ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

Classification of patients using heart sound and MFCC

Seok-Woo Jang[1]and Sang-Hong Lee[2,*]
[1]Dept. of Software, Anyang University, Republic of Korea
[2]Dept. of Computer Science & Engineering, Anyang University, Republic of Korea
*Correspondence

Abstract: Sound processing is emerging as one of the research areas of sound processing. Various research fields of sound processing include speech recognition, speaker recognition, speech synthesis, and speech encoding. In this paper, healthy people and patients were classified using deep learning and Mel Frequency Cepstral Coefficient (MFCC) in the field of sound processing . The study was conducted by collecting heart sound data from healthy people and patients. Heart sound was imaged by MFCC. In traditional MFCC feature calculations, the 0th coefficient was excluded because it was considered somewhat unreliable. The accuracy of 89% was calculated by using the image extracted by MFCC as an input of deep learning.

Keywords: EEG, Wavelet transform, Hilbert transform, Peak extraction, Feature Selection, NEWFM.

1. Introduction

In recent years, interest in deep learning has been increasing. Convolutional Neural Network (CNN) is one of the various deep learning algorithms, a type of artificial neural network known as ImageNet Large Scale Visual Acceptance Composition (ILS) [1]. As CNN has achieved results in various fields, many studies are being conducted in the medical field. Studies on the Deep Neural Network-based EEG detection system for epileptic patient classification and the EEG classification of dementia patients using one-dimensional convolutional neural networks show that deep learning is involved in the development of the medical field [2][3]. In this study, the medical field also conducts a study to classify patients by heart sound. Heart sound is one of the most important pathological and physiological signals in the human body and contains the information of the human body about the heart. They directly reflect the mechanical state of motion of the large blood vessels and heart systems. Detection and analysis of heart sound signals have positive implications for early diagnosis of cardiovascular disease. However, since heart sound signals are a type of weak signal in the human body, they are vulnerable to external noise interference during signal acquisition and processing, which greatly affects the diagnosis of heart disease. So we have a problem diagnosing only with the collected heart sound. Some filtering techniques are neededto make the collected signals cleaner [4]. In speech recognition systems, it is generally recognized as Mel Frequency Cephal Coefficient (MFCC), an algorithm that characterizes acoustic/voice data input [5]. Patients are classified by inputting an image that is characterized by MFCC.

This paper was composed as follows. Chapter 2 describes experimental data, MFCC feature vectorization image extraction, and CNN. Chapter 3 explains the performance of patient classification as a result of the experiment. Finally, we conclude in Chapter 4.

2. Description of dataset and preprocessing model

As shown in the patient classification model in Fig. 1, the characteristics of heart sound were extracted through MFCC. These extracted features were put into CNN as input to classify healthy people and patients.

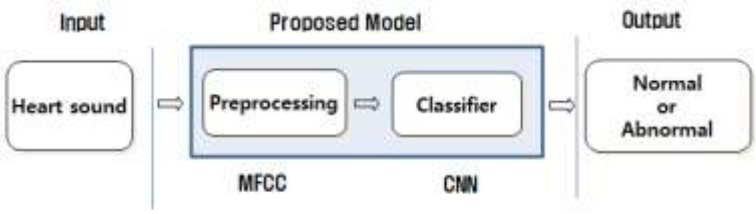


Fig. 1 Proposed model of patient classification.

2.1. Experimentdata

In this paper, normal heartbeat and abnormal heartbeat were classified using heart sound dataset from the 2016 PhysioNet/CinC Challenge [6]. The experimental data consists of two experimental groups. One experimental group is a normal heartbeat obtained through a healthy people and the other is an abnormal heartbeat obtained through a subject with heart disease. The experimental data used in this paper, such as normal heart rate and abnormal heart ratein Fig. 2 and Fig. 3. Each experimental group was divided into 2048 consecutive points in each heart sound data and adjusted to the same length for each sound data. Table 1 describes the experimental data.

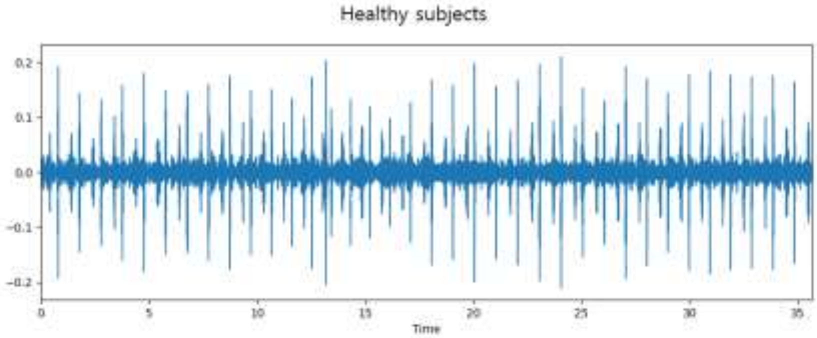


Fig. 2 Example of heart sound of healthy people.

