Automatic Detection of Heartbeats in Heart Sound Signals Using Deep Convolutional Neural Networks

Grega Vrbancic, Iztok Jr. Fister, Vili Podgorelec
Faculty of Electrical Engineering and Computer Science, University of Maribor,
Koroska cesta 46, SI-2000 Maribor, Slovenia
grega.vrbancic@um.si

Abstract—The analysis of non-stationary signals commonly includes the signal segmentation process, dividing such signals into smaller time series, which are considered stationary and thus easier to process. Most commonly, the methods for signal segmentation utilize complex filtering, transformation and feature extraction techniques together with various kinds of classifiers, which especially in the field of biomedical signals, do not perform very well and are generally prone to poor performance when dealing with signals obtained in highly variable environments. In order to address these problems, we designed a new method for the segmentation of heart sound signals using deep convolutional neural networks, which works in a straightforward automatic manner and does not require any complex pre-processing. The proposed method was tested on a set of heartbeat sound clips, collected by non-experts with mobile devices in highly variable environments with excessive background noise. The obtained results show that the proposed method outperforms other methods, which are taking advantage of using domain knowledge for the analysis of the signals. Based on the encouraging experimental results, we believe that the proposed method can be considered as a solid basis for the further development of the automatic segmentation of highly variable signals using deep neural networks.

Index Terms—Artificial neural networks; Machine learning; Segmentation.

I. INTRODUCTION

Generally speaking, signals can be divided into two main categories: deterministic signals and random or statistical signals. Based on the behavior, statistical signals are divided into a group of stationary signals, which statistical characteristics do not change over time, and into a group of non-stationary signals, which may have variable quantities in time, such as mean value, dispersion or frequency spectrum [1]. Based on their properties, stationary signals are easier to process than non-stationary signals. When dealing with non-stationary signals, it is common to divide such signals into smaller time series, making those time series considered as stationary signals. The described process is known as segmentation [2].

Signal segmentation is considered to be one of the fundamental problems of digital signal processing in various information, monitoring, prediction and control systems in a wide range of fields [3]. In general, signal segmentation could be performed in two different ways, namely constant segmentation and adaptive segmentation. Constant segmentation is quite simple, as it basically divides the signal into fixed length segments, but generally provides an extremely weak performance. With such a straightforward approach, it is only by chance that the segment boundaries will fall at the points of change of the signal; therefore, even considering the fixed length segments may annul the purpose of segmentation itself. On the other hand, adaptive segmentation is more complex, usually delivers higher performance, and most importantly detects signal boundaries automatically, depending on the statistical characteristics of the signal [4].

There are various segmentation methods presented in the literature [5]–[7] applied in a wide range of domains, from music [8], and seismology [9], to astrology [10], and medicine [1]–[3]. Most commonly these methods utilize more or less complex conventional techniques like wavelet transformation [4], and frequency analysis [11], etc. to extract features from the signal and perform some kind of classification over them like is presented in [12]–[14]. In recent work, the authors have also presented various kinds of hybrid methods tackling the signal segmentation tasks, including the heart sound signal segmentation tasks [15]. Such hybrid methods are most commonly composed of artificial neural networks, specifically developed for time series processing, and some kind of classifier like Random Forest, Support Vector Machine, feed-forward neural network, convolutional neural network, etc.

However, these approaches, especially in the field of medicine, have not performed very well, due to the inter-patient variations of biomedical signals, leading to the inconsistent performance of such methods when, for instance, one is classifying a new subject's biomedical signals [16]. In recent years, the authors [17], [18] have started addressing the problem with the use of feature extraction capabilities of various types of deep neural networks and also for classification tasks. Not only have such approaches demonstrated significant performance improvements over conventional methods, but they are also less computationally demanding as more conventional methods. The extraction of several features using the

conventional methods, especially in the transform domains along with the post-processing methods, may significantly increase the computational complexity of the overall process, which can reduce the possibility of their use in lightweight applications such as wearable heart monitoring devices, smart watches, etc. [19].

