Modeling Temporal Relationships in Audio Signals Using MFCC And GRU to Enhance Voice Recognition Accuracy

Mahmoud Muhammad [*]
Dr. Ali Darwisho [**]
Dr. Fadi Mutawaj [***]

(Received 25 / 2 / 2026. Accepted 4 / 6 / 2026)

□ ABSTRACT □

Voice recognition is a prominent research trend in signal processing, given its growing role in intelligent interaction applications and biometric systems. This research presents a methodological framework that combines signal spectral characterization using Mel-Frequency Cepstral Coefficients (MFCC) with deep learning models based on Gated Recurrent Units (GRUs). GRUs are advanced models for temporal data analysis due to their ability to represent sequential dependencies with high efficiency and lower computational complexity compared to traditional recurrent models. The methodology begins by converting the audio signal into a compressed spectral representation that reflects its frequency structure. MFCCs provide an accurate description of the underlying acoustic characteristics. GRUs then utilize this representation to extract temporal variations within the signal, relying on a gating mechanism that enables them to retain relevant information over time and process long sequences without loss of context.

Experimental results showed that integrating MFCC with GRU networks significantly improved the performance of voice recognition systems, particularly in environments with time fluctuations or high noise levels. The proposed framework also demonstrated its ability to adapt to speaker differences and diverse acoustic environments.

keywords: MFCC-GRU-Deep learning.

[*]**Ph.D. Student- - Faculty of Science- Latakia University(Formerly Tishreen) - Latakia- Syria. mhmoud.mhmad@latakia-univ.edu.sy**
****Professor - Faculty of Science- Latakia University(Formerly Tishreen) - Latakia- Syria. Dr.darwisho@gmail.com**
*****Associate Professor- Faculty of Mechanical and Electrical Engineering– Latakia University(Formerly Tishreen) - Latakia- Syria. fadimotawej@latakia-univ.edu.sy**

نمذجة العلاقات الزمنية في الإشارات الصوتية باستخدام MFCC وشبكات GRU لتعزيز دقة التعرف على الصوت

محمود محمد [*]

د .علي درويشو [**]

د .فادي متوج [***]

(تاريخ الإيداع 25 / 2 / 2026. قبل للنشر في 4 / 6/ 2026)

□ ملخّص □

يعد التعرف على الصوت أحد الاتجاهات البحثية البارزة في مجال معالجة الإشارات، نظراً لدوره المتنامي في تطبيقات التفاعل الذكي، والأنظمة البيومترية. يقدم البحث إطار منهجي يجمع بين استخلاص الخصائص الطيفية للإشارة باستخدام معاملات ميل (MFCC) Mel–Frequency Cepstral Coefficients وبين نماذج التعلم العميق القائمة على وحدات البوابة الدورية(GRU) Gated Recurrent Unit ، التي تُعد من النماذج المتقدمة في تحليل البيانات الزمنية بفضل قدرتها على تمثيل الاعتماديات المتتابعة بكفاءة عالية وبمستوى تعقيد حسابي أقل مقارنة بالنماذج المتكررة التقليدية. تبدأ المنهجية بتحويل الإشارة الصوتية إلى تمثيل طيفي مضغوط يعكس بنيتها الترددية، حيث توفر معاملات MFCC وصفاً دقيقاً للخصائص الصوتية الأساسية. بعد ذلك، تستفيد شبكات GRU من هذا التمثيل لاستخلاص الأنماط الزمنية المتغيرة داخل الإشارة، معتمدة على آلية البوابات التي تمكّنها من الاحتفاظ بالمعلومات ذات الصلة عبر الزمن ومعالجة التتابعات الطويلة دون فقدان السياق.

أظهرت النتائج التجريبية أن دمج MFCC مع شبكات GRU أسهم في تحسين أداء أنظمة التعرف على الصوت بشكل ملحوظ، خصوصاً في البيئات التي تتسم بتقلبات زمنية أو مستويات مرتفعة من الضجيج. كما أثبت الإطار المقترح قدرته على التكيف مع اختلافات المتحدثين وتنوع البيئات الصوتية.

