Regional Signal Recognition of Body Sounds Vücut Seslerinden Bölgesel Sinyal Teşhisi

Osman Balli[1], Yakup Kutlu[1]
[1]Department of Computer Engineering, Iskenderun Technical University, Turkey
ORCIDs: 0000-0001-6757-3717, 0000-0002-9853-2878
E-mails: osman.balli50@gmail.com, yakup.kutlu@iste.edu.tr
*Corresponding author.

Abstract—**One of the most important signals in the field of biomedicine is audio signals. Sound signals obtained from the body give us information about the general condition of the body. However, the detection of different sounds when recording audio signals belonging to the body or listening to them by doctors makes it difficult to diagnose the disease from these signals. In addition to isolating these sounds from the external environment, it is also necessary to separate their sounds from different parts of the body during the analysis. Separation of heart, lung and abdominal sounds will facilitate digital analysis, in particular. In this study, a dataset was created from the lungs, heart and abdominal sounds. MFCC (Mel Frekans Cepstrum Coefficient) coefficient data were obtained. The obtained coefficients were trained in the CNN (Convolution Neural Network) model. The purpose of this study is to classify audio signals. With this classification, a control system can be created. In this way, erroneous recordings that may occur when recording physicians' body voices will be prevented. When looking at the results, the educational success is about 98% and the test success is about 85%.**

Keywords—body sounds; mel frequency cepstrum coefficients; deep learning

Özetçe—**Biyomedikal alanındaki en önemli sinyallerden birisi ses sinyalleridir. Vücuttan elde edilen ses sinyalleri bize vücudun genel durumu hakkında bilgi verir. Ancak, vücuda ait ses sinyallerini kaydederken veya doktorlar tarafından dinlenirken farklı seslerin algılanması, hastalığın bu sinyallerden teşhis edilmesini zorlaştırır. Bu sesleri dış ortamdan izole etmenin yanı sıra, analiz sırasında vücudun farklı bölgelerinden gelen seslerini de ayırmak gerekir. Kalp, akciğer ve karın seslerinin ayrılması, özellikle dijital analizi kolaylaştıracaktır. Bu çalışmada akciğerler, kalp ve karın seslerinden bir veriseti oluşturulmuştur. MFCC katsayı verileri alınmıştır. Elde edilen katsayılar CNN modelinde eğitilmiştir. Bu çalışmanın amacı ses sinyallerini sınıflandırmaktır. Bu sınıflandırma ile bir kontrol sistemi oluşturulabilecektir. Bu sayede hekimlerin vücut seslerini kaydederken oluşabilecek hatalı kayıtların önüne geçilecektir. Sonuçlara bakıldığında eğitim başarısı %98, test başarısı ise %85 civarındadır.**

Anahtar Kelimeler—vücut sesleri; mel frekansı spektrum katsayıları; derin öğrenme

I. Introduction

The processing and analysis of biomedical signals is of great importance for technologies in the field of health. Sound signals, on the other hand, are an important signal that allows us to have general information about the body [1], [2]. Sound signals in the body are provided by auscultation [3]. Auscultation is the process of listening to intra-body sounds with the help of a stethoscope for the diagnosis of a disease, for checking the functioning of organs [4]. Auscultation is a procedure that provides basic information about the condition of the body due to an ailment or during a routine examination. The sounds of the heart, lungs, abdomen and other organs are tested in a kind of resting way. As a result of listening, a general knowledge of the person's condition is obtained.

We can listen to many sounds from the body. But although we know a little about which region they belong to, we may not be able to understand how much other sounds in the body interfere with the sounds we listen to. On the other hand, it is more convenient and advantageous to understand and distinguish body sounds in a digital environment than to diagnose with a stethoscope.

