



Paper Type: Original Article

Edge-Cloud Enabled Multimodal Framework for Real-Time Pneumonia Detection Using Wearable Sensors

Emmanuel Onwuka Ibam¹, Johnson Bisi Oluwagbemi^{2*}, Kayode Abiodun Oladapo²

¹ Department of Information Technology, Federal University of Technology, Akure, Nigeria; eoibam@futa.edu.ng.

² Department of Computer Science, McPherson University, Seriki-Sotayo, Ogun State, Nigeria; oluwagbemijb@mcu.edu.ng; oladapoka@mcu.edu.ng.

Citation:

Received: 12 January 2025

Revised: 24 March 2025

Accepted: 19 April 2025

Ibam, E. O., Oluwagbemi, J. B., & Oladapo, K. A. (2025). Edge-cloud enabled multimodal framework for real-time pneumonia detection using wearable sensors. *Information sciences and technological innovations*, 2(2), 88-101.


Abstract


Pneumonia continues to be a significant reason for sickness and death worldwide in places with limited resources and among older people. Spotting it is key to stepping in and getting better results for patients. However, the usual ways to diagnose it often take too long and aren't easy to access. This study proposes and validates a novel edge cloud integrated framework that leverages multimodal wearable sensors and deep learning for the pre-symptomatic detection of pneumonia. The system continuously acquires and analyzes high-frequency physiological data (Respiratory Rate (RR), Heart Rate (HR), SpO₂, body temperature) and event-driven acoustic biomarkers (cough sounds) through a distributed architecture. An intelligent edge module performs local preprocessing and anomaly triage, selectively transmitting only flagged anomalous data to a cloud-based multimodal deep learning model, which then performs sophisticated risk stratification. We trained and validated our framework on a composite dataset including public repositories (MIMIC-III, Coswara) and a clinically supervised deployment in two Nigerian hospitals, totaling over 12,000 patient hours. The model achieved an area under the curve of 0.947, with a sensitivity of 94.3% and a specificity of 90.1%, demonstrating its potential as a scalable, interpretable, and privacy-preserving system for proactive respiratory health monitoring.

Keywords: Pneumonia, Deep learning, Edge-cloud computing, Convolutional neural network, Wearable sensor.

1 | Introduction

Pneumonia is a lung infection caused by viruses and bacteria. It generally causes inflammation of the alveoli or the air sacs of the lungs, which usually results in breathing problems, consistent coughing, and chest pain that presents itself with vague, ambiguous, and imprecise symptoms. Pneumonia is responsible for over 800,000 hospitalizations per year in Nigeria [1]. Pneumonia usually starts as an acute infection, which, if not

 Corresponding Author: oluwagbemijb@mcu.edu.ng

 <https://doi.org/10.48314/isti.vi.39>



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

diagnosed and treated on time, can become chronic pneumonia. If diagnosed early, pneumonia can be treated, whereas severe cases of pneumonia usually result in mortality, especially among children less than five years old.

Pneumonia is an acute infection of the lungs, posing a significant global health problem. It is the single major infectious cause of death in children worldwide and a significant cause of hospitalization and mortality among the elderly [2]. The clinical challenge of pneumonia is its insidious onset, where early symptoms, such as fatigue and mild cough, can be easily misinterpreted for less severe respiratory ailments.

Previous research in Artificial Intelligence (AI) driven pneumonia detection has primarily focused on analyzing static medical images, such as CXRs or CT scans [3]. While powerful, these approaches are confined to clinical settings. Other studies have explored unimodal analysis of wearable data, such as classifying cough sounds. Or detecting fever [4], [5]. These single modality systems are susceptible to high false positive rates and lack the contextual richness to distinguish between pneumonia and other conditions like bronchitis or the common cold. There is a critical gap in the synthesis of multiple, continuous biosignals into a cohesive, longitudinal health record for robust, early-stage disease detection in a real-world, ambulatory setting.

This paper addresses this gap by presenting a comprehensive, edge-cloud-enabled framework for early pneumonia detection. Our approach makes the following key contributions:

- I. A multimodal sensing fusion framework: we integrate continuous physiological time series data with event-driven acoustic biomarkers to create a holistic view of a patient's respiratory health.
- II. A hybrid-edge cloud architecture: we design a distributed system that leverages edge computing for real-time preprocessing, low-latency alerts, and data efficiency, while utilizing the cloud for complex deep learning inference and model training.
- III. A robust deep learning model: we develop a hybrid neural network that combines Convolutional Neural Networks for spatial feature extraction from audio spectrograms and Long Short Term Memory networks for capturing temporal dependencies in physiological signals.
- IV. Rigorous evaluation and interpretability: we validate the framework using a large-scale, composite dataset and incorporate explainable AI techniques to provide clinical transparency.

2 | Literature Review

Vidhya et al. [6] developed a software tool for diagnosing pneumonia using AI and lung sound analysis to minimize diagnosis time and reduce dependency on imaging techniques. Lung sounds recorded through a digital stethoscope, Noise reduction and filtering with Audacity software, feature extraction using Librosa (MFCC features), classification with SVM, KNN, and Gradient. In their findings, their proposed AI tool using Gradient Boosting was accurate, fast, non-invasive, and economical for diagnosing pneumonia. However, its performance is limited in a noisy environment.

