شرح المشروع بالعربية

تحليل إشارة الصوت باستخدام تعلّم الآلة للكشف المبكر عن مرض باركنسون

دليل دراسي من تسعة عشر جزءاً

من المبادئ الأولى إلى تحضير المناقشة

سهيل محمد الحمالي

رقم القيد 2200208522

المشرف: الدكتور عادل أضياف

قسم الهندسة الطبية الحيوية، كلية الهندسة، جامعة طرابلس

كل رقم في هذا الدليل مأخوذ من مجلد `docs/` ومجلد `reports/` في المشروع.

المحتويات

الجزء 0: خريطة الطريق ٣

الجزء 1: المشكلة الطبية وكيف يُنتَج الصوت ٦

الجزء 2: الصوت كإشارة رقمية ١٠

الجزء 3: البيانات المستخدمة، مجموعة MDVR-KCL ١٤

الجزء 4: المعالجة المسبقة ١٨

الجزء 5: التقطيع إلى مقاطع من عشر ثوانٍ ٢١

الجزء 6: الخصائص الصوتية (1) النبرة والانتظام ٢٤

الجزء 7: الخصائص الصوتية (2) الجودة والإيقاع والطيف ٢٨

الجزء 8: ما هو تعلّم الآلة، وما هي نماذجنا ٣٣

الجزء 9: أهم فكرة في المشروع، التحقّق المستقل عن المتحدّث ٣٦

الجزء 10: المقاييس، كيف نحكم على النموذج ٣٩

الجزء 11: نتائج المرحلة الرابعة وفحص العامل المُربك ٤٢

الجزء 12: تجارب التحسين وقاعدة الحماية من الإفراط في الملاءمة ٤٤

الجزء 13: النموذج النهائي وأهمية الخصائص ٤٧

الجزء 14: النموذج الأولي (التطبيق) **٤٩**

الجزء 15: التحقّق الخارجي على البيانات الإيطالية **٥٣**

الجزء 16: القيود والأخلاقيات **٥٧**

الجزء 17: بنية المشروع وكيفية إعادة تشغيل كل شيء **٦٠**

الجزء 18: تحضير العرض والأسئلة المتوقّعة **٦٣**

الجزء 0: خريطة الطريق

أولاً: ما هو مشروعك في جملة واحدة

مشروعك يجيب على سؤال واحد محدّد:

«هل نستطيع أن نقيس تسجيلاً صوتياً لشخص ما، ونقول إن نمطه الصوتي يشبه أصوات المصابين بمرض باركنسون أم يشبه أصوات الأشخاص الأصحّاء، باستخدام تعلّم آلة كلاسيكي قابل للتفسير فقط؟»

لاحظ الصياغة بدقّة: السؤال هو «**يشبه أي مجموعة**»، وليس «هل هذا الشخص مريض». هذا الفرق ليس تجميلاً لغوياً، بل هو جوهر المشروع علمياً وأخلاقياً، وسنعود إليه كثيراً.

المخرج النهائي هو **أداة دعم فرز بحثية غير تشخيصية** (research screening-support prototype): برنامج تُعطيه ملفاً صوتياً، فيعطيك مؤشراً حذراً غير طبّي، مع كامل الأدلة العلمية على مدى جودته وأين يفشل.

المصدر: `docs/01_overview_and_journey.md` القسم *1.1*

ثانياً: لماذا الصوت أصلاً؟

مرض باركنسون يقلّل مادة الدوبامين في مناطق الدماغ المسؤولة عن الحركة. وأعراضه المعروفة، وهي الرعاش والتيبّس وبطء الحركة، لا تصيب اليدين والقدمين فقط، بل تصيب **كل عضلة**، بما فيها العضلات الصغيرة جداً التي تتحكّم بالحبال الصوتية واللسان والشفتين والتنفّس.

ولأن هذه العضلات صغيرة ودقيقة جداً، فإن تأثّرها قد يظهر **مبكّراً**، أحياناً قبل أن تصبح أعراض الحركة واضحة بما يكفي لدفع الشخص لزيارة الطبيب. وهنا تكمن الفكرة كلّها: الميكروفون رخيص، وغير جارح، ولا يحتاج مستشفى.

المصدر: `docs/02_scientific_background.md` القسم *2.2*

ثالثاً: الأرقام التي يجب أن تحفظها

هذه هي أهم أرقام المشروع، وكلّها من `docs/README.md`:

- عدد الأشخاص في بيانات التدريب: 37 شخصاً (21 سليماً و16 مريضاً).
- عدد التسجيلات الصوتية: 73 تسجيلاً.

- عدد المقاطع بعد التقطيع إلى 10 ثوانٍ: 1001 مقطع.
- عدد الخصائص الصوتية المستخرجة من كل مقطع: 74 خاصية.
- عدد التكوينات النموذجية التي جُرّبت وقورنت: 19 تكويناً.
- الدقة المتوازنة داخلياً على مستوى الشخص: 0.822 مع انحراف معياري 0.032.
- مساحة تحت المنحنى (ROC-AUC) داخلياً: 0.864.
- الدقة المتوازنة خارجياً على البيانات الإيطالية: 0.629.
- مساحة تحت المنحنى خارجياً: 0.701.
- النتيجة عند استخدام تقسيم خاطئ متعمَّد: 0.909، وهي **رقم مضخَّم وليس نتيجة**.

لا تقلق إن لم تفهم معنى «الدقة المتوازنة» أو «المساحة تحت المنحنى» الآن، فالجزء 10 كامل مخصّص لشرحها بأمثلة رقمية بسيطة.

رابعاً: النقطة التي تميّز مشروعك

هناك أمر يجب أن تفهمه من البداية لأنه سيتكرّر في كل جزء تقريباً، وهو أهم ما ستدافع عنه أمام الأساتذة.

كثير من الأبحاث المنشورة على هذه النوعية من البيانات تُعلن نتائج بين 95٪ و99٪. ومشروعك يُعلن 82٪. للوهلة الأولى يبدو هذا ضعفاً. لكنه في الحقيقة **العكس تماماً**.

السبب أن تلك الأبحاث غالباً تخلط تسجيلات الشخص نفسه بين بيانات التدريب وبيانات الاختبار. فيصبح النموذج قادراً على «التعرّف على الشخص» بدل «التعرّف على المرض»، تماماً كطالب رأى أسئلة الامتحان قبل الامتحان.

وأنتم لم تكتفوا بقول ذلك، بل **قِستموه**: نفس الأنبوب البرمجي بالضبط، ونفس النموذج بالضبط، يعطي 0.909 بالتقسيم الخاطئ و0.775 بالتقسيم الصحيح. الفارق ليس قدرة على الكشف، بل حفظ.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.4*

خامساً: أربع نقاط قوة ستقولها في المناقشة

- بنيتم **القياس** نفسه، لا النموذج فقط. استخراج 74 خاصية صوتية من الصوت الخام هو المساهمة الهندسية الأساسية، ولم تُحمَّل جاهزة من أي مكان.
- تحقّقتم **بأمانة**: لا يظهر أي شخص في التدريب والاختبار معاً، وهذا مفروض بسطر في الكود يُوقف البرنامج فوراً إن حدث.
- **قِستم** قدرة التعميم بدل افتراضها: اختبرتم النموذج مجمّداً على بيانات إيطالية بلغة أخرى، فأعطى 0.701. قلّة قليلة من مشاريع البكالوريوس تفعل هذا.
- بنيتم نظاماً **يعترف بعدم اليقين**: عندما تكون الدرجة قريبة من الحدّ الفاصل يقول «غير حاسم» بدل أن يعطي تصنيفاً واثقاً.

المصدر: `docs/12_presentation_qa.md` القسم *12.1*

سادساً: مسار المشروع الزمني باختصار

لكي تتخيّل الصورة الكبيرة، هذا ما حدث بالترتيب:

أولاً بُنيت بنية المشروع وبيئة العمل، ثم فُحصت البيانات فحصاً كاملاً **قبل** أي تدريب، ثم بُني أنبوب المعالجة واستخراج الخصائص، ثم دُرّبت النماذج وقيّمت تقييماً نزيهاً فأعطت 0.780، ثم بُني التطبيق، ثم طلب المشرف تحسين الدقة فأُجريت دراسة من 19 تكويناً رفعت النتيجة إلى 0.822، ثم سجّل المشرف صوته وصوت صديقه، وكلاهما سليم، فحصل كلاهما على درجة قريبة من 0.65. وهذه الحادثة كشفت ضعفاً حقيقياً وأدّت إلى إضافة «النطاق غير الحاسم». وأخيراً جاء التحقّق الخارجي على البيانات الإيطالية.

المصدر: `docs/01_overview_and_journey.md` القسم *1.3*

لاحظ شيئاً مهماً هنا: أهم تحسينات المشروع لم تأتِ من نجاح، بل من **ملاحظة فشل**. هذه قصة جيدة جداً لتحكيها في المناقشة.

خلاصة الجزء

- المشروع يقيس تشابه نمط صوتي مع مجموعة بيانات، ولا يشخّص مرضاً إطلاقاً.
- الأرقام الأساسية: 37 شخصاً، و73 تسجيلاً، و74 خاصية، ونتيجة 0.822 داخلياً و0.701 خارجياً.
- قوة المشروع ليست في ارتفاع الرقم، بل في نزاهة طريقة قياسه، والدليل المقيس على ذلك هو الفرق بين 0.909 و0.775.

الجزء 1: المشكلة الطبية وكيف يُنتَج الصوت

قبل أن نتحدث عن أي كود أو أي نموذج، يجب أن تفهم شيئاً واحداً فهماً عميقاً: **لماذا يمكن للصوت أصلاً أن يحمل معلومة عن مرض في الدماغ؟**

إن فهمت هذا الجزء جيداً، فسيصبح كل ما بعده منطقياً وطبيعياً. وإن لم تفهمه، فسيبدو المشروع كأنه أرقام عشوائية.

أولاً: الصوت البشري يُنتَج على ثلاث مراحل

الكلام ليس عملية واحدة، بل ثلاث مراحل متتابعة. والتشبيه الأدقّ هو **آلة موسيقية نفخية**، مثل المزمار.

المرحلة الأولى: مصدر الطاقة (the power source) --- الرئتان. الرئتان (lungs) تدفعان الهواء إلى أعلى عبر القصبة الهوائية (trachea). هذا الهواء المتدفّق هو الطاقة التي ستتحوّل لاحقاً إلى صوت. لا هواء يعني لا صوت إطلاقاً. في تشبيه الآلة الموسيقية: الرئتان هما **الشخص الذي ينفخ**.

المرحلة الثانية: المُهتزّ (the vibrator) --- الحبال الصوتية. في الحنجرة (larynx) توجد طيّتان عضليّتان صغيرتان تُسمّيان الحبال الصوتية (vocal folds). عندما تقرّبهما من بعضهما، يصطدم الهواء الصاعد بهما فيفتحهما بالقوّة، ثم تعود مرونتهما الطبيعية فتُغلقهما بسرعة، ثم يفتحهما الهواء مرة أخرى، وهكذا **مئات المرات في الثانية الواحدة**. كل دورة فتح وإغلاق تُطلق «نفخة» صغيرة من الهواء، وتتابع هذه النفخات السريعة هو الصوت الخام الطنّان للإنسان. في تشبيه الآلة: الحبال الصوتية هي **ريشة المزمار**.

المرحلة الثالثة: المرشِّح (the filter) --- مجرى الصوت. الصوت الخام الخارج من الحنجرة يمرّ بعد ذلك في البلعوم والفم واللسان والشفتين، وهذه كلها معاً تُسمّى مجرى الصوت (vocal tract). هذه المنطقة تعمل كأنبوب رنين: تحريك اللسان والشفتين يُقوّي بعض الترددات ويُضعف أخرى، فيتحوّل الطنين الخام إلى حروف متحرّكة (vowels) وحروف ساكنة (consonants). في تشبيه الآلة: مجرى الصوت هو **جسم الآلة وفتحاتها**.

المصدر: *docs/02_scientific_background.md* القسم *2.1*

ثانياً: نتيجتان مهمّتان جداً لمشروعك

النتيجة الأولى: سرعة اهتزاز الحبال الصوتية هي «طبقة الصوت». عدد مرات فتح وإغلاق الحبال الصوتية في الثانية يُسمّى التردد الأساسي (fundamental frequency) ويُرمز له بـF0، ويُقاس بالهرتز (Hz). وأنت تسمعه كـ«حدّة» أو «غِلَظ» الصوت.

- الرجل البالغ: تقريباً بين 85 و180 هرتز.
- المرأة البالغة: تقريباً بين 165 و255 هرتز.

انتبه جيداً لهذه النقطة، فهي مصدر مشكلة سنواجهها لاحقاً: طبقة الصوت المطلقة تحمل معلومة عن **جنس المتحدّث**، لا عن مرضه. وهذا ما يُسمّى «عاملاً مُربكاً» (confounder)، أي متغيّر خفي مرتبط بالجواب ويسمح للنموذج بأن «يغشّ». وفي الجزء 11 سأشرح كيف اختبرتم هذه المشكلة بدل تجاهلها.

النتيجة الثانية: انتظام الاهتزاز هو «جودة الصوت». في الصوت السليم، الاهتزاز **دوري بدرجة عالية** (highly periodic): كل دورة تكاد تكون نسخة طبق الأصل من الدورة السابقة، نفس المدّة الزمنية تقريباً، ونفس القوة تقريباً. وهذه الجملة تحديداً هي الأساس العلمي لأهم خصائص مشروعك: jitter وshimmer وHNR. فكلّها تقيس شيئاً واحداً بصور مختلفة: **إلى أي درجة تشبه كل دورة الدورة التي قبلها؟**

ثالثاً: ماذا يفعل باركنسون بالضبط؟

مرض باركنسون (Parkinson's disease) مرض تنكّسي عصبي سببه نقص مادة الدوبامين (dopamine) في مناطق من الدماغ مسؤولة عن التحكّم بالحركة. وأعراضه الكلاسيكية أربعة:

- الرعاش (tremor): اهتزاز لا إرادي.
- التيبّس (rigidity): تصلّب العضلات.
- بطء الحركة (bradykinesia): الحركة تصير أبطأ.
- قِلّة سعة الحركة (hypokinesia): الحركة تصير **أصغر** في المدى.

والنقطة الجوهرية التي يجب أن تقولها في المناقشة: **هذه الأعراض تصيب كل عضلة في الجسم، لا اليدين فقط.** والعضلات التي تتحكّم بالحبال الصوتية واللسان والشفتين والتنفّس هي عضلات صغيرة جداً وتتطلّب تحكّماً دقيقاً جداً، ولهذا فهي من أوائل ما يتأثّر.

واضطراب الكلام الناتج عن هذا له اسم طبي محدّد: **عُسر التلفّظ ناقص الحركة** (hypokinetic dysarthria).

- عُسر التلفّظ (dysarthria): اضطراب في النطق سببه ضعف أو سوء تحكّم في عضلات الكلام.
- عُسر الصوت (dysphonia): اضطراب في جودة الصوت نفسه، كالبحّة والهَمْس.

رابعاً: الجسر بين العَرَض الطبي والرقم في الحاسوب

هذه أهم قائمة في الجزء كلّه. كل بند فيها يربط بين ثلاثة أشياء: ما يراه الطبيب، وما يحدث فيزيائياً، وما نستطيع نحن قياسه بالحاسوب.

- **انخفاض الصوت والكلام الرتيب.** السبب الفيزيائي: قِلّة سعة الحركة وضعف دعم التنفّس. ما نقيسه: **انخفاض التغيّر** في F0 وفي السعة، أي أن الانحراف المعياري لطبقة الصوت ومداها يصغران.

- **صوت مهموس أو مبحوح.** السبب: الحبال الصوتية لا تنغلق انغلاقاً كاملاً، فيتسرّب هواء على شكل هواء مضطرب. ما نقيسه: **ضجيج أكثر مقابل نغمة أقل**، أي انخفاض HNR وCPPS.
- **صوت غير مستقرّ.** السبب: الرعاش وعدم انتظام التحكّم العضلي. ما نقيسه: **زيادة عدم الانتظام بين دورة وأخرى**، أي ارتفاع jitter وshimmer.
- **حروف ساكنة غير دقيقة.** السبب: حركات اللسان والشفتين صارت أبطأ وأصغر. ما نقيسه: **شكل طيفي مختلف وأبطأ تغيّراً**، أي MFCC ومشتقّاتها delta-MFCC.
- **تردّد وتوقّفات في غير محلّها.** السبب: صعوبة في بدء الحركة وفي الاستمرار بها. ما نقيسه: **توقّفات أكثر وأطول**، أي خصائص التوقّفات.

المصدر: *docs/02_scientific_background.md* القسم *2.2*

إن سألك أستاذ في المناقشة «لماذا اخترتم هذه الخصائص تحديداً؟»، فجوابك هو هذه القائمة بالضبط. كل خاصية في مشروعك لها مبرّر فسيولوجي، ولم تُختَر لأنها متوفّرة في مكتبة برمجية.

خامساً: لماذا «الكشف المبكّر» تحديداً؟

التغيّرات الصوتية التي ذكرناها كثيراً ما تظهر **مبكّراً**، أحياناً قبل أن تصبح الأعراض الحركية واضحة بما يكفي لتدفع الشخص إلى زيارة طبيب أعصاب. والسبب منطقي: عضلات الحنجرة واللسان أصغر بكثير من عضلات الذراع، وتحتاج دقّة تحكّم أعلى بكثير. فالنقص البسيط في التحكّم الحركي قد لا يظهر في مشيتك، لكنه يظهر في اهتزاز يتكرّر 150 مرة في الثانية.

ولهذا فإن فكرة الفرز الصوتي (voice-based screening) جذّابة عملياً:

- الميكروفون رخيص جداً ومتوفّر في كل هاتف.
- الفحص غير جارح (non-invasive) تماماً، فلا إبَر ولا أشعّة ولا مستشفى.
- يمكن تكراره بسهولة وبتكلفة شبه معدومة.
- يمكن، نظرياً، أن يُشير إلى الأشخاص الذين يُستحسن أن يراجعوا طبيب أعصاب.

لاحظ الصياغة الأخيرة: «يُشير إلى من يُستحسن أن يراجع طبيباً»، وليس «يُشخّص». هذا هو التعريف الدقيق لكلمة **فرز** (screening): خطوة تسبق التشخيص وتُوجّه إليه، ولا تحلّ محلّه أبداً.

سادساً: ما لا يدّعيه مشروعك

يجب أن تكون هذه الجملة حاضرة في ذهنك دائماً: **المشروع لا يدّعي أن هذه القياسات تُشخّص مرض باركنسون.**

والسبب بسيط ومقنع جداً: هناك حالات كثيرة أخرى تُغيّر جودة الصوت بنفس الاتجاه تقريباً: الزكام والتهاب الحنجرة (laryngitis)، والتقدّم في العمر، والتدخين، والإرهاق وقلّة النوم، وأمراض عصبية أخرى غير باركنسون.

فإذا قاس النظام ارتفاعاً في jitter وانخفاضاً في HNR، فهو لا يستطيع أن يميّز إن كان السبب باركنسون أم زكاماً. كل ما يستطيع قوله هو: «**هذا النمط الصوتي يشبه نمط المجموعة المرضية في مجموعة البيانات التي تدرّبتُ عليها.**»

المصدر: `docs/02_scientific_background.md` نهاية القسم *2.2*

خلاصة الجزء

- الصوت يُنتَج على ثلاث مراحل: الرئتان مصدر الطاقة، والحبال الصوتية هي المُهتزّ الذي يحدّد F0، ومجرى الصوت هو المرشِّح الذي يصنع الحروف.
- باركنسون يُضعف التحكّم في كل العضلات بما فيها عضلات الكلام الدقيقة، فينتج `hypokinetic dysarthria`، ولكل عَرَض فيه مقابل رقمي نقيسه.
- الفكرة كلّها فرز مبكّر رخيص وغير جارح، وليست تشخيصاً، لأن حالات كثيرة أخرى تُغيّر الصوت بنفس الطريقة.

الجزء 2: الصوت كإشارة رقمية

في الجزء السابق عرفنا كيف يُنتج الجسم الصوت. الآن السؤال التالي: **كيف يدخل هذا الصوت إلى الحاسوب ويصير أرقاماً؟**

أولاً: الصوت في الأصل موجة ضغط

عندما تهتزّ حبالك الصوتية، فهي تضغط الهواء المجاور لها ثم تُرخيه، بالتناوب وبسرعة كبيرة. وهذا التتابع من الضغط والتخلخل ينتشر في الهواء كموجة، وتُسمّى موجة ضغط (pressure wave). تخيّل أنك رميت حجراً في بركة ماء: تنتشر دوائر من الارتفاع والانخفاض. الصوت هو الشيء نفسه لكن في الهواء وبثلاثة أبعاد.

فالصوت إذن، في أصله، ظاهرة **مستمرة** (continuous): لا توجد لحظة لا قيمة فيها. وهذا يمثّل مشكلة للحاسوب، لأنه لا يستطيع تخزين عدد لا نهائي من القيم.

