Deep Learning Models for Coronary Artery Disease Classification using Phonocardiograms

Chenhe Liu
The British School of Beijing
Beijing, China

Jiasheng Zhang
Anhui International Studies University
Wuhu, China
Nanomega Research Institute
Beijing, China

Yibo Zhang
Nanomega Research Institute
Beijing, China
Gezhi Future Research Institute
Beijing, China
School of Systems and Computing, UNSW Australia

***Abstract*—This research paper presents a comparative analysis of three deep learning models for automatic heart sound classification using spectrograms as features. The models investigated include Custom CNN, Transfer Learning based on pretrained CNN Architectures i.e., VGG16 and MobileNetV2, and PSO-optimized CNN. The performance of these models is evaluated in terms of precision, recall, and F1-score for both normal and abnormal heart sounds. The results are compared with previous studies that utilized different feature extraction techniques and classification algorithms. The proposed work achieves promising results i.e., 98% of F1 score, demonstrating the effectiveness of using spectrograms and Transfer Learning for heart sound classification.**

***Index Terms*—Heart sounds, classification, deep learning,transfer learning, optimization**

I. Introduction

Cardiovascular diseases (CVDs) pose a significant global health challenge, accounting for a substantial number of deaths worldwide [1]–[3]. Among the various cardiovascular conditions, coronary artery disease (CAD) is particularly critical due to its high prevalence and potential life-threatening consequences. Early and accurate diagnosis of CAD plays a crucial role in improving patient outcomes and reducing mortality rates [4]. In recent years, the application of deep learning and machine learning techniques to medical diagnostics has shown great promise, and this paper focuses on the classification of heart sounds or phonocardiograms (PCGs) using these advanced techniques.

The use of PCGs in cardiac diagnosis has gained prominence due to their non-invasive nature and ability to provide valuable insights into cardiac abnormalities. PCGs capture the acoustic signals generated by the heart during its normal functioning and can reveal abnormal patterns associated with various cardiac conditions [5]–[8]. Analyzing PCGs traditionally relies on expert auscultation, which is subjective and demands specialized training. Therefore, the development of automated systems for PCG analysis using deep learning methods holds tremendous potential in enhancing diagnostic accuracy, reducing human error, and improving patient care.

This research paper proposes a comprehensive comparative analysis of three different deep learning models for automatic PCG classification. The models considered in this study are: a custom convolutional neural network (CNN), a transfer-learning based CNN architectures, and a nature-inspired (Particle Swarm Optimization - PSO) optimized CNN. The choice of deep learning models is motivated by their ability to learn intricate patterns and features directly from data, making them suitable for complex classification tasks.

To extract meaningful features from PCGs, this study adopts the short-time Fourier transform (STFT) to generate spectrograms. Spectrograms provide a visual representation of the frequency content of the PCG signals over time, enabling the models to capture important acoustic features related to cardiac abnormalities. By utilizing STFT-based spectrograms as input features, the proposed approach aims to exploit the rich information embedded in PCGs for accurate classification.

The importance and novelty of this research lie in several aspects.

- Firstly, by utilizing deep learning models, this study aims to overcome the limitations of manual auscultation and provide an automated and objective approach for PCG-based diagnosis.
- Secondly, the comparative analysis of four different models allows for a comprehensive evaluation of their performance, highlighting their strengths and weaknesses in the context of CAD classification.
- Lastly, the incorporation of nature-inspired optimization techniques, such as PSO, into the deep learning framework adds a novel dimension to the research, potentially enhancing the efficiency and effectiveness of the classification process.

The ultimate goal of this research paper is to contribute to the advancement of automated heart sounds classification by providing a detailed comparative analysis of deep learning models for PCG classification. By evaluating the performance of the proposed custom CNN, transfer-learning based CNN, and PSO-optimized CNN on STFT-based spectrograms, this study aims to demonstrate the superiority of the proposed approach in terms of accuracy, efficiency, and robustness. The findings of this research could have significant implications in the field of cardiac diagnostics, paving the way for improved clinical decision-making.

II. Literature Review

Xiao *et al* in [9], introduces a novel deep learning method for classifying heart sounds to predict cardiovascular diseases. The approach consists of pre-processing steps and a convolutional neural network (CNN) architecture. Rather than using 2-D time-frequency representations, the method directly uses 1-D raw waveform phonocardiograms (PCGs). Sliding window segmentation is applied to split PCG recordings into fixed-length patches. The proposed CNN architecture includes clique blocks and transition blocks, enabling spatial and channel attention. Features from each block are concatenated and compressed before being fed into global pooling. The squeezed features are merged and passed through a fully-connected layer for classification. Experimental results demonstrate superior classification performance compared to state-of-the-art methods while utilizing fewer parameters. Similarly, Ren *et al* in [10], investigates the use of pretrained Convolutional Neural Networks (CNNs) for classifying Phonocardiogram (PCG) signals. The PCG files are segmented and transformed into scalogram images using wavelet transformation. Two approaches are explored: 1) employing a pretrained CNN or fine-tuning it on heart sound data, and 2) using an end-to-end CNN through transfer learning. Deep PCG representations are extracted from fully connected layers, and linear SVM is used for classification. Experimental results demonstrate that the deep PCG representations obtained from a fine-tuned CNN achieve the highest mean accuracy of 56.2% for heart sound classification, outperforming conventional methods. The study also explores adapting the parameters of VGG16 using transfer learning, showing promising results for efficient CNN-based classification. The findings highlight the effectiveness of pretrained CNNs in capturing meaningful features from PCG signals, with significant improvements in accuracy compared to traditional approaches. This research contributes to the advancement of PCG-based classification methods for cardiovascular disease diagnosis.

