Journal of
Engineering Sciences
Faculty of Engineering
Assiut University
JES
مجلة
العلوم الهندسية
كلية الهندسة
جامعة أسيوط
journal homepage: http://jesaun.journals.ekb.eg

Ten2Zero: A Balanced Audio Dataset to Teach Machine Learning

عشرة لصفر: مجموعة بيانات صوتية متوازنة الأصناف للأرقام العربية المنطوقة

Received 24 August 2023; Revised 18 October 2023; Accepted 19 October 2023

Ghassan F. Bati[1]
غسان بن فاروق باتي[1]

Saudi is interested in artificial intelligence and machine learning. Governmental interest appears in several forms, most notably creating a generation that masters the skills of artificial intelligence and machine learning through the approval of the Saudi Ministry of Education to teach artificial intelligence, machine learning, and data science skills in public schools and universities. This interest makes it imperative for researchers to develop Arabic datasets for research and educational purposes, especially with the popularity of English sources and the absence of Arabic sources. This study attempts to fill this gap by creating a dataset for the Arabic spoken digits from ten to zero and analysing it using Orange, which requires no coding. The importance of the study is as follows: first Arabic work to establish a balanced audio dataset of spoken Arabic digits from ten to zero; the dataset contains audio files and the tabular data generated using deep learning from the spectrograms of the audio files; it is the first Arabic scientific work that uses traditional machine learning and deep learning models to create good-performing models for classifying spoken Arabic digits without coding, which enables researchers and those interested in various fields to develop machine learning applications to classify Arabic audio, especially in mobile phones or in microcontrollers, to stimulate IoT applications and Tiny machine learning.

الملخص

تهتم السعودية بالذكاء الاصطناعي وتعلم الآلة. ويظهر الاهتمام الحكومي في عدة أشكال أبرزها إنشاء جيل متقن لمهارات الذكاء الاصطناعي وتعلم الآلة من خلال إقرار وزارة التعليم السعودية لتدريس مهارات الذكاء الاصطناعي وتعلم الآلة وعلم البيانات في مدارس التعليم العام والجامعات. وهذا الاهتمام يحتم على الباحثين تطوير مجموعات بيانات عربية؛ للأغراض البحثية والتعليمية خاصة مع شهرة المصادر الإنجليزية، وشح المصادر العربية. تحاول هذه الدراسة ملء الفراغ من خلال إنشاء مجموعة بيانات عربية، وتحليلها باستخدام برنامج أورانج والذي لا يحتاج إلى برمجة. تكمن أهمية الدراسة في التالي: أول عمل علمي عربي محكم ينشئ ويحلل مجموعة بيانات صوتية متوازنة الأصناف لتصنيف الأرقام العربية المنطوقة من عشرة لصفر، وتتميز مجموعة البيانات باحتوائها على الملفات الصوتية المستخرجة من تسجيل الأرقام العربية المنطوقة، وكذلك احتوائها على الصور الطيفية لتصنيف الأرقام العربية الصوتية والمستخرجة من الملفات الصوتية، واحتوائها على البيانات المجدولة ذات الخصائص المولدة باستخدام التعلم العميق للأرقام العربية المنطوقة والمستخرجة من الصور الطيفية. يعد البحث أول بحث منشور باللغة العربية يستخدم نماذج تعلم الآلة التقليدية والتعلم العميق لإنشاء نماذج ذات أداء عال لتصنيف الأرقام العربية الصوتية من عشرة لصفر بدون برمجة، مما يمكن الباحثين والمهتمين من كافة المجالات من تطوير تطبيقات تعلم آلة لتصنيف الأصوات العربية خاصة في الجوالات أو في المتحكمات الدقيقة تفعيلاً لتطبيقات إنترنت الأشياء وتعلم الآلات الصغيرة.

Keywords: Supervised Learning, Deep Learning, Image Classification, Image Embedding, Audio Classification

الكلمات الرئيسية
التعلم الموجه، التعلم العميق، تصنيف الصور، تضمين الصور، تصنيف الأصوات

[1] أستاذ مساعد بقسم هندسة الحاسب والشبكات بكلية الحاسبات جامعة أم القرى (gfbati@uqu.edu.sa)

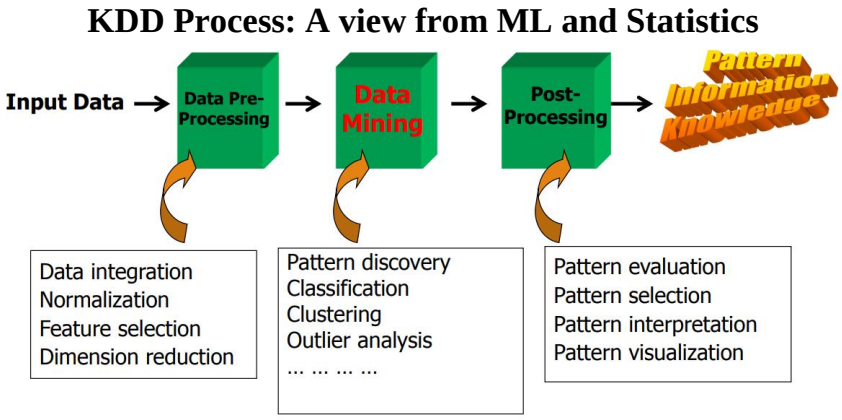
https://doi.org/10.21608/JESAUN.2023.231628.1254

Journal of
Engineering Sciences
Faculty of Engineering
Assiut University
JES
مجلة
العلوم الهندسية
كلية الهندسة
جامعة أسيوط
journal homepage: http://jesaun.journals.ekb.eg

١. المقدمة

تولي المملكة العربية السعودية كبير الاهتمام للذكاء الاصطناعي (*Artificial Intelligence "AI"*) وتعلم الآلة (*Machine Learning "ML"*). يبرز الاهتمام الحكومي في عدة أمور أحدها إنشاء الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) [1]. أمر آخر لا يقل أهمية عن سدايا وهو اهتمام وزارة التعليم السعودية بإعداد جيل يبتكر حلولاً مبدعة من خلال استحداث مناهج عصرية تدرس أبرز تقنيات المستقبل كالذكاء الاصطناعي وعلم البيانات وإنترنت الأشياء [2]. هذا الاهتمام يحث الباحثين من مختلف التخصصات لإبداع مجموعات بيانات عربية تساعد الدارسين والباحثين في مسيرتهم نحو إتقان الذكاء الاصطناعي وتعلم الآلة، وهذا ما تسعى الورقة لإنشائه والمساهمة فيه من خلال مجموعة البيانات عشرة لصفر، خاصة مع وفرة المصادر الإنجليزية وندرة المصادر العربية على مستوى الأبحاث ومجموعات البيانات [4], [3].

