



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO
CARRERA DE BIOMEDICINA**

**COMPARACIÓN DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO EN LA
IDENTIFICACIÓN DE LA ENFERMEDAD DE PARKINSON EN PERSONAS
MAYORES DE 60 AÑOS**

Trabajo de titulación previo a la obtención del título de:
Ingeniero Biomédico

AUTOR: DAVID FERNANDO VEGA SOTO

TUTOR: MAYERLY ELVIRA SÁENZ CASALLAS

Quito-Ecuador

2024

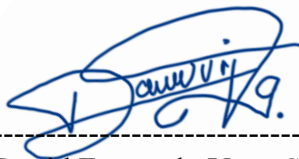
**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, David Fernando Vega Soto con documento de identificación N° 1750833442 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 07 de agosto del año 2024

Atentamente,



David Fernando Vega Soto
1750833442

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, David Fernando Vega Soto con documento de identificación No. 1750833442, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Artículo Académico: Comparación de Técnicas de Aprendizaje Automático en la Identificación de la Enfermedad de Parkinson en Personas Mayores de 60 Años, el cual ha sido desarrollado para optar por el título de: Ingeniero Biomédico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 07 de agosto del año 2024

Atentamente,



David Fernando Vega Soto

1750833442

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Mayerly Elvira Sáenz Casallas con documento de identificación N° 1758414393, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: Comparación de Técnicas de Aprendizaje Automático en la Identificación de la Enfermedad de Parkinson en Personas Mayores de 60 Años, realizado por David Fernando Vega Soto con documento de identificación N° 1750833442, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 07 de agosto del año 2024

Atentamente,



Mayerly Elvira Sáenz Casallas

1758414393

COMPARACIÓN DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO EN LA IDENTIFICACIÓN DE LA ENFERMEDAD DE PARKINSON EN PERSONAS MAYORES DE 60 AÑOS

COMPARISON OF MACHINE LEARNING TECHNIQUES IN THE IDENTIFICATION OF PARKINSON'S DISEASE IN PEOPLE OVER 60 YEARS OF AGE

David Fernando Vega Soto ¹ , Mayerly Elvira Sáenz Casallas ¹ 

Universidad Politécnica Salesiana Quito, Ecuador

Resumen

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo que afecta principalmente a personas mayores de 60 años. La detección precoz de esta enfermedad permite ofrecer tratamientos más eficaces y optimizar la calidad de vida de los pacientes. En esta investigación, se compara la eficacia de tres técnicas de aprendizaje automático: Máquina de Vectores de Soporte (SVM), Bosques Aleatorios (RF) y Bayesiano Ingenuo (NB) en la identificación de la EP. Para la evaluación se utilizan grabaciones de voz, datos de marcha y patrones de escritura, con el objetivo de determinar la precisión y consistencia de estos algoritmos. Los resultados muestran que RF y SVM son los más efectivos, con una precisión media de 94.40% y 92.30%, respectivamente. SVM destacó en la clasificación de datos de marcha, mientras que RF mostró una gran robustez en múltiples métricas, incluyendo especificidad y sensibilidad. Aunque NB es útil, presentó limitaciones con una precisión media de 70.3%, afectada por suposiciones simplificadas sobre la independencia de características.

Palabras clave: Enfermedad de Parkinson, Diagnóstico temprano, Aprendizaje automático, Datos biomédicos, Preprocesamiento de datos

Abstract

Parkinson's disease (PD) is a neurodegenerative disorder that primarily affects individuals over the age of 60. Early detection of this disease allows for the provision of more effective treatments and optimization of patients' quality of life. In this research, the effectiveness of three machine learning techniques—Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB)—in identifying PD is compared. Voice recordings, gait data, and handwriting patterns are used for evaluation, aiming to determine the accuracy and consistency of these algorithms. The results show that RF and SVM are the most effective, with an average accuracy of 94.40% and 92.30%, respectively. SVM excelled in classifying gait data, while RF demonstrated great robustness across multiple metrics, including specificity and sensitivity. Although NB is useful, it showed limitations with an average accuracy of 70.3%, affected by simplified assumptions about feature independence.

Keywords: Parkinson's disease, Early diagnosis, Machine learning, Biomedical data, Data preprocessing

1. Introducción

En los últimos años, los desarrollos en medicina y tecnología han incrementado el bienestar de la población global. La expectativa de vida ha aumentado, llevando consigo un mayor interés en las enfermedades asociadas con el envejecimiento. Entre estas enfermedades, las neurodegenerativas destacan por su impacto en la calidad de vida de los pacientes y sus familias [1]. Dentro de este grupo, la enfermedad de Parkinson (EP) es la segunda causa más común de enfermedad neurodegenerativa crónica progresiva [1], afectando a cerca de 10 millones de personas en todo el mundo. Aproximadamente el 1% de las personas mayores de 60 años y el 4-5% de las personas mayores de 85 años padecen esta enfermedad [2]. Globalmente, la prevalencia de la EP se estima entre 100 y 300 casos por cada 100,000 personas [1].

La EP es una enfermedad neurodegenerativa de carácter crónico y progresivo que impacta el nivel de vida de los pacientes. Se caracteriza por una variedad de síntomas, tanto motores como no motores. Los síntomas motores incluyen temblores y rigidez, mientras que los síntomas no motores abarcan problemas de sueño y cambios en el estado de ánimo. El desarrollo patológico de la EP puede iniciarse muchos años o incluso décadas antes de que se manifiesten los primeros síntomas [3], lo que resalta la complejidad de esta enfermedad. La EP ocurre cuando las estructuras cerebrales comienzan a deteriorarse y finalmente mueren, resultando en una disminución de la producción de dopamina, una molécula esencial para la coordinación y generación de movimientos musculares [1].

El diagnóstico de la EP se enfoca en el síndrome motor o parkinsonismo, que incluye síntomas como temblores, rigidez muscular, lentitud de movimientos y problemas de equilibrio [2]. Los criterios de la Sociedad Internacional de Trastornos del Movimiento de 2015, empleados por los médicos para diagnosticar la enfermedad, abarcan criterios centrales, señales de alerta, criterios de exclusión absoluta y criterios de apoyo. Además de los síntomas motores, estos criterios reconocen la importancia de los síntomas no motores y las pruebas paraclínicas en el diagnóstico de la EP [3].

El aprendizaje automático es una herramienta útil en el campo de la medicina, especialmente en la detección y diagnóstico de enfermedades neurodegenerativas como la EP [1, 3, 4]. Se ha explorado su aplicación en varias facetas de la EP, desde la detección temprana hasta el seguimiento del progreso de las intervenciones terapéuticas.

En esta investigación, se comparan tres técnicas de aprendizaje automático para evaluar su eficacia en la identificación de la EP en personas mayores de 60 años. La detección precoz en este grupo facilita la implementación de intervenciones que pueden disminuir la progresión de la enfermedad y mejorar el

nivel de vida de los pacientes. Los algoritmos evaluados son: Máquina de Vectores de Soporte (SVM), Bosques Aleatorios (RF) y Bayesiano Ingenuo (NB).

Se revisan diversos estudios que han utilizado diferentes fuentes de datos y métodos para identificar la EP en personas mayores de 60 años mediante técnicas de aprendizaje automático. Estos trabajos han proporcionado una base de evidencia sobre la utilidad y precisión de estas técnicas en el contexto del diagnóstico de la EP.

En estudios como los de [5–7], para la evaluación de los pacientes se obtuvieron las grabaciones de voz mediante entrevistas estructuradas donde los participantes leyeron pasajes de texto específicos y realizaron ejercicios vocales. De estas grabaciones, los autores extrajeron características como la frecuencia fundamental, la intensidad y medidas de calidad vocal como el jitter y el shimmer. En los trabajos de [8, 9], los patrones de escritura se capturaron usando tabletas digitalizadoras que registraron la presión, la velocidad del trazo y la aceleración durante la escritura de frases y palabras específicas. Las investigaciones de [10–12] recolectaron datos de la marcha con sensores de movimiento colocados en las extremidades inferiores de los participantes, capturando parámetros como la velocidad, la cadencia y la estabilidad del paso. Además, los estudios de [13, 14] obtuvieron neuroimágenes mediante resonancia magnética funcional (fMRI) y tomografía por emisión de positrones (PET), proporcionando información detallada sobre la estructura y función cerebral. Finalmente, trabajos como los de [15] recopilaban características clínicas a través de cuestionarios de evaluación de síntomas y exámenes neurológicos, incluyendo datos sobre temblores, rigidez, bradicinesia y otros síntomas motores y no motores. En los estudios revisados, tales como: [1, 16–18], los autores realizaron el preprocesamiento de los datos para asegurar la calidad y consistencia. Este proceso incluyó la limpieza de datos para eliminar registros incompletos y datos anómalos, la normalización para garantizar que todas las variables tuvieran un rango comparable, y la extracción de características para reducir la dimensionalidad de los datos y seleccionar las variables más relevantes. Además, el trabajo de [11] utiliza métodos de reducción de dimensionalidad como el análisis de varianza (ANOVA) y de componentes principales (PCA).

En los estudios mencionados anteriormente, se implementaron tres algoritmos de aprendizaje automático para la categorización de pacientes con EP y sujetos sanos. En el trabajo de [13] las Máquinas de Vectores de Soporte (SVM) fueron entrenadas utilizando diferentes kernels (lineal, polinómico, RBF y sigmoide) y se ajustaron los parámetros del margen suave y la penalización (C) para optimizar la precisión del modelo. En el trabajo de [19] el Clasificador Bayesiano Ingenuo (NB) asumió una distribución gaussiana para

las características continuas y calculó las probabilidades condicionales para cada clase. Por último, el trabajo de [17] muestra que los Bosques Aleatorios (RF) se entrenaron con un conjunto de 100 árboles de decisión, evaluando la importancia de las características y ajustando los parámetros del modelo para minimizar el sobreajuste.

La validación de los algoritmos de aprendizaje automático en los diferentes trabajos revisados se llevó a cabo mediante técnicas de validación cruzada de k-pliegues, específicamente utilizando 10 pliegues. Además, calculan diversas métricas de rendimiento, incluyendo la exactitud, que es la proporción de predicciones correctas sobre el total de predicciones; la sensibilidad, que mide la capacidad del modelo para identificar correctamente los casos positivos; la especificidad, que evalúa la capacidad del modelo para identificar correctamente los casos negativos; y el AUC-ROC, que mide el área bajo la curva ROC para evaluar la capacidad del modelo de discriminar entre casos positivos y negativos.

Este artículo se divide en 5 secciones. La introducción del artículo proporciona un contexto sobre la importancia del estudio de la enfermedad de Parkinson y el papel del aprendizaje automático en su identificación temprana. El marco conceptual contextualiza el funcionamiento de los modelos de aprendizaje automático utilizados en el estudio, describiendo las técnicas SVM, RF y NB. La sección de materiales y métodos detalla las fuentes de datos utilizadas, el proceso de preprocesamiento de datos y la implementación de los algoritmos de aprendizaje automático. También explica los métodos de validación y las métricas de rendimiento empleadas. Los resultados y la discusión presentan los resultados obtenidos al aplicar las técnicas de aprendizaje automático en la identificación de la EP, comparando las métricas de rendimiento de los algoritmos evaluados y discutiendo sus implicaciones. Las conclusiones resumen los hallazgos del estudio, destacando cuál algoritmo es más efectivo y sugiriendo posibles direcciones para futuras investigaciones y aplicaciones clínicas.

