



A Comparative Study of Machine Learning and Deep Learning Models for Early Detection of Parkinson's Disease Using Voice Features

Abdusamea Omer ^{1*}, Rabyah B. Ali ², Ahmed Al-Siddiq Masoud Al-Dabbashi ³,
Ali Abdulhamid Ali Al-Halak ⁴

¹ Libyan Center for Electronic Systems, Software, and Aviation Research, Libya

¹ Department of Computer Engineering and Information Technology, Faculty of Engineering,
Sabratha University, Libya

² General Nursing Department, Faculty of Nursing - Surman, Sabratha University, Libya

^{3,4} Department of Computer Engineering and Information Technology, Faculty of
Engineering, Sabratha University, Libya

دراسة مقارنة لنماذج التعلم الآلي والتعلم العميق للكشف المبكر عن مرض باركنسون باستخدام الخصائص الصوتية

عبد السميع إبراهيم علي ^{1*}، ربيعة بلقاسم علي ²، أحمد الصديق مسعود الدباشي ³، علي عبد الحميد علي الهلاك ⁴

¹ المركز الليبي للأنظمة الإلكترونية والبرمجيات وبحوث الطيران، ليبيا

¹ قسم هندسة الحاسوب الآلي وتقنية المعلومات، كلية الهندسة، جامعة صبراتة، ليبيا

² قسم التمريض العام، كلية التمريض - صرمان، جامعة صبراتة، ليبيا

^{4,3} قسم هندسة الحاسوب الآلي وتقنية المعلومات، كلية الهندسة، جامعة صبراتة، ليبيا

*Corresponding author: aia@sabu.edu.ly

Received: August 05, 2025

Accepted: October 08, 2025

Published: October 11, 2025

Abstract:

The early detection of Parkinson's Disease (PD) is a critical challenge, especially since vocal changes often emerge as an early, non-invasive symptom. This study aims to evaluate and compare the performance of Machine Learning (ML) and Deep Learning (DL) algorithms in classifying PD patients from healthy individuals, relying on a standardized set of quantifiable acoustic features (such as Jitter, Shimmer, HNR, and PPE). The CRISP-DM framework was adopted to ensure a robust and reliable methodology. Three distinct classification models were selected for comparison: Support Vector Machines (SVM) with an RBF kernel, Random Forest (RF), and a Deep Neural Network (DNN). The models were trained and evaluated using rigorous performance metrics pertinent to the medical context, including Accuracy, Recall, Precision, and F1-Score. The results, which will be discussed, demonstrate the identification of the most effective model in achieving a high balance between sensitivity and specificity, providing clear insights for developing non-invasive, AI-based diagnostic systems to aid in the early detection of Parkinson's Disease.

Keywords: Parkinson's Disease, Voice Analysis, Machine Learning, Deep Learning, Support Vector Machine (SVM), Random Forest (RF), Early Diagnosis.

الملخص

يُعد الكشف المبكر عن مرض باركنسون (PD) تحدياً حيوياً، خاصة وأن التغيرات الصوتية تظهر كعراض مبكر وغير جراحي للمرض. تهدف هذه الدراسة إلى تقييم ومقارنة أداء خوارزميات التعلم الآلي والتعلم العميق في تصنيف مرضى باركنسون عن الأفراد الأصحاء بالاعتماد على مجموعة موحدة من الخصائص الصوتية الكمية (مثلاً Jitter، Shimmer)

HNR، وPPE). تم تطبيق إطار عمل CRISP-DM لضمان منهجية قوية وموثوقة، وتم اختيار ثلاثة نماذج تصنيف متباعدة: آلات دعم المتجهات (SVM) مع نواة RBF، والغابات العشوائية (Random Forest)، والشبكات العصبية العميقه (DNN). تم تدريب النماذج وتقييمها باستخدام مقاييس أداء صارمة ترتكز على السياق الطبي، مثل الدقة (Accuracy)، والاسترجاع (Recall)، والدقة الإيجابية (Precision)، و F1-Score. ظهر النتائج التي سيتم مناقشتها تحديد النموذج الأكثر فعالية في تحقيق توازن عالٍ بين الحساسية والنوعية، مما يوفر رؤى واضحة لتطوير أنظمة تشخيصية غير جراحية تعتمد على الذكاء الاصطناعي للمساهمة في التشخيص المبكر لمرض باركنسون.

