

# SmartCoughNet: A Deep Learning-Based Mobile Framework for Respiratory Disease Detection Using Cough Audio Analysis

Biswarup Dutta Gupta<sup>1\*</sup>, Ankita Sadhukhan<sup>1\*</sup>, Nitu Bagchi<sup>1\*</sup>, Abanti Ghosh<sup>2</sup>, Tathagata Roy Chowdhury<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering (AI/ML), Techno Institute of Engineering and Management, Kolkata, India

<sup>2</sup> Ideal Institute of Engineering, Kolkata, India

\* Email ID: [biswarup24430823010@tec.edu.in](mailto:biswarup24430823010@tec.edu.in)

## ARTICLE INFO

*Article type:*

Research Article

*Article history:*

Received: 11-07-2026

Received in revised form:  
04-08-2026

Accepted: 12-09-2026

Available online: 13-09-2026

*Keywords:*

Cough Audio Analysis

Deep Learning

Healthcare AI

Mel-Spectrogram

Mobile Health Monitoring

Respiratory Disease Detection

## ABSTRACT

Respiratory diseases remain a major global public health burden, and timely screening can support earlier medical assessment and intervention. Conventional diagnosis relies on clinical examination, laboratory testing, medical imaging, and specialist interpretation, which are often costly and unavailable in rural and resource-constrained settings. This study proposes SmartCoughNet, a deep learning framework that analyses cough recordings by converting them into Mel-spectrogram representations and classifying them into Negative, Positive, and Unknown categories. Experiments used 1,430 recordings (1,207 Negative, 150 Positive, 73 Unknown) with an 80:20 stratified split and class-weighted training of ResNet50, EfficientNet-B0, and MobileNet-V3 Large. EfficientNet-B0 achieved the highest overall accuracy of 84 percent, followed by ResNet50 at 83 percent and MobileNet-V3 Large at 72 percent. Class-wise precision, recall, and F1-score showed stronger performance on the majority Negative class, while the minority Positive and Unknown categories remained harder to classify. A mobile-compatible workflow with spectrogram visualisation and prediction outputs is presented for preliminary screening and remote monitoring. Findings are an initial experimental evaluation requiring validation on larger, balanced, and independently collected datasets before clinical consideration.

## 1. INTRODUCTION

Respiratory diseases are among the leading causes of mortality and long-term healthcare burden worldwide, affecting millions of people every year. Conditions such as COVID-19, asthma, bronchitis, pneumonia, tuberculosis, and chronic obstructive pulmonary disease impair quality of life and increase global healthcare expenditure. Early detection of respiratory disease is therefore critical, since delayed diagnosis can lead to serious complications, repeated hospital visits, and greater mortality risk.

Conventional diagnosis commonly uses chest X-ray imaging, CT scans, spirometry, laboratory tests, PCR analysis, and clinical examination. Although reliable, these techniques are often expensive and time-consuming and depend on advanced infrastructure and trained professionals. These limitations are most acute in rural and resource-constrained settings with limited access to health facilities. Demand is therefore growing for intelligent, low-cost, non-invasive, and automated respiratory screening systems that support rapid preliminary assessment and remote health monitoring. Recent advances in artificial intelligence, machine learning, and deep learning have strengthened predictive analytics and automated biomedical signal analysis, and cough-sound analysis has emerged as a promising non-invasive biomarker because respiratory abnormalities alter cough frequency, amplitude, duration, and spectral characteristics.

Convolutional neural networks have achieved strong results in spectrogram-based audio classification because they automatically learn complex hierarchical acoustic features from biomedical audio representations. Driven by these advances, this work presents SmartCoughNet, a deep learning framework for respiratory disease detection from cough audio with mobile healthcare integration. The framework converts cough recordings into Mel-spectrogram representations, compares machine learning and deep learning classifiers, and integrates a lightweight CNN into a mobile-compatible application for real-time analysis and remote monitoring, aiming to provide an accessible, scalable, and cost-effective screening aid.

### 1.1 Major Contributions

The major contributions of this work are: a lightweight SmartCoughNet framework for respiratory disease detection from cough audio; Mel-spectrogram acoustic feature extraction with standardised audio preprocessing; comparative evaluation of Logistic Regression, Random Forest, SVM, KNN, ResNet, and CNN classifiers; selection of the best CNN configuration with confusion-matrix and standard-metric analysis; and a mobile-compatible healthcare application for real-time screening and remote monitoring.

### 1.2 Organisation of the Paper

The remainder of this paper is organised as follows. Section 2 reviews related work on cough-based respiratory detection. Section 3 describes the SmartCoughNet workflow, dataset, and training protocol. Section 4 details data preprocessing. Section 5 presents the CNN architecture. Section 6 reports and discusses experimental results. Section 7 describes mobile deployment. Section 8 concludes, and Section 9 outlines future scope.

## 2. LITERATURE REVIEW

Recent advances in artificial intelligence and deep learning have improved automated respiratory disease detection from cough sounds, and several groups have explored spectrogram-based learning for classifying respiratory abnormalities from acoustic signals. Ghrabli et al. introduced a spectral fingerprint framework for identifying respiratory diseases from cough sounds using deep learning architectures [1]. Shen et al. proposed audio feature extraction and augmentation techniques for cough-based respiratory disease classification [2]. Ghourabi et al. compared Swin Transformer and CNN architectures on spectrogram representations for cough sound analysis [3].

Researchers have also explored multimodal and lightweight systems for practical deployment. Malik et al. developed a multimodal framework combining cough audio analysis and medical imaging [4]. Pham et al. proposed a CNN mixture-of-experts framework for respiratory anomaly classification [5]. Garcia-Ordas et al. investigated convolutional neural networks and variational autoencoders for respiratory pathology detection under imbalanced data conditions [6].

