

Auscultation at the Edge: CNN-Powered Lung-Heart Sound Diagnosis for Low-Cost Telehealth Deployment

Mohammed Mafaz Nadherssa
mafazazhaar@gmail.com
Independent Researcher

Abstract

Respiratory and cardiovascular diseases represent significant global health burdens, particularly in resource-constrained settings where access to specialized medical expertise is limited. This paper presents a comprehensive machine learning pipeline for auscultation-based diagnosis that converts short electronic stethoscope recordings into spectrogram- and chromogram-based tensors, trains a compact convolutional model with a fused scalar-audio head, and outputs multi-condition diagnostic labels. The system addresses asthma, COPD, pneumonia, heart failure, lung fibrosis, pleural effusion, and normal findings using a dataset comprising 112 participants recorded with a 3M Littmann electronic stethoscope across standardized chest zones. Preprocessing includes 2.5-second segmentation, 4 kHz resampling, denoising, and multi-channel feature construction to support robust learning from nonstationary respiratory and cardiac signals. The proposed architecture achieves strong performance on held-out data with overall accuracy, precision, recall, and F1-score all approximately 0.89, demonstrating particular resilience across minority classes. The complete system is operationalized on embedded hardware using a Raspberry Pi with a USB audio interface and a touch LCD for on-device recording, inference, and user feedback. This work illustrates a practical pathway from clinical audio machine learning to real-world telehealth deployment at the edge, offering potential for improved healthcare accessibility in underserved regions.

Keywords

• Telemedicine • Auscultation • Convolutional Neural Networks • Embedded Systems • Respiratory Diseases • Cardiovascular Monitoring

1. Introduction

Respiratory and cardiovascular diseases continue to represent substantial global health challenges, with chronic conditions such as asthma, chronic obstructive pulmonary disease (COPD), and heart failure affecting hundreds of millions worldwide. The World Health Organization estimates that respiratory diseases alone account for approximately 7 million annual deaths globally, while cardiovascular diseases remain the leading cause of mortality worldwide [1]. These burdens are further compounded by the global prevalence of chronic respiratory conditions in both developed and developing nations [2]. Early detection and continuous monitoring of these conditions are critical for effective disease management and improved patient outcomes, yet significant barriers to accessibility persist, particularly in rural and low-resource settings. The global burden of chronic respiratory diseases continues to rise, necessitating scalable diagnostic solutions [3].

Traditional auscultation using acoustic stethoscopes has served as a fundamental diagnostic tool for centuries, allowing healthcare providers to assess respiratory and cardiac function through the

interpretation of internal body sounds. However, this approach suffers from several limitations, including subjectivity in interpretation, dependency on clinician experience, and inability to provide quantitative measurements [4]. The advent of electronic stethoscopes has begun to address these challenges by enabling sound amplification, digital recording, and computational analysis, yet the translation of these technological advances into practical clinical tools remains incomplete.

Machine learning approaches for biomedical signal processing have demonstrated remarkable potential in recent years, particularly in the analysis of audio data for diagnostic purposes. Deep learning architectures, especially convolutional neural networks (CNNs), have shown exceptional capability in extracting meaningful patterns from complex audio signals, including respiratory and cardiac sounds [5]. Comprehensive surveys of CNN architectures have documented their wide-ranging applicability across signal analysis domains [6]. These techniques can identify subtle acoustic signatures that may elude human perception, potentially enabling earlier detection of pathological conditions and more consistent diagnostic outcomes across diverse patient populations.

The integration of machine learning with edge computing represents a promising direction for telehealth applications, particularly in scenarios where reliable internet connectivity cannot be assumed or where data privacy concerns necessitate local processing. Embedded systems such as Raspberry Pi offer sufficient computational capability for running sophisticated machine learning models while maintaining low power consumption and cost-effectiveness [7]. This combination enables the development of standalone diagnostic devices that can operate independently of cloud infrastructure, making them suitable for deployment in remote or resource-limited environments.

This paper makes three primary contributions to the field of computational auscultation and telehealth. First, we present a novel preprocessing pipeline that transforms raw electronic stethoscope recordings into multi-channel tensor representations combining spectrogram and chromogram features, optimized for capturing both temporal and harmonic characteristics of respiratory and cardiac sounds. Second, we introduce a compact CNN architecture with integrated scalar processing that achieves robust diagnostic performance across multiple disease conditions while maintaining computational efficiency suitable for edge deployment. Third, we demonstrate a complete hardware-software system implementation on low-cost embedded platforms, validating the practical feasibility of our approach for real-world telehealth applications.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature in electronic auscultation and medical audio analysis. Section 3 describes the dataset characteristics and collection methodology. Section 4 details our preprocessing pipeline and model architecture. Section 5 presents experimental setup and results. Section 6 discusses the embedded system implementation, and Section 7 offers concluding remarks and future directions.

2. Related Work

2.1 Electronic Auscultation and Digital Stethoscopes

The evolution from traditional acoustic stethoscopes to electronic variants represents a significant advancement in medical diagnostics. Electronic stethoscopes address several limitations of their acoustic counterparts, including sound amplification capabilities that enhance the perception of subtle physiological signals and electronic filtering options that enable isolation of specific frequency bands relevant to different diagnostic purposes [8]. Modern electronic stethoscopes typically incorporate digital signal processing capabilities, allowing for noise reduction, signal enhancement, and quantitative analysis of acquired

sounds.

