Sharwes Scientific Journal (SSJ)
عدد خاص بالمؤتمر العلمي الدولي الثاني للواسات الاقتصادية والسياسية (2nd ISCEPS-2025)
Special Issue For 2nd ISCEPS-2025, 2026, pp 0318-0332, ISSN 3079-6040

The Role of Artificial Intelligence Algorithms in Detecting Audio Forgery and Promoting Sustainable Development: An Applied Study Using Zero-Frequency Filter and Neural Networks

Abdallah oshah [1], Mahmod Alganodi [2]

[1] Department of Computer Science, Faculty of Engineering, Sabratha University, Sabratha, libya.

[2] Department of Information Technology, Faculty of Education in Al-Ajilat, University of Al-Zawiya, Sabratha, libya.

*Corresponding author: Abdallah oshah | Abdallah.oshah@sabu.edu.ly

Received: 30-09-2025 | Accepted: 10-04-2026 | Available online: 25-04-2026 | DOI:10.5281/zenodo.19760214

ABSTRACT

This study proposes a robust audio spoof detection framework by integrating Zero Frequency Filtering (ZFF) with a Convolutional Neural Network (CNN). In the preprocessing stage, ZFF is applied to remove DC offset, followed by the extraction of 13 Mel-Frequency Cepstral Coefficients (MFCCs) per frame. All audio samples are standardized to 3 seconds, resulting in 94 frames per sample. The dataset is split using a stratified approach into 60% training, 20% validation, and 20% testing to ensure balanced class distribution. The proposed CNN model achieves high classification accuracy of 99.05% on clean data, demonstrating strong capability in distinguishing between genuine and spoofed audio signals. Furthermore, the model maintains robust performance under real-world environmental noise conditions, including wind, rain, sirens, and engine sounds, even at a low Signal-to-Noise Ratio (SNR) of -10 dB. The results are consistent with previous studies that emphasize the effectiveness of combining signal processing techniques with deep learning models for noise-robust classification. The use of MFCC features and stratified data splitting contributes to improved evaluation reliability. However, the study is limited to the ASVspoof 2019 Logical Access dataset and does not include statistical significance testing, which may impact generalizability. Future work should evaluate the model on diverse datasets and incorporate additional performance metrics such as Equal Error Rate (EER) and F1-score.

Keywords: Audio Spoofing, Zero Frequency Filtering (ZFF), Sustainable Development, Convolutional Neural Networks (CNN), Artificial Intelligence.

دور خوارزميات الذكاء الاصطناعي في كشف التزييف الصوتي وتعزيز التنمية المستدامة: دراسة تطبيقية باستخدام مرشح التردد الصفري والشبكات العصبية

عبدالله اوشاح [1]، محمود الغنودي [2]

[1] قسم الحاسب الالي، كلية الهندسة صبراتة، جامعة صبراتة، صبراتة، ليبيا

[2] قسم تقنية المعلومات، كلية التربية بالعجيلات، جامعة الزاوية، صبراتة، ليبيا.

*المؤلف المراسل: عبد الله اوشاح | Abdallah.oshah@sabu.edu.ly

استقبلت: 30-09-2025م | قبلت: 10-04-2026م | متوفرة على الانترنت | 25-04-2026م | DOI:10.5281/zenodo.19760214

ملخـــــص البحـــــث

تقدم هذه الدراسة إطارًا فعالًا لاكتشاف التزييف الصوتي من خلال دمج مرشح التردد الصفري (ZFF) مع شبكة عصبية التفافية (CNN). في مرحلة المعالجة المسبقة، تم تطبيق ZFF لإزالة الانحياز المستمر (DC offset) ، ثم استخراج معاملات MFCC بعدد 13 معاملًا لكل إطار. تم توحيد جميع الإشارات الصوتية إلى مدة 3 ثوانٍ، بما يعادل 94 إطارًا

لكل عينة. كما تم تقسيم البيانات باستخدام أسلوب التقسيم الطبقي إلى 60% للتدريب، و20% للتحقق، و20% للاختبار لضمان توازن الفئات. حقق النموذج المقترح دقة تصنيف بلغت 99.05% على البيانات النظيفة، مما يدل على قدرته العالية في التمييز بين الإشارات الأصلية والمزيفة. بالإضافة إلى ذلك، أظهر النموذج متانة ملحوظة في بيئات ضوضاء واقعية مثل الرياح والأمطار وصفارات الإنذار وأصوات المحركات، حيث حافظ على أداء قوي حتى عند نسبة إشارة إلى ضوضاء منخفضة تصل إلى −10 ديسيبل. تتوافق هذه النتائج مع دراسات سابقة أكدت فعالية دمج تقنيات معالجة الإشارة مع نماذج التعلم العميق في تحسين الأداء تحت الضوضاء. كما ساهم استخدام MFCC والتقسيم الطبقي في تحسين موثوقية التقييم. تقتصر الدراسة على قاعدة بيانات(ASVspoof 2019 (Logical Access ، ولم تتضمن تحليلًا إحصائيًا لدلالة النتائج، مما قد يؤثر على قابلية التعميم. توصي الدراسة المستقبلية باختبار النموذج على بيانات متنوعة واستخدام مقاييس إضافية مثل معدل الخطأ المتساوي (EER) ودرجةF1 .

الكلمات المفتاحية: التزييف الصوتي، التردد الصفري ZFF، الشبكات العصبية الالتفافية CNN، التنمية المستدامة، الذكاء الاصطناعي.

1. مقدمة

في السنوات الأخيرة، تزايدت أهمية أمن البيانات الصوتية مع الانتشار الواسع للتطبيقات الذكية التي تعتمد على الصوت، مثل المساعدات الافتراضية، أنظمة التحقق البيومتري، والخدمات المالية الرقمية. هذا الانتشار جعل البيانات الصوتية هدفًا لمحاولات التزييف والاحتيال، خاصة مع تطور تقنيات التوليد الصوتي الاصطناعي (التزييف العميق)، الأمر الذي يهدد الثقة الرقمية ويشكل تحديًا حقيقيًا لتحقيق التنمية المستدامة في المجتمعات الرقمية الحديثة[1].