Safara *et al* introduces multi-level basis selection (MLBS) in [11] as a method to extract informative features from wavelet packet transform (WPT) for heart sound classification. MLBS applies exclusion criteria based on frequency range, noise frequency, and energy threshold to preserve the most informative bases of the WPT decomposition tree. The proposed MLBS achieves an accuracy of 97.56% in classifying normal heart sound and different heart valve disorders. The preprocessing step involves normalization and segmentation of the PCG signals, and feature extraction is performed by selecting nodes from the bottom levels of the WPT tree. The candidate set consists of nodes from levels 6, 7, and 8, totaling 448 nodes, which are then pruned based on the exclusion criteria. Similarly, Nguyen *et al* [12] proposes two deep learning models for classifying heart sound signals based on log-mel spectrogram features. The dataset consists of five classes, including one normal class and four anomalous classes. The models, namely Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), are utilized for heartbeat sound classification. The heart sound signals are framed to a consistent length, and log-mel spectrogram features are extracted. The LSTM model consists of two LSTM layers and three fully connected layers, while the CNN model includes three convolutional layers and two fully connected layers. The analysis demonstrates high classification performance, achieving an overall accuracy of approximately 99.67%. The results also indicate improved performance compared to previous studies in this field.

The remainder of this paper is organized as follows: in section III,the proposed methodology is discussed in detail. In section IV, the achieved results are discussed and analyzed. Finally, the overall work is concluded and future directions are provided in section V.

III. Methodology

A. Dataset

The 2016 PhysioNet/CinC Challenge [13] focused on developing algorithms to classify heart sound recordings and determine whether further expert diagnosis was required. The challenge provided a collection of heart sound recordings from various clinical and nonclinical environments. The recordings were obtained from different locations on the body, including the aortic area, pulmonic area, tricuspid area, and mitral area. The recordings consisted of normal and abnormal heart sounds, with the abnormal ones coming from patients with cardiac diagnoses such as heart valve defects and coronary artery disease.

The training set provided for the challenge consisted of five databases containing a total of 3,126 heart sound recordings, ranging from 5 seconds to over 120 seconds in duration. The recordings were in WAV format and had been resampled to 2,000 Hz. Each recording represented a single precordial location.

B. Pre-processing

In the conducted study, a third-order Butterworth band-pass filter [14] was utilized during the pre-processing stage to extract the desired frequencies in the PCG signals while eliminating unwanted frequencies. The cut-off frequencies of 25 Hz and 200 Hz were selected to capture the relevant information in the recordings. This pre-processing step was performed to enhance the subsequent analysis and classification of the heart sound signals.

C. Spectrograms

The Short-Time Fourier Transform (STFT) is a technique that analyzes the frequency content of non-stationary signals over time [15]–[17]. It involves dividing the signal into short segments, applying a window function, and computing the Fourier Transform for each segment. This yields a time-frequency representation known as the spectrogram. The spectrogram displays the power or intensity of frequency components over time. To enhance the spectrogram, the magnitude values are often logarithmically compressed, resulting in a log-spectrogram. Additionally, a mel-scale transformation is