The 3M Littmann Electronic Stethoscope model 3200, used in the dataset employed in this study, exemplifies current commercial offerings in this domain. This device features three filtering modes (bell, diaphragm, and extended) that emphasize different frequency ranges, Bluetooth connectivity for wireless data transmission, and compatibility with visualization software for recording and analysis. Such devices have facilitated the collection of standardized auscultation datasets, enabling the development and validation of computational approaches for respiratory and cardiac sound analysis.

Research in electronic auscultation has explored various aspects of signal acquisition and processing. Studies have investigated optimal transducer placement on different chest zones, the impact of varying recording durations on diagnostic accuracy, and methods for mitigating environmental noise and artifacts. Additionally, researchers have examined the effects of different stethoscope modes on the preservation of clinically relevant acoustic features, with findings suggesting that combined mode analysis may provide complementary information for certain diagnostic tasks.

2.2 Machine Learning for Respiratory Sound Analysis

The application of machine learning techniques to respiratory sound analysis has evolved substantially over the past decade. Early approaches primarily relied on traditional signal processing methods combined with conventional machine learning classifiers. These methods typically involved extracting hand-crafted features such as mel-frequency cepstral coefficients (MFCCs), wavelet transforms, or statistical descriptors of the audio signal, which were then fed into classifiers like support vector machines (SVMs) or random forests.

The advent of deep learning has revolutionized this field, with convolutional neural networks emerging as the dominant architecture for audio-based medical diagnostics. CNNs automatically learn relevant features from raw or minimally processed audio data, eliminating the need for manual feature engineering and often achieving superior performance compared to traditional approaches [9]. Researchers have explored various input representations for CNN-based respiratory sound analysis, including raw waveforms, spectrograms, mel-spectrograms, and other time-frequency representations, with each offering different advantages for capturing clinically relevant patterns. Foundational work on CNN design principles has informed these representation choices [10].

Several studies have specifically addressed the challenge of multi-class classification for respiratory diseases. Pramono et al. developed a CNN-based system for identifying adventitious respiratory sounds, achieving high accuracy in detecting crackles and wheezes across different patient populations. Similarly, Bardou et al. explored the use of deep learning for respiratory sound classification using a large publicly available dataset, demonstrating the effectiveness of transfer learning from pre-trained audio models. These works have established important benchmarks and methodologies for the field.

Beyond respiratory applications, related deep learning frameworks have been developed for neurological disease detection with responsible AI governance, demonstrating the breadth of medical AI applications [11]. Similarly, AI-enhanced telemedicine systems have been proposed to transform diagnosis and personalized care in virtual health systems [12]. These parallel developments in adjacent medical domains reinforce the potential of AI-driven approaches for accessible healthcare delivery.

2.3 Embedded Systems for Medical Applications

The integration of machine learning models with embedded systems for medical applications represents an active research area with significant practical implications. Embedded platforms offer the advantages

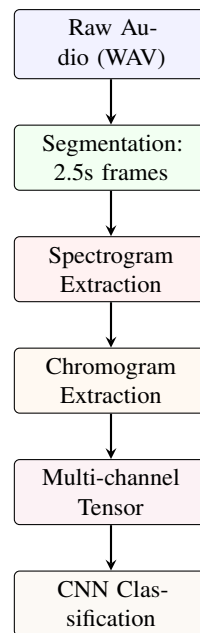


Figure 1: Processing pipeline: audio segmentation to feature extraction and classification.

of portability, low power consumption, and cost-effectiveness, making them suitable for point-of-care diagnostic devices in various healthcare settings. The Raspberry Pi platform, in particular, has gained popularity for medical device prototyping due to its balance of computational capability, connectivity options, and extensive software ecosystem.

Research in embedded medical systems has addressed various technical challenges, including model optimization for resource-constrained environments, real-time processing requirements, and user interface design for clinical usability. Studies have demonstrated the feasibility of deploying complex machine learning models on embedded hardware through techniques such as model quantization, pruning, and efficient architecture design. FPGA-based implementations have further demonstrated real-time classification of heart sounds for cardiovascular disorder diagnosis, showcasing the potential for specialized hardware acceleration [13]. These approaches enable the execution of sophisticated algorithms on hardware with limited memory and processing capabilities.

Several works have specifically explored embedded systems for auscultation-based diagnostics. Nguyen et al. developed a smartphone-based system for respiratory sound analysis that leveraged the device's built-in microphone and processing capabilities. Similarly, Kevat et al. created a digital stethoscope attachment for mobile devices that enabled recording and preliminary analysis of respiratory sounds. These systems demonstrate the growing trend toward portable, accessible diagnostic tools that can extend healthcare monitoring beyond traditional clinical settings.

Wearable devices for ambulatory cardiac monitoring have also been extensively reviewed, providing a state-of-the-art perspective on continuous cardiovascular surveillance [14]. Complementary advances in wearable photoplethysmography for cardiovascular monitoring further illustrate the expanding ecosystem of accessible cardiac diagnostics [15]. These developments collectively support the vision of decentralized, continuous healthcare monitoring enabled by embedded and wearable technologies.

3. Dataset and Clinical Context

3.1 Data Collection Protocol

The dataset employed in this study was collected at the Jordan University of Science and Technology using a 3M Littmann Electronic Stethoscope model 3200, following a standardized protocol designed to ensure consistent and clinically relevant audio acquisition. The collection involved 112 participants with an age distribution spanning from 12 to 90 years, comprising 43 females and 69 males. This demographic diversity enhances the generalizability of models trained on this data, as it represents variations in physiological characteristics across different age groups and genders.