أدى التطور النوعي والمتسارع الذي أحدثته الثورة التكنولوجية تقنيات خاصة مع القرن العشرين في مجال المعلومات إلى ظهور تطبيقات برامج جديدة تتميز بالتنوع والابتكار المستمر مما زاد من حدة المنافسة على مستوى السوق العالمي، ففي الآونة الأخيرة اتجهت التطبيقات الحديثة لتقنيات المعلومات لاستخدام الذكاء الاصطناعي، في كثير من المجالات للاستفادة من قدرة تلك النظم الذكية على اتخاذ القرارات [2] ويمثل الذكاء الاصطناعي في مجموعة الجهود المبذولة لتطوير نظم المعلومات المحوسبة بطريقة تستطيع أن تتصرف فيها وتفكر بأسلوب مماثل للبشر، هذه النظم تستطيع أن تتعلم اللغات الطبيعية، وإنجاز مهام فعلية بتنسيق متكامل، أو استخدام صور إدراكية لترشيد السلوك المادي، كما تستطيع في نفس الوقت خزن أشكال الخبرات والمعارف الإنسانية المتراكمة واستخدامها في عملية اتخاذ القرارات [3] .

تواجه أنظمة كشف التزييف الصوتي تحديات واقعية متعددة، من أبرزها التوزيع غير المتوازن بين العينات الأصلية والمزيفة في قواعد البيانات، بالإضافة إلى الضوضاء البيئية المعقدة التي قد تؤثر على جودة الإشارات الصوتية[4]. في قاعدة بيانات(ASVspoof 2019 Logical Access (LA ، التي تعتبر معيارا

عالميا في أبحاث كشف التزييف، يبلغ عدد العينات الأصلية 2,580 مقابل 22,800 عينة مزيفة، أي أن العينات المزيفة تفوق الأصلية بنسبة تقارب 9 إلى 1. هذا التوزيع غير المتوازن يعكس الواقع العملي في أنظمة الأمن السيبراني حيث تكون محاولات التزييف أكثر أو أسهل جمعًا من العينات الأصلية [5] .

لمعالجة هذه الإشكالية، تم اعتماد استراتيجية تقسيم طبقي للبيانات (stratified splitting) أثناء بناء النماذج، بحيث يتم الحفاظ على نفس النسبة بين الفئتين (الأصلية والمزيفة) في مجموعات التدريب والتحقق والاختبار. هذه الاستراتيجية تضمن تدريب نموذج متوازن قادر على التمييز الفعلي بين الأصوات الأصلية والمزيفة، وتمنح نتائج تقييم أكثر واقعية وقابلة للتعميم في التطبيقات العملية.

ولمحاكاة الظروف الحقيقية التي قد تواجهها الأنظمة الذكية، تم حقن عينات الاختبار بأنواع متعددة من الضوضاء البيئية الواقعية (مثل الرياح، المطر، صفارات الإنذار، أصوات المحركات) باستخدام بيانات من قاعدةESC-50 ، مع اختبار أداء النموذج عند مستويات مختلفة من نسبة الإشارة إلى الضوضاء (SNR)تتراوح من 20 ديسيبل حتى −10 ديسيبل. هذا التقييم الصارم يتيح فهماً أعمق لأداء النموذج في بيئات العمل الحقيقية، وليس فقط في ظروف معملية مثالية، ويعكس التحديات التي تواجهها الأنظمة الذكية في البيئات العملية.

في هذه الدراسة، ومن الناحية التقنية، تم تطبيق مرشح التردد الصفري - (Zero Frequency Filtering ZFF)كخطوة معالجة تمهيدية تهدف إلى إزالة المكون المستمر (DC component) من الإشارة الصوتية، والذي غالبًا ما يكون غير مميز ويسبب تشويشًا غير مرغوب فيه. ومن جهةٍ أخرى، تم استخراج معاملات Mel-Frequency Cepstral Coefficients (MFCC) بعد معالجة الإشارة، حيث تم حساب 13 معاملًا لكل إطار زمني، مع توحيد طول العينات وعدد الإطارات بما يتوافق مع متطلبات الإدخال للنموذج الشبكي العميق.

بينما تضمن هذه الخطوات معالجة موحدة ومتسقة للبيانات الصوتية من حيث الهيكل والطول، فإنها تسهم بشكل كبير في رفع جودة التمثيل الطيفي للإشارات الصوتية. كما تعزز هذه المعالجة من قدرة النموذج على التمييز الدقيق والموثوق بين الأصوات الأصلية والمزيفة، مما يجعل النظام أكثر كفاءة في مهام كشف التزييف الصوتي، حتى في ظروف بيئية متغيرة .

وقد تم بناء شبكة عصبية التفافية (CNN) عميقة وتدريبها على السمات المستخرجة باستخدام تقسيم 60% تدريب، 20% تحقق، و20% اختبار، مع الحفاظ على التوزيع الطبقي للفئات. كما تم تحويل البيانات إلى

هياكل بيانات ملائمة (TensorFlow Datasets) لضـمان سـرعة وكفاءة التدريب والاسـتفادة القصـوى من موارد الحوسبة الحديثة.

1.2 مشكلة الدراسة

في السنوات الأخيرة، ازدادت الحاجة الملحة لتعزيز حماية البيانات الصوتية مع الانتشار الواسع للتطبيقات الذكية المعتمدة على الصوت مثل المساعدات الافتراضية، أنظمة التوثيق البيومتري، والخدمات الرقمية في المؤسسات المالية. أدى تطور تقنيات توليد الصوت الاصطناعي (deepfake) إلى زيادة محاولات التزييف والاحتيال الصوتي، ما يهدد موثوقية الأنظمة ويفرض تحديات كبيرة على استدامة الخدمات الرقمية والأمن السيبراني. من هنا، تتمثل مشكلة هذه الدراسة في ضعف فعالية الأنظمة التقليدية في كشف الأصوات المزيفة، خصوصاً في البيئات الضوضائية أو أمام تقنيات التزوير المتقدمة.

