Journal of Libyan Academy Bani Walid 2025
مجلة الأكاديمية الليبية بني وليد
e-ISSN: 3104-3860
Volume 1, Issue 4 , 2025, Pages: 01-08
Website: https://journals.labjournal.ly/index.php/Jlabw/index

A Comprehensive Methodology for Data Preparation and Feature Evaluation in Parkinson's Disease Detection: Applying the CRISP-DM Framework to Voice Signal Analysis

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
*Corresponding author: aia@sabu.edu.ly

المنهجية الشاملة لتحضير البيانات وتقييم الخصائص في الكشف عن مرض باركنسون: تطبيق إطار عمل CRISP-DM على تحليل الإشارات 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
[1] قسم هندسة الحاسب الآلي وتقنية المعلومات، كلية الهندسة، جامعة صبراتة، ليبيا
[2] قسم التمريض العام، كلية التمريض - صرمان، جامعة صبراتة، ليبيا
[4،3] قسم هندسة الحاسب الآلي وتقنية المعلومات، كلية الهندسة، جامعة صبراتة، ليبيا

Received: 21-07-2025	Accepted: 18-09-2025	Published: 12-10-2025
Copyright: © 2025 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract
his study aims to present a detailed and reliable methodology for preparing and analyzing acoustic data for the early detection of Parkinson's Disease (PD). The research adopted the standard CRISP-DM framework to structure the workflow, starting from integrating a multi-source voice dataset, through essential preprocessing steps such as standardization and data splitting. To overcome the "black box" problem and enhance clinical trust, the research focused on applying Explainable AI (XAI) principles. The winning Random Forest (RF) model from the initial study was utilized to determine and assess the relative importance of 16 acoustic features. The analysis demonstrated that non-linear features, specifically PPE (Pitch Period Entropy) and RPDE (Recurrence Period Density Entropy), are the most influential factors in the diagnostic decision. This confirms that changes in voice signal complexity are powerful biomarkers for the disease. This work provides a clear, systematic, and interpreted roadmap, recommending prioritizing these features for developing effective and safe automated diagnostic systems.

Keywords: Parkinson's Disease, Voice Signals, Machine Learning, CRISP-DM Methodology, Explainable AI (XAI), Feature Importance, PPE, RPDE.

الملخص

يهدف هذا البحث إلى تقديم منهجية تفصيلية وموثوقة لتحضير وتحليل البيانات الصوتية للكشف المبكر عن مرض باركنسون (PD). اعتمدت الدراسة على إطار العمل القياسي CRISP-DM لتنظيم مراحل العمل، بدءاً من دمج مجموعة بيانات صوتية من مصادر متعددة، مروراً بمراحل المعالجة المسبقة الضرورية مثل توحيد البيانات (Standardization) وتقسيمها لاختبار النماذج. ولتجاوز مشكلة "الصندوق الأسود" وتعزيز الثقة السريرية، ركز البحث على تطبيق مبادئ الذكاء الاصطناعي القابل للتفسير (XAI). تم استخدام نموذج الغابات العشوائية (RF) الفائز من الدراسة الأولى لتحديد وتقييم الأهمية النسبية لـ 16 خاصية صوتية. أظهر التحليل أن الخصائص غير الخطية، وتحديداً PPE (انتروبيا فترة النبرة) و RPDE (كثافة فترة التكرار)، هي الأكثر تأثيراً في قرار التشخيص، مما يؤكد أن التغيرات في تعقيد الإشارة الصوتية هي مؤشرات حيوية قوية للمرض. يقدم هذا العمل خريطة طريق منهجية واضحة ومفسّرة، توصي بالتركيز على هذه الخصائص كأولوية قصوى لتطوير أنظمة تشخيصية آلية فعالة وآمنة.

الكلمات المفتاحية: مرض باركنسون، الإشارات الصوتية، التعلم الآلي، منهجية CRISP-DM، الذكاء الاصطناعي القابل للتفسير (XAI)، أهمية الخصائص، PPE، RPDE.

