IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 4, 2019, accepted May 23, 2019, date of publication May 27, 2019, date of current version October 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919485

Early and Remote Detection of Possible Heartbeat Problems With Convolutional Neural Networks and Multipart Interactive Training

KRZYSZTOF WOŁK AND AGNIESZKA WOŁK
Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland
Corresponding author: Krzysztof Wołk (kwolk@pja.edu.pl)

ABSTRACT In this study, the convolutional neural network (CNN) and multipart interactive training were used to create a state-of-the-art classifier for the early detection of cardiac pathologies. The classification was performed using three sets of data samples; the first was recorded using a digital stethoscope and the others were recorded with mobile smart devices. These data were part of the competition on the Kaggle platform and used on the CNN in the form of audio samples as well as in spectrogram format, and experiments were conducted using both methods. Moreover, this method was used not only to distinguish healthy from unhealthy heart rhythm but also to attempt reaching some initial diagnosis. It was possible to identify a set of problems with the dataset, make corrections, and share them with the scientific community. The experiments were conducted using the so-called multi-part interactive training and an additional ResNet pre-trained network, resulting in more than 93% precision. This allows anyone to undertake prophylactical diagnosis using a smartphone alone. It also has a great educational potential for young doctors and students. Greater improvement is possible by supplying more data to the described method.

INDEX TERMS Convolutional neural network, early detection, heartbeat classification, interactive training, ResNet, telemedicine.

I. INTRODUCTION

In the past decades, heart attack has been a significant health issue worldwide. Records from the World Health Organization (WHO) show that the number of deaths due to heart disease was higher in the twenty-first century. For example, in Poland, the Central Statistical Office (GUS) reports show that approximately 46% of all deaths are caused by cardiovascular disease [1]. Early diagnosis of cardiovascular disease is essential for timely implementation of preventive measures. This is especially important in the third-world countries where access to medical care and medical devices is limited or in any other place where telemedical help comes in handy.

Any measures that can help contain the negative effects of cardiovascular disease should be seriously considered. Electrocardiograms have been producing accurate assessments of heart function and recording data, which have attracted attention from the community of computer scientists [2]–[5]. In recent years, convolutional neural networks (CNNs) have found their application in medicine, including mostly neuroradiology, vascular and interventional radiology, and interventional cardiology [6]. This does not only decrease physicians' workload physicians but also increases diagnostic accuracy and lead to the early detection of disease [7]. Using CNN for the electrocardiogram (ECG) data classification [8] and arrhythmia detection (AD) is the most popular research topic in the field of cardiology, and a comprehensive survey of ECG-based heartbeat classification for AD [9] has reported that many studies rely on the publicly available MIT-BIH3 database (which is recommended by ANSI/AAMI [10] for the validation of medical equipment) and AHA.

II. CHALLENGES AND OPEN ISSUES

In recent years, several methods of CNN utilization in ECG signal classification were proposed. Most of them were

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

using signal processing techniques such as wavelet transformation [11]–[13], frequency analysis [14], bandpass filters [15], [16], statistics [17], the hidden Markov model [18], neural networks, and mixed approaches [19]–[24]. Every method has its own limitations, and the most common problem is the variation of ECG signals from patient to patient that leads to lower accuracy when classifying signals.

The primary limitations of these methods are the variations of QRS complex frequencies that lead to a large number of inaccurate results. Methods using the hidden Markov model are time-consuming due to the use of complex algorithms. Researchers often use the wavelet transform to represent ECG signals [25]–[27], including continuous wavelet transform (CWT), which is used to extract features from ECG signals [28], and scattering wavelet transform (SWT) [29]. However, the accuracy of these techniques remains insufficient [30]. The use of CNNs in detecting heartbeat problems relies on the publicly available MIT-BIH and AHA databases. Nevertheless, the use of popular benchmarks to perform screening and diagnosis of heart diseases has some disadvantages, as they do not contain enough amount of data. If this could be in the domain of deep learning, these benchmarks could be utilized in an efficient manner and thus reasonable, and best performance results could be attained. Andrew Ng's group obtained state-of-the-art results in both recall and precision cardiology performance using CNN with 34 layers and a dataset with more than 500 times the data than previously studied corpora [31]. Some research groups used ECG [32] to obtain an ideal data pattern to identify atrial fibrillation, proving that the use of learned expedition and representation can effectively be a good replacement of the user's hand-crafted features.

Thus, the ECG classification utilizing CNNs heavily depends on the appropriate selection of features that represent the data [6], [7], [31]. Combining the self-proclaimed learned features and different classifiers show a significant improvement in the performance of the patient screening systems.

