الجامعة السورية الخاصة

كلية الهندسة

قسم الذكاء الصنعي وعلوم البيانات

أعدت هذه الأطروحة

لإنجاز مقرر المشروع الفصلي في اختصاص الذكاء الصنعي وعلوم البيانات

Voice Command Recognition

إعداد الطلاب:

رنيم ربيع أيهم السالم

أشراف:

د.ميساء أبو قاسم م.وسام السحلي

شهادة المشرفين:

الاسم:

التاريخ:

التوقيع:

في تاريخ: 2025\12\18

الجامعة السورية الخاصة

Syrian private university

نشئت الجامعة السورية الخاصة في ٢٠٠٥ من قبل عدد من الأكاديميين العرب والأجانب ورجال الأعمال. بالرغم من عمر الجامعة القصير إلا أنها تتعرب الآن كواحدة من أفضل الجامعات في سوريا.

يمنح المعهد العالي درجة الإجازة في الهندسة في الاتصالات والمعلوماتية والنظم الالكترونية والميكاترونيكس وعلوم وهندسة المواد وهندسة الطيران يقبل المعهد العالي لدراسة هذه الاختصاصات شريحة منتقاة من المتفوقين في الشهادة الثانوية من الفرع العلمي يتيح المعهد العالي أيضاً برامج ماجستير أكاديمي في نظم الاتصالات وفي التحكم والروبوتيك وفي نظم البيانات الكبيرة ونظم المعلومات ودعم القرار وفي علوم وهندسة المواد وعلوم وهندسة البصريات .وأخيراً، يمنح المعهد العالي درجة الدكتوراه في الاتصالات والمعلوماتية ونظم التحكم والفيزياء التطبيقية . تحدث في المعهد العالي اختصاصات جديدة بحسب متطلبات سوق العمل وتوجهات البحث والتطوير المحلية والعالمية

ينشر المعهد العالي كتب ا علمية عالية المستوى من نتاج أطره، منها ما هو تدريسي يوافق المناهج في المعهد العالي ويفيد شريحة واسعة من الطلاب الجامعيين عموماً، ومنها ما هو علمي ثقافي يتيح المعهد العالي بعضاً من منشوراته على موقعه على الشابكة، كما يتيح إمكانية الاطلاع على رسائل الماجستير والدكتوراه المنفذة في المعهد العالي وعلى بعض منشورات طلابه وأطره من المقالات العلمية.

تصريح

أصرح بأن:

- الأعمال والنتائج المعروضة في هذه الأطروحة هي نتيجة جهودي الشخصية وبتوجيه من المشرفين، وأنّ ما عدا ذلك من معلومات ونتائج قد نسبت إلى مصادرها ومؤلفيها، وأشير إلى ذلك في متن النص وفي قائمة المراجع.
- البيانات والمعلومات المستخدمة في هذه الأطروحة جرى تحصيلها بطرائق سليمة ومشروعة ونسبت إلى بصادرها في المواضع الملائمة.
- كل مكوّن من هذه الأطروحة (مقطع نصّي، صورة، مخطط).. مقتبس من عمل آخر جرى تمييزه بوضوح وتحسب إلى مصدره.
- الأعمال والنتائج المعروضة في هذه الأطروحة لم تستخدم سابقاً وليست قيد الاستخدام للحصول على أي شهادة اكاديمية أخرى

.

التوقيع

أيهم السالم رنيم ربيع

2025/12/18

الفهرس:

الخلاصة:

- تهدف هذه الدراسة إلى تطوير نظام لفهم وتصنيف الأوامر الصوتية من خلال دمج المعلومات الصوتية والنصية لمجموعة بياناتFluent Speech Commands ، بالاستناد أيضًا إلى دراسة تحليلية لست أوراق بحثية حديثة تناولت تصنيف النيات الصوتية وتقنيات التعلم الآلي المتعلقة بها. يعتمد النظام على تحويل كل أمر صوتي إلى نية موحدة على شكل زوج(action → object) ، مثل التحكم في مستوى الصوت، الموسيقى، اللغة أو الإضاءة.

- أظهرت النتائج أن دمج الميزات الصوتية المستخرجة عبر نموذج Whisper-small مع التمثيل النصي الناتج عنTF-IDF + SVD ، ثم تدريب نموذج Logistic Regressionمتعدد الفئات على هذه الميزات المدمجة، أدى إلى أداء استثنائي في تصنيف النيات، حيث بلغ Test Accuracy ≈ 99.86% و Macro F1 ≈ 99.82%، مع وجود أخطاء قليلة جدًا (2 فقط من أصل 1425 عينة).