المقدمة

الخلفية: التحدي التشخيصي لمرض باركنسون

يُعد مرض باركنسون (Parkinson's Disease - PD) ثاني أكثر الأمراض العصبية التنكسية شيوعاً في العالم، ويمثل تحدياً صحياً متزايداً مع ارتفاع متوسط الأعمار (Tsanas et al., 2020). ينجم المرض عن الفقدان التدريجي للخلايا العصبية المنتجة للدوبامين، وتتطور أعراضه ببطء وتدريج. ونظراً لأن التشخيص السريري التقليدي لا يمكن تأكيده بشكل قاطع إلا بعد ظهور الأعراض الحركية المتقدمة، فإن الحاجة إلى أدوات كشف مبكرة وغير جراحية أصبحت ضرورية لتحسين جودة حياة المريض وتحديد التدخل العلاجي في أنسب الأوقات.

برزت الإشارات الصوتية كـ "مؤشر حيوي رقمي "فعال للكشف عن المرض، حيث يعاني ما يصل إلى 90% من مرضى باركنسون من اضطرابات النطق والصوت(Dysphonia) ، الناتجة عن تدهور التحكم العضلي في الجهاز الصوتي والتنفسي. يمكن قياس هذه الاضطرابات كمياً من خلال استخلاص خصائص صوتية دقيقة تتجاوز قدرة الأذن البشرية على التمييز.(Little et al., 2019)

الفجوة المنهجية والحاجة إلى التنظيم (CRISP-DM)

على الرغم من النجاحات الباهرة التي حققتها خوارزميات التعلم الآلي (ML) في تصنيف الإشارات الصوتية، لا تزال هناك فجوة كبيرة في المنهجيات المتبعة. تفتقر العديد من الأبحاث المنشورة في هذا المجال إلى إطار عمل منظم يضمن الشفافية وقابلية التكرار في مراحل المشروع. إن عملية بناء نموذج ذكي لا تقتصر على اختيار الخوارزمية الأفضل فحسب، بل تبدأ من كيفية تجميع البيانات، ومعالجتها، وتقسيمها بطريقة تضمن الحيادية والفعالية.

لذلك، فإن تبني منهجية CRISP-DM (Cross Industry Standard Process for Data Mining) يصبح أمراً بالغ الأهمية. توفر CRISP-DM إطاراً صارماً ومُحدداً لجميع المراحل، بدءاً من فهم العمل وانتهاءً بالتقييم، مما يضمن معالجة متسقة للبيانات المُجمعة من مصادر مختلفة ويمنع التحيز المنهجي الذي قد يؤدي إلى نتائج غير موثوقة (Zhao et al., 2023). تعد هذه الدراسة امتداداً لضرورة تطبيق هذه المنهجية لتوحيد عملية البحث وتجهيز البيانات.

مشكلة "الصندوق الأسود" ومتطلبات الذكاء الاصطناعي القابل للتفسير(XAI)

تمثل مشكلة "الصندوق الأسود (Black Box) "تحدياً كبيراً لقبول نماذج الذكاء الاصطناعي في البيئات السريرية. لا يكفي أن يحقق النموذج دقة عالية؛ يجب على الطبيب أن يفهم لماذا اتخذ النموذج قراراً معيناً، وعلى أي أساس علمي استند. هذا هو جوهر مفهوم الذكاء الاصطناعي القابل للتفسير(Explainable AI - XAI).

في سياق التشخيص، يعزز XAI ثقة المستخدم والمساءلة، حيث يسمح بتتبع قرار النموذج وتحديد الخصائص الصوتية التي ساهمت بأكبر قدر في تصنيف الحالة كـ "مريض باركنسون" (Kim et al., 2024). إن تحديد أهمية الخصائص (Feature Importance)ليس مجرد خطوة تقنية، بل هو خطوة ضرورية لفهم الآليات البيولوجية الكامنة وراء التدهور الصوتي المرتبط بالمرض. وبناءً على نتائج الدراسة السابقة التي أكدت تفوق نموذجRandom Forest ، يصبح تحليل أهمية الخصائص باستخدام هذا النموذج أمراً حتمياً.

الهدف البحثي والمساهمات

في ضوء ما سبق، تسعى هذه الورقة البحثية إلى تجاوز المقارنات السطحية بين الخوارزميات، والتركيز بدلاً من ذلك على تأسيس منهجية سليمة ومفسّرة لعملية اكتشاف مرض باركنسون.