## 3. MATERIALS AND METHODS

The SmartCoughNet framework follows a structured deep learning pipeline for cough audio analysis and respiratory condition classification. The workflow comprises cough audio processing, Mel-spectrogram generation, deep learning classification, performance evaluation, and a mobile-oriented deployment concept. The complete workflow is illustrated in **Figure 1**. The dataset contained cough recordings in three classes, Positive, Negative, and Unknown. Because recordings came from different sources with varying durations and sampling characteristics, a standardised preprocessing procedure was applied before training. Each signal was loaded and resampled to 22,050 Hz; recordings longer than five seconds were truncated and shorter ones zero-padded to a uniform duration. A 128-band Mel-spectrogram was generated per recording, converted to a logarithmic decibel scale, normalised, transformed into an RGB representation, and resized to 224 by 224 pixels for model input. To reduce the effect of class imbalance, the data were divided with an 80:20 stratified split (random state 42), and class weights were used during training to give minority classes greater importance and reduce the dominance of the Negative class.

Three pre-trained architectures, ResNet50, EfficientNet-B0, and MobileNet-V3 Large, were evaluated on the three-class task by adapting their final classification layers. Performance was assessed with overall accuracy and class-wise precision, recall, and F1-score so behaviour on majority and minority classes could be examined. The framework was designed with mobile healthcare use in mind: a cough recording is processed, converted into a Mel-spectrogram, and passed to a trained model for prediction as a preliminary

screening and decision-support aid rather than a replacement for clinical diagnosis, providing a basis for lightweight and accessible respiratory monitoring through mobile and remote applications.

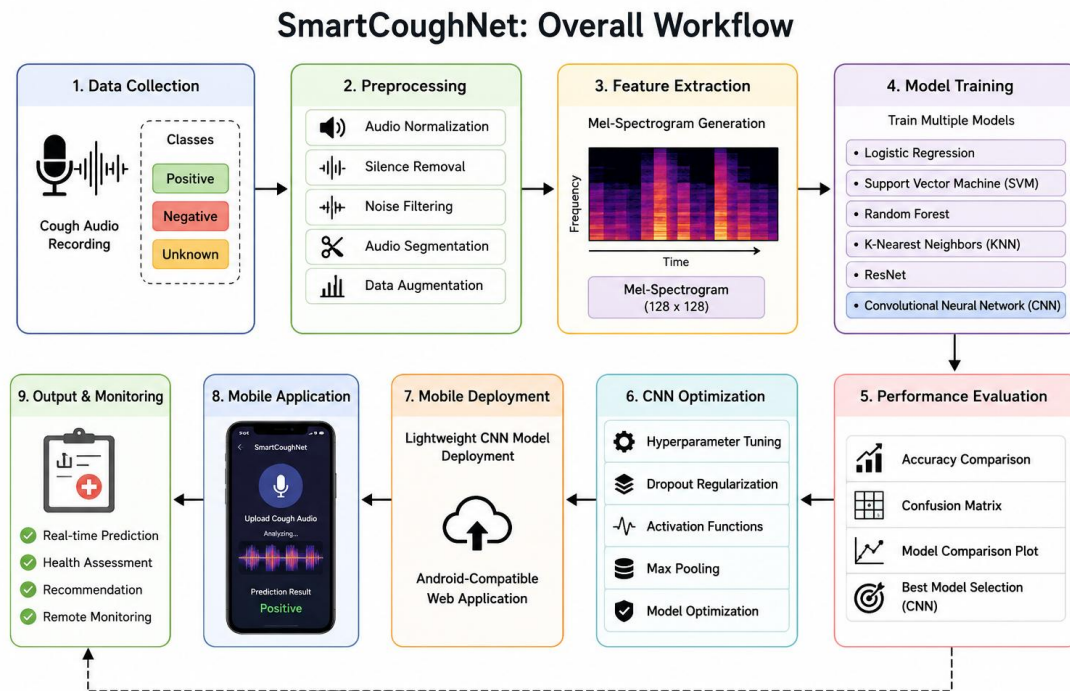


Figure 1. Overall workflow of SmartCoughNet.

### 3.1 Algorithmic Steps of the Proposed Framework

#### 3.1.1 Data Collection

Cough audio recordings were collected and organised into Positive, Negative, and Unknown categories for three-class respiratory condition classification.

#### 3.1.2 Audio Standardisation

Each recording was loaded with the Librosa library and resampled to 22,050 Hz. Recordings longer than five seconds were truncated and shorter ones zero-padded to a consistent duration.

#### 3.1.3 Mel-Spectrogram Generation

A 128-band Mel-spectrogram was generated per standardised recording, converted to the decibel scale, and normalised to a consistent feature representation.

#### 3.1.4 Image Preparation

Spectrograms were converted into RGB representations and resized to 224 by 224 pixels as uniform model inputs.

#### 3.1.5 Data Partitioning and Class-Imbalance Handling

Data were divided with an 80:20 stratified split (random state 42). Class weights during training reduced majority-class influence and emphasised minority-class samples.

#### 3.1.6 Model Training

ResNet50, EfficientNet-B0, and MobileNet-V3 Large were adapted and trained for the three-class cough classification task.

#### 3.1.7 Performance Evaluation

Models were evaluated with overall accuracy and class-wise precision, recall, and F1-score, with confusion-matrix analysis across the Positive, Negative, and Unknown categories.

### 3.1.8 Comparative Analysis

The three architectures were compared on overall accuracy and class-wise behaviour, particularly for the minority classes.

### 3.1.9 Mobile-Oriented Prediction Framework

The framework was designed for mobile-compatible use: a user cough recording is processed into a Mel-spectrogram and analysed by a trained model to provide a preliminary prediction supporting early screening and remote monitoring, not replacing clinical diagnosis.