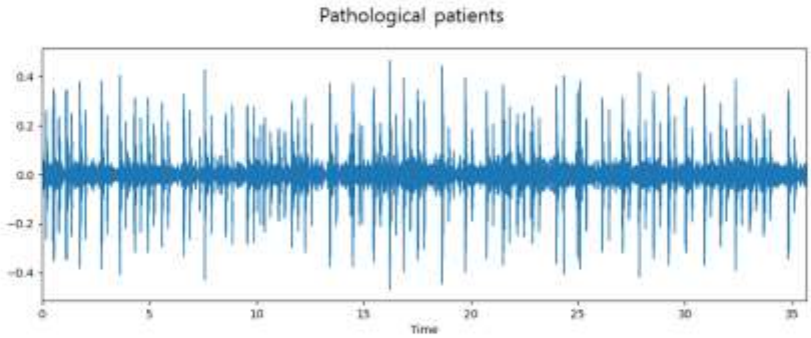


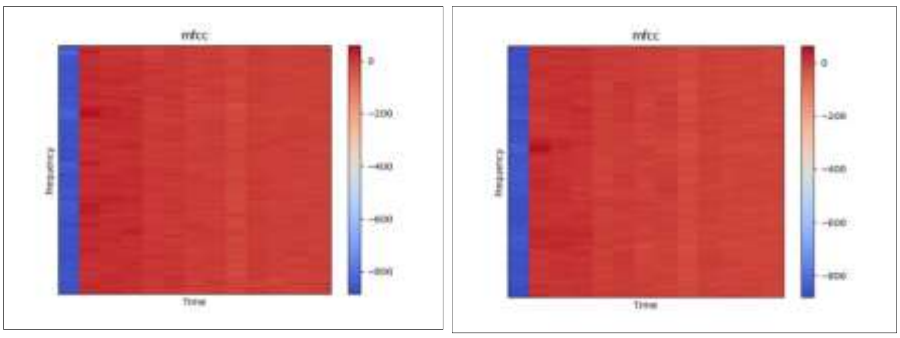
Fig. 3 Example of heart sound of patients.

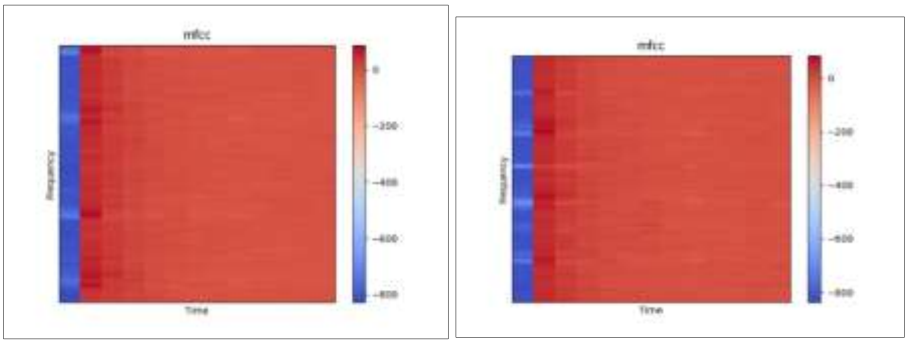
Table 1. Number of training and test datasets.

Class	Training dataset	Test dataset	Total dataset
Patients	1400	600	2000
Healthy people	1400	600	2000
Total	2800	1200	4000

2.2. Mel Frequency Cephal Coefficient (MFCC) for feature extraction

Mel Frequency Cephal Coefficient (MFCC) is one of the various methods for extracting speech signals [7]. Sound is expressed in two ways: linear cepstral and nonlinear cepstral. MFCC was derived from the nonlinear Cepstral representation of sound. In this process, discrete Fourier transform (DFT) transforms the original signal from the time domain to the frequency domain. A Hamming window is used to reduce frequency distortion by division before DFT. After this process, the filter bank wraps the frequencies from the Hertz to the Mel. In conclusion, DFT is used to extract shape vectors in the logarithm of mel-scale power spectra [8]. Each MFCC feature of the sound data is extracted. The extracted features are used and saved as a graph image. This image is needed for classification when evaluating performance. The graphed image using the MFCC features is shown in Fig. 4.





(a) Healthy people (b) patients

Fig. 4 Example of MFCC feature image.