ثانياً: كيف يحوّل الميكروفون الموجة إلى كهرباء

داخل الميكروفون (microphone) غشاء رقيق جداً. وعندما تصطدم به موجة الضغط يهتزّ بنفس نمط الموجة، وهذا الاهتزاز الميكانيكي يُحوَّل إلى **جهد كهربائي متغيّر**. فالميكروفون في جوهره مترجم: يترجم «ضغط هواء» إلى «جهد كهربائي». والجهد الكهربائي لا يزال إشارة مستمرة، أي تناظرية (analogue).

ثالثاً: العيّنة، أهم مفهوم في هذا الجزء

هنا يأتي دور جهاز يُسمّى المحوّل التناظري الرقمي (analogue-to-digital converter، اختصاراً ADC)، وهو موجود في كل هاتف وكل حاسوب. وعمله بسيط جداً في الفكرة: **يقيس قيمة الجهد الكهربائي، ويسجّلها كرقم، ثم ينتظر جزءاً صغيراً جداً من الثانية، ثم يقيس مرة أخرى، ويكرّر ذلك آلاف المرات في الثانية**. وكل قياس منفرد من هذه القياسات يُسمّى **عيّنة** (sample).

المصدر: *docs/11_glossary.md* قسم معالجة الصوت والإشارة

تشبيه: كاميرا الفيديو. الحركة في الواقع مستمرة، لكن الكاميرا لا تُسجّل «الحركة»، بل تلتقط **صوراً ثابتة متتابعة** بسرعة كبيرة، مثلاً 30 صورة في الثانية. وعندما تُعرَض بنفس السرعة، تخدع عينك فتراها حركة متصلة. العيّنة الصوتية هي «الصورة الثابتة»، والصوت الرقمي شريط من ملايين الصور الثابتة المتتابعة.

رابعاً: معدّل العيّنات

عدد العيّنات المأخوذة في الثانية الواحدة يُسمّى معدّل العيّنات (sample rate)، ويُقاس بالهرتز. وفي مشروعك، كل ملفات مجموعة MDVR-KCL مسجّلة بمعدّل **44100 هرتز**، أي 44100 قياس في كل ثانية من الصوت.

المصدر: *docs/03_dataset.md* القسم *3.3*

مثال رقمي صغير. لنحسب حجم تسجيل واحد نموذجي في مشروعك:

- طول التسجيل تقريباً: دقيقتان، أي 120 ثانية.
- معدّل العيّنات: 44100 عيّنة في الثانية.
- الناتج: 120 × 44100 = **5,292,000 عيّنة**.

أي أن التسجيل الواحد في مشروعك ليس «صوتاً» بالنسبة للحاسوب، بل هو ببساطة **قائمة من حوالي 5.3 مليون رقم**.

المصدر: *docs/04_pipeline.md* القسم *4.1*

وهذا الرقم يشرح لك أمرين عمليّين: لماذا استغرق استخراج الخصائص من كل الملفات حوالي 32 دقيقة، ولماذا يحتاج تحليل ملف واحد في التطبيق حوالي دقيقة كاملة.

خامساً: لماذا 44100 هرتز تحديداً؟

هذا سؤال ممتاز وقد يُسأل في المناقشة. والجواب يعتمد على قانون فيزيائي اسمه **نظرية نايكوِست** (the Nyquist theorem)، وتقول باختصار: **لكي تمثّل تردداً ما تمثيلاً أميناً، يجب أن يكون معدّل العيّنات ضعف ذلك التردد على الأقل.**

ولنفهم لماذا: تخيّل موجة تصعد وتنزل. إن أخذتَ عيّنة واحدة فقط في كل دورة، فقد تلتقط دائماً قمّة الموجة وتظنّها خطاً مستقيماً. تحتاج على الأقل قياسين في كل دورة، واحداً للقمّة وواحداً للقاع، لتعرف أنها تتذبذب أصلاً.

والتطبيق على مشروعك: معدّل العيّنات 44100 هرتز، إذن أعلى تردد يمكن تمثيله بأمانة هو 44100 ÷ 2 = **22050 هرتز**. وأذن الإنسان تسمع تقريباً حتى 20000 هرتز. فمعدّل 44100 هرتز يغطّي كامل مدى السمع البشري مع هامش بسيط.

المصدر: *docs/04_pipeline.md* القسم *4.1*

وسنرى في الجزء 15 أن هذه النقطة بالذات كانت فخّاً خطيراً كاد يُفسد التحقّق الخارجي، لأن ملفاً بمعدّل 16000 هرتز **لا يمكن فيزيائياً** أن يحتوي أي صوت فوق 8000 هرتز.

سادساً: السعة

لو كان معدّل العيّنات يجيب على سؤال «متى نقيس؟»، فإن السعة (amplitude) تجيب على سؤال «**ماذا** قِسنا؟». السعة هي قيمة العيّنة الواحدة، وهي تمثّل شدّة الضغط في تلك اللحظة، أي ما تُدركه أذنك على أنه عُلُوّ الصوت في تلك اللحظة. وعادةً تُخزَّن هذه القيم كأرقام عشرية بين 1.0- و1.0+.

فمقطع من الصوت في الحاسوب يبدو هكذا فعلياً:

```
0.0021, 0.0834, 0.1520, 0.0912, -0.0334, -0.1201, ...
```

وكل خصائص مشروعك الـ74 مستخرَجة من قائمة أرقام كهذه، لا من شيء آخر.

مفهوم مرتبط: القَصّ. إذا كان الصوت أعلى من المدى الذي يستطيع الجهاز تسجيله، فإن القمم تُقصّ وتصير مسطّحة عند القيمة القصوى. وهذا يُسمّى القَصّ (clipping)، وهو تشويه حقيقي للإشارة يُفسد قياسات جودة الصوت لأنه يُدخل عدم انتظام مصطنَعاً لم تُنتجه الحنجرة. ولهذا فإن تطبيقك يفحص نسبة العيّنات المقصوصة ويُظهر تحذيراً إذا تجاوزت 0.1٪.

المصدر: *docs/07_prototype.md* القسم *7.4*

سابعاً: القنوات

القناة (channel) هي تسجيل مستقلّ واحد. فأحادي (mono) يعني قناة واحدة، أي قائمة أرقام واحدة، وثنائي (stereo) يعني قناتين. وكل ملفات MDVR-KCL أحادية القناة أصلاً. ومع ذلك فإن أول خطوة في أنبوب المعالجة تُحوّل أي ملف إلى أحادي، لأن التطبيق قد يستقبل من المستخدم ملفاً ثنائياً.

المصدر: *docs/03_dataset.md* القسم *3.3* و*docs/04_pipeline.md* القسم *4.2*

والسبب العلمي في اشتراط الأحادي: خصائصنا تصف **مصدراً صوتياً واحداً** هو الحنجرة. ولو حلّلنا قناتين مختلفتين قليلاً لحصلنا على قيمتين مختلفتين لنفس الشيء، وهذا لا معنى له.

ثامناً: ملف WAV ولماذا هو المهم هنا

صيغة WAV صيغة صوتية **غير مضغوطة** (uncompressed): تخزّن العيّنات كما هي، رقماً رقماً، دون أي فقدان في الجودة.

وهنا نقطة علمية مهمة جداً: صيغ مثل MP3 تُصغّر حجم الملف بأسلوب ذكي، إذ تحذف أجزاءً من الصوت **تظنّ أن الأذن البشرية لن تلاحظها**. وهذا مقبول تماماً للموسيقى، لكنه كارثة لمشروعك. لماذا؟ لأن ما تحذفه تلك الخوارزميات هو غالباً التفاصيل الخافتة وغير المنتظمة، وهي بالضبط ما نقيسه نحن. فعدم الانتظام الدقيق بين دورة وأخرى قد يكون غير مسموع للأذن، لكنه **هو الإشارة الطبية نفسها**.

لذلك كل ملفات مجموعتَي البيانات في مشروعك بصيغة WAV. ولهذا السبب نفسه رفض مشروعك لاحقاً إزالة الضجيج، وهي قصّة الجزء 4.

تاسعاً: الصورة الكاملة في سطر واحد

اهتزاز الحبال الصوتية يُنتج موجة ضغط في الهواء، ثم يحوّلها الميكروفون إلى جهد كهربائي متغيّر، ثم يقيس المحوّل ADC هذا الجهد 44100 مرة في الثانية، ثم تُخزَّن هذه القياسات كقائمة أرقام في ملف WAV، وهذه القائمة هي كل ما يراه برنامجك.

خلاصة الجزء

- العيّنة قياس واحد لشدّة الصوت في لحظة واحدة، ومعدّل العيّنات عددها في الثانية، وهو 44100 في مشروعك، أي أن تسجيلاً من دقيقتين قائمة من نحو 5.3 مليون رقم.
- اختيار 44100 هرتز يعود لنظرية نايكوِست: معدّل العيّنات يجب أن يكون ضعف أعلى تردد نريد تمثيله، و22050 هرتز تغطّي كامل السمع البشري.
- صيغة WAV غير المضغوطة ضرورية لأن الضغط يحذف التفاصيل الدقيقة غير المنتظمة، وهي بالتحديد الإشارة التي يقيسها المشروع.

الجزء 3: البيانات المستخدمة، مجموعة MDVR-KCL

أي مشروع تعلّم آلة يقوم على شيئين: البيانات والنموذج. والبيانات هي الأهم بفارق كبير. فنموذج ممتاز على بيانات سيّئة يعطي نتائج سيّئة، لكن نموذجاً بسيطاً على بيانات مفهومة جيداً يعطي نتائج يمكن الدفاع عنها. ولهذا كانت **المرحلة الثانية** في مشروعك كلّها مخصّصة لفحص البيانات، **قبل** تدريب أي نموذج.

أولاً: ما هي هذه المجموعة؟

اسمها الكامل Mobile Device Voice Recordings at King's College London، واختصارها MDVR-KCL، أي «تسجيلات صوتية بجهاز محمول في كِنغز كوليدج لندن». تحتوي على تسجيلات لأشخاص مُشخَّصين بمرض باركنسون، ولأشخاص أصحّاء يُسمّون مجموعة الضبط الصحيحة (healthy controls، اختصاراً HC)، سُجّلت جميعها بهاتف محمول في بيئة مضبوطة. وهي متاحة للعموم على منصّة Zenodo.

المصدر: *docs/03_dataset.md* القسم *3.1*

ملاحظة عملية: الملفات الصوتية نفسها **غير مرفوعة** إلى مستودع GitHub، لأنها كبيرة ولها رخصة خاصة. والمستودع يحتوي بدلاً منها على تعليمات التحميل ومكان الوضع الصحيح: `data/raw/mdvr_kcl/`.

ثانياً: بنية المجلدات ومهمّتان مختلفتان

تنقسم المجموعة إلى مجلدين، وكل مجلد يمثّل **مهمّة تسجيل** مختلفة:

- **قراءة نصّ** (ReadText): كل متحدّث يقرأ **النصّ نفسه**. وميزة هذا أن المحتوى مضبوط تماماً، فأي فرق بين المتحدّثين يأتي من **كيف** يتكلّمون، لا من **ماذا** يقولون. وهذا مثالي علمياً.
- **حوار عفوي** (SpontaneousDialogue): كلام حرّ غير مقيّد. ميزته أنه أقرب إلى الواقع، وعيبه أن المحتوى يختلف من شخص لآخر.

المصدر: *docs/03_dataset.md* القسم *3.2*

وهذا التقسيم سيفيدك لاحقاً في فهم نقطة دقيقة: التسجيلان اللذان يخصّان الشخص الواحد ليسا متشابهين جداً، لأنهما مهمّتان مختلفتان، وسنستخدم هذه التفصيلة في الجزء 9.

ثالثاً: الأرقام الدقيقة

عن الملفات:

- إجمالي الملفات الصوتية: 73 ملفاً.
- الملفات الصالحة للاستخدام: 73 ملفاً، أي **كلّها**.
- الصيغة: WAV فقط، بلا استثناء.
- معدّل العيّنات: 44100 هرتز في كل الملفات.
- القنوات: أحادي في كل الملفات.

عن المدد الزمنية:

- أقصر تسجيل: 73.04 ثانية.
- الوسيط (median): 141.29 ثانية.
- المتوسّط (mean): 141.35 ثانية.
- أطول تسجيل: 220.64 ثانية.

لاحظ أن الوسيط والمتوسّط متقاربان جداً، وهذا يدلّ على أن التوزيع متوازن ولا توجد ملفات شاذّة تجرّ المتوسّط. وهذه ملاحظة إحصائية جيدة تُظهر أنك فهمت الأرقام لا حفظتها.

عن الأشخاص:

- المجموعة السليمة HC: 21 شخصاً، و42 تسجيلاً.
- المجموعة المريضة PD: 16 شخصاً، و31 تسجيلاً.
- المجموع: **37 شخصاً** و**73 تسجيلاً**.

المصدر: *docs/03_dataset.md* القسم *3.3*

سؤال قد يُطرح عليك: لماذا 73 وليس 74؟ تقريباً كل شخص ساهم بتسجيلين اثنين، فـ$37 \times 2 = 74$. لكن العدد الفعلي 73. السبب أن الشخص المريض صاحب المعرّف ID18 له **تسجيل واحد فقط**. لا يوجد لغز، لكن معرفتك للإجابة تُظهر أنك فحصت بياناتك فعلاً.

رابعاً: كيف عرفنا التسمية؟

التسمية (label) هي «الجواب الصحيح» لكل ملف: هل صاحبه HC أم PD. والنقطة المهمة: **التسمية في مشروعك لم تُخمَّن، بل قُرئت من مصدرين مستقلّين وقُورن بينهما.**

- المصدر الأول: **اسم المجلد** الذي يقع فيه الملف.
- المصدر الثاني: **وسم داخل اسم الملف** نفسه (_hc_ أو _pd_).

مثال على اسم ملف حقيقي: `ID00_hc_0_0_0.wav`، وآخر: `ID02_pd_2_1_1.wav`.

والقاعدة المكتوبة في الكود: **لو اختلف المصدران في أي ملف، يُستبعَد ذلك الملف** ويُسجَّل في التقرير على أنه «تعارض تسمية»، ولا يُسنَد له تصنيف بصمت. والنتيجة الفعلية: **لم يُوجد أي تعارض**.

المصدر: *docs/03_dataset.md* القسم *3.4*

لماذا هذه النقطة مهمة؟ لأنها تُظهر منهجاً: أنت لم تفترض صحّة البيانات، بل **تحقّقت منها**. وهذا فرق جوهري بين مشروع طالب ومشروع باحث.

خامساً: معرّف المتحدّث، أهم معلومة في المشروع كلّه

معرّف المتحدّث (subject ID) هو الرمز الذي يجمع كل تسجيلات الشخص الواحد تحت هوية واحدة. وطريقة استخراجه بسيطة: هو الجزء الأول من اسم الملف. فالملف `ID00_hc_0_0_0.wav` يخصّ الشخص ID00.

والتوثيق يصف هذا المعرّف بأنه **أهم قطعة بيانات وصفية في المشروع بأكمله**. والسبب أنه ما يجعل **التحقّق المستقل عن المتحدّث** (subject-independent validation) ممكناً أصلاً. فالتسجيلان الخاصّان بالشخص ID00 يجب أن يسافرا معاً دائماً: إمّا كلاهما في بيانات التدريب، وإمّا كلاهما في بيانات الاختبار.

واعلم أنه لو لم نستطع استخراج معرّفات المتحدّثين بثقة، لكان **يجب إيقاف المشروع** ومراجعة الأمر، لأن أي نتيجة بعدها ستكون بلا معنى. والقاعدة هنا أيضاً: أي ملف بلا معرّف واضح يُستبعَد ويُذكَر في التقرير. والنتيجة: **لم يُوجد أي ملف كهذا**.

المصدر: *docs/03_dataset.md* القسم *3.5*

سادساً: فحوصات السلامة السبعة

قبل أي تدريب، أُجريت سبعة فحوصات، وكانت نتيجتها كلّها سليمة:

- ملفات تالفة أو غير قابلة للقراءة: لا يوجد.
- ملفات فارغة أو بطول صفر: لا يوجد.
- ملفات مكرّرة المحتوى بمقارنة بصمة MD5: لا يوجد.
- تعارض في التسمية بين المجلد واسم الملف: لا يوجد.
- ملفات بلا معرّف متحدّث: لا يوجد.
- أشخاص ظهروا في أكثر من فئة واحدة: لا يوجد.
- تسجيلات قصيرة جداً أو صامتة: لا يوجد.

المصدر: *docs/03_dataset.md* القسم *3.6*

لماذا فحص التكرار مهم بشكل خاص؟ بصمة MD5 رقم يُحسب من **محتوى** الملف لا من اسمه. فلو وُجد ملفان باسمين مختلفين لكن محتواهما الصوتي متطابق، لأعطيا نفس البصمة تماماً. ولماذا يهمّنا؟ لأن ملفاً مكرّراً قد تقع نسخة منه في بيانات التدريب ونسخة في

بيانات الاختبار، فيكون النموذج قد **رأى الجواب مسبقاً**، فترتفع النتيجة ارتفاعاً كاذباً. لم يحدث هذا في مجموعتك، لكن الجملة المهمة في التوثيق هي: «**جرى التحقّق منه بدل افتراضه**». وهي عبارة ممتازة لاستخدامها في مناقشتك.

سابعاً: المخاطر التي سُجّلت بصراحة

التقرير لم يكتفِ بالنتائج الجيدة، بل سجّل خمسة مخاطر حقيقية، وذكرك لها في المناقشة يزيد مصداقيتك ولا ينقصها:

- **المجموعة صغيرة.** سبعة وثلاثون شخصاً عدد قليل، ولذلك يجب دائماً ذكر التباين بين الطيّات مع كل نتيجة، لا رقم واحد واثق.
- **ظروف تسجيل مشتركة داخل الشخص الواحد.** تسجيلا الشخص الواحد يشتركان في نفس الصوت والميكروفون والجلسة، ولهذا فإن التحقّق المستقل عن المتحدّث **إلزامي وليس اختيارياً**.
- **عدم توازن الفئات** (`class imbalance`): 42 تسجيلاً سليماً مقابل 31 مريضاً، وهذا يجعل «الدقّة» العادية مضلّلة، ولهذا استُخدمت «الدقّة المتوازنة» بدلاً منها.
- **لا توجد بيانات ديموغرافية.** الملفات لا تحمل عمراً ولا جنساً موثّقاً لكل شخص. وبما أن طبقة الصوت تعكس جنس المتحدّث بقوّة، فهذا عامل مُربك محتمل، عولج بتجربة حذف سنشرحها في الجزء 11.
- **قاعدة التحذير من الدقّة العالية.** كُتبت **قبل** ظهور أي نتيجة: أي نتيجة تبلغ 95٪ أو أكثر تُعامَل على أنها **عَرَض من أعراض التسريب يجب التحقيق فيه**، لا نجاحاً يُحتفَل به.

المصدر: `docs/03_dataset.md` القسم *3.7*

هذه القاعدة الأخيرة تُلخّص روح المشروع كلّه: الشكّ في النتيجة الجيدة جداً سلوك علمي، لا تشاؤم.

ثامناً: مجموعة بيانات ثانية للاختبار فقط

هناك مجموعة ثانية تماماً استُخدمت لاحقاً، هي **المجموعة الإيطالية للصوت والكلام في باركنسون**. والنقطة الحاسمة التي يجب أن تكرّرها: **النموذج لم يُدرَّب عليها أبداً**. استُخدمت للاختبار فقط، والنموذج كان مجمَّداً تماماً. تفاصيلها في الجزء 15.

خلاصة الجزء

- مجموعة MDVR-KCL فيها 73 تسجيلاً بصيغة WAV أحادية بمعدّل 44100 هرتز، من 37 شخصاً: 21 سليماً بـ42 تسجيلاً و16 مريضاً بـ31 تسجيلاً، في مهمّتين.
- التسمية قُرئت من مصدرين مستقلّين وقُورن بينهما، ومعرّف المتحدّث استُخرج من بداية اسم الملف، وهو أهم معلومة وصفية في المشروع.
- أُجريت سبعة فحوصات سلامة ونجحت كلّها، وسُجّلت خمسة مخاطر بصراحة، مع قاعدة مكتوبة مسبقاً بأن أي نتيجة تبلغ 95٪ تُعامَل كإنذار تسريب.

الجزء 4: المعالجة المسبقة

الآن صار عندنا 73 ملفاً، وكل ملف قائمة من ملايين الأرقام. هل نبدأ القياس فوراً؟ لا. أولاً يجب **تهيئة** هذه الأرقام، وهذه المرحلة تُسمّى المعالجة المسبقة (preprocessing).

لماذا نحتاجها أصلاً؟ تخيّل أنك تقارن طول شخصين في صورتين. الصورة الأولى مُلتقطة من قريب والثانية من بعيد، والأولى مائلة والثانية مستقيمة. لا يمكنك المقارنة قبل أن **توحّد الظروف**. والمعالجة المسبقة هي بالضبط هذا: توحيد ظروف كل التسجيلات حتى تصبح الاختلافات بينها اختلافات في **الصوت نفسه**، لا في طريقة التسجيل. والملف المسؤول عن هذه المرحلة هو `src/pvoice/preprocess.py`.

أولاً: الخطوات الخمس بترتيبها الثابت

تُطبَّق هذه الخطوات على كل تسجيل، **بنفس الترتيب دائماً**، لأن الترتيب يؤثّر على النتيجة.

