Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.063427

ARTICLE

Cardiovascular Sound Classification Using Neural Architectures and Deep Learning for Advancing Cardiac Wellness

Deepak Mahto[1], Sudhakar Kumar[1], Sunil K. Singh[1], Amit Chhabra[1], Irfan Ahmad Khan[2], Varsha Arya[3,4], Wadee Alhalabi[5], Brij B. Gupta[6,7,8,9,*] and Bassma Saleh Alsulami[10]

[1]Department of CSE, Chandigarh College of Engineering and Technology, Panjab University, Chandigarh, 160019, India
[2]Department of ECE, Chandigarh College of Engineering and Technology, Panjab University, Chandigarh, 160019, India
[3]Department of Electronic Engineering and Computer Science, Hong Kong Metropolitan University, Hong Kong SAR, 999077, China
[4]Center for Interdisciplinary Research, University of Petroleum and Energy Studies (UPES), Dehradun, 248007, India
[5]Immersive Virtual Reality Research Group, Department of Computer Science, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[6]Department of Computer Science and Information Engineering, Asia University, Taichung, 413305, Taiwan
[7]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, 40447, Taiwan
[8]Symbiosis Centre for Information Technology (SCIT), Symbiosis International University, Pune, 411057, India
[9]School of Cybersecurity, Korea University, Seoul, 02841, Republic of Korea
[10]Faculty of computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

*Corresponding Author: Brij B. Gupta. Email: gupta.brij@gmail.com

Received: 14 January 2025; Accepted: 23 May 2025; Published: 30 June 2025

ABSTRACT: Cardiovascular diseases (CVDs) remain one of the foremost causes of death globally; hence, the need for several must-have, advanced automated diagnostic solutions towards early detection and intervention. Traditional auscultation of cardiovascular sounds is heavily reliant on clinical expertise and subject to high variability. To counter this limitation, this study proposes an AI-driven classification system for cardiovascular sounds whereby deep learning techniques are engaged to automate the detection of an abnormal heartbeat. We employ FastAI vision-learner-based convolutional neural networks (CNNs) that include ResNet, DenseNet, VGG, ConvNeXt, SqueezeNet, and AlexNet to classify heart sound recordings. Instead of raw waveform analysis, the proposed approach transforms preprocessed cardiovascular audio signals into spectrograms, which are suited for capturing temporal and frequency-wise patterns. The models are trained on the PASCAL Cardiovascular Challenge dataset while taking into consideration the recording variations, noise levels, and acoustic distortions. To demonstrate generalization, external validation using Google's Audio set Heartbeat Sound dataset was performed using a dataset rich in cardiovascular sounds. Comparative analysis revealed that DenseNet-201, ConvNext Large, and ResNet-152 could deliver superior performance to the other architectures, achieving an accuracy of 81.50%, a precision of 85.50%, and an F1-score of 84.50%. In the process, we performed statistical significance testing, such as the Wilcoxon signed-rank test, to validate performance improvements over traditional classification methods. Beyond the technical contributions, the research underscores clinical integration, outlining a pathway in which the proposed system can augment conventional electronic stethoscopes and telemedicine platforms in the AI-assisted diagnostic workflows. We also discuss in detail issues of computational efficiency, model interpretability, and ethical considerations, particularly concerning algorithmic bias stemming from imbalanced datasets and the need for real-time processing in clinical settings. The study describes a scalable, automated system combining deep learning, feature extraction using spectrograms, and external validation that can assist healthcare providers in the early and accurate detection of cardiovascular disease. AI-driven solutions can be viable in improving access, reducing delays in diagnosis, and ultimately even the continued global burden of heart disease.

KEYWORDS: Healthy society; cardiovascular system; spectrogram; FastAI; audio signals; computer vision; neural network

1 Introduction

Cardiovascular disease is the leading cause of death worldwide. In 2019, approximately 17.9 million individuals succumbed to cardiovascular disease, making up 32% of all global deaths. Cardiovascular attacks and strokes accounted for 85% of these deaths. By 2030, this number is projected to increase to over 23 million annually [1]. The high prevalence and economic cost of cardiovascular diseases impose a substantial social and financial burden on society. For instance, the annual combined direct and indirect expenses associated with cardiovascular disease in the United States are estimated to be $378.0 billion, based on data from the Medical Expenditure Panel Survey 2017–2018 [2–4]. This amount encompasses $226.2 billion in direct expenses as well as $151.8 billion in lost potential productivity (indirect costs) linked to premature cardiovascular disease deaths between 2017 and 2018.

While the predicted number of individuals with cardiac conditions and associated healthcare expenses is significant, it is essential to remember that many cardiovascular diseases are manageable and even curable. Yet, achieving successful outcomes relies on early detection and suitable treatment. As a result, there is a pressing demand for advancements in technologies that enable intensive monitoring and analysis of physiological data associated with cardiac function, all while being both timely and cost-effective [4–6].

Normal cardiovascular sounds are classified as S1 ('lub') and S2 ('dub') (Fig. 1). The S1 sound corresponds to the closure of the atrioventricular valves during systole, while the S2 sound corresponds to the closure of the semilunar valves during diastole. Healthcare professionals use stethoscopes to listen to these heart sounds and identify cardiovascular disorders [7]. According to the Mayo Clinic [8], adults should typically have a heart rate ranging from 60 to 100 beats per minute. The cardiovascular system's characteristic sound pattern, often described as 'lub' 'dub', 'dub' 'lub', represents a normal and healthy sequence of a heartbeat, with the period of 'dub' to 'lub' being longer than 'lub' to 'dub'. However, when there is a loud sound occurring between the 'lub' and 'dub', it can be an indication of cardiovascular problems, such as murmurs [9,10].

Recently, computer-aided analysis of cardiovascular sounds has complemented traditional stethoscope-based interpretation. However, for this to be feasible, algorithms capable of transferring the burden of interpreting signals from physicians to technology are crucial. The sheer volume of generated information would otherwise be overwhelming in a practical setting. This is why the field of automated analysis and interpretation of cardiovascular sounds is gaining momentum and attracting increasing attention [11,12]. Machine-learning models are utilized for analyzing pulse audio signal datasets; nevertheless, these machine-learning approaches are time-consuming and prone to variability and computational inefficiency. To overcome these limits, neural network models capable of automatic feature extraction and classification are used [13].

The significant contributions of this research work are:

1. Automate early detection of abnormal heart rhythms: This research aims to develop a technique using a neural network system that automatically identifies irregularities in heart rhythms, enabling earlier diagnosis of cardiovascular diseases and improving patient outcomes.
2. Enhance diagnostic accuracy through visual representation: By converting heart sound recordings into visual representations in the form of spectrograms, this study provides a novel approach for cardiologists to analyze cardiac health, potentially leading to more accurate diagnoses.

3. Revolutionize cardiac diagnosis with FastAI vision-learner-based neural network architectures: This research explores the potential of various neural network architectures, including ResNet, DenseNet, and others, to automate the analysis of cardiovascular audio signals. This could significantly improve diagnostic efficiency and empower cardiologists to focus on complex cases.
4. Empower individuals for a healthier society: The ultimate goal is to create an accessible and automated system for the early detection of heart disease. This empowers individuals to take control of their cardiac health and fosters a society with better overall cardiovascular wellness.

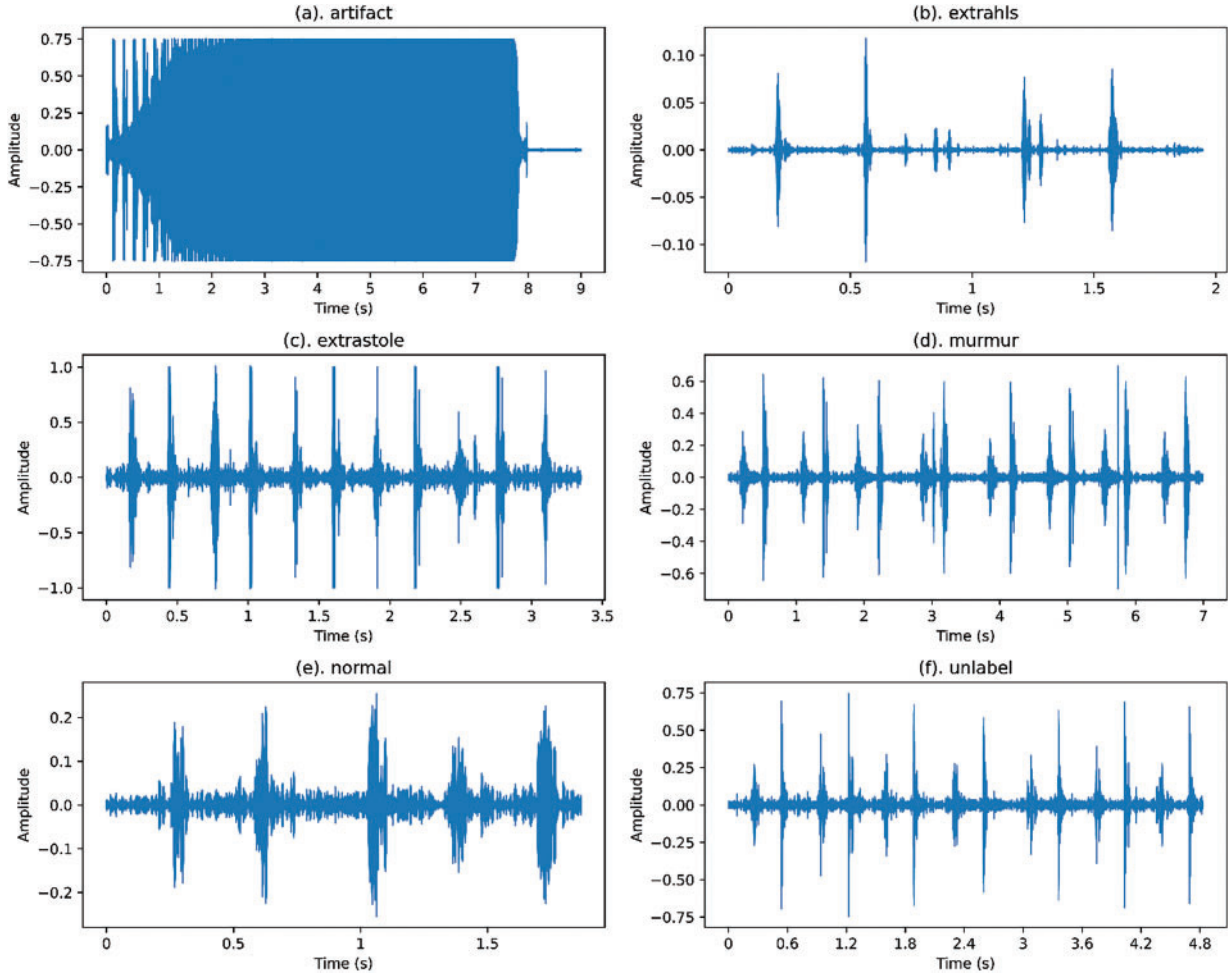

Figure 1: Waveforms of various categories of cardiovascular audio signals from the public PASCAL Challenge benchmark dataset

