



Automatic Classification of Heart Sounds Utilizing Hybrid Model of Convolutional Neural Networks

Methaq A. Shyaa^{1(✉)}, Ayat S. Hasan², Hassan M. Ibrahim³,
and Weam Saadi Hamza³

¹ Department of Information and Communication Systems,
Iraqi Ministry of Interior, Baghdad, Iraq

² Department of Computer Science, College of Education for Pure Science,
University of Diyala, Diyala, Iraq
ayat.saleh.hassan@uodiyala.edu.iq

³ University of Information Technology and Communication, Baghdad, Iraq
hassan.m@uoitc.edu.iq

Abstract. This paper introduces the employment of a deep learning model to develop an automated classification of heart sound into normal and abnormal. The Cardiology Challenge database 2016 (PhysioNet) has been employed in this work. This dataset is established with the assumption that the phonocardiograms (PCG) are extracted concurrently with the electrocardiograms (ECG). For each heartbeat, the PCG record on a time-domain basis is segmented and is converted to images. Then, these images feed the proposed convolutional neural network (CNN) model to perform the classification. Most previous methods used traditional machine learning techniques to classify the heartbeats. These methods depend on feature extraction and selection methods that are sensitive to different colors, shapes, and sizes. CNN has the ability to achieve feature extraction and classification automatically, unlike traditional machine learning methods which require manual feature extraction, and the performance relies on the chosen features. The proposed is designed by integrating multiple ideas including parallel convolutional layers and residual links. To prove that the proposed model is beneficial in terms of feature extraction and classification, the extracted features have been used to train the SVM classifier. The results show that the model can offer acceptable and comparable accuracy of classification. It has outperformed the previous methods by achieving an accuracy of 93.49% and 94.66% with the SVM classifier. High performance could help in saving many people's lives.

Keywords: Heart sound classification · Deep learning · CNN · PCG

1 Introduction

Most of the excellent achievements in the task of image classification are due to the recent progress in Artificial Intelligence. In the real world, the utilization of deep learning techniques like Recursive Neural Network (RNN), Long-Short Term Memory

(LSTM), as well as the Convolutional Neural Network (CNN), generates impressive applications in the fields of biomedical and Bioinformatics [1, 2]. Applications such as breast cancer classification [3] and skin cancer classification [4] are reported in the past three years, as an outcome of CNN algorithm developments, which perform well than the traditional techniques. Heart Sounds have been classified as either normal or abnormal by employing image processing and traditional machine learning methods [5, 6]. These methods depend on feature extraction and selection methods that are sensitive to different colors, shapes, and sizes. Thus, they achieved low performance in the task of heartbeats. On the other hand, deep learning models including CNN have solved the issues of image processing and traditional machine learning methods. Deep learning has shown tremendous performance in a variety of applications [2–4, 7, 8]. One of the best advantages of employing deep learning is the automatic feature extraction and classification unlike machine learning methods [9].

Therefore, this paper introduces a deep learning model for early heart disease detection. Certainly, early, and accurate classification of heartbeats will highly save the lives of the patients. Also, automated monitoring and recording of the ECG and PCG at home, rather than a hospital or clinic, performs as a practical indicative tool for diagnosing any abnormal sign of the heart. Consequently, anybody, who observes any abnormality in his/her ECG or PCG through daily life, could visit a doctor for additional tests. The CNN model proposed has a better feature representation since it uses parallel convolutional neural networks along with the residual connection. We have considered the advantage of previous CNN models. The proposed model was inspired by the design of the model in [3] which has shown impressive performance in different tasks such as diabetic foot ulcer classification [10, 11]. Lastly, to speed up the training process of CNN models, GPU and FPGA have been used to reduce the training time [3, 9, 12]. In this paper, we have utilized GPU to implement the heart sounds classification task. With the recent development in computational tools including a chip for neural networks and a mobile GPU, this work can be extended in the future to be a mobile application. The contributions of this paper can be sum up in two major points (i) novel CNN architecture (ii) Applying a support vector machine classifier (iii) achieving an accuracy of 93.49% and 94.66% with the SVM classifier.

