Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 2, November 2024, pp. 846∼853
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i2.pp846-853

Development of explainable machine intelligence models for heart sound abnormality detection

Jeevakala Siva Ramakrishna[1], Sonagiri China Venkateswarlu[2], Kommu Naveen Kumar[3], Parikipandla Shreya[2]
[1]Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), Institute of Aeronautical Engineering, Hyderabad, India
[2]Department of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Hyderabad, India
[3]Department of Electrical and Computer Engineering, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India

Article Info

Article history:

Received Mar 8, 2024
Revised Jul 10, 2024
Accepted Jul 14, 2024

Keywords:

Explainable learning models
Heart abnormalities
Heart sound detection
LIME
SHAP

ABSTRACT

Developing explainable machine intelligence (XAI) models for heart sound abnormality detection is a crucial area of research aimed at improving the interpretability and transparency of machine learning algorithms in medical diagnostics. In this study, we propose a framework for building XAI models that can effectively detect abnormalities in heart sounds while providing interpretable explanations for their predictions. We leverage techniques such as SHapley additive exPlanations (SHAP) and local interpretable model-agnostic explanations (LIME) to generate explanations for model predictions, enabling clinicians to understand the rationale behind the algorithm's decisions. Our approach involves preprocessing heart sound data, training machine learning models, and integrating XAI techniques to enhance the interpretability of the models. We evaluate the performance of our XAI models using standard metrics and demonstrate their effectiveness in accurately detecting heart sound abnormalities while providing insightful explanations for their predictions. This research contributes to the advancement of XAI in medical applications, particularly in the domain of cardiac diagnostics, where interpretability is crucial for clinical decision-making.

Corresponding Author:

Jeevakala Siva Ramakrishna
Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)
Institute of Aeronautical Engineering
Hyderabad, India
Email: j.sivaramakrishna@iare.ac.in

1. INTRODUCTION

The human heart produces characteristic sounds, known as heart sounds, that result from the closure of heart valves and the movement of blood within the heart chambers. These sounds are crucial indicators of cardiac health and can provide valuable diagnostic information about the condition of the heart. Heart sounds are typically categorized into two main types: S1 (the first heart sound) and S2 (the second heart sound). S1 is produced by the closure of the mitral and tricuspid valves at the beginning of systole, while S2 is produced by the closure of the aortic and pulmonary valves at the beginning of diastole. The purpose of detecting heart sound abnormalities lies in the early identification of cardiac pathologies, such as valvular disorders, murmurs, and other structural abnormalities. The heart sound detetion mechanism requires an accurate and reliable detection algorithms, the interpretation of complex heart sound patterns, and the integration of these algorithms into clinical practice. The use of machine learning (ML) algorithms in medical diagnostics has shown promising results in detecting various health conditions [1]-[3]. The complexity of ML algorithms often makes it difficult for clinicians to trust and understand the decisions made by these models, which can hinder their adoption in real world healthcare settings [4]. In the context of cardiac diagnostics, the interpretation of heart sounds plays a crucial role in identifying abnormalities and making informed clinical decisions. Traditional methods of heart sound analysis are often subjective and rely heavily on the expertise of clinicians. ML-based approaches have the potential to automate and improve the accuracy of the developed models [5] but their lack of interpretability limits their practical utility in clinical settings.

Several studies [6], [7] have explored the application of ML algorithms, such as neural networks and decision trees, in heart sound analysis for detecting abnormalities. The studies have demonstrated promising results in terms of accuracy, they often fall short in providing transparent explanations for their predictions. This limitation has led to growing interest in the development of explainable machine intelligence (XAI) models that can not only achieve high accuracy but also provide interpretable explanations for their decisions. In this work, we propose a novel approach to address the interpretability challenge in ML-based heart sound abnormality detection. Our approach involves leveraging state-of-the-art XAI techniques, such as SHapley additive exPlanations (SHAP) and local interpretable model agnostic explanations (LIME) [8] to generate explanations for the predictions of our ML models.