الكلمات المفتاحية: GRU– MFCC–التعلم العم

[*] **طالب دكتوراه– كلية العلوم– جامعة اللاذقية (تشرين سابقاً) اللاذقية– سورية. mhmoud.mhmad@latakiauniv.edu.sy**

[**] **أستاذ– كلية العلوم – جامعة اللاذقية (تشرين سابقاً)–اللاذقية– سورية..Dr.darwisho@gmail.com**

[***] **أستاذ مساعد– كلية الهندسة الميكانيكية والكهربائية– جامعة اللاذقية(تشرين سابقاً) – اللاذقية– سورية.**

fadimotawej@latakia–univ.edu.sy

مقدمة:

شهد مجال معالجة الإشارات الصوتية خلال السنوات الأخيرة تطوراً ملحوظاً بسبب الطلب المتزايد على أنظمة قادرة على تحليل الإشارة الصوتية وفهمها وتوظيفها في تطبيقات عملية تشمل المساعدات الذكية، وأنظمة التفاعل الصوتي، وتقنيات الأمن البيومتري. يعتمد التعرف على الصوت بصورة جوهرية على جودة تمثيل الإشارة الصوتية واستخلاص خصائصها البنيوية والزمنية بدقة وكفاءة، بما يضمن قدرة الأنظمة على التمييز بين الأنماط الصوتية المختلفة في البيئات الواقعية[1].

وتُعد تقنية **MFCC** من أكثر تقنيات استخلاص الميزات اعتماداً في مجال معالجة الإشارة الصوتية، نظراً لقدرتها على محاكاة آلية السمع البشري وتوفير تمثيل طيفي مضغوط يعكس البنية الترددية للإشارة الصوتية. وقد أثبتت الدراسات أن MFCC توفر أساساً قوياً لبناء نماذج التعرف على الصوت والمتحدث، سواء في البيئات الهادئة أو المليئة بالضجيج، بفضل قدرتها على التقاط السمات الصوتية الأكثر دلالة[2] .

ومع التقدم المتسارع في تقنيات التعلم العميق، برزت الشبكات العصبية المتكررةRNN وخاصة **نموذج GRU** كأحد النماذج الفعّالة في التعامل مع الطبيعة الزمنية المتتابعة للإشارة الصوتية. تتميز GRU ببنية مبسّطة مقارنة بالنماذج المتكررة التقليدية، مع احتفاظها بقدرة عالية على نمذجة الاعتماديات الزمنية طويلة المدى، مما يجعلها مناسبة لالتقاط الأنماط الديناميكية داخل الإشارة الصوتية دون زيادة كبيرة في التعقيد الحسابي[3].

أهمية البحث وأهدافه:

تكمن أهمية البحث من الحاجة إلى تطوير نماذج أكثر دقة وموثوقية قادرة على التعامل مع التحديات العملية، مثل مستويات الضجيج المرتفعة، وتنوع المتحدثين، والتغير المستمر في البيئات الصوتية.يركز هذا البحث على تعزيز جودة استخلاص الميزات الصوتية باستخدام معاملات **MFCC**، لما توفره من تمثيل طيفي مضغوط يعكس الخصائص البنيوية للإشارة. كما يعتمد على توظيف **شبكات GRU** بوصفها أحد النماذج المتقدمة في تحليل البيانات الزمنية، نظراً لقدرتها على نمذجة العلاقات الديناميكية داخل الإشارة بكفاءة عالية وبمستوى تعقيد حسابي أقل مقارنة بالنماذج المتكررة التقليدية. ويهدف هذا التكامل إلى تحسين أداء أنظمة التعرف على الصوت من خلال الاستفادة من قوة MFCC في تمثيل الطيف الصوتي، وقدرة GRU على التقاط الأنماط الزمنية المتتابعة.

طرائق البحث ومواده:

يعتمد البحث على منهجية تجريبية تهدف إلى معالجة الإشارات الصوتية وتقييم أداء النماذج في بيئات متنوعة تحاكي الظروف الواقعية. وقد شملت منهجية البحث عدداً من الخطوات المترابطة، بدءاً من جمع البيانات وانتهاءً بتحليل النتائج.

في مرحلة جمع البيانات، تم استخدام مجموعة بيانات صوتية معيارية تضم تسجيلات متعددة لمتحدثين مختلفين، مع تنوع في المحتوى اللفظي والبيئات الصوتية، إضافة إلى تضمين عينات تحتوي على مستويات متفاوتة من الضجيج لتعزيز واقعية الاختبارات. أما في مرحلة معالجة الإشارة، فقد تم تطبيق معاملات MFCC لاستخلاص الخصائص الطيفية الأساسية، وذلك عبر تقسيم الإشارة إلى إطارات زمنية، وتطبيق نافذة هامنغ، وحساب الطيف، ثم تحويله إلى مقياس ميل واستخلاص المعاملات النهائية.