The problem of developing a method capable of providing high-quality diagnostics using the parameters of all 12 ECG leads to detect heartbeat disturbances remains an urgent issue, since most of currently used methods has its own limitations. The use of CNNs in cardiology is currently limited to the development of diagnostic systems to recognize a narrow class of pathologies with high accuracy. However, to solve the problems of differential diagnostics during mass screening of the population, the list of recognizable diseases should be expanded. With this approach, one can use the advantages of CNNs related to their ability to process large amounts of fuzzy information and provide a high percentage of pathological detection.

This research extends into the scope of the above-mentioned methods and considers another important method. The proposed approach based on using only sound recordings from stethoscopes or smart mobile devices to check and classify heart rhythm. This is of great interest and of great potential for the detection of early heart rhythm disturbances, including self-diagnosis by smartphones users and telemedicine use [32]. The initial detection of health problems usually comes from doctors using stethoscopes, making this method highly suitable for the early detection of heart problem. Let us imagine a time when everyone could possibly perform a prophylactical self-diagnosis and, if necessary, send samples to a doctor or visit him/her in person.

In this research, data from the Kaggle[1] competition, originating from the PASCAL Classifying Heart Sounds Challenge, were used. In some studies [34]–[36], the authors classified these data directly from its Fourier transformation with CNN, omitting the segmentation phase to detect the fundamental physical characteristics of the heartbeat. A similar approach is presented in a study [37] that described a framework based on the autocorrelation feature and diffusion maps, which further provided the SVM classifier. Unfortunately, the datasets were very noisy with numerous mistakes due to the wrong annotation. We have re-annotated samples of phone heart recordings as a part of this research. The correctly annotated data are shared with the research community.

TABLE 1. The number of files in dataset A and B with respect to categories.

Dataset A, 44100 Hz, 16 bit		Dataset B, 4000 Hz, 16 bit	
Category	Train	Category	Train
Normal	45	Normal	336
Murmur	48	Murmur	105
Extra sound	27	Extra systole	66
Artifact	56	–	–

III. DATA DESCRIPTION AND PRE-PROCESSING

The Kaggle (previously PASCAL) Classifying Heart Sounds Challenge provided two datasets used for the analysis [30]. The first dataset, named A, consists of data recorded using a digital stethoscope. The data were gathered in real-world situations and frequently contained background noise such as speech, traffic, and brushing the microphone against cloth or skin, among others. The lengths of audio files were also different, varying between 1 and 30 s. Originally, the Pascal competition datasets were divided into training and testing sets with 4 categories for Dataset A (normal, murmur, extra heart sound, and artifact) and 3 for Dataset B (normal, murmur, and extrasystole). The following table (Table 1) shows the statistics for both datasets.

As mentioned in the introduction, these data were full of errors and incorrectly classified; thus, it should not be used in any research. Alternatively, in this research, Set A was used (manually corrected on the Kaggle) and Set B was corrected by a team of doctors within the scope of this research. After the corrections Set A was divided into 3 categories: artifacts, murmurs, and normal beat. Set B contained heterogeneous data and was divided into five categories: artifacts, extrasystolic arrhythmia, murmurs, tachyarrhythmia,

[1] https://www.kaggle.com/kinguistics/heartbeat-sounds

TABLE 2. The number of files in dataset A and B after corrections in annotation.

Dataset A, 44100 Hz, 16 bit		Dataset B, 4000 Hz, 16 bit	
Category	Samples	Category	Samples
Normal	65	Normal	24
Murmur	53	Murmur	46
Artifact	58	Artifact	24
		Extrasystolic Arrhythmia	99
		Tachyarrhythmia	174

and normal beats. All recordings shorter than 4.5 s were excluded from the dataset due to uselessness.

Table 2 provides detailed statistics for these datasets.

From each category, 10% of data was randomly selected as a test sample. This ensured that each category had a represented number of data and that experiments were compared to each other.

The datasets described are quite limited in terms of the number of training and testing examples. They were used to divide into many categories. There was another dataset, but containing only two classes: the PhysioNet and the Computing in Cardiology Challenge. They assembled the largest public heart sound database, aggregated from eight sources obtained by seven independent research groups worldwide. This comparison aimed to determine, from a single short recording (10–60 s) from a single precordial location and whether the subject of the recording should be referred for an expert diagnosis. The dataset comprised two classes: abnormal sounds (665 training samples and 151 test samples) and normal sounds (2575 training samples and 150 test samples). The task was easier not only because of only two classes and more data but also because recording was taken from the same spot.

