Transfer Learning and Deep Learning for Valvular Heart Disease Detection from Heart Sound Recordings

Saman Parvaneh, Zaniar Ardalan

Edwards Lifesciences, Irvine, California, USA

Abstract

Abnormal heart sounds may indicate Valvular Heart Disease (VHD), and a phonocardiogram (PCG) could serve as an automated screening tool in resource-limited areas. In this paper, we evaluated the feasibility of using a deep learning algorithm and transfer learning for the automated detection of VHD from PCG. A normal category (N), aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP) from a public data set were used in this study (200 audio files per category). Furthermore, 80% of the data was split into a training set, with the remaining 20% used for testing. This study employed YAMNet, a pre-trained deep convolutional neural network, for transfer learning. This model takes an audio waveform as input and makes independent predictions for 521 audio events. For transfer learning of YAMNet in VHD detection, the last three layers (fully connected, SoftMax, and classification layers) were removed and replaced with a 5-class classification layer. Our model achieved 100% and 99.5% accuracy in the training and test sets, respectively. In the test set, only one MR was misclassified as MS. This study's promising results indicate the potential of deep learning models for automated VHD detection using heart sounds. However, external validation on larger multicenter data is necessary.

1. Introduction

Valvular Heart Disease (VHD) involves damage or defects in one or more of the heart's four valves: the aortic, mitral, pulmonary, and tricuspid valves [1]. This damage leads to either a reduced flow (stenosis) or backward flow (regurgitation) of blood through the heart valves. The overall prevalence of VHD in the US adult population is estimated at 2.5% with aortic and mitral valvular disease as the two most diagnosed diseases [2]. VHD significantly impacts a patient's quality of life and is associated with high morbidity and mortality rates [1, 2].

The echocardiogram is a well-accepted diagnostic screening tool for detecting VHD [3]; however, its cost and long examination time limit it, and it should be performed by a certified echocardiographer [4]. Abnormal heart sounds often indicate VHD, and the phonocardiogram (PCG) enables accessible screening of heart sounds, particularly in resource-constrained environments where access to echocardiograms is limited. The presence of VHD is often indicated by abnormal heart sounds, known as murmurs, which can occur at different points in the cardiac cycle depending on the specific valve affected [5]. Heart auscultation is a fundamental clinical skill, and a systematic review of 23 articles found that auscultation sensitivity ranged from 30% to 100% and specificity ranged from 28% to 100% [6]. Davidsen et al. highlighted that the diagnostic accuracy of auscultation is highly dependent on clinical expertise [6].

Deep learning-based approaches have been successfully applied to the detection of normal/abnormal heart sounds [7, 8] and VHD [4, 9, 10]. This paper aims to assess the practicality of applying deep learning and transfer learning techniques for the automated detection of VHD using PCG signals.

2. Materials and Methods

2.1. Data

The dataset used in this study consists of PCG audio recordings representing five distinct heart sound categories: normal (N), aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP). Each category contains 200 audio files, totalling 1,000 samples. All recordings were sourced from a combination of publicly available databases, including resources from institutions such as Washington, Texas, 3M, and Michigan [11].

To ensure data quality and consistency, files with excessive noise were excluded following a cross-check procedure by the dataset providers. The remaining audio files were standardized by resampling to 8,000 Hz and converting to mono. Each recording captures three periods of heart sound signals, and preprocessing steps, including data sampling, conversion, and editing, were performed using Cool Edit software [11].

2.2. Methods

2.2.1. Deep Learning and Transfer Learning Approach

Deep learning, particularly convolutional neural networks (CNNs), has become a powerful tool for biomedical signal analysis, including heart sound classification [12, 13]. However, the effectiveness of deep learning models is often constrained by the size of available labeled datasets, as large datasets are typically required to achieve robust generalization and avoid overfitting [14]. In clinical research, data scarcity is a common challenge that can limit the performance of deep learning models.

To address this limitation, transfer learning is employed. Transfer learning enables a model pre-trained on a large, diverse dataset to be adapted to a specific task using a smaller dataset [15]. This approach leverages rich feature representations learned from extensive data, thereby improving model performance and generalizability in domains with limited labeled data.

