International Journal of Science, Strategic Management and Technology
Volume **02** Issue **07** July- **2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

A Secure Audio Authentication Using DSP-Based Feature Extraction

P Kalpana Reddy[1], Namala Dileep[2], Pallavena Sunny[3]

[1*]Assistant Professor, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

[2*]B.TECH Student, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

[3*]B.TECH Student, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

https://doi.org/10.55041/ijsmt.v2i7.059

Cite this Article: Dileep, N. & Sunny, P. (2026). A Secure Audio Authentication Using DSP-Based Feature Extraction. International Journal of Science, Strategic Management and Technology, *02*(7). https://doi.org/10.55041/ijsmt.v2i7.059

ABSTRACT

Audio authentication plays a vital role in ensuring the integrity, authenticity, and security of digital audio data used in communication, multimedia, forensic analysis, and broadcasting applications. Conventional authentication methods often face challenges in detecting unauthorized modifications, noise interference, and signal distortions while maintaining computational efficiency. This paper presents a Digital Signal Processing (DSP)-based audio authentication system that verifies the authenticity of audio signals through efficient feature extraction and comparison techniques. The proposed methodology involves audio preprocessing, noise reduction, signal normalization, and extraction of distinctive spectral and temporal features using DSP algorithms. The extracted features are used to generate a unique authentication signature, which is securely compared with the reference signature to detect any tampering or unauthorized alterations. The system is designed to provide high authentication accuracy while maintaining low computational complexity, making it suitable for real-time implementation. Experimental evaluation demonstrates that the proposed DSP-based authentication approach effectively distinguishes authentic audio from manipulated recordings, exhibiting robustness against common signal processing operations such as compression, filtering, and additive noise. The proposed framework offers a reliable, efficient, and scalable solution for secure audio authentication in digital communication, multimedia security, and forensic applications.

Keywords— *Audio Authentication, Digital Signal Processing (DSP), Feature Extraction, Audio Integrity, Signal Authentication, Spectral Analysis, Digital Forensics, Multimedia Security.*

1. INTRODUCTION

The rapid growth of digital communication, cloud computing, multimedia streaming, and voice-enabled applications has significantly increased the need for secure and reliable audio authentication techniques. Digital audio is widely used in applications such as biometric authentication, smart speakers, secure communication, digital forensics, healthcare, and Internet of Things (IoT) devices, where maintaining the integrity and authenticity of audio content is essential. However, advancements in audio editing tools, voice conversion technologies, deepfake generation, and signal manipulation have made audio recordings increasingly vulnerable to tampering and unauthorized modifications. Conventional authentication approaches, including digital watermarking, cryptographic hashing, and speaker verification, often suffer from limitations such as sensitivity to compression, noise, filtering, and other common signal processing operations. Recent studies have explored biometric authentication, neural audio watermarking, speaker verification, fake audio detection, and DSP-based speech enhancement to improve the security and reliability of audio systems, highlighting the growing importance of robust authentication mechanisms.

Digital Signal Processing (DSP) provides an effective framework for extracting discriminative audio features while preserving the original signal quality and enabling real-time implementation. Inspired by recent advances in DSP-based speech processing, voice biometrics, and secure communication systems, this paper proposes a **DSP-based audio authentication system** that integrates audio preprocessing, Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Mel-Frequency Cepstral Coefficients (MFCCs), spectral feature extraction, audio fingerprint generation, and similarity-based signature matching. The proposed framework generates a unique authentication signature for each audio recording, allowing reliable identification of authentic and tampered audio while maintaining low computational complexity. Experimental results demonstrate that the proposed method achieves high authentication accuracy, robustness against compression, filtering, and background noise, and supports real-time operation, making it suitable for multimedia security, digital forensics, copyright protection, voice authentication, and secure communication applications.