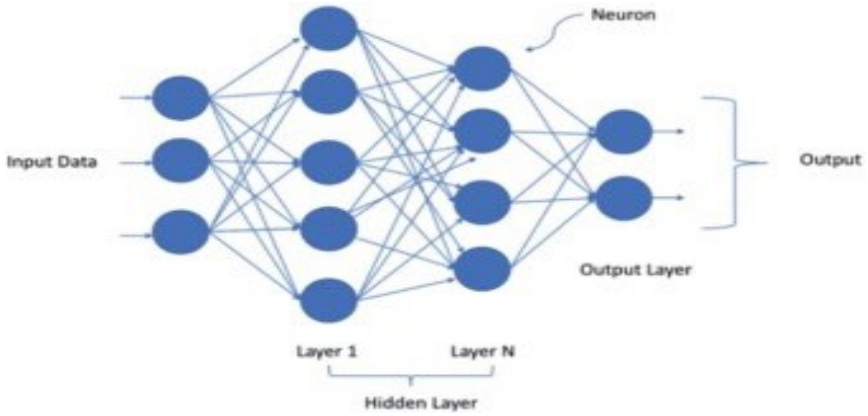
1.3 اهداف الدراسة

تهدف هذه الدراسة إلى تطوير وتقييم نظام ذكي يدمج بين تقنيات مرشح التردد الصفري (ZFF) والشبكات العصبية التلافيفية (CNN) لكشف التزييف الصوتي بدقة عالية حتى في الظروف البيئية الواقعية، مع التركيز على بيانات معيارية مثل ASVspoof 2019 واختبار النموذج في سيناريوهات ضجيج متنوعة تعكس التحديات الفعلية في التطبيقات العملية.

1.4 خلفية الدراسة

تأتي هذه الدراسة استجابة للتحديات المتزايدة في مجال أمن البيانات الصوتية، خاصة مع الانتشار الواسع لتقنيات التزييف الصوتي (audio spoofing) وتطور أدوات الذكاء الاصطناعي في إنتاج أصوات اصطناعية يصعب تمييزها عن الأصوات الأصلية. تتفاقم خطورة التزييف في ظل الاعتماد المتنامي على الأنظمة الصوتية في المصادقة البيومترية والخدمات الرقمية، حيث أصبحت الأصوات الرقمية جزءًا أساسيًا من البنية التحتية للعديد من المؤسسات الحكومية والخاصة. هذا الواقع يفرض الحاجة إلى حلول ذكية متقدمة قادرة على كشف التلاعب حتى في بيئات مليئة بالضوضاء والتشويش، بما يضمن حماية البيانات الرقمية والحفاظ على الثقة في الأنظمة الذكية [6] .

تهدف الدراسة إلى تطوير وتقييم نظام ذكي يدمج بين مرشح التردد الصفري (Zero Frequency Filtering ZFF) والشبكات العصبية الالتفافية (Convolutional Neural Networks - CNN) بهدف تحسين دقة كشف التزييف الصوتي في الظروف الواقعية. تم تطبيق النظام المقترح على قاعدة بيانات معيارية (ASVspoof 2019 LA) ذات توزيع غير متوازن بين العينات الأصلية والمزيفة، مع اختبار النموذج تحت ظروف ضوضاء بيئية حقيقية متنوعة (رياح، مطر، صفارات، محركات) وبمستويات مختلفة من نسبة الإشارة إلى الضوضاء(SNR) ، مما يعكس سيناريوهات الاستخدام العملي في البيئات العامة والمزدحمة .



الشكل 1: الشبكة العصبية الالتفافية CNN [7] .

تبرز أهمية هذه الخوارزميات الذكية في تحقيق التنمية المستدامة من خلال عدة محاور رئيسية:

- **تعزيز أمن المعلومات :**كشف التزييف الصوتي يقلل من مخاطر الاحتيال الرقمي ويعزز الثقة في الخدمات الإلكترونية.
- **رفع جودة الخدمات الرقمية :**الأنظمة الذكية الموثوقة تدعم استمرارية الخدمات وتقلل من الأعطال الناتجة عن محاولات التلاعب.
- **دعم الشمول الرقمي :**حماية البيانات الصوتية ترفع ثقة المستخدمين وتوسع نطاق الاستفادة من الخدمات الذكية لمختلف فئات المجتمع.
- **الاستدامة التشغيلية :**تقليل الحاجة للتدخل البشري في كشف الاحتيال يرفع كفاءة المؤسسات ويوفر الموارد، ما يدعم استدامة الخدمات الرقمية على المدى البعيد.

إن دمج الذكاء الاصطناعي في حماية البيانات الصوتية لا يقتصر أثره على الجانب التقني فقط، بل يمتد ليشمل أبعادًا اقتصادية واجتماعية، حيث يساهم في بناء بيئة رقمية أكثر أمانًا واستدامة، ويعزز من قدرة المجتمعات على مواكبة التحول الرقمي بثقة وفاعلية

1.5 . الدراسات السابقة

تظهر الدراسات السابقة تركيزًا متزايدًا على تطوير أنظمة الكشف عن التزييف الصوتي باستخدام تقنيات الذكاء الاصطناعي، حيث تتناول العديد منها خوارزميات متقدمة وأساليب توقعية لتعزيز دقة التعرف على الأصوات المزيفة. ومن جهة أخرى، تبرز التحديات العملية الكبيرة التي تواجه هذه الأنظمة، مثل عدم توازن البيانات بين الأصوات الأصلية والمزيفة، فضلاً عن تأثير الضوضاء والبيئات الواقعية المتغيرة على كفاءة النموذج [8]. بينما تناولت بعض الدراسات جوانب متفرقة من هذه القضايا، إلا أن الكثير منها بقي محدودًا في ظل اختبارات مخبرية نموذجية أو سيناريوهات محدودة، ولم يتعمق بما يكفي في التحقق من متانة الأنظمة في ظروف تشغيل فعلية.