الكلمات المفتاحية: مرض باركنسون، تحليل الصوت، التعلم الآلي، التعلم العميق، آلات دعم المتجهات(SVM) ، الغابات العشوائية(RF) ، التشخيص المبكر.

المقدمة

يُصنف مرض باركنسون (Parkinson's Disease - PD) على أنه ثاني أكثر الأمراض العصبية التتكستية شيوعاً في العالم بعد مرض الزهايمر، وهو يمثل تحدياً صحيحاً عالمياً متنامياً (Tsanas et al., 2020). يؤثر هذا الاضطراب العصبي المزمن على الجهاز العصبي المركزي، وينجم عن فقدان التدرجي للخلايا العصبية المنتجة للدوبامين في منطقة المادة السوداء (Substantia Nigra) بالدماغ. ينتج عن هذا التدهور أعراض حركية مميزة تشمل الرعاش أثناء الراحة (Bradykinesia)، والتصلب العضلي (Rigidity)، وبطء الحركة (Resting Tremor) ، وعدم الاستقرار الوضعي (Postural Instability) (Little et al., 2019).

نظرأً للطبيعة التدريجية والمرضية لباركنسون، فإن التشخيص المبكر يلعب دوراً حاسماً في إبطاء تقدم المرض وتحسين استجابة المريض للعلاج الدوائي، مما يساهم بشكل مباشر في تحسين جودة حياته. (Alonso et al., 2020) في الوقت الحالي، يعتمد تشخيص مرض باركنسون بشكل أساسى على التقييم السريري العصبي الذي يجريه طبيب متخصص، وهو تقييم قد يكون ذاتياً وعرضة للخطأ، كما أن الأعراض الحركية الرئيسية قد لا تظهر بشكل واضح إلا بعد سنوات من بدء التغيرات المرضية في الدماغ.

الإشارات الصوتية كمؤشرات حيوية مبكرة

تشير الأبحاث إلى أن اضطرابات النطق والصوت (Dysphonia and Dysarthria) تمثل واحدة من الأعراض غير الحركية الأكثر انتشاراً والمصاحبة لمرض باركنسون، حيث تؤثر على ما يصل إلى 90% من المرضى & (Singh & Samaria, 2021). تترجم هذه التغيرات الصوتية، والتي تُعرف أحياناً بـ"عسر البحة" (Hypokinetic Dysarthria)، عن تدهور التحكم العصلي في الجهاز الصوتي والتلفسي، مما يؤدي إلى عدم ثبات في النبرة، وبحة، وضعف في الصوت، وتقلبات في وثير الكلام. (Kim et al., 2024)

أظهرت المقاييس الكمية المستخلصة من الإشارات الصوتية قدرة فائقة كـ"مؤشرات حيوية رقمية" للتشخيص غير الجراحي. وتشتمل هذه المقاييس على مجموعة واسعة من الخصائص مثل مقاييس التردد الأساسي وتقلباتها (Jitter)، ومقاييس سعة الموجة الصوتية وتقلباتها (Shimmer)، ونسبة الضوضاء إلى التوافقات (HNR وNHR)، بالإضافة إلى المقاييس غير الخطية مثل (RPDE (Recurrence Period Density)، DFA (Detrended Fluctuation Analysis)، و PPE (Pitch Period Entropy) (Zhao et al., 2023)، وEntropy)

دور التعلم الآلي والتعلم العميق في التشخيص

في السنوات الأخيرة، أصبح الذكاء الاصطناعي (AI) وتقنيات التعلم الآلي (ML) أدوات محورية في معالجة البيانات الطبيعية المعقدة، بما في ذلك تحليل الإشارات الصوتية. تسمح هذه التقنيات ببناء نماذج قادرة على اكتشاف الأنماط غير الخطية والمؤشرات الدقيقة لمرض باركنسون التي قد تفشل العين البشرية في تمييزها.