We are witnessing a massive increase in various tracking activity devices and mobile activity applications as well as various wearable health monitoring devices that would benefit significantly from general, widely applicable, robust and straightforward segmentation methods. In general, the majority of existing methods were developed and tested on datasets obtained by experts under highly controlled conditions and are prone to mistakes in the presence of noise in the signal [20]. Encouraged by the results of utilizing convolutional neural networks (CNNs) for the purpose of extracting features from EEG signals to classify neural disorders in our previous work [21], in this paper, we are proposing a new method for the automatic segmentation of heart sound signals collected in highly variable conditions using deep CNNs.

As such, in contrast to the majority of existing methods, the proposed method does not require any expert intervention in the sense of using domain knowledge to guide the classification process. The main goal of our research is to study whether the proposed method can extract important features from signals captured in highly variable environments and also successfully detects the heartbeats from it with a level of performance that is at least comparable to other methods utilizing deep neural networks. The main advantage of the proposed method is its direct use of sound signals without any complex feature extraction processes or pre-processing, nor does it require any additional domain expert knowledge and it is designed to work in a straightforward, automatic manner.

The rest of the paper is structured as follows. In Section II, the proposed method is presented in depth. Section III presents experimental settings, while Section IV presents the results of the conducted experiment. Lastly, Section V concludes the paper with our final thoughts and some future work possibilities.

II. PROPOSED METHOD

The proposed method for signal segmentation utilizing deep CNNs, consists primarily of three phases. The first phase includes the processing of the recorded sound signal. In the second phase, the training of CNN for the task of identifying the signal anomalies (in our case heartbeats) is performed and in the last phase or post-processing, the labeling of detected anomalies (heartbeats) is performed. Each of these phases is elaborated upon in the following sections.

A. Signal Processing

At the input, our method receives raw heartbeat activity recordings in the form of raw stereo sound signals with a sampling rate of 44,100 Hz. The obtained signal data is element-wise standardized using the well-known standardization method z-score normalization, presented in (1) where z_i denotes the calculated standardized value for

the element x_i from a group of elements x, while $\text{mean}(x)$ and $\text{std}(x)$ represents the statistical average value and the standard deviation of elements in group x respectively

$$z_i = \frac{x_i - \text{mean}(x)}{\text{std}(x)}. \tag{1}$$

Standardization is quite a common procedure in the pre-processing phase in the field of machine learning [22]. In our case, it also enables us to transform the signal data to a more uniform scale, because the heartbeat signal recordings were obtained in various uncontrolled environments with different scale ranges. Different scale ranges could, in the phase of training the CNN, cause difficulties and result in a poor performance of the CNN model. The outcome of applying the standardization method to raw signals is shown in Fig. 1.
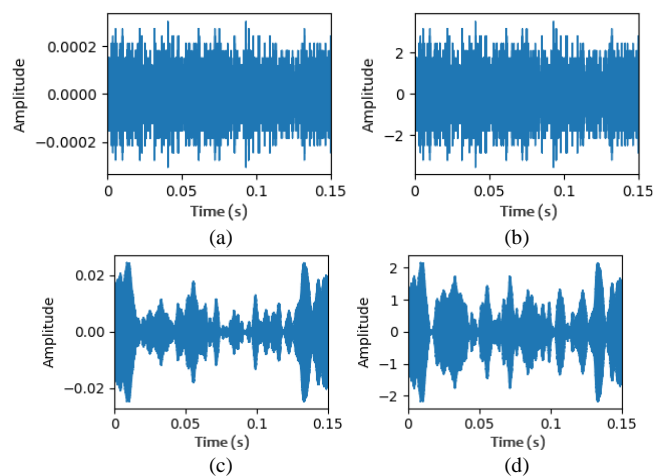

Fig. 1. Sound signal images before and after applied standardization: (a) and (c) represent the splits of two different sound clips in a raw form, while (b) and (d) represent the same sound clips after standardization.

