 1–6.
- [44] Y. Aghzal and M. Louri, "Parkinson's disease detection using hog features and cnns," *Computers in Biology and Medicine*, vol. 118, 2020.
- [45] A. Thomas and M. Kumar, "Handwriting analysis for parkinson's disease diagnosis using deep learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2053–2062, 2017.
- [46] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, Sep. 2022. [Online]. Available: <http://dx.doi.org/10.3390/app12188972>
- [47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [48] V.-T. Hoang and K.-H. Jo, "Practical analysis on architecture of efficientnet," in *2021 14th International Conference on Human System Interaction (HSI)*, 2021, pp. 1–4.
- [49] D. Singh, V. Kumar, and M. Kaur, "Densely connected convolutional networks-based covid-19 screening model," *Applied Intelligence*, vol. 51, no. 5, p. 3044–3051, Feb. 2021. [Online]. Available: <http://dx.doi.org/10.1007/s10489-020-02149-6>
- [50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [51] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [52] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15.
- [53] D. M. W. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *CoRR*, vol. abs/2010.16061, 2020. [Online]. Available: <https://arxiv.org/abs/2010.16061>
- [54] D. G. Altman and J. M. Bland, "Diagnostic tests 1: sensitivity and specificity," *BMJ*, vol. 308, no. 6943, p. 1552, 1994.
- [55] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*, 2017, pp. 4765–4774.
- [56] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017. [Online]. Available: <https://arxiv.org/abs/1711.05101>
- [57] Z. Wang, Y. Huang, Y. Zhu, B. Xu, and L. Chen, "Dynamic layer-span connecting spiking neural networks with backpropagation training," *Complex amp; Intelligent Systems*, vol. 10, no. 2, p. 1937–1952, Oct. 2023.

5 Referencias

- [58] J. Ding, B. Dong, F. Heide, Y. Ding, Y. Zhou, B. Yin, and X. Yang, "Biologically inspired dynamic thresholds for spiking neural networks," 2022.
- [59] L. Herranz-Celotti and J. Rouat, "Stabilizing spiking neuron training," 2022.
- [60] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," 2013.
- [61] F. Wang, Y. Liang, and Q.-W. Wang, "Interpretable machine learning-driven biomarker identification and validation for alzheimer's disease," *Scientific Reports*, vol. 14, no. 1, Dec. 2024. [Online]. Available: <http://dx.doi.org/10.1038/s41598-024-80401-6>
- [62] O. O. Bifarin, "Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification," *PLOS ONE*, vol. 18, no. 5, p. e0284315, May 2023. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0284315>
- [63] H. Chen, S. M. Lundberg, and S.-I. Lee, "Explaining a series of models by propagating shapley values," *Nature Communications*, vol. 13, no. 1, Aug. 2022. [Online]. Available: <http://dx.doi.org/10.1038/s41467-022-31384-3>
- [64] Z. Qi, S. Khorram, and F. Li, "Visualizing deep networks by optimizing with integrated gradients." in *AAAI*, vol. 34, 2020, pp. 11 890–11 898.
- [65] A. Bitar, R. Rosales, and M. Paulitsch, "Gradient-based feature-attribution explainability methods for spiking neural networks," *Frontiers in Neuroscience*, vol. 17, Sep. 2023. [Online]. Available: <http://dx.doi.org/10.3389/fnins.2023.1153999>
- [66] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [67] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3026–3037, 2014.
- [68] D. Impedovo and G. Pirlo, "Handwriting analysis for diagnosis and prognosis of parkinson's disease: A review," *Pattern Recognition Letters*, vol. 128, pp. 110–117, 2019.
- [69] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh)," *Neurocomputing*, vol. 307, pp. 72–77, 2020.
- [70] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.
- [71] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

Apéndices

Anexo A: Visualización de la importancia de características

Este anexo presenta las visualizaciones de explicabilidad generadas mediante SHAP para los modelos de aprendizaje automático clásicos utilizados en el estudio. Las figuras muestran la importancia relativa de las características extraídas, diferenciando el impacto por clase: sujetos sanos (clase 0) y pacientes con enfermedad de Parkinson (clase 1). Se incluyen tanto gráficos de tipo *beeswarm* (puntos) como gráficos de barras agrupadas por clase.

