EfficientNet Optimization on Heartbeats Sound Classification

Husni Fadhilah Dhiya Ul Haq
Dept of Computer Science/Informatics
Universitas Diponegoro
Semarang, Indonesia
husnifd@students.undip.ac.id

Rifky Ismail
Dept of Mechanical Engineering/
Center for Bio Mechanics-Material-Mechatronics-Signal Processing
Universitas Diponegoro
Semarang, Indonesia
ismail.rifky@gmail.com

Suhartini Ismail
Dept of Nursing/
Center for Bio Mechanics-Material-Mechatronics-Signal Processing
Universitas Diponegoro
Semarang, Indonesia
suhartini.ismail@fk.undip.ac.id

Satriawan Rasyid Purnama
Dept of Computer Science/Informatics
Universitas Diponegoro
Semarang, Indonesia
srp21.if@gmail.com

Budi Warsito
Dept of Statistics/
School of Postgraduate Studies
Universitas Diponegoro
Semarang, Indonesia
budiwarsito@lecturer.undip.ac.id

Joga Dharma Setiawan
Dept of Mechanical Engineering/
Center for Bio Mechanics-Material-Mechatronics-Signal Processing
Universitas Diponegoro
Semarang, Indonesia
joga.setiawan@ft.undip.ac.id

Adi Wibowo
Dept of Computer Science/Informatics
Center for Bio Mechanics-Material-Mechatronics-Signal Processing
Universitas Diponegoro
Semarang, Indonesia
bowo.adi@undip.ac.id

2021 5th International Conference on Informatics and Computational Sciences (ICICoS) | 978-1-6654-3807-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICOS53627.2021.9651818

***Abstract*— Currently, most deaths are caused by heart disease. Heartbeat sound analysis is a straightforward way to diagnose heart disease and potential for early detection of heart diseases. However, echocardiographic classification from heartbeat sound has always been a challenge in feature extraction. In this paper, EfficientNet is optimized for heartbeat sound classification. We optimized four EfficientNet architectures and dense layers number and dropout settings. Heartbeat sounds were converted in spectrogram format as input. Data set B was applied in this study consisted of three categories: normal, murmur, and extrasystole heart sound. The best architecture was EfficientNet B0 with four dense layers and dropout in the test dataset with 82% accuracy.**

Keywords— heartbeat sound; classification; EfficientNet

I. INTRODUCTION

According to the World Health Organization (WHO), heart-associated cardiovascular diseases (CVD) are the main reason of demise withinside the world. Heart disease can be identified by describing the sound data of the heartbeat. Before the 19th century, before the stethoscope was invented, doctors listened to the sound of a patient's heartbeat directly to try to diagnose disease. However, this method is considered unethical and scientific for doctors to do. This continued until the invention of the stethoscope in 1816. The stethoscope is now broadly used withinside the clinical area to pay attention to the heartbeat [1]. Heart rate alone can be used to diagnose disease. However, not all diseases can be diagnosed through a stethoscope because doctors need sufficient experience to produce accuracy [2].

Cardiovascular auscultation is the primary diagnostic method used to assess and analyze cardiac surgery and function. The main source of the generation of heart sounds is due to an unstable blood moment called blood turbulence. The process of auscultation of heart abnormalities is carried out using an electronic stethoscope which produces a digital recording of heart sounds called PCG [3].

Heart sound signals can vary concerning different types of heart disease. It may be found that there is a large difference in pattern between normal heart sound signals and abnormal heart sound signals as PCG signals vary with each other concerning time, amplitude, intensity, homogeneity, spectral content, etc.

The normal heartbeat sound in humans that has been observed so far consists of two sound components, namely the S1 sound component (lub) and the S2 sound component (dub). This component is related to the closing and opening of the valve in the diastole [4]. This sound has a “lub dub, dub lub” pattern with a rate of 60-100 beats per minute each. The sound of the heartbeat takes the form of hissing, roaring, rumbling between pattern “lub” to “dub” or vice versa. This sound indicates the symptoms of various heart diseases. The extra systolic heartbeat sound has a “lub-lub dub, lub dub-dub” pattern and is common in adults and children [5].

