A Peak Detection Algorithm for Localization and Classification of Heart Sounds in PCG Signals using K-means Clustering

Shrey Agarwal, Yashaswi Upmon, Muhammad Zubair, *Student Member, IEEE*, and Shaik Rafi Ahamed, *Member, IEEE*

Abstract—Traditionally, the heart sound classification process is performed by first finding the elementary heart sounds of the phonocardiogram (PCG) signal. After detecting sounds S1 and S2, the features like envelopograms, Mel frequency cepstral coefficients (MFCC), kurtosis, etc., of these sounds are extracted. These features are used for the classification of normal and abnormal heart sounds, which leads to an increase in computational complexity. In this paper, we have proposed a fully automated algorithm to localize heart sounds using K-means clustering. The K-means clustering model can differentiate between the primitive heart sounds like S1, S2, S3, S4 and the rest of the insignificant sounds like murmurs without requiring the excessive pre-processing of data. The peaks detected from the noisy data are validated by implementing five classification models with 30 fold cross-validation. These models have been implemented on a publicly available PhysioNet/Cinc challenge 2016 database. Lastly, to classify between normal and abnormal heart sounds, the localized labelled peaks from all the datasets were fed as an input to the various classifiers such as support vector machine (SVM), K-nearest neighbours (KNN), logistic regression, stochastic gradient descent (SGD) and multi-layer perceptron (MLP). To validate the superiority of the proposed work, we have compared our reported metrics with the latest state-of-the-art works. Simulation results show that the highest classification accuracy of 94.75% is achieved by the SVM classifier among all other classifiers.

Index Terms—Heart sounds, K-means Clustering, Machine learning models, Multi Layer Perceptron and PCG signals

I. INTRODUCTION

As per Centers for Disease Control and Prevention (CDC), the major cause of death is heart disease which has caused

6,59,041 deaths in U.S [28]. To detect heart disease with the use of computers, one have to first understand the four fundamental heart sounds S1, S2, S3 and S4. The first heart sound S1 is produced by the closing of the mitral and tricuspid valves. The second heart sound S2, results from the closure of the aortic and pulmonic valves. The third heart sound S3 also termed as ventricular gallop, is observed just after S2 when the mitral valve opens, to allow passive filling of the left ventricle (LV). Lastly, the fourth heart sound S4, also known as the atrial gallop, comes just before S1 when the atria contract to force blood into the LV [27].

The quickest, simplest and most economical initial line of screening is cardiac auscultation. It is performed for the patients with heart problems, including valve disease, arrhythmia, heart failure, pulmonary hypertension, etc. Contrary to this, heart sounds are tough to recognize and examine by physician, because the important events are closely positioned in time, the sounds are feeble, and their frequency content is at the bottom end of the hearable frequency span [1]. Large-scale tutoring is needed to properly grasp cardiac auscultation. It was found that, on an average, only twenty percent of the medical trainees can successfully detect heart problems using auscultation [2].

In the area of biomedical engineering, computerized abnormal heart sound localization can be reviewed as a crucial preceding step to cardiovascular disease recognition [3]. The procedure of recognizing whether a given heart sound is normal or abnormal, can be split into 3 important steps: localization, feature extraction, and classification [4]. First of all, the localization technique locates the elementary heart sounds of the PCG signal S1 and S2 as demonstrated in Fig. 1.

But predicting the elementary heart sounds is altogether a composite task, and can be altered by other sounds such as the existence of S3 and S4 heart sounds and noise [5]. But we have also tried to classify S3 and S4 heart sounds which are present in the abnormal signal as shown in Fig. 2. Previously, Messener et al. [23] have extracted the envelope and spectral features using the same dataset, and for detecting the sequence they have implemented deep recurrent neural network (DRNN). In [24] by employing ensemble empirical mode decomposition (EEMD) along with the kurtosis features is used to localize the presence of S1 and S2 heart sounds.

This paragraph of the first footnote will contain the date on which you submitted your paper for review.

Shrey Agarwal is with the Department of Computer Science Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Orissa, India (e-mail: shreyagarwal1111@gmail.com).

