

Combining Spectral Analysis with Artificial Intelligence in Heart Sound Study

Dariusz Kucharski¹, Marcin Kajor¹, Dominik Grochala¹,
Marek Iwaniec², Joanna Iwaniec^{2*}

¹ Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland

² Faculty of Mechanical Engineering and Robotics AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland

* Corresponding author's e-mail: joanna.iwaniec@agh.edu.pl

ABSTRACT

The auscultation technique has been widely used in medicine as a screening examination for ages. Nowadays, advanced electronics and effective computational methods aim to support the healthcare sector by providing dedicated solutions which help physicians and support diagnostic process. In this paper, we propose a machine learning approach for the analysis of heart sounds. We used the spectral analysis of acoustic signal to calculate feature vectors and tested a set of machine learning approaches to provide the most effective detection of cardiac disorders. Finally, we achieved 91% of sensitivity and 99% of positive predictivity for a designed algorithm based on convolutional neural network.

Keywords: deep learning, heart sound classification, convolutional neural network, machine learning, signal processing.

INTRODUCTION

Nowadays, heart diseases are certainly one of the most frequent causes of death all over the world. Recently, it has been a growing issue, especially in the well-developed countries, where such factors as intensive lifestyle, stress and aging society play a key role in the context of serious cardiac dysfunctions. It has been stated that almost 2000 people die every day because of a heart disease. This means one death every 44 seconds [2]. That is one of the reasons why a deployment of easily accessible medical services is currently of such a great importance. Today's healthcare must take advantage of the recent technological achievements to effectively follow up the demand for speeding up the diagnosis process, especially in terms of scanning examination. Machine learning, in conjunction with advanced signal processing, constitute a field of science

which is currently in the process of rapid development to meet the requirements of modern medical diagnostics. Auscultation is one of the most popular methods of quick patient examination. It has been widely and successfully used for centuries as a method of non-invasive scanning, especially in the context of cardiac disorders. This technique aims at recording the acoustic signal generated by the beating heart and blood flowing through it to determine any pathological sounds. In its original form auscultation is based on a pneumatic stethoscope which transfers vibrations through metal head and rubber tubes to the physician's ears. On the other hand, the today's medical engineering offers electronic stethoscopes capable of recording digital acoustic waveforms which can be later processed by means of digital signal processing and machine learning algorithms. Such an approach leads to the feature extraction and classification model which ultimately supports

physicians and may provide an automatic initial diagnosis. Combining spectral and time-frequency analysis with neural networks has been widely used by many researchers all over the world. Te-chung Issac Yang and Haowei Hsieh used recurrent neural network to detect the anomalies in heart sound. This approach is often applied in deep learning for time series as RNN system is capable of ‘memorizing’ the history associated with each data sample. Such a network is able to detect relevant events, even if there is a mutable time lag between their occurrences. During the research, after the segmentation and preprocessing step, the DFT was calculated for each data sample with the window size of 256 points and a 50% overlap to extract more augmented features for the input vectors of a classifier. As additional augmented information, the standard deviation and variance of used time windows were taken into account. It turned out that these parameters are good estimates of loudness in acoustic signals. The final results showed that the proposed system was able to classify the Physionet database signals with the accuracy of about 80% and may be further improved when more teaching data is provided [9]. Apart from standard FFT based spectral analysis, more sophisticated time-frequency domain methods like Short Time Fourier Transform or Wavelet Decomposition are also widely used for heart sound feature extraction. Nilanjan Dey et. al. used a Stationary Wavelet Transformation to develop a tele-diagnostic system for cardiac diseases based on a watermarked normal and abnormal heart sounds. The classification model was based on the Euclidean minimum distance between the test and training sample feature vectors. Finally, the accuracy of the proposed method was equal to 78% [3].

METHODS

During the research, we used an open access database for the evaluation of heart sounds algorithms. It consists of nearly 2500 heart sound recordings containing different types of heart disorders, such as heart valve disease and coronary artery disease [7]. Having consulted specialists, we selected 66 signals, half of them containing the valve defect and the rest representing healthy patients. The signals were used for verification of the prepared algorithm and designed approaches for detecting a valve defect in the heart’s acoustic

signal. Each signal was 8 seconds long and sampled with the frequency of 2000 Hz.

We started our research with the detection algorithms based on Fourier Transform and simple machine learning classifiers, such as multilayer perceptron and random forest. The analysis of frequencies typical for a particular disorder led us to use more sophisticated and complex neural networks, such as convolutional neural networks and some combination of different kinds of techniques. While looking for the best possible accuracy, we still kept in mind the fact that the developed algorithm was to be implemented on embedded devices. Hence, we were trying to find a compromise between high efficiency and relatively low computational complexity of the proposed solution.