2. Marco conceptual

En la identificación de la enfermedad de Parkinson, se han aplicado diversas técnicas de aprendizaje automático más allá de las tradicionales SVM, NB y RF [5, 8, 20, 21], que incluyen análisis de imágenes y análisis de datos de voz y movimiento.

2.1. Análisis de Imágenes

El uso de técnicas de análisis de imágenes, como RM y tomografía por emisión de positrones (PET), combinado con algoritmos de aprendizaje automático, permite la identificación de anomalías en el cerebro asoci-

adas con Parkinson. Por ejemplo, la combinación de datos de RM y PET con técnicas de aprendizaje profundo ha resultado en diagnósticos con alta precisión y especificidad [22].

2.2. Análisis de Datos de Voz y Movimiento

El análisis de datos de voz ha sido ampliamente utilizado en la identificación de Parkinson debido a la prevalencia de síntomas de alteración del habla en las primeras etapas de la enfermedad. Técnicas de aprendizaje automático, como el análisis de componentes principales (PCA) y el análisis discriminante lineal (LDA), se han aplicado para extraer características relevantes de las grabaciones de voz y han demostrado ser eficaces en la detección temprana de Parkinson [23, 24].

2.3. Ensembles de Modelos

El uso de ensembles de modelos, que combinan múltiples técnicas de aprendizaje automático, ha demostrado mejorar la precisión del diagnóstico de Parkinson. Por ejemplo, la combinación de modelos de SVM, RF y redes neuronales ha llevado a una mejora en la identificación de la enfermedad al aprovechar las fortalezas de cada método individual [25].

2.4. Maquinas de vectores de soporte

El algoritmo SVM, se emplea en el análisis de neuroimágenes [26, 27]. Debido a la capacidad de manejar la clasificación en datos de alta dimensionalidad, como los generados en estudios de neuroimagen. Este método se ha aplicado en la investigación de trastornos neurológicos, aplicándose en el Análisis de Patrones Multivoxel (MVPA) para identificar patrones complejos asociados con diversas condiciones cerebrales [26]. En la psiquiatría, las SVM se han utilizado para la predicción de diagnósticos y pronósticos en enfermedades como el Alzheimer, Parkinson, esquizofrenia y depresión [26, 27].

Las SVM optimizan el margen entre las clases en el espacio de características [27]. Para datos linealmente separables, este proceso implica la optimización del margen geométrico mediante la minimización de la norma del vector de pesos. En casos donde los datos no son linealmente separables, las SVM utilizan funciones kernel para proyectar los datos a un espacio de mayor dimensión donde la separación es posible. Los kernels más utilizados incluyen el lineal, polinómico, RBF y sigmoide, cada uno con sus propias ventajas según la naturaleza del problema.

Para abordar situaciones donde los datos presentan solapamientos y no permiten una separación perfecta, las SVM implementan un margen suave, introduciendo variables de holgura y un parámetro de penalización conocido como C [27]. Esto permite flexibilidad en la clasificación, aceptando un grado controlado de errores.

Además, la formulación dual del problema de optimización, que involucra multiplicadores de Lagrange, simplifica los cálculos y es especialmente útil cuando se combinan con funciones kernel.

Las SVM se basan en la determinación de un hiperplano que maximice el margen de separación entre las clases en el espacio de características. Este margen se establece como la distancia entre el hiperplano y los puntos más próximos de cada categoría [26], denominados vectores de soporte. La optimización de este margen se realiza minimizando una función de costo que penaliza tanto los errores de categorización como las infracciones del margen.

Las SVM son capaces de manejar eficientemente grandes cantidades de características y datos de alta dimensionalidad sin caer en el sobreajuste, lo que es fundamental en el análisis de datos biomédicos complejos [26]. Esta capacidad se ve potenciada por el uso de kernels que permiten transformar los datos originales en espacios de mayor dimensión, donde la clasificación puede ser más sencilla. Por ejemplo, el kernel RBF es especialmente útil para problemas donde las relaciones entre las variables no son lineales y el kernel polinómico puede capturar interacciones entre características que no son evidentes en el espacio original.

Además, las SVM son robustas frente a problemas de multicolinealidad y pueden manejar eficientemente situaciones donde hay más características que muestras, un escenario común en la bioinformática y otros campos de la biomedicina. Esto se logra mediante la regularización, que controla la complejidad del modelo y evita el sobreajuste, asegurando que el modelo generalice bien en datos no vistos.

2.5. Bayesiano Ingenuo

El modelo NB, fundamentado en el teorema de Bayes, es un algoritmo de categorización empleado en el aprendizaje automático y la minería de datos. [28]. Se utiliza para estimar la probabilidad de que una observación pertenezca a una categoría particular, basándose en características o atributos conocidos.

El teorema de Bayes ofrece un método para ajustar las probabilidades de una hipótesis a medida que se obtiene nueva evidencia [28]. En el contexto del NB, este teorema se utiliza para calcular la probabilidad posterior de una clase dada una combinación de características. El modelo NB asume que las características son independientes entre sí, una suposición que rara vez se cumple en la práctica, pero que simplifica considerablemente el cálculo de las probabilidades.

El NB se basa en la fórmula del teorema de Bayes, que se expresa como la ecuación (1):

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

donde $P(C|X)$ es la probabilidad posterior de la

clase C dado el vector de características X , $P(C|X)$ es la probabilidad de observar el vector de características X dado que la clase es C , $P(C)$ es la probabilidad previa de la clase C y $P(X)$ es la probabilidad marginal de observar el vector de características X .

Para calcular $P(X|C)$, el NB asume la independencia condicional de las características, es decir, que la presencia de una característica es independiente de la presencia de cualquier otra característica, dado la clase. Esta suposición permite descomponer $P(X|C)$ en el producto de las probabilidades individuales de las características según la ecuación (2):

$$P(X|C) = \prod_{i=1}^n P(x_i|C) \quad (2)$$

donde x_i es el valor de la i -ésima característica.

El NB se utiliza en una amplia gama de aplicaciones debido a su simplicidad y eficacia [19, 29]. Es particularmente útil en situaciones donde se necesita una rápida toma de decisiones, como en el filtrado de spam, la clasificación de texto y la detección de fraudes. Su capacidad para manejar tanto datos categóricos como continuos lo hace versátil en diferentes contextos.

Para aplicar el NB a datos continuos, como los biomarcadores en estudios médicos, se suele asumir una distribución gaussiana de las características. En este caso, la probabilidad $P(x_i|C)$ se calcula utilizando la función de densidad de probabilidad de una distribución normal. Para datos continuos [26], $P(x_i|C)$ se calcula según la ecuación (3):

$$P(x_i|C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} \exp\left(-\frac{(x_i - \mu_C)^2}{2\sigma_C^2}\right) \quad (3)$$

donde μ_C y σ_C son la media y la desviación estándar de la característica x_i para la clase C .

Una de las características del NB es su eficiencia computacional [30, 31], lo cual lo hace apropiado para el análisis de grandes volúmenes de datos. A pesar de su suposición de independencia condicional, el NB ha demostrado ser sorprendentemente efectivo en muchos escenarios prácticos.

2.6. Bosques Aleatorios

El RF es una técnica de aprendizaje automático basada en conjuntos, empleada para resolver problemas de regresión y clasificación [32]. Esta técnica se aplica tanto a problemas de clasificación como de regresión y se fundamenta en la construcción de múltiples árboles de decisión durante el entrenamiento [26]. Cada árbol se entrena con un subconjunto aleatorio de datos y características. La aleatoriedad en el proceso de entrenamiento ayuda a prevenir el sobreajuste y mejora la capacidad de generalización del modelo.

Para generar una predicción, cada árbol de decisión en el bosque emite un voto, y la decisión final se toma mediante votación por mayoría en problemas de clasificación o promediando las predicciones en problemas de regresión [17, 33, 34]. Este enfoque de conjunto, donde múltiples modelos se combinan para producir una única predicción, reduce la variabilidad y aumenta la precisión del modelo final.

Los RF incorporan aleatorización en el entrenamiento mediante el uso de bootstrap samples, donde cada árbol se entrena con un subconjunto aleatorio de datos. Además, en cada nodo de un árbol, se elige aleatoriamente un subconjunto de características para identificar la mejor división. Esta doble aleatorización introduce diversidad entre los árboles y contribuye a la robustez del modelo.

- **Árboles de decisión:** La base de los Bosques Aleatorios son los árboles de decisión, los cuales se construyen dividiendo iterativamente el espacio de características en regiones homogéneas con respecto a la variable de respuesta [17]. Cada división se realiza seleccionando la característica y el punto de división que minimizan la impureza en las particiones resultantes, según una métrica como la ganancia de información o el índice de Gini [33].
- **Bootstrap Aggregating (Bagging):** Este método utiliza la técnica de bagging para entrenar cada árbol con una muestra bootstrap del conjunto de datos original. Una muestra bootstrap se obtiene seleccionando aleatoriamente con reemplazo un subconjunto de los datos originales [34]. Esto significa que algunos datos pueden aparecer varias veces en una muestra, mientras que otros pueden no aparecer en absoluto. El bagging ayuda a reducir la varianza del modelo final.
- **Selección aleatoria de características:** En cada nodo de un árbol, se selecciona aleatoriamente un subconjunto de características disponibles y se determina la mejor división sólo entre este subconjunto [34]. Esta estrategia, junto con el bagging, introduce diversidad en los árboles, lo cual es fundamental para la efectividad de los Bosques Aleatorios.
- **Votación por mayoría y promedio:** En problemas de clasificación, cada árbol del conjunto vota por la clase más probable, y la clase final se determina por mayoría de votos. En problemas de regresión, las predicciones de todos los árboles se promedian para obtener la predicción final. Esta combinación de múltiples árboles reduce la varianza y mejora la precisión del modelo.

- **Importancia de características:** Los RF permiten calcular la importancia de las características, lo cual es útil para entender cuáles son los factores más influyentes en la predicción. La importancia de una característica se mide por la cantidad de reducción de impureza que dicha característica aporta a lo largo de todos los árboles del bosque.

Una de las ventajas de este modelo es su capacidad para reducir el sobreajuste. En comparación con los árboles de decisión individuales, los RF tienden a ser menos propensos al sobreajuste, lo que mejora la capacidad de generalización del modelo. Además, RF proporciona una medida de la importancia de cada característica en la predicción. Esta medida se obtiene observando la mejora en la precisión del modelo al dividir los nodos basados en una característica particular a lo largo de todos los árboles, lo cual es útil para la selección de características y la comprensión de los factores que más influyen en las predicciones.