Yashaswi Upmon is with the Department of Computer Science Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Orissa, India (e-mail: yashaswiupmon2407@gmail.com).

Muhammad Zubair is with the Department of Electronics and Communication Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, AP, India (e-mail: muhammadzub17@gmail.com).

Prof. Shaik Rafi Ahamed is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology, Guwahati, Assam, India (e-mail: rafiahamed@iitg.ac.in).

Fig. 1. Position of S1 and S2 heart sounds in a normal signal

Fig. 2. Position of S1, S2, S3 and S4 in an abnormal signal

Springer et al. [25] have proposed a method for correct segmentation of the first and second heart sound within PCG signals using an hidden semi-markov model (HSMM) and logistic regression. In [26], the Hilbert vibration decomposition system is used to decompose the heart sound into a fixed number of sub-components while conserving the phase information to further expedite the localization and identification of the signal. However, the accuracies reported in these two works are relatively low.

In order to improve the classification accuracy and automate the process, we proposed an algorithm in which we have used the labelled peaks as the features for the five classifiers, that have been implemented to recognize normal and abnormal heart sounds. Our work presents a novel approach using unsupervised learning model which has been implemented on the PhsioBank ATM Challenge 2016 for the localization of heart sounds. Being unsupervised it may be able to recognize the fundamental sounds in real time noisy PCG signals. Also it saves time by using the labelled peaks as the feature for the classification models, instead of manually finding the features of the peaks. We have attained trailblazing outcomes compared with the start-of-the-art methods for numerous databases present in PhsioBank ATM Challenge 2016.

II. LITERATURE REVIEW

Much of the prior work was focused on PCG segmentation using deep learning and machine learning algorithms. In [6], Fernando et al. proposed a technique to detect the heart state from the PCG signals by using recurrent neural networks (RNNs). They have used attention based learning for the extraction of salient features from noisy and irregular heart signals. Their proposed algorithm was implemented on two databases, which are “PhysioNet/CinC challenge 2016 (PCC)” and “M3dicine database”, which is a privately collected dataset consisting of both human and animal heart sounds. For segmentation tasks, numerous feature combinations were evaluated.

Dissanayake et al. in [7] exploited segmentation of heart sound before proceeding for classification by using three different deep learning architectures. In [8], the proposed procedure uses shapley additive explanations (SHAP) and occlusion map algorithm which aided the segmentation to play

a vital role in classification and detection of abnormal and normal heart sounds. Further, the use of MFCC features can easily identify the location of S1 and S2 in PCG signals. The new proposed classifiers in the above work can classify the heart signals with an accuracy of approximately 100%.

Renna et. al. [9] have used a deep convolutional neural network in concurrence with the underlying hidden markov models (HMM), and HSMM, to deduce emission distribution. The outcome of the work has given an average sensitivity of 93% along with an average true positive rate of 94%.

Gojoreski et al. [10] have described a novel approach for the prediction of chronic heart failure using PCG signals. In this method, classical machine learning models have been trained from numerous expert designed features, whereas, deep learning models have been trained from a spatiotemporal and spectral representation of signals. The model has then been implemented on their chronic heart failure (CHF) database along with an open-source training dataset available at “PhysioNet Cardiology Challenge”. This methodology attains 89.3% accuracy compared to the baseline model. They infer, 15 expert features that can be useful for the detection of CHF phases with a claimed accuracy of 93.2%.

Naami et al. [11] describe adaptive network-based fuzzy inference system (ANFIS) for the classification of abnormal signals. The de-noising of the PCG signals has been performed using notch and butterworth filtering, and the denoised signal is then fed to discrete Fourier transform (DFT) and from third cumulant using high order spectral (HOS) analysis, five features were extracted. These five features have been used with neural network systems to classify heart signals into abnormal and normal signals. The accuracy of their proposed model ranges from 63-89% for the classification task.