وفيما يتعلق بتصميم النموذج العميق، تم بناء نموذج يعتمد على الشبكات العصبية المتكررة من نوعLSTM ، نظراً لقدرتها على نمذجة العلاقات الزمنية داخل الإشارة الصوتية. شمل تصميم النموذج تحديد عدد طبقاتLSTM ، وعدد الوحدات في كل طبقة، إضافة إلى ضبط المعاملات الفائقة مثل معدل التعلم وحجم الدفعة التدريبية لتحقيق أفضل أداء ممكن. بعد ذلك، تم تدريب النموذج باستخدام خوارزمية الانتشار العكسي عبر الزمن(BPTT) ، مع تقسيم البيانات إلى مجموعات تدريب واختبار لضمان تقييم موضوعي للأداء. وقد تم اعتماد مقاييس تقييم متعددة مثل الدقة، ومعدل الخطأ، ومصفوفة الالتباس لقياس فعالية النموذج.

وفي المرحلة الأخيرة، تم تحليل النتائج من خلال مقارنة أداء النموذج المقترح مع نماذج تقليدية تعتمد على خوارزميات مثل HMM أو ميزات طيفية بديلة، بهدف تحديد مدى التحسن الذي يحققه دمج معاملاتMFCC مع نموذج LSTM في مهام التعرف على الصوت.

خوارزمية MFCC

تُعدّ معاملات الترددات الميليّة (Mel-Frequency Cepstral Coefficients – MFCC) من أكثر الأساليب استخداماً وموثوقية في تحليل الإشارات الصوتية، نظراً لقدرتها على تمثيل الخصائص الطيفية للإشارة بطريقة تتوافق مع آلية إدراك الأذن البشرية للترددات. تعتمد هذه الخوارزمية على تحويل الإشارة الصوتية الخام إلى مجموعة من المعاملات التي تعكس البنية الطيفية الأساسية، مما يجعلها مناسبة لتطبيقات التعرف على الكلام والمتحدث في مختلف البيئات[4] .

1 التقنيات الأساسية في خوارزميةMFCC

تعتمد خوارزمية MFCC على مجموعة من التقنيات المتتابعة التي تهدف إلى تحويل الإشارة من شكلها الزمني الخام إلى تمثيل طيفي مضغوط وذي دلالة عالية، وتشمل[5] :

1. **تحويل فورييه المتقطع(DFT)** :يُستخدم لتحويل الإشارة من المجال الزمني إلى المجال الترددي، مما يسمح بتحليل مكونات التردد المختلفة داخل الإشارة.
2. **مقياس الميل(Mel Scale)** :يعتمد على نماذج سيكوأكوستيكية تهدف إلى تقريب الترددات بما يتوافق مع حساسية الأذن البشرية، حيث يتم توزيع المرشحات بشكل غير خطي يعكس إدراك الإنسان للصوت.
3. **حساب المعاملات الطيفية** :بعد تطبيق مرشحات الميل، يتم استخراج المعاملات الطيفية التي تمثل الخصائص الجوهرية للطيف الصوتي، والتي تُعدّ الأساس في بناء ميزات فعّالة للتعرف على الإشارة.

2 مزايا خوارزميةMFCC

تتميز خوارزمية MFCC بعدد من الخصائص التي جعلتها معياراً أساسياً في معالجة الإشارات الصوتية، من أبرزها:

1. **الفعالية** :تُعدّ MFCC من أكثر الطرق قدرة على استخلاص السمات المميزة للإشارة الصوتية، مما يجعلها مناسبة لمختلف تطبيقات التعرف.
2. **الدقة** :توفر الخوارزمية تمثيلاً دقيقاً للبنية الطيفية للإشارة، مما ينعكس إيجاباً على أداء النماذج المعتمدة عليها.
3. **البساطة** :تتميز ببنية حسابية بسيطة نسبياً وسهلة التنفيذ، مما يسهم في اعتمادها على نطاق واسع في الأنظمة العملية.[6]

3 مراحل خوارزميةMFCC

تمر خوارزمية MFCC بعدة مراحل متتابعة تهدف إلى تحويل الإشارة الصوتية إلى مجموعة من المعاملات القابلة للاستخدام في النماذج التحليلية، وتشمل:

1. **تحويل الإشارة من المجال الزمني إلى المجال الترددي :**يتم تقسيم الإشارة إلى إطارات زمنية قصيرة، ثم تطبيق تحويل فوريه المتقطع (DFT) للحصول على الطيف الترددي الذي يمثل مكونات الإشارة من حيث التردد والمطال.
2. **تطبيق مقياس الميل :**يُمرّر الطيف عبر مجموعة من المرشحات الموزعة وفق مقياس الميل، وهو مقياس سيكوأكوستيكي يعكس حساسية الأذن البشرية للترددات المختلفة.[7]
3. **حساب المعاملات الطيفية :**بعد الحصول على طيف الميل، يتم تطبيق اللوغاريتم ثم التحويل الجيبي المنفصل (DCT)لاستخلاص معاملات MFCC النهائية. تمثل هذه المعاملات الخصائص الأكثر ارتباطاً بإدراك الكلام البشري، مع تجاهل المعلومات الأقل أهمية، مما يجعلها مناسبة لمهام مثل التعرف على المتحدث وتحويل الكلام إلى نص[8] .