Los RF son robustos frente a datos ruidosos y la ausencia de valores en el conjunto de datos [32]. La construcción de múltiples árboles con diferentes subconjuntos de datos permite que el modelo maneje mejor las anomalías y los valores atípicos.

3. Materiales y Métodos

3.1. Estrategias de búsqueda

Se realizó una revisión sistemática de la literatura para identificar estudios relevantes sobre la identificación de la EP utilizando técnicas de aprendizaje automático. La revisión siguió las directrices PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) para asegurar la exhaustividad y la reproducibilidad del proceso [35–37].

Para llevar a cabo esta revisión, se consultaron diversas bases de datos electrónicas, incluyendo PubMed, IEEE Xplore, ScienceDirect y Elsevier. Se utilizaron términos de búsqueda como "Parkinson's disease", "machine learning", "support vector machines", "Naive Bayes", "random forests" y "classification". La revisión se restringió a estudios publicados en los últimos 5 años para garantizar la actualidad y relevancia de los datos. Además, se dio prioridad a que las bases de datos empleadas para entrenar los modelos de aprendizaje automático incluyeran pacientes mayores de 60 años.

3.2. Criterios de inclusión y exclusión

Los estudios fueron seleccionados si cumplían con ciertos criterios de inclusión. En particular, los estudios debían (1) utilizar técnicas de aprendizaje automático para la identificación de la EP, (2) incluir datos de pacientes mayores de 60 años, y (3) presentar resultados

de evaluación de modelos. En contraste, se excluyeron aquellos estudios que no proporcionaban suficientes detalles sobre los datos utilizados o la metodología empleada.

3.3. Revisión de la literatura

Con base en los criterios de búsqueda, se recuperaron 37 artículos de PubMed, 21 artículos de IEEE Xplore, 18 artículos de ScienceDirect y 12 de Elsevier, obteniendo un total de 88 artículos. Se seleccionaron 60 publicaciones para la revisión de títulos y resúmenes, tras lo cual se excluyeron 19 artículos según los criterios de exclusión y se examinaron 37 textos completos. Finalmente, se incluyeron 28 artículos de investigación para la extracción de datos y métricas de rendimiento, como se muestra en la Tabla 1. Todos los artículos se publicaron a partir del año 2019 en adelante.

3.4. Objetivos del estudio

Para delinear las diferentes metas y objetivos de los estudios incluidos, se han categorizado de acuerdo con el tipo de diagnóstico y su objetivo general. Desde la perspectiva del diagnóstico, estos estudios se pueden dividir en tres categorías: (a) diagnóstico o detección de EP, que compara datos recopilados de pacientes con EP y controles sanos; (b) diagnóstico diferencial, que discrimina entre pacientes con EP idiopática y aquellos con EP atípica, como el parkinsonismo; y (c) subtipificación, que discrimina entre diferentes subtipos de EP.

Los estudios incluidos también se analizaron según su objetivo general. Aquellos centrados en el desarrollo de enfoques técnicos novedosos para el diagnóstico de la EP, como nuevos modelos y arquitecturas de aprendizaje automático, dispositivos de adquisición de datos y algoritmos de extracción de características no presentados o utilizados previamente, se definieron como estudios de "metodología". Por otro lado, los estudios que validan e investigan la aplicación de métodos previamente publicados y validados de aprendizaje automático, así como la viabilidad de introducir modalidades de datos no comúnmente utilizadas en el diagnóstico de la EP basado en aprendizaje automático (por ejemplo, datos del LCR), se definieron como "estudios de aplicación clínica".

3.5. Entradas de los modelos de aprendizaje automático

Las entradas para los modelos de aprendizaje automático utilizados en los estudios analizados incluyen:

1. Características Demográficas:

- **Edad:** Edad del paciente en años, mayores a 60 años.

- **Género:** Género del paciente (masculino/femenino).

2. Características Clínicas

- **Historia Médica:** Información sobre antecedentes familiares de Parkinson o trastornos neurológicos.
- **Duración de los Síntomas:** Tiempo transcurrido desde la aparición de los primeros síntomas de Parkinson.

3. Evaluaciones Neurológicas

- **UPDRS (Unified Parkinson's Disease Rating Scale):** Puntajes de las distintas secciones de esta escala que evalúan la severidad de los síntomas motores y no motores.
- **Evaluaciones Cognitivas:** Resultados de pruebas cognitivas que miden la función ejecutiva, memoria y atención.

4. Pruebas de Imagen

- **Imágenes de Resonancia Magnética (RM):** Datos cuantitativos extraídos de las imágenes de RM, tales como el volumen de ciertas estructuras cerebrales.

5. Características de Voz y Movimiento

- **Análisis de Voz:** Parámetros acústicos obtenidos de grabaciones de voz, como jitter, shimmer y pausas.
- **Análisis de Movimiento:** Datos obtenidos mediante sensores de movimiento que miden la bradicinesia y la rigidez.

En estudios previos, las características de voz han sido utilizadas extensivamente. Por ejemplo, [20] y [5] emplearon grabaciones de voz obtenidas de diferentes fuentes, como el UCI Machine Learning Repository y colecciones de participantes específicos.

3.6. Fuentes de datos y tamaño de la muestra

En 4 de 28 estudios (14.28%), los datos originales se recopilaron de participantes humanos. En 19 estudios (67.86%), los datos utilizados procedían de repositorios y bases de datos públicos, incluyendo el Repositorio de aprendizaje automático de la Universidad de California en Irvine (UCI), la iniciativa de marcadores de progresión de Parkinson (PPMI), PhysioNet, conjunto de datos PC-GITA, base de datos mPower y otras 4 bases de datos. En 5 estudios (17.86%), las fuentes de datos no fueron mencionadas explícitamente. Los participantes en estos estudios tenían un rango de edad que iba desde los 46 años hasta los 85 años.

Los 28 estudios tuvieron un tamaño de muestra promedio de 111.8 (122.4), con un tamaño de muestra más pequeño de 22 y un tamaño de muestra más grande de 584. El estudio con el tamaño de muestra más pequeño fue realizado por [38], con 22 personas, de las cuales 10 tenían EP y 12 eran sanas. Por otro lado, el estudio con el tamaño de muestra más grande fue realizado por [13], con 584 personas, de las cuales 401 tenían EP y 183 eran sanas.

3.7. Tipos de datos

En los estudios revisados, se utilizaron diversos tipos de datos para la identificación de la EP, los cuales se detallan en la Tabla 1. Estos tipos de datos incluyen grabaciones de voz, patrones de marcha, patrones de escritura, datos de movimiento ocular rápido, pérdida olfativa, datos de líquido cefalorraquídeo, marcadores de imágenes dopaminérgicas, datos de movimiento, señales de electroencefalograma y datos demográficos combinados con resultados de cuestionarios y pruebas psicofísicas.

Las grabaciones de voz fueron el tipo de dato más comúnmente utilizado, apareciendo en 17 estudios. Estos estudios incluyen trabajos como [20] y [39], que se centraron en analizar las características vocales de los participantes para identificar la enfermedad de Parkinson.

Los patrones de marcha fueron utilizados en 4 estudios, como se observa en los trabajos de [6] y [40], donde se analizaron las características de la marcha de los participantes para detectar anomalías asociadas con la enfermedad.

En 2 estudios, se emplearon patrones de escritura para el análisis, como en el estudio de [5], donde se evaluaron las características de la escritura de los participantes para identificar signos de Parkinson.

Un estudio utilizó datos de movimiento ocular rápido, pérdida olfativa, datos de líquido cefalorraquídeo y marcadores de imágenes dopaminérgicas. Este enfoque integral permitió una evaluación más completa de los diferentes aspectos de la enfermedad.

Además, 3 estudios se centraron en datos de movimiento, recopilando información sobre la actividad motora de los participantes. Otro estudio utilizó señales de electroencefalograma para evaluar la actividad cerebral y detectar patrones asociados con la enfermedad de Parkinson.

Finalmente, en dos estudios se combinaron datos demográficos con resultados de cuestionarios y pruebas psicofísicas, proporcionando una visión holística del estado de salud de los participantes.

Esta diversidad en los tipos de datos utilizados en los estudios revisados destaca la complejidad de la EP y la necesidad de múltiples enfoques para su identificación y diagnóstico efectivo.

3.8. Métricas de Rendimiento

En el presente estudio, se utilizaron la precisión, la sensibilidad, la especificidad y el AUC-ROC para comparar el rendimiento de los modelos de aprendizaje automático. Para cada tipo de datos, se resumió el tipo de modelo de aprendizaje automático que condujo a la mayor precisión en cada estudio.

Para los estudios que informaron la precisión tanto del entrenamiento como de las pruebas/validación, se consideró la precisión de las pruebas o la validación. En los estudios con más de un conjunto de datos o problemas de clasificación, por ejemplo, HC versus EP y HC versus hiposmia idiopática versus EP, se promedió la precisión entre los diferentes conjuntos de datos o problemas de clasificación. Para los estudios que informaron la precisión de la clasificación para cada grupo de sujetos individualmente, se promedió la precisión entre los grupos. En aquellos estudios que informaron una variedad de precisiones o precisiones obtenidas mediante diferentes métodos de validación cruzada o combinaciones de características, se consideraron las precisiones más altas.

3.8.1. Tablas Comparativas de Resultados

En las tablas comparativas de resultados (ver Tabla 1), se presenta una comparación de los resultados de diversos estudios que han utilizado modelos de aprendizaje automático para la identificación de la EP. Estas tablas incluyen información sobre:

- **Referencia:** La fuente bibliográfica del estudio.
- **Algoritmo:** Los algoritmos de aprendizaje automático utilizados, como SVM, RF, NB, k-NN, entre otros.
- **Tipo de datos:** Los tipos de datos empleados en cada estudio, que pueden incluir grabaciones de voz, patrones de escritura, datos de marcha, neuroimágenes, etc.
- **Exactitud:** La proporción de predicciones correctas realizadas por el modelo [26].
- **Precisión:** La proporción de verdaderos positivos sobre el total de positivos predichos [41].
- **Sensibilidad:** La capacidad del modelo para identificar correctamente los casos positivos [41].
- **Especificidad:** La capacidad del modelo para identificar correctamente los casos negativos [41].
- **F1-Score:** La media armónica de la precisión y la sensibilidad [41].
- **AUC:** El área bajo la curva ROC, que evalúa la capacidad del modelo para discriminar entre casos positivos y negativos [41].

Estas métricas permiten la evaluación del rendimiento de los diferentes modelos y algoritmos en la identificación de la EP. A continuación se presenta una descripción más detallada de los estudios incluidos:

La Tabla 1 presenta una comparación detallada de los resultados obtenidos en diferentes estudios utilizando diversos algoritmos de aprendizaje automático para la clasificación de la EP partir de grabaciones de voz, patrones de escritura, datos de marcha, datos de movimiento, datos de líquido cefalorraquídeo, marcadores de imágenes dopaminérgicos, resultados de cuestionarios y pruebas psicofísicas. Se incluyen métricas de rendimiento como exactitud, precisión, sensibilidad, especificidad, F1-Score y AUC. Por ejemplo, el estudio de [42] reporta una exactitud y sensibilidad del 99.00% y 100.00%, respectivamente, utilizando SVM, destacando la alta eficacia de este algoritmo en su aplicación. Otro ejemplo es el trabajo de [43], que reporta una exactitud del 92.30% y un F1-Score de 92.10%.