Xu and Wang [7] developed a robust and flexible multimodal deep learning framework that integrates image and textual data to enhance pneumonia detection, even when one modality is wholly or partially missing. The concept used was Multimodal learning, Masked attention, Stacking Mixture of Experts (MOE), Transfer learning with BERT, and ResNet 50 Multitask learning. They developed the Flexible Multimodal Transformer (FMT), Combined BERT (for text) and ResNet-50 (for X-ray images). Used dynamic masked attention to simulate real-world data loss. They applied a stacked MOE architecture for refined prediction, trained and tested on a private small multimodal pneumonia dataset (43 samples). Also, they conducted ablation studies comparing FMT against ResNet, BERT, and CheXMed. The FMT model effectively handles multimodal diagnostic challenges such as data scarcity and missing modalities. It improves pneumonia detection performance with fewer parameters and better robustness compared to existing benchmarks, offering clinical applicability and scalability. However, a small dataset was used, and simulated modality loss may not fully represent clinical complexity.

Rancea et al. [8] present edge computing in healthcare: innovations, opportunities, and challenges. They systematically reviewed and analyzed recent research on edge computing in healthcare, focusing on privacy and security, AI optimization methods, and edge offloading strategies. In their findings, they discovered that Edge computing holds robust potential to reform healthcare by addressing latency, scalability, and privacy challenges. It enables intelligent and adapted care, enhances patient monitoring, and optimizes healthcare resource use. However, limitations exist in interoperability, standardization, and resource management.

In Rashid et al [9], an enhanced deep learning framework for pneumonia detection in chest X-rays, they aim to develop an enhanced deep learning model for pneumonia detection that integrates densenet-121 with the Convolutional Block Attention Module (CBAM) to improve diagnostic accuracy and interpretability in chest X-ray images. They used the Kermany chest X-ray dataset (5,856 images), integrated CBAM with DenseNet-121 using additive fusion, applied three-phase training with selective freezing of 40 layers, which used Mish activation, data augmentation, and transfer learning. The enhanced DenseNet-121 + CBAM framework significantly improves pneumonia detection from chest X-rays. It offers better feature focus through attention, high accuracy, and better interpretability. However, no external validation dataset was used, and the generalizability to other imaging modalities was not assessed.

Mbata et al. [10] in development of ai-assisted wearable devices for early detection of respiratory diseases, they aim to design and evaluate AI-assisted wearable devices for real time detection of respiratory diseases, enable continuous, non-invasive monitoring of key physiological markers like oxygen saturation, Respiratory Rate (RR), and Heart Rate (HR) and to assess their utility in managing conditions such as asthma, COPD, and pneumonia. The method used was prototyping wearable devices (wristbands, chest straps, patches), and they deployed AI models (CNN/RNN) for pattern recognition and prediction. In their findings, high sensors accurately detect early signs of respiratory distress, AI algorithms achieved strong performance in predicting exacerbations, real-time feedback enabled proactive management and timely interventions, and participants demonstrated improved awareness and engagement in health self-monitoring. Their AI-assisted wearables represent a breakthrough in respiratory care by offering scalable, cost-effective, and continuous monitoring solutions. However, Sensor performance is affected by environmental and physical factors, and it has limited battery life and a need for frequent recharging.

Sathupadi et al. [11] present edge-cloud synergy for AI-enhanced sensor network data: a real-time predictive maintenance framework. They developed a hybrid AI framework that combines edge-based anomaly detection with cloud-based failure prediction. They designed and implemented a two-tier system: lightweight KNN on Raspberry Pi Zero 2 W (edge) and LSTM on AWS Lambda (cloud), dataset preprocessing, anomaly detection, and failure prediction using statistical and DL models, evaluated on 20 industrial sensors deployed across machines (rollers, conveyors), and used real data and compared with a cloud-only method. In their findings, the hybrid model outperformed fog and static load-balancing methods. However, incorporating detection of additional failure modes and other environments like healthcare or smart grids will go a long way.

Al Waisy et al. [12] in COVID-DeepNet: hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images, developed a hybrid deep learning system (COVID-DeepNet) combining two robust architectures, DBN and CDBN, to enhance COVID-19 pneumonia detection in chest X-ray (CX-R) images. They created a large dataset called COVID19-vs-Normal using public X-ray image repositories (Cohen's GitHub, RSNA, Radiopaedia), preprocessing with CLAHE (for contrast enhancement) and a Butterworth filter (for denoising). Training DBN and CDBN from scratch using augmented images (24,000 samples) and fused output using Weighted Sum Rule (WSR), and compared with other fusion strategies. COVID-DeepNet was an accurate, interpretable, and practical tool for diagnosing COVID-19 pneumonia from CX-R images. It combines the strengths of both DBN and CDBN, ensuring high confidence for medical practitioners and minimal false classifications. However, they focused only on binary classification (COVID-19 vs Normal); validation on multi-class pneumonia datasets is needed.

Rajaraman et al. [13] emphasize the need for interpretability in CNN-based medical diagnosis due to the “black box” nature of these models. Their study integrates interpretability methods like Grad-CAM and LIME into CNN predictions, making the model’s decision process more transparent. The research compares an optimized custom CNN and a pre-trained VGG16 to distinguish between typical, bacterial, and viral pneumonia cases. The pre-trained ResNet-50 model performs best, achieving 96.2% accuracy and an interpretability score of 91.8% (measured by MCC), while the custom CNN reaches 94.1% accuracy and 87.3% interpretability. This advancement in CNN interpretability holds.