نعيش اليوم في عالم مليء بالبيانات الصادرة من مختلف الأجهزة حولنا، فمن هواتفنا الجوالة ومستشعراتها المختلفة، ومن الساعات الذكية والأجهزة القابلة للارتداء عمومًا، ومن المساعدات الصوتية الذكية مثل سيري وأليكسا، ومن كاميرات المراقبة،... -على سبيل المثال لا الحصر-. دائمًا ما نحرص على تنقيب هذه البيانات ((*Data Mining*) (*Knowledge Discovery from Data "KDD") or*)) لاستخراج أبرز الأنماط داخلها ولتحويل البيانات إلى معلومات ومعرفة تساعدنا على اتخاذ القرارات المختلفة وتحسين جودة حياتنا بشكل عام. وللقيام بعملية التنقيب عن البيانات باستخدام تعلم الآلة، فيلزمنا القيام بثلاث خطوات رئيسية. الخطوة الأولى: هي معالجة البيانات (*Preprocessing*) ويتم فيها تهيئة البيانات المراد تعليمها للآلة (الحاسوب) لاستخراج أفضل ما فيها، وهي خطوة مهمة؛ لأن البيانات السيئة لن تنتج إلا معرفة سيئة، والعكس صحيح، فالبيانات المتميزة، ستنتج معرفة متميزة. بالإضافة إلى اختيار أبرز الخصائص التي يمكن استخدامها لتعليم الحاسوب الشيء المطلوب. أما الخطوة الثانية فهي المعالجة (*Processing*) ويتم فيها نمذجة البيانات بطرق متعددة، تهتم هذه الورقة بالتصنيف (*Classification*)، والذي سيبين معناه لاحقًا. أما الخطوة الثالثة والأخيرة فهي المعالجة اللاحقة (*Postprocessing*) ويتم فيها معالجة مخرجات النموذج بعد تشغيله في الخطوة الثانية وتقييمها والتأكد من صحة أداء النموذج وتفسير نتائجه [7]–[5]، كما يظهر في شكل رقم١.



شكل (١) عملية تنقيب البيانات واكتشاف المعرفة داخلها بعيون باحثي الإحصاء وتعلم الآلة [7].

للتأكيد على أهمية وجود وإنشاء مجموعات بيانات (*Datasets*) عربية، يمكن تصفح أحد أشهر مواقع الإنترنت المتخصصة في الذكاء الاصطناعي وعلم الآلة Huggingface وهو يحوي عند كتابة البحث 5,033 مجموعة بيانات إنجليزية، بينما بلغت مجموعات البيانات العربية ٢٦١ مجموعة بيانات، منها ٧ مجموعات بيانات مختصة بتصنيف الأصوات (*Audio Classification*). الملاحظة ذاتها موجودة في موقع آخر شهير كذلك في المجال ذاته وهو paperswithcode، حيث يحوي عند كتابة هذا البحث ٢٤٦١ مجموعة بيانات إنجليزية، في مقابل ٧٨ مجموعة بيانات

https://doi.org/10.21608/JESAUN.2023.231628.1254

عربية، منها ٧ مجموعات بيانات صوتية عربية. مما يبين أهمية إنشاء مجموعات بيانات عربية تساعد الدارسين والباحثين والمبتكرين في شتى مجالات الحياة [9] ,[8].

نظرًا لكثرة المصطلحات العربية وما يقابلها بالإنجليزية في هذا البحث، وخشية الإطالة، فإن كل المصطلحات العربية كتب ما يقابلها بالإنجليزية في ثنايا البحث بين قوسين *وبخط مائل*، ولم يفرد لها قسم خاص. جل المصطلحات العربية المستخدمة في هذه الورقة مستقاة من "معجم البيانات والذكاء الاصطناعي عربي - إنجليزي" المعد من قبل سدايا ومجمع الملك سلمان العالمي للغة العربية [١٠]. تعلم الآلة -بحسب المعجم- هو فرع من فروع الذكاء الاصطناعي الذي يرتكز على تعلم الأنماط من مجموعة البيانات للتنبؤ بأصناف مجموعة البيانات أو اتخاذ قرارات مبنية على بيانات جديدة دون برمجة صريحة [١٠].

تكمن أهمية الدراسة في التالي:

١- أول بحث -في حدود علمي- محكم ومنشور باللغة العربية ينشئ مجموعة بيانات صوتية متوازنة الأصناف للأرقام العربية من عشرة لصفر باللغة العربية الفصحى وبعض اللهجات المحلية السعودية من المنطقة الغربية خاصة لهجة مدينة مكة المكرمة.

٢- أول مجموعة بيانات صوتية متوازنة الأصناف لا تكتفي بنشر البيانات (الملفات الصوتية) فحسب، بل تجمع بين البيانات الصوتية، والبيانات المجدولة (*Tabular Data*) ذات الخصائص أو السمات (*Features or Attributes*) المستقاة من خصائص الصور الطيفية (*Spectrograms*) المستخرجة من الملفات الصوتية باستخدام تقنيات تضمين الصور (*Image Embedding*) والتعلم العميق (*Deep Learning*).

تحاول هذه الدراسة الإجابة على السؤالين البحثيين التاليين:

١- ما أفضل نموذج تعلم آلة لتصنيف الأرقام العربية المنطوقة من عشرة لصفر؟

٢- من الأفضل مضمن الصور SqueezeNet أو Inception V3 والذي يعطي نتائج أفضل لتصنيف أصوات الأرقام العربية المنطوقة من عشرة لصفر من خلال الصور الطيفية؟

تعرض الأقسام المتبقية في البحث أبرز مجموعات البيانات المنشورة في الإنترنت للأرقام العربية المنطوقة، ومن ثم تفصيل لمجموعة بيانات هذه الورقة عشر لصفر وطريقة جمعها وإعدادها، وتختم بعرض أبرز النتائج وخلاصة للبحث ودراسات مستقبلية مقترحة.