Además, se observa que los estudios utilizan una variedad de algoritmos como SVM, RF, NB y ANN, indicando un enfoque multidisciplinario en la búsqueda del modelo más efectivo. La diversidad en los tipos de datos utilizados, como señales de electroencefalograma y otros métodos de aprendizaje automático, también se refleja en la tabla, lo que resalta la amplitud de técnicas empleadas en la investigación de la enfermedad de Parkinson.

3.9. Comparación de modelos de aprendizaje automático

Para optimizar la comparación de resultados entre los artículos analizados, se propone un enfoque multifacético que abarca análisis cuantitativos y cualitativos.

3.9.1. Análisis Descriptivo

Un análisis descriptivo de la métrica de exactitud, calculando la media y la desviación estándar de las diferentes métricas de rendimiento, se realizó para proporcionar una visión general del rendimiento promedio y la variabilidad de los modelos. La media se calculó utilizando la fórmula (4), mientras que la desviación estándar se determinó mediante la fórmula (5). Esto permitió obtener una medida precisa de la centralidad y dispersión de los datos, facilitando así la comparación entre los diferentes estudios y algoritmos presentados en la Tabla 1.

- **Media** (μ): La media se calcula como la suma de todas las observaciones dividida por el número total de observaciones [41], como se muestra en la ecuación (4).

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (4)$$

donde x_i representa cada observación y N es el número total de observaciones.

- **Desviación Estándar** (σ): La desviación estándar se calcula como la raíz cuadrada de la varianza, que es la media de los cuadrados de las diferencias de cada observación respecto a la media [41], como se muestra en la ecuación (5).

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (5)$$

donde μ es la media de las observaciones.

3.9.2. Análisis Cualitativo

Un análisis cualitativo de los artículos con resultados extremos. Examinar los métodos, datos y características específicas empleados en estos estudios. Es importante considerar otros factores, como el tamaño de la muestra, el tipo de datos y el diseño del estudio, para contextualizar las diferencias observadas en la exactitud.

3.9.3. Análisis de Subgrupos

Se propone un análisis de subgrupos, agrupando los artículos según características relevantes como el tipo de algoritmo y los datos.

4. Resultados y Discusión

Los resultados obtenidos en esta investigación se derivan del análisis y comparación del rendimiento de tres técnicas de aprendizaje automático: SVM, NB y RF. Estos modelos fueron evaluados utilizando conjuntos de datos que incluyen grabaciones de voz y datos clínicos de pacientes con Parkinson y un grupo control.

Los resultados presentados en la Tabla 1 se basan en el rendimiento de los modelos reportado en los estudios revisados. Las gráficas se generaron calculando la media y la desviación estándar de cada métrica de rendimiento para los diferentes modelos, proporcionando una representación visual de su desempeño.

4.1. Métodos de aprendizaje automático aplicados en la identificación de la EP

En los estudios revisados, se emplearon diversos algoritmos de aprendizaje automático para la identificación de la enfermedad de Parkinson.

El algoritmo más comúnmente utilizado fue el SVM, apareciendo en 20 estudios. Entre estos estudios se encuentran los trabajos de [20] y [39], que utilizaron SVM por su capacidad para gestionar grandes conjuntos de características y determinar un hiperplano óptimo que separa las clases de datos con la máxima distancia.

RF fue el segundo algoritmo más utilizado, apareciendo en 6 estudios. Este algoritmo permite manejar datos con alta dimensionalidad y es robusto ante el sobreajuste. Un ejemplo notable es el estudio de [13], donde se utilizó RF para evaluar múltiples características y mejorar la precisión del diagnóstico.

NB fue utilizado en 2 estudios. Este algoritmo es conocido por su simplicidad y eficiencia en clasificaciones de alta velocidad, especialmente útil en situaciones donde se necesita una rápida toma de decisiones. El estudio de [38] utilizó NB para clasificar los datos de los participantes y evaluar la presencia de la enfermedad.

La selección del algoritmo depende de las características del conjunto de datos y del objetivo particular de cada estudio. Mientras que SVM se utiliza ampliamente debido a su efectividad en espacios de alta dimensionalidad, se prefiere RF por su capacidad para gestionar características correlacionadas y ofrecer una estimación de la importancia de cada característica. NB, por otro lado, es elegido por su simplicidad y rapidez en la clasificación.

4.2. Métricas de rendimiento

En los estudios revisados, se utilizaron diversas métricas de rendimiento para evaluar la efectividad de los algoritmos de aprendizaje automático en la identificación de la EP. Estas métricas incluyen exactitud, precisión, sensibilidad, especificidad, F1-Score y área bajo la curva (AUC), las cuales se detallan en la Tabla 1.

La exactitud fue reportada en 19 estudios. Por ejemplo, [20] y [39] informaron una exactitud de 86% y 95%, respectivamente. La exactitud evalúa el porcentaje de predicciones correctas efectuadas por el modelo.

La precisión fue mencionada en 10 estudios. En el estudio de [38], se informó una precisión de 93.84%. La precisión calcula la proporción de verdaderos positivos sobre el total de positivos predichos, ofreciendo una medida de la exactitud entre las instancias clasificadas como positivas.

La sensibilidad (o recall) fue utilizada en 11 estudios. [6] informaron una sensibilidad de 98.26%, lo

cual refleja la capacidad del modelo para identificar correctamente las instancias positivas.

La especificidad fue reportada en 8 estudios. Por ejemplo, [40] informaron una especificidad de 99.87%, lo que demuestra la capacidad del modelo para reconocer correctamente las instancias negativas.

El F1-Score se utilizó en 6 estudios. Esta métrica representa la media armónica de la precisión y la sensibilidad, ofreciendo un equilibrio entre ambas medidas. En el estudio de [13], se reportó un F1-Score de 99.5%.

El área bajo la curva (AUC) se informó en 4 estudios. Esta métrica evalúa la capacidad del modelo para diferenciar entre clases. Por ejemplo, [38] informaron un AUC de 0.9657, lo que indica una excelente capacidad discriminadora del modelo.

Estas métricas ofrecen una visión completa del rendimiento de los algoritmos empleados y facilitan la comparación de la efectividad de distintos enfoques en la identificación de la EP.

A continuación, se calcula la media y la desviación estándar de las métricas de rendimiento para realizar un análisis descriptivo, cualitativo y por subgrupos. Para calcular la media y la desviación estándar de cada métrica, se procede de la siguiente manera:

1. Recolección de datos: Se extraen los valores reportados para cada métrica de los estudios disponibles.
2. Limpieza de datos: Se eliminan los valores faltantes y se aseguran que los datos sean numéricos y consistentes.
3. Cálculo de la media: Los valores de cada métrica se suman y luego se dividen por el número de estudios que informaron dicha métrica.
4. Cálculo de la desviación estándar: Se obtiene la raíz cuadrada de la varianza, la cual es la media de los cuadrados de las diferencias entre cada observación y la media.

Este proceso permite obtener una visión general del rendimiento promedio de los algoritmos utilizados y facilita el análisis comparativo entre diferentes enfoques y subgrupos de estudios.

Tabla 1. Tabla comparativa de los resultados de cada estudio

Referencia	Algoritmo	Tipo de datos	Exactitud	Precisión	Sensibilidad	Especificidad	F1-Score	AUC
Karapinar Senturk, 2020 [20]	SVM	Grabaciones de voz	-	93,84%	-	-	-	-
Sakar et al., 2019 [39]	SVM	Grabaciones de voz	86,00%	-	-	-	-	-
Ouhmida, A., 2021 [5]	SVM, K-NN, DT	Grabaciones de voz	-	98,26%	-	-	-	-
Aich et al., 2019 [40]	SVM, RF	Grabaciones de voz	-	-	97,56%	99,87%	-	96,57%
Suppa et al., 2022 [6]	SVM	Grabaciones de voz	98,30%	-	-	96,60%	100%	99,90%
Haq et al., 2019 [42]	SVM	Grabaciones de voz	99,00%	-	100%	99,00%	-	-
Yaman et al., 2020 [44]	SVM, k-NN	Grabaciones de voz	91,25%	91,15%	91,25%	-	91,25%	-
Sztahó et al., 2019 [45]	SVM - RBF, k-NN, ANN	Grabaciones de voz	84,62%	-	-	-	-	-
Mostafa et al., 2019 [17]	RF, NB, ANN, DT, SVM	Grabaciones de voz	99,49%	99,50%	99,50%	-	-	100%
Pramanik et al., 2021 [33]	ForestPA, SysFor, RF	Grabaciones de voz	94,12%	-	94,00%	94,00%	-	-
Sharanyaa et al., 2020 [7]	RF, LR, K-NN, NB	Grabaciones de voz	87,27%	70,60%	92,80%	-	85,80%	-
Rana et al., 2022 [18]	SVM, NB, K-NN, ANN	Grabaciones de voz	87,17%	-	62,50%	93,54%	66,19%	-
Avuçlu et al., 2020 [29]	NB, RF, K-NN, SVM	Grabaciones de voz	70,26%	-	-	-	-	-
Elshewey et al., 2023 [43]	SVM, LR, NB, RC, DT	Grabaciones de voz	92,30%	92,10%	92,30%	-	92,10%	-
Tadse et al., 2021 [46]	SVM, DT, LR, k-NN	Grabaciones de voz	82,05%	100%	42,00%	100%	59,00%	-
Bharath y Rajagopalana, 2023 [34]	RF, DT, LR, SVM, ANN	Grabaciones de voz	96,56%	-	98,26%	96,06%	93,00%	-

Referencia	Algoritmo	Tipo de datos	Exactitud	Precisión	Sensibilidad	Especificidad	F1-Score	AUC
Nayak et al., 2023 [47]	SVM	Grabaciones de voz	98,70%	98,90%	-	-	98,70%	-
Fabian Maass et al., 2020 [8]	SVM, AdaBoost, K-NN	Patrones de escritura	-	-	81,30%	-	-	-
Segovia et al., 2019 [9]	SVM	Patrones de escritura	94,20%	-	-	-	-	-
J. P. Félix et al., 2019 [10]	SVM	Marcha	96,80%	-	-	-	-	-
Andrei et al., 2019 [11]	SVM	Marcha	100%	-	-	-	-	-
Ricciardi et al., 2020 [12]	RF	Marcha	-	84,60%	-	94,90%	94,90%	-
Balaji et al., 2021 [19]	NB, SVM, Bagging, K-NN	Marcha	-	98,40%	-	-	-	-
Wang et al., 2020 [13]	SVM	Movimiento ocular rápido, pérdida olfativa, datos de líquido cefalorraquídeo, marcadores de imágenes dopaminérgicos	95,96%	96,90%	97,24%	93,16%	97,06%	98,70%
Ricci et al., 2019 [48]	SVM, K-NN, NB	Movimiento	95,00%	95,10%	-	-	-	95,00%
Zhang et al., 2023 [14]	SVM, DT, k-NN, NB, ANN, PLR, RF	Demográficos, exámenes de admisión hospitalaria, evaluaciones clínicas, puntuación de riesgo poligénico, biomarcadores	88,60%	-	93,32%	73,61%	-	83,90%
Oliveira de Santana et al., 2020 [38]	RF, SVM, MLP, J48, DT, ELM, mELM, NB	Señales de electroencefalograma	-	99,22%	-	-	-	-
Martinez-Eguiluz et al., 2022 [15]	SVM, MLP, DT, k-NN, NB, RF, AdaBoost, Bagging, RIPPER	Demográficos, resultados de cuestionarios y pruebas psicofísicas	87,50%	-	-	-	86,90%	-

