Wadi Alshatti University Journal of
Pure and Applied Sciences

مجلة جامعة وادي الشاطئ
للعلوم البحتة والتطبيقية

Volume 3, No. 1, January-June 2025 Online ISSN: 3006-0877 المجلد 3، الاصدار 1، يناير-يونيو 2025

RESEARCH ARTICLE ELECTRONIC ENGINEERING

Improving Accuracy and Efficiency of Speaker Identification Using K-means and MFCC Algorithms in Noisy Environments

AL-Salem Al-Suwaid Nasr[1,*], Ali Ukasha [2]

Department of Electrical and Electronic Engineering, Engineering Faculty, Wadi Alshatti University, Brack Alshatti, Libya

ARTICLE HISTORY
Received 13 January 2025
Revised 18 February 2025
Accepted 21 February 2025
Online 23 February 2025

KEYWORDS
K-means Algorithm;
Speaker identification;
Artificial Intelligence;
Signal to noise ratio (SNR);
Mel-frequency cepstral;
Coefficients (MFCC);
Mean Squared Error (MSE).

ABSTRACT

Speaker identification is a critical challenge in audio processing, with significant applications in security and authentication systems. Efforts focus on developing fast and efficient AI-based techniques to identify speakers using features such as pitch and frequency. A speaker recognition system consists of two main stages: feature extraction and matching. This research presents an innovative model aimed at enhancing the accuracy of speaker recognition using K-means and MFCC algorithms. The results demonstrate that the K-means algorithm reduced the error rate from 20% to 0.85%, while the MFCC features achieved an accuracy range between 80% and 99.15%. Additionally, recognition time was significantly improved, decreasing from 0.4092 seconds to 0.0438 seconds, thereby increasing the system's efficiency. Moreover, the system's performance in noisy environments was evaluated using the Signal-to-Noise Ratio (SNR), while the Mean Squared Error (MSE) metric was employed to ensure reliability and confidence in the recognition results. These findings highlight the effectiveness of the proposed algorithms and underscore the system's potential for applications in voice-controlled systems and personal assistants.

تحسين دقة وكفاءة التعرف على هوية المتحدث باستخدام خوارزميات K-means وMFCC في البيئات الصاخبة

السالم الصويد نصر[1,*]، علي عبدالرحمن عكاشة [1]

الملخص

تعد عملية التعرف على المتحدث من التحديات الرئيسية في مجال معالجة الصوت، حيث تلعب دورًا حيويًا في تطبيقات الأمان والمصادقة. تسعى الجهود البحثية إلى تطوير تقنيات تعتمد على الذكاء الاصطناعي لتحديد هوية المتحدثين بناءً على خصائص صوتية مثل النغمة والتردد. يتألف نظام التعرف على المتحدث من مرحلتين أساسيتين: استخراج الخصائص الصوتية ومقارنتها. في هذا البحث، تم تقديم نموذج مبتكر يهدف إلى تحسين دقة التعرف على الصوت باستخدام خوارزميات(K-means) و (MFCC) أظهرت النتائج أن استخدام خوارزمية (K-means) قللت معدل الخطأ من 20% إلى 0.85%، بينما حققت ميزات (MFCC) نسبة دقة تتراوح بين 80% و 99.15% كما تم تحسين سرعة التعرف بشكل ملحوظ، حيث انخفض الزمن المطلوب من 0.4092 ثانية إلى 0.0438 ثانية، مما يعزز كفاءة النظام. بالإضافة إلى ذلك، تم تقييم جودة النظام في البيئات الصاخبة من خلال حساب نسبة الإشارة إلى الضوضاء (SNR)، بينما ساهم استخدام متوسط الخطأ التربيعي (MSE) معيار لتقييم الجودة في زيادة الثقة في نتائج التعرف. تؤكد هذه النتائج فعالية الخوارزميات المقترحة وإمكانية تطبيق النظام في مجالات متعددة، مثل أنظمة التحكم الصوتي والمساعدات الشخصية.

الكلمات المفتاحية

خوارزمية (K-means)
التعرف على الصوت
الذكاء الاصطناعي
نسبة الإشارة إلى الضوضاء (SNR)
معامل الطيف الترددي (MFCC)
متوسط مربع الخطأ (MSE)

المقدمة

تعد أنظمة التعرف على الصوت من أبرز تطبيقات الذكاء الاصطناعي وتعلم الآلة في العصر الحالي، حيث تساهم في تحسين التفاعل بين الإنسان والآلة في مجالات متعددة كالأمن، والاتصالات، والمساعدات الصوتية الذكية. يهدف النظام المقترح للتعرف على المتحدث إلى تحقيق تحسينات جوهرية في دقة التعرف على الصوت وسرعة الاستجابة، مما يجعله ملائمًا للاستخدام في بيئات حقيقية تتسم بالتحديات، مثل الضوضاء العالية أو الحاجة إلى أداء عالٍ.

يركز النظام على تعزيز قدراته من خلال معالجة دقيقة للإشارات الصوتية واستخدام تقنيات مبتكرة لتحليل البيانات واستخراج الخصائص الصوتية الفريدة. كما يهدف إلى توفير حلول أمنية فعالة تعتمد على التعرف على المتحدث كوسيلة لتحديد الهوية، مما يعزز من موثوقية النظام ويتيح إمكانية تطبيقه في مجالات متعددة، مثل أنظمة التحكم الصوتي والمساعدات الشخصية.