٢. الدراسات السابقة

لم يجد الباحث أي دراسة علمية سابقة مكتوبة ومنشورة باللغة العربية لتصنيف الأرقام العربية المنطوقة باستخدام أي من تقنيات الذكاء الاصطناعي المختلفة وهذا الأمر يبين الحاجة الماسة إلى هذه الدراسة، على الرغم من وجود العديد من الأبحاث باللغة الإنجليزية في الموضوع ذاته. تم محاولة الوقوف على أبرز مجموعات البيانات المتعلقة بالأرقام العربية المنطوقة المنشورة في الإنترنت والمتاحة مجانًا للتنزيل. واحدة من أقدم مجموعات البيانات في هذا المجال هي الرقم العربي المنطوق (*Spoken Arabic Digit*). تحتوي هذه المجموعة على بيانات مجدولة (*Tabular Data*) تجمع ٨٨٠٠ عينة (*Instance*) تم جمعها من خلال ٤٤ متطوع و٤٤ متطوعة قاموا بتسجيل الأرقام العربية من صفر إلى تسعة عشر مرات لكل رقم. لم يتم نشر التسجيلات الصوتية، نشرت فقط البيانات المجدولة التي تحوي ١٣ سمة أو صفة (*Feature or Attribute*) تمثل البيانات الطيفية لكل ملف صوتي والتي يمكن استخدامها لتصنيف الأرقام العربية المنطوقة من صفر لتسعة [11].

مجموعة بيانات أخرى وهي الأوامر الصوتية العربية (*Arabic Speech Commands Dataset*) قام جامعوها بتسجيل ٤٠ أمرًا صوتيًا عربيًا شاملة الأرقام العربية من صفر لتسعة. قام بعملية التسجيل ٣٠ متطوعًا، ١٩ رجلاً و١١ امرأة من أعمار ومدن سورية مختلفة. شارك مؤلفوها كافة التفاصيل المتعلقة بها من ملفات صوتية وبيانات مجدولة وأكواد [12]. توجد دراسات أخرى للأرقام العربية المنطوقة، لكن مجموعات بياناتها لم تنشر، مثل [17]–[13] - على سبيل المثال لا الحصر-.

٣. مجموعة البيانات عشرة لصفر

قام المؤلف بجمع مجموعة البيانات عشرة لصفر بمساعدة طلابه في قسم (هندسة الحاسب والشبكات) بجامعة (أم القرى) حين قام بتدريسهم مادة تعلم الآلة لمهندسي الحاسب في الفصل الدراسي الثاني لعام ١٤٤٤هـ مقابل الحصول على خمس درجات إضافية حال إنجاز المطلوب بدقة. يمكن الوصول لمجموعة البيانات عشرة لصفر من خلال الرابط التالي: https://huggingface.co/datasets/gfbati/Ten2Zero.

تم إعطاء الطلبة الإرشادات التالية؛ لجمع البيانات بدقة:

١- يجب تسجيل الأصوات التالية للأرقام العربية من صفر إلى عشرة باللغة العربية الفصحى وبلهجتك العامية:

أ. صفر
ب. واحد، واحدة
ت. اثنان، اثنين
ث. ثلاث، ثلاثة
ج. أربع، أربعة
ح. خمس، خمسة
خ. ست، ستة
د. سبع، سبعة
ذ. ثمان، ثمانية
ر. تسع، تسعة
ز. عشر، عشرة
س. ملف واحد لصوت مكان التسجيل (لمدة ٢٠ ثانية).

٢- يجب أن تكون صيغ الملفات "wav". كل رقم يتم تسجيله خمس مرات = خمس ملفات صوتية. يجب ألا يزيد وقت الملف الواحد عن ثانية واحدة فقط، ما عدا ملف صوت غرفة التسجيل = ٢٠ ثانية. مجموع الملفات = ١١ كلاس (رقم) * ٥ ملفات لكل رقم + ملف واحد لصوت غرفة التسجيل = ٥٦ ملفًا صوتيًا بصيغة "wav".

٣- يمكن استخدام أي برنامج أو تطبيق تسجيل أصوات في حاسوبك أو جوالك، مثل -على سبيل المثال لا الحصر-: Audacity.

٤- يجب أن يتم التسليم في مجلد مضغوط واحد باسم الطالب الرباعي + الرقم الجامعي وبصيغة zip، يحوي ١٢ مجلدًا. لا تنس ملء ملف معلوماتي بدقة! يمكن استخدام برنامج 7-zip للضغط أو برنامجك المفضل.

أعطي الطلبة أسبوعًا كاملاً لإنجاز المهمة. بلغ عدد الطلبة في الشعبة ٣١ طالبًا، قام ٢٤ منهم بتسليم الملف المطلوب، قام الباحث بالتأكد من الالتزام بالتعليمات أعلاه، لذا تم الاقتصار فقط على بيانات ١٩ طالبًا؛ وحذفت ٧ محاولات؛ لنقص الملفات الصوتية المطلوبة وعدم الالتزام بالتعليمات. يبرز جدول رقم ١ أسماء الطلبة المشاركين، ومعلومات عن أجهزة تسجيلهم ولهجاتهم.

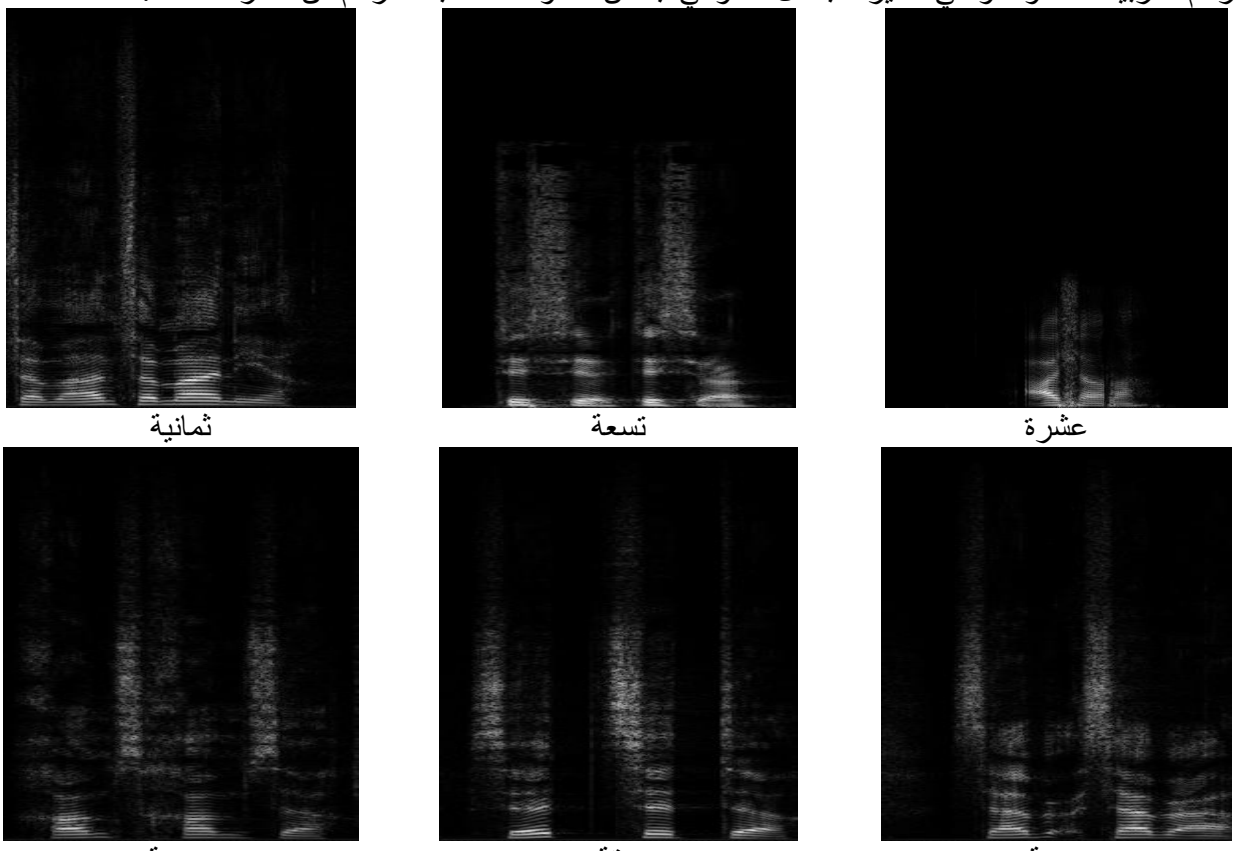
جدول (١) الطلبة المشاركون في جمع مجموعة البيانات عشرة لصفر وخصائص أجهزة تسجيلهم ولهجاتهم.