The convergence of engineering and medical expertise has rapidly advanced the health care domain. In this context, we review the progression of heart sound classification and disease diagnosis. The combination of deep convolutional neural networks and mel-frequency cepstral coefficients (MFCC) employed to classify normal and abnormal phonocardiography signals. The model has achieved a competitive score but did not address the generalizability of their approach to different subjects [9]. The automatic recognition of heart rate variations from phonocardiograms (PCG) using transfer learning with MFCC features has achieved high accuracy. However the limited availability of PCG recordings raises concerns about the generalizability of their method [10]. The heart sound classification using improved MFCC features and CRNN, provides an improvement in classification accuracy. However, the model suffers with computational complexity, need for further investigation into generalizability and other experiment conditions [11]. The reliance on traditional classifiers may limit their ability to capture complex patterns, and the evaluation based on the datasets [12]. The recurrent neural network (RNN) based framework provides promising results. However, the computational complexity of RNNs and their performance dependence on dataset used for the analysis [13], [14]. The segmented and unsegmented PCG signals have been explored for heart sound classification [15]. The heartbeat sound classification using normal, murmur, and extra-systole heartbeat sounds, achieves competitive performance [16]. The deep learning and ensemble learning techniques for phonocardiogram classification, achieves competitive performance [17]. The classification of normal and abnormal heart sounds using an ensemble approach along with handcrafted features may limit its ability to capture complex patterns [18].

Heart sound classification using deep learning based CNNs produce an improved classification accuracy [19]-[21]. The learnable filter banks is another approach for heart sound detection [22]. The interpretable machine learning techniques applied on medical image provides better explanations for the analysis [23], [24]. In this we have integrated the XAI techniques into the development of ML models for heart sound analysis. The proposed model has the potential to enhance the trust and adoption of ML technologies in cardiac diagnostics.

2. PROPOSED METHOD

The proposed methodology for heart sound abnormality detection using explainable machine learning (XML) models is shown in Figure 1. The method comprises a series of well-defined steps aimed at developing a robust and interpretable model. The methodology can be summarized as following steps:

- Step 1: Data collection: the data collection process involves gathering heart sound recordings for normal and abnormal conditions. These recordings are typically obtained from medical databases [25].
- Step 2: Data preprocessing: preprocessing begins with a thorough inspection of the collected data to identify and remove any outliers, artifacts, or corrupted recordings that could adversely affect model training. Noise reduction techniques, such as filtering or denoising algorithms, are applied to improve the quality of the recordings.
- Step 3: Feature extraction: feature extraction involves transforming raw heart sound recordings into a set of relevant features that can be used as a fetures while building the model. The features include spectral

characteristics and statistical measures that capture the underlying patterns in the heart sounds.
- Step 4: Dataset splitting: the dataset is split into training, validation, and testing sets using a stratified approach to ensure that each set contains a proportional representation of normal and abnormal heart sounds.
- Step 5: Feature scaling: feature scaling is applied to normalize the extracted features to a consistent range, preventing certain features from dominating the model training process due to their larger magnitudes.
- Step 6: Model training: multiple XML algorithms such as random forest, gradient boosting (GBoost), AdaBoost, K-nearest neighbors (KNN), and logistic regression (LR) are trained using the preprocessed and scaled features. The training process involves optimizing each algorithm's parameters through techniques like grid search or randomized search to find the best-performing configuration.
- Step 7: LIME explanation: LIME are applied to provide local explanations for individual predictions made by the XML model. LIME generates interpretable explanations by perturbing the input features around a specific instance and observing the changes in the model's predictions.
- Step 8: SHAP explanation: SHAP is used to provide global explanations for the XML model's predictions, highlighting the contributions of each feature to the overall predictions. SHAP values quantify the impact of each feature on the model's output and help understand the model's decision-making process at a global level.
- Step 9: Visualization: the results of LIME and SHAP explanations are visualized using various techniques such as bar plots, and summary plots. Visualization aids in presenting the explanations in an intuitive and understandable manner, facilitating their interpretation by domain experts and stakeholders.