Despite several advantages, deep learning approaches face challenges related to dataset representativeness. The PASCAL Cardiovascular Challenge dataset has been widely used since its conception, but the lack of demographic diversity and real-world variability raises concerns about model generalizability. As such, the external validation methodology uses the Google Audio Set Heartbeat Sound dataset [14] to further show its wide representation from different conditions in cardiovascular anomalies. Additionally, bias introduced by AI diagnostics must be considered. Machine-learning models trained on imbalanced datasets may underperform on underrepresented populations, potentially leading to unfair clinical outcomes. It is vital that bias mitigation strategies be employed, such as data augmentation, fairness-aware training, and

post-hoc model calibration, to ensure equitable performance across diverse populations. On top of this, the real-world deployment introduces practical challenges around computational efficiency, interpretability, and clinical approval. Despite their accuracy, current deep learning models pose integration challenges with existing diagnostic workflows, such as real-time processing requirements, regulatory approval, and gaining clinician trust. The field must appreciate future research that looks at further lightweight, edge-compatible models capable of operating efficiently on portable devices to ease the deployment in telemedicine and point-of-care settings.

Using the public PASCAL Challenge benchmark dataset, the proposed model employs a FastAI vision-learner-based neural network architecture [15–17] and is validated on the Google Audio Set. In Section 2, a literature study is presented on the categorization of abnormal heartbeat sounds using machine learning techniques and artificial neural networks. Section 3 discusses the methodology, data, and the suggested framework. Section 4 presents the experimentation findings and analysis of these results. It includes a comparative analysis of different FastAI vision-learner-based approaches like ResNet, DenseNet, AlexNet, VGG, and ConvNext. Section 5 provides a detail on clinical implications. This article concludes in Section 6, followed by acknowledgments and references.

2 Motivation and Literature Review

In recent years, there has been significant progress in neural network architectures and end-to-end systems across diverse industries (Table 1). Deep learning, a widely adopted technique, has found applications ranging from speech recognition to autonomous vehicle driving. The success of deep learning solutions in challenging domains can be attributed to their ability to autonomously learn practical tasks, contrasting with manual, handcrafted functionalities.

A systematic review conducted by Dwivedi et al. [18] analyzed 1347 research publications from 1963 to 2018, identifying 117 peer-reviewed articles related to cardiovascular sound-based model development. The literature covered segmentation (53 publications), feature extraction techniques (72 publications), classification (88 publications), databases, and cardiovascular sound acquisition (56 publications) [19]. While automated analysis has seen substantial research, developing robust methods for identifying and classifying cardiac events remains a priority [20]. This is crucial for effective integration with wearable mobile technologies to enhance cardiovascular disease diagnosis and management [21].

Malik et al. proposed a recurrent neural network (RNN) model utilizing Long Short-Term Memory (LSTM) on the PASCAL Challenge and PhysioNet competition databases [22]. The model demonstrated high classification accuracy. However, a few limitations included reliance on only two databases, potential generalizability issues, and concerns about the loss of temporal and frequency information due to fixed sampling frames and down-sampling techniques. Narvaez et al. employed the modified empirical wavelet transform (EWT) for the preprocessing and automatic segmentation of cardiac sound signals. While achieving comparable results with state-of-the-art methodologies, challenges were noted in handling high-amplitude ambient noise, potential false positives during segmentation, and limited generalization due to specific datasets [11].

Li et al. utilized a convolutional neural network (CNN) with 497 features from eight domains. Despite achieving favorable results, the study faced challenges related to the limited dataset's impact on Deep Neural Network (DNN) performance and uncertainties about real-world effectiveness due to a lack of external validation [23,24]. Chao et al., categorized audio cardiac recordings using six machine learning models, showing favorable results but lower precision in specific cases. Concerns arose from interpretability issues, lack of detailed explanations for classifier performance, and insufficient information on feature extraction methods and handling unbalanced datasets [25].

Zeng et al. studied phonocardiogram (PCG) recordings without segmenting cardiovascular sound signals. Their experiments, using a 10-fold cross-validation approach, demonstrated excellent classification results with a dynamic neural network-based classifier. Limitations included a small database size, challenges in parameter regulation, absence of patient-specific grouping, and lack of detailed patient information [26,27]. Table 1 summarizes various vision-based methods for sound signal classification.

Table 1: A comprehensive summary of various methods in usage of vision-based methods for classification sound signals

Title	Description	Limitation
Classification of cardiovascular Sounds Using a Convolutional Neural Network [23]	The first 497 features were extracted from eight domains. Then, these features were fed into the designed convolutional neural network (CNN). Stratified five-fold cross-validation was used to evaluate the performance of the proposed method. The proposed algorithm achieves a balanced trade-off between sensitivity and specificity.	A reliance on a limited dataset for training Deep Neural Networks (DNNs) hinders their performance due to the substantial data requirements. The achieved results' sensitivity to variations in frequencies and coefficients suggests potential instability and reduced generalizability. The absence of external validation and the lack of broader context raise questions about the method's real-world effectiveness and general relevance.
ImageNet Classification with deep convolutional neural network [28]	Trained in a large, deep convolutional neural network. Classified the 1.2 million high-resolution images in the ImageNet LSVRC-2010 competition into 1000 different classes.	The network's current depth and scale remain far from matching the complexity of the human visual system's infero-temporal pathway. The absence of unsupervised pre-training and the challenge of effectively utilizing temporal information from video sequences further delineate the network's limitations.

(Continued)

Table 1 (continued)

Title	Description	Limitation
Urban Sound Tagging using Convolutional Neural Networks [29]	A framework for Environmental Sound Classification in a Low-data Context. Using pre-trained image classification models along with the usage of data augmentation techniques results in higher performance over alternative approaches.	Area Under the Precision-Recall Curve (AUPRC) non-correlation with cross-entropy poses optimization challenges. The reliance on diverse input representations and an ensemble approach, while beneficial, hints at the original model's potential shortcomings. The context of a low-resource scenario and a narrow challenge scope further suggests a need for a broader applicability assessment.
Automatic tagging using deep convolutional neural networks [30]	A content-based automatic music tagging algorithm using a fully convolutional neural net-work (FCN). A 4-layer architecture shows state-of-the-art performance with Mel-spectrogram inputs.	The presented automatic tagging algorithm doesn't address potential algorithmic limitations. It lacks exploration into the broader applicability of the proposed approach to diverse datasets and music genres, leaving room for a more comprehensive evaluation of its performance across varied contexts.
Classification of cardiovascular Sound Signal Using Multiple Features [31]	An enhanced, automated classification algorithm for cardiac disorders using cardiovascular sound signals. Extracts feature from phonocardiogram signals and then process those features using machine-learning techniques for classification.	The small dataset size could compromise the result generalizability. The study's reliance on MFCCs and DWT might not optimally manage data features. The system's narrow focus is on 4 abnormal cardiovascular disease types limiting its applicability to broader conditions. Potential enhancements lie in introducing new features for more accurate cardiovascular sound signal analysis.

Heart murmur detection and classification using machine learning techniques have gained significant attention. This is due to their potential in early diagnosis and automated cardiovascular disease detection. Fernando et al. presented a machine learning framework for heart murmur detection using phonocardiogram signals, achieving an 81.08% accuracy for murmur presence and 68.23% for clinical outcomes [32]. This study highlights the potential of machine learning in cardiovascular sound classification. Liu et al. explored deep learning models, including CNNs and transfer learning approaches, for coronary artery disease classification using phonocardiograms, achieving a 98% F1 score [33]. Mains integrated PCG and electrocardiogram (ECG) data for heart sound detection, demonstrating improved performance through multimodal fusion [34]. Behera et al. focused on machine learning models for cardiovascular disease classification but did not specifically address heart sound classification [35]. Singh et al. utilized harmonic and percussive spectral features with a deep Artificial Neural Networks (ANN) approach, achieving 93.40% accuracy [36]. Liyong and Haiyan developed a CNN-based classifier using BI spectral feature extraction, achieving an accuracy of 91.0%, sensitivity of 88.4%, and specificity of 94.0% [37]. Shakhovska and Zagorodniy applied machine learning methods like CNNs, Random Forests (RFs), and Support Vector Machines (SVMs) for acoustic tone and heart murmur classification [38]. Nkereuwem et al. proposed a system for early heart disease detection using audio signal processing with Mel-Frequency Cepstral Coefficients (MFCCs), achieving high classification accuracy through an ensemble model [39].

While various studies have explored the categorization of cardiovascular disorders using deep learning on medical datasets, challenges persist, such as interpretability, dataset limitations, and generalization concerns. The utilization of convolutional and recurrent neural networks in the context of sound signal classification presents promising avenues for further research and development in the field.

3 Research Methodology

Manual visual approaches for audio data interpretation have become widely adopted across various disciplines and applications. This fueled the notion of experimenting with image-based models on the audio dataset, and several prior works have demonstrated promising results using this approach. This section details this study's research methods.

3.1 Dataset Description

The PASCAL Cardiovascular Challenge dataset comprises labelled and unlabeled cardiac audio recordings collected using digital stethoscopes and the i-Stethoscope Pro app. The recordings range from 1 to 30 s and belong to different heart sound categories. To further evaluate the model's generalizability and real-world applicability, an external validation dataset, the Audio Set Heartbeat Sound Dataset from Google, is included. This dataset contains a variety of cardiovascular sounds, which facilitates evaluating representativeness, which is crucial given the differences in environmental, patient, and clinical conditions during recording.

The dataset is divided into two parts:

- *Dataset A* collected using the i-Stethoscope Pro app, includes the categories Normal, Murmur, Extra Heartbeat Sound, and Artifact. Since these recordings are obtained via mobile devices, ambient noise levels are generally higher.
- *Dataset B* consists of three categories: Extrasystole, Murmur, and Normal. This dataset is clinically validated but significantly imbalanced, with Normal constituting 70% of the data.