الخطوة 1: التحويل إلى أحادي. إن كان الملف يحتوي على قناتين تأخذ متوسّطهما فتصير قناة واحدة. والسبب أن الأنبوب كلّه مبني على إشارة واحدة، وكل ملفات MDVR-KCL أحادية أصلاً، لكن هذه الخطوة موجودة لأن المستخدم في التطبيق قد يرفع ملفاً ثنائي القناة.

الخطوة 2: إعادة أخذ العيّنات إلى 44100 هرتز. خصائص مثل `jitter` تعتمد على **الدقّة الزمنية** في قياس طول كل دورة اهتزاز. فلو كان ملف بمعدّل 16000 هرتز وآخر بمعدّل 44100، لكانت دقّة قياس الزمن في الأول أقلّ بكثير، ولصارت أرقامهما غير قابلة للمقارنة. فتوحيد معدّل العيّنات يعني توحيد «المسطرة» التي نقيس بها.

الخطوة 3: إزالة الانزياح المستمر. تحسب متوسّط كل قيم الإشارة، ثم تطرح هذا المتوسّط من كل عيّنة. والانزياح المستمر (DC offset) إزاحة كهربائية ثابتة تضيفها أجهزة التسجيل غير المثالية إلى الإشارة كلّها، فبدل أن تتأرجح الإشارة حول الصفر، تتأرجح حول 0.02 مثلاً. ولماذا يضرّ؟ لأن خاصية `shimmer` تقيس اختلاف **قوّة** كل دورة عن التي تليها، فإذا كانت كل القيم مرفوعة بمقدار ثابت، انحرفت حسابات القوّة كلّها. **تشبيه:** ميزان يشير إلى 2 كيلوغرام وأنت لم تقف عليه بعد. كل وزن ستقيسه سيكون خاطئاً بمقدار 2، والحلّ أن تُصفّر الميزان أولاً.

الخطوة 4: قصّ الصمت من الأطراف. تحذف الصمت في بداية التسجيل ونهايته فقط، والعتبة المستخدمة: أي جزء تقلّ شدّته عن **30 ديسيبل** تحت أعلى قمّة يُعتبَر صمتاً. والسبب أن الصمت في البداية والنهاية لا يحمل أي معلومة عن المتحدّث، بل **يشوّه** إحصاءات التوقّفات. فلو ترك المتحدّث خمس ثوانٍ صمتاً في البداية، لظهر ذلك كأنه «توقّف طويل في الكلام». **وانتبه لدقّة الصياغة:** الحذف من **الأطراف فقط**، أما الصمت **داخل** الكلام فيبقى كما هو تماماً، لأنه معلومة طبية حقيقية، وهو بالضبط ما نقيسه في خصائص التوقّفات.

الخطوة 5: التطبيع إلى القمّة عند 0.95. تضرب الإشارة كلّها في عدد ثابت بحيث تصبح أعلى عيّنة فيها مساوية لـ0.95 بالضبط. والسبب أن المتحدّث قد يكون قريباً من الميكروفون أو بعيداً عنه، وقد يكون كسب الميكروفون عالياً أو منخفضاً. هذه فروق في **ظروف التسجيل**، لا في **الصوت**. ولماذا 0.95 وليس 1.0؟ لترك هامش أمان بسيط يمنع القَصّ.

المصدر: *docs/04_pipeline.md* القسم *4.2*

ثانياً: نقطة دقيقة قد تُسأل عنها

قد يقول أستاذ: «أنتم تقيسون shimmer وهو يقيس تغيّر **قوّة** الصوت، ثم تأتون وتضربون الإشارة كلّها في عدد ثابت لتطبيعها. ألا يُفسد هذا القياس؟»

والجواب، وهو مذكور صراحة في التوثيق، **لا، والسبب أن shimmer مقياس نسبي** (a relative measure). ولنفهم بمثال رقمي:

- افترض دورتين متتاليتين قوّتهما 0.40 و0.44.
- الفرق النسبي بينهما = 0.04 ÷ 0.40 = 10٪.
- الآن ضربنا كل شيء في 2 بسبب التطبيع: صارتا 0.80 و0.88.
- الفرق النسبي = 0.08 ÷ 0.80 = **10٪ أيضاً، لم يتغيّر.**

فالضرب في ثابت لا يغيّر **النسبة** بين دورة وأخرى، وهذا هو ما يقيسه shimmer فعلاً. لذلك التطبيع آمن تماماً.

ثالثاً: القرار الأهم، لماذا رفضنا إزالة الضجيج

هذا هو القرار العلمي الأبرز في هذه المرحلة، وأنصحك بشدّة أن تجعله شريحة مستقلّة في عرضك.

الفكرة التي تبدو ممتازة. إزالة الضجيج (noise reduction) تقنية معروفة ومتوفّرة في كل مكتبات الصوت. ومن البديهي أن يخطر ببالك: «التسجيلات فيها ضجيج خلفية، فلماذا لا ننظّفها لتتحسّن النتائج؟» والأساتذة **سيسألونك هذا السؤال بالتحديد**، فاستعدّ له.

لماذا هي في الحقيقة كارثة هنا. افهم كيف تعمل خوارزميات إزالة الضجيج: هي تبحث في الإشارة عن الأجزاء **غير المنتظمة وغير المتوقّعة**، وتعتبرها ضجيجاً، ثم **تُنعّمها وتزيلها**. والآن تذكّر ما الذي نقيسه نحن أصلاً:

- خاصية jitter تقيس **عدم انتظام** في توقيت الدورات.
- خاصية shimmer تقيس **عدم انتظام** في قوّة الدورات.
- خاصية HNR تقيس نسبة الجزء **غير المنتظم** إلى الجزء المنتظم.
- خاصية CPPS تقيس مدى وضوح البنية المنتظمة فوق الخلفية.

هل ترى المشكلة؟ **خصائصنا كلّها مقاييس لعدم الانتظام، وإزالة الضجيج وظيفتها حذف عدم الانتظام.** فلو طبّقناها، لكانت النتيجة أن صوت المريض المهموس غير المنتظم، وهو بالضبط الدليل الطبي الذي نبحث عنه، سيُنعَّم ويُحذَف، **فيبدو صوت المريض سليماً بشكل مصطنَع.**

المصدر: *docs/04_pipeline.md* القسم *4.2*

تشبيه يُثبّت الفكرة. تخيّل طبيباً يفحص صورة أشعّة، فوجد فيها بقعاً غير منتظمة، فقرّر أن «ينظّف» الصورة ببرنامج تحرير ليجعلها أوضح وأجمل. النتيجة؟ سيحذف الورم نفسه. عدم الانتظام هنا **ليس تشويشاً على الإشارة، بل هو الإشارة**.

ولهذا السبب اقتصر المشروع على **خمس عمليات آمنة فقط**، ولم يُطبَّق أي ترشيح عدواني على الإطلاق. وهذا القرار مكتوب أصلاً كقيد ثابت في ملف قواعد المشروع `CLAUDE.md`.

رابعاً: لماذا كل الإعدادات في ملف واحد؟

تخيّل أن قيمة «44100» مكتوبة في عشرة أماكن مختلفة في الكود، ثم قرّرتَ تغييرها فغيّرتها في تسعة أماكن ونسيتَ العاشر. النتيجة: النموذج تدرّب على إعداد، والتنبّؤ يستخدم إعداداً آخر، وهذا **خطأ صامت** لن يظهر كرسالة خطأ، بل كنتائج خاطئة لا تعرف سببها.

ولهذا وُضعت **كل** الإعدادات في `src/pvoice/config.py`: معدّل العيّنات، ومدى البحث عن الطبقة، وإعدادات `MFCC`، والمسارات، وكل شيء. فلا يمكن للتدريب والتنبّؤ أن ينحرفا عن بعضهما بصمت.

المصدر: *docs/01_overview_and_journey.md* القسم *1.3*

وهذا يخدم **القاعدة الذهبية للمشروع**: نفس الأنبوب بالضبط يُستخدم في التدريب وفي التقييم وفي التنبّؤ على ملف جديد. وسنرى في الجزء 14 أن هذه القاعدة ليست وعداً، بل **ضمانة ميكانيكية** مفروضة بالكود عبر ملف `models/pipeline_config.json`.

خلاصة الجزء

- المعالجة المسبقة توحّد ظروف التسجيل في خمس خطوات: التحويل إلى أحادي، ثم توحيد معدّل العيّنات، ثم إزالة الانزياح المستمر، ثم قصّ الصمت من الأطراف عند 30 ديسيبل، ثم التطبيع إلى قمّة 0.95.
- التطبيع آمن لخاصية `shimmer` لأنها مقياس **نسبي**: الضرب في ثابت لا يغيّر النسبة بين دورة وأخرى.
- رُفضت إزالة الضجيج رفضاً تاماً، لأن خصائص المشروع كلّها تقيس **عدم الانتظام**، وإزالة الضجيج وظيفتها حذفه، أي أنها ستمحو الدليل الطبي نفسه.

الجزء 5: التقطيع إلى مقاطع من عشر ثوانٍ

عندنا الآن تسجيل مهيّأ طوله دقيقتان تقريباً. لماذا لا نقيس الخصائص على التسجيل كاملاً دفعة واحدة ونستريح؟ في الواقع هذا **بالضبط** ما فعله المشروع في البداية، وفي المرحلة الرابعة كانت النتيجة 0.780. ثم جاء التقطيع فرفعها إلى 0.822. لنفهم لماذا.

أولاً: ما هو التقطيع بالضبط؟

بعد المعالجة المسبقة، يُقطَّع كل تسجيل إلى **مقاطع متتابعة طول كل منها عشر ثوانٍ**. والمقطع يُسمّى `chunk`. والقواعد المطبَّقة:

- طول المقطع: 10 ثوانٍ.
- المقطع الأخير إن كان أقصر من **5 ثوانٍ** يُهمَل ويُرمى.
- إن كان التسجيل كلّه أقصر من 5 ثوانٍ، يصير مقطعاً واحداً.

المصدر: `docs/04_pipeline.md` القسم *4.3*

مثال ملموس. لنأخذ تسجيلاً طوله 145 ثانية: المقاطع الكاملة 145 ÷ 10 = 14 مقطعاً تستهلك 140 ثانية، والمتبقّي 5 ثوانٍ بالضبط وهي عند الحدّ الأدنى فتُقبَل. ولو كان المتبقّي 3 ثوانٍ فقط لأُهمِل. ولماذا نهمله؟ لأن مقطعاً قصيراً جداً يعطي قياسات غير مستقرّة، وقد لا يحتوي على كلام كافٍ ليقيس `Praat` طبقة الصوت فيه أصلاً.

ثانياً: الأرقام الفعلية في مشروعك

- عدد التسجيلات: 73.
- عدد المقاطع الناتجة: **1001 مقطع**.
- عدد المقاطع لكل تسجيل: بين 7 و22 مقطعاً.

والتباين بين 7 و22 منطقي تماماً، لأن أقصر تسجيل 73 ثانية وأطول تسجيل 220 ثانية.

ثالثاً: لماذا يساعد التقطيع؟ ثلاثة أسباب

السبب الأول: قياسات أكثر استقراراً. عندما تقيس شيئاً مرة واحدة، فقياسك عرضة للصدفة. أما إن قِسته 14 مرة وأخذت المتوسّط، فالتقلّبات العشوائية يُلغي بعضها بعضاً، فيصير المتوسّط أقرب إلى القيمة الحقيقية. **تشبيه:** لو أردتَ معرفة ضغط دم

شخص، فهل تقيسه مرة واحدة، أم تقيسه في 14 لحظة مختلفة وتأخذ المتوسّط؟ والأمر نفسه هنا: لو صادف أن المتحدّث سعل في الثانية 40، فقياس التسجيل كاملاً سيتأثّر كلّه بذلك، أما مع التقطيع فمقطع واحد من 14 سيتأثّر، ويكاد أثره يختفي في المتوسّط.

السبب الثاني: مادة تدريب أكثر. تحوّلت 73 صفّاً من البيانات إلى **1001 صفّ**، أي أكثر من ثلاثة عشر ضعفاً. وهذا مهم جداً لأن أكبر مشكلة في مشروعك هي صِغَر العيّنة.

السبب الثالث: التقطيع يحترم تجميع المتحدّثين. وهنا **أخطر نقطة في هذا الجزء كلّه.** كل مقطع **يرث معرّف المتحدّث** من التسجيل الذي جاء منه. فالمقاطع الأربعة عشر المستخرَجة من تسجيل الشخص ID00 كلّها موسومة بأنها تخصّ ID00.

لماذا هذا حرج إلى هذه الدرجة؟ تخيّل أننا نسينا هذا الوسم، وتعاملنا مع الـ1001 مقطع كأنها 1001 عيّنة مستقلّة، ثم قسّمناها عشوائياً بين التدريب والاختبار. سيقع المقطع رقم 3 من تسجيل ID00 في بيانات التدريب، والمقطع رقم 7 من **نفس التسجيل** في بيانات الاختبار. وهما مقطعان من نفس الصوت، ونفس الميكروفون، ونفس الجلسة. عندها لن يتعلّم النموذج «كيف يبدو صوت مريض باركنسون»، بل سيتعلّم «كيف يبدو صوت هذا الشخص تحديداً»، ثم يجده في الاختبار فيصيب.

والتوثيق يصف نسيان هذه النقطة بأنه **كارثي**. وليس هذا كلاماً نظرياً، بل **مقيس بالأرقام**: نفس الأنبوب ونفس النموذج يعطي 0.909 بالتقسيم الخاطئ على مستوى المقاطع، مقابل 0.775 بالتقسيم الصحيح.

المصدر: *docs/04_pipeline.md* القسم *4.3* و*docs/05_validation_and_metrics.md* القسم *5.4*

رابعاً: كيف نعود من 1001 مقطع إلى 73 تسجيلاً؟

القاعدة المستخدمة بسيطة: **متوسّط الخصائص عبر المقاطع**، خاصية خاصية. وتُسمّى في ملف الإعدادات chunk_feature_mean.

مثال رقمي كامل. لنأخذ تسجيلاً له 14 مقطعاً: كل مقطع يُنتج 74 رقماً، إذن لدينا 14 × 74 = 1036 رقماً لهذا التسجيل. ثم نأخذ متوسّط كل خاصية عبر المقاطع الأربعة عشر، فيكون الناتج النهائي **74 رقماً فقط** تمثّل التسجيل كلّه. ولنُوضّح بخاصية واحدة، مثلاً f0_mean_hz: لو كانت قيمها في المقاطع 152 و148 و155 و150 وهكذا، فالقيمة النهائية للتسجيل هي متوسّط هذه الأرقام الأربعة عشر.

المصدر: *docs/04_pipeline.md* القسم *4.5*

والنقطة الحاسمة: عندما ترفع ملفاً جديداً في التطبيق، يحدث **الشيء نفسه بالضبط.** لا يوجد مسار مختلف للتنبّؤ. هذه هي «قاعدة الأنبوب الواحد».

خامساً: ملاحظة صادقة عن القيم الناقصة

في جدول المقاطع، هناك **عدد قليل من المقاطع**، بين 2 و16 مقطعاً حسب الخاصية، تحتوي على **قيم ناقصة** في خصائص الاضطراب مثل jitter وshimmer. والسبب أن هذه المقاطع لا تحتوي على صوت مُجهَّر مستقرّ بما يكفي ليتمكّن Praat من قياس الدورات فيها.

وماذا فُعل بها؟ لم تُملأ بقيمة عشوائية، ولم يُحذَف المقطع، بل تُركت ناقصة، ثم تُملأ لاحقاً بالوسيط، **لكن بشرط حاسم:** يُحسَب الوسيط من بيانات **التدريب فقط داخل كل طيّة**، ولا يُنظَر إلى بيانات الاختبار إطلاقاً.

المصدر: *docs/04_pipeline.md* القسم *4.7*

وللمقارنة: في جدول التسجيلات الكامل عدد القيم الناقصة **صفر**، لأن المتوسّط عبر 14 مقطعاً يعوّض المقاطع الناقصة.

سادساً: نتيجة التقطيع بالأرقام

- قبله (التسجيل كاملاً، 43 خاصية): الدقّة المتوازنة على مستوى الشخص = 0.780.
- بعده (مقاطع بمتوسّط، 74 خاصية): الدقّة المتوازنة = **0.822**.
- التحسّن: **+0.042**.

والملاحظة المهمة: هذا التحسّن جاء من **الخصائص الموسّعة المحسوبة على المقاطع ثم المتوسَّطة**، ولم يأتِ من ضبط المعاملات الفائقة الذي لم يفد إطلاقاً.

المصدر: `docs/06_experiments_and_results.md` القسم *6.4*

سابعاً: لماذا عشر ثوانٍ تحديداً؟

هذا سؤال مذكور صراحة في قائمة الأسئلة المتوقّعة. والإجابة: عشر ثوانٍ **طويلة بما يكفي** لإعطاء قياسات صوتية مستقرّة، فالمقطع يحتوي على مئات دورات الاهتزاز وعلى عدّة كلمات وتوقّفات. وهي في الوقت نفسه **قصيرة بما يكفي** لإنتاج حوالي 14 مقطعاً من التسجيل الواحد، وهو عدد جيد لأخذ متوسّط ذي معنى. ولو اخترنا ثانيتين لكانت القياسات غير مستقرّة، ولو اخترنا 60 ثانية لما حصلنا إلا على مقطعين أو ثلاثة، فضاعت فائدة المتوسّط.

ثم تُختَم الإجابة بالجملة الأهم: «**والأمر الحاسم أن المقاطع ترث معرّف المتحدّث، فيبقى التحقّق المجمَّع صحيحاً، لأن التقسيم على مستوى المقاطع بدون هذا التجميع هو بالضبط التسريب الذي أثبتناه بالقياس.**»

المصدر: `docs/12_presentation_qa.md`

خلاصة الجزء

- كل تسجيل يُقطَّع إلى مقاطع من عشر ثوانٍ والحدّ الأدنى المقبول خمس، فنتج من 73 تسجيلاً **1001 مقطع**، بين 7 و22 مقطعاً لكل تسجيل.
- التقطيع يفيد لثلاثة أسباب: استقرار أعلى بفضل المتوسّط، ومادة تدريب أكثر بثلاثة عشر ضعفاً، مع الحفاظ على معرّف المتحدّث في كل مقطع.
- تُجمَع المقاطع بقاعدة `chunk_feature_mean`، ونفس القاعدة تُطبَّق عند التنبّؤ على ملف جديد. والنتيجة تحسّن من 0.780 إلى 0.822.

الجزء 6: الخصائص الصوتية (1) النبرة والانتظام

وصلنا إلى قلب المشروع الهندسي. فبعد التهيئة والتقطيع، السؤال: **ماذا نقيس بالضبط في كل مقطع؟** تذكّر أن هذه هي **المساهمة الأساسية** لمشروعك: أنت لم تُحمّل جدول خصائص جاهزاً من الإنترنت، بل بنيت القياس بنفسك من الصوت الخام.

في هذا الجزء نغطّي **17 خاصية** من أصل 74، وكلّها تدور حول سؤال واحد: **كيف تهتزّ الحبال الصوتية، وهل اهتزازها منتظم؟** والأداة المستخدمة لحسابها كلّها هي `Praat`، وهو البرنامج المعياري عالمياً في تحليل الكلام العلمي، ونتعامل معه من `Python` عبر مكتبة `praat-parselmouth`.

أولاً: التردد الأساسي F0 وإحصاءاته السبع

F0 هو عدد مرات فتح وإغلاق الحبال الصوتية في الثانية. لكن انتبه: F0 **ليس رقماً واحداً للتسجيل**. فأثناء الكلام تتغيّر طبقة صوتك باستمرار، ترتفع عند السؤال وتنخفض عند نهاية الجملة. لذلك يقوم `Praat` **بتتبّع** قيمة F0 عبر التسجيل كلّه لحظة بلحظة، فينتج **منحنى** لا رقماً. ثم نلخّص هذا المنحنى في سبعة أرقام:

- `f0_mean_hz`: المتوسّط، أي الطبقة الوسطى لصوت الشخص.
- `f0_median_hz`: الوسيط، وهو أقلّ تأثّراً بالقيم الشاذّة من المتوسّط.
- `f0_std_hz`: الانحراف المعياري، أي مقدار **تذبذب** الطبقة حول متوسّطها.
- `f0_min_hz`: أدنى طبقة وصل إليها.
- `f0_max_hz`: أعلى طبقة وصل إليها.
- `f0_range_hz`: المدى، أي الفرق بين الأعلى والأدنى.
- `voiced_fraction`: النسبة المُجهَّرة، أي نسبة الوقت الذي اهتزّت فيه الحبال الصوتية أصلاً.

ومدى البحث المستخدم من 75 إلى 500 هرتز، وهو يغطّي أصوات الرجال والنساء البالغين جميعاً.

المصدر: *docs/04_pipeline.md* القسم *4.4*

ما معنى `voiced_fraction` تحديداً؟ ليس كل الكلام «مُجهَّراً». فبعض الحروف تُنطَق باهتزاز الحبال الصوتية مثل حرف الزاي، وبعضها بلا اهتزاز أصلاً بل بمجرّد احتكاك هواء مثل حرف السين. جرّب: ضع إصبعك على حنجرتك وقل «ززززز» ثم «سسسسس»، ستشعر بالاهتزاز في الأولى فقط. فهذه الخاصية تقيس كم نسبة التسجيل التي وُجد فيها اهتزاز أصلاً، وانخفاضها قد يدلّ على صوت مهموس أو ضعيف أو متقطّع.

