P-4

Digital Biomarker for Post-stroke Dysphagia Diagnosis via a Smart Device using Voice Spectograms

Hae-Yeon Park[1*], Geun- Young Park[1], Seungchul Lee[2,3], Heekyu Kim[2,3], DoGyeom Park[2,3], Hyemi Hwang[1], Sun Im[1†]

[1] Department of Rehabilitation Medicine, Bucheon St. Mary`s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea
[2] Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Korea
[3] Graduate School of Artificial Intelligence, Pohang University of Science and Technology (POSTECH), Pohang, Korea

Background and Objective

Patients with dysphagia show changes in articulation and voice quality, and recent studies using machine learning models have been employed to help in the classification. This study aimed to apply a novel deep learning method using only the patient`s voice to classify normal controls from dysphagia patients and determine whether this new deep learning method may help provide a rapid and accurate means to supplement the existing clinical methods in dysphagia screening and assessment.

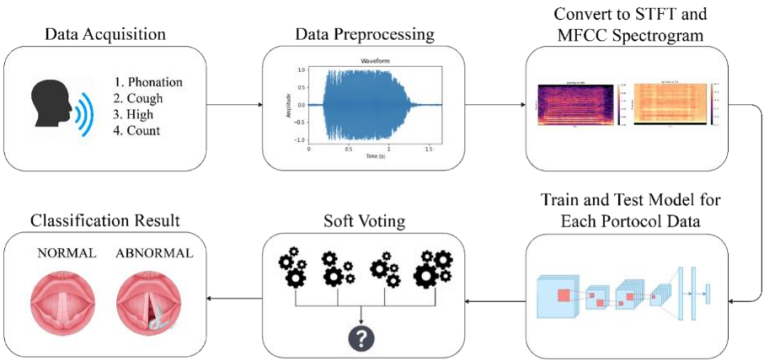
Method

Voice samples from 299 healthy controls and 290 patients with post-stroke dysphagia; who performed four simple phonation tasks; were obtained in a prospective manner at a university-affiliated hospital using a smart digital device. For this study, the participants were required to produce the following four tasks. 1) sustained vowel phonation ‘e’ for at least 3 seconds, 2) a voluntary “cough” with maximal effort as if to remove secretion, 3) pitch elevation with phonation of the “eee” with effort moving from a low to a high pitch and 4) counting from 1 to 5. Deep learning methods were employed as follows: firstly, a spectrogram is obtained through Short Time Fourier Transform (STFT) and Mel-frequency cepstral coefficients (MFCC) on a sound signal, respectively. Secondly, the STFT and MFCC spectrograms obtained for each protocol are fed to each multibranch model. Finally, during the test, each model is ensembled in a soft voting method to distinguish normal and dysphagia classes.

Result

Five evaluation metrics are used to evaluate the performance of the model: AUC, Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Among the performance metrics, sensitivity and specificity levels are compared with the existing diagnostic tools. The ensemble model incorporating all four tasks showed an AUC of 0.97±0.01, with sensitivity and specificity levels as high as 92.6% and 88.7%, respectively.

Figure 1.
Overall workflow that incorporated 4 phonation tasks and the deep learning algorithm flows to help classify dysphagia vs. control



Conclusion

The model did not require extracted voice features but instead received the spectrogram itself as an input. The model required input from 4 different phonation tasks. The methods were non-invasive, low-cost, and simple, with an automated workflow that can be used immediately in the clinical field. Our results show that the ensemble method used in this study may be utilized as a convenient and rapid biomarker of dysphagia in a non-invasive and automated manner.

Corresponding Author: Sun Im (lafolia@catholic.ac.kr)

Acknowledgments
This research was supported in part by Ministry of Trade, Industry and Energy (MOTIE) (Development of Meta Soft Organ Module Manufacturing Technology without Immunity Rejection and Module Assembly Robot System, 20012378), in part by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03047902), in part by the NRF (no. 2020R1F1A106581412), and in part by the Po-Ca Networking Groups funded by the Postech-Catholic Biomedical Engineering Institute (No. 5-2020-B0001-00050)