Los números corresponden a las siguientes características: **0:** Altitude Permutation entropy dimension 7 tau 1, **1:** Pressure change quantiles f agg "mean" isabs True qh 0.6 ql 0.0, **2:** Azimuth sum of reoccurring data points, **3:** Pressure change quantiles f agg "mean" isabs True qh 0.4 ql 0.2, **4:** Altitude change quantiles f agg "var" isabs True qh 1.0 ql 0.2, **5:** Azimuth number cwt peaks n 1, **6:** Altitude FFT aggregated aggtpe "kurtosis", **7:** Pressure maximum, **8:** Altitude sample entropy, **9:** Altitude quantile q 0.2.

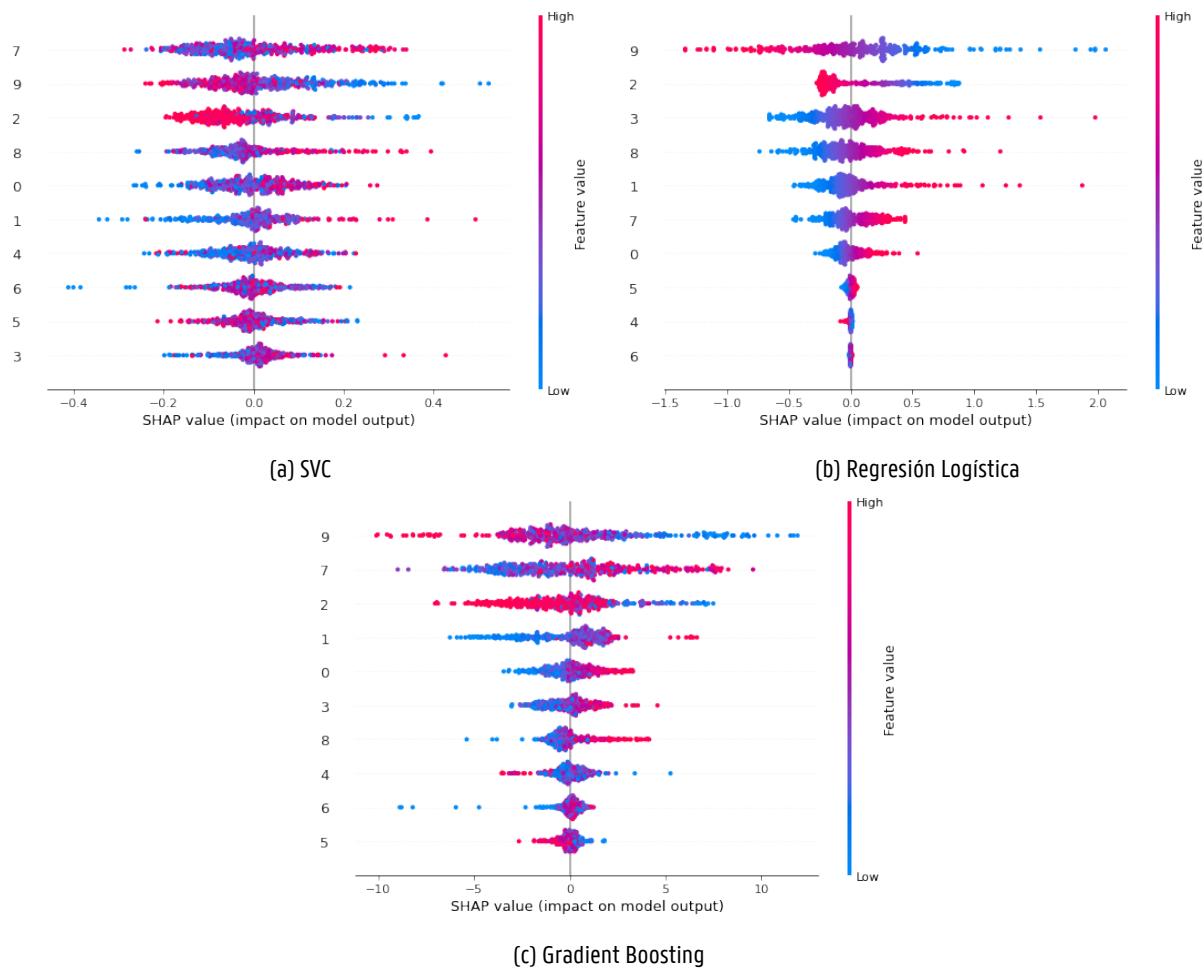


Figura 1: Importancia de características determinada por los modelos SVC, Regresión Logística y Gradient Boosting.