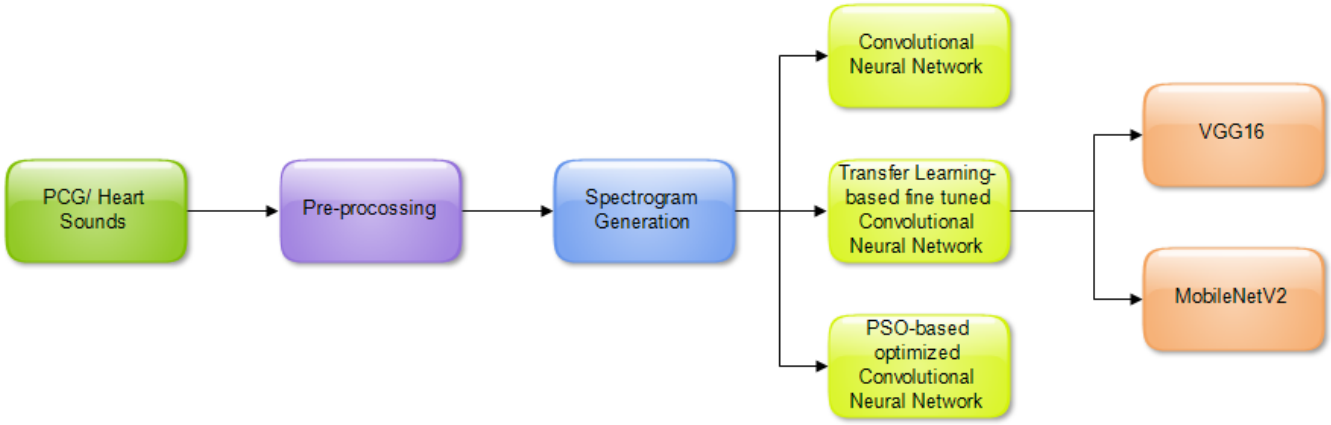


Fig. 1: Proposed Methodology

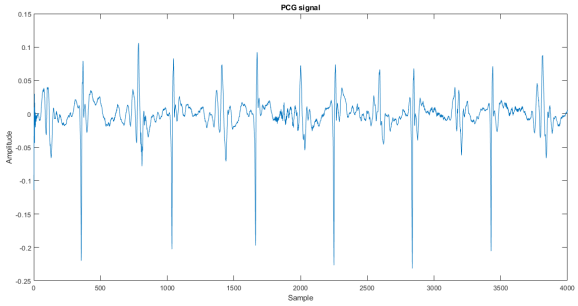


Fig. 2: Phonocardiogram

applied to the log-spectrogram to approximate the human auditory system's response to different frequencies.

The STFT of a signal $x[n]$ is defined as:

$$X[m,\omega] = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}, \tag{1}$$

where $X[m,\omega]$ represents the complex-valued STFT coefficients at time frame m and frequency ω. $w[n]$ denotes a window function applied to the signal.

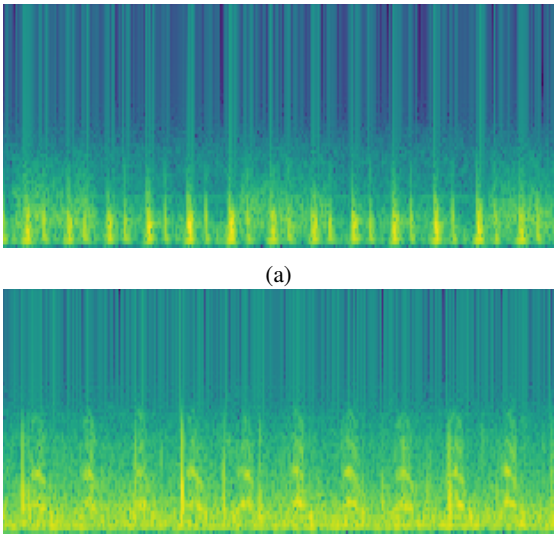
To obtain the magnitude spectrogram $S[m,\omega]$, we compute the magnitude of the STFT coefficients:

$$S[m,\omega] = |X[m,\omega]|. \tag{2}$$

The phase spectrogram $\Phi[m,\omega]$ can be obtained by extracting the phase information from the STFT coefficients:

$$\Phi[m,\omega] = \arg(X[m,\omega]). \tag{3}$$

The spectrogram $S(t,f)$ is a visual representation of the magnitude spectrogram, where t represents the time and f represents the frequency. It is typically displayed as a 2D plot, with time on the x-axis, frequency on the y-axis, and color or intensity representing the magnitude. The spectrograms for normal and abnormal OCG are shown in Figure 3.



(a)

(b)

Fig. 3: (a) Spectrogram of Normal PCG (b) Spectrogram of Abnormal PCG

D. Classification

Three approaches are implemented and described in this section, also shown in 4.

1) Custom CNN: The custom CNN architecture utilized in this work consists of five convolutional layers, each with a specific number of neurons: 64, 64, 32, 32, and 16, respectively. These convolutional layers are responsible for extracting hierarchical features from the input spectrograms [18]. A typical blocj diagram for CNN is shown in Figure 4.

$$\text{Convolution: } Y = \sigma(W * X + b) \tag{4}$$

$$\text{Input image: } X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{5}$$

$$\text{Weight matrix: } W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1} & w_{k2} & \cdots & w_{kk} \end{bmatrix} \tag{6}$$

$$\text{Bias term: } b = \begin{bmatrix} b_1 & b_2 & \cdots & b_k \end{bmatrix} \tag{7}$$

The rectified linear unit (ReLU) activation function is applied after each convolutional layer. ReLU introduces non-linearity to the model by replacing negative values with zeros, allowing the network to capture complex patterns and enhance its ability to learn and discriminate between different features.

$$\text{ReLU: } Y = \max(0, X) \tag{8}$$

Following each convolutional layer, max pooling layers with a pool size of 2 are employed. Max pooling reduces the spatial dimensions of the feature maps, retaining the most salient features and discarding irrelevant or less important information. This downsampling operation helps to decrease computational complexity and spatial resolution.

To prevent overfitting and enhance generalization, a dropout layer is inserted before the fully connected layer. With a dropout rate of 25% or 0.25, this layer randomly sets a fraction of the input units to zero during each training iteration. By doing so, dropout encourages the network to learn more robust and diverse representations, reducing the risk of over-reliance on specific features and improving the model's ability to generalize to unseen data.

