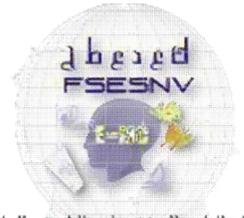




République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la
recherche scientifique



كلية العلوم الدقيقة وعلوم الطبيعة والحياة
FACULTÉ DES SCIENCES EXACTES
ET DES SCIENCES DE LA NATURE ET DE LA VIE

Université Larbi Tébessi - Tébessa
Faculté des Sciences Exactes et des Sciences de la
Nature et de la vie
Département : Mathématique et informatique

Mémoire de fin d'étude
Pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et informatique

Filière : Informatique

Option : Système et multimédia

Thème

إستعمال مصنف التعلم العميق لتعرف الآلي علي قراء القرآن الكريم
Deep Learning pour la classification des récitateurs du
Saint Coran

Présenté Par :

Tahar Ramzi

Devant le jury :

Mr M.Laouar	PR	Université de Larbi Tébessi	Président
Mr I.Bendib	MCB	Université de Larbi Tébessi	Examineur
Mr M.Gasmi	MCB	Université de Larbi Tébessi	Encadreur
Mr A.Gahmous	MAA	Université de Larbi Tébessi	Co-Encadreur

Année universitaire :2020/2021

ملخص

أخذ التعلم العميق (Deep Learning) في السنوات القليلة الماضية مسألة التعرف الآلي على الصوت إلى مستويات جديدة من الدقة . حيث حاز على أعلى نسب التعرف ، سواء على الكلمات أو على المتكلم .
نقدم في هذا البحث نظام للتعرف الآلي على قراء القرآن الكريم اعتمادا على الشبكات العصبونية الالتفافية Convolutional neural network (CNN) و قد تم فيه استخدام خوارزمية استخلاص السمات ((Mel Frequency Cepstral Coefficients (MFCC) حيث يجري تحويل الإشارة الصوتية إلى مصفوفة ثنائية الأبعاد ضمن الفضاء الترددي ومن ثم يتم إدخال هذه المصفوفة كدخل للشبكة العصبونية الالتفافية قمنا باختبار المصنف المقترح على قاعدة بيانات تحتوي على عشرة قراء للقرآن الكريم . حيث بلغت نسبة التعرف في المصنف المقترح 89 % .

Abstract

Deep learning in the past few years has taken automatic voice recognition to new levels of accuracy. Where he has the highest rates of recognition, In this research, we present a system , whether on the words or on the speaker for automatic identification of the reciters of the Holy Qur'an based on Convolutional Neural Network (CNN), in which the Mel Frequency Cepstral Coefficients (MFCC) algorithm was used, where the audio signal is converted into a two-dimensional matrix within space. Then, this matrix is entered as input to the convolutional neural network. We tested the proposed classifier on a database containing ten reciters of the Holy Qur'an, where the identification rate in the proposed classifier was 89%.

Résumé

Au cours des dernières années, l'apprentissage en profondeur a porté la reconnaissance vocale automatique à de nouveaux niveaux de précision. Là où il a les taux de reconnaissance les plus élevés, que ce soit sur les mots ou Dans cette recherche, nous présentons un système , sur le locuteur d'identification automatique des récitants du Saint Coran basé sur le réseau neuronal convolutif (CNN), dans lequel l'algorithme des coefficients cepstraux de fréquence de Mel (MFCC) a été utilisé, où le signal audio est converti en une matrice bidimensionnelle dans l'espace. Ensuite, cette matrice est entrée en entrée du réseau de neurones convolutifs. Nous avons testé le classificateur proposé sur une base de données contenant dix récitants du Saint Coran, où le taux d'identification dans le classificateur proposé était de 89%.

الإهداء

إلى تلك الروح التي أنارت لي العالم بأسره ، تلك الروح التي كلما ذكرت رفعت رأسي عاليا ، تلك الروح التي علمتني الصبر والقوة ، تلك الروح التي أعطتني مفاتيح الحياة ، تلك الروح التي ما بخلت وجادت ، تلك الروح التي أدفع عمري ثمنا لرجوعها ، تلك الروح التي منحتني كل الأمان فكانت دولتي وجيشي وكل حراسي ، شكرا يا سندي...

والدي المرحوم طاهر عبد القادر

إلى تلك الحنونة التي وقفت معي في كل لحظاتي ، في ضعفي قبل قوتي ، في مرضي قبل صحتي ، تلك الحنونة التي تخشى علي أكثر من نفسي ، تلك الحنونة التي جعلت الجنة تحت قدميها، تلك الحنونة الصابرة، تلك الحنونة التي ربّنتي، تلك الحنونة التي لا تجعلني أشعر بمرارة الحياة، شكرا.

أمي الغالية

إلى جبالي الرواسي.. وراياتي الشامخة ..

إلى من كانوا ومازالوا درعا في ظهري.. وصخرة أتكى عليها دون مخافة انزلاقها.

إلى أخوتي: هشام -أسامة - عائشة

إلى أصدقائي الذين كانوا ومازالوا أخوة لي وسند، أصحاب الضحكة والفرح والحزن والألم والحلو والمر والعقل والجنون..

حميد-جلال-مبروك-إسحاق-خالد-علي

إلى كل أهلي وأصدقائي وزملائي في العمل والدراسة، وإلى كل من وضع منارة في طريقي، وإلى كل معلّم علمني حرفا..

كلمة شكر

أقدم بجزيل الشكر إلى عموم كوادر قسم رياضيات والإعلام الآلي جامعة الشيخ العربي التبسي أساتذة ومشرفين وإداريين. وأقدم بالشكر إلى الأساتذة المؤطرين الأستاذ قحموص عبد اللطيف والأستاذ قاسمي محمد على جهودهم وملاحظاتهم القيّمة. التي كانت كشمعة تضيء لي الطريق. كما أقدم بشكر إلى كل كوادر مختبر الأبحاث قسم رياضيات والإعلام الآلي وأخص بالذكر الأستاذ بن جنة الذي كان عوناً لي في كل مراحل تربصتي.

قائمة المحتويات

2	ملخص
4	الإهداء
5	كلمة شكر
8	قائمة الأشكال
10	قائمة الجداول
11	الإختصارات
12	مقدمة عامة
15	الفصل الأول سياق الدراسة و أحدث ما توصلت إليه التقنية
15	1. مقدمة لمجال التعرف الآلي على المتكلم (Automatic Speaker Recognition)
15	2. نبذة تاريخية على التعرف على المتكلم.
15	3. التعرف الآلي على المتكلم
17	1.3 التعرف الآلي على المتكلم (ASI)
17	2.3 التحقق التلقائي من المتكلم (ASV)
18	4. مجالات التطبيق التعرف الآلي على المتكلم
20	5. الهدف من التعرف الآلي على المتكلم
20	6. صعوبات التعرف الآلي على المتكلم
21	7. التعرف الآلي على قراء القراءان الكريم
22	8. منشورات بحثية في التعرف الآلي على قراء القراءان الكريم
24	9. الخاتمة
26	الفصل الثاني: المقاربة المقترحة لمعالجة البيانات الصوتية واستخلاص السمات (دراسة منهجية وتطبيقية)
26	1. مقدمة
26	2. قاعدة البيانات (Dataset)
27	3. تنقية البيانات (Data Cleaning)
29	4. معالجة البيانات (Data Pre-processing)
29	1.4 توحيد صيغ الملفات الصوتية (Standardization of audio file formats)
30	1.1.4 التحويل من صيغ mp3 إلى صيغ wav:
30	2.1.4 معدل اخذ العينات (Sampling rate)
31	3.1.4 توحيد قنوات الإخراج

32	دمج الملفات الصوتية (الصور).....
34	3.4 نزع الصمت من القراءان الكريم.....
34	4.4 اختيار العينات من الملفات الصوتية (Sample Random Selection).....
35	5. خوارزمية إستخلاص السمات (Feature Extraction).....
41	الفصل الثالث : التعلم العميق، إنشاء و عمل تجريبي.....
41	1. مقدمة.....
41	2. التعلم العميق (Deep Learning).....
42	3. الشبكة العصبونية التلافيفية (Convolutional Neural Network).....
42	1.3 تدريب الشبكة العصبونية التلافيفية (CNN).....
43	2.3 المعمارية المقترحة في الشبكة العصبونية التلافيفية (CNN) لتعرف الآلي على قراء القراءان الكريم.....
43	3.3 تقييم أداء المعمارية المقترحة.....
44	4. البرامج ومكتبات والبيئات المستعملة في هذا البحث.....
45	5. التجارب والنتائج العملية.....
59	1.5 مناقشة النتائج العملية.....
60	6. الآفاق المستقبلية.....
60	7. خاتمة.....
61	الخاتمة عامة.....
62	المراجع.....

قائمة الأشكال

- 17 الشكل 1 مبدأ تحديد المتكلم
- 18 الشكل 2 مبدأ التحقق من المتكلم
- 26 الشكل 3 مراحل المقارنة المقترحة
- 29 الشكل 4 صورة توضح طول التكرار في المقطع الصوتي
- 31 الشكل 5 تمثيل أخذ عينات من الإشارة
- 32 الشكل 6 الفرق بين **mono + stéréo**
- 33 الشكل 7 مرحلة الدمج العشوائي لملفات القراء
- 34 الشكل 8 عدد مقاطع الصمت بلون الأحمر
- 36 الشكل 9 مخطط خوارزمية الـ **MFCC**
- 37 الشكل 10 حزمة مرشحات **Mel**
- 38 الشكل 11 المصفوفة النهائية ثلاثية الأبعاد لكل العينات (1280,13,216)
- 38 الشكل 12 نتائج هذه المرحلة
- 41 الشكل 13 معمارية نظام التعرف الآلي على قراء القراءان الكريم
- 41 الشكل 14 رسم تخطيطي لعصب بيولوجي (2)
- 42 الشكل 15 رسم تخطيطي لطبقات (2) **CNN**
- 48 الشكل 16 نسبة التعرف مع تغيير عدد سمات باستعمال تقسيم **split**
- 48 الشكل 17 نسبة التعرف مع تغيير عدد السمات باستعمال **kfold**
- 48 الشكل 18 نسبة الفقدان مع تغيير عدد السمات باستعمال **split**
- 49 الشكل 19 نسبة الفقدان مع تغيير عدد السمات باستعمال تقسيم **kfold**
- 49 الشكل 20 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال **mfcc13**
- 51 الشكل 21 نسبة التعرف مع تغير معدل اخذ العينات باستعمال **split**
- 51 الشكل 22 نسبة التعرف مع تغير معدل اخذ العينات باستعمال **kfold**
- 52 الشكل 23 نسبة الفقدان مع تغير معدل اخذ العينات باستعمال تقسيم **split**
- 52 الشكل 24 نسبة الفقدان مع تغير معدل اخذ العينات باستعمال تقسيم **kfold**
- 52 الشكل 25 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال **22050Hz**
- 54 الشكل 26 نسبة التعرف مع تغير عدد العينات باستعمال تقسيم **split**
- 55 الشكل 27 نسبة التعرف مع تغير عدد العينات باستعمال تقسيم **kfold**
- 55 الشكل 28 نسبة الفقدان مع تغير عدد العينات باستعمال تقسيم **split**
- 55 الشكل 29 نسبة الفقدان مع تغير عدد العينات باستعمال تقسيم **kfold**

- الشكل 30 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال 128 عينة..... 56
- الشكل 31 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم **kfold** 57
- الشكل 32 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم **split** 58
- الشكل 33 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال 5 ثواني..... 58

قائمة الجداول

27	الجدول 1 قاعدة البيانات المتحصل عليها
33	الجدول 2 الحجم الساعي للملف النهائي المتحصل عليه لكل مقرر
43	الجدول 3 الطبقات المقترحة ودوال التنفيع الخاصة بالمعمارية
45	الجدول 4 نتائج اختيار 13 mfcc باستعمال kfold
45	الجدول 5 نتائج اختيار mfcc13 باستعمال split
46	الجدول 6 نتائج اختيار 14 mfcc باستعمال kfold
46	الجدول 7 نتائج اختيار 14 mfcc باستعمال split
46	الجدول 8 نتائج اختيار 16 mfcc باستعمال kfold
46	الجدول 9 نتائج اختيار 16 mfcc باستعمال split
47	الجدول 10 نتائج اختيار 18 mfcc باستعمال kfold
47	الجدول 11 نتائج اختيار 18 mfcc باستعمال split
47	الجدول 12 نتائج اختيار 20 mfcc باستعمال kfold
47	الجدول 13 نتائج اختيار 20 mfcc باستعمال split
50	الجدول 14 نتائج اختيار 22050 Hz باستعمال kfold
50	الجدول 15 نتائج اختيار 22050 Hz باستعمال split
50	الجدول 16 نتائج اختيار 16000 Hz باستعمال kfold
50	الجدول 17 نتائج اختيار 16000 Hz باستعمال split
51	الجدول 18 نتائج اختيار 11000 Hz باستعمال kfold
51	الجدول 19 نتائج اختيار 11000 Hz باستعمال kfold
53	الجدول 20 نتائج اختيار 128 عينة باستعمال kfold
53	الجدول 21 نتائج اختيار 128 عينة باستعمال split
53	الجدول 22 نتائج اختيار 256 عينة باستعمال kfold
53	الجدول 23 نتائج اختيار 256 عينة باستعمال split
54	الجدول 24 نتائج اختيار 512 عينة باستعمال kfold
54	الجدول 25 نتائج اختيار 512 عينة باستعمال split
56	الجدول 26 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم kfold
57	الجدول 27 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم split

CNN	Convolutional Neural Network	الشبكة العصبونية التلافيفية
ASR	Automatic Speaker Recognition	التعرف الآلي علي المتكلم
NN	Neural Network	الشبكة العصبونية
MFCC	Mel Frequency Cepstral Coefficients	تحويل التردد إلى تردد ميل
ANN	Artificial neural network	شبكة عصبونية اصطناعية
DL	Deep Learning	التعلم العميق
	Feature	السمات
ASV	Automatic speaker vérification	التحقق التلقائي من المتكلم
	Cepstrum	طيف الكلام
	Access control	التحكم في الوصول
	Transaction authentication	مصادقة المعاملات
	Speech data management	إدارة البيانات الصوتية
	Data Pre-processing	المعالجة المسبقة للبيانات
	Sampling rate	معدل اخذ العينات
	Feature Extraction	خوارزمية إستخلاص السمات
	Window Hamming	نافذة هامينغ
FFT	Transform Fourier Fast	تحويل فورييه السريع
DFT	Transform Fourier Discrete	تحويل فورييه المتقطع
	Mel- Frequency Warping	تزييف التردد الخاصة بميل
	Labels Vector	عمود الوسوم
	Convolutional Layer	الطبقة التلافيفية
	confusion matrix	مصفوفة الالتباس
	Activating function	طبقة التفعيل
	Pooling Layer	طبقة التجميع
	Flatten	طبقة التسوية
kfold	Cross-validation	التصديق متقاطع
	Fully-connected Layer	طبقة الاتصال الكامل

مقدمة عامة

النمو الهائل لتكنولوجيا المعلومات وما يوفره من سهولة في مجالات الحياة المتعددة جعل الإنسان يسعى دائما لتطوير ودمج هذه التكنولوجيات في كل محاور عمله ومن بين هذه المحاور والتي ساهمت التكنولوجيا في تطويرها هي نظام التعرف الآلي على الكلام والمتكلم هذا النظام الذي أصبح رائجا وذا طلب واسع في يوميات كل فرد لما يوفره من أمان وسهولة في التعاملات اليومية .

سعي الإنسان للتطور الدائم جعله يطمح لإيجاد وسائل عملية تساعد في زيادة أمان الأنظمة و موثوقية المبادلات وابتكار بدائل لوسائل الأمان الشخصية و التخلص من وحدات الإدخال و لوحق الأجهزة الذكية مما دفع العلماء لبحث عن طرق و وسائل جديدة من شأنها تسهيل هذا الأمر , و من بين الحلول التي اقترحها الباحثين و عملوا على تطويرها هي نظام التعرف الآلي على المتكلم .

في العديد من تطبيقات الكلام ، من الصعب تحديد ما إذا كان يمكن محاكاة الأداء البشري بواسطة الآلات ولكن في التعرف على المتكلمين تشير الأدلة التجريبية الحالية في المؤلفات والمنشورات العلمية السابقة إلى أن الآلات لا سيما في مجال التعرف الآلي على المتكلم تحقق أداء أفضل من المستمعين من البشر .

إن التعرف الآلي على المتكلم هو مجال تم العمل عليه و توفير حلول تكنولوجية كثيرة لتطويره من طرف الباحثين و لكن هذه الأبحاث تبقى مقتصرة على لغات و لهجات محددة و لا تتوفر أبحاث كثيرة في موضوع التعرف الآلي على المتكلم العربي و لهجات اللسان العربي .

نركز في هذا البحث على التعرف الآلي على قراء القرآن الكريم وهو عبارة عن جزء صعب من مجال التعرف الآلي على المتكلم. تكمن صعوبة التعرف الآلي على قراء القرآن الكريم في اختلاف صوت القارئ عن صوت المتكلم العادي بسبب التنوع الكبير في أنواع وطرق قراءة القرآن الكريم من تلاوة وتجويد ومقامات .

تم في هذا البحث العمل على 10 قراء من المغرب العربي وتم اقتراح مقارنة تتكون من جزئين الجزء الأول معالجة البيانات الصوتية بنزع الصمت منها واختيار عينات عشوائية لكل قارئ ثم العمل على استخلاص السيمات الصوتية باستخدام خوارزمية ((Mel Frequency Cepstral Coefficients (MFCC)) التي تم ذكرها في الأوراق البحثية ذات صلة ببحثنا. أم في الجزء الثاني من مقاربتنا المقترحة تم العمل على الخلايا العصبية في التعلم العميق (Deep Learning) وتم اقتراح نموذج في الشبكة العصبونية الالتفافية (CNN) ثم القيام بتجارب عملية لتحسين دقة ونسبة تعرف الخاصة بالشبكة .

الغاية من هذا البحث هو إنشاء نظام متكامل للتعرف الآلي على قراء القرآن الكريم دون الحاجة إلى مستمع بشري.