الأهمية الطبية، وهنا المفاجأة. من أشهر علامات باركنسون **الكلام الرتيب**، وترجمة هذا رقمياً واضحة: **انخفاض `f0_std_hz`**

وانخفاض f0_range_hz. والآن الرقم المذهل في مشروعك: عندما سألنا النموذج النهائي «أي خاصية اعتمدتَ عليها أكثر؟»، كانت الإجابة:

- المرتبة الأولى: f0_range_hz بأهمية 0.069.
- المرتبة الثانية: f0_max_hz بأهمية 0.058.

المصدر: *docs/06_experiments_and_results.md* القسم *6.5*

مدى النبرة هو الخاصية الأهم في المشروع كلّه، وهذا يطابق تماماً أشهر وصف سريري لكلام مريض باركنسون. نتيجة مُرضية جداً لمشروع يشترط قابلية التفسير، ويجب أن تجعلها شريحة في عرضك.

تحذير: الطبقة المطلقة مسألة حسّاسة. انتبه للفرق بين نوعين من إحصاءات F0: إحصاءات **التغيّر** (f0_std_hz وf0_range_hz) لها معنى سريري مباشر، أما إحصاءات **القيمة المطلقة** فتعكس أساساً **جنس المتحدّث**. وفي بياناتك فعلاً: متوسّط F0 للمجموعة السليمة حوالي 180 هرتز، وللمجموعة المريضة حوالي 154 هرتز. هل هذا فرق مرضي؟ أم أن المجموعة السليمة فيها نساء أكثر؟ **لا نستطيع الجزم**، لأن البيانات لا تحتوي على جنس موثّق لكل شخص. ولهذا أُجريت تجربة خاصة لقياس مدى **اعتماد** النموذج على هذه الخصائص، وتفاصيلها في الجزء 11.

ثانياً: الارتجاف jitter، عدم انتظام التوقيت

خاصية jitter تقيس **عدم الانتظام في توقيت** دورات الحبال الصوتية. ولنفهم بمثال رقمي: تخيّل أن حبالك الصوتية تهتزّ، فكانت مدد الدورات المتتالية 5.0 مللي ثانية ثم 5.4 ثم 5.1. هذه الفروق الصغيرة بين دورة والتي تليها هي jitter. وتُعبَّر عادةً كنسبة مئوية لا كزمن مطلق، **حتى لا تعتمد على طبقة صوت المتحدّث**، فالرجل ذو الصوت الغليظ دوراته أطول أصلاً، ولو قِسنا بالمللي ثانية لصار مقياساً غير عادل بين الرجال والنساء.

التشبيه الأفضل: طبّال يضرب إيقاعاً ثابتاً. الطبّال الماهر تكون المسافات بين ضرباته متساوية تماماً، أما لو ارتجفت يده فبعض الضربات ستأتي مبكّرة قليلاً وبعضها متأخّرة. وjitter يقيس مقدار هذا الخلل.

الأنواع الأربعة في مشروعك:

- jitter_local: الفرق بين كل دورة والدورة التي تليها مباشرة، كنسبة.
- jitter_local_abs: الشيء نفسه لكن بالثواني لا كنسبة.
- jitter_rap: يأخذ المتوسّط على **ثلاث دورات** متجاورة.
- jitter_ppq5: يأخذ المتوسّط على **خمس دورات** متجاورة.

لماذا أربعة أنواع وليس نوعاً واحداً؟ الأنواع التي تأخذ متوسّطاً على عدّة دورات تكون **أقلّ حساسية للتقلّبات العشوائية القصيرة**. تخيّل أن دورة واحدة فقط شذّت بسبب ضجيج عابر: مقياس local سيلتقط هذا الشذوذ كاملاً، أما ppq5 فسيوزّعه على خمس دورات فيخفّ أثره. فنحن نقيس نفس الظاهرة بمستويات تنعيم مختلفة، وندع النموذج يقرّر أيّها أنفع.

المصدر: *docs/02_scientific_background.md* القسم *2.3*

والصوت السليم الممدود يكون jitter فيه عادةً **أقلّ من 1٪** تقريباً. وفي نموذجك النهائي جاءت خاصية jitter_local_abs في **المرتبة التاسعة** بقيمة 0.027.

ثالثاً: التذبذب shimmer، عدم انتظام القوّة

الفكرة نفسها تماماً، لكن مطبَّقة على **السعة** بدل **التوقيت**: كم تختلف **قوّة** كل دورة اهتزاز عن الدورة التي تليها؟ وبالعودة لتشبيه الطبّال: jitter هو «هل تأتي الضربات في مواعيدها؟»، وshimmer هو «هل كل ضربة بنفس القوّة؟». والطبّال المرتجف يخلّ بالاثنين معاً.

الأنواع الخمسة: shimmer_local، وshimmer_local_db بوحدة الديسيبل، وshimmer_apq3 بمتوسّط على ثلاث دورات، وshimmer_apq5 على خمس، وshimmer_apq11 على إحدى عشرة دورة وهو الأكثر تنعيماً.

الأهمية الطبية: الإغلاق الضعيف أو غير المتّسق للحبال الصوتية ينتج نفخات هواء متفاوتة القوّة، والأذن تسمع هذا على شكل صوت **خشن أو غير ثابت**. وهنا تظهر أهمية ما شرحناه في الجزء 4: لولا إزالة الانزياح المستمر لكانت كل حسابات shimmer منحرفة، ولولا كونه مقياساً **نسبياً** لأفسده التطبيع.

رابعاً: نسبة التوافقيات إلى الضجيج HNR

هذه خاصية واحدة فقط اسمها hnr_mean_db، لكنها من أهم مقاييس جودة الصوت. والفكرة أن صوت الإنسان ليس نقيّاً تماماً، بل هو **خليط من مكوّنين**:

- مكوّن **منظّم ودوري**: وهو الاهتزاز المنتظم للحبال الصوتية ومضاعفاته، ويُسمّى التوافقيات (harmonics). هذا هو «النغمة».
- مكوّن **غير منظّم**: وهو الهواء المضطرب المتسرّب. هذا هو «الضجيج».

وخاصية HNR هي **نسبة الأول إلى الثاني**، مقيسة بالديسيبل. فصوت نظيف قوي قد يبلغ 20 ديسيبل أو أكثر، وصوت مهموس أو مبحوح أقلّ بكثير.

المصدر: *docs/02_scientific_background.md* القسم *2.3*

الأهمية الطبية: هذه هي الخاصية التي تقيس **الهَمْس** مباشرة. تذكّر الآلية الفيزيائية من الجزء 1: في باركنسون قد لا تنغلق الحبال الصوتية انغلاقاً كاملاً بسبب ضعف التحكّم العضلي، فتبقى فتحة صغيرة يتسرّب منها هواء بلا اهتزاز، وهذا الهواء المتسرّب **ضجيج خالص**. فالنتيجة نغمة أقلّ وضجيج أكثر، أي **انخفاض HNR**. وهذه الخاصية مثال ممتاز على الترابط بين الفسيولوجيا والرقم: عيب ميكانيكي محدّد في إغلاق الحنجرة يؤدّي إلى انخفاض قابل للقياس في رقم واحد.

خامساً: قيد صادق يجب أن تذكره بنفسك

هذه النقطة مهمة جداً لمصداقيتك، وأنصحك أن **تذكرها أنت قبل أن يسألك أحد**.

خصائص jitter وshimmer وHNR صُمّمت أصلاً لـ**حرف الممدود** (sustained vowel)، أي أن يقول الشخص «آآآآآه» بثبات لعدّة ثوانٍ. وفي هذه الحالة **يُفترَض** أن تكون كل دورة مطابقة للتي قبلها، فأي عدم انتظام يكون ذا معنى طبي مباشر.

لكن بياناتك تحتوي على **كلام متّصل** (continuous speech): قراءة نصّ وحوار. وفي الكلام المتّصل تتغيّر الطبقة والقوّة باستمرار **لأسباب لغوية طبيعية تماماً**، فأنت ترفع صوتك عند التشديد وتخفضه عند نهاية الجملة. والنتيجة أن هذه المقاييس الثلاثة تصير **أكثر ضجيجاً وأقلّ نقاءً سريرياً** ممّا هي عليه في عيادة الصوت.

المصدر: *docs/02_scientific_background.md* القسم *2.4*

وهذا القيد ليس عيباً مخفيّاً، بل هو **السبب المباشر** في إضافة خصائص أخرى في مرحلة التحسين: CPPS وإحصاءات التوقّفات والخصائص الطيفية، وهي قصّة الجزء القادم.

خلاصة الجزء

- إحصاءات F0 السبع تصف طبقة الصوت، وأهمّها سريرياً مقاييس **التغيّر** لأن الرتابة علامة باركنسونية كلاسيكية، وفعلاً جاء `f0_range_hz` **الخاصية الأهم في النموذج كلّه** بأهمية 0.069.
- `jitter` بأنواعه الأربعة يقيس عدم انتظام **التوقيت**، و`shimmer` بأنواعه الخمسة يقيس عدم انتظام **القوّة**، وتعدّد الأنواع سببه اختلاف درجة التنعيم.
- HNR يقيس نسبة النغمة المنتظمة إلى الضجيج المضطرب، وانخفاضه هو الترجمة الرقمية المباشرة لعدم اكتمال إغلاق الحبال الصوتية. والقيد الصادق أن هذه المقاييس الثلاثة صُمّمت للحرف الممدود لا للكلام المتّصل.

الجزء 7: الخصائص الصوتية (2) الجودة والإيقاع والطيف

في الجزء السابق غطّينا 17 خاصية تدور حول اهتزاز الحبال الصوتية. والآن نغطّي **57 خاصية** متبقّية، وكلّها أُضيفت أو صُمّمت لأنها تعمل جيداً على **الكلام المتّصل**، وهو بالضبط القيد الذي ختمنا به الجزء الماضي. وتنقسم إلى ثلاث عائلات: CPPS خاصية واحدة، وإحصاءات التوقّفات أربع، ومعاملات MFCC ومشتقّاتها اثنتان وخمسون.

أولاً: CPPS، أذكى مقياس في المشروع

اسمها الكامل بروز القمّة الكبسترالية المُنعَّم (smoothed Cepstral Peak Prominence)، واسمها في الكود cpps_db. والتوثيق يصفها بأنها **أكثر المقاييس تقدّماً تقنياً في المشروع، وأنسبها للكلام المتّصل**.

الخطوة 1: ما هو الطيف؟ الطيف (spectrum) هو وصف الصوت بدلالة **كمّية الطاقة عند كل تردد**. تخيّل معادِل الصوت في مشغّل موسيقى: أعمدة ترتفع وتنخفض، كل عمود يمثّل تردداً. والآن النقطة المهمة: عندما يهتزّ إنسان بصوت منتظم عند 150 هرتز، فإن طيفه **لا يحتوي على قمّة واحدة عند 150**، بل على سلسلة قمم عند 150 و300 و450 و600، أي عند مضاعفات التردد الأساسي. وهذه القمم هي **التوافقيات**. والملاحظة الحاسمة أن هذه القمم **متباعدة بانتظام**، والمسافة بينها ثابتة تساوي F0.

الخطوة 2: ما هو الكبستروم؟ الكبستروم (cepstrum) هو ناتج تحليل **طيف الطيف**. أي أننا نأخذ الطيف، الذي هو نفسه رسم بياني، ونعامله كأنه إشارة جديدة، ثم نحلّله مرة أخرى. لاحظ حتى الاسم: كلمة cepstrum هي كلمة spectrum بعد قلب أحرفها الأولى، وهي دعابة متعمّدة من مخترعي المفهوم. ولماذا نفعل هذا؟ لأن التحليل الطيفي وظيفته اكتشاف **التكرار المنتظم**، والطيف يحتوي على قمم متكرّرة بانتظام، فتحليل الطيف يكشف هذا الانتظام وينتج عنه **قمّة واحدة حادّة** في الكبستروم.

الخطوة 3: ما الذي تقيسه CPPS بالضبط؟ تقيس **كم ترتفع تلك القمّة فوق الخلفية المحيطة بها؟** فصوت قوي منتظم توافقياته واضحة ومتباعدة بانتظام تام، فالقمّة الكبسترالية **حادّة وبارزة**، أي CPPS مرتفعة. وصوت مهموس أو مليء بالضجيج انتظامه مشوّش، فالقمّة **باهتة وضائعة في الخلفية**، أي CPPS منخفضة.

المصدر: *docs/02_scientific_background.md* القسم *2.3*

لماذا هي متفوّقة على jitter وshimmer هنا؟ لكي يحسب Praat قيمة jitter، يجب أولاً أن **يتعرّف على كل دورة اهتزاز على حدة** ويقيس طولها بدقّة، وهذه العملية **تفشل كثيراً في الكلام الجاري** حيث الطبقة تتغيّر باستمرار والحروف الساكنة تقطع الاهتزاز. أما CPPS فلا تحتاج ذلك إطلاقاً، فهي لا تسأل «كم طول الدورة رقم 47؟»، بل تسأل سؤالاً إجمالياً: «هل يوجد انتظام توافقي واضح في هذا الصوت أم لا؟». ولهذا تُعتبَر في الأدبيات السريرية **أمتن مؤشّر صوتي على الهَمْس وعلى عُسر الصوت عموماً**.

والإعداد المستخدم في مشروعك: أرضية الطبقة عند 60 هرتز. وملاحظة عملية: حساب CPPS هو **الجزء الأبطأ** في المشروع كلّه،

وهو السبب الرئيسي في أن تحليل ملف واحد في التطبيق يستغرق حوالي دقيقة.

المصدر: `docs/07_prototype.md` القسم *7.2*

ثانياً: إحصاءات التوقّفات، قياس الإيقاع لا الصوت

هذه العائلة مختلفة جوهرياً عن كل ما سبق: فهي لا تقيس **جودة** الصوت، بل **إيقاع** الكلام وتنظيمه الزمني.

كيف تُحسَب؟ باستخدام كشف قائم على الطاقة، يُقسَّم التسجيل إلى فترات كلام وفترات صمت، وأي فجوة صمت طولها **0.2 ثانية أو أكثر** تُعتبَر توقّفاً، والعتبة المستخدمة 25 ديسيبل تحت القمّة.

الخصائص الأربع:

- `pause_count_per_min`: عدد التوقّفات في الدقيقة.
- `pause_mean_s`: متوسّط طول التوقّف بالثواني.
- `pause_ratio`: نسبة الوقت الكلّي المقضيّ في التوقّف.
- `speech_seg_mean_s`: متوسّط طول فترة الكلام المتّصل بين توقّفين.

المصدر: `docs/04_pipeline.md` القسم *4.4*

الأهمية الطبية: من أعراض باركنسون صعوبة **بدء** الحركة والاستمرار فيها، وهذا يظهر في الكلام على شكل تردّد وتوقّفات في غير مواضعها الطبيعية، وإيقاع كلام مفكّك.

ميزة إضافية ثمينة جداً: خصائص التوقّفات تعتمد على **التوقيت**، لا على **محتوى التردد**، ولهذا فهي **متينة نسبياً أمام تغيّر الميكروفون**. لماذا؟ لأن ميكروفوناً رخيصاً وآخر غالياً قد يختلفان كثيراً في تسجيل الترددات العالية، لكن كليهما سيتّفقان تماماً على أن المتحدّث صمت لمدة 0.4 ثانية. الزمن هو الزمن. احفظ هذه النقطة، لأنها ستفسّر لك جزءاً مهماً من نتائج الجزء 15.

ثالثاً: معاملات MFCC، وصف شكل الطيف

هذه أكبر عائلة عدداً: **52 خاصية** من أصل 74، أي أكثر من ثلثَي المشروع. واسمها الكامل معاملات الكبستروم الميلية التردد (Mel-Frequency Cepstral Coefficients)، ووظيفتها وصف **شكل الطيف** وصفاً مضغوطاً.

ما معنى «شكل الطيف»؟ عد إلى الجزء 1 واستعد المراحل الثلاث: الحبال الصوتية تُنتج طنيناً خاماً، ثم يُرشّحه **مجرى الصوت** فيصير حرفاً مفهوماً. والآن الفكرة: عندما تقول «آ» ثم «إي»، فإن **الطنين الخام من حنجرتك لم يتغيّر كثيراً**، لكن **شكل فمك تغيّر**، فغيّر أي الترددات تُقوّى وأيّها تُضعَّف. أي أن الفرق بين «آ» و«إي» فرق في **شكل الطيف**. فخصائص MFCC تلتقط بالضبط **الأثر الترشيحي لمجرى الصوت**، أي وضعية اللسان والفم والشفتين.

لماذا كلمة «مِيلي»؟ مقياس Mel مقياس تردد مُعدَّل ليحاكي **السمع البشري**. فالأذن لا تسمع الترددات بشكل خطّي: الفرق بين 100 و200 هرتز تسمعه فرقاً كبيراً، أما الفرق بين 10000 و10100 هرتز فلا تكاد تميّزه، رغم أن الفرق العددي واحد. فمقياس Mel يعطي تفصيلاً أكبر للترددات المنخفضة وأقلّ للعالية، تماماً كما تفعل أذنك.

كيف تصير 52 رقماً؟ يُحسَب **13 معاملاً**، وهو العدد المعياري في هذا المجال. ولكل معامل نحسب المتوسّط والانحراف المعياري

عبر الوقت، أي 13 × 2 = **26 رقماً**. ثم نحسب **مشتقّاتها** delta-MFCC بنفس الطريقة، أي 26 رقماً آخر. والمجموع **52 خاصية**. والإعدادات المستخدمة: نافذة تحليل 2048 عيّنة (حوالي 46 مللي ثانية)، وخطوة 512 عيّنة (حوالي 12 مللي ثانية). والأداة المستخدمة هنا مكتبة librosa لا Praat.

المصدر: *docs/04_pipeline.md* القسم *4.4*

ما هي delta-MFCC تحديداً؟ إن كان MFCC يصف **شكل** الطيف في لحظة، فإن delta-MFCC يصف **سرعة تغيّر** ذلك الشكل من لحظة لأخرى. **تشبيه:** لو صوّرنا سيارة، فإن MFCC يقول «أين هي الآن»، وdelta-MFCC يقول «بأي سرعة تتحرّك». وترجمتها الفسيولوجية أنها تصف **ديناميكية النطق**، أي بأي سرعة يتحرّك لسانك وشفتاك من وضعية حرف إلى وضعية الحرف التالي.

الأهمية الطبية: تذكّر عَرَض «الحروف الساكنة غير الدقيقة» من الجزء 1. سببه أن حركات اللسان والشفتين صارت **أبطأ وأصغر**، فلا تصل إلى الوضع الصحيح تماماً قبل أن تنتقل إلى الحرف التالي. وهذا يظهر في مكانين: في MFCC شكل طيفي **مختلف** لأن اللسان لم يصل إلى موضعه، وفي delta-MFCC تغيّر طيفي **أبطأ** لأن الحركة نفسها أبطأ.

ماذا قال النموذج فعلاً؟ من قائمة أهمية الخصائص:

- المرتبة الثالثة: mfcc2_std بأهمية 0.054.
- المرتبة الرابعة: mfcc6_mean بأهمية 0.048.
- المرتبة الخامسة: mfccD2_std بأهمية 0.036.
- المرتبة السادسة: mfccD6_std بأهمية 0.036.
- المرتبة الثامنة: mfcc4_std بأهمية 0.028.

المصدر: *docs/06_experiments_and_results.md* القسم *6.5*

لاحظ نمطاً واضحاً: خصائص **الانحراف المعياري** (_std) تحتلّ مراتب أعلى من خصائص **المتوسّط** (_mean). وترجمة ذلك أن النموذج يهتمّ بـ**مدى تنوّع** شكل الطيف أكثر من اهتمامه بشكله الوسطي، وهذا **متّسق تماماً** مع قِلّة سعة الحركة النطقية: المريض لا ينطق «خطأً»، بل ينطق بتنوّع **أقلّ**. وهذه ملاحظة ممتازة لتقولها في المناقشة، لأنها تربط رقماً في جدول بعَرَض سريري.

رابعاً: تحذير صريح بشأن MFCC

خصائص MFCC **حسّاسة جداً** لثلاثة أشياء ليس لها علاقة بالمرض:

- **الميكروفون:** لكل ميكروفون بصمته الترددية الخاصة.
- **الغرفة:** صدى الغرفة يغيّر الطيف.
- **اللغة:** لأن كل لغة لها مجموعة أصوات ووضعيات لسان مختلفة.

والتوثيق يقول صراحة إن هذا **سبب رئيسي في انخفاض الأداء على البيانات الإيطالية**.

المصدر: *docs/08_external_validation.md* القسم *8.7*

وهنا تكتمل الصورة: مشروعك يحتوي على نوعين من الخصائص. خصائص **متينة عبر البيئات** مثل التوقّفات التي تعتمد على

الزمن، وإلى حدّ ما jitter وshimmer وHNR وCPPS التي تعتمد على فسيولوجيا الحنجرة لا على اللغة. وخصائص **هشّة عبر البيئات** مثل MFCC وdelta-MFCC التي تعتمد على القناة واللغة. وهذا التقسيم يفسّر لك مسبقاً لماذا انخفضت النتيجة من 0.822 داخلياً إلى 0.629 خارجياً، ولماذا لم تنهَر إلى 0.5 أيضاً.