الصوت هو إشارة تتكون من نغمة أو مجموعة من النغمات المتتالية التي تحمل معانٍ محددة، وتستخدم كوسيلة للتواصل بين البشر والكائنات الحية

*Corresponding author: al.abdallah@wau.edu.ly
https://dooi.org/

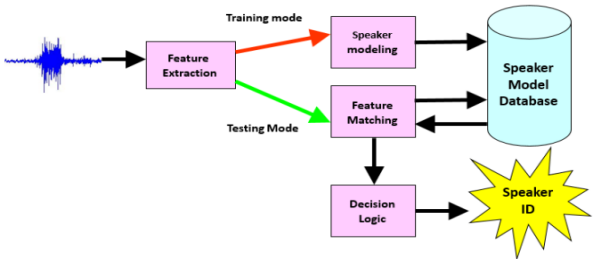
للتعبير عن الأفكار والرغبات، ويعرف الإحساس الناتج عن هذه الموجات بالسمع. تلعب الأصوات دورًا حيويًا في إثراء تجارب البشر، إذ كانت في الماضي تتنوع بين الأصوات الصادرة من الحنجرة وأصوات أدوات أخرى مثل الطبول والمزامير التي تصدر ذبذبات تُستخدم للتواصل. تبلغ سرعة الصوت في الهواء 343 مترًا في الثانية، وتعتمد على كثافة الوسط الذي تنتقل عبره [11,10,1]. يتمثل التغير المستمر للإشارات الصوتية بمرور الوقت في تغييرات صغيرة متتابعة، مما يتطلب معالجة دقيقة لفهم محتوى الصوت وللتعرف على الصوت، حيث يميز بين التعرف على الصوت الذي يحدد هوية المتحدث، والتعرف على الكلام الذي يحدد المحتوى المنطوق، وغالبًا ما يُستخدم مصطلح "التعرف على الصوت" للإشارة إلى كلا العمليتين. يعد التحقق من هوية المتحدث جزءًا رئيسيًا من تطبيقات الأمان، حيث يعتمد النظام المقترح على تحويل الصوت إلى بيانات رقمية، مما يسهم في تحسين دقة التحقق واستجابة النظام [2,1] .

تهدف الأنظمة الصوتية الحديثة إلى تخصيص إجراءات الأمان لتلائم احتياجات المستخدمين، ويشمل ذلك إنشاء ملفات شخصية دقيقة تعتمد على تحليل صوت المستخدم كما يظهر في الشكل (1).

ويعتمد هذا النظام على تقنيات الذكاء الاصطناعي، التي تمكن النظام من تعزيز قراراته عبر تحليل البيانات الصوتية، مما يسهم في تطوير أساليب أمنية أكثر تطورًا وموثوقية [2].

ومن أجل التعرف على المتحدث في نظام تحديد المتحدث، يجب أن تمر الإشارة بعدة مراحل، كالتالي:

1. تسجيل ومعالجة الإشارة.
2. استخراج الخصائص.
3. مطابقة الخصائص. كما هو موضح بالشكل (1).



الشكل 1: الهيكل الأساسي لنظام تحديد هوية المتحدث.

تعد معالجة الإشارات الصوتية والتحليل السريع للبيانات الصوتية عناصر حيوية في تطوير أنظمة التحقق من الهوية، حيث يجب أن تمر الإشارة بعدة مراحل لتحقيق التعرف الدقيق والسريع. تعتمد الأنظمة الحديثة للتعرف على المتحدث على دمج تقنيات معالجة الصوت والذكاء الاصطناعي، مما يجعلها ملائمة لبيئات العمل الحقيقية التي تتطلب أداءً عاليًا واستجابة سريعة، وتؤدي هذه التقنيات دورًا مهمًا في تعزيز التفاعل بين الإنسان والآلة، خصوصًا في مجالات الأمن، والمساعدات الصوتية الذكية، والاتصالات [3,2].

يعتمد النظام المطور على تقنيات حديثة لتحليل الإشارات الصوتية بفعالية. تبدأ العملية بتسجيل الصوت وتحويله إلى صيغة رقمية باستخدام مكتبة Sound Device، تليها استخراج معاملات (MFCC) باستخدام مكتبة (Librosa)، تعتبر معاملات (MFCC) من أهم الخصائص التي تمثل الإشارات الصوتية، حيث تعكس السمات الترددية الفريدة لأصوات المستخدمين، مما يساعد النظام في التعرف على الخصائص الصوتية الخاصة بكل فرد.

بعد ذلك، يتم تدريب نموذج التكميم المتجهي (VQ) باستخدام خوارزمية (K-means) لإنشاء (Code Book) مخصص لكل متحدث، تتيح هذه التقنية ضغط بيانات الصوت، مما يسهل المقارنة والتصنيف، كما يتم تسجيل نمط صوتي فريد لكل مستخدم يتم استخدامه لاحقًا في عملية المطابقة والتحقق من الهوية. يتم تخزين (Code Book) في قاعدة بيانات باستخدام مكتبة (Pickle)، مما يضمن الوصول السريع والآمن إلى بيانات المستخدمين عند الحاجة [15,14].

تستعرض هذه الورقة كيفية دمج هذه التقنيات لإنشاء نظام أمني يعتمد على بصمة الصوت للتعرف على المستخدمين، مما يعكس كفاءة النظام في تحقيق مستوى عالٍ من الأمان ويعزز من موثوقية الحلول الأمنية المعاصرة.

المنهجية

تقدم الدراسة الحالية نظامًا شاملاً للتعرف على الصوت يجمع بين تقنيات معالجة الإشارات المتقدمة وخوارزميات التعلم الآلي لتعزيز قدرات التعرف على المتحدث. يوضح الشكل (2) مراحل بناء هذا النظام، بدءًا من تسجيل الصوت وتحويله إلى تمثيل رقمي، مرورًا باستخراج الميزات باستخدام معامل الترددات (MFCC)، ثم استخدام خوارزمية K-Means لتدريب نموذج (VQ). يتم بعد ذلك تخزين البيانات المستخرجة لاستخدامها في عملية التعرف، حيث يتم تصنيف المدخلات الجديدة بناءً على النموذج المدرب، وأخيرًا عرض النتائج وفقًا لعملية المطابقة.