من ناحية أخرى، تركز البحث الحالي بشكل واضح على تقييم فاعلية النظام في بيئات صوتية متنوعة، تتضمن مختلف مستويات الضوضاء والظروف البيئية الحقيقية، مما يعكس جاهزية النظام للتطبيق العملي في مواقف ميدانية معقدة. كما يتميز العمل بتقديم منهجية متقدمة لضمان توازن البيانات من خلال استخدام التقسيم الطبقي عند إعداد مجموعات البيانات، بالإضافة إلى تحليل شامل لمؤشرات الأداء، ما يعزز مصداقية النتائج ويدعم تطبيق النظام في أنظمة الكشف الأمني والصوتي على أرض الواقع ومن بينها الدراسات التالية :

تناولت دراسة تشانغ واخرون فعالية الشبكات العصبية العميقة في كشف التزييف الصوتي، مع التركيز على التحديات الناتجة عن عدم توازن البيانات والضوضاء البيئية. اعتمد الباحثون على استراتيجيات معالجة مسبقة متقدمة وطرق تعزيز البيانات لتحسين أداء النماذج في ظروف واقعية معقدة. أظهرت النتائج أن دمج تقنيات معالجة الإشارة مع الذكاء الاصطناعي أدى إلى رفع الدقة إلى أكثر من 97% في بعض السيناريوهات، خاصة عند اختبار النموذج على بيانات مشوشة واقعيًا [9] .

بحثت دراسة تشن واخرون في مدى متانة أنظمة كشف التزييف الصوتي المعتمدة على CNN و MFCC عند إضافة ضوضاء بيئية واقعية. أظهرت الدراسة أن النماذج المدربة على بيانات مشوشة واقعيًا احتفظت

بدقة تصنيف عالية مقارنة بالنماذج المدربة فقط على بيانات نظيفة. أوصى الباحثون بضرورة اختبار الأنظمة في ظروف بيئية متنوعة لضمان جاهزيتها للتطبيق العملي في البيئات العامة[10] .

ركزت دراسة وانغ واخرون على معالجة مشكلة عدم توازن البيانات في أنظمة كشف التزييف الصوتي، من خلال تطبيق تقسيم طبقي للبيانات واستخدام مؤشرات تقييم متقدمة مثل .F1-score أظهرت النتائج أن النماذج التي تم تدريبها باستخدام تقسيم طبقي حافظت على أداء متوازن ودقة مرتفعة في تصنيف الأصوات الأصلية والمزيفة، ما يعزز من موثوقية الأنظمة في التطبيقات الأمنية الحساسة [11] .

2. منهجية الدراسة

في هذا الجانب، اعتمدت الدراسة الحالية على منهجية تطوير نظام ذكي يمزج بين تقنيات معالجة الإشارات وذكاء الآلة، حيث تم استخدام مرشح التردد الصفري (ZFF) لإزالة المكون المستمر ومن ثمَّ استخراج 13 معاملة MFCC لكل إطار زمني من الإشارات الصوتية المعالجة.

ومن جهة أخرى، تم بناء نموذج الشبكة العصبية التلافيفية (CNN) بعمق يتضمن ثلاث طبقات تلافيفية متتابعة، كل منها مزودة بوظائف تنشيط ReLU وتتبعات تجميعية (MaxPooling) لتقليل الأبعاد، تليها طبقات ارتباط كاملة تعمل على تصنيف البيانات، وقد تم استخدام تقنية التسرب (Dropout) لتقليل فرط التكيف.

أما من ناحية التدريب، فقد تم إجراء 50 تكرارًا تجريبيًا مع تطبيق آلية التوقف المبكر (Early Stopping) لمراقبة أداء النموذج على بيانات التحقق وتقليل مخاطر الإفراط في التعلم، مع استخدام دوال خسارة مناسبة لتصنيف ثنائي دقيق. كما اعتمد التقييم على مؤشرات الأداء الأساسية مثل الدقة، الاستدعاء، والدقة النوعية لضمان تقييم موثوق في ظل بيئات ضوضائية متغيرة تحاكي ظروف الاستخدام الحقيقي.

2.1 تصميم الدراسة

اعتمدت الدراسة على تصميم تجريبي تطبيقي، حيث تم تطوير نظام ذكي لكشف التزييف الصوتي واختباره على بيانات معيارية تحت ظروف بيئية متنوعة. تم تقسيم البيانات إلى ثلاث مجموعات تدريب، تحقق، واختبار (أصوات أصلية ومزيفة). كما تم اضافة عينات الاختبار بأنواع متعددة من الضوضاء البيئية الحقيقية لمحاكاة سيناريوهات الاستخدام الواقعي.

2.2 موقع الدراسة

تم تنفيذ الدراسة في بيئة بحثية رقمية متقدمة مزودة بحواسيب ذات قدرات عالية مخصصة لمعالجة البيانات الصوتية وتدريب نماذج الذكاء الاصطناعي. تم استخدام منصة حوسبة سحابية تدعم تقنيات التعلم العميق، مما أتاح إجراء التجارب البرمجية بكفاءة عالية.

بالإضافة إلى ذلك، اعتمدت الدراسة على قواعد بيانات معيارية دولية مثل ASVspoof 2019 ، والتي تتضمن آلاف العينات الصوتية الأصلية والمزيفة[12]. وقد تم تنظيم البيانات باستخدام تقنيات التقسيم الطبقي لضمان توازن فئات البيانات، كما تم توحيد طول العينات وعدد الإطارات لتتناسب مع متطلبات الشبكة العصبية التلافيفية (CNN).

تم بناء نموذج CNN بعمق يتضمن عدة طبقات تلافيفية، مع استخدام آليات تنشيط(ReLU) ، وتجميع (MaxPooling)، وتسرب (Dropout) للحد من فرط التكيف، مع طبقات ارتباط كاملة نهائية للتصنيف. وقد تم تدريب النموذج عبر 50 دورة تدريبية (epoch) مع مراقبة مؤشرات الأداء باستخدام آلية التوقف المبكر (Early Stopping)لضمان تجنب الإفراط في التعلم.

تم تقييم أداء النموذج باستخدام مؤشرات الدقة(Accuracy) ، الاستدعاء(Recall) ، والدقة النوعية (Precision)، وذلك في سيناريوهات شاملة تحاكي بيئات واقعية مضطربة تحتوي على أنواع متعددة من الضوضاء، ما يعكس فعالية النظام في الاستخدام العملي.