Tien-En Chen et al. [12] have proposed a DNN based method for the detection of S1 and S2 heart sounds. The sounds have been converted to MFCCs. The MFCCs features have been clustered into two categories which are then fed into DNN, that further classifies S1 and S2 with an accuracy of 91%. Mishra et. al. [13] have considered morphological characteristics, derived for the PCG signals for recognition of S1 and S2. This method is implemented on publicly available databases along with which they have implemented this on experimentally collected HSS recording in the supervision of the physicians. Quantifying the non-linear and non-stationary

Fig. 3. Normal PCG Signal

Fig. 5. Localised peaks of normal PCG Signal using K-means clustering

Fig. 4. Abnormal PCG Signal

Fig. 6. Localised peaks of abnormal PCG Signal using K-means clustering

nature of PCG signals was inferred from variational mode decomposition (VMD), which is based on quantifying the non-stationary and non-linear nature of PCG signals. Experiments conducted by them conclude a notable improvement in the classification task as compared to the prior works.

All the prior works discussed above are based on the classification of heart sounds in PCG Signals. Most of them are focused on supervised learning for signal localization. Moreover, deep learning requires an enormous amount of data for training purposes and is also a time-consuming task. To overcome this limitation, in this work we have proposed a novel algorithm using unsupervised learning which is time efficient and reports high accuracy. Additionally, most of the other works have used feature extraction technique, which makes the process more computationally expensive and laborious. Therefore, in our work we have not statistically calculated the features and have instead used an automated process for finding the peaks in the signal which further acts as features for the classification models. Simulation results show that the proposed algorithm exhibits better performance as compared

to the above state-of-the-art works.

III. PROPOSED METHODOLOGY

Firstly, we have collected the PCG data from PhsioBank ATM. Then we computed the absolute value of the amplitude, which is applied as the input to K-means clustering algorithm. This machine learning algorithm clusters the amplitude of the PCG signals into two clusters and labels it as 0 and 1. Further, to identify the cluster having peaks, we compare the size of the cluster. Whichever cluster is smaller in size, we have considered it to contain the peaks of the signals. Then we have labelled those peaks using a pattern of -1 and 1, and the rest of the signal is labelled as 0. Next, we have balanced this dataset using synthetic minority over-sampling technique (SMOTE). After which we have fed this pattern to classification models, which predict the normal and abnormal heart sounds. The system architecture of the proposed methodology is shown in Fig. 7 and is described in the following subsections.

A. Peak detection using K-means clustering

For a normal cardiac signal, the cycle consists of four major regions that are the first heart sound (S1), the systole pause interval (SPI), the second heart sound (S2) and the diastole pause interval [13]. But the pathological signals may contain a third heart sound (S3) and a fourth heart sound (S4) [14]. To identify the four sounds using machine learning, we have used an unsupervised K-means clustering method. We have considered both abnormal and normal datasets for the implementation of the proposed algorithm. The representation of one of the normal signals and abnormal signals present in the dataset is shown in Fig. 3. and Fig. 4. respectively.

Next, to identify the peaks in PCG signals, we have proposed a novel algorithm which is based on K-means clustering as described in Algorithm 1. In Algorithm 1, we first take the absolute value or modulus of the amplitude of the PCG signals. Then, we implement K-means clustering where the value of k is taken as 2. The equation of K-means clustering is as follows:

Algorithm 1 Peak Detection Algorithm using K-means Clustering

```

A = Amplitude of signal
X = abs(A)
zero = 0
one = 0
switch = -1
K-means clustering (number of clusters = 2)
for value in K-means output do
    if value == 1 then
        one = one + 1
    else
        zero = zero + 1
    end if
end for
if one > zero then
    non-peak = one
else
    non-peak = zero
end if
for each signal in dataset do
    if switch == -1 then
        switch = 1
    else
        switch = -1
    end if
    if value of signal == non-peak then
        signal = 0
    else
        signal = switch
    end if
end for

```

$$J = \sum_{j=1}^k \sum_{i=1}^n ||x_i^j - c_j||^2 \quad (1)$$

Where J is the objective function, k is the number of clusters, n is the number of cases, x_i is the i^{th} case and c_j is

the centroid for cluster j . The euclidean distance is calculated by the function $||x_i^j - c_j||^2$ whose sum is for all the amplitudes of the PCG signal. The distance is summed up for both the centroids. Then we consider the data points for the centroid which has the minimum distance.