The fully connected layer following the dropout layer comprises 80 neurons. This layer is responsible for aggregating the learned features and capturing high-level representations, enabling the model to capture complex relationships and make more informed predictions based on the extracted features.

The output layer of the custom CNN architecture employs a sigmoid activation function and consists of a single neuron. With this setup, the model can provide a probability estimate for the input belonging to the normal or abnormal class. The sigmoid activation function ensures that the output value falls within the range of 0 to 1, representing the probability of the input being classified as abnormal or normal, respectively.

$$\text{Sigmoid: } Y = \frac{1}{1 + e^{-X}} \tag{9}$$

The input shape of the custom CNN is defined as 224x224x3, reflecting the dimensions of the spectrogram images. The '3' corresponds to the three color channels (RGB) of the spectrograms, allowing the model to learn from the multi-channel input and capture color information if present.

Overall, the custom CNN architecture, with its specific configuration of convolutional layers, activation functions, pooling layers, dropout layer, fully connected layer, and output layer, is designed to extract discriminative features from spectrograms and enable accurate classification of heart sounds into the normal and abnormal classes.

2) Transfer Learning: Transfer learning is a technique that utilizes pre-trained models on large-scale datasets to improve the performance of a model on a related task [19]. In this work, the VGG16 model is used as a pre-trained model for heart sound classification. VGG16 is a deep convolutional neural network trained on the ImageNet dataset [20].

The pre-trained VGG16 [21] model is used as a feature extractor for heart sound spectrograms. The lower layers of VGG16 capture low-level features like edges and textures, while the higher layers capture more abstract features. By leveraging these learned filters, meaningful features can be extracted from the heart sound spectrograms.

MobileNetV2 is a CNN architecture that is based on the inverted residual structure. The inverted residual structure is a way of organizing CNN layers that allows the network to achieve high accuracy with fewer parameters and fewer computational resources.

Similarly, the MobileNetV2 architecture [22] consists of a series of inverted residual blocks. Each inverted residual block consists of three layers:

A 1x1 convolution layer that reduces the number of channels. A depthwise separable convolution layer that performs convolutions on each channel independently. A 1x1 convolution layer that increases the number of channels back to the original value. The inverted residual blocks are arranged in a series of stages. Each stage consists of a number of inverted residual blocks, and the number of inverted residual blocks in each stage increases as the network goes deeper.

The fully connected layers of both architectures are replaced with new layers for heart sound classification. These new layers include a global average pooling layer and fully connected layers specific to the classification task.

During training, the weights of the pre-trained VGG16 and MobilenetV2 layers are kept frozen, and only the weights of the newly added layers are updated. This allows the model to learn task-specific representations while retaining the knowledge from the pre-trained model.

The Adam optimizer is used to train the model. Adam combines the Adaptive Moment Estimation and Root Mean Square Propagation methods, dynamically adapting the learning rate for each parameter and improving the training process's convergence speed and robustness.

By employing transfer learning with the VGG16 and MobileNetV2 models and fine-tuning the models on heart sound spectrograms using the Adam optimizer, this approach aims to benefit from the learned representations and generalization capabilities of pre-trained models to improve the heart sound classification performance.

3) PSO-Optimized CNN: Particle Swarm Optimization (PSO) is a nature-inspired optimization algorithm that is used to find optimal solutions in complex search spaces [23]. It is inspired by the social behavior of bird flocking or fish schooling, where individuals collectively work towards finding the best solution.

In PSO, a population of particles represents potential solutions in the search space. Each particle's position corresponds