تتمثل المساهمات الرئيسية لهذه الورقة في:

1. التطبيق المنهجي لإطار CRISP-DM: توضيح مراحل تطبيق منهجية CRISP-DM بشكل مفصل، مع التركيز على عمليات تحضير البيانات المُجمعة لضمان سلامة النماذج وجودة المدخلات.
2. التحليل العميق لأهمية الخصائص (XAI): استخدام تقنيات الذكاء الاصطناعي القابل للتفسير لتحديد وترتيب الأهمية النسبية لـ 16 خاصية صوتية قياسية، بهدف تحديد الخصائص غير الخطية الأكثر تأثيراً في التشخيص (مثل PPE و RPDE).
3. توفير خريطة طريق منظمة: تقديم دليل إجرائي واضح يمكن للباحثين والمطورين في المجال الطبي اتباعه لبناء أنظمة تشخيصية آلية تتسم بالموثوقية وقابلية التفسير السريري.

وبالتالي، توفر هذه الدراسة ليس فقط نموذجاً فعالاً، بل توفر أساساً منهجياً لتفسير النتائج وزيادة الاعتماد على حلول الذكاء الاصطناعي في المجال الصحي.

الدراسات السابقة

تهدف هذه الدراسات إلى توفير الأساس النظري وتبرير الأدوات المنهجية والتحليلية التي تم استخدامها في الدراسة، خاصة فيما يتعلق بضمان جودة البيانات وقابلية تفسير نتائج التعلم الآلي في السياق السريري.

المنهجية المنظمة: ضرورة تطبيقCRISP-DM

في مجال التنقيب عن البيانات والتعلم الآلي، خاصة في التطبيقات الحساسة مثل الرعاية الصحية، لا يمكن الاكتفاء بالاختيار العشوائي للخوارزميات. هناك حاجة ماسة إلى إطار عمل منظم يضمن الشمولية، قابلية التكرار، والشفافية في جميع مراحل المشروع. وقد أظهرت الأبحاث أن منهجية CRISP-DM (Cross Industry Standard Process for Data Mining)هي الإطار المعياري الأكثر اعتماداً في الصناعة والأوساط الأكاديمية.(Zhao et al., 2023)

يكمن الدور المحوري لـ CRISP-DM في:

توحيد المراحل: تقسيم الدراسة إلى مراحل واضحة (فهم العمل، فهم البيانات، إعداد البيانات، النمذجة، التقييم، والتنفيذ)، مما يقلل من الغموض ويسهل إدارة المشاريع المعقدة.

دمج البيانات المتعددة المصادر: عند التعامل مع بيانات صوتية مُجمعة من مصادر مختلفة (مثل UCI و Kaggle)، تضمن CRISP-DM تطبيق معايير متسقة لتوحيد البيانات وتحقيق الاتساق المطلوب قبل مرحلة النمذجة.

التركيز على فهم الأعمال: تفرض المنهجية أن يكون الهدف النهائي للنموذج متوافقاً مع الاحتياجات السريرية الفعلية، مثل ضمان أن يكون النموذج حذراً بشأن الأخطاء من النوع الثاني (False Negatives) لتقليل إخفاق التشخيص.

إن اعتماد CRISP-DM في هذه الدراسة يبرر عملية المعالجة المسبقة للبيانات الشاملة ويؤكد على أن المنهجية المتبعة ليست مجرد تطبيق تقني، بل هي عملية تحليلية منظمة.

أهمية المعالجة المسبقة للبيانات وتحليل الخصائص

تُعد مرحلة إعداد البيانات (Data Preparation) التي تشمل التوحيد والمعالجة المسبقة هي المرحلة الأكثر استهلاكاً للوقت والأكثر أهمية في عملية CRISP-DM (Zhao et al., 2023). وفي سياق الإشارات الصوتية، تصبح هذه المرحلة حاسمة للأسباب التالية:

توحيد البيانات (Standardization): تتسم الخصائص الصوتية المستخلصة (مثل Jitter و Shimmer و NHR) بوجود مقاييس مختلفة وقيم متباينة. إن توحيد البيانات لمتوسط صفر وانحراف معياري واحد ضروري لضمان أن النماذج الحساسة للمقاييس (مثل SVM و DNN) تعطي وزناً متساوياً لجميع الخصائص ولا تتحيز للميزات ذات المدى الأكبر.