Many researchers use exclusive strategies along with denoising, feature extraction, down-sampling, and classification [6]. These strategies are used to are expecting coronary heart disease. There is an interesting problem for machine learning

researchers to identify the type of human heartbeat. That's because the sound of a heartbeat in the real world often contains noise and the differences can be very subtle and hard to tell apart. This can be done by filtering noise with the help of a computer. Classification algorithms in machine learning or deep learning are applied to the extracted features to identify different heart sound signals that are indicative of different heart problems [6]. In deep learning, there are several architectures that can be used, one of which is EfficientNet has recently been used as the foremost deep architecture [7].

Elsa Ferreira Gomes et al. [5] describes the PASCAL challenge echocardiography classification. They used a set of rules to decide the S1 and S2 of the coronary heart sounds. Here, S1 is lub and S2 is dub. They applied the MATLAB decimation function to the original audio signals and applied a bandpass filter to denoise those signals. Then apply the average Shannon energy to help the researcher easily identify the peaks on the echocardiogram. Also the use of algorithms for finding the minimum and maximum points of the voice signal and segmenting the sound of the heartbeat. They trained an audio signal prediction model using J48 and MLP algorithms. These heart sound signals are Normal, Murmur, Extrasystole, and artifacts.

ShiWen Deng et al. [8] show a framework for classifying echocardiograms based on autocorrelation functions without the use of segmentation. The autocorrelation function is extracted from the subband coefficients of the cardiac signal using Discrete Wavelet Transform (DWT) after the autocorrelation has been merged to obtain the final features, with these final properties and Dataset-B [9]. Applies to Support Vector Machines (SVMs). It is not very good to get the accuracy of normal (77%), shuffle (76%), and extrasystole (50%).

Raza et al [10] also generates a deep learning model to classify heart rate sounds based on data frames, downsampling, and RNN for Dataset-B. Heart rate signal cleaned by noise filtering. Dataframing converts the sampled frame rate of each audio file to a fixed frame rate, downsampling to reduce the waveform size of the heartbeat audio signal and extracting more distinctive features. The model proposed by RNN was applied to Dataset-B in this study and achieved the highest accuracy of 80.8%.

Based on Banerjee et al. [11] research, they use A three-layered CNN to classify the heartbeat sound audio clips. In the pre-processing stage, they extracted features from audio clips using MFCC. The proposed model achieves an overall test accuracy of 83%.

TABLE I. SUMMARY OF PREVIOUS RESEARCH USING THE . DATASET B

Authors	Model	Results
[8]	Support Vector Machine	2.03 (precision)
[10]	RNN-LSTM	0.80 (accuracy)
[11]	2D-Convolutional Neural Network	0.83 (accuracy)
[12]	Multi Layer Perceptron	1.67 (precision)
[13]	Support Vector Machine	0.74 (accuracy)

Based on Raza et al [10] research, they want to use CNN and apply spectrogram to get more discriminative features The Convolutional Neural Network (CNN) is a part of deep learning and has impressive records in the application of image analysis and interpretation, including medical images. Compared to manual extraction techniques, CNN s can more accurately understand the low-level and high-level characteristics of the image classification process. Based on previous research, broadly speaking, the architecture of CNNs is, and the performance is created by using a neural network layer (width), deeper layers (depth), and a larger input image resolution.

Therefore, in this study, we want to try to apply the spectrogram classification technique of the converted heartbeat sound signal using CNN with EfficientNet architecture. Our contribution to this research is:

- Perform a heart rate classification of dataset-B [9] which is normal, murmur, and extrasystole.
- Use EfficientNet to classify spectrograms from heartbeat.
- Comparing the accuracy results of EfficientNet B0-B3 classification.
- Adding multiple dropouts on multiple dense in fully connected layers.

In section I, it contains the background of the problem, the problem formulation, and the purpose of writing the paper. Section II contains the methodology used in heartbeat spectrogram classification research using EfficientNet starting from data collection, pre-processing data, model training to model testing. In addition, discuss theories related to the topic of research problems. In section III, present the results of experimental analysis with various scenarios along with their discussion and comparison with the methods that have been done. Section IV contains the conclusions of the studies that have been conducted. Finally, section V contains input materials and plans for future research.