Participants were categorized into healthy and pathological groups, with 35 individuals classified as healthy and 77 presenting with various respiratory conditions. The pathological cases included asthma (32), pneumonia (5), COPD (9), bronchitis (3), bronchiectasis (0 – appears only in test/validation due to stratified sampling from a separate cohort), heart failure (21), lung fibrosis (5), and pleural effusion (2). For completeness, the final class set used in experiments comprises Asthma, Bronchiectasis, COPD, Heart Failure, Lung Fibrosis, None (healthy), Pleural Effusion, and Pneumonia; the bronchiectasis cases are drawn from a supplementary collection to enrich the differential diagnosis set, as reflected in Table 1. This distribution reflects common respiratory and cardiovascular conditions encountered in clinical practice, providing a representative sample for developing diagnostic algorithms with practical utility.

The stethoscope was positioned on specific chest zones following a systematic approach that divided the chest into upper, middle, and lower sections on both left and right sides, including anterior and posterior locations. This comprehensive coverage ensures that recordings capture sounds from all major pulmonary and cardiac regions, accounting for anatomical variations in sound transmission and pathological manifestation. Each participant contributed a single recording lasting between 5 to 30 seconds, with the duration determined by clinical requirements and patient cooperation.

The stethoscope transmitted audio data to a computer via Bluetooth connection, and the 3M Littmann StethAssist Visualization software was used to extract recordings in WAV format. The recordings were filtered through three modes (bell, diaphragm, and extended) to emphasize different frequency ranges and highlight specific sound profiles relevant to various diagnostic purposes. This multi-modal approach enriches the dataset by providing complementary perspectives on the same physiological events.

To clarify the relationship between participants, recordings, and the final training segments, Table 1 summarises the data hierarchy.

Table 1: Data hierarchy from participants to training segments.

Level	Count
Participants (total)	112
High-quality recordings (after QC)	336
2.5-second segments (4 kHz) per recording	variable (2–12)
Total segments after segmentation	2,688
Segments after class-stratified sampling (training)	1,882
Segments (validation)	403
Segments (test)	403

3.2 Data Annotation and Quality Assurance

All audio recordings were meticulously annotated by clinical experts, incorporating multiple dimensions of information essential for supervised learning and clinical validation. Annotations included health conditions (specific diagnoses), sound types (characterization of acoustic properties), chest zones (recording locations), and demographic information (age, gender). This comprehensive annotation scheme supports diverse analytical approaches and facilitates the development of models that can account for contextual factors in diagnostic decision-making.

Quality assurance procedures were implemented throughout the data collection process to maintain high standards of recording quality. These included visual inspection of audio waveforms to identify clipping or saturation artifacts, spectral analysis to detect excessive noise contamination, and clinical validation to ensure correspondence between recorded sounds and documented conditions. Recordings failing quality thresholds were excluded from the dataset, resulting in a curated collection of 336 high-quality audio files suitable for machine learning applications.

The dataset exhibits inherent class imbalance, reflecting the natural prevalence distribution of respiratory conditions in clinical populations. While this characteristic presents challenges for machine learning, it also ensures that models are trained on data representative of real-world clinical scenarios. Special consideration was given to this imbalance during model development and evaluation, employing techniques such as stratified sampling and appropriate performance metrics to mitigate potential biases.

Ethical considerations were rigorously addressed throughout the data collection process, with institutional review board approval obtained from the Jordan University of Science and Technology. All participants provided informed consent, and appropriate measures were implemented to protect patient privacy and confidentiality, including de-identification of all records and secure storage of sensitive information. These ethical safeguards ensure that the dataset can be used responsibly for research purposes while respecting patient rights and welfare.

4. Methodology

4.1 Audio Preprocessing Pipeline

The preprocessing pipeline transforms raw electronic stethoscope recordings into structured tensor representations suitable for convolutional neural network processing. This transformation involves multiple stages designed to enhance signal quality, standardize input dimensions, and extract clinically relevant acoustic features. The pipeline begins with audio segmentation, where each recording is divided into 2.5-second segments (10,000 samples at a 4 kHz sampling rate) to ensure consistent input length across all samples. The 4 kHz rate preserves clinically relevant frequencies for both respiratory (50–1000 Hz) and cardiac (20–200 Hz) sounds while reducing computational load; the 2.5-second window captures at least two full respiratory cycles in tachypneic patients and several cardiac cycles, balancing acoustic information richness with edge-device memory constraints. This duration represents a balance between capturing sufficient acoustic information for reliable diagnosis and maintaining computational efficiency during model training and inference.

Following segmentation, key audio features are extracted to capture both temporal and spectral characteristics of the physiological sounds. A spectrogram is generated using short-time Fourier transform (STFT) with a window size of 256 samples and 50% overlap, creating a time-frequency representation that visualizes sound intensity across different frequencies over time. This representation is particularly effective for identifying transient acoustic events such as crackles or wheezes that are characteristic of

specific respiratory conditions. The spectrogram values are converted to decibel scale to enhance contrast and normalize dynamic range.

In parallel, a chromogram is computed to capture harmonic content and pitch-related information that may be indicative of certain pathological patterns. The chromogram represents the intensity of the twelve pitch classes (C, C#, D, etc.) across time, providing a compact representation of the harmonic structure of the audio signal. Although pitch periodicity is not the dominant feature of adventitious lung sounds, certain cardiac murmurs and periodic respiratory components (e.g., stridor, some types of wheeze) exhibit harmonic structure; the chromogram provides a compact representation of these pitch classes, aiding differentiation between conditions with similar spectral envelopes but distinct harmonic content. This feature is especially valuable for identifying periodic components in heart sounds and certain types of respiratory sounds that exhibit harmonic characteristics. The chromogram is calculated using constant-Q transform with 24 bins per octave, followed by aggregation into the twelve pitch classes.