As there are two clusters, the cluster with more data points is suppressed as it does not contain the peaks. This can be inferred by looking at the signal, that the data points of peaks in the signal is less than the data points in the rest of the signal.

To suppress the signal we substitute all the non-peak signals with 0 and peaks with a pattern of 1 and -1. The representation of 1 and -1 depends on representation of latest peak example, if 1 has already been assigned to previous peak then -1 would be assigned to current peak.

After running the peak detection algorithm for both normal and abnormal PCG signals, we have considered the new values of the peak as mentioned in step 5 of Algorithm 1 for the classification of the signals. The detected peaks of normal and abnormal PCG signals are shown in Fig. 5. and Fig. 6. respectively.

B. Classification of Peaks using Machine Learning models

For the classification of the PCG signals, using labelled peaks, we have used KNN, logistic regression, SVM, SGD and MLP. But before inputting the data into the classifiers we have used the SMOTE algorithm for oversampling the labelled peaks synthetically. As the data provided in the dataset was imbalanced.

To implement the models, we have taken X as peaks detected through algorithm 1 and Y as the target value. Where the target value is either normal or abnormal heartbeat. We have divided the training and testing data in the ratio of 8:2.

1) *K-Nearest Neighbors*: To find in which class the signal will lie using KNN we calculate the distance between the testing data and for each training data one by one.

2) *Logistic Regression*: For implementing logistic regression a weight matrix is created for all the values of input and is then multiplied with it. The result of which is used to predict the class label.

3) *Support Vector Machine*: SVM uses hyperplane to classify the data into normal and abnormal PCG signals.

4) *Stochastic Gradient Descent*: The goal of stochastic gradient descent for binary classification is to put input which is the labelled peaks in the linear scoring function and then learn from it. To find the model parameters we minimize $E(w,b)$ given in the equation below:

$$E(w, b) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha * R(w) \quad (2)$$

where we have taken L as:

$$L(y_i, f(x_i)) = \begin{cases} \max(0, 1 - y_i f(x_i))^2 & \text{if } y_i f(x_i) > 1 \\ -4y_i f(x_i) & \text{otherwise} \end{cases} \quad (3)$$

Fig. 7. System hierarchy for implementation of proposed algorithm

and the value of R as:

$$R(w) = \frac{1}{2} \sum_{j=1}^m w_j^2 \quad (4)$$

5) *Multi Layer Perceptron* : For MLP we have used X as the input in the input layer after which it is passed to the first hidden layer from which the output is calculated as:

$$O_j = \max(0, \sum_{i=1}^{40} w_i^j * X_i) \quad (5)$$

For the second hidden layer:

$$O_j = \max(0, \sum_{i=1}^{80} w_i^j * O_i) \quad (6)$$

For the third hidden layer:

$$O_j = \max(0, \sum_{i=1}^{20} w_i^j * O_i) \quad (7)$$

For the fourth hidden layer:

$$O_j = \max(0, \sum_{i=1}^{10} w_i^j * O_i) \quad (8)$$

For the fifth hidden layer:

$$O_j = \max(0, \sum_{i=1}^5 w_i^j * O_i) \quad (9)$$

And finally for the output layer:

$$O_j = \frac{1}{1 + e^{-a}} \quad (10)$$

where $a = \sum_{i=1}^2 w_i^j * X_i$.

IV. EXPERIMENTAL RESULTS

A. Dataset description

The proposed methodology is implemented and evaluated on the Challenge 2016 training set A, B, C and F from the PhysioBank database [15]. The data has been collected from both healthy subjects and unhealthy patients who are either in clinical or non-clinical environment, and the heart sound signals are taken from diverse locations on the body. Also, the normal signals were from healthy patients and the abnormal recordings were from subjects with an established cardiac problems. Both healthy patients and unhealthy subjects included adults and children. One important point to be noted is that the dataset is unbalanced, which means that the count of the normal signals does not match that of abnormal signals. Also as the recordings were taken in a non-controlled environment most of the recordings are disrupted by various

