IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 1, January, 2026, pp. 21-26
Journal homepage: http://iieta.org/journals/isi

# Detection of Real, AI-Synthesized, and Human-Mimicked Voices in Forensic Audio Using One-Dimensional Convolution Neural Networks

Marem H. Abdulabas[1], Ruaa Kadhim Khalaf[2], Azha Talal Mohammed Ali[3*], Noor D. AL-Shakarchy[3]

[1] Department of Information Technology, Faculty of Computer Science and Information Technology, Kerbala University, Kerbala 56001, Iraq
[2] Department of Central Library, Kerbala University, Kerbala 56001, Iraq
[3] Department of Computer Science, Faculty of Computer Science and Information Technology, Kerbala University, Kerbala 56001, Iraq

Corresponding Author Email: azha.t@uokerbala.edu.iq



https://doi.org/10.18280/isi.310103

**Received:** 13 July 2025
**Revised:** 15 September 2025
**Accepted:** 18 January 2026
**Available online:** 31 January 2026

***Keywords:***
*forensic audio analysis, voice spoofing detection, deepfake speech detection, human voice mimicry, synthesized speech detection, one-dimensional convolutional neural networks, audio forensics*

**ABSTRACT**

The rapid development of speech synthesis and voice manipulation technologies has raised growing concerns regarding the authenticity and reliability of audio evidence in forensic investigations. Advanced deep learning techniques can generate highly realistic synthetic voices, while skilled individuals may imitate voices with remarkable accuracy, making reliable forensic voice verification increasingly challenging. This study presents a lightweight deep learning framework for forensic audio classification capable of distinguishing between real, AI-synthesized, and human-mimicked voices. The proposed approach employs a one-dimensional convolutional neural network (1D CNN) architecture designed to learn discriminative temporal and spectral representations directly from audio waveforms. A curated version of the DEEP-VOICE dataset was used in this study, further augmented with additional mimicry recordings to balance the class distribution and increase data diversity. The final dataset contains 6,780 audio samples, divided into training, validation, and testing sets using a 70:15:15 split. Experimental evaluation demonstrates that the proposed model achieves an overall classification accuracy of 93% with a loss value of 0.0218. Class-specific results show F1-scores of 0.88 for real voices, 0.98 for AI-synthesized voices, and 0.92 for human-mimicked voices. Additional experiments under noisy conditions and cross-dataset evaluation confirm the robustness and generalization capability of the proposed architecture. These results indicate that lightweight one-dimensional convolutional neural networks provide an effective and computationally efficient solution for forensic voice authentication and detection of manipulated speech in legal and security contexts.

## 1. INTRODUCTION

Speech is among the most widely disseminated data in the digital realm. Through speech, the recipient can discern not only the message's content (i.e., plain text) but also additional attributes such as intonation, rhythm, and genre, among others. Nonetheless, these signals are readily susceptible to manipulation to mislead the recipient of the voice message [1]. However, these signals are easily manipulated for the purpose of deceiving the listener of the voice message [2]. A person may imitate another individual's voice so professionally that it is difficult for the listener to distinguish or discover the true identity of the speaker [3].

Recent advances in AI-generated speech and spoofing attacks [4-6] introduced new challenges in the field of forensic science have been introduced due to the advent of advanced audio manipulation technologies, such as deep learning-based voice synthesis and human mimicry [7, 8]. These technologies create realistic fake voices with alarming accuracy that can be used to impersonate individuals. This phenomenon poses a significant threat to the authenticity of forensic voice evidence. Voice manipulation can be categorized into two primary types: AI-generated voices, which are generated using deep learning models, including GANs, which stand for generative adversarial networks, and human-mimicking voices, which are generated by skilled individuals who imitate the voices of others. These two manipulation types are increasingly utilized in various malicious activities, including fraud, misinformation, and identity theft [9].