اسم الطالب	جهاز التسجيل	برنامج التسجيل	اللهجة
عبد الله بن أحمد الزهراني	IdeaPad Gaming 3 15ARH05	Audacity	العربية الفصحى
عبد الرحمن بن هاني خالدي	Huawei Matebook	Audacity	العربية الفصحى ومدينة مكة المكرمة
أحمد بن عمر اللهيبي	PC	Windows Sound Recorder	العربية الفصحى ومدينة مكة المكرمة
رياض بن سلطان الشهري	Laptop	Audacity	العربية الفصحى ومدينة النماص
فهد بن شاكر الفضلي	Samsung S9	Audio Recorder	العربية الفصحى ومدينة مكة المكرمة
عبد الله بن فايز البركاتي	مايك احترافي	Audacity	العربية الفصحى ومدينة مكة المكرمة
سعود بن مشعل القرشي	iPad	المذكرات الصوتية	العربية الفصحى ومدينة مكة المكرمة
عبد الرحمن بن منصور حلواني	PC External Microphone	Audacity	العربية الفصحى ومدينة مكة المكرمة
عبد الرحمن بن نجم الوارث بتي وارث	TOSHIBA SATELLITE (L50-A668)	Audacity	العربية الفصحى ومدينة مكة المكرمة
عبد الله بن محمد كديش	Dell Optiplex 9010	Audacity	العربية الفصحى
متعب بن محمد الحارثي	Acer Laptop	Audacity	العربية الفصحى ومدينة الطائف
محفوظ بن خالد فلمبان	PC	Audacity	العربية الفصحى ومدينة مكة المكرمة
عبد الله بن محمد العتيبي	iPhone 14 Pro	المذكرات الصوتية	العربية الفصحى ومدينة مكة المكرمة
مشاري بن أحمد الجيزاني	BM-800 Microphone	Audacity	العربية الفصحى والحجازية
غسان بن نبيل حداد	Microphone of HyperX cloud 2 headsets	Audacity	العربية الفصحى ومدينة مكة المكرمة
محمد بن عاطف الشبراوي	Acer Aspire A315-55G	Google Teachable Machine	العربية الفصحى ومدينة غزة الفلسطينية
مجاهد بن سعيد باوزير	Aspire F 15	Audacity	العربية الفصحى ومدينة مكة المكرمة
سعد بن ياسر المطرفي	iPhone 11	المذكرات الصوتية	العربية الفصحى ومدينة مكة المكرمة
عمر بن حسن مصطفى	Dell Inspiron 3521	Audacity	العربية الفصحى ومدينة مكة المكرمة

يحوي مجلد مجموعة البيانات عشرة لصفر ٤ مجلدات. المجلد الأول "Dataset" يحوي الملفات الصوتية بصيغة wav من عشرة لصفر، وكذلك الصور الطيفية (*spectrograms*)، كل رقم في مجلد خاص به. المجلد الثاني "Students" يحوي أسماء الطلبة المشاركين في جمع الملفات الصوتية ومعلومات تفصيلية عنهم وعن أجهزة التسجيل المستخدمة، كل طالب من الطلبة التسعة عشر في مجلد خاص به. المجلد الثالث "Testing" يحوي محاولات الطلبة غير المكتملة أو الذين قدموا ملفات أكثر من المطلوب، يمكن استخدام هذه الملفات في عمليات مختلفة من أبرزها -على سبيل المثال لا الحصر- اختبار نماذج الآلة المختلفة. المجلد الرابع "audio2spec-master" مأخوذ من الإنترنت ويحوي الكود البرمجي المكتوب بلغة بايثون والمعتمد على مكتبة librosa والذي يقوم بتحويل الملفات الصوتية بصيغة wav إلى صور طيفية (*spectrograms*) [19] ,[18]. قامت الأداة بتحويل ٨٥ ملفًا صوتيًا إلى صور طيفية بصيغة png من أصل ٩٥ ملفًا صوتيًا لكل رقم من عشرة لصفر. هذا يعني أن عدد الملفات لكافة الأرقام العربية المنطوقة من عشرة لصفر = ٨٥ صورة * ١١ رقمًا = ٩٣٥ صورة طيفية. سميت المجلدات بالإنجليزية ليتم نشر مجموعة البيانات عشرة لصفر في كافة المواقع ذات العلاقة بمجموعات البيانات وعلم الآلة؛ لإثراء المحتوى العربي ونشر العلم والمعرفة لأوسع نطاق ممكن. تم استخدام الصور الطيفية في النمذجة؛ لشهرتها في هذا المجال [4]. كما يحوي المجلد الرئيس العديد من الملفات التي تم استخدامها لاستخراج خصائص الأرقام المنطوقة من مضمني الصور (Inception v3 and SqueezeNet)؛ للتصنيف، وكذلك ملفات برنامج أورانج لتنقيب البيانات (الإصدار ٣,٣٦) والتي تم استخدامها لبناء نماذج تعلم الآلة لتصنيف الأرقام العربية وتقييمها [20]، كما سيبين في القسم القادم.