ضمن هذه البحث سنتناول مايلي:

في الفصل الأول نستعرض الأساس النظري الذي يركز عليه العمل, من تعرف علي الكلام والمتكلم والهدف من هذا البحث والأعمال المنجزة في مجال التعرف الآلي على القراءان الكريم .

في الفصل الثاني نستعرض الجزء الأول من المقاربة المقترحة وهي عبارة عن معالجة الملفات الصوتية بدمجها ونزع الصمت واختيار عينات عشوائية منها وتطبيق خوارزمية استخلاص السمات.

في الفصل الثالث نستعرض الجزء الثاني من المقاربة المقترحة و المتمثلة في المعمارية المقترحة والتجارب العملية لتحسين نسبة تعرف الشبكة والأفاق المستقبلية لهذا المجال .

الفصل الأول

سياق الدراسة وأحدث ما توصلت إليه التقنية



الفصل الأول سياق الدراسة و أحدث ما توصلت إليه التقنية

1. مقدمة لمجال التعرف الآلي على المتكلم (Automatic Speaker Recognition).

في العقد الماضي تأثر تصميم وإنشاء أنظمة الاتصال بين الإنسان والآلة و بين الإنسان و الإنسان بالنمو الهائل لتكنولوجيا المعلومات حيث أصبحت ذات أهمية متزايدة في مجالات عديدة من جوانب الحياة . من بين مجالات معالجة الصوت والكلام تم تكريس قدر كبير من الاهتمام لتطوير أنظمة تحدد وتعرف علي الأشخاص من أصواتهم أين أصبح مشروع تصميم وإنشاء أنظمة التعرف الآلي علي المتكلم مشروعاً رائجاً يستقطب اهتمام ومتابعة الباحثين ومطوري الأنظمة علي مدي عقود عديدة ،نتطرق في هذا الفصل إلى مجال التعرف على المتكلم بشكل عام والتعرف الآلي علي قارئ القرءان بشكل خاص .

2. نبذة تاريخية على التعرف على المتكلم.

كان موضوع التعرف على المتكلم قيد التطوير منذ منتصف القرن العشرين. كانت أقدم الأوراق البحثية المعروفة حول هذا الموضوع ، والتي نُشرت في خمسينيات القرن الماضي ، تبحث عن اكتشاف سمات (Feature) شخصية للمتكلمين ، من خلال تحليل كلامهم ، مع بعض الأسس الإحصائية. مع ظهور الشبكات المعلوماتية للاتصال ، أشار بولاك (Bolac) وباحثون آخرون إلى الحاجة إلى تحديد المتكلم. على الرغم من أنهم وظفوا مستمعين بشريين للقيام بتحديد الأفراد ودرسوا أهمية مدة الكلام والجوانب الأخرى التي تساعد في التعرف على المتكلم [1].

في معظم البحوث العلمية المبكرة ، تم إجراء تحليل يعتمد على النص ، من أجل تبسيط مهمة تحديد المتكلم. في عام 1959 ، بعد وقت قصير من تحليل بولاك (Bolac) ، قام شيرم (Shearme) بمقارنة صيغ الكلام من أجل تسهيل عملية تحديد المتكلم [1] ومع ذلك ، لا يزال الخبير البشري يقوم بالتحليل. تم استخدام هذا التقليد المتجسد في التعرف الآلي على المتكلم ، أي باستخدام الخبرة البشرية. حتى الآن من أجل التعامل مع تحديد المتكلم. تم تحسين هذه الفئة من الأساليب واستخدامها في مجموعة متنوعة من التحليلات الجنائية والطب الشرعي من قبل خبراء قانونيين. ومازال هذا البحث مستمر ليومنا هذا [1].

3. التعرف الآلي على المتكلم

تلقي تطبيقات تحديد الأشخاص عن طريق إشاراتهم الصوتية اهتماماً كبيراً لدي الباحثين في مجال التعرف البيومتري تسمى القدرة على التعرف على شخص ما من خلال صوته باسم التعرف على المتكلم ، وقد حظي هذا المجال مؤخراً باهتمام بين الباحثين في مجال دراسة النطق.

نظرًا لأن أداء الإنسان في التمييز بين المتكلمين معروف منذ فترة طويلة ، فإن الهدف الأكثر أهمية في هذا المجال هو معرفة ما إذا كان يمكن برمجة أجهزة الكمبيوتر للتعرف على المتكلمين من أصواتهم كما يفعل البشر في العديد من تطبيقات الكلام ، من الصعب تحديد ما إذا كان يمكن محاكاة الأداء البشري بواسطة الآلات ولكن في التعرف على المتكلمين تشير الأدلة التجريبية الحالية في المؤلفات والمنشورات العلمية السابقة إلى أن الآلات ، لا سيما في مهام التحقق من المتكلم (speaker verification) ذات النطق القصير وعدد كبير من المتكلمين ، فالآلات لديها القدرة على أداء أفضل من المستمعين من البشر في وقت التدريب على التعرف على متكلمين غير مألوفين. [2] .

التعرف و/أو التحقق من المتكلم هو مطلب أساسي للتحكم في الوصول إلى الموارد المحمية. يمكن المطالبة بالهوية الشخصية بواسطة مفتاح أو كلمة مرور أو شارة ، ويمكن بسهولة سرقتها أو فقدها أو تزويرها أو تمويهها. ومع ذلك ، هناك بعض السمات الفريدة للأفراد والتي لا يمكن لأي شخص آخر تقليدها. مثل بصمات الأصابع وهندسة اليد ونمط شبكية العين والسمات الشخصية كالكتابة اليدوية والبصمة الصوتية [2].

على الرغم من أن بصمات الأصابع أو أنماط شبكية العين عادة ما تكون طرقًا أكثر موثوقية للتحقق من هوية الشخص. إلا أن البصمة الصوتية تعتبر من أهم طرق التعرف على هوية الشخص [2].

التحقق من الهوية على أساس صوت الشخص له مزايا خاصة للنشر العملي مثل سهولة جمع البيانات عبر الهاتف. على وجه الخصوص ، يمكن أن يكون تحديد المتكلم الموثوق عن طريق الصوت مفيدًا للغاية عندما تكون الأدلة الأخرى لهوية المتكلم إما مفقودة أو غامضة للغاية [2].

بالنسبة للتعرف الآلي على المتكلم ، فإن مهمة النظام هي إما التحقق من المتكلم المطالب به ، أو تحديد المتكلم من بعض المجموعات المعروفة. اعتمادًا على التطبيق ، يمكن تقسيم التعرف على المتكلمين إلى منطقتين فرعيتين متصلتين ولكن مختلفتين على النحو التالي: التحقق التلقائي من المتكلم (Automatic speaker verification (ASV)) والتعرف التلقائي على المتكلم (Automatic speaker identification (ASI)).

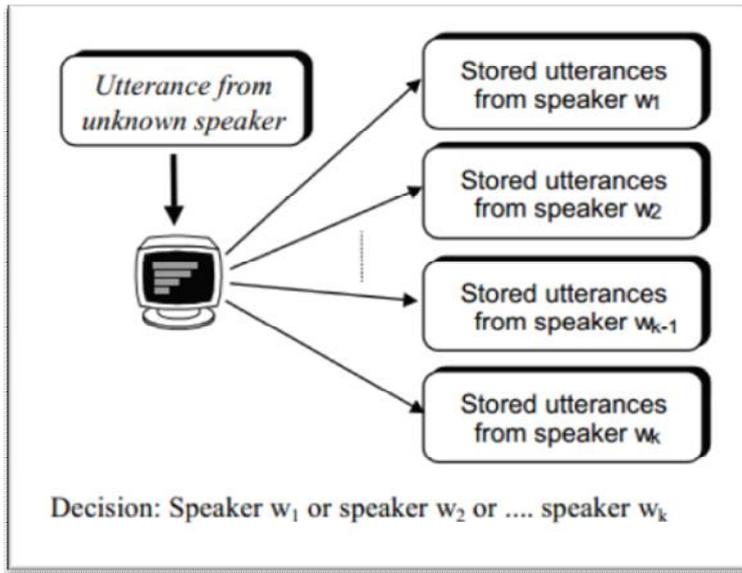
الهدف من التعرف على المتكلم هو التعرف على الشخص من خلال صوته عن طريق استخراج المعلومات من النص المنطوق ، والإجابة تلقائيًا على السؤال "من المتكلم؟". إذا لزم الأمر ، يمكن أيضًا توسيع النظام للإجابة على سؤال "ما يقال؟ من خلال استخراج معلومات مختلفة. في الواقع ، يرتبط التعرف على المتكلم إلى حد ما بالتعرف على الكلام [2].

تحديد المتكلم هو عملية تحديد هوية المتكلم من بين مجموعة الأصوات المخزنة في قاعدة البيانات يحاول نظام تحديد المتكلم الإجابة على السؤال ، أي متكلم هو هذا؟ بينما يعني التحقق من المتكلم تحديد ما إذا كان الصوت المتكلم

المجهول يطابق الصوت المعروف للمتكلم الذي تتم المطالبة بهويته. بينما يجيب نظام التحقق من المتكلم علي السؤال هل هذا الشخص هو من يدعي؟ [2]

1.3 التعرف الآلي على المتكلم (ASI)

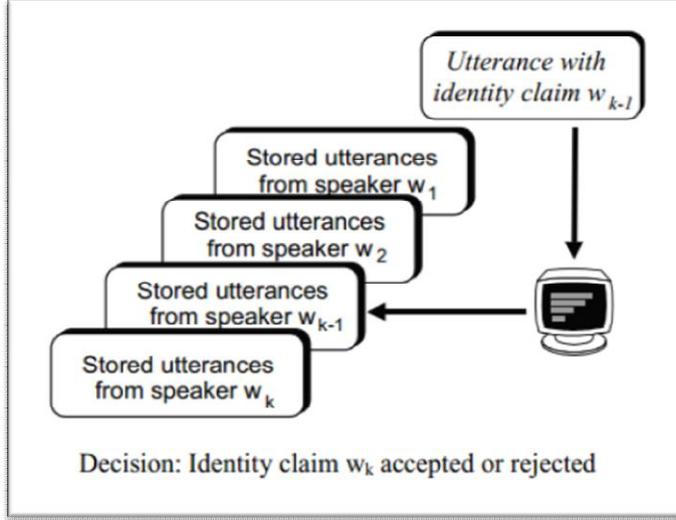
في تحديد المتكلم ، يجب أن يُنسب الكلام من متكلم غير معروف ، أو لا ، إلى مجموعة من المتكلمين المعروفين الذين تتوفر مراجع لهم. بشكل عام ، مع إعطاء مجموع الأصوات N ، يتطلب تحديد المتكلم اختيار أي من الأصوات N المعروفة للنظام تتطابق بشكل أفضل مع نمط المتكلم غير المعروف [2] ، كما هو موضح في الشكل 1



الشكل 1 مبدأ تحديد المتكلم [2]

2.3 التحقق التلقائي من المتكلم (ASV)

في التحقق من المتكلم ، يتم تقديم مطالبة هوية أو تأكيدها للمتكلم المجهول. في هذه المهمة ، يتم إعطاء نطق المتكلم غير المعروف مع الهوية المطالب بها والهدف هو تحديد ما إذا كان الكلام مشابهاً بدرجة كافية لإشارة النمط المرتبط بالهوية المزعومة لقبول هذا الإدعاء. في هذه الحالة ، يلزم إجراء مقارنة واحدة فقط للأنماط بغض النظر عن حجم المتكلمين [2] ، كما هو موضح في الشكل 2



الشكل 2 مبدأ التحقق من المتكلم [2]

على الرغم من وجود الكثير من القواسم المشتركة بين مهام تحديد الهوية والتحقق ، إلا أن الإجراءات المستخدمة في كل منها يمكن أن تكون مختلفة تمامًا. تستخدم مهام التحقق وتحديد الهوية قاعدة بيانات مخزنة يمكن استخدام الأنماط المرجعية للمتكلمين N المعروفين وتقنيات التحليل والقرار المماثلة. ومع ذلك ، يكمن الاختلاف الأكثر أهمية بين تحديد الهوية والتحقق في عدد بدائل القرار. يمكن القول أن التحقق من المتكلم هو مهمة أبسط لأنه لا يتطلب سوى قرار ثنائي ، أي قبول أو رفض الهوية المزعومة من الكلام و أداء مستقل عن عدد المتكلمين [2].

4. مجالات التطبيق التعرف الآلي على المتكلم

يتضمن التعرف الآلي على المتكلم استخدام منظومة لتحديد المتكلم من جملة منطوقة. في عملية التعرف على المتكلم ، كلما عرفت الشخص بشكل أفضل ، كان من الأسهل التعرف على الآخرين من خلال حديثهم. مثل البشر ، تحتاج أنظمة التعرف التلقائي على المتكلم إلى فترة تدريب لمعرفة شكل الكلام.

بدأ البحث في التعرف على المتكلم في الستينيات ، عندما حاول العلماء استخدام مخطط طيف الكلام (cepstrum) كأداة للتعرف على المتكلمين لم تكن التكنولوجيا في ذلك الوقت متقدمة بما يكفي لاستكمال العمل اليدوي لعلماء الصوتيات في تفسير مخططات الطيف، في الوقت الراهن هنالك عاملان رئيسيان يجعلان من إنشاء نظام للتعرف أو التحقق من المتكلم ممكنا هما [3]:

- ❖ أولاً: التطور الكبير في المعالجات CPU وبطاقات الشاشة GPU.
- ❖ ثانياً: يوفر نظام الهاتف حاليًا شبكة واسعة الانتشار ومألوفة من أجهزة الاستشعار للحصول على الإشارة الصوتية وتسليمها. لذلك ، يوجد في الوقت الحاضر مجموعة كبيرة ومتنوعة من الأصوات المخزنة والجهازية للاستعمال.

ومن أهم مجالات تطبيق التعرف الآلي علي المتكلم :

❖ **التحكم في الوصول (Access control):** من المحتمل أن تكون التطبيقات المتعلقة بالوصول الآمن هي الأكثر شيوعًا. تتمتع هذه التطبيقات بميزة أنه على عكس كلمات المرور والمفاتيح الشخصية ، لا يمكن سرقة الأصوات. ومع ذلك ، يمكن نسخها باستخدام أجهزة التسجيل. من أجل حماية أنظمة الأمان من هذا الخطر ، تتم دعوة المتكلم الذي يرغب في الوصول إلى مكان آمن بشكل عام للإعلان عن نص معين. في هذه الحالة ، يتم أخذ كل من المتكلم ومحتوى الخطاب في الاعتبار [3].

❖ **مصادقة المعاملات (Transaction authentication):** بالإضافة إلى التحكم في الوصول ، قد تكون هناك حاجة إلى مستويات أعلى من التحقق للمعاملات المصرفية عبر الهاتف ، على سبيل المثال لتحقيق معاملات أكثر أمانًا. تركز بعض التطبيقات الحديثة علي التعرف علي المتكلم للمصادقة علي عمليات البيع والشراء [3].

❖ **تطبيق القانون والتعرف الجنائي (Law enforcement and forensics):** يشمل العديد من التطبيقات مثل الإشراف على الإفراج المشروط حيث يتم استدعاء المفرج عنهم للتحقق من أنهم يقيمون في المنزل، مراقبة مكالمات السجن وتطبيقات الطب الشرعي ، والجمارك ، وما إلى ذلك من طلبات إنفاذ القانون ، ولاسيما تلك المتعلقة بموقع المفقودين والتعرف الجنائي [3].

❖ **إدارة البيانات الصوتية (Speech data management):** يعد استخدام التعرف على المتكلم أيضًا أمرًا أوليًا في العديد من التطبيقات مثل التنقل في محرك البحث أو إعطاء الأوامر للأنظمة الذكية ، حيث يتم تمييز الأصوات الواردة باستخدام التعرف على المتكلم. يمكن أيضًا العثور على إدارة البيانات الصوتية في الغرف الذكية لتتبع من قال ماذا ، على سبيل المثال ، في غرفة الاجتماعات [3].

❖ **الشخصنة (Personalisation):** يمكن العثور على مجموعة متنوعة من الأجهزة والأنظمة لتنظيم وتسهيل حياتنا اليومية. في الأساس ، ستعمل هذه الأجهزة التي يتم التحكم فيها بالصوت بشكل أفضل مع التخصيص الشخصي الجيد. بالإضافة إلى ذلك ، هناك أيضًا اهتمام أولي باستخدام توصيف المتكلم من أجل توفير معلومات لاستخدامها في الإعلانات أو الخدمات الأخرى [3].

يمكن استخدام جميع التطبيقات المذكورة أعلاه تقريبًا في بيئة متعددة الوسائط أي ، يمكن دمج مهمة التعرف على المتكلم مع تقنيات أخرى التي تتضمن طرائق أخرى. على سبيل المثال ، يمكن استخدام بصمات الأصابع أو مسح قزحية العين

للتحكم في الوصول ، ويمكن استخدام ميزات الوجه على سبيل المثال ، في غرف الاجتماعات الذكية وتخصيص الخدمة [3].

5.الهدف من التعرف الآلي على المتكلم

الهدف الرئيسي من التعرف الآلي على المتكلم هو الحصول على طرق فعالة لتسهيل حياة الإنسان و التي تتلخص في

- يمكن استعمالها كنوع من أنواع أنظمة الأمان.
- إضفاء أكثر موثوقية للمعاملات و المبادلات التجارية.
- تسهيل المراقبة الشخصية للأفراد و تطبيقات التعرف و تحديد الأشخاص.
- الوصول للأنظمة الذكية و تسهيل التعامل الشخصي معها .
- التخلص من وحدات الإدخال و لواحق الأجهزة الذكية

6.صعوبات التعرف الآلي على المتكلم

كل المتكلمين لديهم أصواتهم الخاصة ، بسبب الحبال الصوتية والحنجرة وشخصيتهم الفريدة. لا يختلف الصوت فقط بين المتكلمين ، بل هناك أيضًا اختلافات واسعة في متكلم معين ، وسنقوم بإدراج بعض هذه الاختلافات في الأقسام الفرعية أدناه.