II. METHODOLOGY

This study aimed to classify the Mel spectrogram of various heart rate signals by optimizing the deep learning model with EfficientNet architecture. We will also compare the results with existing research. These results can provide indicating the type of heartbeat as an initial diagnosis to help determine further procedures in medicine. The dataset implemented on this look at is dataset-B [9], which incorporates 3 categories: normal, murmur, and extrasystole.

The main structure of this research is shown in Figure 1. It consists of three stages: preprocessing, feature extraction, and model classification. We used dataset B from the PASCAL Cardiac Acoustic Challenge classification [12]. Dataset-B was collected from a hospital clinical trial using a digital stethoscope (DigiScope. Then we convert the heartbeat sound signal to digital spectrum Mel spectrogram. After preprocessing the data, we divide the data into 80% for training data and 20% for test data. Then the training data is divided back into 80% for training data and 20% for data validation. We build the model with EfficientNet by optimizing the output layer and adding dense layers.

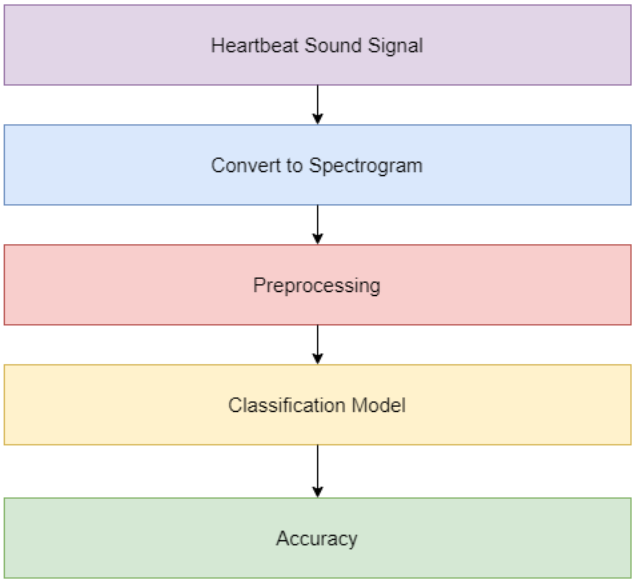


Fig. 1. The Steps of this study

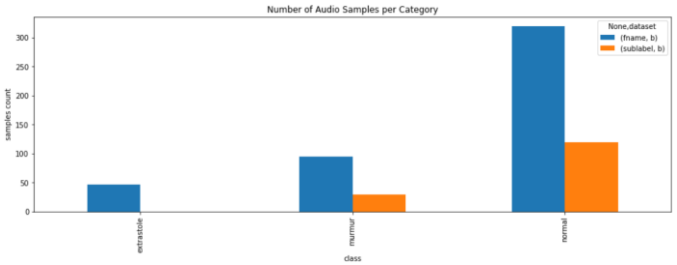


Fig. 2. Data representation of Heartbeat sound signal

A. *Data Understanding*

Dataset-B classifies heart sounds in the PASCAL Challenge. Data have been accumulated from a sanatorium scientific trial the use of a DigiScope virtual stethoscope. Dataset B includes 461 samples divided into 3 groups: Normal, Murmur, and Extrasystole [12].

TABLE II. NUMBER OF SAMPLE DATASET-B

Class	Number of Samples
Murmur	95
Normal	320
Extrasystole	46

A normal heartbeat is heard such as for example, lub dub or dub lub, and a heartbeat murmur has a sound between dub to lub or lub to dub and an extra systolic heart sound not beating for example, lub-lub dub or lub dub-dub [5].

Figure 3 (a) shows a normal sound wave with the correct shape of lub without noise. Noise sound wave indicates that the sound between lub-to-dub or dub-to-lub and extrasystole is in the format of different sounds (such as lub lub dub or lub dub dub) that appear in extrasystole sound waves. The spectrogram of extrasystole heartbeat sound has shaped as shown at Figure 4.