The heartbeat sound datasets are primarily audio-based: all heartbeat sounds are stored as WAV files that record either normal or abnormal heartbeats. In general, uncompressed audio is stored as a sequence of numbers that indicate the amplitude of the recorded sound pressure at each time point. In the WAV standard, these numbers are packed into a byte-string, which is interpreted depending primarily on two factors: first, the sampling rate, usually given in Hertz, which indicates how many number samples comprise the second's worth of data; and second, the bit depth (or sample width), which indicates how many bits comprise a single number. These parameters, along with other parameters such as the number of channels (e.g., whether the audio is mono or stereo) are stored in the header of the WAV file.

The so-called frames hold the entire byte-string representing all the audio samples in the sound file. It was necessary to unpack this byte-string into an array of numbers that could be useful. The first question was: how many bytes represent a single observation? In voice recording, 16-bit and 24-bit are the most common sample widths. Additionally, powers of 2 tend to be the easiest to work with. Fortunately, the heartbeat audio samples were 16-bit. The waveforms were generated

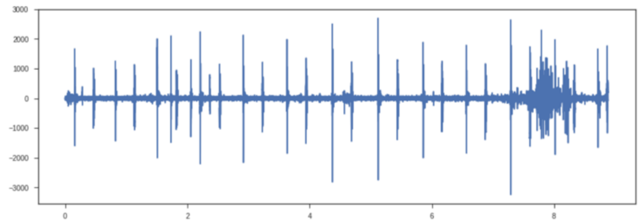

FIGURE 1. Sample waveform.

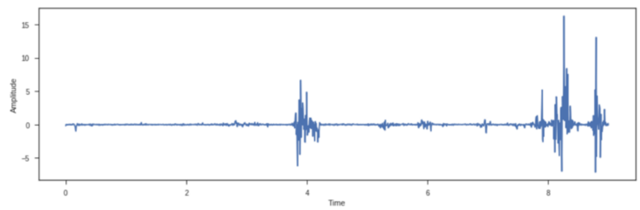

FIGURE 2. An example of a downsampled waveform.

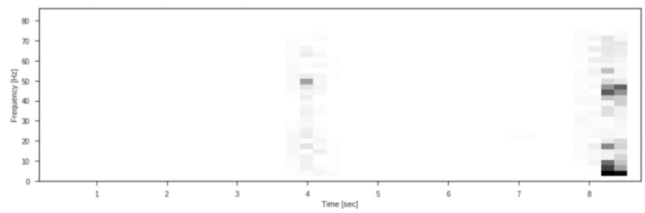

FIGURE 3. Sample generated spectrogram.

based on this information and sampling rate. A sample is shown in Figure 1.

Subsequently, the data were downsampled with a very aggressive low pass filter, which is not needed for a computational time, but seems to improve the generalization of this dataset. With more data, this step can be removed or reduced, and 2-5 extra convolution layers can be added. The reason this works better is probably that what is heard in the stethoscope is almost exclusively low-frequency sounds, especially murmurs.

Figure 2 shows a downsampled file, and Figure 3 shows a spectrogram from the same file.

For easier processing, the spectrogram was reduced to fit into the 224 × 224 px size and was saved in colored. An example of this is shown in Figure 4:

This is exactly the file format t used with CNN for training and prediction. Experiments with different resolutions were also conducted but was not proven to be worthwhile to increase. In the second method, a standard approach was used (simply analyzing audio samples).

IV. EXPERIMENTS

Sampling and experiment design for training CNNs was conducted to classify the real heartbeats as described in Section 2. In general, two main steps were performed in order to train a world-class image classifier; the system used transfer learning and multipart interactive training. In such type of training, the procedure is interrupted and human decision is required

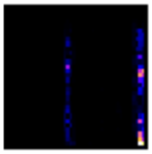

FIGURE 4. Example of data after pre-processing used in training.

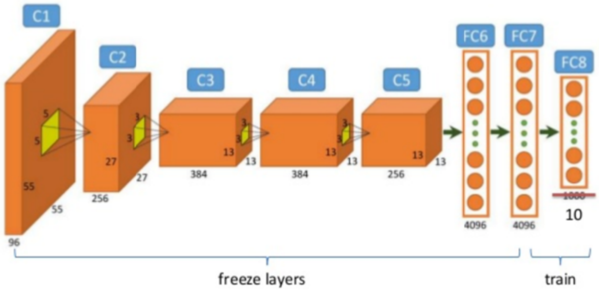

FIGURE 5. CNN topology.

