

PRESS RELEASE

GIST Scientists Advance Voice Pathology Detection via Adversarial Continual Learning

The proposed adversarial task adaptive pretraining mitigates the overfitting problem associated with pretrained models used to detect voice problems

Voice pathology detection (VPD) is a non-invasive procedure that detects voice problems, due to conditions such as cancer and cysts, by differentiating abnormal vibrations in the vocal cords. But pretrained VPD models can show a dip in performance due to overfitting when they are self-supervised. Recently, researchers from Gwangju Institute of Science and Technology (GIST) and Massachusetts Institute of Technology (MIT) have developed an adversarial task adaptive pretraining approach, which enhances the performance of a VPD model by over 15%, compared to the state-of-the-art ResNet50 model.

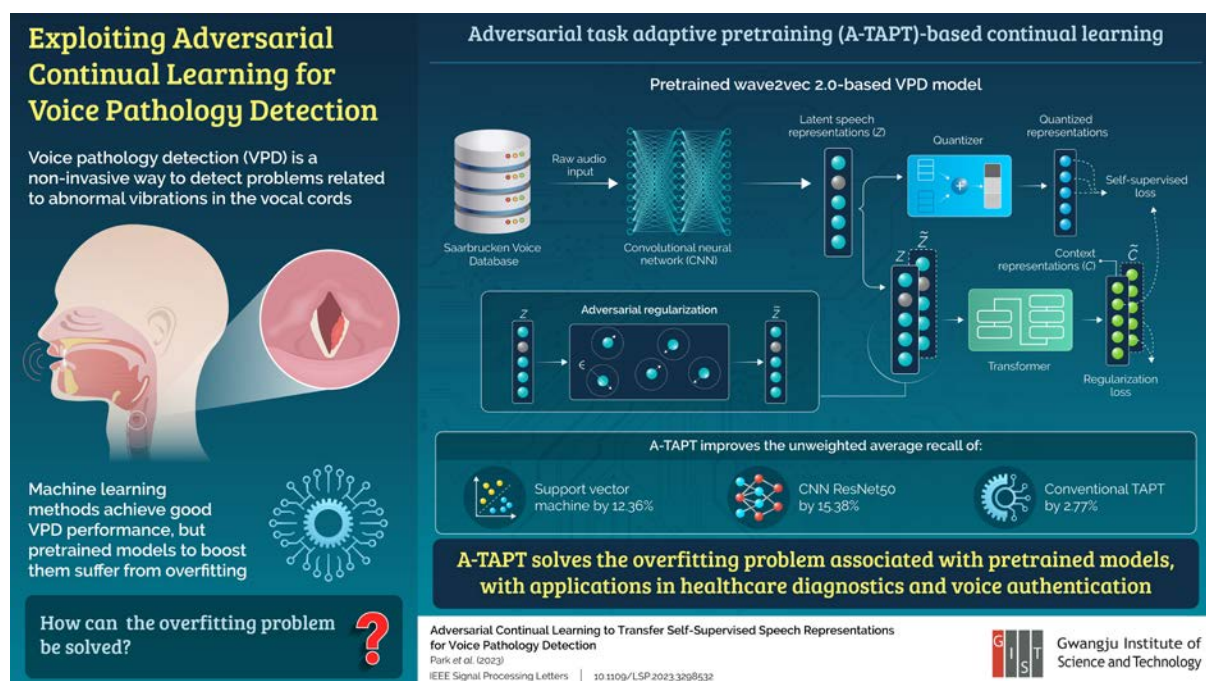


Image title: Diagram of the wave2vec 2.0 model on adversarial continual learning

Image caption: The proposed adversarial task adaptive pretraining approach outperforms state-of-the-art alternatives for voice pathology detection, paving the way for advanced healthcare.

Image credit: *IEEE Signal Processing Letters*

Image source link: <https://ieeexplore.ieee.org/document/10192328>

License type: Licensed content; Copyright of IEEE

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 groundbreaking 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 says who is the first author of this article: *“In a span of five to 10 years, our pioneering research in VPD, developed in collaboration with MIT, may fundamentally transform healthcare, 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 made available online on 24 July 2023 and published in Volume 30 of the journal [IEEE Signal Processing Letters](#). Their research, performed as part of a GIST funded project entitled ‘Extending Contrastive Learning to New Data Modalities and Resource-Limited Scenarios’ in collaboration with the MIT, Cambridge, MA, USA, embarks on a path that promises to redefine the landscape of VPD and artificial intelligence in medical applications. The project team includes Hong Kook Kim (EECS, GIST) and Dina Katabi (EECS, MIT) as Principal Investigators (PIs) as well as Jeany Son (AI Graduate School, GIST), Moongu Jeon (EECS, GIST), and Piotr Indyk (EECS, MIT) as co-PIs.

Prof. Kim points out: *“Our partnership with MIT has been instrumental in this success, facilitating ongoing exploration of contrastive learning. The collaboration is more than a mere partnership; it’s a fusion of minds and technologies that strive to reshape not only medical applications but various domains requiring intelligent, adaptive solutions.”*

Furthermore, it is promising for health monitoring in vocally demanding professions like call center agent, ensuring robust voice authentication in security systems, making artificial intelligence voice assistants more responsive and adaptive, and developing tools for voice quality enhancement in the entertainment industry.

Here's hoping for further innovation in the field of self-supervised learning and contrastive learning!

Reference

Title of original paper: Adversarial Continual Learning to Transfer Self Supervised Speech Representations for Voice Pathology Detection

Journal: *IEEE Signal Processing Letters*

DOI: 10.1109/LSP.2023.3298532

*Corresponding author's email: hongkook@gist.ac.kr

Your Press Release Source

Gwangju Institute of Science and Technology

About the Gwangju Institute of Science and Technology (GIST)

The Gwangju Institute of Science and Technology (GIST) was founded in 1993 by the Korean government as a research-oriented graduate school to help ensure Korea's continued economic growth and prosperity by developing advanced science and technology with an emphasis on collaboration with the international community. Since that time, GIST has pioneered a highly regarded undergraduate science curriculum in 2010 that has become a model for other science universities in Korea. To learn more about GIST and its exciting opportunities for researchers and students alike, please visit: <http://www.gist.ac.kr/>.

Website: <http://www.gist.ac.kr/>

About Prof. Hong Kook Kim

Prof. Hong Kook Kim received a B. S. degree in Control and Instrumentation Engineering from Seoul National University, Korea, in 1988. He then received both M.S. and Ph.D. degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea in 1990 and 1994, respectively. He is currently an IEEE Senior Member and an affiliate member of IEEE Speech and Language Technical Committee. His current research interests include large vocabulary speech recognition, audio coding and speech/audio source separation, and embedded algorithms and solutions for speech and audio processing for handheld devices.