Cardiovascular sound recordings vary based on recording mode, location, background noise, and patient movement during auscultation. To standardize the dataset and ensure consistent model performance, pre-processing steps are applied, including:

- Volume normalization reduces variations in loudness from different devices.
- Noise is reduced using a low-pass filter.
- Resampling all recordings to 44.1 kHz.

AI-based diagnostics can introduce bias toward overrepresented classes in the training set. To address this, data augmentation techniques such as time-stretching, pitch shifting, and Gaussian noise injection are employed. Additionally, appropriate class weighting is applied during training to ensure balanced learning and to mitigate bias in favor of the Normal class.

Normal cardiovascular sounds in both datasets exhibit conventional patterns with a distinct 'lub'—'dub' rhythm and minimal background noise. The diastolic period exceeds the systolic period, following typical cardiac sound characteristics, and includes cardiovascular audio patterns from both active and resting individuals [40].

Murmur sounds indicate potential cardiac issues, occurring between S1 and S2 sounds and sometimes overlapping with genuine cardiovascular sounds. Extra cardiovascular sounds involve a consistent additional 'lub' or 'dub' at the end of S1 or S2, potentially signaling a medical condition or a benign variation.

Extrasystole sounds feature irregular 'lub'—'dub' sequences with additional or missing beats. They are more prevalent in children but also occur in adults (Fig. 2). Artifact sounds, consisting of background noise without distinct audible components, are primarily captured by mobile devices in uncontrolled environments, typically exceeding 195 Hz [41].

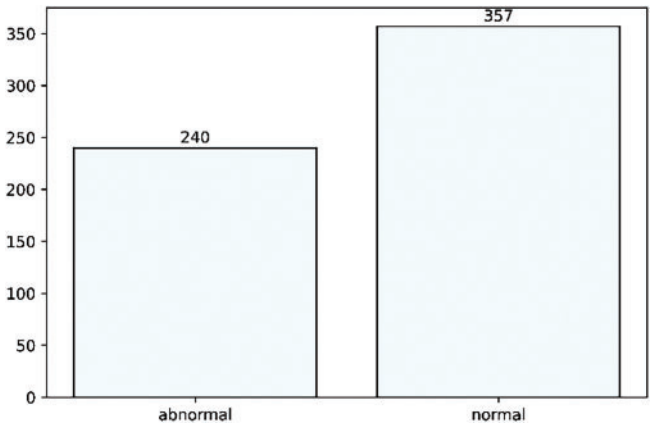

Figure 2: Dataset description of both abnormal and normal classes

This research categorizes data into normal and abnormal classes, combining all other categories under the abnormal class, resulting in a highly unbalanced dataset, as detailed in Table 2. External validation with Google's Audio Set ensures that the model generalizes beyond a single dataset. The PASCAL database is a well-known benchmark for cardiac sound classification, integrating diverse heart sound recordings from various sources. This dataset is publicly available and widely used for training machine learning models in cardiovascular disease detection, promoting fairness within the equality-of-opportunity framework.

Table 2: Dataset description of data used in this study

Category		Recording (Data A)	Recording (Data B)
Normal	Normal	31	320
Abnormal	Murmur	34	95
	Extra-Systole	19	46
	Artifacts	40	–
	Unlabeled	52	195
Total		176	656

Key advantages of this dataset include:

1. Comprehensive coverage of cardiovascular anomalies: Normal heart sounds within the S1—S2 cycle. Pathological conditions such as murmurs (indicative of turbulent blood flow due to valve defects) and extrasystoles (premature, irregular beats suggesting heart disease). Artifacts containing background noise to test model robustness.
2. Variability in real-life recording conditions: Dataset A was collected using the i-Stethoscope Pro app in uncontrolled environments (e.g., homes), resulting in ambient noise such as background conversations, breathing sounds, and electronic interference. This diversity supports robust telemedicine applications. Dataset B was recorded using a digital stethoscope in controlled clinical settings, ensuring high-fidelity recordings and medically validated detection of abnormalities.
3. Class imbalance reflecting real-life distributions: With normal recordings constituting 70% of Dataset B, the dataset mirrors real-world conditions, where healthy individuals significantly outnumber those with cardiovascular abnormalities.
4. Benchmarking capabilities for fair comparisons: The dataset is well-structured and fully annotated, supporting supervised deep-learning research. It enables standardized benchmark studies, allowing performance comparisons across different models.

In real-world applications, heart sounds are recorded under non-ideal conditions with background noise, including breathing sounds, ambient noise, and device interference. The dataset includes labeled artifacts, helping the model distinguish between noise and actual heart sounds, thereby improving real-world applicability. While the PASCAL dataset is crucial for ensuring generalization, additional validation using Google's Audio Set strengthens the model's reliability. Google's Audio Set covers a broader range of patient demographics and recording conditions, preventing overfitting to any single dataset.

3.2 Dataset Description: Preprocessing and Transformation from Audio Signal Data into Image Data

This dataset poses challenges, including acoustic noise, unreliable information, and class imbalance. Background noise is usually addressed by a low-pass filter that cuts off audio frequencies above 195 Hz, thereby reducing noise. In an optimal situation, noise and the original signal can be separated based on their frequency and amplitude components using the Fast Fourier Transform (FFT). In this work, the low-pass filter was implemented programmatically via the use of the Librosa Python module, a library for audio and music analysis. Main augmentations are time and pitch-shifting-based audio augmentations as shown in Fig. 3. Various techniques, such as the Fourier Transform and Filter-banks, are available to transform audio signals into different spectrograms. Most of the studies state that Mel-scale spectrograms outperform visual domain models. Therefore, in this paper, log/Mel-scale spectrograms were built using the Short-time

Fourier Transform (STFT) method. The basic idea is to compute an STFT using Eqs. (1) and (2), divide the raw signal into overlapping frames, and apply the window function denoted by 'w'.

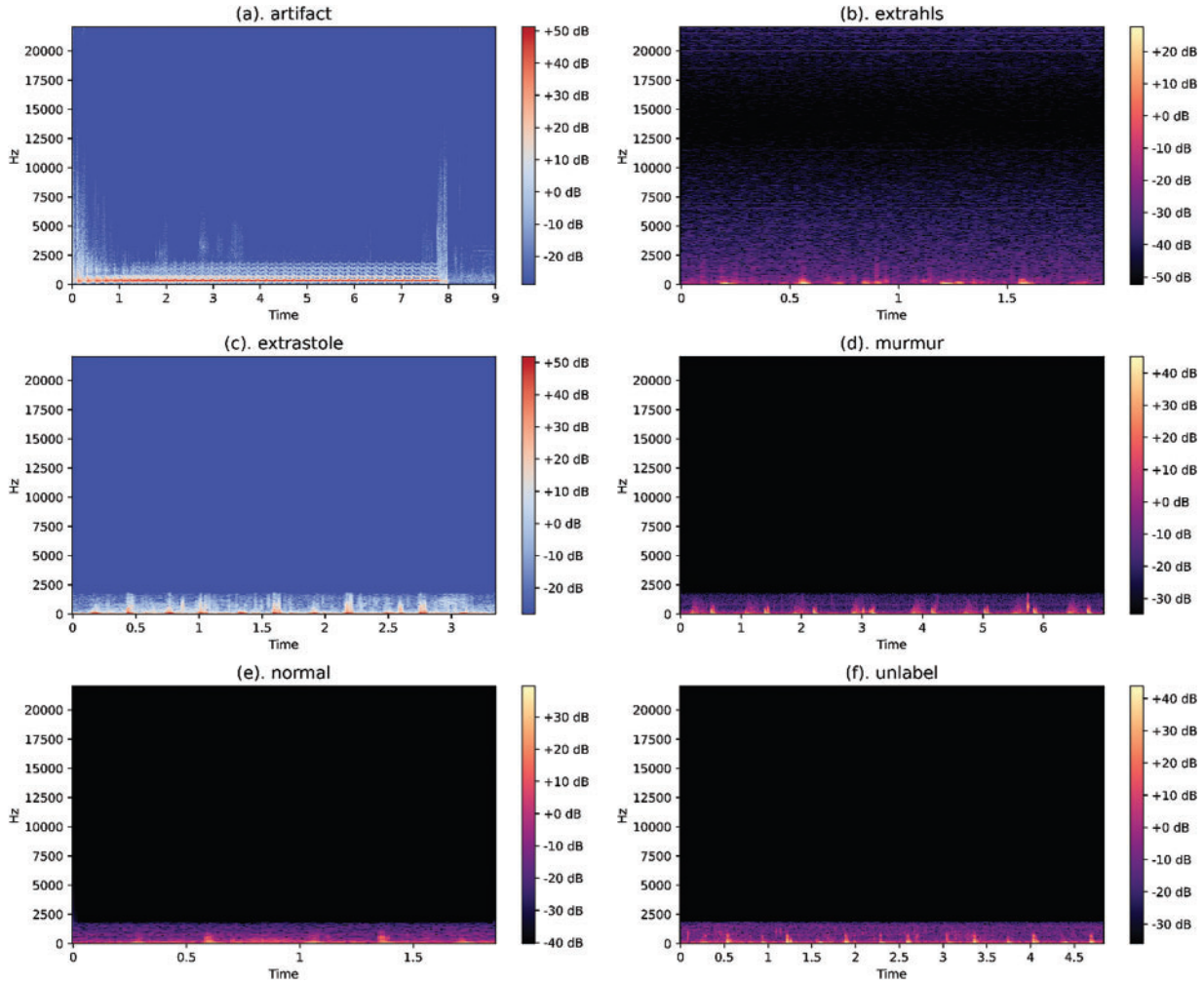

Figure 3: Examples of Normal Spectrograms on the dataset

Noise Reduction: To mitigate differences caused by various recording environments, a multi-step noise reduction pipeline is employed. A low-pass filter (cut-off = 195 Hz) is used to remove high-frequency background noise and unwanted sounds from the recording, such as microphone interference. Spectral subtraction suppresses environmental noise while preserving essential heart sound components. An adaptive volume normalization ensures standardization of recordings of varying loudness, thereby preventing discrepancies in data arising from device-specific volume variations.