✓ أسلوب الحديث

كل البشر يتحدثون بشكل مختلف ، إنها طريقة للتعبير عن شخصيتهم. إنهم لا يستخدمون المفردات الشخصية فحسب ، بل لديهم طريقة فريدة للإعلان والتأكيد. يختلف أسلوب التحدث أيضًا باختلاف المواقف ، فنحن لا نتحدث بنفس الطريقة في البنك ، كما هو الحال مع والدينا أو مع أصدقائنا ، كما ينقل البشر مشاعرهم من خلال الكلام. نتحدث بشكل مختلف خاصة عندما نكون سعداء ، حزينين ، محبطين ، متوترين ، غاضبين ، مندفعين ، إلخ. إذا كنا حزينين فيمكننا خفض أصواتنا والتحدث ببطء أكثر ، وإذا شعرنا بالإحباط فيمكننا التحدث بصوت أقوى [4]

✓ جنس المتحدث

تختلف أصوات الرجال والنساء ، والسبب الرئيسي هو أن النساء عادة ما يكون لديهن حبال صوتية أقصر من الرجال. النغمة الأساسية للأصوات الأنثوية تبلغ ضعف نبرة الرجال بسبب هذا الاختلاف [4].

✓ السمات الجسدية

لكل متكلم سماته الجسدية الفريدة ، وهذا يؤثر على حديثه.

شكل وطول الحبال الصوتية ، وتشكيل التجاويف ، وحجم الرئتين ، وما إلى ذلك. تتغير هذه السمات بمرور الوقت ، على سبيل المثال في صحة المتحدث أو عمره [4].

✓ سرعة التحدث

نحن نتحدث في أنماط مختلفة من السرعة وفي أوقات مختلفة. إذا كنا متوترين ، فإننا نميل إلى التحدث بشكل أسرع ، وإذا كنا متعبين ، فإن سرعتنا تميل إلى الانخفاض. نتحدث أيضاً بسرعات مختلفة سواء كنا نتحدث عن شيء معروف أو شيء غير معروف [4].

✓ تغير القناة (Channel variability)

أحد جوانب التباين هو السياق الذي تنبعث فيه الموجة الصوتية. هنا لدينا مشكلة الضوضاء التي تتغير بمرور الوقت وأنواع مختلفة من الميكروفونات وأي شيء يؤثر على محتوى الموجة الصوتية من المصدر إلى المتلقي. وتؤدي إلى تمثيل منفصل في الكمبيوتر تسمى هذه الظاهرة تغير القناة [4].

✓ الضوضاء

عادة يتم التكلم في بيئة من الأصوات ، أو دقائق الساعة ، أو طنين الكمبيوتر ، أو راديو يتم تشغيله في مكان ما في الردهة ، أو متكلم آخر في الخلفية ، وما إلى ذلك. يشار إلى هذا بشكل عام باسم الضوضاء ، أي المعلومات غير المرغوب فيها في إشارة الكلام ، نحتاج إلى تحديد وتصفية هذه الضوضاء من الإشارة الصوتية [4].

✓ الصدى

نوع آخر من الضوضاء هو تأثير الصدى ، وهو إشارة الكلام التي ترتد عن الجسم المحيط ، ويحدث ذلك في الميكروفون بعد بضعة ثواني. إذا كان المكان الذي تم فيه إصدار الإشارة الصوتية له صدى قوي ، فقد يؤدي ذلك إلى ظهور ظاهرة تسمى الصدى ، والتي يمكن أن تستمر حتى بضع ثوانٍ [4].

7. التعرف الآلي على قراء القرءان الكريم

الكلام هو وسيلة مهمة وأساسية للتواصل بين البشر. إنهما الميزة الفريدة التي يتمتع بها كل شخص ويختلف بها عن الآخر. لذلك ، استفاد العلماء من هذه الميزة الفريدة وطوروا العديد من الأنظمة التي يمكنها التعرف والتمييز بين البشر من خلال أصواتهم ، اليوم مع التقدم الملحوظ في مجال النمذجة الإحصائية للكلام ، فإن أنظمة التعرف على المتكلم لديها العديد من التطبيقات في العديد من المجالات في حياتنا والتي تتطلب التفاعل بين الآلة والإنسان [5].

كل شخص في هذا العالم له صوت وأسلوب كلام مختلف وفريد. وذلك ينطبق أيضا على قراء القرآن الكريم الذين يميلون تمامًا إلى الاختلاف فيما بينهم في الصوت وأسلوب التلاوة علي الرغم من أنهم يتلون نفس الآية القرآنية .
تلاوة القرآن هي القراءة (الترتيل أو التجويد أو التعبير) بصوت عالٍ أو تلاوة أو ترديد أجزاء من القرآن. يُطلق على القارئ اسم التالي ، أو المرتل ، أو المجود ، أو القاري وهي الأكثر شيوعًا [5].
يجب أن يتم التلاوة وفقًا لقواعد النطق والتجويد والقيصورات التي وضعها نبي الإسلام محمد (عليه الصلاة والسلام). يوجد اليوم الآلاف من القراء الذين يتلون القرآن ولكل منهم صوته الفريد الذي يميزه عن غيره رغم أنه يقرأ نفس السورة أو الآية [5].

منذ اليوم الذي انتشر فيه الإسلام في جميع أنحاء العالم ، زادت الحاجة إلى تطبيقات القرآن على الإنترنت والتسجيلات الصوتية لكثير من القراء المشهورين والموثوقين الذين يتلون القرآن الكريم. بما أن القرآن الكريم هو الكتاب المقدس للمسلمين الذي يرشدتهم في حياتهم ، فقد قاموا بنشره عبر وسائل الإعلام للوصول إلى المسلمين وغير المسلمين في جميع أنحاء العالم. على الرغم من أن نشر القرآن الكريم عبر وسائل الإعلام يعد ميزة كبيرة لتعليم المسلمين في جميع أنحاء العالم الرسالة الحقيقية التي يحملها القرآن ، إلا أن هناك أشخاصًا ضد الإسلام يحاولون تغيير وتنسيق الآيات من خلال جلب قراء مزيفين لتلاوة القرآن بطريقة خاطئة [5].

لذلك تأتي الحاجة إلى تطوير نظام تمييز يتعرف على قراء القرآن الكريم المشهورين والموثوقين ويميزهم عن غيرهم.

8. منشورات بحثية في التعرف الآلي على قراء القرآن الكريم .

في هذا الجزء من الفصل سنستعرض أهم الأعمال في التعرف الآلي على قراء القرآن الكريم وذكر أهم ما توصلوا إليه منقولاً بألسنتهم نبدأ بورقة البحثية ذات عنوان (A Machine Learning Approach for Recognizing the Holy Quran Recite) في هذه المقالة [6]، تم اقتراح منهج للتعرف الآلي على قراء القرآن باستخدام مصنفات KNN و ANN. تم تسجيل ومقارنة أداء كلا المصنفين ANN و KNN. في الأساس ، تم استخراج ميزات MFCC وعمل بها في ANN لكل من التدريب والاختبار. تم اختيار فصلين مشتركين من قبل عشرة قراء مشهورين في مكة والمدينة.

كان معدل التعرف الذي تم الحصول عليه لجميع القراء للصور 18 و 36 هو 97.6% و 96.7% باستخدام مصنفات ANN. ومع ذلك ، كان معدل التعرف الذي تم الحصول عليه لجميع القراء للصورتين 18 و 36 هو 97.03 و 96.08% على التوالي باستخدام KNN ومنه حققت كلا المصنفات الخاصة ب RNN و KNN نسبة تعرف جيدة علي قراء القرآن الكريم وتم إثبات فعالية هذه المصنفات في التعرف علي المتكلم ويقترح الباحث انه في

المستقبل من أجل تحسين أداء النموذج وزيادة دقة التعرف ، يُقترح تطبيق خوارزميات استخراج سمات آخري علي غرار (Vector Quantization) على إشارات الموجة الصوتية واستعمال خوارزميات التصنيف الأخرى مثل نماذج ماركوف المخفية (HMMs). [6]

أما في الورقة البحثية ذات عنوان (Quranic Reciter Recognition: A Machine Learning Approach) يقول الباحث في الخاتمة [7] انه باستخدام 120 تلاوة إجمالية ، قمنا بتحليل تلاوات ل12 قارئ. استخدمنا طريقتين لمعالجة التلاوات الصوتية. الأول هو MFCC و Pitch ، والثاني باستخراج الصور القائمة على Spectrogram. [7]. يتم التعرف على الميزات من خلال Naïve Bayes و J48 و Random Forest ، حيث يتم اختيارها نظرًا لأدائها الجيد الشامل في أحدث التقنيات.

يُظهر التحليل التجريبي أن المصنفين يمكن أن يتعرفوا علي قارئ القرآن بكفاءة إذا كانت ميزات MFCC و Pitch تمثل الصوت. في مثل هذه الحالة ، نحصل على دقة التعرف بنسبة 88٪ مع Naïve Bayes و Random Forest التي توضح أنه يمكن التعرف على قارئ بشكل فعال من تلاوة الآيات القرآنية.

أما في البحث العلمي المنشور بعنوان (A Holy Quran Reader/Reciter Identification System Using Support Vector Machine) يقول الباحث انه في هذا البحث قام بتطوير نظام التعرف على قارئ القرآن الكريم من الموجات الصوتية التي تمثل نطق "الآيات". تم استخراج ميزات MFCC من الإشارات الصوتية. تكشف النتائج التي تم الحصول عليها أن دقة تحديد الهوية باستخدام SVM أفضل بكثير مقارنة بشبكة ANN. يعطي 96.59٪ الدقة باستخدام SVM بينما تم الحصول على دقة 86.17٪ باستخدام ANN. النظام المقترح هو أول نظام للتعرف على قارئ القرآن الكريم. يقترح في المستقبل توسيع نطاق عمله مع أكثر من 15 قارئًا للقرآن الكريم واستخدام المزيد من تقنيات الاستخراج وخوارزميات التصنيف مثل Gaussian Mixture لزيادة دقة التعرف [8].

أما في البحث العلمي المنشور تحت عنوان (An Automatic Qari Recognition System) [9] تم تصميم وتنفيذ نظام التعرف على السماعيات البسيط لاستخدامه في التعرف على قراء القرآن الكريم.

قام الباحث بتصميم وتنفيذ نظام تعرف على السماعيات استعمل فيه (mfcc) و (LBG Vector Quantization) لاستخدامه في التعرف على قراء القرآن الكريم وقد تحصل علي نتائج رائعة لاسيما باستعمال (VQ) حيث بلغت نسبة التعرف 95% [9].

أما في البحث العلمي المنشور بعنوان (Quran Reciter Identification: A Deep Learning Approach) [10] حيث قام الباحث باختبارات علي (Recurrent Neural Networks (RNNs))

و الذاكرة طويلة المدى ثنائية الاتجاه (BLSTM) للتعرف على قارئ القرآن الكريم حيث تم الحصول على نتائج واعدة علي وجه الخصوص باستعمال BLSTM اعلي من تلك المتحصل عليها باستخدام الشبكات العصبية الاصطناعية التقليدية (ANNs) بالإضافة إلى نماذج أساسية مختلفة أخرى.

تظهر نتائج المتحصل عليها أن شبكات BLSTM أكثر فاعلية من الشبكات العصبية العميقة الأخرى (DNNs) لنمذجة الديناميكيات الزمنية لتلاوات القرآن. لقد تفوق النموذج المقترح باستخدام BLSTM من طبقتين بشكل كبير على أحدث الحلول القائمة على ANN (والتي حققت دقة 91.28% مع 40 طبقة مخفية) من خلال تحقيق دقة تصل إلى 99.89% [10]. مما يدل على أن النموذج المقترح أكثر دقة ، بينما يكون أيضاً غير معقد من الناحية الحسابية.

في النهاية نلاحظ أن كل الباحثين اشتركوا في أمور منها استخراج مميزات MFCC وتم الوصول إلي نسب عالية في تحديد والتعرف علي قارئ القرآن الكريم برغم استخدامهم لشبكات عصبية مختلفة ك ANN و SVM و RNN

9. الخاتمة

تكلمنا في هذا الفصل علي نشأة مجال علمي مهم وهو التعرف الآلي علي المتكلم ورواد هذا المجال ودوافع البحث فيه وتطويره والمجالات التي يستخدم فيها وفعاليتها وهل يمكن أن يطبق التعرف علي المتكلم في مجال صعب مثل قراءة القرآن و التعرف علي قارئ والأعمال والبحوث التي تمت في هذا المجال وهل توصلوا إلي نتيجة مرضية وفعالة .

الفصل الثاني

المقاربة المقترحة لمعالجة البيانات

الصوتية واستخلاص السمات (دراسة منهجية وتطبيقية)

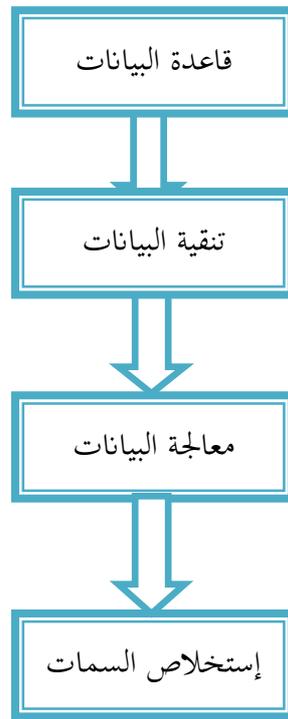


الفصل الثاني: المقاربة المقترحة لمعالجة البيانات الصوتية واستخلاص السمات (دراسة منهجية

وتطبيقية)

1. مقدمة

في عملية التعرف علي المتكلم هنالك كثير من الأمور التي تتحكم في نسبة نجاعة النظام المراد إنشاؤه منها سلامة الملفات وهل هي ذات فاعلية ومعرفة إذا ما كانت مراحل تنقية ومعالجة البيانات واستخراج خصائص وسمات الصوت تساعد علي تحسين قدرة النظام المراد إنشاؤه علي التعرف وزيادة فاعليته في ما يلي نفتح مقاربة لمعالجة البيانات الصوتية الشكل (3)



الشكل 3 مراحل المقاربة المقترحة

2. قاعدة البيانات (Dataset)

في هذا البحث تم جمع قاعدة البيانات وملفات الصوتية بشكل يدوي علي مدار أسبوعين وذلك لعدم توفر تسجيلات صوتية لقراء قرآن جزائريين الذين كان البحث مقتصرًا عليهم فقط وذلك يعود لقلة استوديوهات التسجيل في الجزائر خاصة لمثل هذا النوع من التسجيلات وحتى ولو توفرت تسجيلات فهي مقتصرة علي سور معينة مثل سورة البقرة وسورة الكهف أو تواجد تسجيلات ذات جودة سيئة جدا نظرا لأنها سجلت بأجهزة غير احترافية مثل أجهزة الهاتف

النقل أو سجلت في فعاليات دينية مثل جلسات الذكر أو صلاة التراويح ومنه قررنا توسيع مجال بحثنا علي أشهر قراء القرآن في المغرب العربي الكبير. تم الحصول علي قاعدة البيانات التالية الجدول (1) من موقع (tvquran) ومن وزارة الشؤون الدينية والأوقاف لولاية تبسة

تتكون قاعدة البيانات المجمعة مما مجموعه 1140 ملف صوتي لسور قرآنية لعشرة قراء من المغرب العربي الكبير ، كل قارئ يتوفر علي 114 سورة ، بأطوال متفاوتة و 60 حزب 6236 أية بمعني قرآن كامل في مجلد يتراوح حجمه بين 1.3 و 2 جيجا تكون هذه الملفات بصيغة mp3.

الجدول 1 قاعدة البيانات المتحصل عليها

حجم المجلد (Folder)	متوسط زمن الملفات الصوتية	صيغة الملفات الصوتية	عدد الملفات الصوتية	جنسية القارئ	الاسم بالعربي	
1.7gb	د23	Mp3	114	جزائري	ياسين الجزائري	01
1.9gb	د 22	Mp3	114	ليبي	مُحَمَّد العالم الدوكالي	02
2gb	د 19	Mp3	114	مغربي	عمر القزابري	03
1.9gb	د 27	Mp3	114	تونسي	علي البراق	04
1.4gb	د 16	Mp3	114	جزائري	رياض الجزائري	05
1.6gb	د 17	Mp3	114	جزائري	رشيد بلعالية	06
1.7gb	د 19	Mp3	114	موريتاني	الناجي ولد بلال	07
1.4gb	د 16	Mp3	114	مغربي	العيون الكوشي	08
1.8gb	د 18	Mp3	114	ليبي	إبراهيم كشيديان	09
1.7gb	د 22	Mp3	114	تونسي	أبو عبد الله منير المظفر التونسي	10

3.تنقية البيانات (Data Cleaning)

في عملنا هذا وبعد عملية جمع البيانات في المرحلة السابقة و الإستماع لعينات متفرقة منها, تم ملاحظة وجود شوائب وتكرارات تستدعي القيام بعملية إضافية وهي عملية تنظيف البيانات (data cleaning)

تنظيف البيانات هي عملية إصلاح أو إزالة البيانات غير الصحيحة أو التالفة أو المنسقة بشكل غير صحيح أو المكررة أو الغير مكتملة ضمن مجموعة البيانات

هناك العديد من الأسباب و التي تجعل تنظيف البيانات أمرا ضروريا منها:

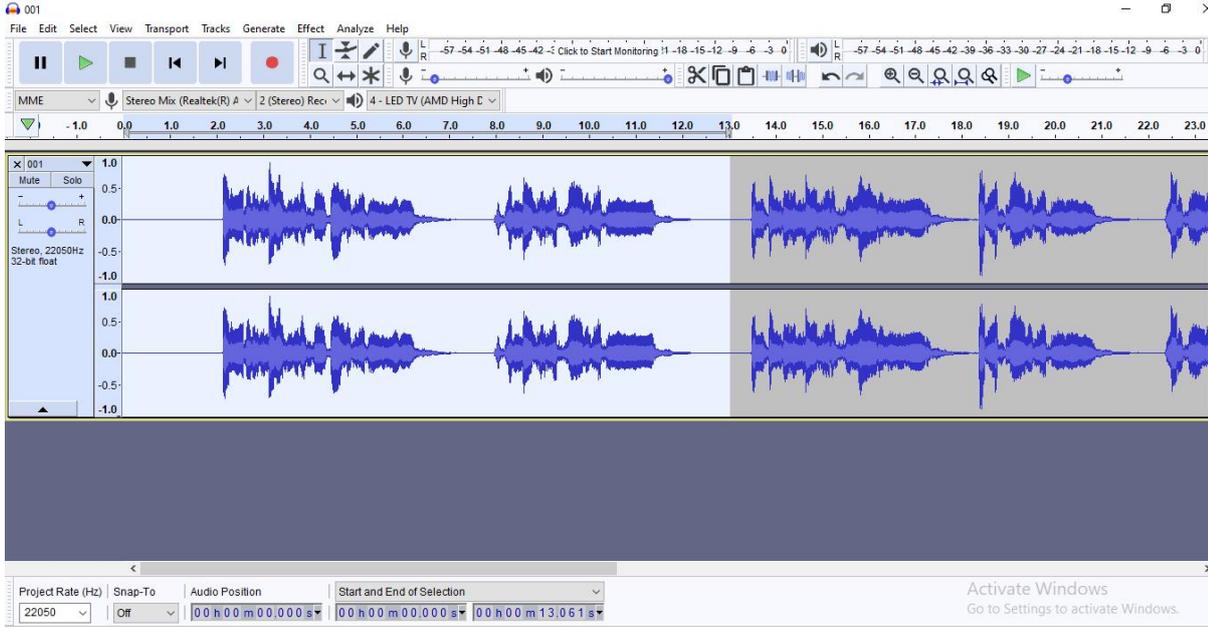
- تقليل هامش الخطأ .
- تحديد جودة البيانات .
- زيادة صلاحية البيانات .