الخصائص غير الخطية كمؤشرات حيوية: أشارت الأبحاث الحديثة إلى أن المقاييس التقليدية لاضطراب النبرة (Jitter/Shimmer) قد لا تكون كافية للكشف المبكر. لذلك، تحول الاهتمام نحو الخصائص غير الخطية وخصائص التعقيد مثل PPE، RPDE، و DFA (Little et al., 2019). هذه الخصائص توفر رؤى أعمق حول عدم انتظام النظام الصوتي الناتج عن التدهور العصبي، مما يبرر ضرورة تحليل أهميتها بشكل منفصل.

الأساس النظري للذكاء الاصطناعي القابل للتفسير (XAI) وتحليل الأهمية

مع تزايد دقة نماذج التعلم الآلي، تزداد أيضاً حاجتها إلى القابلية للتفسير. تعالج تقنيات الذكاء الاصطناعي القابل للتفسير (XAI)مشكلة "الصندوق الأسود" لضمان أن قرارات النموذج يمكن تبريرها ومنطقية بالنسبة للمختصين السريريين (Kim et al., 2024).

في التشخيص الطبي، يخدم تحليل أهمية الخصائص (Feature Importance) غرضين رئيسيين:
تعزيز الثقة والمساءلة: يوفر تفسيراً مباشراً لسبب تصنيف النموذج لحالة ما، مما يسمح للأطباء بالثقة في النظام واتخاذ قرار قائم على البيانات.(Kim et al., 2024)
الرؤية البيولوجية: تحديد الخصائص التي تحمل أعلى قيمة تنبؤية (مثل PPE و RPDE) يوجه الأبحاث المستقبلية نحو الآليات البيولوجية الأكثر تأثراً بالمرض، مما يمكن من تطوير أدوات تشخيصية أكثر تركيزاً وفعالية. إن استخدام نماذج مثل Random Forest في هذا السياق مفيد بشكل خاص، لأنه يوفر مقياساً داخلياً وموثوقاً لأهمية الخصائص دون الحاجة إلى أدوات XAI خارجية أكثر تعقيداً.

منهجية البحث

أولاً: إطار العمل المنهجي: عملية CRISP-DM

تم اعتماد منهجية CRISP-DM (Cross Industry Standard Process for Data Mining) كإطار عمل رئيسي لتنظيم مراحل البحث الست المتتابعة. وقد تم تطبيق هذه المنهجية لضمان أن عملية التحليل والنمذجة تتسم بالشفافية وقابلة للتكرار، وكونها تركز على ضمان جودة البيانات في كل مرحلة.(Zhao et al., 2023)
يوضح الشكل رقم 1 المخطط الانسيابي التفصيلي لتطبيق هذه المراحل على عملية الكشف عن مرض باركنسون.

الشكل رقم 1: المخطط الانسيابي لمنهجية CRISP-DM المطبقة في الدراسة.