2 Related Work

There were very few techniques or tools based on deep learning available for automatic heart sound diagnosis [13]. The PhysioNet Cardiology Challenge 2016 (abbreviated to “PhysioNet”) is the first approach for applying deep learning in biomedical fields (up to the authors’ knowledge). However, earlier approaches are based on traditional classifiers of supervised machine learning with pre-extracting feature algorithms. The extracted features from the heart periods are input into support vector machines [14] and Artificial Neural Network (ANN) [15], which are complexity-based features,

wavelet features, and time-frequency features. Moreover, earlier works utilized Hidden Markov models in PCG signal segmentation [16] and classification [17].

However, evaluating the success of the earlier works is extremely difficult, due to the variations in the length of the recorded signals, testing algorithms, the available number of recordings for training, dataset quality, as well as the collected data environments. Furthermore, some works did not execute suitable train-test data sets and recorded the results on validating or training data, which is very expected for producing positive results because of overfitting [13]. In addition, to overcome the overfitting problem, similar subject recordings are not involved in both training and validating. A collection of noisy and clean PCG records, which showed extra weak signal quality, is involved in enhancing the development of robust and accurate algorithms.

This paper introduces a deep learning technique, as one of the earliest challenges for heart sound classification. There were several works for applying deep learning techniques to other forms of physiological signal classification. Martinez, et al. [18] described their work, which used deep learning in the Psychophysiology domain. They support the utilization of partiality deep learning for identifying the influence of bodily inputs (e.g., blood volume pulse and skin conductor) inside a study of the game-based user. Besides, they go over the manual use of ad-hoc feature selection and extraction in emotional modeling because it reduces the feature design creativity to them. Furthermore, the reason why this works differently from their work is that they execute a primary unsupervised step of pre-training based on stacked convolutional auto-encoders. In contrast, this work eliminates the need for this step and is trained in an end-to-end supervised fashion instead. Similarly, some works process physical signals using deep learning in human activity diagnosis [19]. PhysioNet 2016 dataset has been used in this paper. Initially, five sub-folders, which are A, B, C, D, and E, are included in the training set of the dataset. Each folder comprises 3126 PCGs. Each PCG record has a time interval of 5 to 120 s [13]. There is the largest number of techniques that transform PCGs into images of exploiting spectrogram methods. For instance, Rubin et al. [20], utilized a logistic regression model based on hidden semi-Markov to segment the beginning of every beat. Next, these beginnings are converted to spectrograms exploiting MFCCs (Mel-Frequency Cepstral Coefficients) method. Later, spectrograms are identified as either normal or abnormal exploiting 2-layer CNN. This CNN has a modified loss function, which augments specificity and sensitivity, accompanied by a regularization parameter. Finally, the last signal classification is the mean probability of the total segment probabilities. This model gets an eighth place at the PhysioNet challenge, with a total score of 83.99%. Kucharski et al. [21], exploited a 5-layer CNN with dropout after transforming the segments by an eight-second spectrogram. This method attained 91.6% in specificity and 99.1% insensitivity. The result is equivalent to the most recent available methods. Another technique by Dominguez et al. [22] segmented each input signal and pre-processing it exploiting the neuromorphic auditory sensor for decomposing the acoustic data into frequency bands. Next, the

spectrograms are calculated and fed to a modified AlexNet version. This model attained a vital enhancement contrasted to the prizewinning model of PhysioNet. It achieved an accuracy of 94.16%. Furthermore, Potes et al. [23] exploited Adaboost and CNN. The spectrogram features were fed to Adaboost. Also, cardiac cycles, which were decomposed into four frequency bands, were used for training the CNN. The end outcome is obtained by combining the outputs of the Adaboost and CNN with an overall accuracy of 89%. This model gets the first in the formal aspect of PhysioNet.

On the other hand, models give the impression that has lower performance if there are no conversions from PCGs to spectrograms. For example, Ryu et al. [24] employed a Window-sinc Hamming filter for noise-reduction, signal scaling, and segmentation with a constant window. The used CNN consists of four layers and the achieved accuracy was 79.5%. Shortly later, Chen et al. [25] employed PCG for recognizing the segments S1 and S2. In their work, the PCG signals are converted into a series of MFCCs. Next, the K-means method is applied for clustering the features of the MFCC into two clusters aimed at distinguishing capability as well as refining their representation. Lastly, the S1 and S2 features are classified using a DBN. For obtaining the best results, the researchers compared their technique with SVM, logistic regression, Gaussian mixture models, and KNN.