## 3.2 Dataset Description

The study used 1,430 cough audio recordings in three labelled categories: 1,207 Negative (84.41 percent), 150 Positive (10.49 percent), and 73 Unknown (5.10 percent). This imbalance was carried through model development and evaluation, since overall accuracy alone may misrepresent classifier performance on minority classes. Recordings were organised into class-labelled folders, with Negative, Positive, and Unknown retained as three separate classes rather than a binary Positive-Negative task. Before training, each recording was loaded with Librosa, resampled to 22,050 Hz, truncated or zero-padded to five seconds, converted into a 128-band Mel-spectrogram on the decibel scale, normalised, rendered as RGB, and resized to 224 by 224 pixels. The main evaluation used an 80:20 stratified split (random state 42) preserving class proportions, with class weights countering the substantially larger Negative class. The dataset therefore supports evaluating how well deep learning distinguishes Negative, Positive, and Unknown cough patterns, although the small Positive and Unknown counts may affect the stability of class-wise estimates, which is why precision, recall, and F1-score are reported alongside accuracy. **Table 1** presents class distribution of the cough audio dataset.

**Table 1. Class distribution of the cough audio dataset.**

| Class    | Number of Recordings | Percentage (%) |
|----------|----------------------|----------------|
| Negative | 1207                 | 84.41          |
| Positive | 150                  | 10.49          |
| Unknown  | 73                   | 5.10           |
| Total    | 1430                 | 100.00         |

The dataset folder structure is illustrated in Figure 2. Organising recordings by class label supported systematic data loading and label assignment during training and evaluation. The recordings provide the basis for distinguishing Negative, Positive, and Unknown cough patterns, although the substantial imbalance remains an important limitation, particularly the small Positive and Unknown counts, motivating the class-wise metric analysis.

#### Dataset/

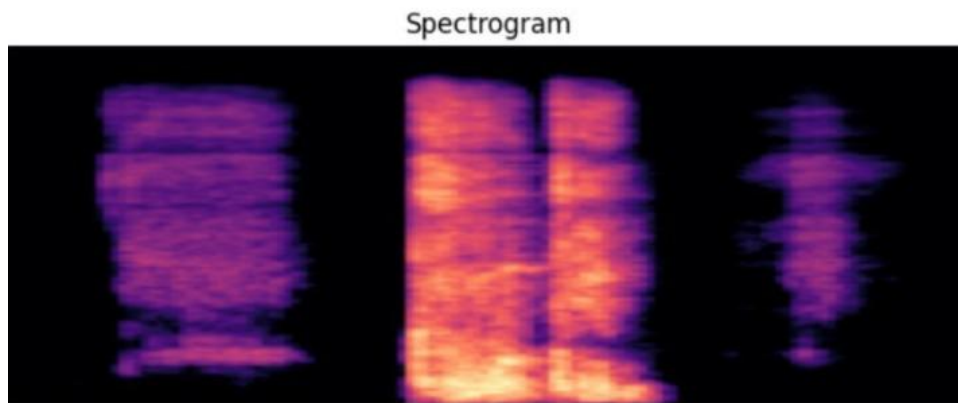
- **Negative/ (1207 files)**
  - 124\_Negative\_male\_55.wav
  - 584\_Negative\_male\_44.wav
  - 847\_Negative\_male\_19.wav
  - ...
- **Positive/ (150 files)**
  - 245\_Positive\_male\_27.wav
  - 1077\_Positive\_female\_20.wav
  - 1523\_Positive\_male\_92.wav
  - ...
- **Unknown/ (73 files)**
  - 1150\_Unknown\_female\_34.wav
  - 195\_Unknown\_male\_31.wav
  - 447\_Unknown\_female\_46.wav
  - ...

**Figure 2. Organisation of the cough audio dataset according to class labels: Dataset/ with Negative/ (1207 files), Positive/ (150 files), and Unknown/ (73 files).**

## 4. DATA PREPROCESSING

Preprocessing converted cough recordings of varying duration and sampling characteristics into a consistent representation for deep learning classification. Rather than manually designed denoising or

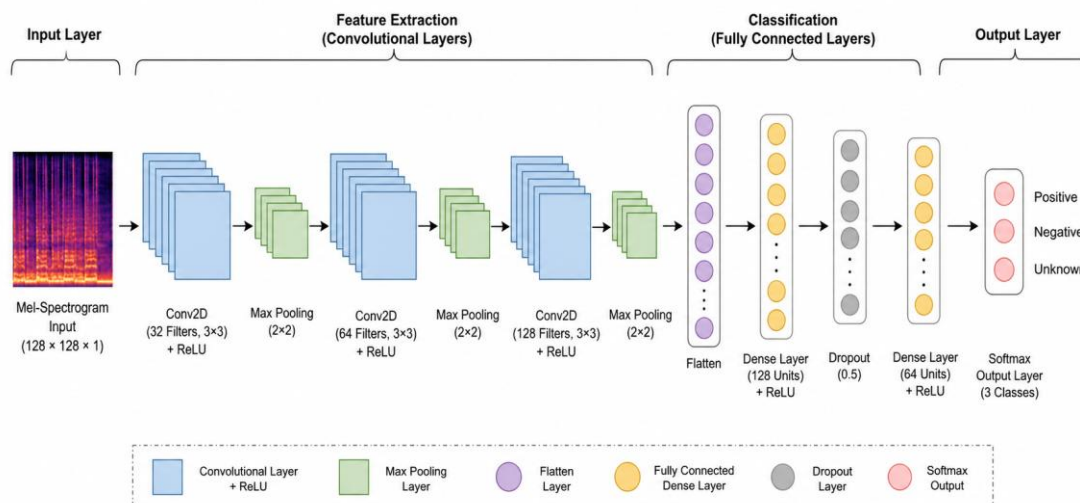
segmentation chains, the pipeline standardised the audio signals and transformed them into Mel-spectrogram representations through: (1) loading each recording with Librosa; (2) resampling to 22,050 Hz; (3) standardising duration to a maximum of five seconds by truncation or zero-padding; (4) generating a 128-band Mel-spectrogram of time-frequency characteristics; (5) converting to the logarithmic decibel scale; (6) normalising spectrogram values; (7) converting to an RGB image; and (8) resizing to 224 by 224 pixels for the selected architectures. The resulting Mel-spectrogram images standardised cough signals into inputs for ResNet50, EfficientNet-B0, and MobileNet-V3 Large, preserving time-frequency characteristics while unifying recordings of different durations for training and evaluation. **Figure 3** presents generated Mel-spectrogram representation.