At the very beginning of the development process, we split the whole recordings dataset into three independent parts, to achieve reliable output of designed experiments. Afterwards, we obtained the following subsets: training set, validation set and test set with proportion of 0.6:0.2:0.2. Then, depending on the used approach, further preprocessing methods, like filtering, extracting features and data augmentation were applied.

Fourier transform on whole signal

After splitting the dataset into subsets a mean value was subtracted from each recording and then a Fourier Transform was calculated. Fourier transform decomposes a function of time (an acoustic heart signal in our case) into a frequency function, which represents frequency components of the signal. The result of Fourier transform is simply calculated using the formula (1):

$$F[S] = \int_{\mathbb{R}} f(x) e^{-2\pi i(x,S)} dx \quad (1)$$

where: i is an imaginary unit and (x, S) is a dot product.

This resulted in 8000 points of usable spectrum values (from 0 to $\frac{1}{2}$ sampling rate). Next, we extracted the values for the frequencies between 640 and 840 Hz. Using this particular spectrum section was justified as being the most informative one in the context of detecting heart disorders [6]. That was our input for a classifier.

Firstly, we decided to try a MLP (Multilayer Perceptron) classifier because it has been widely described in literature in the context of applying regularization technique called dropout which

allows user to prevent a model from overfitting. On the other hand MLP can be used to easily simulate nonlinearity by adding a hidden layer. Finally, the resulting model consisted of two layers, each was built as follows: first hidden layer was built of 256 neural units, then the second layer comprised 128 neurons.

Multilayer perceptron can be easily represented as matrix multiplication:

$$y=f(x \cdot W)+b \tag{2}$$

where: x is an input vector,
 W is a weight matrix and
 f is the activation function.

If the x is a vector $1 \times n$, then the weight matrix W is of size $n \times m$, where n is a number of features, and m is a number of neuron in a layer.

Moreover, dropout was added before each fully-connected layer. Dropout is a regularization technique which helps to prevent overfitting. At each epoch of training, individual neurons are dropped with probability $1 - p$, where p is a hyperparameter defined at the beginning of the learning. The output consisted of 2 neurons clipped together with softmax activation function:

$$\text{softmax}(x)=\frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \tag{3}$$

$$\text{for } j=1 \dots K, z=(x_1, \dots, x_K) \in \mathbb{R}$$

where x is an output vector from the last layer of neural network. Simply speaking, softmax function binds output values so that their sum sums up to 1 and each of them represents probability of occurrence of each class.

In our case m equals 256 in the first layer and 128 in the seconds one. The F symbol represents an activation function and b is a bias added to the result. The output y constitutes an input for the next hidden layer. Dropout was also added before each fully-connected layer. The output consisted of 2 neurons clipped together with softmax activation function. After each hidden layer, a hyperbolic activation function was used. The approach with neural network resulted in accuracy of 75% [4]. The obtained result seemed reliable however many matrix multiplications discouraged us from using MLP as our final classifier on embedded system. The presented approach

proved that there was a way to distinguish physiological heart recording from pathological one utilizing Fourier transform. Hence we decided to try another powerful nonlinear classifier called Random Forest. This method features self-overfitting prevention achieved by using tools such as bagging and boosting. What is more, the classifier has a form of a forest – a data structure, which is computationally much less demanding than a multilayer perceptron (logn in comparison with $m1 \cdot m^2$, where n is a depth of the tree and $m1$ and $m2$ are as follows: size of the dataset and a number of hidden units). While training the Random Forest algorithm we took advantage of scikit-learn framework [8]. The result was 71% accuracy using 5-fold cross validation technique [4]

Analysis of a single heartbeat

Promising results achieved by applying the Fourier Transform on a whole signal encouraged us to try a frequency analysis on a single heartbeat section. We assumed that the Fourier Transform calculated on a very specific region which represents a single heart systolic event would contain much more useful information concerning a disorder. For that reason we designed a method for manually cutting out the regions of interest from the whole signal. As a result we got 616 subsections each about 1000 samples long. Next, we filled each subsection with zeros on the beginning and the end to achieve a length of 2048. Finally, we performed the Fourier Transform at each calculated systolic event. As in the previous paragraph, we extracted a part of the spectrum between 640 and 876 Hz. The resulting feature vectors constituted an input of another neural network.

Again we started with a simple, single hidden layer MLP with the 241 samples input vector. The hidden layer consisted of 256 neural units. A 0.5 dropout was applied subsequently.