خامساً: الجرد الكامل للخصائص الـ74

مشروعك يميّز بين مجموعتين. **المجموعة الأساسية** (base feature set) وفيها 43 خاصية، استُخدمت في المرحلتين الثالثة والرابعة:

- إحصاءات F0: 7
- الارتجاف jitter: 4
- التذبذب shimmer: 5
- نسبة HNR: 1
- معاملات MFCC: 26

و**المجموعة الموسّعة** (extended feature set) وفيها 74 خاصية، وهي المستخدمة في النموذج النهائي، وتضيف 31 خاصية:

- خاصية CPPS: 1
- إحصاءات التوقّفات: 4
- مشتقّات delta-MFCC: 26

المصدر: *docs/04_pipeline.md* القسم *4.4*

وانتبه للترتيب الزمني هنا، فهو يحكي قصّة: الخصائص الإحدى والثلاثون الإضافية **لم تكن موجودة في البداية**، بل أُضيفت في مرحلة التحسين تحديداً لمعالجة قيد «الكلام المتّصل»، ونتيجتها كانت رفع الأداء من 0.780 إلى 0.822.

سادساً: فحص جودة جدول الخصائص

بعد بناء الجدول فُحص آلياً، وكانت النتائج:

- قيم ناقصة في جدول التسجيلات: صفر.
- قيم لا نهائية: صفر.
- قيم خارج المدى الفيزيائي المعقول: صفر.
- خصائص ثابتة لا تتغيّر: صفر.
- متّجهات خصائص مكرّرة: صفر.

المصدر: *docs/04_pipeline.md* القسم *4.7*

وهناك تفصيلة لطيفة تستحقّ الذكر لأنها تُظهر دقّة الفحص: أحياناً يُبلّغ Praat عن قيمة f0_min_hz أقلّ قليلاً من أرضية البحث 75 هرتز، مثل 74.95 هرتز. وهذا ليس خطأً، بل نتيجة أن الخوارزمية **تُقحِم** قيماً بين إطارات التحليل. ولهذا يسمح الفاحص بهامش تسامح 2٪ بدل أن يرفض القيمة.

خلاصة الجزء

- CPPS تحلّل «طيف الطيف» لتقيس مدى بروز الانتظام التوافقي، وهي متفوّقة على jitter وshimmer في الكلام المتّصل لأنها **لا تحتاج التعرّف على كل دورة اهتزاز**.
- إحصاءات التوقّفات الأربع تقيس إيقاع الكلام لا جودته، وميزتها أنها تعتمد على **الزمن** لا على التردد، فهي متينة نسبياً أمام تغيّر الميكروفون.
- معاملات MFCC وdelta-MFCC تصف شكل الطيف وسرعة تغيّره، وقد جاءت خصائص std_ أهمّ من mean_، لكنها **حسّاسة للميكروفون والغرفة واللغة**، وهذا سبب رئيسي في انخفاض الأداء الخارجي.

الجزء 8: ما هو تعلّم الآلة، وما هي نماذجنا

أولاً: كيف «يتعلّم» البرنامج أصلاً؟

في البرمجة التقليدية أنت تكتب القاعدة: «إذا كان jitter أكبر من 1٪ فقُل مريض». والمشكلة أن أحداً لا يعرف القاعدة الصحيحة، ولا يعرف كيف تتفاعل 74 خاصية معاً.

في تعلّم الآلة (machine learning) نعكس الاتجاه: نعطي البرنامج **أمثلة مع أجوبتها**، ونتركه يستنتج القاعدة بنفسه. وفي مشروعك، المثال الواحد هو **متّجه خصائص** (feature vector): قائمة من 74 رقماً تمثّل تسجيلاً واحداً. والجواب هو **التسمية**: إمّا HC أو PD. وهذه المسألة تُسمّى **تصنيفاً ثنائياً** (binary classification)، أي فئتان فقط.

تشبيه: طبيب متدرّب يرى ألف حالة مع تشخيصاتها. لا أحد أعطاه قاعدة مكتوبة، لكنه بعد ألف حالة صار يميّز الأنماط. هذا هو التدريب (training).

مرحلتان منفصلتان تماماً:

- **التدريب**: نُري النموذج أمثلة مع أجوبتها، فيضبط أرقامه الداخلية.
- **الاختبار** (testing): نُعطيه أمثلة **لم يرَها أبداً** بلا أجوبة، ونقارن تنبّؤه بالحقيقة.

وكل الجزء 9 مخصّص لشرح كيف نضمن أن الاختبار نزيه فعلاً.

ثانياً: الأنبوب، ثلاث خطوات معاً

كل نموذج في مشروعك ليس نموذجاً وحده، بل **أنبوب** (pipeline) من ثلاث خطوات متسلسلة: سدّ القيم الناقصة بالوسيط، ثم التقييس، ثم المصنِّف.

- **سدّ القيم الناقصة** (median imputation): يملأ القيم المفقودة بالوسيط. تذكّر من الجزء 5 أن بعض المقاطع لم يستطع Praat قياس jitter فيها.
- **التقييس** (standardisation): يعيد ضبط الخصائص إلى مدى متقارب.
- **المصنِّف** (classifier): النموذج نفسه.

لماذا التقييس ضروري؟ انظر إلى أرقامك: قيمة f0_mean_hz حوالي 180، وقيمة jitter_local حوالي 0.02. فلو حسب النموذج «المسافة» بين شخصين، لطغى فرق الهرتزات على فرق الارتجاف طغياناً كاملاً، ليس لأن الطبقة أهم، بل لأن **أرقامها أكبر عددياً** فقط. والتقييس يزيل هذا الظلم.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.5*

ثالثاً: الخط الأساسي، نموذج غبيّ عمداً

قبل النماذج الحقيقية، هناك نموذج اسمه **خط الأساس بفئة الأغلبية** (`majority-class baseline`). ووظيفته أنه **لا يتعلّم شيئاً إطلاقاً**: يقول دائماً اسم الفئة الأكثر شيوعاً. وفي بياناتك سيقول «سليم» في كل مرة.

لماذا نبنيه؟ لأنه يعطينا **أرضية للمقارنة**. فلو حقّق نموذجك الحقيقي 0.60 وحقّق الغبيّ 0.575، لعرفت أن نموذجك بلا قيمة تقريباً.

والنتيجة الفعلية كانت درساً بليغاً: الخط الأساسي حقّق **0.575 دقّة عادية** بينما التقط **صفراً من المرضى**. وهذا أوضح برهان ممكن على أن مقياس «الدقّة» وحده مضلّل.

المصدر: `docs/06_experiments_and_results.md` القسم *6.2*

رابعاً: النماذج الأربعة

الانحدار اللوجستي (`Logistic Regression`): أبسط نموذج وأكثرها قابلية للتفسير. يرسم **حدّاً مستقيماً** بين المجموعتين، ويعطي كل خاصية وزناً يعبّر عن أهميتها واتجاه تأثيرها. قوّته أنك تستطيع أن تنظر في أوزانه وتقول «هذه الخاصية تدفع نحو PD وتلك تدفع نحو HC». وضعفه أنه إن كان الفصل الحقيقي منحنياً فلن يستطيع تمثيله.

آلة المتّجهات الداعمة (SVM مع نواة RBF): تبحث عن الحدّ الفاصل الذي يترك **أوسع هامش ممكن** بين المجموعتين. **التشبيه:** لو أردت رسم خطّ يفصل بين قريتين، فلن ترسمه ملاصقاً لبيوت إحداهما، بل في منتصف المسافة تماماً، لأن الخطّ الأبعد عن الطرفين هو الأمتن أمام حالات جديدة. و**النواة** (`RBF kernel`) حيلة رياضية تسمح لهذا الحدّ بأن يكون **منحنياً** بدل مستقيم.

الغابة العشوائية (`Random Forest`) وهو نموذجك النهائي: تُبنى من **300 شجرة قرار**. وشجرة القرار سلسلة أسئلة نعم ولا: «هل `f0_range_hz` أقلّ من 90؟ إذا نعم، فهل `hnr_mean_db` أقلّ من 12؟» حتى تصل إلى جواب. والفكرة الذكية أن كل شجرة تُدرَّب على **جزء عشوائي** من البيانات و**جزء عشوائي** من الخصائص، فتصير كل شجرة مختلفة عن أختها، ثم **يصوّتن جميعاً**. **التشبيه:** بدل استشارة طبيب واحد، تستشير 300 طبيباً درس كل منهم جانباً مختلفاً، ثم تأخذ رأي الأغلبية، فأخطاء الأفراد تُلغي بعضها. **وميزتها الحاسمة لمشروعك** أنها تستطيع أن تخبرك **أي الخصائص كانت الأهم**، وهذا بالضبط ما أنتج قائمة الجزء 13.

الشبكة العصبية الصغيرة (MLP): شبكة عصبية بسيطة بطبقتين مخفيّتين فيهما 32 و16 وحدة. أُدرجت للاكتمال فقط. وسنرى في الجزء 11 أنها **أُسقطت** من المشروع بعد أن كشفت تجربة أنها كانت تعتمد على عامل مُربك.

المصدر: `docs/06_experiments_and_results.md` القسم *6.1*

خامساً: وزن الفئات المتوازن

كل النماذج عدا الخط الأساسي تستخدم إعداد `class_weight="balanced"`. ومعناه أنه بما أن المرضى أقلّ عدداً (31 مقابل 42)، فإن الخطأ في تصنيف مريض **يُحسَب بوزن أكبر** أثناء التدريب. وهذا يمنع النموذج من الميل الكسول نحو «قُل سليم دائماً».

سادساً: لماذا لم نستخدم التعلّم العميق؟

ثلاثة أسباب جاهزة:

- **سبعة وثلاثون شخصاً قليل جداً** لتدريب شبكة عميقة دون إفراط شديد في الملاءمة.
- **المشروع يشترط قابلية التفسير**: نستطيع أن نقول إن مدى النبرة هو الخاصية الأهم ونربطه بعلامة سريرية معروفة. الشبكة العميقة لا تعطيك ذلك.
- **تجربتنا نفسها أثبتت الخطر**: حتى الشبكة الصغيرة MLP انهارت من 0.671 إلى 0.526 عند حذف خصائص الطبقة، ممّا كشف اعتمادها على عامل مُربك.

المصدر: *docs/12_presentation_qa.md*

خلاصة الجزء

- تعلّم الآلة يعطي البرنامج أمثلة مع أجوبتها فيستنتج القاعدة، والمثال عندك متّجه من 74 رقماً وجوابه HC أو PD.
- كل نموذج يعمل داخل أنبوب من ثلاث خطوات، والتقييس ضروري لأن الخصائص بمقاييس متباعدة جداً.
- الخط الأساسي الغبيّ يحقّق 0.575 دقّة عادية بصفر مرضى مكتشفين، والنماذج الأربعة هي الانحدار اللوجستي وSVM والغابة العشوائية وMLP.

الجزء 9: أهم فكرة في المشروع، التحقّق المستقل عن المتحدّث

هذا هو **قلب المشروع العلمي**. إن لم تفهم إلا جزءاً واحداً فليكن هذا.

أولاً: الفكرة الأساسية

نتيجة أي نموذج لا معنى لها إلا بقدر نزاهة الطريقة التي اختُبر بها.

التشبيه المرجعي: لو أعطى أستاذ طلابه أسئلة الامتحان وأجوبتها قبل الامتحان، فحصلوا كلّهم على الدرجة الكاملة، فهل يعني ذلك أنهم فهموا المادة؟ إطلاقاً. الدرجة الكاملة هنا **لا تُثبت شيئاً**. واختبار نموذج على بيانات تدرّب عليها هو **هذا بالضبط**.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.1*

ثانياً: التحقّق المتقاطع

عندنا 37 شخصاً فقط. لو عزلنا مجموعة اختبار ثابتة، لضيّعنا بيانات ثمينة، ولاعتمدت النتيجة اعتماداً كبيراً على **من وقع في تلك المجموعة صدفة**. والحلّ هو **التحقّق المتقاطع بـK طيّات** (`K-fold cross-validation`):

- نقسّم البيانات إلى 5 أجزاء تُسمّى طيّات (`folds`).
- ندرّب على 4 ونختبر على الخامسة.
- نكرّر 5 مرات، بحيث تكون كل طيّة مجموعة اختبار **مرة واحدة بالضبط**.
- فيحصل كل مثال على تنبّؤ من نموذج **لم يرَه أبداً**، وتُسمّى هذه **تنبّؤات خارج الطيّة** (`out-of-fold predictions`).

ثم نكرّر العملية كلّها بـ**ثلاث بذور عشوائية** مختلفة (`seeds`): 42 و7 و2025. والبذرة رقم يتحكّم بالاختيارات العشوائية، فتغييره ينتج توزيعاً مختلفاً للأشخاص على الطيّات. **ولماذا ثلاث بذور؟** لنعرف كم من النتيجة **حظّ** وكم منها **قدرة حقيقية**.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.2*

ثالثاً: الجزء الحاسم، التجميع حسب المتحدّث

كل شخص في بياناتك ساهم بتسجيلين. ولو انفصل تسجيلاه، واحد في التدريب وواحد في الاختبار، فماذا يحدث؟ سيتعرّف النموذج على **الشخص**: جَرْس صوته الفريد، وميكروفونه، وغرفته. لا على **المرض**. والنتيجة: نتيجة ممتازة على هذه البيانات، وفشل تامّ على أي شخص جديد. وهذا يُسمّى **تسريب البيانات** (`data leakage`)، وهو الخطر المركزي الذي يحرسه المشروع كلّه.

الحلّ: نُعلن معرّف المتحدّث على أنه **المجموعة** (`group`)، فيضمن المقسِّم `StratifiedGroupKFold` أن كل تسجيلات الشخص الواحد تبقى معاً على جانب واحد من كل تقسيم. وكلمة `Stratified` تعني أنه يحافظ أيضاً على **نسبة متقاربة** بين HC وPD في كل طيّة.

الحزام والحمّالة معاً. والنقطة التي أنصحك بذكرها في المناقشة: **الكود لا يكتفي بالثقة في المكتبة**. ففي كل طيّة يحسب الكود التقاطع بين أشخاص التدريب وأشخاص الاختبار، وإن لم يكن فارغاً **يُوقف البرنامج كلّه** برسالة خطأ. وهذا التأكيد **لم يُفعَّل ولا مرة واحدة**، لكن وجوده يحوّل النزاهة من وعد إلى **ضمانة ميكانيكية**.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.3*

رابعاً: الدليل المقيس، لا المُدّعى

وهنا أقوى ورقة في مشروعك كلّه. شُغِّل **نفس الأنبوب بالضبط** و**نفس النموذج بالضبط** مرتين، ولم يتغيّر إلا **طريقة التقسيم**:

- التقسيم الصحيح، مجمَّع حسب المتحدّث: الدقّة المتوازنة = **0.775**.
- التقسيم الخاطئ، مقاطع عشر ثوانٍ عشوائية: الدقّة المتوازنة = **0.909**.

الفارق **0.13+**. وهذا الفارق **ليس قدرة على الكشف**، بل هو النموذج وهو يتعرّف على أن هذا المقطع يخصّ متحدّثاً حفظ جزءاً من تسجيله. والصورة `reports/figures/validation_comparison.png` هي أهم صورة في المشروع، واجعلها شريحة مستقلّة.

وهذا هو السبب المباشر في أن كثيراً من الأبحاث المنشورة تُعلن 95٪ إلى 99٪ على هذا النوع من البيانات: **هي تقسّم على مستوى التسجيل أو النافذة، لا على مستوى الشخص**.

تفصيلة صادقة تزيد مصداقيتك. هناك أمر تحقّقنا منه بأمانة: التقسيم العشوائي على مستوى **التسجيل** لا يضخّم النتيجة تقريباً (0.807 مقابل 0.806). **لماذا؟** لأن تسجيلَي الشخص الواحد مهمّتان مختلفتان، قراءة نصّ وحوار عفوي، فليسا متشابهين جداً. فالتسريب الكارثي يقع تحديداً على مستوى **المقاطع**. والمشروع يذكر هذه النسخة الدقيقة بدل الادّعاء الأكثر إثارة والأقلّ دقّة.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.4*

خامساً: حماية ثانية، المعالجة داخل الطيّة

هناك نوع أخفى من التسريب، وفهمك له سيُبهر أي أستاذ. خطوتان في الأنبوب **تتعلّمان من البيانات**: سدّ القيم الناقصة يحتاج حساب الوسيط، والتقييس يحتاج حساب المتوسّط والانحراف المعياري.

فلو حسبنا هذا الوسيط على **كل** البيانات دفعة واحدة، لكان الوسيط قد «رأى» بيانات الاختبار، فتسرّب معلومة عنها إلى مرحلة التدريب، بشكل خفيّ لكنه **حقيقي**. والحلّ: وُضعت الخطوتان **داخل** كائن `Pipeline` في `scikit-learn`، ممّا يضمن إعادة حسابهما من **جزء التدريب في كل طيّة فقط**. كما لم يُجرَ أي انتقاء للخصائص على البيانات الكاملة.

المصدر: *`docs/05_validation_and_metrics.md`* القسم *5.5*

سادساً: مستويان للنتائج

- **مستوى التسجيل**: تنبّؤ واحد لكل ملف صوتي.
- **مستوى الشخص**: تُجمَع درجات تسجيلات الشخص بالمتوسّط، ثم تُقارَن بالعتبة 0.5. تنبّؤ واحد لكل إنسان.

ومستوى الشخص هو **الأهم**، لأنك في الواقع العملي تهتمّ بحكم النظام على **شخص**، لا على ملف. والمشروع يذكر الاثنين دائماً.

خلاصة الجزء

- التحقّق المتقاطع بخمس طيّات وثلاث بذور يعطي كل مثال تنبّؤاً من نموذج لم يرَه، والتجميع حسب معرّف المتحدّث يمنع ظهور الشخص على جانبي أي تقسيم.
- الدليل مقيس لا مُدّعى: 0.775 بالتقسيم الصحيح مقابل 0.909 بالخاطئ، وفارق 0.13 كلّه حفظ لا كشف.
- هناك تسريب أخفى عبر سدّ القيم الناقصة والتقييس، وعولج بوضعهما داخل الأنبوب فيُحسبان من بيانات التدريب في كل طيّة فقط.

الجزء 10: المقاييس، كيف نحكم على النموذج

أولاً: مصفوفة الالتباس

كل تنبّؤ يقع في واحدة من أربع خانات، وهذه هي **مصفوفة الالتباس** (confusion matrix):

- **سلبي صحيح** (True Negative، TN): سليم صنّفه النظام سليماً.
- **إيجابي كاذب** (False Positive، FP): سليم صنّفه النظام مريضاً، أي **إنذار كاذب**.
- **سلبي كاذب** (False Negative، FN): مريض صنّفه النظام سليماً، أي **حالة فائتة**.
- **إيجابي صحيح** (True Positive، TP): مريض صنّفه النظام مريضاً.

ومصفوفة نموذجك النهائي على مستوى التسجيل بالبذرة 42: من 42 تسجيلاً سليماً 36 صُنّفت سليمة و6 صُنّفت مريضة خطأً، ومن 31 تسجيلاً مريضاً 24 التُقطت و7 فاتت.

المصدر: *docs/06_experiments_and_results.md* القسم *6.4*

ثانياً: الدقّة، ولماذا لا نكتفي بها

الدقّة (accuracy) = عدد التنبّؤات الصحيحة ÷ العدد الكلّي.

والآن المثال الذي يجب أن تحفظه لأنه يُقنع أي لجنة: عندنا 42 تسجيلاً سليماً و31 مريضاً. تخيّل نموذجاً بليداً يقول «سليم» دائماً:

- سيصيب في 42 حالة من 73.
- دقّته = 42 ÷ 73 = **0.575**، أي 57.5٪.
- وعدد المرضى الذين التقطهم = **صفر**.

فالدقّة تعطي 57.5٪ لنموذج **عديم الفائدة تماماً وخطير طبياً**. لهذا لا نعتمد عليها.

المصدر: *docs/05_validation_and_metrics.md* القسم *5.7*

ثالثاً: الحساسية والنوعية

الحساسية (sensitivity، وتُسمّى أيضاً recall) = TP ÷ (FN + TP). والسؤال: من بين كل المرضى فعلاً، كم نسبة من التقطناهم؟ ونموذجك: **0.771**، أي حوالي 3 من كل 4.

النوعية (specificity) = TN ÷ (FP + TN). والسؤال: من بين كل الأصحّاء فعلاً، كم نسبة من برّأناهم بصحّة؟ ونموذجك: **0.873**، أي أن حوالي 1 من كل 8 أصحّاء يُنذَر عنه خطأً.

وفي سياق الفرز، **فوات المريض هو الخطأ الأغلى ثمناً**، لأن الإنذار الكاذب ينتهي بفحص إضافي، أما الحالة الفائتة فقد تتأخّر سنوات.

رابعاً: الدقّة المتوازنة، مقياسك الأساسي

الدقّة المتوازنة (balanced accuracy) = (الحساسية + النوعية) ÷ 2.

