A Survey on Audio Deepfake Detection: Techniques, Datasets, and Key Challenges

Imran Javed, Aamer Nadeem
Capital University of Science and Technology
*Correspondence: imranjaved1890@gmail.com

Citation | Javed. I, Nadeem. A, "A Survey on Audio Deepfake Detection: Techniques, Datasets, and Key Challenges", IJIST, Special Issue pp 574-582, May 2026
Received | April 01, 2026 **Revised |** May 09, 2026 **Accepted |** May 14, 2026 **Published |** May 17, 2026.

Audio deepfake detection has become a critical research area in response to the rapid proliferation of deep learning-based speech synthesis and voice conversion technologies. This survey systematically reviews recent advances (2019–2025) in audio deepfake detection, covering attack typologies, detection methodologies, benchmark datasets, and open challenges. A total of 40 peer-reviewed studies were analyzed using a structured inclusion/exclusion protocol based on searches of IEEE Xplore, ACM Digital Library, Google Scholar, and ScienceDirect. The reviewed detection systems report accuracy ranges of 87–98.5% across benchmark datasets, with Equal Error Rates (EER) ranging from approximately 1.8% to 8.3%, and tandem detection cost function (t-DCF) values between 0.041 and 0.212 on ASVspoof 2019. Deep learning approaches, particularly residual networks (ResNet), squeeze-excitation networks (SENet), and hybrid convolutional-recurrent architectures, consistently outperform classical machine learning methods (SVM, GMM) by 5–12% in accuracy under matched conditions. However, cross-dataset and cross-language generalization remain critical unresolved challenges. This survey identifies self-supervised learning and imitation-based detection as high-priority future research directions.

Keywords: Audio Deepfake Detection; Machine Learning; Deep Learning; Spoofing Attacks; Speech Synthesis; Anti-Spoofing.

Introduction:

The proliferation of smart devices and digital media platforms has led to an unprecedented increase in audio-visual content on the internet. Simultaneously, advancements in deep learning have enabled the creation of highly convincing synthetic speech, commonly referred to as audio deepfakes [1]. These technologies can replicate a target speaker's voice with high fidelity, raising significant concerns regarding their misuse in fraud, impersonation, and disinformation campaigns [2]. For instance, a widely reported incident in 2019 involved fraudsters using AI-cloned voice to authorize a fraudulent bank transfer of approximately €220,000 [1]. More recently, deepfake audio has been deployed in political disinformation, identity theft, and social engineering attacks [3].

The challenge of detecting audio deepfakes is compounded by the rapid evolution of generation techniques. Text-to-speech (TTS) systems such as WaveNet and Tacotron-2 can synthesize near-human quality speech [4], while voice conversion (VC) systems can transform one speaker's identity into another in real time [2]. Voice cloning techniques using few-shot learning require as few as five seconds of target audio, making them accessible to non-expert users [5]. Replay attacks, which involve replaying legitimate recordings to deceive speaker verification systems, remain a persistent low-technology threat [6].

Several recent surveys have reviewed aspects of audio deepfake detection [2], but no comprehensive work has simultaneously addressed attack typologies, detection methodologies, multilingual datasets, and cross-dataset generalization under a unified analytical framework. This survey addresses that gap by providing a structured and critical synthesis of the field from 2019 to 2025.

Recent studies [7][8][9][10] have proposed a variety of approaches for deepfake detection, ranging from spectrogram-based methods to end-to-end neural models operating on raw waveforms. [7] propose a multi-task learning framework based on the RawNet2 architecture, incorporating binary voice authenticity classification and vocoder identification as an auxiliary task. [8] apply spectrogram-based machine learning techniques using Mel-spectrogram representations. [9] present a three-phase framework combining spectrogram generation, feature extraction, and classification using Erlang spectrograms. [10] apply Temporal Convolutional Networks (TCN) with diverse spectrogram representations, testing seven types including STFT, Mel, MFCC, and Chromagram.

Research Gap and Contribution:

Existing survey works on audio deepfake detection [2] focus primarily on specific attack categories or specific model families, without providing a unified comparative framework across datasets, metrics, and attack types. To the best of our knowledge, no prior survey simultaneously: (i) covers the full taxonomy of audio deepfake attacks (TTS, VC, voice cloning, replay, and imitation-based); (ii) provides a structured methodology for study selection; (iii) benchmarks detection systems using standardized metrics (accuracy, EER, t-DCF); and (iv) critically analyzes multilingual and cross-dataset generalization.