- كما تبيّن أن طبيعة المهمة البسيطة نسبيًا (عدد محدود من الفئات، نصوص قصيرة وواضحة، تسجيلات صوتية نظيفة) ومعالجة البيانات الدقيقة، تلعب دورًا محوريًا في تعزيز دقة النموذج وقدرته على التعميم. وتؤكد الدراسة المرجعية أن استخدام التمثيلات متعددة الوسائط وتقنيات تنظيف البيانات يرفع بشكل ملحوظ من دقة أنظمة تصنيف الأوامر الصوتية.

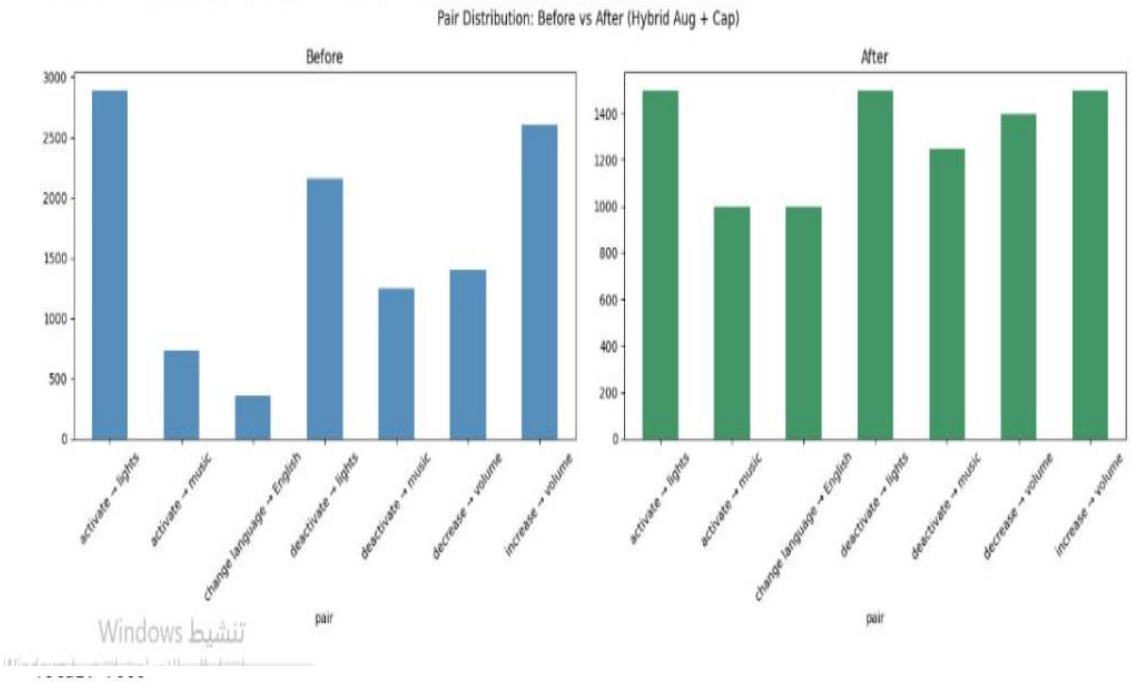
- توفر هذه النتائج أساسًا قويًا لتطوير أنظمة فهم صوتية موثوقة، مع إمكانية توسيعها مستقبلًا لدعم أوامر صوتية أكثر تعقيدًا أو دمجها ضمن أنظمة متعددة الوسائط في التطبيقات العملية.