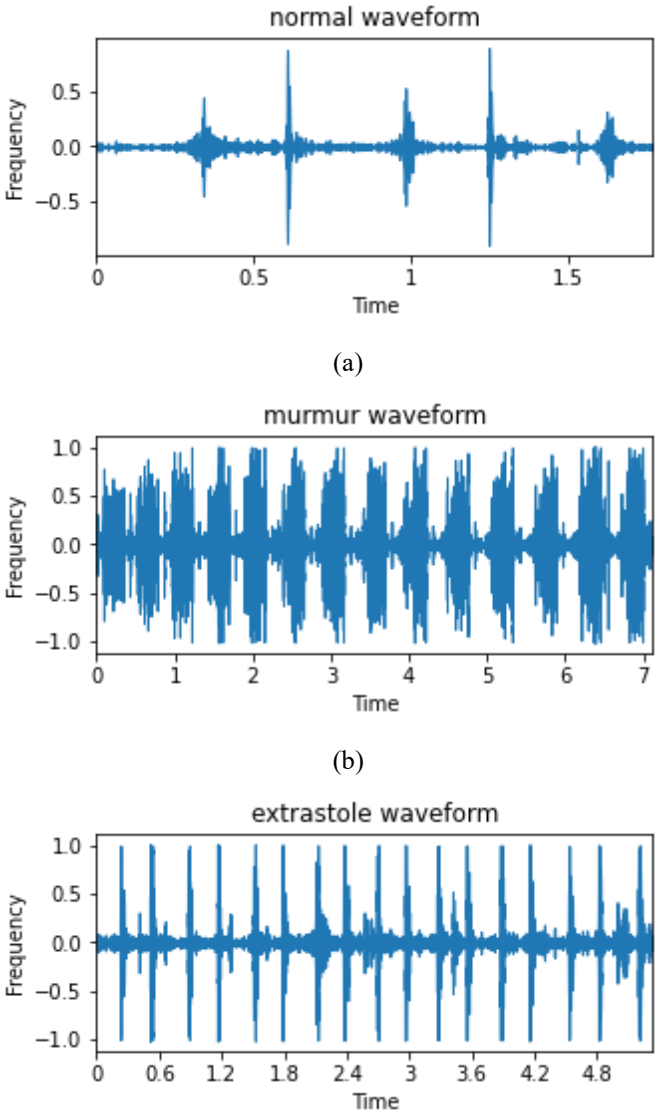


Fig. 3. Waveform representation of heartbeat sound signal. (a) Normal waveform, (b) murmur waveform, and (c) extrastole waveform

B. *Data Preparation*

Collected data is processed in several steps. The first step is to convert the data into a digital spectrum. Then, the Mel spectrogram was removed while maintaining the audio sample along the frequency axis, then filtering was performed using the preferred PerChannel energy normalization (PCEN) method. We convert audio into an image with an image size of 112 x 224 x 3.

C. *Proposed Method*

The author creates a model of deep learning to classify heartbeat sounds based on Mel spectrogram data. The deep learning model architecture is shown in Figure 5.

We adopted the EfficientNet B0 model [11] to perform feature extraction from spectrogram images. EfficientNet B1-B3 is also used as a comparison, but in this case, it is not more

optimal than EfficientNet B0. EfficientNet is a model that has a very complex network but is computationally efficient. As the number of models increases, the EfficientNet group contains eight models between B0 and B7. The number of calculation parameters has not increased significantly, but the accuracy has increased significantly. This technique is called compound scaling. The composite coefficient φ is used to uniformly scale the depth d, width w, and resolution r according to (1).

$$\begin{aligned} d &= \alpha^{\varphi} \\ w &= \beta^{\varphi} \\ r &= \gamma^{\varphi} \\ \alpha &\geq \beta \geq \gamma \end{aligned} \tag{1}$$

where, α, β, γ are the coefficients determined by the grid search assigned to the width, depth, and resolution of the network, respectively. φ is a coefficient defined by the user that controls available to extend the model. In a regular convolution process, FLOPS are proportional to d, w^2, r^2. Since the computational overhead of the convolutional network is specially because of the convolution operation, scaling convolution network as given in Eq. (1) will increase the FLOPS of the network via approximately $(\alpha, \beta^2, \gamma^2)\ \varphi$ in total [14].