ولنطبّقها على النموذج البليد: حساسيته 0.000 ونوعيته 1.000، فدقّته المتوازنة = **0.500 بالضبط**. وهكذا يفضحه المقياس فوراً بدل أن يمنحه 57.5٪. **لهذا هو المقياس الأساسي في مشروعك كلّه**. ونموذجك النهائي: (0.771 + 0.873) ÷ 2 ▯ **0.822**.

خامساً: الضبط ومقياس F1

الضبط (precision) = TP ÷ (FP + TP). والسؤال: حين يقول النظام «مريض»، كم مرة يكون محقّاً؟

مقياس F1 هو المتوسّط التوافقي بين الضبط والحساسية، أي رقم واحد يوازن بين «التقاط المرضى» و«عدم إطلاق إنذارات كاذبة كثيرة».

سادساً: ROC-AUC، أهم مقياس عبر المجموعات

النموذج لا يُخرج تسمية فقط، بل **درجة** (score) بين 0 و1. والعتبة 0.5 مجرد اختيار. ومنحنى ROC يرسم المقايضة بين الحساسية والإنذارات الكاذبة **عبر كل العتبات الممكنة**، وAUC هي المساحة تحته.

التعريف الأبسط والأجمل: اختر مريضاً عشوائياً وشخصاً سليماً عشوائياً. قيمة AUC هي احتمال أن يُعطي النموذج المريضَ درجة أعلى.

- القيمة 0.5: عديم الفائدة، كرمي عملة.
- القيمة 0.7: تمييز مفيد.
- القيمة 1.0: مثالي.

ونموذجك: **0.864 داخلياً** و**0.701 على البيانات الأجنبية**. وقيمة AUC ثمينة بشكل خاص عبر المجموعات، لأنها تقيس **قدرة الترتيب** بغضّ النظر عن موقع العتبة.

المصدر: *docs/05_validation_and_metrics.md* القسم *5.7*

سابعاً: لماذا نذكر التباين دائماً؟

عندنا 37 شخصاً، فكل طيّة اختبار تحتوي على **7 أو 8 أشخاص فقط**. فتغيّر تصنيف **شخص واحد** يحرّك النتيجة عدّة نقاط. لذلك تُكتَب كل نتيجة على صورة **متوسّط ± انحراف معياري**، مثل **0.822 ± 0.032**. وذِكر رقم واحد بلا مداه عند هذا الحجم يكون **غير أمين**.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.8*

ثامناً: قاعدة النتيجة المشبوهة

كُتبت في الكود **قبل وجود أي نتيجة**: إن بلغت أي نتيجة مستقلّة عن المتحدّث **0.95 أو أكثر**، يطبع سكربت التدريب تحذيراً، ويجب التحقيق في التكوين بحثاً عن تسريب أو تكرار أو عامل مُربك **قبل** أن يُسمح لأحد بتسميته نجاحاً. والنتيجة: **لم يُفعّل هذا التحذير في أي تكوين صحيح**. والرقم الوحيد الذي اقترب (0.909) كان العرض التوضيحي المسرَّب عمداً.

المصدر: `docs/05_validation_and_metrics.md` القسم *5.9*

خلاصة الجزء

- الدقّة العادية مضلّلة: نموذج يقول «سليم» دائماً يحقّق 0.575 بصفر مرضى مكتشفين، بينما الدقّة المتوازنة تعطيه 0.500 فتفضحه.
- نموذجك: حساسية 0.771 أي 3 من كل 4 مرضى، ونوعية 0.873 أي إنذار كاذب لواحد من كل 8 أصحّاء.
- AUC هي احتمال أن يُعطي النموذج مريضاً عشوائياً درجة أعلى من سليم عشوائي، وهي 0.864 داخلياً و0.701 خارجياً.

الجزء 11: نتائج المرحلة الرابعة وفحص العامل المُربك

أولاً: النتائج

43 خاصية، التسجيل كاملاً، خمس طيّات، بتحقّق مستقل عن المتحدّث. الدقّة المتوازنة:

- **الخط الأساسي**: 0.500 لكل شيء.
- **الانحدار اللوجستي**: 0.651 لكل تسجيل، 0.662 لكل شخص، وAUC = 0.717.
- **SVM-RBF**: 0.768 لكل تسجيل، **0.780** لكل شخص، وAUC = 0.741.
- **الغابة العشوائية**: 0.739 لكل تسجيل، 0.725 لكل شخص، وAUC = 0.775.
- **MLP**: 0.671 لكل تسجيل، 0.686 لكل شخص، وAUC = 0.736.

المصدر: *reports/metrics_report.md* و*docs/06_experiments_and_results.md* القسم *6.2*

ثانياً: قاعدة الاختيار المستخدمة

لم يُختَر النموذج بأعلى متوسّط، بل بأعلى **درجة متينة** (robust score)، وهي متوسّط الدقّة المتوازنة عبر الطيّات ناقص انحرافها المعياري. **لماذا؟** لأن هذا يعاقب النموذج الذي كان **محظوظاً في طيّة واحدة** ومتقلّباً في البقيّة. نحن نريد أداءً ثابتاً لا وميضاً. وفاز SVM-RBF بدرجة متينة 0.637.

ثالثاً: فحص العامل المُربك، جنس المتحدّث

هذه أنضج تجربة في المرحلة، واذكرها في مناقشتك.

المشكلة. تذكّر من الجزء 1 و6: الطبقة المطلقة تعكس جنس المتحدّث بقوّة. وفي بياناتك متوسّط F0 للمجموعة السليمة حوالي **180 هرتز**، وللمجموعة المريضة حوالي **154 هرتز**. فهل هذا فرق مرضي؟ أم أن المجموعة السليمة فيها نساء أكثر؟

لماذا لم نستطع الفحص المباشر. لأن البيانات **لا تحمل جنساً موثّقاً** لكل شخص، فلا يمكن فحص **تركيب** المجموعتين.

الحلّ الذكي: قِس الاعتماد بدل التركيب. بما أننا لا نستطيع معرفة التركيب، نطرح سؤالاً آخر قابلاً للقياس: **هل يعتمد النموذج على هذه الخصائص أصلاً؟** فأُعيد التحقّق المتقاطع كاملاً بعد **حذف خصائص F0 المطلقة الأربع**، وهذه تُسمّى تجربة **الحذف** (ablation).

النتائج:

- الانحدار اللوجستي: 0.651 صار **0.699** (تحسّن).
- SVM-RBF: 0.768 صار **0.780** (تحسّن).
- الغابة العشوائية: 0.739 صار 0.707 (تراجع طفيف).
- MLP: 0.671 صار **0.526** (انهيار شبه تامّ).

المصدر: `docs/06_experiments_and_results.md` القسم *6.2*

كيف نقرأ هذا؟ بالنسبة لSVM والانحدار اللوجستي: لم يتأثّرا بل تحسّنا، ومعنى ذلك أن الإشارة التي يستخدمانها تعيش في **خصائص الاضطراب والطيف**، لا في الطبقة المطلقة. فالنموذج المختار **لا يعتمد على العامل المُربك**. أما MLP فانهار إلى 0.526، أي قريباً من الصدفة المحضة، وهذا يكشف أنه **كان يتّكئ** على العامل المُربك المحتمل، أي أن جزءاً كبيراً من نتيجته 0.671 لم يكن كشفاً للمرض.

النتيجة العملية: أُسقطت الشبكة العصبية MLP من كل الأعمال اللاحقة. وتأمّل ما حدث: لم يُسقَط النموذج لأنه الأضعف رقمياً، بل لأن تجربة كشفت أن نتيجته **مبنيّة على أساس غير موثوق**. هذا سلوك علمي ناضج جداً لمشروع بكالوريوس.

وصياغة الإجابة الجاهزة إن سُئلت «كيف تعرفون أن النموذج لا يميّز الرجال من النساء فقط؟»: «اختبرناه. أعدنا التحقّق كاملاً بعد حذف خصائص الطبقة المطلقة، فصمد الأداء (0.768 صار 0.780 للنموذج المختار). ونذكر هذا لأن البيانات لا تحمل جنساً موثّقاً، فلم نستطع فحص التركيب مباشرة، فقِسنا الاعتماد بدلاً منه.»

المصدر: `docs/12_presentation_qa.md`

خلاصة الجزء

- في المرحلة الرابعة فاز SVM-RBF بدقّة متوازنة 0.780 على مستوى الشخص، واختير بقاعدة «الدرجة المتينة» التي تعاقب التقلّب بين الطيّات.
- العامل المُربك المحتمل هو جنس المتحدّث، ولم نستطع فحص التركيب لغياب البيانات الديموغرافية، فقِسنا **الاعتماد** بحذف خصائص الطبقة المطلقة.
- صمد SVM بل تحسّن، بينما انهارت MLP من 0.671 إلى 0.526 فأُسقطت، لأن نتيجتها كانت مبنيّة على أساس غير موثوق.

الجزء 12: تجارب التحسين وقاعدة الحماية من الإفراط في الملاءمة

أولاً: الخلفية

رأى المشرف أن 0.78 منخفضة لعرض عامّ، فطلب تحسين الدقّة. وهنا يظهر الخطر: **البحث عن رقم أعلى هو أسرع طريق للغشّ غير المتعمّد**. فإن جرّبت 50 تكويناً واخترت الأعلى، فأنت غالباً لم تجد نموذجاً أفضل، بل **أفضل حظّ**.

ثانياً: القواعد أُعلنت قبل أي تجربة

وهذا الترتيب الزمني هو ما يجعل الدراسة جديرة بالثقة:

- البروتوكول: 5 طيّات `StratifiedGroupKFold` × 3 بذور، **بنفس التقسيمات لكل المتغيّرات**.
- المقياس الأساسي: الدقّة المتوازنة لكل شخص.
- **هامش التبنّي**: أي متغيّر **أعقد** يجب أن يفوز بأكثر من **0.02** ليُعتمَد.
- أي نتيجة تبلغ 0.95 توقف الدراسة للتحقيق في التسريب.
- **كل** التكوينات تُذكَر، بما فيها الخاسرة.

المصدر: *docs/06_experiments_and_results.md* القسم *6.3*

ثالثاً: المتغيّرات السبعة

- V0: التسجيل كاملاً، 43 خاصية.
- V1: مقاطع بمتوسّط لكل تسجيل، 74 خاصية.
- V2: تصنيف على مستوى المقطع ثم متوسّط الدرجات، 43 خاصية.
- V3: تصنيف على مستوى المقطع ثم متوسّط الدرجات، 74 خاصية.
- V4: الأفضل ممّا سبق مع ضبط معاملات فائقة متداخل.
- V5: الأفضل مع تجميع SVM وRF بتصويت ليّن.

- V6: مقاطع ملخَّصة بالمتوسّط **والانحراف المعياري**، 148 خاصية.

وبضربها في ثلاثة نماذج نحصل على **19 تكويناً**.

رابعاً: أبرز النتائج

الدقّة المتوازنة لكل شخص، متوسّط ثلاث بذور:

- V0 مع SVM: 0.780 ± 0.019
- **V1 مع الغابة العشوائية: 0.822 ± 0.032** وهو الفائز
- V1 مع SVM: 0.816 ± 0.026
- V2 مع SVM: 0.817 ± 0.010
- V4 مع غابة مضبوطة: 0.814 ± 0.024
- **V5 التجميع: 0.840 ± 0.026** وهو الأعلى رقمياً
- V6 مع SVM: 0.824 ± 0.015

المصدر: `docs/06_experiments_and_results.md` القسم *6.3*

خامساً: القرار، ولماذا رفضنا الأعلى

المُعتمَد: V1 مع الغابة العشوائية بنتيجة 0.822. والمرفوضون رغم تساويهم أو تفوّقهم:

- **التجميع V5 بنتيجة 0.840**: هامشه فوق الفائز **0.018+ فقط**، وهو **أقلّ من 0.02**. مرفوض. ويضيف تعقيداً مضاعفاً: ضبط معاملات **ونموذجين**.
- **V6 مع SVM بنتيجة 0.824**: هامشه **0.002+** فقط، ويضاعف عدد الخصائص إلى 148. مرفوض.
- **V4 غابة مضبوطة بنتيجة 0.814**: هامشه **0.008-**، أي أن الضبط لم يبلغ حتى الإعدادات الافتراضية.

سادساً: لماذا هذا الرفض قوّة لا فرصة ضائعة؟

وهنا الحجّة الإحصائية التي يجب أن تتقنها. عندك 37 شخصاً، والتباين بين بذرة وأخرى حوالي **±0.03**. والفارق بين 0.840 و0.822 هو **0.018**، أي **أصغر من الضجيج نفسه**. فهما إحصائياً **متعادلان**. واختيار **الأعقد** من بين خيارات متعادلة ليس تحسيناً، بل هو **ملاءمة لإجراء التحقّق نفسه**، وهذه هي الطريقة الكلاسيكية التي تُفرِط بها المشاريع في الملاءمة بهدوء دون أن تشعر. والأهم: **القاعدة وُضعت مسبقاً تحديداً حتى لا يمكن تبرير هذا القرار بعد رؤية الأرقام.**

المصدر: `docs/06_experiments_and_results.md` القسم *6.3*

والإجابة الجاهزة إن سُئلت «تجميعكم حقّق 0.840، فلماذا استخدمتم 0.822؟»: «ثبّتنا القاعدة قبل إجراء التجارب: المتغيّر الأعقد يجب أن يفوز بأكثر من 0.02. وقد فاز التجميع بـ0.018 فقط، وهي أصغر من تباين البذور البالغ 0.03، أي أنه إحصائياً تعادل.

واختيار الأعقد من بين خيارات متعادلة يعني ملاءمة إجراء التحقّق لا المسألة. ونحن **نذكر 0.840 علناً**، لكننا لم ندَعها تقود القرار.»

خلاصة الجزء

- أُعلنت خمس قواعد **قبل** إجراء التجارب، أهمّها هامش تبنٍّ قدره 0.02 لأي متغيّر أعقد، والإبلاغ عن كل التكوينات بما فيها الخاسرة.
- من 19 تكويناً فاز `V1` مع الغابة العشوائية بنتيجة 0.822، ورُفض التجميع صاحب 0.840 لأن هامشه 0.018 فقط.
- الرفض قوّة لا ضعف: الفارق أصغر من تباين البذور البالغ 0.03، واختيار الأعقد من بين متعادلين يعني ملاءمة إجراء التحقّق لا المسألة.

الجزء 13: النموذج النهائي وأهمية الخصائص

أولاً: مواصفات النموذج

غابة عشوائية بـ300 شجرة، بأوزان فئات متوازنة، وبعمق افتراضي، على **74 خاصية موسّعة** مستخرَجة من كل مقطع عشر ثوانٍ ثم مأخوذة بالمتوسّط لكل تسجيل. ولاحظ: **بإعدادات افتراضية**. لم يُضبَط شيء، لأن الضبط لم يفد.

ثانياً: الأداء عبر البذور الثلاث

- **البذرة 42**: 0.816 لكل تسجيل، 0.780 لكل شخص، حساسية 0.750، نوعية 0.810، AUC = 0.872.
- **البذرة 7**: 0.816 لكل تسجيل، 0.827 لكل شخص، حساسية 0.750، نوعية 0.905، AUC = 0.826.
- **البذرة 2025**: 0.788 لكل تسجيل، 0.859 لكل شخص، حساسية 0.812، نوعية 0.905، AUC = 0.893.

المتوسّط النهائي: لكل تسجيل **0.806 ± 0.013**، ولكل شخص **0.822 ± 0.032**، وحساسية **0.771**، ونوعية **0.873**، وAUC لكل شخص **0.864**.

المصدر: `docs/06_experiments_and_results.md` القسم *6.4*

لاحظ شيئاً مفيداً: تباين مستوى الشخص (±0.032) أكبر من تباين مستوى التسجيل (±0.013). والسبب بسيط: عدد الأشخاص أقلّ من عدد التسجيلات، فكل شخص يزن أكثر.

ثالثاً: التحسّن الكلّي

من **0.780** إلى **0.822**، أي **+0.042**. ومصدر التحسّن مذكور بدقّة: **الخصائص الموسّعة** (CPPS وإحصاءات التوقّفات وdelta-MFCC) محسوبة على المقاطع ومأخوذة بالمتوسّط. **وليس** من ضبط المعاملات الفائقة الذي لم يفد إطلاقاً.

رابعاً: أي الخصائص كانت الأهم؟

من درجات أهمية الغابة العشوائية:

- المرتبة 1: `f0_range_hz` بأهمية 0.069, أي **مدى النبرة**.

- المرتبة 2: `f0_max_hz` بأهمية 0.058, أي أعلى نبرة بُلغت.
- المرتبة 3: `mfcc2_std` بأهمية 0.054, أي تنوّع شكل الطيف.
- المرتبة 4: `mfcc6_mean` بأهمية 0.048, أي متوسّط شكل الطيف.
- المرتبة 5: `mfccD2_std` بأهمية 0.036, أي تنوّع تغيّر الطيف.
- المرتبة 6: `mfccD6_std` بأهمية 0.036.
- المرتبة 7: `f0_mean_hz` بأهمية 0.031, أي متوسّط النبرة.
- المرتبة 8: `mfcc4_std` بأهمية 0.028.
- المرتبة 9: `jitter_local_abs` بأهمية 0.027.
- المرتبة 10: `f0_median_hz` بأهمية 0.022.

المصدر: *`docs/06_experiments_and_results.md`* القسم *6.5*

خامساً: لماذا هذه النتيجة مُرضية علمياً؟

هذه أجمل لحظة في المشروع، ويجب أن تكون شريحة كاملة.

الخاصية الأهم على الإطلاق هي مدى النبرة. وهذا يطابق **أشهر وصف سريري** لكلام مريض باركنسون: الكلام الرتيب، وقِلّة التلوين الصوتي. أي أن النموذج، الذي لم يُخبره أحد بأي شيء عن الطبّ، اكتشف بنفسه العلامة التي يعرفها أطباء الأعصاب منذ عقود.

والملاحظة الثانية: خصائص التنوّع (`std_`) تحتلّ مراتب أعلى من خصائص المتوسّط. وهذا متّسق مع **قِلّة سعة الحركة النطقية:** المريض لا ينطق خطأً، بل ينطق بتنوّع أقلّ.

وهذا بالضبط ما تعنيه عبارة «مشروع قابل للتفسير» (`explainable`): لا نقول «النموذج قال» فحسب، بل نستطيع أن نفتحه ونرى أنه يستخدم ما نتوقّع طبياً أن يستخدمه.

خلاصة الجزء

- النموذج النهائي غابة عشوائية بـ300 شجرة وبإعدادات افتراضية على 74 خاصية، بنتيجة 0.822 ± 0.032 على مستوى الشخص و`AUC` = 0.864.
- التحسّن 0.042+ جاء من الخصائص الموسّعة والتقطيع، لا من ضبط المعاملات.
- الخاصية الأهم هي `f0_range_hz` أي مدى النبرة، وهي المقابل المباشر لرتابة الكلام، وخصائص التنوّع تتفوّق على خصائص المتوسّط.

الجزء 14: النموذج الأولي (التطبيق)

أولاً: واجهتان ودالّة واحدة

- **تطبيق المتصفّح**: ملف `app.py` مبني بـ`Streamlit`، يُشغَّل بالأمر `streamlit run app.py`، ويتيح رفع ملف أو التسجيل بالميكروفون.
- **واجهة سطر الأوامر**: ملف `scripts/predict_file.py`.

وكلاهما ينادي **دالّة واحدة مشتركة** هي `predict_file()` في `src/pvoice/predict.py`. **ولماذا هذا مهم؟** لأنهما بذلك **لا يمكن أن يختلفا أبداً**، ولأن اختبار تلك الدالّة يختبر الواجهتين معاً.

المصدر: *`docs/07_prototype.md` القسم 7.1*

ثانياً: ما يحدث عند تحليل ملف، ثماني خطوات

- **1. التحقّق من توافق الأنبوب**: يُقارَن ملف `models/pipeline_config.json` بإعدادات الكود الحالية. وعند الاختلاف **يُرفض التنبّؤ** ويُطلَب إعادة التدريب.
- **2. التحقّق من الصوت**: موجود، غير فارغ، قابل للفكّ، فيه عيّنات، طوله ثانية على الأقل، وليس صامتاً. وعند الفشل رفض برسالة مفهومة.
- **3. المعالجة المسبقة**: الخطوات الخمس من الجزء 4.
- **4. التقطيع** إلى مقاطع عشر ثوانٍ.
- **5. استخراج 74 خاصية** لكل مقطع ثم أخذ المتوسّط.
- **6. تطبيق النموذج**: ينتج تصنيفاً ودرجة لفئة PD.
- **7. تفسير الدرجة**: تطبيق النطاق غير الحاسم.
- **8. جمع التحذيرات**: تسجيل قصير، اختلاف معدّل عيّنات، قَصّ، نسبة إشارة إلى ضجيج منخفضة.

والزمن حوالي دقيقة لتسجيل من دقيقتين، والجزء الأبطأ هو CPPS.

المصدر: *`docs/07_prototype.md` القسم 7.2*

لماذا الخطوة الأولى بالغة الأهمية؟ هي التي تحوّل **قاعدة الأنبوب الواحد** من وعد إلى ضمانة. فعند تدريب النموذج يُكتَب بجانبه ملف يسجّل **كل** إعداد: معدّل العيّنات، وعتبة القصّ، ومدى الطبقة، وإعدادات MFCC، ومجموعة الخصائص، وقاعدة التجميع،

وطول المقطع، **وقائمة أسماء الخصائص كاملة ومرتّبة**. وقبل **كل** تنبّؤ يُقارَن هذا الملف بالكود الحالي، وإن اختلف أي شيء، **حتى ترتيب خاصية واحدة**، يُرفض التنبّؤ. وغياب الملف أو تعذّر قراءته يُعامَل كعدم توافق أيضاً.