The integration of adversarial training and explainability techniques was investigated to enhance the performance and interpretability of CNN models. According to the findings, the architecture is based on ResNet-50 and uses methods such as Layer Relationship Propagation (LRP) and mechanisms. The findings showed that adversarial training significantly improved the robustness of the model and achieved accuracies between 82.8% (lowest) and 92.4% (highest) under different experimental conditions. The interpretability of the model was also evaluated, with results indicating a range of 79.6% to 87.3% in interpretability. These results indicate that adversarial techniques, when combined with interpretability methods such as LRP and attention to detail, not only provide an accurate model but also provide transparency in the model decision-making process.

However, previous works such as Rashid et al. [9] and Vidhya et al. [6] have demonstrated the potential of unimodal audio analysis and static lung sound classification for pneumonia detection; these approaches suffer from high false positives, environmental sensitivity, and lack of real-time, contextual integration. Our framework addresses these limitations by blending continuous physiological time series with acoustic biomarkers within a hybrid edge-cloud architecture delivering both accuracy and deployability in ambulatory settings.

3 | Methodology

This section describes a hybrid cloud-edge computing framework designed for intelligent, real-time health monitoring. The architecture is composed of four primary modules, each with a distinct role.

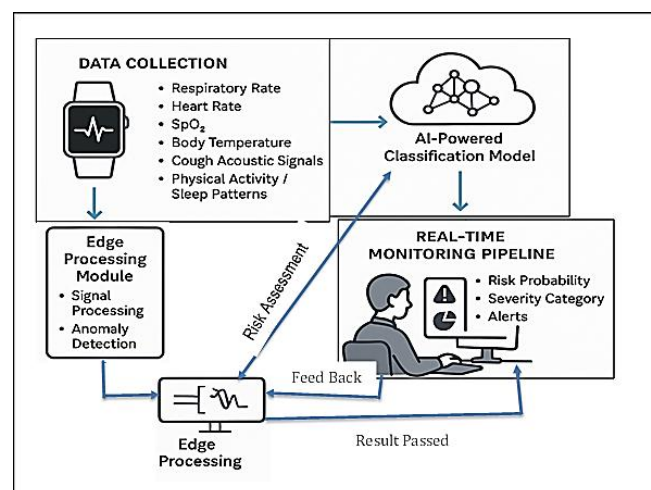


Fig. 1. System architecture of the edge-cloud-enabled multimodal framework for real-time pneumonia detection using wearable sensors.

3.1| Data Collection: The Wearable Sensor Ecosystem

This is the starting point of the entire system. A user wears one or more sensors (represented by the smart watch) that continuously collect a variety of physiological and behavioral data. This data includes:

- I. Respiratory Rate: number of breaths per minute
- II. Heart Rate: number of heartbeats per minute.
- III. SpO₂: blood oxygen saturation level.
- IV. Body temperature: the user's core or skin temperature.
- V. Cough acoustic signals: audio recordings and analysis of coughs, which can be a key symptom.
- VI. Physical activity or sleep patterns: data on movement, exercise, and rest cycles.

This raw data is then sent along two parallel paths for processing.

3.2| The Dual Processing Approach: Edge and Cloud

The system splits the processing tasks to get the best of both worlds: the speed of local processing and the power of cloud computing.

- I. The cloud path: AI-powered classification

The raw sensor data is sent to the cloud. Then, a Powered Classification Model (deep learning model) analyzes the comprehensive dataset. This model is trained to recognize complex patterns that may indicate a specific health risk or condition. Because it runs on powerful cloud servers, it can perform computationally intensive tasks that would be too slow or power hungry for a local device. The output of this model is a sophisticated classification (risk assessment), which is then sent to the final stage.

- II. The edge path: immediate local processing

The raw data is also sent to an Edge Processing Module. Edge means the processing happens locally, on or near the user's device (smartphone, a dedicated hub, or wearable device). The module does two main tasks: i. Signal Processing cleans up the raw sensor data, sifting out noise and making it suitable for analysis and ii. Anomaly detection runs simpler, quicker algorithms to look for immediate anomalies or deviations from the user's normal baseline (a sudden, sharp drop in SpO₂ or a spike in HR at rest). This provides a rapid, first-pass analysis, and results are then passed to the user's local monitoring interface and also sent to the final pipeline.

3.3| Synthesis: The Real-Time Monitoring Pipeline

This is the central hub where information from both the cloud and the edge converges. It receives the deep, analytical insights from the AI-powered classification model (75% probability of a respiratory infection). Also, it receives the immediate findings from the edge processing module (anomaly detected: high RR for the last 10 minutes). By combining these inputs, the system generates a holistic and reliable output, which includes:

- I. Risk Probability calculated score representing the user's risk level.
- II. Severity Category classifying the condition's severity (mild, moderate, severe).
- III. Alerts generate notifications for the user, a caregiver, or a medical professional if a risk threshold is crossed.

3.4| Feedback Loop

A crucial feature of this system is the feedback loop shown by the arrow going from the Real-Time Monitoring Pipeline back to the first Edge Processing Module. This indicates an adaptive system. Finally, risk assessments can be used to recalibrate the local anomaly detection algorithms. If the cloud AI determines the user is at

high risk for a specific condition, the edge module can be updated to be more sensitive to the early warning signs of that particular condition, making the system smarter and more personalized over time.