In the current digital age and the explosion of online social networking sites, falsehoods may be propagated faster than truths. As well as the advancements in deep learning algorithms and the rise of deep fakes techniques within reach have further exacerbated the misinformation and pose an increasingly sophisticated threat to individuals' livelihoods and reputations. Fake videos and audio with high production values are widespread and can pose a serious threat to people's lives. When you look at the numbers, you will see that in this post-truth age, people of all ages, backgrounds, and levels of education are eager to receive this information [1]. It is

important to recognize the value of authentic multimedia content for reporting on world events and bringing those who are responsible to justice. Audiovisual media [10] is one of the best ways to get people to pay attention to global atrocities and genocides and to prove guilt in a specific case.

In forensics, it is very important to be able to tell the difference between real, AI-generated, and imitated voices. In criminal investigations, legal proceedings, and security systems, the integrity of voice evidence is very important. Not being able to find fake impersonators can lead to wrong convictions, less secure systems, and less trust in the law [2]. It is still difficult to detect human audio mimicry, even when done by professional voice actors. This is especially true in blind mimicry, where the attributed person's actual recorded voice is either partially or completely absented, making it difficult to detect tested audio without any reference, whether real, fake, or mimicked [10]. This difficulty is at the heart of the study's research question. Specifically, determining whether a given audio is authentic or fake and for fake voice directly determines whether it is human-mimicked or AI-generated, without making assumptions about the identity of the supposed speaker.

Assuming the purported speaker has never been encountered by the system, it is possible to train models that can distinguish between genuine recordings and forgeries (human-mimic generated/AI generated) by utilizing the advances in machine learning, specifically deep neural networks, to extract meaningful features from audio data. The ability of CNNs, which symbolizes Convolutional Neural Networks, to effectively process and analyze sequential data, including time series, they are extensively used across various applications [11, 12]. A 1D CNN, which represents One-dimensional Convolutional Neural Network, with a lightweight architecture is offering several advantages, such as [13-15]:

- The model can quickly learn local patterns in the input sequence owing to the 1D convolution operation.
- The model's trainability and resistance to overfitting are both improved when the number of parameters is reduced.
- The 1D CNN model is capable of recognizing patterns regardless of features position within the input sequence because it learns translation-invariant features.
- The model can learn hierarchical representations of the input by stacking multiple convolutional layers with increasing filter sizes, and it can capture complex patterns at various scales in the data [16].

This paper addresses the pressing need for reliable forensic detection mechanisms capable of identifying manipulated voices in forensic evidence. The proposed model implemented a lightweight convolutional neural network architectures to analyze and classify voice samples. By extracting intricate audio features and patterns, our system aims to provide robust and accurate detection of voice status (original\fake) as well as the impersonators status (human -mimic generated \ AI generated), enhancing the reliability of forensic audio analysis. This research aims to address the growing concern of security and trustworthiness in voice-based technologies by providing a reliable method for voice authentication to Real, Fake, and Mimic. To reflect three categories detected in this study and enhance the accuracy of voice classification, the DEEP-VOICE Dataset enrichment.

The following is the outline for the rest of the paper: we review related work in the field of voice manipulation detection, outline the methodology and data set used in our study, present the results of our experiments, and discuss the implications and future directions for forensic audio analysis.

## 2. LITERATURE REVIEW

Previous studies have addressed replay and synthetic voice attacks using deep learning approaches [17-20].

Using a CNN, which symbolizes Convolutional Neural Network, with image augmentation and dropout, Ballesteros et al. [9] developed Deep4SNet to differentiate between authentic and fake speech recordings acquired through Deep Voice and Imitation. There are three convolution layers, with a pooling layer coming after each one, then a flatten layer, a hidden layer, and the output layer. To prevent overfitting, the architecture makes use of dropouts in the hidden layer. 2092 histograms of authentic and fraudulent voice recordings were used to train the suggested architecture, and 864 histograms were used for cross-validation. Using 476 new histograms for external validation resulted in a global accuracy of 98.5%.

