Automatic Acoustic Diagnosis of Heartbeats

Simone Mastrangelo and Stavros Ntalampiras
Department of Computer Science, University of Milan, via Celoria 18, Milan, Italy
simone.mastrangelo1@studenti.unimi.it

Keywords: Heartbeat classification, heartbeat features extraction, heart sounds, machine learning.

Abstract: Automatic identification of heart irregularities based on the respective acoustic emissions is a relevant research field which receives ever-increasing attention over the last years. Devices such as digital stethoscope and smartphones can record the heartbeat sounds and are easily accessible, making this method more appealing. This paper presents different automatic procedures to classify heartbeat sounds coming from such devices into five different labels: *normal*, *murmur*, *extra heart sound*, *extrasystole* and *artifact* so that even people without medical knowledge can detect heart irregularities. The data used in this paper come from two different datasets. The first dataset is collected through an iPhone application whereas the second one is collected from a digital stethoscope. To be able to classify heartbeat sounds, time and frequency domain features are extracted and modeled by different machine learning algorithms, i.e. *k*-NN, random forest, SVM and ANNs. We report the achieved performances and a thorough comparison.

1 INTRODUCTION

The leading cause of global death is represented by cardiovascular diseases (CVDs), whose death rate is estimated at 17.9 million people every year. One third of these deaths occur prematurely in people under 70 years old[1].It is of fundamental importance to be able to promptly identify the symptoms of these diseases in order to ensure the patient the most suitable medical care and avoid possible premature death. However, medical experts and physicians may not always be available to provide an accurate diagnosis (Schneiderman, 2001; Roy et al., 2002). At the same time, tools such as smartphones and digital stethoscopes are easily accessible and can provide a first evaluation in finding any CVDs very quickly even by people without specific medical knowledge or, as in the case of digital stethoscopes, they can be of support to medical staff to facilitate the diagnosis. The aim of this study is to create machine learning models that can autonomously identify a CVD using heartbeat sounds from applications and digital stethoscopes. The data come from a challenge (Bentley et al., 2011) in which there are two distinct datasets: the first contains data collected by iStethoscope Pro, an iPhone app that allows the user to record the sound of his heartbeat, while the second contains recordings of heartbeat sounds coming from DigiScope, a digital stethoscope. These data are divided into 5 categories: normal, murmur extra-heart sound, artifact and extrasystole.

From and audio analysis point of view, a heartbeat includes two sound events: the first (S1 or lub) marks the beginning of a systole (the contraction movement of the myocardium), while the second (S2 or dub) marks the end of systole and the beginning of diastole (the relaxation phase after contraction). The beat of a healthy heart is formed by the succession of S1 and S2 sounds i.e. the succession of systole and diastole. There may be two other sounds, S3 and S4 which are called extra heart sounds which are not part of the normal heart sound. They can be found either individually or together, while they are typically located between S2 and S1. It is not certain that it comprises a sign of a specific disease; nonetheless, they can reveal different clinical conditions[2]. Murmurs are other types of heartbeat sounds that may appear during auscultation; they arise from the flow of blood within the heart or large vessels and can be caused by structural abnormalities of the heart or by an increase in blood flow. They are classified according to their occurrence within the normal cardiac cycle, so they can be systolic, diastolic or continuous. Systolic murmurs are not necessarily a sign of disease

[1]World Health Organization, Cardiovascular Diseases, https://www.who.int/health-topics/cardiovascular-diseases

[2]University of Washington School of Medicine, Technique: Heart Sounds Murmurs, https://depts.washington.edu/physdx/heart/tech.html