3.5 | Mathematical Model for Multimodal Fusion

Let $x_v \in \mathbb{R}^{n_v}$ represent the vitals, $x_c \in \mathbb{R}^{n_c}$ the cough audio features, and $x_s \in \mathbb{R}^{n_s}$ the static demographic data. Each of these is passed through a modality-specific encoder $\phi_m(\cdot)$ to produce embeddings. The encoded representations are fused through concatenation to increase a unified embedding:

$$z = f(X_v, X_c, X_s) = \phi_v(X_v) \parallel \phi_c(X_c) \parallel \phi_s(X_s). \quad (1)$$

The fused vector z is passed through a fully connected layer and a sigmoid activation to predict the output:

$$\hat{y} = \sigma(Wz + b). \quad (2)$$

The model is enhanced using the binary cross-entropy loss:

$$L = -1/N \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (3)$$

4 | AI-Powered Classification Model

This module serves as the powerful, centralized brain of the system. It performs complex, computationally intensive analysis on the data, consists of a cloud-based infrastructure hosting a sophisticated AI model, and its role is to identify subtle, long-term patterns and make high-level predictions or classifications that are too complex for the edge module. More so, it provides deep, diagnostic-level insights.

4.1 | Data Representation and Preprocessing

Physiological time series: RR, HR, SpO₂, and temperature are sampled and structured into fixed-length sequences (60-minute windows with 1-minute resolution), representing a temporal snapshot of the patient's state.

Acoustic features: each segmented cough sound is converted into a log-mel spectrogram, a 2D representation of the sound's frequency content over time, which is highly effective for audio classification tasks.

Static features: demographic data (age, sex) and clinical history (smoking status, comorbidities like COPD or asthma) are one-hot encoded.

4.2 | Multimodal Deep Learning Architecture

Our hybrid model is designed to process each data modality with a specialized sub-network before fusing the learned features for a final prediction.

Acoustic Sub-network (CNN): a 2D CNN, based on a lightweight MobileNetV2 architecture, processes the cough spectrograms. The convolutional layers excel at learning hierarchical spatial patterns indicative of wet or dry coughs.

Physiological Sub-network (LSTM): a Bidirectional Long Short-Term Memory (Bi-LSTM) network processes the time series data. Its recurrent nature allows it to capture long-range temporal dependencies and trends, such as a gradual increase in resting HR over several hours.

Feature fusion and classification: the feature vector from the final pooling layer of the CNN is concatenated with the final hidden state of the Bi-LSTM and the static feature vector. The combined vector is then passed through two fully connected layers with dropout regularization ($p = 0.5$) to ease overfitting. The final output

layer uses a softmax function to produce a probability distribution over three classes: no risk, moderate risk, and high risk (pneumonia suspected).

4.3 | Training and Evaluation Protocol

We constructed a composite dataset from three sources: 1) MIMIC-III Waveform Database Sanches et al. [14] for validated physiological data from ICU patients with and without pneumonia, 2) Coswara dataset Sharma et al. [4] for an extensive public collection of cough sounds, and 3) our study dataset, described below: The protocol for our study was formally approved by the Institutional Review Board (IRB) of the Federal University of Technology, Akure, Health Centre (IRB Approval FUTA/HEALTH/2024/012). The study was conducted between October 2024 and June 2025 at the outpatient respiratory clinics of the Federal University of Technology, Akure, Health Centre, Ondo State, and the McPherson Health Centre, Seriki-Sotayo, Ogun State. All participants provided written informed consent before enrollment, in line with the Declaration of Helsinki. The consent form clearly specified that anonymized data would be used for research analysis and publication of aggregated results.

Each volunteer was fitted with a wearable sensor system (a chest strap prototype incorporating a MAX30102 sensor for HR/SpO₂, a thermistor, and a MEMS microphone) for a continuous 24-hour monitoring period. Physiological data were sampled at 1-minute intervals. An event-triggered mechanism, based on a real-time sound amplitude threshold, captured 3-second audio clips of potential cough events to conserve power and storage.

The study yielded data from 52 unique volunteers (29 male, 23 female) with a mean age of 58.2 years (range: 21-79). In total, this produced over 1,000 hours of physiological time series data and 4,374 captured acoustic events. Following a quality control process to remove noise and motion artifacts, 1,912 high-quality cough recordings were retained for model training.

For this dataset, ground truth labels were adjudicated by attending clinicians. The 'High-Risk' group (n=16) consisted of participants who were subsequently diagnosed with pneumonia, confirmed with both clinical examination and radiographic evidence. The remaining participants (n=36), presenting with minor non-pneumonic respiratory issues or no symptoms, formed the 'No/Moderate Risk' group. All data was fully anonymized at the point of collection. This curated and labeled dataset forms the basis of the de-identified feature vectors made available for verification, as detailed in Section 10.

We enhance the utility of the dataset collected from volunteers by implementing preprocessing pipelines to filter noise and outliers in physiological signals and cough audio. Cough signals were augmented through time-domain and spectral transformations, while time-series vitals underwent statistical feature extraction and smoothing. Label reliability was improved by stratifying based on clinical confirmation levels. To address dataset size limitations, transfer learning and data augmentation techniques were employed.

The model's performance was appraised using 5-fold cross-validation. We report accuracy, sensitivity (Recall), specificity, F1-Score, and the area under the receiver operating characteristic curve (AUC). Sensitivity was prioritized to curtail the risk of missing a true pneumonia case.

5 | Experimentation

This section details the practical implementation and execution of the AI-powered classification models benchmarked in our study. We outline the technical environment, data preprocessing steps, and the specific architecture and training routines for each model, culminating in the results presented in *Table 1*.