تسجيل الصوت

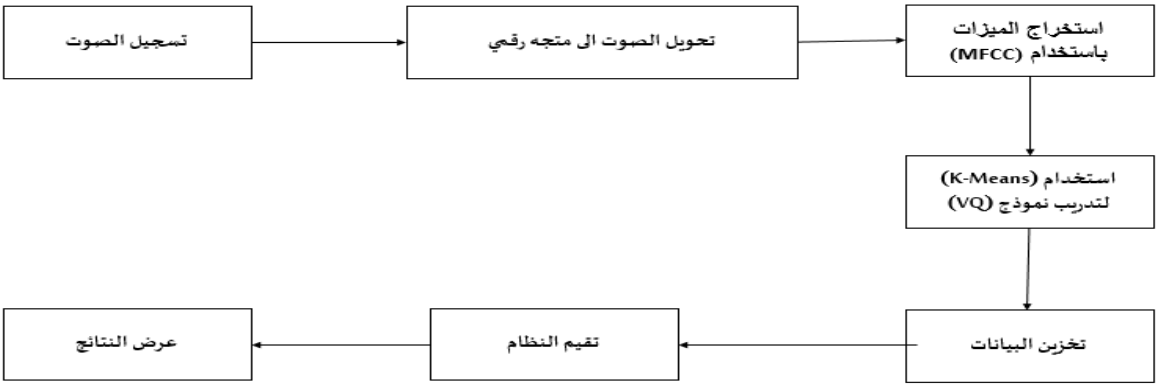
تبدأ العملية بالتقاط الإشارة الصوتية الأصلية، وهي ببساطة الصوت الخام الذي يتم تسجيله مباشرة باستخدام ميكروفون أو جهاز تسجيل. هذه الإشارة الصوتية تتكون من موجات صوتية عادية، كما يمكن أن تشمل الضجيج المحيط وأي أصوات أخرى غير مرغوبة.

تحويل الإشارة إلى إشارة رقمية

بعد تسجيل الصوت، يتم تحويله من شكل تناظري إلى شكل رقمي، وهي عملية تعرف بالتكميم. يتم هذا التحويل عن طريق أخذ عينات من الإشارة الصوتية وتحويل كل عينة إلى قيمة رقمية، وبالتالي يمكن تمثيل الموجة الصوتية كسلسلة من الأرقام. هذا يسمح للحاسوب بالتعامل مع الصوت بشكل مباشر ومعالجته باستخدام خوارزميات مختلفة[15] [3] [7] [11] .

استخراج الميزات (MFCC)

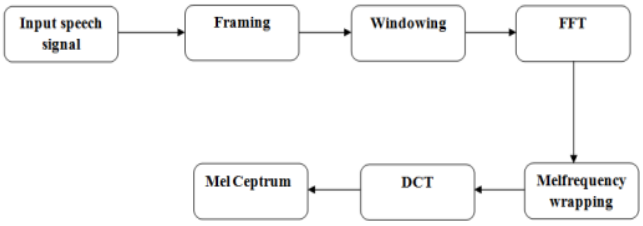
في هذا النظام بمجرد تحويل الصوت إلى شكل رقمي، يتم تطبيق تقنية معاملات (MFCC) وهي واحدة من أكثر الطرق شيوعًا في معالجة الصوت. في هذه الخطوة، يتم تحويل الإشارة الصوتية إلى مجموعة من الميزات الصوتية التي تعكس الخصائص الفريدة للصوت مثل النغمة، القوة، والترددات الرئيسية التي تميز أصوات الأشخاص المختلفين (MFCC) يقوم بتجزئة الصوت إلى أجزاء زمنية صغيرة ويطبق تحليل طيفي على كل جزء للحصول على تمثيل رقمي لهذه الخصائص. في معالجة الصوت، تُستخدم معاملات تردد (MFCC) لتمثيل طيف الطاقة الصوتي بناءً على تطبيق تحويل جيب



الشكل 2: المخطط العام للتعرف على الصوت

التمام الخطي على مقياس تردد غير خطي يُعرف بمقياس (Mel)، وتتألف (MFCC)من مجموعة من القيم التي تُعبّر عن هذا الطيف بتوزيع ترددي يتناسب مع مقياس ميل، مما يجعلها تختلف عن الطيف العادي حيث تكون الترددات موزعة بشكل متساوي على مقياس ميل(Mel).

يساعد هذا التوزيع على تحسين دقة تمثيل الصوت، وهو ما يُفيد في تطبيقات مثل ضغط الصوت. يتم اشتقاق (MFCC)بشكل عام وفق الخطوات التالية كما هو موضح بالشكل (3) [11] .



الشكل 3: خطوات اشتقاق معاملات MFCC

بالنظر إلى أن قيم (MFCC)ليست فعالة جدًا في وجود ضوضاء إضافية، لذا يُفضل توحيدها في أنظمة التعرف على الصوت لتقليل تأثير الضوضاء. يقترح بعض الباحثين تعديلات على خوارزمية (MFCC)، مثل زيادة سعة لوغاريتم مقياس (Mel) قبل تطبيق تحويل جيب التمام المتقطع (DCT) لتحسين المتانة وتقليل تأثير العناصر ذات الطاقة المنخفضة. يتم الاحتفاظ بعدد محدود من عينات (DCTحوالي 12 إلى 14) بينما يتم تجاهل الباقي والشكل (4) يوضح المخطط الانسيابي لكتلة(MFCC).

يتم تحويل التردد إلى قيمة متناسبة مع مقياس (Mel) والعكس باستخدام المعادلتين التاليتين [9,5,4].

1-مقياس (Mel):

$$m(f) = \ln(1 + (\frac{f}{700})) \quad (1)$$

2- التحويل العكسي لمقياس ميل (Inverse Mel scale transformation):

$$M^{-1}(m) = 700(\exp\left(\frac{m}{1125}\right) - 1) \quad (2)$$

استخدام K-Means لتدريب نموذج VQ

نموذج (Vector Quantization)

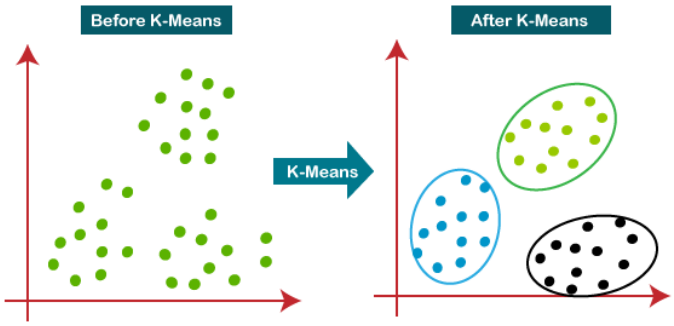
الشكل 4: المخطط الانسيابي لكتلة (MFCC).

