# PCG Sound Quality Classification Based Mel-Frequency Cepstral Coefficients Features and Convolutional Neural Network

Edwin M. Chambi[1] , Erasmo Sulla-Espinoza[1] , and Jorge Rendulich[1]
[1] Universidad Nacional de San Agustin de Arequipa, Perú, emendozacham@unsa.edu.pe, esullae@unsa.edu.pe, jrendulich@unsa.edu.pe

*Abstract–PCG signals are important to give diagnoses of any disease, specifically the problem of incorrect data collection is addressed, either by an inexperienced doctor or by the patient himself. In this article we seek to classify the signals by their quality labeled as 'Acceptable and Unacceptable' with balanced data and the proposed method-based Mel frequency cepstral coefficients (MFCC) to feature extraction and convolutional neural networks (CNN) for their classification. The results showed us a high performance of our model with 95.3% precision, 95.1% recall, 95.2% F1-score and 95.3% accuracy and it will be compared with already existing works.*

*Keywords—PCG, Mel frequency cepstral coefficients (MFCC), Convolutional Neural Networks (CNN), Quality, Classification, Features.*

## I. INTRODUCTION

The heartbeat sound is a fundamental part of the study in the field of cardiology when we record them digitally, they are called a phonocardiogram (PCG) [1] in addition to being an important element for diagnosing diseases to patients who may suffer from a disease. In the last SARS-CoV-2 pandemic, the number of deaths from cardiovascular problems increased due to confinement and lack of patient care; According to data from the Peruvian government, 10 out of every 100 deaths from this strain suffered from cardiovascular problems, thus exacerbating the symptoms and making it, together with the virus, deadly risk factors [2].

Telemedicine has become a solution to the problem of not attending to the patient personally and quickly with heart difficulties, but the field is opening up to new challenges such as the inexperience of the treating physician and erroneous readings of the PCG signals taken by the patient himself or failure of the device that captures these signals. Since good quality readings will not always be obtained.

In recent years, different proposals for deep learning algorithms have been developed for the classification of heart signals with different pre-processing methods and extraction of signal features. Hong et al. [3] classifies the signal quality of a public dataset in a binary way using SVM. In Potes et al. [4] extracts 126 characteristics from audio signals and uses a variation of AdaBoost for classification of normal and abnormal PCG signals. Reference [5] they use three stages for their classification where first they extract the spectrograms, then deep features are extracted from three different pre-trained convolutional neural network models such as AlexNet, VGG16,

and VGG19. SVM is used for further classification. Combine a conventional feature engineering method with deep learning algorithms to automatically classify normal and abnormal heart sounds [6]. ANN to classification [8].

Mel-frequency cepstral coefficient (MFCC) is an efficient representation of the acoustic information of an audio signal. In Reference [7] the authors use MFCC and deep neural networks to classify phonocardiography signals as normal and abnormal using the PhysioNet.org heart sound database.

Taking advantage of the CNN's ability to learn audio characteristics from automatically, while at the same time using MFCC to provide a compact and efficient representation information base on that, the aim is classifying the quality of PCG signals of balanced database based on the extraction of the MFCC features for its subsequent training with a CNN with binary classification and reach better performance against previous works. This paper is part of the research development in the area of biomedical engineering of the Universidad Nacional de San Agustin de Arequipa [12-15] for the Think Health project which seeks to improve medical assistance in the auscultation process.

## II. MATERIALS AND METHODS

### A. Dataset

In this article we use the database compiled by Hong et al. [3]. Their database is a total accumulated of 4 databases which are: CINC challenge 2016 [9], Pascal classifying heart sound challenge 2012 [10], PHHS 2015 [11] and heart sounds Catania 2011. Having a total of 7893 audios of heart sounds.

### B. Signals Labeling

In order to have a correct classification of the signals, we balanced database and was necessary to be able to label them and our base will be the labels made by Hong et al. [3], the signals that have a quality rating of 1 to 3 we label as Unacceptable, while those with a quality rating of 4 to 5 we label them as Acceptable. The total data and respective labeling for each one is shown in the Table I.