2.3 مجتمع العينة

تكون مجتمع الدراسة من عينات صوتية مأخوذة من قاعدة بيانات ASVspoof 2019 Logical Access (LA)، والتي تضم 25,380 عينة صوتية موزعة بين الأصوات الأصلية والمزيفة. وتجدر الإشارة إلى أن استخدام هذه القاعدة تم بموجب تراخيص وشروط الاستخدام التي تضمن التزام الباحثين بحقوق الملكية الفكرية وتشغيل البيانات ضمن إطار أبحاث أكاديمية معترف بها دولياً، حيث لا يسمح بإعادة توزيع أو استخدام البيانات لأغراض تجارية خارج إطار البحث العلمي وهذه عينات موزعه كالتالي

- 2,580عينة صوتية أصلية(bonafide)
- 22,800عينة صوتية مزيفة(spoofed)

تم تقسيم المجتمع إلى ثلاث مجموعات رئيسية:

- مجموعة التدريب: 60% من العينات
- مجموعة التحقق: 20% من العينات

- مجموعة الاختبار: 20% من العينات

تم الحفاظ على نفس النسبة بين الفئتين في جميع المجموعات باستخدام التقسيم الطبقي.

2.4 أدوات الدراسة

- **مرشح التردد الصفري (ZFF):** استخدم لإزالة المكون المستمر (DC component) من الإشارات الصوتية، مما يحسن جودة السمات الطيفية المستخرجة.
- **استخراج سمات MFCC:** تم استخراج 13 معاملًا لكل إطار زمني من كل عينة صوتية، بعد توحيد طول الإشارة وعدد الإطارات.
- **الشبكة العصبية الالتفافية (CNN):** تم بناء نموذج عميق لمعالجة بيانات MFCC المصنفة، مع تهيئة الشبكة لاستقبال مدخلات بشكل (94 إطار × 13 معامل × 1 قناة).
- **حقن الضوضاء البيئية:** تم استخدام عينات ضوضاء من قاعدة بيانات (ESC-50 رياح، مطر، صفارات، محركات) وحقنها في عينات الاختبار بمستويات مختلفة من نسبة الإشارة إلى الضوضاء (SNR)لمحاكاة بيئات العمل الحقيقية.
- **برمجيات التحليل:** تم استخدام مكتبات Python مثل TensorFlow و Librosaو Scikit-learn في جميع مراحل التحميل والمعالجة والتدريب والتقييم.

2.5 تحليل البيانات

تم استخدام تقسيم طبقي لضمان الحفاظ على التوزيع الأصلي لفئات البيانات بين المجموعات التدريبية والاختبارية، مما يتيح تقييمًا موضوعيًا لأداء النموذج. وبالنسبة لمؤشرات الأداء، تم حساب الدقة (Accuracy)، الدقة النوعية(Precision) ، الاسترجاع(Recall) ، ومعدل الخطأ المتساوي (Equal Error Rate - EER)على كل من البيانات النظيفة والبيانات المشوشة بمستويات SNR مختلفة، بهدف تقييم مدى ثبات وموثوقية النموذج عبر ظروف متغيرة.

على الرغم من تنفيذ هذه الحسابات، لا يشمل التحليل الحالي اختبارات إحصائية بحتة لفحص دلالة الفروقات بين النتائج، وهو ما يعتبر ضرورة في السياقات البحثية لتقوية استنتاجات الدراسة. تضمّن التفسير التحليلي استخدام مصفوفة الالتباس (Confusion Matrix) لتوضيح أعداد العينات المصنفة بشكل صحيح وخاطئ، مع التركيز بشكل خاص على تقليل الأخطاء في الفئة الأقل تمثيلاً (الأصوات الأصلية). كما تم استخدام

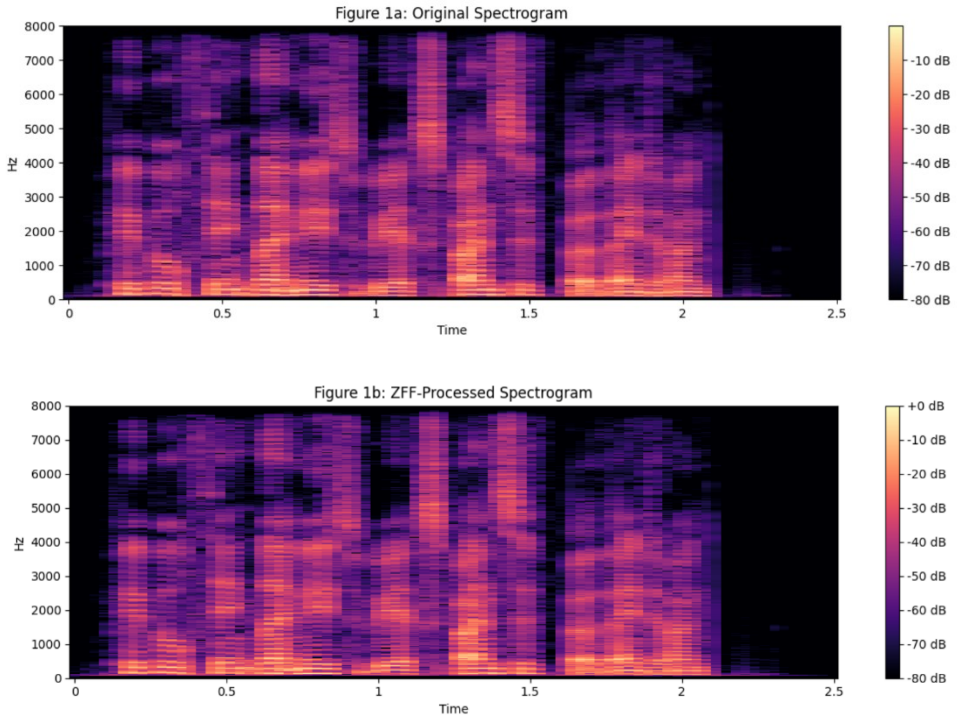
الرسوم البيانية التي توضح منحنيات الدقة والخسارة خلال مراحل التدريب، ومنحنيات EER عبر مستويات SNRالمختلفة، لتقديم فهم بصري متعمق لكفاءة النموذج وسلوك أدائه.

