Voice-based Artificial Intelligence Algorithm for Diagnosing Voice Disease

Hye Rim Chae, SLP, MS[1]; Jae-Yol Lim, MD, PhD[1]; Seung Jin Lee, SLP, PhD[2]; SangJae Lee, PhD[3]; ChangHyun Kim, PhD[3]; KwangSeok Lee, PhD[3]; HanSu Cho, PhD[3]; Seung-Mo Cho, PhD[3]; Young Min Park, MD, PhD[1]
[1]Department of Otorhinolaryngology, Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul, Korea
[2]Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea
[3]AIX SK Telecom

Introduction

The diagnosis of voice diseases requires assessment procedures that include objective voice tests using acoustic analysis, as well as subjective voice tests such as auditory-perceptual evaluation, questionnaires, and visual inspection using laryngeal stroboscopy (1). Although various acoustic features and subjective measures have been used to evaluate patient's vocal status, there are limitations in diagnosing voice disorders solely through these procedures.

Artificial intelligence technology has been useful in diagnosing various and predicting diseases, and in laryngology, it can be used to diagnose or predict voice diseases based on various data (2-4). Recently, ensemble learning, which combines different data or artificial intelligence models, is widely used to enhance the accuracy of artificial intelligence models for diagnosing voice diseases (5, 6).

The purpose of this study was to improve classification accuracy by using voice-based artificial intelligence algorithms for the diagnosis of voice diseases.

Methods and Materials

We used voice data (N = 12,112) collected from the Department of Otorhinolaryngology, Gangnam Severance Hospital, Yonsei University College of Medicine from 2003 to 2020 (Table 1). The data used was as follows: 1) Medical record, 2) Sustained vowel /a/ of 1 to 4 seconds long, 3) Second sentence of the Korean standard passage 'Ga-eul'.

Before learning the model, a data preprocessing process was conducted, and then built a deep learning-based CNN using the collected data. Gangnam severance Hospital dataset were used to establish classifiers and to verify the classifier's performance in the generated model.

Classification accuracy was obtained by performing ensemble learning using CNN classification algorithm, transformer, spectral analysis, and so on.

Table 1. Details of the Voice Dataset

Voice diseases	Number (N)
Normal	671
Laryngeal cancer	322
Spasmodic dysphonia	851
Vocal fold nodules	965
Vocal fold polyp	1915
Intracordal Cyst	522
Reinke's edema	391
Vocal fold Granuloma	211
Laryngeal papilloma	255
Vocal fold leukoplakia	84
Sulcus vocalis	836
Vocal folds paralysis	1264
Other	3825
Total	**12112**