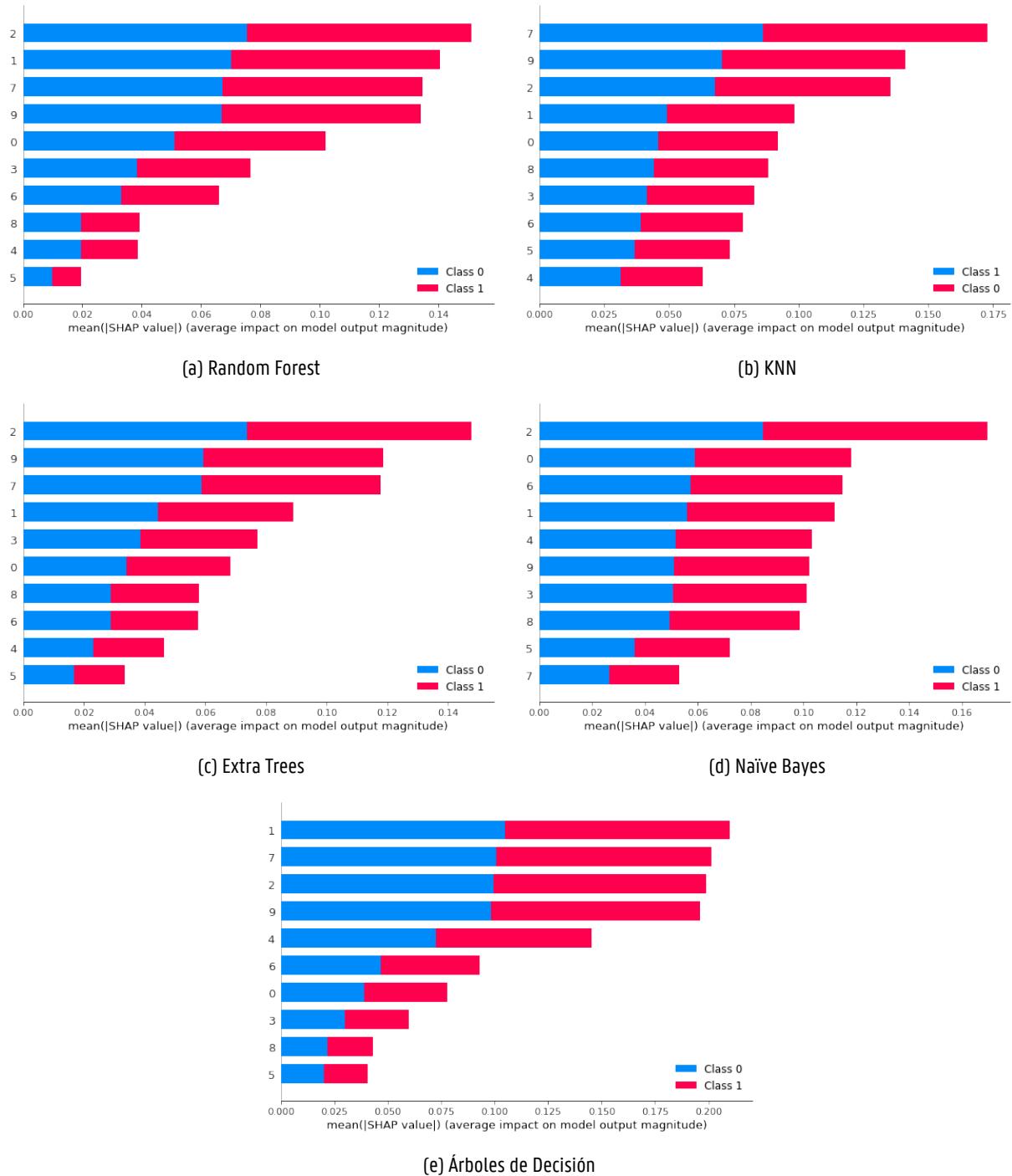


Figura 2: Importancia de características determinada por los modelos Random Forest, KNN, Extra Trees, Naïve Bayes y Árboles de Decisión.