Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Study on Performance of Patient Classification Using Heart Sound and Deep Learning

Seok-Woo Jang[1], Sang-Hong Lee[2*]
[1]*Dept. of Software, Anyang University, Republic of Korea, swjang7285@gmail.com*
[2]*Dept. of Computer Science & Engineering, Anyang University, Republic of Korea, shleedosa@gmail.com*
** Corresponding Author*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 30 Jan 2025

Accepted: 08 Feb 2025

ABSTRACT

Speech processing is emerging as an important application area of digital signal processing. In this paper, we present a performance comparison evaluation for patient classification based on Mel Frequency Cepstrum Coefficient (MFCC) using deep learning in the field of speech recognition. We conduct research by heart sound data of patients and healthy people. Each MFCC feature and heart sound feature are extracted by imaging them. We extract only MFCC features and compare the performance. In addition, we perform wavelet transformation to solve the noise problem of data and learn the extracted heart sound information using Gramian Angular Fields (GAF) and Phrase Space Reconstruction (PSR) techniques using deep learning and conduct a comparative evaluation. The accuracy results were 89.31%, 97.53%, and 100%, respectively, by learning with MFCC, GAF, and PSR techniques. We confirmed that good performance evaluations can be obtained by distinguishing patients using MFCC features.

Keywords: Wavelet transform, Phase space reconstruction, Deep learning, Speech processing, Signal processing.

INTRODUCTION

In recent years, deep learning has increased in various fields such as computer engineering and medical engineering [1][2]. Among various deep learning models, the most famous algorithm is the Convolutional Neural Networks (CNN), an artificial neural network as the ImageNet Large Scale Visual Acceptance Competition (ILS) [3]. CNNs are being widely studied in various fields such as computer engineering and medical engineering. In particular, in the medical field, it is shown that deep learning is widely used in research on a CNN-based epilepsy classification system for classifying epilepsy patients [4] and classification of dementia patients [5].

In this paper, we conduct research on classifying or detecting patients using heart sounds. Heart sounds are one of the important physiological signals generated in the human body. In addition, heart sounds contain a lot of pathological information about the human body, especially the heart. Heart sounds directly or indirectly reflect the repetitive activity of blood vessels or the heart. Analysis of heart sound signals has a positive meaning in the early diagnosis of patients with cardiovascular disease. However, heart sounds are weak signals among the various signals generated by the human body. Therefore, in the process of signal acquisition and signal processing for heart sounds, they are unintentionally vulnerable to external noise, which has a great impact on the diagnosis of cardiovascular diseases. Therefore, there is a big problem for researchers to diagnose cardiovascular diseases using only the collected heart sounds.

A filtering technique is needed to remove noise from the collected heart sounds [6]. There is a Mel-Frequency Cepstral Coefficient (MFCC) technique, which is an algorithm that extracts features from voice data collected in some systems that recognize human voice [7]. Wavelet transform is used in the filtering process. Wavelet transform refers to transforming data using wavelet basis functions. Here, the wavelet basis function refers to a function that becomes 0 when integrated and converges to 0 in amplitude when vibrating [8]. Wavelet transforms decrease the frequency resolution and increase the time resolution for signals with high-frequency components. On the other hand, it increases the frequency resolution and decreases the time resolution for signals with low-frequency components [9].

In this study, images that are preprocessed by wavelet transform are reconverted and used for research by using the Gramian Angular Field (GAF) technique and the Phrase space reconstruction (PSR) technique as features. Learning is performed with the images used in the research. The performance for patient classification is compared and evaluated based on the learning results.

RELATED RESEARCH

A. CNN (Convolutional Neural Networks)

CNN is an optimal method for extracting high-level abstract features from images or processing texture information. It is an algorithm that optimizes neural networks to learn images well because it can input two-dimensional structures instead of vector-type input data [10]. The following Figure 1 expresses the CNN algorithm. The final layer, which is the output layer that can be classified into four classes from the input layer, an image patch of 36 X 36 pixels, represents an artificial neural network. There are two hidden layers between the output layer and the input layer. Hidden layer 1 is the result of the convolution of the previous layer, and hidden layer 2 is a maximum step, which significantly reduces the number of units by maintaining only the maximum response of several units in the first step. After several hidden layers, the output layer is usually a completed layer [11].