TABLE I
TYPE LABELS

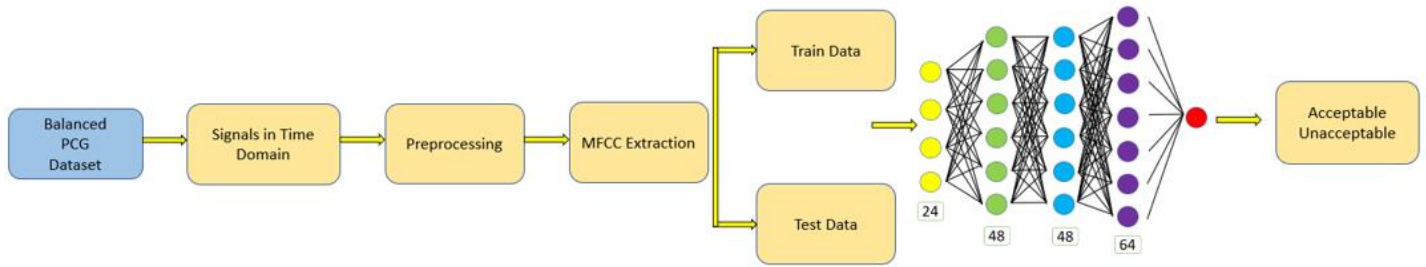| Label | Total data |
|---|---|
| Unacceptable | 4386 |
| Acceptable | 3507 |

### C. Methods

Fig. 1 Block diagram of PCG sound classification model

In Fig. 1 we can see the diagram that was used to carry out this investigation. The first part we have the database in which we have 2 types of labeled signals, to be able to visualize them it's necessary to show them in the time domain for Acceptable signals Fig. 2 and Unacceptable signals Fig. 3.

It's possible to observe a line at the maximum points of noise to be able to identify them and thus apply a Chebyshev filter of order 5 with a cut-off frequency of 2Hz. As part of the preprocessing, it's also necessary to have the duration of the signals normalized for their analysis, in this case it was indicated for 3 seconds of duration for all signals.

The extraction of the MFCC for these signals was established with a number of mels of 128 and a sampling frequency of 2kHz. Having the signals ready, it's necessary for the entire dataset to be divided into 2 parts for this case: Train data being 80% of the signals and Test data being 20% of the signals.
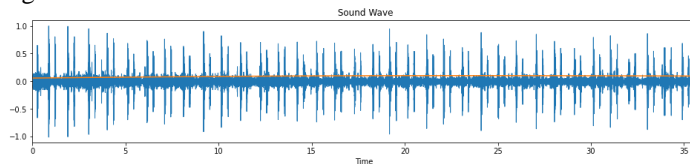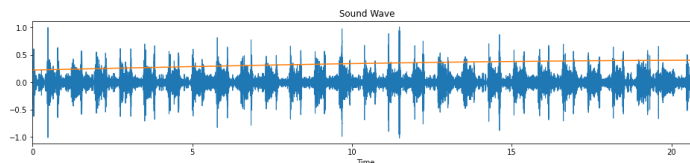

Fig. 2 Acceptable Signal.


Fig. 3 Unacceptable Signal.

The algorithm of our proposed convolutional network is composed of 4 layers in which there are 24 neurons in the first, the next 2 composed of 48 neurons, the third layer with 64 neurons and the last layer with Sigmoid activation for binary classification. The other parameters can be seen in Table II.

TABLE II
PARAMETER SETTING OF CNN MODEL

| CNN Parameter | Value |
|---|---|
| Learning rate | 1e-4 |
| Number of Epochs | 100 |
| Batch size | 128 |
| Dropout rate | 0.5 |

| Hidden layers | 24-48-48-64 |
|---|---|
| Hidden layer activation | ReLU |
| Optimizer | ADAM |
| Loss | Binary cross-entropy |
| Output activation | Sigmoid |

Precision (1), recall (2), F1-score (3), Accuracy (4) are criteria used to evaluate the performance of our algorithm. These parameters are determined by classifier's true positive (TP), true negative (TN), false positive (FP) and false negative. The confusion matrix Fig.4 will show us the information about these values in order to evaluate our classifier.