Based on the literature, the most neural networks applied for diagnosing PCG tasks are the CNNs. In addition, like ECG, the spectrogram technique is used for converting PCG signals to images in several deep learning techniques [20, 22, 23] which is the same concept used in this paper. However, the proposed model is more developed than the previous models in aspects of feature extraction and classification.

3 Methodology

3.1 Main Components of CNN

Convolution The organization of the CNN components has an important responsibility in developing innovative structures for attaining improved performance. The following describes the role of CNN components in its structure.

A) Convolutional Layer

This layer consists of a group of convolutional kernels (actually, each kernel is a neuron). The receptive field is defined as a small area of the image, where each kernel is associated with it. Initially, the image is divided into several receptive fields (so-called blocks). Next, each block is multiplying by its corresponding weight (element of filter). This operation can be expressed as in Eq. 1:

$$F_l^k = (I_{x,y} * K_l^k) \quad (1)$$

Where: $I_{x,y}$ = image input, x,y = spatial locality, and $K_l^k = l^{th}$ convolutional kernel of the k^{th} layer.

The process of image division into receptive fields facilitates in determining locally the values of the correlated pixel. In addition, this locally accumulated information is so-called feature patterns. However, when moving the convolutional kernel over the image along with its weights, various groups of features are extracted. The convolution process can extra be classified into various types based on the path of convolution, the padding type, and the filter size and type.

B) Pooling Layer

As mentioned earlier, the outcome of the convolution process, which is called feature patterns, can happen at various image positions. Since a feature is extracted, its inexact location comparative to others is saved. In contrast, its exact position will not be as important. Downsampling or pooling is similar to convolution as motivating the local process. It accumulates similar information around the reception field and outcomes the main response inside this local area. The pooling process can be expressed as in Eq. 2.

$$Z_l = f_p \left(F_{x,y}^l \right) \quad (2)$$

Where: $Z_l = l^{th}$ output feature map, f_p = pooling process type, and $F_{x,y}^l = l^{th}$ input feature map.

The benefit of the pooling process facilitates the extraction of a feature set, which is unchanged to minor distortions and translational shifts. In addition, it may assist in enlarging the generalization by decreasing the overfitting. Moreover, the decrease in feature map size controls the network complexity. Various pooling formulation types like overlapping, L2, average, and max-pooling are employed to extract translational unchanged features.

C) Activation Function

This function assists in learning a complex pattern and works as a decision function. In general, selecting a proper activation function accelerates the learning operation. Equation 3 defines the function of the convolved feature map.

$$T_l^k = f_A \left(F_l^k \right) \quad (3)$$

Where: $T_l^k =$ transformed output for k^{th} layer, $f_A =$ activation layer, and $F_l^k =$ output of a convolution process.

From literature, several activation functions like ReLU, max-out, tanh, sigmoid, as well as, ReLU alternatives like PReLU, ELU, and leaky ReLU are utilized for training nonlinear feature sets. However, ReLU and its alternatives are chosen, since it facilitates solving the problem of vanishing gradient.

D) Batch Normalization

It is employed for addressing the problems associated with the internal covariance shift inside the feature maps. This internal shift can be defined as a variation in the distribution of the values of the hidden units, which imposing the rate of learning to a lower value (i.e., slowing down the convergence) and necessitates well-thought-out parameter initialization. Equation 4 represents the batch normalization of the transformed feature map.

$$N_l^k = \frac{T_l^k}{\sigma^2 + \sum_l T_l^k} \quad (4)$$

Where: N_l^k = normalized feature map, T_l^k = input feature map, and σ = variation in a feature map.

However, batch normalization brings together the value distribution of the feature map via carrying these values to zero mean and unit variance. In addition, it makes the gradient flow smoother and works as a regulating factor, which enhances the network generalization with no dependence on dropout.