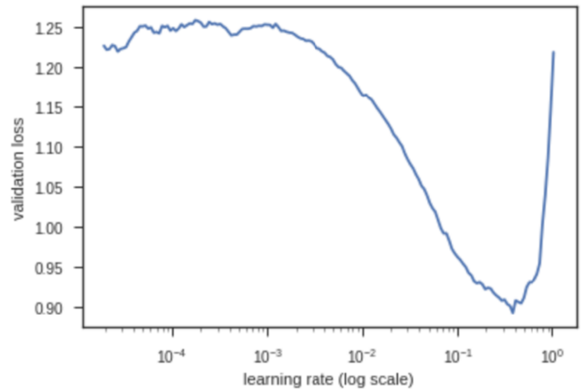

FIGURE 6. Learning rate finding.

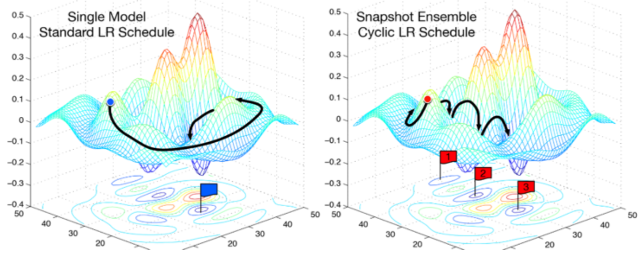

FIGURE 7. Cyclic SGDR example.

in most important parts (e.g., decision on learning rate). This can be extended to 8 more dialed steps:

- Enable data augmentation and activation cache
- Use algorithm to find the highest learning rate where loss is still improving
- Train the last layer of pre-computed activations for 1-2 epochs
- Train the last layer with data augmentation for 2-3 epochs
- Unlock all layers
- Set a smaller learning step for previous layers 3–10 times higher than for the next layers
- Reuse the learning rate to find the algorithm
- Train the entire network until it matches the best possible results

In this research, ResNet-34 and ResNet-50 pre-trained networks were used to facilitate transfer learning; in other words, the system was built using a previously trained model. With this approach, the last layer of the model must be replaced with a layer of appropriate dimensions to the dataset. The model used had been trained for 1000 classes, and the final layer provides a vector with a probability of 1000. The model for heartbeat classification must derive a less-dimensional vector. Figure 5 shows an example of how this was done. The "FC8" layer was replaced with a new layer, with the number of outputs equal to the number of possible classes, and freeze layers are in fact ResNet pre-trained networks.

In this approach, parameters are learned by adjusting the model to the data. Hyperparameters are another type of parameter that cannot be learned directly from the normal training process. These parameters express the properties of a "higher level" model, such as its complexity or speed of learning. Two examples of hyperparameters are the learning rate and number of epochs.

During the iterative neural network training, a batch or mini-batch is a subset of training samples used in one iteration of Stochastic Gradient Descent (SGD). Epoch is a single passage through the entire training set, which consists of multiple SGD iterations.

Then, the model can be fit, that is, it can use gradient descent to find the best parameters for the fully connected layer added, which can separate images. Two hyperthermers are required: learning rate (generally 1e-2 or 1e-3 was a good starting point) and the number of epochs.

In this research, a technique called stochastic gradient descend with restart (SGDR) [38] was used. It is a variant of annealing the learning rate, which gradually reduces the learning rate as you progress through training. This is helpful because when approaching optimal weights, smaller steps should be taken [39].

However, it can possibly be found in a part of the weight space that is not very resistant, i.e., small changes in weight can cause large changes in weight loss. The model should be enhanced to find parts of the weight space that are both accurate and stable. Therefore, from time to time, system increases the speed of learning (a "restart" in "SGDR"), which will force the model to move to another part of the weighing space if the current area is "pointed." Figure 7 illustrates what this can look as if the learning rates are restarted 3 times (in this research, it is called a "cyclic LR schedule"):

After all these steps, all layers of the ResNet CNN can be unlocked; therefore, a smaller learning step should be set for previous layers, i.e., 3–10 times higher than the next layers. Having a well-trained final layer, the remaining layers can be refined. Note that the remaining layers have already been