تحويل فوريه المتقطع DFT:

يُعدّ تحويل فوريه المتقطع أداة رياضية أساسية في مجال معالجة الإشارات. يُستخدم هذا التحويل لتحويل إشارة من النطاق الزمني إلى النطاق الترددي. يُمكن تمثيل الإشارة في النطاق الزمني كمجموعة من النقاط، بينما تُمثل الإشارة في النطاق الترددي كمجموعة من الترددات والمطالات.

يتمّ تعريف تحويل فوريه المتقطع لإشارة زمنية محددة x[n]، حيث n = 0, 1, 2, ..., N-1، بالصيغة التالية[9]:

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-j\frac{2\pi}{N}kn\right)$$

حيث:

X[k]هو طيف الترددات في النطاق الترددي. x[n]هي الإشارة في النطاق الزمني, Nهو عدد العينات في الإشارة, k هو تردد التردد, jهي الوحدة التخيلية

تطبيق تقنية MEL

الإشارة الصوتية من الإشارات المتغيرة مع الزمن , لذلك لابد من تقطيع الإشارة إلى اطر , وليكن بطول N نقطة , اقصر من ms25 تقريبا لضمان استقرار الإشارة على كامل الاطار , نفضل الضرب بنافذة hamming لتخفيف حدة الانقطاعات بين الأطر لضمان نتائج افضل , لذلك وللتعويض عن تخميد المطالات على الأطراف نأخذ نوافذ متداخلة بمقدار M عينة تكون من مرتبة نصف طول عينات الاطار N , تعطى عبارة نافذة hamming بالعلاقة التالية[10]

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

حيث تمثل N عدد عينات الإطار ,

تكون إشارة الناتج جداء الاطار بالنافذة هي :

بعد الضرب بالنافذة يتم تطبيق تحويل فوريه السريع FFTلايجاد فوريه المتقطع DFT لكل اطار , ونبقي النصف الأول من الإشارة الناتجة (الموافق للترددات الموجبة من الإشارة لأنها ستكون متناظرة كون إشارة الدخل حقيقة) . بالتالي نكون حصلنا على الطيف الموافق من اجل كل اطار , لكن هذا الطيف يحوي الكثير من المعلومات التي لن نحتاجها من اجل مرحلة مطابقة السمات . لذلك نوزع ترددات الطيف إلى مجموعات قليلة لنرى كمية الطاقة المتواجدة

ضمن كل مجموعة .تتم هذه العملية معيارياً بضرب طيف كل اطار بمجموعة مرشحات تكون على شكل مثلثات فلاتر mel [11].

التعلم العميق والشبكات العصبية المتكررة (LSTM) في التعرف على الصوت

شهد مجال التعرف على الصوت تطوراً كبيراً مع ظهور تقنيات التعلم العميق، التي تجاوزت قدرات النماذج الإحصائية التقليدية مثل **HMM** و **GMM** بفضل قدرتها على التعلم التلقائي للميزات واستخلاص الأنماط المعقدة من الإشارات الصوتية. تعتمد هذه النماذج على بنية متعددة الطبقات قادرة على تمثيل البيانات بطريقة هرمية، حيث تُستخلص السمات الأولية في الطبقات السطحية، بينما تُلتقط العلاقات الزمنية العميقة في الطبقات المتقدمة، مما يمنحها قدرة عالية على التعامل مع التغيرات الزمنية وتنوع المتحدثين.[12]

تنفيذ إطار العمل المقترح

من اجل تحقيق إطار العمل تم دمج التمثيل الطيفي المضغوط الذي توفره MFCC مع قدرة GRU على نمذجة العلاقات الزمنية داخل الإشارة الصوتية، مما يتيح بناء نموذج قادر على تتبع تطور الخصائص الصوتية عبر الزمن بكفاءة عالية , ويمر اطار العمل بالخطوات التالية