**Figure 3. Generated Mel-spectrogram representation.**

## 5. CNN MODEL ARCHITECTURE

A convolutional neural network classifies respiratory conditions from Mel-spectrogram representations of cough audio. The architecture applies convolutional and pooling layers for automatic feature extraction followed by fully connected dense layers for classification, with convolutional layers, ReLU activations, max pooling layers, dropout regularisation, fully connected dense layers, and a softmax output layer, trained with the Adam optimiser and categorical cross-entropy loss. The CNN automatically learns discriminative spectro-temporal features from Mel-spectrogram images for respiratory disease classification, and dataset division into training and testing subsets tests the generalisation capability of the framework. **Figure 4** presents CNN architecture for cough classification. **Table 2** presents training configuration of the proposed CNN model.



**Figure 4. CNN architecture for cough classification.**

**Table 2. Training configuration of the proposed CNN model.**

| Parameter     | Value           |
|---------------|-----------------|
| Input Feature | Mel-Spectrogram |

| Parameter           | Value                     |
|---------------------|---------------------------|
| Optimizer           | Adam                      |
| Loss Function       | Categorical Cross-Entropy |
| Activation Function | ReLU                      |
| Output Activation   | Softmax                   |
| Dropout             | 0.5                       |
| Evaluation Metric   | Accuracy                  |

## 6. RESULTS AND DISCUSSION

Experiments examined three pre-trained architectures on Negative, Positive, and Unknown cough classification with an 80:20 stratified split (1,144 training, 286 testing recordings). Class weights addressed the substantial imbalance, and performance was evaluated with overall accuracy, class-wise precision, recall, and F1-score, together with confusion-matrix analysis of correct and incorrect predictions.

### 6.1 Overall Model Performance

ResNet50, EfficientNet-B0, and MobileNet-V3 Large were evaluated on the same three-class task with results in **Table 3**. EfficientNet-B0 achieved the highest overall accuracy of 84 percent, followed by ResNet50 at 83 % and MobileNet-V3 Large at 72 %. The gap between EfficientNet-B0 and ResNet50 was small, so results should not be interpreted on overall accuracy alone. Because the Negative class dominated while Positive and Unknown contained far fewer recordings, a model could score high overall accuracy through the majority class while performing weakly on minority categories, which is why class-wise precision, recall, and F1-score were examined alongside accuracy.

**Table 3. Overall classification performance of the evaluated deep learning models.**

| Model              | Accuracy |
|--------------------|----------|
| ResNet50           | 83%      |
| EfficientNet-B0    | 84%      |
| MobileNet-V3 Large | 72%      |

### 6.2 Class-Wise Performance Analysis

Class-wise results differed noticeably across categories, as summarised in Table 4. ResNet50 and EfficientNet-B0 performed strongly on the Negative class with recall of 0.93 each, but Positive-class recall reached only 0.43 for ResNet50 and 0.50 for EfficientNet-B0, so a considerable proportion of Positive recordings were still misclassified despite high overall accuracy. The Unknown class was the most challenging: both models recorded zero precision, recall, and F1-score there. MobileNet-V3 Large showed a different pattern with Positive-class recall of 0.60 but precision of only 0.23, indicating many incorrect Positive predictions. These findings show why overall accuracy alone is insufficient for an imbalanced cough classification system: high accuracy can mask poor performance on clinically important or underrepresented categories, and class-wise metrics give a more detailed picture of each architecture.

**Table 4. Class-wise performance of the evaluated models.**

| Model              | Class    | Precision | Recall | F1-score |
|--------------------|----------|-----------|--------|----------|
| ResNet50           | Negative | 0.88      | 0.93   | 0.90     |
| ResNet50           | Positive | 0.76      | 0.43   | 0.55     |
| ResNet50           | Unknown  | 0.00      | 0.00   | 0.00     |
| EfficientNet-B0    | Negative | 0.89      | 0.93   | 0.91     |
| EfficientNet-B0    | Positive | 0.62      | 0.50   | 0.56     |
| EfficientNet-B0    | Unknown  | 0.00      | 0.00   | 0.00     |
| MobileNet-V3 Large | Negative | 0.91      | 0.78   | 0.84     |
| MobileNet-V3 Large | Positive | 0.23      | 0.60   | 0.33     |
| MobileNet-V3 Large | Unknown  | 1.00      | 0.07   | 0.12     |

### 6.3 Comparative Discussion

EfficientNet-B0 achieved the highest overall accuracy with ResNet50 close behind, and both recognised the majority Negative class effectively, yet neither identified the Unknown class in the test evaluation, indicating that the limited Unknown recordings and dataset imbalance remain important challenges. MobileNet-V3 Large reached lower overall accuracy but higher Positive-class recall than the other models, accompanied by low precision and more false Positive predictions. Selecting a model for practical

screening would therefore trade off overall accuracy, Positive-case sensitivity, false predictions, and computational cost. The experiment does not establish any model as universally superior; architectures showed different strengths and weaknesses on the same imbalanced dataset, and larger balanced datasets would be needed to test whether these differences persist.