Hamza et al. [13] in 2022 use the technique of MFCCs, which symbolizes Mel-frequency cepstral coefficients in combination with machine learning and deep learning-based approaches to identify deepfake audio. Part of the larger Fake-or-Real (FoR) dataset, which includes the for-norm, for-2-sec, for-rece, and for-original subsets, is used in this investigation. In terms of accuracy, the outcomes of the experiment demonstrate that the SVM, which symbolizes support vector machine, performed better than the other machine learning (ML) models on the for-rece dataset (98.83%) and the for-2-sec dataset (97.57%), while the gradient boosting model demonstrated exceptional performance on the for-norm dataset (92.63%). When applied to the for-original dataset, the VGG-16 model achieved an impressive 93% accuracy, which is very encouraging.

Almutairi and Elgibreen [2] in 2023, a new audio deepfake detection system known as Arabic-AD was presented, which uses self-supervised learning procedures to recognize imitation and synthetic sounds. By making the first synthetic dataset of a single speaker who speaks MSA (Modern Standard Arabic) fluently, it has added to the body of literature. Also, the accent was taken into account when collecting Arabic recordings from people who did not speak Arabic to see how well Arabic-AD worked. Arabic-AD did these things with very little training, and it had the lowest EER rate (0.027%) and the highest detection accuracy (97%) of any state-of-the-art method.

The Convolutional Block Attention Module (CBAM) and a deep layered model called VGGish were introduced by Kanwal et al. [12] in 2024 for the purpose of spoofing detection. The attention block is used to split the input audio into two groups: fake and real. Then, the audio is turned into mel-spectrograms, and the most representative features are taken out. The model's simple, layered structure makes it a great tool for finding audio spoofing. The attention module can pick up on complex relationships in audio signals because it has spatial and channel features. The suggested method got 99.5% accuracy when it used the ASVspoof 2019.

In 2024, Al Ajmi et al. [10] present a deep neural network technique for developing a classifier that would blindly categorize input audio as mimicked or real; a three-hidden-layer deep neural network that applies a sequential model with drop-out and dense layers alternated. The model's performance is proven by its accuracy, which routinely exceeds 94%

(94.2% for the mixed dataset and 94.1% for the English dataset, compared to 85% accuracy for human observers).

Antony and Gopikakumari [20] in 2018 proposed a speaker identification system that combines MFCC, which symbolizes Mel-frequency cepstral coefficients and UMRT, which symbolizes Unique Mapped Real Transform features for text dependent and text independent system. MLP, which denotes Multi-layer perception with the back propagation algorithm is used to classify the features. The confusion matrix is used to determine the accuracy. For speech-dependent systems, the accuracy attained is approximately 97.91%, whereas for speech-independent systems, it is approximately 94.44%.

While the reviewed studies have demonstrated substantial progress in voice spoofing and deepfake detection, several limitations remain unaddressed. Most existing methods, such as Deep4SNet and Arabic-AD, primarily focus on binary classification (real vs. fake) and do not account for human-mimicked voices, which are increasingly prevalent in forensic scenarios. Furthermore, many models rely on spectrogram-based 2D CNN architectures, which are computationally expensive and sensitive to noise, limiting their applicability in real-time forensic analysis. In contrast, the proposed study introduces a lightweight 1D CNN architecture capable of processing raw audio waveforms directly, thereby reducing preprocessing complexity while maintaining high accuracy. Moreover, by extending the classification task to include three distinct categories—real, AI-synthesized, and human-mimicked voices— this work advances beyond previous binary detection frameworks and provides a more comprehensive forensic detection system suitable for real-world evidence authentication. Table 1 shows summary of related work.