The result from a first approach was quite mediocre as we achieved only 52% of accuracy. The further step aiming to tune neural network involved extending the network's memory by multiplying the number of hidden units by 2. This gave us 512 neurons, because of overfitting occurring in the validation set. That upgrade resulted in 57% accuracy and still did not indicate any overfitting. That is why the next step included introducing some more nonlinearity by adding another hidden layer with 128 neurons and 0.5 dropout.

After receiving even worse result we decided to play a little with a different number of hidden units and layers. Nonetheless, surprisingly a network with two hidden layers performed worse than the one with only one layer. For that reason we focused more on extracting some more representative feature vectors. What is more, a heart-beat detection algorithm would add some deterioration in the overall performance, hence we finally decided to try a model with more sophisticated feature extraction mechanism.

Combining Convolutional Neural Network with a spectrogram

While testing different approaches, aside from a Fourier Transform, we were investigating a spectrogram generated for a whole signal. The STFT (Short-Time Fourier Transform) algorithm was used, since it has been widely applied in the speech recognition systems based on a time-frequency domain [1]. After a visual analysis of the spectrogram, we found that there is a significant intensity of some higher frequencies in the case of pathological beats, whilst physiological signals lack this particular spectrum section. Combining these conclusions with the power of CNNs (Convolutional Neural Networks) in terms of extracting features from images, we decided to adapt such a technique in our research. Figure 1 represents differences between feature maps generated for physiological and pathological signals.

Convolutional neural networks mainly consist of convolutional layer. Convolutional layer, in comparison with fully connected layer, uses convolutional operation to calculate the layer's

results. Simple discrete convolution in two dimensions is defined as(4),

$$c[m,n] = (a * h)[m,n] = \sum_j \sum_k h[j,k] a[m-j,n-k] \tag{4}$$

where m and n defines the size of a matrix.

Owing to the convolutional layer combined with decimation layer (e.g. maxpooling) applied on the spectrogram, a single heartbeat detection step can be skipped because a maxpooling would choose local maximum values at each layer anyway. In the case of stacking this technique would lead to a model independent from the heartbeat position. In other words, the model would always choose maximum values, wherever they are placed on the image.

At the very beginning, we applied a data augmentation technique on each subset (as mentioned above, subsets – training, validation and test sets had been selected from the dataset before). From each recording, a couple of 4 seconds long sections were cut out with a shift equal to 0.5 second. Thus, we achieved more examples for training neural networks. What is more, due to extracting beat sections from a single recording with overlap, we received similar signals but with systolic events shifted relatively to one another. The first step of the preprocessing part involved the calculation of the spectrogram with STFT window of 500 samples and a shift of the window equal to 50 samples. Next, we subtracted a global minimum value from each spectrogram atom and divided the result by the global maximum. In order to emphasize the values at higher frequencies,

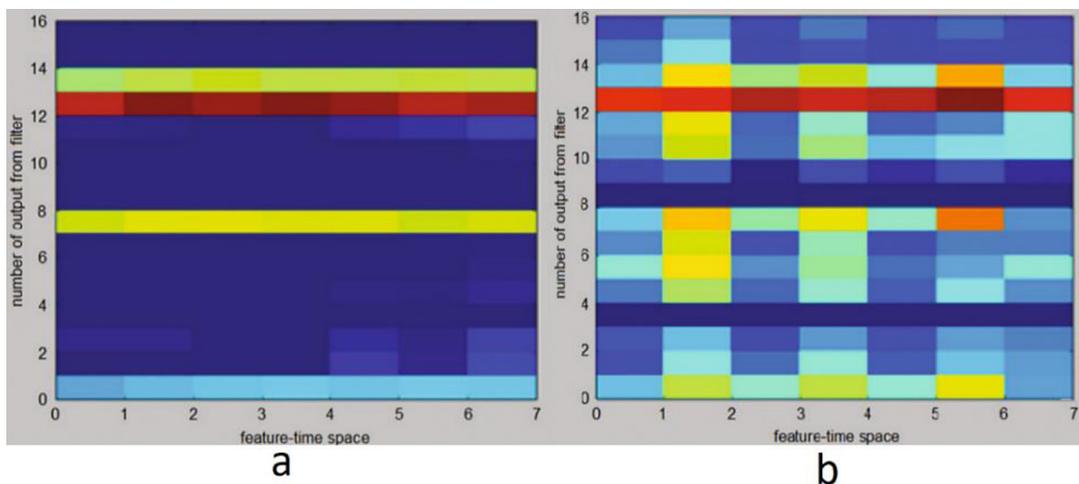


Fig. 1. Feature map for a) physiological signal and b) pathological one

a spectrogram was transformed to the logarithmic scale. Then, we extracted only the frequencies between 640 and 876 Hz. Finally the preprocessed data resulted in an image of size 60 by 151 pixels [6]. Preprocessing steps described above are depicted in Figure 2.