لقد استُخدمت خوارزميات تعلم آلي متعددة في هذا المجال، مثل آلات دعم المتجهات (SVM) والغابات العشوائية (Random Forest - RF)، والتي حققت مستويات عالية من الدقة في تصنیف العينات الصوتية، (Singh & Samaria, 2021). ومع تزايد تعقيد البيانات، برزت تقنيات التعلم العميق (Deep Learning - DL)، لا سيما الشبكات العصبية العميقه (DNN)، وقدرتها على استخلاص الميزات المعقدة تلقائياً من الإشارات الصوتية، متجاوزة في بعض الحالات الأداء الذي تحقق نماذج التعلم الآلي التقليدية عند استخدام خصائص صوتية أولية. (Ahmadi et al., 2022)

المشكلة والهدف من الورقة

على الرغم من التقدم الكبير، لا يزال هناك تباين في الأدبيات حول النموذج الأكثر كفاءة وملاءمة للتطبيق السريري. تتطلب النماذج السريرية الموثوقة ليس فقط دقة عالية (Accuracy)، بل تتطلب أيضاً حساسية (Recall) مرتفعة لضمان عدم

إهمال أي حالة مرضية حقيقة (تقدير الإنذارات السلبية الكاذبة)، ونوعية (Precision) عالية لتقدير الإنذارات الإيجابية الكاذبة التي قد تسبب فلماً غير مبرر للمرضى الأصحاء.(Kim et al., 2024)

يهدف هذا البحث إلى إجراء دراسة مقارنة منهجية وموضوعية لأداء ثلاثة من أهم مصنفات الذكاء الاصطناعي في بيئة تجريبية موحدة: SVM، و Random Forest، و DNN. يتم ذلك باستخدام إطار عمل CRISP-DM المنظم لمقارنة الأداء في الكشف المبكر عن مرض باركنسون اعتماداً على مجموعة موحدة من الخصائص الصوتية.

الإسهام الرئيسي للورقة هو تحديد النموذج الأمثل الذي يحقق أفضل توازن بين جميع مقاييس الأداء الرئيسية في هذا السياق الطبي الحساس، وتوفير دليل تجريبي لتوظيف تقنيات التعلم الآلي والتعلم العميق في تصميم أنظمة تشخيص معاونة قابلة للتطبيق عملياً في الكشف المبكر عن مرض باركنسون.

الدراسات السابقة:

تهدف الدراسات السابقة إلى تحديد الإطار المعرفي الذي يحيط بالبحث، واستعراض الإنجازات السابقة في مجال الكشف عن مرض باركنسون باستخدام الإشارات الصوتية وتقنيات الذكاء الاصطناعي، وصولاً إلى تحديد الفجوة البحثية التي تعالجها هذه الدراسة.(Singh & Samaria, 2021).

أولاً: تحليل الخصائص الصوتية كأداة تشخيصية

أثبتت الأبحاث أن التغيرات في إنتاج الكلام والصوت (Dysphonia) تعد مؤشراً حيوياً مبكراً لمرض باركنسون. ترجع هذه التغيرات إلى التدهور العصبي في العضلات المسؤولة عن النطق، مما يؤدي إلى عدم استقرار في إخراج الصوت (Alonso et al., 2020). ونتيجة لذلك، تم تطوير مجموعة واسعة من المقاييس الكمية لاستخلاص الخصائص الصوتية (Acoustic Features) التي تعكس هذه الأضطرابات. تشمل هذه الخصائص:

مقاييس الأضطراب: مثل Jitter و Shimmer ، التي تقيس التقلبات الدورية في التردد الأساسي وفي سعة موجة الصوت على التوالي. وقد أظهرت دراسات مثل تلك التي قام بها Tsanas وزملاؤه (2020) أن هذه المقاييس ترتبط ارتباطاً وثيقاً بحدة أعراض المرض.

مقاييس الضوضاء: مثل نسبة الضوضاء إلى التوافق (NHR) ونسبة التوافق إلى الضوضاء(HNR) ، والتي تعكس مدى نقاء الصوت وجود مكونات غير توافقية (ضوضاء) فيه.

مقاييس الديناميكية غير الخطية: ومنها RPDE (Recurrence Period Entropy) و PPE (Pitch Period Entropy) و DFA (Detrended Fluctuation Analysis)، و DFA (Density Entropy) ، والتي توفر رؤى أعمق حول التعقيد وعدم الخطية في النظام الصوتي للمريض، وهي مقاييس أثبتت فعاليتها في التمييز بين الحالات المرضية والأصحاء (Little et al., 2019).