After the standardization of the raw signal data, each signal recording is divided into multiple frames using the sliding window algorithm. For a window size, we chose 6,615 samples which translates to a length equal to 150 ms and step size set as one-third of the window – 2,205 samples or 50 ms. In the case of the last frame being shorter than the chosen windows size, the frame is at the end padded to the same length with zero (0.0) values. The basis for selecting the windows size are the ANSI/AAMI EC38 [23] and EC57 [24] standards which state that the estimated location of the heartbeat is deemed to be accurate if it is no further than 150 ms from the corresponding annotated (real) location of a heartbeat. This gives us a time window in total of 300 ms (150 ms before and after annotated location) inside which any detected heartbeat is deemed accurate. Based on this, the mentioned time windows of 300 ms is equal to the size of our two frames. The step size, one-third of the window size, is chosen because this way each data point in any frame (except the first two frames) is captured in three windows, which increases the possibility of detecting the potential heartbeat correctly.

B. Convolutional Neural Network

The CNNs were firstly developed and presented in

Fukushima's paper [25] in the 1980s. The proposed deep learning approach, known as neocognitron, was based on hierarchical layers trained with the use of the stochastic gradient descent algorithm. The real breakthrough in the field of machine learning happened later in 1998 with the proposal of LeCun's LeNet5 [26] architecture, which is one of the very first CNNs and is considered to be a factor that propelled the field of deep learning. Currently, the CNNs are generally known to have a great ability to extract features from various kinds of signals (e.g., image, video, etc.) while achieving near-human performance [27], [28].

The deep CNNs were initially developed as feed-forward and 2D constrained neural networks with alternating convolutional and subsampling layers fully connected at the end. They combine three architectural ideas: local receptive fields, shared weights and spatial and temporal subsampling. The CNN ensures some degree of shift and distortion invariance [26]. A convolutional layer can be considered as a fuzzy filter, which enhances the features of the original signal and reduces the noises. Basically, it models the cells in the human visual cortex [29]. A convolutional layer is most commonly composed of several feature maps, which are calculated with different weight vectors, enabling us to extract multiple features from each location. The convolution operation is performed between feature maps of the previous layer and convolution kernel of the current layer in addition with the activation function, which gives the results of convolution calculations. The formal characterization of the output of the convolutional layer is presented in (2) where X_j^l stands for the characteristic vector corresponding to the j-th convolution kernel of the l-th layer and M is the receptive field of the current neuron. W_{ij}^l indicates the weight and the b_j^l denotes the bias coefficient appropriated to the j-th convolution kernel of the l-th layer while f is a nonlinear function such as ReLU [30], Softmax [31], etc.,

$$X_j^l = f(\sum_{i \in M_j} X_j^{l-1} \times W_{ij}^l + b_j^l). \qquad (2)$$

A subsampling layer reduces the dimension of feature maps while preserving the important extracted features, performing local averaging and subsampling. A subsampling is possible because the real locations of the extracted features are not important as long as their approximate positions relative to others remain the same.

While the initial CNNs were not intended to work with one-dimensional signals, we had to make small adjustments to the architecture of our CNN to ingest a one-dimensional signal. Generally, the one-dimensional convolutional layers can be regarded as conventional two-dimensional layers, with the exception of the second dimension of layers being equal to one. In this manner, the feature maps are calculated utilizing the convolution operation on the subsection of the previous layer with the sliding kernel only in one direction.

For the purpose of detecting the heartbeat in the sound signal, we designed a simple, straightforward CNN architecture presented in Fig. 2. The architecture consists of two one-dimensional convolutional layers, each followed by a one-dimensional subsampling layer applying the maximization function and at the end one fully connected layer with an output layer.