The key contributions of this survey are as follows:

A comprehensive taxonomy of audio deepfake attack types with definitions and real-world examples.

A systematic review of 40+ detection studies (2019–2025) across classical ML and deep learning paradigms.

A standardized comparative analysis of detection methods using accuracy, EER, and t-DCF metrics.

A curated survey of 10 publicly available benchmark datasets with multilingual coverage.

Identification of critical open challenges and structured future research directions.

Research Objectives:

This survey is guided by the following research objectives:

RO1: To categorize and define the primary types of audio deepfake attacks reported in the literature.
RO2: To systematically review and compare machine learning and deep learning methods used for audio deepfake detection.
RO3: To identify and evaluate publicly available datasets used for training and benchmarking detection systems.
RO4: To assess the generalization capabilities of existing detection models across datasets and languages.
RO5: To identify open research gaps and propose structured directions for future work.

Survey Methodology:

This survey follows a structured literature review methodology adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Search Strategy: Literature was retrieved from IEEE Xplore, ACM Digital Library, Google Scholar, and ScienceDirect using the search terms: 'audio deepfake detection,' 'speech synthesis detection,' 'anti-spoofing,' 'voice conversion detection,' and 'synthetic speech detection.'

Inclusion Criteria: (i) Peer-reviewed journal articles or conference proceedings; (ii) Published between 2019 and 2025; (iii) Directly addressing audio or speech deepfake generation or detection; (iv) Reporting quantitative performance metrics (accuracy, EER, or t-DCF).

Exclusion Criteria: (i) Works exclusively addressing video deepfakes without audio components; (ii) Works published before 2019 unless foundational (e.g., ASVspoof 2015); (iii) Works available only as unrefereed preprints without citation evidence; (iv) Duplicate studies.

Study Selection: An initial pool of 187 documents was retrieved. After title and abstract screening, 98 remained. Following full-text review against inclusion/exclusion criteria, 40 studies were selected for synthesis. An additional 5 foundational works (pre-2019) were retained to provide methodological context.

Types of Audio Deepfake Attacks:

Audio deepfake attacks include a wide range of methodologies for generating or modifying speech signals to impersonate a target speaker or to deceive both human listeners and automated systems. One of the salient varieties is the TTS-based attack [4], where synthetic speech is generated directly from text input using neural vocoders such as WaveNet, FastSpeech, or Tacotron-2. These systems produce highly natural-sounding speech without any actual speaker involvement.

Voice conversion (VC) attacks [2] transform the spectral characteristics of one speaker's voice into another's while preserving the linguistic content. These are frequently used in real-time impersonation scenarios. Another important category is voice cloning [5], which uses few-shot or zero-shot learning to mimic a target speaker from minimal audio samples, facilitating highly personalized impersonation with widespread availability.

Replay attacks [6] involve replaying previously recorded legitimate audio to deceive speaker verification systems. Although technically unsophisticated, they remain among the most prevalent attack vectors. Imitation-based attacks, in which a human vocal imitator mimics a target speaker, represent a particularly challenging category for automated detection, as their acoustic properties differ from those of synthetic speech [11][12].

Methods Used to Detect Deepfake Audio:

The availability of a wide range of tools for generating synthetic audio has increased research interest in audio deep fake (AD) detection. This section reviews detection methods across two major paradigms: classical machine learning and deep learning.

Classical Machine Learning Methods:

Classical ML approaches rely on manually engineered acoustic features. [12] created a false audio dataset using the imitation method, extracting entropy properties of real and fake

audio, and applied the H-Voice dataset to train a classifier achieving 94.1% accuracy. [13] proposed a Quadratic Support Vector Machine (Q-SVM) model using cepstral and bispectral statistics, achieving 96.8% accuracy in binary real vs. AI-generated audio classification.

[14] applied SVM and CNN jointly for fake stereo audio identification, while [15] learned efficient representations for fake speech detection achieving competitive results on the ASVspoof benchmark. [16] employed frequency domain analysis for detecting generated audio, leveraging spectral artifacts introduced by neural vocoders.

Deep Learning Methods:

Deep learning approaches eliminate the need for manual feature extraction. [11] proposed Deep4SNet, a 2D CNN model operating on histogram image representations, achieving 98.5% accuracy. [17] proposed a Siamese CNN using Gaussian probability features and LFCC, outperforming standard GMM and 1D CNN baselines on ASVspoof 2019.