1 مراحل معالجة الإشارة الصوتية

الجدول 1: المراحل التفصيلية لمعالجة الإشارة الصوتية واستخلاص ميزاتMFCC

المرحلة	الوصف العلمي	الهدف	المخرجات
جمع البيانات الصوتية	تسجيل عينات متعددة لمتحدثين مختلفين ببيئات متنوعة	ضمان تنوع البيانات وتحسين التعميم	ملفات صوتية خام
إزالة الضجيج	تطبيق مرشحات مثل Spectral Subtraction أو Wiener Filtering	تحسين نسبة الإشارة إلى الضجيج	إشارة صوتية أنقى
التقسيم إلى إطارات	تقسيم الإشارة إلى إطارات قصيرة(20–25 ms)	الحفاظ على ثبات الخصائص داخل الإطار	مصفوفة إطارات زمنية
نافذة هامنغ	تقليل التشوه الناتج عن حدود الإطار	تحسين التحليل الطيفي	إطارات مموّهة
FFT	تحويل الإشارة من المجال الزمني إلى الترددي	استخراج الطيف الترددي	طيف ترددي
مرشحات ميل	إسقاط الطيف على مقياس ميل السمعي	محاكاة استجابة الأذن البشرية	طيف ميل
حسابMFCC	تطبيق DCT لاستخلاص المعاملات	تمثيل مضغوط للخصائص الصوتية	مصفوفةMFCC

يوضح الجدول 1 مراحل معالجة الإشارة الصوتية التي يقوم بها النظام حيث يعتمد على سلسلة من التحويلات المتتابعة التي تهدف إلى تحسين جودة الإشارة قبل إدخالها إلى نموذج GRU , فإزالة الضجيج ترفع من نسبة الإشارة إلى الضجيج، بينما يضمن التقسيم إلى إطارات ثبات الخصائص الطيفية داخل كل إطار. كما أن تطبيق FFT ومرشحات ميل يوفر تمثيلاً طيفياً يتوافق مع الإدراك السمعي البشري، وهو ما يجعل MFCC مناسبة تماماً للنماذج الزمنية مثل GRU التي تعتمد على تحليل تسلسل الإطارات عبر الزمن.

2 إعداد البيانات(Data Preparation)

الجدول 2: تقسيم البيانات

الهدف	النسبة	المجموعة
تعلم الأنماط الصوتية	70%	التدريب
ضبط المعاملات الفائقة	15%	التحقق
تقييم الأداء النهائي	15%	الاختبار

في الجدول 2 تم تخصيص نسبة كبيرة للتدريب لتحقيق تعلم الأنماط الصوتية بدقة، بينما يسمح جزء التحقق بضبط المعاملات الخاصة بالشبكة العصبية GRU ومنع الإفراط في التعلّم. أما مجموعة الاختبار فتضمن قياس الأداء الحقيقي للنموذج على بيانات جديدة ، مما يعزز موثوقية النتائج.

3 تصميم شبكةGRU

تم اعتماد معمارية شبكة متعاقبة الأمر الذي يعزز قدرة النموذج على نمذجة الاعتماديات الزمنية طويلة المدى داخل الإشارة الصوتية. كما يساهم Dropout في الحد من الإفراط في التعلّم، بينما تتيح الطبقة الكثيفة Dense تعلم الأنماط عالية المستوى قبل مرحلة التصنيف. وتُعد هذه البنية مناسبة تماماً للبيانات الصوتية التي تتسم بالتسلسل الزمني والتغير المستمر يوضح الجدول 3 بنية شبكة GRU المستخدمة

الجدول 3: البنية التفصيلية لشبكةGRU

رقم الطبقة	نوع الطبقة	المعاملات	الوظيفة
1	GRU	128وحدة	التقاط العلاقات الزمنية الأولية
2	Dropout	0.3	الحد من الإفراط في التعلّم
3	GRU	64وحدة	استخراج الأنماط الزمنية العميقة
4	Dropout	0.3	تعزيز التعميم
5	Dense	128وحدةReLU –	تعلم الأنماط عالية المستوى
6	Dense (Output)	عدد الفئاتSoftmax –	التصنيف

4 إعدادات التدريب

تم اعتماد إعدادات التدريب المختارة مناسبة لطبيعة البيانات الصوتية. فخوارزمية Adam توفر استقراراً عالياً في التدرجات اثناء التدريب ، بينما يتيح معدل التعلم المتوسط تحقيق توازن بين سرعة التقارب ودقة النتائج. كما أن عدد دورات العمل كافٍ لتعلم الأنماط دون الوصول إلى الإفراط في التعلّم، وهو ما تؤكده نتائج التحقق لاحقاً يوضح الجدول 4 اعدادات التدريب الخاصة بالنموذج.