٤. نمذجة مجموعة البيانات عشرة لصفر ونتائجها

كما ذكر في القسم السابق، تم تحويل الملفات الصوتية إلى صور طيفية (*Spectrograms*)؛ لانتشار هذه الطريقة في الدراسات السابقة التي قامت بتصنيف الأرقام المنطوقة [4]. هذه العملية تحول عملية تصنيف الأصوات (*Audio Classification*) إلى عملية تصنيف الصور (*Image Classification*). يظهر شكل رقم ٢ الصور الطيفية لبعض الأرقام العربية المنطوقة والتي اختيرت بشكل عشوائي لبعض محاولات الطلبة للأرقام من عشرة لخمسة.

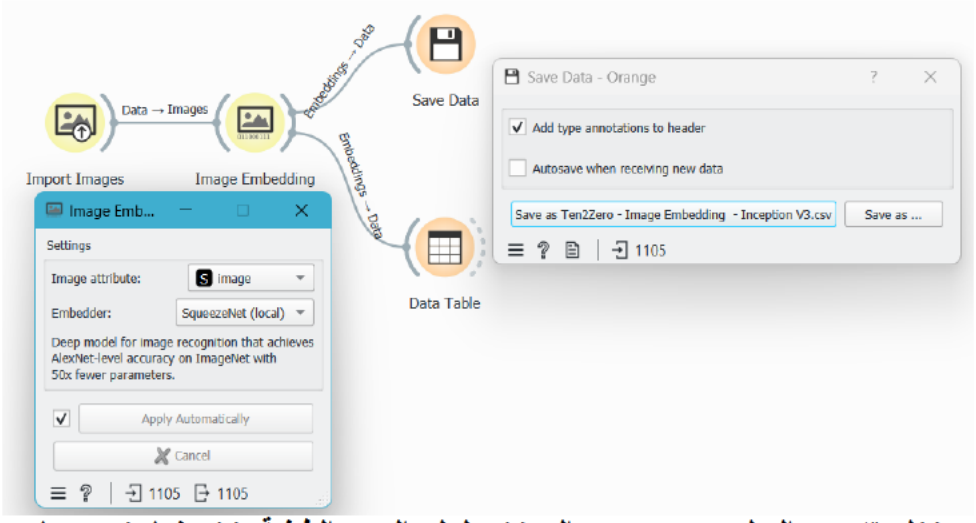


شكل (2) الصور الطيفية لبعض الأرقام العربية المنطوقة من عشرة لخمسة.

تم استخدام برنامج أورانج لتنقيب البيانات (الإصدار ٣,٣٦) في كل ما سيذكر في هذا القسم [20]. للقيام باستخراج خصائص أو سمات مميزة لكل رقم من الأرقام العربية من صفر لعشرة للقيام بعملية التصنيف، تم الاستفادة من الأداتين جلب الصور (*Import Image*) وتضمين الصور (*Image Embedding*). الأداة الأولى تمكننا من جلب أي مجموعة بيانات في شكل صور، سواء أكانت مخزنة في الحاسوب أو في الإنترنت، والأداة الثانية تمكننا من تضمين الصور. عملية تضمين الصور تحول كل صورة في مجموعة البيانات إلى متجهات عددية (خصائص أو سمات) تصف الصورة وأجزائها المختلفة باستخدام خوارزميات التعلم العميق وتمكننا من إنشاء نماذج تعلم الآلة المختلفة للتنبؤ بأصناف صور مجموعات البيانات المختلفة. جدول رقم ٢ يبين الخوارزميتين المستخدمة في هذه الدراسة وعدد المتجهات التي تم توليدها لكل صورة في مجموعة البيانات عشرة لصفر. يمكن تصدير الملف الذي ينتجه المضمن باستخدام الأداة "حفظ البيانات" (*Save Data*) بصيغة CSV [٢١]، [٢٢]. يبرز شكل رقم ٣ سير العمل المستخدم في برنامج أورانج لاستخراج الملفات من المضمن.

جدول (٢) المضمنان (*Embedders*) المستخدمان وخصائصهما.

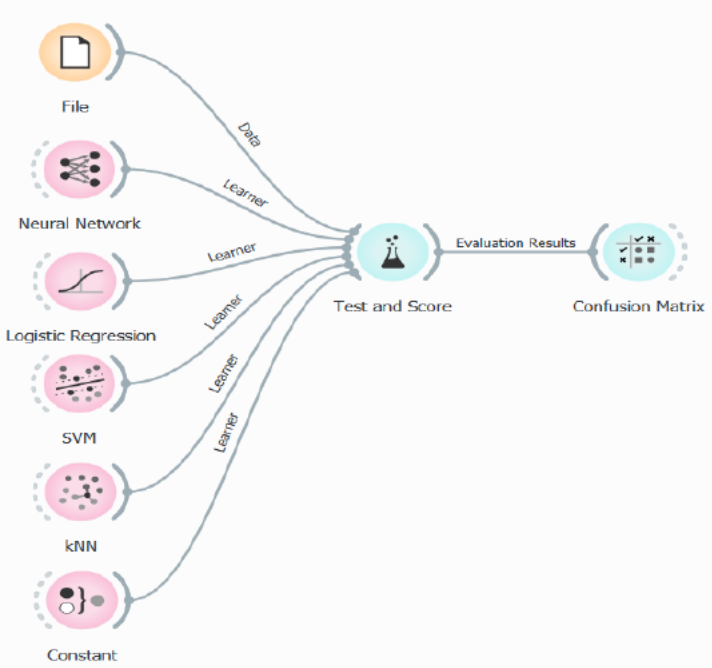
Embedder المضمن	Vectors المتجهات	Description وصفه
SqueezeNet	١٠٠٠ متجه (خاصية)	مضمن صغير الحجم يتم تشغيله وتدريبه في حاسوب المستخدم، ويستخدم للتعرف على الصور، تم تدريبه على مجموعة البيانات الشهيرة ImageNet.
Inception v3	٢٠٤٦ متجه (خاصية)	مضمن أكبر من سابقه حجمًا يتم تشغيله وتدريبه من الإنترنت، تم إنشاؤه من قبل شركة قوقل، يستخدم للتعرف على الصور، تم تدريبه على مجموعة البيانات الشهيرة ImageNet.



شكل (٣) سير العمل (*Workflow*) المستخدم لجلب الصور الطيفية وتضمينها وتصديرها.