In the first convolutional layer, 20 convolution kernels with a length of 661 sampling points are distributed over the input audio signal with 6,615 sampling points. The output, calculated using a ReLU activation function of the first convolutional layer, is then connected to the neurons in the first subsampling layer with a pool size of 20 and stride of 10 samples, producing 20 feature maps with a length of 659, which are connected to the second convolutional layer. In the second convolutional layer, 50 kernels with a length of 440 are distributed over the input, producing the output calculated with the same activation function as in the first convolutional layer. There follows the second subsampling layer with the same kernel size and stride as the first one, producing 50 feature maps with a length of 64, which are fully connected to the layer with a length of 3,200 sampling points. Finally, the output layer, with Sigmoid activation function, is applied, which classifies the extracted features of the fed input audio signal as heartbeat (as 1) or as no heartbeat (as 0).
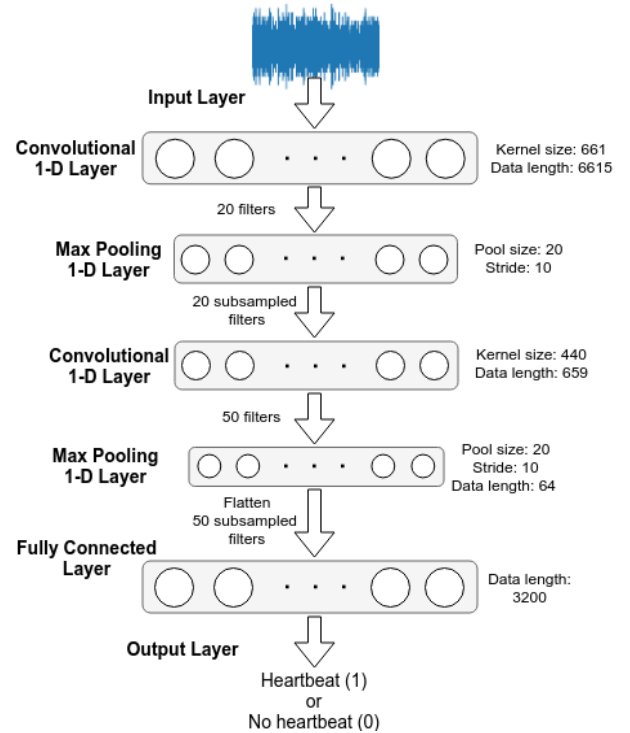

Fig. 2. The architecture of the proposed CNN method.

C. *Post-processing*

The post-processing phase is a procedure in which we determine the locations (samples) in the heartbeat sound clip in which the heartbeats occurred. From the splits of the original sound clips, using the proposed CNN model, we obtain probabilities and predicted classes for each split. In order to determine the exact location of a heartbeat, we propose a three-step procedure. In the first step, each occurrence or series of occurrences of predicted class 1 among the obtained predictions is transformed into a "candidate" ranges of splits. After the construction of the so-called candidate ranges, the selection of the single split (from each candidate range) is performed in the second step. The selection is based on the obtained probability for each split, in our case, the split with the highest probability is

selected, forming a group of candidate splits (CS). Having selected one split from each candidate range, in the third step, the selection of a sample from each candidate split was done following (3), where TS denotes the calculated location of a heartbeat, i represents the single selected split from the group of candidate splits CS, and L represents the length of a split

$$TS_{i \in CS} = \left| \frac{L_i}{2} \right|. \tag{3}$$

III. EXPERIMENTAL SETTINGS

To test the proposed method for automatic signal segmentation, we used a collection of annotated heartbeat sound clips, initially prepared for Classifying Heart Sounds Challenge (Dataset A) [32]. An annotated collection of sound clips contains 21 recordings captured with iStethoscope Pro iPhone app with a sampling rate of 44,100 Hz and stored in the form of .wav files. The recordings were captured in uncontrolled environments. Associated with those recordings is a CSV file that contains the annotations of heartbeats for each recording. The annotations are presented in the form of a sample number for each heartbeat. The lengths of the recordings vary between 1 seconds and 30 seconds, some of the recordings were also clipped in order to reduce excessive noise.

For the purpose of fair and in-depth validation of our experimental results, we performed the well-established methodology of 10-fold cross-validation. Because of the task we are addressing and also because of the nature of the data we are working on, we implemented the 10-fold cross-validation as follows. First, we randomly divided 21 audio recordings into 10 disjointed parts (folds). In the next step, we trained our CNN model on 9 out of the 10 disjointed parts and tested the performance of the model on the remaining one. We repeated this procedure for a total of 10 times, each time testing the performance with the different remaining disjointed part.

The training parameters were carefully picked, based on our previous experience with CNNs and practical recommendations [22]. The training was done utilizing efficient mini-batch training [33] with a batch size of 256 epochs and 150 epochs. The Adam [34] was picked as the optimization function with a learning rate of 10^{-3}.

Training of our CNN model was performed on a machine with an Intel Core i7-6700K CPU running at 4.0 GHz, 64 GB of RAM and three dedicated Nvidia Titan X graphics cards each with 12 GB of dedicated GDDR5 memory.