**Table 1.** The related works summary

| Authors | Year | Method | Dataset | Accuracy |
|---|---|---|---|---|
| Ballesteros et al. [9] | 2021 | Convolutional Neural Network (CNN), using image augmentation and dropout | The LJ Speech Dataset | 98.5% |
| Hamza et al. [13] | 2022 | (MFCCs combination with SVM model) (MFCCs combination with VGG-16 model) | Fake-or-Real (FoR) dataset | 98.83%, 93% |
| Almutairi and Elgibreen [2] | 2023 | Arabic-AD based on self-supervised learning techniques to detect both synthetic and imitated voices | Arabic Diversified Audio (Ar-DAD), Arabic Speech Corpus (ASC), Arabic-CAPT dataset | 97% |
| Kanwal et al. [12] | 2024 | a deep layered model (VGGish) | ASVspoof 2019 dataset | 99.5% |
| Al Ajmi et al. [10] | 2024 | deep neural network following a sequential model | English dataset, mixed dataset | 94.1%, 94.2% |
| Antony and Gopikakumari [20] | 2018 | For feature extractor both MFCC and UMRT, for classification of the features is done using (MLP) | Speech dependent systems, for speech independent systems | 97.91%, 94.44% |

Note: CNN = Convolutional Neural Network; MFCCs = Mel-Frequency Cepstral Coefficients; SVM = Support Vector Machine; VGG-16 = Visual Geometry Group 16-layer Convolutional Neural Network; AD = Audio Deepfake; UMRT = Uniform Multi-Resolution Texture; MLP = Multi-Layer Perceptron.

## 3. PROPOSED REAL, FAKE AND MIMIC DETECTION MODEL (RFMD MODEL)

By examining the data in an audio file and identifying important aspects to produce feature maps, the suggested model generates a pattern of genuine voice characteristics. The same model then tracks these feature maps over time to identify whether the audio is mimic (human mimicking), fake (AI-synthesized), or real. The preprocessing stage and the classification stage are the two primary phases of the suggested model's operation. Every stage has multiple steps, each of which serves a distinct purpose, Figure 1 shows the proposed audio classification model.

### 3.1 Dataset

DEEP-VOICE, this publicly accessible dataset, found in the "KAGGLE" directory, includes real-world human speech samples from eight recognized people that have had their speech translated using retrieval-based voice conversion. "Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion" is the study that produced this record. You may find the raw audio in the "AUDIO" directory. Within the "REAL" and "FAKE" class directories, they are organized. The audio filenames identify the speakers who delivered the original speech and the voices they were transformed into. For instance, "Obama-to-Biden" indicates that Joe Biden's voice has been used to mimic Barack Obama's words. Since there isn't a publicly available dataset (benchmark dataset) that satisfies every need of the suggested system, the DEEP-VOICE dataset is enhanced by adding mimic voices for the same people with matching identities.

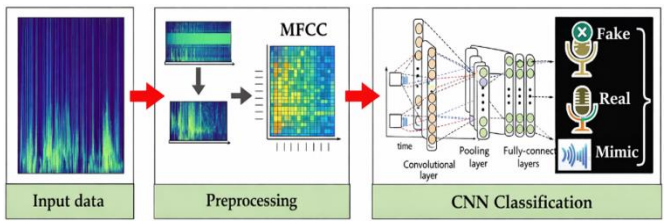



**Figure 1.** The proposed audio classification model

### 3.2 Data preprocessing

All necessary preparations are made to input data during the pre-processing stage so that it can be used in the proposed classification stage. As unprocessed audio could include unwanted background noise or inappropriate information, the pre-processing stage prepares the input audio to ensure suitability for the classification stage while maintaining system generalization across diverse inputs. This phase

involves the following steps:

- Standardizing Audio File Extensions: All audio files are converted to a uniform common file format to ensure consistency across the dataset. This standardization streamlines the handling and processing of the audio data, allowing for more efficient batch processing and compatibility with various audio analysis tools.
- Resampling: To ensure consistency in sample rate across different audio files in the dataset, resampling is applied. The sample rate of all audio recordings is adjusted to a desired rate with the same number of samples per second in order to avoid discrepancies due to differing sampling rates each audio file. This step adopted Fourier method along the given axis to resample audio files to 60000 Hertz.
- Trimming and Padding: They are essential preprocessing functions to ensure all inputs have the same length by removing silence at the beginning and end of audio or by padding the shorter one.
- Noise Reduction: This step involves the removal of non-informative parts and unwanted sounds to ensure that only the useful audio content remains for analysis such as background noise, clapping or environmental disturbances. Energy-based filtering techniques are employed to identify and discard these irrelevant portions based on a spectral gating approach.
- Normalization and Scaling: By analyzing the audio signal's amplitude to find its peak amplitude—the loudest part of the audio—and then modifying the audio to make the loudest part reach a predetermined target level, the normalization process optimizes the dynamic range of audio to a consistent and standardized amplitude level. The large integer value input causes disrupted or slowed down the learning process, therefore the Neural network models need to deal with small weight values. By utilizing Loudness Units Full Scale, a scaling is employed to fine-tune a gain that is directly proportional to the disparity between the initial peak level and the desired level.
- Segmentation: To analyze specific sections and uniform processed, this step split long audio files into manageable smaller segments
- Sampling Dataset: According to imbalanced dataset, adjusting the size of the dataset by balancing class distributions based on Down sampling it. This measure guarantees that models remain impartial to the majority class.
- Splitting Dataset: The final preprocessing step divides the dataset into training, validation, and test sets.

### 3.3 Classification stage

This work presents One-dimensional Convolutional Neural Network (1D CNN) model to recognize important characteristics in the audio data, with pooling layers to reduce complexity and prevent overfitting, and fully connected layers for final classification to three categories: "Real", "Fake", and "Mimic" using a Soft-Max activation function. The model is structured lightly architecture. This architecture is composed of the following layers each specific function done:

- Input Layer: The input layer receives the audio file after pre-processing as a group of classes indicating the waveform (the audio signal in the time domain) and the sampling rate (the number of samples per second).
- Convolutional Layers: The model begins with a series of Conv1D layers, which are designed to capture local patterns in sequential data. The first Conv1D layer has 64 filters that are 5 pixels wide. It then uses ReLU activation to make the output non-linear. The model can learn more and more complex features from the audio data by stacking more Conv1D layers. These layers are essential for extracting crucial attributes, including rhythm, pitch, and other acoustic features from the audio signal.
- Pooling Layers: To avoid overfitting and decrease the data's dimensionality, the model includes several MaxPooling1D layers. These layers down-sample the feature maps, assisting in the retention of only the utmost crucial information. Max-Pooling also aids in making greater invariance of the model to slight changes in the input data.
- Dropout Layer: To further combat overfitting, a 0.5-rate dropout layer is added after one of the convolutional layers. Dropout randomly deactivates half of the neurons during training, which forces the model to learn more robust features and improves generalization. The above layers are done together to extract the salient features from audio file.
- Flattening Layer: The Flatten layer flattens the feature maps created by the pooling and convolutional layers into a one-dimensional vector. The data is then ready to be loaded into the fully connected layers.
- Fully Connected Layers: The model includes two Dense layers with 1000 and 300 units, respectively. These layers help the model learn high-level, abstract features and perform the final classification. ReLU activation is employed to present non-linearity and enhance learning.
- Output Layer: All three output classes—"Real," "Fake," and "Mimic"—are represented by the three units in the last layer, a Dense layer. In this layer, the model's output is guaranteed to be a distribution over the classes by using a Soft-Max activation function to generate probabilities for each.