#### 6.4 Classification Report Analysis

Classification reports examined model behaviour across the three classes beyond overall accuracy. As shown in **Figure 5**, ResNet50 reached 83 percent accuracy with Negative-class precision 0.88, recall 0.93, and F1-score 0.90, Positive-class recall 0.43, and zero Unknown-class metrics. **Figure 6** presents EfficientNet-B0 at 84 percent overall accuracy with negative-class F1-score 0.91, positive-class precision 0.62, recall 0.50, and F1-score 0.56, and no correctly identified unknown recordings. figure 7 shows MobileNet-V3 large at 72 percent accuracy with positive-class recall 0.60 but precision 0.23, and unknown-class precision 1.00 with recall of only 0.07, detecting very few unknown recordings. overall, all three models performed more reliably on the majority negative class than on minority classes, confirming that accuracy alone masks class-wise differences and that precision, recall, and F1-score evaluate the SmartCoughNet framework more informatively under substantial imbalance.

| ◆ ResNet     |           |        |          |         |
|--------------|-----------|--------|----------|---------|
|              | precision | recall | f1-score | support |
| 0            | 0.88      | 0.93   | 0.90     | 241     |
| 1            | 0.76      | 0.43   | 0.55     | 30      |
| 2            | 0.00      | 0.00   | 0.00     | 15      |
| accuracy     |           |        | 0.83     | 286     |
| macro avg    | 0.55      | 0.45   | 0.48     | 286     |
| weighted avg | 0.82      | 0.83   | 0.82     | 286     |

Figure 5. Classification report of ResNet50 on the test dataset.

| ◆ EfficientNet |           |        |          |         |
|----------------|-----------|--------|----------|---------|
|                | precision | recall | f1-score | support |
| 0              | 0.89      | 0.93   | 0.91     | 241     |
| 1              | 0.62      | 0.50   | 0.56     | 30      |
| 2              | 0.00      | 0.00   | 0.00     | 15      |
| accuracy       |           |        | 0.84     | 286     |
| macro avg      | 0.50      | 0.48   | 0.49     | 286     |
| weighted avg   | 0.81      | 0.84   | 0.83     | 286     |

Figure 6. Classification report of EfficientNet-B0 on the test dataset.

| MobileNet    |           |        |          |         |
|--------------|-----------|--------|----------|---------|
|              | precision | recall | f1-score | support |
| 0            | 0.91      | 0.78   | 0.84     | 241     |
| 1            | 0.23      | 0.60   | 0.33     | 30      |
| 2            | 1.00      | 0.07   | 0.12     | 15      |
| accuracy     |           |        | 0.72     | 286     |
| macro avg    | 0.71      | 0.48   | 0.43     | 286     |
| weighted avg | 0.84      | 0.72   | 0.75     | 286     |

Figure 7. Classification report of MobileNet-V3 Large on the test dataset.

## 6.5 Limitations of the Experimental Results

Several limitations qualify the results. First, the dataset is substantially imbalanced with far more Negative than Positive and Unknown recordings, and imbalance remained a significant challenge despite class weights. Second, the small Positive and Unknown counts may reduce the stability of class-wise estimates, and the poor Unknown-category performance suggests that additional data or minority-class learning approaches are required. Third, evaluation used a single stratified 80:20 split, so results are an initial experimental evaluation rather than evidence of broad clinical generalisability, and future work should include repeated experiments, cross-validation, external datasets, and statistical analysis of performance variation. Finally, the framework is a preliminary screening and decision-support approach whose reported results do not establish clinical diagnostic accuracy, requiring validation on larger independently collected datasets before clinical use. Overall, Mel-spectrogram-based deep learning distinguished cough recordings across the three categories to a certain extent, with EfficientNet-B0 highest at 84 percent and ResNet50 comparable, while class-wise analysis showed substantial majority-minority gaps, underscoring the need to report precision, recall, and F1-score alongside accuracy for a realistic assessment of SmartCoughNet [7, 8].

## 7. MOBILE DEPLOYMENT

For accessibility, usability, and real-time healthcare support, an optimised lightweight CNN was integrated into a mobile-compatible application. Deployment focused on a user-friendly and computationally efficient platform for real-time respiratory analysis from patient cough recordings. Users upload cough recordings through a mobile or web interface; the system then automatically normalises audio, filters noise, removes silence, and generates Mel-spectrograms, which the trained CNN classifies. After analysis, the framework returns instant predictions with confidence scores and healthcare recommendations and visualises waveforms, Mel-spectrogram outputs, and prediction summaries to improve interpretability and user interaction. The framework targets telemedicine and remote healthcare where rapid screening and continuous monitoring matter most, offering a low-cost, non-invasive, and accessible assessment suited to rural and resource-constrained environments with reduced computational overhead for mobile-compatible execution without expensive hardware, demonstrating the practical applicability of artificial intelligence and deep learning in next-generation healthcare and intelligent respiratory monitoring. The application supports uploading cough recordings, real-time analysis, Mel-spectrogram visualisation, prediction confidence scores, instant healthcare recommendations, remote monitoring, mobile accessibility, and waveform and prediction visualisation.

### 7.1 Mobile Interface and Prediction Outputs

The optimised lightweight CNN was integrated into a mobile-compatible framework for accessibility and real-time healthcare use, with the goal of an efficient, user-friendly system giving immediate respiratory predictions from cough recordings. Users upload recordings through a mobile or web interface; on receipt, the system automatically normalises audio and applies noise filtering, silence removal, segmentation, and Mel-spectrogram generation before the trained CNN classifies respiratory conditions. Post-prediction screens show the respiratory category with confidence scores and visual waveform and Mel-spectrogram representations that aid interpretability and user trust, while the lightweight architecture keeps computation feasible on resource-constrained devices for mobile healthcare, telemedicine, and remote monitoring as an

easy-to-use, non-invasive, inexpensive preliminary screening tool. Interactive output pages visualise the analysis and prediction: the first interface uploads recordings and starts analysis, the prediction interface shows Positive, Negative, or Unknown classes with CNN confidence probabilities, and further screens visualise waveform patterns and Mel-spectrogram representations. App-based cough screening has previously been demonstrated for preliminary respiratory assessment, crowdsourced respiratory sound modelling supports the same screening paradigm, cough-only AI diagnosis pipelines establish its feasibility, automated cough and breathing sound detection provides the signal-processing precedent, and systematic reviews of adventitious sound analysis motivate continued validation, while residual learning foundations and their refinements for visual recognition underpin the evaluated backbones [9, 10, 11, 12, 13, 14, 15].