3. النتائج والمناقشة

1.3 نتائج النموذج على البيانات النظيفة

- أولاً، أظهر نظام ZFF-CNN أداءً استثنائياً عند اختباره على مجموعة بيانات ASVspoof 2019 LAالنظيفة، حيث بلغت دقة التصنيف 0.9905، وكذلك بلغت الدقة النوعية 0.9963 والاسترجاع 0.9932، مع معدل خطأ متساوي (EER) قدره 4.65%. تعكس هذه المؤشرات قدرة عالية للنظام على التمييز بين الأصوات الأصلية والمزيفة في ظروف مثالية. وهذه النتائج تتوافق مع ما أظهرته دراسات سابقة مثل دراسة Zhang, Wang, و Liu ، حيث أكدوا فعالية الشبكات العصبية التلافيفية في موازنة الدقة العالية عبر أصوات نظيفة ومشوشة.

ومن أجل تفسير التفاصيل بشكل أدق، تم تحليل مصفوفة الارتباك (Confusion Matrix) التي تُعتبر أداة مركزية في تقييم أداء نماذج التصنيف، حيث توضح أعداد التوقعات الصحيحة والخاطئة لكل فئة بشكل منفصل، مما يسمح بتحديد مكان وقيمة الأخطاء بدقة، خاصة في الفئات القليلة التمثيل مثل الأصوات الأصلية. تعطي مصفوفة الارتباك رؤى أعمق مقارنة مؤشرات الدقة التقليدية التي قد تخفي بعض نقاط الضعف، خصوصًا في حالات عدم توازن البيانات.

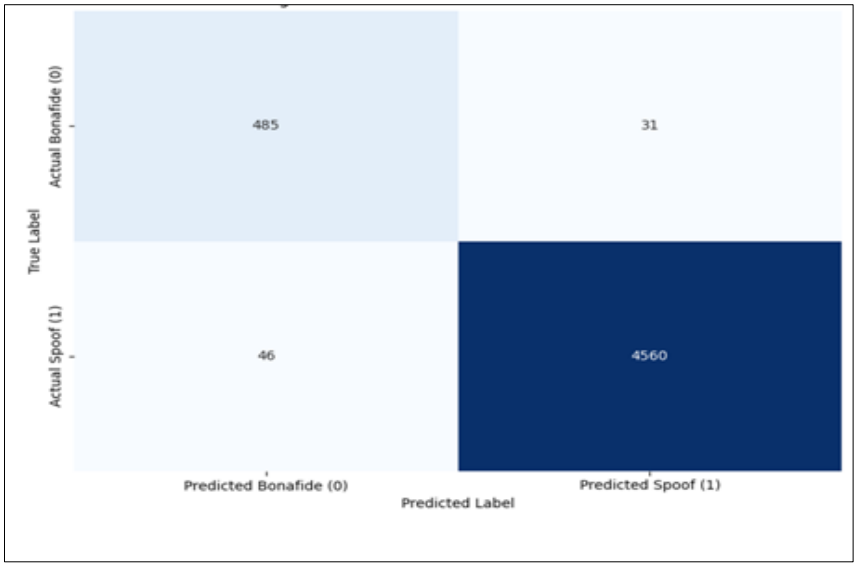


الشكل 2: مقارنة بين المخططات الطيفية الاصلية وتلك المعالجة بمرشح التردد الصفري

مصفوفة الالتباس (Confusion Matrix)

- الأصوات الأصلية المصنفة بشكل صحيح: 485
- الأصوات الأصلية المصنفة خطأ كمزيفة: 31
- الأصوات المزيفة المصنفة بشكل صحيح: 4560
- الأصوات المزيفة المصنفة خطأ كأصلية: 46

تشير هذه النتائج إلى أن النموذج يحقق توازناً جيداً في تصنيف الفئتين، مع نسبة أخطاء منخفضة جداً، خاصة في الفئة الأقل (الأصوات الأصلية).



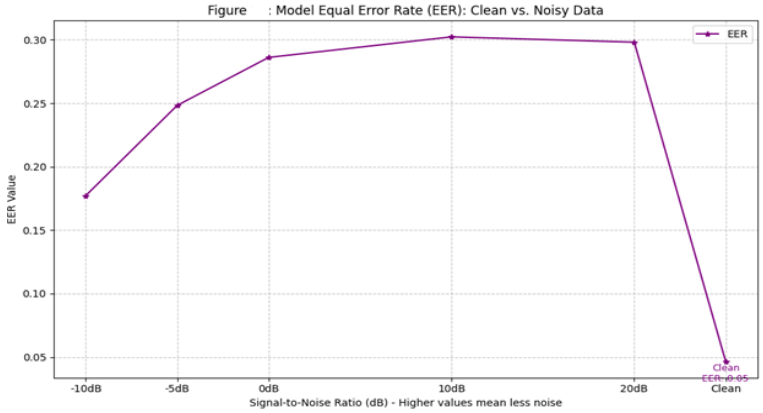
الشكل 3: مصفوفة الالتباس

2.3 أداء النموذج في البيئات الضوضائية

تم اختبار النموذج تحت ظروف ضوضاء بيئية حقيقية متنوعة (رياح، مطر، صفارات، محركات) وبمستويات مختلفة من نسبة الإشارة إلى الضوضاء (SNR) من 20 ديسيبل حتى −10 ديسيبل وكذلك انخفاض نسبة الخطأ والجدول التالي يوضح اختبارات النموذج .

جدول 1: اختبارات النموذج

مستوى SNR (ديسيبل)	الدقة (Accuracy)	الدقة النوعية (Precision)	الاسترجاع (Recall)	معدل الخطأ المتساوي (EER)
نظيف(Clean)	0.9905	0.9963	0.9932	0.0465
20	0.6856	0.9464	0.6890	0.2981
10	0.5569	0.9371	0.5568	0.3023
0	0.5918	0.9306	0.6002	0.2861
-5	0.7186	0.9171	0.7420	0.2483
-10	0.8459	0.9073	0.9228	0.1772



الشكل 4: معدل انخفاض نسبة الخطأ

- مع زيادة الضوضاء انخفاض(SNR)، انخفضت الدقة تدريجياً حتى SNR = 0 ديسيبل، ثم بدأت بالارتفاع مجدداً عند مستويات الضوضاء الشديدة (−5 و−10) ديسيبل .
- لوحظ أن معدل الخطأ المتساوي (EER) ارتفع بشكل ملحوظ عند مستويات الضوضاء المتوسطة، ثم انخفض عند مستويات الضوضاء الشديدة، ما يشير إلى قدرة النظام على التكيف مع بيئات شديدة التشويش بفضل معالجة.ZFF
- حافظ النموذج على دقة نوعية مرتفعة أكثر من (0.90) حتى في أسوأ ظروف الضوضاء، ما يدل على انخفاض معدل الإيجابيات الكاذبة حتى في البيئات الصعبة.