يحوي برنامج أورانج أكثر من ضمن. تم الاقتصار على المضمنين SqueezeNet وInception v3؛ لأنهما مناسبين لغرض الدراسة ولأنهما يستخرجان آليًا عددًا معقولاً من المتجهات يناسب للتدريس في المدارس والجامعات. يوجد مضمنات أخرى أكثر دقة، ولكنها تستخرج عددًا كبيرًا جدًا من المتجهات يزيد من وقت التدريب والاختبار، ويمكن أن يتم بحثها كدراسة مستقبلية. ويلاحظ من جدول رقم ٢ أن المضمن الأول استخرج ١٠٠٠ خاصية لكل صورة من الصور الطيفية للأرقام العربية المنطوقة من عشرة لصفر، وقد تم استخدام الرموز من n0 إلى n999 للتعبير عن المتجهات الألف. بينما تسمى من n0 إلى n2046 للمضمن الآخر. تقوم كل المضمنات في أورانج باستخراج خمس خصائص وصفية (*Meta Features*) تبرز اسم ملف الصورة، ومسار المجلد المخزنة داخله، وحجمها، وعرضها، وطولها.

الهدف الآن هو إنشاء نموذج تعلم آلة قادر على التنبؤ بتصنيف الصوت المنطوق من خلال الصور الطيفية. للقيام بهذه المهمة تم استخدام التحقق المتقاطع على عشرين جزءًا (*20-Fold Cross-Validation*) لتدريب النموذج وتقييمه، يتم في هذه العملية تقسيم مجموعات البيانات عشوائيًا إلى عشرين مجموعة فرعية متساوية لتدريب النموذج وتقييمه (*Training and Testing*). تستخدم الأداة "التقييم والنتائج" (*Test and Score*) في برنامج أورانج للقيام بعمليتي التدريب والتقييم (الاختبار) للنموذج. تم استخدام العديد من خوارزميات التصنيف الشهيرة كما يظهر في شكل رقم ٤.



شكل (٤) سير العمل (*Workflow*) المستخدم لتدريب وتقييم نماذج تنبؤ تصنيف الصور الطيفية للأرقام العربية المنطوقة.

الخوارزميات أو النماذج المستخدمة هي: الشبكة العصبية (*Neural Network*) وهي نموذج حاسوبي يحاول محاكاة عقول الكائنات الحية وهو النموذج الشهير في scikit-learn والمسمى البيرسبترون متعدد الطبقات (*Multilayer Perceptron*)، ويحوي في أورانج افتراضيًا 100 طبقة مخفية (*Hidden Layers*)، الانحدار اللوجستي (*Logistic Regression*) ويتم فيها استخدام دالة تحول النموذج الخطي إلى تصنيف، آلة المتجهات الداعمة (*Support Vector Machine "SVM"*) يتم فيها وضع حدود قصوى للحصول على التصنيف، أقرب عدد من الجيران (*K Nearest Neighbors "kNN"*) يتم فيها مقارنة الصورة الطيفية المراد تصنيفها بناءً على أبرز خمس نقاط قريبة منها في المسافة ويرجح رأي الأغلبية، الثابت (*Constant*) ويتم فيها اختيار الصنف الأكثر شيوعًا في مجموعة البيانات لتصنيف كافة الأرقام العربية المنطوقة، تستخدم هذه الخوارزمية عادة لمقارنة أداء الخوارزميات الأخرى؛ نظرًا لكونها دائمًا ما تصنف كافة بيانات المجموعة بالاعتماد على صنف واحد ألا وهو "صنف الأغلبية". ونظرًا لأن مجموعات البيانات عشرة لصفر متوازنة؛ فيمكن استخدام أي من الأرقام من عشرة لصفر كصنف أغلبية. تم استخدام كافة الخوارزميات المذكورة أعلاه في الشكل رقم ٤ بإعداداتها الافتراضية في برنامج أورانج دون أي تعديل أو تغيير [10], [21].

لتقييم أداء النماذج الخمسة سالفة الذكر، تم القيام بتجربتين اثنتين، تجربة لكل مضمن من المضمنين المستخدمين كما ذكر في جدول رقم ٢. يظهر جدول رقم ٣ نتائج التجربة الأولى والخاصة بتقييم النماذج الخمسة باستخدام البيانات المستخرجة من المضمن SqueezeNet. تم استخدام معيار من أشهر معايير تقييم نماذج تعلم الآلة خاصة عند كثرة وتعدد الأصناف وهو المساحة تحت منحنى دقة الأداء (*Area Under ROC Curve "AUC"*) [23].

جدول (٣) نتائج تقييم نماذج تعلم الآلة المختلفة باستخدام المضمن (SqueezeNet).

النموذج	AUC
Neural Network	92.1%
Logistic Regression	90.7%
SVM	88.1%
kNN	84.5%
Constant	47.5%

يلاحظ أن أداء النموذج الشبكة العصبية (*Neural Network*) هو الأعلى من حيث الـ AUC بنسبة وصلت إلى ٩٢,١٪ متفوقًا على كافة النماذج الأخرى. كما يلاحظ أن هذه الزيادة كبير جدًا وبنسبة تزيد على ٩٣٪ إذا قورنت بالنموذج الثابت (*Constant*) الذي يصنف كافة الأرقام إلى صنف واحد فقط "صنف الأغلبية".

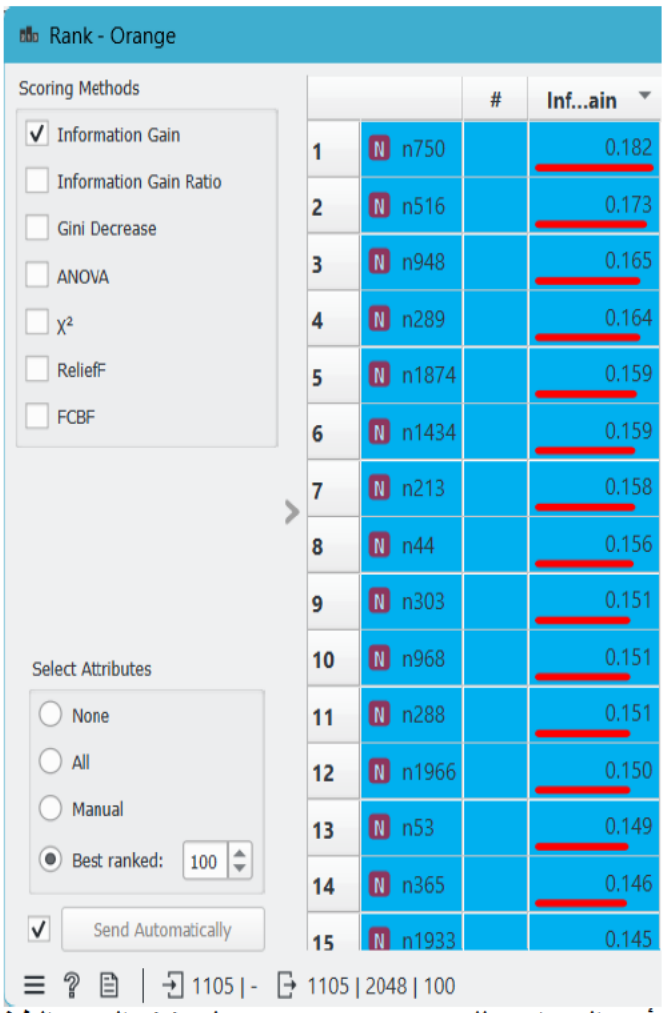
يظهر جدول رقم ٤ نتائج التجربة الثانية والخاصة بتقييم النماذج الخمسة باستخدام البيانات المستخرجة من المضمن Inception v3. تم استخدام نفس المعيار السابق للتقييم وهو المساحة تحت منحنى دقة الأداء (*Area Under ROC Curve "AUC"*).