There are many studies on body sounds in the literature. In the study conducted on the diagnosis of Chronic Obstructive Pulmonary Disease (COPD) using lung sounds, the Second Degree Difference Plot (SODP) analysis method was used. Deep Belief Networks (DBNs) are combined with this method. The binding of 3D-SODP quantification properties together with DBN separated lung sounds from different COPD levels with high classification performance rates of 95.84%, 95.84%, 93.34% and 93.65% for accuracy, sensitivity and specificity, respectively [5]. Another study focused on the most common lung sounds, wheeze and crackle. The data was collected with a custom mobile phone application and an electronic stethoscope. 284 Data received from patients with was detected ROC (Receiver Operating Characteristic) curves with AUCs (Area Under the Curve) of 0.86 for wheeze and 0.74 for crackle. according to the data taken from 284 patients, 0.86 of the results were obtained. It has been shown that semi-supervised deep learning can be successful in large datasets [6]. Aykanat et al., on the other hand, have shown how different classifications in lung sounds will affect the results. there are 2 types of methods used; Mel Frequency Spectral Coefficient (MFCC) features in a Support Vector Machine (SVM) and spectrogram images in the Convolutional Neural Network (CNN). He achieved the best result in the training of healthy and pathologically classified voices. He achieved 86% success in SVM and 86% success in CNN [7]. Altan et al. analyzed the severity of COPD from 12 channel lung sounds. The Deep ELM (Extreme Learning Machine) model was used and has separated five COPD severities with classification performance rates of 94.31%, 94.28%, 98.76%, and 0.9659 for overall accuracy, weighted-sensitivity, weighted-specificity, and area under the curve (AUC) value, respectively [8]. In another study, 100 diseased and 100 normal heart sound data were used. The Time-Frequency Distribution (TFD) analysis of these data and the heart sound analysis based on Mel Frequency Spectrum Coefficient (MFCC) were performed. The coefficients obtained from these methods are trained in artificial neural networks. The system is able to produce the accuracy up to 90% using the TFD and 80% using the MFCC [9]. In another study on heart sounds, it was found that MFCC was applied to 1381 data sets of real and simulated, normal and abnormal domains. The classification rates for normal and abnormal heart sounds were found to be 95.7% for continuous murmurs, 96.25% for systolic murmurs, and 90% for diastolic murmurs using a probabilistic comparison approach [10]. In a study conducted on respiratory sounds, a dataset consisting of normal and abnormal sounds was used. KNN was applied to the attributes obtained as a result of MFCC application and 93.21% success was achieved [11]. In a study conducted on COPD, normal and asthma sounds, the feature extraction method was applied. The obtained data were trained in SVM, KNN, DT structures and compared with the CNN model results. CNN and SVM also achieved 96% success [12]. There are also many studies on the analysis of bowel sounds [13]. MFCC was applied to the framed bowel sounds and LSTM-based deep learning was applied to the resulting attributes. The data were recorded from 45 different people and the patients' age varies from 42 to 77, and the patients' gastrointestinal conditions contain astriction, dyspepsia or normal [14].

In this study, a dataset consisting of sounds taken from the heart, lungs and abdomen was prepared. To be able to classify audio signals received from different regions of the body. In this way, the accuracy of data sets built with body audio signals can be tested. In addition, with a control system to be prepared, the signals that doctors will record will be analyzed and any errors that may occur in creating labels will be prevented. MFCC was first applied to audio data. The coefficients obtained as a result of Mfcc were trained in the CNN model.

II. Materials & Methods

A. Database

In this study, audio recordings were taken from 12 people. A total of 12 records were obtained from 1 person, including four from different regions of the lungs, heart and abdomen. The recordings are 20s and have a sampling frequency of 4000. The sounds were divided into 2s 10 audio data in total were obtained from 1 audio recording.

B. Mel Frequency Spectral Coefficient (MFCC)

MFCC serves to create attributes using audio signals. There are many studies in which attribute extraction is performed using MFCC in audio signal [9], [15], [16]. It is very important to create meaningful attributes of body sounds for educational work to be done. In this study, Mel Frequency Cepstrum Coefficients were used. Fig. 1 visualizes the processes used in MFCC.

