ejpmr, 2022,9(4), 111-116

SJIF Impact Factor 6.222

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Review Article

ISSN 2394-3211

EJPMR

BLOOD PRESSURE ESTIMATION FROM VOICE SPECTRUM WITH CONVOLUTIONAL NEURAL NETWORKS

***Motoki Sakai Ph.D.**

Associate Professor, College of Engineering, Nihon University, 1 Tokusada Aza-nakagawara, Tamuramachi, Koriyama-shi, Fukushima 963-8642 Japan.

***Corresponding Author: Motoki Sakai**
Associate Professor, College of Engineering, Nihon University, 1 Tokusada Aza-nakagawara, Tamuramachi, Koriyama-shi, Fukushima 963-8642 Japan.

Article Received on 11/02/2022 **Article Revised on 03/03/2022** **Article Accepted on 24/03/2022**

ABSTRACT

Blood pressure (BP) is an important vital index for predicting the risk of brain or cardiac infarction. This study proposes a method for estimating BP from voice signals without a BP gauge. The voice can be recorded using common tools, such as smartphones, which contributes to the ease of BP monitoring. In our experiment, systolic BP (SBP), diastolic BP (DBP), and voice data were obtained from 14 male and six female subjects. Five convolutional neural network (CNN) structures were used to estimate BPs. Consequently, no CNNs could estimate SBP and DBP in both men and women. However, there were effective CNN structures for male SBP, female SBP, and female DBP estimation, which could estimate BPs with a mean absolute error (MAE) of less than 10 mmHg.

KEYWORD: Blood pressure, Blood pressure estimation, Voice, Convolutional neural network, preventive healthcare.

1. INTRODUCTION

Blood pressure (BP) is a vital sign of respiration, heart rate, and body temperature. In particular, high BP, which is caused by sleep apnea syndrome, excessive salt intake, lack of exercise, etc., is a serious symptom because it leads to brain infarct or cardiac infarct.[1] Therefore, we must measure our own BP daily. Currently, there are many ambulatory BP cuffs and monitors. However, these are not necessarily appropriate for healthy people. I-can, a general device, is ideal for daily BP measurements.

Cuffless BP estimation methods have been presented in previous studies.[2-6] In many of them, BPs were estimated using both the ECG signal and pulse wave, based on the relationship between the heart and cardiovascular systems. However, ECG and pulse wave measurement devices, like BP gauges, are not generic for ordinary people. In papers.[3–6], Voice signal-based BP estimation methods have been proposed in previous studies.[3–6] Voice can be recorded by common devices, such as a smartphone microphone; it is easy to use compared to the BP cuff, ECG, and pulse wave.

There are two reasons for voice signal-based BP estimation: one is the influence of vagus nerve activity, and the other is a change in the hardness of blood vessels around a vocal fold. In general, it is known that the vagus nerve regulates the vocal fold and heart, and mental stress is often evaluated with speech or BP analysis.[7–9] as mental stress stimulates the vagus nerve, and the vocal fold and heart are affected by the activated vagus nerve. On the other hand, literature[10,11] reported that the vocal fold is affected by blood flow, and fluctuations synchronized with heart activity can be seen in a voice spectrogram. From this report, we can hypothesize that the hardness of blood vessels varies according to changes in the BP level, which influences fluctuations in the vocal fold, and then changes the features of the vocal spectrum. In fact, studies[3–6] have shown the possibility of BP estimation using vocal analysis.

In one study[3], patients with high BP could be distinguished from subjects without high BP by using vocal analysis to a certain degree, but quantitative BP values were not obtained. On the other hand, study[4] attempted to estimate quantitative BP values from vocal signals using a deep learning algorithm. Consequently, an accurate estimation of the BP values was obtained. However, this study constructed a plurality of networks for each specific patient group, and evaluations were performed for their groups. For example, neural networks have been generated for normal-high BP, normal-prehypertension-hypertension, and low BP-normal-prehypertension-hypertension groups. Such artificial grouping can lead to overfitting. Moreover, users must know in advance which group is most suitable to obtain accurate estimation results. It is not ideal for the datasets to be learned for practical use. In our previous studies[5,6], we proposed vocal analysis-based BP estimation methods using kernel ridge regression (KRR) and polynomial models, and accurate estimation results were obtained. However, the number of subjects was two

(one male and one female), and the parameters of the KRR and polynomial model were selected for each individual. Thus, this study proposes a common convolutional neural network (CNN)- based BP estimation method with more subjects.