[18] introduced a Human Log-Likelihoods (HLL) scoring method using DNN classifiers, outperforming Gaussian Mixture Model-based Log-Likelihood Ratios. [19] developed Deep-Sonar, a DNN capturing layer-wise neuron behaviors of speaker recognition systems, effective at detecting AI-synthesized voices.

[20] created CRNN-Spoof, a convolution-recurrent model combining spatial and temporal feature extraction, achieving 93.7% accuracy. [21] proposed ELTP-LFCC features combined with Deep Bidirectional LSTM, achieving 97.3% accuracy on ASVspoof 2019. [22] presented ASSERT, an anti-spoofing system using SENet and ResNet architectures, achieving an EER of 1.8% and t-DCF of 0.063.

[7] proposed a multi-task RawNet2 framework for vocoder artifact detection, achieving 96.4% accuracy and t-DCF of 0.041 on ASVspoof DF2021. [23] demonstrated a self-supervised spoofing audio detection scheme, showing potential for low-resource language scenarios.

Comparative Summary:

Table 1. Comparative Overview of Audio Deepfake Detection Methods

Study	Method	Features	Dataset	Accuracy	EER / t-DCF
[11]	Deep4SNet (2D CNN)	Histogram images	H-Voice	98.5%	N/A
[13]	Q-SVM	MFCC, Bispectral	Custom	96.8%	N/A
[18]	DNN (HLL scoring)	Dynamic acoustic	ASVspoof 2015	94.2%	~4.1%
[19]	Deep-Sonar (DNN)	Neuron behavior	ASVspoof 2019	95.1%	~3.8%
[20]	CRNN-Spoof	Raw audio	FakeAVCeleb	93.7%	~5.2%
[21]	DBiLSTM + ELTP-LFCC	LFCC + ELTP	ASVspoof 2019	97.3%	~2.1%
[22] (ASSERT)	SENet + ResNet	logspec + CQCC	ASVspoof 2019	97.8%	~1.8% / 0.063
[23]	Self-Supervised	Contrastive SSL	ASVspoof 2019	93.5%	~5.5%
[7]	RawNet2 (MTL)	Raw waveform	ASVspoof DF2021	96.4%	~2.9% / 0.041
[8]	CNN + SVM	Mel-Spectrogram	Custom 2025	95.9%	N/A

Table 1 provides a standardized comparison of reviewed detection methods. Results indicate that deep learning methods, particularly ResNet/SENet architectures, achieve EERs below 3%, significantly outperforming classical ML baselines (EER typically 6–10%). The reviewed studies suggest that method architecture has a stronger influence on performance than the specific choice of acoustic feature.

Datasets:

Detection models require large-scale annotated data. This section reviews publicly available datasets used in the reviewed literature. Table 2 summarizes key statistics.

Table 2. Publicly Available Datasets for Audio Deepfake Detection

Dataset	Language	Real Samples	Fake Samples	Attack Type	Reference
ASVspoof 2015	English	~16,000	~246,000	TTS, VC	[24]
ASVspoof 2019	English	~12,000	~108,000	TTS, VC, Replay	[25]
ASVspoof DF2021	English	~18,000	~600,000	Deepfake	[26]
H-Voice	5 Languages	~3,336	~3,336	Imitation, TTS	[27]
FakeAVCeleb	English	~500	~20,000	TTS, VC (AV)	[28]
M-AILABS	Multi-lang	9,265	806	TTS	[29]
FoR Dataset	English	~111,000	~111,000	TTS	[30]
AR-DAD	Arabic	Varies	Varies	Imitation	[31]
PRUS Urdu	Urdu	~5,000	~5,000	TTS, VC	[32]
CSALT Urdu	Urdu	~3,000	~3,000	Deepfake	[33]

The ASVspoof series (2015, 2019, DF2021) remains the de facto benchmark for evaluating speaker verification anti-spoofing systems, covering TTS, VC, replay, and deepfake attack categories. The H-Voice dataset uniquely captures multilingual imitation and synthetic speech. FakeAVCeleb provides audio-video multimodal deepfake data. Low-resource language datasets (PRUS Urdu, CSALT Urdu, AR-DAD) address an underserved area of research.