E) Dropout

It presents regularization inside the network. It eventually enhances generalization via arbitrarily moving over some connections and units including a particular probability. Multi-connection occasionally becomes co-adapted and may produce overfitting if it learned non-linear relation. In contrast, various thinned network architectures are produced due to the randomly dropping of some units or connections. One of these architectures is selected with small weights. It is considered a representative network, which represents an approximation of the whole of the intended networks.

F) Fully Connected Layer

It is typically utilized for classification purposes as a last layer in the network. In general, it receives and analyzes the output of all preceding layers. Furthermore, it generates a non-linear grouping of the chosen features, which are employed for the data classification. Unlike convolution and pooling, it is a universal task.

3.2 The Proposed Model

The proposed model is designed to have better feature representation and classification of the heart sound. It starts by employing traditional convolution layers to reduce the size of input images. The first traditional convolutional has a filter size of 5×5 . Then it is followed by batch normalization and ReLU layers to speed up the training process and to avoid gradient vanishing issues. Both batch normalization and ReLU layers come after every single convolutional layer in the model. The second traditional convolutional has a filter size of 7×7 . After the traditional convolution layers, three blocks of parallel have been employed to extract good features. Each block consists of three branches. Each branch has a convolutional layer followed by batch normalization and ReLU layers. The first convolutional layer in the first branch of all three blocks has a filter size of 3×3 . The second convolutional layer in the second branch of all three blocks has a filter size of 5×5 . The third convolutional layer in the third branch of all three blocks has a filter size of 7×7 . The output of three branches is combined in the concatenation layer then it pushes to the next block. We have also used residual connections in blocks one and three for better feature representation. At the end of the model, an average pooling layer is added to reduce the dimensionality and decrease the effect of overfitting. Two fully connected layers have been added and between them, the dropout layer is to avoid overfitting issues. All heart sounds converted to signal images in the size of 512×512 . We have divided the dataset into 70% for training and 30% for testing. The model has been trained with 100 epochs using MatLab as software and GPU as hardware. Figure 1 shows an example of a learned filter by the first convolutional layer while Fig. 2 shows the design of the proposed model.

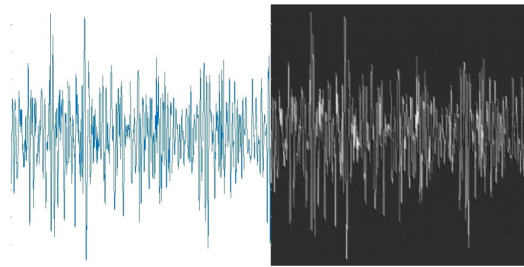


Fig. 1. Example of the learned filter from the first convolutional layer.