TABLE I
EVALUATION FOR TRAINING SET A

Models	Accuracy	Precision	Recall	F-1 Score
SVM	92	0.87	0.88	0.86
Logistic regression	87	0.81	0.8	0.82
KNN	88	0.88	0.87	0.88
SGD	85	0.82	0.81	0.81
MLP	87	0.88	0.87	0.86

TABLE II
EVALUATION FOR TRAINING SET B

Models	Accuracy	Precision	Recall	F-1 Score
SVM	94	0.9	0.88	0.88
Logistic regression	90	0.87	0.87	0.87
KNN	55	0.76	0.54	0.42
SGD	89	0.82	0.81	0.81
MLP	83	0.82	0.82	0.82

TABLE III
EVALUATION FOR TRAINING SET C

Models	Accuracy	Precision	Recall	F-1 Score
SVM	100	1	1	1
Logistic regression	100	1	1	1
KNN	60	0.3	0.5	0.37
SGD	100	1	1	1
MLP	100	1	1	1

noises. Further, we divided the dataset into 80% and 20%. Where 80% of the dataset is considered for training the models and the rest 20% for validation.

B. Peak Localization

In this work, K-means algorithm has been implemented on the modulus of the amplitude of the signals for localization task. Two clusters are formed from the signals. It can be inferred from the data points of the clusters that a large number of data points are not representing the fundamental heart sounds in the signals. We have labelled the amplitudes of the signal into the pattern of 0, -1 & 1. Allocation of 1 & -1 is totally dependent on the their prior existence in the signal and changes alternatively wherever the amplitude of a peak occurs. For the insignificant amplitudes in PCG signals, we have labelled them as 0. Then machine learning models were implemented to classify the signals into its true positive class.

C. Classification

Five models are implemented for classification of the heart sound signals. Before feeding the features into the classification model SMOTE algorithm was used to over-sample the data synthetically. The main reason behind this step is imbalanced nature of the available data set. The pattern which has been formed after the clustering of the signals are fed as the features to perform classification on signals and the target value is whether the provided signal is normal or abnormal signal, for which abnormal signal is consider as 1 and normal signal is consider as 0. All the data is divided into training and test data in the ratio of 8:2. The models used for classification are SVM, Logistic Regression, KNN, SGD, MLP algorithm. The steps considered in implementing the models are as follows:

- The data set was annotated and then it was shuffled so each point contributes independently and there is no biasing.

TABLE IV
EVALUATION FOR TRAINING SET F

Models	Accuracy	Precision	Recall	F-1 Score
SVM	93	0.9	0.87	0.87
Logistic regression	97	0.97	0.97	0.97
KNN	50	0.44	0.44	0.41
SGD	88	0.81	0.77	0.77
MLP	88	0.88	0.87	0.87

TABLE V
AVERAGE OF METRICS OF TRAINING SETS A,B,C & F

Models	Accuracy	Precision	Recall	F-1 Score
SVM	94.75	0.92	0.91	0.9
Logistic regression	93.5	0.91	0.91	0.92
KNN	63.25	0.6	0.59	0.52
SGD	90.5	0.86	0.85	0.85
MLP	89.5	0.9	0.89	0.89

- The data set was divided into training and testing data set. The size of training data set is 80% of the whole data set and test is about 20% of the whole data set.
- After the deduction of pattern from the signal, we have implemented five classification models to classify the signal into the abnormal and normal signal.

D. Classification Results

In this work, the classification task has been evaluated on various metrics i.e. Precision, Accuracy, F-1 Score and Recall. To further validate and visualize the performance, we have plotted receiver operating characteristic (ROC) curve for the SVM classifier because it achieved highest overall accuracy, shown in Fig 8. The evaluation is conducted on all the datasets whose result is given in Table I, Table II, Table III, Table IV and Table V.

Fig. 8. Comparison of ROC Curve of all dataset

The clustering and classification of abnormal and normal signals are implemented using python 3.8.5 on a PC having 8GB RAM and powered with an i5 processor.