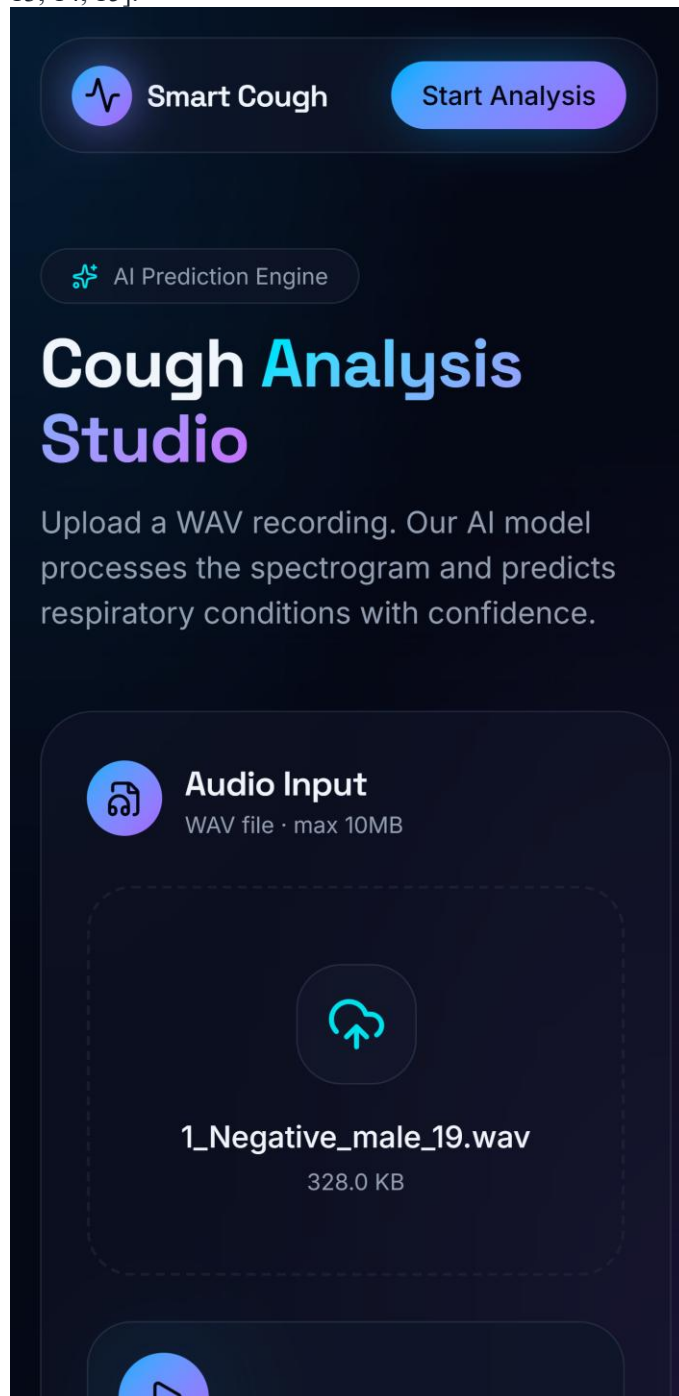


Figure 8. SmartCoughNet mobile and web main interface.

Figure 8 illustrates the main mobile and web interface used to upload cough audio recordings and initiate respiratory analysis.