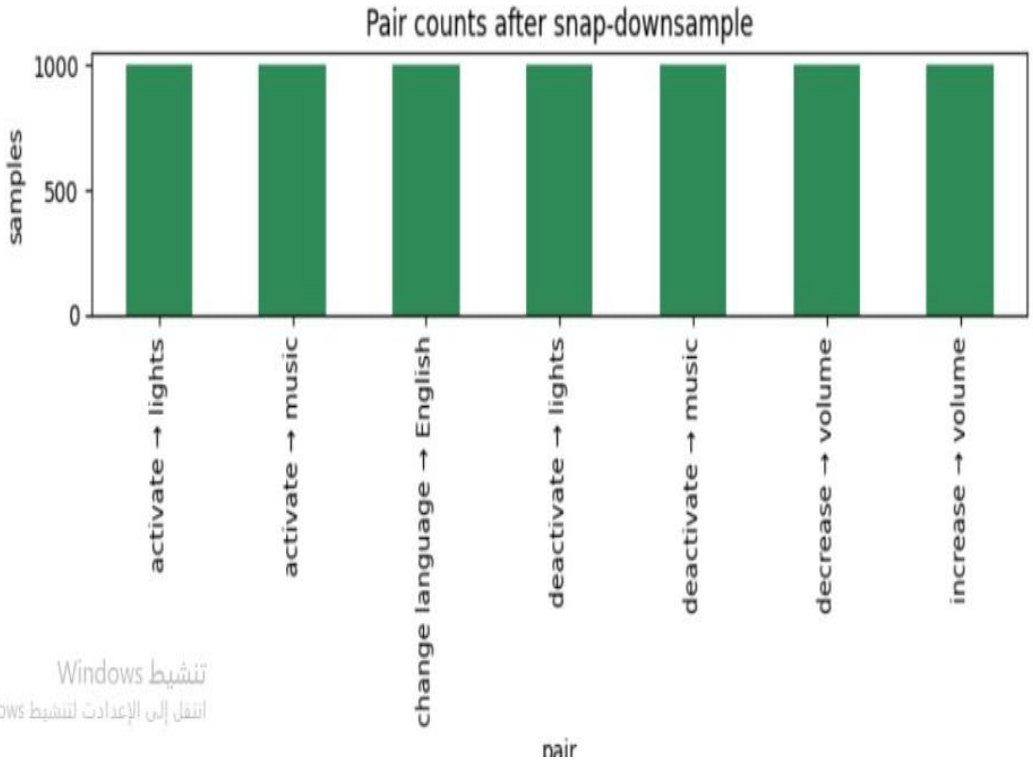
قائمة بأهم المصطلحات

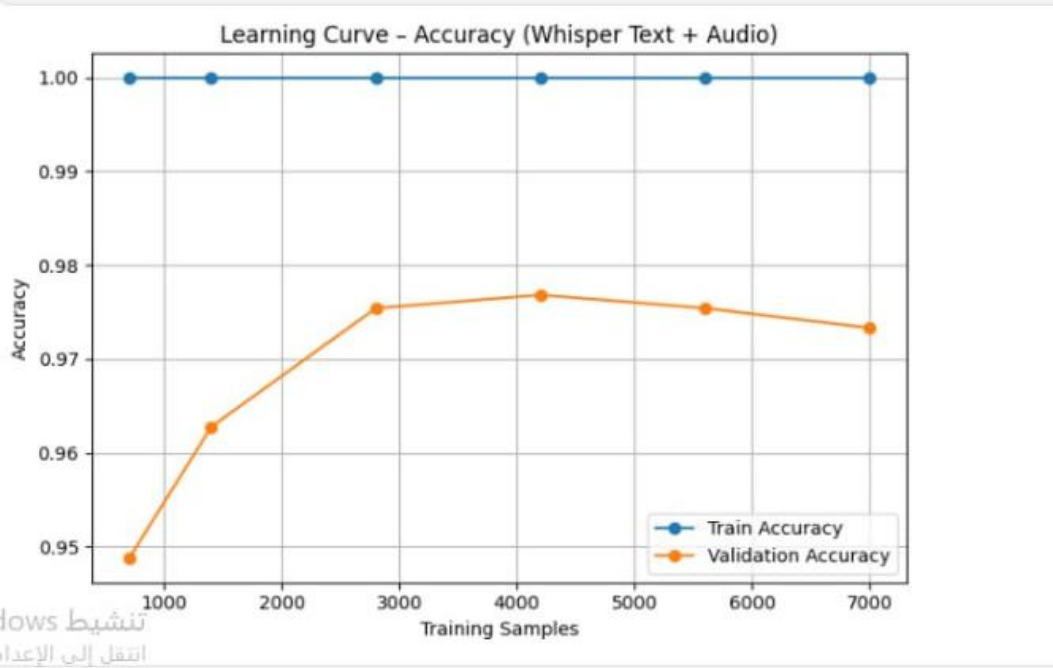
Voice Command Recognition

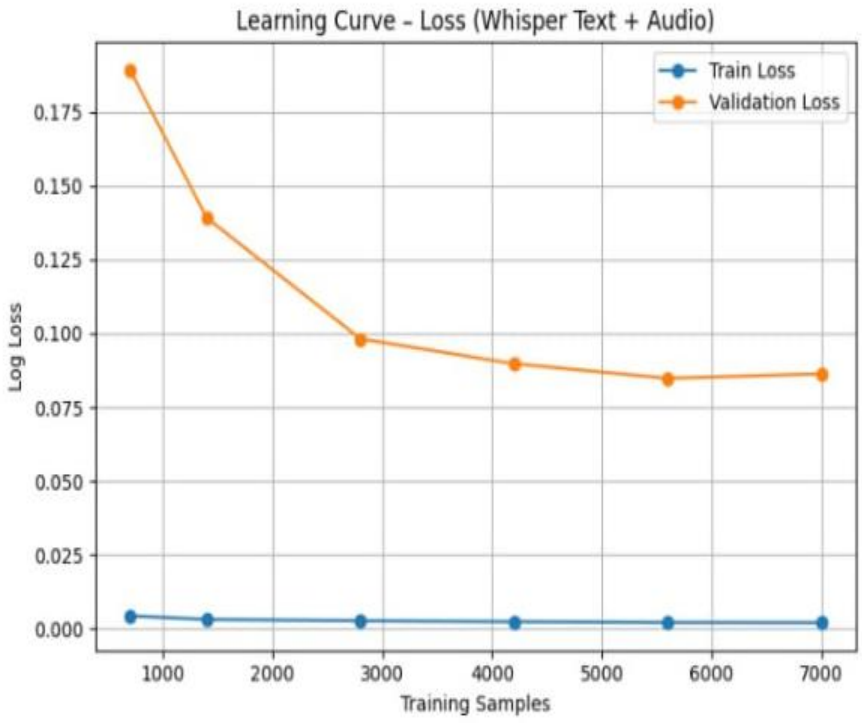
Voice Command Recognition

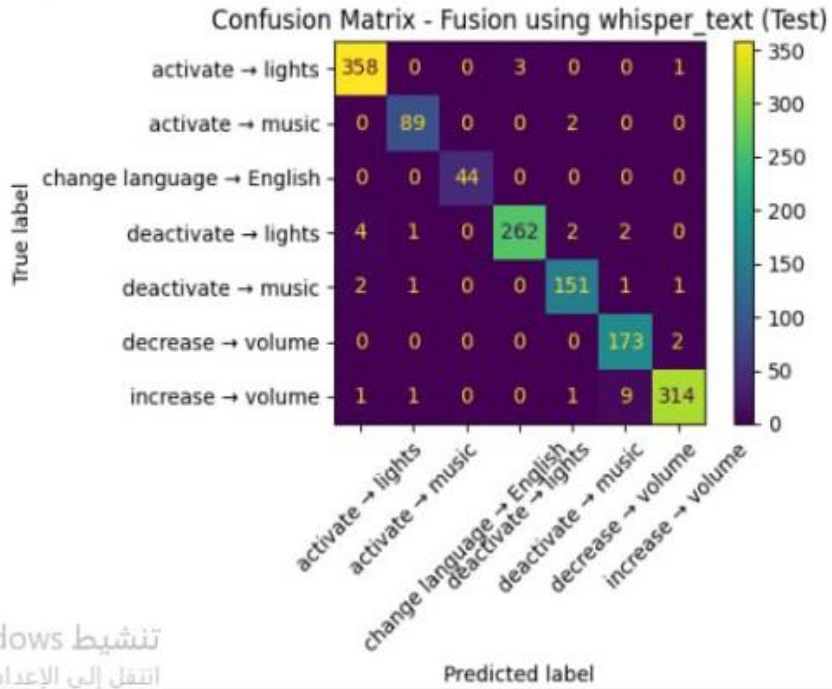
قائمة بأهم الأشكال











Classification Report (Test):

	precision	recall	f1-score	support
activate → lights	0.98	0.99	0.98	362
activate → music	0.97	0.98	0.97	91
change language → English	1.00	1.00	1.00	44
deactivate → lights	0.99	0.97	0.98	271
deactivate → music	0.97	0.97	0.97	156
decrease → volume	0.94	0.99	0.96	175
increase → volume	0.99	0.96	0.98	326
accuracy			0.98	1425
macro avg	0.98	0.98	0.98	1425
weighted avg	0.98	0.98	0.98	1425

الفصل الأول: مقدمة عامة

1.1 مقدمة عن المشروع :

شهد التفاعل بين الإنسان والحاسوب تطورًا ملحوظًا في السنوات الأخيرة مع التقدم السريع في مجالات الذكاء الاصطناعي ومعالجة الإشارات الصوتية. ومن بين أهم هذه التطورات أنظمة التحكم الصوتي التي تتيح للمستخدم تنفيذ الأوامر والتحكم بالأجهزة باستخدام الصوت بدلاً من وسائل الإدخال التقليدية مثل لوحة المفاتيح أو الفأرة. يركّز المشروع على فهم الأمر الصوتي وظيفيًا وليس فقط تحويل الصوت إلى نص، مما يجعله أكثر ملاءمة للتطبيقات العملية مثل أنظمة التحكم بالحاسوب، المنازل الذكية، والمساعدات الصوتية.