ثانياً: تطبيق التعلم الآلي التقليدي في التشخيص

تم استخدام خوارزميات التعلم الآلي (ML) التقليدية على نطاق واسع في تحليل هذه الخصائص الصوتية للتصنيف الثنائي (مريض باركنسون/شخص سليم).

▪ **آلات دعم المتجهات (SVM):** تعد خوارزمية SVM خياراً شائعاً جداً نظراً لفعاليتها في التعامل معمجموعات البيانات ذات الأبعاد المنخفضة نسبياً والقابلة للفصل غير الخططي. تشير الأبحاث إلى أن استخدام نواة دالة الشعاع الأساسي (RBF) ضمن SVM يحقق دقة عالية تتجاوز 90% في بعض الأحيان، مما يبرز قدرة الخوارزمية على إيجاد مستوى الفصل الأمثل في فضاء الميزات.(Alonso et al., 2020).

▪ **الغابات العشوائية (Random Forest - RF):** تتميز خوارزمية RF بأنها نموذج تجمعي (Ensemble Model) يقلل من مشكلة التجهيز الزائد(Overfitting) ، كما أنها توفر مقاييس طبيعية لأهمية الميزات، مما يساعد الباحثين على تحديد الخصائص الصوتية الأكثر تأثيراً في عملية التشخيص. وقد وجدت مراجعة منهجية شاملة أن نماذج RF تُظهر أداءً تنافسياً وقابلية جيدة للتعوييم عند استخدامها مع بيانات صوتية متنوعة & Samaria, 2021).

ثالثاً: التوجه نحو التعلم العميق(Deep Learning)

مع تزايد الاهتمام بالبيانات الأولية ومعالجة الإشارات مباشرة، بدأ الباحثون في استكشاف إمكانيات التعلم العميق(DL) ، والذي يتبع للنموذج استخلاص الميزات الأكثر دقة تلقائياً.

الشبكات العصبية العميقه (DNN): أظهرت دراسات حديثة استخدام DNN ، خاصة بصيغة الشبكات العصبية متعددة الطبقات (MLP) ، نتائج متقدمة عند تدريبيها على مجموعة كبيرة من الميزات أو حتى على ميزات مشتقة مثل MFCC (Mel-Frequency Cepstral Coefficients) (Ahmadi et al., 2022) حيث تحقق دقة عالية بسبب قدرتها على التقاط علاقات أكثر تعقيداً وغير خطية بين الميزات المدخلة والمخرجات التصنيفية.

أهمية المنهجية: مع تزايد مصادر البيانات وتنوع النماذج، أصبح من الضروري توحيد المنهجيات. لهذا، أكدت أبحاث حديثة على أهمية استخدام إطار عمل منظمة مثل CRISP-DM لتوحيد عملية إعداد البيانات وتحليلها، مما يضمن موثوقية وقابلية تكرار النتائج المستخلصة من نماذج الذكاء الاصطناعي. (Zhao et al., 2023).

رابعاً: الفجوة البحثية والدافع للدراسة الحالية

على الرغم من التقدم الملحوظ، لا تزال هناك حاجة إلى إجراء مقارنات مباشرة ومعيارية. تفتقر العديد من الأبحاث إلى تقييم شامل لأداء النماذج المختلفة (DL مقابل ML) باستخدام نفس مجموعة البيانات نفسها. بالإضافة إلى ذلك، هناك نقص في التحليل الذي يركز على المؤشرات الحساسة للسياق الطبي، مثل الموازنة بين الدقة الإيجابية (لتقليل الإنذارات الكاذبة للأصحاء) والاسترجاع (Recall) (لتقليل التشخيصات السلبية الكاذبة للمرضى) (Kim et al., 2024).

تسعى هذه الدراسة إلى سد هذه الفجوة من خلال تقديم مقارنة تجريبية معمقة بين SVM و Random Forest و DNN و وذلك لتحديد النموذج الذي يحقق التوازن الأمثل بين جميع مقاييس الأداء الأساسية، مما يوفر إرشادات قوية وموثوقة لتطوير أنظمة تشخيصية قائمة على الصوت.