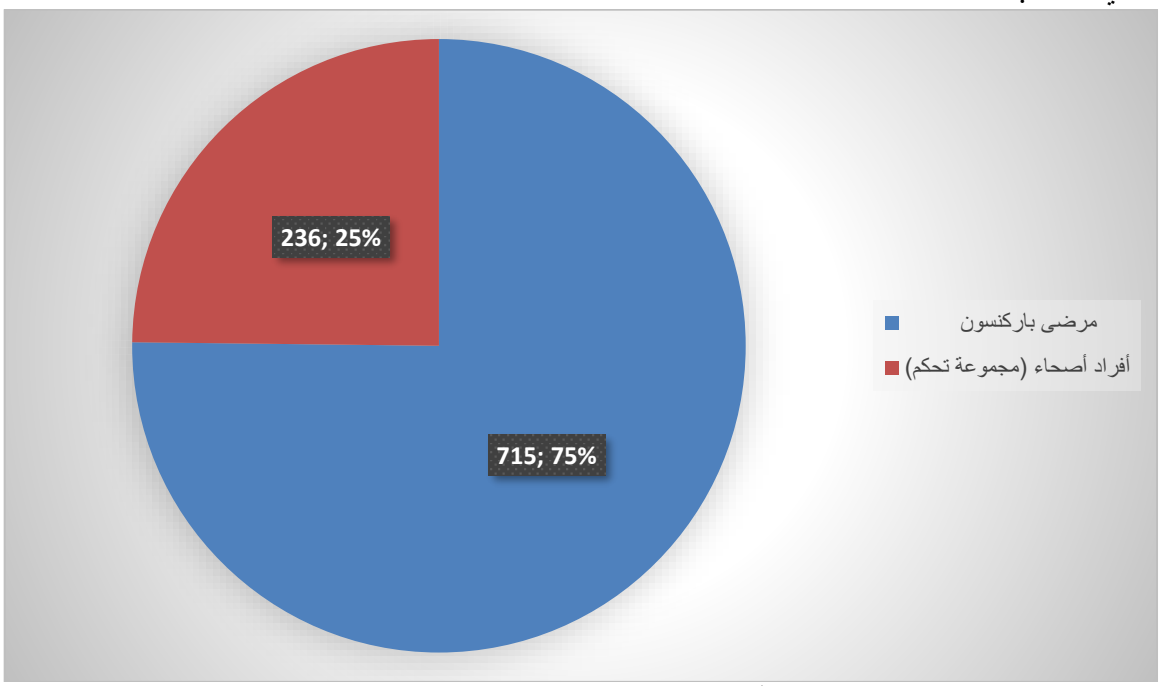
ثانيًا: إعداد البيانات والخصائص المُحللة

اعتمدت الدراسة على مجموعة بيانات صوتية موحدة تتألف من 951 سجلًا، تحتوي على 16 سمة صوتية كمية تُعد مؤشرات حيوية لاضطرابات الكلام (Little et al., 2019).

تحليل توزيع البيانات

تمثل مرحلة فهم البيانات وإعدادها جزءًا جوهريًا من المنهجية. أظهر تحليل توزيع العينات في مجموعة البيانات أن التوازن بين فئتي التشخيص كان مقبولاً، وهو أمر أساسي لضمان تدريب نماذج غير متحيزة.

الشكل رقم 2 التوزيع الفعلي للعينات في مجموعة البيانات، حيث يشير إلى وجود توازن مقبول يسمح بتدريب نماذج تصنيف فعالة.



الشكل رقم 2: التوزيع الفعلي للعينات في مجموعة البيانات.

المعالجة المسبقة والتعديل القياسي

لضمان الأداء العادل والفعال لجميع الخوارزميات (خاصة SVM و DNN)، تم تطبيق عملية التوحيد القياسي (Standardization) على جميع الخصائص الـ 16. وقد ضمنت هذه الخطوة أن تكون البيانات في نطاق قياسي بمتوسط يساوي صفر وانحراف معياري يساوي واحد، مما يمنع النماذج من الانحياز للميزات ذات المدى الكبير.

