ISSN: 2093-5374 / E-ISSN: 2093-5382

Heart Sound Analysis Using SAINet Incorporating CNN and Transfer Learning for Detecting Heart Diseases

S. Sathyanarayanan[1*], and K. Srikanta Murthy[2]

[1*]Assistant Professor and IT Head, Department of Mathematical and Computaional Sciences, Sri Sathya Sai University for Human Excellence, Gulbarga, Karnataka, India.
Sathyanarayanan.brn@gmail.com, sathyanarayanan.s@sssuhe.ac.in,
https://orcid.org/0000-0003-0739-3452

[2]Professor and Vice-Chancellor, Department of Mathematical and Computaional Sciences, Sri Sathya Sai University for Human Excellence, Gulbarga, Karnataka, India.
srikantamurthy.k@sssuhe.ac.in, https://orcid.org/0000-0003-2744-777X

Received: February 10, 2024; Revised: March 25, 2024; Accepted: May 03, 2024; Published: June 29, 2024

Abstract

Cardiovascular disease (CVD) is the leading cause of death worldwide. Accurate and early diagnosis of cardiovascular disease (CVD) is essential for its timely treatment and management. However, this is challenging because traditional techniques for detecting heart diseases, such as auscultation, are highly subjective and prone to error. This study addresses this issue by building a novel customised deep learning architecture, SAINet, for automated CVD detection through heart sound analysis. Research is being conducted on the application of artificial intelligence (AI) to analyse phonocardiograms to detect CVD. This study aims to address this challenge by detecting heart disease using a novel customised neural network consisting of transfer learning techniques and convolutional neural networks to analyse heart sounds with increased accuracy, precision and recall and reduced computational complexity compared when compared to others. Approximately 1000 recordings of heart sounds were used to train and test the model. Data augmentation was performed to increase the size of the training data. Two combinations of datasets were used in the experiments. The first combination consisted of two categories of heart sound recording: normal and abnormal. The second combination consisted of one normal and four different abnormal categories of heart sounds. An accuracy of 99.68% was achieved with the first combination, and 99.58% with the second combination. Both combinations yielded values above 99% for precision, recall, specificity, and the F1-score. The method proposed in this study is suitable for embedding CVDs in real-time devices such as an electronic stethoscope.

Keywords: Cardiovascular Diseases, Phonocardiogram, Transfer Learning, Convolutional Neural Networks, Deep Learning.

1 Introduction

The use of artificial intelligence (AI) in healthcare has been studied to automate repetitive tasks and make healthcare cheaper, more accessible, effective, accurate, and affordable (Sathyanarayanan & Chitnis, 2022; Chatterjee et al., 2024; Mumtaj Begum, 2022). Heart disease is a major cause of death

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 2 (June), pp. 152-169. DOI: *10.58346/JOWUA.2024.I2.011*

*Corresponding author: Assistant Professor and IT Head, Department of Mathematical and Computaional Sciences, Sri Sathya Sai University for Human Excellence, Gulbarga, Karnataka, India.

worldwide. Effective treatment is possible if the heart disease is detected early. Cardiovascular diseases (CVD) affect half a billion people worldwide. It has caused 20.5 million deaths in 2021. Almost 80% of strokes and heart attacks can be prevented by timely detection and treatment (World Heart Federation, 2023; Sofiene et al., 2023; Trivedi et al., 2023).

The last 50 years of advancements in cardiovascular medicine have provided the world with the knowledge and tools necessary to alleviate the harm to cardiovascular health (Kodric et al., 2021). Communities that have the greatest need for tools to diagnose, prevent, and treat CVDs do not have them (Watrianthos et al., 2020). Most CVD-related deaths occur in middle- and low-income countries. Most cardiovascular care is provided in high-income countries. This inequality needs to be addressed by integrating AI into healthcare to make cardiac care and healthcare, in general, more accessible to underserved segments of the population. Studies have been conducted to develop automatic heart sound classification algorithms based on numerous deep learning (DL) and machine learning (ML) techniques (Jelena et al., 2023; Sakthivel et al., 2019; Arora et al., 2024).

Existing approaches rely on manual feature extraction or traditional ML techniques, which can be affected by noise, data variability, and limited generalisation (Bobir et al., 2024). DL techniques, which have been attempted by researchers, require high computational power (Kutlu et al., 2021). Distinguishing between specific heart diseases remains a challenge. This study aims to address this research gap by using advanced DL techniques, including transfer learning and convolutional neural networks (CNNs), to build a novel customised neural network, Sonic AI Net (SAINet), with an efficient architecture with minimal computation requirements that can perform multiclass classification accurately and efficiently, and has the possibility of real-time implementation by integrating it into a device such as an electronic stethoscope(Choi et al., 2022; Jelena et al., 2023).

Heart Structure and Function

The heart consists of two atria chambers and two ventricular chambers. The right atrium (RA) receives deoxygenated blood from the body and pumps it into the right ventricle (RV). The right ventricle pumps blood into the lungs for oxygenation. Oxygenated blood then returns to the left atrium (LA) and flows into the left ventricle (LV). The left ventricle pumps blood to the remainder of the body. This intricate structure ensures a unidirectional blood flow. Figure 1 shows a diagram of the human heart.

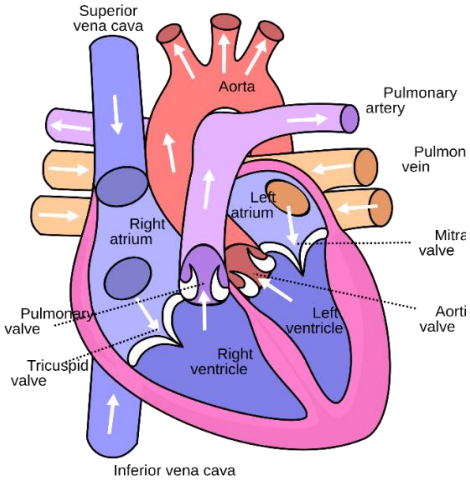

Figure 1: Structure of the Human Heart – the Four Chambers and the Valves (human heart)

Heart Sounds: The closure of heart valves during the cardiac cycle generates heart sounds. S1 (the first sound) and S2 (the second sound) are the primary heart sounds. S3 and S4 are additional heart sounds that may be heard under certain conditions. S3, also known as the ventricular gallop, is common in children, athletes, and pregnant women but may indicate heart failure in adults. S4, atrial gallop occurs just before the next S1 and is a sign of cardiac ailment (Sathyanarayanan et al., 2023).