في هذه المرحلة وبعد الاستماع لعينات لعشر قراء لاحضنا وجود تكرارات تتمثل في البسملة (بسم الله الرحمن الرحيم) في بداية كل سورة و صدق الله العظيم في نهاية كل منها أين تتراوح مدة البسملة في المقاطع الصوتية للقراء بين 5 ثوان إلى نصف دقيقة على حساب مقامات تلاوة كل قارئ بينما تتراوح مدة صدق الله العظيم في نفس المقاطع بين 2 و 7 ثواني.

كذلك تمت ملاحظة وجود شوائب في هذه المقاطع الصوتية و المتمثلة في ذكر استديو التسجيل أو التعريف بصاحب هذه التلاوة و عادة ما تكون هذه الشوائب بصوت مختلف عن صوت القارئ بعد تحديد هذه الشوائب و التكرارات قمنا بتحميل و استعمال برنامج audacity وهو برنامج مجاني على نظام تشغيل الويندوز يتيح لنا العمل على المقاطع الصوتية و تحويلها إلى ملفات wav كما يعطينا القدرة على قص أو لصق مقاطع صوتية من الملف الأصلي بطريقة يدوية كما يوضح الشكل (4)

الهدف من قيامنا بعملية نزع الشوائب هو الحصول على مقاطع صوتية تحتوي فقط على صوت القارئ و تقليل نسبة الخطأ وزيادة صلاحية البيانات, أما الهدف من عملية نزع التكرارات التي تتمثل في البسملات و الصدقات هو تقليل سعة تخزين الملفات لزيادة جودة و فعالية البيانات و تقليل العبء على الجهاز.

بعد نهاية عملية التنظيف نكون قد تحصلنا على مقاطع صوتية منزوعة الشوائب و التكرارات بصوت القارئ فقط ذات مدة و سعة اقل من الملفات الأصلية.



الشكل 4 صورة توضح طول التكرار في المقطع الصوتي مأخوذة من برنامج audacity

4. معالجة البيانات (Data Pre-processing)

في هذه المرحلة نعمل علي استخراج و تحويل البيانات المتحصل عليها في مرحلة تنقية البيانات و تمريرها بعدة عمليات للحصول على بيانات منسقة و مفهومة ويسهل العمل عليها
تمر مرحلة معالجة البيانات في عملنا هذا بعدة مراحل هي:

- توحيد صيغ الملفات الصوتية (Standardization of audio file formats)
- دمج السور
- نزع الصمت من القراءان الكريم
- إختيار عينات

1.4 توحيد صيغ الملفات الصوتية (Standardization of audio file formats)

قبل الحديث عن أنواع الصيغ علينا أولاً معرفة طريقة تخزين الصوت في الملفات الرقمية بعد تسجيل الصوت من خلال أدوات التسجيل هناك خيارين لتخزين هذه البيانات, إما تخزينها دون أي عملية ضغط مباشرة إلى ملفات صوتية أو تخزينها بعد عملية ضغطها اعتماداً على إحدى الترميزات أو الصيغ الشائعة و المتوافقة مع العتاد الموجود.
في هذه المرحلة تمت ملاحظة اختلاف في صيغ و ترميزات و قنوات إخراج الصوتية في الملفات التي نعمل عليها ما توجب عنه ضرورة توحيد هذه الصيغ و الترميزات و القنوات لكل الملفات الصوتية و ذلك عبر ثلاث عمليات هي:

1.1.4. التحويل من صيغ mp3 إلى صيغ wav:

هناك الكثير من صيغ الملفات الصوتية ولكل منها مميزات و عيوب تختلف من صيغة لأخرى ومن هذه الصيغ:

- صيغ ملفات الصوت عديمة الضغط (Uncompressed Audio File Formats): وهي ملفات يبقى فيها الصوت بالجودة التي تم تسجيله فيها دون أي تغيير ورغم ذلك تبقى هذه الملفات هي الأكبر حجما من بين صيغ الملفات الأخرى من أشهر هذه الصيغ (wav, pcm, aff) تستخدم غالبا من طرف المنتجين [11].

- صيغ ملفات الصوت المضغوطة بلا فقدان (Lossless Compressed Audio File Formats): هذه الملفات تكون مضغوطة و ذات حجم اقل من السابقة و تحافظ على جودتها من أشهرها (flac, alac) يستخدمها المنتجون في بعض الأحيان [11]

- صيغ ملفات الصوت المضغوطة جدا مع فقدان (Very compressed audio file formats with lossy): هذه الملفات تكون مضغوطة جدا حتى تصبح اقل جودة وأكثر تشتتا و هذا ما يجعل هذا النوع من الملفات غير ملائم للإنتاج أو العمل عليها لكنه يجعله مثاليا لخدمات البث الموسيقي و من أشهرها (mp3 و acc) [11]

في عملنا هذا وبما أن الملفات الصوتية التي نعمل عليها تشترك في صيغة mp3 وهي صيغة غير ملائمة في أعمال معالجة البيانات

قمنا بتحويل صيغ الملفات من mp3 إلى wav باستعمال مكتبة (pydub) وهي مكتبة مفتوحة المصدر على بايثون.

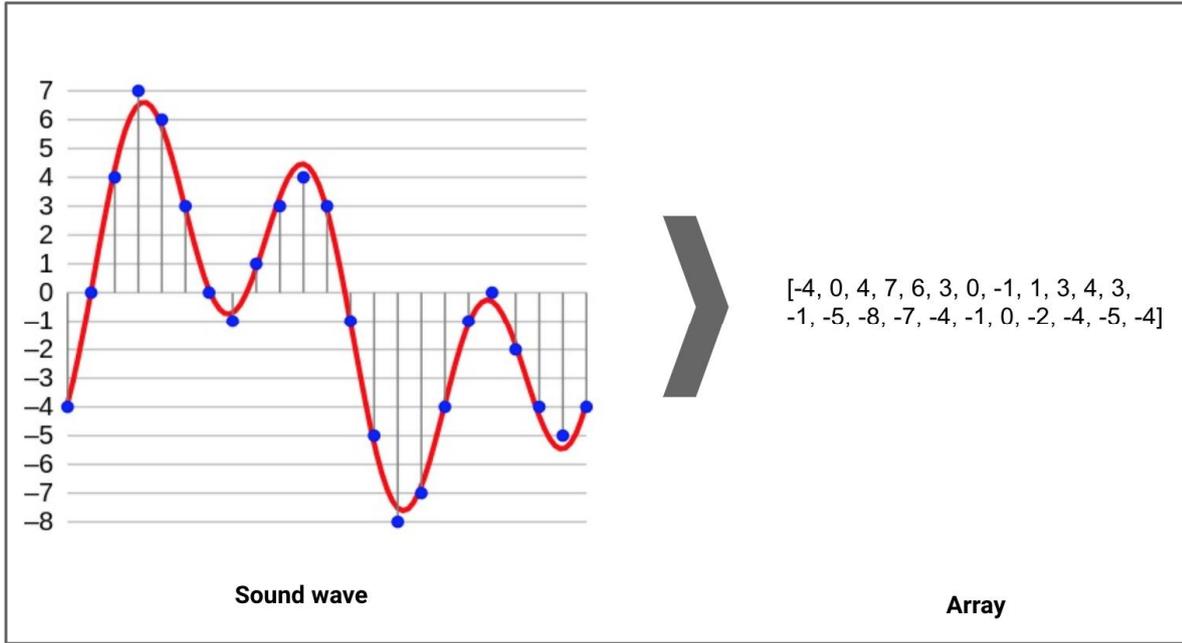
الهدف من هذه العملية هو الحصول على ملفات صوتية بصيغة wav وهي الصيغة الأكثر ملائمة في أعمال معالجة البيانات في المراحل القادمة , هذا التحويل ليس الهدف منه زيادة جودة الصوت ولا إكساب الملف أية خصائص إضافية.

2.1.4. معدل اخذ العينات (Sampling rate)

الإشارات التي نستخدمها في العالم الحقيقي مثل أصواتنا تسمى الإشارات التناظرية (Analogique) لمعالجة هذه الإشارات في أجهزة الكمبيوتر نحتاج إلى تحويل الإشارات إلى شكل رقمي في حين إن الإشارات التناظرية مستمرة في كل من الوقت و السعة فان الإشارة الرقمية (Numérique) منفصلة في كل منهما و لتحويل الإشارة من وقت

مستمر إلى وقت منفصل يتم استخدام عملية تسمى التكميم (QUantification) يتم قياس الإشارة على فترات زمنية معينة يشار إلى كل قياس على انه عينة [12] الشكل (5).

يتم تمثيل الإشارة المستمرة بخط أخضر ملون بينما يشار إلى العينات المنفصلة بخطوط عمودية زرقاء



الشكل 5 تمثيل أخذ عينات من الإشارة. [13]

يقاس معدل اخذ العينات ب الكيلوهيرتز (kHz) أي عدد العينات في الثانية, توصي العديد من السلطات المعنية بحفظ التسجيلات الصوتية مثل الرابطة الدولية للمحفوظات الصوتية والسمعية البصرية (IASA) [14] بمعدلات حفظ ب:

- من 22 إلى 44 كيلوهيرتز بالنسبة للموسيقى
- من 16 إلى 22 كيلوهيرتز بالنسبة للكلام

في عملنا هذا و بما أن الملفات التي نعمل عليها تختلف في معدل اخذ العينات مقاطع الصوتية قمنا بتوحيد معدل اخذ العينات بين كل المقاطع الصوتية باستعمال مكتبة. (Pydub) بتحديد 22050 Hz لكل المقاطع الصوتية,الهدف من هذه العملية هو التحصل على قاعدة بيانات متماثلة في معدل اخذ العينات.

3.1.4. توحيد قنوات الإخراج

كل مزاج (Mixer) يحتوي علي قنوات للإدخال تدخل منها الإشارات الصوتية ليتم خلطها ومعالجتها ثم إخراجها عبر قناة واحدة أو قناتين الشكل(6).



الشكل 6 الفرق بين stereo + mono الصور مأخوذة من برنامج audacity

عندما تجتمع قناتي الإدخال في قناة واحدة نقول عندها أن إشارة المزج مونو (MONO) يعني المزج الصوتي في القناة المركزية واحدة [15].

عندما تجتمع كل قنوات الإدخال في قناتين يعني ويسرى نقول عندها أن إشارة المزج ستيريو (STEREO) يعني إن الخليط الصوتي يمر علي قناتين مركزيتين يمكن سماع هذا الخليط باستعمال مكبرين لصوت [15].

في هذه العملية تمت ملاحظة أن التسجيلات الصوتية لبعض القراء تختلف من حيث قنوات الإخراج ما يستدعي توحيد قنوات الإخراج لكل المقاطع بطريقة مماثلة سواء للمونو أو ستيريو باستعمال مكتبة (pydub) قمنا بتوحيد كل المقاطع الصوتية علي قناة واحدة مونو. الهدف من هذه العملية هو توحيد كل المقاطع علي قناة إخراج واحدة لتفادي التكرار وتقليل العبء علي الآلة وخفض حجم الملف .

الهدف من مرحلة توحيد الصيغ هو التحصل علي ملفات متساوية في عدد اخذ العينات 22 كيلو هرتز ذات

صيغ مشتركة wav وقناة واحدة مونو

2.4 دمج الملفات الصوتية (السور)

يتكون القرآن الكريم من مجموعه 114 سورة علي الترتيب من سورة الفاتحة إلي سورة الناس أي أن لكل قارئ 114 ملف صوتي وبما أن قاعدة البيانات تحتوي علي 10 قراء فان مجموع الملفات الصوتية في قاعدة البيانات هو 1140 ملف صوتي. في هذه المرحلة نقوم بدمج المقاطع صوتية لمقرئ معين باستعمال مكتبة (pydub) لقراءة الملفات الصوتية باستخدام خاصية العشوائية في انتقاء الملف تقوم باسترجاع قيمة عشوائية من بين عناصر أي سلسلة نمررها لها وفي حالتنا هذه نقوم بإعطائها رقم بين 1 و 114 وهو عدد المقاطع ,لتقوم بإرجاع رقم عشوائي المتمثل في رقم المقطع المراد دمجه لنقوم بالدمج وتحقيق دمج عشوائي لكل المقاطع الصوتية فتتحصل علي ملف صوتي شامل المتمثل في قرآن كريم كامل لمقرئ معين, الجدول (2).

الهدف من هذه المرحلة هو إنشاء ملف واحد لكل قارئ سيكون ملف مرجعي في هذا العمل وهو إجراء احترازي وتنظيمي تم اعتماده في بروتوكول المقاربة المقترحة الشكل (7) يوضح عملية دمج الملفات الأولية في ملف صوتي شامل .



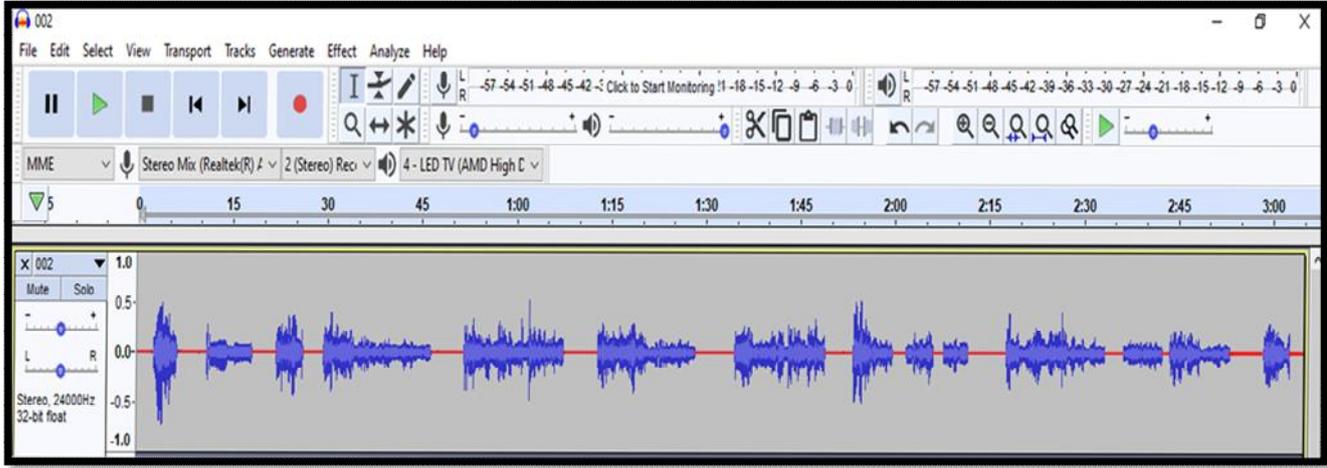
الشكل 7 مرحلة الدمج العشوائي لملفات القراء

الجدول 2 الحجم الساعي للملف النهائي المتحصل عليه لكل مقرئ

إسم القارئ	الحجم الساعي للملف النهائي
ياسين الجزائري	25 سا
مُحَمَّد العالم الدوكالي	26 سا
عمر القزابري	24 سا
عمر القزابري	27 سا
رياض الجزائري	30 سا
رشيد بلعالية	25 سا
الناجي ولد بلال	25 سا
العيون الكوشي	27 سا
إبراهيم كشيدان	28 سا
أبو عبد الله منير المظفر التونسي	26 سا

3.4.3 نزع الصمت من القراءان الكريم

يعرف الصمت على أنه جزء من الصوت غير محسوس بما في ذلك الضوضاء غير الملحوظة والنقرات القصيرة جدا في المقطع الصوتي وهي عبارة عن إشارات صوتية غير معرفة وليس لها معنى شكل(8).



الشكل 8 عدد مقاطع الصمت بلون الأحمر الصورة مأخوذة من برنامج audacity

مقطع صوتي لأحادي المقرئين علي برنامج Audacity يظهر عدد مقاطع الصمت سيتم تحديد وإزالة مقاطع الصمت من التسلسل الصوتي، سينقسم التابع الصوتي المعالج إلى مقاطع صوتية منفصلة دون صمت ثم يتم دمج هذه المقاطع إلى مقطع صوتي واحد منزوع صمت. في هذه المرحلة وبعد الحصول تجمع البيانات المرجعي للقارئ وبعد استعمال مكتبة (pydub) وباستعمال (silence.detect_silence) التي تتوفر على ثلاث إعدادات (Settings) هي : المقطع الصوتي المراد نزع الصمت منه ، العتبة (threshold) التي تعرف علي أنها صمت وطول المدة التي تعتبر صمت ، عند تمرير الملف الصوتي عليها نتحصل كنتيجة على لائحة ثنائية الأبعاد ببداية ونهاية كل صمت ثم نقوم بنزع الصمت المعلم في اللائحة، ولصق مقاطع الكلام للحصول على ملف تجمع البيانات المرجعي للقارئ منزوع الصمت .