Nota: Máquina de vectores de soporte (SVM), k vecinos más cercano (K-NN), Árbol de decisión (DT), Regresión logística (LR), Potenciación del gradiente (GB), Bosque aleatorio (RF), Modelo de Markov Oculto (HMM), Bayesiano ingenuo (NB), Red neuronal artificial (ANN), Bosque de desarrollo sistemáticamente (SysFor), Bosque de decisiones mediante atributos de penalización (ForestPA), Clasificador Ridger (RC), Máquina de gradiente luminoso (LightGBM), Aumento extremo del gradiente(XGBoost), Refuerzo adaptativo (AdaBoost), Agregación de Bootstrap (Bagging), Perceptrón multicapa (MLP), Regresión logística penalizada (PLR), Máquina de Aprendizaje Extremo (ELM), Máquina de Aprendizaje Extremo Morfológico (mELM), Poda incremental repetida para producir reducción de errores (RIPPER).

Para proporcionar una comprensión más detallada de las métricas de rendimiento evaluadas en los estudios revisados, se ha incluido una tabla de estadísticas descriptivas en los anexos (ver Tabla 3). Esta tabla presenta las formulas empleadas para calcular la media y la desviación estándar de varias métricas de rendimiento clave: exactitud, precisión, sensibilidad, especificidad, F1-Score y área bajo la curva (AUC).

La media de cada métrica (exactitud, precisión, sensibilidad, especificidad, F1-Score y AUC) se calcula como la suma de todas las observaciones dividida por el número total de observaciones. Esto proporciona una medida general del rendimiento promedio de los modelos de aprendizaje automático evaluados en los estudios. Por ejemplo, la media de la exactitud indica el porcentaje promedio de predicciones correctas reali-

zadas por los modelos a lo largo de todos los estudios revisados.

La desviación estándar de cada métrica mide la dispersión de las observaciones en relación con la media. Esto proporciona una medida de la consistencia de los resultados entre los diferentes estudios. Por ejemplo, una desviación estándar baja en la precisión indicaría que los modelos evaluados tendieron a tener un rendimiento consistente en términos de precisión, mientras que una desviación estándar alta indicaría una mayor variabilidad en los resultados.

4.2.1. Análisis Descriptivo

En primer lugar, se realiza el análisis descriptivo de la exactitud de los algoritmos. Para esto, se calcula la media de la exactitud utilizando la fórmula (4), que consiste en sumar todas las observaciones de exactitud y dividir el resultado entre el número total de observaciones. Este cálculo permite obtener un valor promedio de la exactitud de 0.917.

Posteriormente, se calcula la desviación estándar mediante la fórmula (5), la cual mide la dispersión de las observaciones respecto a la media. Este cálculo implica determinar las diferencias de cada observación con respecto a la media, elevar esas diferencias al cuadrado, promediar esos valores y, finalmente, tomar la raíz cuadrada del resultado. De esta manera, se obtiene la desviación estándar de 0.073.

Estos valores indican que, en promedio, los algoritmos de aprendizaje automático logran una exactitud del 91.7% en la identificación de la EP. La desviación estándar de 0.073 sugiere una variabilidad en el rendimiento de los algoritmos. Esta variabilidad puede atribuirse a diversos factores, como el tipo de datos utilizados, las técnicas de preprocesamiento y las características específicas de los algoritmos.

La exactitud promedio refleja la eficacia de estos métodos en el diagnóstico de la EP. La variabilidad observada destaca la importancia de considerar cuidadosamente el diseño del estudio, incluyendo la selección de datos y las metodologías aplicadas. Factores como la calidad y la heterogeneidad de los datos, así como las configuraciones de los modelos, pueden influir en los resultados.

4.2.2. Análisis Cualitativo

En segundo lugar tenemos el análisis de los artículos con métricas extremas como se observa en la figura 1.

El análisis de los artículos muestran una disparidad en la exactitud de los algoritmos empleados para la identificación de la enfermedad de Parkinson. La SVM ha demostrado una exactitud del 100% en el estudio de [11]. Este resultado sugiere que el algoritmo SVM, particularmente cuando se aplica a datos específicos como los de la marcha, puede ser extremadamente

eficaz. Por otro lado, el algoritmo NB mostró la menor exactitud con un 70.26% en el estudio de [29], lo que podría indicar limitaciones del algoritmo cuando se aplica a ciertos tipos de datos o en combinación con técnicas de preprocesamiento específicas.

Los datos de marcha utilizados en el modelo SVM parecen ser efectivos para la identificación de la EP.

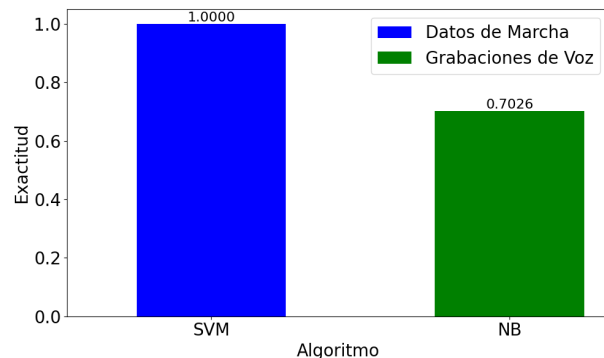


Figura 1. Algoritmos con mayor y menor exactitud

Esto podría deberse a la alta calidad y especificidad de las características extraídas de las señales de marcha, que proporcionan una representación detallada del patrón de movimiento del individuo afectado. En contraste, las grabaciones de voz, aunque comunes en varios estudios, presentan una variabilidad en la efectividad que depende significativamente del algoritmo empleado y de las técnicas de preprocesamiento aplicadas.

El Laboratorio de Marcha y Neurodinámica proporciona datos de alta calidad que, combinados con algoritmos robustos como el SVM, pueden conducir a un rendimiento excepcional en la identificación de la EP. Por otro lado, el UCI Machine Learning Repository, aunque es una fuente de datos fiable y ampliamente utilizada, puede mostrar variaciones en la exactitud de los resultados dependiendo del algoritmo y el preprocesamiento de datos utilizados.

4.2.3. Análisis de Subgrupos

En tercer lugar se realiza el análisis de subgrupos por algoritmo, donde se realiza el cálculo de la media y desviación estándar de las métricas de exactitud, especificidad y sensibilidad de los 3 modelos de aprendizaje automático. En la figura 2 y la figura 3 se observan los resultados de estos cálculos.

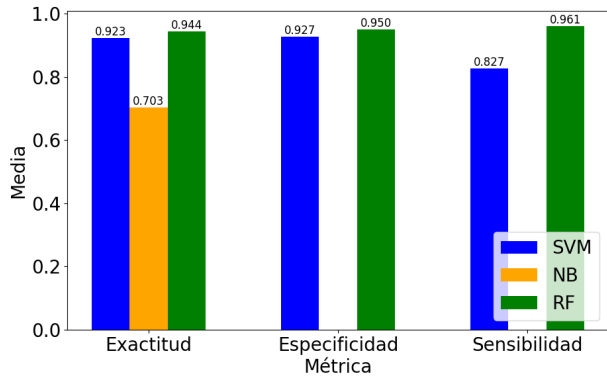


Figura 2. Media de las diferentes métricas por algoritmos

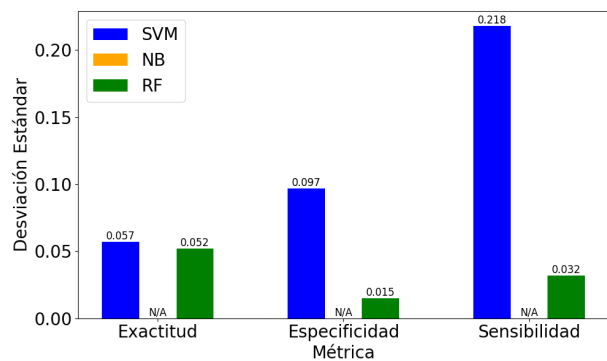


Figura 3. Desviación estándar de las diferentes métricas por algoritmos

El modelo de SVM muestra un buen rendimiento en términos de exactitud y especificidad, alcanzando una media de especificidad cercana al 93%. Estos resultados indican una alta capacidad del modelo para identificar correctamente los casos negativos, reduciendo así los

falsos positivos. Sin embargo, la sensibilidad presenta una mayor variabilidad, con una desviación estándar considerablemente alta. Esto sugiere que el modelo puede tener inconsistencias en la detección de casos positivos de enfermedades neurodegenerativas, lo que podría limitar su aplicación en entornos clínicos donde la identificación precisa de estos casos es crucial.

El modelo NB presenta una exactitud inferior en comparación con los modelos SVM y RF, con una media de 0.703 basada en un único artículo disponible para el análisis. La falta de datos sobre especificidad y sensibilidad impide una evaluación completa del rendimiento del modelo. No obstante, la baja exactitud observada sugiere limitaciones significativas en el rendimiento de NB en este contexto, lo que podría deberse a las asunciones simplificadas del modelo sobre la independencia de las características.

Si bien NB es reconocido por su simplicidad y eficiencia, presentando un rendimiento notable en diversas áreas como la clasificación de texto y el manejo de problemas con alta dimensionalidad, es importante

considerar ciertas limitaciones [28] inherentes al modelo:

- **Supuesto de independencia condicional:** El algoritmo asume que la presencia de una característica en una clase es independiente de otras características. Sin embargo, en aplicaciones reales, esta suposición a menudo no se cumple, lo que puede impactar negativamente en la precisión del modelo.
- **Sensibilidad a datos irrelevantes:** La inclusión de características irrelevantes en el conjunto de datos puede introducir ruido y distorsionar el proceso de aprendizaje del modelo, afectando su desempeño.
- **Distribución de datos:** NB asume una distribución gaussiana (normal) de los datos. En situaciones donde los datos no se ajustan a esta distribución, la efectividad del modelo puede verse comprometida.

Es importante tener en cuenta estas limitaciones al implementar NB, evaluando cuidadosamente la naturaleza de los datos y las relaciones entre las características para asegurar un rendimiento óptimo del modelo.