Fig. 4 Confusion Matrix.

$$Precision = \frac{TP}{TP+FP} \qquad (1)$$

$$Recall = \frac{TP}{TP+FN} \qquad (2)$$

$$F1-score = \frac{2*Precision*recall}{Precision+recall} \qquad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

Experimental work was performed using Keras installed on a computer with an AMD Ryzen(TM) 7-1700 processor, NVIDIA GeForce 3060 GPU and 16GB RAM.

III. BINARY CLASSIFICATION RESULTS

To evaluate the performance of the algorithm we will use the precision, recall and accuracy metric to compare with previous works. Table III show a comparison with binary classifications works, SVM by Hong et al. [3] and ANN by Carlos [8]. The Fig. 4 shows the accuracy model and loss model for CNN model using as input the MFCC of the database. In Fig. 4(a), during a 100-epoch training, peaks of 96.3% training

accuracy were reached while in Fig. 4(b) we see overfitting given that the testing loss is higher than the training loss. The performance of the algorithm for classification with the proposed metrics to evaluate it for 95.3% precision, 95.1% recall, 95.2% F1-score and 95.3% accuracy. For this case we will emphasize the values obtained for sensitivity and specificity, "Recall" measures the proportion of Unacceptable signals that are correctly identified as Unacceptable. A high recall 95.1% indicates that the model is good at identifying Unacceptable signals. "Precision" measures the proportion of Acceptable signals that are correctly identified as Acceptable. A high precision 95.3% indicates that the model is good at identifying Acceptable signals.

TABLE III
COMPARISON BETWEEN THE PROPOSED METHOD AND STATE-OF-THE ART METHODS

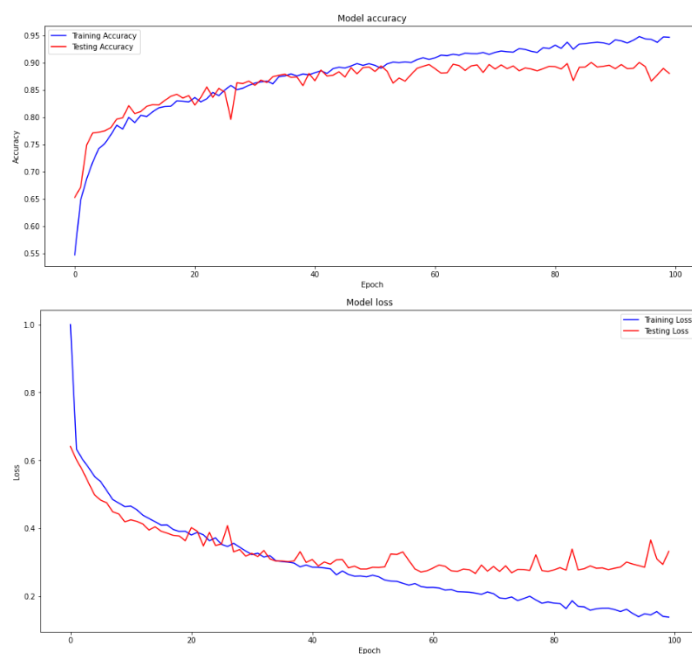| Method | Pr. | Rc. | Acc. |
|--------|-----|-----|------|
| MFCCs | 95.3% | 95.1% | 95.3% |
| ANN [18] | 95.4% | 93.2% | 94.5% |
| SVM [11] | 96.1% | 92.2% | 94.3% |



Fig. 4 a) Model Acuracy. b) Model Loss.

### III. CONCLUSIONS

Taking readings of PCG signals with different qualities is possible and can cause the patient to be given an erroneous diagnosis, which is why it was necessary to classify them according to their quality. The proposed CNN model taking as input the MFCC of the PCG signals in the precision part is surpassed by 0.1% compared to the ANN model [8]. On the other hand, in the other recall and accuracy metrics we have a higher percentage with 1.9% and 0.8% respectively against the ANN model [8].

It's important to balance recall and precision to maximize the accuracy of the model and obtain an accurate classification.

The error that can be seen between the training and testing loss curves in Fig. 4(b) can be reduced by changing the density and dropout parameters of algorithm, but the computational cost would increase given the magnitude of signals in the database.