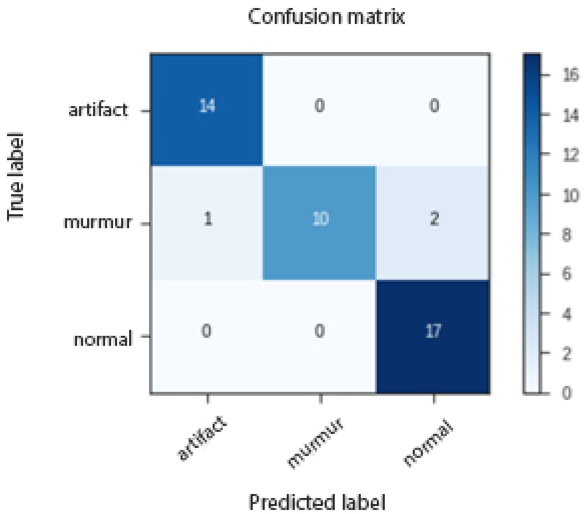

FIGURE 8. Set A result confusion matrix.

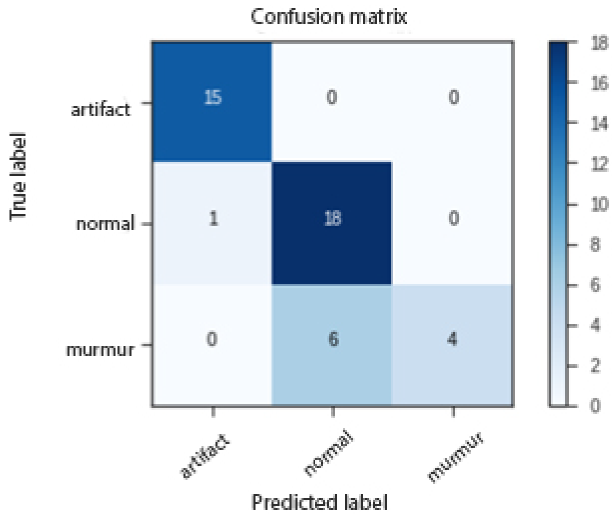

FIGURE 9. Set B result confusion matrix.

trained to recognize images (while final layers have been randomly initiated); therefore, the solution is to be careful not to destroy the accurately adjusted scales that are already present.

In general, earlier layers have more general-purpose functions. Therefore, they can be fine-tuned to the new datasets. For this reason, different learning rate levels for different layers were used: the first few layers were trained at level 1e-4, the middle layers at 1e-3, and the FC layers were reduced to 1e-2, as before. This can be named as a differentiated learning indicator. Another trick used here is that cycles were measured in epochs. A cycle length of 1 would mean a constant decrease in the speed of learning during one epoch and then jumping back up. Finally, the learning rate and descent algorithms for fine-tuning were repeated.

V. RESULTS

Owing to the method described in Section 3, 93.18% prediction accuracy of Set A can be achieved using spectrograms. Figure 8 presents the confusion matrix for this result. Only 3 murmurs were incorrectly classified: one as artifact and 2 as normal heart beat sounds.

In contrast, when trained on plain audio samples with similar optimizations, the system obtained only 90.90% precision. This resulted in 8 mistakes as shown in the confusion matrix in Figure 9.

Regarding the Set B data, using the spectrogram audio representation, the system obtained 59.78% accuracy. It should be noted that those two experiments cannot be directly compared only by accuracy. Set B contained more classes, which made the classification more difficult, and the quality of audio was low. It was tricky even for experienced doctors to make judgements while annotating. The confusion matrix is shown in Figure 10.

In this case, using audio samples provided a slightly better accuracy of 63.04%. This is most likely because images were very noisy, and distinguishing noise from relevant

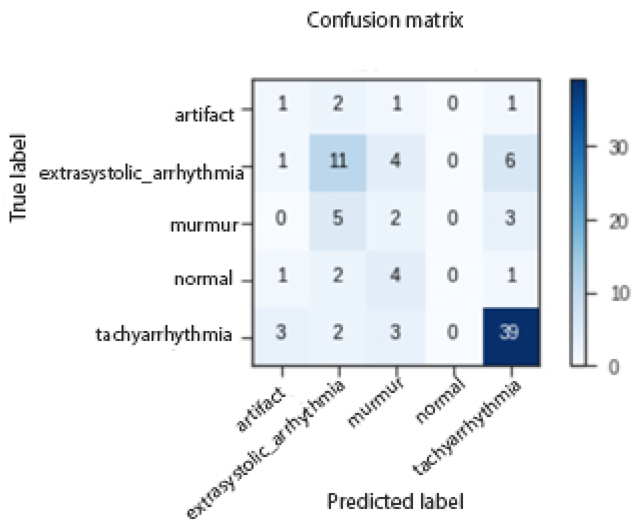

FIGURE 10. Confusion matrix for set A results.

information on images was difficult. The confusion matrix is shown in Figure 11.