IV. RESULTS

The reported results were all obtained performing 10-fold cross-validation. Given the specifics of the problem which we were solving we also used, in addition to the conventional classification metrics such as classification accuracy, the performance metrics – averaged distance from real heartbeat locations, from the previously mentioned Classifying Heart Sounds Challenge. The latter is formally defined as presented in (4) and (5), where δ_k denotes the calculated averaged distance from real heartbeat locations of k-th sound clip, N_k is the total number of heartbeats present in the k-th sound clip, RS_i indicates the real location of the i-th heartbeat and the TS_i indicates the predicted location of the i-th heartbeat, while δ_{total} represents the total sum distance over all sound clips in the each fold, where j is denoting the total number of clips:

$$\delta_k = \frac{\sum_{i=1}^{N_k} |RS_i - TS_i|}{N_k}, \tag{4}$$

$$\delta_{total} = \sum_{k=1}^{j} \delta_k, \tag{5}$$

$$\delta_{average} = \frac{1}{l} \sum_{k=1}^{l} \delta_k. \tag{6}$$

In addition, to present the performance of average distance over sound clips in each fold, we are also introducing the $\delta_{average}$ metric which is formally presented in (6), where l denotes the number of sound clips in a fold.

Beside the before presented heart beat distance metrics, we are also presenting the classification performance of our proposed method, which is utilized for detection of heart beats. The classification performance is measured using well known classification metrics: accuracy, sensitivity or recall, specificity, precision and f-1 score.

The results of the conducted experiment are presented in Table I and Table II. All of the measured metrics are reported for each fold separately as well as overall for all 10 folds. The distance metrics are reported as a number of samples.
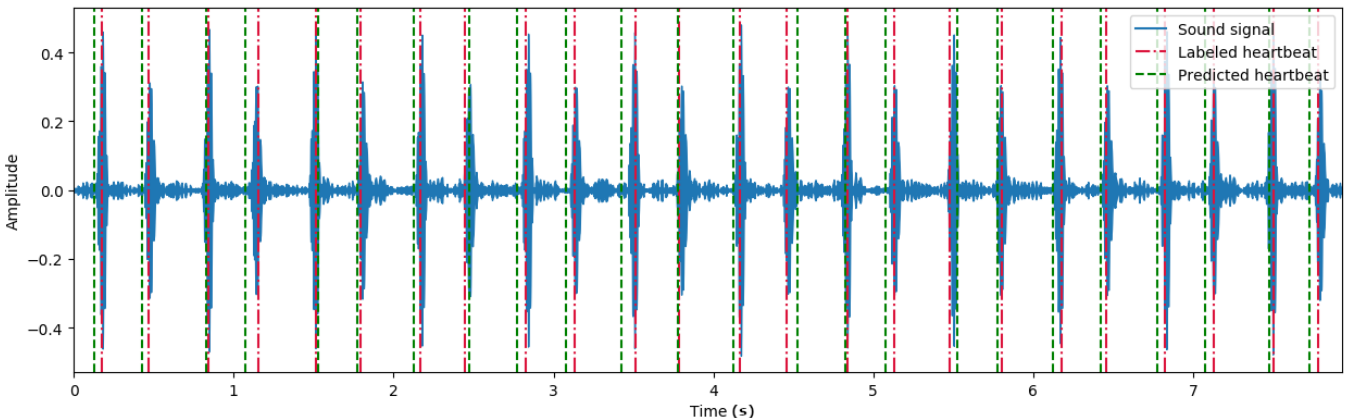

Fig. 3. Sound wave plot with real, labeled heartbeats and predicted heartbeats obtained as a result of our proposed method.

TABLE I. SIGNAL SEGMENTATION PERFORMANCE.