Data Augmentation for Robustness: To deal with dataset imbalance while improving model generalization, we apply data augmentation techniques. Time-stretching, which slows down or speeds up the audio but doesn't affect pitch. Pitch shifting, altering the modified frequency characteristics while maintaining the shape of the waveform. Injection of Gaussian noise, which simulates real-world distortions of signals to make the model more robust to noise. Random time masking, inspired by Spec Augment, a randomized dropout introduced in spectrograms to make the model more resistant to variation within the dataset.

Spectrogram Conversion: The cardiovascular audio signals, which are preprocessed through the application of an STFT, are transformed into image-like formats suitable for deep learning in the following process. This involves computing magnitude spectrograms using STFT. Applying Mel-scale transformations in capturing the human-audible frequency feature. Normalization of spectrogram pixel values from 0 to 1 to stabilize training.

Following the computation, magnitude spectrograms are generated, and frequencies are warped to the Mel-scale. The Mel-frequency bins are created by combining FFT bins, resulting in a Mel-spectrogram. The Mel-scaled power spectrogram is then obtained by squaring the magnitude spectrogram and multiplying it by the Mel-filter bank, as per Eqs. (1)–(3).

$$STFT\left(\{x(t)\}(\tau,\omega)\right) \equiv x\left(\tau,\omega\right) = \sum_{-\infty}^{\infty}\left(s(t)w(t-\tau)e^{-i\omega t}\right)dt \tag{1}$$

$$STFT\left(\{x[n]\}(m,\omega)\right) \equiv x\left(m,\omega\right) = \sum_{-\infty}^{\infty}\left(x[n]w(n-m)e^{-i\omega n}\right)dt \tag{2}$$

$$m = 2595\ log\ 10(1+f/700) \tag{3}$$

The Mel-spectrograms, derived through this operation, are displayed as Red Green Blue (RGB) images to visualize frequency changes over time and amplitude variations, as shown in Fig. 4.

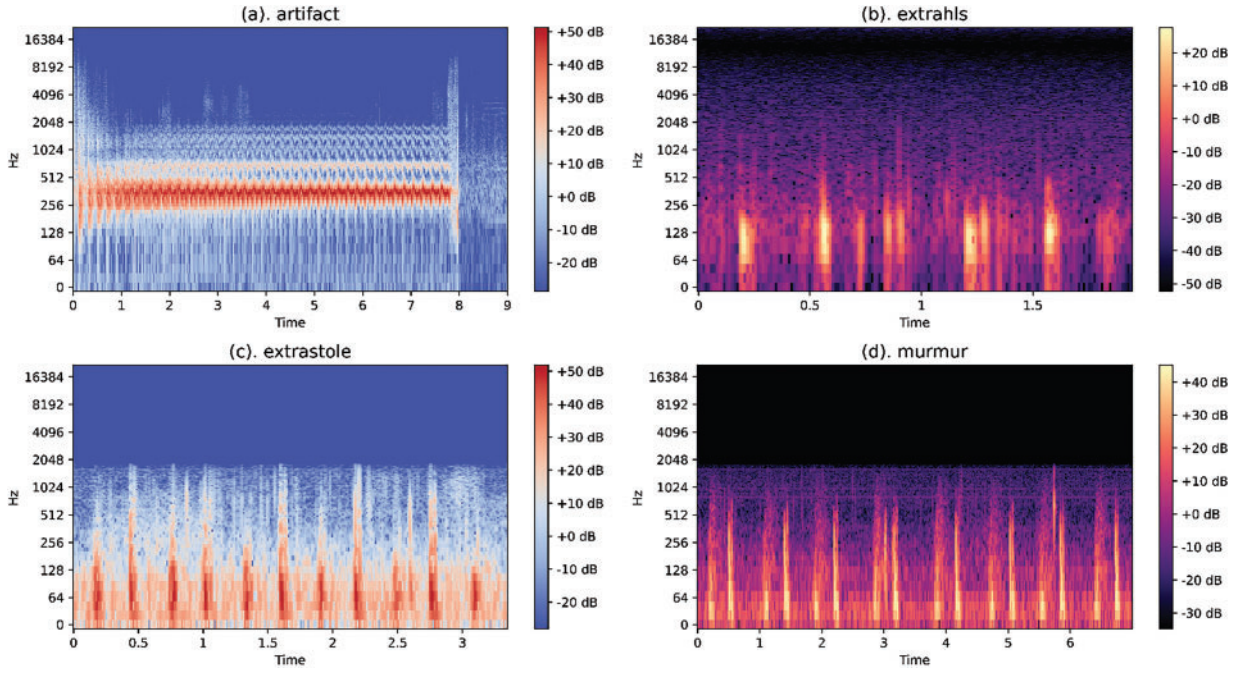

Figure 4: (Continued)

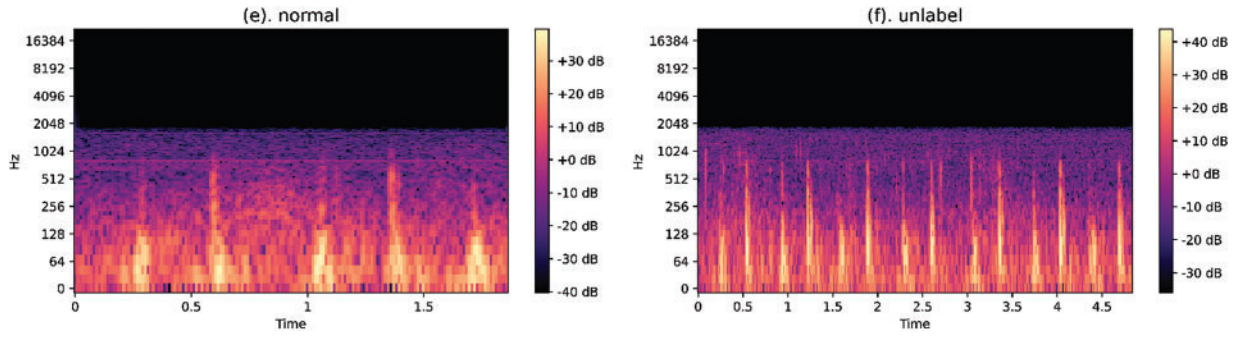

Figure 4: Examples of log/Mel-spectrogram on the dataset

This study uses spectrogram augmentation techniques to enhance data availability for training and validation, thereby improving generalizability. Before integration with deep learning models, all enhanced colored spectrogram images undergo pixel value scaling, ensuring values fall within the 0 to 1 range. A simple division by 255 scales pixel values, which ranged from 255 to 1 in the research images, to the required range. The normalized images are resized to 128*128 dimensions, optimizing results within the hardware capacity used during the training process.

3.3 FastAI Architecture

FastAI, a deep learning library, is meticulously designed to provide high-level components that enable rapid experimentation and learning, facilitating rapid development of state-of-the-art models in standard deep learning domains. It also offers customizable low-level components that support the development of novel techniques. The library operates under the principles of accessibility, rapid productivity, deep hackability, and flexibility, striking a balance between usability and performance without compromise [42].

DenseNet, or Dense Convolutional Network, proposes a unique CNN architecture in which each layer connects directly to all subsequent layers to maximize information flow. Unlike conventional networks, DenseNet optimizes feature integration by concatenating them, which reduces the number of parameters and avoids training redundant feature maps (Fig. 5). The DenseNet architecture includes Transition Layers and Dense Blocks, beginning with an initial convolutional and pooling layer, followed by an alternating sequence of dense blocks and transition layers. This repeated architectural pattern culminates in a dense block succeeded by a classification layer.

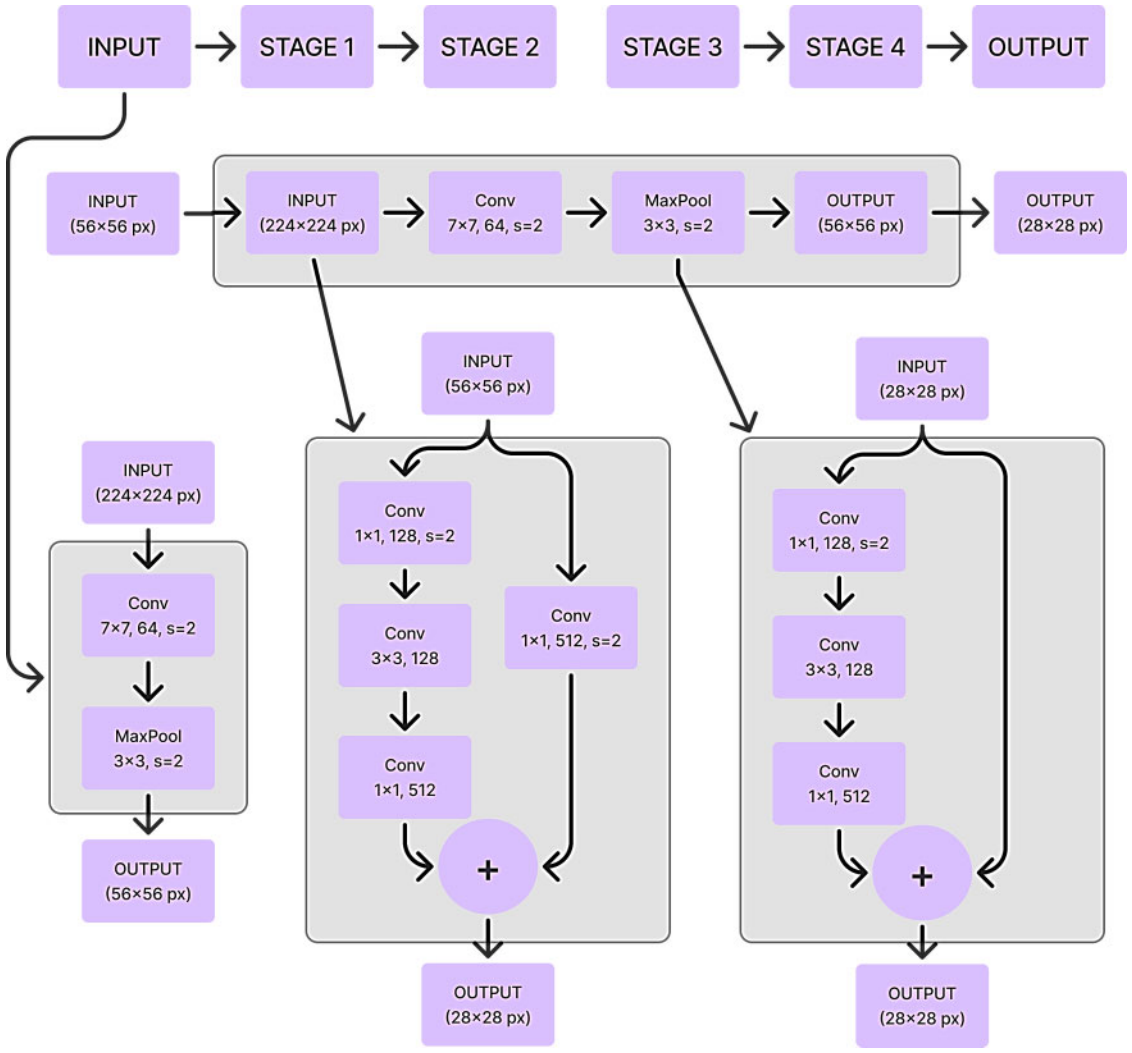

Figure 5: DenseNet121 the fastai architecture description

ResNet, short for Residual Network, is a powerful deep neural network known for its excellent generalization in recognition tasks, making it a popular choice across many computer vision tasks (Fig. 6). The FastAI library incorporates ResNet models with configurable depths ranging from 18 to 152 layers [43].