TABLE VI
COMPARISON OF OUR WORK WITH OTHER STATE-OF-THE ART IMPLEMENTATIONS

Author	Method	Accuracy (%)
Humanyu et. al. [17]	CNN layers having tConv unit	82.27
Renna et. al. [9]	Deep convolutional neural network	93.70
Xiao et. al. [18]	Convolutional Neural Network	93.00
Gjoreski et. [10]	Deep Learning Model Learn from Spectro-temporal portrayal	92.90
Noman et. al. [16]	Markov-switching auto-regressive (MSAR)	91.20
Al-Naami et. al. [11]	Adaptive Neuro-Fuzzy Inference System (ANFIS),	89.00
Shuvo et. al. [19]	CardioXNet, a new light end-to-end CRNN structure	86.57
Boulares et al. [20]	Pre-trained CNN mode	89.00
Chen et al. [21]	CNN and modified frequency slice wavelet transform (MFSWT)	93.91
Li et al. [22]	CNN	86.80
Proposed Work	K-means Clustering and SVM	94.75

V. DISCUSSION

The accuracy of SVM given in Table V is compared with other state-of-the-art methods, which have been implemented to classify PCG signals as normal or abnormal. In [16] Noman et al. have extended the MSAR model to a switching linear dynamic system (SLDS). They have also made a new algorithm via mixing of duration-dependent Viterbi algorithm and the switching Kalman filter. To evaluate the results of their methodology they have used the same dataset and have achieved an accuracy of 91.2%.

Other authors such as Humanyu et al. [17] have proposed a new CNN layer, having time-convolutional (tConv) units, that copy Finite Impulse Response (FIR) filters. They have evaluated their work on publicly available multi-domain datasets and their results have shown the competency of the proposed learnable filterbank CNN architecture in getting durability towards sensor/domain variability in PCG signals by attaining an accuracy of 82.27%. Further, a new heart sound classification system based on technologies that use deep learning, have been implemented for cardiovascular disease prediction by Xiao et al. [18]. To evaluate their model they have also used the same dataset that has been used in this paper. They have achieved state-of-the-art results, accuracy of 93%. In paper [11], to scrutinize the execution of ANFIS, which is started by spectral analysis features used to detect abnormal heart sounds, Al-Naami et al. have used the same dataset mentioned in this paper. They have got a classification accuracy ranging from 63-89%. Using the same dataset Shuvo et al. [19] have implemented CardioXNet, a new light end-to-end CRNN structure for automatic recognition of 5 target values of cardiac auscultation using unprocessed PCG signals. The proposed model has achieved an accuracy of 86%.

Boulares et al. [20] have shown experimental outcomes based on fine-tuning pre-trained CNN models. The accuracy they have got is 89%. In paper [21] Chen et al. have proposed a new system that takes together CNN and modified frequency slice wavelet transform (MFSWT) for the classification of heart sounds as normal and abnormal. A concealed Markov model is taken into consideration to detect the location of every cardiac cycle in the signal. They have combined two CNN models for training by using sample entropy (SampEn) to see which model is used for classification. For evaluation, they have also used PhysioNet Computing in Cardiology Challenge

2016 and have achieved efficiency and a high accuracy of 93.91% for the classification of normal and abnormal signals. A combination of standard feature engineering method with deep learning algorithms has been implemented by Li et al. [22] for automatic classification of normal and abnormal heartbeats. Stratified five-fold cross-validation was used on the PhysioNet/CinC Challenge 2016 dataset for evaluation, giving an accuracy of 86.8%.

From the above discussed work, it can be contrasted that the proposed method can produce a higher accuracy and is working without taking any features of the fundamental heart sounds for classification. Moreover, the proposed approach has used an unsupervised machine learning method for the localization of heart sounds which does not need to know the location of signals before hand nor requires any pre-processing of the data.

VI. CONCLUSION

In this paper, we have introduced a novel approach for peak localization using K-means Clustering. We demonstrated how the above mentioned algorithm is more efficient and also predicts the peaks accurately. By only using the knowledge of the peaks detected, it is enough for the machine learning and deep learning models to classify normal and abnormal PCG signals. The presented method not only attains high-level performance compared with state-of-the-art baseline systems but is also remarkably more efficient in terms of the steps it requires to classify the signals as abnormal or normal. Moreover, we have used an algorithm using an unsupervised machine learning model which is faster and can automatically localize the fundamental heart sounds without calculating any features. As per our experimental reports, SVM achieved an highest accuracy of 94.75% mentioned in Table V. To validate the performance of our proposed algorithm, we have compared our work with other state-of-the-art works shown in Table VI. The high-level outcomes produced by the proposed method avers its usefulness in computer supported heart sound examination and gives a stage for applications having localization of murmurs and ejection clicks and a variety of subsequent examination. In future, we can use this algorithm on real-time data to check its efficiency and robustness.