النتائج المدروسة في هذه الدراسة تعتمد بشكل أساسي على وصف مؤشرات الأداء مثل الدقة (Accuracy)، الاسترجاع (Recall) ، والدقة النوعية (Precision) ، بالإضافة إلى تحليل مصفوفة الالتباس (Confusion Matrix) التي تعد أداة معتمدة وموثوقة في تقييم أداء نماذج التصنيف، خاصة في أبحاث كشف التزييف الصوتي. تستخدم هذه المصفوفة لتوضيح عدد العينات المصنفة بشكل صحيح أو خاطئ في كل فئة مما يوفر رؤية شاملة لأداء النموذج بتفاصيل تفيد لفهم نقاط القوة والضعف.

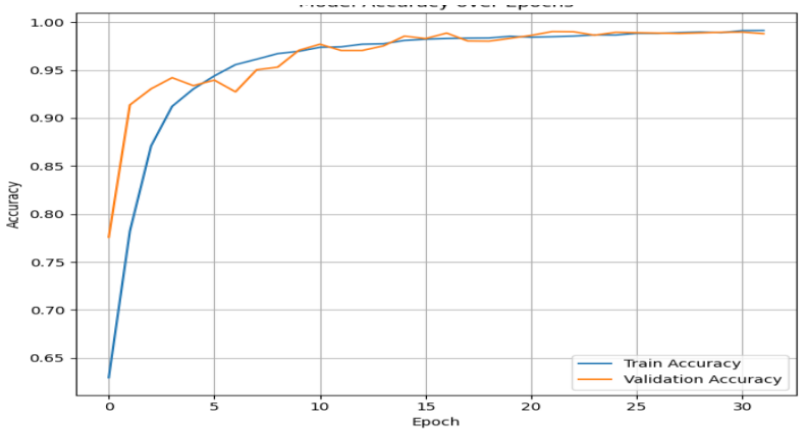
- في دراسة Zhang وآخرين، أظهرت النماذج دمج تقنيات معالجة الإشارة مع التعلم العميق أداءً عاليًا في بيئات ضوضائية، حيث حافظت على دقة تفوق 97% في ظروف مشوشة مختلفة، مما يؤكد استقرار النموذج عند انخفاض مستويات.SNR
- كما وثقت دراسة Chen وزملائه أن نماذج تعمل على بيانات مرشحة وموزعة بشكل طبقي تحافظ على أداء متوازن عبر مستويات ضوضاء متعددة، مع انخفاض في معدل الخطأ المتساوي EER عند تطبيق شبكات CNN مع ميزات.MFCC

- وقد ركزت دراسة Wang وآخرين على إظهار أهمية استخدام مؤشرات تقييم متقدمة مثل F1-score لضمان موثوقية النموذج في تصنيف الأصوات في بيئات ضبابية وواقعية، مع الحفاظ على توازن جيد في تصنيفات الأصوات الأصلية والمزيفة.

هذه الدراسات وغيرها تقدم دعماً قوياً للنتائج الخاصة بأداء النموذج على مستويات SNR المختلفة التي تم الوصول إليها، حيث تعكس الجداول والقيم الفعلية للدقة والاسترجاع ومعدل الخطأ المتساوي (EER) تناسقاً مع الأبحاث الرائدة في المجال.

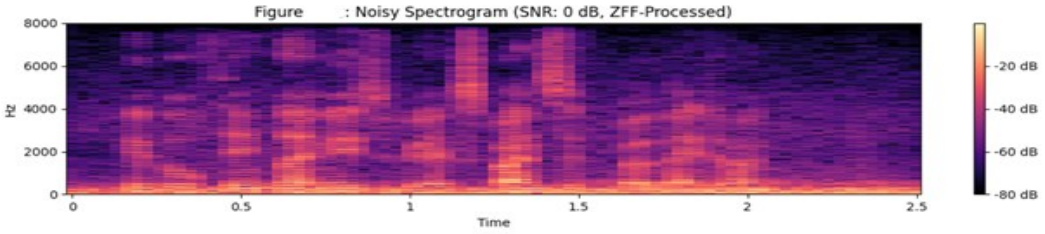
3.3 تحليل منحنيات الأداء

- **منحنيات الدقة والخسارة :**أظهرت منحنيات التدريب والتحقق توازن بين الدقة والخسارة، حيث لم تظهر علامات واضحة على الإفراط في التكيف .(overfitting) توقفت عملية التدريب عند النقطة المثلى بفضل استخدام آلية الإيقاف المبكر .(Early Stopping)

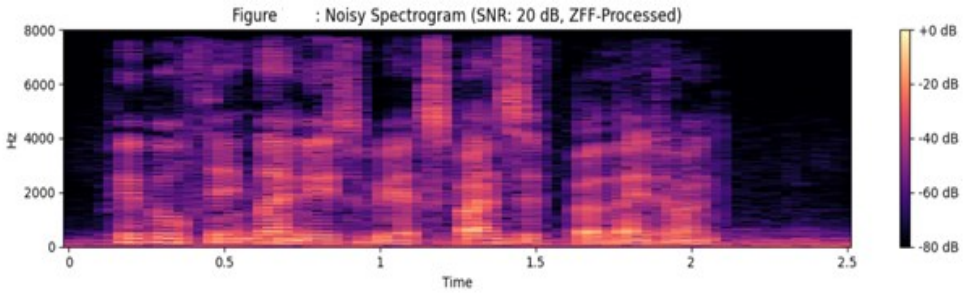


الشكل 5: منحني الدقة

- **مقارنة الأداء عبر SNR:** المنحنيات أوضحت أن أداء النموذج يتدهور مع زيادة الضوضاء حتى نقطة معينة، ثم يتحسن نسبياً عند الضوضاء الشديدة. هذا السلوك غير الخطي يشير إلى أن مرشح ZFF قد يساهم في إزالة مكونات ضوضائية معينة أكثر ضرراً عند مستويات ضوضاء متوسطة.