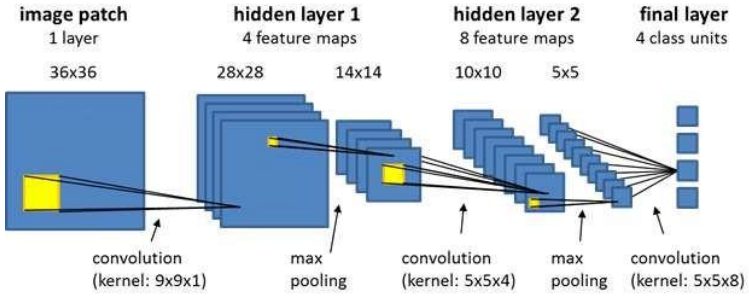

Figure 1. Structure of CNN

B. Study on heart sound noise removal

The heart sound noise removal algorithm adopts a multi-stage wavelet decomposition threshold to remove noise based on the fact that the wavelet coefficients of the signal and sound have different mechanisms at different scales. The heart sound noise was removed through wavelet transform, and in this paper, the heart sound data is also subjected to wavelet transform to extract features.

C. MFCC (Mel-Frequency Cepstral Coefficient)

Mel-Frequency Cepstral Coefficient (MFCC) is an algorithm for characterizing speech recognition data. Values may be obtained through cepstral analysis from the Mel Spectrum. The Mel Spectrum performs a function of extracting features by dividing sound source data into a predetermined period frame and analyzing a spectrum.

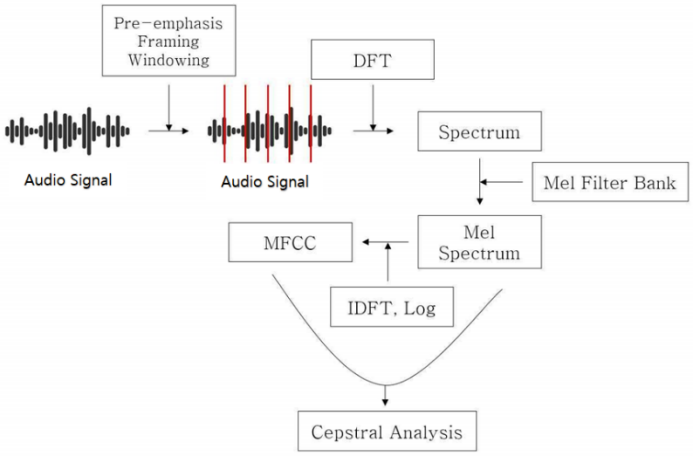

Figure 2. MFCC block diagram

First, the spectrum is extracted by applying Fast Fourier Transform (FFT) to a frame divided by 20 to 40 ms length of the sound source data. The frequency domain can be expressed by performing the role. Since the frequency contains unnecessary information for sound source recognition, the filter bank technique is applied to identify the low frequency range. The Mel Scale filter used informs the interval value to be divided and can determine the amount of energy generated in each section. The MFCC progress is proposed as shown in Figure 2 [12][13].

D. GAF (Gramian Angular Fields)

GAF (Gramian Angular Fields) converts sound signals into image data obtained from time series and provides temporal relationships between each point in time. Time series data uses statistical techniques based on similarity, probability, bound-aries, and clustering for classification prediction. It generally performs the role of extracting and selecting important features from data [14]. GAF is expressed as one of the deep learning utilization methods based on neural networks that can be used for prediction through feature learning.

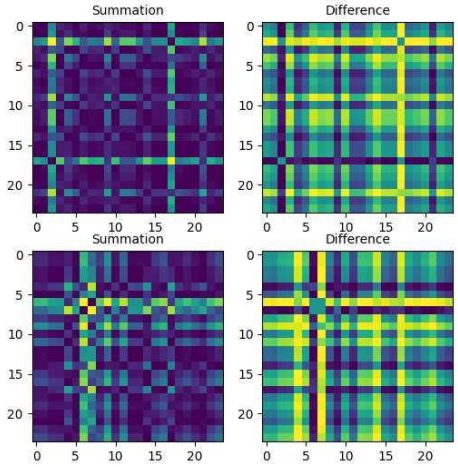

Figure 3. GAF image data

Figure 3 shows GAF image data. Summation, which is the value of adding data, and Difference, which is the value of subtracting data, are provided in the image. In Figure 3, GAF's D3 and D4 data are expressed as four images, and two images are generated for one data value. GAF makes the length of the input sound data constant and performs the feature extraction process using it as the input value.