VGG-19_bn is a deep 19-layer CNN, known for its simple yet effective architecture. It consists of 16 convolutional layers followed by 3 fully connected layers and with Batch Normalization (BN) applied after each convolutional layer, to improve convergence and training stability, as shown in Table 3.

Model Performance Comparison: To compare the strengths and weaknesses of these architectures, we evaluate key aspects such as efficiency, accuracy, and computational cost, as shown in Table 4.

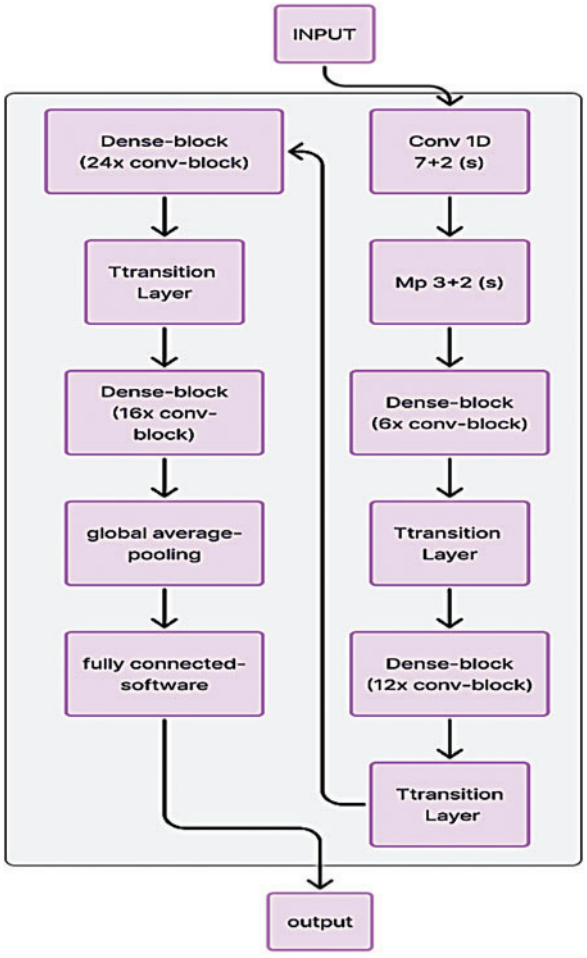

Figure 6: Resnet50 of the fastai architecture description

Table 3: VGG-19 layers

Layer type	Count
Convolutional layer	16
Max pooling layers	5
Full connected layers	3
SoftMax output layer	1

Table 4: Model performance comparison

Model	Strengths	Weaknesses
DenseNet-201	Efficient feature reuse, deep feature extraction.	High Video Random Access Memory usage, slower training.
ConvNeXt-Large	Modernized CNN, strong, large-scale dataset performance.	High number of parameters, needs strong Graphics Processing Units (GPUs).
ResNet-152	Deep network with skip connections prevents vanishing gradients.	Computationally expensive for training.

(Continued)

Table 4 (continued)

Model	Strengths	Weaknesses
VGG-19_bn	Simple architecture, easy to interpret, stable with BatchNorm.	High memory consumption, poor generalization.
DenseNet-121	Fewer parameters than ResNet, efficient gradient flow.	Limited expressiveness due to shallower depth.
DenseNet-169	Balanced depth vs. efficiency, good feature reuse.	Slower inference compared to ResNets.
ResNet-101	Strong generalization ability, deep feature extraction.	Requires more training time than ResNet-50.
ResNet-50	Faster training and inference compared to deeper ResNets.	Slightly lower performance than deeper variants.
SqueezeNet	Lightweight, optimized for mobile deployment.	Lower accuracy than deeper models.
AlexNet	High-speed inference, useful for benchmarking.	Outdated, lower accuracy than modern CNNs.

3.4 Proposed Algorithm and Methodology

The proposed algorithm and methodology involve processing cardiovascular audio classification data through a structured pipeline, as outlined in Algorithm 1. The dataset (nha) undergoes a series of preprocessing steps, resulting in a processed dataset (P). Each data instance is iterated through to generate spectrograms (spec), which are further split into training and validation sets in an 80–20 ratio. These spectrograms are prepared using data loaders with transformations like resizing. A model is then trained using the FastAI framework by initializing a vision learner and optimizing it over multiple training cycles. Metrics such as accuracy, F1-score, and Area Under the Curve (AUC) are computed to evaluate the model's performance, and the training process is summarized to provide insights into the learning dynamics. This approach provides an efficient pipeline for cardiovascular audio classification. The flow diagram of the proposed system can be found in Fig. 7.

Algorithm 1: FastAI-based cardiovascular audio classification using spectrograms

a. nha ← Load cardiovascular audio classification data
b. Apply preprocessing steps to nha → preprocessed_data
c. For each recording record in preprocessed_data
d. Generate spectrogram(s) from record → spectrogram
e. Append spectrogram to spectrograms
f. Split spectrograms` into training and validation sets:
g. train_set, val_set ← train_val_split(spectrograms, ratio = 0.8)
h. Load image data using FastAI:
i. dataloaders ← ImageDataLoaders.from_folder(...) with resize/item_tfms
j. Initialize a CNN model:
k. learner ← fastai.vision_learner(dataloaders, pretrained_model)
l. Train the model using 1-cycle policy:
m. For cycle from 1 to N: → learner.fit_one_cycle(cycle_size)

(Continued)

Algorithm 1 (continued)

n. Evaluate model:
o. preds, targets ← learner.get_preds()
p. Accuracy ← accuracy(preds, targets)
q. F1 Score ← f1_score(preds, targets)
f. AUC ← auc_score(preds, targets)

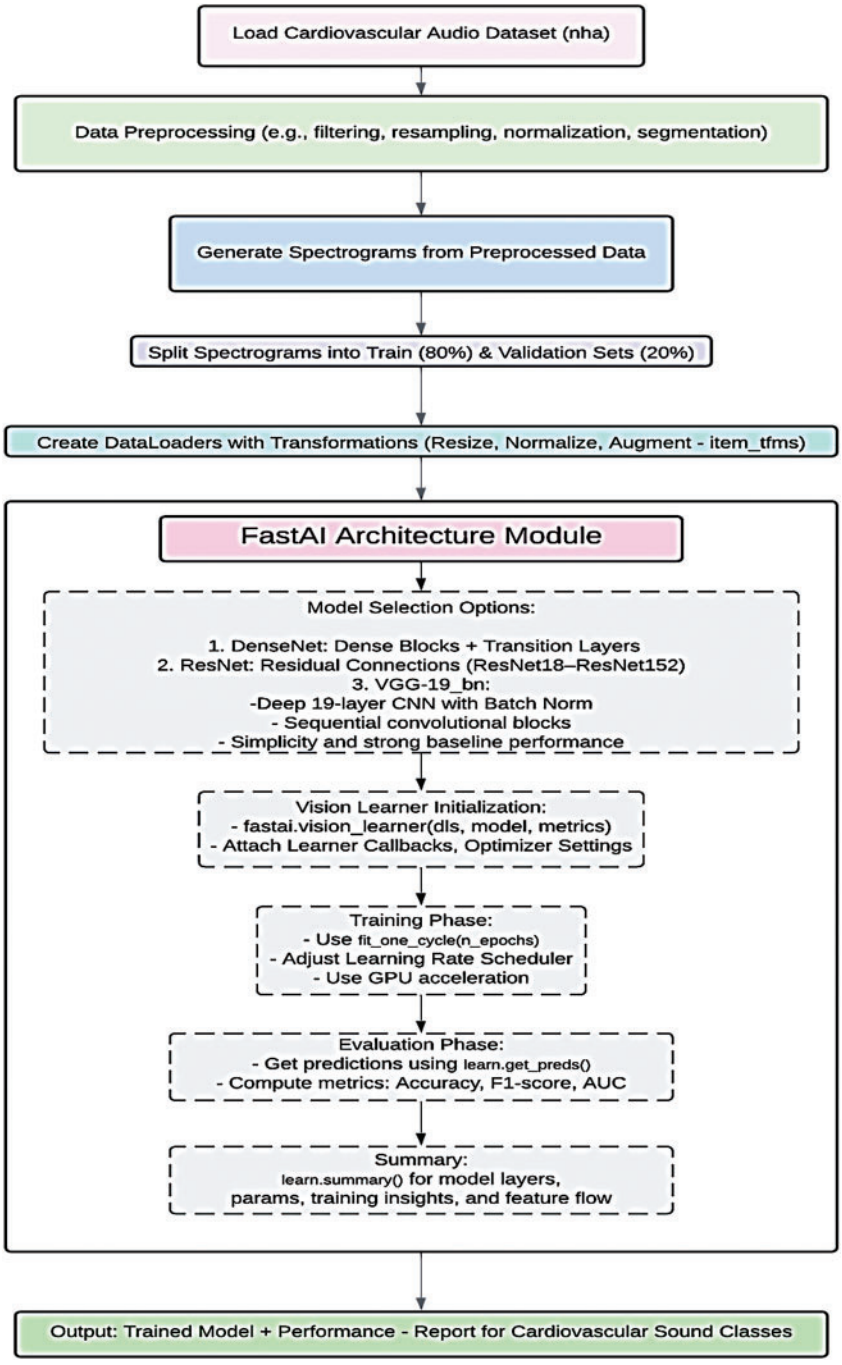

Figure 7: Flow diagram of the proposed system

3.5 Evaluation Criteria

The computational modeling for this experiment was conducted using a high-performance system with an Intel Core i7-12700H processor, NVIDIA RTX 3060 GPU, and 16 GB RAM. The system ran on Linux (Ubuntu 22.04 LTS) [44]. Software tools included Python 3.2 for computations, along with TensorFlow, PyTorch, and Scikit-learn for machine learning, while CUDA 11.7, OpenMP 5.1, and MPI 4.0 enabled parallel computing. This configuration ensured efficient and reliable execution of computational experiments.