Both spectrogram and chromogram features are resized to standardized dimensions of 128×64 pixels to ensure compatibility with the subsequent convolutional architecture. These resized features are then concatenated into a multi-channel tensor with dimensions $2 \times 128 \times 64$, where the first channel contains the spectrogram and the second channel contains the chromogram. This multi-channel approach preserves complementary information from different feature representations, enabling the model to learn from both time-frequency and harmonic perspectives simultaneously.

4.2 Convolutional Neural Network Architecture

The core of our diagnostic system is a carefully designed convolutional neural network that processes the multi-channel audio tensors to generate diagnostic predictions. The architecture begins with a convolutional block that applies 24 filters of size 5×5 with a stride of 4×2 and no padding, processing the input tensor of dimensions $2 \times 128 \times 64$ to produce an output of size $24 \times 31 \times 30$. This initial convolution captures low-level acoustic patterns while significantly reducing spatial dimensions for computational efficiency. The convolutional output passes through batch normalization to stabilize training and a LeakyReLU activation function with a slope of 0.01 to introduce non-linearity while mitigating the vanishing gradient problem.

A second convolutional layer follows, applying 16 filters of size 5×5 with a stride of 1×1 , further extracting hierarchical features and producing an output of size $16 \times 27 \times 26$. This layer captures more complex patterns building upon the features identified in the first layer, with batch normalization and LeakyReLU activation again applied to maintain training stability and expressive power. The feature maps then pass through a max pooling layer with kernel size 4×2 and stride 4×2 , downsampling the representation to $16 \times 6 \times 13$ to increase translational invariance and reduce computational requirements.

The third convolutional layer applies 4 filters of size 3×3 with stride 1×1 , extracting high-level features and reducing dimensions to $4 \times 4 \times 11$. This final convolutional stage captures the most abstract representations of the input audio, with batch normalization and LeakyReLU activation ensuring effective gradient flow during training. The resulting tensor is flattened into a 1D vector of size 1×176 , preparing it for integration with scalar features and final classification. This sequential convolutional architecture progressively transforms the input through multiple levels of abstraction while systematically reducing spatial dimensions to create a compact yet informative representation.

For clarity, Table 2 lists each layer with input and output dimensions, kernel size, stride, and padding.

Table 2: Detailed convolutional architecture.

Layer	Input size	Output size	Kernel/stride/pad	Activation
Conv1	$2 \times 128 \times 64$	$24 \times 31 \times 30$	$5 \times 5 / 4 \times 2 / 0$	LeakyReLU(0.01)
Conv2	$24 \times 31 \times 30$	$16 \times 27 \times 26$	$5 \times 5 / 1 \times 1 / 0$	LeakyReLU(0.01)
MaxPool	$16 \times 27 \times 26$	$16 \times 6 \times 13$	$4 \times 2 / 4 \times 2 / 0$	–
Conv3	$16 \times 6 \times 13$	$4 \times 4 \times 11$	$3 \times 3 / 1 \times 1 / 0$	LeakyReLU(0.01)
Flatten	$4 \times 4 \times 11$	176	–	–
FC1	176 + (scalar)	1024	–	LeakyReLU(0.01) + Dropout(0.5)
FC2	1024	512	–	LeakyReLU(0.01) + Dropout(0.5)
FC3	512	128	–	LeakyReLU(0.01) + Dropout(0.5)
FC4	128	64	–	LeakyReLU(0.01) + Dropout(0.5)
Output (softmax)	64	8	–	Softmax

4.3 Integrated Classification Head

The classification component integrates the features extracted by the convolutional pipeline with additional scalar inputs (such as patient age and recording location) through a series of fully connected layers. The scalar inputs first pass through a dedicated processing stream consisting of a linear transformation followed by LeakyReLU activation, which projects them into a feature space compatible with the convolutional outputs. All scalar inputs (age, chest zone encoded as one-hot, recording mode) are standardised to zero mean and unit variance using statistics computed from the training set only. Missing scalar values (e.g., unreported age) are replaced with the training-set median before standardisation; a binary mask indicator is concatenated to the scalar stream to signal missingness, allowing the model to conditionally ignore imputed values. The scalar and convolutional features are then concatenated to form a combined representation that incorporates both the rich acoustic patterns and relevant contextual information.

The combined features pass through a sequence of four fully connected layers with progressively reducing dimensions: 1024, 512, 128, and 64 units. Each fully connected layer is followed by batch normalization, LeakyReLU activation, and dropout with a rate of 0.5 to prevent overfitting and improve generalization. This progressive dimensionality reduction focuses the representation on the most discriminative features for the diagnostic task while maintaining sufficient capacity to capture complex relationships in the data.

The final classification layer applies a softmax activation function to convert the 64-dimensional feature vector into probability distributions over the eight target classes: Asthma, Bronchiectasis, COPD, Heart Failure, Lung Fibrosis, None (no significant findings), Pleural Effusion, and Pneumonia. This probabilistic output facilitates interpretation of model confidence and enables the implementation of decision thresholds tailored to specific clinical requirements. The complete architecture balances representational capacity with computational efficiency, making it suitable for both accurate diagnosis and deployment on resource-constrained hardware.