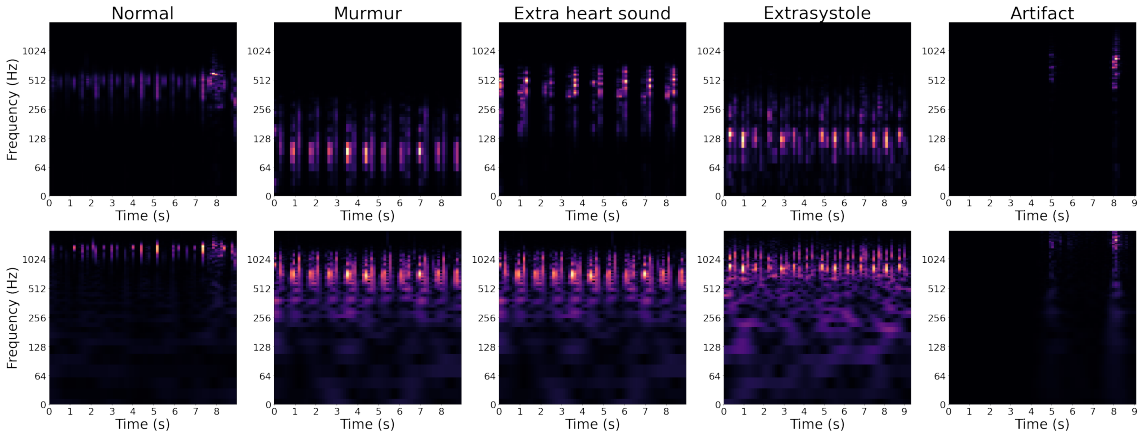

Figure 1: Representative Mel-spectrograms and Constant-q transforms extracted from normal, murmur, extra heart sound, extrasystole and artifact classes existing in the datasets A and B.

and are often perceived in patients with normal heart structure. The diastolic and continuous murmurs, on the other hand, always indicate a structural disease (Davey et al., 2018b). An extrasystole is often a premature heart impulse that is not part of the normal heart cycle. More frequently, they originate from the ventricles taking the name of ventricular extrasystole or premature ventricular complexes; less frequently, they originate from atria, the atrioventricular junction or, rarely, from the sinus node. Extrasystoles can appear after every second or third beat (Davey et al., 2018a). The following section describes the existing works in the area of automatic classification of heartbeat sounds.

2 RELATED WORK

Several studies have been conducted designing methods able to correctly identify cardiovascular diseases, while several of them have used data from phonocardiograms (PCGs), i.e. a plot of the heartbeat sound generated by a phonocardiograph which is accurate but, at the same time, also expensive (Ntalampiras, 2020). In their study, OH, Shu Lih, et al. (Oh et al., 2020) use this type of data achieving satisfactory results. Such PCG signals have been used to train a deep WaveNet model, which is an artificial neural network composed of several layers of neurons, managing to obtain a training accuracy of 97%.

There is a series of studies using data coming from recordings made from smartphones or digital stethoscopes, where various audio feature extraction techniques have been used towards training different machine learning methods. In (Chao et al., 2018) the following classifiers were used: Naive Bayes, support vector machine (SVM), Decision Trees, AdaBoost, Random Forest and Gradient Boosting, managing to obtain an f1-score of 71.37% from the SVM using data from smartphone recordings and an f1-score of 71.26% obtained through a random forest, using data from digital stethoscope recordings. Further improvements were achieved for heartbeat sound recordings from the digital stethoscope thanks to the use of recurrent neural networks reaching an accuracy of 80.8% (Raza et al., 2019).

3 AUDIO FEATURES

From these two datasets several audio features are extracted: energy, zero-crossing rate, spectral rolloff, spectral centroid, mel-frequency cepstral coefficients (MFCCs), mel spectrogram and constant-Q transform. These features are then used for training different classifiers, and finally, the results are compared in terms of precision, recall and f1-score evaluation metrics. It should be mentioned that for the extraction we used Librosa, a Python package for processing audio and music signals (McFee et al., 2015).

During the feature extraction process, the signal is divided into equal-sized frames of 25ms with an overlap of 50%. The extraction can take place in the time domain or in the frequency domain; in the latter case, the Discrete Fourier Transform (DFT) is used to generate the spectrogram of the signal.

Signal energy is calculated for each frame using the root mean square. *Zero crossing rate* indicates the number of times the signal changes sign i.e. goes from positive to negative or negative to positive, divided by the length of the frame. *Spectral rolloff* is the frequency below which a certain percentage of the

magnitude distribution is concentrated. *Spectral centroid* is the center of gravity of the signal spectrum. *Chroma vector* is a twelve-element representation of the spectral energy and is calculated by grouping the DFT coefficients of a short term window in twelve bins. Each bin represents one of the twelve tones of Western music. *MFCCs* comprise a short-term power spectrum signal representation, where the frequency bands are distributed according to a Mel-scale instead of the linearly-spaced approach. This type of feature is widely used in the field of audio analysis due to its discriminating power (Ntalampiras, 2016). The *Mel-scale* filter bank maps the powers of the spectrum using triangular overlapping windows. Mel-scale has a distortion effect of the frequencies in order to conform to the human auditory system which is able to more easily distinguish the low-frequency region (Giannakopoulos and Pikrakis, 2014). Representative Mel-spectrograms and Constant-q transforms characterizing the employed datasets are illustrated in Fig. 1. Last but not least, the calculation of *constant-Q transform* is similar that of the Fourier transform, with the difference that it has a constant ratio between center frequency and resolution. As such, a constant pattern in the frequency domain is obtained for the sounds whose harmonic frequency has been plot, unlike the standard DFT where the spacing between the frequencies is constant (Brown, 1991).

