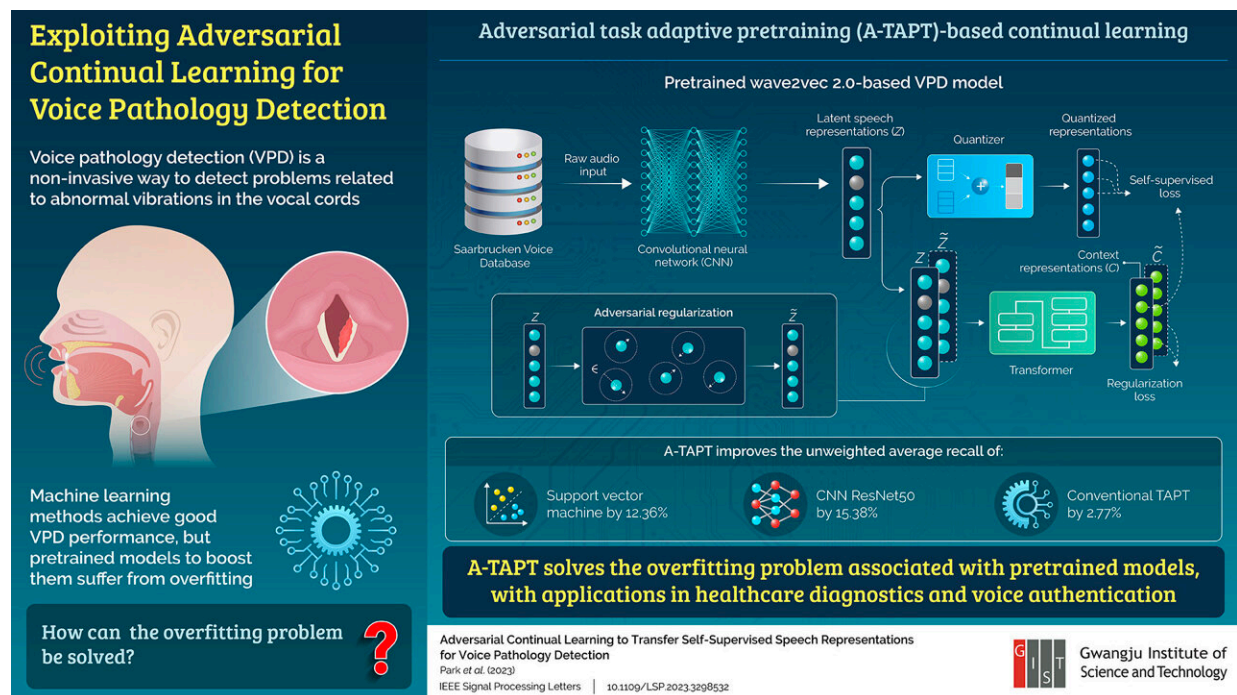


# Scientists advance voice pathology detection via adversarial continual learning

October 16 2023



Adversarial Continual Learning to Transfer Self Supervised Speech Representations for Voice Pathology Detection. Credit: Gwangju Institute of Science and Technology (GIST)

Voice pathology refers to a problem arising from abnormal conditions, such as dysphonia, paralysis, cysts, and even cancer, that cause abnormal vibrations in the vocal cords (or vocal folds). In this context, voice pathology detection (VPD) has received much attention as a non-

invasive way to automatically detect voice problems. It consists of two processing modules: a feature extraction module to characterize normal voices and a voice detection module to detect abnormal ones.

Machine learning methods, like support vector machines (SVM) and [convolutional neural networks](#) (CNN) have been successfully utilized as pathological voice detection modules to achieve good VPD performance. Also, a self-supervised, pretrained model can learn generic and rich speech feature representation, instead of explicit speech features, which further improves its VPD abilities.

However, fine-tuning these models for VPD leads to an overfitting problem, due to a domain shift from conversation speech to the VPD task. As a result, the pretrained model becomes too focused on the [training data](#) and does not perform well on new data, preventing generalization.

To mitigate this problem, a team of researchers from Gwangju Institute of Science and Technology (GIST) in South Korea, led by Prof. Hong Kook Kim, has proposed a contrastive learning method involving Wave2Vec 2.0—a self-supervised pretrained model for speech signals—with a novel approach called adversarial task adaptive pretraining (A-TAPT). Herein, they incorporated adversarial regularization during the continual learning process.

The researchers performed various experiments on VPD using the Saarbrücken Voice Database, finding that the proposed A-TAPT showed a 12.36% and 15.38% improvement in the unweighted average recall (UAR), when compared to SVM and CNN ResNet50, respectively. It also achieved a 2.77% higher UAR than the conventional TAPT learning. This shows that A-TAPT is better at mitigating the overfitting problem.

Talking about the long-term implications of this work, Mr. Park, the first author of this article, says, "In a span of five to 10 years, our pioneering research in VPD, developed in collaboration with MIT, may fundamentally transform health care, technology, and various industries. By enabling early and accurate diagnosis of voice-related disorders, it could lead to more effective treatments, improving the quality of life of countless individuals."

Their article was published in [\*IEEE Signal Processing Letters\*](#).

**More information:** Dongkeon Park et al, Adversarial Continual Learning to Transfer Self-Supervised Speech Representations for Voice Pathology Detection, *IEEE Signal Processing Letters* (2023). [DOI: 10.1109/LSP.2023.3298532](#)

Provided by Gwangju Institute of Science and Technology

Citation: Scientists advance voice pathology detection via adversarial continual learning (2023, October 16) retrieved 12 May 2024 from <https://techxplore.com/news/2023-10-scientists-advance-voice-pathology-adversarial.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--