2. Recording experiment for BP and voice signal

In this study, BP and voice signals were recorded from 20 participants (14 men and six women). The ages of male subjects ranged from 20 to 24 years, and those of female subjects ranged from 20 to 60 years.

We conducted these experiments after obtaining ethical approval from the ethical review committee of Tokyo Denki University. Before the experiments, informed consent was obtained from the subjects. First, the purpose of this study and the experimental procedure was explained to them. Next, the subjects signed a letter of consent if they understood the purpose of this research. One trial was conducted according to following steps.

I. Voice was recorded using an voice recorder: an ICD-TX50 produced by SONY. The voice recorder was kept 5 cm from the subject's mouth, and the subject was asked to sustain the sound [a] for 7 s (voice signals were sampled at 44.1 kHz, 16-bit resolution).
II. Immediately following voice recording, the subject's diastolic BP (DBP) and systolic BP (SBP) were measured using a BP gauge (HEM-1010, OMRON).

The trial was repeated several times for each participant. The number of trials differed for each subject, ranging from 10 to 60.

3. Vocal signal-based BP estimation methods

3.1 Generation of Input data

As described above, the voice data was recorded for 7 s. These raw data were divided into seven 1 s-segments, of which, only six segments were used as learning and test data, except for the first 1 s-segment. Each separated segment was transformed into a Mel-scale, and a 32×98 RGB image of the Mel-spectrogram was generated (each pixel was normalized by dividing by 256.). Finally, 2258 images for males and 1976 images for females were generated and used as input data for the CNNs described below.

From the data of 14 males, seven were selected as learning data, and the remaining seven were used as test data. For the female data, three data points were selected for learning, and the remaining three were used for testing. In sum, the number of learning Mel-spectrogram images was 1462, and that of test images was 796 for males, 992 for learning images, and 984 for females.

3.2 CNN structures

In this research, four existing and one proposed CNN structures were adopted (however, as described below, parts of the existing CNN structures were modified.). The information on the layers is presented in Table. 1. Four existing CNN architectures have been proposed in the literature.[12–15] The CNN architecture proposed in[12] (hereafter referred to as net1) was proposed to identify a certain speaker. Reference[13] provided the CNN structure (net2) to separate the singing voice. For the CNN presented in[14] (net3), the recognition of utterance content was attempted. Finally, the CNN presented in[15] (net4) was developed for voice-activity detection. In fact, there are more CNN structures for voice signals besides these four CNNs, but this paper presents abstracts for only these four architectures and their evaluation results because these CNNs yielded better evaluation results to a certain degree.

In this research, CNNs were constructed with the images described in Section 3.1, and transfer learning and fine-tuning were not performed. In addition, the output layers were changed to regression layers, unlike the CNNs in the literature[12] through[15], because these CNNs were not developed to estimate BP values.

To train the CNN models, the batch size was set to 128, 400 epochs were used to train the CNN models, and the Adam optimization algorithm was selected (learning rate = 1×10^{-3}). These parameters and optimizers were common in the five CNNs listed in Table. 1.

The proposed net was devised by referring to net1 (12), which was constructed using two convolutional/max pooling, one dropout, flattened, and dense layers. In the proposed CNN, the pooling layers were replaced with max pooling layers, and the number of feature maps of the convolutional layer and stride sides of the pooling layers were changed.

Table 1: Adopted CNN structures.

Model	Layer	Structure
net1	1	Convution_2D: number of feature map=32, filter size = (5,5), activation function='relu'
	2	Average Pooling_2D: pool size= (2, 2), strides=(1,1)
	3	Convution_2D: number of feature map=32, filter size = (5,5), activation function='relu'
	4	Average Pooling_2D: pool size= (2, 2), strides=(1,1)
	5	Dropout: dropout rate=0.2
	6	Flatten
	7	Dense: number of outputs=1
net2	1	Convution_2D: number of feature map=64,