For appraising a binary classification model, key metrics include accuracy, precision, F1 score, and the AUROC (Area under the Receiver Operating Characteristic curve) score. Precision, focusing on True Positives, offers insights into imbalanced datasets. The AUROC score, derived from true positive and false positive rates, assesses a model's class discrimination ability through the Receiver Operating Characteristic (ROC) curve. This metric remains effective in handling imbalanced datasets. Eqs. (4)–(7) detail the calculation of accuracy, precision, recall, and F1-Score on the specific dataset.

$$Accuracy\ (Ac) = \frac{t(P) + t(N)}{t(P) + t(N) + f(P) + f(N)} \tag{4}$$

$$Precision\ (Pr) = \frac{t(P)}{t(P) + f(P)} \tag{5}$$

$$Recall\ (Re) = \frac{t(P)}{t(P) + f(N)} \tag{6}$$

$$F1 - Score\ (F1S) = 2 * \frac{Pr * Re}{Pr + Re} \tag{7}$$

4 Results and Analysis

Various classification models are compared based on evaluation metrics such as accuracy, precision, F1 score, AUROC curve, number of epochs, and loss (Figs. 8 and 9, Tables 5–7). Models exhibiting higher precision demand significant processing resources and place a substantial burden on the GPU. Consequently, a trade-off emerges between the precision required and the number of parameters to be configured. This evaluation sheds light on the intricate balance needed in setting up models, considering computational intensity and precision as interconnected factors.

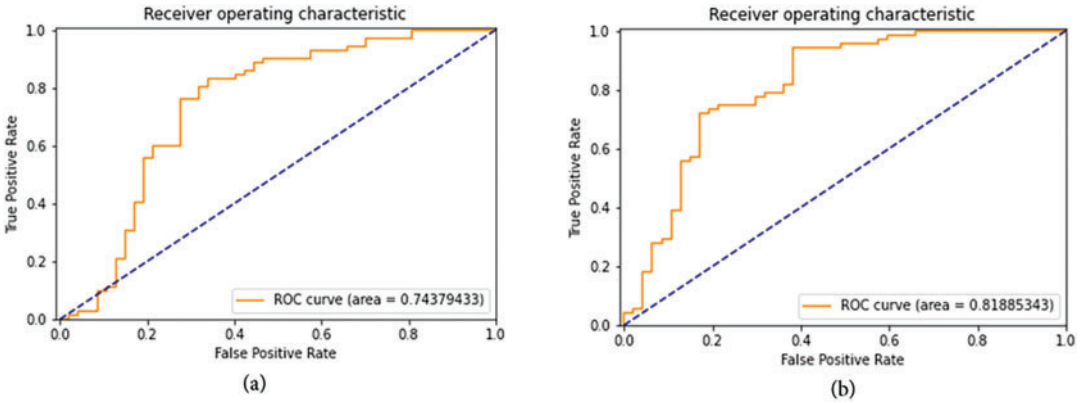

Figure 8: (Continued)

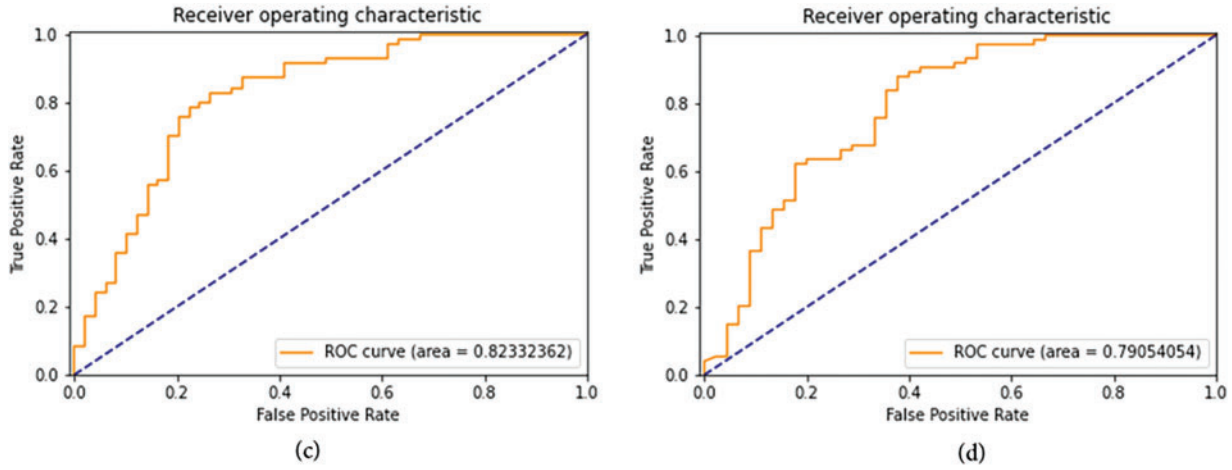

Figure 8: ROC curve and AUROC (Area under ROC Curve) scores for the first set of FastAI vision-learner-based models. (**a**) ResNet50 ROC curve and AUROC (**b**) ResNet152 ROC curve and AUROC (**c**) VGG-19_bn ROC curve and AUROC (**d**) DenseNet169 ROC curve and AUROC

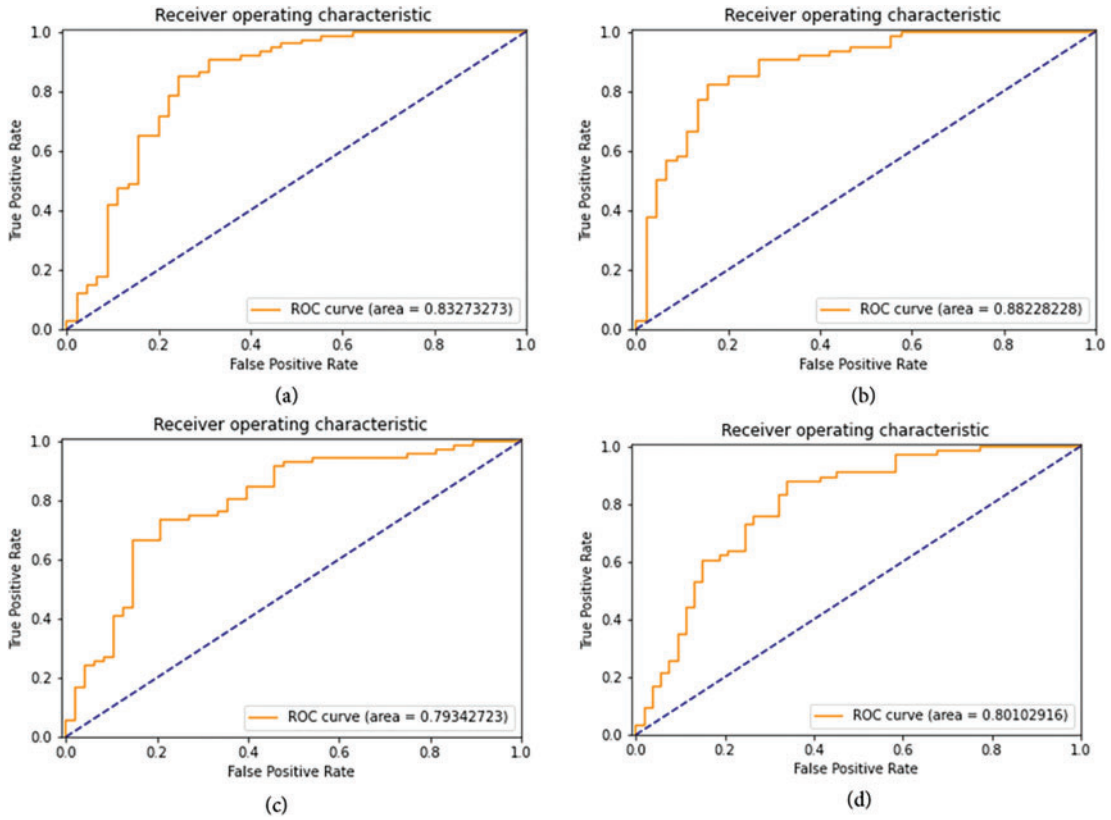

Figure 9: ROC curve and AUROC (Area Under ROC Curve) scores for the second set of FastAI vision-learner-based models. (**a**) DenseNet201 ROC curve and AUROC (**b**) ConvNeXt-Large ROC curve and AUROC (**c**) SqueezeNet 1.1 ROC curve and AUROC (**d**) ResNet101 ROC curve and AUROC

Table 5: Comparative summarization of training time and VRAM usage of all the FastAI vision-learner-based models applied over the dataset

Model name	Trainable parameters	VRAM usage	Memory footprint	Training time (per epoch)
DenseNet201	18 M	7.8 GB	12 GB	~3 min
ConvNeXt-Large	197 M	12.3 GB	18 GB	~3 min
ResNet152	60 M	10.2 GB	15 GB	~3 min
VGG19_bn	144 M	9.5 GB	14 GB	~2.5 min
DenseNet121	8 M	6.5 GB	10 GB	~2.5 min
DenseNet169	14 M	7.2 GB	11 GB	~2 min
ResNet101	44 M	9 GB	13 GB	~3 min
ResNet50	26 M	8.5 GB	12 GB	~3 min

Table 6: Comparative summarization results of all the FastAI vision-learner-based models applied over the dataset

Model name	Accuracy (Ac)	Precision (Pr)	F1 score (F1)	AUROC	Train loss	Valid loss	Total epochs
DenseNet201	0.82	0.82	0.82	0.82	0.82	0.82	60
ConvNeXt-Large	0.8	0.82	0.82	0.82	0.82	0.82	60
ResNet152	0.8	0.83	0.83	0.83	0.83	0.83	60
VGG19_bn	0.79	0.03	0.03	0.03	0.03	0.03	40
DenseNet121	0.77	1.1	1.1	1.1	1.1	1.1	60
DenseNet169	0.76	0.8	0.8	0.8	0.8	0.8	25
ResNet101	0.76	0.8	0.8	0.8	0.8	0.8	60
ResNet50	0.76	0.88	0.88	0.88	0.88	0.88	50