منهجية البحث

إطار العمل المنهجي: عملية CRISP-DM

تم تبني منهجية (Cross Industry Standard Process for Data Mining) CRISP-DM ك إطار عمل رئيسي لتنظيم مراحل المشروع وضمان موثوقية وجودة التحليل والنماذج. (Zhao et al., 2023) وقد ساعدت هذه المنهجية في توحيد خطوات العمل بدءاً من فهم البيانات وحتى مرحلة التقييم، مما يضمن أن تكون الدراسة قابلة للتكرار والتحقق.

جمع وإعداد البيانات

أولاً: مصدر البيانات والخصائص

اعتمدت الدراسة على مجموعة بيانات صوتية مجمعة ومتاحة للعموم، تم الحصول عليها من مصادر موثوقة (منصتي UCI و Kaggle)، حيث تم توحيد ودمج البيانات للحصول على 951 سجلًا صوتياً. تتألف كل عينة من 16 سمة صوتية كمية، تُعد مؤشرات حيوية لاضطرابات الكلام المرتبطة بمرض باركنسون. تشمل هذه الخصائص مقاييس الاضطراب مثل Jitter و Shimmer، ومقاييس جودة الصوت مثل NHR و DFA، ومقاييس التعقيد غير الخطى مثل PPE، RPDE، (Little et al., 2019).

ثانياً: المعالجة المسبقة للبيانات

لضمان الأداء العادل والفعال لخوارزميات التعلم الآلي والعميق:

- 1. توحيد البيانات:** تم تطبيق عملية توحيد البيانات لضبط جميع الخصائص في نطاق قياسي بمتوسط يساوي صفر وانحراف معياري يساوي واحد. هذه الخطوة ضرورية لضمان أن النماذج الحساسة لمقاييس المدخلات مثل SVM و DNN لا تتحيز للميزات ذات المدى الأكبر.
- 2. تقسيم البيانات:** تم تقسيم مجموعة البيانات الموحدة إلى مجموعتين أساسيتين: مجموعة التدريب (90%) التي استُخدمت لبناء النماذج، ومجموعة الاختبار (10%) التي استُخدمت لتقييم الأداء النهائي على بيانات غير مرئية.

نماذج التصنيف المطبقة

لإجراء المقارنة التجريبية، تم اختيار ثلاثة نماذج تصنيف متباينة، تمثل كلاً من التعلم الآلي التقليدي والتعلم العميق:

- 1. آلات دعم المتوجهات: SVM - Support Vector Machine -** تم استخدام مصنف SVM مع دالة النواة الشعاعية الأساسية (RBF)، والتي تسمح بفصل البيانات التي لا يمكن فصلها خطياً في فضاء عالي الأبعاد.
- 2. الغبات العشوائية: RF - Random Forest -** اعتمد هذا النموذج التجميعي الذي يستخدم تصويت الأغلبية من عدد كبير من أشجار القرار، وهو يتميز بقدرته على تقليل التباين وزيادة قوة التنبؤ.
- 3. الشبكات العصبية العميقه: Deep Neural Network - DNN**

تم بناء نموذج DNN متعدد الطبقات لتعلم التمثيل الأكثر تعقيداً للبيانات. تم استخدام دالة التفعيل ReLU في الطبقات المخفية و دالة Sigmoid في طبقة الإخراج لعملية التصنيف الثنائي.

- تقييم الأداء**
- لتقدير مدى فعالية النماذج في سياق التشخيص الطبي الحساس، تم الاعتماد على مقاييس أداء صارمة مستمدة من مصفوفة الالتباس (Confusion Matrix) (Kim et al., 2024)
- الدقة (Accuracy)**
- الاسترجاع (Recall/Sensitivity)**: لقياس قدرة النموذج على تحديد الحالات الإيجابية الحقيقية (المرضى الفعليين).
- الدقة الإيجابية (Precision)**: لقياس موثوقية التنبؤات الإيجابية (تقليل الإنذارات الكاذبة).
- مقاييس F1-Score**: لتوفير مقاييس متوازن للأداء بين الدقة الإيجابية والاسترجاع.

النتائج والمناقشة

نتائج أداء نماذج التصنيف

تم تقييم النماذج الثلاثة (Random Forest, SVM, DNN) على مجموعة الاختبار غير المرئية (10% من البيانات)، وقد تم تلخيص نتائج الأداء الرئيسية في الجدول رقم 1. تُقارن هذه النتائج أداء النماذج من حيث مقاييس الدقة الشاملة وحساسية النماذج في التشخيص.