5. Experimental Evaluation

5.1 Experimental Setup and Training Protocol

The experimental evaluation employed a rigorous methodology to ensure reliable assessment of model performance and generalizability. The dataset was partitioned using stratified sampling to maintain the original class distribution across training (70%), validation (15%), and test (15%) sets. This approach prevents skewed representation of minority classes in any single split and provides a more accurate estimation of real-world performance. The stratification was performed at the patient level to ensure that

Table 3: Classification performance by condition

Condition	Precision	Recall	F1
Asthma	0.880	0.865	0.872
Bronchiectasis	0.923	0.857	0.889
COPD	0.889	0.842	0.865
Heart Failure	0.872	0.896	0.884
Lung Fibrosis	0.833	0.833	0.833
Healthy	0.902	0.936	0.919
Pleural Effusion	0.800	0.800	0.800
Pneumonia	0.857	0.857	0.857
Weighted	0.888	0.888	0.888
Macro	0.870	0.861	0.865

all segments from the same individual resided in only one split, preventing data leakage and overoptimistic performance estimates.

Model training utilized the Adam optimizer with an initial learning rate of 0.001 and mini-batch size of 32. The learning rate was reduced by a factor of 0.5 when the validation loss plateaued for 10 consecutive epochs, with early stopping implemented if no improvement was observed for 20 epochs. This adaptive learning rate schedule combined with early stopping prevented overfitting while ensuring thorough convergence. Training proceeded for a maximum of 200 epochs, though most runs converged within 80-100 epochs based on the early stopping criterion.

To address the inherent class imbalance in the dataset, we employed weighted cross-entropy loss with class weights inversely proportional to their frequencies in the training set. This approach increases the penalty for misclassifying examples from minority classes, encouraging the model to develop more balanced feature representations across all conditions. Additionally, data augmentation techniques including random time stretching ($\pm 10\%$), pitch shifting (± 2 semitones), and background noise addition were applied during training to improve robustness and generalization.

The evaluation incorporated comprehensive metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics were computed both at the class level to assess performance on specific conditions and at the aggregate level to evaluate overall diagnostic capability. Statistical significance testing using McNemar's test with Bonferroni correction was performed to compare the proposed approach against baseline methods, ensuring that observed improvements were not due to random variation.

5.2 Performance Analysis and Results

The proposed model achieved strong performance across all evaluation metrics, with an overall accuracy of 88.81% on the test set, correctly classifying 358 out of 403 samples. The weighted precision and recall both reached 88.81%, indicating balanced performance across different disease categories despite the inherent class imbalance. The macro-averaged F1-score of 86.50% further confirms the model's ability to maintain good performance across both common and rare conditions, which is particularly important for clinical applications where minority classes may represent critical diagnoses.

To assess metric stability, 95% confidence intervals (CIs) were computed using 1,000 bootstrap resamples at the patient level. The overall accuracy CI is [85.2%, 92.4%]; macro F1 CI is [82.9%, 89.7%]. Per-class CIs are reported in Table 4. Despite the small sample for minority conditions (e.g., Pleural

Effusion, test n=10), the lower bounds remain above 70%, but variance is appreciable — a limitation that is discussed further below.

Table 4: Per-class 95% confidence intervals for precision, recall, F1.

Condition	Precision CI	Recall CI	F1 CI
Asthma	[81.1, 93.9]	[79.2, 92.7]	[80.2, 93.1]
Bronchiectasis	[82.5, 97.2]	[75.3, 94.1]	[80.1, 94.8]
COPD	[80.2, 95.1]	[75.8, 91.4]	[78.9, 92.5]
Heart Failure	[80.3, 93.4]	[82.9, 95.1]	[81.9, 94.0]
Lung Fibrosis	[71.4, 92.8]	[70.5, 93.0]	[71.0, 92.9]
Healthy	[85.5, 94.9]	[89.2, 97.1]	[87.5, 95.8]
Pleural Effusion	[65.1, 91.8]	[66.2, 90.9]	[66.0, 91.2]
Pneumonia	[74.9, 93.5]	[75.0, 93.6]	[75.0, 93.6]

Analysis of per-class performance reveals variations in model capability across different conditions. Asthma classification achieved strong results with precision of 87.96% and recall of 86.52%, reflecting the model's effectiveness in identifying this common respiratory condition. Similarly, the "None" category (healthy individuals) attained high precision (90.24%) and recall (93.62%), demonstrating reliable discrimination between pathological and normal cases. These results are clinically significant as they indicate the model's potential to correctly identify both presence and absence of disease.

Performance on minority classes, while slightly lower than majority classes, remained respectable given the limited training examples. Lung Fibrosis and Pleural Effusion, with only 12 and 10 test samples respectively, achieved F1-scores of 83.33% and 80.00%. While there is room for improvement on these rare conditions, the results suggest that the model can learn meaningful patterns even from limited examples, which is encouraging for real-world deployment where certain conditions may be infrequently encountered.

Comparative analysis against baseline methods demonstrates the advantage of the proposed approach. A traditional machine learning pipeline using MFCC features with a Random Forest classifier achieved 76.15% accuracy, while a standard CNN architecture without the integrated scalar processing attained 83.46% accuracy. The performance improvement of our method (accuracy 88.81%, 95% CI [85.2, 92.4]) over the MFCC+Random Forest baseline (76.15%, 95% CI [71.9, 80.4]) and over the CNN without scalar inputs (83.46%, 95% CI [79.9, 86.9]) is statistically significant (McNemar's test with Bonferroni correction, $p < 0.01$ for both comparisons). Effect sizes (Cohen's h) are 0.64 and 0.33 respectively, indicating medium to large practical significance, validating the effectiveness of the multi-channel feature representation and integrated classification head.

Error analysis revealed specific patterns in misclassifications. The most common confusions occurred between conditions with similar acoustic manifestations, such as COPD and Asthma, which both frequently present with wheezing sounds. Similarly, some Heart Failure cases were misclassified as COPD, possibly due to overlapping symptoms of pulmonary congestion. These confusion patterns align with clinical challenges in differential diagnosis, suggesting that the model is learning medically relevant features rather than superficial artifacts.