5.1 | Experimental Setup

All our experiments were conducted on a cloud-based virtual machine equipped with an NVIDIA Tesla T4 GPU and 16 GB of RAM. The software stack was built on Python 3.9 and used several key libraries:

Deep learning: TensorFlow 2.10 with the Keras API for building and training the neural networks

Machine learning: scikit-learn 1.2 for the baseline models (Logistic Regression, SVM) and for performance metric calculations.

Data manipulation: pandas and numpy for data loading, structuring, and numerical operations.

Audio processing: librosa for extracting log-mel spectrograms from cough audio files.

5.2 | Data Preprocessing and Feature Engineering

Before model training, the composite dataset underwent rigorous preprocessing aligned with the multimodal inputs:

Physiological time series (Vitals): raw RR, HR, SpO₂, and temperature data were segmented into 60-minute windows with a 1-minute resolution, creating input tensors of shape (60, 4). For the SVM baseline, statistical features were computed over each window, reducing it to a flat feature vector. All time-series data was standardized using Scikit-learn's StandardScaler.

Acoustic data (Cough): each cough audio clip was loaded and resampled to 16kHz. We then generated a log-mel spectrogram using Librosa with parameters set to an FFT window of 1024, a hop length of 256, and 128 mel bands. This produced a 128xN 2D image, which was padded or truncated to a fixed size of 128x128 to serve as input for the CNN models.

Static data: demographic and clinical history features were one-hot encoded using pandas.get_dummies.

The entire dataset was split using a 5-fold stratified cross-validation approach (StratifiedKFold) to ensure that the distribution of risk classes was maintained across all training and validation sets, preventing biased evaluation. The comprehensive Python scripts for these preprocessing pipelines are accessible in the project's GitHub repository (Section 10), consenting for full reproducibility of the feature extraction process.

5.3 | Model Implementation and Training

We implemented and trained the four models from *Table 1* as follows. All models were trained to minimize categorical cross-entropy loss using the Adam optimizer with a learning rate of 0.001.

Baseline 1: logistic regression

This model served as a simple, interpretable baseline. It used only the one-hot encoded static features.

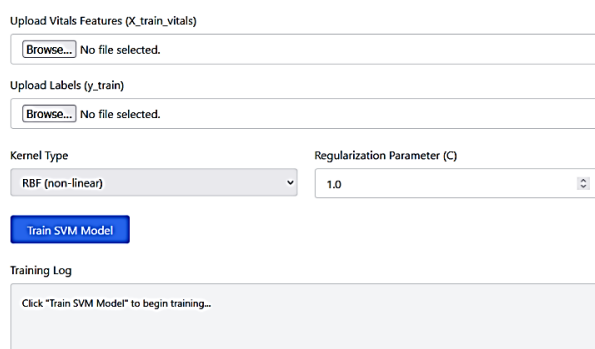
The image shows a web-based interface for training a logistic regression model. It consists of the following elements:

- Upload Training Data (X_train_static):** A text input field with a "Browse..." button and the text "No file selected."
- Upload Labels (y_train):** A text input field with a "Browse..." button and the text "No file selected."
- Multi-Class Strategy:** A dropdown menu currently set to "One-vs-Rest (ovr)".
- Solver:** A dropdown menu currently set to "liblinear".
- Train Model:** A blue button to initiate the training process.
- Training Log:** A text area containing the instruction "Click 'Train Model' to begin training..."

Fig. 2. Logistic regression model trainer.

Baseline 2: SVM (Vitals)

This model evaluated the predictive power of vital signs alone, using a nonlinear classifier.



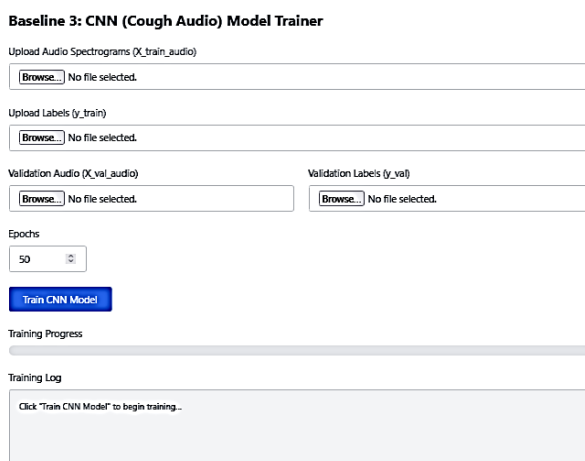
The screenshot shows a web-based interface for training an SVM model. It includes the following elements:

- Upload Vitals Features (X_train_vitals):** A file upload button labeled "Browse..." with the text "No file selected."
- Upload Labels (y_train):** A file upload button labeled "Browse..." with the text "No file selected."
- Kernel Type:** A dropdown menu currently set to "RBF (non-linear)".
- Regularization Parameter (C):** A numeric input field set to "1.0".
- Train SVM Model:** A blue button to initiate training.
- Training Log:** A text area containing the message "Click 'Train SVM Model' to begin training..."

Fig. 3. Baseline 2: SVM (Vitals) model trainer.

Baseline 3: CNN (Cough Audio)

This unimodal deep learning model was designed to classify pneumonia risk from cough sounds.