1.2 الهدف من المشروع :

يهدف هذا المشروع إلى تحقيق مجموعة من الأهداف، من أبرزها:

1. تصميم نظام ذكي للتعرف على الأوامر الصوتية اعتمادًا على الصوت الطبيعي للمستخدم.
2. استخراج النص من الصوت باستخدام نموذج Whisper ASR ومعالجة أخطاء التحويل الصوتي.
3. تصنيف الأوامر الصوتية إلى فئات واضحة *(Action → Object)* بدل الاكتفاء بالنص الخام.
4. بناء نموذج Fusion يجمع بين:
    - الميزات الصوتية المستخرجة منWhisper
    - الميزات النصية المستخرجة من نصWhisper
5. معالجة مشكلة عدم توازن البيانات باستخدام تقنيات Data Augmentation و Downsampling.
6. تحقيق دقة عالية في تصنيف الأوامر مع الحفاظ على قابلية التعميم.
7. تمهيد الطريق لإضافة مرحلة Reasoningلاحقًا باستخدام LLM أو قواعد منطقية

1.3 المشكلة التي يقوم المشروع بحلها

تواجه طرق التفاعل التقليدية مع أجهزة الحواسيب، مثل استخدام لوحة المفاتيح والفأرة، العديد من التحديات التي تؤثر على كفاءة المستخدمين وراحتهم ومن بين أبرز هذه التحديات، الصعوبات التي يواجهها ذوو الاحتياجات الخاصة، خصوصًا أولئك الذين يعانون من إعاقات حركية تمنعهم من استخدام الوسائل التقليدية للتحكم بالجهاز. في هذا السياق، تسعى أنظمة التحكم الصوتي إلى تقديم بديل عملي يمكنهم من التعامل مع الكومبيوتر بسهولة واستقلالية أكبر

إلى جانب ذلك، تسهم هذه الأنظمة في تحسين الإنتاجية، حيث تتيح للمستخدمين تنفيذ الأوامر بشكل أسرع، دون الحاجة إلى التنقل بين القوائم أو الضغط على أزرار متعددة كما تلعب دورًا مهمًا في تقليل الجهد البدني والإجهاد الناتج عن الاستخدام الطويل للأدوات التقليدية، مما يساعد على تجنب بعض المشاكل الصحية مثل آلام المعصم أو الكتف. وفي بيئات العمل متعددة المهام أو في الظروف التي تكون فيها اليدان مشغولتين، يصبح التحكم الصوتي وسيلة مثالية للتفاعل مع الحاسوب، إذ يوفّر للمستخدم مرونة أكبر ويجعل تجربة الاستخدام أكثر سلاسة وفعالية.

الفصل الثاني: الدراسة المرجعية

2.1 مقدمة

شهد مجال فهم اللغة المنطوقة (Spoken Language Understanding – SLU) تطورًا كبيرًا في السنوات الأخيرة نتيجة التقدم في تقنيات الذكاء الاصطناعي، ومعالجة الإشارات الصوتية، والنماذج اللغوية العميقة. وقد ركزت الأبحاث الحديثة على تحسين قدرة الأنظمة الصوتية على فهم نية المستخدم بدقة، والتعامل مع أخطاء التعرف التلقائي على الكلام، بالإضافة إلى دمج أكثر من وسيط (الصوت، النص، السياق) لتحقيق أداء أفضل في البيئات الواقعية.

2.2 الدراسة الأولى – SpeechVerse :نموذج صوت–لغة واسع النطاق

تناولت الدراسة الأولى تطوير نموذج متكامل يجمع بين معالجة الإشارة الصوتية والنمذجة اللغوية العميقة ضمن إطار موحد، أُطلق عليه اسم .**SpeechVerse** يهدف هذا النموذج إلى فهم الكلام مباشرة دون الاعتماد على مراحل منفصلة للتعرف على الصوت ثم تحليل النص.

اعتمد الباحثون على مجموعات بيانات صوتية كبيرة ومتنوعة، شملت لهجات وبيئات مختلفة، وتم تدريب النموذج على مهام متعددة مثل التعرف على الكلام، فهم النية، والتفاعل مع المستخدم. أظهرت نتائج الدراسة تحسنًا ملحوظًا في الدقة تراوح بين 9% و23% مقارنة بالنماذج التقليدية، خاصة في معالجة الأوامر القصيرة والمتشابهة صوتيًا.

2.3 الدراسة الثانية Word Confusion Networks :لتحسين متانةSLU

ركزت الدراسة الثانية على مشكلة أخطاء أنظمة التعرف التلقائي على الكلام (ASR) وتأثيرها السلبي على فهم النية. اقترح الباحثون استخدام **شبكات التباس الكلمات (Word Confusion Networks)** كتمثيل وسيط يدمج عدة فرضيات ناتجة عن ASR بدل الاعتماد على نص واحد فقط.

تم تقييم النموذج على بيانات صوتية متنوعة تحتوي على مستويات عالية من الضوضاء، وأظهرت النتائج تحسنًا واضحًا في مقاومة الأخطاء مقارنة بالأساليب التقليدية. تعكس هذه الدراسة أهمية التعامل مع عدم يقين مخرجات Whisper أوASR

2.5الدراسة الرابعة: توليد البيانات باستخدام LLM لتصنيف النية باللغة الألمانية

استعرضت الدراسة الرابعة دور النماذج اللغوية الكبيرة في تحسين تصنيف النية للغات محدودة الموارد، مع التركيز على اللغة الألمانية. اعتمد الباحثون على توليد أمثلة لغوية جديدة تغطي تراكيب وأساليب مختلفة، ثم تدريب نماذج تصنيف تقليدية على البيانات الأصلية والموسعة.