Heart murmurs: Turbulent blood flow in the heart during the cardiac cycle generates sounds called heart murmurs. Murmurs can be classified as systolic or diastolic. Systolic murmurs occur between S1 and S2. Diastolic murmurs occur between the S2 of the current cardiac cycle and the S1 of the subsequent cardiac cycle. Understanding the causes and types of murmurs is important for the diagnosis and management of cardiovascular conditions (Sathyanarayanan et al., 2023).

Healthcare professionals use auscultation (Lubaib et al., 2015), which involves listening to heart sounds using a stethoscope and recognising potential abnormalities in heart sounds. This process is highly subjective and requires several years of clinical experience to master it. Even after that, there is a possibility of a faulty diagnosis. A phonocardiogram (PCG) is a plot of the audio waveform recorded in a non-invasive manner and provides information about heart function. The use of PCG and DL techniques has demonstrated significant promise for accurate and efficient diagnosis of cardiovascular diseases. The use of DL techniques to accurately analyse PCG to detect heart disease is promising.

2 Background

Heart sound classification algorithms: Traditional heart sound classification algorithms typically comprise pre-processing, segmentation, feature extraction, and classification steps using ML algorithms. These algorithms segment the signals into appropriate time intervals after removing noise and artefacts from heart sound signals and extracting informative features for classification. However, these methods require manual extraction of features.

DL Techniques for Heart Sound Analysis

DL techniques for heart sound analysis eliminate the feature extraction phase and do not require predefined rules. DL models automatically learn features to capture patterns and variations in heart sounds, which are often difficult to capture using traditional feature extraction methods. These models can improve performance with more data. DL can also handle large amounts of data, enabling the more accurate detection and classification of heart diseases. In addition, DL models can employ a transfer learning layer to improve their performance further.

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) that can learn hierarchical representations from data and extract features have been used for the analysis of heart sound classification.

Spectrograms

A spectrogram is a pictorial depiction of the frequencies of a signal generated over time by applying the short-time Fourier transform (STFT) technique to the audio signal (Arora et al., 2020). Spectrograms are useful for identifying patterns and irregularities associated with specific physiological events from heart sounds.

Mel-spectrograms, a sample of which is depicted in Figure 2 use the mel scale, which is based on the non-linear response to frequency by the human auditory system, unlike the normal spectrogram, which uses a linear frequency scale (Zhou et al., 2022). The frequency axis is compressed at the lower end and expanded at the higher end, simulating human loudness perception. The logarithmic intensity scale emphasises changes in apparent loudness rather than absolute amplitude. It captures temporal changes in sound. Mel-spectrograms are robust to noise. The mel scale frequency is derived from equation (1).

$$\text{Mel}(f) = 2595 \log(1 + f / 100) \quad (1)$$

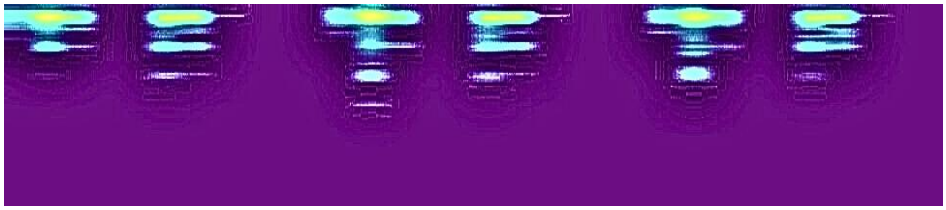

Figure 2: Sample Mel-spectrogram of a Heart Sound Audio Sample

Convolutional Neural Networks (CNNs)

CNNs are the preferred deep learning method for analysing image data, particularly for classifying heart sounds (Johnson et al., 2020). They use a multilayered architecture with specialised filters to learn relevant features and capture subtle differences within complex timbral and temporal structures. Multilayered architectures make them robust, making them excellent in various environments and with background noise or individual variations. Higher-level features are extracted as data pass through the neural network layers, capturing finer variations within heart sound categories.

CNNS have the advantage of spatial invariance, because of which they are good at capturing and analysing heart sounds, regardless of variations in recording conditions or patient characteristics. CNNs can focus on small regions of input data simultaneously using local receptive fields, allowing them to detect patterns and variations in audio signals. They also use parameter sharing, reducing the number of parameters in the model and encouraging the network to learn spatially invariant features.

Transfer Learning

Transfer learning is the use of the knowledge learned from one task to learn a related task. Machine learning practitioners leverage pre-trained models on one task to build models faster and more effectively for similar or related tasks (Kora et al., 2022). Medical tasks can be handled efficiently using transfer learning. In the context of heart sound analysis, transfer learning enables the use of DL models trained on large image datasets, such as ImageNet, to extract relevant features from spectrograms. Using the knowledge acquired from these pre-trained models, transfer learning reduces the need for vast amounts of labelled data and accelerates the training process (Kora et al., 2022).

Combination of Transfer Learning and CNNs

CNNs are suitable for hierarchical representations of data. CNN, along with transfer learning for detecting diseases from mel-spectrograms, enables the use of a model with a pre-trained transfer learning layer that can be trained further with CNN to build a heart sound signal classification model. This reduces the training time of the model.

3 Related Work

An in-depth survey by Sathyanarayanan et al., (2023) explored the application of machine learning techniques in heart disease detection. Dey et al., (2012) applied discrete wavelet transform (DWT) to spectrograms to classify heart sounds as normal or abnormal. Khan et al., (2022) proposed a completely automatic residual neural network model using power spectrograms of PCG audio samples as input for the diagnosis of multiple heart disorders. A CNN architecture developed by (Baghel et al., 2020) for multiclass classification of cardiac diseases provided very high accuracy in diagnosing multiple cardiac diseases on the test set. Shabbir et al., (2023) focussed on the classification of heart murmurs using CNNs trained on different signal representations, such as mel-frequency cepstral coefficients (MFCC) and spectrograms of phonocardiograms. SpectroCardioNet, an attention-based (DL) network that uses triple-spectrograms, that is, spectrograms, delta spectrograms, and double-delta spectrograms of PCG signals, for valvular disease detection, was built by (Chowdhury et al., 2022), and the researchers claimed to have obtained satisfactory results. Carter et al., (2023) combined deep CNNs to extract temporal signatures in heart recordings, enabling multi-label classification and severity determination, and incorporated explainable AI algorithms.