For this study, we utilized YAMNet, a deep convolutional neural network based on the MobileNetV1 architecture [16]. YAMNet is pre-trained on the Google AudioSet dataset, which contains over 2.1 million human-annotated sound clips spanning 521 audio event categories, including heart sounds, respiratory sounds, coughing, snoring, and environmental noises [17]. This extensive pre-training enables the model to learn generalizable audio features applicable to biomedical signal analysis.

2.2.2. Transfer Learning Adaptation

To adapt YAMNet for automated detection of VHD from PCG recordings, transfer learning was performed by modifying the final layers of the pre-trained model. Specifically, the last three layers (fully connected, SoftMax, and classification layers) were replaced with a new classification layer tailored to five target classes: N, AS, MS, MR, and MVP. The architecture modification is illustrated in Figure 1. The model was then fine-tuned on a curated dataset of heart sound recordings, enabling it to adapt to the specific characteristics of cardiac sounds relevant to VHD detection.

2.2.3. Model Training

To ensure robust model evaluation and mitigate overfitting, stratified splitting was used to split the dataset into training and test sets, with 80% of samples from each category allocated to the training set and 20% reserved for the test set. Hyperparameters and training procedures were selected based on established practices in deep learning for biomedical signal processing.

2.3. Model Evaluation

Model evaluation was conducted to objectively assess the performance and generalizability of the deep learning algorithm for automated detection of VHD from PCG recordings. To evaluate the classification model, overall accuracy and sensitivity are used.

All preprocessing (resampling and feature extraction) and model development (training, fine-tuning, and evaluation) were conducted in MATLAB R2024a.

3. Results

The performance of the proposed deep learning model for automated detection of VHD from PCG recordings was evaluated on a curated dataset comprising 1,000 audio samples, evenly distributed across five diagnostic categories. Stratified sampling was employed to partition the dataset into training (80%, n= 800) and test (20%, n = 200) subsets, ensuring proportional representation of class distributions across both subsets.

Figure 2 shows the learning curve and loss function of the proposed model across 10 epochs. The model achieves rapid convergence within the first epoch, where accuracy quickly rises and the loss sharply decreases before both stabilize, indicating efficient learning and strong generalization.

The model achieved promising classification accuracy on the training set, correctly identifying all 800 cases. On the test set, the overall accuracy exceeded 99%. The only error observed in the test set was the misclassification of a single MR case as MS. No systematic misclassification patterns were identified. Sensitivity for AS and MCP was 100%, while sensitivity for MS and MR was 97.5%. These results indicate robust detection capability across the major valvular disease categories.

Table 1 compares the proposed method with several recent studies on multi-class classification of VHDs using the same dataset. The proposed approach, using mel-spectrogram features and transfer learning with YAMNet, achieved an overall accuracy of 99.5%, comparable to or exceeding that of other state-of-the-art methods.

4. Discussion

This study demonstrates the feasibility and effectiveness of a deep learning approach, specifically transfer learning with YAMNet, for automated detection of VHD from PCG recordings. The model achieved high

accuracy and sensitivity across all major VHD categories, with only a single misclassification observed in the test set. These findings underscore several important considerations for both clinical application and future research. The near-perfect classification performance observed in both the training and test sets suggests that deep learning models, when properly adapted and fine-tuned, can reliably distinguish between normal and pathological heart sounds. The use of transfer learning enabled the model to leverage rich feature representations learned from large-scale audio datasets, thereby overcoming the limitations imposed by the relatively small amount of labeled data. The model's high sensitivity in correctly identifying aortic stenosis, mitral stenosis, mitral regurgitation, and mitral valve prolapse is particularly notable, given the clinical importance of these conditions and the variability in heart sound presentations. A comparative analysis with recent studies in the field reveals that the proposed method matches or exceeds the performance of other state-of-the-art approaches for multi-class classification of VHDs. The overall accuracy and sensitivity metrics are consistent with, or superior to, those reported for models utilizing alternative architectures and feature extraction techniques. This supports the utility of mel-spectrogram features and transfer learning with YAMNet for this application.