4 CLASSIFICATION TECHNIQUES

Aiming at evaluating the performance of diverse classifiers on the present problem, we used the following five techniques: *k*-nearest neighbors (k-NN), random forest (RF), support-vector machines (SVM), and Artificial neural networks (ANN) including convolutional neural network (CNN).

k-NN the specific classifier has been thoroughly applied to audio classification problems, such as discrimination of an audio stream in speech, music, ambient sound and silence (Lu et al., 2001). The *k*-NN implementation used for this project is that of scikit-learn (Pedregosa et al., 2011); the algorithm generates several *k*-NN instances in order to search for the optimal number of neighbors and ultimately, the one offering the highest f1-score is chosen.

Random forest This is a popular classifier which consists in an ensemble of decision trees where their outcomes are averaged in order to improve the overall accuracy. In this work, the number of used decision trees is equal to 200.

Support Vector Machine This is another popular classifier used in audio analysis (Lu et al., 2003); SVM aims at discovering the optimal hyperplane separator that minimizes the classification error on a validation set of data. The radial basis function was employed as a kernel during the SVM learning (Zhu et al., 2007).

Artificial neural network ANNs are structures inspired by the animal and human brain and they have been quite successful in diverse applications such as commerce, industry and finance (Kruse et al., 2016). ANNs have been shown to produce good results in the classification of abnormal heart sound using data coming from phonocardiogram (Ari and Saha, 2009). The model used in this project is a multilayer perceptron network with three hidden layers trained with the standard version of the back-propagation algorithm (Rojas, 1996).

Convolutional Neural Network Following the recent success of CNNs in audio pattern recognition applications (Purwins et al., 2019; Ntalampiras, 2020). Here, two different methods have been used: for the first model uses the same features as the ANN. The net is composed of three convolutional layers activated by a rectified linear unit (ReLU). Each convolutional layer is followed by one max-pooling layer and one dropout layer. The last two hidden layers are standard fully-connected ones, while the output layer employed softmax as the activation function. For the second method, the already computed features log-mel spectrogram and constant-Q transform have been used as inputs. They were downsampled to 177x44px and converted to RGB. It should be mentioned that the two features are then evaluated separately. For each model hyperparameter tuning is based on random search, while maximizing the obtained accuracy comprises the overall objective. Finally, CNN's structure is similar to the first method with the only differences being the input shape, which fits the input image, and the hyperparameters.

It should be mentioned that log-mel spectrogram and constant-Q transform were employed only during CNN training as dictated by the related literature (Purwins et al., 2019).

Table 1: The obtained results for all considered classifiers when applied on Dataset A. The highest rate per figure of merit is emboldened.

Classifier	Precision	Recall	f1-score
k-NN	**0.86**	**0.80**	**0.82**
Random forest	0.81	0.75	0.73
SVM	0.77	0.75	0.73
ANN	0.79	0.78	0.76
CNN+MFCC	0.78	0.67	0.69
CNN+log-mel spectrogram	0.79	0.67	0.65
CNN+constant-Q transform	0.57	0.58	0.55

5 EXPERIMENTAL SET-UP AND RESULTS

5.1 Data

The present set of data originates from a challenge where the purpose is to classify heartbeat sounds in order to promptly help diagnose CVDs (Bentley et al., 2011). Two different datasets are available: in the first (dataset A) the data were gathered using IStethoscope Pro, an iPhone app, while in the second (dataset B) the data were gathered in a hospital setting during a clinical trial using a digital stethoscope called DigiScope. Dataset A contains 124 audio files sampled at 44100 Hz while dataset B contains 312 audio files sampled at 4000 Hz. The files are labeled with the normal and murmur classes for both datasets, while dataset A additionally contains the artifact (various types of sounds unrelated to the heartbeat) and extra heart sounds classes. Interestingly, dataset B contains the extrasystole class as well. It should be noted that all classifiers operated on identical test, validation, and train sets of data so as to obtain a reliable comparison. The division was 70% for training, 10% for validation and 20% for testing.