الهدف من هذه العملية هو الحصول على مقاطع صوتية نهائية لقراءان كريم منزوع الصمت ، لأن تسجيلات القراءان الكريم عبارة عن مقاطع صوتية تحتوي الكثير من فترات الصمت وذلك راجع أحكام التجويد ومقامات النلاوة ، ولأن الصمت عبارة عن مقاطع غير معرفة قد تشكل عائقا أثناء مرحلة استخراج السمات.

4.4 اختيار العينات من الملفات الصوتية (Sample Random Selection)

إن الملفات الصوتية المتحصل عليها في المرحلة السابقة تكون كبيرة جدا في الحجم الساعي قد تتراوح بين 23 و 27 ساعة على حسب المقرئ وطريقة القراءة ولأن المرحلة القادمة هي مرحلة استخلاص سمات من الصوت فلا يمكن العمل

على ملف تصل مدته إلى 25 ساعة. ومنه يجب اختيار عينات فقط من الملف الكلي بطريقة عشوائية على أن تكون من كل الملف الصوتي و نقوم بالتحكم في طولها (Sample Duration) وعددها (Number of Samples) في هذه المرحلة الأخيرة من معالجة البيانات قمنا باستعمال مكتبة (pydub) للعمل على المقطع الصوتي منزوع الصمت وذلك باختيار عينات عشوائية عن طريق خاصية (Random) نحن نتحكم في طول العينات L وعدد العينات N . الهدف من هذه العملية هو تقليل العبء على الآلة لأننا نعمل على عينات ولا نعمل على الملف الكلي. باختيار عينات بطريقة عشوائية نضمن أن تكون العينات من كل المقطع الصوتي لأن طريقة القراءة في سورة معينة تختلف عن قراءة سور أخرى نظرا لأحكام تلاوتها ومشاعر المقرئ فيها من آيات حزن وفرح وترغيب وترهيب. إن البيانات المتحصل عليها في المرحلة السابقة قد تصل في حجمها إلى 20 جيجا وبعد المرور على مرحلة معالجة البيانات ومن توحيد للصيغ ودمج ونزع للصمت ومن اختيار لعينات عشوائية تم الحصول على نوع من البيانات ذات صيغة WAV لا يتجاوز حجمها 300 ميجا عبارة عن عينات لصوت قراء منزوع الشوائب جاهز للعمل عليها في المرحلة القادمة.

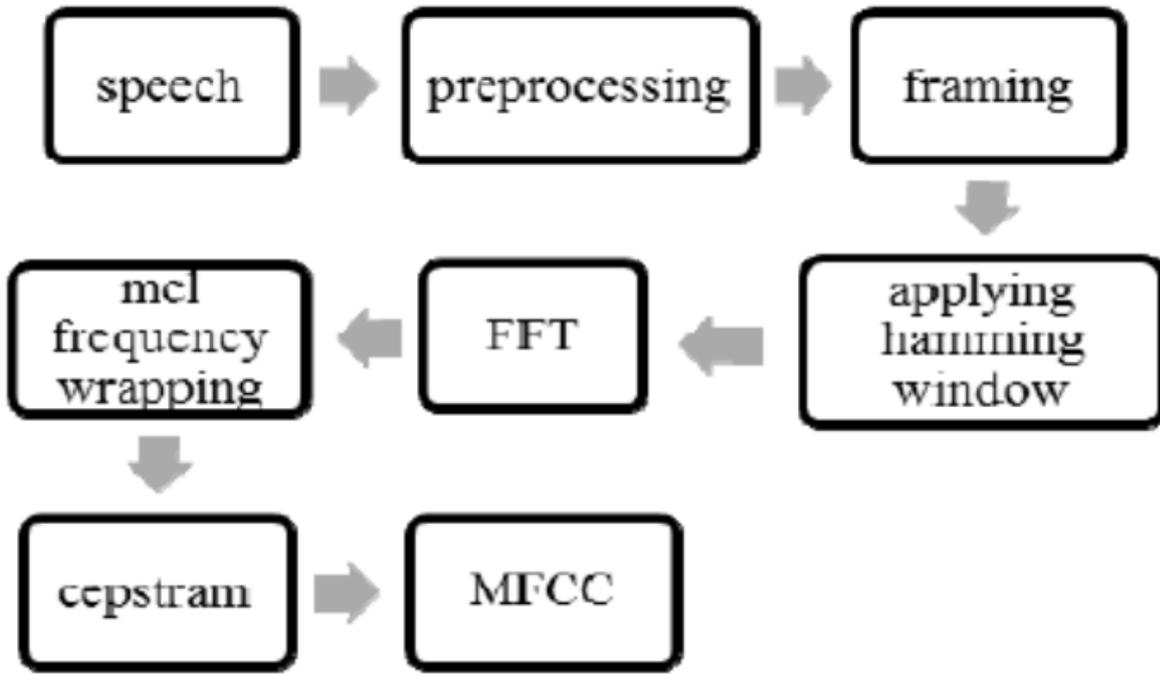
5. خوارزمية إستخلاص السمات (Feature Extraction)

يمكن القول أن مرحلة استخلاص السمات من أهم مراحل المقاربة المقترحة ، لما لها من تأثير علي نسبة التعرف الكلية للنظام. حيث يجب أن تكون السمات المستخلصة من العينات الخاصة بالقراء مستقلة قدر الإمكان ومتباينة، ومفيدة (معبرة عن إشارة الكلام الأصلية المستخلصة منها). نورد فيما يلي أهم خوارزمية لاستخلاص السمات في مجال التعرف علي الكلام والمتكلم.

Mel Frequency Cepstral Coefficients (MFCC)

تعتبر خوارزمية MFCC من أهم خوارزميات إستخلاص السمات حسب النتائج المعروفة من مجال البحث يوضح الشكل (9) مخطط لهذه العملية تمر هذه الخوارزمية علي 5 مراحل.

- ✓ تحديد أطر كل قسم من أقسام الإشارة (Frame Blocking) .
- ✓ نافذة هامينغ (Window Hamming) (الحصول على طيف الاتساع لكل إطار).
- ✓ تحويل فورييه السريع (FFT) (استخراج عناصر التردد).
- ✓ تردد الاعوجاج ميل (Mel- Frequency Warping): (التحويل إلى طيف ميل)
- ✓ حصول علي معاملات [5] Cepstrum



الشكل 9 مخطط مراحل خوارزمية الـ MFCC [16]

قبل أن تبدأ خوارزمية استخلاص السمات، يجب تحويل إشارة الصوت التماثلية إلى إشارة رقمية. نعلم أن طيف الإشارة الكلامية يعاني من تناقص وتحامد بشكل ملحوظ عند الترددات العالية وذلك بسبب الإشعاع عند الشفاه. لذلك نحتاج إلى تعزيز هذه الترددات عبر تمرير الإشارة إلى مرشح تمرير مرتفع تعطي علاقة هذا المرشح بالمعادلة تالية :

$$y[n] = x[n] - k \cdot x[n - x], \quad 0 < k < 1$$

حيث $x[n]$ هي مخرج المرشح، بينما $x[n]$ هي المدخل. و k من رتبة 95.0 تقريباً تعتبر الإشارة الكلامية من الإشارات المتغيرة مع الزمن، لذلك لا بد من تقطيع الإشارة إلى أطر، وليكن بطول n نقطة، أقصر من 25ms تقريباً لضمان استقرار الإشارة الكلامية على كامل الإطار (ضمن أحد الصوتيات) نفضل الضرب بنافاذة على شكل جرس مثل Window Hamming لتخفيف حدة الانقطاعات بين الأطر، لضمان نتائج أفضل. لذلك، نأخذ نوافذ متداخلة بمقدار M عينة تكون من رتبة نصف طول عينات الإطار N . تعطي

عبارة نافذة Hamming بالعلاقة [16]

$$w[n] = 0.54 - 0.46 \cos\left[\frac{2\pi n}{N - 1}\right]$$

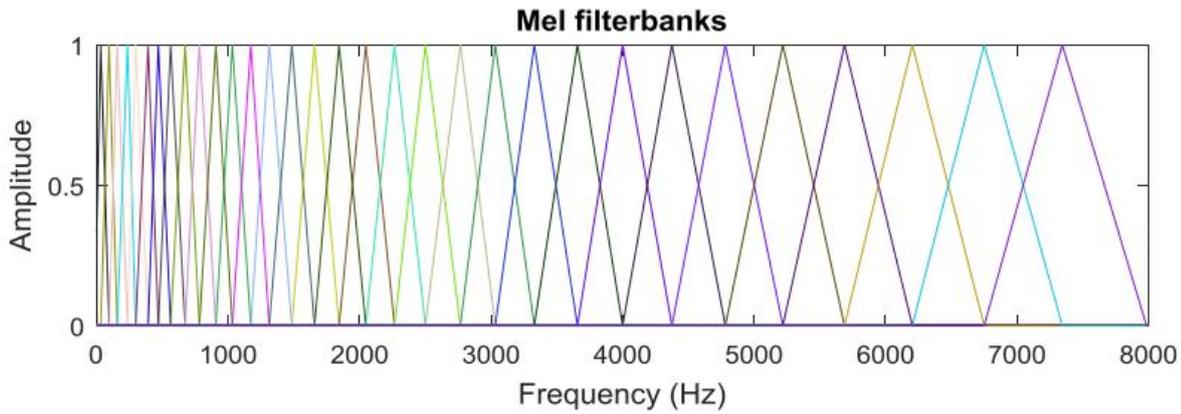
حيث N عدد عينات الإطار.

تكون عبارة إشارة ناتج جداء الإطار بالنافذة هي [16] :

$$y[n] = x[n] \cdot w[n]$$

بعد الضرب بالنافذة، يتم تطبيق خوارزمية تحويل فورييه السريع ((Transform Fourier Fast(FFT)) لإيجاد تحويل فورييه المتقطع ((Transform Fourier Discrete(DFT)) لكل إطار، ونبقي النصف الأول من

الإشارة الناتجة, بالتالي نكون حصلنا على الطيف الصحيح من أجل كل إطار، ولكن هذا الطيف يحتوي علي الكثير من المعلومات التي لن نحتاجها من اجل مرحلة مطابقة السمات,دائماً بالأخذ بالاعتبار خصائص وطريقة عمل الأذن البشرية. لذلك نوزع ترددات الطيف إلى مجموعات قليلة لنرى كمية الطاقة المتواجدة ضمن كل مجموعة. [16] تتم هذه العملية معيارياً بضرب طيف كل إطار بمجموعة مرشحات تكون على شكل مثلثات ضمن هذه الخوارزمية MFCC. يوضح الشكل (10) هذه المرشحات. [16].



الشكل 10 حزمة مرشحات Mel [16]

نلاحظ أن المرشح الأول ضيق جداً ويعطينا تلميحاً عن كمية الطاقة المتواجدة قرب 0 Hz مع زيادة التردد تصبح المرشحات أعرض حيث لا نستطيع تمييز الصغيرة التغيرات في الترددات. [16] تعطي العلاقة التي تمكننا من حساب ترددات MEL انطلاقاً من التردد كما يلي:

$$F(MEL) = 2595 \times \log_{10} \left[1 + \frac{f}{700} \right]$$

نضرب استجابة كل مرشح بطيف طاقة الإطار ثم نجمع المعاملات لنحصل على الطاقة ضمن كل مرشح. بعد هذه العملية يبقى لدينا عدد معاملات مساوٍ لعدد المرشحات المستخدمة والتي تعبر عن توزيع الطاقة ضمن كل حزمة. بعد ذلك نطبق تابع اللوغاريتم على هذه المعاملات ثم نطبق تحويل (DCT)، حسب العلاقة التالية [16]

$$C_n = \sum_{k=1}^k (\log S_k) \cos \left[\frac{\left(k - \frac{1}{2} \right) n \pi}{k} \right]$$

حيث $n=1,2,3,\dots,k$ و S_k معاملات تحويل فورييه.

في هذه المرحلة استعملنا مكتبة (librosa) وهي مكتبة مفتوحة المصدر علي بايثون تتيح لنا العمل علي مقاطع wav وتحليل الصوت أو إضافة مؤثرات أو استخراج سمات وقد قمنا بتطبيق خوارزمية MFCC علي مقاطع العينات من المرحلة السابقة باستعمال (librosa.feature.mfcc) علي كل العينات والتي تتوفر علي إعدادات

مهمة منها عدد السمات في الصوت وطول المقفزة بين كل استخلاص وقد تحصلنا علي مصفوفة ثلاثية الأبعاد (عدد العينات , Number of Samples , طول العينة MFCC Length , عدد السمات Number of MFCC) كما نلاحظ في الشكل (11) المصفوفة النهائية ثلاثية الأبعاد لكل العينات.



الشكل 11 المصفوفة النهائية ثلاثية الأبعاد لكل العينات (1280,13,216) الصورة مأخوذة من برنامج python

في نهاية هذه المرحلة نتحصل علي ملفين بصيغة CSV الأول يحتوي علي مصفوفة الكلية لMFCC لكل القراء والملف الثاني يحتوي علي عمود الوسوم (Labels Vector) لأصحاب العينات مشفرين من 1 إلي 10 كما في الشكل (12) النواتج لهذه المرحلة



الشكل 12 نتائج هذه المرحلة

6. خاتمة

في نهاية هذه المقاربة وبعد العمل على معالجة البيانات تم الحصول على ملفين ملف أول متمثل في مصفوفة MFCC لعينات كل القراء وملف ثاني متمثل في عمود الوسوم أو المسميات (Labels Vector) هذه الملفات ستكون محل دراسة تجريبية وتطبيقية في الفصل التالي.

الفصل الثالث

التعلم العميق، إنشاء و عمل تجريبي



الفصل الثالث : التعلم العميق، إنشاء و عمل تجريبي

1. مقدمة

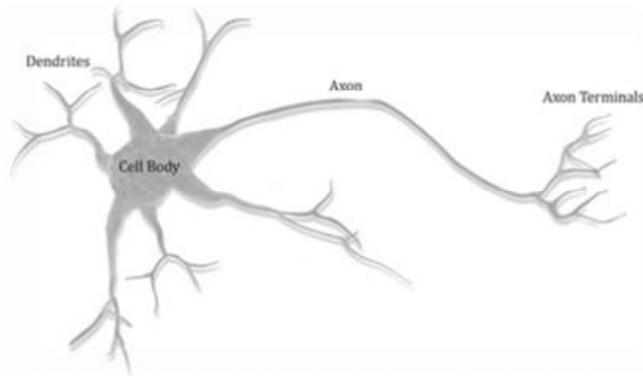
ستعرض في هذا الفصل وتكملة لمقاربتنا المقترحة إلى التعلم العميق (Deep Learning) و النموذج المقترح والطبقات الخاصة به، وباستعمال نتائج المرحلة السابقة وهي مصفوفة mfcc و عمود الوسوم من اجل تعليم (train) واختبار (test) معماريتنا المقترحة وتقييم أدائها ونسبة التعرف الخاصة بها وكذلك القيام بتجارب عملية للوصول إلى أفضل إعداد ونتيجة يمكن الحصول عليها. الشكل (13)



الشكل 13 النموذج المقترح لنظام التعرف الآلي على قراء القرآن الكريم

2. التعلم العميق (Deep Learning)

يعد مجال التعلم العميق (Deep Learning) أحد أكثر المفاهيم التي يتم الحديث عنها في التعلم الآلي (machine learning) في كل من المجتمع الأكاديمي ووسائل الإعلام. تستمد فكرة التعلم العميق والشبكات العصبية من التمثيل البيولوجي للدماغ البشري الشكل (14) يتناول إيجاد نظريات وخوارزميات تتيح للآلة أن تتعلم بنفسها عن طريق محاكاة الخلايا العصبية في جسم الإنسان. تركز معظم أبحاث التعلم العميق على إيجاد أساليب استنباط من المتجردات بتحليل مجموعة بيانات ضخمة باستخدام تصميمات مختلفة لشبكات عصبية. [17]



الشكل 14 رسم تخطيطي لعصب بيولوجي [17]

3. الشبكة العصبونية التلافيفية (Convolutional Neural Network)

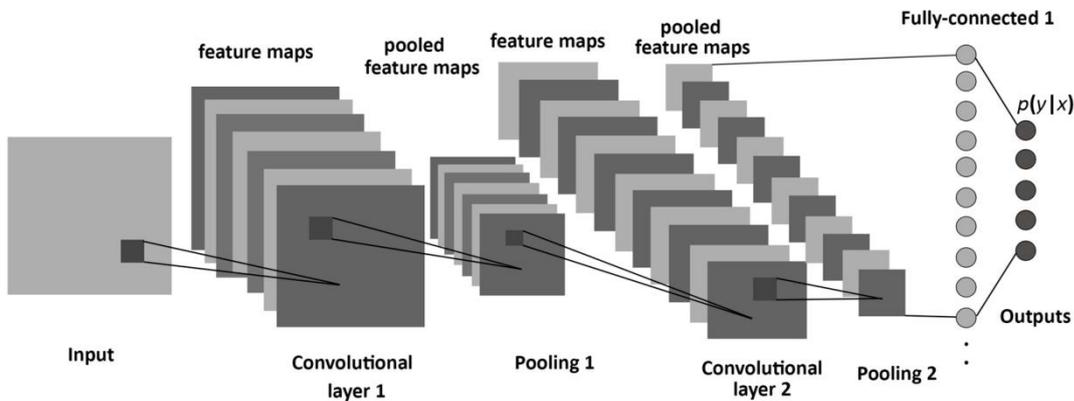
الشبكات العصبونية الإلتفافية أحد أنواع الشبكات العصبونية العادية. إلا أن اختلافها عن الشبكات العصبونية الأخرى يكمن في ما تتعلمه الشبكة، وكيف تُبنى، وماهية الهدف المنشود منها. عادةً ما تُستخدم شبكات CNN في تطبيقات الرؤية الحاسوبية (Computer vision)، وتحليل المشاهد المرئية؛ إذ تتميز بوجود طبقة مخفية واحدة أو أكثر، من شأنها استخراج السمات الموجودة في الصور أو مقاطع الفيديو. [17]

1.3 تدريب الشبكة العصبونية التلافيفية (CNN)

يتم تدريب الشبكة عن طريق إنشاء طبقات متصلة تتمثل في:

1. الطبقة التلافيفية (Convolutional Layer) يُطبَّق في طبقة التلافيفية مُرشِّحًا (Filter) ويُعرف أيضًا بـ (kernel) من شأنه تحديد وجود سمات أو أنماط مُعينة في المصفوفة الأصلية .
2. طبقة التنفيع (Activating function) تتحكم هذه الطبقة في كيفية انتقال الإشارة من طبقة إلى أخرى ، على غرار الخلايا العصبية في دماغنا.
3. طبقة التجميع (Pooling Layer) تقوم باختزال المصفوفة الناتجة من الطبقة السابقة، وذلك باستخدام تابع ما (تابع أكبر قيمة (Max-pooling)، أو تابع المتوسط (Average -pooling)).
4. طبقة التسوية (Flatten) تقوم هذه المرحلة بجعل نتائج المرحلتين على شكل شعاع خطي.
5. طبقة الارتباط الكامل (Fully-connected Layer) في الطبقة يتم توصيل جميع طبقات الشبكة بكل خلية عصبية من طبقة سابقة إلى خلايا عصبية من الطبقة التالية.