Abubakar et al., (2021) proposed a hybrid CNN model for diagnosing heart conditions by analysing heart sound signals to classify heart sounds into three classes: normal, extrasystole, and murmur. Tao et al., (2021) designed a lightweight end-to-end 2D-CNN neural network with features from the frequency domain as input and reported an accuracy of 86%. Feng Li et al., (2022) reported an accuracy of 94.43% using improved MFCCs and ResNet. Proposed a hybrid model that uses spectrograms and interpolation for accurate training and data augmentation, combined with the Relief feature selection method, to optimise the feature maps.

Yaseen et al., (2018) proposed an algorithm for heart sound classification using features extracted from phonocardiogram signals, such as MFCCs and DWT, and ML techniques, such as deep neural networks (DNN), support vector machines (SVM), and k-nearest neighbours (kNN) and claimed an accuracy of up to 97.9%. Spectrograms were generated from the heart sound signals using continuous wavelet transform (CWT), which were subsequently used as input to ten transfer learning networks by (Wang et al., 2022). Four transfer learning networks (ResNet101, GoogleNet, DarkNet19 and darkNet201) gave the best results, with an accuracy of 98%. The centroid frequency, a time-varying spectral feature, was used to classify three heart sounds by (Upretee et al., 2019), who reported 96.50% accuracy for multi-class classification and 99.6% accuracy for binary classification using both SVM and kNN classifiers.

Taneja et al., (2023) reported an accuracy of 94.87 % using a combination of textural features extracted from the spectrogram and chromagram representations of an audio dataset. Milani et al., (2021) extracted the frequency domain, time domain, and statistical features of the audio dataset and input them to LDA and ANN to obtain an accuracy of up to 93.33% on the PhysioNet dataset.

LBP and HOG features of the spectrograms generated from the audio dataset were used by (Sathyanarayan et al., 2024) as input for ML classification modelsto obtain excellent results.

Gap in research: These studies collectively showcase the depth and breadth of research in the domain of heart sound analysis using AI. Most researchers have used very complex neural network layers with a large number of layers. Several methods require segmentation (identifying S1 and S2 in the cardiac cycle) as one of the pre-processing tasks. Most studies reported working with high-end systems. The

number of epochs required for DL-based models was quite high because of which the computation requirement is huge.

The purpose of this work is to build a model with a customised novel neural network for heart sound classification with increased accuracy, precision, and recall using the Yaseen dataset with reduced computational complexity. The proposed model has a minimal footprint and requires minimal computational resources. Therefore, the model can be embedded in electronic diagnostic equipment with minimal computational hardware.

4 Dataset Details

Dataset: The dataset used in this study consisted of heart sound recordings related to valvular diseases that were processed and collected by (Yaseen et al., 2018). The data were collected from various random sources, such as books like "Auscultation Skills CD" and "Heart Sound Made Easy", as well as from multiple websites (a total of 48 websites provided the data, including Washington, 3M, and Michigan). The samples containing excessive noise were eliminated. The dataset was filtered and converted to a mono channel consisting of 1000 audio samples, including 200 samples from each of the five classes. The dataset consists of single-channel audio with a 128-kbps bit rate, sampling at 8 KHz, and 16 bits per sample. The recordings were made between 1 s and 3 s, with most of the audio data lasting 2 s. We excluded audio samples shorter than two seconds and trimmed longer audio samples to 2 s before training to prevent possible errors due to nonuniform training data.

Table 1: Details of the Dataset

Heart Diseases	Number of Heart Sounds
Normal	200
Mitral regurgitation	184
Aortic stenosis	200
Mitral valve prolapse	187
Mitral stenosis	186
Total	957

According to (Bao et al., 2022), 2 s recordings of heart sounds are ideal because longer samples would not necessarily yield higher accuracy and would waste computing resources. Audio samples of less than 2 s duration may not be sufficient to identify patterns and reduce random errors. Finally, 957 audio samples were used, as listed in Table 1.

The four types of valvular heart disease have distinct characteristics. Aortic stenosis occurs due to the narrowing of the aortic valve, which regulates the blood flow from the left ventricle to the aorta. Mitral regurgitation is caused by the reverse flow of blood from the left ventricle to the left atrium. The narrowing of the valve between the left ventricle and left atrium leads to mitral stenosis. Finally, the term "mitral valve prolapse" describes a medical condition characterised by the protrusion of the mitral valve leaflets into the left atrium during each cardiac cycle. One PCG for each category of heart sounds is shown in Figure 3.

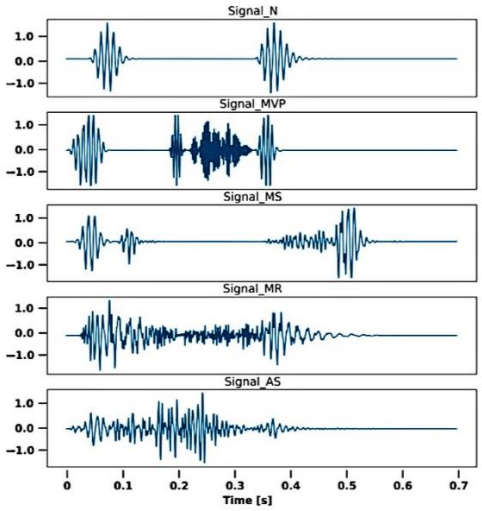

Figure 3: Sample PCG of Each Category of Heart Sounds (Aljohani et al., 2023)

5 Methodology

Steps in the Classification System

The Yaseen dataset is used in this study. The dataset comprised audio recordings of heart sounds. Following the procedures described previously, the dataset was pre-processed to ensure uniformity. Experiments were conducted using two combinations of datasets. The first combination comprised both the normal and abnormal categories. The second combination of datasets involved one normal category of heart sound recordings and four different categories of abnormal heart sound recordings, each of which consisted of audio samples from one valvular disease. Ultimately, 957 mel-spectrograms were generated from the dataset.

This study was performed in five steps, as shown in Figure 4, for both combinations of datasets.