5.2 Preprocessing

This phase precedes the feature extraction phase and serves to prepare the dataset so that the data are processed effectively by the various machine learning models. Following the challenge's guidelines, audio files lasting less than two seconds have been eliminated as they could not represent a heartbeat full cycle. Subsequently, files coming from dataset A were downsampled to 4000 Hz, while files belonging to both datasets the files were trimmed to the same length. It should be mentioned that standard normalization techniques including mean removal and variance scaling ($z = \frac{x-\mu}{\sigma}$, where μ is the mean of the training samples and σ is the standard deviation of the training samples) have been applied on the extracted features before being used by the classifiers (Ntalampiras, 2021).

5.3 Results

To thoroughly measure the classifiers performances, we employed standardized figures of merit, i.e. *precision*, *recall* and *f1-score*. Table 1 includes the results with respect to dataset A. In this case, the best-performing classifier is k-NN which offers the highest scores for precision, recall and f1-score ($k = 120$). More in detail, classes *artifact* and *mumrur* are identified with significantly higher rates with respect to the rest of classes, having an f1-score of 0.91 and 0.90 respectively. Interestingly, the present approach outperforms the state of the art reported in (Chao et al., 2018), where the obtained *f1-score* is equal to 0.71[3]. The remaining classifiers offer similar performances, managing to better classify data associated with *artifact* and *murmur*. Unfortunately, performances associated with deep CNNs are worse than shallow approaches, while performances are better for *artifact* and *murmur* classes. This could be an indication that more data is needed in order to allow the deep network to learn the distributions associated with the specific classes of heartbeat sounds.

Table 2 illustrates the results for every classifier when applied on dataset B. In general, we see that the classifiers performed poorer in this case, while there is no winning classifier for all figures of merit. The random forest approach performed better in *recall* and *f1-score*, while k-NN achieved the highest precision rate ($k = 264$). Overall, every classifier demonstrates similar results in terms of correct and incorrect classifications. It should be noted that for Dataset B, the presented rates are lower with respect to the ones shown in (Chao et al., 2018), where f1-score is 0.71. Moreover, it is often that *extrasystole* are *murmur* are misclassified as *normal*. At the same time, *normal* samples are correctly classified, while random forest

[3]https://github.com/lindawangg/Classifying-Heartbeats

Table 2: The obtained results for all considered classifiers when applied on Dataset B. The highest rate per figure of merit is emboldened.

Classifier	Precision	Recall	f1-score
k-NN	**0.64**	0.69	0.60
Random forest	0.62	**0.73**	**0.66**
SVM	0.61	0.69	0.60
ANN	0.59	0.61	0.60
CNN+MFCC	0.59	0.72	0.64
CNN+log-mel spectrogram	0.58	0.72	0.64
CNN+constant-Q transform	0.60	0.68	0.57

reached *f1-score* of 0.82 in this case. However, log-mel spectogram as input offered the highest *f1-score* (0.85) for the normal class. We conclude that the task represented by Dataset B is more challenging due to noises/interferences associated with sounds emitted from other human organs. Similarly to dataset A, increasing data quantity could be particularly useful especially for the deep learning based solutions.

6 CONCLUSIONS

In this paper, a great variety of machine learning methods has been extensively evaluated on heartbeat sound classification, with the aim being the detection of abnormalities such as extrasystole, extra heart sound and murmurs. To this end, several temporal and spectral audio features have been exploited. Such an automatic framework aims at supporting decisions made by healthcare professionals, as well as early diagnosis, e.g. using a smartphone, so as to quickly check for any existing heartbeat abnormalities and contact an expert physician. It was shown that the recognition rates reached by such audio pattern recognition methods differ significantly between dataset A (smartphone) and dataset B (stethoscope). In the first datasets, the methods achieved quite good results in distinguishing artifacts and murmurs, while in the second the results were worse, especially for the extrasystole class where no model was able to classify correctly.