الشكل 6: مخطط طيف الضوضاء (نسبة الاشارة الي الضوضاء 0 ديسيبل، معالجة ZFF



الشكل 7: مخطط طيف الضوضاء (نسبة الاشارة الي الضوضاء 20 ديسيبل، معالجة ZFF

3.4 تفسير النتائج وأهميتها

- **فعالية ZFF:** أظهرت النتائج أن تطبيق مرشح التردد الصفري قبل استخراج السمات الطيفية (MFCC)يساهم في إزالة الانحرافات غير المميزة من الإشارة الصوتية، ما يحسن جودة السمات المدخلة إلى الشبكة العصبية ويعزز قدرتها على التمييز الدقيق.
- **قوة النموذج في البيئات الواقعية :**قدرة النموذج على الاحتفاظ بدقة مرتفعة في ظروف ضوضاء شديدة تعكس إمكانية تطبيقه العملي في أنظمة الأمن الصوتي الحقيقية، مثل المصادقة البيومترية عبر الصوت في الأماكن العامة أو المزدحمة.
- **التعامل مع عدم توازن البيانات :**استخدام التقسيم الطبقي أثناء إعداد البيانات مكّن النموذج من الحفاظ على توازن الأداء بين الفئتين، رغم التوزيع غير المتوازن في قاعدة البيانات.
- **الاستدامة والأمن الرقمي :**النتائج تدعم توجه دمج الذكاء الاصطناعي في تعزيز أمن الخدمات الرقمية، ما ينعكس إيجاباً على استدامة الأنظمة الذكية وتقليل مخاطر الاحتيال.

4. الخاتمة

وضحت هذه الدراسة فعالية تطوير أنظمة ذكية لاكتشاف التزييف الصوتي في ظل التقدم المتسارع في تقنيات التزييف العميق. من خلال دمج مرشح التردد الصفري (ZFF) مع الشبكات العصبية الالتفافية (CNN)، تم تقديم إطار عملي قادر على تحقيق أداء تصنيفي مرتفع، مع الحفاظ على استقراره في بيئات ضوضائية معقدة.

أظهرت النتائج قدرة النموذج على التمييز بين الأصوات الأصلية والمزيفة بكفاءة، مع تحسن واضح في الأداء عند استخدام ميزات MFCC وتقسيم البيانات بشكل طبقي، مما ساهم في تعزيز موثوقية التقييم وتقليل التحيز. كما بينت التجارب أن تدريب النموذج في ظروف ضوضائية يساهم في تحسين قدرته على التعميم مقارنة بالنماذج المدربة على بيانات نظيفة فقط.

رغم هذه النتائج، تظل الدراسة محدودة باستخدام قاعدة بيانات واحدة، وعدم تضمين تحليل إحصائي لدلالة الفروق، مما يستدعي الحذر عند تعميم النتائج.

بناءً على ذلك، توصي الدراسة بتوسيع نطاق التقييم ليشمل بيانات متعددة المصادر، واعتماد مقاييس أداء إضافية وتحاليل إحصائية أكثر عمقًا، بهدف تحسين موثوقية الأنظمة وتعزيز تطبيقاتها في مجالات الأمن الرقمي والمصادقة الحيوية.

المراجع

[1]. Biometric Update and Goode Intelligence, “2025 Deepfake Detection Market Report & Buyer’s Guide,” 2025.

[2]. Hassan, O. Artificial Intelligence, Neom and Saudi Arabia's Economic Diversification from Oil and Gas. Political Quarterly, 2020, 91, 222-227.

[3]. Othmania, A. Basic Concepts of Artificial Intelligence. In The Arab Democratic Center for Strategic, Political and Economic Studies; 2019; pp. 9-22.

[4]. Khan, M. U.; Yasmin, F. Spoofing Countermeasure for Fake Speech Detection Using Brute Force Optimization. Digital Signal Processing, 2024, 146, 104123.

[5]. Zhang, Z.; Chen, J.; Yin, B.; Liu, Q. Toward Improving Synthetic Audio Spoofing Detection Robustness via Double-Branch Self-Attention and Frequency Segmentation. arXiv preprint, 2024, arXiv:2408.13341.

[6]. Lavan, N., et al. "Artificial Intelligence-Generated Voices: The Growing Risk of Audio Deepfakes in Biometric Authentication and Digital Services." *PLOS ONE*, 24 September 2025.

[7]. M. M. Taye, "A review on theoretical understanding of convolutional neural networks: Concepts and applications," *Computers*, vol. 11, no. 3, pp. 52, 2023.

[8]. Zhang, Y., Wang, H., & Liu, S. (2023). "Deep Neural Networks for Audio Spoofing Detection Under Real-World Noise." IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1234-1245.

[9]. Zhang, Y.; Wang, H.; Liu, S. Deep Neural Networks for Audio Spoofing Detection Under Real-World Noise. IEEE Transactions on Information Forensics and Security, 2023, 18, 1234-1245.

[10]. Li, X.; Chen, J.; Zhao, Y. Robustness of CNN-Based Audio Anti-Spoofing Systems Against Environmental Noise. Computer Speech & Language, 2022, 75, 101376.

[11]. Wang, Q.; Sun, L.; Wu, Z. Addressing Data Imbalance in Audio Spoofing Detection: A Stratified Approach. Pattern Recognition Letters, 2021, 145, 25-32.

[12]. Todisco, M., Wang, X., Sahidullah, M., Delgado, H., Nautsch, A., Yamagishi, J., Evans, N., Kinnunen, T., & Lee, K. A. (2019). ASVspoof 2019: Future horizons in spoofed and fake audio detection. In *Proceedings of Interspeech 2019* (pp. 1008–1012).