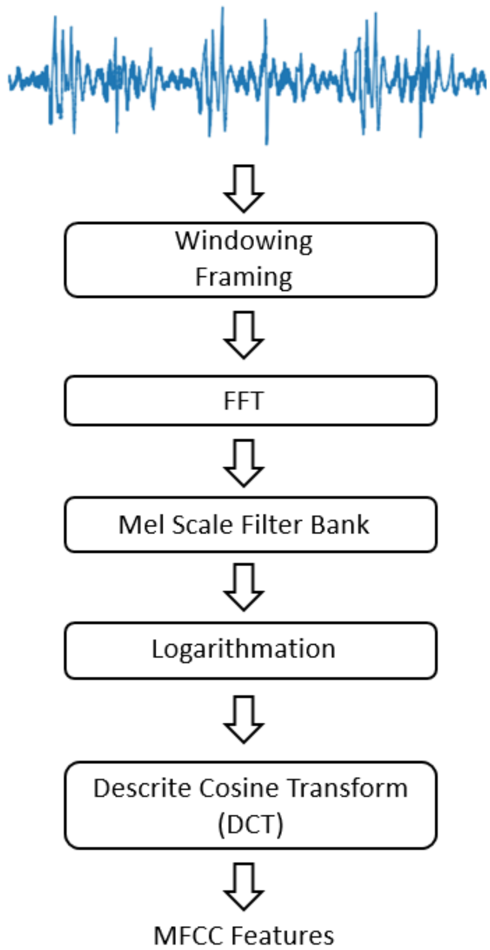

Figure 1: Block diagram of MFCC

MFCC applies the window size to the signal first. Each window implements a fft and allows it to pass to the frakans domain. Then convert it from the frequency scale to the Mel logarithmic scale with the Mel filter bank. Logarithmic applies DCT to the data obtained as a result of the conversion, and the results obtained are converted to a time scale. As a result of these operations, 1287 attributes were obtained from every 2-second signal. The obtained data were trained in the CNN model.

C. Convolution Neural Network (CNN)

Deep learning is a branch of artificial intelligence that has started to be used frequently with the spread of high-performance hardware. One of the most popular methods of deep learning is CNN. It is a method used in different fields such as signal processing, classification, recognition.

The data provided to CNN must be in matrix format. It consists of layers called convolution and pooling. The number of these layers is determined by the person whose method is to be applied. The Convolution layer is the layer where the number and size of filters to be applied to the data are determined. The results obtained from this layer are again given to the convolution layer or the pooling layer. Different methods such as max, min, average can be used in the pooling layer. After the pooling process is applied to the data coming to the pooling layer, the number of data decreases. According to the method we have determined, it moves meaningful data to the next layer. After a certain number of convolution and pooling operations are performed, the data is moved to the flatten layer. Here the results are transformed into a one-dimensional matrix. The matrix is transferred to artificial neural networks. Weights in training are constantly updated. The highest performance gives the optimal weight values.

In this study, the data were converted into a 35x35 square matrix. 64-, 128- and 256-layer Convolution layers were applied. In addition, 3x3 average pooling was performed 2 times [17], [18]. Fig. 2 shows the general scheme of CNN used in this study.

III. Results and Discussion

The creation of ready-made models is effective in faster operation of diagnostic systems. It is very important to achieve high performance values in such models. We focus on 2 achievements: training and test success. As a result of this study, these 2 parameters were examined and analyzed.

As a result of CNN training with attributes obtained by applying MFCC to audio data, train and test achievements were examined. A model was prepared that used 1440 voice data consisting of 2s and determined which of these sounds belong to the heart, lung and abdominal regions. A processing time of 130 epoch was determined in the training. It is seen that when the educational achievement reaches around 98%, the test success is around 85% (Fig. 3).

It is seen that regional classification has been successful with MFCC and CNN. But in order for the results to be even more reliable, the number of data must be increased. In addition, increasing the test performance will be effective in making the model work more accurately.

References

[1] Telceken M, Kutlu Y. Detecting abnormalities in heart sounds. Journal of Intelligent Systeams with Applications 2021; 4(2): 137-143.

[2] Cancioglu E, Sahin S, Isler Y. Heart sounds analysis and classification based on long-short term memory. Journal of Intelligent Systeams with Applications 2020; 3(1): 25-28.

[3] Altan G, Kutlu Y, Pekmezci AO, Nural S. The diagnosis of asthma using hilbert-huang transform and deep learning on lung sounds. Journal of Intelligent Systeams with Applications 2019; 2(2): 100-105.