		filter size = (3,3), activation function='relu'
	2	Max Pooling_2D: pool size= (2, 2), strides=(1,1)
	3	Convution_2D: number of feature map=32, filter size = (3,3), activation function='relu'
	4	Max Pooling_2D: pool size= (2, 2), strides=(1,1)
	5	Convution_2D: number of feature map=128, filter size = (3,3), activation function='relu'
	6	Convution_2D: number of feature map=64, filter size = (3,3), activation function='relu'
	7	Flatten
	8	Dense: number of output=1
net3	1	Convution_2D: number of feature map=64, filter size = (3,3), activation function='relu'
	2	Max Pooling_2D: pool size= (2, 2), strides=(2,2)
	3	Dropout: dropout rate=0.2
	4	Flatten
	5	Dense: number of output=1
net4	1	Convution_2D: number of feature map=40, filter size = (5,5), strides=(2,2), activation function='relu'
	2	Convution_2D: number of feature map=20, filter size = (5,5), strides=(2,2), activation function='relu'
	3	Convution_2D: number of feature map=10, filter size = (5,5), strides=(2,2), activation function='relu'
	4	Dropout: dropout rate=0.2
	5	Flatten
	6	Dense: number of output=1, kernel_initializer='normal'
proposed net	1	Convution_2D: number of feature map=60, filter size = (5,5), strides=(1,1), activation function='relu'
	2	Max Pooling_2D: pool size= (2, 2), strides=(2,2)
	3	Convution_2D: number of feature map=16, filter size = (5,5), strides=(2,2), activation function='relu'
	4	Max Pooling_2D: pool size= (2, 2), strides=(2,2)
	5	Dropout: dropout rate=0.2
	6	Flatten
	7	Dense: number of output=1, kernel_initializer='normal'

4. Evaluations and results

In this study, the mean absolute error (MAE) was adopted as the evaluation index. The definition of MAE is shown in Eq. (1).

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n}|x_{true} - x_{est}|, \quad (1)$$

where x_{true} and x_{est} are the measured and estimated one, respectively. n means the number of test datasets.

Considering the difference between male and female voice pitches, CNN models were constructed based on gender. Additionally, CNN models constructed using both male and female datasets were constructed to indicate the effectiveness of the CNN models constructed by gender. Tables 2–4 show the MAE between the measured and estimated BPs for all subjects, female, and male.

Table. 2: MAE errors between true and estimated BP values for all subjects.

	net1	net2	net3	net4	proposed net
SBP	11.8	12.2	16.7	12.8	12.6
DBP	13.2	18.1	14.6	13.7	13.8

Table. 3: MAE errors between true and estimated BP values for female subjects.

	net1	net2	net3	net4	proposed net
SBP	11.9	12.1	8.5	12.7	10.7
DBP	10.4	10.5	10.31	8.6	11.0

Table. 4: MAE errors between true and estimated BP values for male subjects.

	net1	net2	net3	net4	proposed net
SBP	8.8	6.4	13.7	8.3	6.8
DBP	16.6	16.1	17.3	18.1	18.1

Figures 1 and 2 show examples of mel-spectrogram plots for two male and two female subjects. In Figures 1 and 2, the top shows high SBPs and the bottom shows low SBPs.

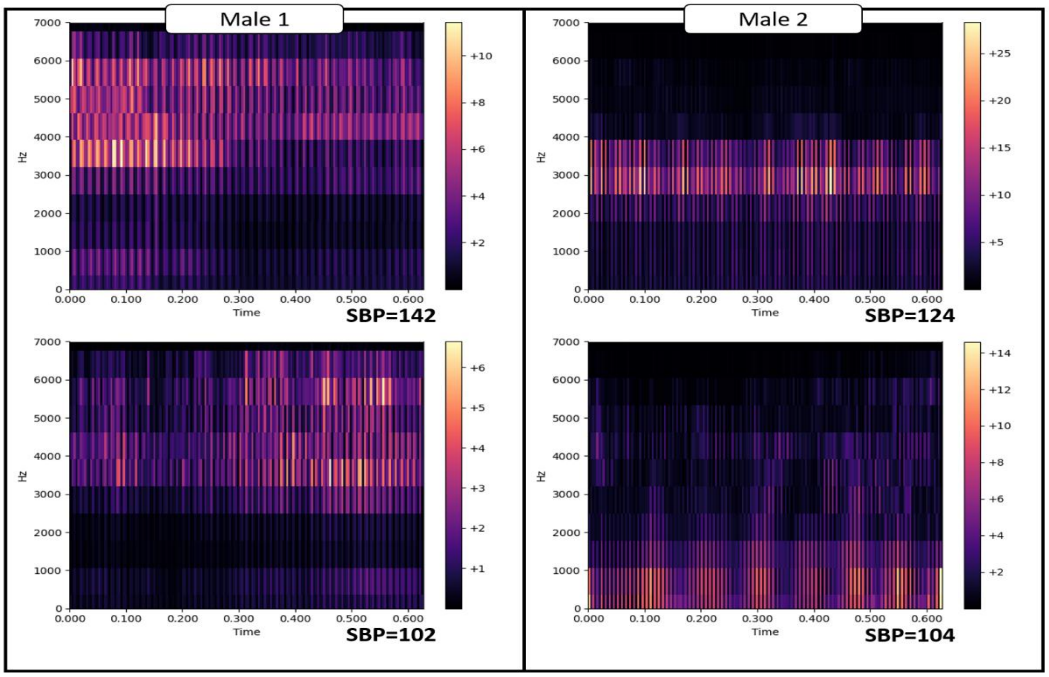

Figure. 1: Examples of Mel-Spectrogram plot for male subjects.