While composing the architecture of our neural network, we based on the similar research described in the literature as well as our own experience in this field.

The final system was built up with three convolutional layers, each of which consisted of as follows: 16 filters of size 7×7, 32 filters of size 5×5 and 32 filters of size 3×3. Additionally, each layer was followed by max pooling layer of size 2×2. Then, after the mentioned convolution-max pooling block, we implemented another convolutional layer with 32 filters of a size 3×3 and another convolutional layer with 16 filters, but this time with size of 1×1.

The main classification part consisted of a fully connected layer (with 32 neurons) and a softmax layer as an output. What is more, before both the hidden layer and the output layer, we

applied dropout of 0.5 and 0.2 respectively. Such architecture resulted in a final score of 91% accuracy and is presented in Figure 3.

RESULTS

In the end we calculated sensitivity and positive predictivity. Finally, the deep neural network model achieved 91% sensitivity (Se) and 99% positive predictivity (PP) where the positive result was considered as pathological class and the negative as a physiological one.

Table 1. Summarization of results achieved during training of deep neural models

	Se	PP	F1-SCORE
CNN + MLP	91	99	94

CONCLUSIONS

In this research we proposed a novel algorithm for detection of cardiac valve

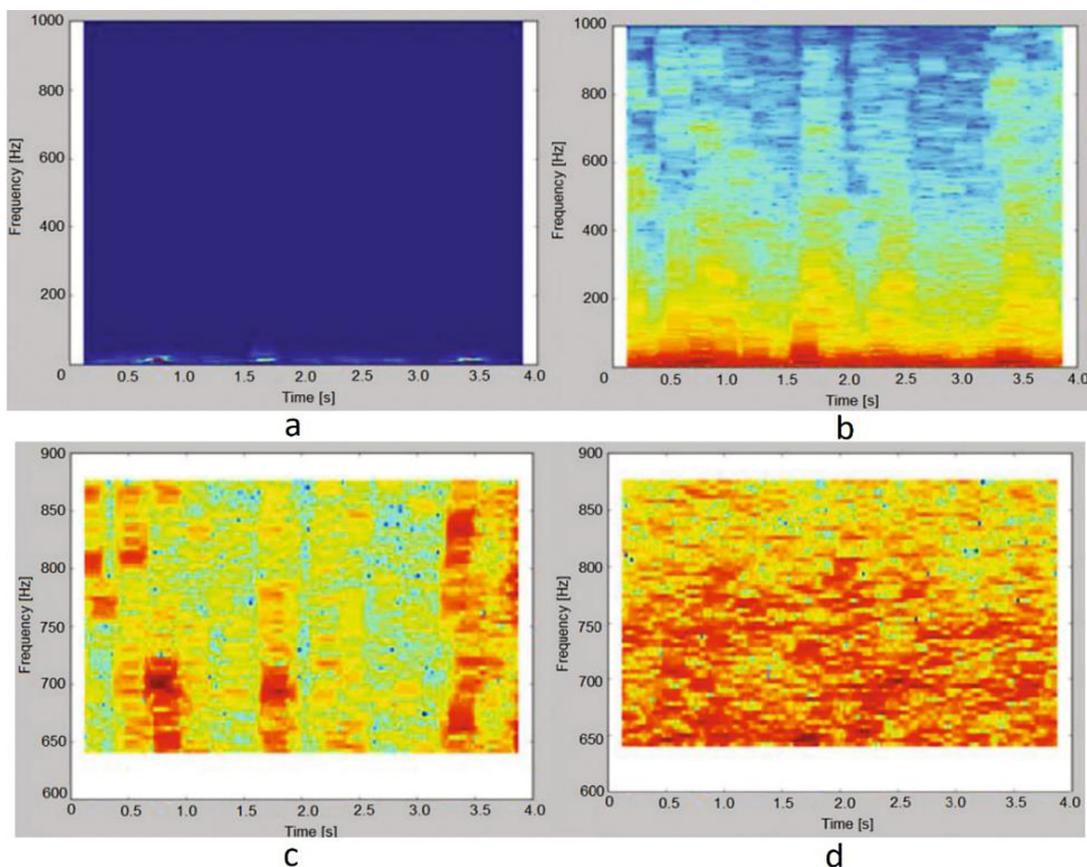


Fig. 2. Preprocessing steps for calculated spectrogram: a) after calculation b) after normalization and logarithm function application c) after snipping out between 640 and 876 Hz (extracted from pathological example) d) after snipping out between 640 and 876 Hz (extracted from physiological example)