تقنية (VQ)التكميم المتجه تُعتبر إحدى الأساليب المستخدمة في تقليل الأبعاد في معالجة الإشارات، وتجد استخداماً واسعاً في مجال التعرف على الصوت. تهدف هذه التقنية إلى تقليص حجم البيانات اللازمة لتمثيل الإشارة الصوتية مع الحفاظ على المعلومات الأساسية. آلية العمل: تستند (VQ)إلى تقسيم الفضاء العددي الذي تمثله ميزات الصوت، مثل (MFCC)، إلى عدة مناطق (Clusters) تُحدد بواسطة مراكز (Code words)، يتم تعيين كل نقطة بيانات إلى أقرب مركز، مما يسهل تقليل حجم البيانات المستخدمة أثناء عملية التعرف.

خوارزمية (K-Means)

(K-Means) هي خوارزمية فعالة ومستخدمة على نطاق واسع لتقليل الأبعاد في معالجة الإشارات، مثل ما تم استخدامه في النظام الحالي. (VQ) تساعد هذه الخوارزمية على تقسيم البيانات إلى مجموعات، مما يسمح بتقليل كمية البيانات التي تُستخدم في تمثيل الإشارة الصوتية، مع الحفاظ على المعلومات الأساسية. تعد (K-Means) خطوة أساسية في تدريب نموذج (VQ)، حيث تحدد المراكز التي تُستخدم لتقليل البيانات وتسهيل عملية التعرف على الصوت. وفي سياق عملية التجميع باستخدام خوارزمية (K-Means)، يقدم الشكل (5) تمثيلًا بصريًا لنتائج تطبيق الخوارزمية على مجموعة بيانات معينة، الصورة الأولى تُظهر حالة البيانات قبل تطبيق (K-Means). حيث

تكون نقاط البيانات موزعة بشكل عشوائي دون أي تشكيلات واضحة. أما الصورة الثانية، فتوضح نتائج عملية التجميع، حيث نجحت الخوارزمية في تنظيم البيانات إلى مجموعات متميزة، مما يعكس قدرتها على تحليل البيانات واستخراج الأنماط [15,9].



الشكل 5: مقارنة بصرية لتجميع البيانات باستخدام (K-Means)

تخزين البيانات (تحديد المتحدث)

تحديد المتحدث هو مرحلة حرجة في نظام التعرف على الصوت، حيث يتم خلالها تحديد هوية المتحدث من خلال تحليل الميزات الصوتية المستخرجة. سنستعرض هنا بشكل مفصل كيفية تنفيذ هذه الخطوة وأهمية المعايير المستخدمة فيها.

1-حساب المسافات

بعد استخراج ميزات (MFCC) من العينة الصوتية الجديدة، يتم حساب المسافة الإقليدية بين هذه الميزات وميزات المتحدثين المخزنة في قاعدة البيانات. المسافة الإقليدية هي مقياس يُستخدم لقياس الفرق بين نقطتين في الفضاء. في سياق التعرف على المتحدث، تُعبر المسافة الإقليدية عن مدى اختلاف الميزات الصوتية للعينة الجديدة عن ميزات كل متحدث في قاعدة البيانات. حيث يتم حساب المسافة الاقليدية حسب المعادلة التالية [10].

$$D = \sum_{j=1}^{n} \sqrt{(x_i - y_i)^2} \quad (3)$$

حيث x_i تمثل ميزات العينة الجديدة و y_i تمثل ميزات المتحدثين في قاعدة البيانات. كلما كانت المسافة أقل، زادت درجة التشابه بين العينة الجديدة والمتحدث في قاعدة البيانات.

2- تحديد المتحدث المعترف به

بعد حساب المسافات لكل المتحدثين، يتم تحديد المتحدث الذي حقق أقل مسافة إقليدية. يُعتبر هذا المتحدث هو المعترف به، حيث يعني ذلك أن الميزات الصوتية لعينة الاختبار تتشابه بشكل كبير مع ميزات المتحدث. هذه الخطوة تعتمد على مفهوم أن كل متحدث له بصمة صوتية فريدة، وبالتالي، إذا كانت الميزات قريبة من بعضها، فهذا يشير إلى أن العينة الجديدة تخص المتحدث نفسه.

تقييم النظام

تقييم النظام هو خطوة حاسمة في تطوير أي نظام للتعرف على الصوت، حيث يتم من خلالها قياس كفاءة النظام في أداء المهمة المحددة. إليك كيفية تقييم نظام التعرف على المتحدث بشكل شامل:

أولا: تحديد معايير الأداء

تتطلب عملية تقييم نظام التعرف على الصوت مجموعة من المعايير أو مؤشرات الأداء التي تعكس فعالية النظام. من بين هذه المعايير التي تم استخدامها في النظام الحالي [11,6,4,1].

1-دقة التعرف (Recognition Accuracy)

تشير إلى النسبة المئوية للحالات التي تم التعرف عليها بشكل صحيح بالنسبة إلى إجمالي عدد الحالات التي تم اختبارها.

$$RA = \left(\frac{\text{Number of Correct Recognitions}}{\text{Total Number of Tests}}\right) * 100 \quad (4)$$

2-معدل الخطأ (Error Rate)

$$ER = \left(\frac{\text{Number of Errors}}{\text{Total Number of Tests}}\right) * 100 \quad (5)$$

3-متوسط نسبة الإشارة إلى الضوضاء (SNR)

$$SNR = 10Log\left(\frac{\text{Total Signal Power}}{\text{Total Noise Powe}}\right) dB \quad (6)$$

4-زمن استخراج الميزات (Feature Extraction Time – MFCC)

$$FET = \text{Time After Extraction} - \text{Time Before Extraction} \quad (7)$$

5-زمن التعرف (Recognition Time)

يحسب زمن التعرف من خلال قياس الوقت المستغرق في التعرف بعد استخراج الميزات.

$$RT = \text{Ti After Recognition} - \text{Ti After Feature Extraction} \quad (8)$$

6- متوسط الخطأ التربيعي(MSE)

يتم حساب متوسط الخطأ التربيعي بقياس الفرق المتوسط بين القيم المتوقعة والقيم الفعلية، ويُستخدم بشكل شائع في قياس دقة التوقعات أو الأداء في نماذج التحليل. يُحسب (MSE) باستخدام القانون التالي [14]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad (9)$$

حيث أن: x_iتمثل القيمة الفعلية للعنصر i و y_i تمثل القيمة المتوقعة للعنصر i ، N تمثل عدد العناصر (البيانات).