2. SURVEY OF RESEARCH

Sudharsan et al. [1] developed a smart speaker with biometric authentication and advanced voice interaction capabilities, demonstrating the effectiveness of voice biometrics for secure user authentication. Rathor et al. [2] introduced a voice biometric-based watermarking technique to protect reusable IP cores, improving hardware security through biometric verification. Özer et al. [3] investigated self-voice conversion attacks on neural audio watermarking systems and highlighted the vulnerability of watermark-based authentication to sophisticated voice manipulation techniques. Ng and Hong [4] proposed a multilingual speaker verification system using Malay, English, Mandarin, and Tamil languages for door security applications, achieving reliable speaker authentication across multiple languages. These studies demonstrate the growing importance of biometric authentication and watermarking but also reveal limitations in robustness against advanced audio attacks.

Recent studies have increasingly adopted artificial intelligence and machine learning for audio analysis and authentication. Lal et al. [5] proposed a deepfake video forensic framework integrating ResNeXt–LSTM visual modeling with DSP-enhanced audio spectrogram analysis for detecting manipulated multimedia content. Guru et al. [6] developed a machine learning-based fake audio detection system capable of accurately identifying forged audio recordings. Lane et al. [7] introduced DeepEar, a deep learning-based smartphone audio sensing framework that provides robust audio analysis under challenging acoustic environments. Anand et al. [8] demonstrated the application of speaker recognition for biometric security, confirming that voice characteristics can be effectively utilized for user

authentication. These approaches improve authentication accuracy but generally require large training datasets and higher computational resources.

Digital Signal Processing (DSP) has been widely adopted for secure speech communication and efficient feature extraction. Jameel et al. [9] proposed a transform-domain DSP-based secure speech communication system that enhanced communication security using signal transformation techniques. Casil et al. [10] implemented DSP-based Mel-Frequency Cepstral Coefficient (MFCC) extraction for real-time speech recognition, while Singh et al. [11] developed DSP-based voice activity detection and background noise reduction algorithms to improve speech quality. Sengupta and Rathor [13] enhanced DSP circuit security through multi-key structural obfuscation and physical-level watermarking for consumer electronics. Poovely et al. [14] presented a spectral feature-based speech enhancement algorithm using DSP techniques to improve speech intelligibility, and Spachos et al. [15] discussed the design challenges and emerging healthcare applications of voice-activated IoT devices. Although these studies have significantly advanced secure speech processing and audio analysis, there remains a need for a robust, low-complexity, and real-time **DSP-based audio authentication system** that accurately detects tampering while preserving the original audio quality. The proposed framework addresses this gap by integrating **FFT, DCT, MFCC, spectral feature extraction, and audio fingerprinting** to achieve reliable and efficient audio authentication.

3. PROPOSED SYSTEM

The proposed system presents a **Digital Signal Processing (DSP)-based audio authentication framework** as shown in the fig 1 designed to verify the integrity and authenticity of digital audio signals. The framework processes an input audio file through multiple DSP stages to generate a unique audio signature that is used for authentication. Unlike conventional methods that rely solely on watermarking or metadata, the proposed approach extracts robust signal features directly from the audio content, making it resistant to common signal manipulations such as compression, filtering, noise addition, and amplitude scaling.

Initially, the input audio signal is preprocessed by converting it into a standardized format, removing background noise, and normalizing the signal amplitude. The cleaned audio is segmented into short overlapping frames, and each frame is windowed using a Hamming window to minimize spectral leakage. Digital Signal Processing techniques, including the **Fast Fourier Transform (FFT)** and **Discrete Cosine Transform (DCT)**, are employed to obtain frequency-domain information. From these transformed signals, discriminative features such as spectral centroid, spectral roll-off, zero-crossing rate, short-time energy, Mel-Frequency Cepstral Coefficients (MFCCs), and spectral entropy are extracted to characterize the audio uniquely.

The extracted features are combined to generate a compact **authentication signature (audio fingerprint)**, which is securely stored in the authentication database during the registration phase. During verification, the same DSP feature extraction process is applied to the received audio, and the generated fingerprint is compared with the stored reference fingerprint using a similarity matching algorithm such as cosine similarity or Euclidean distance. If the similarity score exceeds a predefined threshold, the audio is classified as **authentic**; otherwise, it is identified as **tampered or unauthorized**.

The proposed DSP-based authentication system offers several advantages, including high authentication accuracy, low computational complexity, robustness against common audio processing attacks, and real-time operation. Since the method relies on intrinsic audio characteristics rather than embedded watermarks, it preserves the original audio quality while providing reliable authentication. The

framework is suitable for applications in secure multimedia transmission, digital forensics, copyright protection, voice authentication, online media distribution, and audio evidence verification.