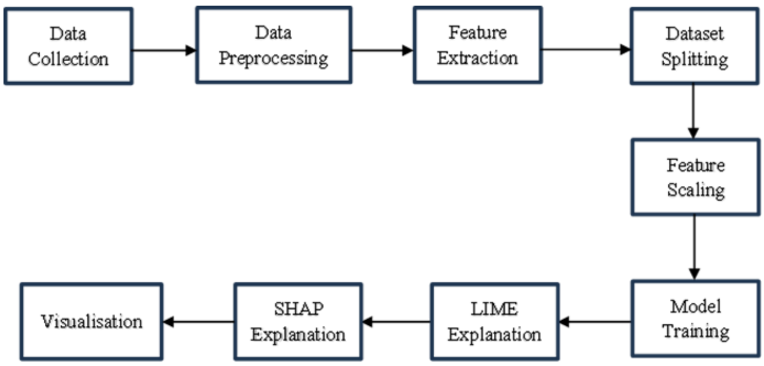

Figure 1. Proposed block diagram for heart sound classification

3. RESULTS AND DISCUSSION

The obtained results demonstrate that employing LIME and SHAP significantly enhances the interpretability of machine learning models used for detecting heart sound abnormalities. These techniques offer valuable insights into the decision-making processes of the models by pinpointing the key factors influencing their classifications. Through visualizing the impact of different variables, we have achieved a deeper understanding of the models' behavior, thereby enhancing transparency and interpretability

3.1. LIME explanations

Across the many classifiers used in our analysis, LIME consistently identified the most important variables for heart sound categorization. LIME is crucial in our analysis because it consistently identifies the most important variables for heart sound categorization across various classifiers. For example, in the random forest classifier, LIME explanations demonstrated how key spectral and statistical properties, such as spectral centroid, spectral bandwidth, and zero-crossing rate, were critical in discriminating between normal and pathological heart sounds.

3.1.1. Random forest classifier

Figure 2 illustrates LIME visualization using a random forest classifier, emphasizing features such as consistent spectral patterns and statistical properties that play a crucial role in the classification process. These include the regularity in spectral centroid and the low variability in zero-crossing rate. Figure 3 illustrates how LIME visualization reveals that irregularities in spectral patterns and statistical features, such as high variation in spectral centroid and pronounced changes in zero-crossing rate, are crucial factors influencing the classification.

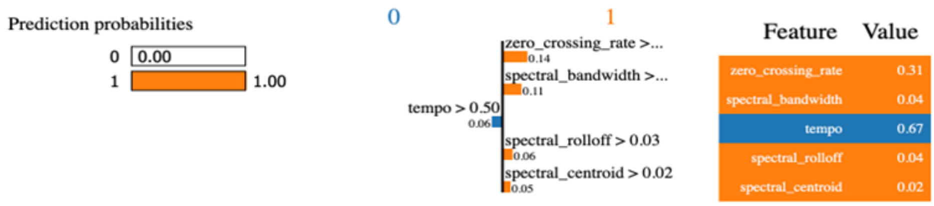

Figure 2. LIME visualization where the person has normal heartbeat using random forest

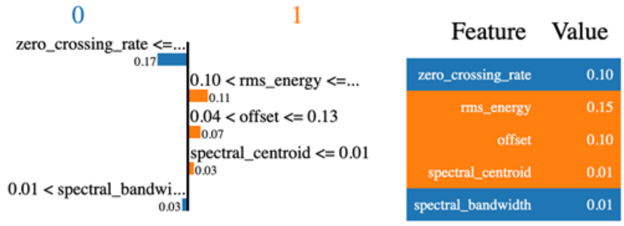

Figure 3. LIME visualization where the person has abnormal heartbeat using random forest

3.1.2. GBoost classifier

Figure 4 illustrates the LIME visualization with a GBoost classifier, elucidating how distinct spectral and statistical features contribute to a consistent and predictable pattern. This includes stable values in spectral centroid and smooth transitions in zero-crossing rate. Figure 5 illustrates how LIME visualization with a GBoost classifier can uncover irregularities in spectral and statistical features, such as abrupt shifts in spectral centroid and significant variability in zero-crossing rate, which play a crucial role in distinguishing abnormal heart sounds.