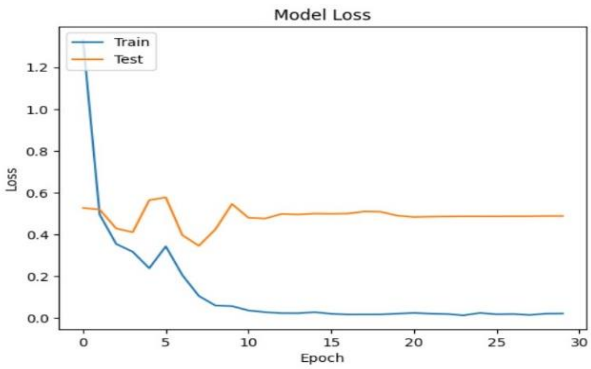
Categorical cross-entropy, the loss function used in the model's compilation, is perfect for multi-class classification tasks. The Adam optimizer is used for efficient training, and the accuracy metric is tracked to assess the model's performance. The TensorFlow framework and the Keras API were used to build the proposed 1D CNN model. There are three convolutional blocks in the architecture. Each one has a Conv1D layer, a MaxPooling1D layer, and a Dropout layer. Specifically, the first, second, and third Conv1D layers contain 64, 128, and 256 filters, respectively, each with a kernel size of 5 and ReLU activation. To lower the number of features, all MaxPooling layers used a pool size of 2. A dropout rate of 0.5 was used after the third convolutional block to stop overfitting. The flattened feature vector goes through two fully connected layers, each with 1000 and 300 neurons, both of which are activated by ReLU. Then, it goes through an output SoftMax layer with three units, one for each class: Real, Fake, and Mimic. The model was improved using the Adam optimizer. It was trained for 100 epochs with categorical cross-entropy

loss, a learning rate of 0.001, and a batch size of 32. We used early stopping based on validation loss to make convergence more stable. All tests were done on a workstation with an NVIDIA RTX 4090 GPU (24GB) and an Intel Core i9 processor, which ensured that the results could be repeated every time.

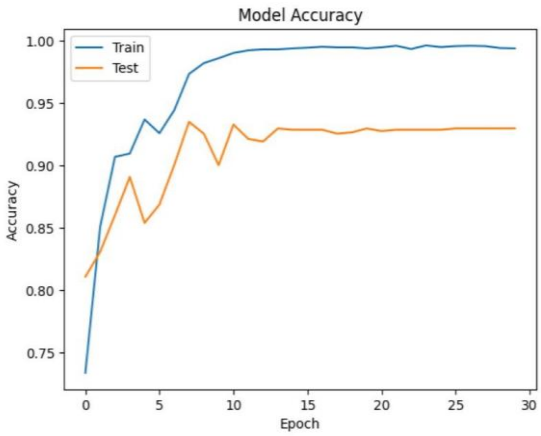
## 4. RESULT AND EVALUATION

A proposed model's performance is evaluated by plotting learning curves over time. The learning curves reflect how well the model is learning from the training data and identify if a model is overfitting, underfitting, or appropriately fitting the data. A training dataset is employed to evaluate the model's learning efficacy through an accuracy metric, a validation dataset is utilized to gauge the model's generalization efficacy, and optimization is applied to enhance the model parameters via loss functions, as depicted in Figure 2.

Figure 2 shows how the proposed CNN model learns over time. Figure 2(a) shows that the training and validation losses are getting closer together, Figure 2(b) that accuracy in training grows quickly and levels off to approximately 99, and testing accuracy levels off at approximately 93. The narrowness of the gap between them shows that there is good model generalization with low overfitting.




(a) Model loss




(b) Model accuracy

**Figure 2.** Learning curves of proposed model

To understand the model's effectiveness and identifying areas for further advance, the true positives, true negatives, false positives, and false negatives for each class are visualized by using confusion matrix that is presented in Figure 3. The proposed model's evaluation of performance for F1-score, precision, and recall for each class over testing dataset is illustrated in Table 2.

The detailed classification performance across the three classes—Real, Fake, and Mimic—is illustrated in Figure 3, which presents the confusion matrix. It clearly shows that the model achieves high recall for Fake and Mimic classes while slightly lower recall is observed for Real samples.