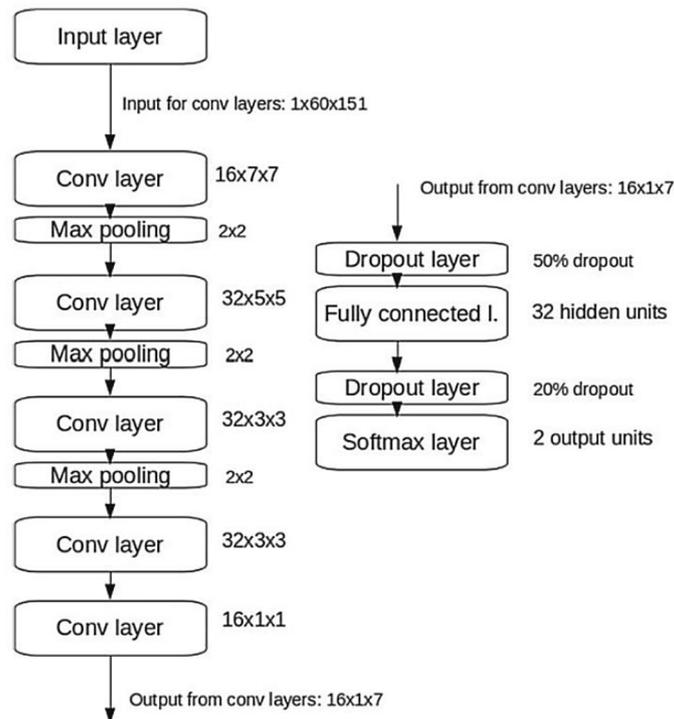


Fig. 3. Visualization of convolutional neural network architecture used in the research

abnormalities based on heart sounds. Presented method can be implemented as improvement of auscultation diagnosis, making it more objective and sensitive. An important conclusion is that achieved accuracy of our model justifies a potential of the method to be used in real clinical applications to support pre-diagnose of heart disorders. A final target of our idea is to implement the designed algorithms on embedded platform with a real time analysis of acoustic heart signals provided as a system input. While the Fast Fourier Transform is well established in a literature, using it as a feature extractor gave us a bit worse results in comparison to deep neural networks approach. On the other hand, when it comes to CNN, many convolution operations might imply significant power consumption which is not desirable in the context of low energy embedded system. Taking into consideration both the efficiency of calculations and the performance in terms of a signal analysis, there is still a place to find a classifier which combines fast execution and robust feature extraction. In the future, we are planning to integrate the presented algorithm with our own hardware prototype [5] and start a pilot research to create a dedicated heart sound database. This will allow us to extend the training sets and verify the model generalization on

new signal examples. Such an approach is necessary for adapting and tuning neural network parameters so that the algorithm is effective in the context of a wider spectrum of cases.

REFERENCES

1. Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., et al. (2016). Deep speech 2: End-to-end speech recognition in english and mandarin. In: International Conference on Machine Learning, pp. 173–182.
2. Azra'ai, R.A., bin Taib, M.N., Tahir, N.M. (2008). Artificial neural network for identification of heart problem. Signal Processing and Communication Systems, 2008. In: ICSPCS IEEE 2nd International Conference on, pp. 1–6.
3. Dey, N., Mishra, G., Nandi, B., Pal, M., Das, A., Chaudhuri, S.S. (2012). Wavelet based watermarked normal and abnormal heart sound identification using spectrogram analysis. In: Computational Intelligence & Computing Research (ICIC), IEEE International Conference on, pp. 1–7. IEEE.
4. Grochala, D., Kajor, M., Kucharski, D., Iwaniec, M., Kantoch, E. (2018). A Novel Approach in Auscultation Technology-New Sensors and Algorithms. In: IEEE 11th International Conference on Human System Interaction (HSI), pp. 240–244..
5. Kajor, M., Grochala, D., Iwaniec, M., Kantoch, E., Kucharski, D. (2018). A prototype of the mobile

- stethoscope for telemedical application. In: IEEE XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), pp. 5–8.
6. Kucharski, D., Grochala, D., Kajor, M., Kańtoch, E. (2017). A Deep Learning Approach for Valve Defect Recognition in Heart Acoustic Signal. In: International Conference on Information Systems Architecture and Technology, pp. 3–14, Springer, Cham.
 7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
 8. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
 9. Yang, T.C.I., Hsieh, H. (2016). Classification of acoustic physiological signals based on deep learning neural networks with augmented features. In: 2016 IEEE Computing in Cardiology Conference (CinC), pp. 569–572.