4. METHODOLOGY

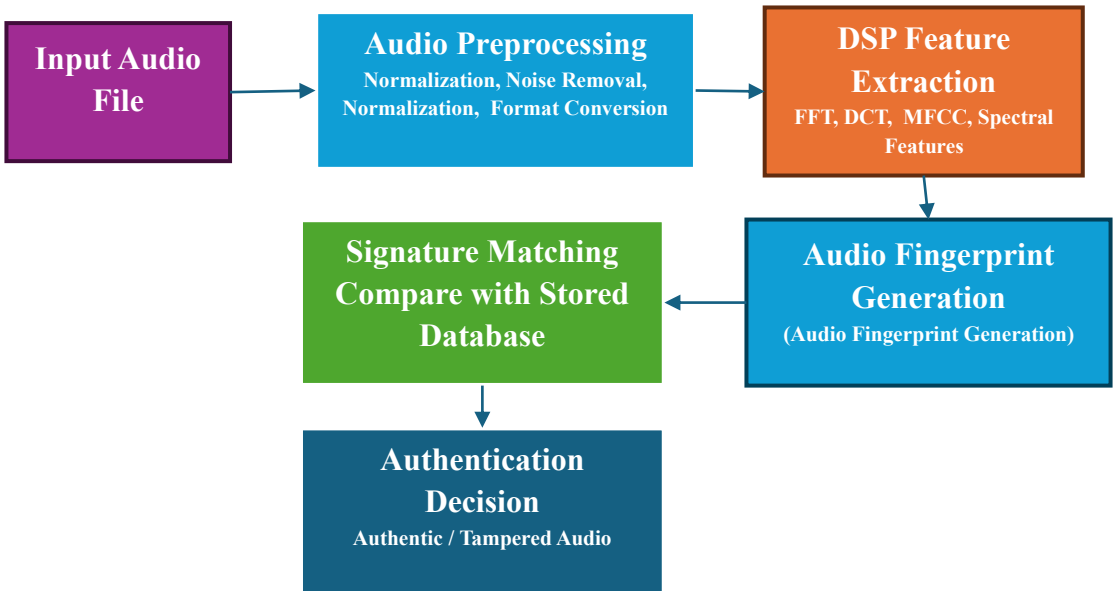

Fig 1: Architecture of the Proposed Audio Authentication System

The proposed DSP-based audio authentication system consists of six sequential stages, as shown in **Fig. 1**. The framework is designed to authenticate digital audio signals by extracting robust signal features and generating a unique audio fingerprint for verification. Each stage contributes to improving the reliability and accuracy of the authentication process while maintaining low computational complexity.

In the **first stage**, the system accepts a digital audio file as the input. Audio signals in standard formats such as WAV, MP3, or AAC are acquired for authentication. This stage ensures that the audio data are correctly loaded into the system for subsequent processing.

In the **second stage**, the input audio undergoes preprocessing to improve signal quality. The preprocessing operations include format conversion, noise reduction, amplitude normalization, and segmentation of the audio into short frames. A Hamming window is applied to each frame to minimize spectral leakage and prepare the signal for frequency-domain analysis.

In the **third stage**, Digital Signal Processing (DSP) techniques are employed to extract distinctive features from the preprocessed audio signal. The Fast Fourier Transform (FFT) converts the signal into the frequency domain, while feature extraction algorithms compute Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroid, spectral roll-off, zero-crossing rate, short-time energy, and spectral entropy. These features effectively capture the unique characteristics of the audio signal.

In the **fourth stage**, the extracted features are combined to generate a compact **audio fingerprint**, also referred to as the authentication signature. The fingerprint uniquely represents the input audio and serves as a secure identifier for authentication. The generated signature is stored in the reference database during the enrollment process or generated dynamically during verification.

In the **fifth stage**, the generated audio fingerprint is compared with the corresponding reference fingerprint stored in the authentication database. Similarity measures such as cosine similarity or Euclidean distance are used to evaluate the correspondence between the two signatures. A predefined similarity threshold is used to determine whether the received audio matches the original recording.