Table 7: Comparative summarization of results for all the FastAI vision-learner-based models applied over the dataset on different classes

Model	Normal (Acc, Pr, Rc, F1)	Murmur (Acc, Pr, Rc, F1)	Extrasystole (Acc, Pr, Rc, F1)	Artifacts (Acc, Pr, Rc, F1)
Densenet201	0.85, 0.83, 0.84, 0.83	0.80, 0.78, 0.79, 0.78	0.79, 0.75, 0.76, 0.75	0.82, 0.80, 0.81, 0.80
ConvNext_Large	0.83, 0.82, 0.81, 0.81	0.81, 0.79, 0.78, 0.78	0.78, 0.74, 0.75, 0.74	0.81, 0.79, 0.80, 0.79
ResNet152	0.84, 0.83, 0.82, 0.82	0.82, 0.80, 0.79, 0.79	0.80, 0.76, 0.77, 0.76	0.83, 0.81, 0.82, 0.81
VGG19_BN	0.79, 0.03, 0.04, 0.03	0.78, 0.02, 0.03, 0.02	0.77, 0.02, 0.03, 0.02	0.78, 0.02, 0.03, 0.02

(Continued)

Table 7 (continued)

Model	Normal (Acc, Pr, Rc, F1)	Murmur (Acc, Pr, Rc, F1)	Extrasystole (Acc, Pr, Rc, F1)	Artifacts (Acc, Pr, Rc, F1)
Densenet121	0.81, 1.1, 1.2, 1.1	0.80, 1.1, 1.2, 1.1	0.78, 1.0, 1.1, 1.0	0.80, 1.1, 1.2, 1.1
Densenet169	0.82, 0.80, 0.79, 0.79	0.79, 0.78, 0.77, 0.77	0.76, 0.74, 0.73, 0.73	0.80, 0.78, 0.77, 0.77
ResNet101	0.82, 0.80, 0.79, 0.79	0.79, 0.78, 0.77, 0.77	0.76, 0.74, 0.73, 0.73	0.80, 0.78, 0.77, 0.77
ResNet50	0.83, 0.88, 0.87, 0.87	0.81, 0.85, 0.84, 0.84	0.80, 0.83, 0.82, 0.82	0.82, 0.86, 0.85, 0.85

Within the spectrum of FastAI vision-learner models, DenseNet201, ConvNeXt-Large, and ResNet152 exhibited exceptional performance, as detailed in Tables 6 and 7. However, the VGG19_bn model showed unusually low precision, recall, and F1-score values (all at 0.03), indicating potential limitations. One reason could be its shallow architecture, lacking skip connections and efficient feature reuse, which are crucial for capturing detailed cardiovascular sound patterns. Additionally, overfitting and poor generalization might have affected its performance, possibly due to ineffective batch normalization layers or small batch sizes. The vanishing gradient issue further compounds the problem, as the deep layers without residual connections result in weak learning in later layers. Moreover, VGG-based models are originally designed for image classification, making them less adaptable to time-series or spectrogram data, which may hinder their ability to process cardiovascular sound variations. Suboptimal hyperparameter tuning, such as improper learning rates or weight decay, could have also prevented proper convergence. Lastly, imbalanced data handling might have biased the model towards the dominant 'Normal' class, leading to near-zero precision and recall for minority classes. These factors highlight the importance of selecting more advanced architectures and refining hyperparameters to enhance cardiovascular sound classification.

4.1 External Validation on Google AudioSet Dataset

To assess model generalizability, we tested our trained models on Google's Audio Set Heartbeat Sound dataset, which contains a broader range of cardiovascular sound samples across varied recording conditions and patient demographics shown in Table 8 shows the external validation on different datasets other than Pascal shows promising results.

Table 8: Performance comparison over different datasets: PASCAL challenge vs. google audio set

Model	PASCAL accuracy (%)	Google audio set accuracy (%)	Precision	Recall	F1-score
DenseNet-201	82	79.5	80.2	78.8	79.5
ResNet-152	80	77.2	78.5	76.5	77.4
ConvNeXt-L	80	78.1	79	77	78
VGG19_bn	79	67.4	70	64.5	67.1
ResNet-50	76	74.3	75.5	72.8	74.1

4.2 Ablation Study: Evaluating Model Components

To analyze the impact of different components of the proposed model, an ablation study was conducted by systematically removing or modifying key processing steps and observing their effects on model performance. This analysis helps in understanding the contribution of noise reduction, data augmentation, and model depth to overall classification accuracy and generalization.

4.3 Impact of Noise Reduction

Noise reduction is a critical step in pre-processing, as heart sound recordings contain background interference such as breathing sounds, stethoscope friction, and ambient noise. To evaluate its impact, the model was trained both with and without noise reduction techniques, including low-pass filtering and spectral subtraction. The results indicate that removing noise reduction significantly decreases accuracy and other evaluation metrics, leading to a higher misclassification rate, particularly for abnormal heart sounds (Table 9).

Table 9: Impact of noise reduction

Configuration	Accuracy (%)	Precision	F1-score	AUROC
With noise reduction	82	82.5	81.8	82.4
Without noise reduction	76.8	74.2	73.9	75.5

5 Clinical Implications Discussion

The integration of deep learning-based cardiovascular sound classification into clinical practice can improve early detection, diagnosis, and monitoring of cardiovascular diseases. However, real-world deployment faces challenges such as ethical concerns, dataset biases, and practical implementation barriers.

5.1 Clinical Implications and Real-World Integration

The clinical integration of machine learning for diagnosing cardiovascular sounds adds learning capabilities that can positively impact the process of disease identification, prognostication, and cardiovascular disease monitoring in both community and hospital settings. The largest limitation to the real-world utilization of these findings is ethical concerns, dataset biases, and challenges in the practical implementation of such AI-based diagnosis. The AI-based model for early detection and automated diagnosis can infer and analyze heart sounds with a rapid efficiency that supersedes manual auscultation. Early identification enables timely intervention for heart murmurs, arrhythmias, and other abnormalities, potentially preventing severe cardiac events. Augmenting Physician Expertise: The system will assist General Practitioners (GPs) and non-cardiologists in making the diagnosis of heart conditions. An AI-based screening test would serve as a pre-diagnostic level check before consultation with a cardiologist. Telemedicine and Remote Monitoring: Integration of the AI model with digital stethoscopes and mobile applications. This would be valuable for telehealth consultation and also widen accessibility across rural or underserved areas. Decision Support in Hospitals: AI could complement Electrocardiogram (ECG) and echocardiography for a multi-modal diagnostic approach. AI-driven alerts could prioritize high-risk patients and improve the efficacy of triaging.

5.2 Ethical Considerations and Bias in AI-Based Diagnosis: Dataset Bias and Fairness of the Model

There are few existing datasets that adequately reflect population diversity (age, gender, ethnicity, conditions) on a global scale. The existing datasets primarily represent normal populations, reducing sensitivity to rare abnormalities. Mitigation Strategies: Use of diverse datasets from many hospitals and patient demography. Employing bias correction techniques such as re-weighted loss functions and adversarial debiasing. Continuously training the model with real in-hospital data to enhance performance. Transparency and Explainability: Black-box AI give clinicians difficulty in understanding the reasoning behind clinical decisions. Saliency maps and attention visualization on spectrograms can assist clinicians in understanding the AI predictions. Data Privacy and Security: AI diagnosis, being dependent on acoustic data and patient health records, raises privacy-related concerns. Encryption and federated learning allow modelling based on always-decentralized hospital data with patient confidentiality protected.

5.3 Real-World Deployment Challenges: Cost and Infrastructure Constraints

AI inference generally requires high-end GPUs and cloud computing, the cost of which may be too high for small clinics to afford. Developing countries may need lightweight AI models optimized for mobile devices. Optimizing the model through quantization and knowledge distillation can help reduce the computational cost. Deploying AI diagnostics on edge devices (portable stethoscopes and mobile apps) will enhance their accessibility. AI models trained on one dataset often struggle to generalize across diverse hospital environments. Variations in heart sound recordings (e.g., stethoscope quality, patient movement, background noise) can affect accuracy. Domain adaptation techniques and periodic retraining with new patient data can improve real-life model performance. Physicians may be reluctant to trust AI decisions without prior clinical validation. AI models must also comply with regulatory standards (FDA, CE, HAPS) before clinical use. Combined with prospective clinical trials and the integration of human-in-the-loop systems, is crucial for making AI trustworthy and safe.

6 Conclusion and Future Research Direction

This research presents a cardiovascular sound classification system based on deep learning that uses FastAI vision-learner architectures to detect abnormal heart sounds. Models were created from the preprocessed heart sounds in these spectrograms, employing techniques associated with CNN models such as DenseNet-201, ConvNeXt-Large, and ResNet-152 to achieve high classification accuracy. It is a proposed system that could play an important role in the early diagnosis of cardiovascular diseases, benefiting healthcare personnel as well as applications in telemedicine. In performance evaluations, DenseNet-201 on the PASCAL Cardiovascular Challenge dataset achieved the highest classification accuracy, followed by ConvNeXt-Large and ResNet-152. Whereas, in the second phase, the real-world applicability was assessed through the model's validation over Google's Audio Set Heartbeat Sound dataset, which showed a slight drop in accuracy (~2%–5%), highlighting the variability and noise conditions from datasets in real-world applications.

Future research in cardiovascular sound classification can explore several promising directions to enhance diagnostic accuracy and clinical applicability. One key area is multimodal learning, where heart sounds are combined with other physiological signals such as ECG or patient metadata to provide richer context for diagnosis. Advanced data augmentation strategies, such as Spec Augment or Generative Adversarial Networks (GANs) for audio, can further enrich training datasets. Establishing open benchmarks and community challenges would facilitate standardized evaluation, while federated learning enables privacy-preserving, distributed model training across institutions. Furthermore, cross-domain generalization and domain adaptation should be investigated to ensure robustness across diverse clinical environments and

datasets. Class imbalance can be addressed using techniques like focal loss, or synthetic data generation using GANs can improve model fairness.

