Non-Invasive Voice Disorder Detection

Lim Kee Boon
Temasek Polytechnic
Singapore
1906371H@student.tp.edu.sg

Gina Tan Ci En
Temasek Polytechnic
Singapore
1906315E@student.tp.edu.sg

Muhammad Adam
Temasek Polytechnic
Singapore
muhdadam@tp.edu.sg

1. INTRODUCTION

Dysphonia is an alteration of voice production due to a morphological or functional alteration of the pneumo-articulatory apparatus. Moreover, dysphonia affects great number of people, with about one third of adults suffering from this disorder at least once in their lifetime [1].

The conventional method used to diagnose dysphonia is via the use of laryngoscopy. However, the laryngoscopy examination is an invasive procedure performed by an experienced laryngologist. Furthermore, the equipment required is costly and not commonly found in primary care units [2].

The aim of this study is to provide an alternative approach, which is efficient and non-invasive. The project proposed to develop a computer aided diagnosis (CAD) system to classify pathological and healthy voices from voice signal using Multi-layer Perceptron (MLP) and Convolutional neural network (CNN).

1.1 Literature Review

The studies on voice disorder detection using voices as listed in Table 1.

2. MATERIALS AND METHODS

The following are the materials used:

- VOICED (VOice ICar fEDerico II) database [1]
- 58 healthy and 150 pathological voices

The following are the methods used:

- Noise removal and downsampling
- Multi-Layer Perceptron (MLP)
- 1-Dimensional Convolutional Neural Network (1D CNN)

3. RESULTS

The healthy and pathological voice signals initially undergone noise removal and downsampled from 8000 Hz to 1000 Hz. This study experimented on both raw and preprocessed voice signals to separately train and test the baseline MLP and 1D CNN models of various layers. Further, the models employed model improvement methods, such as the callbacks, dropout and weight regularization, together with K-fold cross validation and confusion matrix. Overall, 1D CNN model of one layer with callbacks of patience=50 and weight regularization yielded the best validation accuracy of 82.76% and training accuracy of 68.97% at epoch 121 as shown in Figure 1.

Table 1. The studies on voice disorder detection using voices.

Author/Year	Method	Best Result
Ji-Yeoun Lee et al., 2021 [3]	CNN with linear prediction cepstrum coefficients (LPCCs)	82.69% accuracy
Wu et al., 2018 [4]	CNN	66.2% accuracy

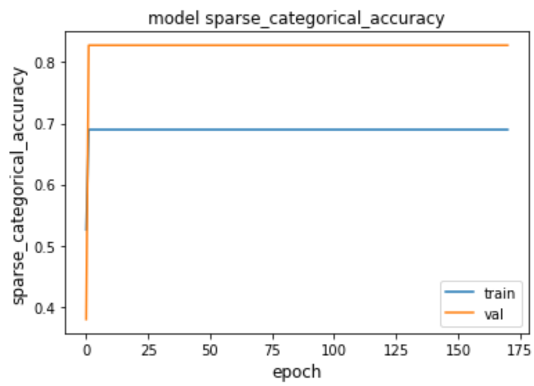

Figure 1. The classification performance of 1D CNN one-layer model with callbacks and weight regularization using denoised voices.

ACKNOWLEDGMENTS

We deeply appreciate for the sponsors of HPC Asia 2022.

REFERENCES

[1] Cesari, U., Pietro, G.D., Marciano, E., Niri, C., Sannino, G., & Verde, L. (2018). A new database of healthy and pathological voices. Comput. Electr. Eng., 68, 310-321.

[2] Fang, S., Tsao, Y., Hsiao, M., Chen, J., Lai, Y., Lin, F., & Wang, C. (2018). Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. Journal of voice: official journal of the Voice Foundation.

[3] Lee, Ji-Yeoun. "Experimental Evaluation of Deep Learning Methods for an Intelligent Pathological Voice Detection System Using the Saarbruecken Voice Database." Applied Sciences 11.15 (2021): 7149.

[4] Wu, H., Soraghan, J., Lowit, A., & Di Caterina, G. (2018, July). Convolutional neural networks for pathological voice detection. In 2018 40th annual international conference of the ieee engineering.