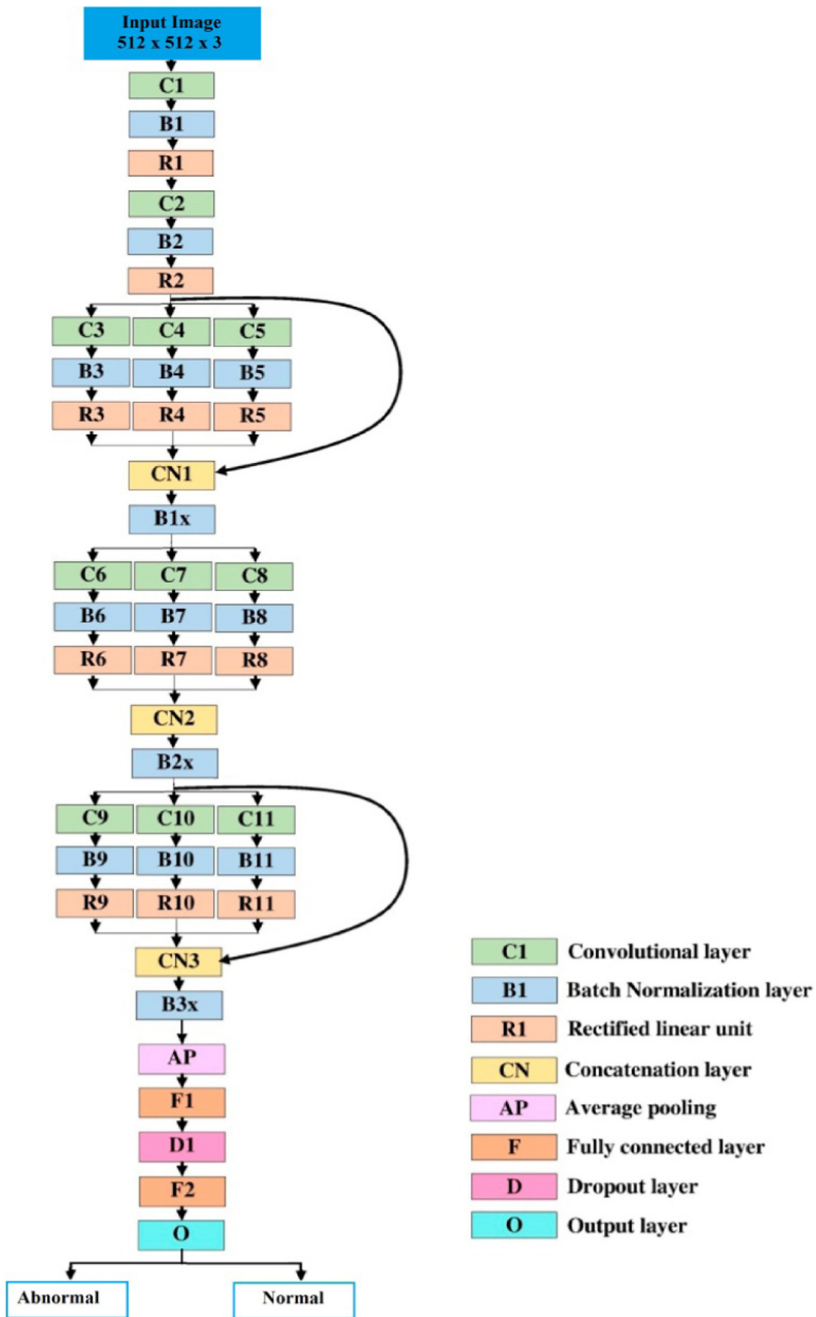


Fig. 2. The proposed model structure.

4 Experimental Results

We have assessed the proposed models in terms of sensitivity, specificity, precision, MAcc as described in [11]. We have worked on PhysioNet 2016 datasets as mentioned earlier. We have compared the proposed model with previous methods that used the same dataset as listed in Table 1. It outperformed the previous methods by achieving a sensitivity of 89.51%, specificity of 97.48%, and MAcc of 93.49%. To improve the results further and to prove the ability of the proposed model in terms of feature extraction, we have used the extracted features by the proposed model to train the SVM classifier. the proposed model with SVM has achieved 91.44%, 97.89%, 94.66% for sensitivity, specificity, MAcc, respectively. The proposed model with the SVM classifier has improved the results due to the good features extracted by the proposed model.

Table 1. Performance comparison between methods using the same dataset.

Methods	Sensitivity (%)	Specificity (%)	MAcc (%)
Potes et al. [23]	94.24	77.81	86.02
Zabihi et al. [26]	86.91	84.90	85.90
Plesinger et al. [27]	76.96	91.25	84.11
Rubin et al. [20]	72.78	95.21	83.99
Langley et al. [28]	94%	65%	80%
Krishnan PT et al. [29]	87%	85%	86%
The proposed model	89.51	97.48	93.49
The proposed + SVM	91.44	97.89	94.66

We have tested some heart sound with the proposed model and effectively classified them correctly.

5 Conclusion

This paper presented a hybrid CNN model that combining multiple ideas including parallel convolutional layers and residual links for the task of automated classification of heart sound into normal and abnormal. The parallel convolutional layers have different filter sizes to obtain better feature representation. PhysioNet 2016 has been used in this work which is a very challenging dataset. Each heartbeat sound was converted to images then these images were utilized to train the proposed model. The proposed model has shown excellent results by achieving an accuracy of 93.49%. Furthermore, we have utilized the features that have been extracted by the proposed model to train the SVM classifier. The proposed model with SVM has achieved an accuracy of 94.66% which outperformed the previous methods. The proposed model has proved that it is effective in terms of feature extraction and classification. The plan is to use it to classify ECG heartbeats sound with the employment of transfer learning plus build it as a mobile application. as shown in Fig. 3.

