

by Syllable-based Speech Analysis with Daily Conversations

Kyeong Eun Uhm¹, Seokhyeon Heo¹, Doyoung Yuk¹, Bo Mi Kwon¹, Byounghyun Yoo², Jisoo Kim³, Jongmin Lee¹

¹Department of Rehabilitation Medicine, Konkuk University School of Medicine, Seoul, Korea; ²Center for Artificial Intelligence, Korea Institute of Science and Technology, Seoul, Korea; ³Department of Artificial Intelligence, Jeju National University, Jeju, Korea

Introduction

- Dysphagia, a disorder affecting the ability to swallow, has a high incidence rate among the elderly and can lead to serious health complications. Consequently, early detection of dysphagia is important. The current gold standard for diagnosing dysphagia is the Video Fluoroscopic Swallowing Study (VFSS). However, it is invasive due to radiation exposure and risk of aspiration. Recent advances in artificial neural networks have significantly improved the ability to detect dysphagia using specific acoustic signals such as phonation and coughing sound.
- In this study, we investigated accuracy of newly developed deep learning model for diagnosing dysphagia by syllable-based speech analysis using daily conversations.

Methods

- The audio data of daily conversations were obtained from 16 patients with dysphagia and 24 controls. The presence of dysphagia was determined by VFSS.
- A deep learning model was developed to effectively detect dysphagia by segmenting daily conversation into syllables using a Speech-to-Text (STT) model. These syllable-segmented data were analyzed with a Convolutional Neural Network (CNN) to perform binary classification between the dysphagia patients and control groups.
- Additionally, individual-based evaluation was performed. The determination of dysphagia was made by a proportion of syllable-segmented data classified as dysphagia in each individual's conversation, threshold was set at 0.5.

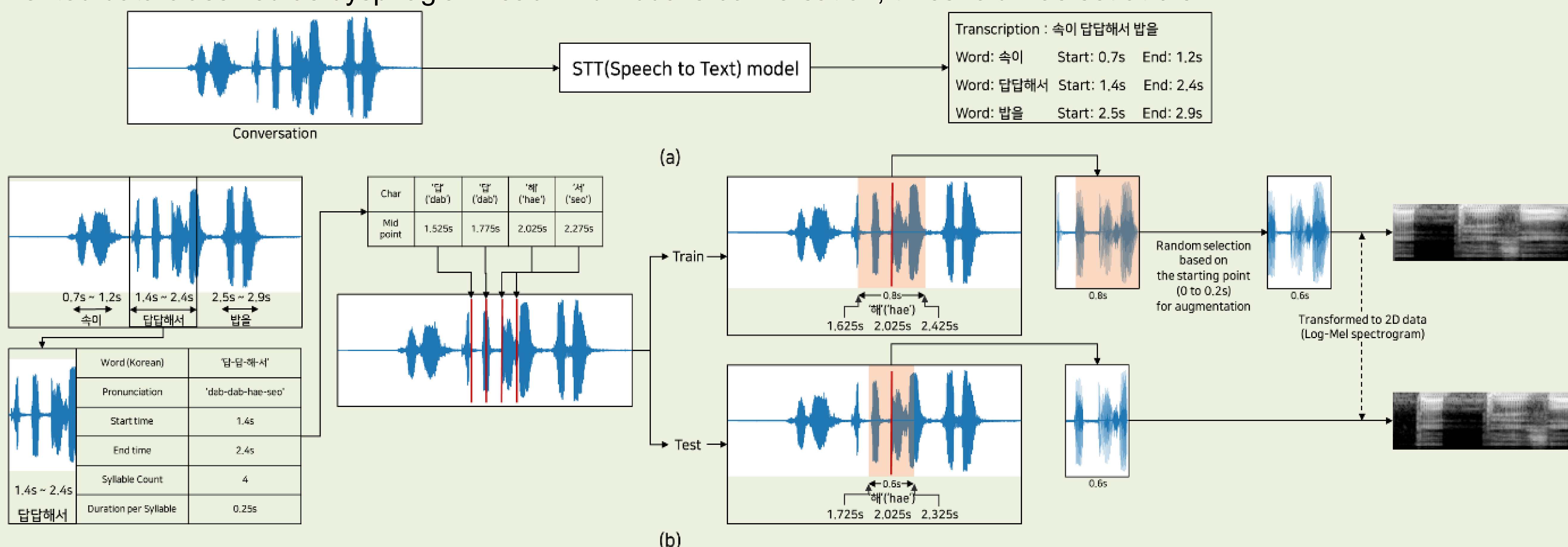


Figure 1. Method of segmenting conversations into syllable units

Results

Table 1. Demographic features of participants and collected data

	Enrolled control	Collected control (YouTube)	Dysphagia
N	16	8	16
Age (years), mean $\pm$ SD	66.3 $\pm$ 12.6	61.4 $\pm$ 9.1	67.7 $\pm$ 16.1
Male:female (N)	9:7	8:0	11:5