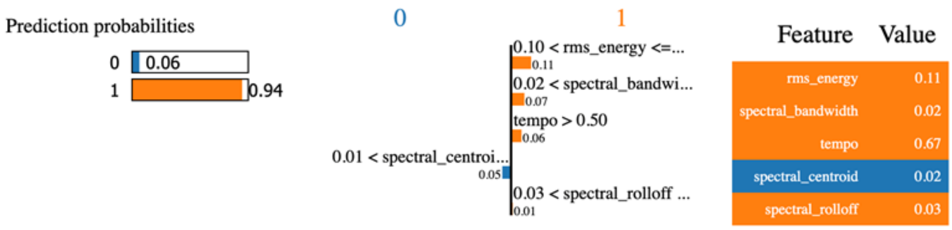

Figure 4. LIME visualization where the person has normal heartbeat using GBoost classifier

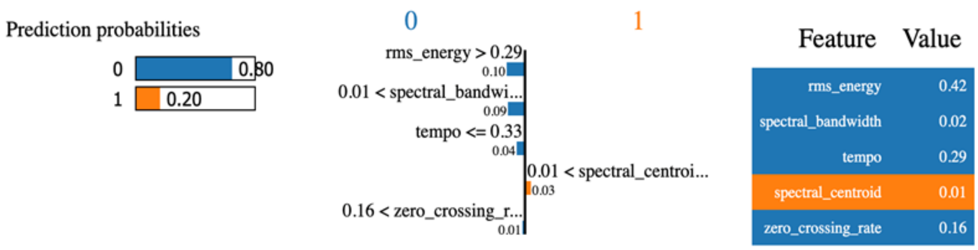

Figure 5. LIME visualization where the person has normal heartbeat using GBoost classifier

3.1.3. XGB classifier
Figure 6 illustrates how LIME visualization with an extreme gradient boosting (XGBoost) classifier might show how certain consistent spectral patterns and statistical features, like stable spectral centroid and minimal zero-crossing rate, contribute to the classification. Figure 7 illustrates how LIME visualization for abnormal heart sounds with an XGBoost classifier could reveal how irregularities in spectral patterns and statistical features, such as fluctuations in spectral centroid and high zero-crossing rate, play a crucial role in the classification. Table 1 shows the accuracies of the classifiers using LIME, with the XGB and GBoost achieving the highest accuracy for heart sound classification, followed by the random forest and decision tree. KNN also showed respectable accuracies but were slightly lower compared to the XGB and GBoost classifiers.

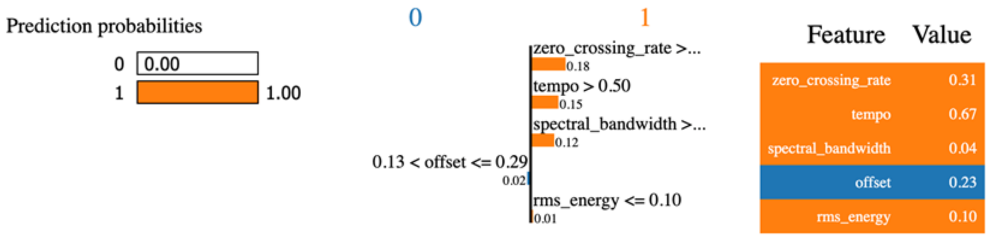

Figure 6. LIME visualization where the person has normal heartbeat using XGB classifier

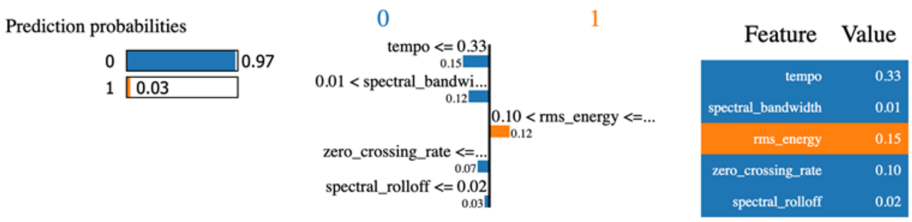

Figure 7. LIME visualization where the person has abnormal heartbeat using XGB classifier