1) Acquire the heart sounds dataset.
2) Pre-process the audio data and perform data augmentation. All audio samples were trimmed to a uniform length of 2 s and samples with shorter lengths were eliminated.
3) Convert one-dimensional audio data to three-dimensional mel-spectrograms that give the time, frequency, and strength of the signal by applying STFT to the audio signal.
4) Build a customised neural network including a ResNet50 transfer learning layer and a CNN to train the model for classifying the mel-spectrograms.
5) Train the model and evaluate performance based on various ML metrics.

Figure 4: Steps in the Classification System

Model Architecture

The classification head designed for this task shown in Figure 5 comprises a ResNet50 pretrained layer followed by a convolutional layer with 64 filters and a (3,3) kernel size. This is followed by a global average pooling layer. A dense layer using the ReLU activation function succeeds the global pooling layer. The final layer is a dense layer that uses Softmax function to calculate class probabilities.

The CNN layer extracts task-specific characteristics from the high-level representations acquired using the ResNet50 model. The global average pooling layer reduces the dimensions of the features to a single vector. Only the essential information necessary for classification is preserved. A dense layer consisting of 256 nodes and the ReLU function was used to process the extracted features and learn the decision boundaries for separating the classes.

The ReLU function is defined as given in equation (2).

$$f(x) = \max(0, x) \quad (2)$$

If x is positive, it returns x, and returns zero otherwise. RELU introduces sparsity in the network, is computationally efficient compared to Sigmoid and tanh function. ReLU function does not saturate unlike the Sigmoid and tanh function. It also helps in more efficient optimistion during training and avoids vanishing gradients. Due to the above reasons, it also helps in faster convergence which will result in reduced training time.

A dropout layer with a 50% dropout rate was introduced to ensure the learning of robust features and reduce overfitting. This layer randomly drops connections during the training. The output layer is a dense layer with several nodes that are equivalent to the number of classes present in the dataset and utilises the Softmax activation function.

The Softmax function is defined as given in equation (3).

$$Softmax\ (z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (3)$$

where K represents the total number of classes, z_i is the raw score (logit) for class i and Softmax $(z)_i$ is the probability of class i for input z.

The final classification task outputs the probability distributions over classes for each input image. Table 2 provides the details of the layers of the customised architecture.

Table 2: The SAINet Model Architecture

S. No.	Layer type	Activation function	Kernel size/value/nodes	Filters
1	Input			
2	Data augmentation			
3	ResNet50			
4	Conv2D	ReLU	(3,3)	64
5	GlobalAveragePooling2D			
6	Dense	ReLU	256	
7	Dropout		0.5	
8	Output	Softmax	Number of classes	

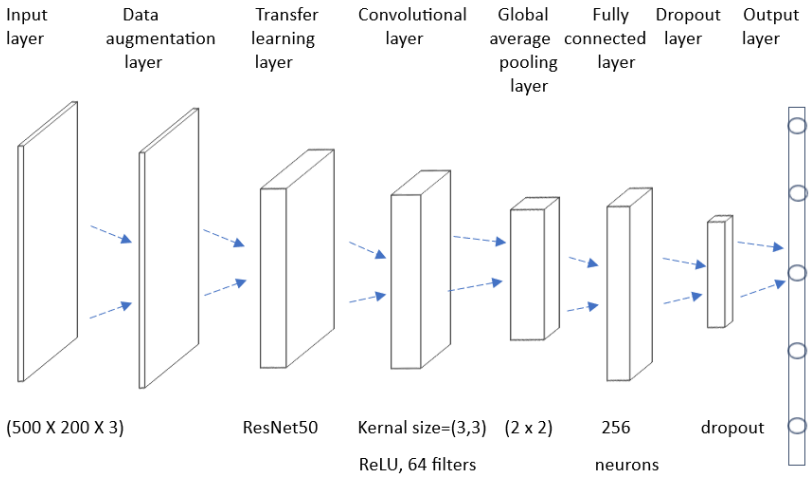

Figure 5: Architecture of the Proposed Customised Neural Network SAINet for 5-class Classification

Role of Transfer Learning and CNN in Classification

SAINet leverages the combined power of transfer learning and CNNs to achieve high accuracy in detecting valvular heart diseases through heart sound analysis.

A pre-trained layer is incorporated in SAINet to take advantage of the knowledge already learnt by these layers. This allows SAINet to efficiently extract features from heart sounds without having to train its convolutional layers from scratch, thus reducing the training time and the requirement of computational resources. In SAINet, this knowledge is transferred and adapted to the specific domain of heart sound analysis. By fine-tuning these pre-trained layers with labelled heart sound data, SAINet learns to extract features that are relevant to distinguishing normal from abnormal heart sounds and identifying disease characteristics.

Convolutional layers extract features from localised segments of the heart sound signal and learn patterns related to valve opening and closing events. The subsequent convolutional layers in the SAINet were specifically designed for heart sound analysis. The layers refine the features extracted by the pre-trained layers and focus on differentiating between normal and abnormal heart sounds. These refined features are then fed into fully connected layers for classification, enabling SAINet to categorise heart sounds accurately.

Transfer learning provides a fast method for extracting meaningful features, while CNNs ensure that this process is customised to the domain of heart sound analysis. This combination allows SAINet to identify minute variations in heart sounds and differentiate between different types of valvular heart diseases.

Model Training and Optimisation

The Adam optimiser was used for gradient descent updates during training. Combining the benefits of AdaGrad and RMSProp, Adam often converges faster than other optimisers. It employs adaptive learning rates that are adjusted based on the parameters, thereby improving convergence and stability. A lower learning rate of 0.0001 was used compared with typical values to stabilise the training process, improve convergence, and avoid overfitting. ReLU function is used for the hidden layers.

The batch size was fixed at 24 after experimentation with 16, 32 and 64 to have a balance between training speed and memory usage and offered a good balance between training time and validation accuracy. This is smaller than the dataset size, allowing for more frequent updates and potentially better generalisation. The validation accuracy of the model improved up to 95 epochs, after which it stagnated or started to decrease, indicating potential overfitting. Hence, the number of epochs was fixed at 95.

The last layer computed the accuracy of the model. The loss in multi-class classification was calculated using the categorical cross-entropy function. This model combines the strengths of transfer learning with a custom classification head to achieve efficient and accurate performance in classification tasks. Data augmentation increased the model`s generalisability. The model architecture, training configuration, and hyperparameters were selected to balance the effectiveness and robustness. Table 3 lists the values used for hyperparameters.