أظهرت النتائج تفوقًا واضحًا للنماذج المدربة على البيانات الموسعة، مما يبرز قدرة LLM على معالجة مشكلة محدودية الموارد اللغوية

2.6 الدراسة الخامسة: توليد عبارات جديدة باستخدام LLM لتوسيع بياناتSLU قدمت الدراسة الخامسة منهجية جديدة تعتمد على توليد عبارات لغوية متنوعة باستخدام LLM بهدف زيادة تنوع بيانات فهم اللغة المنطوقة. تم تطبيق المنهجية على مجموعات بيانات صغيرة الحجم، وأظهرت النماذج المدربة على البيانات الموسعة أداءً أفضل في التعامل مع العبارات غير المرئية سابقًا.

2.8الدراسة السادسة: فهم النية متعدد الوسائط(EMNLP 2025)

استعرضت الدراسة السابعة أحدث الاتجاهات في نماذج فهم النية متعددة الوسائط، مع التركيز على دمج الصوت والنص والسياق الدلالي ضمن نموذج موحد. أظهرت النتائج أن النماذج المعتمدة على LLM متعددة الوسائط تحقق أعلى مستويات الدقة، خاصة في السيناريوهات المعقدة التي تتطلب فهمًا عميقًا للسياق.

الفصل الثالث: منهجية العمل

3.1 مقدمة

تُعد أنظمة التعرف على الأوامر الصوتية من أكثر تطبيقات الذكاء الاصطناعي تحدّيًا، وذلك بسبب الطبيعة المعقدة للإشارات الصوتية التي تتأثر بعوامل متعددة مثل الضوضاء، اختلاف اللهجات، سرعة النطق، ونبرة الصوت.

يتطلب فهم الأوامر الصوتية بدقة الجمع بين تحليل الإشارة الصوتية ومعالجة اللغة الطبيعية، حيث إن الاعتماد على الصوت فقط أو النص فقط قد يؤدي إلى ضعف في التعميم عند العمل في بيئات واقعية.

يساهم تطوير أنظمة فعّالة للتعرف على الأوامر الصوتية في تحسين التفاعل بين الإنسان والحاسوب، خاصة في تطبيقات التحكم بالأجهزة، الأنظمة الذكية، ودعم المستخدمين ذوي الاحتياجات الخاصة.

في هذا الفصل، يتم استعراض المنهجية المتبعة لبناء نظام ذكي قادر على التعرف على الأوامر الصوتية وتنفيذها، بدءًا من جمع البيانات الصوتية، مرورًا بمرحلة التحضير والمعالجة، وصولًا إلى بناء نموذج الدمج (Fusion) وتقييم أدائه باستخدام مقاييس علمية دقيقة.

3.2 منهجية العمل العامة

تم بناء النظام المقترح باتباع منهجية متكاملة شملت:

جمع البيانات الصوتية والنصية ،إعداد البيانات وتحسين جودتها، توحيد الأوامر وربطها بالنية الصحيحة(Mapping) ، توسيع البيانات (Augmentation) ثم موازنتها، استخراج الميزات الصوتية والنصية ، بناء نموذج دمج بين الصوت والنص، تقييم النموذج باستخدام بيانات اختبار واقعية. اعتمدت المنهجية على التحليل العملي والتجريبي مع مقارنة النتائج في كل مرحلة، لضمان بناء نموذج موثوق وقابل للتعميم

3.3 مجموعة البيانات المستخدمة

تم استخدام **Fluent Speech Commands Dataset**، وهي مجموعة بيانات مخصصة للأوامر الصوتية، تحتوي على تسجيلات حقيقية لمستخدمين متعددين.

تتضمن البيانات أوامر صوتية قصيرة مرتبطة بأفعال محددة مثل:

- تشغيل/إيقاف الإضاءة.
- التحكم بمستوى الصوت.
- تشغيل/إيقاف الموسيقى.
- تغيير لغة النظام.

تتميز البيانات بتنوع المتحدثين وصيغ الأوامر، مما يجعلها مناسبة لتقييم أداء النموذج في بيئات قريبة من الواقع.

3.4تحضير البيانات(Data Preparation)

3.4.1تحميل البيانات ودمج الملفات الأصلية

تم الاعتماد على ملفات البيانات الأصلية ضمن Fluent Speech Commands والمقسّمة مسبقًا إلى train_data.csv / valid_data.csv / test_data.csv

جرى تحميل هذه الملفات ثم دمجها ضمن DataFrame موحّد لتوحيد المعالجة والتحضير على كامل البيانات.