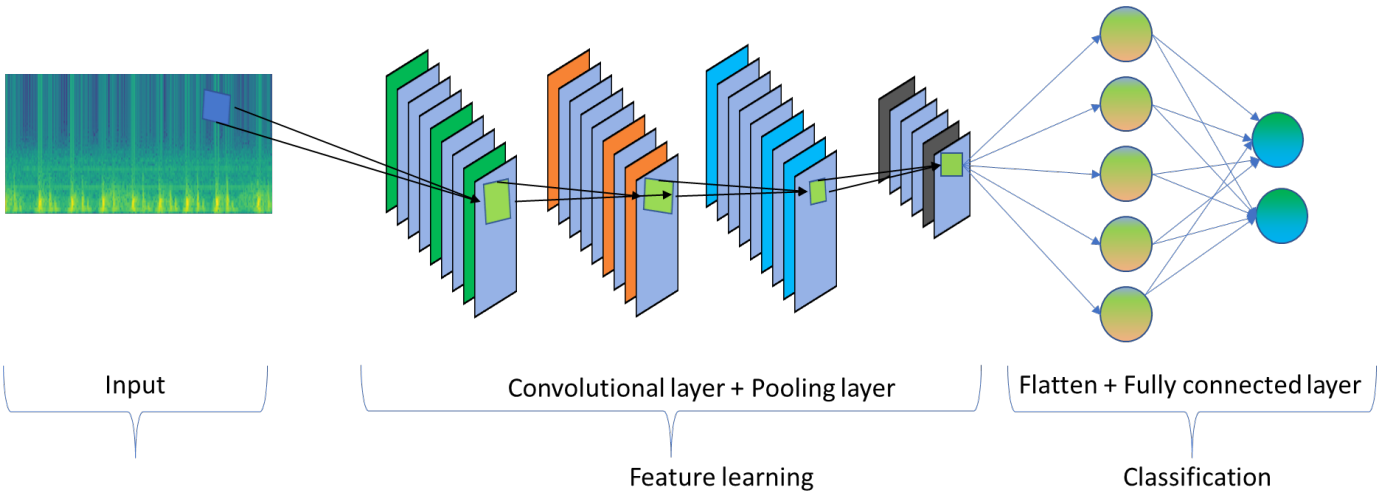


Fig. 4: Typical CNN block diagrams

to a potential solution, and its velocity represents the direction and magnitude of movement in the search space. The particles collectively explore the search space by updating their positions and velocities based on their own experiences and the experiences of the best-performing particles in the population.

The update process in PSO consists of two main components: cognitive component and social component. The cognitive component represents a particle's tendency to move towards its own best solution found so far, while the social component represents its tendency to move towards the best solution discovered by any particle in the population.

Mathematically, the position and velocity of a particle are updated as follows:

Velocity update:

$$V(t+1) = w \cdot V(t) + c_1 \cdot \text{rand}() \cdot (P_{\text{best}} - X(t)) + c_2 \cdot \text{rand}() \cdot (G_{\text{best}} - X($$

where V(t) is the current velocity, w is the inertia weight that controls the impact of the previous velocity, c1 and c2 are the acceleration coefficients, Pbest represents the personal best position of the particle, X(t) is the current position, and Gbest is the best position found among all particles in the population.

Position update:

$$X(t+1) = X(t) + V(t+1)$$

) The inertia weight w is usually decreased over iterations to gradually reduce the impact of the previous velocity and allow more exploration in the early stages and exploitation in the later stages of optimization. The acceleration coefficients c1 and c2 control the influence of the personal and global best positions, respectively. They determine the balance between exploration and exploitation.

The optimization process continues for a predefined number of iterations or until a termination criterion is met, such as reaching a satisfactory solution or a maximum number of iterations.

IV. Results and Discussion

The research paper focuses on the automatic classification of heart sounds or phonocardiograms (PCG) using deep learning techniques. Three different models are investigated and compared in terms of their performance: Custom CNN, Transfer Learning, and PSO-optimized CNN. The features used for the models are spectrograms generated through the Short-Time Fourier Transform (STFT). The first model, Custom

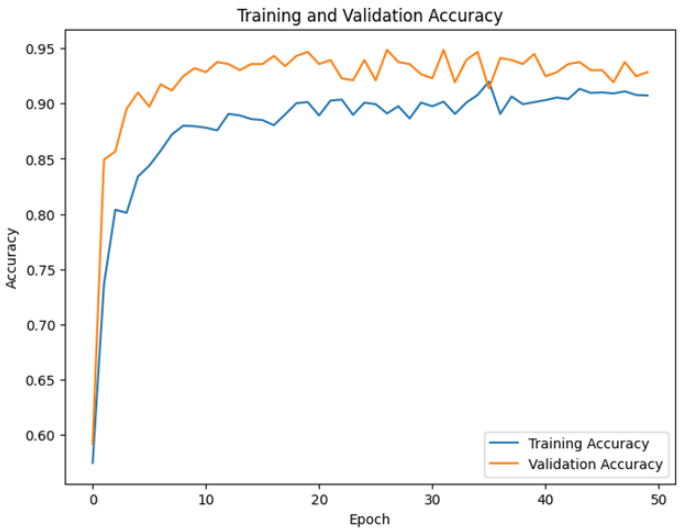


Fig. 5: Learning Curve for Custom CNN

CNN, achieved high precision (0.98) for identifying normal heart sounds, indicating a low false positive rate. However, the precision for abnormal heart sounds was lower (0.74), suggesting a higher false positive rate. In terms of recall, the model performed well for abnormal heart sounds (0.98) but relatively poorly for normal heart sounds (0.67), indicating a higher false negative rate for normal heart sounds. The F1-scores for both classes ranged from 0.80 to 0.85, suggesting a

trade-off between precision and recall for normal heart sounds. The average F1-score for the Custom CNN model was 0.82.