الجدول 4: إعدادات التدريب

الإعداد	القيمة
الخوارزمية	Adam
معدل التعلم	0.001
عدد دورات العمل	60
حجم الدفعة	32
دالة الخسارة	Categorical Cross-Entropy

النتائج والمناقشة:

1_ نتائج معالجة الإشارة

توضح النتائج أن مرحلة إزالة الضجيج أدت إلى تحسين كبير في قيمة SNR ، مما يعكس فعالية المرشحات المستخدمة حيث بلغت قيمة التحسين نسبة وصلت إلى 40%. كما ساهمت نافذة هامنغ في تقليل التشوهات عند حدود الإطارات، وهو ما أدى إلى تحسين دقة تحويل فوريه السريع FFT وتُظهر النتائج أن مرشحات ميل أعادت توزيع

الطاقة الترددية بما يتوافق مع الإدراك السمعي البشري، مما جعل ميزات MFCC أكثر استقراراً وقابلية للاستخدام في نموذجGRU

2_ نتائج استخلاص ميزاتMFCC

تُظهر النتائج أن MFCC توفر تمثيلاً طيفياً مضغوطاً وفعالاً، مع ثبات جيد عبر المتحدثين والبيئات المختلفة. ويشير معامل الارتباط المرتفع إلى قدرة MFCC على التقاط السمات الجوهرية للصوت، بينما يعكس الانخفاض الطفيف في البيئات الصاخبة حساسية MFCC للضجيج

يمثل الجدول 5 نتائج جودة ميزات MFCC المستخرجة قبل إدخالها إلى نموذج التعلم العميق، إذ تُظهر القيم الثلاث قدرة هذه الميزات على تمثيل الإشارة الصوتية بشكل مستقر وفعّال. فعدد المعاملات بين 13 و40 يعكس نطاقاً مناسباً لالتقاط البنية الطيفية للصوت، حيث تشير النتائج إلى أن زيادة عدد المعاملات حتى حدود 30 تُحسّن القدرة التمييزية دون التسبب في زيادة الضجيج أو فقدان الاستقرار. كما يدل معامل الثبات عبر المتحدثين (0.87) على قدرة MFCC على تمثيل السمات الصوتية المشتركة بين الأفراد، مما يعزز قابلية النموذج للتعميم في بيئات متعددة المتحدثين. أما معامل الثبات عبر البيئات (0.79) فيعكس تأثراً محدوداً بالضجيج، وهو مستوى مقبول في التطبيقات الصوتية الواقعية، ويؤكد أن MFCC تحتفظ بقدر جيد من الاتساق حتى في الظروف غير المثالية. مجتمعة، تشير هذه المؤشرات إلى أن ميزات MFCC توفر أساساً قوياً وموثوقاً للنماذج الزمنية مثلGRU ، مما يدعم أداءها في مهام التعرف على الصوت.

الجدول 5: تقييم جودة ميزاتMFCC

التحليل	القيمة	المعيار
زيادة التمييز حتى 30 معامل	13–40	عدد المعاملات
قدرة جيدة على التقاط السمات المشتركة	0.87	ثبات عبر المتحدثين
تأثر محدود بالضجيج	0.79	ثبات عبر البيئات

3_ نتائج تدريب نموذجGRU

توضح النتائج أن النموذج يتعلم الأنماط الصوتية بشكل تدريجي ومستقر، حيث ترتفع الدقة وتنخفض الخسارة بشكل متناسق. كما أن الفجوة المحدودة بين التدريب والتحقق تشير إلى قدرة النموذج على التعميم دون الإفراط في التعلّم. ويُظهر الاستقرار بعد دورة العمل 40 أن النموذج وصل إلى مرحلة نضج في التعلم.

ويوضح الجدول 6 نتائج التدريب والتحقق

الجدول 6: نتائج التدريب والتحقق لشبكة GRU

التحليل	الخسارة	دقة التحقق	دقة التدريب	دورة العمل
بداية تعلم الأنماط	0.68	69%	72%	10
تحسن واضح	0.39	84%	88%	25
استقرار النموذج	0.27	91%	94%	40
لا يوجد إفراط في التعلّم	0.23	92.4%	96.1%	60