E. PSR (Phase space reconstruction)

Phase space reconstruction (PSR) is a technique for analyzing dynamic waveforms based on phase space [15]. It plays a role in extracting important features from data, and in this paper, a one-dimensional signal of music data is made into a two-dimensional signal and imaged. To obtain a PSR image, preprocessing is first required. In order to obtain good accuracy during the deep learning process, preprocessing is performed to remove noise and have a high contrast. In this paper, wavelet transform is used as a preprocessing operation.

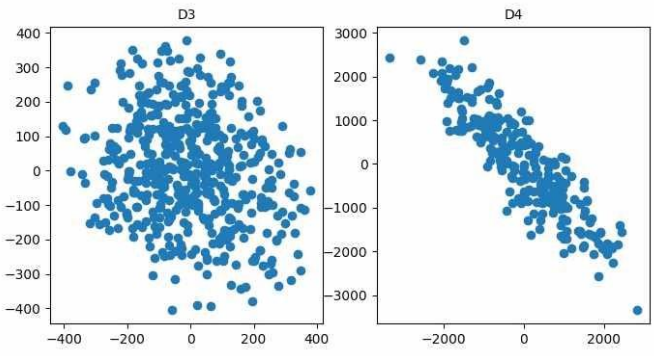

Figure 4. PSR image data

Wavelet transformation improves subjective image quality by decomposing and synthesizing it according to human visual characteristics. It can be composed of a set of signals and is used in various applications such as improvement through image matching and restoration and noise removal. Figure 4 represents PSR image data. PSR makes the length of the input one-dimensional signal data constant and converts it into a two-dimensional signal by performing a feature extraction process as an input value. In this paper, the signal data is divided into 4000 pieces each and the above process is performed to represent D3 and D4 resultant images for each music genre.

ARTIFICIAL NEURAL NETWORK DESIGN AND EXPERIMENTAL RESULT

In this paper, we evaluate the performance of patient classification by extracting features through four processing steps using heart sounds as input and generating images. The performance evaluation of feature extraction based on heart sound recognition is proposed as shown in Figure 5. For patient classification, the heart sound set is divided into 'healthy' and 'unhealthy'. The collected heart sounds all have different lengths. For good accuracy, all sounds should be made of the same length. The sounds are made of signal data. The signal data is divided into 4000 pieces each to have the same length. Features are extracted using sounds of the same length as input. In this paper, we compare the best performance of three methods for extracting features.