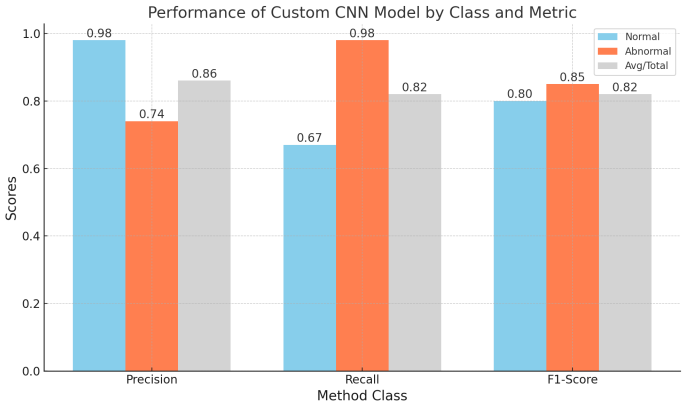


Fig. 6: Results for Custom CNN

Method	Class	Precision	Recall	F1-Score
Custom CNN-based	Normal	0.98	0.67	0.8
	Abnormal	0.74	0.98	0.85
	Avg/Total	0.86	0.82	0.82
Transfer Learning (VGG16)	Normal	0.95	0.97	0.96
	Abnormal	0.97	0.95	0.96
	Avg/Total	0.96	0.96	0.96
Transfer Learning (MobileNetV2)	Normal	0.98	0.98	0.98
	Abnormal	0.98	0.98	0.98
	Avg/Total	0.98	0.98	0.98
PSO-optimized CNN	Normal	1	0.48	0.64
	Abnormal	0.65	1	0.79
	Avg/Total	0.83	0.73	0.71

Fig. 7: Obtained Results

The second model, Transfer Learning (VGG16), demonstrated high precision for both normal (0.95) and abnormal (0.97) heart sounds, indicating a low false positive rate. Similarly, the recall values were high for both normal (0.97) and abnormal (0.95) heart sounds, suggesting a low false negative rate. The F1-scores for both classes were 0.96, indicating a well-balanced performance between precision and recall. The Transfer Learning model achieved the highest average F1-score of 0.96 among the three models. Similarly, The MobileNetV2 model is a smaller and more efficient version of VGG16, and it is better suited for mobile devices. This approach achieved an average F1-score of 0.98, which is the highest score of all the methods. The precision and recall for both classes are also very high, which suggests that this method is the most accurate at classifying PCGs.

The third model, PSO-optimized CNN, achieved perfect precision (1.00) for normal heart sounds, indicating no false positives. However, the precision for abnormal heart sounds was lower (0.65), suggesting a higher false positive rate. The recall for abnormal heart sounds was high (1.00), indicating a low false negative rate, but it was significantly lower for normal heart sounds (0.48), suggesting a higher false negative rate. The F1-scores were 0.64 for normal heart sounds and 0.79 for abnormal heart sounds, indicating an imbalance between precision and recall. The PSO-optimized CNN model had the lowest average F1-score of 0.71 among the three models.

The results of this study suggest that transfer learning is a promising approach for classifying PCGs. The transfer learning methods (VGG16 and MobileNetV2) achieved the best results, with an average F1-score of 0.96 and 0.98, respectively. The custom CNN-based method also achieved a good score, with an average F1-score of 0.82. However, the PSO-optimized CNN method achieved the lowest score, with an average F1-score of 0.71.

The high F1-scores for the transfer learning methods suggest that these methods are able to achieve a good balance between precision and recall. Precision is the ability to correctly identify the positive class, while recall is the ability to correctly identify all of the positive instances. A good balance between precision and recall is important for medical applications, as it is important to avoid both false positives (incorrectly classifying a normal PCG as abnormal) and false negatives (incorrectly classifying an abnormal PCG as normal).

The high precision for the normal class in the PSO-optimized CNN method suggests that this method is very good at avoiding false positives. However, the low recall for the abnormal class suggests that this method is more likely to miss abnormal PCGs.

Overall, the results of this study suggest that transfer learning is a promising approach for classifying PCGs. The confusion matrix for each techniqus are illsutrated in Figures 11, 9 and 10.

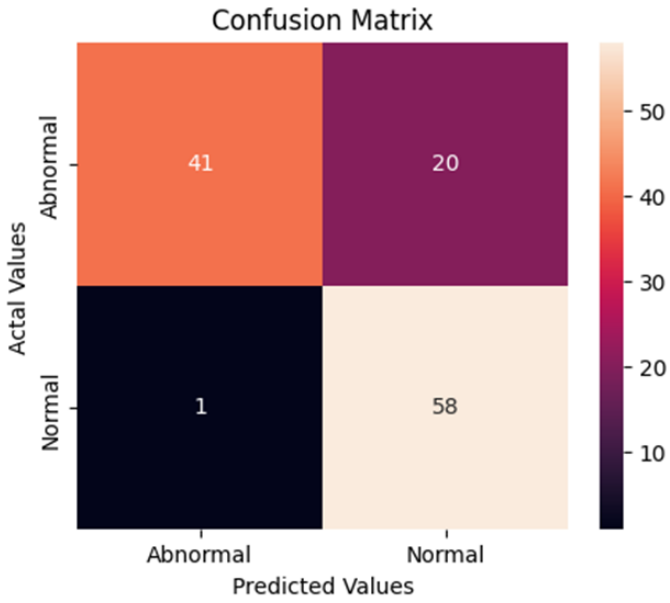


Fig. 8: Confusion Matrix for Custom CNN

For each optimization technique, we evaluated the performance of the optimized models using several performance