3. Convolutional Neural Network (CNN)

This paper is constructed based on CNN and classifies into two classes using the feature-extracted image as an input node. For this classification, the convolutional neural network and convnet are configured as follows. The network of the convnet is configured as described below. Basically, the composition of the layer consists of Conv2D and MaxPooling2D layers. The convnet must be set to process input in size (image_height, image_width, image_channels). Here, image_height means the height of the image, image_width means the width of the image, and image_channels means the depth of the image. The shape of the input layer varies depending on the size of the feature image used as the input value.

The size of the image extracted using the MFCC feature is 640 X 480 pixels. The number of channels is adjusted by the first parameter transferred to the Conv2D layer. Accordingly, the input node is set to (640, 480, 3). The activation function uses the relu function. The relu function is a function that is 0 if the input value is less than 0, and the input value if the input value is greater than 0. When a network is configured with Conv2D and MaxPooling2D layers, a Dense layer is added. The output size of the last layer is set to 1 and the softmax activation function is used. Next [Figure 2] represents the structure of a convolutional neural network.

After completing the network structure, the neural network is compiled. In the compilation step, an 'rmsprop' optimizer is used. The network used binary cross entropy as a loss because the last layer contains a sigmoid activation function.

The JPEG image data extracted as an MFCC feature must be preprocessed with a floating point type tensor before being injected into the network. The preprocessing process is as described below. First, read the image. Second, the JPEG image is decoded into RGB pixel values. Third, this RGB pixel value is converted into a floating-point type tensor. The RGB pixel value contains between 0 and 255. It is too large to use as it is as an input to the neural network. Therefore, the pixel value is initially adjusted to a normalized value between 0 and 1. Keras can handle these steps automatically. The ImageData Generator class automatically replaces the image file with a preprocessed batch tensor.

A generator is set before training the data model. A generator is made by adjusting the size of each batch to 20 on the 640 X 480 RGB image. The generator makes this batch infinite.

After that, the task of learning is performed using the fit_generator function. After drawing as many batches of step_per_epoch from the generator, the training process moves on to the next epoch. The epoch is the number of times to learn, and the task is repeated by setting the amount of learning. In the following structure, 100 classes are grouped and taught 20 times.

4. Experiment results

In common, the experiment in this paper is conducted by cutting the amount of learning 20 times and 100 times. Accuracy was used as the performance evaluation of the algorithm to evaluate classification performance. The accuracy of learning with images made of MFCC characteristics was different depending on the amount of learning. In general, all of them showed high accuracy. The accuracy is shown in Table 2 and evaluation accuracy learned with the MFCC technique.

Table 2. Performance results.

	Loss	Accuracy
Train	0.1330	94.2%
Test	0.4506	81.6%

5. Concluding remarks

This paperconducts an experiment using the MFCC characteristics, which are used in the field of sound processing. Basically, MFCC characteristics were extracted using normal heart sound and abnormal heart sound as data. Accuracy was confirmed by putting the extracted feature data into an artificial neural network designed using it as input data. The results showed high performance accuracy, and the results of this experiment can implement a system that can classify normal and abnormal hearts by obtaining heart sounds in real time.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1F1A1055423).

References

[1] Marcus Stoffel, Franz Bamer, Bernd Markert, Deep convolutional neural networks in structural dynamics under consideration of viscoplastic material behaviour, Mechanics Research Communications 108, 103565, 2020.

[2] Ratnaprabha Ravindra Borhade, Manoj S. Nagmode, Modified Atom Search Optimization-based Deep Recurrent Neural Network for epileptic seizure prediction using electroencephalogram signals, Biocybernetics and Biomedical Engineering 40, 1638-1653, 2020.

[3] Manuel Lopez-Martin, Angel Nevado, Belen Carro, Detection of early stages of Alzheimer’s disease based on MEG activity with a randomized convolutional neural network, Artificial Intelligence in Medicine 107, 101924, 2020.

[4] Cheng Xiefeng, Zhang Zheng, Denoising method of heart sound signals based on self-construct heart sound wavelet, AIP Advances 4, 087108, 2014.

[5] Zhiguang Wang, Tim Oates, Imaging Time-Series to Improve Classification and Imputation, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 3939-3945, 2015.

[6] The 2016 PhysioNet/CinC Challenge, https://physionet.org/content/challenge-2016/1.0.0/.

[7] C. Sunitha, Evania Haris Chandra. Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm, International Journal of Engineering and Technology 7, 1685-1692, 2015.

[8] Sugandha Agarwal, O. P. Singh, Deepak Nagaria, Analysis and Comparison of Wavelet Transforms For Denoising MRI Image, Biomedical and Pharmacology Journal 10, 831-836, 2017.