Fold	δ_{total}	$\delta_{average}$	Predicted heartbeats in range ±150 ms
Fold 1	5,272.71	1,748.6	92.98 %
Fold 2	7,877.47	3,666.38	81.00 %
Fold 3	6,996.14	3,439.73	75.00 %
Fold 4	27,128.24	14,477.48	85.00 %
Fold 5	9,621.66	4,634.54	75.89 %
Fold 6	46,516.4	27,046.11	64.58 %
Fold 7	29,618.51	17,249.76	59.17 %
Fold 8	6,847.81	3,447.66	96.43 %
Fold 9	8,209.55	4,258.08	83.33 %
Fold 10	17,806.52	9,566.91	86.11 %
Average	16,589.50	7,687.89	79.95 %
Sum	165,895.01	-	-

When comparing the average distance of the predicted location from the real heartbeat location, our proposed method outperforms the existing methods [20], [35]. The average distance error over the folds varies from 1,748.6 to 27,046.11 with a standard deviation of 8,221.53 and averages at 7,687.89. Focusing on the overall average distance of predicted heartbeat locations from real heartbeat locations, we can see that our predicted results are, on average, in the defined range of 300 ms (13,230 samples) by ANSI/AAMI standards. The rate of correctly predicted heartbeats within the mentioned range is also quite promising and averages at 79.95 %.

In contrast to the existing method from [36], when analyzing the results of our proposed method we cannot observe and detect any performance issues based on the length of the sound clip.

TABLE II. METHOD CLASSIFICATION PERFORMANCE.

Fold	Accuracy	Sensitivity (Recall)	Specificity	Precision	f-1 score
1	82.66 %	83.13 %	82.22 %	81.18 %	82.14 %
2	75.17 %	70.63 %	78.41 %	70.08 %	70.36 %
3	70.73 %	59.04 %	75.49 %	49.49 %	53.85 %
4	82.60 %	69.44 %	92.31 %	86.96 %	77.22 %
5	77.16 %	72.22 %	80.28 %	69.89 %	71.04 %
6	77.60 %	55.56 %	92.15 %	82.35 %	66.35 %
7	73.60 %	60.00 %	78.38 %	49.37 %	54.17 %
8	87.77 %	80.95 %	90.21 %	74.73 %	77.71 %
9	75.98 %	70.37 %	78.67 %	61.29 %	65.52 %
10	75.30 %	77.78 %	73.91 %	62.50 %	69.31 %
Avg.	77.86 %	69.91 %	82.20 %	68.78 %	68.77 %
Std. dev.	5.06 %	9.35 %	6.86 %	13.13 %	9.37 %

The classification accuracy over each fold varies from 70.73 % to 87.77 % with a standard deviation of 5.06 %, which is slightly better than the [13] and significantly better than the [15] where the score for the dataset A is 55 %. Observing the other classification metrics, such as sensitivity and specificity, and comparing them with the results from [13], [15], and [37], we can see that our method achieved better balance of sensitivity vs. specificity, unlike the proposed method in [15], which produces great sensitivity of 99 % but fails to deliver specificity higher than only 11 %. Based on the folds, the f-1 score, the measure that considers both precision and recall, is averaging at

68.77 % with the standard deviation of 9.37 %.

When observing the Fig. 3, we can see an example of a heartbeat sound clip, with its amplitude and with vertical lines showing the real and the predicted heartbeat samples. The predicted locations of the heartbeats in the presented sound clip are all in the defined range of 300 ms versus the real location.

V. CONCLUSIONS

In this paper, we presented a new method for the automatic segmentation of heart sound signals using deep CNNs. The heart sound signals, used to evaluate our proposed method, were captured in highly variable environments with excessive background noise. Although there are existing methods which use CNNs for classification in signal segmentation, such methods are commonly utilizing excessive pre-processing techniques to extract features from the signal based on domain knowledge. In contrast, our proposed method does not require these demanding feature engineering tasks. In our approach, we defined a multi-layered deep neural network architecture tailored to the process of signal segmentation. In this manner, the proposed method performs the segmentation and detects heartbeats in a fully automatic manner. The results obtained from the conducted experiment are very promising when compared to the existing methods, especially considering the nature of the used sound recordings and the totally straightforward, automatic manner of our proposed method.

Based on these promising results, we would like to extend our work and apply more advanced algorithms for the calculation of predicted locations of sound beats, as well as improve the classification performance of our CNN by introducing different learning strategies and utilize some neural network hyperparameter optimization methods.