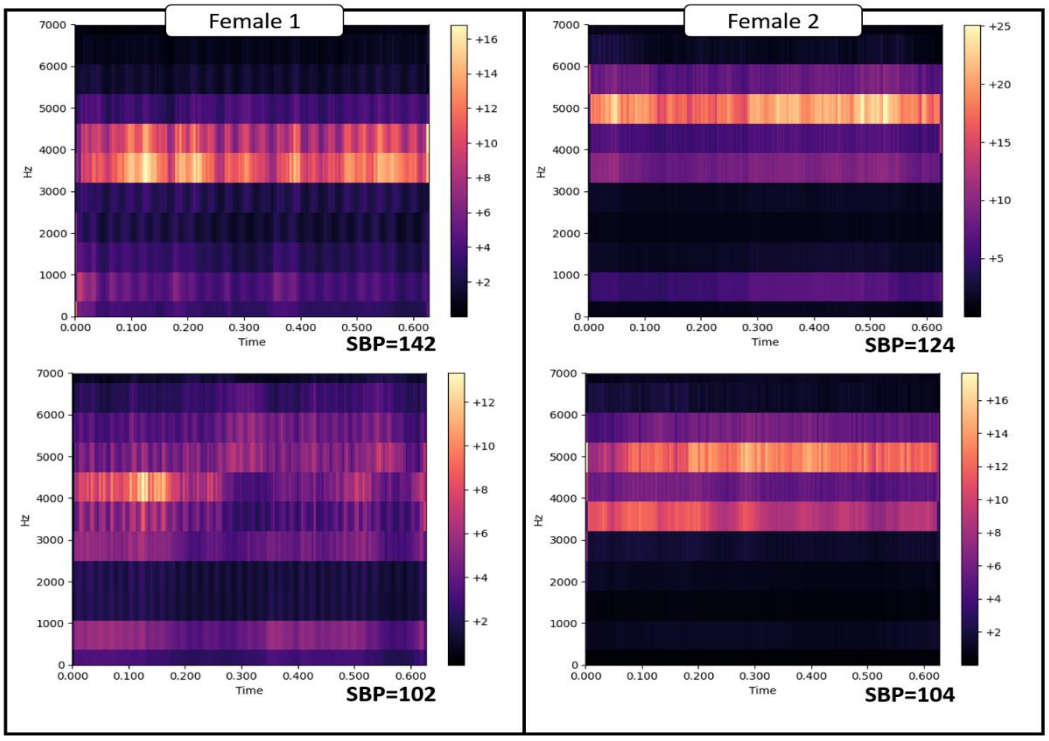

Figure. 2: Examples of Mel-Spectrogram plots for female subjects.

5. DISCUSSION

From Figures 1 and 2, we cannot find clear and common features to distinguish high SBP from low SBP. However, overall, figures 1 and 2 show that Mel-spectrograms of low SBP do not have frequency bands that have higher energy compared to those of high SBP.

Tables 2 through 3 show that there are no CNN models that effectively estimate male SBP, male DBP, female SBP, and female DBP. In SBP estimations, the net3 yielded lowest MAE (8.5 mmHg) for females, and the proposed net yielded the lowest MAE (6.8 Â mmHg) for males. In DBP estimations, the net4 yielded lowest MAE (8.6 mmHg) for females, and the net2 yielded lowest MAE (16.1 mmHg) for males. On the other hand, Table 2 shows that there were no CNN models that could estimate an MAE of less than 10.0 mmHg, which indicates that the CNN model should be constructed not for both male and female data but separately for male and female data.

In this study, male DBPs could not be properly estimated. The reason for this might be that the CNN structures and hyperparameters were selected to increase the estimation accuracy of male SBP. In short, the CNN structures and hyperparameters were not optimized for male DBP, female SBP, and DBP. Therefore, in future work, we plan to determine the CNN architectures separately for male SBP, male DBP, female SBP, and female DBP.

