

Audio Deepfake Detection Methods: A Review

W.V.M. Janani, N.B.A.S.C. Nissanka, and K.G.A. Thisaravi

Department of ICT, Faculty of Technological Studies, University of Vavuniya, Sri Lanka

Abstract

The emergence of audio deepfake technologies has raised new concerns in the area of digital security, privacy, and trust in media. Audio deepfakes are AI-generated synthetic audio clips that are capable of accurately imitating a real human voice and can be used for malicious applications like voice phishing, impersonation, and misinformation. This research presents a detection system based on Convolutional Neural Networks (CNNs) trained on multiple engineered audio features, including Mel-Frequency Cepstral Coefficients, mel-spectrograms, and chroma features. The system is evaluated with public datasets including ASVspoof 2019, WaveFake, and FoR and utilizes preprocessing techniques like normalization, resampling, and fixed-length trimming to standardize the input. The CNN model is constructed using several convolutional layers, pooling layers, and fully-connected layers trained with binary cross entropy loss, and tested and validated using a cross-validation framework. In testing, the system demonstrated high accuracy and strong generalization over several spoof types. Overall, through experimentation, this study illustrated the potential of deep learning-based audio feature analysis to achieve efficient scaling of audio deepfake detection for real-time deployment in security, forensic, and media verification applications.

Keywords: audio deepfake, machine learning, deep learning, features, voice cloning