Table 3: Hyperparameters of the Proposed Model

Hyperparameter	Value
Loss function	Categorical cross-entropy
Epoch	24
Batch size	95
Optimization algorithm	Adam
Learning rate	0.0001
Dropout	0.5
Activation function	ReLU and Softmax

6 Results and Discussion

The following metrics were derived from the confusion matrix for performance analysis of the model:

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

$$\text{Precision} = \frac{TP}{FP+TP} \quad (5)$$

$$\text{Sensitivity} = \text{Recall} = \text{True Positive Rate} = \frac{TP}{FN+TP} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (7)$$

$$F1-\text{score} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \quad (8)$$

where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively shows in equation (4)-(8).

Table 4: Performance Evaluation Metrics

Metric	Two classes	Five classes
Training Accuracy	1	1
Validation Accuracy	0.98958	0.979166
Test Accuracy	0.994791	1
Overall Accuracy	**0.996865**	**0.99582**
Precision	0.996168	0.995779
Recall	0.994339	0.995924
Specificity	0.998678	0.995
F1-Score	0.995252	0.995852
AUC	0.99	1

The implementation of the SAINet model incorporating a CNN and transfer learning on the heart sound dataset yielded noteworthy results. This approach reduces reliance on labelled data by using pre-trained models and CNNs' ability to capture relevant frequency patterns. The generalisation provided by transfer learning combined with CNNs performed well. The performance metrics are listed in Table 4 and displayed in a chart in Figure 6.

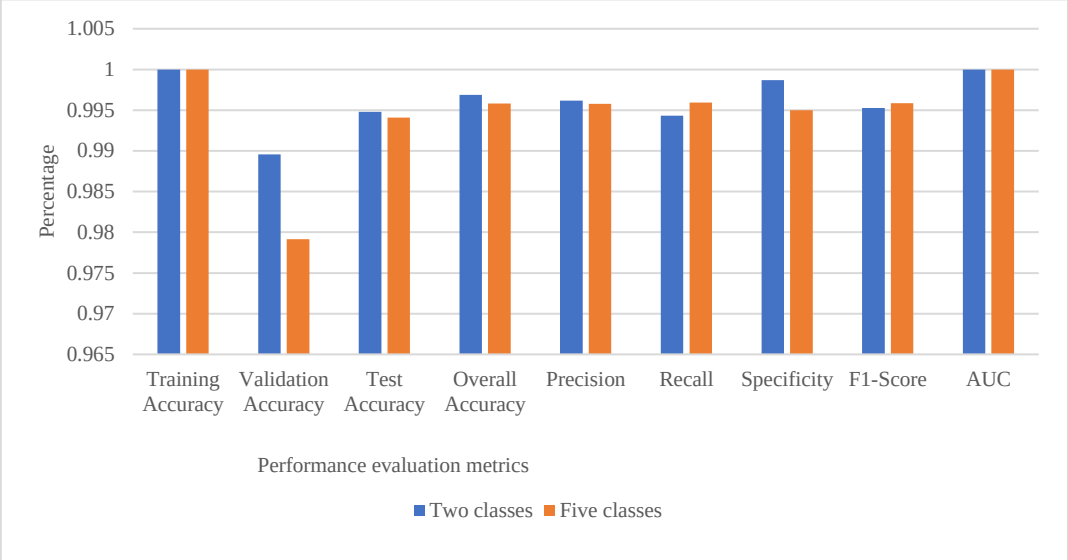

Figure 6: Performance Metrics for the Two-class and Five-class Combination

Two-class classification: The results of the binary classification model, which distinguished between abnormal and normal heart sounds, indicated excellent performance across various evaluation metrics. The model consistently achieves accuracy scores of approximately 99%, indicating its ability to correctly classify heart sounds into their respective categories and indicates that the model is robust and has generalisation capabilities. The high precision and recall values indicate the model's ability to perform classification with minimal errors. The high F1- score (> 99%) indicates the model's ability to achieve a balance between precision and recall. High specificity (> 99%) indicates that the false positives are very low.

Specificity, which is a measure of accurately identifying negative instances, is very high, indicating that the model effectively identifies normal heart sounds with a very low rate of false positives. AUC represents the ability of the model to discriminate between abnormal and normal heart sounds across all possible thresholds. A near perfect AUC score of 0.99 signifies that the model's predicted probabilities rank true positive instances higher than false positive instances, regardless of whether the threshold chosen is perfect. Figure 7 depicts the ROC curve for two-class classification

Combination 1: Two Classes - Normal and Abnormal

Table 5: Confusion Matrix for the Two-class Combination

Predicted Class	Actual abnormal	Actual normal
Abnormal	756	1
Normal	2	198

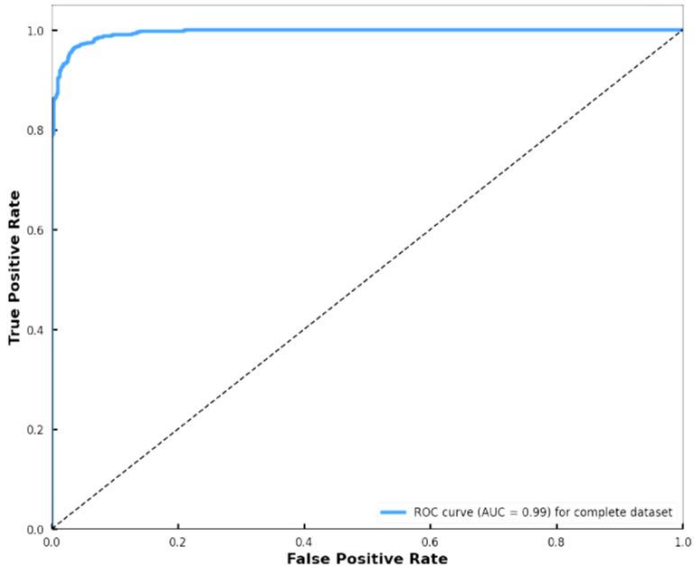

Figure 7: ROC Curve for the Two-class Dataset (AUC = 0.99)

The confusion matrix in Table 5 shows only three instances of incorrect classifications out of the 957 samples. Most of the errors occur when predicting normal heart sounds as abnormal, with only one instance of an abnormal heart sound misclassified as normal. The 2 x 2 confusion matrix reveals excellent performance. With 756 correctly classified abnormal cases and only 2 misclassified normal instances, the model demonstrates a strong ability to distinguish between abnormal and normal heart sounds. The high true positive (756) and true negative (198) values, along with minimal false positives (1) and false negatives (2), indicate a well-balanced and accurate classification.