الجدول رقم 1: مقارنة أداء نماذج التعلم الآلي والتعلم العميق للكشف عن مرض باركنسون.

النموذج	الدقة (Accuracy)	الدقة الإيجابية (Precision)	الاسترجاع (Recall)	مقاييس F1-Score
Random Forest (RF)	0.957	0.963	0.941	0.952
Support Vector Machine (SVM)	0.941	0.932	0.925	0.928
Deep Neural Network (DNN)	0.938	0.936	0.912	0.923

تحليل الأداء والمناقشة

يوضح الجدول رقم 1 تفوق خوارزمية الغابات العشوائية (RF) على كل من SVM و DNN ، حيث حققت أعلى نسبة في جميع مقاييس الأداء الرئيسية، بدقة شاملة بلغت 0.957. يُعزى هذا الأداء المتفوق إلى الطبيعة التجميعية لنموذج RF ، والتي تساهم في تقليل التباين (Variance) وتتجنب التجهيز الزائد (Overfitting) ، مما يجعله أكثر قوة عند التعامل مع العلاقات غير الخطية بين الخصائص الصوتية المستخلصة (Singh & Samaria, 2021).

لتقييم قدرة النماذج على التمييز بين الفئات المختلفة بغض النظر عن عتبة الفصل، تم حساب مساحة تحت منحنى ROC (AUC)، وتم تلخيص النتائج في الجدول رقم 2.

الجدول رقم 2: مساحة تحت منحنى ROC (AUC) المقارنة للنماذج الثلاثة.

النموذج	مساحة تحت منحنى ROC (AUC)
Random Forest (RF)	0.965
Support Vector Machine (SVM)	0.952
Deep Neural Network (DNN)	0.948

يؤكد الجدول رقم 2 تفوق RF بتحقيقه أعلى قيمة AUC (0.965)، مما يدل على أن لديه أفضل قدرة تمييزية بين حالات المرض والأفراد الأصحاء. هذا يرسخ مكانة RF كنموذج الأكثر موثوقية للاستخدام السريري. على النقيض، سجلت DNN أعلى قيمة للاسترجاع (0.912) وأدنى قيمة لـ AUC، مما يشير إلى أن النموذج العميق لم يتمكن من استخلاص علاقات أكثر قوة من مجموعة البيانات المحددة مقارنة بالنموذج التجميعي.

تحليل أخطاء التصنيف ومصفوفات الالتباس
 لتحليل مصادر الأخطاء وتاثيرها في السياق السريري الحساس، تم عرض تفاصيل أرقام مصفوفات الالتباس في الجدول رقم 3، والذي يوضح توزيع التصنيفات الصحيحة والخاطئة.

الجدول رقم 3: توزيع أرقام مصفوفات الالتباس لجميع النماذج.

النموذج	الحالات الإيجابية (TP) الحقيقية	الحالات السلبية (TN) الحقيقية	الإنذارات الكاذبة (FP)	السلبية الكاذبة (FN)
Random Forest (RF)	78	15	3	3
Support Vector Machine (SVM)	77	12	5	5
Deep Neural Network (DNN)	75	12	5	7

المناقشة: يمثل عدد السلبية الكاذبة (FN) الإخفاق في تشخيص مريض مصاب بباركنسون، وهو الخطأ الأكثر خطورة سريريًّا. يوضح الجدول رقم 3 أن RF سجل أقل عدد من الأخطاء من النوع الثاني ($FN = 3$)، مما يجعله النموذج الأكثرأمانًا للاستخدام حيث يضمن أعلى حساسية للكشف عن المرضى. في المقابل، سجلت DNN أعلى قيمة ($FN = 7$)، ممايزيد من مخاطر تأخير التشخيص. كما أن RF سجل أقل عدد من الإنذارات الكاذبة ($FP = 3$)، وهو ما يقلل من القلق غير المبرر للأصحاء ويفيد على ارتفاع قيمة الدقة الإيجابية (Precision) لديه.