حيث انه كلما كانت قيمة (MSE) أقل، كان التطابق بين القيم الفعلية والمتوقعة أقرب، مما يعني أداءً أفضل للنظام [14,1].

ثانيا: تصميم التجارب:

في النظام الحالي تم تقسيم مجموعة البيانات إلى مجموعتين رئيسيتين: مجموعة مخصصة للتدريب وأخرى للاختبار. حيث تحتوي هذه البيانات على تنوع في أصوات المتحدثين وظروف التسجيل، بما في ذلك اختلاف مستويات الضوضاء الخلفية. بلغت بيانات التسجيل حوالي 350 عينة صوتية، تم تخصيص 250 عينة منها للتدريب و100 عينة للاختبار.

الدراسات السابقة

هناك العديد من الدراسات السابقة التي لها صلة بهذا النظام، ومن اهم هذه الدراسات التي تم من خلالها تصميم هذا النظام هي:

1- في عام 2023 قام كلا من H, Zhang واخرون بدراسة لتحسين أداء ميزات (MFCC) في أنظمة الطاقة الذكية من خلال تعديل عدد المرشحات وترتيبها، حيث يؤثر الترتيب بشكل أكبر على دقة التعرف مقارنةً بعدد المرشحات. أظهرت النتائج أيضًا قدرة النموذج على تحقيق معدل لدقة التعرف يصل إلى 85% حتى مع SNR = 0، مما يعكس فعالية النموذج في ظروف ضوضاء صعبة. وفي ظروف الضوضاء العالية، مما يبرز فعاليته في مواجهة التحديات البيئية.

2- وفي عام 2021 قام كلا من M, AlDujaili واخرون بدراسة الضوء على تحسين نظام التعرف على المشاعر الكلامية من خلال معالجة خصائص الصوت مثل التردد الأساسي والطاقة ومعدل العبور الصفري، باستخدام خوارزميات تحليل المكونات الرئيسية (PCA) وتقنيات

التصنيف مثل (SVM) و (KNN) وأظهرت النتائج التجريبية تحقيق دقة تصل إلى 90.83% عند استخدام قاعدة بيانات اللغة الإنجليزية، مما يعكس فعالية النموذج في التعرف على المشاعر بدقة عالية وسرعة تنفيذ جيدة.

3- اما في عام 2021 قام كلا من H,Abdullah واخرون بتصميم نظام يهدف الي اكتشاف المحتالين باستخدام تقنيات تعلم الآلة المختلفة لتحديد أفضل مجموعة تناسب التعرف على المتحدثين، حيث تم تطبيق أساليب معالجة الصوت مثل تقليل الضوضاء وتعزيز الصوت، واستخراج معاملات (MFCC) ، مما أدى إلى تحقيق دقة بلغت 97.9% باستخدام خوارزمية الغابة العشوائية.

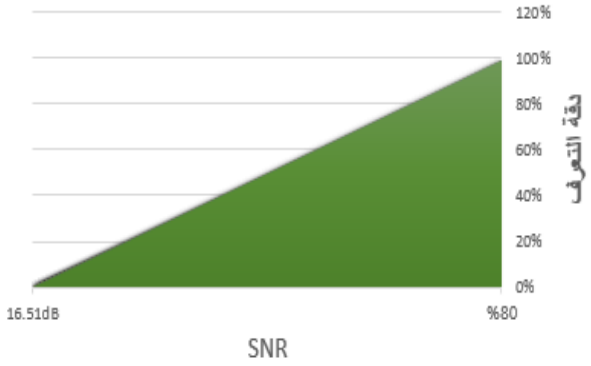
4- وفي عام 2022 قام كلا من Y,Albagory واخرون بتطوير نظام تحويل الكلام إلى نص باستخدام الشبكات العصبية التلافيفية (CNN)لتعرف إشارات الكلام النغمية الخاصة بترانيم الغوربراني، حيث حقق النظام دقة تصل إلى 89.15%مع معدل خطأ في الكلمات يبلغ10.65%.

النتائج والمناقشة

في هذه الورقة، تم تقييم كفاءة نظام التعرف القائم على تقنيات الذكاء الاصطناعي من خلال تحليل مجموعة متنوعة من المؤشرات، بما في ذلك دقة التعرف، معدل الخطأ، نسبة الإشارة إلى الضوضاء (SNR)، زمن استخراج الميزات، وزمن التعرف. تهدف هذه النتائج إلى تقديم رؤية شاملة حول فعالية النظام في معالجة البيانات في سياقات تطبيقية متعددة. تشير النتائج إلى تحسين ملحوظ في دقة التعرف، حيث ارتفعت من 80% إلى 99.15 % بينما انخفض معدل الخطأ من 20%إلى 0.85% هذا التحسن يعكس كفاءة النظام في تقليل الأخطاء على مدار فترة الاختبار للمتحدث. وهذا مما يدل على فعالية الأساليب المستخدمة في معالجة البيانات والتعلم الآلي. بالإضافة إلى ذلك، لوحظ تحسن كبير في زمن استخراج الميزات وزمن التعرف، حيث انخفض زمن التعرف إلى 0.0438s مع الحفاظ على مستوى عالٍ من الدقة.