[4] Selek MB, Duyar MC, Isler Y. Electronic stethoscope design. Journal of Intelligent Systeams with Applications 2020; 3(1): 10-16

[5] Altan G, Kutlu Y, Pekmezci AO, Nural S. Deep learning with 3D-second order difference plot on respiratory sounds. Biomedical Signal Processing and Control 2018; 45: 58–69.

[6] Chamberlain D, Kodgule R, Ganelin D, Miglani V, Fletcher RR. Application of semi-supervised deep learning to lung sound analysis. In 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 16-20, 2016, Orlando, FL, USA, pp. 804–807.

[7] Aykanat M, Kilic O, Kurt B, Saryal S. Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing 2017; 65(2017).

[8] Altan G, Kutlu Y, Gokcen A. Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds. Turkish Journal of Electrical Engineering and Computer Sciences 2020; 28(5): 2979–2996.

[9] Kamarulafizam I, Salleh SH, Najeb JM, Ariff AK, Chowdhury A. Heart sound analysis using MFCC and time frequency distribution. IFMBE Proceedings 2007; 14(1): 946–949.

[10] Chauhan S, Wang P, Sing Lim C, Anantharaman V. A computer-aided MFCC-based HMM system for automatic auscultation. Computers in Biology and Medicine 2008; 38(2): 221–233.

[11] Balli O, Kutlu Y. Effect of window size for detection of abnormalities in respiratory sounds. Natural and Engineering Sciences 2019; 4(3): 124-129.

[12] Balli O, Kutlu Y. Effect of deep learning feature inference techniques on respiratory sounds. Journal of Intelligent Systems with Applications 2020; 3(2): 134-137.

[13] Nowak JK, Nowak R, Radzikowski K, Grulkowski I, Walkowiak J. Automated bowel sound analysis: An overview. Sensors 2021; 21(16): 1–16.

[14] Liu J, Yin Y, Jiang H, Kan H, Zhang Z, Chen P, Zhu B, Wang Z. Bowel sound detection based on MFCC feature and LSTM neural network. In 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), October 17-19, 2018, Cleveland, OH, USA, pp. 1-4.

[15] Palaniappan R, Sundaraj K, Sundaraj S. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. BMC Bioinformatics 2014; 15(1): 1–8.

[16] Nogueira DM, Ferreira CA, Gomes EF, Jorge AM. Classifying heart sounds using images of motifs, MFCC and temporal features. Journal of Medical Systems 2019; 43(6): 186–203.

[17] Kutlu Y, Camgozlu Y. Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. Natural and Engineering Sciences 2021; 6(1): 60–74.

[18] Narin A, Isler Y. Detection of new coronavirus disease from chest x-ray images using pre-trained convolutional neural networks, Journal of the Faculty of Engineering and Architecture of Gazi University 2021; 36(4): 2095-2107.