جدول (٤) نتائج تقييم نماذج تعلم الآلة المختلفة باستخدام المضمن (Inception v3).

النموذج	AUC
Neural Network	٩٦.٩%
SVM	٩٦.١%
Logistic Regression	٩٦.٠%
kNN	٨٩.٧%
Constant	47.٧%

يلاحظ أن أداء النموذج الشبكة العصبية (*Neural Network*) ما زال مميزًا وهو الأعلى من حيث الـ AUC بنسبة وصلت إلى ٩٦,٩٪ متفوقًا على كافة النماذج الأخرى. كما يلاحظ أن هذه الزيادة كبير جدًا وبنسبة تزيد على ١٠٣٪ إذا قورنت بالنموذج الثابت (*Constant*) الذي يصنف كافة الأرقام إلى صنف واحد فقط "صنف الأغلبية". يلاحظ أيضًا زيادة واضحة في أداء كافة النماذج الأربعة (دون النموذج الثابت) المدربة والمقيمة على البيانات المجدولة المستخرجة من المضمن Inception v3 إذا ما قورنت بالنتائج المستخرجة من المضمن SqueezeNet. قد تعود هذه الزيادة المتوقعة لوفرة المتجهات المستخرجة من المضمن Inception v3 والبالغ عددها ٢٠٤٦ والتي تصف الصور الطيفية للأصوات بشكل أدق من متجهات المضمن SqueezeNet البالغ عددها ١٠٠٠ متجه. حتى يتم التأكد من التفوق الملاحظ للمضمن Inception v3 على SqueezeNet، تم إجراء تجربة أخرى باستخدام الأداة ترتيب (*Rank*) والتي تهتم بانتقاء أفضل الخصائص التي تحقق أفضل وأسرع نموذج تخمين لتصنيف الأرقام العربية المنطوقة باستخدام الصور الطيفية. تم اختيار أفضل ١٠٠ خاصية لكلا المضمنين لتدريب النماذج الخمسة المذكورة أعلاه باستخدام خوارزمية اكتساب المعرفة (*Information Gain*). وقد تم استخدام التحقق المتقاطع على عشرين جزءًا (*20-Fold Cross-Validation*) كما تم سابقًا. يبين شكل رقم ٥ نبذة عن أبرز الخصائص المئة المختارة للمضمن Inception v3، ويكتفى بهذه الصورة دون صورة المضمن الآخر؛ للتشابه الكبير بين الصورتين.

جدول رقم ٥ ورقم ٦ يبينان نتيجة التصنيف بعد استخدام الأداة ترتيب (*Rank*) واختيار أفضل ١٠٠ متجه لتصنيف الأرقام العربية المنطوقة. يلاحظ كما لوحظ سابقًا بأن نتائج التصنيف باستخدام المضمن Inception v3 أفضل من حيث الـ AUC لكل الخوارزميات المستخدمة دون استثناء؛ مما يؤكد تفوق المضمن Inception v3 على المضمن الآخر SqueezeNet. الجدير بالذكر أن زمن تدريب وتقييم النماذج بعد استخدام الأداة ترتيب (*Rank*) أسرع بكثير من تدريب نماذج باستخدام متجهات تربو على الألف مع وجود نقص في نتائج الـ AUC يمكن الاستغناء عنه في تصنيف الأرقام العربية المنطوقة مقابل الزمن. ويمكن الخلوص إلى أن الزيادة واضحة من حيث الـ AUC في أداء كافة النماذج المدربة والمقيمة على البيانات المجدولة المستخرجة من المضمن Inception v3 إذا ما قورنت بالنتائج المستخرجة من المضمن SqueezeNet في كلا الحالتين: عند اختلاف أعداد المتجهات المستخدمة في تدريب واختبار النماذج، وعند الالتزام بعدد متجهات واحد لكلا المضمنين.



شكل (٥) جزء من أبرز الخصائص للمضمن Inception v3 لتصنيف الصور الطيفية للأرقام العربية.

جدول (٥) نتائج تقييم نماذج تعلم الآلة المختلفة باستخدام أفضل ١٠٠ خاصية للمضمن (SqueezeNet).

AUC	النموذج
90.4%	Neural Network
88.0%	Logistic Regression
85.7%	SVM
79.7%	kNN
47.5%	Constant

جدول (٦) نتائج تقييم نماذج تعلم الآلة المختلفة باستخدام أفضل ١٠٠ خاصية للمضمن (Inception v3).

AUC	النموذج
94.0%	Neural Network
93.9%	SVM
90.0%	Logistic Regression
87.5%	kNN
47.٧%	Constant

٥. الخلاصة والدراسات المستقبلية

قدمت هذه الدراسة مجموعة بيانات جديدة "عشرة لصفر" وهي مجموعة بيانات صوتية متوازنة الأصناف للأرقام العربية المنطوقة باللغة العربية واللهجة المكية لتدريس تعلم الآلة. تحوي مجموعة البيانات بيانات مجدولة (*Tabular Data*) وصور طيفية (*Spectrograms*) وملفات صوتية (*wav files*)، تتيح للمهتمين من كافة التخصصات القيام بدراسات مستقبلية في مجالات عدة باستخدام مجموعة البيانات هذه، بالإضافة إلى إعادة إجراء تجارب هذه الدراسة وإنتاج نتائجها للأغراض التعليمية المختلفة في كافة المراحل الدراسية. تعد هذه الدراسة أول دراسة علمية باللغة العربية -بحسب علمي- لتصنيف الأرقام العربية المنطوقة باستخدام تعلم الآلة والتعلم العميق. خلصت الدراسة إلى أن أفضل نموذج لتصنيف الصور الطيفية المستقاة من الملفات الصوتية في مجموعة البيانات عشرة لصفر هو الشبكة العصبية (*Neural Network*) المبني على البيانات المستخرجة من مضمن الصور Inception v3، كما خلصت إلى تميز كافة النماذج الأخرى المدربة والمقيمة على البيانات المجدولة المستخرجة من المضمن Inception v3 إذا ما قورنت بالنتائج المستخرجة من المضمن SqueezeNet سواء عند اختلاف أعداد المتجهات أو تساويهما. تفتح هذه الدراسة الكثير من الآفاق للدارسين والباحثين العرب لتطوير تطبيقات جوالات أو نماذج تعلم آلات صغيرة (*TinyML*) خاصة مع انتشار الأنظمة المبنية على المساعدات الصوتية مثل سيري وأليكسا، وانتشار العديد من المنصات التي لا تتطلب برمجة أو تتطلب برمجة قليلة (*No Code or Low Code Tools*) مثل: Liner.ai [24] و PictoBlocks [25] -على سبيل المثال لا الحصر-.