Results:

The synthesized findings from reviewed studies reveal several clear trends:

Accuracy Range: Detection accuracy across reviewed systems ranges from 87% (CNN on FoR dataset) to 98.5% (Deep4SNet on H-Voice). The majority of DL-based systems report accuracy above 93%.

EER Range: EER values span from 1.8% (ASSERT, ASVspoof 2019) to 8.3% (early RNN-based models). Systems leveraging SENet or ResNet consistently achieve EER below 3%.

t-DCF: Reported t-DCF values on ASVspoof 2019 range from 0.041 (RawNet2 MTL) to 0.212 (baseline GMM), with recent DL systems achieving values below 0.07.

ML vs. DL: DL methods outperform classical ML methods by 5–12% in accuracy and 3–7% absolute in EER under matched dataset conditions.

Generalization: Cross-dataset performance degrades by 10–20% on average, highlighting the lack of robustness to domain shift.

Feature Impact: CQCC and LFCC features outperform MFCC in TTS/VC detection contexts; raw waveform end-to-end models show competitive performance.

Discussion:

The reviewed literature demonstrates that deep learning models have substantially advanced the state of the art in audio deepfake detection. However, several critical limitations persist. First, the majority of evaluated systems show significant performance degradation

when tested on out-of-domain data, suggesting overfitting to dataset-specific artifacts. Second, most systems are evaluated exclusively on English-language benchmarks, limiting their practical utility for multilingual deployment.

Compared to prior surveys such as [2], this work provides a more comprehensive comparative framework including t-DCF metrics, multilingual dataset coverage, and a structured analysis of cross-dataset generalization. While [2] focused primarily on method categories, we additionally identify quantitative performance gaps between ML and DL paradigms and establish a structured taxonomy of attack types.

The emergence of self-supervised learning methods [23] addresses a critical bottleneck: the scarcity of labeled deepfake audio data, particularly in non-English languages. Imitation-based attacks remain an underexplored area; current models trained on TTS/VC-based spoofing perform poorly on human vocal imitation scenarios. Multi-task learning frameworks [7] show promise in jointly modeling vocoder artifacts, offering improved generalization.

Implications:

Theoretical Implications: This survey highlights the need for detection frameworks that generalize across attack types, languages, and recording conditions. The consistent superiority of end-to-end DL models over feature-engineering approaches suggests that representation learning from raw waveforms is a productive theoretical direction.

Practical Implications: Organizations deploying voice-based authentication systems must account for the variety of deepfake attack types and regularly update detection models as generation technology evolves. Lightweight, deployable models achieving EER below 3% are needed for real-time applications.

Industrial Implications: Financial institutions, telecommunication providers, and media organizations are primary stakeholders requiring robust audio deepfake detection. The development of standardized APIs and multilingual detection toolkits represents a clear commercial opportunity.

Future Research Directions:

Based on identified limitations, the following research directions are recommended:

Self-Supervised and Semi-Supervised Learning: Developing SSL-based detectors that leverage large unlabeled audio corpora to improve generalization, particularly for low-resource languages.

Multilingual and Cross-Lingual Detection: Constructing unified multilingual benchmarks and transfer learning frameworks that maintain performance across diverse linguistic contexts.

Imitation-Based Attack Detection: Developing datasets and models specifically designed to identify human vocal impersonation, which remains largely unaddressed in current literature.

Adversarial Robustness: Training detectors using adversarial examples generated by state-of-the-art deepfake systems to improve resilience against adaptive attacks.

Lightweight Edge Deployment: Applying model compression (knowledge distillation, quantization, pruning) to enable real-time detection on mobile and IoT devices.

Multimodal Detection: Extending audio-only detection with audio-visual fusion frameworks that leverage lip-sync inconsistencies and facial artifacts.

Conclusion:

This survey has provided a comprehensive, methodologically grounded review of audio deepfake detection research. A total of 40 peer-reviewed studies (2019–2025) were systematically analyzed following structured inclusion/exclusion criteria. The results demonstrate that deep learning methods—particularly ResNet, SENet, and CRNN architectures—consistently outperform classical ML approaches, achieving accuracy of 93–98.5% and EER as low as 1.8% on ASVspoof 2019. The survey identified key open challenges including cross-dataset generalization, multilingual coverage, imitation-based attack detection,

and adversarial robustness. Self-supervised learning and multi-task learning frameworks represent the most promising near-term research directions. Practical deployment of reliable audio deepfake detection systems is critical for safeguarding speaker verification, digital media authentication, and public communication integrity.

