Convolutional neural network approach for heart MurMur detection in auscultation signals using wavelet transform based features

Robertas Petrolis [1,2], Renata Paukstaitiene[1], Gabriele Rudokaite[3], Andrius Macas[4], Arturas Grigaliunas[1], Algimantas Krisciukaitis[1,2]

[1] Department of Physics, Mathematics and Biophysics, Lithuanian University of Health Sciences.
[2] Neuroscience Institute, Lithuanian University of Health Sciences.
[3] Department of Cardiology, Medical Academy, Lithuanian University of Health Sciences.
[4] Department of Anaesthesiology, Medical Academy, Lithuanian University of Health Sciences.

Abstract

The algorithm for classification of cardiac auscultation signals uses key-features calculated by wavelet transform and the decision is made by convolutional neural network trained on the annotated George Moody PhysioNet Challenge 2022 dataset.

The heart auscultation signal contains strong beats representing cardiac valve closures and the MurMur sounds (if present) following, or even partially overlapping with the beats. These key-components of the signal differ in time and frequency, therefore continuous wavelet transform (CWT) was proposed as the method for primary features formation. The result of CWT of randomly taken excerpt of the signal is two-dimensional array containing bold areas of high value estimates representing the strong beats and some areas of moderate values representing MurMur sounds, in case they are present. Strong beat representations in these arrays give the time marks for the eventual representations of sought MurMur sounds. Therefore, we did not do the cardiocycle delineation, but use the whole signal calculating CWT results of sliding-overlapping windows of at least one cardiocycle length instead. For final analysis we use only few selected CWT-results per recording, having lowest, non-zero entropy. In this way we get rid of noisy or corrupted signal parts. The convolutional neural network does the final classification.

We used the same convolutional neural network and CWT features to classify patient's clinical outcomes. Outcomes indicate the normal or abnormal clinical outcome diagnosed by a medical expert.

The proposed algorithm was tested on the George B. Moody PhysioNet Challenge 2022 hidden validation set. Our team's, LSMU, MurMur detection classifier received a weighted accuracy score of 0.599 and Challenge cost score of 13075.

1. Introduction

Cardiac auscultation is the simplest and most cost-effective method of screening for a large number of heart disorders, including arrhythmia and valve disease. The method could be effectively used even for late postoperative diagnostics after valve replacements [1]. However, heart sounds are difficult to identify and analyse because significant events are closely spaced or even overlapped in time, and their frequency content is at the lower end of the audible frequency range [2]. Experienced cardiologists classify heart auscultation signals with great agreement between each other, at the same time facing difficulties to precise verbally the key features used. Therefore on average, only 20% of medical interns can effectively detect heart conditions using auscultation [3]. Machine learning algorithms trained on experts annotated data could be a valuable tool for clinical decision support increasing reliability of cardiac diagnostics. The George B. Moody PhysioNet Challenge 2022 [4] invites participants to identify murmurs and clinical outcomes using heart sound recordings collected from multiple auscultation locations. We propose here a convolutional neural network approach for heart MurMur detection and clinical outcome prediction using wavelet transform based features extracted from auscultation signals.

2. Methods

2.1. Data preparation

The time-frequency estimates of 1 second long (4000 samples) consequent partially overlapping excerpts of heart auscultation signals form George Moody PhysioNet

Challenge 2022 dataset [5] were obtained by means of continuous wavelet transform (CWT) (MatLab function "*cwt*") using Morse wavelets [6]. The estimate of ordinary excerpt was consisting of 91 x 4000 array, where 91 rows were representing estimates at particular central frequencies ranging from 3.3 Hz till 1.840 Hz in logarithmic scale. The frequency range was covering all expected frequency components, which could be diagnostically important for detecting cardiac valve closure sounds and sought MurMur sounds. The length of each particular central frequency estimate initially was 4000, but we observed that the precision of time representation of the highest frequency components could be much less then initial 1/4000 of sec. So, we reduced its length till 91 making final estimate array representing one signal excerpt of 91 x 91. It saved a lot of computation resources. The representation of CWT estimate array as grey scale image was very useful for preliminary visual evaluation of the diagnostic usefulness of the features. Therefore, we used some image analysis algorithms for features preparation and evaluation.