المصدر: `docs/04_pipeline.md` القسم *4.6*

ثالثاً: النطاق غير الحاسم

الدرجات بين **0.35** و**0.65** لا تُعرَض كتصنيف إطلاقاً، بل يقول النظام: «لم يستطع النموذج البحثي إسناد هذا التسجيل بوضوح إلى أي من الفئتين (الدرجة تقع في المدى الأوسط غير الحاسم).»

لماذا وُجد هذا النطاق؟ القصّة حقيقية وتستحقّ أن تُروى في العرض: سجّل المشرف صوته وصوت صديق له، **وكلاهما سليم**. فحصل كلاهما على درجة حوالي 0.65، فصنّفهما النظام PD. والتحقيق كشف أن السبب **ليس مرضاً**، بل **انزياح المجال** (`domain shift`): ميكروفون مختلف، وغرفة مختلفة، ولغة مختلفة عن بيانات التدريب. وقرب العتبة 0.5 يكون الدليل **ضعيفاً فعلاً**، وطباعة تصنيف واثق هناك **تبالغ فيما يعرفه النموذج**.

وصورة توزيع الدرجات `reports/figures/score_distribution.png` تبرّر النطاق تحديداً: التداخل بين توزيعَي الفئتين متركّز في هذه المنطقة بالذات.

المصدر: `docs/07_prototype.md` القسم *7.3*

رابعاً: تحذيرات ظروف التسجيل

هذه **لا تمنع** التنبّؤ، بل تُرفق ملاحظات بلغة بسيطة:

- **تسجيل قصير**: إذا قلّ الصوت المحلَّل عن 10 ثوانٍ. والسبب أن بيانات التدريب كانت 70 ثانية فأكثر.
- **اختلاف معدّل العيّنات**: إذا لم يكن المعدّل الأصلي 44100 هرتز، فقد أُعيد أخذ عيّناته وهذا قد يزيح الخصائص قليلاً.
- **قَصّ** (`clipping`): إذا تجاوزت 0.1٪ من العيّنات المدى الأقصى.
- **تسجيل مشوّش**: إذا قُدّرت نسبة الإشارة إلى الضجيج بأقلّ من 15 ديسيبل.

وتُقدَّر نسبة الإشارة إلى الضجيج بمقارنة المئين التسعين بالمئين العاشر لمستويات RMS على إطارات 50 مللي ثانية، وهو تقدير تقريبي لكنه فعّال.

المصدر: `docs/07_prototype.md` القسم *7.4*

خامساً: وضع الميكروفون

أُضيف **بعد موافقة المشرف**، إذ كانت قواعد المشروع تؤجّله حتى تتحقّق نتائج جيدة على البيانات. ويعرض التطبيق قبل التسجيل:

«ملاحظة عن الميكروفون: دُرِّب النموذج البحثي على تسجيلات بهاتف واحد محدّد في غرفة هادئة واحدة. وميكروفونك وغرفتك ولغتك مختلفة، لذا فإن هذا الوضع يوضّح أساساً كيف يتعامل النظام مع التسجيلات غير المألوفة، وتوقّع نتيجة غير حاسمة أو غير

موثوقة. تحدّث بلا توقّف 30 إلى 60 ثانية على الأقل للحصول على أثبت القياسات.»

ويظهر تحذير ثانٍ مع النتيجة. والصوت المسجَّل يسلك **نفس مسار الملف المؤقّت** تماماً كالمرفوعات، ويُحذَف فور المعالجة.

المصدر: `docs/07_prototype.md` القسم *7.5*

سادساً: الصياغة الإلزامية

تظهر بوضوح **قبل وبعد** كل نتيجة، في الواجهتين:

«هذه أداة أوّلية لدعم الفرز البحثي وليست أداة تشخيص طبي. ولا يجوز أن تحلّ نتيجتها محلّ تقييم مختصّ رعاية صحية مؤهّل.»

وجملة النتيجة مصاغة عمداً كعبارة عن **النموذج**، لا عن **الشخص**: «صنّف النموذج البحثي النمط الصوتي على أنه أقرب إلى فئة PD أو فئة HC.» والدرجة كذلك: «درجة النموذج لفئة PD: 0.16 (حيث 0 تعني أقرب إلى HC و1 أقرب إلى PD). وهذه الدرجة خاصيّة **للنموذج البحثي**، وليست خطراً طبياً على شخص.»

المصدر: `docs/07_prototype.md` القسم *7.6*

سابعاً: الخصوصية ومعالجة الأخطاء

- **لا يُحفَظ أي صوت.** الملف المرفوع أو المسجَّل يُكتَب في ملف مؤقّت يُحذَف داخل كتلة `finally`، أي أنه يُحذَف **حتى لو انهار التحليل**.
- **لا تصل أخطاء تقنية للمستخدم.** المشاكل المتوقّعة تُنتج رسائل ودودة، وغير المتوقّعة تُنتج رسالة عامّة بينما تذهب التفاصيل التقنية إلى سجلّ الطرفية أو عبر خيار `debug`--.
- **رموز خروج ذات معنى** في سطر الأوامر: 0 نجاح، 1 مشكلة ملف، 2 مشكلة نموذج أو إعدادات، 3 خطأ غير متوقّع.

ثامناً: الاختبارات الآلية

ملف `tests/test_prediction.py` وكلّها ناجحة:

- تنبّؤ ناجح على ملف WAV اصطناعي، مع التحقّق من ظهور الصياغة الحذرة.
- حدود النطاق غير الحاسم: 0.349 تعطي HC، والقيم 0.35 و0.50 و0.65 تعطي «غير حاسم»، و0.651 تعطي PD.
- عدم توافق ملف الإعدادات، وغيابه، وقائمة خصائص مبتورة: كلّها تُرفَض.
- صوت مفقود أو فارغ أو تالف أو صامت أو قصير جداً: كلّها تُرفَض برسالة مقروءة.

وملف `tests/test_pipeline.py` يتحقّق من المعالجة والاستخراج على **نغمة اصطناعية 150 هرتز**: تُتتبَّع الطبقة قرب 150، ويكون `jitter` و`shimmer` منخفضَين جداً، و`HNR` مرتفعاً، وترتيب الخصائص ثابتاً.

وهناك تشغيل شامل على تسجيل حقيقي من MDVR-KCL، **لكن التوثيق ينصّ صراحة** على أن هذا **اختبار وظيفي فقط**، لأن الملف كان ضمن بيانات التدريب، فليس دليلاً على الدقّة.

المصدر: `docs/07_prototype.md` القسم *7.8*

خلاصة الجزء

- واجهتان تنادِيان دالّة تنبّؤ واحدة، فلا يمكن أن تختلفا، ويُتحقَّق من عقد الإعدادات قبل كل تنبّؤ فيُرفض التنبّؤ عند أي اختلاف.
- الدرجات بين 0.35 و0.65 تُعرَض «غير حاسمة»، وهذه الإضافة جاءت بعد حادثة حقيقية سجّل فيها المشرف صوته فصنّفه النظام مريضاً.
- لا يُحفَظ أي صوت، والصياغة دائماً عن **النموذج** لا عن **الشخص**، والاختبارات الآلية تغطّي كل حالات الرفض وحدود النطاق.

الجزء 15: التحقّق الخارجي على البيانات الإيطالية

أولاً: لماذا هو أقوى دليل في المشروع؟

كل النتائج السابقة جاءت من مجموعة بيانات واحدة. والتحقّق المتقاطع يُثبت أن النموذج يعمل على **أشخاص جدد**، لكن كل أولئك الأشخاص سُجّلوا **بنفس الهاتف، في نفس الغرفة، بنفس اللغة**. والنموذج قد يجتاز هذا الاختبار ثم **يفشل تماماً** مع ميكروفون آخر أو لغة أخرى، لأنه ربما تعلّم **إعداد التسجيل** لا **المرض**. والطريقة الوحيدة للمعرفة: اختباره على بيانات من مصدر مختلف جذرياً.

والتوثيق يقول بصراحة: **قلّة قليلة جداً من مشاريع البكالوريوس تفعل هذا، وهو أمر محفوف بالمخاطر لأن الرقم يهبط دائماً تقريباً. وقد فعلناه، ونذكر الهبوط.**

المصدر: `docs/08_external_validation.md` القسم *8.1*

ثانياً: المجموعة الإيطالية

اسمها `Italian Parkinson's Voice and Speech`، وحُصل عليها من مرآة عامّة برخصة `CC-BY-4.0`.

- الملفات: **831 ملفاً**، كلّها `WAV` وكلّها أحادية.
- المتحدّثون: **61 متحدّثاً**.
- أصحّاء شباب: 15 متحدّثاً، 45 ملفاً.
- أصحّاء كبار السنّ: 22 متحدّثاً، 349 ملفاً.
- مرضى باركنسون: 24 متحدّثاً، 437 ملفاً.
- معدّلات العيّنات: 16000 هرتز لـ671 ملفاً، و44100 هرتز لـ160 ملفاً.
- المدد: من 3.4 ثانية إلى 250.3 ثانية، بوسيط 11.6 ثانية.
- ملفات تالفة: لا يوجد.

ولم يُدرَّب النموذج ولا يُضبَط على أي جزء منها إطلاقاً.

المصدر: `docs/08_external_validation.md` القسم *8.2*

ثالثاً: الفخّ الذي اكتُشف قبل تشغيل أي شيء

كشف تقرير الفحص نمطاً خطيراً جداً:

- الأصحّاء الشباب: 45 ملفاً بـ16000 هرتز، و**صفر** بـ44100.
- الأصحّاء كبار السنّ: 349 ملفاً بـ16000 هرتز، و**صفر** بـ44100.
- المرضى: 277 ملفاً بـ16000 هرتز، و**160 ملفاً بـ44100**.

أي أن **كل ملف بمعدّل 44.1 كيلوهرتز يخصّ مجموعة باركنسون، وكل ملفات الأصحّاء بـ16 كيلوهرتز**.

لماذا هذا خطير؟ عد إلى نظرية نايكوِست في الجزء 2. الملف بـ16000 هرتز **لا يمكن فيزيائياً** أن يحتوي أي صوت فوق 8000 هرتز، أما الملف بـ44100 فيصل إلى 22050 هرتز. وخصائص MFCC عندنا **ترى الطيف كلّه**. فكان بإمكان النموذج أن يفصل بين المجموعتين **فصلاً شبه تامّ** بمجرّد كشف **عرض النطاق الترددي للتسجيل**، وهي خاصيّة **للميكروفون لا للإنسان**، ولكنّا أعلنّا نتيجة مبهرة وبلا معنى.

العلاج. كل ملف مشمول يُخفَّض أولاً إلى 16000 هرتز، فيُحذَف كل ما فوق 8000 هرتز **من الجميع بالتساوي**، ثم يُعاد رفعه إلى 44100 هرتز وهو معدّل الأنبوب المعياري. وتُسمّى هذه العملية **مواءمة عرض النطاق** (bandwidth harmonisation). وبعدها **لا يمكن تمييز أي مجموعة بعرض نطاقها**.

المصدر: `docs/08_external_validation.md` القسم *8.3*

وهذا مثال ممتاز على انضباط المشروع عموماً: **افحص قبل أن تُقيّم، ولا تفترض شيئاً**. وهذه إجابة جاهزة لسؤال «ما أصعب مشكلة تقنية واجهتكم؟».

رابعاً: تصميم التقييم، اتُّفق عليه قبل التشغيل

- **قاعدة الإدراج**: التسجيلات التي طولها **30 ثانية فأكثر** فقط، أي مهمّات القراءة المتّصلة، لأن الحروف الممدودة والمقاطع القصيرة لا تناسب افتراضات أنبوبنا. والنتيجة **216 تسجيلاً من 61 متحدّثاً**.
- **مواءمة عرض النطاق** كما سبق.
- **النموذج مجمَّد**: لا إعادة تدريب، ولا تعديل عتبة، ولا أي تكيّف.
- **التجميع**: درجة لكل تسجيل، ثم متوسّط لكل متحدّث، بعتبة 0.5، مطابق للتقييم الداخلي تماماً.
- **المقارنة الرئيسية**: الأصحّاء كبار السنّ مقابل المرضى، لأن مجموعة المرضى كبيرة السنّ، ومقارنتهم بشباب أصحّاء **تخلط العمر بالمرض**. والشباب يُذكَرون منفصلين.

ولم تحدث **أي** حالة فشل في الاستخراج عبر الـ216 تسجيلاً.

المصدر: `docs/08_external_validation.md` القسم *8.4*

خامساً: النتائج

لكل متحدّث، أصحّاء كبار السنّ مقابل المرضى:

- الدقّة المتوازنة: **0.822 داخلياً** صارت **0.629 خارجياً**.
- قيمة AUC: **0.864 داخلياً** صارت **0.701 خارجياً**.
- الحساسية: 0.771 صارت **0.667**.
- النوعية: 0.873 صارت **0.591**.

ومصفوفة الالتباس على مستوى المتحدّث: من 22 سليماً كبير السنّ بُرِّئ 13 وأُنذِر عن 9 خطأً، ومن 24 مريضاً التُقط 16 وفات 8.

المصدر: `docs/08_external_validation.md` القسم *8.5*

الخبر الجيد. قيمة AUC تساوي 0.701 تعني أنك لو أخذت مريضاً عشوائياً وسليماً عشوائياً، لَرتّب النموذجُ المريضَ أعلى في **70٪** من الحالات، رغم اختلاف **البلد واللغة والميكروفون والغرفة وعرض النطاق**، وبصفر تكيّف. ومعنى ذلك أن ما تعلّمه النموذج **ليس حفظاً خالصاً لمجموعة البيانات**، بل هناك معلومة صوتية **حقيقية وقابلة للانتقال**.

الخبر الصادق. التصنيف الحاسم قرب العتبة 0.5 **يتدهور بشدّة**. فدقّة متوازنة 0.629 ونوعية 0.591 تعنيان أنه في هذه البيئة الأجنبية، حوالي **4 من كل 10 أصحّاء** سيُنذَر عنهم خطأً. أي أن النموذج بحالته هذه **غير صالح للاستخدام خارج ظروف تدريبه**، وهذا بالضبط سبب رفض النموذج الأوّلي إعطاء أجوبة واثقة هناك.

سادساً: نتيجتان أثمن من الرقم الرئيسي

النتيجة الأولى: النطاق غير الحاسم عمل تماماً كما صُمّم. نسبة الوقوع في النطاق غير الحاسم:

- كل التسجيلات: **71.3٪**.
- الأصحّاء كبار السنّ: **100٪**.
- المرضى: 62.5٪.
- الأصحّاء الشباب: 53.3٪.

كل متحدّث سليم كبير السنّ، بلا استثناء، وقع في النطاق غير الحاسم. أي أن النظام **لم يُصنّف الأصحّاء الأجانب خطأً بثقة**، بل قال «لا أستطيع الجزم»، وهو **الجواب الصحيح** لمُدخَل بعيد إلى هذا الحدّ عن خبرته.

النتيجة الثانية: شذوذ الشباب الأصحّاء يُثبت انزياح المجال. متوسّط درجة PD لكل مجموعة:

- الأصحّاء كبار السنّ: **0.489**.
- المرضى: **0.566**.
- الأصحّاء الشباب: **0.626**.

أي أن **الشباب الأصحّاء حصلوا على درجات باركنسونية أعلى من المرضى أنفسهم**، ولم يُصنَّف منهم سليماً إلا 3 من 15، أي

نوعية 0.200 فقط.

وهذه النتيجة **مستحيلة طبياً**: لا يمكن أن يكون شاب سليم «أكثر باركنسونية» من مريض فعلي. ولذلك فهي تُثبت أن الدرجة، في ظروف خارج المجال، **مدفوعة إلى حدّ كبير بقناة التسجيل وأسلوب الكلام، لا بالمرض**. وبدل إخفاء هذه النتيجة، **يذكرها المشروع صراحة** كدليل مباشر على ضرورة التصميم الحذر للواجهة.

المصدر: `docs/08_external_validation.md` القسم *8.6*

سابعاً: ملاحظات تفسيرية

- مواءمة عرض النطاق تجعل المقارنة عادلة **داخل** المجموعة الإيطالية، لكنها **تحذف أيضاً** محتوى ترددياً عالياً كان النموذج قد رآه أثناء التدريب. وهذا يزيح الدرجات المطلقة، ولهذا فإن **AUC هي أهم رقم عبر المجموعات**، لأنها تقيس الترتيب بمعزل عن موقع العتبة.
- عدد التسجيلات المشمولة يختلف بين المتحدّثين، فنتائج **مستوى المتحدّث** هي الأساسية.
- اللغة مختلفة، وخصائص الاضطراب وإحصاءات التوقّفات **مستقلّة نسبياً عن اللغة**، أما MFCC **فلا**، وهذا يفسّر على الأرجح جزءاً كبيراً من الهبوط.

المصدر: `docs/08_external_validation.md` القسم *8.7*

خلاصة الجزء

- التحقّق الخارجي أقوى دليل لأنه يختبر النموذج مجمَّداً على بلد ولغة وميكروفون مختلفة، والنتيجة 0.629 دقّة متوازنة و0.701 مساحة تحت المنحنى.
- اكتُشف فخّ قبل التقييم: كل ملفات 44.1 كيلوهرتز للمرضى فقط، فعولج بمواءمة عرض النطاق عند 16 كيلوهرتز للجميع.
- النتيجتان الأثمن: 100٪ من الأصحّاء كبار السنّ وقعوا في النطاق غير الحاسم، والشباب الأصحّاء حصلوا على 0.626 مقابل 0.566 للمرضى، وهو دليل مباشر على أن الدرجة خارج المجال تعكس القناة لا المرض.

الجزء 16: القيود والأخلاقيات

أي نظام فرز يمسّ قرارات صحّية **يجب** أن يكون صريحاً فيما لا يستطيع فعله. وكل ما يلي مذكور في تقارير المشروع، ويجب أن يظهر في التقرير النهائي وفي عرض المناقشة.

أولاً: قيود البيانات، ستّة

- **عيّنة صغيرة جداً:** 37 شخصاً، وكل طيّة تختبر 7 أو 8 أشخاص فقط، فتصنيف شخص واحد يحرّك النتيجة عدّة نقاط. ولهذا تُذكَر كل نتيجة كمتوسّط ± انحراف معياري.
- **إعداد تسجيل واحد:** كل بيانات التدريب من جهاز محمول واحد في بيئة واحدة. فالنموذج **لا يستطيع التمييز** بين «هذا الصوت غير سليم» و«هذا التسجيل غير مألوف».
- **لغة واحدة:** الإنجليزية فقط، والاختبار الخارجي أظهر حجم الهبوط الناتج.
- **لا بيانات ديموغرافية:** لا عمر ولا جنس موثّق لكل شخص. وبما أن الطبقة تعكس الجنس، وباركنسون مرتبط بالعمر، فإن الإرباك الديموغرافي **لا يمكن استبعاده تماماً، بل تحديد حدوده فقط،** إذ غيّرت تجربة الحذف النتيجة بأقلّ من 0.02 للنموذج المختار.
- **عدم توازن الفئات:** 42 مقابل 31، عولج بأوزان متوازنة وبالدقّة المتوازنة، لكنه يبقى قيداً.
- **لا معلومات عن شدّة المرض:** مريض في بداية المرض ومريض متقدّم يُعامَلان كفئة واحدة، رغم أن صوتيهما قد يختلفان كثيراً. **فحساسية النظام للحالات المبكّرة تحديداً غير مقيسة،** وهذا مهم لأن الكشف المبكّر هو دافع المشروع المُعلن.

المصدر: `docs/09_limitations_and_ethics.md` القسم *9.2*

ثانياً: قيود المنهج، أربعة

- **خصائص الاضطراب مستخدَمة خارج سياقها الكلاسيكي:** `jitter` و`shimmer` و`HNR` صُمّمت للحروف الممدودة، وعلى الكلام المتّصل تصير بدائل أكثر ضجيجاً. و`CPPS` وإحصاءات التوقّفات تعوّض جزئياً.
- **التفاصيل الزمنية مُهمَلة:** متوسّط الخصائص عبر المقاطع يحذف عمداً معلومة **متى** حدث كل شيء، مقابل قابلية التفسير والاستقرار.
- **الدرجات ليست احتمالات معايَرة:** درجة 0.7 **لا تعني** «احتمال 70٪ للإصابة». هي مخرج نموذج، ومعناها يعتمد كلياً على توزيع التدريب.
- **مقطعي فقط** (`cross-sectional`): النظام يقارن الشخص بمجموعة، ولا يقارنه أبداً بتسجيلاته السابقة. وتتبّع التغيّر عبر

الزمن سيكون على الأرجح أكثر حساسية بكثير، لكنه يحتاج بيانات طولية.