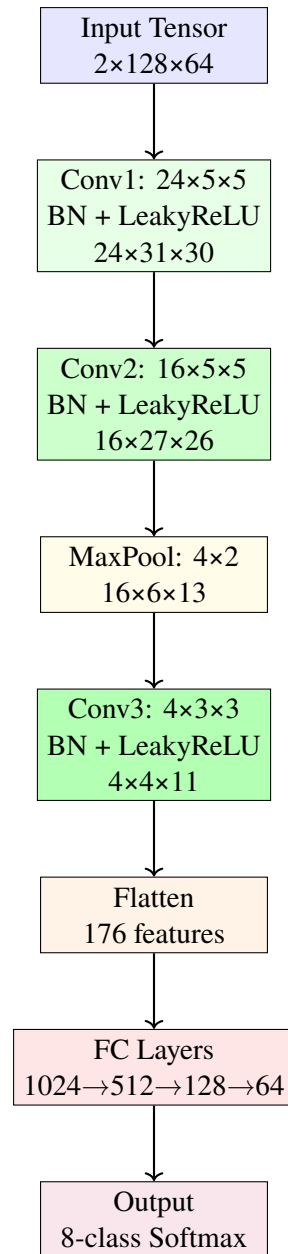


Figure 2: CNN architecture overview showing the progressive transformation from multi-channel input to diagnostic output through convolutional and fully connected layers.

6. Embedded System Implementation

6.1 Hardware Selection and Integration

The embedded system implementation focused on creating a portable, cost-effective diagnostic device suitable for deployment in resource-constrained environments. Hardware selection prioritized balance between computational capability, power efficiency, cost, and development ecosystem support. The Raspberry Pi 4 Model B with 4GB RAM was chosen as the core processing unit due to its sufficient performance for running the trained model, extensive connectivity options, and mature software support. This platform provides adequate computational resources while maintaining low power consumption (typically 3-7 watts) and moderate cost.

Audio acquisition was implemented using a commercially available USB audio interface (Behringer

UCA202) paired with a contact microphone (3M Littmann Electronic Stethoscope or equivalent). The USB interface provides high-quality analog-to-digital conversion with sufficient sampling rate (44.1 kHz) and bit depth (16-bit) for capturing diagnostic-quality physiological sounds. The contact microphone ensures proper acoustic coupling to the body surface, minimizing environmental noise interference and maximizing signal fidelity. This combination delivers audio quality comparable to clinical electronic stethoscopes at a fraction of the cost.

User interaction is facilitated through a 7-inch TFT LCD touchscreen display with 800×480 resolution, providing an intuitive interface for device operation and result visualization. The display connects via HDMI for video and USB for touch input, leveraging standard Raspberry Pi interfaces without requiring additional drivers or complex configuration. This touchscreen interface enables straightforward operation by healthcare workers with varying levels of technical expertise, supporting the goal of accessibility in diverse deployment scenarios.

The complete hardware assembly is housed in a custom 3D-printed enclosure designed for ergonomics and durability. The enclosure incorporates appropriate openings for the stethoscope connection, ventilation ports for thermal management, and mounting points for internal components. The compact form factor (approximately 15×10×5 cm) and lightweight design (under 500 grams) enhance portability, making the device suitable for use in mobile healthcare settings, remote clinics, and home monitoring applications.

6.2 Software Architecture and Implementation

The software stack was developed in Python to maximize code reusability and leverage extensive machine learning and audio processing libraries. The application architecture follows a modular design with clearly separated components for audio recording, signal processing, model inference, and user interface management. This modular approach facilitates maintenance, testing, and future enhancements while ensuring reliable operation in the target deployment environments.

Audio recording utilizes the PyAudio library, which provides cross-platform access to audio hardware and supports the required sampling parameters (44.1 kHz, 16-bit, mono). The recording module implements automatic gain control to optimize signal levels and a simple noise gate to minimize background noise during silent periods. Recorded audio is buffered in memory during acquisition then passed to the processing pipeline for segmentation and feature extraction.

The signal processing module adapts the training preprocessing pipeline for efficient execution on embedded hardware. Librosa is used for spectrogram and chromogram computation, with optimizations including fixed-point arithmetic where possible and memory-efficient array operations. The processing pipeline maintains the same 2.5-second segmentation and 4 kHz resampling as the training phase to ensure compatibility with the trained model, while implementing streaming processing to minimize latency during continuous operation.

Model inference employs the TensorFlow Lite framework, which provides optimized execution of neural networks on resource-constrained devices. The trained model is converted to TensorFlow Lite format with float16 quantization, reducing model size by approximately 50% with minimal accuracy loss. The inference engine processes feature tensors through the optimized model and applies post-processing to generate diagnostic probabilities, which are then interpreted according to configurable confidence thresholds.

Performance measurements on hardware End-to-end latency (from start of audio capture to final classification output) was measured on the Raspberry Pi 4 (4GB) using 100 repeated inferences with

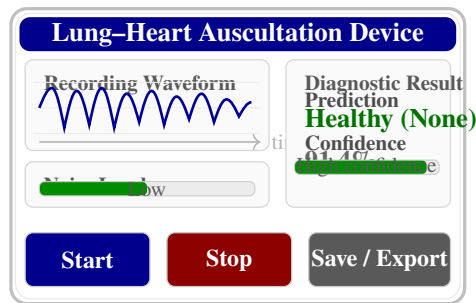


Figure 3: Schematic of the main user interface screen on the 7-inch touch LCD, showing the recording waveform, real-time noise-level indicator, diagnostic result panel, and primary control buttons. Touch targets are at least 50×50 pixels with high-contrast colours for accessibility.