Five-class classification: The performance of the five-class classification model is excellent. The model consistently achieved accuracy scores exceeding 99%, indicating the model is robust. It learns patterns effectively and generalises to unseen data. High precision and recall values (>0.99) suggest the model accurately identifies most instances it predicts as a particular class (precision) and captures most actual cases of each class (recall). The high specificity values (>0.99) indicate the model effectively avoids classifying negative instances (e.g., normal heart sounds) as positive (abnormal). The high F1-score (>0.99) shows the balanced performance of the model in correctly identifying positive cases and avoiding false positives. A perfect AUC of 1 signifies the model's ideal performance in distinguishing between classes. Figure 8 depicts the ROC curve for the five-class combination and the AUC values for each category.

Combination 2: Five Classes – one Normal and Four Abnormal Classes

Table 6: Confusion Matrix for the Five-class Combination

Predicted Class	Actual AS	Actual MR	Actual MS	Actual MVP	Actual Normal
AS	199	1	0	0	0
MR	0	184	0	0	0
MS	1	0	185	0	0
MVP	0	0	0	187	0
Normal	0	0	2	0	198

There are a very small number of misclassifications across the five classes as seen from the confusion matrix. The 5x5 confusion matrix shows satisfactory performance across all classes. The diagonal element represents correctly classified instances. There are 199 true positives in the AS category, 184 true positives in the MR category, 185 true positives in the MS category, 187 true positives in the MVP category and 198 true positives in the Normal category. These high values indicate the model effectively identifies most instances within each class. Instances of misclassification are minimal.

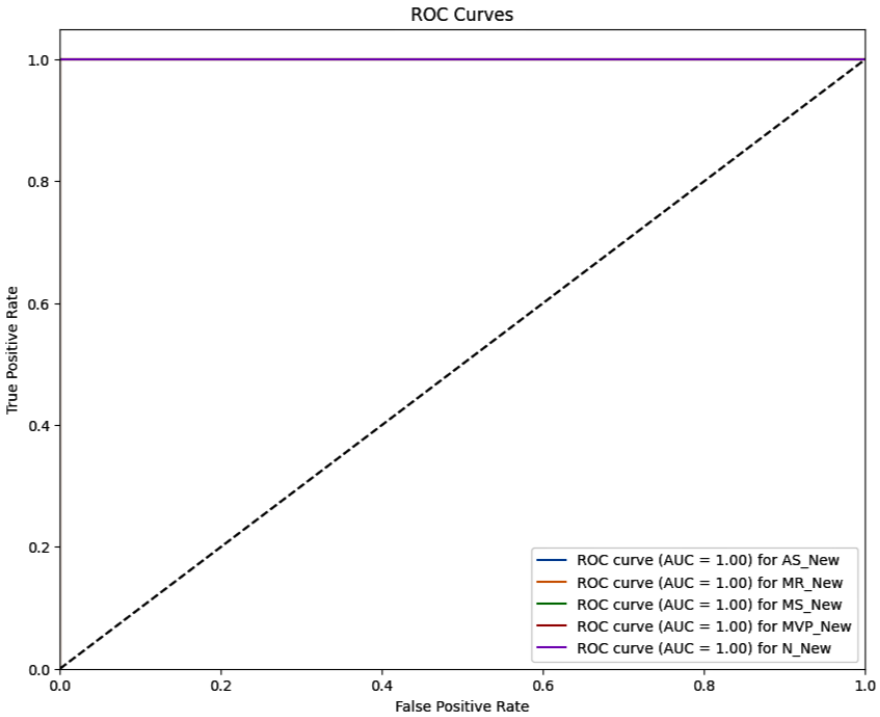

Figure 8: ROC Curve for the Five-class Dataset (AUC = 1)

Further investigation is required to identify the underlying factors that have caused the misclassification and ways can be found for further improvements.

The heart sound classification model demonstrates notable performance based on the metrics and confusion matrix. The high accuracy, precision, recall, specificity, F1-score, and AUC across both two and five-class scenarios indicate the model's efficacy in correctly classifying various heart conditions.

The performance of the proposed model is presented in Table 7 along with the results obtained from other studies. Several factors are responsible for the high accuracy and performance of the SAINet model. The combination of CNN and transfer learning enabled the model to perform robustly even with background noise and variations in the recording quality of heart sounds. The CNN's ability to extract, coupled with pre-trained knowledge from transfer learning, contributed significantly to the model's performance.

The model’s generalisation ability has been enhanced because of the data augmentation techniques used in this study. By increasing the diversity of the training data, the model became more adept at handling real-world variations in heart sound recordings.

Table 7: Comparison of the Results

Author(s)	Methodology	Dataset used	Accuracy (%)
5-class classification			
Yaseen et al.	MFCC + Discrete wavelet transform features combined with SVM, DNN and centroid displacement kNN	Yaseen dataset	97.9
Chowdhury et al.	DNN	Yaseen dataset	97.77
Wang et al.	Transfer learning	Yaseen dataset + additional category (6 classes)	98%
Baghel et al.	CNN with augmentation	Yaseen dataset	98.6
Yadav et al.	Statistical features	Private dataset	97.78
Upretee et al.	Spectral centroid frequency with kNN and SVM	Yaseen dataset	96.50
Proposed method	**Mel-spectrograms + CNN and transfer learning**	**Yaseen dataset**	**99.58**
Binary classification			
Upretee et al.	Spectral centroid frequency with kNN and SVM	Yaseen dataset	99.60
Taneja et al.	LBP + chromagram	PhysioNet 2016	94.87
Takezaki et al.	Data augmentation and CNN	PhysioNet 2016	93.7
Milani et al.	Time-domain features and ANN	PhysioNet 2016	93.33
T. Li, & Yin et al.	Frequency domain features and 2D-CNN	PhysioNet 2016	86%
F.Li & Zhang et al.	Improved MFCC and ResNet	PhysioNet 2016	94.43
Proposed method	**Mel-spectrograms + CNN and transfer learning**	**Yaseen dataset**	**99.68**

7 Conclusion

In conclusion, the findings presented in this paper demonstrate the remarkable performance of heart sound classification models in effectively differentiating between various valvular ailments. Both models yielded exceptional performance. The binary classification model yielded high levels of accuracy, precision, recall, specificity, and AUC score despite the dataset being imbalanced. These results indicate the accuracy of the model in detecting abnormal heart sounds effectively.