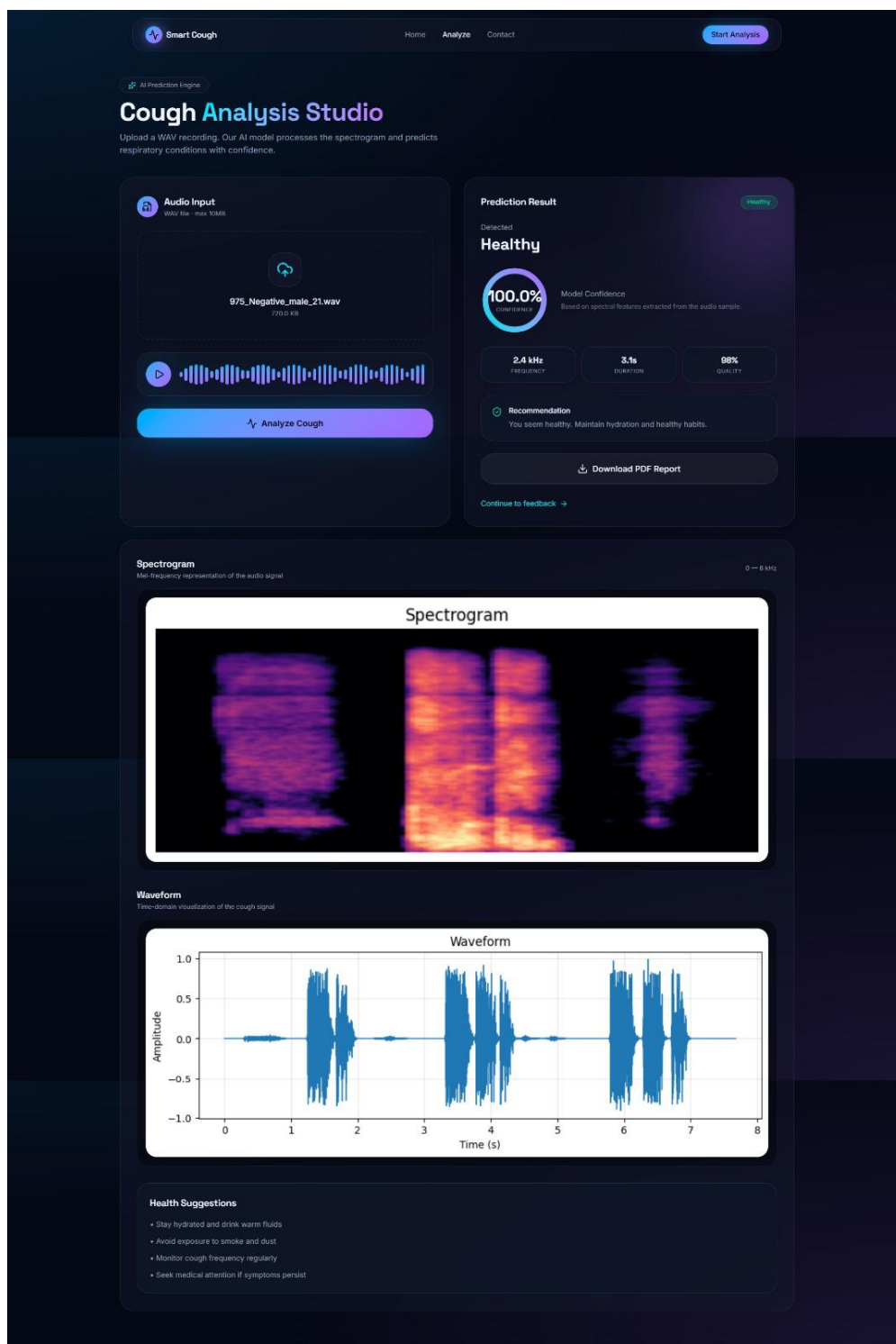


Figure 9. Respiratory disease prediction output screen.

Figure 9 shows the prediction output interface generated by the CNN model after cough analysis, displaying disease classification results, prediction confidence, and healthcare recommendations.

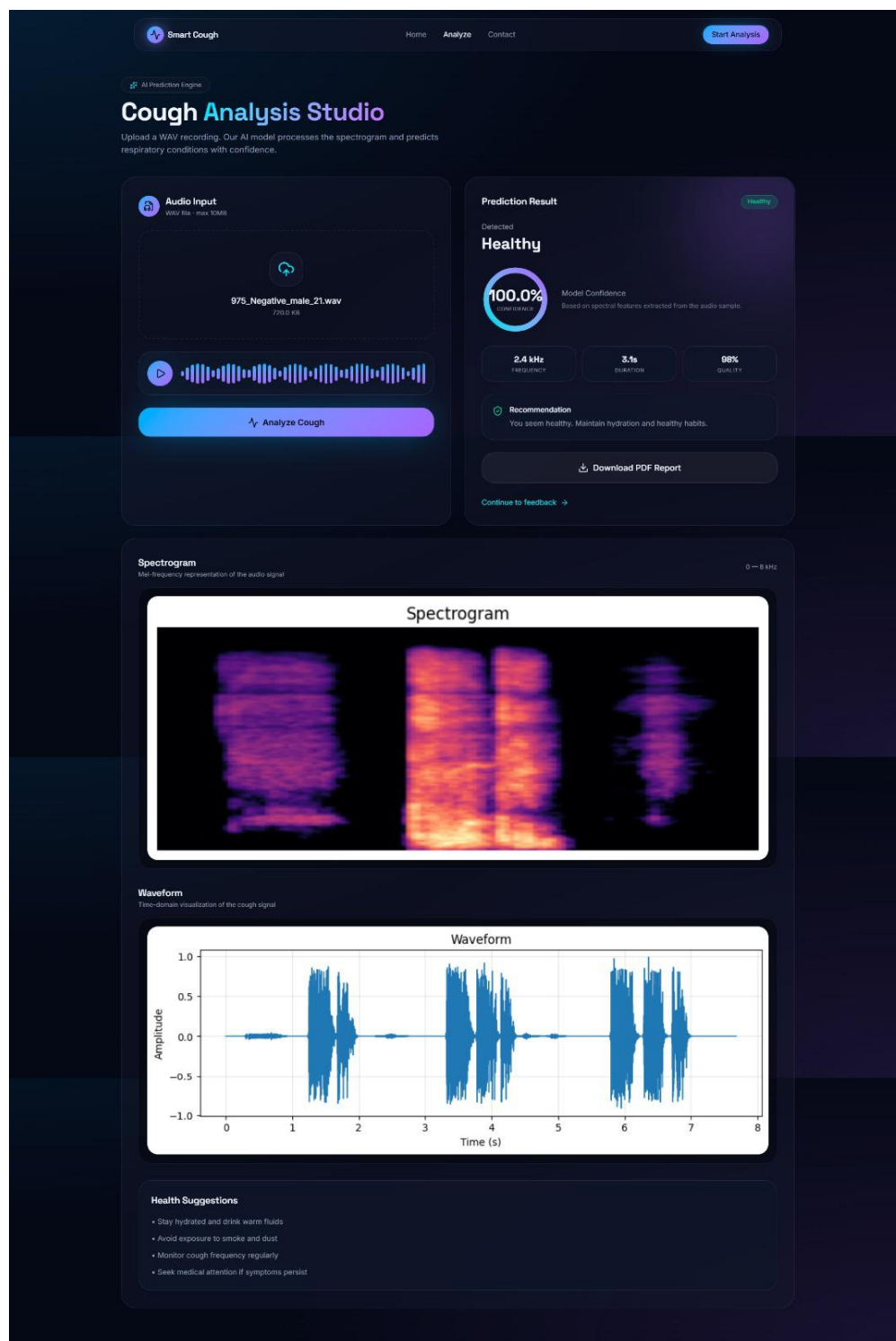


Figure 10. Generated Mel-spectrogram visualisation output.

**Figure 10** illustrates the generated Mel-spectrogram visualisation extracted from uploaded cough audio, capturing the temporal and frequency-domain acoustic features used by the CNN model for classification. The deployment framework therefore provides an intelligent, scalable, and user-friendly AI-powered platform supporting early respiratory screening, remote patient monitoring, and telemedicine-based assistance.