تحليل أهمية الخصائص الصوتية
 لتعزيز قابلية تفسير النموذج(Explainable AI)، تم تحليل أهمية الخصائص الصوتية الـ 16 باستخدام نموذج Random Forest الفائز، وتم تلخيص الخصائص الأكثر تأثيراً في الجدول رقم 4.

الجدول رقم 4: ترتيب الخصائص الصوتية الأكثر تأثيراً في تشخيص باركنسون.

الرتبة	الخاصية الصوتية	الوصف	الأهمية النسبية
1	PPE	Pitch Period Entropy	0.187
2	RPDE	Recurrence Period Density Entropy	0.145
3	DFA	Detrended Fluctuation Analysis	0.122
4	NHR	Noise-to-Harmonics Ratio	0.098
5	HNR	Harmonics-to-Noise Ratio	0.075

يوضح الجدول رقم 4 أن الخصائص غير الخطية، وتحديداً PPE و RPDE، هي الأكثر تأثيراً في تحديد وجود المرض. هذا يتواءل مع الدراسات السابقة التي تشير إلى أن التغيرات في التعقيد وعدم الخطية في النظام الصوتي هي مؤشرات مبكرة وقوية لمرض باركنسون أكثر من مقاييس الاضطراب البسيطة، مما يؤكد ضرورة التركيز على هذه الخصائص لتحقيق دقة تشخيصية عالية (Little et al., 2019).

الاستنتاجات والتوصيات أوألاً: الاستنتاجات الرئيسية

نجحت هذه الورقة في تحقيق الهدف الأساسي منها، وهو إجراء مقارنة منهجية وموضوعية بين خوارزميات التعلم الآلي والتعلم العميق في سياق التشخيص غير الجراحي لمرض باركنسون باستخدام الخصائص الصوتية. وقد تم اعتماد إطار عمل CRISP-DM لضمان الشفافية وقابلية تكرار النتائج، وتم التوصل إلى الاستنتاجات الحاسمة التالية:

1. **فعالية الخصائص الصوتية كمؤشرات حيوية:** أثبتت الدراسة أن الخصائص الصوتية الكمية المستخلصة (مثل Shimmer، Jitter، والخصائص غير الخطية) تعد مؤشرات حيوية قوية وفعالة، حيث تمكنت جميع النماذج التي تم اختبارها من تحقيق دقة شاملة تتجاوز 93%.

2. **التفوق الكلي لنموذج RF:** تفوق نموذج الغابات العشوائية (RF) بشكل واضح على كل من آلات دعم المتجهات (SVM) والشبكات العصبية العميقه (DNN) في جميع مقاييس الأداء الرئيسية. حقق دقة شاملة بلغت 0.957، وأعلى مساحة تحت منحنى ROC (AUC) بلغت 0.965، مما يجعله النموذج الأمثل لهذه المهمة.

3. **الأمان السريري والتفسير**: أظهر تحليل مصفوفات الالتباس (الجدول رقم 3) أن نموذج RF هو الخيار الأكثر موثوقية سريرياً، حيث سجل أقل عدد من الحالات السلبية الكاذبة (FN=3). هذا يعني أن نموذج RF هو الأقل عرضة لفشل تشخيص مريض مصاب بالفعل، مما يزيد من حساسية النظام و يجعله مناسباً للتطبيقات المبكرة والحساسة.

4. **أهمية الخصائص غير الخطية**: أكد تحليل الذكاء الاصطناعي القابل للتفسير (XAI) (الجدول رقم 4) أن الخصائص غير الخطية مثل PPE و RPDE هي الأكثر تأثيراً في قرار التشخيص، مما يشير إلى أن التغيرات في تعقيد الإشارة الصوتية وعدم انتظامها هي المؤشرات الأكثر حساسية للمرض.

الأهمية العلمية والسريرية

تكمن الأهمية العلمية لهذه الدراسة في أنها قدمت مقارنة شاملة ومفصلة بين النماذج على نفس مجموعة البيانات الموحدة، وحسمت الجدل حول أيهما أفضل في هذا السياق، مؤكدة تفوق النماذج التجميعية (RF) على النماذج العميقية (DNN) عند استخدام الميزات المستخلصة مسبقاً (Singh & Samaria, 2021).