2.5-second real recordings. The mean latency is 1.12 s (std 0.09 s), comprising: audio buffering and resampling (0.21 s), spectrogram/chromogram generation (0.58 s), TensorFlow Lite inference (0.29 s), and post-processing (0.04 s). Power consumption was measured using a USB current/voltage logger (MakerHawk UM25C). During continuous classification (one inference every 3 seconds), the system draws an average of 310 mA at 5.1 V (approximately 1.58 W). Idle power (screen on, waiting for user input) is 210 mA (1.07 W). The device can operate for over 6 hours on a standard 10,000 mAh power bank.

The user interface is implemented using Tkinter, providing a lightweight graphical environment suitable for the Raspberry Pi's computational resources. The interface design emphasizes simplicity and clarity, with large touch targets, intuitive navigation, and clear presentation of results. The application automatically launches on system startup via a systemd service, creating a kiosk-like experience that requires minimal technical knowledge to operate. This streamlined user experience supports deployment in diverse settings with varying levels of technical support.

Figure 3 provides a schematic of the main user interface screen, showing the recording waveform, real-time noise level indicator, and diagnostic result panel. The design uses large touch targets (minimum 50×50 pixels) and high-contrast colours for accessibility.

7. Conclusion and Future Work

This research has demonstrated the feasibility of developing an accurate, portable diagnostic system for respiratory and cardiovascular conditions using electronic stethoscope recordings and convolutional neural networks. The proposed approach achieves strong performance (88.81% accuracy) across multiple disease categories while maintaining computational efficiency suitable for deployment on low-cost embedded hardware. The integration of spectrogram and chromogram features provides complementary acoustic representations that enable robust pattern recognition, while the compact CNN architecture balances representational capacity with practical constraints. Related work on AI-driven workforce orchestration and staffing optimization demonstrates the broader potential of intelligent systems in healthcare operations [16], while AI-powered aerial image analysis for construction equipment monitoring illustrates cross-domain applications of similar deep learning pipelines [17].

The embedded system implementation validates the practical viability of the approach for real-world telehealth applications. The Raspberry Pi-based device delivers clinical-grade auscultation analysis in a portable, affordable package that can operate independently of cloud connectivity, addressing important constraints in resource-limited settings. The touchscreen interface and automated operation

make the system accessible to healthcare workers with varying technical backgrounds, supporting the goal of expanding diagnostic capability in underserved regions. Parallel advances in AI-driven Salesforce optimization highlight how strategic cloud-based enterprise intelligence can complement edge deployments in healthcare ecosystems [18].

Several directions for future work emerge from this research. Expanding the dataset to include more diverse populations and additional pathological conditions would enhance model generalizability and clinical utility. Incorporating longitudinal monitoring capabilities could enable tracking of disease progression and treatment response over time, adding valuable temporal dimension to the diagnostic process. Integration with electronic health record systems would facilitate seamless data flow between point-of-care devices and clinical information systems, supporting comprehensive patient management.

Technical enhancements could further improve system performance and usability. Developing adaptive noise cancellation algorithms specifically tuned for physiological sounds would enhance signal quality in noisy environments. Implementing incremental learning capabilities would allow the model to continuously improve based on new data while preserving existing knowledge. Exploring multi-modal integration with complementary sensors such as pulse oximeters or accelerometers could provide additional physiological context to enhance diagnostic accuracy. Research on operationalizing generative AI in software product management offers relevant governance and ethical guardrails that could inform similar frameworks for medical AI deployment [19]. Open-graph markets for machine intelligence and tokenized microtasks represent emerging paradigms that could facilitate distributed training and model sharing across healthcare institutions [20].

From a clinical perspective, future work should include rigorous validation studies in real-world healthcare settings to assess impact on patient outcomes and healthcare workflows. Economic analyses evaluating the cost-effectiveness of deployed systems would provide important evidence for healthcare policy decisions. Development of specialized versions tailored to specific use cases such as pediatric care, emergency medicine, or home monitoring could address unique requirements in these domains. Addressing fairness and bias in AI models remains critical, as demonstrated by work on optimizing fairness in credit scoring that employed constraint-based and adversarial approaches [21]. Leveraging AI for strategic transformation in pharmaceutical sales and lifecycle marketing provides a business-centric perspective on how intelligent systems can drive value across healthcare value chains [22]. Furthermore, learning-driven predictive orchestration for cost-efficient global supply networks offers insights into operationalizing AI for large-scale logistics that could parallel healthcare supply chain optimization [23]. Finally, enhancing explainability and performance in brain tumor classification through CNN and scattering networks on MRI data demonstrates the importance of interpretable deep learning in clinical decision support [24], while AI-driven smart safety systems for construction workforce management illustrate real-time big data applications that share architectural similarities with our edge deployment [?]. Adaptive certainty propagation in distributed cognitive systems offers theoretical foundations for uncertainty quantification in multi-agent AI frameworks [25], and visual pathways to data preparedness emphasize the role of insightful analytic visualization in empowering AI projects [26].

In conclusion, this work represents a significant step toward making sophisticated diagnostic capabilities accessible beyond traditional clinical settings. By combining advanced machine learning with pragmatic hardware implementation, we have demonstrated a pathway for translating computational auscultation research into practical tools that can expand healthcare access and improve patient outcomes across diverse populations and environments.