6. CONCLUSION

This study aimed to estimate SBP and DBP from vocal signals using CNN models. Vocal and BP data were recorded from 20 subjects and four CNN architectures, whose output layers were modified from the existing ones, and one proposed architecture was used to estimate BPs. As a result, there were no CNN models that could effectively estimate both SBP and DBP for men and women, but this study showed that there were appropriate CNN structures for male SBP, male DBP, female SBP, and female DBP estimation.

ACKNOWLEDGEMENTS

This work was supported by a Science Research Grant, Grant-in-Aid for Scientific Research (B) (General). We would like to thank Editage (www.editage.jp) for their English language editing services.

REFERENCES

1. Myung-Sun Song, Yeon Joo Choi, Hyunjin Kim, Myung Ji Nam, Chung-Woo Lee, Kyungdo Han, Jin-Hyung Jung, Yong-Gyu Park, Do-Hoon Kim and Joo-Hyun Park, "Relationship Between Blood Pressure Levels and Ischemic Stroke, Myocardial Infarction, and Mortality in Very Elderly Patients Taking Antihypertensives: A Nationwide Population-Based Cohort Study," BMC Geriatrics, 2021; 21(1): 620.
2. Manuja Sharma, Karinne Barbosa, Victor Ho, Devon Griggs, Tadesse Ghirmai, Sandeep K. Krishnan, Tzung K. Hsiai, Jung-Chih Chiao and Hung Cao, "Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review," Technologies, 2017; 5(2).
3. R. K. Saloni, Sharma and Anil K. Gupta Classification of high blood pressure persons vs normal blood pressure persons using voice analysis I.*J. Image, Graphics and Signal Processing*, 2014; 47–52.
4. H. Ankışhan, Blood Pressure Prediction from Speech Recordings, Biomedical Signal Processing and Control, 2020; 58: 101842.
5. M. Sakai, "Feasibility Study on Blood Pressure Estimations from Voice Spectrum Analysis," International Journal of Computer Applications, January 2015; 109(7): 39–43.
6. M. Sakai, "A Case Study on Analysis of Vocal Frequency to Estimate Blood Pressure" Proceedings of the 2015 IEEE Congress on Evolutionary Computation, May 2015; 2336–2340.
7. K. Chang, D. Fisher and J. Canny, "Ammon: A Speech Analysis Library for Analyzing Affect, Stress, and Mental Health on Mobile Phones" Proceedings of the 2nd International Workshop on Sensing Applications on Mobile Phones, 2011.
8. Amie M. Gordon and Wendy Berry Mendes, "A Large-Scale Study of Stress, Emotions, and Blood Pressure in Daily Life Using a Digital Platform," Proceedings of the National Academy of Sciences of the United States of America, 2021; 118: 31.
9. Daniela Gasperin, Gopalakrishnan Netuveli, Juvenal Soares Dias-da-Costa and Marcos Pascoal Pattussi, "Effect of Psychological Stress on Blood Pressure Increase: A Meta-Analysis of Cohort Studies," Cadernos de Saúde Pública, 2009; 25(4): 715–726.
10. D. Skopin and S. Baglikov, "Heartbeat Feature Extraction from Vowel Speech Signal Using 2D Spectrum Representation" Proceedings of the 4th International Conference Information Technology, June 2009.
11. A. Mesleh, D. Skopin, S. Baglikov and Anas Quteishat, "Heart Rate Extraction from Vowel Speech Signals," Journal of Computer Science and Technology, 2012; 27(6): 1243–1251.
12. Parashar Dhakal, Praveen Damacharla, Ahmad Y. Javaid and Vijay Devabhaktuni 'A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface,' machine learning & knowledge extraction, 2019; 1: 504–520.
13. Rupak Vignesh Swaminathan and Alexander Lerch, "Improving Singing Voice Separation Using Attribute-Aware Deep Network" Proceedings of the 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), 2019.
14. Jui-Ting Huang, Jinyu Li and Yifan Gong, "An Analysis of Convolutional Neural Networks for Speech Recognition" *Proceedings of the 2015 IEEE International Conference on Acoustics*, Speech, *and* Signal Processing (*ICASSP*), 2015.

15. Abhishek Sehgal and Nasser Kehtarnavaz, "A Convolutional Neural Network Smartphone App for Real-Time Voice Activity Detection," IEEE Access, 2018; 6: 9017–9026.