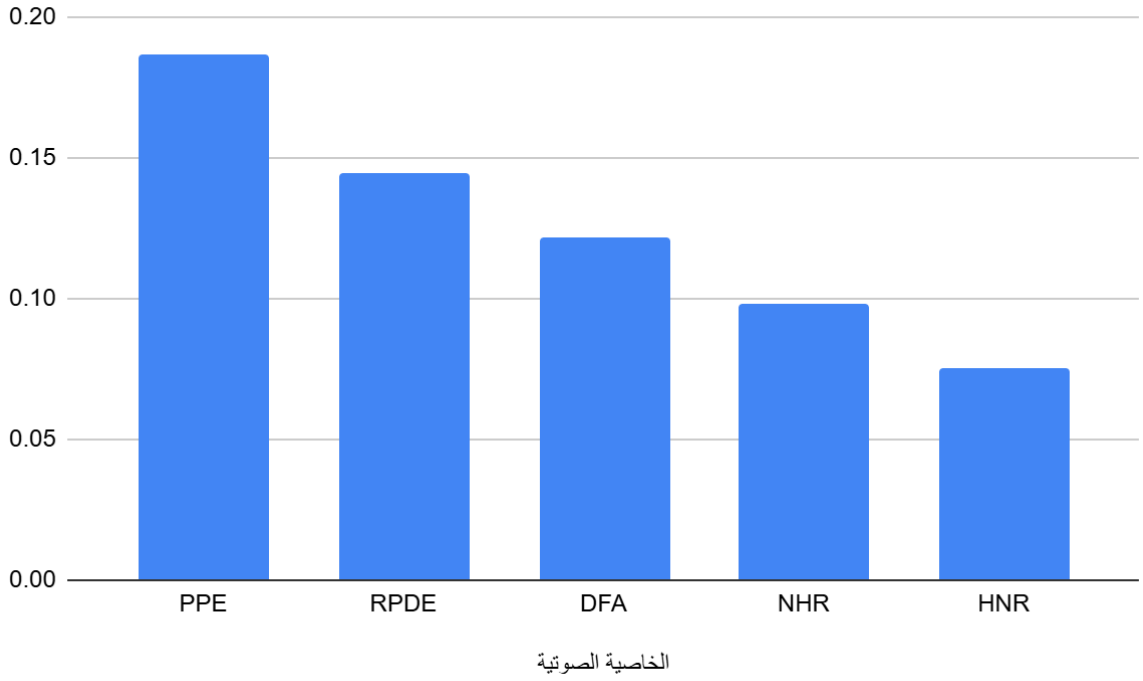
تحليل أهمية الخصائص باستخدامXAI

لتحقيق أهداف الذكاء الاصطناعي القابل للتفسير(XAI) ، تم استخدام نموذج Random Forest (RF) الفائز من مرحلة التقييم (الذي حقق أعلى دقة واسترجاع) لتحديد الأهمية النسبية لكل خاصية من الخصائص الصوتية الـ 16 (Kim et al., 2024).

يهدف هذا التحليل إلى:

1. تحديد الخصائص الصوتية الأكثر ارتباطاً بالتدهور العصبي الناتج عن مرض باركنسون.
2. توجيه الأبحاث المستقبلية للتركيز على هذه الخصائص ذات الأهمية العالية.

تم تلخيص الخصائص الخمسة الأكثر تأثيراً في الشكل رقم 3، والذي يُظهر قيمة الأهمية لكل خاصية. حيث يوضح الشكل تفوق الخصائص غير الخطية (PPE و RPDE) في قرار التصنيف، مما يبرهن على أن التغيرات في تعقيد الإشارة الصوتية هي المؤشرات الأقوى للكشف عن المرض.



الشكل رقم 3: تفوق الخصائص غير الخطية (PPE وRPDE) في قرار التصنيف.

النتائج والمناقشة

في هذه الورقة البحثية، لا يتمحور النقاش حول دقة النموذج فحسب، بل يمتد ليشمل تقييم فعالية المنهجية المُتبعة وتفسير أساس قرار النموذج من خلال تحليل أهمية الخصائص.

فعالية المنهجية المنظمة: إطار عمل CRISP-DM

يوضح الشكل رقم 1 (المخطط الانسيابي للمنهجية) أن تطبيق إطار عمل CRISP-DM ساهم في إضفاء مستوى عالٍ من التنظيم والشفافية على المشروع. إن الالتزام بالمراحل المتتابعة، خاصة مرحلتي فهم البيانات وإعداد البيانات، كان حاسماً في تحقيق الموثوقية العالية للنتائج التي تم التوصل إليها في الدراسة الأولى.

- **التعامل مع مصادر البيانات المتعددة:** سمحت CRISP-DM بدمج البيانات المأخوذة من منصات مختلفة (UCI وKaggle) ومعالجتها بمعايير موحدة، مما ضمن أن اختلافات المصدر لا تؤثر على نزاهة البيانات أو دقة النماذج.
- **توثيق عملية التحضير:** وفرت المنهجية مساراً واضحاً لتوثيق جميع خطوات المعالجة المسبقة، بما في ذلك التوحيد القياسي وتقسيم البيانات (90/10)، مما يجعل الدراسة قابلة للتكرار والتحقق من قبل باحثين آخرين (Zhao et al., 2023).