REFERENCES

[1] H. Azami, S. Sanei, K. Mohammadi, "A novel signal segmentation method based on standard deviation and variable threshold", *International Journal of Computer Applications*, vol. 34, no. 2, pp. 27–34, 2011.
[2] H. Azami, K. Mohammadi, H. Hassanpour, "An improved signal segmentation method using genetic algorithm", *International Journal of Computer Applications*, vol. 29, no. 8, pp. 5–9, 2011. DOI: 10.5120/3586-4967.
[3] A. Prochazka, M. Kolinova, J. Stribrsky, "Signal segmentation using time-scale signal analysis", *9th European Signal Processing Conf. (EUSIPCO 1998)*, Rhodes, Greece, 1998, pp. 1–4.
[4] H. Hassanpour, M. Shahiri, "Adaptive segmentation using wavelet transform", *Int. Conf. Electrical Engineering (ICEE 2007)*, Lahore, Pakistan, 2007, pp. 1–5. DOI: 10.1109/ICEE.2007.4287348.
[5] S. Choi, Z. Jiang, "Comparison of envelope extraction algorithms for cardiac sound signal segmentation", *Expert Systems with Applications,* vol. 34, no. 2, pp. 1056–1069, 2008. DOI: 10.1016/j.eswa.2006.12.015.
[6] E. Punskaya, C. Andrieu, A. Doucet, W. J. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation", *IEEE Trans. Signal Processing,* vol. 50, no. 3, pp. 747–758, 2002. DOI: 10.1109/78.984776.
[7] H. Hassanpour, S. M. Anisheh, "An improved adaptive signal segmentation method using fractal dimension", *10th Int. Conf. Information Sciences, Signal Processing and their Applications, (ISSPA 2010)*, Kuala Lumpur, Malaysia, 2010, pp. 720–723. DOI: 10.1109/ISSPA.2010.5605569.
[8] O. Nieto, M. M. Farbood, "Identifying polyphonic musical patterns from audio recordings using music segmentation techniques", in *Proc. 15th Int. Society for Music Information Retrieval Conference (ISMIR 2014)*, 2014, pp. 411–416.

[9] T. D. Popescu, D. Aiordachioaie, "Signal segmentation in time-frequency plane using Rényi entropy - Application in seismic signal processing", *Conf. Control and Fault-Tolerant Systems (SysTol 2013)*, Nice, France, 2013, pp. 312–317. DOI: 10.1109/SysTol.2013.6693812.

[10] N. Dobigeon, J. Tourneret, J. D. Scargle, "Change-point detection in astronomical data by using a hierarchical model and a bayesian sampling approach", *IEEE 13th Workshop on Statistical Signal Processing (SP 2005)*, Bordeaux, France, 2005, pp. 369–374. DOI: 10.1109/SSP.2005.1628623.

[11] K. I. Minami, H. Nakajima, T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with fourier-transform neural network", *IEEE Trans. Biomedical Engineering*, vol. 46 no. 2, pp. 179–185, 1999. DOI: 10.1109/10.740880.

[12] C. C. Bali, M. C. C. Sobrepena, P. C. Naval, "Classification of heart sounds using discrete and continuous wavelet transform and random forests", *IAPR Asian Conf Pattern Recognition (ACPR 2015)*, Kuala Lumpur, Malaysia, 2015, pp. 655–659. DOI: 10.1109/ACPR.2015.7486584.

[13] K. Lifu, W.Weilian, "Heart sound signals based on CNN classification research", in *Proc. 6th Int. Conf. Bioinformatics and Biomedical Science*, Singapore, Singapore, 2017, pp. 44–48. DOI: 10.1145/3121138.3121173.

[14] E. F. Gomes, E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction", *Workshop Classifying Heart Sounds*, 2012.

[15] C. Thomae, A. Dominik, "Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound", *Computing in Cardiology Conference (CinC 2016)*, 2016, pp. 625–628. DOI: 10.22489/CinC.2016.183-214.

[16] P. D. Chazal, R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features", *IEEE Trans. Biomedical Engineering*, vol. 53, no. 12, pp. 2535–2543, 2006. DOI: 10.1109/TBME.2006.883802.