Acknowledgement: The authors thanks the deanship of scientific research (DSR), King Abdulaziz University, Jeddah for supporting this study.

Funding Statement: The project was funded by the deanship of scientific research (DSR), King Abdulaziz University, Jeddah, under grant No. (G-1436-611–309). The authors, therefore acknowledge with thanks DSR technical and financial support.

Author Contributions: Conceptualization, study conception and design: Deepak Mahto, Sudhakar Kumar, Sunil K. Singh, Amit Chhabra; data curation, formal analysis, methodology, writing—original draft preparation: Deepak Mahto, Sudhakar Kumar, Sunil K. Singh, Amit Chhabra; formal analysis: Irfan Ahmad Khan; validation, visualization: Varsha Arya, Wadee Alhalabi; investigation, project administration: Sunil K. Singh, Brij B. Gupta; writing—review and editing: Sudhakar Kumar, Sunil K. Singh; funding acquisition: Brij B. Gupta, Bassma Saleh Alsulami, Wadee Alhalabi. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All data generated or analyzed during this study are collected from the following public repository: https://istethoscope.peterjbentley.com/heartchallenge/index.html (accessed on 22 May 2025), https://research.google.com/audioset/dataset/heart_sounds_heartbeat.html (accessed on 22 May 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation. 2022 Feb;145(8):e153–639. doi:10.1161/cir.0000000000001074.
2. Shuvo SB, Ali SN, Swapnil SI, Al-Rakhami MS, Gumaei A. CardioXNet: a novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. IEEE Access. 2021;9:36955–67. doi:10.1109/ACCESS.2021.3063129.
3. Riegel B, Moser DK, Buck HG, Dickson VV, Dunbar SB, Lee CS, et al. Self care for the prevention and management of cardiovascular disease and stroke. J Am Heart Assoc. 2017 Sep;6(9):e006997.
4. WHO. Cardiovascular diseases (CVDs); 2021 [Internet]. [cited 2025 May 22]. Available from: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
5. Nath M, Srivastava S, Kulshrestha N, Singh D. Detection and localization of S1 and S2 heart sounds by 3rd order normalized average Shannon energy envelope algorithm. Proc Inst Mech Eng Part H: J Eng Med. 2021;235(6):615–24. doi:10.1177/0954411921998108.
6. Ahamed J, Manan Koli A, Ahmad K, Alam Jamal M, Gupta BB. CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning. Int J Interact Multimed Artif Intell. 2022;7(4):78–86.
7. Lekic M, Lekic V, Riaz IB, Mackstaller L, Marcus FI. The cardiovascular physical examination—is it still relevant? Am J Cardiol. 2021 Jun;149:140–4. doi:10.1016/j.amjcard.2021.02.042.
8. Heart disease; 2022 [Internet]. [cited 2025 May 22]. Available from: https://www.mayoclinic.org/diseases-conditions/heart-disease/diagnosis-treatment/drc-20353124.
9. Montinari MR, Minelli S. The first 200 years of cardiac auscultation and future perspectives. J Multi-discip Healthc. 2019 Mar;12:183–9.
10. Liu Y, Chen J, Bao N, Gupta BB, Lv Z. Survey on atrial fibrillation detection from a single-lead ECG wave for internet of medical things. Comput Commun. 2021;178:245–58.

11. Narváez P, Gutierrez S, Percybrooks WS. Automatic segmentation and classification of heart sounds using modified empirical wavelet transform and power features. Appl Sci. 2020 Jul;10(14):4791. doi:10.3390/app10144791.
12. da Silva-Oolup SA, Giuliano D, Stainsby B, Thomas J, Starmer D. Evaluating the baseline auscul-tation abilities of second-year chiropractic students using simulated patients and high-fidelity manikin simulators: a pilot study. J Chiropr Educ. 2022 Oct;36(2):172–8. doi:10.7899/jce-21-1.
13. Ren H, Jin H, Chen C, Ghayvat H, Chen W. A novel cardiac auscultation monitoring system based on wireless sensing for healthcare. IEEE J Transl Eng Health Med. 2018;6:1–12. doi:10.1109/jtehm.2018.2847329.
14. AudioSet. (n.d.). Retrieved March 27, 2025 [Internet]. [cited 2025 May 22]. Available from: https://research.google.com/audioset/dataset/heart_sounds_heartbeat.html.
15. Liu C, Springer D, Moody B, Silva I, Johnson A, Samieinasab M, et al. Classification of heart sound recordings: the PhysioNet/computing in cardiology challenge. Physiol Meas. 2016 Dec;37(12):2181–213.
16. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. Physiol Meas. 2016 Dec;37(12):2181–213. doi:10.1088/0967-3334/37/12/2181.
17. Bentley P, Nordehn G, Coimbra M, Mannor S. The PASCAL classifying heart sounds challenge 2011 (CHSC2011) results; 2011 [Internet]. [cited 2025 May 22]. Available from: http://www.peterjbentley.com/heartchallenge/index.html.
18. Dwivedi AK, Imtiaz SA, Rodriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds-A systematic review. IEEE Access. 2019;7:8316–45. doi:10.1109/access.2018.2889437.
19. Nogueira DM, Ferreira CA, Gomes EF, Jorge AM. Classifying heart sounds using images of motifs, MFCC and temporal features. J Med Syst. 2019 Jun;43(6):168. doi:10.1007/s10916-019-1286-5.
20. Raza A, Mehmood A, Ullah S, Ahmad M, Choi GS, On B-W. Heartbeat sound signal classification using deep learning. Sensors. 2019 Nov;19(21):4819. doi:10.3390/s19214819.
21. Gomes EF, Bentley PJ, Pereira E, Coimbra MT, Deng Y. Classifying heart sounds—approaches to the PASCAL challenge. In: Proceedings of the International Conference on Health Informatics (HEALTHINF-2013); 2013; Barcelona, Spain. p. 337–40.
22. Malik H, Bashir U, Ahmad A. Multi-classification neural network model for detection of abnormal heartbeat audio signals. Biomed Eng Adv. 2022 Dec;4(100048):100048. doi:10.1016/j.bea.2022.100048.
23. Li F, Tang H, Shang S, Mathiak K, Cong F. Classification of heart sounds using convolutional neural network. Appl Sci. 2020 Jun;10(11):3956. doi:10.3390/app10113956.
24. Li F, Liu M, Zhao Y, Kong L, Dong L, Liu X, et al. Feature extraction and classification of heart sound using 1D convolutional neural networks. EURASIP J Adv Signal Process. 2019 Dec;2019(1):117. doi:10.1186/s13634-019-0651-3.
25. Chao A, Ng S, Wang L. Listen to your heart: feature extraction and classification methods for heart sounds; 2023 [Internet]. [cited 2025 May 22]. Available from: https://lindawangg.github.io/projects/classifying-heartbeats.pdf.
26. Zeng W, Yuan J, Yuan C, Wang Q, Liu F, Wang Y. A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks. Artif Intell Rev. 2021 Mar;54(3):1613–47. doi:10.1007/s10462-020-09875-w.
27. Oh SL, Jahmunah V, Ooi CP, Tan RS, Ciaccio EJ, Yamakawa T, et al. Classification of heart sound signals using a novel deep WaveNet model. Comput Methods Programs Biomed. 2020 Nov;196(105604):105604. doi:10.1016/j.cmpb.2020.105604.
28. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural net-works. Commun ACM. 2017 May;60(6):84–90. doi:10.1145/3065386.
29. Adapa S. Urban Sound Tagging using Convolutional Neural Networks. In: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019); 2019; New York, NY, USA. p. 5–9.
30. Choi K, Fazekas G, Sandler M. Automatic tagging using deep convolutional neural networks. In: 17th International Society for Music Information Retrieval Conference; 2016 Aug 7–11; New York, NY, USA. p. 805–11.
31. Yaseen, Son G-Y, Kwon S. Classification of heart sound signal using Multiple features. Appl Sci. 2018 Nov;8(12):2344. doi:10.3390/app8122344.

32. Fernando I, Kannangara D, Kodituwakku S, Sirithunga A, Gayan S, Herath T, et al. Machine learning based heart murmur detection and classification. Biomed Phys Eng Express. 2024;11(1):015052.
33. Liu C, Zhang J, Zhang Y. Deep learning models for coronary artery disease classification using phonocardiograms. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2024; Lisbon, Portugal. p. 4973–80. doi:10.1109/BIBM62325.2024.10822772.
34. Mains T. A machine learning approach for integrating phonocardiogram and electrocardiogram data for heart sound detection [Doctoral dissertation]. Wichita, KS, USA: Wichita State University; 2024.
35. Behera TK, Sathia S, Panigrahi S, Naik PK. Revolutionizing cardiovascular disease classification through machine learning and statistical methods. J Biopharm Stat. 2024;10(2):1–23. doi:10.1080/10543406.2024.2429524.
36. Singh A, Arora V, Singh M. Heart sound classification using harmonic and percussive spectral features from phonocardiograms with a deep ANN approach. Appl Sci. 2024;14(22):10201. doi:10.3390/app142210201.
37. Liyong P, Haiyan Q. Heart sound classification algorithm based on bispectral feature extraction and convolutional neural networks. J Biomed Eng. 2024;41(5):977–85, 994. doi:10.7507/1001-5515.202310016.
38. Shakhovska N, Zagorodniy I. Classification of acoustic tones and cardiac murmurs based on digital signal analysis leveraging machine learning methods. Computation. 2024;12(10):208. doi:10.3390/computation12100208.
39. Nkereuwem EE, Ansa GO, Umoh UA, Essien UD, Nkereuwem EE, Asuquo MP. Machine learning based system for early heart disease detection and classification using audio signal processing approach. Glob J Eng Technol Adv. 2024;21(1):87–104.
40. Amiriparian S, Schmitt M, Cummins N, Qian K, Dong F, Schuller B. Deep unsupervised representation learning for abnormal heart sound classification. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018 Jul 18–21; Honolulu, HI, USA. p. 4776–9.
41. Mukherjee U, Pancholi S. A visual domain transfer learning approach for heartbeat sound classifi-cation. arXiv:2107.13237. 2021.
42. Fast AI. Welcome to fastai; 2023 [Internet]. [cited 2025 May 22]. Available from: https://docs.fast.ai/.
43. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI, USA. p. 4700–8.
44. Singh SK. Linux yourself: concept and programming. Boca Raton, FL, USA: Chapman and Hall/CRC; 2021.