Table 1. Classification accuracy of various classifiers for normal and abnormal heart sound detection

Classifier	Accuracy (%)
Random forest	88.2
GBoost	89.2
XGB	91.8
KNN	70
Decision tree	88.2

3.2. SHAP explanations
In our analysis, SHAP consistently identified key variables for heart sound categorization across various classifiers. For example, in the GBoost classifier, SHAP highlighted the importance of features such as spectral rolloff, RMS energy, and tempo in distinguishing normal from pathological heart sounds. These insights provided a clear understanding of feature contributions, enhancing model interpretability. In addition, SHAP analysis revealed the consistent impact of these features across different classifiers, reinforcing their significance in heart sound categorization and providing valuable insights for model refinement and clinical decision-making.

3.2.1. Random forest
In SHAP the features are ranked based on their importance in the model's predictions. For normal heart sounds, features with higher SHAP values, such as spectral centroid and spectral bandwidth, indicate a strong influence on the classification, highlighting their importance in distinguishing normal heart sounds. In a SHAP force plot for a random forest classifier, the plot shows how individual feature values contribute

to the model's prediction for a specific instance. For normal heart sounds, features like spectral centroid and spectral bandwidth with positive SHAP values indicate their contribution to classifying the sound as normal, providing a detailed understanding of the model's decision process. An explanation of SHAP for the random forest classifier can be seen in Figure 8.

Figure 8. SHAP explanations for random forest classifier

3.2.2. Gradient boosting

In the summary plot as shown in Figure 9, SHAP values are displayed horizontally, representing the impact of each feature on the model's output. Features are sorted by importance, with color indicating the value of the feature (red for high, blue for low). This plot provides a clear overview of which features are most influential in the model's decision-making process. In a force plot, SHAP explanations for individual predictions are visualized. For example, for normal heart sounds, the force plot might show how certain spectral and statistical features lead to a classification of normal, providing a detailed breakdown of the contribution of each feature to the final prediction. In comparison with [15] the proposed technique provides an alternate perspective on heart sound abnormality detection. While the preceding publication focused on time-frequency approaches and deep learning for representation learning, our research focuses on the construction of XAI models. We combine diverse machine learning algorithms with explainable AI approaches such as LIME and SHAP, resulting in improved accuracy and clear explanations for model predictions. The potential applications of ECG (Electrocardiography) data analysis in healthcare, wellness monitoring, and beyond are vast and promising. Likewise, explainable machine learning models are poised to be pivotal in guaranteeing transparency, trustworthiness, and accountability in AI-driven ECG analysis systems.

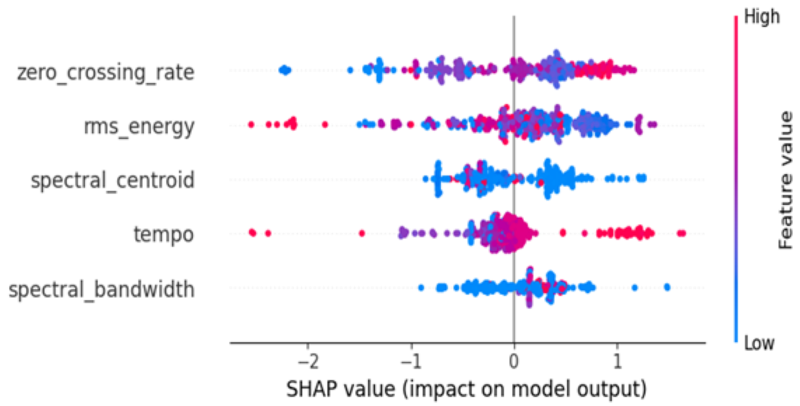

Figure 9. Summary plot of SHAP using GBoost

4. CONCLUSION

Our study successfully created machine intelligence models for heart sound anomaly identification utilising methods such as random forest, GBoost, XGBoost, and decision trees. The use of explainable AI approaches such as LIME and SHAP improved model interpretability, offering insights into predictions. These models present a viable method for automated and transparent heart sound analysis, with potential applications in early identification and monitoring of cardiac diseases. Future research might concentrate on increasing the dataset and conducting real world trials to confirm their beneficial effects. Overall, our effort highlights the utility of explainable machine intelligence models in cardiac diagnosis.