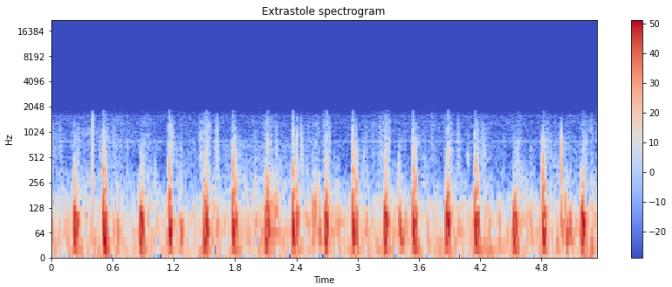


Fig. 4. Spectrogram of Heartbeat sound signal

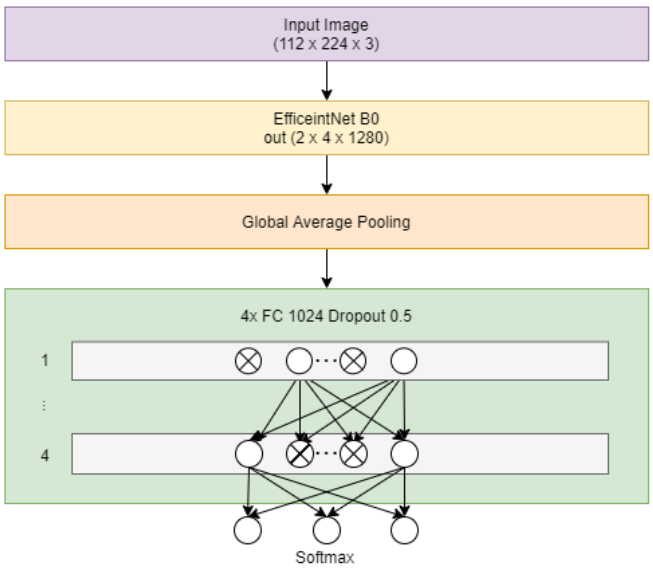


Fig. 5. Architecture of proposed model

Unlike different CNN models, EfficientNet makes use of a brand new trigger function known as Swish rather than the ReLU feature. With the transfer learning technique, EfficientNet has been previously trained using the ImageNet dataset. Its weights can be used for training in this task to obtain better and faster generalizations. After the final convolution, Global Average Pooling and several fully connected layers with dropout layers produce a generalized classification. The final classification layer has three output nodes using the softmax activation function to generate a multi-class model.

Some of the algorithms used by the author include:

1.) Global Average Pooling 2D Layer

Global average pooling 2D layer is an extreme of average pooling that can reduce dimensions of a tensor input with the size of w × w × d to 1 × 1 × d.

2.) Dropout Layer

Overfitting is a major problem in deep learning. This takes place whilst the classification algorithm trains the data to offer great results. Then follow the classification algorithm to the test data, where the data gives no match. This takes place whilst two or more neurons constantly stumble on the identical end result c and it's miles important to forestall the neurons from affecting the end result. Equation (2) suggests the dropout layer

$$z_i^L = w_i^L y^L + b_i^L \tag{2}$$

where L is a hidden layer wherein L ∈ l1, l2, l3, ..., ln, zL is the enter layers, yL is the vector output, wL is the weights and bL is the biases.

3.) *Softmax Activation Function*

Softmax is a completely crucial activation function in artificial neural networks and determines whether or not neurons are active. Softmax is a good manner to resolve the multi-elegance type problem, in which the output is expressed in phrases of classification [15]. The essential intention of Softmax is to emphasise the most cost of neurons. The most weight assigned to one neuron is 1 and the burden assigned to different neurons is zero. Softmax function is described as (3).

