Electronics, vol.15, no.13, pp.2846, July, 2026.

The Generalization Gap: Do Audio Deepfake Detectors Actually Protect Against Modern Vishing?

V. García Martíinez-Echevarríia; R. Palacios Hielscher; G. López López; A. Gupta

Abstract-

Voice phishing, commonly known as vishing, has become one of the fastest-growing threats in social engineering. The rapid advancement and accessibility of AI voice cloning tools have enabled attackers to produce highly convincing synthetic speech at minimal cost, driving a sharp increase in impersonation fraud. Accordingly, automatic detection of synthetic voices could contribute, as one component of a broader defense, to mitigating vishing attacks. This paper studies the automatic detection of AI-generated speech, with a particular focus on how well such detectors generalize beyond their training data to modern, unseen synthesis methods. Two detection approaches are evaluated: a Residual CNN (convolutional neural network) trained as a binary classifier on three different time–frequency representations and a one-class learning strategy with a ResNet-18 backbone, yielding four models in total. Models were trained on the well-known ASVspoof 2019 Logical Access dataset and tested on its standard partitions. Then, models were tested on the SONAR benchmark, which gathers voices generated with state-of-the-art synthesis techniques unseen during training. Experimental results show that, on the modern systems gathered in SONAR, all four configurations fall close to chance. The LFCC one-class detector generalizes comparatively best, but the apparently higher accuracy of some models reflects a tendency to label most speech as spoofed. These findings indicate that the evaluated detectors can provide, at most, a partial security layer against vishing driven by current and emerging speech-synthesis technologies, although continuous model updates are recommended.

Index Terms- AI-generated speech; spoofing detection; residual CNN (convolutional neural network); one-class learning; generalization; vishing

Citation:
García Martíinez-Echevarríia, V.; Palacios, R.; López, G.; Gupta, A. "The Generalization Gap: Do Audio Deepfake Detectors Actually Protect Against Modern Vishing?", Electronics, vol.15, no.13, pp.2846, July, 2026.