ثالثاً: قيود الأدلّة، ثلاثة

- **مجموعة خارجية واحدة:** قيمة 0.701 نقطة بيانات واحدة عن التعميم، لا صورة كاملة.
- **التحقّق المتقاطع يقيس المجموعة لا الفرد:** دقّة 0.822 تصف الأداء على **هذه الفئة السكّانية**، ولا تقول شيئاً موثوقاً عن مستخدم مستقبلي بعينه.
- **هامش 0.02 حكم اجتهادي:** ثبّت مسبقاً وطُبّق باتّساق، لكن عتبة معقولة أخرى كانت ستختار التجميع بدلاً منه.

وهذه النقطة الأخيرة تُظهر نضجاً كبيراً: المشروع ينتقد **قاعدته الخاصة** التي انتصر بها.

رابعاً: قواعد الصياغة

المستخدَم دائماً:

- «أداة أوّلية لدعم الفرز البحثي».
- «مؤشّر غير تشخيصي».
- «تغيّرات صوتية مرتبطة بباركنسون».
- «صنّف النموذج البحثي النمط الصوتي على أنه أقرب إلى فئة PD أو HC».
- «يتطلّب تقييماً من مختصّ رعاية صحية مؤهّل».

الممنوع تماماً:

- «يُشخّص مرض باركنسون».
- «يكشف باركنسون بيقين».
- «نظام قرار طبي».
- «أداة مُثبتة سريرياً».
- أي صياغة تقدّم الدرجة على أنها خطر طبي على شخص أو احتمال إصابته.

خامساً: لماذا تهمّ الصياغة عملياً؟

هذه ليست شكليات قانونية، بل **منع ضررين ملموسين:**

- **الطمأنة الكاذبة:** شخص لديه أعراض حقيقية يرى HC فيؤجّل زيارة الطبيب. وحساسية النظام 0.771 داخلياً، أي أن **حوالي 1 من كل 4 مرضى يفوت.**
- **الإنذار الكاذب:** شخص سليم يرى PD فيصاب بقلق حقيقي. وخارجياً **حوالي 4 من كل 10 أصحّاء أُنذِر عنهم خطأً.**

والنطاق غير الحاسم موجود تحديداً **لتقليل الضررين معاً** في الحالات غير المؤكّدة.

سادساً: أقوى صياغة للدفاع

هذه الفقرة أنصحك بحفظها شبه حرفياً، فهي خلاصة المشروع كلّه:

«لا يدّعي هذا المشروع كشف مرض باركنسون. إنه يُظهر أن خصائص صوتية قابلة للقياس تختلف بين مجموعة من مرضى باركنسون ومجموعة من الأصحّاء، وأن نموذج تعلّم آلة كلاسيكياً يستطيع استثمار هذه الفروق بدقّة متوازنة تقارب 82٪ على **أشخاص لم يرهم** من نفس المجموعة، وأن نحو 70٪ من قدرة الترتيب **تصمد** عند الانتقال إلى مجموعة بيانات ولغة مختلفتين تماماً، وأن النموذج الأولي المسؤول ينبغي أن **يُبلّغ عن عدم اليقين** بدل إعطاء تصنيف واثق حين يقع المُدخَل خارج خبرته.»

المصدر: `docs/09_limitations_and_ethics.md` القسم *9.6*

خلاصة الجزء

- القيود ثلاث عائلات: ستّة للبيانات وأربعة للمنهج وثلاثة للأدلّة، وذكرها يزيد مصداقيتك ولا ينقصها.
- الصياغة ليست شكليات: هي تمنع ضررين مقيسين، فوات مريض من كل أربعة داخلياً، وإنذار كاذب لأربعة من كل عشرة أصحّاء خارجياً.
- أقوى صياغة للدفاع تعترف بحدود المشروع بوضوح، وهذا ما يجعلها قوية.

الجزء 17: بنية المشروع وكيفية إعادة تشغيل كل شيء

أولاً: خريطة المجلدات

- app.py: تطبيق المتصفّح.
- requirements.txt: قائمة المكتبات.
- README.md: الإعداد والنتائج وأوامر التشغيل.
- CLAUDE.md: قواعد المشروع الثابتة.
- /src/pvoice: كود الأنبوب القابل لإعادة الاستخدام، وفيه config.py لكل الإعدادات، وdataset.py لاكتشاف الملفات والتسميات، وpreprocess.py لتهيئة الصوت والتقطيع، وfeatures.py لحسابات الخصائص، وmodeling.py لتعريفات النماذج، وevaluate.py للتحقّق المجمَّع والمقاييس، وpredict.py للتنبّؤ وفحوصات الأمان.
- /scripts: سكربت لكل مرحلة.
- /tests: الاختبارات الآلية.
- /docs: التوثيق.
- /reports: التقارير والصور المولَّدة.
- /models: النموذج المدرَّب وإعدادات الأنبوب.
- /data: البيانات الخام والمعالَجة، وهي خارج git.

ثانياً: ماذا يفعل كل سكربت وكم يستغرق

- inspect_dataset.py: يفحص MDVR-KCL ويكتب تقرير الفحص. ثوانٍ.
- build_features.py: 43 خاصية لكل تسجيل كامل. حوالي 12 دقيقة.
- validate_features.py: فحوصات الجودة. ثوانٍ.
- train_models.py: تقييم المرحلة الرابعة. حوالي دقيقتين.
- build_segment_features.py: 74 خاصية لكل مقطع. حوالي **32 دقيقة**.
- run_experiments.py: دراسة الـ19 تكويناً. حوالي 30 دقيقة.
- train_final_model.py: تدريب النموذج النهائي وحفظه. حوالي دقيقة.
- make_figures.py: كل الصور. حوالي دقيقتين.

- `inspect_italian_dataset.py`: فحص المجموعة الخارجية. حوالي دقيقة.
- `evaluate_external.py`: التحقّق الخارجي. حوالي **75 دقيقة**.
- `predict_file.py`: تنبّؤ لملف واحد. حوالي دقيقة.

ملاحظة عملية مهمة: السكربتات الطويلة **تحفظ نتائجها لكل ملف** في ذاكرة مؤقّتة. فإن انقطع التشغيل، أعِد تشغيله فقط وسيُكمل من حيث توقّف. واحذف مجلد الذاكرة لفرض إعادة استخراج كاملة بعد تغيير الإعدادات.

المصدر: *docs/10_repository_and_reproduction.md* القسم *10.2*

ثالثاً: إعادة الإنتاج من الصفر

كل الأوامر تُنفَّذ **داخل مجلد المشروع** ومع تفعيل البيئة الافتراضية أولاً بالأمر `venv\Scripts\activate`. والناتج المتوقّع أن يبدأ سطر الأوامر بـ`(venv)`.

- **الخطوة 1**: تثبيت المكتبات بـ`pip install -r requirements.txt`.
- **الخطوة 2**: تحميل MDVR-KCL ووضعها في `data/raw/mdvr_kcl/`.
- **الخطوة 3**: التأكّد أن الأنبوب يعمل بـ`python tests/test_pipeline.py`، والناتج المتوقّع `All tests passed.`
- **الخطوة 4**: فحص البيانات بـ`python scripts/inspect_dataset.py`.
- **الخطوة 5**: استخراج خصائص المقاطع بـ`python scripts/build_segment_features.py`. وهي الخطوة الطويلة.
- **الخطوة 6**: تدريب النموذج النهائي بـ`python scripts/train_final_model.py`.
- **الخطوة 7**: توليد الصور بـ`python scripts/make_figures.py`.
- **الخطوة 8**: تشغيل النموذج الأولي بـ`streamlit run app.py`.

وإن فشل أي أمر، انسخ **الأمر الذي شغّلته** و**رسالة الخطأ كاملة** والصقهما في جلسة مساعدة جديدة.

رابعاً: الصور المولَّدة

- `validation_comparison.png`: التقسيم الصحيح 0.775 مقابل المسرَّب 0.909.
- `feature_importance.png`: أهم 15 خاصية.
- `roc_curve.png`: منحنى ROC على مستوى التسجيل والشخص.
- `score_distribution.png`: توزيع الدرجات لكل فئة مع النطاق غير الحاسم.
- `external_scores.png`: درجات المتحدّثين الـ61 الإيطاليين.
- `confusion_*.png`: مصفوفات الالتباس لنماذج المرحلة الرابعة.

خامساً: إدارة الإصدارات

المرفوع إلى المستودع: كل الكود، والاختبارات، والتوثيق، والتقارير، والصور، وإعدادات الأنبوب، وقائمة المكتبات.

غير المرفوع عمداً: مجموعات الصوت لأنها كبيرة ولها رخص خاصة، والبيئة الافتراضية، والذواكر المؤقّتة، وجداول CSV المولَّدة، وملفات النموذج الثنائية، **وكلّها قابلة لإعادة التوليد** من السكربتات.

ومعنى ذلك أن أي شخص يستطيع استنساخ المستودع وإعادة إنتاج **كل رقم** من الصفر، وهذه في ذاتها مساهمة تستحقّ الذكر في التقرير.

المصدر: *docs/10_repository_and_reproduction.md* القسم *10.6*

سادساً: البيئة والمكتبات

Python 3.14.5 في بيئة افتراضية محلّية، **بلا أي تثبيت عام**. والمكتبات الأساسية: praat-parselmouth لـF0 وjitter وshimmer وHNR وCPPS، وlibrosa لـMFCC وإعادة أخذ العيّنات وكشف الصمت، وscikit-learn للنماذج والتحقّق، وpandas وnumpy للجداول والحسابات، وmatplotlib للصور، وstreamlit للواجهة، وsoundfile لقراءة الصوت.

خلاصة الجزء

- كل الإعدادات في src/pvoice/config.py، وكل مرحلة لها سكربت واحد يُعيد توليد تقريرها ومخرجاتها من الصفر.
- أطول خطوتين هما استخراج خصائص المقاطع (32 دقيقة) والتحقّق الخارجي (75 دقيقة)، وكلاهما يحفظ نتائجه فيُكمل بعد أي انقطاع.
- المستودع يحتوي على كل شيء عدا الصوت، فأي شخص يستطيع إعادة إنتاج كل رقم، وهذه مساهمة في ذاتها.

الجزء 18: تحضير العرض والأسئلة المتوقّعة

أولاً: الرسائل الأربع التي يجب أن تصل

- **بنينا القياس، لا النموذج فقط.** استخراج 74 قياساً صوتياً من الصوت الخام، مبنيّة على فسيولوجيا الكلام، هو المساهمة الهندسية الأساسية، وليس جدول خصائص محمَّلاً جاهزاً.
- **تحقّقنا بأمانة.** لا يظهر أي شخص في التدريب والاختبار معاً، ومفروض بتأكيد يُوقف البرنامج. ونفس الأنبوب يعطي 0.909 بتقسيم مسرَّب عمداً مقابل 0.775 بالتقسيم الصحيح، وهو برهان **مقيس** على أن كثيراً من نتائج 95٪ فأكثر المنشورة غير جديرة بالثقة.
- **قِسنا التعميم بدل افتراضه.** اختُبر النموذج بلا تكيّف على مجموعة أجنبية بلغة أخرى: AUC = 0.701. ومعظم مشاريع البكالوريوس لا تحاول هذا أصلاً.
- **بنينا نظاماً يعترف بعدم اليقين.** الدرجات قرب الحدّ تُبلَّغ كـ«غير حاسمة». وعلى المجموعة الأجنبية وقع **100٪** من الأصحّاء كبار السنّ في هذا النطاق.

المصدر: `docs/12_presentation_qa.md` القسم *12.1*

ثانياً: ترتيب الشرائح المقترح

- الدافع: باركنسون، وأعراض الكلام، ولماذا يهمّ الفرز المبكّر.
- كيف يُنتَج الصوت وماذا يغيّره المرض.
- البيانات: 37 شخصاً، 73 تسجيلاً، ونتائج الفحص.
- رسم الأنبوب: صوت، ثم معالجة مسبقة، ثم مقاطع، ثم 74 خاصية، ثم نموذج.
- عائلات الخصائص مع الفسيولوجيا وراء كل عائلة.
- **شريحة التسريب**، وهي أهم شريحة عندك.
- النتائج: جدول المقاييس ومنحنى ROC.
- أهمية الخصائص: ما الذي يستخدمه النموذج فعلاً.
- دراسة التجارب: 19 تكويناً، ولماذا رفضنا الأعلى نتيجة.
- التحقّق الخارجي: التصميم، وفخّ عرض النطاق، والنتائج.
- النموذج الأولي: عرض حيّ.

- القيود والأخلاقيات.
- الخلاصة والعمل المستقبلي.

ثالثاً: نصائح العرض الحيّ

- **الأكثر أماناً:** تسجيل من مجموعة البيانات عبر التطبيق أو سطر الأوامر، فهو داخل المجال، والنتيجة ذات معنى وسريعة الشرح.
- **الأكثر إبهاراً:** سجّل حيّاً بالميكروفون ودع النظام يُعيد «**غير حاسم**» مع تحذيرات الظروف، ثم اشرح أن **هذا هو النظام وهو يعمل بشكل صحيح، لا وهو يفشل.** هذه الحيلة تحوّل أضعف نقطة تقنية إلى **دليل على نضج علمي**.
- استعدّ لزمن التحليل البالغ دقيقة: ابدأ التحليل ثم اشرح شريحة الأنبوب أثناء عمله.
- **جهّز لقطات شاشة** احتياطية تحسّباً لتعثّر العرض الحيّ.

رابعاً: الأسئلة المتوقّعة وإجاباتها

س: دقّتكم 82٪ بينما الأبحاث المنشورة تقول 95 إلى 99٪ على نفس البيانات. لماذا نتيجتكم أقلّ؟

لأن تلك الأبحاث تقسّم عادةً على مستوى التسجيل أو النافذة، فيظهر صوت الشخص نفسه في التدريب والاختبار. وقد **قِسنا هذا الأثر مباشرة:** بتقسيم مقاطع مسرَّب يعطي أنبوبنا 0.909، وبتجميع صحيح حسب المتحدّث يعطي 0.775. الفارق **حفظ لا كشف.** نتيجتنا أقلّ **لأنها صادقة.**

س: لماذا لم تستخدموا التعلّم العميق؟

ثلاثة أسباب. أولاً 37 شخصاً قليل جداً لتدريب شبكة عميقة دون إفراط شديد في الملاءمة. ثانياً المشروع يشترط قابلية التفسير، فنستطيع إظهار أن مدى النبرة هو الخاصية الأهم وربطه بعلامة سريرية معروفة. ثالثاً تجربتنا نفسها أظهرت أن شبكة صغيرة MLP انهارت من 0.671 إلى 0.526 عند حذف خصائص الطبقة، كاشفةً اعتمادها على عامل مُربك.

س: لماذا الدقّة المتوازنة بدل الدقّة؟

بياناتنا 42 سليماً و31 مريضاً. ونموذج يقول «سليم» دائماً يحقّق 57.5٪ دقّة بينما يلتقط **صفراً** من المرضى. والدقّة المتوازنة تعطيه 50٪، فتفضحه بشكل صحيح.

س: تجميعكم حقّق 0.840، فلماذا استخدمتم نموذج 0.822؟

ثبّتنا القاعدة قبل التجارب: المتغيّر الأعقد يجب أن يفوز بأكثر من 0.02. وقد فاز التجميع ب0.018، وهي أصغر من تباين البذور البالغ نحو 0.03، أي تعادل إحصائي. واختيار الأعقد من بين خيارات متعادلة يعني ملاءمة **إجراء التحقّق** لا المسألة. **نذكر 0.840 علناً،** لكننا لم ندَعها تقود القرار.

س: هل يستطيع هذا تشخيص باركنسون؟

لا، وهو غير مصمَّم لذلك. إنه يُبلّغ عمّا إذا كان النمط الصوتي لتسجيل ما يشبه مجموعة بيانات أم أخرى. وحالات كثيرة أخرى، كزكام أو التهاب حنجرة أو تقدّم في العمر أو تدخين، تُغيّر جودة الصوت أيضاً. وأي قلق يستوجب تقييم طبيب مؤهّل. وهذه العبارة تظهر **قبل وبعد كل نتيجة** في البرنامج.

س: ماذا يحدث لو سجّلتُ صوتي الآن؟

على الأرجح نتيجة «غير حاسمة»، وربما مع تحذيرات ظروف. **وهذا سلوك صحيح:** ميكروفونك وغرفتك وربما لغتك تختلف عن بيانات التدريب. وقد أثبتنا هذا بالضبط على المجموعة الإيطالية، حيث وقع 100٪ من الأصحّاء كبار السنّ في النطاق غير الحاسم.

س: أليس 37 شخصاً قليلاً جداً؟

بلى، ونقول ذلك في كل تقرير. إنه القيد الرئيسي. ونخفّفه بذكر التباين بدل الأرقام المفردة، وبتكرار التحقّق بثلاث بذور عشوائية، وباستخدام تحقّق مجمَّع يمنع التضخيم، وباختبار خارجي على مجموعة ثانية من 61 متحدّثاً.

س: كيف تعرفون أن النموذج لا يكشف الرجال من النساء فقط؟

اختبرناه. خصائص الطبقة المطلقة هي الأرجح لترميز جنس المتحدّث، فأعدنا التحقّق كاملاً بعد حذفها. وصمد الأداء (0.768 صار 0.780 للنموذج المختار)، ممّا يُظهر أن الإشارة تعيش في خصائص الاضطراب والطيف. ونذكر ذلك لأن البيانات لا تحمل جنساً موثّقاً، فلم نستطع فحص التركيب مباشرة.

س: ما أصعب مشكلة تقنية واجهتكم؟

كشف العوامل المُربكة التي كانت **ستجعل النتائج تبدو أفضل**. وأوضح مثال: في المجموعة الإيطالية، كل ملف بـ44.1 كيلوهرتز يخصّ مجموعة باركنسون وكل ملفات الأصحّاء بـ16 كيلوهرتز، فكان بإمكان نموذج أن «يكشف باركنسون» بكشف عرض نطاق الميكروفون. اكتشفناه أثناء الفحص، وخفّضنا كل شيء إلى 16 كيلوهرتز قبل التقييم.

س: لماذا مقاطع من عشر ثوانٍ؟

طويلة بما يكفي لقياسات صوتية مستقرّة، وقصيرة بما يكفي لإنتاج نحو 14 مقطعاً لكل تسجيل. ومتوسّط قياسات كثيرة أثبت من قياس واحد على تسجيل كامل، وقد رفع الدقّة المتوازنة من 0.780 إلى 0.822. والأهم أن المقاطع ترث معرّف المتحدّث فيبقى التحقّق المجمَّع صحيحاً، فالتقسيم على مستوى المقاطع بلا هذا التجميع هو بالضبط التسريب الذي أثبتناه.

س: ماذا كنتم ستفعلون لو توفّر وقت أكثر؟

بترتيب القيمة المتوقّعة: جمع أو الحصول على أشخاص أكثر، خصوصاً بتنوّع في ظروف التسجيل؛ ثم إضافة بيانات الجنس والعمر لكل شخص حتى يُضبَط الإرباك بدل تحديد حدوده فقط؛ ثم اختبار طولي بمقارنة الشخص بتسجيلاته السابقة، وهو على الأرجح أكثر حساسية بكثير؛ ثم تطبيق تقنيات تكيّف المجال ليعبر النموذج بين الميكروفونات واللغات بشكل أفضل.

س: هل العمل قابل لإعادة الإنتاج؟

نعم. كل رقم في كل تقرير يُعاد توليده بسكربت في المستودع العام. والصوت نفسه غير موزَّع لأسباب ترخيص، لكن تعليمات التحميل مرفقة، وإعدادات الأنبوب محفوظة مع النموذج فلا يمكن للتنبّؤات أن تستخدم إعدادات مختلفة عن التدريب بصمت.

المصدر: `docs/12_presentation_qa.md` القسم *12.4*

خلاصة الجزء

- أربع رسائل يجب أن تصل: بنينا القياس، وتحقّقنا بأمانة، وقِسنا التعميم، وبنينا نظاماً يعترف بعدم اليقين.
- العرض الحيّ الأكثر إبهاراً هو التسجيل بالميكروفون والحصول على «غير حاسم»، ثم شرح أن هذا هو النظام وهو يعمل بشكل صحيح.
- لكل سؤال متوقّع إجابة جاهزة، وأهمّها سؤال «لماذا نتيجتكم أقلّ من المنشور»، وجوابه أن الفرق حفظ لا كشف، ونتيجتنا أقلّ لأنها صادقة.

خاتمة الدليل

المشروع يبني القياس من الصوت الخام: 74 خاصية لها كلّها مبرّر فسيولوجي، عبر أنبوب واحد ثابت يُستخدَم في التدريب والتقييم والتنبّؤ.

والنزاهة العلمية هي المساهمة الحقيقية: تحقّق مستقل عن المتحدّث مفروض بالكود، ودليل مقيس على التسريب (0.909 مقابل 0.775)، وقواعد اختيار مُعلنة مسبقاً أدّت إلى **رفض** أعلى نتيجة، وتحقّق خارجي على بيانات أجنبية أظهر هبوطاً صادقاً إلى AUC = 0.701.

والنتيجة النهائية 0.822 ± 0.032 داخلياً مع AUC = 0.864، وأهم خاصية هي مدى النبرة f0_range_hz، وهو ما يطابق أشهر علامة سريرية لكلام باركنسون، ونظام أوّلي **يعترف بعدم اليقين** بدل أن يعطي تصنيفاً واثقاً خارج خبرته.