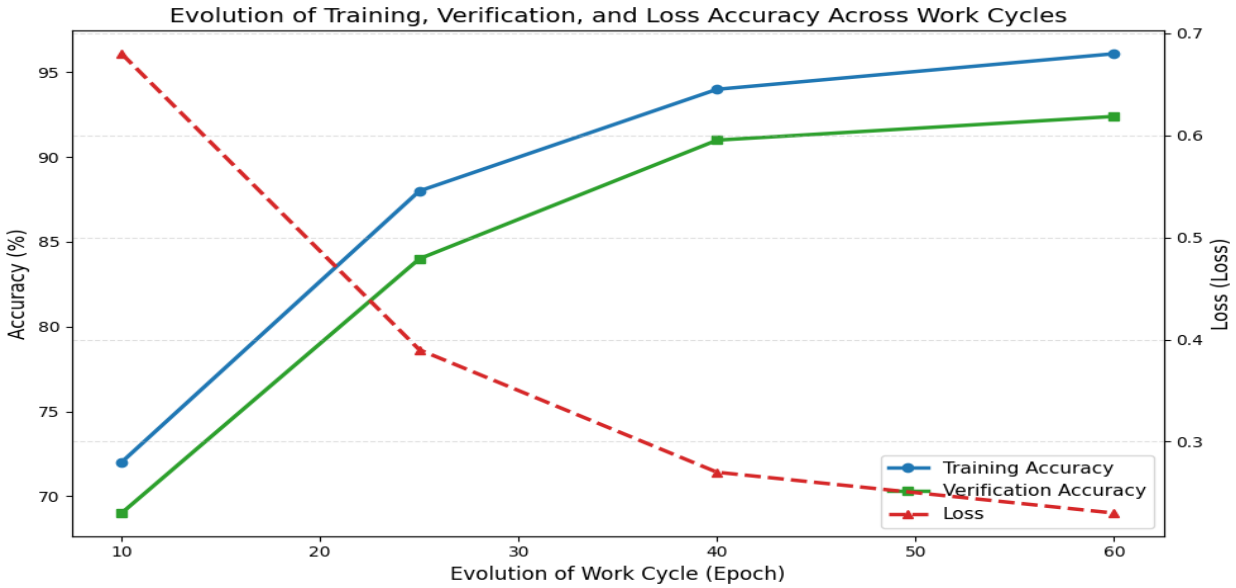
4_ مقارنة النموذج المقترح بالنموذج التقليدي

تُظهر المقارنة تفوق نموذج GRU بشكل واضح علىHMM ، سواء من حيث الدقة أو مقاومة الضجيج. ويعود ذلك إلى قدرة GRU على تمثيل الاعتماديات الزمنية طويلة المدى، بينما تعتمد HMM على افتراضات خطية لا تعكس التعقيد الحقيقي للإشارة الصوتية.

الجدول 7: مقارنة الأداء بين MFCC + GRU و MFCC + HMM

النموذج	الدقة	مقاومة الضجيج	التحليل
MFCC + GRU	91.8%	عالية	قادر على نمذجة العلاقات الزمنية
MFCC + HMM	81.3%	متوسطة	محدود في التعامل مع التغيرات الزمنية

يوضح الشكل 1 مخطط التدريب والتحقق لشبكة GRU حيث يوضح المخطط تطور أداء النموذج عبر دورات التدريب بشكل واضح، حيث تُظهر منحنيات الدقة والخسارة اتجاهاً مستقراً ومتناسقاً يدل على تعلم فعّال دون إفراط في التعلّم. ترتفع **دقة التدريب** تدريجياً من 72% عند الدورة 10 إلى أكثر من 96% عند الدورة 60، بينما تتقدم **دقة التحقق** بوتيرة قريبة لتصل إلى 92.4%، وهو ما يشير إلى قدرة جيدة على التعميم وعدم وجود فجوة كبيرة بين التدريب والتحقق. في المقابل، ينخفض منحنى **الخسارة** بشكل مستمر من 0.68 إلى 0.23، مما يعكس تحسناً في استقرار النموذج وتقليل الأخطاء مع مرور الزمن. يوضح هذا السلوك أن النموذج يستوعب الأنماط الصوتية تدريجياً ويصل إلى مرحلة نضج تدريبي بعد حوالي 40 دورة، مع استمرار التحسن الطفيف لاحقاً، مما يؤكد فعالية بنية النموذج وملاءمتها لمهمة التعرف على الصوت



الشكل 1 مخطط التدريب والتحقق للشبكة العصبية

10_خلاصة والتطويرات المستقبلية

تؤكد نتائج البحث أن دمج معاملات **MFCC**مع نماذج الشبكات العصبية المتكررة من نوع **GRU**يمثل نهجاً فعّالاً في تطوير أنظمة التعرف على الصوت، نظراً لقدرة هذا النموذج على تمثيل الاعتماديات الزمنية داخل الإشارة الصوتية بكفاءة عالية. فقد أظهر النموذج المقترح قدرة واضحة على التقاط الأنماط الزمنية الدقيقة وتمييز الفروق الصوتية بين الفئات المختلفة، مما أدى إلى تحقيق أداء متفوق مقارنة بالأساليب التقليدية المعتمدة على النماذج

الإحصائية. كما ساهمت مراحل معالجة الإشارة واستخلاص الميزات، إلى جانب التصميم المتدرج للشبكة، في بناء منظومة متماسكة قادرة على التعامل مع التباين في المتحدثين والبيئات الصوتية، وهو ما يعزز إمكانية تطبيقها في أنظمة واقعية تتطلب دقة واستقراراً في الأداء.