REFERENCES

- [1] A. Pease, "If the heart could speak," Pictures Future, pp. 60–61, 2001.

- [2] S. Mangione, "Cardiac auscultatory skills of physicians-in-training: A comparison of three english-speaking countries," *Amer. J. Med.*, vol. 110, no. 3, pp. 210–216, 2001.
- [3] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami and C. Fookes, "A Robust Interpretable Deep Learning Classifier for Heart Anomaly Detection Without Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2162–2171, June 2021, doi: 10.1109/JBHI.2020.3027910.
- [4] V. Arora, R. Leekha, R. Singh, and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *Modern Phys. Lett. B*, vol. 33, Aug. 2019, Art. no. 1950321.
- [5] V. G. Sujadevi, K. P. Soman, R. Vinayakumar, and A. U. Prem Sankar, "Anomaly detection in phonocardiogram employing deep learning," in *Proc. Adv. Intell. Syst. Comput.*, 2019, vol. 711, pp. 525–534.
- [6] Fernando, Tharindu, Houman Ghaemmaghami, Simon Denman, Sridha Sridharan, Nayyar Hussain, and Clinton Fookes. "Heart sound segmentation using bidirectional lstms with attention." *IEEE journal of biomedical and health informatics* 24, no. 6 (2019): 1601-1609.
- [7] Dissanayake, Theekshana, Tharindu Fernando, Simon Denman, Sridha Sridharan, Houman Ghaemmaghami, and Clinton Fookes. "A robust interpretable deep learning classifier for heart anomaly detection without segmentation." *IEEE Journal of Biomedical and Health Informatics* (2020).
- [8] Lundberg, Scott, and Su-In Lee. "A unified approach to interpreting model predictions." *arXiv preprint arXiv:1705.07874* (2017).
- [9] Renna, Francesco, Jorge Oliveira, and Miguel T. Coimbra. "Deep convolutional neural networks for heart sound segmentation." *IEEE journal of biomedical and health informatics* 23, no. 6 (2019): 2435-2445.
- [10] Gjoreski, Martin, Anton Gradišek, Borut Budna, Matjaž Gams, and Gregor Poglajen. "Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds." *IEEE Access* 8 (2020): 20313-20324.
- [11] Al-Naami, Bassam, Hossam Fraihat, Nasr Y. Gharaibeh, and Abdel-Razzak M. Al-Hinnawi. "A Framework Classification of Heart Sound Signals in PhysioNet Challenge 2016 Using High Order Statistics and Adaptive Neuro-Fuzzy Inference System." *IEEE Access* 8 (2020): 224852-224859.
- [12] Chen, Tien-En, Shih-I. Yang, Li-Ting Ho, Kun-Hsi Tsai, Yu-Hsuan Chen, Yun-Fan Chang, Ying-Hui Lai, Syu-Siang Wang, Yu Tsao, and Chau-Chung Wu. "S1 and S2 heart sound recognition using deep neural networks." *IEEE Transactions on Biomedical Engineering* 64, no. 2 (2016): 372-380.
- [13] K. A. Babu, B. Ramkumar and M. S. Manikandan, "Automatic Identification of S1 and S2 Heart Sounds Using Simultaneous PCG and PPG Recordings," in *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9430-9440, 15 Nov.15, 2018, doi: 10.1109/JSEN.2018.286941
- [14] Varghees, V. Nivitha, and K. I. Ramachandran. "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope." *IEEE Sensors Journal* 17, no. 12 (2017): 3861-3872.
- [15] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13)
- [16] F. Noman, S. -H. Salleh, C. -M. Ting, S. B. Samdin, H. Ombao and H. Hussain, "A Markov-Switching Model Approach to Heart Sound Segmentation and Classification," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 705-716, March 2020, doi: 10.1109/JBHI.2019.2925036.
- [17] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng and T. Hasan, "Towards Domain Invariant Heart Sound Abnormality Detection Using Learnable Filterbanks," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2189-2198, Aug. 2020, doi: 10.1109/JBHI.2020.2970252.
- [18] Xiao, Bin, Yunqiu Xu, Xiuli Bi, Junhui Zhang, and Xu Ma. "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption." *Neurocomputing* 392 (2020): 153-159.
- [19] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami and A. Gumaci, "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings," in *IEEE Access*, vol. 9, pp. 36955-36967, 2021, doi: 10.1109/ACCESS.2021.3063129.
- [20] M. Boulares, T. Alafif and A. Barnawi, "Transfer Learning Benchmark for Cardiovascular Disease Recognition," in *IEEE Access*, vol. 8, pp. 109475-109491, 2020, doi: 10.1109/ACCESS.2020.3002151.
- [21] Chen, Y., Wei, S. and Zhang, Y. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. *Med Biol Eng Comput* 58, 2039–2047 (2020). <https://doi.org/10.1007/s11517-020-02218-5>
- [22] Li, Fan, Hong Tang, Shang Shang, Klaus Mathiak, and Fengyu Cong. "Classification of heart sounds using convolutional neural network." *Applied Sciences* 10, no. 11 (2020): 3956.
- [23] E. Messner, M. Zöhrer and F. Pernkopf, "Heart Sound Segmentation—An Event Detection Approach Using Deep Recurrent Neural Networks," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964-1974, Sept. 2018, doi: 10.1109/TBME.2018.2843258.
- [24] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient Heart Sound Segmentation and Extraction Using Ensemble Empirical Mode Decomposition and Kurtosis Features," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1138-1152, July 2014, doi: 10.1109/JBHI.2013.2294399.
- [25] D. B. Springer, L. Tarassenko and G. D. Clifford, "Logistic Regression-HSMM-Based Heart Sound Segmentation," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822-832, April 2016, doi: 10.1109/TBME.2015.2475278.
- [26] S. Barma, B. Chen, W. Ji, S. Rho, C. Chou and J. Wang, "Detection of the Third Heart Sound Based on Nonlinear Signal Decomposition and Time-Frequency Localization," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1718-1727, Aug. 2016, doi: 10.1109/TBME.2015.2500276.
- [27] <https://www.healio.com/cardiology/learn-the-heart/cardiology-review/topic-reviews/heart-sounds>
- [28] <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>