[17] C. Potes, S. Parvaneh, A. Rahman, B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds", *Computing in Cardiology Conf. (CinC 2016)*, 2016, pp. 621–624. DOI: 10.22489/CinC.2016.182-399.

[18] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, K. Sricharan, "Recognizing abnormal heart sounds using deep learning", in *Proc. 2nd Int. Workshop on Knowledge Discovery in Healthcare Data Co-located with the 26th Int. Joint Conf. Artificial Intelligence (IJCAI 2017)*, 2017, pp. 13–19, 2017.

[19] S. Kiranyaz, T. Ince, R. Hamila, M. Gabbouj, "Convolutional Neural Networks for patient-specific ECG classification", in *Proc. Annual International Conf. IEEE Engineering in Medicine and Biology Society (EMBS 2015)*, Milan, Italy, 2015, DOI: 10.1109/EMBC.2015.7318926.

[20] Y. Deng, P. J. Bentley, "A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram", *Workshop Classifying Heart Sounds*, 2012.

[21] G. Vrbancic, V. Podgorelec, "Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks", *Elektronika ir Elektrotechnika*, vol. 24, no. 4, pp. 1–7, 2018. DOI: 10.5755/j01.eie.24.4.21469.

[22] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures", *Lect. Notes Comput. Sci. (including Subser. Lect.*

[23] *Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700 LECTU, pp. 437–478, 2012. DOI: 10.1007/978-3-642-35289-8_26.

AMMI ECCAR, "Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms", *Association for the Advancement of Medical Instrumentation*, pp. 69, 1987.

[24] ANSI/AMMI, EC57, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms", 2012.

[25] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980. DOI: 10.1007/BF00344251.

[26] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[27] S. Dodge, L. J. Karam, "A study and comparison of human and deep learning recognition performance under visual distortions", *Computer Communication and Networks (ICCCN 2017)*, 2017, pp. 1–7. DOI: 10.1109/ICCCN.2017.8038465.

[28] E. A. Hay, R. Parthasarathy, "Performance of convolutional neural networks for identification of bacteria in 3D microscopy datasets", *Computational Biology*, pp. 273–318, 2018. DOI: 10.1371/journal.pcbi.1006628.

[29] D. H. Wiesel, T. N. Hubel, "Receptive fields of single neurones in the cat's striate cortex", *Journal of Physiology*, vol. 148, pp. 574–591, 1959. DOI: 10.1113/jphysiol.1959.sp006308.

[30] X. Glorot, A. Bordes, Y. Bengio, "Deep sparse rectifier neural networks", *JMLR W&CP*, no. 15, pp. 315–323, 2011.

[31] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition", *Neurocomputing*, vol. 68, pp. 227–236, 1990. DOI: 10.1007/978-3-642-76153-9_28.

[32] Bentley, P., G. Nordehn, M. Coimbra, S. Mannor, R. Getz, "The pascal classifying heart sounds challenge 2011 (chsc2011) results". [Online]. Available: http://www.peterjbentley.com/heartchallenge/index.html

[33] M. Li, T. Zhang, Y. Chen, A. J. Smola, "Efficient mini-batch training for stochastic optimization", in *Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD 2014)*, New York, New York, USA, 2014, pp. 661–670. DOI: 10.1145/2623330.2623612.

[34] D. Kingma, J. Ba, "Adam: a method for stochastic optimization", *arXiv preprint*, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980v8.

[35] N. Marques, R. Almeida, A. P. Rocha, M. Coimbra, "Exploring the Stationary Wavelet Transform detail coefficients for detection and identification of the S1 and S2 heart sounds", *Computing in Cardiology*, pp. 891–894, 2013.

[36] S. L. Strunic, F. Rios-Gutierrez, R. Alba-Flores, G. Nordehn, S. Burns, "Detection and classification of cardiac murmurs using segmentation techniques and artificial neural networks", *IEEE Symposium on Computational Intelligence and Data Mining*, 2007, pp. 128–133. DOI: 10.1109/CIDM.2007.368902.

[37] M. Singh, A. Cheema, "Heart sounds classification using feature extraction of phonocardiography signal", *International Journal of Computer Application*, vol. 77, no. 4, 2013. DOI: 10.5120/13381-1001.