The presented method allowed to achieve an accuracy of 99.66% in the PhysioNet challenge, which is 13% better results than the best one from the challenge. The system after few iterations of training made only one mistake on the official testing set. The confusion matrix is shown in Figure 12.

Please note that the entire solution was implemented on the free-to-use Google Collaboratory platform. Implementation, data, and results for Set A[2] can be accessed from the footnote links. The same applies for Set B[3] and PhysioNet challenge.[4]

[2]https://bit.ly/2ROkX4O
[3]https://bit.ly/2CP2xq8
[4]http://bit.do/eHq3t

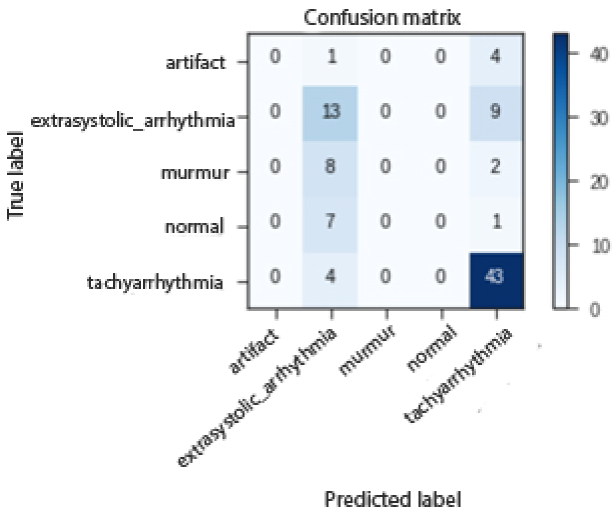

FIGURE 11. Confusion matrix for set B results.

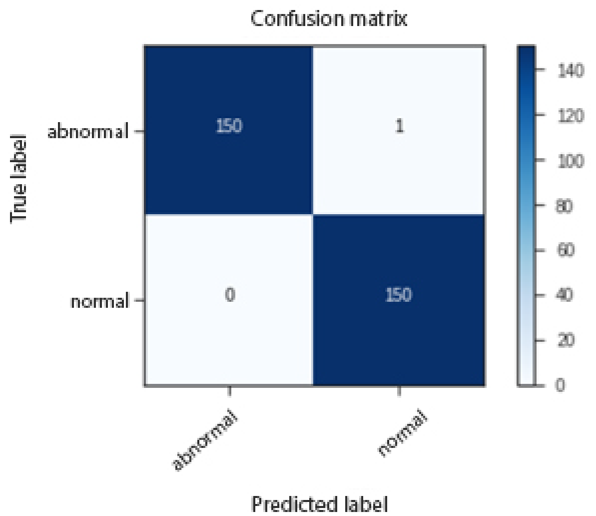

FIGURE 12. Confusion matrix for PhysioNet results.

VI. CONCLUSION

In this research, the utilization of CNN for preliminary screening of heart rhythm disturbances by classifying the heartbeat sounds is proposed. The heartbeat sounds used by CNN can be recorded using a digital stethoscope or mobile phone.

The proposed method deals well with the challenge, especially if audio recordings are transformed into a spectrogram form. The solution outperformed the recent wavelet scattering approach (80%) as described by Kleć [40], (84%) on CNN as described by Rubin *et al.* [5] and even the ECG approach (80.9%) by Rajpurkar *et al.* [31] or by Pyakillya et al. (85%) [4]. The system also consisted 99.66% accuracy on the PhysioNet challenge, whereas the best score after this one is 86.02%. The experiments reveal the optimum CNN preparation and fine-tuning procedure required to obtain the state-of-the-art method to identify heartbeat problems. However, datasets used for these experiments are quite limited in terms of the number of training and testing examples. Increasing it by a factor of ten would most likely result in far greater accuracy improvements in order to allow us to divide datasets into more categories. Another problem is phone recordings.

The accuracy of Set B was lower because the audio samples were of low quality, full of noise, and with low sampling rate. Users of Set B were not properly instructed on how to perform recording, from the chest area, and the length of time. All this resulted in very noisy and varied datasets; however, even on using such datasets with this method, the system can produce a good classifier with more classes than in concurrent research on the same dataset. Performing simple instructions by smartphone users (recording sound in one of four positions used for heart auscultation by cardiologists [38], as well as limiting the level of extraneous noise) will further increase the accuracy of the analysis.