All recordings of heart auscultation signals unfortunately contained corrupted or noisy episodes. Visually, CWT estimates of corrupted or noisy signal excerpts were much more motley than ones from clear and not corrupted signal parts. Image entropy (MatLab function "*entropy*") was a reliable estimate to identify corrupted or noisy excerpts. So, for further analysis we used not all CWT estimates of recording excerpts, but only CWT estimates with the lowest entropy. Since PhysioNet Challenge 2022 dataset contained unequal proportions of 942 patients with MurMur related labels (absent – 695, present – 179 and unknown – 68) and the goal was to have similar amounts of training data we took 5 CWT estimates from auscultation recordings with the lowest entropy from patients with MurMur label absent, 20 CWT estimates with the lowest entropy from patients with MurMur label present and 80 CWT estimates with the lowest entropy from patients with MurMur label unknown.

So finally, we prepared 30587 CWT (11955 – absent, 12118 – present and 6514 – unknown) estimates of signal excerpts labelling them according to status of MurMur, clinical outcome (11721 – normal and 18866 – abnormal) and patient identifier.

2.2. Data selection

Typical excerpts of heart auscultation signal containing and not containing the MurMur sounds are presented in Fig. 2 together with their CWT estimates. As one can notice the white areas representing valve closure caused beats in the signal are represented by pyramidic shape bold areas. In case of presence of MurMur sounds, the peaks of them are surrounded by sparsely distributed spots. The example of representation of noisy part of the signal is in the left-hand side of part B, at the beginning of the signal excerpt. As one can see, the shape of bold area in this case is different when compared to the regular beats.

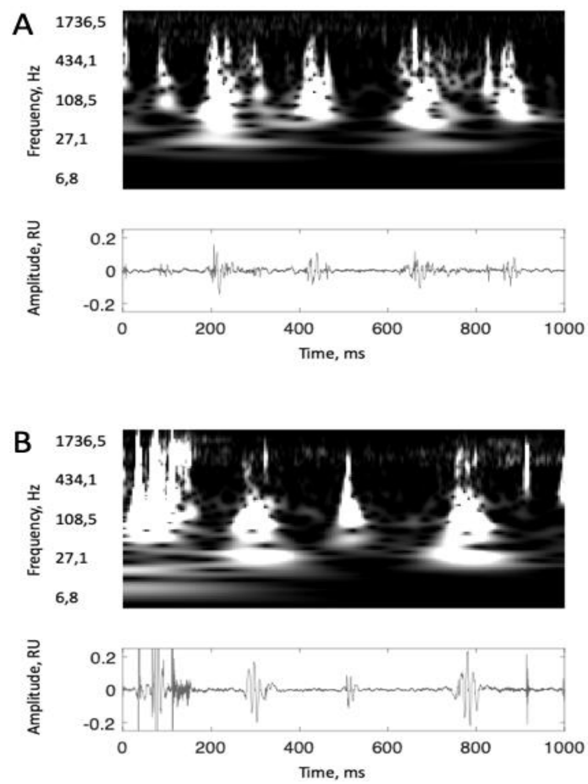

Fig.2 Typical example of CWT estimates of signal excerpts containing MurMur sounds (A) and not (B). The image of CWT estimate is stretched to fit the length of the signal below to align with the time scale of the signal.

As for clinical outcome prediction the same visual CWT estimate appearance that was seen in MurMur sounds presented in Fig.2 was noticed. Normal clinical outcome's auscultation signals CWT should look like part A in Fig. 2, as for abnormal clinical outcome – part B in Fig. 2.

After tests with the Challenge 2022 dataset and official phase entries scores, we trained our straightforward architecture convolutional neural network for detection of MurMur sounds only with the CWT estimates with MurMur labels of absent and present. By doing this our MurMur detection classifier was ignoring cases which were labelled unknown, and we entered 0 as probability of a murmur unknown status.