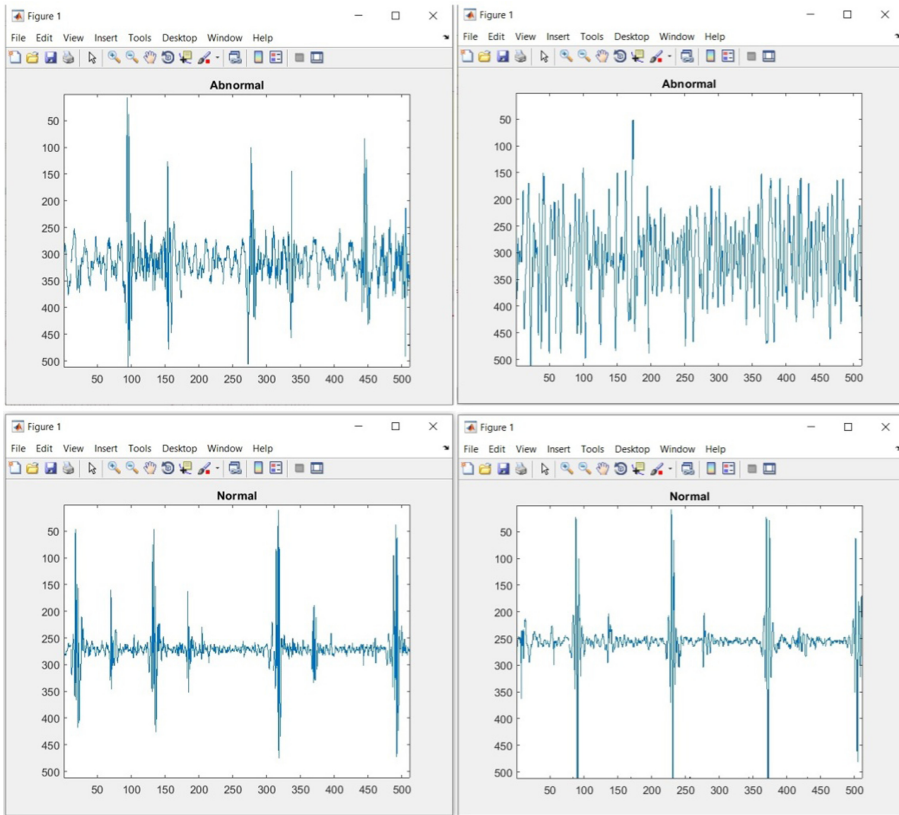


Fig. 3. Heart sound is classified by the proposed model. The first row is abnormal cases, the second row is normal cases.

References

1. Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., Gao, X.: Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods* **166**, 4–21 (2019)
2. Alzubaidi, L., Fadhel, M.A., Al-Shamma, O., Zhang, J., Duan, Y.: Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. *Electronics* **9**(3), 427 (2020)
3. Alzubaidi, L., Al-Shamma, O., Fadhel, M.A., Farhan, L., Zhang, J., Duan, Y.: Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics* **9**(3), 445 (2020)
4. Pacheco, A.G.C., Krohling, R.: An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE J. Biomed. Health Inform.* **25**, 3554–3563 (2021)
5. Fadhel, M.A., Al-Shamma, O., Olewi, S.R., Taher, B.H., Alzubaidi, L.: Real-time PCG diagnosis using FPGA. In: Abraham, A., Cherukuri, A.K., Melin, P., Gandhi, N. (eds.) *ISDA 2018* 2018. AISC, vol. 940, pp. 518–529. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-16657-1_48