الشكل (15)



الشكل 15 رسم تخطيطي لطبقات CNN [17]

2.3 النموذج المقترح في الشبكة العصبونية التلافيفية (CNN) لتعرف الآلي على قراء القراءان الكريم

إن النموذج المقترح يحتوي على طبقة إدخال واحدة (input layer) حجمها (216,13,1) وطبقة إخراج (output layer) و 8 طبقات مخفية (hidden layers) وهذا ما يوضحه الجدول (3) من عدد طبقات و دوال التفعيل وعدد المرشحات ونسبة التسرب .

الجدول 3 الطبقات المقترحة ودوال التفعيل الخاصة بالمعمارية

ملاحظات	دالة التفعيل	نوع الطبقة
64 مرشح	relu	2Conv التلافيفية
/	/	BatchNormalization
نواة بأبعاد 2x2	/	Max-pooling
32 مرشح	relu	2Conv التلافيفية
/	/	BatchNormalization
نواة بأبعاد 2x2	/	Max-pooling
/	/	التسوية Flatten
32	relu	dense
نسبة 30 %	/	Dropout
10	softmax	dense

3.3 تقييم أداء المعمارية المقترحة

البروتوكول التجريبي :

من اجل اختبار نسبة التعرف الخاصة بالمعمارية المقترحة تم العمل بطريقتين لتقسيم مصفوفة mfcc الكلية وذلك باستعمال طريقة (train_test_split) وهي تقسم المصفوفة علي حسب النسب إلى مجموعة تدريب ومجموعة اختبار , وطريقة (Cross-validation) وهي طريقة مقترحة في معظم الأوراق البحثية.

يتمحور البروتوكول التجريبي على اختبار المعمارية المقترحة وتم العمل على الإعدادات التالية : خصائص الصوت في mfcc و اختيار معدل اخذ البيانات (SAMPLING RATE) وعدد العينات المأخوذة من كل قارئ والمدة الزمنية للعينه الواحدة. تم تثبيت إعدادات الشبكة في كل التجارب على (Batch_Size : 32) و (Epoques : 50) و (Learning_rate : 0,001).

4. البرامج ومكتبات والبيئات المستعملة في هذا البحث

أجريت الدراسة والعمل باستخدام أدوات مختلفة في كلا المرحلتين (معالجة الصوت وإنشاء الشبكة)

➤ بايثون (Python)

هي لغة برمجة، عالية المستوى سهلة التعلم مفتوحة المصدر قابلة للتوسيع، تعتمد أسلوب البرمجة الكائنية (POO). لغة بايثون هي لغة مُفسَّرة ومُتعدِّدة الاستخدامات وتستخدم بشكل واسع في العديد من المجالات، كبناء البرامج المستقلة باستخدام الواجهات الرسومية وفي تطبيقات الويب، ويمكن استخدامها كلغة برمجة نصية للتحكم في أداء العديد من البرمجيات مثل بلندر. بشكل عام، يمكن استخدام بايثون لعمل البرامج البسيطة للمبتدئين، ولإنجاز المشاريع الضخمة في الوقت نفسه في بايثون قمنا باستيراد بعض المكتبات التي سهلت تنفيذ نموذجنا:

➤ مكتبة (NumPy)

هي إضافة على لغة البرمجة بايثون، تُستخدم للتعامل مع المصفوفات الكبيرة والحقول متعددة المستوى، وكذلك توفر مكتبة كبيرة من الاقتارات الرياضية عالية المستوى للعمل على هذه الحقول والمصفوفات.

➤ مكتبة (Pydub)

هي إضافة على لغة البرمجة بايثون، تُستخدم للتعامل مع الملفات الصوتية بتحويل صيغها وقيام بعمليات رياضية عليها بعد تحويلها إلى سجل (list) من البيانات

➤ مكتبة (librosa)

هي إضافة على لغة البرمجة بايثون لتحليل الموسيقى والصوت. توفر اللبئات الأساسية اللازمة لإنشاء أنظمة المعلومات الصوتية.

➤ مكتبة (TensorFlow)

هي مكتبة ذكاء اصطناعي مفتوحة المصدر ، تستخدم الرسوم البيانية لتدفق البيانات لبناء النماذج. يسمح للمطورين بإنشاء شبكات عصبية واسعة النطاق مع العديد من الطبقات.تستخدم هذه الحزمة بشكل أساسي في: التصنيف

(Classification) ، والإدراك (Perception) ، والفهم (Understanding)، والاكتشاف (Discovering) ، والتنبؤ (Prediction).

➤ مكتبة (Keras)

هي مكتبة ذكاء اصطناعي مفتوحة المصدر ، صُممت لتمكين إجراء التجارب على الشبكات العصبية العميقة بشكل سريع، وهي تركز على أن تكون سهلة الاستخدام ومرنة وقابلة للتوسيع.

➤ برنامج (Spyder)

هي بيئة تطوير متكاملة مفتوحة المصدر ومتعددة المنصات للبرمجة العلمية بلغة Python. يتكامل مع Spyder مع عدد من الحزم البارزة في كومة Python العلمية.

5. التجارب والنتائج العملية

التجربة الأولى

في هذه التجربة سنقوم باختبار تغير نسبة التعرف عند التغير في عدد خصائص الصوت الموجودة في MFCC وهل عند زيادة في عدد الخصائص ستزداد نسبة التعرف أم لا , نتائج التجربة تتمثل في الجداول الآتية :

❖ نقوم باختبار 13 خاصية في mfcc

الجدول 4 نتائج اختبار mfcc 13 باستعمال kfold

loss	accuracy	تقسيم (Cross-validation)
%0.562	%87.12	5 مجلدات
%0.756	%83.98	10 مجلدات

الجدول 5 نتائج اختبار mfcc13 باستعمال split

loss	accuracy	تقسيم (train_test_split)
%0.488	%86.32	التقسيم الأول
%0.771	%84.76	التقسيم الثاني

❖ نقوم باختبار 14 خاصية في mfcc

الجدول 6 نتائج اختبار mfcc 14 باستعمال kfold

loss	accuracy	تقسيم (Cross-validation)
%0.328	%87.12	5 مجلدات
%0.521	%85.93	10 مجلدات

الجدول 7 نتائج اختبار mfcc 14 باستعمال split

loss	accuracy	تقسيم (train_test_split)
%0.328	%87.12	التقسيم الأول
%0.521	%85.93	التقسيم الثاني

❖ نقوم باختبار 16 خاصية في mfcc

الجدول 8 نتائج اختبار mfcc 16 باستعمال kfold

loss	accuracy	تقسيم (Cross-validation)
%0.435	%85.54	5 مجلدات
%0.441	%85.93	10 مجلدات

الجدول 9 نتائج اختبار mfcc 16 باستعمال split

loss	accuracy	تقسيم (train_test_split)
%0.298	%89.06	التقسيم الاول
%0.521	%85.98	التقسيم الثاني

❖ نقوم باختبار 18 خاصية في mfcc

الجدول 10 نتائج اختبار mfcc 18 باستخدام kfold

loss	accuracy	تقسيم (Cross-validation)
%0.435	%88.67	5 مجلدات
%0.441	%87.56	10 مجلدات

الجدول 11 نتائج اختبار mfcc 18 باستخدام split

loss	accuracy	تقسيم (train_test_split)
%0.477	%89.84	التقسيم الاول
%0.728	%89.06	التقسيم الثاني

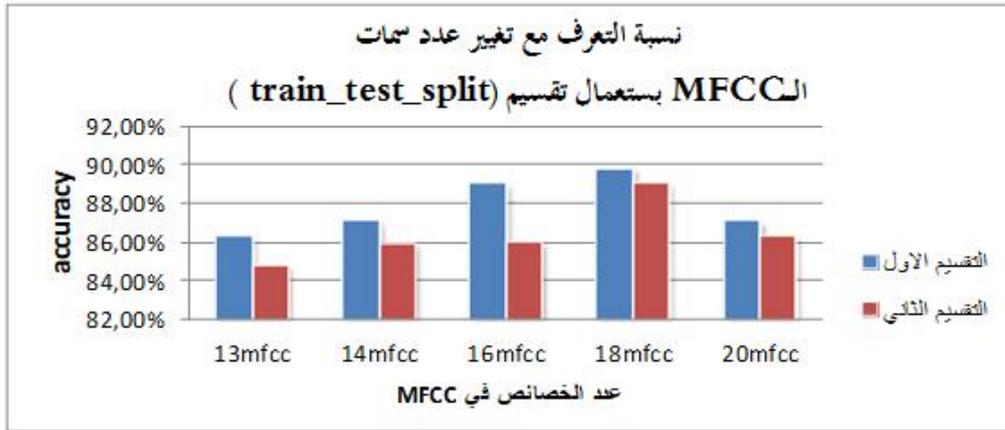
نقوم باختبار 20 خاصية في mfcc

الجدول 12 نتائج اختبار mfcc 20 باستخدام kfold

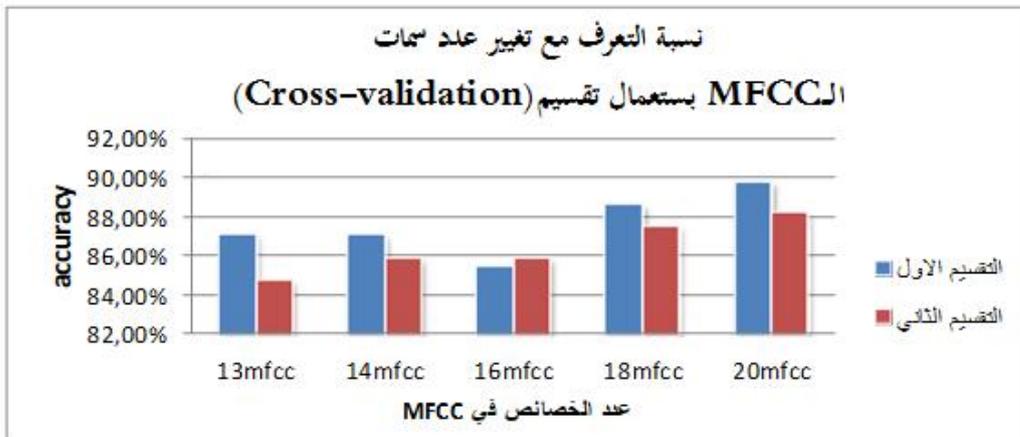
loss	accuracy	تقسيم (Cross-validation)
%0.332	%89.84	5 مجلدات
%0.519	%88.28	10 مجلدات

الجدول 13 نتائج اختبار mfcc 20 باستخدام split

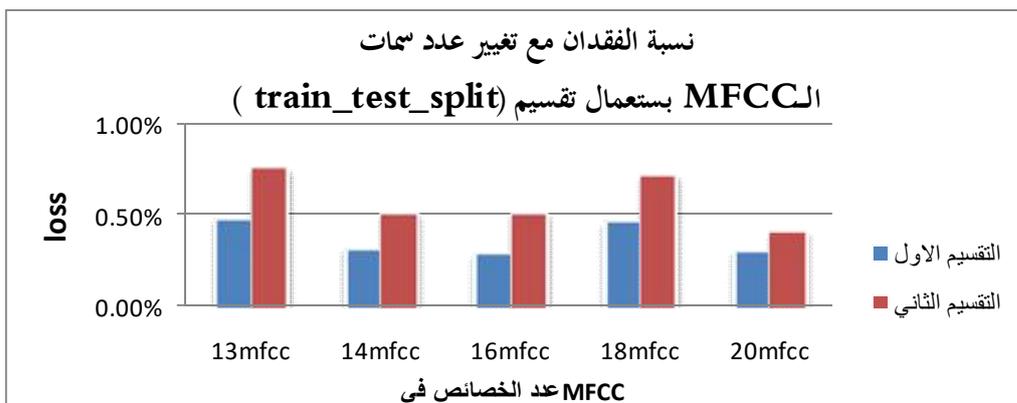
loss	accuracy	تقسيم (train_test_split)
%0.318	%87.10	التقسيم الاول
%0.425	%86.32	التقسيم الثاني



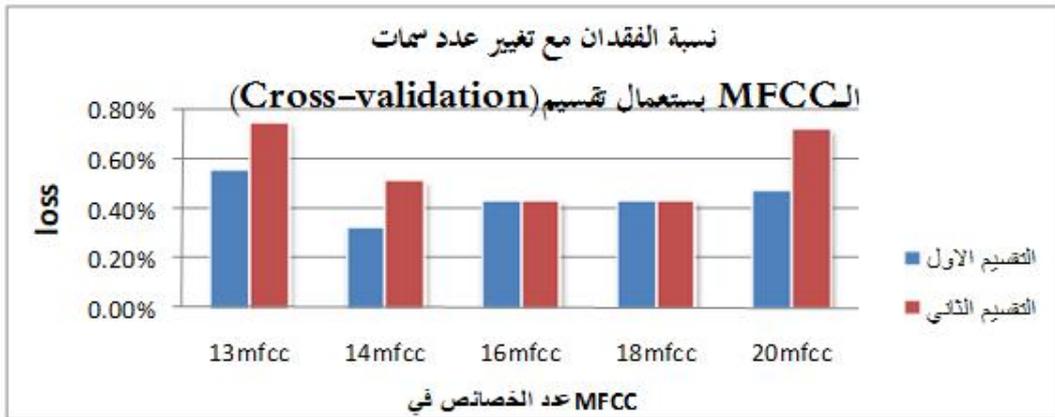
الشكل 16 نسبة التعرف مع تغيير عدد سمات باستخدام تقسيم split



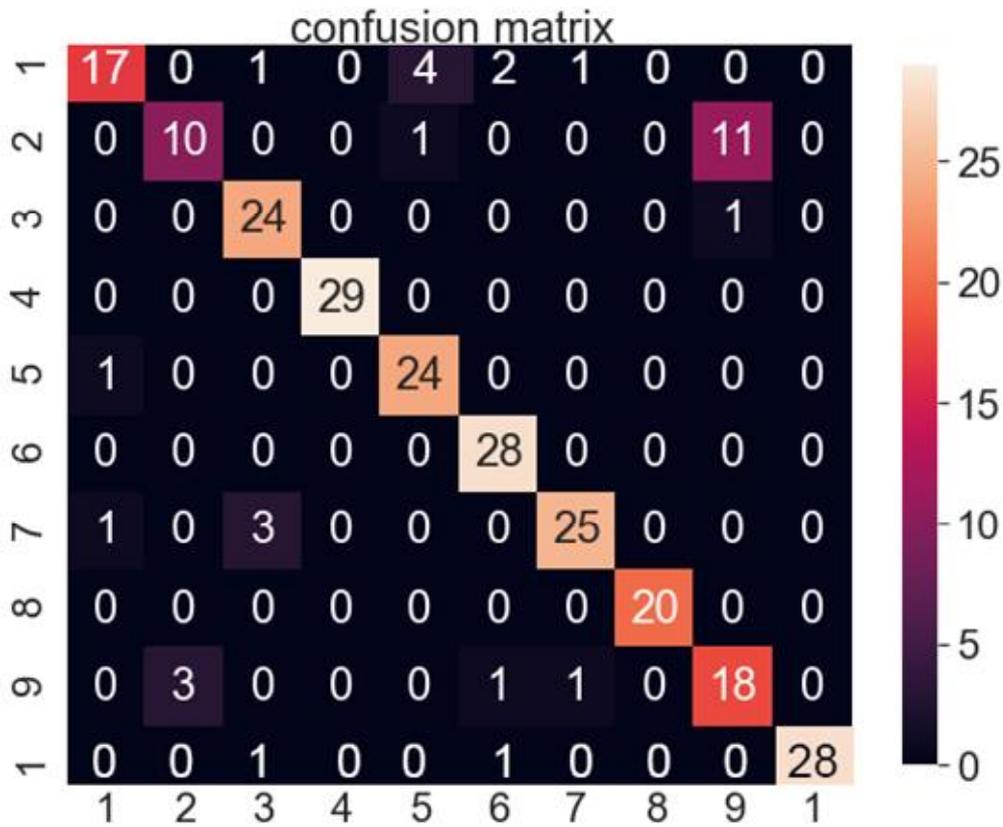
الشكل 17 نسبة التعرف مع تغيير عدد السمات باستخدام kfold



الشكل 18 نسبة الفقدان مع تغيير عدد السمات باستخدام تقسيم split



الشكل 19 نسبة الفقدان مع تغيير عدد السمات باستخدام تقسيم kfold



الشكل 20 مصفوفة الالتباس لنتيجة التصنيف باستخدام 13 mfcc بتقسيم (Cross-validation)

التجربة الثانية

في هذه التجربة سنقوم باختبار معدل أخذ العينات (SAMPLING RATE) مع العلم أن المدة الزمنية لكل عينة في هذا الإختبار هي خمسة ثواني و عدد العينات هو 128 عينة وعدد خصائص صوت هو 13 mfcc