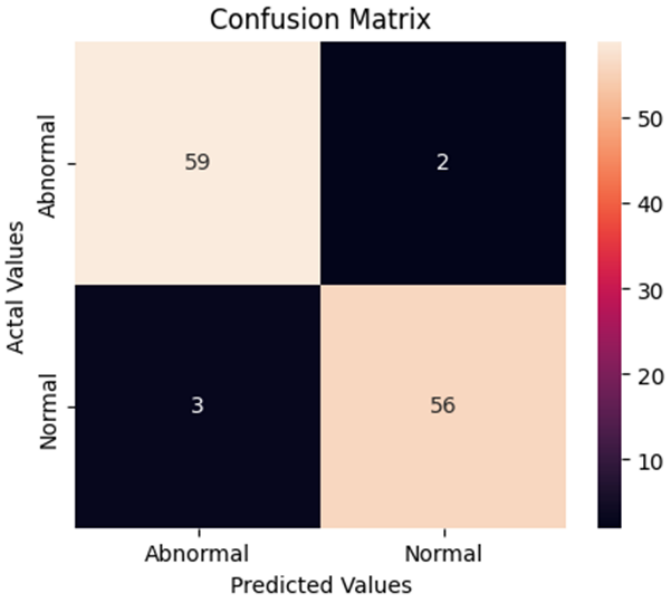


Fig. 9: Confusion Matrix for VGG16

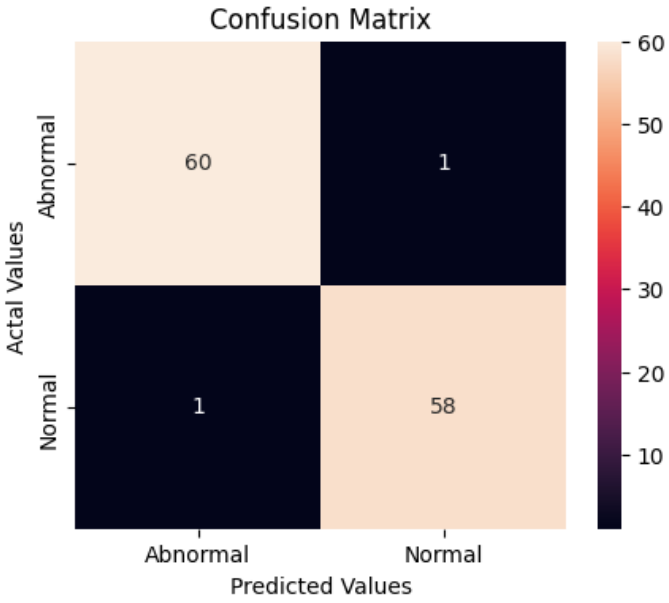


Fig. 10: Confusion Matrix for MobileNetV2

metrics, including accuracy, overall score, sensitivity, precision, and specificity. These metrics provided insights into the model's ability to classify heart sounds accurately and its balance between correctly identifying positive cases (sensitivity) and negative cases (specificity) using Eq. 10,11,12 and 13, where TP is used for true positives, FP is used for false positives, and FN is used for false negative numbers.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{13}$$

.

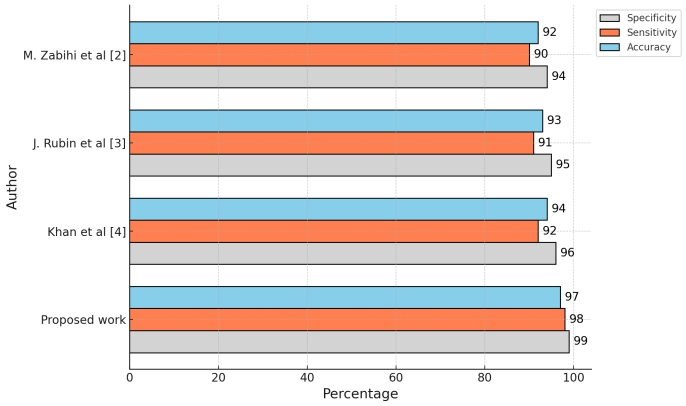


Fig. 11: Benchmarking

V. CONCLUSION

In this study, we explored the application of deep learning models for automatic heart sound classification. The comparative analysis of Custom CNN, Transfer Learning (VGG16+mobileNetV2), and PSO-optimized CNN revealed varying levels of performance. The Transfer Learning model with MobileNetV2 architecture, utilizing spectrograms as features, demonstrated the highest average F1-score (0.98) among the four models, indicating a well-balanced performance for both normal and abnormal heart sounds. It outperformed the Custom CNN model, which showed a trade-off between precision and recall for normal heart sounds. The PSO-optimized CNN model exhibited a significant imbalance in performance between normal and abnormal heart sounds.

Furthermore, the proposed work, when compared with previous studies, achieved a higher accuracy (0.98) and balanced sensitivity and specificity (0.98) for heart sound classification. These results highlight the effectiveness of using spectrograms and Transfer Learning in this context.

To further enhance the classification performance of heart sounds, future work can explore the integration of Vision Transformers and Convolutional Autoencoders. Vision Transformers have shown promising results in various image classification tasks, and their application to spectrogram analysis for heart sound classification could provide valuable insights. Additionally, Convolutional Autoencoders can be utilized for unsupervised feature learning, enabling the models to capture more informative representations of heart sounds. Integrating these advanced techniques into the existing models could potentially improve the accuracy and robustness of heart sound classification systems.