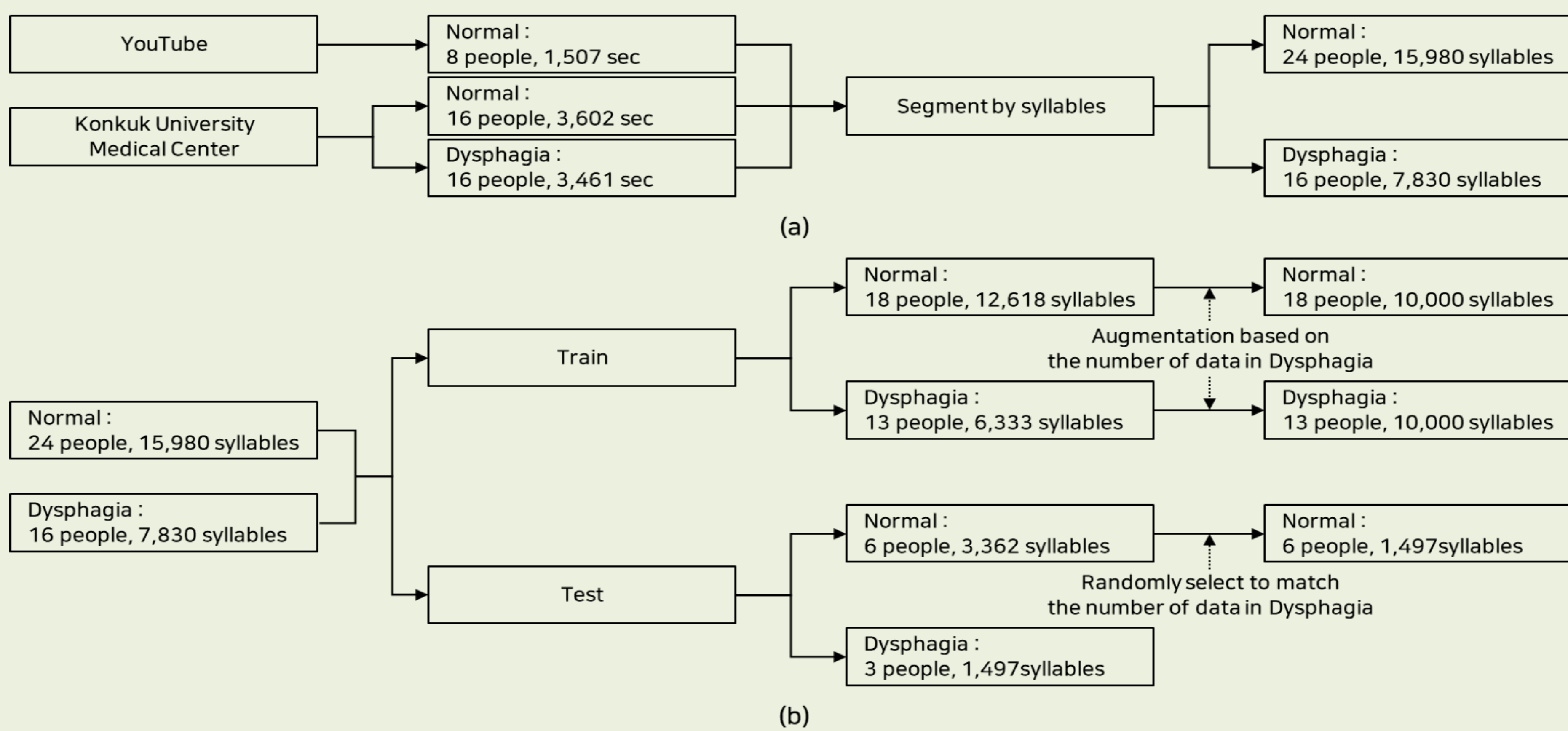


Figure 2. Construction of dataset (a) syllable-based segmentation (b) train and test set