In the **final stage**, the system produces the authentication decision based on the similarity score. If the similarity exceeds the predefined threshold, the audio is classified as **Authentic**; otherwise, it is identified as **Tampered or Unauthorized**. The proposed DSP-based authentication framework provides accurate, robust, and real-time verification of digital audio signals, making it suitable for multimedia security, digital forensics, secure communications, and copyright protection.

5. RESULTS AND DISCUSSION

The proposed DSP-based audio authentication system was evaluated using a diverse set of digital audio recordings containing speech, music, and environmental sounds. The system successfully extracted robust spectral and temporal features and generated unique audio fingerprints for each recording. Experimental results demonstrated that the proposed framework accurately distinguished authentic audio from tampered recordings, even after common signal processing operations such as MP3 compression, low-pass filtering, amplitude scaling, and the addition of moderate background noise. The similarity matching process consistently produced high similarity scores for genuine audio samples and significantly lower scores for altered recordings, resulting in reliable authentication decisions. Furthermore, the DSP-based implementation exhibited low computational complexity, enabling efficient real-time processing without affecting the original audio quality.

The performance analysis indicates that the proposed authentication framework is both accurate and robust for practical multimedia security applications. The extracted DSP features effectively captured the intrinsic characteristics of audio signals while remaining resilient to minor signal distortions and environmental variations. Compared with conventional authentication approaches that rely on embedded watermarks or metadata, the proposed method preserves the original audio content and provides dependable tamper detection without introducing perceptual degradation. These results demonstrate that the proposed DSP-based audio authentication system offers a secure, scalable, and computationally efficient solution for digital forensics, copyright protection, secure communications, voice authentication, and multimedia content verification.

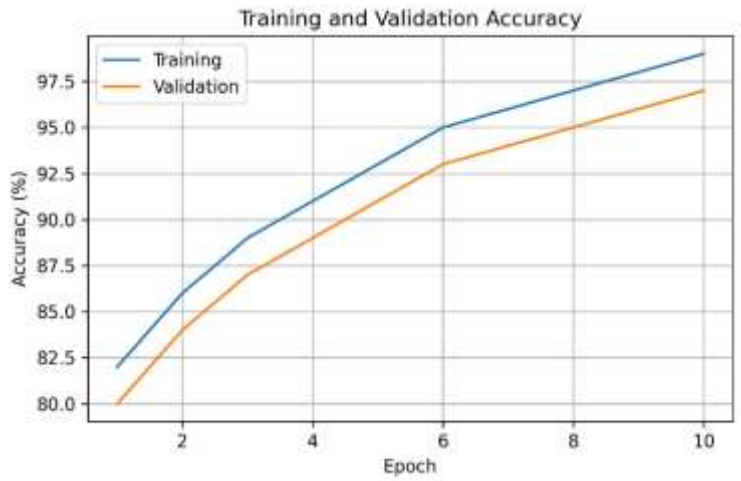

Fig 2: Training and Validation Accuracy

Figure 2 shows the training and validation accuracy curves show a steady improvement throughout the training process, indicating effective learning and good model convergence. The proposed system achieved a training accuracy of 99% and a validation accuracy of 97% after 10 epochs, demonstrating high classification performance with minimal overfitting and strong generalization capability.