REFERENCES

[1] M. A. H. Islam, S. M. Shahriyar, M. J. Alam, M. Rahman, and M. R. K. R. Sarker, "Skin disease detection employing transfer learning approach- a fine-tune visual geometry group-19," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 31, no. 1, pp. 321–328, Jul. 2023, doi: 10.11591/ijeecs.v31.i1.pp321-328.

[2] R. Battur and J. Narayana, "Classification of medical X-ray images using supervised and unsupervised learning approaches," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 3, p. 1713, Jun. 2023, doi: 10.11591/ijeecs.v30.i3.pp1713-1721.

[3] B. J. Khadhim, Q. K. Kadhim, W. K. Shams, S. T. Ahmed, and W. A. Wahab Alsiadi, "Diagnose COVID-19 by using hybrid CNN-RNN for chest X-ray," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 29, no. 2, pp. 852–860, Feb. 2023, doi: 10.11591/ijeecs.v29.i2.pp852-860.

[4] N. Chinnamadha, R. Z. Ahmed, and K. Kalegowda, "Development of health monitoring system using smart intelligent device," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 3, p. 1381, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1381-1387.

[5] S. C. Venkateswarlu, S. R. Jeevakala, N. U. Kumar, P. Munaswamy, and D. Pendyala, "Emotion recognition from speech and text using long short-term memory," *Engineering, Technology and Applied Science Research*, vol. 13, no. 4, pp. 11166–11169, Aug. 2023, doi: 10.48084/etasr.6004.

[6] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2435–2445, Nov. 2019, doi: 10.1109/JBHI.2019.2894222.

[7] A. Rath, D. Mishra, G. Panda, and M. Pal, "Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal," *Biomedical Signal Processing and Control*, vol. 76, p. 103730, Jul. 2022, doi: 10.1016/j.bspc.2022.103730.

[8] A. Holzinger, A. Sarant, C. Molnar, P. Biecek, and W. Samek, "Explainable AI methods-a brief overview," in *International workshop on extending explainable AI beyond deep models and classifiers*, 2022, pp. 13–38, doi: 10.1007/978-3-031-04083-2 2.

[9] V. Maknickas and A. Maknickas, "Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiological Measurement*, vol. 38, no. 8, pp. 1671–1684, Jul. 2017, doi: 10.1088/1361-6579/aa7841.

[10] T. Alafif, M. Boulares, A. Barnawi, T. Alafif, H. Althobaiti, and A. Alferaidi, "Normal and abnormal heart rates recognition using transfer learning," in *Proceedings - 2020 12th International Conference on Knowledge and Systems Engineering, KSE 2020*, Nov. 2020, pp. 275–280, doi: 10.1109/KSE50997.2020.9287514.

[11] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Networks*, vol. 130, pp. 22–32, Oct. 2020, doi: 10.1016/j.neunet.2020.06.015.

[12] Z. Abduh, E. A. Nehary, M. A. Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers," *Biomedical Signal Processing and Control*, vol. 57, p. 101788, Mar. 2020, doi: 10.1016/j.bspc.2019.101788.

[13] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9393–9400, Nov. 2018, doi: 10.1109/JSEN.2018.2870759.

[14] T. C. I. Yang and H. Hsieh, "Classification of acoustic physiological signals based on deep learning neural networks with augmented features," *Computing in Cardiology*, vol. 43, pp. 569–572, 2016, doi: 10.22489/cinc.2016.163-228.

[15] F. A. Khan, A. Abid, A. Abid, M. S. Khan, and M. S. Khan, "Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features," *Physiological Measurement*, vol. 41, no. 5, p. 055006, Jun. 2020, doi: 10.1088/1361-6579/ab8770.