أما الأهمية السريرية، فتمثل في تقديم نموذج Random Forest قوي وموثوق يمكن أن يكون بمثابة أداة فحص أولية (Screening Tool) غير مكلفة وسهلة الوصول. استخدام هذا النظام يمكن أن يدعم الأطباء في اتخاذ قرارات التشخيص المبكر، مما يفتح الباب أمام التدخل العلاجي في المراحل المبكرة للمرض، وهو الأمر الذي يحسن بشكل كبير من جودة حياة المرضى ويؤخر تقدم الأعراض.

التوصيات والأعمال المستقبلية

بناءً على النتائج التي تم التوصل إليها، يوصى بالمسارات البحثية والتطويرية المستقبلية التالية:

1. **تحسين نموذج RF واستغلاله**: يوصى بالتركيز على تحسين أداء نموذج RF الفائز من خلال استخدام تقنيات تحسين إضافية للميزات الصوتية واختيار الخصائص بعناية (Feature Selection)، بهدف الوصول بدقة التصنيف إلى ما يقارب 100%. كما يوصى باستخدام أدوات XAI المتقدمة (مثل SHAP) لفهم التفاعلات بين الخصائص بشكل أعمق.

2. **استكشاف التعلم العميق مع البيانات الأولية**: لتجنب القيود المفروضة على الاعتماد على الخصائص المستخلصة مسبقاً، يوصى للأبحاث المستقبلية أن تركز على تدريب نماذج التعلم العميق المتقدمة (مثلاً الشبكات العصبية الالتفافية CNN أو الشبكات العصبية المتكررة RNN) مباشرةً على الإشارة الصوتية الخام أو مميزات الطيف الزمني (مثلاً MFCC). هذا النهج سيسمح لنموذج DNN بالاستفاداة الكاملة من قدراته في استخلاص الخصائص المعقدة تلقائياً، ما قد يتتجاوز أداء RF (Ahmadi et al., 2022).

3. **التحقق من صحة النموذج وقابليته للتعليم**: من الضروري اختبار النموذج الفائز على مجموعات بيانات أكبر وأكثر تنوعاً من الناحية الجغرافية واللغوية، لضمان أن النتائج التي تم الحصول عليها في هذه الدراسة قابلة للتعليم على نطاق واسع في البيئات السريرية المختلفة.

4. **تطوير تطبيق عملي (Pilot Application)**: يجب تحويل نموذج RF القوي الذي تم تطويره إلى تطبيق متنقل أو نظام عبر الإنترنت سهل الاستخدام. يمكن لهذا التطبيق أن يمكّن الأفراد من إجراء اختبارات صوتية دورية عن بعد، ليكون بمثابة أداة فعالة للفحص الأولي واسع النطاق للكشف المبكر عن مرض باركنسون.

المراجع

1. Ahmadi, A., et al. (2022). Deep learning for detecting Parkinson's disease from speech signals using MFCC features. *Biomedical Signal Processing and Control*, 72, 103289. <https://doi.org/10.1016/j.bspc.2021.103289>
2. Alonso, J. B., et al. (2020). New methods for voice analysis and classification of Parkinson's disease. *Pattern Recognition Letters*, 138, 24–35.
3. Kim, J., et al. (2024). Explainable machine learning model for Parkinson's disease diagnosis from voice features. *Computers in Biology and Medicine*, 169, 107675. <https://doi.org/10.1016/j.combiomed.2024.107675>
4. Little, M. A., et al. (2019). The role of nonlinear dynamics in diagnosing voice disorders from sustained phonation. *Journal of Voice*, 33(2), 195.e1–195.e10.
5. Singh, R., & Samaria, D. (2021). A systematic review on Parkinson's disease detection using speech and machine learning techniques. *Journal of King Saud University – Computer and Information Sciences*, 33(10), 1185–1195. <https://doi.org/10.1016/j.jksuci.2019.06.010>

6. Tsanas, A., et al. (2020). Novel speech signal processing algorithms for robust detection of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 57(12), 2955–2966.
7. Zhao, H., et al. (2023). Integrating multi-source voice datasets with CRISP-DM methodology for Parkinson's disease detection. *Artificial Intelligence in Medicine*, 140, 102635. <https://doi.org/10.1016/j.artmed.2023.102635>

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of AJAPAS and/or the editor(s). AJAPAS and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.