The results of the present experiments could be primarily improved by expanding the datasets. More specifically, it would be especially useful to have available more heartbeat samples representing the extra heart sound classes, i.e. extrasystole and murmur. Further improvements could be obtained by correctly extracting and labeling the S1 and S2 sounds of the heartbeat, and use them as an additional input feature for the different classifiers. From a machine learning perspective, it would be interesting to experiment with a) data augmentation methods, including transfer learning (Ntalampiras and Potamitis, 2018), b) modeling temporal properties of heartbeats, using e.g. temporal convolutional networks (Yan et al., 2020), and c) employed one-shot learning techniques (Lake et al., 2015) accommodating scarce data availability.

REFERENCES

Ari, S. and Saha, G. (2009). In search of an optimization technique for artificial neural network to classify abnormal heart sounds. *Applied Soft Computing*, 9(1):330–340.

Bentley, P., Nordehn, G., Coimbra, M., and Mannor, S. (2011). The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. http://www.peterjbentley.com/heartchallenge/index.html.

Brown, J. C. (1991). Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, 89(1):425–434.

Chao, A., Ng, S., and Wang, L. (2018). Listen to your heart: Feature extraction and classification methods for heart sounds.

Davey, P., Sprigings, D., El-Kadri, M., and Hart, G. (2018a). *Extrasystoles*. Oxford University Press, Oxford, UK.

Davey, P., Sprigings, D., Timperley, J., and Hothi, S. (2018b). *Murmur*. Oxford University Press, Oxford, UK.

Giannakopoulos, T. and Pikrakis, A. (2014). Chapter 4 - audio features. In Giannakopoulos, T. and Pikrakis, A., editors, *Introduction to Audio Analysis*, pages 59 – 103. Academic Press, Oxford.

Kruse, R., Borgelt, C., Braune, C., Mostaghim, S., Steinbrecher, M., Klawonn, F., and Moewes, C. (2016). *Computational Intelligence: A Methodological Introduction*. Springer Publishing Company, Incorporated, 2nd edition.

Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.

Lu, L., Jiang, H., and Zhang, H. (2001). A robust audio classification and segmentation method. In *Proceedings of*

the ninth ACM international conference on Multimedia, pages 203–211.

Lu, L., Zhang, H.-J., and Li, S. Z. (2003). Content-based audio classification and segmentation by using support vector machines. *Multimedia systems*, 8(6):482–492.

McFee, B., Raffel, C., Liang, D., Ellis, D., Mcvicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. pages 18–24.

Ntalampiras, S. (2016). Automatic analysis of audiostreams in the concept drift environment. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6.

Ntalampiras, S. (2020). Deep learning of attitude in children's emotional speech. In *2020 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 1–5.

Ntalampiras, S. (2020). Identification of anomalous phonocardiograms based on universal probabilistic modeling. *IEEE Letters of the Computer Society*, (01):1–1.

Ntalampiras, S. (2021). Speech emotion recognition via learning analogies. *Pattern Recognition Letters*, 144:21–26.

Ntalampiras, S. and Potamitis, I. (2018). Transfer learning for improved audio-based human activity recognition. *Biosensors*, 8(3):60.

Oh, S. L., Jahmunah, V., Ooi, C. P., Tan, R.-S., Ciaccio, E. J., Yamakawa, T., Tanabe, M., Kobayashi, M., and Acharya, U. R. (2020). Classification of heart sound signals using a novel deep wavenet model. *Computer Methods and Programs in Biomedicine*, page 105604.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S., and Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2):206–219.

Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., and On, B.-W. (2019). Heartbeat sound signal classification using deep learning. *Sensors*, 19(21):4819.

Rojas, R. (1996). *Neural Networks*. Springer Berlin Heidelberg.

Roy, D., Sargeant, J., Gray, J., Hoyt, B., Allen, M., and Fleming, M. (2002). Helping family physicians improve their cardiac auscultation skills with an interactive CD-ROM. *Journal of Continuing Education in the Health Professions*, 22(3):152–159.

Schneiderman, H. (2001). Cardiac auscultation and teaching rounds: how can cardiac auscultation be resuscitated? *The American Journal of Medicine*, 110(3):233–235.

Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y. (2020). Temporal convolutional networks for the advance prediction of ENSO. *Scientific Reports*, 10(1).

Zhu, Y., Ming, Z., and Huang, Q. (2007). SVM-based audio classification for content- based multimedia retrieval. In *Multimedia Content Analysis and Mining*, pages 474–482. Springer Berlin Heidelberg.