شكر وتقدير

خالص الشكر والتقدير للطلبة المهندسين الذين ساهموا في جمع مجموعة البيانات عشرة لصفر.

المراجع

[1] "الهيئة السعودية للبيانات والذكاء الاصطناعي, *الهيئة السعودية للبيانات والذكاء الاصطناعي*, ٢٠٢٣." https://sdaia.gov.sa/ar/default.aspx (accessed Aug. 08, 2023).

[2] "ملامح تطوير المناهج السعودية." Accessed: Jul. 26, 2023. [Online]. Available: https://moe.gov.sa/ar/education/generaleducation/StudyPlans/Documents/Features-of-the-development-of-the-Saudi-curriculum.pdf

[3] A. M. A. Alqadasi, R. Abdulghafor, M. S. Sunar, and Md. S. B. H. J. Salam, "Modern Standard Arabic Speech Corpora: A Systematic Review," *IEEE Access*, vol. 11, pp. 55771–55796, 2023, doi: 10.1109/ACCESS.2023.3282259.

[4] A. Dhouib, A. Othman, O. El Ghoul, M. K. Khribi, and A. Al Sinani, "Arabic Automatic Speech Recognition: A Systematic Literature Review," *Appl. Sci.*, vol. 12, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/app12178898.

[5] A. Hassan, S. Aftab, R. Khan, and H. Asim, "The Analysis on the usage of the Video Conferencing Rooms using Classification," *KIET J. Comput. Inf. Sci.*, vol. 2, no. 2, pp. 09–09, Jul. 2019, Accessed: Sep. 19, 2023. [Online]. Available: https://kjcis.kiet.edu.pk/index.php/kjcis/article/view/28

[6] J. Han, J. Pei, and H. Tong, *Data Mining Concepts and Techniques- 4th Edition*. Morgan Kaufmann, 2022. Accessed: Sep. 19, 2023. [Online]. Available: https://shop.elsevier.com/books/data-mining/han/978-0-12-811760-6

[7] "Chapter 1. Introduction by Jiawei Han, Computer Science, Univ. Illinois at Urbana-Champaign, 2017." Accessed: Sep. 19, 2023. [Online]. Available: http://hanj.cs.illinois.edu/cs412/bk3_slides/01Intro.pdf

[8] "Hugging Face – The AI community building the future.," Jun. 12, 2023. https://huggingface.co/datasets (accessed Sep. 19, 2023).

[9] "Papers with Code - Machine Learning Datasets." https://paperswithcode.com/datasets?mod=audio&lang=arabic (accessed Sep. 19, 2023).

[10] "معجم البيانات والذكاء الاصطناعي." Accessed: Aug. 21, 2023. [Online]. Available: https://sdaia.gov.sa/ar/MediaCenter/KnowledgeCenter/ResearchLibrary/SDAIAPublications15.pdf

[11] N. H. Mouldi Bedda, "Spoken Arabic Digit." UCI Machine Learning Repository, 2008. doi: 10.24432/C52C9Q.

[12] A. Ghandoura, F. Hjabo, and O. Al Dakkak, "Building and benchmarking an Arabic Speech Commands dataset for small-footprint keyword spotting," *Eng. Appl. Artif. Intell.*, vol. 102, p. 104267, Jun. 2021, doi: 10.1016/j.engappai.2021.104267.

[13] Y. Alotaibi, "A Simple Time Alignment Algorithm for Spoken Arabic Digit Recognition," *J. King Abdulaziz Univ.-Eng. Sci.*, vol. 20, no. 1, pp. 29–43, 2009, doi: 10.4197/Eng.20-1.2.
[14] Y. A. Alotaibi, "Investigating spoken Arabic digits in speech recognition setting," *Inf. Sci.*, vol. 173, no. 1, pp. 115–139, Jun. 2005, doi: 10.1016/j.ins.2004.07.008.
[15] A. Ganoun and I. Almerhag, "Performance Analysis of Spoken Arabic Digits Recognition Techniques," vol. 10, no. 2, 2012.
[16] A. S. Mahfoudh BA WAZIR and J. Huang CHUAH, "Spoken Arabic Digits Recognition Using Deep Learning," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, Jun. 2019, pp. 339–344. doi: 10.1109/I2CACIS.2019.8825004.
[17] H. Satori, M. Harti, and N. Chenfour, "Introduction to Arabic Speech Recognition Using CMUSphinx System." arXiv, Apr. 16, 2007. doi: 10.48550/arXiv.0704.2083.
[18] M. Huzaifah, "muhdhuz/audio2spec." Mar. 11, 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://github.com/muhdhuz/audio2spec
[19] B. McFee *et al.*, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, 2015, pp. 18–25.
[20] J. Demšar *et al.*, "Orange: Data Mining Toolbox in Python," *J. Mach. Learn. Res.*, vol. 14, pp. 2349–2353, 2013.
[21] B. L. Ljubljana University of, "Widget catalog." https://orangedatamining.com/widget-catalog/ (accessed Jul. 22, 2023).
[22] د. ع. طعيمة, "كتاب تنقيب البيانات وتعلم الآلة: بدون برمجة - التعلم العميق بالعربي || DL الدكتور علاء طعيمة بالعربي," Mar. 30, 2023., https://dlarabic.com/كتاب-تنقيب-البيانات-وتعلم-الآلة-بدون-ب/ (accessed Sep. 19, 2023).
[23] G. F. Bati and V. K. Singh, "NADAL: A Neighbor-Aware Deep Learning Approach for Inferring Interpersonal Trust Using Smartphone Data," *Computers*, vol. 10, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/computers10010003.
[24] "Liner.ai - Machine Learning without Code." https://liner.ai/ (accessed Sep. 19, 2023).
[25] "Examples Archive," *STEMpedia Education*. https://ai.thestempedia.com/example/ (accessed Sep. 19, 2023).