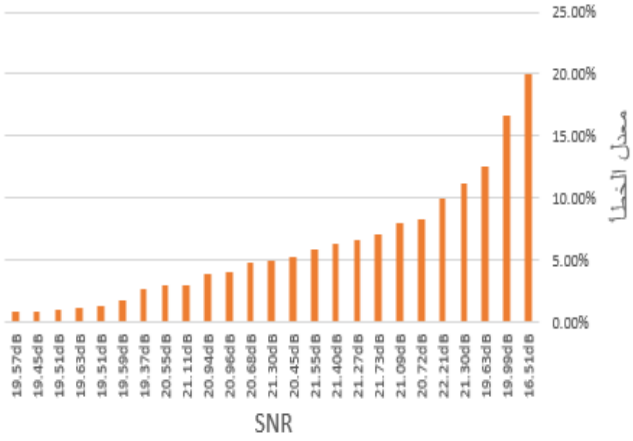
هذه النتائج تدل على قدرة النظام على تحقيق استجابة سريعة وفعالة، مما يعكس نجاح تقنيات التعلم الالي المستخدمة في استخراج الميزات وتحليلها. كما سجلت نسبة الإشارة إلى الضوضاء (SNR)تحسنًا ملحوظًا، لمجموعة قيم تراوحت بين 16.51dBو 19.37dB يشير هذا التحسن إلى جودة البيانات المدخلة، وهو عامل حاسم في قدرة النظام على الأداء الجيد. بشكل عام، توفر هذه النتائج دليلاً قويًا على كفاءة وفاعلية النظام في التعرف على الأنماط، مما يعزز إمكانية تطوير وتحسين مثل هذه الأنظمة في المستقبل. تعكس هذه الإنجازات التقدم المستمر في هذا المجال وتفتح آفاق جديدة لتطبيقات عملية متنوعة.

يوضح الشكل (6) العلاقة البيانية بين دقة التعرف و (SNR). نلاحظ من المخطط أن زيادة قيمة (SNR)تؤدي إلى تحسين ملحوظ في دقة التعرف. فعندما تزداد قيمة (SNR)من 16.51dB الى 19.37dB، نلاحظ أن دقة التعرف تزداد تدريجيًا من حوالي 80% إلى ما يقارب 99.15% هذا يعني أن النظام يصبح أكثر دقة في التعرف على الإشارات أو الصور مع ارتفاع نسبة الإشارة إلى الضوضاء. هذه العلاقة تعكس تأثير جودة الإشارة على أداء النظام.



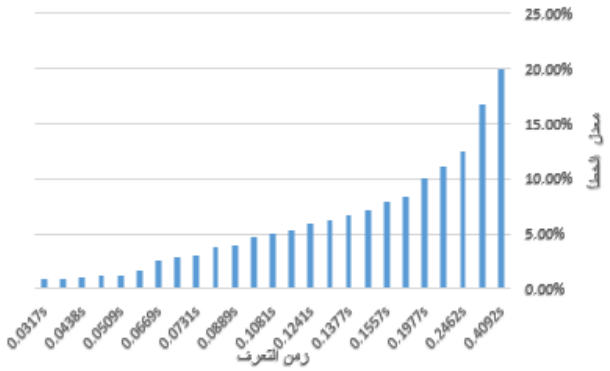
الشكل 6: العلاقة بين دقة التعرف ونسبة الإشارة الى الضوضاء.

حيث تساعد المستويات الأعلى من (SNR) في تقليل التأثير السلبي للضوضاء، مما يؤدي إلى تحسين دقة التعرف. يتيح هذا المخطط فهماً واضحاً لكيفية تأثير التغيرات في (SNR) على أداء النظام من حيث دقة التعرف، ويظهر أن تحسين جودة الإشارة يمكن أن يكون عاملاً أساسياً لتحسين أداء النظام في البيئات ذات الضوضاء العالية، ومن خلال تحليل العلاقة الموضحة في الشكل (7)، يظهر أن هناك علاقة عكسية بين معدل الخطأ ونسبة الإشارة إلى الضوضاء (SNR)فعند زيادة قيمة (SNR)من 16.51dB الى 19.37dB، نلاحظ انخفاضًا ملحوظًا في معدل الخطأ من حوالي 20%إلى أقل من 1%. يشير هذا الانخفاض المستمر إلى أن تحسين جودة الإشارة من خلال زيادة (SNR)يقلل بشكل كبير من معدل الخطأ في أداء النظام. بعبارة أخرى.



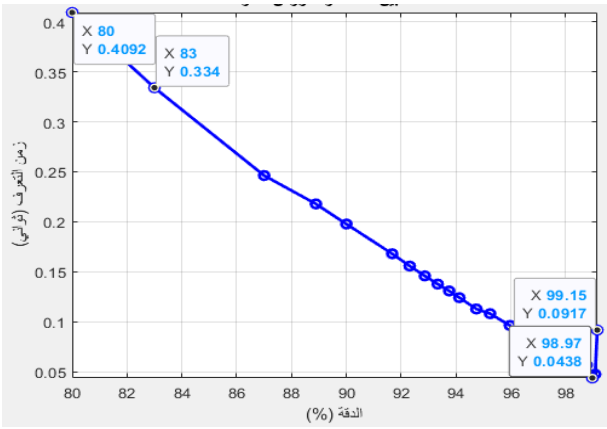
الشكل 7: العلاقة بين معدل الخطأ وSNR

فكلما كانت الإشارة أقوى وأكثر وضوحًا مقارنة بالضوضاء المحيطة، أصبح النظام أقل عرضة للأخطاء. يقدم المخطط تمثيلاً بصريًا واضحًا لهذا التغيير التدريجي، مما يسهم في فهم كيفية تأثر معدل الخطأ بتغيرات نسبة SNR هذه العلاقة تبرز أهمية جودة الإشارة كعامل أساسي في تحسين دقة النظام، مما يمكن أن يكون ذا فائدة كبيرة عند تصميم الأنظمة التي تتطلب دقة عالية في بيئات تحتوي على ضوضاء. وفقًا للعلاقة التالية في الشكل (8) يظهر أن هناك علاقة عكسية بين زمن التعرف ومعدل الخطأ. فعندما يقل زمن التعرف من 0.4092s إلى 0.0317s، ينخفض معدل الخطأ بشكل مطرد من حوالي20% إلى أقل من1%.