تأثير إعداد البيانات على جودة النماذج

أكد تحليل توزيع البيانات، الموضح في الشكل رقم 2، أن مجموعة البيانات الموحدة تتسم بتوزيع مقبول بين فئة مرضى باركنسون (حوالي 75%) ومجموعة التحكم الأصحاء (حوالي 25%). وعلى الرغم من وجود عدم توازن بسيط، إلا أن تطبيقه في نماذج مثل Random Forest أظهر قدرة النموذج على التعامل مع هذا التوزيع بفعالية.

كانت عملية التوحيد القياسي (Standardization) لجميع الخصائص الـ 16 ضرورية لضمان العدالة والنزاهة في مرحلة النمذجة. هذه الخطوة منعت النماذج الحساسة للمقاييس (مثل SVM وDNN) من الانحياز للخصائص ذات المدى العددي الأكبر (مثل DFA)، مما ساهم في أن تعكس نتائج تقييم الأداء الحقيقي للنماذج بناءً على القيمة التنبؤية للخاصية، وليس حجمها.

نتائج تحليل أهمية الخصائص (XAI)

تم تلخيص نتائج تحليل أهمية الخصائص الـ 16 باستخدام تقنيات الذكاء الاصطناعي القابل للتفسير (XAI) على نموذج Random Forest (RF) في الشكل رقم 3.

- **القيادة للخصائص غير الخطية:** يُظهر الشكل رقم 3 تفوقاً واضحاً للخصائص المتعلقة بـ التعقيد وعدم الخطية في النظام الصوتي. فقد احتلت خاصية PPE (انتروبيا فترة النبرة) و RPDE (كثافة فترة التكرار) المرتبة الأولى والثانية على التوالي في الأهمية النسبية. هذا يؤكد الاستنتاجات النظرية السابقة بأن التدهور العصبي لمرض باركنسون يظهر بوضوح في نقص تعقيد الإشارة الصوتية وعدم انتظامها بطريقة غير طبيعية (Little et al., 2019).
- **القدرة التفسيرية لنموذج Random Forest:** يفسر هذا التحليل أهمية نموذج Random Forest في هذا السياق؛ فنموذج RF قادر بطبيعته على تحديد واستغلال العلاقات المعقدة وغير الخطية بين هذه الخصائص ذات الأهمية العالية، بعكس النماذج التقليدية أو الأقل تعقيداً. هذا يبرر اختيار RF تحديداً لاستخدامه كأداة للذكاء الاصطناعي القابل للتفسير وتحديد أولويات الخصائص.
- **الآثار المترتبة على التصميم المستقبلي:** إن تحديد أن PPE هي الخاصية الأكثر تأثيراً في التصنيف (بأهمية نسبية 0.187) يقدم دليلاً حاسماً للمطورين. يجب أن تركز النماذج السريرية المستقبلية على استخلاص وتمثيل هذه الخصائص غير الخطية بفعالية قصوى.

الآثار المترتبة على الأهمية المنهجية والسريرية

تؤكد هذه النتائج على أهميتين متلازمتين: الأولى منهجية، تتمثل في ضرورة تبني إطار عمل CRISP-DM لضمان جودة البحث. والثانية سريرية، تتمثل في أن التركيز على PPE و RPDE كأهم مؤشرات حيوية يجب أن يكون هو المعيار الجديد في أنظمة التشخيص المبكر. ومن خلال الجمع بين المنهجية الصارمة وتقنياتXAI ، تقدم هذه الورقة خريطة طريق واضحة وقابلة للتفسير لتطوير أنظمة الذكاء الاصطناعي التشخيصية الموثوقة.