This work, unlike other types of analysis and classification of PCG signals, the trained model is lightweight and can be easily implemented in a low computing device. For future work, it's proposed to use the Mel and Cochleagram spectrograms towards certain segments of the signal to obtain more promising results.

### REFERENCES

[1] Özkan Arslan, Mustafa Karhan, Effect of Hilbert-Huang transform on classification of PCG signals using machine learning, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 10, Part B, 2022, Pages 9915-9925, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.12.019.

[2] M. d. s. (Minsa), «Plataforma digital única del estado peruano,» 1 Octubre 2021. Available: https://www.gob.pe/institucion/minsa/noticias/542169-minsa-personas-con-enfermedades-cardiovasculares-tienen-tres-veces-mas-riesgo-de-fallecer-por-covid-19.

[3] Hong Tang, Miao Wang, Yating Hu, Binbin Guo, Ting Li, "Automated Signal Quality Assessment for Heart Sound Signal by Novel Features and Evaluation in Open Public Datasets", BioMed Research International, vol. 2021, Article ID 7565398, 15 pages, 2021. https://doi.org/10.1155/2021/7565398

[4] C. Potes, S. Parvaneh, A. Rahman and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 621-624.

[5] Demir, F., Şengür, A., Bajaj, V. et al. Towards the classification of heart sounds based on convolutional deep neural network. Health Inf Sci Syst 7, 16 (2019). https://doi.org/10.1007/s13755-019-0078-0

[6] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, "Classification of Heart Sounds Using Convolutional Neural Network," *Applied Sciences*, vol. 10, no. 11, p. 3956, Jun. 2020, doi: 10.3390/app10113956.

[7] Maknickas, V.; Maknickas, A. Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. Physiol. Meas. 2017, 38, 1671–1684. [Google Scholar] [CrossRef] [PubMed]

[8] Carlos M. Huisa, Edward Figueroa, Supo, E., Rendulich, J., & Sulla, E. (2022)., "PCG heart sound quality classification using neural networks and Smote Tomek Links for the Think Healtj project," *International Conference on Computational Intelligence and Innovative Technologies (ICCIIT – 2022),* in press.

[9] Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016. Gari D. Clifford, Chengyu Liu, Benjamin Moody, David Springer, Ikaro Silva, Qiao Li, Roger G. Mark

[10] Peter Bentley, «Classifying Heart Sounds Challenge,» PASCAL, 13 April 2012. Available: http://www.peterjbentley.com/heartchallenge/. [Último acceso: 7 November 2022].

[11] Padaccini, A., & Beritelli, F. (2013). Performance evaluation of heart sounds biometric systemson an open dataset. 2013 18Th International Conference On Digital Signal Processing (DSP). https://doi.org/10.1109/icdsp.2013.6622835

[12] R. Huamani R., J. R. Talavera, E. A. S. Mendoza, N. M. Dávila and E. Supo, "Implementation of a real-time 60 Hz interference cancellation

algorithm for ECG signals based on ARM cortex M4 and ADS1298," *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON),* Cusco, Peru, 2017, pp. 1-4, doi: 10.1109/INTERCON.2017.8079725.

[13] T. R. Sulla, S. J. Talavera, C. E. Supo and A. A. Montoya, "Non-invasive glucose monitor based on electric bioimpedance using AFE4300," *2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON),* Lima, Peru, 2019, pp. 1-3, doi: 10.1109/INTERCON.2019.8853561.

[14] E. Supo, M. D. D. C, R. H. R., J. Galdos, J. Rendulich and E. Sulla, "PRD as an indicator proposal in the evaluation of ECG signal acquisition prototypes in real patients," *2022 IEEE ANDESCON*, Barranquilla, Colombia, 2022, pp. 1-4, doi: 10.1109/ANDESCON56260.2022.9989674.

[15] Medina, A., Lopez, N., Galdos, J., Supo, E., Rendulich, J., & Sulla, E. (2022). Continuous Blood Pressure Estimation in Wearable Devices Using Photoplethysmography: A Review. International Journal of Emerging Technology and Advanced Engineering, 12(10), 104-113. https://doi.org/10.46338/ijetae1022_12