الشكل 8: العلاقة بين معدل الخطأ وزمن التعرف

تشير هذه العلاقة العكسية إلى أنه كلما أصبح زمن التعرف أقصر، يقل معدل الخطأ في أداء النظام بشكل ملحوظ. بعبارة أخرى، تحسين سرعة التعرف يؤدي إلى تقليل الأخطاء، مما يعكس كفاءة النظام في تقديم نتائج دقيقة خلال فترة زمنية أقل. يوفر المخطط تمثيلًا مرئيًا واضحًا لكيفية تغير معدل الخطأ بتغير زمن التعرف، مما يساعد على فهم التوازن بين سرعة المعالجة ودقة النظام. هذا الفهم يمكن أن يكون ذا قيمة كبيرة في التطبيقات التي تتطلب توازنًا دقيقًا بين السرعة والدقة لضمان أداء موثوق وفعال في البيئات العملية، الشكل يوضح (9) العلاقة بين دقة التعرف (%) وزمن التعرف (بوحدات عشوائية). مع زيادة دقة التعرف من 80%إلى حوالي99.15%، ينخفض زمن التعرف بشكل حاد من حوالي0.4092s إلى 0.0317s.



الشكل 9 : العلاقة بين الدقة و زمن التعرف

تشير هذه العلاقة العكسية إلى أنه كلما زادت دقة النظام في التعرف، تقل المدة الزمنية المطلوبة لإجراء عملية التعرف بشكل ملحوظ. يوفر الرسم البياني تصورًا واضحًا لهذا التوازن بين الدقة وسرعة المعالجة، مما يساعد على فهم خصائص الأداء للنظام وتحديد كيفية تحسين التوازن بين الدقة والسرعة وفقًا لمتطلبات التطبيق. تمكّن هذه النقاط البيانية من قياس هذه العلاقة، مما يسهم في توجيه عملية تصميم وضبط أنظمة التعرف لتحقيق الأداء الأمثل الذي يلبي احتياجات التطبيق بشكل فعال. وتعتبر مؤشرات الأداء جزءًا أساسيًا من تقييم فعالية أي نظام تعرف أو معالجة صوتية. في هذا القسم، سيتم مقارنة نتائج النظام الحالي مع نتائج الدراسات السابقة في المجال، وذلك لتحديد نقاط القوة والضعف في الأداء. كما هو موضح في الجدول (1)، وتتضمن المقارنة تحليل الدقة ومعدل الخطأ في التعرف، بالإضافة إلى تأثير جودة الإشارة SNRعلى أداء النظام. كما تتناول المقارنة زمن استخراج الميزات باستخدام معامل الترددات (MFCC)، من خلال استعراض النتائج المتاحة في الأدبيات العلمية. الأداء جزءًا أساسيًا من تقييم فعالية أي نظام تعرف أو معالجة صوتية. يمكن تقديم رؤية شاملة حول مدى تطور تقنيات التعرف على الصوت وفعالية الأساليب المختلفة المستخدمة. سيتم استخدام هذا التحليل لتقييم مدى قدرة النظام الحالي على مواجهة التحديات وكذلك معرفة كيفية تحسين الأداء في الدراسات المستقبلية. من خلال هذه المقارنة، يمكن تسليط الضوء على التحسينات التي يحققها النظام الحالي مقارنة بالدراسات السابقة، بالإضافة إلى تحديد التحديات التي قد تؤثر على دقته وأدائه العام. يساعد هذا التحليل في تقديم رؤية واضحة حول مدى تطور تقنيات التعرف على الصوت، وتحديد الجوانب التي يمكن تحسينها في المستقبل لتعزيز كفاءة أنظمة التعرف الصوتي، خاصةً في البيئات ذات الضوضاء العالية أو عند التعامل مع بيانات متنوعة. تعكس هذه النتائج تحقيق الأهداف الرئيسة للنظام المتمثلة في تحسين دقة التعرف على الصوت وسرعة الاستجابة، إلى جانب تعزيز فعاليته في البيئات الواقعية ذات الضوضاء المتفاوتة. كما تدعم هذه المؤشرات رؤية النظام كحل متطور يعتمد على تقنيات الذكاء الاصطناعي لتوفير أنظمة أمان موثوقة وقابلة للتطبيق في مجموعة واسعة من المجالات.

الاستنتاجات

في هذا الجانب من الورقة، يتم تقديم الاستنتاجات الرئيسية المستخلصة من تطبيق الخوارزميات المختلفة في نظام التعرف على الصوت، والتي تم تحليل أدائها ونتائجها بناءً على الكود المستخدم.

تلخص الاستنتاجات التالية الفوائد والنجاحات التي حققها النظام، مما يعكس مدى فعاليتها وكفاءتها في تحسين دقة التعرف على الصوت وزمن الاستجابة. إن نتائج هذا البحث تساهم في تعزيز الفهم العام حول كيفية تحسين أنظمة التعرف على الصوت، مما يمهد الطريق لتطبيقات عملية أكثر تقدمًا في هذا المجال. حيث أنه أثبتت الخوارزميات المستخدمة في نظام التعرف على الصوت فعاليتها البارزة وتأثيرها الإيجابي المباشر على الأداء العام ونتائج النظام. حيث أسهمت خوارزمية K-means بشكل حاسم في تحسين دقة التعرف، حيث انخفضت نسبة الخطأ من 20% إلى 0.85%بفضل تجميع الأصوات المتشابهة، مما يعكس قدرة النظام الفائقة على التكيف مع تنوع الأصوات وظروف التسجيل المختلفة. علاوة على ذلك، كانت ميزات MFCC حاسمة في تحقيق دقة تعرف تتراوح بين 80%الى 99.15%، مما مكن النظام من تمييز الأصوات بدقة عالية حتى في البيئات المتغيرة. كما أظهرت النتائج أهمية زمن التعرف، حيث تم تقليله بشكل كبير من 0.4092s إلى 0.0317s، مما يزيد من فعالية النظام في التطبيقات التي تتطلب استجابة فورية. يُعد تحسين زمن التعرف عاملاً أساسيًا في رفع كفاءة النظام، مما يجعله مناسبًا للاستخدام في بيئات تتطلب استجابة سريعة ودقيقة. علاوة على ذلك، يعد حساب نسبة الإشارة إلى الضوضاء SNR ضروريًا لتقييم جودة النظام في بيئات التسجيل الصاخبة. حيث تشير قيم SNR العالية إلى قدرة النظام على التمييز بين الإشارة المفيدة والضوضاء الخلفية، مما يعزز من جودة التعرف على الصوت ويقلل من الأخطاء الناتجة عن الضوضاء. تعكس هذه النتائج تحقيق الأهداف الرئيسة للنظام المتمثلة

في تحسين دقة التعرف على الصوت وسرعة الاستجابة، إلى جانب تعزيز فعاليته في البيئات الواقعية ذات الضوضاء المتفاوتة. كما تدعم هذه المؤشرات رؤية النظام كحل متطور يعتمد على تقنيات الذكاء الاصطناعي لتوفير أنظمة أمان موثوقة وقابلة للتطبيق في مجموعة واسعة من المجالات.