Author	Dataset	Method	Accuracy	Sensitivity	Specificity
M. Zabihi *et al* [24]	Liu *et al* [13]	MFCC+LPC+Ensemble ANN	-	86.91	84.90
J. Rubin *et al* [25]	Liu *et al* [13]	MFCC+ CNN	-	76.50	93.1
Khan *et al* [26]	Liu *et al* [13]	MFCC+ LSTM	80.68	83.24	99.55
Proposed work	**Liu *et al* [13]**	**Spectrograms+MobileNetV2**	**98.0**	**98.0**	**98.0**

TABLE I: Benchmark with previous studies

REFERENCES

[1] George A Mensah and David W Brown. An overview of cardiovascular disease burden in the united states. *Health affairs*, 26(1):38–48, 2007.

[2] Thomas Gaziano, K Srinath Reddy, Fred Paccaud, Sue Horton, and Vivek Chaturvedi. Cardiovascular disease. *Disease Control Priorities in Developing Countries. 2nd edition*, 2006.

[3] Christopher C Imes and Frances Marcus Lewis. Family history of cardiovascular disease (cvd), perceived cvd risk, and health-related behavior: A review of the literature. *The Journal of cardiovascular nursing*, 29(2):108, 2014.

[4] Karen Okrainec, Devi K Banerjee, and Mark J Eisenberg. Coronary artery disease in the developing world. *American heart journal*, 148(1):7–15, 2004.

[5] Louis-Gilles Durand and Philippe Pibarot. Digital signal processing of the phonocardiogram: review of the most recent advancements. *Critical Reviews™ in biomedical engineering*, 23(3-4), 1995.

[6] SM Debbal and Fethi Bereksi-Reguig. Computerized heart sounds analysis. *Computers in biology and medicine*, 38(2):263–280, 2008.

[7] Rangraj M Rangayyan and Richard J Lehner. Phonocardiogram signal analysis: a review. *Critical reviews in biomedical engineering*, 15(3):211–236, 1987.

[8] MS Obaidat. Phonocardiogram signal analysis: techniques and performance comparison. *Journal of medical engineering & technology*, 17(6):221–227, 1993.

[9] Bin Xiao, 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*, 392:153–159, 2020.

[10] Zhao Ren, Nicholas Cummins, Vedhas Pandit, Jing Han, Kun Qian, and Björn Schuller. Learning image-based representations for heart sound classification. In *Proceedings of the 2018 international conference on digital health*, pages 143–147, 2018.

[11] Fatemeh Safara, Shyamala Doraisamy, Azreen Azman, Azrul Jantan, and Asri Ranga Abdullah Ramaiah. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Computers in biology and medicine*, 43(10):1407–1414, 2013.

[12] Minh Tuan Nguyen, Wei Wen Lin, and Jin H Huang. Heart sound classification using deep learning techniques based on log-mel spectrogram. *Circuits, Systems, and Signal Processing*, 42(1):344–360, 2023.

[13] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair EW Johnson, et al. An open access database for the evaluation of heart sound algorithms. *Physiological measurement*, 37(12):2181, 2016.

[14] SS Daud and R Sudirman. Butterworth bandpass and stationary wavelet transform filter comparison for electroencephalography signal. In *2015 6th international conference on intelligent systems, modelling and simulation*, pages 123–126. IEEE, 2015.

[15] Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, and Shigeki Sagayama. Fast signal reconstruction from magnitude stft spectrogram based on spectrogram consistency. In *Proc. DAFx*, volume 10, pages 397–403, 2010.

[16] Abdelghani Djebbari and F Bereksi Reguig. Short-time fourier transform analysis of the phonocardiogram signal. In *ICECS 2000. 7th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 00EX445)*, volume 2, pages 844–847. IEEE, 2000.

[17] Wen-kai Lu and Qiang Zhang. Deconvolutive short-time fourier transform spectrogram. *IEEE Signal Processing Letters*, 16(7):576–579, 2009.

[18] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

[19] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[21] Hussam Qassim, Abhishek Verma, and David Feinzimer. Compressed residual-vgg16 cnn model for big data places image recognition. In *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*, pages 169–175. IEEE, 2018.

[22] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[23] Federico Marini and Beata Walczak. Particle swarm optimization (pso). a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 149:153–165, 2015.

[24] Morteza Zabihi, Ali Bahrami Rad, Serkan Kiranyaz, Moncef Gabbouj, and Aggelos K Katsaggelos. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In *2016 computing in cardiology conference (CinC)*, pages 613–616. IEEE, 2016.

[25] Jonathan Rubin, Rui Abreu, Anurag Ganguli, Saigopal Nelaturi, Ion Matei, and Kumar Sricharan. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In *2016 Computing in cardiology conference (CinC)*, pages 813–816. IEEE, 2016.

[26] Faiq Ahmad Khan, Anam Abid, and Muhammad Salman Khan. Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. *Physiological measurement*, 41(5):055006, 2020.