Shrey Agarwal is presently pursuing a bachelor's degree in information technology at Kalinga Institute of Industrial Technology, Bhubaneswar, India. He is expecting to complete his bachelor's by June 2022. He completed a research internships at IoT Lab KIIT, Bhubaneswar, India and an industrial internship at InfoAxxon, Noida, India. He is presently working on multiple research projects one of which is with Samsung. His research interests include Machine learning and Deep learning.

Yashaswi Upmon is presently pursuing bachelor's degree in Information Technology at Kalinga Institute of Industrial Technology, Bhubaneswar, India. He is expecting to complete his bachelor's by June 2022. He is presently working on multiple research projects. His research interests include Natural Language Processing, Machine Learning and Deep Learning.

Muhammad Zubair (Student Member, IEEE) has completed his bachelor's degree in electronics and communication engineering at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, India; affiliated to JNTU Kakinada, Andhra Pradesh, India.

He has published four research papers in various IEEE conferences and two papers in IEEE Sensors Journal. He completed three internships at IIT Guwahati, Assam, India. He is presently an active student member at IEEE and IETE organisations. His research interests include Biomedical signal and image processing, Time-Frequency Analysis, Machine learning, Deep learning and Brain computer interface.

Shaik Rafi Ahamed received his Ph.D. degree from IIT Kharagpur, India, in 2007. He is currently working as a professor with the Department of Electronics and Electrical Engineering, IIT Guwahati, India. His teaching and research interests include digital, adaptive, biomedical, and VLSI signal processing and machine learning.