التوصيات

1-تحسين دقة التعرف باستخدام تقنيات دمج الميزات (Feature Fusion): دمج ميزات MFCC مع ميزات إضافية مثل PLP (Perceptual Linear Prediction) أو LPC (Linear Predictive Coding) لتعزيز الأداء، حيث يمكن أن يؤدي هذا الدمج إلى تحسين التمييز بين الأصوات المتشابهة.

2- استخدام خوارزميات تحسين الزمن الحقيقي: تطبيق تقنيات مثل Real-Time Streaming Models باستخدام مكتبات TensorFlow Lite أو ONNX لتحسين أداء النظام في الزمن الحقيقي، مما يجعله أكثر فعالية في التطبيقات الفورية.

3- استخدام التعلم العميق لتحليل الميزات: دمج نموذج يعتمد على Deep Feature Extraction باستخدام شبكات مثل CNN أو LSTM لتحليل أنماط الصوت بدقة أكبر، خاصة في البيئات الصاخبة.

4- دمج تحسينات تقنيات الضغط: استخدام تقنيات ضغط بيانات صوتية مثل Linear Predictive Coding (LPC) لتقليل حجم البيانات وتحسين سرعة الاستجابة دون التضحية بالدقة.

5- دمج نظام متعدد الوسائط لتعزيز الأمان: يُنصح بتوسيع النظام ليشمل طرق تعريف إضافية مثل التعرف على الوجه أو التعرف على الإيماءات بجانب الصوت لتعزيز الأمان والموثوقية في البيئات الحساسة.

Author Contributions: "All authors have made a substantial, direct, and intellectual contribution to the work and approved it for publication."
Funding: "This research received no external funding."
Data Availability Statement: "The data are available at request."
Conflicts of Interest: "The author declares no conflict of interest."

References

[1] S. Dua, et al., "Developing a speech recognition system for recognizing tonal speech signals using a convolutional neural network," *Applied Sciences*, 12(12): 6223, 2022.
[2] A. Ali, H. Abdullah, and M. Fadhil, "Voice recognition system using machine learning techniques," *Materials Today: Proceedings*, 20: 1-7, 2021.
[3] M. Al Dujaili, A. Moghadam, and A. Fatlawi, "Speech emotion recognition based on SVM and KNN classifications fusion," *International Journal of Electrical and Computer Engineering*, 11(2): 1259, 2021.
[4] H. Zhang, et al., "Feature extraction of speech signal based on MFCC (Mel cepstrum coefficient)," *Journal of Physics: Conference Series*, 2584(1): IOP Publishing, 2023.
[5] V. Prasad, "Voice recognition system: speech-to-text," *Journal of Applied and Fundamental Sciences*, 1(2): 191, 2015.
[6] D. Prabakaran, and S. Sriuppili, "Speech processing: MFCC based feature extraction techniques-an investigation," *Journal of Physics: Conference Series*, 1717(1): IOP Publishing, 2021.
[7] M. Algabri, et al., "Automatic speaker recognition for mobile forensic applications," *Mobile Information Systems*, 6986391, 2017.
[8] S. Dua, et al., "Developing a speech recognition system for recognizing tonal speech signals using a convolutional neural network," *Applied Sciences*, 12(12): 6223, 2022.
[9] M. Ahmed, R. Seraj, and S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, 9(8): 1295, 2020.
[10] Z. Tu, et al., "A feature fusion model with data augmentation for speech emotion recognition," *Applied Sciences*, 13(7): 4124, 2023.
[12] Z. Li, "Feature extraction optimization method in speaker recognition system," *Journal of Xiamen University (Natural Science)*, 59(6): 995-1003, 2020.
[11] A. Sewunet, An Ensemble of Vector Quantization and CNN for Digital Based Text Independent Amharic Language Speaker. Recognition, Ph.D. dissertation, 2023.
[13] S. Bai, X. Yan, and S. Zhang, "Low signal-to-noise ratio speech endpoint detection based on Meir frequency cepstrum coefficient and short time energy," *Journal of Nanjing Normal University (Natural Sciences)*, 44(2): 117-120, 2021.
[14] Q. Li, et al., "MSP-MFCC: Energy-efficient MFCC feature extraction method with mixed-signal processing architecture for wearable speech recognition applications," *IEEE Access*, 8: 48720-48730, 2020.
[15] R. Hidayat, "Frequency domain analysis of MFCC feature extraction in children's speech recognition system," *Jurnal Infotel*, 14(1): 30-36, 2022.

الجدول 1: مقارنة أداء نموذج التعرف بين النظام الحالي والدراسات السابقة

مؤشرات الأداء	النظام الحالي	النظام الألماني الجديد 2021	النظام الانجليزي الجديد2021	[4]	[3]	[2]	[1]
دقة التعرف	99.15%	87.85%	90.83%	85%	92.21%	97.9%	89.15%
معدل الخطأ	0.85%	12.15%	9.17%	15%	7.79%	2.1%	10.56%
زمن استخراج الميزات (MFCC)	0.0127s	#####	#####	#####	#####	#####	#####
زمن التعرف	0.0317s	0.48S	0.35s	#####	#####	#####	#####
SNR	16.15dB	#####	#####	0dB	#####	#####	#####