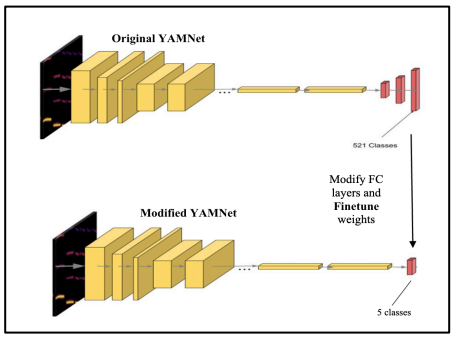

Figure 1. The YAMNet architecture modification for valvular heart disease detection

The results of this study have significant implications for clinical practice, particularly in resource-limited settings where access to echocardiography is restricted. Automated analysis of PCG recordings using deep learning could facilitate early detection and referral for echocardiographic evaluation, potentially improving patient outcomes. The model's robust performance across multiple disease categories suggests that such an approach could reduce reliance on expert clinical interpretation, which is known to be variable and experience-dependent. Furthermore, the learning curve and loss function demonstrated rapid convergence and stable optimization, indicating effective model training and the absence of overfitting. The observed fast convergence results from leveraging a pre-trained YAMNet architecture, which already encodes robust audio features from large-scale datasets, requiring only minor fine-tuning for heart sound classification.

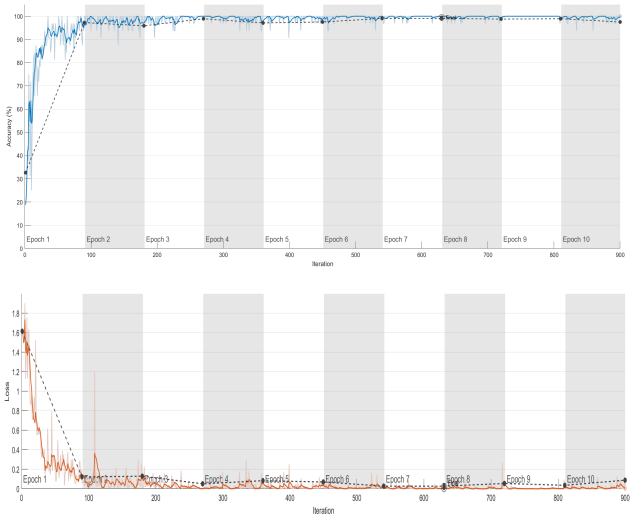

Figure 2. Learning curve (top) and loss function (bottom) for the proposed method across 900 epochs.

Table 1. Evaluation comparison of the proposed method with several recent methods on multi-class classification of valvular heart diseases (CWT: Continuous Wavelet Transform, VMD: Variational Mode Decomposition)

Study	Feature	Architecture	Overall Accuracy
Wang et al. [18]	CWT	Google Net + Transfer Learning	98%
Fakhry et al. [19]	VMD + logarithmic features	CNN-LSTM	98.7%
Divakar et al. [20]	Mel-spectrograms	CNN-LSTM	97.5%
Alotaibi et al. [21]	Log Mel-spectrograms	Whisper	99.3%
Proposed Method	Mel-spectrograms	YAMNet +Transfer Learning	**99.5%**

Despite the promising results, several limitations must be acknowledged. First, the dataset used in this study was clean, balanced, and curated; however, this may not fully reflect the variability and noise encountered in real-world clinical environments. Second, the sample size, while sufficient for initial feasibility assessment, remains modest compared to the diversity of heart sound presentations in the general population. Third, external validation on larger, multicenter datasets is necessary to confirm the generalizability and robustness of the proposed approach.

Future research should focus on validating the model in broader clinical contexts, including heterogeneous patient populations and recordings obtained under variable conditions. Integration of the algorithm into mobile or point-of-care devices could further enhance accessibility and impact. Additionally, expanding the model to detect a broader range of cardiac pathologies and incorporating multimodal data (e.g., patient demographics and clinical history) may improve diagnostic accuracy and utility.

5. Conclusion

In summary, this study provides strong evidence for the potential of deep learning and transfer learning in the automated detection of valvular heart disease from heart sound recordings. While further validation is required, the results support continued development and clinical translation of such technologies to improve cardiovascular screening and diagnosis.

References

[1] A. M. Small, K. E. Yutzey, B. A. Binstadt, K. Voigts Key, N. Bouatia-Naji *et al.*, "Unraveling the mechanisms of valvular heart disease to identify medical therapy targets: a scientific statement from the american heart association," *Circulation,* vol. 150, no. 6, pp. e109–e128, 2024, doi: doi:10.1161/CIR.0000000000001254.