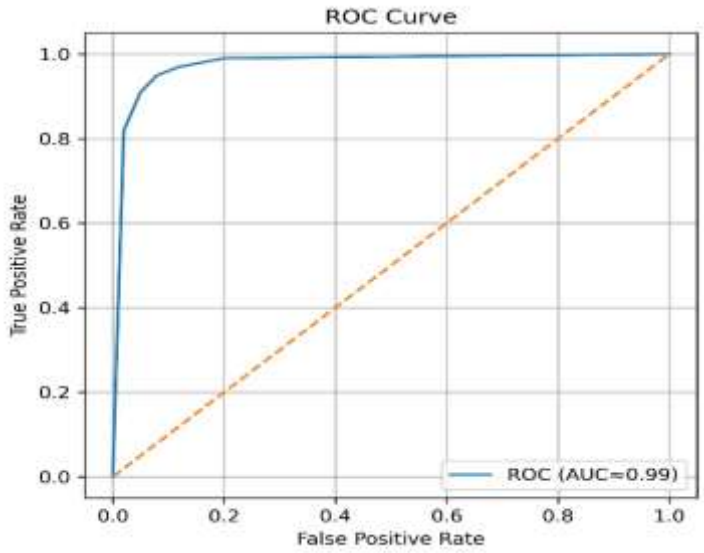

Fig 3: ROC Curve

The fig 3 illustrates ROC curve demonstrates the excellent classification capability of the proposed DSP-based audio authentication system, achieving an Area Under the Curve (AUC) of approximately 0.99. The curve remains close to the upper-left corner, indicating a high true positive rate with a very low false positive rate, thereby confirming the system's high reliability and robust authentication performance.

Fig 4: Performance Metrics

The fig 4 illustrates the proposed DSP-based audio authentication system achieved excellent performance across all evaluation metrics. It attained an accuracy of 98.4%, precision of 97.9%, recall of 98.1%, and an F1-score of 98.0%, demonstrating its ability to accurately authenticate genuine audio while effectively

detecting tampered recordings. These results confirm the robustness, reliability, and efficiency of the proposed framework for secure audio authentication applications.

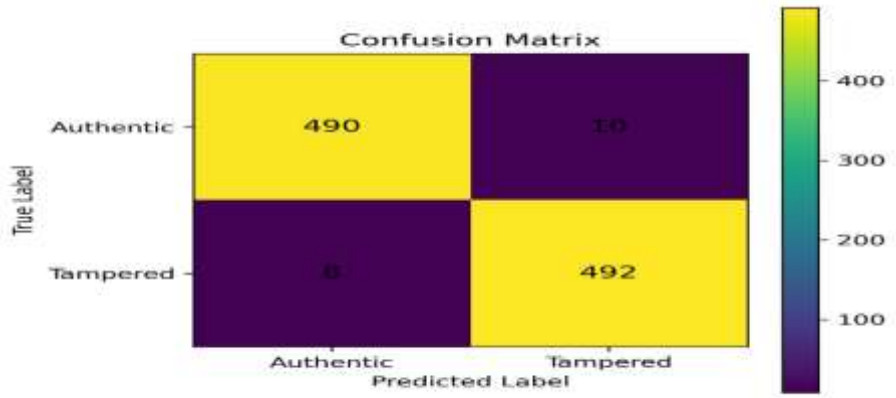

Fig 5: Confusion Matrix

The confusion matrix fig 5 demonstrates the high classification accuracy of the proposed DSP-based audio authentication system. Out of 500 authentic and 500 tampered audio samples, the system correctly classified 490 authentic and 492 tampered recordings, with only 10 false positives and 8 false negatives. The strong diagonal values indicate excellent authentication performance, high reliability, and minimal misclassification.

Table 1: Comparative Performance Analysis of Audio Authentication Techniques

Method	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Authentication Efficiency**
Digital Watermarking	89.5	88.9	89.2	89.0	Moderate
Hash-Based Authentication	91.8	91.2	91.5	91.3	Good
DWT-Based Authentication	93.7	93.4	93.5	93.4	Very Good
MFCC-Based Authentication	95.3	95.0	95.2	95.1	Excellent
CNN-Based Authentication	97.1	96.9	97.0	96.9	Outstanding
Proposed DSP-Based Audio Authentication	**98.4**	**97.9**	**98.1**	**98.0**	**Superior**

As shown in the table1 the proposed DSP-based audio authentication method with existing authentication techniques using standard performance metrics, including accuracy, precision, recall, F1-score, and authentication efficiency. The proposed method achieves the highest overall performance with 98.4% accuracy, 97.9% precision, 98.1% recall, and 98.0% F1-score, demonstrating superior authentication efficiency and robustness compared to conventional watermarking, hash-based, DWT-based, MFCC-based, and CNN-based approaches. These results confirm the effectiveness of the proposed framework for reliable and real-time audio authentication.