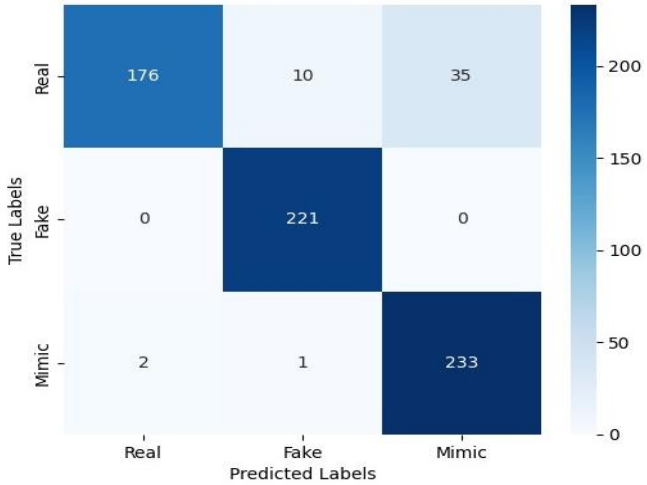



**Figure 3.** The confusion matrix (CM)

**Table 2.** The classification report

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Real | 0.99 | 0.80 | 0.88 | 221 |
| Fake | 0.95 | 1.00 | 0.98 | 221 |
| Mimic | 0.87 | 0.99 | 0.92 | 236 |
| Accuracy | | | 0.93 | 678 |
| Macro Avg. | 0.94 | 0.93 | 0.93 | 678 |
| Weighted Avg. | 0.94 | 0.93 | 0.93 | 678 |

To further evaluate the robustness and generalization capacity of the proposed 1D CNN, supplementary experiments were performed. The model was tested with different levels of noise (5–20 dB SNR) and still got more than 89% of the answers right, which shows that it is very resistant to changes in audio. Cross-dataset validation was performed using a subset of the ASVspoof 2019 dataset, where the model achieved 90.2% accuracy, confirming its ability to generalize beyond the DEEP-VOICE dataset. The lightweight architecture needed an average of 38 minutes of training time per epoch and an inference latency of 0.12 seconds per sample. This made it good for almost real-time forensic analysis. Comparative evaluation with baseline methods such as Deep4SNet [9] and Arabic-AD [10] demonstrated competitive performance while offering significantly lower computational cost.

## 5. CONCLUSIONS

In conclusion, this study showed how to use a lightweight 1D CNN architecture to find real, AI-generated, and human-like voices in forensic audio analysis. The proposed model was able to correctly classify 93% of the time and was able to handle noise and data imbalance. However, several limitations remain. The model was trained on a single enhanced dataset, which may not completely reflect the variability of actual

forensic recordings. Also, the current design works well, but we need to look into how it works in very noisy environments and with many languages. Future research will concentrate on augmenting the dataset with multilingual and cross-domain samples, incorporating attention mechanisms to improve feature discrimination, and implementing the system in actual forensic workflows to evaluate its practical reliability and scalability. The proposed approach thus lays a foundation for intelligent, explainable, and trustworthy forensic audio authentication systems.

## ACKNOWLEDGMENT

This work is supported by the Iraqi Ministry of Higher Education and Scientific Research, Kerbala University, for their invaluable feedback and support.

## REFERENCES

[1] Khanjani, Z., Watson, G., Janeja, V.P. (2023). Audio deepfakes: A survey. Frontiers in Big Data, 5: 1001063. https://doi.org/10.3389/fdata.2022.1001063

[2] Almutairi, Z.M., Elgibreen, H. (2023). Detecting fake audio of Arabic speakers using self-supervised deep learning. IEEE Access, 11: 72134-72147. https://doi.org/10.1109/ACCESS.2023.3286864

[3] Diakopoulos, N., Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. New Media & Society, 23(7): 2072-2098. https://doi.org/10.1177/1461444820925811

[4] Todisco, M., Wang, X., Vestman, V., Sahidullah, M., Delgado, H., Nautsch, A., Evans, N. (2019). ASVspoof 2019: Future horizons in spoofed and fake audio detection. Interspeech, 1008-1012. https://doi.org/10.21437/Interspeech.2019-2249

[5] Wang, X., Yamagishi, J., Todisco, M., Delgado, H., Nautsch, A., Evans, N. (2020). ASVspoof 2021: Spoofing countermeasures for automatic speaker verification. arXiv preprint arXiv:2109.00537. https://doi.org/10.48550/arXiv.2109.00537