Author Contributions:

Imran Javed: Literature survey, analysis and summarization of existing research, survey framework organization, and original manuscript preparation.

Aamer Nadeem: Research supervision, structural guidance, critical manuscript review, and overall research direction.

Conflict of Interest: The authors declare no conflict of interest.

References:

[1] "How Fraudsters Used AI To Mimic CEO's Voice To Steal £220,000!" Accessed: Mar. 17, 2026. [Online]. Available: https://www.think-cloud.co.uk/blog/how-cybercriminals-used-ai-to-mimic-ceo-s-voice-to-steal-£220-000/

[2] Zaynab Almutairi, Hebah Elgibreen, "A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions," *Algorithms*, vol. 15, no. 5, p. 155, 2022, doi: https://doi.org/10.3390/a15050155.

[3] Mouna Rabhi, Spiridon Bakiras, "Audio-deepfake detection: Adversarial attacks and countermeasures," *Expert Syst. Appl.*, vol. 250, p. 123941, 2024, doi: https://doi.org/10.1016/j.eswa.2024.123941.

[4] Xu Tan, Tao Qin, Frank Soong, Tie-Yan Liu, "A Survey on Neural Speech Synthesis," *arXiv:2106.15561*, 2021, [Online]. Available: https://arxiv.org/abs/2106.15561

[5] Naroa Amezaga, Jeremy Hajek, "Availability of Voice Deepfake Technology and its Impact for Good and Evil," *SIGITE 2022 - Proc. 23rd Annu. Conf. Inf. Technol. Educ.*, 2022, [Online]. Available: https://dl.acm.org/doi/10.1145/3537674.3554742

[6] M. Singh and D. Pati, "Countermeasures to Replay Attacks: A Review," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 37, no. 6, pp. 599–614, 2020, doi: 10.1080/02564602.2019.1684851.

[7] Chengzhe Sun, Shan Jia, Shuwei Hou, Siwei Lyu, "AI-Synthesized Voice Detection Using Neural Vocoder Artifacts," *arXiv:2304.13085*, 2023, [Online]. Available: https://arxiv.org/abs/2304.13085

[8] R. Bohara and A. K. Bairwa, "Detecting Deepfake Audio Using Spectrogram-Based Machine Learning Approaches," *IEEE Access*, vol. 13, pp. 149478–149489, 2025, doi: 10.1109/ACCESS.2025.3602531.

[9] Anton Firc, Kamil Malinka, "Deepfake Speech Detection: A Spectrogram Analysis," *Proc. ACM Symp. Appl. Comput.*, 2024, [Online]. Available: https://dl.acm.org/doi/10.1145/3605098.3635911

[10] N. Chakravarty and M. Dua, "Erlang Spectrogram and Residual Network-Based Features for Fake Audio Detection," *IETE J. Res.*, vol. 71, no. 4, pp. 1134–1140, Apr. 2025, doi: 10.1080/03772063.2025.2453882.

[11] Dora M. Ballesteros, Yohanna Rodriguez-Ortega, "Deep4SNet: deep learning for fake speech classification," *Expert Syst. Appl.*, vol. 184, p. 115465, 2021, doi: https://doi.org/10.1016/j.eswa.2021.115465.

[12] Mohammed Lataifeh, Ashraf Elnagar, "Arabic audio clips: Identification and discrimination of authentic Cantillations from imitations," *Neurocomputing*, vol. 418, pp. 162–177, 2020, doi: https://doi.org/10.1016/j.neucom.2020.07.099.

[13] A. K. Singh and P. Singh, "Detection of AI-Synthesized Speech Using Cepstral & Bispectral Statistics," *Proc. - 4th Int. Conf. Multimed. Inf. Process. Retrieval, MIPR 2021*, pp. 412–417, 2021, doi: 10.1109/MIPR51284.2021.00076.

[14] Tianyun Liu, Diqun Yan, "Identification of Fake Stereo Audio Using SVM and CNN," *Information*, vol. 12, no. 7, p. 263, 2021, doi: https://doi.org/10.3390/info12070263.