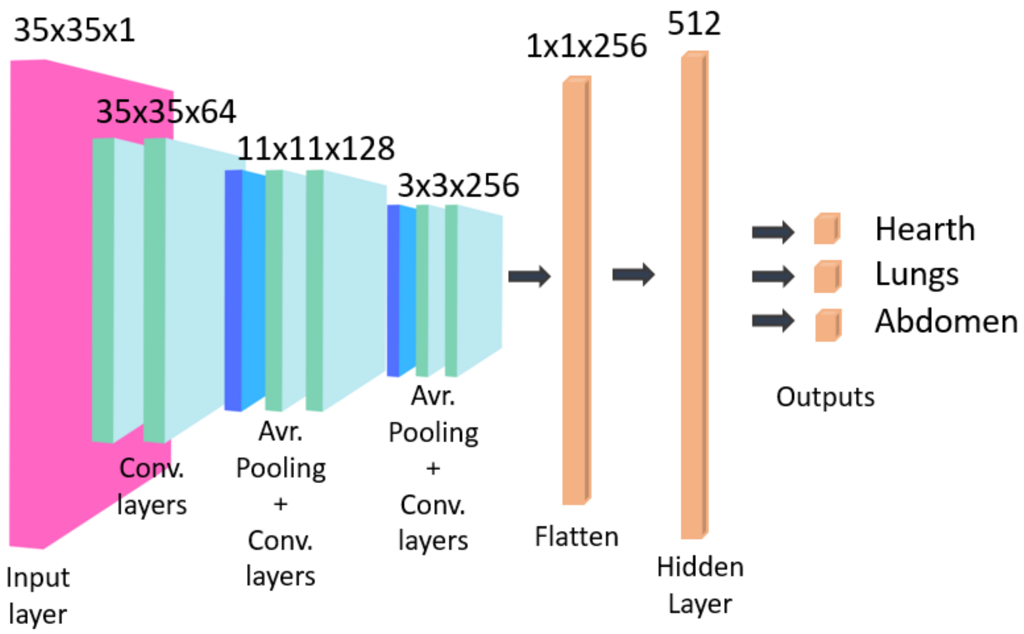

Figure 2: Shema of convolution neural network

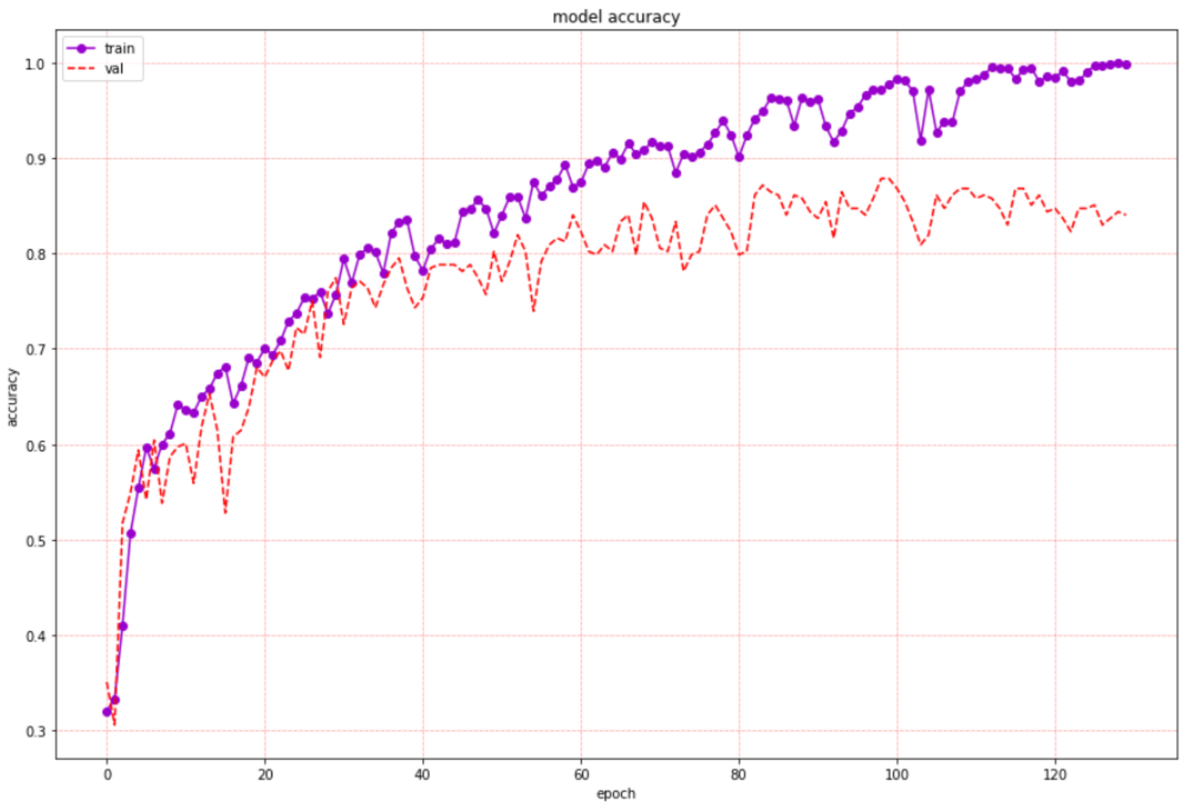

Figure 3: Result of model accuracy