6. Alshamma, O., Awad, F.H., Alzubaidi, L., Fadhel, M.A., Arkah, Z.M., Farhan, L.: Employment of multi-classifier and multi-domain features for PCG recognition. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp. 321–325. IEEE, October 2019
7. Hasan, R.I., Yusuf, S.M., Alzubaidi, L.: Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion. *Plants* **9**(10), 1302 (2020)
8. Alzubaidi, L., Fadhel, M.A., Al-Shamma, O., Zhang, J., Santamaría, J., Duan, Y.: Robust application of new deep learning tools: an experimental study in medical imaging. *Multimed. Tools Appl.* 1–29 (2021)
9. Alzubaidi, L., et al.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**(1), 1–74 (2021)
10. Alzubaidi, L., et al.: Towards a better understanding of transfer learning for medical imaging: a case study. *Appl. Sci.* **10**(13), 4523 (2020)
11. Alzubaidi, L., et al.: Novel transfer learning approach for medical imaging with limited labeled data. *Cancers* **13**(7), 1590 (2021)
12. Al-Shamma, O., Fadhel, M.A., Hameed, R.A., Alzubaidi, L., Zhang, J.: Boosting convolutional neural networks performance based on FPGA accelerator. In: Abraham, A., Cherukuri, A.K., Melin, P., Gandhi, N. (eds.) ISDA 2018 2018. AISC, vol. 940, pp. 509–517. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-16657-1_47
13. Liu, C., et al.: An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* **37**(12), 2181 (2016)
14. Ari, S., Hembram, K., Saha, G.: Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier. *Expert Syst. Appl.* **37**(12), 8019–8026 (2010)
15. Uğuz, H.: Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy. *Neural Comput. Appl.* **21**(7), 1617–1628 (2012)
16. Springer, D.B., Tarassenko, L., Clifford, G.D.: Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* **63**(4), 822–832 (2015)
17. SaraçOğLu, R.: Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction. *Eng. Appl. Artif. Intell.* **25**(7), 1523–1528 (2012)
18. Martinez, H.P., Bengio, Y., Yannakakis, G.N.: Learning deep physiological models of affect. *IEEE Comput. Intell. Mag.* **8**(2), 20–33 (2013)
19. Inoue, M., Inoue, S., Nishida, T.: Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif. Life Robot.* **23**(2), 173–185 (2017). <https://doi.org/10.1007/s10015-017-0422-x>
20. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., Sricharan, K.: Recognizing abnormal heart sounds using deep learning. arXiv preprint [arXiv:1707.04642](https://arxiv.org/abs/1707.04642) (2017)
21. Kucharski, D., Grochala, D., Kajor, M., Kańtoch, E.: A deep learning approach for valve defect recognition in heart acoustic signal. In: Borzowski, L., Świątek, J., Wilimowska, Z. (eds.) ISAT 2017. AISC, vol. 655, pp. 3–14. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67220-5_1
22. Dominguez-Morales, J.P., Jimenez-Fernandez, A.F., Dominguez-Morales, M.J., Jimenez-Moreno, G.: Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. *IEEE Trans. Biomed. Circuits Syst.* **12**(1), 24–34 (2017)
23. Potes, C., Parvaneh, S., Rahman, A., Conroy, B.: Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In: 2016 Computing in Cardiology Conference (CinC), pp. 621–624. IEEE, September 2016
24. Ryu, H., Park, J., Shin, H.: Classification of heart sound recordings using convolution neural network. In: 2016 Computing in Cardiology Conference (CinC), pp. 1153–1156. IEEE, September 2016

25. Chen, T.E., et al.: S1 and S2 heart sound recognition using deep neural networks. *IEEE Trans. Biomed. Eng.* **64**(2), 372–380 (2016)
26. Zabihi, M., Rad, A.B., Kiranyaz, S., Gabbouj, M., Katsaggelos, A.K.: Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In: 2016 Computing in Cardiology Conference (CinC), pp. 613–616. IEEE, September 2016
27. Plesinger, F., Jurco, J., Jurak, P., Halamek, J.: Discrimination of normal and abnormal heart sounds using probability assessment. In: 2016 Computing in Cardiology Conference (CinC), pp. 801–804. IEEE, September 2016
28. Langley, P., Murray, A.: Heart sound classification from unsegmented phonocardiograms. *Physiol. Meas.* **38**(8), 1658 (2017)
29. Krishnan, P.T., Balasubramanian, P., Umapathy, S.: Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Phys. Eng. Sci. Med.* **43**(2), 505–515 (2020). <https://doi.org/10.1007/s13246-020-00851-w>