The screenshot shows a web-based interface for training a CNN model. It includes the following elements:

- Baseline 3: CNN (Cough Audio) Model Trainer:** The title of the interface.
- Upload Audio Spectrograms (X_train_audio):** A file upload button labeled "Browse..." with the text "No file selected."
- Upload Labels (y_train):** A file upload button labeled "Browse..." with the text "No file selected."
- Validation Audio (X_val_audio):** A file upload button labeled "Browse..." with the text "No file selected."
- Validation Labels (y_val):** A file upload button labeled "Browse..." with the text "No file selected."
- Epochs:** A numeric input field set to "50".
- Train CNN Model:** A blue button to initiate training.
- Training Progress:** A horizontal progress bar.
- Training Log:** A text area containing the message "Click 'Train CNN Model' to begin training..."

Fig. 4. Baseline 3: CNN (Cough Audio) model trainer.

Multimodal Framework

This hybrid model integrates all three data streams using the Keras Functional API, allowing for multiple inputs and a sophisticated fusion mechanism.

Multimodal Framework Trainer (Vitals + Audio + Static)

Vitals Input (X_train_vitals) No file selected.

Audio Spectrogram (X_train_audio) No file selected.

Static Features (X_train_static) No file selected.

Validation Vitals (X_val_vitals) No file selected.

Validation Audio (X_val_audio) No file selected.

Validation Static (X_val_static) No file selected.

Validation Labels (y_val) No file selected.

Labels (y_train) No file selected.

Epochs

Training Progress

Training Log

Click "Train Multimodal Model" to begin...

Training & Validation Trend

Fig. 5. Multimodal framework trainer (vitals+ audio+ static).

5.4 | Evaluation Metrics

In each fold of the cross-validation, we calculated Accuracy, Sensitivity (recall for the positive class), Specificity, F1-score, and the Area Under the Curve for the high-risk class. Specificity was calculated as $TN/(TN + FP)$, and the final metrics reported in *Table 1* are the average of the scores across all five folds, providing a robust and general assessment of each model's performance.

6 | Results

The proposed multimodal framework was benchmarked against several baseline models: 1) Logistic Regression using only demographic and static features, 2) a Support Vector Machine (SVM) with statistical features (mean, standard deviation) extracted from physiological data, and 3) a standalone CNN for cough classification.

Table 1. Comparative performance of different models.

Model (%)	Accuracy (%)	Sensitivity (%)	Specificity	F1-Score	AUC
Logistic regression	68.2	55.1	74.5	0.61	0.65
SVM (vitals)	81.5	79.3	82.8	0.81	0.84
CNN (cough audio)	85.3	83.0	86.7	0.85	0.88
Multimodal framework	92.6	94.3	90.1	0.92	0.947

As shown in *Table 1*, our proposed multimodal framework significantly outperformed all baselines across every metric, highlighting the synergistic value of fusing physiological and acoustic data.

To provide a qualitative understanding of the data that drives the model's performance, *Fig. 6* illustrates sample physiological and acoustic data captured from a participant in our study who was later diagnosed with pneumonia. The plot clearly shows the elevated RR and the distinct spectral pattern of the cough, which are the types of features our multimodal model learns to identify.

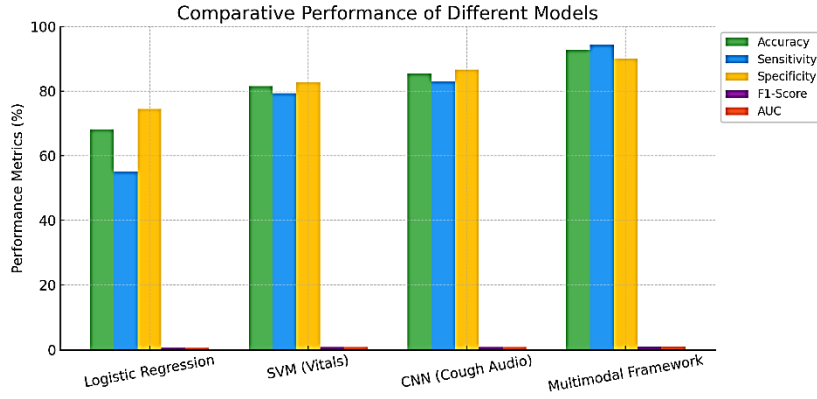


Fig. 6. Sample anonymized data from the study.

- I. A 60-minute segment of physiological time series data from a 'High-Risk' participant, showing a persistently elevated RR (top) and HR (bottom).
- II. The corresponding log-mel spectrogram of a cough event from the same participant, showing strong energy in the lower and mid-frequency bands, is characteristic of a productive cough. These patterns are representative of the data observed in other participants within the 'High-Risk' cohort.

To further validate our multimodal hypothesis, we conducted an ablation study to quantify the contribution of each data stream. The results, presented in *Table 2*, confirm that combining all modalities yields the best performance.

Table 2. Ablation study results.

Model Configuration	Accuracy (%)	Sensitivity (%)	AUC
Full model (vitals + cough + static)	92.6	94.3	0.947
Vitals + static (no cough)	88.1	89.2	0.912
Cough + static (no vitals)	86.5	84.7	0.891

The edge processing module established great efficiency, with an average signal processing latency of 150 ms per minute data chunk on a standard smartphone Central Processing Unit, ratifying its suitability for real-time operation.