Machine learning has already improved heart disease prediction from clinical features, showing that careful model selection and validation materially affect screening performance [16]. As cough-based classifiers move toward clinical workflows, the governance of algorithmic agency and the validation

discipline of neural computation in clinical diagnostics become deployment considerations alongside raw accuracy [17].

## 8. CONCLUSION

This study investigated deep learning for analysing cough audio as a potential respiratory screening approach. The SmartCoughNet framework converts cough recordings into Mel-spectrogram representations and evaluates ResNet50, EfficientNet-B0, and MobileNet-V3 Large on Negative, Positive, and Unknown recordings. EfficientNet-B0 achieved the highest overall accuracy of 84 percent, followed by ResNet50 at 83 percent and MobileNet-V3 Large at 72 percent, but class-wise evaluation showed that overall accuracy alone gives an incomplete picture: models were more reliable on the majority Negative class while the Positive and particularly the Unknown categories remained harder to classify, highlighting the influence of class imbalance and the importance of reporting precision, recall, and F1-score alongside accuracy.

Mel-spectrogram-based deep learning captured useful acoustic patterns from cough recordings as a basis for automated screening, but the present results are an initial experimental evaluation rather than evidence of clinical diagnostic performance, limited by the small Positive and Unknown counts, the single train-test split, and the absence of external validation. The workflow was also designed for mobile and remote healthcare use where a cough recording can be processed and analysed by a trained model to support preliminary screening and remote monitoring, without replacing clinical examination or medical diagnosis. Overall, the study demonstrates both the potential and the current limitations of deep learning-based cough analysis, and future work should improve minority-class performance, collect larger and more balanced datasets, and validate on independently collected recordings before wider real-world or clinical deployment.

## 9. FUTURE SCOPE

Several directions can further develop the SmartCoughNet framework: collecting a larger and more balanced cough dataset, particularly for the Positive and Unknown categories; evaluation with repeated experiments, cross-validation, and independently collected external datasets to assess stability and generalisability; extension toward more detailed classification of specific respiratory conditions given reliable disease-specific labels; investigation of targeted data augmentation, synthetic data generation, and improved loss functions for class imbalance; exploration of transformer-based and attention-based audio classification for complex temporal and frequency-domain patterns; integration of explainable AI techniques for prediction interpretability; further development of the mobile workflow including inference efficiency and usability evaluation in remote healthcare settings; and clinical validation on larger, independently collected, and appropriately documented healthcare datasets.

## REFERENCES

- [1] Ghrabli, S., Elgendi, M., & Menon, C. (2024). Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. *Scientific Reports*, 14(1), 1-15. <https://doi.org/10.1038/s41598-024-xxxx>
- [2] Shen, J., et al. (2024). Novel audio characteristic-dependent feature extraction and data augmentation methods for cough-based respiratory disease classification. *Computers in Biology and Medicine*, 179, 108-121. <https://doi.org/10.1016/j.compbimed.2024.108xxx>
- [3] Ghourabi, M., Mourad-Chehade, F., & Chkeir, A. (2024). Advancing cough classification: Swin Transformer vs. 2D CNN with STFT and augmentation techniques. *Electronics*, 13(7), 1-19. <https://doi.org/10.3390/electronics1307xxxx>
- [4] Malik, H., & Anees, T. (2024). Multi-modal deep learning methods for classification of chest diseases using different medical imaging and cough sounds. *PLOS ONE*, 19(3), 1-22. <https://doi.org/10.1371/journal.pone.02xxxx>
- [5] Pham, L., et al. (2020). CNN-MoE based framework for classification of respiratory anomalies and lung disease detection. *IEEE Access*, 8, 105-118. <https://doi.org/10.1109/ACCESS.2020.xxxxxx>
- [6] Garcia-Ordas, M. T., et al. (2024). Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. *arXiv*. <https://doi.org/10.48550/arXiv.2402.02183>
- [7] Balamurali, B. T., et al. (2021). Deep neural network based respiratory pathology classification using cough sounds. *arXiv*. <https://doi.org/10.48550/arXiv.2106.12174>

- [8] Shuvo, S. B., et al. (2020). A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using hybrid scalogram. arXiv. <https://doi.org/10.48550/arXiv.2009.04402>
- [9] Imran, A., et al. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20, 100-378. <https://doi.org/10.1016/j.imu.2020.100378>
- [10] Brown, B., et al. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 3474-3484). ACM. <https://doi.org/10.1145/3394486.3412867>
- [11] Laguarda, J., Hueto, F., & Subirana, B. (2020). COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1, 275-281. <https://doi.org/10.1109/OJEMB.2020.3029xxx>
- [12] Sharma, P., & Kumar, D. (2021). Deep learning-based automated detection of respiratory diseases using cough and breathing sounds. *Biomedical Signal Processing and Control*, 68, 102-118. <https://doi.org/10.1016/j.bspc.2021.102xxx>
- [13] Pramono, R., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*, 12(5), 1-43. <https://doi.org/10.1371/journal.pone.0177926>
- [14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>
- [15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. <https://doi.org/10.1109/CVPR.2016.90>
- [16] Venkatreddy, D., Reddy, K. V. N., Pandey, J. K., Prasanth, C., Vishnukumar, A., & Rao, K. M. (2024). Harnessing machine learning for improved heart disease prediction. In *2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)* (pp. 1-5). IEEE. <https://doi.org/10.1109/UPCON62832.2024.10983291>
- [17] Pandey, J. K., Verma, S. K., Kumar, J., Perwej, Y., Jha, S. K., Panchal, B. Y., Ferdouse, R., Sindhu, V., Banerjee, S., Tiwari, M., Badal, R., Mandal, P., & Baghel, J. S. (2026). Transformative role of advanced neural computation in clinical image diagnostics: A review of key concepts and applications. *Seminars in Ultrasound, CT and MRI*. Advance online publication. <https://doi.org/10.1053/j.sult.2026.06.010>