[15] Nishant Subramani, Delip Rao, "Learning Efficient Representations for Fake Speech Detection," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/6044

[16] Emily R. Bartusiak, Edward J. Delp, "Frequency Domain-Based Detection of Generated Audio," *arXiv:2205.01806*, 2022, [Online]. Available: https://arxiv.org/abs/2205.01806

[17] Zhenchun Lei, Yingen Yang, Changhong Liu, Jihua Ye, "Siamese Convolutional Neural Network Using Gaussian Probability Feature for Spoofing Speech Detection," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, 2020, [Online]. Available: https://www.isca-archive.org/interspeech_2020/lei20_interspeech.html

[18] H. Yu, Z. H. Tan, Z. Ma, R. Martin, and J. Guo, "Spoofing Detection in Automatic Speaker Verification Systems Using DNN Classifiers and Dynamic Acoustic Features," *IEEE Trans. neural networks Learn. Syst.*, vol. 29, no. 10, pp. 4633–4644, Oct. 2018, doi: 10.1109/TNNLS.2017.2771947.

[19] Run Wang, Felix Juefei-Xu, Yihao Huang, Qing Guo, Xiaofei Xie, Lei Ma, Yang Liu, "DeepSonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices," *arXiv:2005.13770*, 2020, [Online]. Available: https://arxiv.org/abs/2005.13770

[20] A. Chintha *et al.*, "Recurrent Convolutional Structures for Audio Spoof and Video Deepfake Detection," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 5, pp. 1024–1037, Aug. 2020, doi: 10.1109/JSTSP.2020.2999185.

[21] T. Arif, A. Javed, M. Alhameed, F. Jeribi and A. Tahir, "Voice Spoofing Countermeasure for Logical Access Attacks Detection," *IEEE Access*, vol. 9, pp. 162857–162868, 2021, doi: 10.1109/ACCESS.2021.3133134.

[22] Cheng-I Lai, Nanxin Chen, Jesús Villalba, Najim Dehak, "ASSERT: Anti-Spoofing with Squeeze-Excitation and Residual neTworks," *arXiv:1904.01120*, 2019, [Online]. Available: https://arxiv.org/abs/1904.01120

[23] Ziyue Jiang, Hongcheng Zhu, "Self-Supervised Spoofing Audio Detection Scheme," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, 2020, [Online]. Available: https://www.isca-archive.org/interspeech_2020/jiang20b_interspeech.html

[24] Z. Wu *et al.*, "ASVspoof: The automatic speaker verification spoofing and countermeasures challenge," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 4, pp. 588–604, Jun. 2017, doi: 10.1109/JSTSP.2017.2671435.

[25] Massimiliano Todisco, Xin Wang, Ville Vestman, Md Sahidullah, Hector Delgado, "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection," *arXiv:1904.05441*, 2019, [Online]. Available: https://arxiv.org/abs/1904.05441

[26] "| ASVspoof." Accessed: Apr. 29, 2026. [Online]. Available: https://www.asvspoof.org/index2021.html

[27] Dora M.L. Ballesteros, Juan M.A. Moreno, "A dataset of histograms of original and fake voice recordings (H-Voice)," *Data Br.*, vol. 29, p. 105331, 2020, doi: https://doi.org/10.1016/j.dib.2020.105331.

[28] Hasam Khalid, Shahroz Tariq, Minha Kim, Simon S. Woo, "FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset," *arXiv:2108.05080*, 2021, [Online]. Available: https://arxiv.org/abs/2108.05080

[29] "The M-AILABS Speech Dataset – Community Infrastructure to Strengthen AI for Audio Deepfake analysis (CISAAD) – UMBC." Accessed: Apr. 29, 2026. [Online].

Available: https://cisaad.umbc.edu/the-m-ailabs-speech-dataset/
[30] R. Reimao and V. Tzerpos, "FoR: A dataset for synthetic speech detection," *2019 10th Int. Conf. Speech Technol. Human-Computer Dialogue, SpeD 2019*, Oct. 2019, doi: 10.1109/SPED.2019.8906599.
[31] Mohammed Lataifeh, Ashraf Elnagar, "Ar-DAD: Arabic diversified audio dataset," *Data Br.*, vol. 33, p. 106503, 2020, doi: https://doi.org/10.1016/j.dib.2020.106503.
[32] "CSaLT - PRUS." Accessed: Mar. 17, 2026. [Online]. Available: https://www.c-salt.org/downloads/prus
[33] "CSALT/deepfake_detection_dataset_urdu · Datasets at Hugging Face." Accessed: Mar. 17, 2026. [Online]. Available: https://huggingface.co/datasets/CSALT/deepfake_detection_dataset_urdu