[16] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat sound signal classification using deep learning," *Sensors (Switzerland)*, vol. 19, no. 21, p. 4819, Nov. 2019, doi: 10.3390/s19214819.

[17] J. M. T. Wu *et al.*, "Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model," *Applied Soft Computing Journal*, vol. 78, pp. 29–40, May 2019, doi: 10.1016/j.asoc.2019.01.019.

[18] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *Computing in Cardiology*, Sep. 2016, vol. 43, pp. 621–624, doi: 10.22489/cinc.2016.182-399.

[19] F. Li et al., "Feature extraction and classification of heart sound using 1D convolutional neural networks," *Eurasip Journal on Advances in Signal Processing*, vol. 2019, no. 1, p. 59, Dec. 2019, doi: 10.1186/s13634-019-0651-3.

[20] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, "Classification of heart sounds using convolutional neural network," *Applied Sciences (Switzerland)*, vol. 10, no. 11, p. 3956, Jun. 2020, doi: 10.3390/app10113956.

[21] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, Jun. 2020, doi: 10.1016/j.neucom.2018.09.101.

[22] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2189–2198, Aug. 2020, doi: 10.1109/JBHI.2020.2970252.

[23] S. R. Jeevakala and H. Ramasangu, "Identification of significant instants of voxels for cognitive state classification using interpretable machine learning models," *Journal of Medicinal and Chemical Sciences*, vol. 6, no. 6, pp. 1291–1301, 2023, doi: 10.26655/JMCHEMSCI.2023.6.9.
[24] Z. Wang, K. Qian, H. Liu, B. Hu, B. W. Schuller, and Y. Yamamoto, "Exploring interpretable representations for heart sound abnormality detection," *Biomedical Signal Processing and Control*, vol. 82, p. 104569, Apr. 2023, doi: 10.1016/j.bspc.2023.104569.
[25] "PhysioNet Challenge2016:Heart sound classification," *Kaggle*, 2016. https://www.kaggle.com/datasets/bjoernjostein/physionet-challenge (accessed Aug. 05, 2022).

BIOGRAPHIES OF AUTHORS

Jeevakala Siva Ramakrishna is associate professor in the Department of CSE (Artificial Intelligence and Machine Learning), Institute of Aeronautical Engineering, Hyderabad, India. He Holds a Ph.D. degree in Electronics and Communication Engineering with specialization in biomedical signal, image processing and Machine Learning. His research areas include image and signal processing, medical image analysis, and FMRI data analysis. Dr. J. Siva Ramakrishna has over 14 years of experience in teaching, including 4 years as a researcher. He published several research papers in international journals and conferences. He can be contacted at email: j.sivaramakrishna@iare.ac.in.

Sonagiri China Venkateswarlu professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds a Ph.D. degree in Electronics and Communication Engineering with a specialization in Digital Speech Processing. Dr. S. China Venkateswarlu has more than 40 citations and paper publications across various publishing platforms, and expertise in teaching subjects such as microprocessors and microcontrollers, digital signal processing, digital image processing, and speech processing. With 20 years of teaching experience, he can be contacted at email: c.venkateswarlu@iare.ac.in.

Kommu Naveen Kumar is a professor in the Department of Electronics and Communication Engineering, MLRITM, Hyderabad, India. He holds a Ph.D. degree in Electronics and Communication Engineering with specialization in Digital Image Processing. His research areas include digital image processing, and embedded systems. Dr. K. Naveen Kumar has over 15 years of experience in teaching. He published paper in several national and international journals and conferences. He can be contacted at email: naveenkarunya@gmail.com.

Parikipandla Shreya is a student pursuing bachelor's degree in Electronics and Communication Engineering in Institute of Aeronautical engineering. She has strong interest in Artificial Intelligence and Machine Learning as well as cloud computing. She is very keen in research and always eager to learn and apply new concepts. She is determined to contribute to the advancements of technology and excited about the opportunities that lie ahead. She can be contacted at email: shreyaparikipandlla@gmail.com.