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} (i = 1,2, \ldots, N) \tag{3}$$

y and S are the input and output. The Softmax function is used in the last layer of the neural network to get the probability of the category class of each input.

D. Evaluation

There are several types of metrics to assess the performance of the classifier [16]. The evaluation criteria used in Dataset-B is accuracy. Accuracy can be calculated by dividing the evaluation by the total number of evaluations. To degree the accuracy, we want the variety of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) should be described as (4).

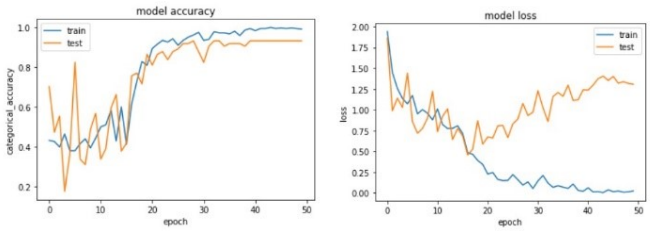
$$A = \frac{(TP + TN)}{TP + TN + FP + FN} \tag{4}$$

III. Result and Discussion

In this study, we use a classification algorithm i.e EfficientNet B0-B3 architecture with CNN algorithm applies to the Dataset-B. As a result, EfficientNet B0 has gained more accuracy displayed in the table. Table III shows the comparison between efficient net models B0 - B3. Table III shows that the EfficientNet model with the best accuracy is EfficientNet B0 with a dropout of 4x where the validation accuracy score is 93%. Meanwhile, the lowest score was obtained by the non-dropout EfficientNet B0 model.

TABLE III. Accuracy Results of EFFNET B0-B3

Model	Dense	Train	Val	Test
B0	1024	98	82	73
B0	1024 DO(0.5) 2x	98	83	75
B0	1024 DO(0.5) 4x	100	93	82
B0	1024 DO(0.5) 6x	99	88	78
B1	1024 DO(0.5) 4x	100	81	80
B2	1024 DO(0.5) 4x	94	80	78
B3	1024 DO(0.5) 4x	99	81	76



(a) (b)

Fig 6. (a) Val Accuracy & (b) Val Loss

The validation accuracy and training loss are shown in Figure 6 help determine if is model performing well. If the val accuracy increases and the val loss decreases, it means that the learning and operation are good. If the loss of validation decreases and the accuracy of validation increases, it means that our model is not in the learning state. If the loss of val and accuracy increases, it means that our model is overfitting. Figure 6 shows the increase in val accuracy and the decrease in validation which is means the model is running as it should.

The confusion matrix can be understood as a tool whose function is to analyze whether a classification model is good at recognizing tuples of values from different classes. From the results of the confusion matrix, the model correctly predicted 61 normal heartbeat spectrograms, 14 murmurs, and 1 extrasystoles. In addition, 3 normal data were wrongly predicted as murmurs, 4 data of murmurs that were wrongly predicted as normal, and 10 extrasystole data that were wrongly predicted as normal.

Comparison between the best results of the EfficientNet B0 model with some previous studies, one of which was Raza, it was found that with the right settings, the EfficientNet B0 model

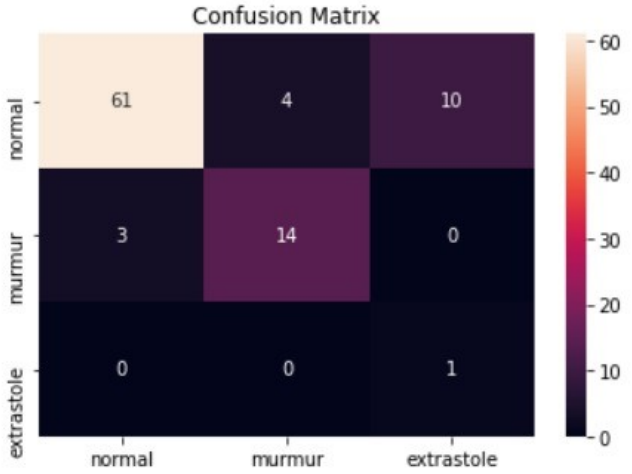


Fig. 7. Confusion Matrix Results

can outperform the results of sound classification with RNN & LSTM.