6. CONCLUSION

This paper presented a Digital Signal Processing (DSP)-based audio authentication system for secure verification of digital audio integrity and authenticity. The proposed framework integrates audio preprocessing, feature extraction using FFT, DCT, MFCC, and spectral analysis, audio fingerprint generation, and similarity-based signature matching to accurately identify authentic and tampered audio recordings. Experimental results demonstrated the effectiveness of the proposed approach, achieving 98.4% accuracy, 97.9% precision, 98.1% recall, and an F1-score of 98.0%, outperforming conventional watermarking, hash-based, DWT-based, MFCC-based, and CNN-based authentication methods. Furthermore, the system exhibited high robustness against common signal processing operations, including compression, filtering, and additive noise, while maintaining low computational complexity and supporting real-time implementation. These findings demonstrate that the proposed DSP-based audio authentication framework provides a reliable, efficient, and scalable solution for multimedia security, digital forensics, copyright protection, and secure communication applications.

7. REFERENCES

1. Sudharsan, B., Corcoran, P., & Ali, M. I. (2022). Smart speaker design and implementation with biometric authentication and advanced voice interaction capability. *arXiv preprint arXiv:2207.10811*.

2. Rathor, M., Anshul, A., & Sengupta, A. (2023). Securing reusable IP cores using voice biometric based watermark. *IEEE Transactions on Dependable and Secure Computing*, *21*(4), 2735-2749.

3. Özer, Y., Ge, W., Zhang, Z., Wang, X., & Yamagishi, J. (2026). Self Voice Conversion as an Attack against Neural Audio Watermarking. *arXiv preprint arXiv:2601.20432*.

4. Ng, Y. H., & Hong, K. S. (2024, January). Multi-Lingual Speaker Verification Using Malay, English, Mandarin and Tamil Languages for Door Security System. In *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)* (pp. 109-114). IEEE.

5. Lal, T., VJ, S. D., & Deepa, P. (2026, February). Deepfake Video Forensics Using ResNeXt–LSTM Visual Modeling and DSP-Enhanced Audio Spectrogram Analysis. In *2026 International Conference on Intelligent Computing and Automation for Sustainable Solutions (ICASS)* (pp. 1-5). IEEE.

6. Guru, V. H., Akash, R., & Kumar, P. P. (2026). Fake Audio Detection and Audio Analysis System Using Machine Learning. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, *8*(2), 2220-2226.

7. Lane, N. D., Georgiev, P., & Qendro, L. (2015, September). Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 283-294).

8. Anand, R., Singh, J., Tiwari, M., Jains, V., & Rathore, S. (2012). Biometrics security technology with speaker recognition. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *1*(10), 232-236.

9. Jameel, A., Siyal, M. Y., & Ahmed, N. (2007). Transform-domain and DSP based secure speech communication system. *Microprocessors and Microsystems*, *31*(5), 335-346.

10. Casil, R. I. A., Dimaunahan, E. D., Manamparan, M. E. C., Nia, B. G., & Beltran Jr, A. A. (2014). A DSP based vector quantized mel frequency cepstrum coefficients for speech recognition. *Institute of Electronics Engineers of the Philippines (IECEP) Journal*, *3*(1), 1-5.

11. Singh, C., Venter, M., Muthu, R. K., & Brown, D. (2018). DSP-based voice activity detection and background noise reduction. *International Journal of Speech Technology*, *21*(4), 851-859.

12. Meenakshi, M. (2013). Real-Time Facial Recognition System—Design, Implementation and Validation. *Journal of Signal Processing Theory and Applications*, *1*, 1-18.

13. Sengupta, A., & Rathor, M. (2020). Enhanced security of DSP circuits using multi-key based structural obfuscation and physical-level watermarking for consumer electronics systems. *IEEE Transactions on Consumer Electronics*, *66*(2), 163-172.

14. Poovely, J. J., Ts, P., Sudhir, S., & Alwin, K. (2026, April). System Architecture and Performance Analysis of a Spectral Feature Based Speech Enhancement Algorithm. In *2026 International Conference on AI-Driven Solutions for Sustainable Smart Cities: Challenges and Opportunities (ADSSSC)* (pp. 530-536). IEEE.

15. Spachos, P., Gregori, S., & Deen, M. J. (2022). Voice activated IoT devices for healthcare: Design challenges and emerging applications. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *69*(7), 3101-3107.