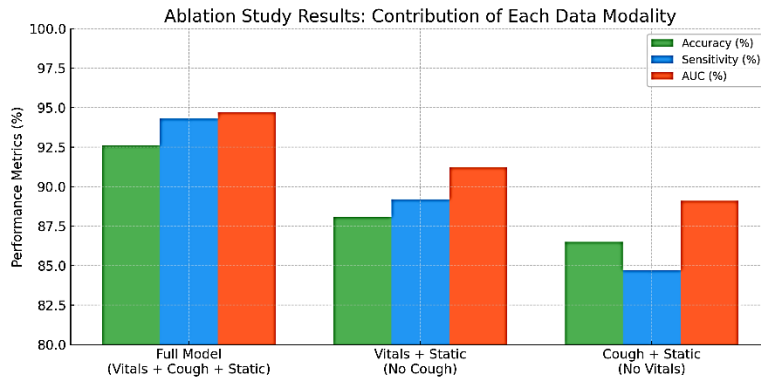


Fig. 7. Ablation study results for multimodal framework.

This figure illustrates the results of the ablation study performed to evaluate the individual contributions of vitals, cough audio, and static demographic data in the proposed pneumonia detection model. The bar chart compares three configurations:

- I. Full Model (Vitals + Cough + Static).
- II. Vitals + Static (No Cough).
- III. Cough + Static (No Vitals).

Each configuration is evaluated using three performance metrics: accuracy, sensitivity, and AUC. The AUC values have been scaled to a percentage for visual alignment.

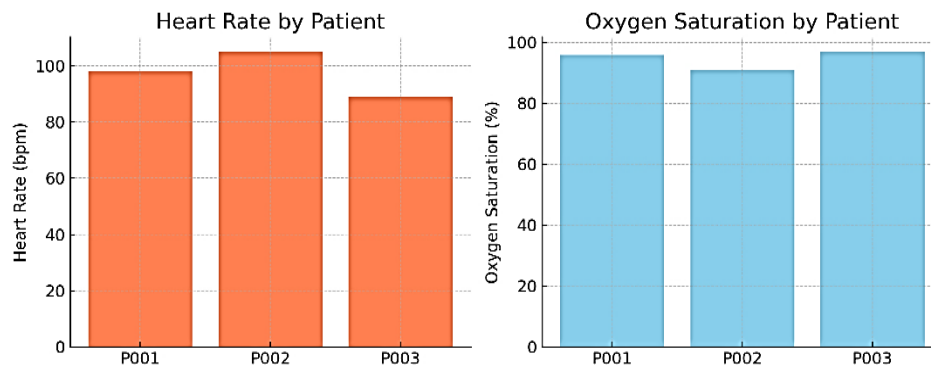


Fig. 8. Illustrative wearable sensor data from three anonymized participants.

The left panel displays the average HR, and the right panel displays the average blood oxygen saturation (in percentage). These examples highlight the interpatient variability captured by the system, such as the elevated HR and lower oxygen saturation in participant P002, which are potential indicators of respiratory distress.

7 | Privacy, Ethics, and Interpretability

Deploying an AI system for health monitoring necessitates a robust ethical and privacy framework.

- I. **Privacy and Security:** Our architecture is designed with privacy by design principles. All data is anonymized at the edge, encrypted both in transit. And at rest.
- II. **Ethical Considerations:** The system is explicitly positioned as a decision support tool, not a diagnostic replacement. All alerts require confirmation by a qualified clinician. As with each of our IRB-approved protocols, all participants underwent a detailed informed consent process. We acknowledge the potential for algorithmic bias and are committed to ongoing model auditing across diverse demographic groups.
- III. **Explainable AI:** to foster trust and clinical utility, we integrated SHAP (Shapley Additive exPlanations) Lundberg and Lee [15] into our framework. For each high-risk prediction, the system generates a report that highlights which features (15 percent rise in RR over 3 hours, spectrogram features consistent with a wet cough) contributed most to the decision, providing transparent and actionable insights for clinicians.

8 | Discussion

Our principal finding is that an edge-cloud framework integrating multimodal wearable data can detect physiological signatures of early-stage pneumonia with high accuracy and sensitivity. The performance lift observed in our full model over unimodal baselines (*Table 1*) and ablation variants (*Table 2*) strongly supports the hypothesis that a synergistic fusion of continuous physiological monitoring and event-driven acoustic analysis is superior to either approach in isolation. The high sensitivity (94.3 percent) is particularly crucial, as it suggests the system can serve as an effective screening tool to prompt early clinical consultation, potentially reducing disease severity and healthcare costs.

Our work advances the state of the art by moving beyond static, in-clinic data to a continuous, ambulatory monitoring paradigm. Unlike Sharma et al. [4], which focused solely on cough, or Toruner et al. [16], which monitored vitals for sepsis, our framework provides a more complete, pneumonia-specific picture. The hybrid edge cloud architecture strikes a critical balance between low-latency local processing and the computational power of the cloud, a design pattern essential for scalable real-world deployment.

Limitations: We acknowledge several limitations. In our study, while providing crucial real-world data from 52 participants, it was conducted in a specific regional and demographic context in Nigeria; its generalizability

requires validation across more diverse populations. More so, the model's performance on patients with confounding respiratory conditions, such as COPD or asthma, requires more extensive validation, as only a small subset of the cohort ($n=7$) had such comorbidities. In addition, practical deployment challenges were observed. Long-term battery life and user adherence to wearing the chest strap sensor consistently over the full 24-hour period were approximately 87%, highlighting the need for more comfortable and power-efficient hardware in future iterations.