Similarly, the multi-class classification model demonstrated exceptional performance across all categories, with perfect AUC scores of 1 for each class. This indicates the capability of the model to detect valvular diseases, thereby providing clinicians with a potentially valuable tool for decision-making.

This investigation suggests the potential of automated diagnostic tools for treating CVDs. The model could be a useful tool for increasing access to healthcare because of its excellent performance in various metrics. In the analysis of mel-spectrograms of heart sounds for disease detection, the use of transfer learning techniques in combination with CNNs has shown promising results. This approach could prove beneficial in constructing models for medical diagnosis given the usual limitations in the dataset.

This automated classification integrated into an electronic stethoscope can serve as the initial stage of screening for cardiovascular disease and facilitate prompt and accurate diagnosis. These models hold promise for enhancing diagnostic capabilities in cardiovascular healthcare, contributing to advancements in patient care and medical research.

Limitations and Future Work: There are certain limitations to this study. The chances of overfitting can be reduced by acquiring more data in each category. Future work should focus on validating the model across more diverse datasets, including sounds recorded from different demographic groups and using various stethoscope types.

Several other tasks must be done before the model can be implemented in clinical practice. The model must be integrated with electronic health record (EHR) systems. User interfaces must be designed for non-technical medical staff members. Patient data security and privacy must be guaranteed. of must be ensured. Efforts must be made to develop an explainable model that enables user acceptance.

In conclusion, the SAINet model demonstrates a significant potential for advancing cardiac healthcare through AI-driven diagnostics. Future research should aim at addressing the current limitations and integrating the technology into practical, user-friendly electronic stethoscope that can detect heart diseases and could be integrated into the EHR systems. Such tools can be used as a first level screening device by the healthcare providers.

Acknowledgements

We thank Sri Sathya Sai University for Human Excellence for funding this study. We thank Mrs Shobha Sathyanarayanan, Dr. Satishkumar Mallappa, Dr. Chandrasekhar Gudada and Sri Sandeep Relan for all their help.

Conflict of Interest

No conflict of interest is involved in this study.

References

[1] Abubakar, M. M., Adamu, B. Z., & Abubakar, M. Z. (2021). Pneumonia classification using hybrid CNN architecture. *In IEEE International Conference on Data Analytics for Business and Industry (ICDABI)*, 520-522.

[2] Aljohani, R. I., Hosni Mahmoud, H. A., Hafez, A., & Bayoumi, M. (2023). RETRACTED: A Novel Deep Learning CNN for Heart Valve Disease Classification Using Valve Sound Detection. *Electronics*, *12*(4), 846. https://doi.org/10.3390/electronics12040846

[3] Arora, G. (2024). Desing of VLSI Architecture for a flexible testbed of Artificial Neural Network for training and testing on FPGA. *Journal of VLSI Circuits and Systems, 6*(1), 30-35.

[4] Arora, V., Ng, E. Y. K., Leekha, R. S., Verma, K., Gupta, T., & Srinivasan, K. (2020). Health of things model for classifying human heart sound signals using co-occurrence matrix and spectrogram. *Journal of Mechanics in Medicine and Biology*, *20*(06), 2050040. https://doi.org/10.1142/S0219519420500402

[5] Baghel, N., Dutta, M. K., & Burget, R. (2020). Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network. *Computer Methods and Programs in Biomedicine*, *197*, 105750. https://doi.org/10.1016/j.cmpb.2020.105750

[6] Bao, X., Xu, Y., & Kamavuako, E. N. (2022). The effect of signal duration on the classification of heart sounds: A deep learning approach. *sensors*, *22*(6), 2261. https://doi.org/10.3390/s22062261

[7] Bobir, A. O., Askariy, M., Otabek, Y. Y., Nodir, R. K., Rakhima, A., Zukhra, Z. Y., Sherzod, A. A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. *Natural and Engineering Sciences*, *9*(1), 72-83.

[8] Carter, T. S., Yang, G. H., Loke, G., & Yan, W. (2023). Deciphering simultaneous heart conditions with spectrogram and explainable-AI approach. *Biomedical Signal Processing and Control*, *85*, 104990. https://doi.org/10.1016/j.bspc.2023.104990

[9] Chatterjee, P., Siddiqui, S., Granata, G., Dey, P., & Abdul Kareem, R. S. (2024). Performance Analysis of Five U-Nets on Cervical Cancer Datasets. *Indian Journal of Information Sources and Services, 14*(1), 17–28.

[10] Choi, J., & Zhang, X. (2022). Classifications of restricted web streaming contents based on convolutional neural network and long short-term memory (CNN-LSTM). *Journal of Internet Services and Information Security, 12*(3), 49-62.

[11] Chowdhury, S., Morshed, M., & Fattah, S. A. (2022). SpectroCardioNet: An attention-based deep learning network using triple-spectrograms of PCG signal for heart valve disease detection. *IEEE Sensors Journal*, *22*(23), 22799-22807.

[12] Dey, N., Mishra, G., Nandi, B., Pal, M., Das, A., & Chaudhuri, S. S. (2012). Wavelet based watermarked normal and abnormal heart sound identification using spectrogram analysis. *In IEEE international conference on computational intelligence and computing research*, 1-7.

[13] Diagram of the human heart, https://upload.wikimedia.org/wikipedia/commons/thumb/f/fa/Diagram_of_the_human_heart.svg/2004px-Diagram_of_the_human_heart.svg.png

[14] Jelena, T., & Srđan, K. (2023). Smart Mining: Joint Model for Parametrization of Coal Excavation Process Based on Artificial Neural Networks. *Archives for Technical Sciences*, *2*(29), 11-22.

[15] Johnson, C., Khadka, B., Basnet, R. B., & Doleck, T. (2020). Towards Detecting and Classifying Malicious URLs Using Deep Learning. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 11*(4), 31-48.

[16] Khan, J. S., Kaushik, M., Chaurasia, A., Dutta, M. K., & Burget, R. (2022). Cardi-Net: A deep neural network for classification of cardiac disease using phonocardiogram signal. *Computer Methods and Programs in Biomedicine*, *219*, 106727. https://doi.org/10.1016/j.cmpb.2022.106727

[17] Kodric, Z., Vrhovec, S., & Jelovcan, L. (2021). Securing edge-enabled smart healthcare systems with blockchain: A systematic literature review. *Journal of Internet Services and Information Security*, *11*(4), 19-32.