بعد الدمج، أصبح حجم البيانات الكلي: (30043, 7)

تضمّن كل سجل: مسار الملف الصوتي، النص(transcription) ، معرف المتحدث، الفعل (action)، الهدف(object) ، والموقع.(location)

3.4.2 تنظيف البيانات وإزالة الأعمدة غير الضرورية

بهدف جعل البيانات أقرب لمجال مشروع التحكم بالحاسوب، تم حذف أعمدة لا تخدم هدف النظام:

Location: غير مؤثر في التحكم بالحاسوب.

speakerId : اختياري، وتمت إزالته لتبسيط البيانات.

Unnamed

إضافةً إلى ذلك، تم استبدال القيمة "none"بقيمة مفقودة NA في سياقات التنظيف.

حجم البيانات بعد التنظيف الأساسي بقي: (30043, 7) وذلك لأن التنظيف هنا استهدف الأعمدة وقيم "none"دون حذف عينات كاملة في هذه المرحلة.

3.4.3إعادة تعريف الأوامر وتحويلها إلى نوايا قياسية Mapping

بسبب وجود اختلافات كبيرة بين صيغ النصوص والحقول الأصلية، تم تصميم دالة Mappingتعتمد على النص الفعلي (transcription) لإعادة تعيين (action, object) إلى نوايا موحدة.

اعتمدت عملية الـ Mapping على قواعد لغوية بسيطة عبر كلمات مفتاحية، مثل:

كلمات الإيقافmusic → deactivate → (pause/stop/mute…)

كلمات التشغيلmusic → activate → (play/resume/start…)

كلمات رفع الصوتvolume → increase → (loud/up/increase…)

كلمات خفض الصوتvolume → decrease → (down/decrease/lower…)

وجود language أوEnglish → change language → english

أوامر الإضاءة تعتمد علىactivate/deactivate lights → (on/off)

بعد ذلك تم إنشاء عمود النية النهائي بصيغة موحدة:

pair = action → object

المشروع تم توجيهه إلى 7فئات Intent ثابتة وواضحة بدل نوايا متفرعة كثيرة.

3.4.4 إزالة فئات غير مرغوبة خارج نطاق المشروع

نظرًا لأن هدف النظام هو التحكم بالحاسوب (وليس التحكم بأشياء منزلية متنوعة)، تم حذف بعض الأهداف غير المناسبة مثل:

heat, temperature, lamp, newspaper, juice, socks, shoes, German, Korean, Chinese

ساعدت هذه الخطوة في جعل البيانات تمثل نطاق مشروعك الحقيقي وتقليل التشويش.

3.4.5 تقسيم البيانات Stratified Split على مستوى النية Pair

بعد الحصول على ملف البيانات النهائي بعد الـ Mapping والتنظيف (حجمه : 14241)، تم تقسيمه إلى: Train: 80% Validation: 10% Test: 10%

تم تنفيذ التقسيم بطريقة Stratified اعتمادًا على عمودpair ، لضمان بقاء نفس النسبة لكل نية ضمن كل.split

الأحجام النهائية: Train: 11392 Valid: 1424 Test: 1425

3.4.6معالجة عدم توازن البيانات داخل التدريب(Hybrid Balancing)

بعد التقسيم، ظهر أن بيانات التدريب غير متوازنة (بعض النوايا أكثر بكثير من غيرها).

لذلك تم اعتماد استراتيجية هجينة تتكون من مرحلتين:

(أ) رفع الفئات الضعيفة عبرAugmentation

تم تحديد حد أدنى للأزواج الضعيفة LOWER_TARGET = 1000

أي فئة تدريب عددها أقل من 1000 يتم دعمها بإنتاج عينات جديدة عبر augmentation.

خفيف يشمل:

Shift زمني

Time-stretch

Pitch shift

إضافةNoise

عدد العينات التي تم توليدها: 913 عينة جديدة

بعد الدمج مع بيانات التدريب الأصلية وإجراء Hybrid balancing أصبح توزيع التدريب:

activate → lights : 1500

activate → music : 1000

change language → English : 1000

deactivate → lights : 1500

deactivate → music : 1250

decrease → volume : 1395

increase → volume : 1500

(ب) توحيد جميع الفئات عبر Downsampling النهائي

تم تنفيذ Snap Downsampling بحيث يتم توحيد جميع الفئات على عدد واحد ثابت يساوي أصغر فئة.