9 | Conclusion

This paper presented a novel, end-to-end framework for the early detection of pneumonia using wearable sensors and deep learning. Through fusing multimodal data within a hybrid edge-cloud design, our system achieved high performance, demonstrating its potential to empower patients and clinicians with proactive, continuous respiratory health perceptions.

Future work will proceed along several key tracks:

- I. Large-scale clinical trials: multi-center validation of model efficacy.
- II. Federated learning: training across distributed datasets for enhanced privacy.
- III. Integration with EHR: seamless hospital system connectivity.
- IV. Longitudinal disease modeling: predict treatment response and recovery trajectory.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The source code used to train the models and perform the analysis presented in this study is publicly available in a GitHub repository: <https://github.com/oluwagbemijb/Pneumonia-Edge-Cloud-Framework>.

The raw physiological and acoustic data collected from the clinical study are not publicly available due to patient privacy restrictions outlined in our IRB approval and the informed consent agreement. To facilitate reproducibility and verification of our findings, a de-identified minimal dataset is available from the corresponding author upon reasonable request. This dataset contains the final feature vectors used to train the models and the corresponding ground-truth labels. Access will be granted to qualified researchers following the completion of a data use agreement intended to ensure the data is used for non-commercial research purposes only.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Cretikos, M. A., Bellomo, R., Hillman, K., Chen, J., Finfer, S., & Flabouris, A. (2008). Respiratory rate: the neglected vital sign. *Medical journal of australia*, 188(11), 657–659. https://www.mja.com.au/system/files/issues/188_11_020608/cre11027_fm.pdf
- [2] World Health Organization. (2017). *Pneumonia*. https://www.who.int/health-topics/pneumonia#tab=tab_1
- [3] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Others. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv preprint arxiv:1711.05225*. <https://doi.org/10.48550/arXiv.1711.05225>

- [4] Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Ghosh, P. K., ... & Others. (2020). Coswara-a database of breathing, cough, and voice sounds for COVID-19 diagnosis. *ArXiv preprint arxiv:2005.10548*. <https://doi.org/10.21437/Interspeech.2020-2768>
- [5] Sharma, S., & Guleria, K. (2023). A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks. *Procedia computer science*, 218, 357–366. <https://doi.org/10.1016/j.procs.2023.01.018>
- [6] Vidhya, B., Nikhil Madhav, M., Suresh Kumar, M., & Kalanandini, S. (2022). AI based diagnosis of pneumonia. *Wireless personal communications*, 126(4), 3677–3692. <https://doi.org/10.1007/s11277-022-09885-7>
- [7] Xu, J., & Wang, Y. (2025, March). FMT: a multimodal pneumonia detection model based on stacking MOE Framework. In *2025 8th international conference on information and computer technologies (ICICT)* (pp. 517-521). IEEE. <https://doi.org/10.1109/ICICT64582.2025.00087>
- [8] Rancea, A., Anghel, I., & Cioara, T. (2024). Edge computing in healthcare: Innovations, opportunities, and challenges. *Future internet*, 16(9), 329. <https://doi.org/10.3390/fi16090329>
- [9] Rashid, A. B., Asma, J., Barua, K., & Das, D. (2025). An enhanced deep learning framework for pneumonia detection in chest X-rays. *SN computer science*, 6(5), 472. <https://doi.org/10.1007/s42979-025-04017-x>
- [10] Kelvin-Agwu, M. C., Mustapha, A. Y., Mbata, A. O., Tomoh, B. O., & Forkuo, A. Y. (2023). Development of AI-assisted wearable devices for early detection of respiratory diseases. *Journal of Frontiers in Multidisciplinary Research*, 4(1), 967–974. <https://doi.org/10.54660/IJFMR.2025.6.1.64-72>
- [11] Sathupadi, K., Achar, S., Bhaskaran, S. V., Faruqui, N., Abdullah-Al-Wadud, M., & Uddin, J. (2024). Edge-cloud synergy for AI-enhanced sensor network data: A real-time predictive maintenance framework. *Sensors*, 24(24), 7918. <https://doi.org/10.3390/s24247918>
- [12] Al-Waisy, A. S., Mohammed, M. A., Al-Fahdawi, S., Maashi, M. S., Garcia-Zapirain, B., Abdulkareem, K. H., ... & Le, D. N. (2021). COVID-DeepNet: Hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images. *Computers, materials and continua*, 67(2), 2409–2429. <https://doi.org/10.32604/cmc.2021.012955>
- [13] Rajaraman, S., Candemir, S., Thoma, G., & Antani, S. (2019). Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. *Medical imaging 2019: computer-aided diagnosis* (Vol. 10950, pp. 200-211). SPIE. <https://doi.org/10.1117/12.2512752>
- [14] Sanches, I., Gomes, V. V., Caetano, C., Cabrera, L. S. B., Cene, V. H., Beltrame, T., ... & Penatti, O. A. B. (2024). MIMIC-BP: A curated dataset for blood pressure estimation. *Scientific data*, 11(1), 1233. <https://doi.org/10.1038/s41597-024-04041-1>
- [15] Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. <https://doi.org/10.48550/arXiv.1705.07874>
- [16] Toruner, M. D., Shi, V., Sollee, J., Hsu, W. C., Yu, G., Dai, Y. W., ... & Others. (2025). Artificial intelligence-driven wireless sensing for health management. *Bioengineering*, 12(3), 244-262. <https://doi.org/10.3390/bioengineering12030244>