El modelo de RF demuestra un rendimiento sobresaliente, con alta exactitud, especificidad y sensibilidad. La exactitud media de 0.944 y la especificidad de 0.950 indican una excelente capacidad del modelo para distinguir entre casos positivos y negativos. Además, la sensibilidad media de 0.961 sugiere que RF es altamente eficaz en la identificación de casos positivos de enfermedades neurodegenerativas. La baja desviación estándar en todas estas métricas indica un rendimiento consistente en diferentes estudios, lo que refuerza la robustez y fiabilidad del modelo RF para aplicaciones clínicas en diversos escenarios y tipos de datos.

No obstante, RF presenta algunas limitaciones [32]. La complejidad del modelo puede dificultar su interpretación, especialmente al aumentar el número de árboles. Además, no es el algoritmo más adecuado para problemas con relaciones lineales simples entre características y la variable objetivo. También requiere un ajuste de hiperparámetros, aunque es menos sensible a este aspecto que otros algoritmos.

Para tener una perspectiva más completa del desempeño de los modelos de aprendizaje automático se analiza por subgrupos todas las métricas de rendimiento y el preprocesamiento realizado a las bases de datos, como se ilustra en la figura 4 y la figura 5.

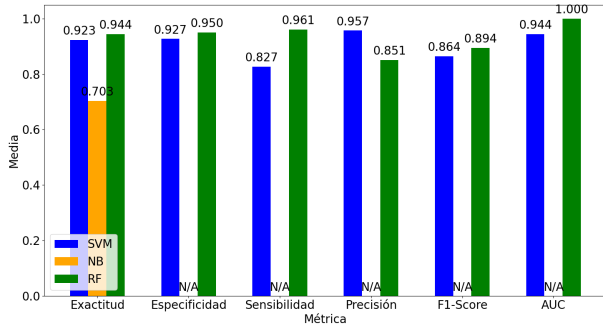


Figura 4. Media de todas las métricas por algoritmos

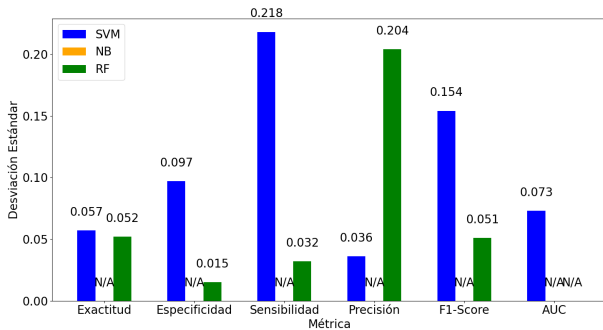


Figura 5. Desviación estándar de todas las métricas por algoritmos

Al considerar métricas adicionales como precisión, F1-score y AUC, junto con el preprocesamiento de datos, se obtiene una evaluación más completa y matizada del rendimiento de los algoritmos. SVM y RF se destacan como los algoritmos más robustos y confiables para la identificación de la EP, con RF mostrando una consistencia notable en métricas críticas. Por otro lado, NB presenta limitaciones que requieren un análisis más detallado y posiblemente un ajuste en las técnicas de preprocesamiento y selección de características para mejorar su eficacia. Dando como resultado que el artículo con mejores resultados de cada subgrupo es:

- **SVM:** Demuestra un rendimiento notable en diversas métricas clave, destacando en precisión y AUC, lo cual sugiere una capacidad robusta para distinguir entre diferentes clases. Sin embargo, la alta desviación estándar en la sensibilidad indica que el rendimiento puede ser variable en la detección de casos positivos de EP. El uso de técnicas avanzadas de preprocesamiento y la cuidadosa selección de características contribuyen significativamente a su alto rendimiento. El artículo de [13] muestra un rendimiento equilibrado y sólido en todas las métricas clave, siendo el mejor en este subgrupo.
- **NB:** Se observa una exactitud inferior en comparación con SVM y RF, lo que indica posibles

limitaciones en su capacidad para clasificar correctamente los casos de EP. La falta de datos sobre especificidad, sensibilidad y otras métricas clave dificulta una evaluación completa de su rendimiento. El preprocesamiento utilizado es básico en comparación con otros algoritmos, lo que puede contribuir a su menor rendimiento. El artículo de [29], aunque el mejor en su subgrupo, presenta limitaciones significativas en comparación con SVM y RF.

- **RF:** Demuestra un rendimiento excelente y consistente en todas las métricas clave, destacándose especialmente en sensibilidad y AUC, lo que lo hace muy adecuado para identificar casos de EP con alta precisión. La variabilidad en la precisión sugiere que, aunque es eficaz en la clasificación general, puede haber algunos falsos positivos. El uso de preprocesamiento avanzado y la selección de características aseguran que solo las características más relevantes sean consideradas, contribuyendo significativamente a su alto rendimiento. El artículo de [17] presenta un rendimiento sobresaliente, especialmente en exactitud, sensibilidad y AUC, destacándose como el mejor en su subgrupo.

4.3. Tendencias y Lagunas Identificadas

El análisis de los estudios revisados permitió identificar varias tendencias predominantes en la investigación actual, se muestran en la tabla 2:

Tabla 2. Tendencias y Lagunas Identificadas en la Literatura

Categoría	Descripción
Tendencias en los Algoritmos	Las técnicas de aprendizaje automático más comúnmente empleadas incluyeron SVM y RF. La combinación de múltiples algoritmos también fue una práctica común para mejorar la precisión del diagnóstico.
Tipos de Datos Utilizados	Los datos de neuroimagen y las grabaciones de voz fueron los tipos de datos más utilizados, seguidos de los patrones de escritura y los datos de la marcha. Los estudios destacaron la importancia de combinar múltiples tipos de datos para mejorar la precisión del diagnóstico.

Categoría	Descripción
Preprocesamiento de Datos	Las técnicas de preprocesamiento más comunes incluyeron la normalización, el análisis de componentes principales (PCA) y la selección de características. La imputación de datos faltantes y el filtrado de ruido también fueron técnicas frecuentemente aplicadas.
Herramientas y Software	Python, MATLAB y R fueron las herramientas más utilizadas para el procesamiento de datos y la implementación de algoritmos de aprendizaje automático.
Lagunas en la Investigación	La necesidad de más estudios longitudinales para evaluar la progresión de la EP y la falta de investigaciones que exploren la combinación de múltiples técnicas de aprendizaje automático para mejorar la precisión del diagnóstico.

5. Conclusiones

En este estudio, se compararon tres técnicas de aprendizaje automático para la identificación de la enfermedad de Parkinson en personas mayores de 60 años. A través de un análisis de los algoritmos utilizados, las fuentes de datos y las métricas de rendimiento, se obtuvieron varios hallazgos.

En primer lugar, se revisaron 88 artículos de varias bases de datos, de los cuales se seleccionaron 28 artículos de investigación para la extracción de datos y métricas de rendimiento. Los estudios se publicaron desde el año 2019 en adelante, reflejando las tendencias más recientes en la investigación de la enfermedad de Parkinson.

En términos de fuentes de datos, se observó que el 67.86% de los estudios utilizaron datos de repositorios públicos, mientras que solo el 14.28% de los estudios recopiló datos de participantes humanos. Esta tendencia destaca la dependencia de la investigación actual en datos accesibles y preexistentes, lo que facilita la replicación y la comparación entre estudios. Los datos de los participantes humanos mostraron un rango de edad entre 46 y 85 años, proporcionando una visión adecuada de la población mayor afectada por la enfermedad.

Los tipos de datos más comunes fueron las grabaciones de voz, utilizadas en 17 estudios, seguidas por los patrones de marcha y escritura. La diversidad de los tipos de datos utilizados resalta la complejidad de

la enfermedad de Parkinson y la necesidad de múltiples enfoques para su identificación y diagnóstico efectivo.

En cuanto a los algoritmos de aprendizaje automático, el SVM fue el más utilizado, apareciendo en 20 estudios, seguido por RF en 6 estudios y NB en 2 estudios. La elección del algoritmo depende de las características del conjunto de datos y del objetivo específico de cada estudio. Los resultados mostraron que SVM y RF fueron los más efectivos en la identificación de la enfermedad de Parkinson, con SVM destacándose en términos de exactitud y especificidad, mientras que RF mostró una alta sensibilidad y consistencia en diferentes estudios.

Las métricas de rendimiento más reportadas fueron la exactitud y la sensibilidad, reflejando la capacidad de los algoritmos para identificar correctamente las instancias positivas y negativas. La media y la desviación estándar de estas métricas proporcionaron una visión integral del rendimiento de los algoritmos, facilitando el análisis comparativo entre diferentes enfoques y subgrupos de estudios.

El modelo de SVM demostró una exactitud promedio del 91.7%, destacándose especialmente en estudios donde se utilizaron datos específicos como las señales de marcha. Esta técnica mostró una capacidad sobresaliente para diferenciar entre individuos con y sin EP, alcanzando en algunos casos una exactitud del 100%. Sin embargo, la variabilidad en la sensibilidad sugiere que su rendimiento puede no ser consistente en todos los contextos, lo que resalta la necesidad de un ajuste cuidadoso de las características y técnicas de preprocesamiento utilizadas.

El modelo NB presentó limitaciones significativas, con una exactitud promedio de 70.3%. Estas limitaciones pueden atribuirse a las suposiciones simplificadas del modelo sobre la independencia de las características y a su sensibilidad a datos irrelevantes y no gaussianos. Aunque NB es eficiente y simple, su aplicabilidad en el diagnóstico de EP parece restringida, requiriendo mejoras en el preprocesamiento y la selección de características para potenciar su rendimiento.

El modelo RF se destacó como el más robusto y consistente entre los tres modelos evaluados, con una exactitud promedio de 94.4% y alta especificidad y sensibilidad. Su capacidad para manejar grandes volúmenes de datos y su resistencia a la variabilidad en las métricas lo hacen particularmente adecuado para aplicaciones clínicas. No obstante, su complejidad puede presentar desafíos en términos de interpretación y ajuste de hiperparámetros.

A continuación, se presenta un análisis descriptivo, cualitativo y por subgrupos de las métricas de rendimiento. El análisis mostró que, en promedio, los algoritmos lograron una exactitud del 91.7% en la identificación de la enfermedad de Parkinson, con una desviación estándar de 0.073. Esta variabilidad puede atribuirse a factores como el tipo de datos utilizados,

las técnicas de preprocesamiento y las características específicas de los algoritmos.

El análisis cualitativo reveló que el algoritmo SVM, cuando se aplica a datos específicos como los de la marcha, puede ser eficaz, alcanzando una exactitud del 100%. Por otro lado, el algoritmo NB mostró la menor exactitud con un 70.26%, lo que indica limitaciones del algoritmo cuando se aplica a ciertos tipos de datos o en combinación con técnicas de preprocesamiento específicas.

El análisis de subgrupos destacó que el modelo SVM mostró un buen rendimiento en términos de exactitud y especificidad, mientras que RF demostró una alta efectividad en la identificación de casos positivos de enfermedades neurodegenerativas. La baja desviación estándar en todas las métricas indica un rendimiento consistente en diferentes estudios, lo que refuerza la robustez y fiabilidad del modelo RF para aplicaciones clínicas en diversos escenarios y tipos de datos.