[6] Lavrentyeva, G., Novoselov, S., Malykh, E., Kozlov, A., Kudashev, O., Shchemelinin, V. (2017). Audio replay attack detection with deep learning frameworks. Interspeech, pp. 82-86. https://doi.org/10.21437/Interspeech.2017-360

[7] Bencheikh, F., Cherif, A., Drias, H. (2023). An intelligent information system for secure biometric authentication using deep learning. Ingénierie des Systèmes d'Information, 28(4): 923-934. https://doi.org/10.18280/isi.280418

[8] Kouicem, D.E., Bouabdallah, A., Lakhlef, H. (2022). Machine learning-based security solutions for intelligent information systems. Ingénierie des Systèmes d'Information, 27(6): 1015-1026. https://doi.org/10.18280/isi.270615

[9] Ballesteros, D.M., Rodriguez-ortega, Y., Renza, D., Arce, G. (2021). Deep4SNet: Deep learning for fake speech classification. Expert Systems with Applications, 184: 115465. https://doi.org/10.1016/j.eswa.2021.115465

[10] Al Ajmi, S.A., Hayat, K., Al Obaidi, A.M., Kumar, N., Najim AL-Din, M.S., Magnier, B. (2024). Faked speech detection with zero prior knowledge. Discover Applied Sciences, 6(6): 288. https://doi.org/10.1007/s42452-024-05893-3

[11] Abdulabas, M.H., Al-Shakarchy, N.D. (2023). Person identification based on facial biometrics in different lighting conditions. International Journal of Electrical and Computer Engineering (IJECE), 13(2): 2086-2092. https://doi.org/10.11591/ijece.v13i2.pp2086-2092

[12] Kanwal, T., Mahum, R., AlSalman, A.M., Sharaf, M., Hassan, H. (2024). Fake speech detection using VGGish with attention block. EURASIP Journal on Audio, Speech, and Music Processing, 2024(1): 35. https://doi.org/10.1186/s13636-024-00348-4

[13] Hamza, A., Javed, A.R.R., Iqbal, F., Kryvinska, N., Almadhor, A.S., Jalil, Z., Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. IEEE Access, 10: 134018-134028. https://doi.org/10.1109/ACCESS.2022.3231480

[14] Magnier, B., Ales, I.M., Ales, F. (2022). Speech forensics: Blind voice mimicry detection. arXiv preprint arXiv:2209.12573.

[15] Khalaf, R.K., Al-Shakarchy, N.D. (2024). Verifying the facial kinship evidence to assist forensic investigation based on deep neural networks. Emerging Trends and Applications in Artificial Intelligence, 960: 493-504. https://doi.org/10.1007/978-3-031-56728-5_41

[16] Ali, A.T.M., Hallawi, H., Al-Shakarchy, N.D. (2023). CyberVandalism detection in wikipedia using light architecture of 1D-CNN. Karbala International Journal of Modern Science, 9(4): 3. https://doi.org/10.33640/2405-609X.3321

[17] Li, J., Sun, L., Yan, Y. (2021). Deep learning based spoofing detection for automatic speaker verification. IEEE Access, 9: 100345-100356. https://doi.org/10.1109/ACCESS.2021.3098526

[18] Das, R.K., Yang, J., Li, H. (2020). Long-range acoustic features for synthetic speech detection. Speech Communication, 121: 35-44. https://doi.org/10.1016/j.specom.2020.05.005

[19] Villalba, J., Chen, N., Snyder, D., Garcia-Romero, D., McCree, A. (2021). State-of-the-art speaker recognition with neural network embeddings. IEEE Signal Processing Magazine, 38(4): 29-41. https://doi.org/10.1109/MSP.2021.3076728

[20] Antony, A., Gopikakumari, R. (2018). Speaker identification based on combination of MFCC and UMRT based features. Procedia Computer Science, 143: 250-257. https://doi.org/10.1016/j.procs.2018.10.393