Clinical outcome classifier, the same architecture convolutional neural network that was used for MurMur sound detection, was trained with CWT estimates which had MurMur label absent and Outcome label normal (7370

CWT estimates), and MurMur label present and Outcome label abnormal (10159 CWT estimates).

All the selected CWT estimates were randomly divided into training and testing subgroups according following percentage: 80%, 20%.

2.3. Architecture of Convolutional Neural Network

Straight forward architecture convolutional neural network with 15 layers was used in proposed algorithm for MurMur detection and clinical outcome prediction.

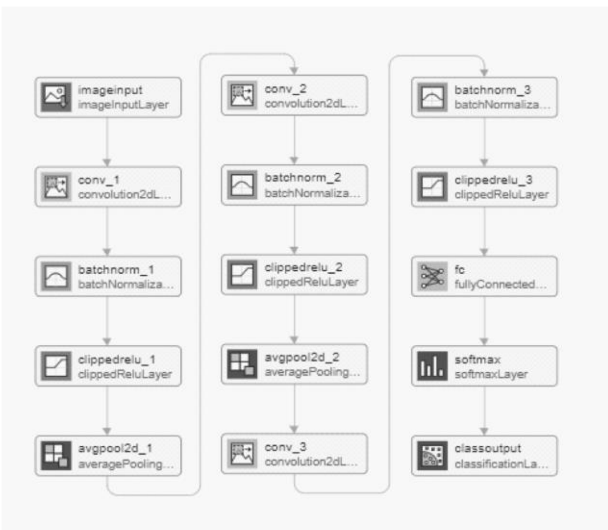

Fig.1 Architecture of convolutional neural network.

Proposed neural network has three 2-D convolutional layers which applies sliding convolutional filters (filter size [3 3]) to the input. The number of neurons in the convolutional layers are increasing by double from 8 to 32. There are also two average pooling layers which are performing down sampling by dividing the input into rectangular pooling regions (region size – [5 5] and [2 2]) and computing the average values of each region. Step size for traversing input 2.

Training was done using default MatLab options, except these parameters - stochastic gradient descent with momentum was chosen, epochs for training was set to 15 and data shuffling was set to every-epoch.

3. Results

Feature extraction and training process of the network was lasting about 100 min. and the model run time was about 10 min with computational resources given by the challenge organizers. Each entry had access to 8 virtual CPUs, 52GB RAM, 50GB local storage, and an optional NVIDIA T4 Tensor Core GPU (driver version 470.82.01) with 16GB VRAM.

Final accuracy reached by training, validation, and testing process of Challenge 2022 dataset of selected data for MurMur and Outcome prediction are presented in tables 1 and 2.

Training	Validation	Test	Ranking
0.777*±0.03	0.599		/81

Table 1. Weighted accuracy metric scores (official Challenge score) for our final selected entry (team LSMU) for the murmur detection task, including the ranking of our team on the hidden test set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set. *Accuracy for predicting absent or present labels in the training set.

Training	Validation	Test	Ranking
12981±346	13075		/81

Table 2. Cost metric scores (official Challenge score) for our final selected entry (team LSMU) for the clinical outcome identification task, including the ranking of our team on the hidden test set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

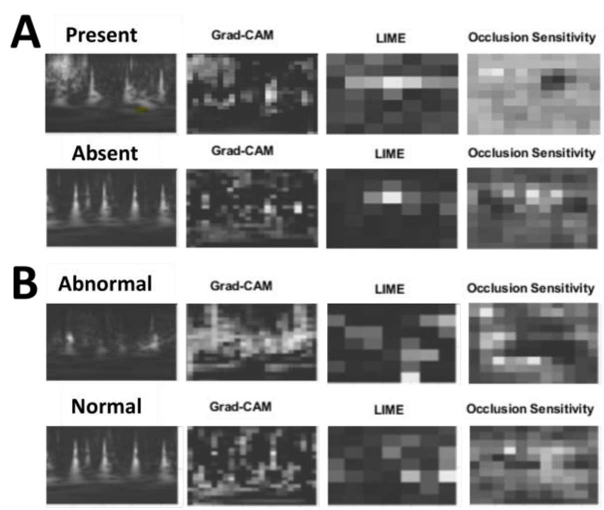

Fig.3 Grad-CAM, LIME and Occlusion sensitivity characteristic maps for MurMur (A) and Outcome (B) prediction.