الاستنتاجات والتوصيات

أولاً: الاستنتاجات

نجحت هذه الدراسة في تحقيق هدفها المتمثل في تأسيس منهجية شاملة ومفسّرة لعملية تحضير البيانات وتقييم الخصائص في سياق الكشف عن مرض باركنسون باستخدام الإشارات الصوتية. وقد تم التوصل إلى الاستنتاجات الحاسمة التالية التي تدعم موثوقية النماذج وتفسيرها:

1. **الفعالية المنهجية لإطار عمل CRISP-DM:** أكدت الدراسة أن تطبيق إطار عمل CRISP-DM يوفر الهيكلة اللازمة لضمان شفافية وقابلية تكرار مشاريع التعلم الآلي في المجال الطبي. وقد ثبتت فعالية المراحل المنهجية، خاصة في توحيد البيانات المجمعة من مصادر متعددة وتوثيق جميع خطوات المعالجة المسبقة.
2. **الاكتشاف التحليلي للخصائص غير الخطية:** كان التطبيق الفعال لتقنيات الذكاء الاصطناعي القابل للتفسير (XAI)حاسماً في فك شيفرة القرار التشخيصي. أظهر تحليل أهمية الخصائص (الشكل رقم 3) أن الخصائص غير الخطية، وتحديداً PPE وRPDE ، هي المؤشرات الحيوية الأكثر قوة وتأثيراً في تحديد وجود المرض، متجاوزة الخصائص التقليدية لاضطراب النبرة.
3. **إثبات الأساس البيولوجي الرقمي:** يوفر تحديد PPE كأكثر الخصائص أهمية (بأهمية نسبية 0.187) دليلاً عملياً قوياً. هذا يشير إلى أن التغيرات في تعقيد الإشارة الصوتية وعدم انتظامها هي الآليات البيولوجية الأكثر تأثراً بالمرض، ويجب أن تكون هي المحور الأساسي لأي نظام تشخيص آلي مستقبلي.

ثانيًا: التوصيات

بناءً على النتائج المنهجية والتحليلية التي توصلت إليها هذه الدراسة، تُوصى بالمسارات البحثية والتطويرية التالية:

1. **التركيز الحصري على الخصائص ذات الأهمية العالية:** يُوصى بأن تركز الأبحاث القادمة على تطوير نماذج تصنيف تعتمد بشكل أساسي على مجموعة من الخصائص المفسّرة والمهمة (مثل PPE, RPDE, DFA) بدلاً من الخصائص الـ 16 كاملة. هذا سيقلل من الأبعاد، ويزيد من كفاءة النموذج الحسابية، ويحسن قابليته للتفسير.
2. **تطوير تقنيات استخلاص غير خطية متقدمة:** يُوصى بالعمل على تطوير تقنيات أكثر دقة لاستخلاص وتعزيز تمثيل خصائص PPE و RPDEمن بيانات الصوت الخام، لضمان استغلال كامل لقوتهما التنبؤية عند استخدام نماذج التعلم العميق.

3. **دمج XAI في التطبيقات التشغيلية:** يجب على المطورين السريريين دمج أدوات XAI في التطبيقات التشخيصية النهائية. يجب أن يقدم النظام تقريراً موجزاً يوضح سبب القرار، مثل: "احتمالية عالية للمرض بسبب ارتفاع قيمة"PPE ، لزيادة الثقة والمساءلة.
4. **توسيع منهجية CRISP-DM:** يُوصى بتوسيع إطار عمل CRISP-DM ليشمل تقييم الآثار السريرية والاجتماعية للنموذج في مرحلة التنفيذ (Deployment)، لضمان أن الحلول التكنولوجية تراعي اعتبارات الخصوصية والأخلاقيات الطبية قبل الاعتماد الكامل عليها.

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. Berry, M. J. A., & Linoff, G. S. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (2nd ed.). Wiley.
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.compbiomed.2024.107675
4. Lage, I., et al. (2019). Interpretable machine learning for healthcare: A comparative study of explanations generated for clinical predictive models. Bioinformatics, 35(17), 2959–2967. https://doi.org/10.1093/bioinformatics/btz084
5. 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.
6. 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
7. Tsanas, A., & Little, M. A. (2009). Accurate automated diagnosis of Parkinson's disease from sustained vowels. In 2009 International Conference on Machine Learning and Applications (pp. 714–719). IEEE.
8. 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.
9. 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
10. Abdusamea Omer, Rabyah B. Ali, Ahmed Al-Siddiq Masoud Al-Dabbashi, & Ali Abdulhamid Ali Al-Halak. (2025). A Comparative Study of Machine Learning and Deep Learning Models for Early Detection of Parkinson's Disease Using Voice Features. *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, *4*(4), 106–113.

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 **JLABW** and/or the editor(s). **JLABW** 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.