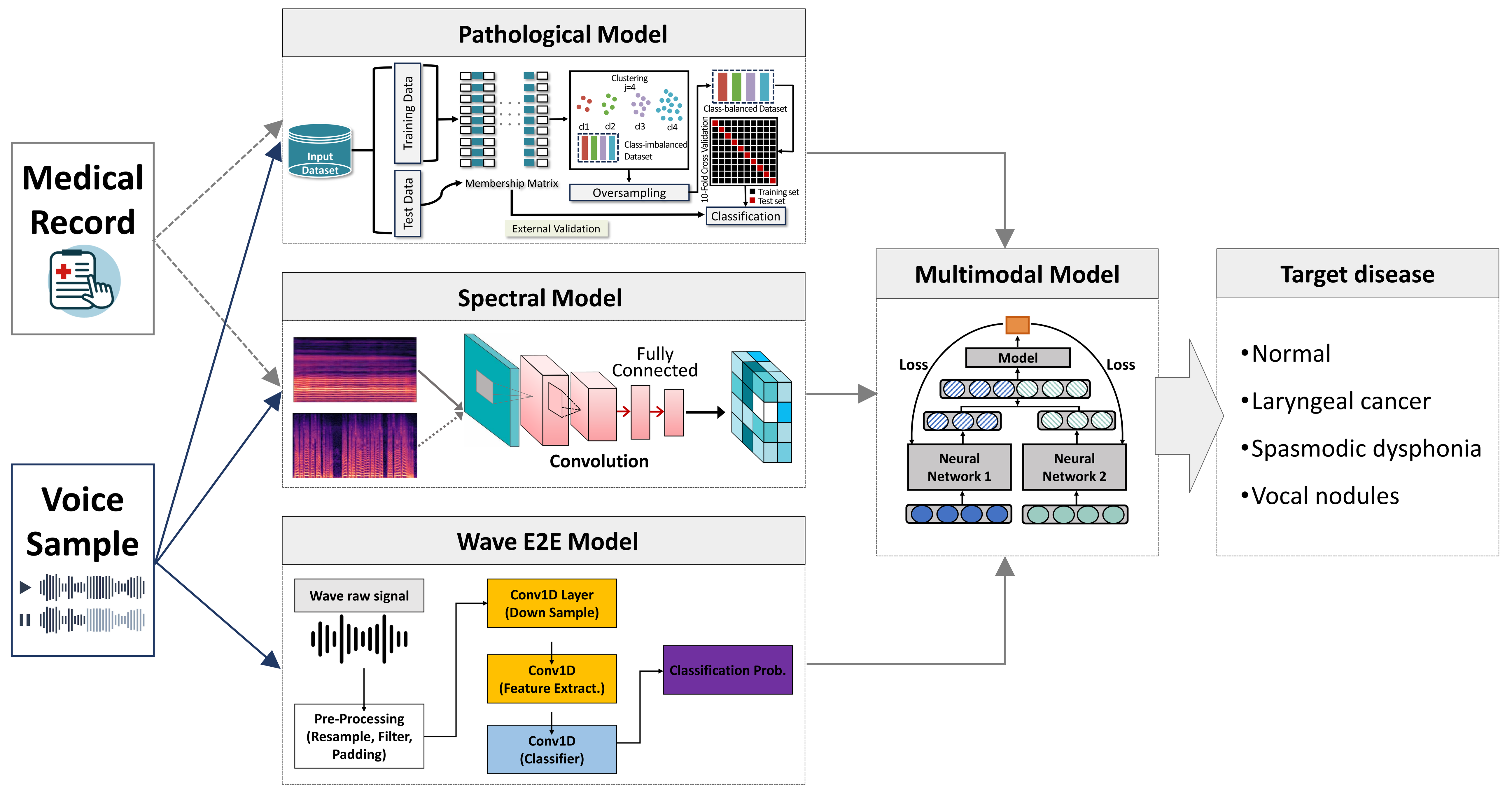

Figure 1. Overview of classifier adaptation process

Results

We used ensemble models to combine various artificial intelligence algorithms, including CNNs, transformers, and spectral analysis algorithms, to enhance the model's performance.

Additionally, during the data preprocessing process, we focused on a few target disease (laryngeal cancer, spasmodic dysphonia, vocal fold nodules) with the enough number of patients.

We obtained classification accuracy of more than 88% in the established voice classification models using our dataset. The accuracy for each target disease is as follows: Normal 98.75%, Laryngeal cancer 93.69%, Spasmodic dysphonia 92.84%, and vocal fold nodules 91.02% (Table 2).

Table 2. Classification Results

	Sensitivity	Specificity	Accuracy
Normal	0.96364	0.99545	0.98750
Laryngeal cancer	0.85455	0.96472	0.93693
Spasmodic dysphonia	0.88636	0.94242	0.92841
Vocal fold nodules	0.82273	0.93939	0.91023

Discussion

In this study, we applied ensemble learning to classify voice diseases by combining various voice-based artificial intelligence algorithms, and it was confirmed that the ensemble method improved classification accuracy.

Most studies have used binary classifier model to classify normal vs laryngeal cancer or normal vs pathological voice. Some studies have also used multi-class classifiers by refining the target diseases. Some studies have used multi-class classifiers by specifying the target diseases, with accuracy as low as 66% and as high as 89% (5, 6).

Our study used a multi-class classifier by combining several voice-based artificial intelligence algorithms and obtained a high average accuracy of over 88%.

Furthermore, the target diseases include diseases that are commonly seen in the clinic, which is clinically useful for diagnosing common voice diseases.

Conclusions

The results of our study suggest that ensemble learning aimed at training multiple classifiers is useful to obtain an increased classification accuracy.

Although a large data amount is essential for artificial intelligence analysis, when an integrated approach is taken by combining various models, high diagnostic classification accuracy can be expected.

Reference

1) Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. Am J Speech Lang Pathol. 2018;27(3):887-905.

2) Fang SH, Tsao Y, Hsiao MJ, et al. Detection of pathological voice using cepstrum vectors: A deep learning approach. J Voice. 2019;33(5):634-41.

3) Mohammed MA, Abdulkareem KH, Mostafa SA, et al. Voice pathology detection and classification using convolutional neural network model. Appl Sci (Basel). 2020;10(11): 3723.

4) Xiong H, Lin P, Yu JG, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine*. 2019;*48*, 92-99.

5) Hu HC, Chang SY, Wang CH, et al. Deep learning application for vocal fold disease prediction through voice recognition: preliminary development study. *J Med Internet Res,* 2021;*23*(6):e25247. 2021.

6) Ariyanti W, Hussain T, Wang JC, et al. Ensemble and multimodal learning for pathological voice classification. IEEE Sens Lett, 2021;5(7):1-4.