IV. Conclusion

This study aims to create a deep learning model of Mel spectrogram of heart rate sound classification based on EfficientNet, and CNN for Dataset-B. This intended method can efficiently classify Mel spectrogram of heart rate signals. Dataset-B falls into three categories: Normal, Murmur and Extrasystole. Audio filtering was performed using the preferred PerChannel energy normalization (PCEN) method. The proposed model of EfficientNet is applied for dataset-B in this study had the highest accuracy of 82. Experiments show a more competitive and efficient method. Research can be developed further by increasing the amount of data and the availability of datasets for heart disease should also be updated so that the model can be better to predict new data.

V. Future Work

The future work of this paper is to preprocess the dataset in other ways or other methods and perform other modeling or apply hyperparameter tuning techniques to improve accuracy.

Acknowledgment

This work was supported by Diponegoro University.

References

[1] I. R. Hanna and M. E. Silverman, "A history of cardiac auscultation and some of its contributors," *Am. J. Cardiol.*, vol. 90, no. 3, pp. 259–267, 2002, doi: 10.1016/S0002-9149(02)02465-7.

[2] Z. Jiang and S. Choi, "A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 286–298, 2006, doi: 10.1016/j.eswa.2005.09.025.

[3] S. K. Randhawa and M. Singh, "Classification of Heart Sound Signals Using Multi-modal Features," *Procedia Comput. Sci.*, vol. 58, pp. 165–171, 2015, doi: 10.1016/j.procs.2015.08.045.

[4] D. Kumar, P. Carvalho, M. Antunes, P. Gil, J. Henriques, and L. Eugénio, "A new algorithm for detection of S1 and S2 heart sounds," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2, pp. 1180–1183, 2006, doi: 10.1109/icassp.2006.1660559.

[5] E. F. Gomes and E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction," *Aistats*, no. 1, pp. 1–5, 2012.

[6] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, no. 6, pp. 1–18, 2021, doi: 10.3390/e23060667.

[7] A. Wibowo *et al.*, "Earthquake Early Warning System Using Ncheck and Hard-Shared Orthogonal Multitarget Regression on Deep Learning," *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, 2021, doi: 10.1109/LGRS.2021.3066346.

[8] S. W. Deng and J. Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Futur. Gener. Comput. Syst.*, vol. 60, pp. 13–21, 2016, doi: 10.1016/j.future.2016.01.010.

[9] Y. Zheng, X. Guo, and X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2710–2721, 2015, doi: 10.1016/j.eswa.2014.10.051.

[10] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat sound signal classification using deep learning," *Sensors (Switzerland)*, vol. 19, no. 21, pp. 1–15, 2019, doi: 10.3390/s19214819.

[11] M. Banerjee and S. Majhi, "Multi-class heart sounds classification using 2D-convolutional neural network," *Proc. 2020 Int. Conf. Comput. Commun. Secur. ICCCS 2020*, 2020, doi: 10.1109/ICCCS49678.2020.9277204.

[12] E. F. Gomes, P. J. Bentley, M. Coimbra, E. Pereira, and Y. Deng, "Classifying heart sounds: Approaches to the PASCAL challenge," *Heal. 2013 - Proc. Int. Conf. Heal. Informatics*, pp. 337–340, 2013, doi: 10.5220/0004234403370340.

[13] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," *Expert Syst. Appl.*, vol. 84, pp. 220–231, 2017, doi: 10.1016/j.eswa.2017.05.014.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.

[15] M. Yu, Q. Huang, H. Qin, C. Scheele, and C. Yang, "Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies," *Int. J. Digit. Earth*, vol. 12, no. 11, pp. 1230–1247, 2019, doi: 10.1080/17538947.2019.1574316.

[16] R. Manikandan, A. M. Barani, R. Latha, and R. Manikandan, "Implementation of Artificial Fish Swarm Optimization for Cardiovascular Heart Disease," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4S5, pp. 134–136, 2020, doi: 10.35940/ijrte.d1004.1284s519.