Additionally, the quality and annotation correction of dataset B are the contribution of research. Currently, any research team can use it for their study purposes. Moreover, our method is ready to be released to the market and could be used, e.g., with StethoMe [41] device for remote diagnosis and educational purposes.

VII. HIGHLIGHTS

1. New approach of detecting heart rhythm disturbances based on heartbeat sound is proposed.
2. The proposed method is simple and allows self-diagnosis by non-experienced smartphones users
3. The study reveals the optimum CNN preparation and fine-tuning procedure required to obtain high accuracy
4. Obtained overall accuracy of 99,66%
5. The entire solution was implemented on the free-to-use Google platform

REFERENCES

[1] M. Cierniak-Piotrowska, G. Marciniak, and J. Stańczak, "Statystyka zgonów i umieralności z powodu chorób układu krążenia," in *Zachorowalność i umieralność na choroby ukladu krążenia a sytuacja demograficzna Polski*. Warsaw, Poland: Rządowa Rada Ludnościowa, 2015, pp. 46–80.
[2] U. R. Acharyam, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
[3] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.
[4] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky, "Deep Learning for ECG Classification," *J. Phys., Conf. Ser.*, vol. 913, no. 1, 2017, Art. no. 012004.
[5] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing abnormal heart sounds using deep learning," Jul. 2017, *arXiv:1707.04642*. [Online]. Available: https://arxiv.org/abs/1707.04642
[6] K. Karako, Y. Chen, and W. Tang, "On medical application of neural networks trained with various types of data," *BioSci. Trends*, vol. 12, no. 6, pp. 553–559, 2018.
[7] F. Renna, J. H. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Inform.*, to be published.
[8] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *Proc. Comput. Cardiol. (CinC)*, 2017, pp. 1–4.

[9] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, Apr. 2016.

[10] *Testing and Reporting Performance Results of Cardiac Rhythm and st Segment Measurement Algorithms*, AAMI, Arlington, VA, USA, 1998.

[11] L.-Y. Shyu, Y.-H. Wu, and W. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1269–1273, Jul. 2004.

[12] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2507–2515, Dec. 2006.

[13] M. F. Amri, M. I. Rizqyawan, and A. Turnip, "ECG signal processing using offline-wavelet transform method based on ECG-IoT device," in *Proc. 3rd Int. Conf. Inf. Technol., Comput., Elect. Eng. (ICITACEE)*, Oct. 2016, pp. 1–6.

[14] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 179–185, Feb. 1999.

[15] A. Turnip, M. A. Suhendra, K. D. Esti, and P. Sihombing, "Design of extraction method of SSVEP brain activity with IIR Chebyshev," In *Proc. 5th Int. Conf. Instrum., Control, Automat. (ICA)*, Aug. 2017, pp. 121–125.

[16] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 192–202, Feb. 1999.

[17] J. L. Willems and E. Lesaffre, "Comparison of multigroup logistic and linear discriminant ECG and VCG classification," *J. Electrocardiol.*, vol. 20, no. 2, pp. 83–92, 1987.

[18] D. A. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 9, pp. 826–836, Sep. 1990.

[19] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, Sep. 1997.

[20] Y. R. Garda, W. Caesarendra, T. Tjahjowidodo, A. Turnip, S. Wahyudati, L. Nurhasanah, and D. Sutopo, "Flex sensor based biofeedback monitoring for post-stroke fingers myopathy patients," *J. Phys., Conf. Ser.*, vol. 1007, no. 1, Apr. 2018, Art. no. 012069.

[21] P. Sihombing, Y. M. Siregar, J. T. Tarigan, I. Jaya, and A. Turnip, "Development of building security integration system using sensors, microcontroller and GPS (Global Positioning System) based Android smartphone," *J. Phys., Conf. Ser.*, vol. 978, no. 1, Mar. 2018, Art. no. 012105.

[22] A. Turnip and K.-S. Hong, "Classifying mental activities from EEG-P300 signals using adaptive neural network," *Int. J. Innov. Comp. Inf. Control*, vol. 8, no. 7, pp. 5839–5850, 2012.

[23] A. Turnip, S. S. Hutagalung, J. Pardede, and D. Soetraprawata, "P300 detection using multilayer neural networks based adaptive feature extraction method," *Int. J. Brain Cognit. Sci.*, vol. 2, no. 5, pp. 63–75, 2013.

[24] A. Turnip, A. I. Simbolon, M. F. Amri, P. Sihombing, R. H. Setiadi, and E. Mulyana, "Backpropagation neural networks training for EEG-SSVEP classification of emotion recognition," *Internetwork. Indonesian J.*, vol. 9, no. 1, pp. 53–57, 2017.