تم أخذ 1000 عينة فقط من كل فئة (للفئات الأكبر)، لتصبح جميع الفئات متساوية تمامًا:

جميع الأزواج 1000 = لكل فئة أصبح إجمالي التدريب النهائي 7000

3.4.8 توحيد صيغة الإشارات الصوتية(Audio Normalization)

تم توحيد الإشارات الصوتية لضمان ثبات المدخلات للنموذج.

تختلف الملفات الصوتية في مجموعة البيانات من حيث:

عدد القنوات(Mono / Stereo)

معدل أخذ العينات(Sampling Rate)

لذلك تم اعتماد دالة مخصصة لمعالجة كل ملف صوتي وفق الخطوات التالية:

تحميل الملف الصوتي باستخدام مكتبة torchaudio، حيث يتم تحميل الإشارة على شكل Tensor.

تحويل الإشارة إلى قناة واحدة (Mono) في حال كانت الإشارة متعددة القنوات، وذلك عبر حساب المتوسط بين القنوات.

إعادة أخذ العينات (Resampling) إلى معدل ثابت قدره 16 kHzفي حال كان الملف الأصلي بمعدل مختلف.

إرجاع الإشارة الصوتية بصيغة موحدة على شكل متجه أحادي البعد (1D Tensor) يمثل العينات الزمنية فقط.

3.5 استخراج الميزات(Feature Extraction)

3.5.1الميزات الصوتية باستخدام Whispe

تم استخدام نموذج Whisper كـ Encoder Feature Extractorعبر:

تجهيز المدخلات باستخدام processorلإنتاج input_featuresبشكل.(1, 80, T)

تمرير input_featuresإلى Encoderفقط.((model.get_encoder()

الحصول على last_hidden_stateثم تطبيق Temporal Mean Pooling(أخذ المتوسط عبر الزمن) للحصول على embedding ثابت الطول.

آلية الـ:pooling

إذا كانت last_hidden_stateبشكل(1, T, D)

نطبق المتوسط على محور الزمن ← الناتج(1, D)

ثم نحوله إلى متجه 1 DبطولD

3.5.3معالجة الملفات الصوتية القصيرةPadding

أثناء الاستخراج ظهرت ملفات قصيرة ينتج عنها طول زمني صغير داخل .input_features

لضمان ثبات المعالجة داخل النموذج، تم اعتماد شرط:

إذا كان 3000 > Tيتم تنفيذ Paddingإلى 3000.

تم تسجيل عدد الملفات التي احتاجت padding ضمن المتغير short_countبهدف التوثيق والتحليل لاحقًا.

3.5.4 بناء ملف الميزات النهائي لكلSplit

بعد استخراج embedding لكل ملف، تم بناء DataFrame ميزات نهائي يتضمن:

) f767 ... f0عددها 768 ميزة(الأعمدة الرقمية

أعمدة تعريفية للمطابقة والتتبع: full_path pair object action

3.6 دمج الميزات(Feature Fusion)

تم دمج الميزات الصوتية مع الميزات النصية باستخدام أسلوب .**Concatenation**

يهدف هذا الدمج إلى الاستفادة من:

المعلومات الصوتية الخام.

المعنى اللغوي المستخرج من النص.

أظهرت هذه الخطوة تحسنًا واضحًا في أداء النموذج مقارنة باستخدام كل وسيط على حدة

3.7 بناء نموذج التصنيف

تم استخدام نموذج **Logistic Regression متعدد الفئات** لتصنيف الأوامر إلى النية الصحيحة.

سبقت عملية التدريب مرحلة **Standardization** لتوحيد نطاق القيم العددية.

اختير هذا النموذج لكونه:

بسيطًا وقابلًا للتفسير.

مناسبًا لحجم البيانات.

فعّالًا كخط أساس (Baseline) للمقارنة.

3.8 تقييم النموذج

تم تقييم النموذج باستخدام:

الدقة.(Accuracy)

مقياس.F1-Score

مصفوفة الالتباس.(Confusion Matrix)

منحنيات التعلم.(Learning Curves)

أُجري التقييم على:

بيانات اختبار قياسية.

بيانات واقعية تعتمد على نصوص Whisper غير المثالية.

أظهرت النتائج أن النموذج يحقق أداءً عاليًا عند تصنيف الأوامر الصحيحة، مع وجود تحديات في الحالات الخارجة عن نطاق الأوامر.(Out-of-Domain)

3.9 آلية رفض الأوامر غير الصالحة(Rejection Mechanism)

تم تطوير آلية بسيطة لاكتشاف الأوامر غير الصالحة أو غير الواضحة، مثل:

الجمل العامة.

النصوص غير المرتبطة بالأوامر.

أخطاء التعرف الشديدة.

ساهمت هذه الآلية في:

تقليل الأخطاء في التنبؤ.

رفع موثوقية النظام في الاستخدام الواقعي

Voice Command Recognition

Voice Command Recognition

مشكلة

Voice Command Recognition