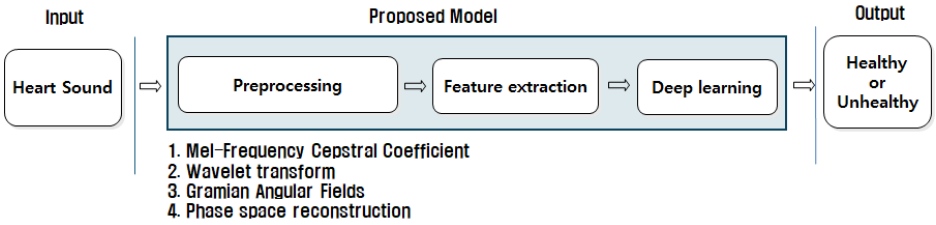

Figure 5. Feature extraction process

A. Design of artificial neural network for MFCC

The image by the MFCC technique is 640 X 480 pixels. In the overall design of the artificial neural network, the network, compilation, preprocessing, generator, and learning tasks are the same as in Figure 6, but only the image size is changed to 640 X 480 pixels. Table 1 shows the accuracy of MFCC.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 638, 478, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 319, 239, 32)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 317, 237, 64)      18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 158, 118, 64)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 156, 116, 64)      36928
_________________________________________________________________
flatten (Flatten)            (None, 1158144)           0
_________________________________________________________________
dense (Dense)                (None, 64)                74121280
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
Total params: 74,177,665
Trainable params: 74,177,665
Non-trainable params: 0
```

Figure 6. The structure of CNN

B. Design of artificial neural network for image data obtained by GAF

The image by the GAF technique is 600 X 600 pixels. In the overall design of the artificial neural network, the network, compilation, preprocessing, generator, and learning tasks are the same as in Figure 6, but only the image size is changed to 600 X 600 pixels. In other words, this means that the target_size is changed when configuring the input node and generator of the input layer. Table 1 shows the accuracy of GAF.

C. Design of artificial neural network for image data obtained by PSR

The image by the PSR technique is 800 X 400 pixels. In the overall design of the artificial neural network, the network, compilation, preprocessing, generator, and learning tasks are the same as in Figure 6, but only the image size is changed to 800 X 400 pixels. Table 1 shows the accuracy of PSR.

Table 1. Comparisons of performance results

	MFCC	GAF	PSR
Accuracy	89.31%	97.53%	100%

CONCLUDING REMARKS

In this paper, we conducted an experiment by MFCC features, that are widely used in the wide field of speech recognition. Afterwards, we conducted additional experiments using GAF and PSR techniques to confirm higher performance accuracy. Basically, we used normal heart sounds and abnormal heart sounds as data, utilized various features of each, and checked the accuracy by putting them into the designed artificial neural network. As a result, MFCC features, GAF, and PSR techniques showed high performance accuracy, and PSR, GAF, and MFCC features showed high accuracy in that order. Based on the results of this experiment, we believe that if we supplement it further, we will be able to identify people's diseases through heart sounds. Sounds can be recorded with a lot of noise. Heart sounds are ultimately recorded data. In the field of deep learning, a preprocessing process for noise removal is performed before learning to extract features. In this paper, we performed wavelet transform to remove noise. Although we confirmed good performance accuracy through this work, we believe that it would be good to find a way to remove noise and supplement it for better performance. In addition, we believe that the data provided is insufficient. Further research is needed to determine whether increasing the amount of heart sound data will lead to better performance accuracy.

REFRENCES

[1] Park, K. B., & Lee, J. Y., “SwinE-Net: hybrid deep learning approach to novel polyp segmentation using convolutional neural network and Swin Transformer,” Journal of Computational Design and Engineering, 9(2), 616-632, 2022.

[2] Qayyum, A. et al. “Medical image retrieval using deep convolutional neural network,” Neurocomputing, Vol. 266, pp. 8-20, 2017.

[3] Russakovsky O, Deng J, Su H et al, “ImageNet Large Scale Visual Recognition Challenge,” Int J Comput Vis 115:211–252, 2015.

[4] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,” Computers in biology and medicine, 2017.

[5] M. Golmohammadi, S. Ziyabari, V. Shah, E. Von Weltin, C. Campbell, I. Obeid, and J, Picone, “Gated recurrent networks for seizure detection,” In Signal Processing in Medicine and Biology Symposium (SPMB), IEEE, pp.1-5, 2017.

[6] 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.

[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] Shyu LY, Wu YH, Hu W, “Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG,” IEEE Trans Biomed Eng 51:1269–1273, 2004.

[9] Minami K, Nakajima H, Toyoshima T, Real-time discrimination of ventricular tachy- arrhythmia with Fourier-transform neural network, IEEE Trans Biomed Eng 46:176– 185, 1999.

[10] Anthimopoulos, M. et al. “Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," IEEE Transactionson MedicalImaging, Vol. 35, No. 5, pp. 1207-1216, 2016.

[11] S. A. Singh, S. Majumder, "A novel approach osa detection using single-lead ECG scalogram based on deep neural network." Journal of Mechanics in Medicine and Biology, Vol. 19, No. 04, June, 2019.

[12] A. Abbaskhah, H. Sedighi, and H. Marvi, "Infant cry classification by MFCC feature extraction with MLP and CNN structures", Biomedical Signal Processing and Control, Vol 86, No. Part B, pp. 105261, Sep. 2023.

[13] A. Pikrakis, T. Giannakopoulos and S. Theodoridis, "A Speech/Music Discriminator of Radio Recordings Based on Dynamic Programming and Bayesian Networks," IEEE Transactions on Multimedia, vol. 10, no. 5, pp. 846-857, Aug. 2008.

[14] R. N. Toma, F. Piltan, K. Im, D. Shon, T. H. Yoon, D.-S. Yoo and J.-M. Kim “A Bearing Fault Classification Framework Based on Image Encoding Techniques and a Convolutional Neural Network under Different Operating Conditions,” sensors, vol.22, no.13, pp.4881, 2022.

[15] Xu, P., "Differential Phase Space Reconstructed for Chaotic Time Series," Applied Mathematical Modelling, Vol. 33, No. 2, pp. 999-1013, 2019.