[18] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., & Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, *42*(1), 79-107.

[19] Kutlu, Y., & Camgözlü, Y. (2021). Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. *Natural and Engineering Sciences*, *6*(1), 60-74.

[20] Li, F., Zhang, Z., Wang, L., & Liu, W. (2022). Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning. *Frontiers in Physiology*, *13*, 1084420. https://doi.org/10.3389/fphys.2022.1084420

[21] Li, T., Yin, Y., Ma, K., Zhang, S., & Liu, M. (2021). Lightweight end-to-end neural network model for automatic heart sound classification. *Information*, *12*(2), 54. https://doi.org/10.3390/info12020054

[22] Lubaib, P., Kv, A. M., & AR, V. (2015). Phonocardiogram based Diagnostic System. *International Journal of Biomedical Science and Engineering, 2*(3), 1–10.

[23] Milani, M. G. M., Abas, P. E., De Silva, L. C., & Nanayakkara, N. D. (2021). Abnormal heart sound classification using phonocardiography signals. *Smart Health*, *21*, 100194. https://doi.org/10.1016/j.smhl.2021.100194

[24] Mumtaj Begum, H. (2022). Scientometric Analysis of the Research Paper Output on Artificial Intelligence: A Study. *Indian Journal of Information Sources and Services*, *12*(1), 52–58.

[25] Sakthivel, M. V., Kesaven, M. P., William, M. J. M., & Kumar, M. S. M. (2019). Integrated

platform and response system for healthcare using Alexa. *International Journal of Communication and Computer Technologies (IJCCTS), 7*(1), 14-22.

[26] Sathyanarayanan, S., & Chitnis, S. (2022). A Survey of Machine Learning in Healthcare. In *Artificial Intelligence Applications for Health Care*, pp. 1-22.

[27] Sathyanarayanan, S., Murthy, S., & Chitnis, S. (2023). A Comprehensive Survey of Analysis of Heart Sounds using Machine Learning Techniques to Detect Heart Diseases. *Journal of Population Therapeutics and Clinical Pharmacology*, *30*(11), 375-384.

[28] Shabbir, M., Liu, X., Nasseri, M., & Helgeson, S. (2023). Heart Murmur Classification in Phonocardiogram Representations Using Convolutional Neural Networks. In *The International FLAIRS Conference Proceedings*, *36.* https://doi.org/10.32473/flairs.36.133189

[29] Sofiene, M., Souhir, C., Yousef, A., & Abdulrahman, A. (2023). Blockchain Technology in Enhancing Health Care Ecosystem for Sustainable Development. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14*(3), 240-252.

[30] Swaminathan, S., Krishnamurthy, S. M., Gudada, C., Mallappa, S. K., & Ail, N. (2024). Heart Sound Analysis with Machine Learning Using Audio Features for Detecting Heart Diseases. *International Journal of Computer Information Systems and Industrial Management Applications*, *16*(2), 131-147.

[31] Takezaki, S., & Kishida, K. (2022). Data Augmentation and the Improvement of the Performance of Convolutional Neural Networks for Heart Sound Classification. *International Journal of Computer Science*, *49*(4).

[32] Taneja, K., Arora, V., & Verma, K. (2023). Classifying the heart sound signals using textural-based features for an efficient decision support system. *Expert Systems*, *40*(6), e13246. https://doi.org/10.1111/exsy.13246

[33] Trivedi, J., Devi, M. S., & Solanki, B. (2023). Step Towards Intelligent Transportation System with Vehicle Classification and Recognition Using Speeded-up Robust Features. *Archives for Technical Sciences, 1*(28), 39-56.

[34] Upretee, P., & Yüksel, M. E. (2019). Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results. *In IEEE Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT)*, 1-4.

[35] Wang, M., Guo, B., Hu, Y., Zhao, Z., Liu, C., & Tang, H. (2022). Transfer learning models for detecting six categories of phonocardiogram recordings. *Journal of Cardiovascular Development and Disease*, *9*(3), 86. https://doi.org/10.3390/jcdd9030086

[36] Watrianthos, R. (2020). A Compact Hybrid Ring Patch Antenna for Fixed Communication Applications. *National Journal of Antennas and Propagation (NJAP), 2*(1), 13-18.

[37] World Heart Federation. (2023). World Heart Report 2023: Confronting the World's Number One Killer.

[38] Yadav, A., Singh, A., Dutta, M. K., & Travieso, C. M. (2020). Machine learning-based classification of cardiac diseases from PCG recorded heart sounds. *Neural Computing and Applications*, *32*(24), 17843-17856.

[39] Yaseen, Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences*, *8*(12), 2344. https://doi.org/10.3390/app8122344

[40] Zhou, G., Chen, Y., & Chien, C. (2022). On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks. *BMC Medical Informatics and Decision Making*, *22*(1), 226. https://doi.org/10.1186/s12911-022-01942-2

Authors Biography

Sathyanarayanan Swaminathan is a Senior Manager at SSSUHE and is leading the IT team. He worked as the Information Scientist for 17 years with Sri Sathya Sai Institute of Higher Learning (SSSIHL) earlier. He has been teaching computer science courses for more than 32 years and is also a research scholar in the department of Mathematics and Computational Sciences. He has published/presented around 25 articles including 7 papers in peer-reviewed journals and two book chapter. His research area is Artificial Intelligence, applications of AI in healthcare, computational finance and free and open-source software (FOSS) and is also guiding PG students. He also has administrative experience including as Director in-charge of the campus several times and has acquired industry certifications.

Prof. K. Srikanta Murthy is currently serving as the Vice-Chancellor of SSSUHE and as the Professor in DMACS. He has 36 years of teaching experience and 16 years of research experience. He has guided 9 research scholars towards Ph.D. and is guiding 2 research scholars at present. He also has several national awards to his credit. He has published/presented more than 100 papers and one book chapter. His area of expertise is Image Processing, Pattern Recognition, Document Image Analysis, Character Recognition, Data Mining, Natural Language Processing and Artificial Intelligence. He also has 27 years of administrative experience including as HoD of CS, Principal and member of various academic bodies.