To investigate what regions of the image are important to the score for the specified class three characteristic maps were produced - Grad-CAM, LIME and Occlusion sensitivity. Grad-CAM uses the gradient of the classification score with respect to the convolutional features determined by the network to understand which parts of the image are most important for classification [7]. The places where there are bright areas in the map gradient is large and final score depends most on this data. LIME approximates the classification behavior of a deep learning network using a simpler, more interpretable model, such as a linear model or a regression tree [7]. The simple model determines the importance of features of the input data as a proxy for the importance of the features to the deep learning network [7]. Occlusion sensitivity — perturbs small areas of the input by replacing them with an occluding mask, typically a gray square [7]. As the mask moves across the image, the change in probability score for analyzed class values, the brighter the area the bigger the change, is measured. The results of primary feature regions' importance are presented in Fig. 3.

4. Discussion

Majority of currently published algorithms of cardiac auscultation signal analysis (e.g. [5,8]) start from signal segmentation, which takes a substantial part of computational resources. Preprocessing part of our algorithm transforms analyzed signal into 2-dimensional array of primary features and image analysis methods are used for further processing and final decision making. As long the sought MurMur sounds are time-linked to the main acoustic events in the signal, their reflection 2-dimensional array of primary features will appear in the arbitrary position, yet always in the same position regarding the reflection of respected major acoustic events. Therefore, our idea was that neural network will learn to detect major acoustic event with or without the sought MurMur sound wherever it appears in time. So, there is no need for primary signal segmentation. The idea was confirmed by the analysis of regions' importance in the primary features.

All versions of our algorithm which used segmentation time marks provided in official training set did not outperform the current version.

We found the straightforward architecture of the neural network as the optimal one regarding comparatively short training time and the best performance. More complex architecture showed similar results, but greater training time.

5. Conclusion

Wavelet transform and image analysis technic based algorithm shows comparatively high accuracy in recognition of normal and abnormal heart sounds using comparatively small computational resources and needs no primary signal segmentation.

References

[1] Jonkaitiene R, Benetis R, Eidukaityte R. [Management of patients with prosthetic heart valves]. Medicina (Kaunas). 2005;41(7):553-60.

[2] A. Pease, "If the heart could speak," Pictures of the Future, pp. 60–61, 2001.

[3] S. Mangione, "Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries," The American journal of medicine, vol. 110, no. 3, pp. 210–216, 2001.

[4] Reyna, M. A., Kiarashi, Y., Elola, A., Oliveira, J., Renna, F., Gu, A., Perez-Alday, E. A., Sadr, N., Sharma, A., Mattos, S., Coimbra, M. T., Sameni, R., Rad, A. B., Clifford, G. D. (2022). Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv, doi: 10.1101/2022.08.11.22278688.

[5] Oliveira, J., Renna, F., Costa, P. D., Nogueira, M., Oliveira, C., Ferreira, C., … & Coimbra, M. T. (2022). The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification. IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2021.3137048.

[6] Lilly, J. M., and S. C. Olhede. "Generalized Morse Wavelets as a Superfamily of Analytic Wavelets." IEEE Transactions on Signal Processing 60, no. 11 (November 2012): 6036–6041. https://doi.org/10.1109/TSP.2012.2210890.

[7] Investigate Audio Classifications Using Deep Learning Interpretability Techniques. Available from:https://de.mathworks.com/help/deeplearning/ug/investigate-audio classifications-using-interpretability-techniques.html [Accessed 19 July 2022].

[8] Springer D.B., Tarassenko L., Clifford G.D. Logistic Regression-HSMM-based Heart Sound Segmentation. IEEE Transactions on Biomedical Engineering 2016, Vol. 63, p. 822 – 832.

Address for correspondence:

Robertas Petrolis
Department of Physics, Mathematics and Biophysics, Lithuanian University of Health Sciences.
Eiveniu str. 4, Kaunas 50103, Lithuania
Robertas.petrolis@lsmuni.lt