[25] I. Güler and E. D. Übeylı, "ECG beat classifier designed by combined neural network model," *Pattern Recognit.*, vol. 38, no. 2, pp. 199–208, 2005.

[26] Y. Kutlu and D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," *Comput. Methods Programs Biomed.*, vol. 105, no. 3, pp. 257–267, 2012.

[27] C.-H. Lin, Y.-C. Du, and T. Chen, "Adaptive wavelet network for multiple cardiac arrhythmias recognition," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2601–2611, 2008.

[28] P. S. Addison, "Wavelet transforms and the ECG: A review," *Physiological Meas.*, vol. 26, no. 5, p. R155, 2005.

[29] J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4114–4128, Aug. 2014.

[30] S. Mannor. (2011). *The PASCAL Classifying Heart Sounds Challenge*. [Online]. Available: http://www.peterjbentley.com/heartchallenge/

[31] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," Jul. 2017, *arXiv:1707.01836*. [Online]. Available: https://arxiv.org/abs/1707.01836

[32] Z. Dokur and T. Ölmez, "ECG beat classification by a novel hybrid neural network," *Comput. Methods Programs Biomed.*, vol. 66, nos. 2–3, pp. 167–181, 2001.

[33] H. P. da Silva, C. Carreiras, A. Lourenço, A. Fred, R. C. das Neves, and R. Ferreir, "Off-the-person electrocardiography: Performance assessment and clinical correlation," *Health Technol.*, vol. 4, no. 4, pp. 309–318, 2015.

[34] W. Zhang and J. Han, "Towards heart sound classification without segmentation using convolutional neural network," *Comput. Cardiol.*, vol. 44, pp. 1–4, Sep. 2017.

[35] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, Feb. 2017.

[36] 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, Oct. 2017.

[37] S.-W. Deng and J.-Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Future Gener. Comput. Syst.*, vol. 60, pp. 13–21, Jul. 2016.

[38] S. Yuenyong, A. Nishihara, W. Kongprawechnon, and K. Tungpimolrut, "A framework for automatic heart sound analysis without segmentation," *Biomed. Eng. Online*, vol. 10, no. 1, p. 13, 2011.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[40] M. Kleć, "Early detection of heart symptoms with convolutional neural network and scattering wavelet transformation," in *Proc. Int. Symp. Methodologies Intell. Syst.* Cham, Switzerland: Springer, 2018, pp. 24–31.

[41] T. Grzywalski, A. Maciaszek, A. Biniakowski, J. Orwat, S. Drgas, M. Piecuch, R. Belluzzo, K. Joachimiak, D. Niemiec, J. Ptaszynski, and K. Szarzynski, "Parameterization of Sequence of MFCCs for DNN-based voice disorder detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5247–5251.

KRZYSZTOF WOŁK received the Ph.D. degree in computer science engineering from the Polish-Japanese Academy of Information Technology, where he is currently an Associate Professor with the Cathedral of Multimedia. He conducts research related to natural language processing and machine learning based on statistical methods and neural networks and deep learning. He eagerly takes up IT challenges and engages in interesting interdisciplinary projects, in particular related to HCI, UX, medicine, and psychology. He was a Lecturer with the Warsaw School of Photography and as an IT Trainer. His specialties as a Teacher are primarily deep learning, machine learning, natural language processing, computational linguistics, multimedia, HCI, UX, mobile applications, HTML 5, Adobe applications, and server products from Apple and Microsoft. As far as, his didactic work is concerned, he leads classrooms with the Faculty of Computer Science, with the New Media Art Department, and with the Polish-Japanese Academy of Information Technology. He also used to lead classes and lectures at the Warsaw School of Photography and Graphic Design. He was the Technical Supervisor of B.A. and M.A. diplomas, a Diploma Reviewer, and a Reviewer for scientific conferences and journals. He is certified Microsoft, Apple, Adobe, w3schools, and EITCA specialist. He is the author of scientific monographs and specialized IT books related to the administration of servers and multimedia. He is also an Editor of the portal in4.pl, pclab.pl, and e-biotechnologia.pl portals and the author of training materials and guides.

AGNIESZKA WOŁK received the master's degree in computer science. She is currently pursuing the Ph.D. degree with the Institute of Literary Research, Polish Academy of Sciences, and the Polish-Japanese Academy of Information Technology, where she is also an Assistant with the Multimedia Department. She holds many certificates of Microsoft, Apple, and Adobe companies. Her research interests include digital humanities and artificial intelligence in the context of natural language processing.