❖ نقوم باختبار 22050Hz كمعدل اختيار للعينات

الجدول 14 نتائج اختيار 22050 Hz باستخدام kfold

loss	accuracy	تقسيم (Cross-validation)
%0.328	%87.12	5 مجلدات
%0.521	%85.93	10 مجلدات

الجدول 15 نتائج اختيار 22050 Hz باستخدام split

loss	accuracy	تقسيم (train_test_split)
%0.331	%87.50	التقسيم الأول
%0.525	%84.89	التقسيم الثاني

❖ نقوم باختبار 16000Hz كمعدل اختيار للعينات

الجدول 16 نتائج اختيار 16000 Hz باستخدام kfold

loss	accuracy	تقسيم (Cross-validation)
%0.553	%89.84	5 مجلدات
%0.934	%93.58	10 مجلدات

الجدول 17 نتائج اختيار 16000 Hz باستخدام split

loss	accuracy	تقسيم (train_test_split)
%0.580	%86.75	التقسيم الأول
%0.610	%85.54	التقسيم الثاني

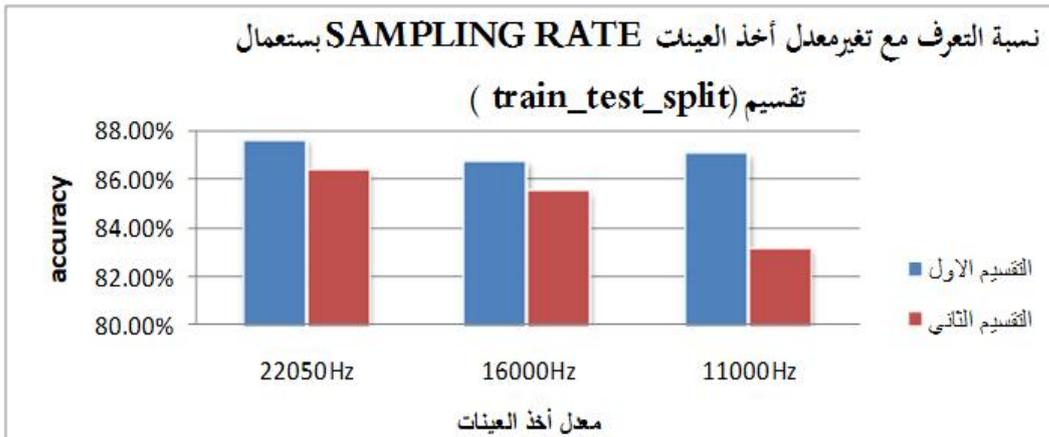
❖ نقوم باختبار 11000Hz كمعدل اختيار للعينات

الجدول 18 نتائج اختيار 11000 Hz باستخدام kfold

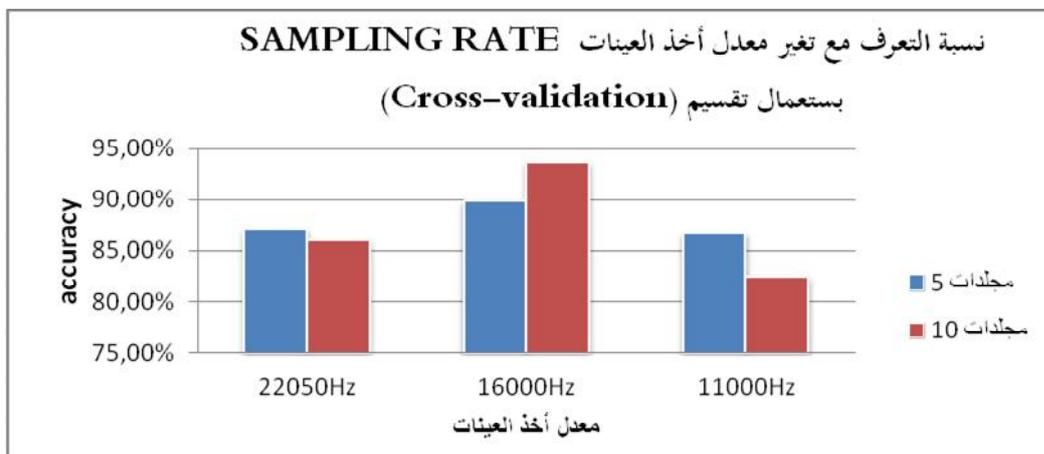
loss	accuracy	تقسيم (Cross-validation)
%0.676	%86.71	5 مجلدات
%0.759	%82.42	10 مجلدات

الجدول 19 نتائج اختيار 11000 Hz باستخدام kfold

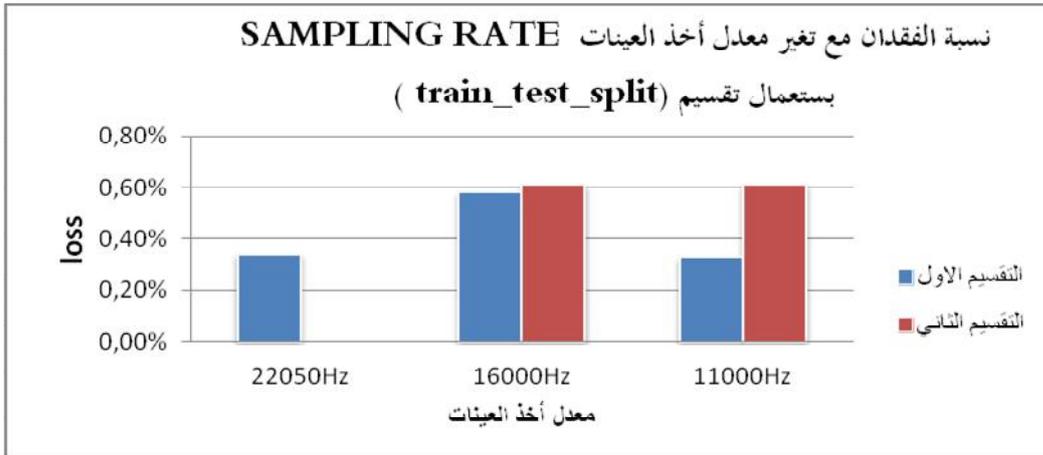
loss	accuracy	تقسيم (train_test_split)
%0.328	%87.10	التقسيم الأول
%0.612	%83.20	التقسيم الثاني



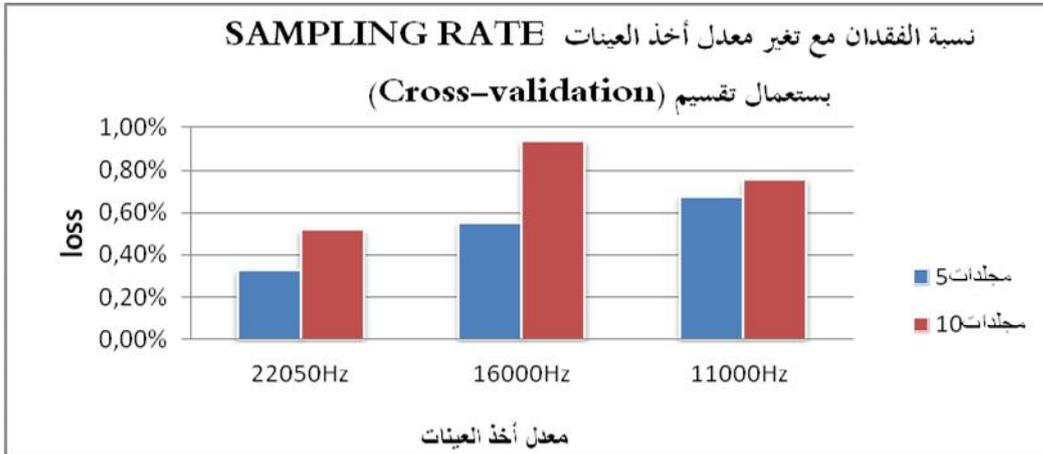
الشكل 21 نسبة التعرف مع تغير معدل اخذ العينات باستخدام split



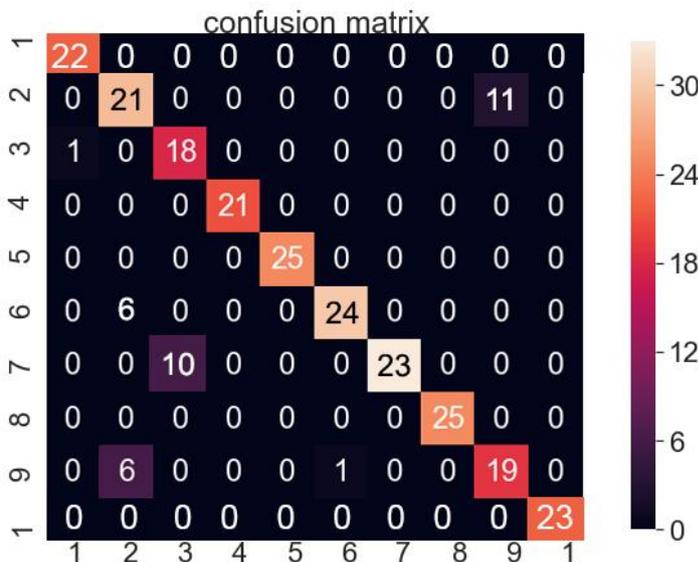
الشكل 22 نسبة التعرف مع تغير معدل اخذ العينات باستخدام kfold



الشكل 23 نسبة الفقدان مع تغير معدل اخذ العينات باستخدام تقسيم split



الشكل 24 نسبة الفقدان مع تغير معدل اخذ العينات باستخدام تقسيم kfold



الشكل 25 مصفوفة الالتباس لنتيجة التصنيف باستخدام 11000Hz بطريقة تقسيم (Cross-validation)

التجربة الثالثة

في هذه التجربة سنقوم بتغيير عدد العينات المأخوذة من المقطع الصوتي الكامل لكل قارئ منزوع الصمت مع العلم أن المدة الزمنية لكل عينة في هذا الإختبار هي 05 ثواني و معدل أخذ العينات هو 22050Hz وعدد خصائص صوت هو 13 mfcc.

❖ نقوم باختيار 128 عينة لكل قارئ

الجدول 20 نتائج اختيار 128 عينة باستعمال kfold

loss	accuracy	تقسيم (Cross-validation)
%0.676	%86.71	5 مجلدات
%0.759	%82.42	10 مجلدات

الجدول 21 نتائج اختيار 128 عينة باستعمال split

loss	accuracy	تقسيم (train_test_split)
%0.328	%87.10	التقسيم الأول
%0.612	%83.20	التقسيم الثاني

❖ نقوم باختيار 256 عينة لكل قارئ

الجدول 22 نتائج اختيار 256 عينة باستعمال kfold

loss	accuracy	تقسيم (Cross-validation)
%0.860	%84.37	5 مجلدات
%0.831	%85.74	10 مجلدات

الجدول 23 نتائج اختيار 256 عينة باستعمال split

loss	accuracy	تقسيم (train_test_split)
%0.645	%87.69	التقسيم الأول
%0.677	%86.58	التقسيم الثاني

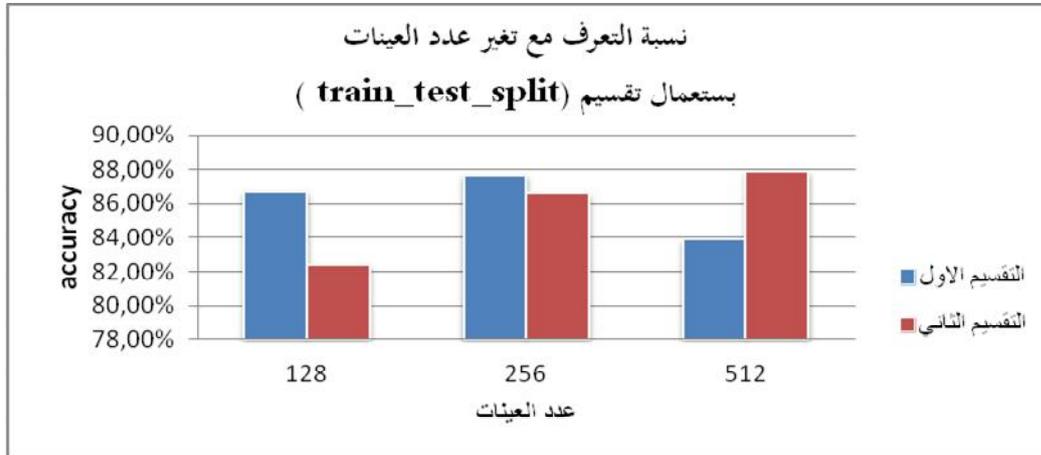
❖ نقوم باختبار 512 عينة لكل قارئ

الجدول 24 نتائج اختبار 512 عينة باستخدام **kfold**

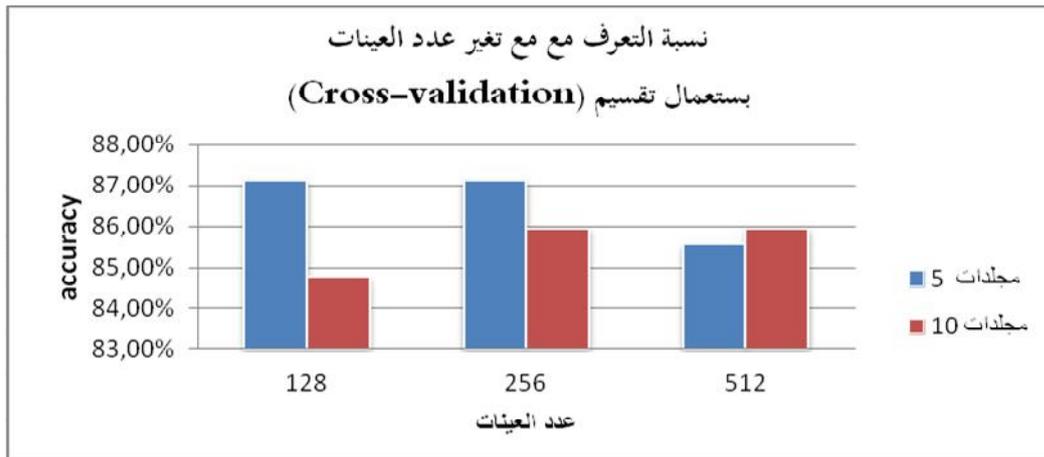
loss	accuracy	تقسيم (Cross-validation)
%1.010	%84.76	5 مجلدات
%0.678	%88.47	10 مجلدات

الجدول 25 نتائج اختبار 512 عينة باستخدام **split**

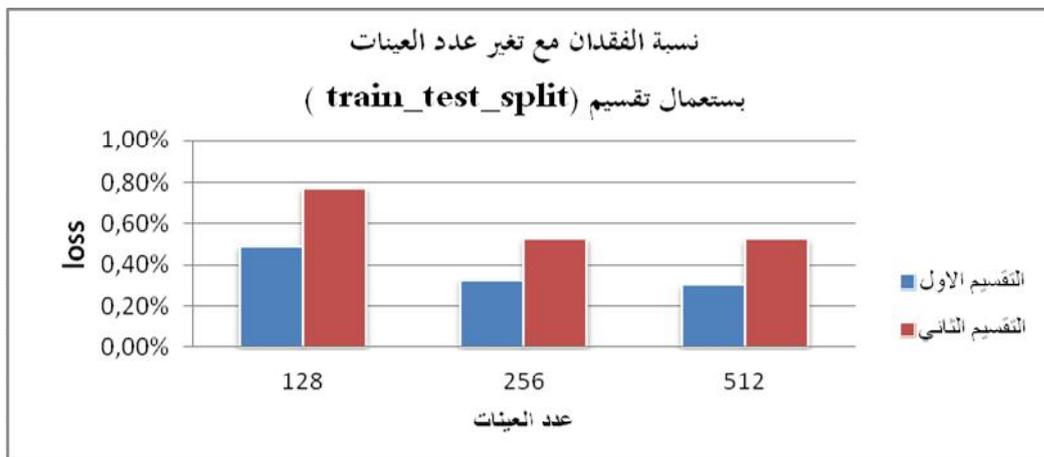
loss	accuracy	تقسيم (train_test_split)
%1.13	%83.88	التقسيم الأول
%0.837	%87.89	التقسيم الثاني



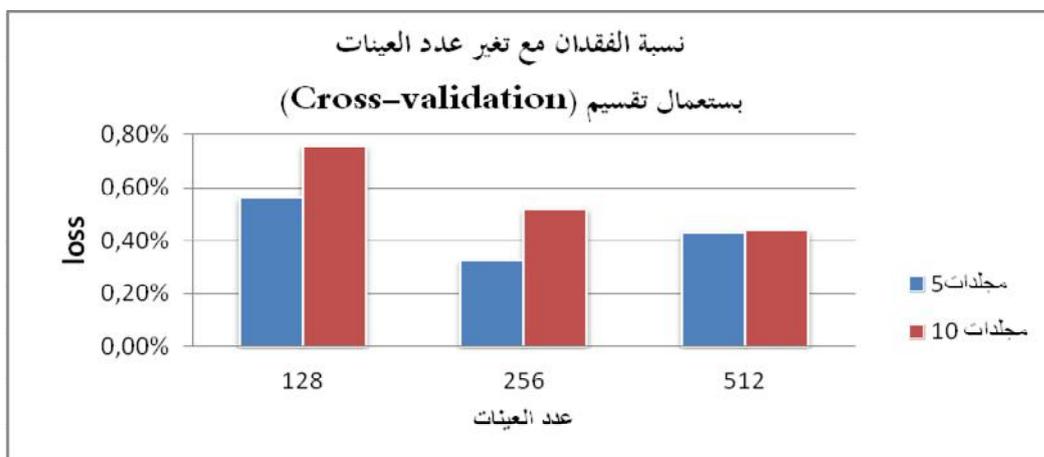
الشكل 26 نسبة التعرف مع تغير عدد العينات باستخدام تقسيم **split**