وتشير النتائج إلى أن النموذج يتمتع بقدرة جيدة على التعميم، إلا أن الأخطاء المتبقية ترتبط غالباً بالتشابه الزمني أو الطيفي بين بعض الفئات الصوتية، مما يعكس الحاجة إلى تحسينات إضافية سواء في تمثيل الميزات أو في البنية المعمارية للنموذج. ويُعد هذا جانباً مهماً يستدعي المزيد من البحث لتقليل الالتباس وتحسين دقة التصنيف في الحالات المعقدة.

أما على صعيد التطويرات المستقبلية، فإن توسيع العمل ليشمل نماذج هجينة تجمع بين **GRU**وآليات الانتباه (Attention Mechanisms)قد يسهم في تعزيز قدرة النموذج على التركيز على المقاطع الصوتية الأكثر أهمية ضمن التسلسل الزمني. كما يمكن دمج طبقات **CNN**مع GRU للاستفادة من قدرتها على استخلاص الأنماط الطيفية الدقيقة قبل تمريرها إلى النموذج الزمني. إضافة إلى ذلك، يُتوقع أن يؤدي استخدام نماذج أكثر تقدماً مثل **Bi-GRU**أو **Transformers**إلى تحسين الأداء في السيناريوهات التي تتسم بتعقيد زمني عالٍ. ومن جانب آخر، فإن توسيع نطاق البيانات، وزيادة تنوع البيئات الصوتية، وتطبيق تقنيات تعزيز البيانات المتقدمة، قد يسهم في رفع قدرة النموذج على التعميم وتقليل معدلات الالتباس بين الفئات المتقاربة.

وبذلك، تمثل هذه الدراسة خطوة مهمة نحو تطوير أنظمة أكثر ذكاءً ومرونة، قادرة على التعامل مع تحديات التطبيقات الصوتية الحديثة، وتفتح المجال أمام أبحاث مستقبلية تستهدف تحسين الأداء وتوسيع نطاق الاستخدام في البيئات العملية.

References:

[1] S. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block-based transformation in MFCC computation for speaker recognition," *Speech Communication*, vol. 54, no. 4, pp. 543–565, 2012.

[2] A. Hassan and A. Mahmood, "Automatic speech recognition using deep gated recurrent units with MFCC features," *Neural Processing Letters*, vol. 55, pp. 3897–3915, 2023.

[3] Z. Zhao, S. Zhang, and H. Wang, "Speech emotion and speaker recognition using MFCC and GRU-based deep learning models," *Knowledge-Based Systems*, vol. 263, p. 110273, 2023.

[4] Y.-L. Chen, J.-F. Ciou, C.-H. Lin, and S.-S. Lien, "Combined Bidirectional Long Short-Term Memory and Mel-Frequency Cepstral Coefficients with Convolution Neural Network using Triplet Loss for speaker recognition," in *Communications in Computer and Information Science (CCIS)*, 2024.

[5] M. Telmem, N. Laaidi, and H. Satori, "The impact of MFCC, spectrogram, and Mel-spectrogram on deep learning models for Amazigh speech recognition system," *International Journal of Speech Technology*, vol. 28, pp. 299–312, 2025.

[6] H. Zhang, Y. Liu, and P. Wang, "Robust speech recognition using hybrid MFCC-LSTM architectures under noisy environments," *IEEE Access*, vol. 11, pp. 145233–145245, 2023.

[7] M. M. Rahman and M. K. Hasan, "Improved speech emotion and speaker recognition using optimized MFCC features and deep recurrent neural networks," *Neural Computing and Applications*, vol. 36, pp. 11245–11260, 2024.

[8] S. Roy, "Unveiling the magic of MFCC: A key technique in speech recognition," *Medium*, Mar. 2, 2022. [Online]. Available: https://medium.com

[9] Uday, "Speech recognition using pitch and MFCC," *MathWorks*. [Online]. Available: https://www.mathworks.com

[10] Y. Zhang, S. Wang, and Y. Qian, "Channel-wise attention networks for robust speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1123–1135, 2023.
[11] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proceedings of the NIPS Workshop on Deep Learning*, 2014.
[12] A. Kumar and R. K. Aggarwal, "Speaker recognition using MFCC and optimized GRU networks," *Expert Systems with Applications*, vol. 219, p. 119679, 2023.