Actual	Normal	1028	469
	Dysphagia	148	1349
		Normal	Dysphagia
		Prediction	

- Accuracy=0.7939
- Sensitivity=0.9011
- Specificity=0.6867
- PPV=0.7420
- NVP=0.8741

Figure 4. Syllable-level classification

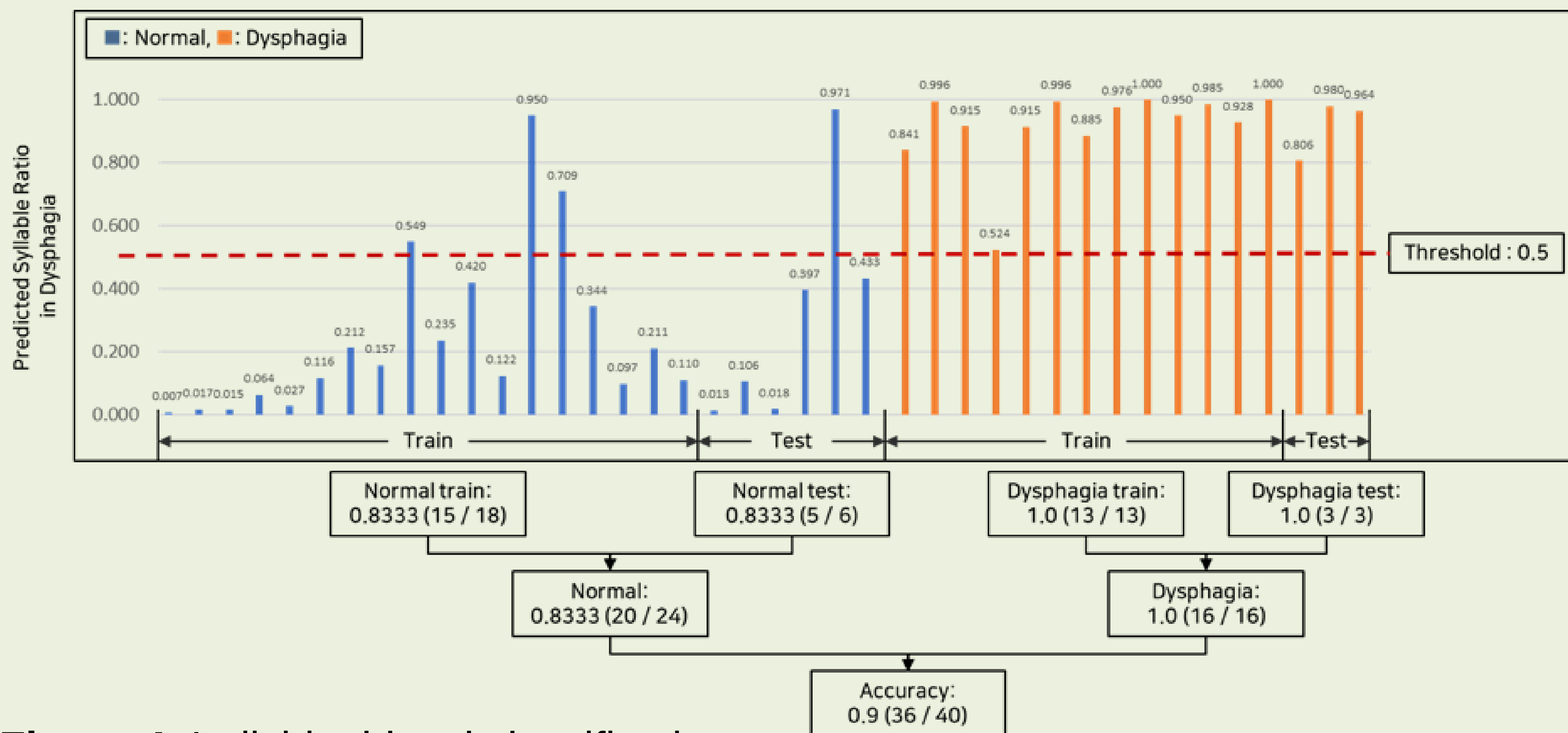


Figure 4. Individual-level classification

Conclusion

- These results suggest that the proposed deep learning model exhibits promising performance in dysphagia classification.
- This research presents the possibility of deep learning model of syllable-based speech analysis with daily conversations as a tool for early, non-invasive, and simple diagnosis for dysphagia in everyday environments.