[2] M. C. Tan, Y. H. Yeo, B. J. San, A. Suleiman, J. Z. Lee *et al.*, "Trends and disparities in valvular heart disease mortality in the united states from 1999 to 2020," *Journal of the American Heart Association,* vol. 13, no. 8, p. e030895, 2024, doi: doi:10.1161/JAHA.123.030895.

[3] S. Wen and T. Z. Naqvi, "Point-of-care Ultrasound in detection, severity and mechanism of significant valvular heart disease and clinical management," *Journal of Clinical Medicine,* vol. 12, no. 20, p. 6474, 2023. [Online]. Available: https://www.mdpi.com/2077-0383/12/20/6474.

[4] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio *et al.*, "Classification of heart sound signals using a novel deep WaveNet model," *Computer Methods and Programs in Biomedicine,* vol. 196, p. 105604, 2020.

[5] J. S. Coviello, *Auscultation skills: breath & heart sounds*. *Lippincott Williams & Wilkins*, 2013.

[6] A. H. Davidsen, S. Andersen, P. A. Halvorsen, H. Schirmer, E. Reierth, and H. Melbye, "Diagnostic accuracy of heart auscultation for detecting valve disease: a systematic review," *BMJ Open,* vol. 13, no. 3, p. e068121, 2023.

[7] S. Parvaneh, Z. Ardalan, J. Song, K. Vyas, and C. Potes, "Heart murmur detection using ensemble of deep learning classifiers for phonocardiograms recorded from multiple auscultation locations," in *2022 Computing in Cardiology (CinC)*, 2022, vol. 498: IEEE, pp. 1–4.

[8] 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 Conference (CinC), 2016*, 2016: IEEE, pp. 621–624.

[9] Z. Jiang, W. Song, Y. Yan, A. Li, Y. Shen *et al.*, "Automated valvular heart disease detection using heart sound with a deep learning algorithm," *IJC Heart & Vasculature,* vol. 51, p. 101368, 2024.

[10] A. Shokouhmand, H. Wen, S. Khan, J. A. Puma, A. Patel *et al.*, "Diagnosis of coexisting valvular heart diseases using image-to-sequence translation of contact microphone recordings," *IEEE Transactions on Biomedical Engineering,* vol. 70, no. 9, pp. 2540–2551, 2023.

[11] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Applied Sciences,* vol. 8, no. 12, p. 2344, 2018.

[12] A. Ameen, I. E. Fattoh, T. Abd El-Hafeez, and K. Ahmed, "Advances in ecg and pcg-based cardiovascular disease classification: a review of deep learning and machine learning methods," *Journal of Big Data,* vol. 11, no. 1, p. 159, 2024.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436–444, 2015.

[14] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo *et al.*, "A guide to deep learning in healthcare," *Nature Medicine,* vol. 25, no. 1, pp. 24–29, 2019.

[15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, no. 10, pp. 1345–1359, 2009.

[16] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen *et al.*, "CNN architectures for large-scale audio classification," presented at *the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[17] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence *et al.*, "Audio set: an ontology and human-labeled dataset for audio events," presented at *the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2017.

[18] M. Wang, B. Guo, Y. Hu, Z. Zhao, C. Liu, and H. Tang, "Transfer learning models for detecting six categories of phonocardiogram recordings," *Journal of Cardiovascular Development and Disease,* vol. 9, no. 3, p. 86, 2022.

[19] M. Fakhry and A. Gallardo-Antolín, "Variational mode decomposition and a light cnn-lstm model for classification of heart sound signals," in *IEEE EUROCON 2023-20th International Conference on Smart Technologies*, 2023: IEEE, pp. 295–300.

[20] C. Divakar, R. Harsha, K. Radha, D. V. Rao, N. Madhavi, and T. Bharadwaj, "Explainable AI for cnn-lstm network in pcg-based valvular heart disease diagnosis," in *2024 14th international Conference on Cloud Computing, Data Science & Engineering (confluence)*, 2024: IEEE, pp. 92–97.

[21] M. Alotaibi, Y. Bazi, M. M. Al Rahhal, N. Ammour, and M. Zuair, "Classification of heart sound signals with Whisper model," *Connection Science,* vol. 37, no. 1, p. 2449943, 2025.

Address for correspondence:

Saman Parvaneh
1 Edwards Way, Irvine, CA, USA
parvaneh@ieee.org