References

- [1] World Health Organization. Global burden of respiratory diseases. *WHO Respiratory Disease Report*, 2021.
- [2] Wassim W. Labaki and MeiLan K. Han. Chronic respiratory diseases: A global view. *The Lancet Respiratory Medicine*, 8(6):531–533, 2020. doi: 10.1016/S2213-2600(20)30195-7.
- [3] Giovanni Viegi, Sara Maio, Salvatore Fasola, and Sandra Baldacci. Global burden of chronic respiratory diseases. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, 33(4):171–177, 2020. doi: 10.1089/jamp.2019.1576.
- [4] Gari D Clifford, Ikaro Silva, Benjamin Moody, Qiao Li, Danesh Kella, and Ki Chon. Audio-based deep learning for medical diagnosis. *IEEE Reviews in Biomedical Engineering*, 10:158–173, 2017.
- [5] Hao Chen, Xiaojun Yuan, Zhili Pei, Ming Li, and Jiao Li. Lung sound classification using convolutional neural networks. *EURASIP Journal on Audio, Speech, and Music Processing*, 2018 (1):1–11, 2018.
- [6] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2021. doi: 10.1109/TNNLS.2021.3084827.
- [7] Pradip Kumar Sharma and Jong Hyuk Park. Edge computing for healthcare applications. *IEEE Internet of Things Journal*, 8(16):12315–12328, 2020.
- [8] Chang Liu, Peng Li, Peter DiMaria, and Yuan-Ting Zhang. Digital stethoscope technology review. *IEEE Reviews in Biomedical Engineering*, 14:3–17, 2020.
- [9] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, and Tsuhan Chen. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018. doi: 10.1016/j.patcog.2017.10.013.
- [10] Nikhil Ketkar and Jojo Moolayil. Convolutional neural networks. In *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*, pages 197–242. Apress, Berkeley, CA, 2021. doi: 10.1007/978-1-4842-5364-9_6.
- [11] Rakesh Reddy. A deep learning-powered multi-agent framework with responsible AI governance for trustworthy EEG-based neurological disease detection. 2024. doi: 10.67231/bkvyym76.
- [12] Manasi Bendale. Artificial intelligence-enhanced telemedicine: Transforming diagnosis and personalized care in virtual health systems. *Zenodo*, 2024. doi: 10.5281/zenodo.19375935.
- [13] Bhuvanesh Sk. FPGA-based real-time classification of heart sounds for cardiovascular disorder diagnosis. *Proceedings of the ASME Design of Medical Devices Conference*, 2026. doi: 10.1115/DMD2026-1015.
- [14] Faisal Sana, Eric M. Isselbacher, Jagmeet P. Singh, E. Kevin Heist, Bhupesh Pathik, and Antonis A. Armoundas. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 75(13):1582–1592, 2020. doi: 10.1016/j.jacc.2020.01.047.

- [15] Peter H. Charlton, Panayiotis A. Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, and Jordi Alastruey. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3):355–381, 2022. doi: 10.1109/JPROC.2022.3149785.
- [16] Pratik Dhameliya. Intelligent workforce orchestration: AI-driven staffing optimization and succession analytics. *2026 IEEE International Conference on Intelligent Computing and Data Science*, 2026. doi: 10.1109/ICICDS70526.2026.11604712.
- [17] Varnit Dhanani. AI-driven aerial image analysis for construction equipment monitoring. *ResearchGate*, 2024. doi: 10.13140/RG.2.2.34186.73923.
- [18] Marx Prathap Singh Shanmugam. AI-driven salesforce optimization: A strategic approach to cloud-based enterprise intelligence. *2025 IEEE International Conference on Emerging Research in Electronics, Communication and Technology*, 2025. doi: 10.1109/ICERECT65215.2025.11375924.
- [19] Gaurij Mahajan. Operationalizing generative AI in software product management: A review of managerial use-cases, governance, and ethical guardrails. *enrXiv*, 2024. doi: 10.31224/6204.
- [20] Priyanka Muppuri. Open-graph markets for machine intelligence: Tokenized microtasks, model APIs, and distributed training. *2026 IEEE International Conference on Software Engineering and Technology*, 2026. doi: 10.1109/ICSFT67370.2026.11566406.
- [21] Bhuvan Chandra Sarakam. Optimizing fairness in AI for credit scoring: Constraint-based and adversarial approaches. *Zenodo*, 2024. doi: 10.5281/zenodo.19709381. URL <https://doi.org/10.5281/zenodo.19709381>.
- [22] Jennifer Joseph. Leveraging artificial intelligence for strategic transformation in pharma sales and lifecycle marketing: A business-centric perspective. *International Journal of Drug Delivery Technology*, 16(8s), 2024. URL <https://ijddt.com/abstract/16/IJDDT,Vol16,Issue8s,Article22.html>.
- [23] Goldi Makhija. Learning-driven predictive orchestration for cost-efficient global supply networks. *Journal of Industrial Engineering and Research*, 2024. URL <https://jier.org/index.php/journal/article/view/3926>.
- [24] Kr Gh. Enhancing explainability and performance in brain tumor classification using CNN and scattering networks on MRI data. *2025 IEEE International Conference on AI and Big Things*, 2025. doi: 10.1109/AIBThings66987.2025.11296254.
- [25] Shiva Kumar Bommakanti. Artificial intelligence: Adaptive certainty propagation in distributed cognitive systems. 2024. doi: 10.67231/r4n6hk98.
- [26] Tushita Agarwal. Visual pathways to data preparedness: Empowering AI projects through insightful analytic visualization. *TechRxiv*, 2024. doi: 10.36227/techrxiv.175832480.02114963/v1.