En general, los resultados indican que tanto SVM como RF son efectivos en la identificación de la EP, con RF mostrando una consistencia notable en diferentes escenarios y tipos de datos. La selección de características y las técnicas de preprocesamiento son factores críticos que influyen significativamente en el rendimiento de los modelos, y deben ser cuidadosamente considerados en el diseño de estudios futuros.

A pesar de los avances logrados en este estudio, existen varias áreas que requieren mayor investigación. En primer lugar, se recomienda explorar el uso de algoritmos de aprendizaje profundo, que han demostrado ser efectivos en otros dominios de la medicina. Además, la integración de datos multimodales, que incluyen no solo grabaciones de voz y patrones de marcha, sino también imágenes cerebrales y datos genéticos, puede ofrecer una visión más holística de la enfermedad.

Otro aspecto importante es la necesidad de estudios longitudinales que sigan a los pacientes a lo largo del tiempo para comprender mejor la progresión de la enfermedad y ajustar los modelos predictivos en consecuencia. Asimismo, es fundamental considerar la diversidad de la población estudiada, incluyendo diferentes grupos étnicos y géneros, para garantizar que los modelos sean generalizables y aplicables a una población más amplia.

Este estudio proporciona una base para futuras investigaciones en la identificación de la enfermedad de Parkinson mediante técnicas de aprendizaje automático. Al continuar explorando y mejorando estos enfoques, se puede avanzar hacia un diagnóstico más temprano y preciso, lo que a su vez puede mejorar significativamente la calidad de vida de los pacientes con Parkinson. Los hallazgos subrayan la importancia de continuar explorando y optimizando las técnicas de aprendizaje automático en el diagnóstico de enfermedades neurodegenerativas. Futuras investigaciones

podrían centrarse en la integración de diferentes tipos de datos y en el desarrollo de enfoques híbridos que combinen las fortalezas de múltiples algoritmos para mejorar aún más la precisión y consistencia en la identificación de la EP.

Futuras investigaciones podrían centrarse en la integración de diferentes tipos de datos y en el desarrollo de enfoques híbridos que combinen las fortalezas de múltiples algoritmos para mejorar aún más la precisión y consistencia en la identificación de la EP. La exploración de técnicas de aprendizaje profundo, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), podría ofrecer nuevas perspectivas en la identificación de patrones complejos en los datos biomédicos.

Además, es esencial desarrollar y validar modelos predictivos en cohortes más grandes y diversas para asegurar la generalizabilidad de los hallazgos. La colaboración entre diferentes instituciones y la creación de consorcios de datos pueden facilitar la obtención de conjuntos de datos más representativos y robustos.

Otro enfoque es la personalización de los modelos predictivos mediante el uso de datos genéticos y biomarcadores específicos de cada paciente. Esto podría permitir la creación de modelos más precisos y adaptados a las características individuales de los pacientes, mejorando así la capacidad de predecir la progresión de la enfermedad y la respuesta al tratamiento.

La incorporación de técnicas de aprendizaje automático en dispositivos portátiles y aplicaciones móviles también podría revolucionar el monitoreo continuo y en tiempo real de los síntomas de la EP. Esto permitiría una detección temprana de cambios en el estado de salud de los pacientes y una intervención oportuna, mejorando potencialmente los resultados clínicos.

Finalmente, la ética y la privacidad de los datos deben ser consideraciones clave en cualquier investigación futura. Es crucial desarrollar frameworks robustos para el manejo seguro y ético de los datos de los pacientes, asegurando que se respeten los derechos de privacidad y se minimicen los riesgos asociados con la recopilación y el uso de datos sensibles.

Referencias

- [1] G. A. Herrera Horta, G. L. Herrera Miranda, and Z. Gutiérrez García, "Calidad de vida de pacientes con enfermedad de parkinson que reciben tratamiento rehabilitador," *Interdisciplinary Rehabilitation / Rehabilitacion Interdisciplinaria*, vol. 3, 2023.
- [2] D. S. Marín, H. Carmona, M. Ibarra, and M. Gámez, "Enfermedad de parkinson: fisiopatología, diagnóstico y tratamiento," *Revista de*

- la Universidad Industrial de Santander. *Salud*, vol. 50, no. 1, pp. 79–92, 2018.
- [3] J. S. Saavedra Moreno, P. A. Millán, and O. F. Buritica Henao, “Introducción, epidemiología y diagnóstico de la enfermedad de parkinson,” *Acta Neurológica Colombiana*, vol. 35, no. 3 Supl 1, pp. 2–10, 2019.
- [4] J. Mei, C. Desrosiers, and J. Frasnelli, “Machine learning for the diagnosis of parkinson’s disease: A review of literature,” *Front Aging Neurosci*, vol. 13, p. 633752, 2021.
- [5] A. Ouhmida, A. Raihani, B. Cherradi, and O. Terada, “A novel approach for parkinson’s disease detection based on voice classification and features selection techniques,” *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 10, 2021.
- [6] A. Suppa, G. Costantini, F. Ascì, P. Di Leo, M. S. Al-Wardat, G. Di Lazzaro, S. Scalise, A. Pisani, and G. Saggio, “Voice in parkinson’s disease: A machine learning study,” *Front Neurol*, vol. 13, p. 831428, 2022.
- [7] S. Sharanyaa, P. N. Renjith, and K. Ramesh, “Classification of parkinson’s disease using speech attributes with parametric and nonparametric machine learning techniques,” pp. 437–442, 2020.
- [8] F. Maass, B. Michalke, D. Willkommen, A. Leha, C. Schulte, L. Tonges, B. Mollenhauer, C. Trenkwalder, D. Ruckamp, M. Borger, I. Zerr, M. Bahr, and P. Lingor, “Elemental fingerprint: Reassessment of a cerebrospinal fluid biomarker for parkinson’s disease,” *Neurobiol Dis*, vol. 134, p. 104677, 2020.
- [9] F. Segovia, J. M. Gorriz, J. Ramirez, F. J. Martinez-Murcia, and D. Castillo-Barnes, “Assisted diagnosis of parkinsonism based on the striatal morphology,” *Int J Neural Syst*, vol. 29, no. 9, p. 1950011, 2019.
- [10] J. P. Félix, F. H. T. Vieira, A. Cardoso, M. V. G. Ferreira, R. A. P. Franco, M. A. Ribeiro, S. G. Araújo, H. P. Corrêa, and M. L. Carneiro, “A parkinson’s disease classification method: An approach using gait dynamics and detrended fluctuation analysis,” in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–4.
- [11] A.-G. Andrei, A.-M. Tăuțan, and B. Ionescu, “Parkinson’s disease detection from gait patterns,” in *2019 E-Health and Bioengineering Conference (EHB)*, 2019, pp. 1–4.
- [12] C. Ricciardi, M. Amboni, C. De Santis, G. Ricciardelli, G. Improta, L. Iuppariello, G. D’Addio, P. Barone, and M. Cesarelli, “Classifying different stages of parkinson’s disease through random forests,” in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, J. Henriques, N. Neves, and P. de Carvalho, Eds. Cham: Springer International Publishing, 2020, pp. 1155–1162.
- [13] W. Wang, J. Lee, F. Harrou, and Y. Sun, “Early detection of parkinson’s disease using deep learning and machine learning,” *IEEE Access*, vol. 8, pp. 147 635–147 646, 2020.
- [14] J. Zhang, W. Zhou, H. Yu, T. Wang, X. Wang, L. Liu, and Y. Wen, “Prediction of parkinson’s disease using machine learning methods,” *Biomolecules*, vol. 13, no. 12, 2023.
- [15] M. Martinez-Eguiluz, O. Arbelaitz, I. Gurrutxaga, J. Muguerza, I. Perona, A. Murueta-Goyena, M. Acera, R. Del Pino, B. Tijero, J. C. Gomez-Esteban, and I. Gabilondo, “Diagnostic classification of parkinson’s disease based on non-motor manifestations and machine learning strategies,” *Neural Computing and Applications*, vol. 35, no. 8, pp. 5603–5617, 2022.
- [16] Z. Ou, J. Pan, S. Tang, D. Duan, D. Yu, H. Nong, and Z. Wang, “Global trends in the incidence, prevalence, and years lived with disability of parkinson’s disease in 204 countries/territories from 1990 to 2019,” *Front Public Health*, vol. 9, p. 776847.
- [17] S. A. Mostafa, A. Mustapha, M. A. Mohammed, R. I. Hamed, N. Arunkumar, M. K. Abd Ghani, M. M. Jaber, and S. H. Khaleefah, “Examining multiple feature evaluation and classification methods for improving the diagnosis of parkinson’s disease,” *Cognitive Systems Research*, vol. 54, pp. 90–99, 2019.
- [18] A. Rana, A. Dumka, R. Singh, M. K. Panda, and N. Priyadarshi, “A computerized analysis with machine learning techniques for the diagnosis of parkinson’s disease: Past studies and future perspectives,” *Diagnostics (Basel)*, vol. 12, no. 11, 2022.
- [19] B. E. B. D, V. K. Elumalai, and U. K., “Data-driven gait analysis for diagnosis and severity rating of parkinson’s disease,” *Medical Engineering Physics*, vol. 91, pp. 54–64, 2021.
- [20] Z. Karapinar Senturk, “Early diagnosis of parkinson’s disease using machine learning algorithms,” *Med Hypotheses*, vol. 138, p. 109603, 2020.