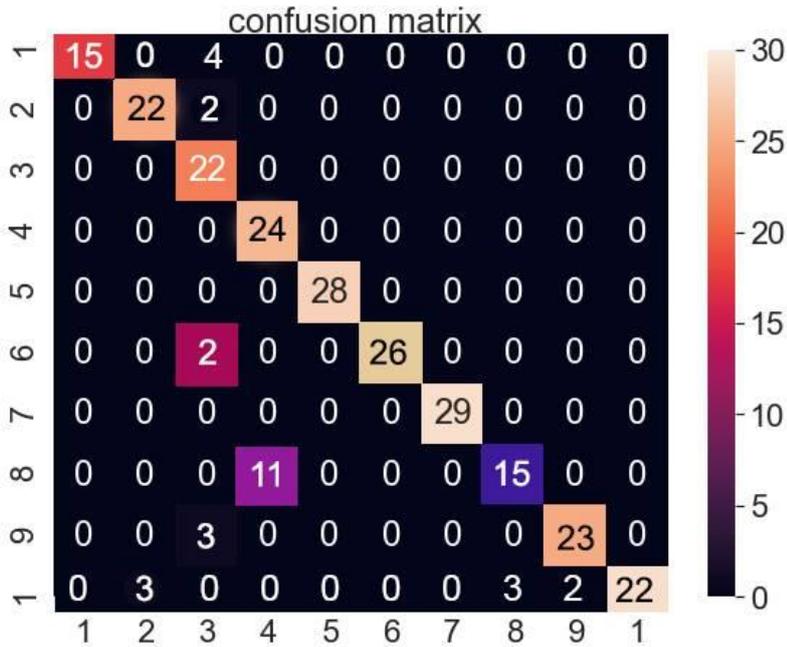
الشكل 27 نسبة التعرف مع تغير عدد العينات باستعمال تقسيم kfold



الشكل 28 نسبة الفقدان مع تغير عدد العينات باستعمال تقسيم split



الشكل 29 نسبة الفقدان مع تغير عدد العينات باستعمال تقسيم kfold



الشكل 30 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال 128 عينة لكل مقررئ

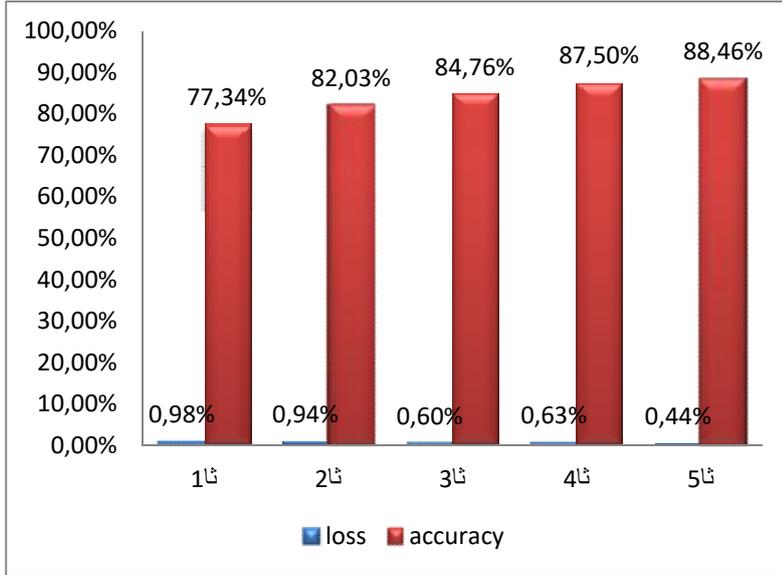
التجربة الرابعة

في هذه التجربة سنقوم بتغيير المدة الزمنية للعينات مع العلم أن عدد العينات في كل إختبار هو 128 عينة و معدل أخذ العينات هو 22050 Hz

❖ باستخدام تقسيم (Cross-validation) ل 5ملفات

الجدول 26 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم kfold

loss	accuracy	تقسيم (Cross-validation)
%0.984	%77.34	٣1
%0.937	%82.03	٣2
%0.602	%84.76	٣3
%0.626	%87.5	٣4
%0.438	%88.46	٣5

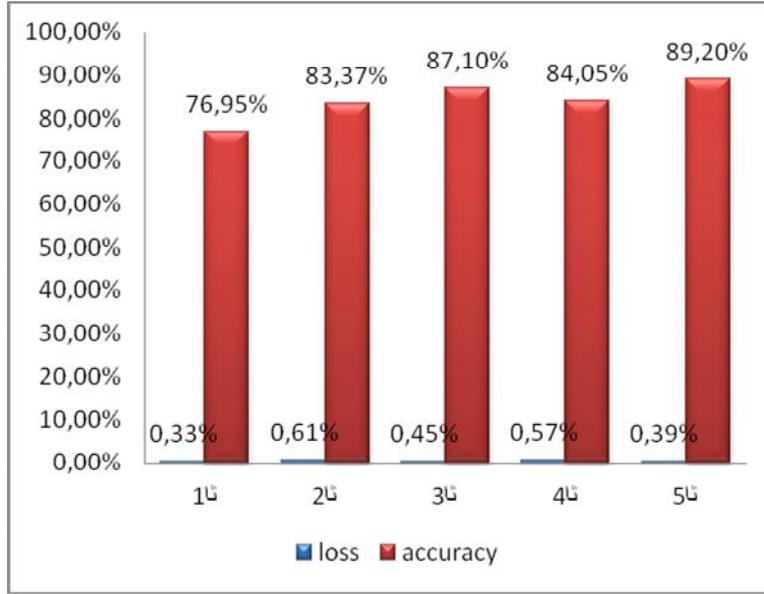


الشكل 31 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم kfold

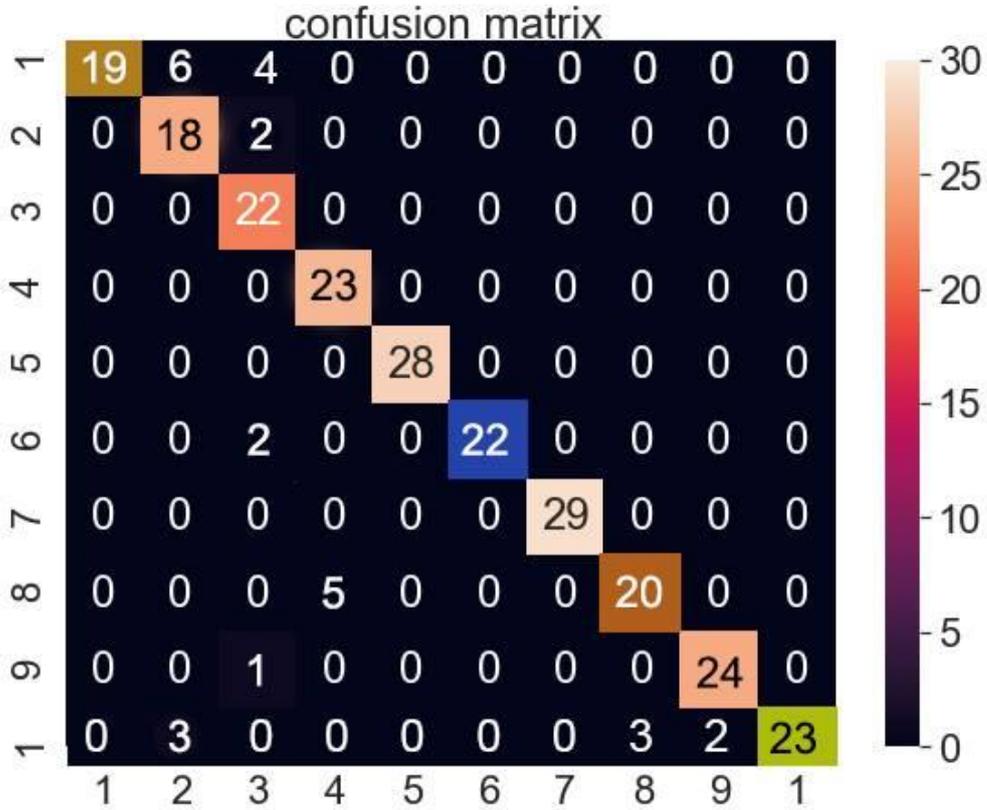
❖ باستخدام تقسيم (train_test_split)

الجدول 27 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم split

loss	accuracy	تقسيم (train_test_split)
%0.328	%76.95	تأ1
%0.612	%83.37	تأ2
%0.451	%87.10	تأ3
%0.568	%84.05	تأ4
%0.389	%89.20	تأ5



الشكل 32 نسبة التعرف والفقدان بتغيير المدة الزمنية باستعمال تقسيم split



الشكل 33 رسم توضيحي لمصفوفة الالتباس لنتيجة التصنيف باستعمال 5 ثواني

1.5 مناقشة النتائج العملية

في ما يخص التجربة الأولى الخاصة بعدد السمات في mfcc وبعد الاطلاع على النتائج المتلخصة في الجداول (4...13) تمت ملاحظة أن التغيير في عدد خصائص mfcc لم يحسن من النتائج باعتماد المصنف المقترح الشكل (20)، وبعد الاطلاع على الأعمدة البيانية الخاصة بتجربة الأولى الشكل (16..19) نلاحظ تغير طفيف في نسبة تعرف النموذج ومنه ننصح باستعمال 13 سمة لأنها تعطي نتائج مقبولة ولان الأوراق البحثية المذكورة في الفصل الأول يتم فيها استعمال 13 سمة عند استعمال mfcc .

أما في التجربة الثانية التي تم فيها اختبار هل معدل أخذ العينات (SAMPLING RATE) قد يؤثر في نسبة تعرف النموذج المقترح، وبعد الاطلاع على النتائج المتمثلة في الجداول (14...19) تمت ملاحظة أن النتائج متقاربة بدرجة كبيرة مع تغير طفيف فقط في نسبة التعرف وذلك ما تم ملاحظته الشكل (25) ومنه نستطيع أن نقول أن معدل أخذ العينات (SAMPLING RATE) لم يؤثر بشكل كبير و ملحوظ على نسبة التعرف الخاصة بالنموذج المقترح والظاهر في الأعمدة البيانية الشكل (21..24). ومنه نقترح تثبيت معدل اخذ العينات عند 16000Hz , كان الهدف من هذه التجربة هو معرفة إذا كان لمعدل اخذ سمات تأثير ملحوظ على نسبة تعرف النموذج.

في التجربة الثالثة الذي كان الهدف منها هو اختيار عدد العينات الذي يحقق أعلى نسبة تعرف في النموذج المقترح، في المرحلة الابتدائية استعملنا 128 عينة لكل مقررئ وتمت زيادتها إلى 256 عينة ثم 512 عينة وتم التحصل على النتائج المسجلة في الجداول (20..25)،

عند اخذ 128 عينة من القارئ الواحد تعطي نسبة تعرف مقبولة أما عند زيادة العينات إلى 256 أو 512 لم تتأثر نسبة تعرف بشكل كبير وملحوظ وذلك ما تم ملاحظته في الأعمدة البيانية الشكل (26..29) ومنه نقترح اخذ 128 عينة لكل قارئ من اجل تقليل حجم المصفوفة النهائية و زيادة سرعة معالجة البيانات.

أما في التجربة الأخيرة الخاصة بالمدة الزمنية للعينة الواحدة و الذي كان الهدف منها الوصول إلى المدة الزمنية التي تحقق أعلى نسبة تعرف، تم الاختبار على 1 ثا و2 ثا و3 ثا و4 ثا و5 ثا النتائج موضحة في الجدولين (26 و27).

تمت ملاحظة الشكل (33) أن أفضل نتيجة تم الوصول إليها هي عند اخذ عينات ذات 5 ثواني ومنه نقترح في الأعمال المشابهة اخذ عينات ذات 5 ثواني وذلك بعد النظر في الأعمدة البيانية لنسبة التعرف و الفقدان الشكل (31 و32).

6. الآفاق المستقبلية

قمنا في هذا البحث بدراسة أشهر خوارزمية استخلاص سمات (mfcc) وتم استعمالها لتدريب الشبكة العصبونية التلافيفية (CNN) ولكن هنالك العديد من خوارزميات استخلاص السمات التي يمكن دمجها أو إضافتها لزيادة نسبة التعرف مع استعمال أكثر من مصنف.

لم يسعنا الوقت لتوسيع مجال البحث أكثر ومنه نذكر الآفاق المستقبلية لهذا البحث :

- ❖ استخدام مصنفات أخرى في التعلم العميق أو دمج أكثر من مصنف أو استعمال التعلم الآلي .
- ❖ استخدام خوارزميات استخلاص سمات أخرى على غرار (Spectral Bandwidth) أو (Chroma feature) أو دمج أكثر من خوارزمية.
- ❖ توسيع قاعدة المعطيات لتشمل عشرات القراء الآخرين.
- ❖ العمل على دمج نظام التعرف الآلي على المتكلم مع نظام التعرف الآلي على الكلام ومنه إنتاج نظام متكامل يقوم بالتعرف على القارئ والقراء مع بعضهم البعض .

7. خاتمة

في ختام الفصل الأخير تم تحقيق النتائج التي كنا نتطلع إلى تحقيقها ، مع تنفيذ معمارية ناجحة خاصة بشبكة CNN. كان الغرض من هذا الفصل عرض الجوانب التطبيقية من عملنا ، من اقتراح معمارية تحقق الأداء المطلوب إلى التجارب ونتائج المتحصل عليها مع ذكر البرامج والبيئات المستعملة و الآفاق المستقبلية . تم عرض جميع الخطوات التي اتبعناها للوصول إلى نهاية هذا الفصل وسجلنا نسبة تعرف بلغت 89%.

الخاتمة عامة

قمنا في هذا البحث بدراسة طرق التعرف الآلي على المتكلم بإنشاء نظام للتعرف الآلي على قراء القرآن الكريم، إعتدنا على التعلم العميق في إنشائه حيث وقع اختيارنا على الشبكة العصبونية التلافيفية CNN التي تمتاز بأعلى نتائج تصنيف عند التعامل مع دخل مصفوفاتي.

يقدم هذا البحث دراسة كافية نظريا وعمليا عن الشبكة العصبونية التلافيفية والتي تم استخدامها للتعرف الآلي على قراء القرآن الكريم باستخدام خوارزمية استخراج السمات mfcc التي تمتاز بمحاكاة استجابة الأذن البشرية للإشارات الكلامية.

في هذا البحث تعرضنا لعدة عراقيل منها : محدودية أداء الأجهزة المستعملة في العمل ، نقص في الأعمال البحثية المتعلقة بموضوعنا

في ختام هذا البحث تم الحصول على نظام للتعرف الآلي على قراء القرآن الكريم تتجاوز نسبة التعرف فيه على 88% وهي نتيجة مقبولة استنادا على الأوراق البحثية ذات الصلة .

يمكن التطلع مستقبلا للوصول إلى نسب تعرف أكبر باستخدام مصنفات أخرى أو خوارزميات إستخلاص السمات أو دمج أكثر من خوارزمية كما نتطلع إلى العمل على نضام تعرف آلي على أحكام التجويد و تلقينها.

المراجع

- [1] H. Beigi, "Speaker Recognition: Advancements and Challenges," *New Trends Dev. Biometrics*, no. April, pp. 2–30, 2012, doi: 10.5772/52023.
- [2] H. Beigi, "Fundamentals of Speaker Recognition," *Fundam. Speak. Recognit.*, 2011, doi: 10.1007/978-0-387-77592-0.
- [3] VDM Verlag Dr. Müller, "Prosody in automatic speaker recognition : applications in biometrics and voice imitation Mireia Farrús," p. 216, 2010.
- [4] M. Forsberg, "Why is Speech Recognition Difficult?," *Technology*, no. March, pp. 1–10, 2003.
- [5] T. M. H. Asda, T. S. Gunawan, M. Kartiwi, and H. Mansor, "Development of Quran reciter identification system using MFCC and neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 1, no. 1, pp. 168–175, 2016, doi: 10.11591/ijeecs.v1.i1.pp168-175.
- [6] J. H. Alkhateeb, "A machine learning approach for recognizing the Holy Quran reciter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 268–271, 2020, doi: 10.14569/IJACSA.2020.0110735.
- [7] R. U. Khan, A. M. Qamar, and M. Hadwan, "Quranic reciter recognition: A machine learning approach," *Adv. Sci. Technol. Eng. Syst.*, vol. 4, no. 6, pp. 173–176, 2019, doi: 10.25046/aj040621.
- [8] K. M. O. Nahar, M. Al-Shannaq, A. Manasrah, R. Alshorman, and I. Alazzam, "A Holy Quran Reader/reciter identification system using Support Vector Machine," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 4, pp. 458–464, 2019, doi: 10.18178/ijmlc.2019.9.4.826.
- [9] M. A. Hussaini and R. W. Aldhaheeri, "An automatic qari1 recognition system," *Proc. - 2012 Int. Conf. Adv. Comput. Sci. Appl. Technol. ACSAT 2012*, pp. 524–528, 2012, doi: 10.1109/ACSAT.2012.48.
- [10] A. Qayyum, S. Latif, and J. Qadir, "Quran Reciter Identification: A Deep Learning Approach," *Proc. 2018 7th Int. Conf. Comput. Commun. Eng. ICCCE 2018*, pp. 492–497, 2018, doi: 10.1109/ICCCE.2018.8539336.
- [11] D. Luo, W. Luo, R. Yang, and J. Huang, "Compression history identification for digital audio signal," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, no. March, pp. 1733–1736, 2012, doi: 10.1109/ICASSP.2012.6288233.
- [12] A. Stan, J. Yamagishi, S. King, and M. Aylett, "The Romanian speech synthesis (RSS) corpus: Building a high quality HMM-based speech synthesis system using a high sampling rate," *Speech Commun.*, vol. 53, no. 3, pp. 442–450, 2011, doi: 10.1016/j.specom.2010.12.002.
- [13] N. S. Chauhan, "Audio Data Analysis Using Deep Learning with Python," *KDnuggets*, 2020. <https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-1.html>.
- [14] "International Association of Sound and Audiovisual Archives." <https://www.iasa-web.org/> (accessed Jul. 17, 2021).
- [15] J. Solà, A. Monin, and M. Devy, "BiCamSLAM: Two times mono is more than stereo," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 4795–4800, 2007, doi: 10.1109/ROBOT.2007.364218.

- [16] A. S. Recognition, “Automatic Speech Recognition Algorithms “خوارزميات تعرف الكلام اليا”, 2017.
- [17] K. Nyodu and S. Vijaya, *Automatic Speech Recognition of Galo*, vol. 686. 2020.