- [21] B. E., B. D., and B. R., “Supervised machine learning based gait classification system for early detection and stage classification of parkinson’s disease,” *Applied Soft Computing*, vol. 94, p. 106494, 2020.
- [22] S. Wang *et al.*, “Mri and pet data in diagnosis,” *NeuroImage: Clinical*, vol. 14, pp. 87–97, 2017.
- [23] C. O. Sakar *et al.*, “Analysis of voice features,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 884–890, 2013.
- [24] L. Ma *et al.*, “Voice analysis for early detection,” *Journal of Voice*, vol. 28, no. 4, pp. 465–471, 2014.
- [25] S. Nuvoli *et al.*, “Model ensembles for parkinson’s diagnosis,” *Computers in Biology and Medicine*, vol. 108, pp. 176–186, 2019.
- [26] D. A. Pisner and D. M. Schnyer, “Chapter 6 - support vector machine,” in *Machine Learning*, A. Mechelli and S. Vieira, Eds. Academic Press, 2020, pp. 101–121.
- [27] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [28] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, “Ai-based smart prediction of clinical disease using random forest classifier and naive bayes,” *The Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, 2020.
- [29] E. Avuclu and A. Elen, “Evaluation of train and test performance of machine learning algorithms and parkinson diagnosis with statistical measurements,” *Med Biol Eng Comput*, vol. 58, no. 11, pp. 2775–2788, 2020.
- [30] R. Priyadarshini, N. Dash, and R. Mishra, “A novel approach to predict diabetes mellitus using modified extreme learning machine,” in *2014 International Conference on Electronics and Communication Systems (ICECS)*, 2014, pp. 1–5.
- [31] S. Sankaranarayanan and P. Perumal T., “Diabetic prognosis through data mining methods and techniques,” in *2014 International Conference on Intelligent Computing Applications*, 2014, pp. 162–166.
- [32] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, “Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308–6325, 2020.
- [33] M. Pramanik, R. Pradhan, P. Nandy, A. K. Bhoi, and P. Barsocchi, “Machine learning methods with decision forests for parkinson’s detection,” *Applied Sciences*, vol. 11, no. 2, 2021.
- [34] E. Bharath and T. Rajagopalana, “Parkinson’s disease classification using random forest kerb feature selection,” *Intelligent Automation Soft Computing*, vol. 36, no. 2, pp. 1417–1433, 2023.
- [35] D. Moher, A. Liberati, J. Tetzlaff, D. Altman, and T. P. Group, “Preferred reporting items for systematic reviews and meta-analyses: The prisma statement,” *PLoS Med*, vol. 6, no. 7, p. e1000097, 2009.
- [36] M. Page, J. McKenzie, P. Bossuyt, I. Boutron, T. Hoffmann, C. Mulrow, L. Shamseer, J. Tetzlaff, E. Akl, S. Brennan, R. Chou, J. Glanville, J. Grimshaw, A. Hróbjartsson, M. Lalu, T. Li, E. Loder, E. Mayo-Wilson, S. McDonald, L. McGuinness, L. Stewart, J. Thomas, A. Tricco, V. Welch, P. Whiting, and D. Moher, “The prisma 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, 2021.
- [37] —, “Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews,” *BMJ*, vol. 372, p. n160, 2021.
- [38] A. P. S. de Oliveira, M. A. de Santana, M. K. S. Andrade, J. C. Gomes, M. C. A. Rodrigues, and W. P. dos Santos, “Early diagnosis of parkinson’s disease using eeg, machine learning and partial directed coherence,” *Research on Biomedical Engineering*, vol. 36, no. 3, pp. 311–331, 2020. [Online]. Available: <https://doi.org/10.1007/s42600-020-00072-w>
- [39] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, “A comparative analysis of speech signal processing algorithms for parkinson’s disease classification and the use of the tunable q-factor wavelet transform,” *Applied Soft Computing*, vol. 74, pp. 255–263, 2019.
- [40] S. Aich, H.-C. Kim, K. younga, K. L. Hui, A. A. Al-Absi, and M. Sain, “A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of parkinson’s disease,” pp. 1116–1121, 2019.
- [41] S. McGrath, X. Zhao, R. Steele, B. D. Thombs, and A. Benedetti, “Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis,” *Statistics in Medicine*, vol. 42, no. 2, pp. 155–170, 2023.

- [42] A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. Shahid, "Feature selection based on l1-norm support vector machine and effective recognition system for parkinson's disease using voice recordings," *IEEE Access*, vol. 7, pp. 37 718–37 734, 2019.
- [43] A. M. Elshewey, M. Y. Shams, N. El-Rashidy, A. M. Elhady, S. M. Shohieb, and Z. Tarek, "Bayesian optimization with support vector machine model for parkinson disease classification," *Sensors (Basel)*, vol. 23, no. 4, 2023.
- [44] O. Yaman, F. Ertam, and T. Tuncer, "Automated parkinson's disease recognition based on statistical pooling method using acoustic features," *Med Hypotheses*, vol. 135, p. 109483, 2020.
- [45] D. Sztahó, I. Valálík, and K. Vicsi, "Parkinson's disease severity estimation on hungarian speech using various speech tasks," in *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2019, pp. 1–6.
- [46] S. Tadse, M. Jain, and P. Chandankhede, "Parkinson's detection using machine learning," pp. 1081–1085, 2021.
- [47] S. S. Nayak, A. D. Darji, and P. K. Shah, "Identification of parkinson's disease from speech signal using machine learning approach," *International Journal of Speech Technology*, vol. 26, no. 4, pp. 981–990, 2023.
- [48] M. Ricci, G. Di Lazzaro, A. Pisani, N. B. Mercuri, F. Giannini, and G. Saggio, "Assessment of motor impairments in early untreated parkinson's disease patients: The wearable electronics impact," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 120–130, 2020.

6. Anexos

6.1. Anexo 1

Tabla 3. Estadísticas Descriptivas de las Métricas de Rendimiento

Métrica	Fórmula
Media de la Exactitud ($\mu_{\text{Exactitud}}$)	$\mu_{\text{Exactitud}} = \frac{1}{N} \sum_{i=1}^N \text{Exactitud}_i \quad (6)$ <p>donde Exactitud_i representa cada observación de exactitud y N es el número total de observaciones.</p>
Media de la Precisión ($\mu_{\text{Precisión}}$)	$\mu_{\text{Precisión}} = \frac{1}{N} \sum_{i=1}^N \text{Precisión}_i \quad (7)$ <p>donde Precisión_i representa cada observación de precisión y N es el número total de observaciones.</p>
Media de la Sensibilidad ($\mu_{\text{Sensibilidad}}$)	$\mu_{\text{Sensibilidad}} = \frac{1}{N} \sum_{i=1}^N \text{Sensibilidad}_i \quad (8)$ <p>donde Sensibilidad_i representa cada observación de sensibilidad y N es el número total de observaciones.</p>
Media de la Especificidad ($\mu_{\text{Especificidad}}$)	$\mu_{\text{Especificidad}} = \frac{1}{N} \sum_{i=1}^N \text{Especificidad}_i \quad (9)$ <p>donde Especificidad_i representa cada observación de especificidad y N es el número total de observaciones.</p>
Media del F1-Score ($\mu_{\text{F1-Score}}$)	$\mu_{\text{F1-Score}} = \frac{1}{N} \sum_{i=1}^N \text{F1-Score}_i \quad (10)$ <p>donde F1-Score_i representa cada observación de F1-Score y N es el número total de observaciones.</p>
Media del AUC (μ_{AUC})	$\mu_{\text{AUC}} = \frac{1}{N} \sum_{i=1}^N \text{AUC}_i \quad (11)$ <p>donde AUC_i representa cada observación de AUC y N es el número total de observaciones.</p>
Desviación Estándar de la Exactitud ($\sigma_{\text{Exactitud}}$)	$\sigma_{\text{Exactitud}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Exactitud}_i - \mu_{\text{Exactitud}})^2} \quad (12)$ <p>donde $\mu_{\text{Exactitud}}$ es la media de las observaciones de exactitud.</p>
Desviación Estándar de la Precisión ($\sigma_{\text{Precisión}}$)	$\sigma_{\text{Precisión}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Precisión}_i - \mu_{\text{Precisión}})^2} \quad (13)$ <p>donde $\mu_{\text{Precisión}}$ es la media de las observaciones de precisión.</p>

Métrica	Fórmula
Desviación Estándar de la Sensibilidad ($\sigma_{\text{Sensibilidad}}$)	$\sigma_{\text{Sensibilidad}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Sensibilidad}_i - \mu_{\text{Sensibilidad}})^2} \quad (14)$ <p>donde μ_{Sensib} es la media de las observaciones de sensibilidad.</p>
Desviación Estándar de la Especificidad ($\sigma_{\text{Especificidad}}$)	$\sigma_{\text{Especificidad}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Especificidad}_i - \mu_{\text{Especificidad}})^2} \quad (15)$ <p>donde μ_{Esp} es la media de las observaciones de especificidad.</p>
Desviación Estándar del F1-Score ($\sigma_{\text{F1-Score}}$)	$\sigma_{\text{F1-Score}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{F1-Score}_i - \mu_{\text{F1-Score}})^2} \quad (16)$ <p>donde $\mu_{\text{F1-Score}}$ es la media de las observaciones de F1-Score.</p>
Desviación Estándar del AUC (σ_{AUC})	$\sigma_{\text{AUC}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{AUC}_i - \mu_{\text{AUC}})^2} \quad (17)$ <p>donde μ_{AUC} es la media de las observaciones de AUC.</p>

6.2. Anexo 2

Plan de Investigación para la Identificación de la Enfermedad de Parkinson

1. Revisión Regular de las Últimas Investigaciones

Objetivo: Mantenerse actualizado con los avances recientes en el campo de la identificación de la enfermedad de Parkinson mediante técnicas de aprendizaje automático.

Frecuencia: Mensualmente

Acciones:

- Suscribirse a revistas científicas y conferencias relevantes (e.g., IEEE, Nature, PLOS ONE).
- Configurar alertas de Google Scholar para términos clave como "Parkinson's Disease Machine Learning", "SVM Parkinson Detection", etc.
- Participar en webinars y seminarios online relacionados con la bioinformática y el aprendizaje automático.
- Resumir los artículos relevantes y discutir sus hallazgos en reuniones periódicas con el equipo de investigación.

2. Actualización Continua de los Modelos y Algoritmos Utilizados

Objetivo: Asegurar que los modelos y algoritmos utilizados sean los más eficientes y actualizados.

Frecuencia: Trimestralmente

Acciones:

- Evaluar nuevas técnicas y algoritmos que aparezcan en la literatura científica.
- Implementar mejoras en los modelos existentes basadas en las nuevas investigaciones.
- Realizar experimentos comparativos entre los modelos actuales y las nuevas versiones para determinar su efectividad.
- Documentar los cambios y resultados obtenidos en un registro de actualizaciones.

3. Validación Periódica de las Soluciones contra Nuevos Datos

Objetivo: Garantizar la robustez y precisión de los modelos frente a nuevos conjuntos de datos.

Frecuencia: Semestralmente

Acciones:

- Recopilar nuevos datos de pacientes mayores de 60 años con diagnóstico de Parkinson y sin él.
- Dividir los datos en conjuntos de entrenamiento y prueba, asegurando la diversidad y representatividad de las muestras.
- Validar los modelos actuales utilizando estos nuevos datos y analizar su rendimiento.
- Ajustar los modelos en base a los resultados obtenidos para mejorar su precisión y sensibilidad.
- Presentar los hallazgos en reuniones del equipo y considerar su publicación en revistas científicas.

4. Implementación de un Ciclo de Retroalimentación

Objetivo: Incorporar un proceso iterativo de mejora continua.

Frecuencia: Continuamente

Acciones:

- Establecer un ciclo de retroalimentación donde los resultados de la validación se utilicen para refinar los modelos.
- Fomentar la colaboración con otros investigadores y profesionales del campo para recibir opiniones y sugerencias.
- Mantener una base de datos centralizada de resultados, cambios de modelo y análisis para facilitar el seguimiento y la revisión.

5. Documentación y Publicación de Resultados

Objetivo: Asegurar la transparencia y reproducibilidad de la investigación.

Frecuencia: Anualmente

Acciones:

- Redactar informes detallados sobre los avances, metodologías y resultados obtenidos.
- Publicar artículos en revistas científicas y presentaciones en conferencias para compartir los hallazgos con la comunidad académica.
- Actualizar una página web o repositorio de la investigación con los datos y modelos disponibles para otros investigadores.