

Respiratory Disease Classification based on Audio Sounds



Presented by

Ishrat Bombaywala, Gracy Tanna, Pahul Singh

Index

- 1. Problem Statement**
- 2. Literature Review**
- 3. Dataset Collection & Preprocessing**
- 4. ML Methodology**
- 5. Performance & Deployability**
- 6. Future Works**

The Problem.

Respiratory diseases represent a major global health challenge, causing millions of deaths and affecting hundreds of millions of people worldwide.

Asthma

01

260M+

lives affected

Causing significant respiratory distress and widespread hospitalizations globally.

COPD

02

3M+

annual deaths

A leading cause of death globally according to the World Health Organization.

COVID-19

7M+

total deaths

Demonstrated how rapidly respiratory diseases can overwhelm healthcare systems.

Symptoms of cough, wheezing, and breathlessness overlap across various diseases, making accurate diagnosis difficult.

Early and accurate differentiation between respiratory conditions is essential for timely treatment and disease management.

Why is this important?

Delhi Records Over 9,000 Deaths Due To Respiratory Diseases In 2024

Circulatory diseases were found to be the top most cause of death in 2024, followed by infectious and parasitic diseases. The overall death rate in the national capital also showed a rise.

Press Trust of India | [India News](#) | Jan 15, 2026 22:03 pm IST ⓘ

Source: [Delhi Government](#)

STATISTICS ABOUT CHRONIC RESPIRATORY DISEASES

3m

people die of COPD each year

6%

of deaths worldwide is due to COPD

300m

people suffer with asthma worldwide

250k

people die from asthma each year

Chronic respiratory diseases: more than 80 million affected and many more undiagnosed, warns new WHO and European Respiratory Society report

12 June 2025 | Media release | Reading time: 3 min (855 words)

Source: [WHO](#)



Problem Statement.

Classification of Respiratory Diseases
Through Audio-Based Physiological
Signal Analysis.

LITERATURE REVIEW



Article

Deep Neural Network-Based Respiratory Pathology Classification Using Cough Sounds

B T Balamurali ^{1,*}, Hwan Ing Hee ^{2,3}, Saumitra Kapoor ¹, Oon Hoe Teoh ⁴, Sung Shin Teng ⁵, Khai Pin Lee ⁵, Dorien Herremans ⁶ and Jer Ming Chen ¹

- ¹ Science, Mathematics and Technology, Singapore University of Technology and Design, Singapore 487372, Singapore; saumitrakapoor@gmail.com (S.K.); jerming_chen@sutd.edu.sg (J.M.C.)
² Department of Paediatric Anaesthesia, KK Women's and Children's Hospital, Singapore 229899, Singapore; hee.hwan.ing@singhealth.com.sg
³ Anaesthesiology and Perioperative Sciences, Duke-NUS Medical School, 8 College Road

scientific reports

Check for updates

OPEN Screening COVID-19 by Swaasa AI platform using cough sounds: a cross-sectional study

Padmalatha Pentakota¹, Gowrisree Rudraraju^{2✉}, Narayana Rao Sripada², Baswaraj Mamidgi², Charishma Gottipulla², Charan Jalukuru², Shubha Deepti Palreddy², Nikhil Kumar Reddy Bhoge², Priyanka Firmal², Venkat Yechuri², Manmohan Jain², Venkata Sudhakar Peddireddi¹, Devi Madhavi Bhimarasetty¹, S. Sreenivas¹, Kesava Lakshmi Prasad K¹, Niranjan Joshi³, Shibu Vijayan⁴, Sanchit Turaga⁵ &

2019 19th International Conference on Control, Automation and Systems (ICCAS 2019)
Oct. 15~18, 2019; ICC Jeju, Korea

Automatic Classification of Large-Scale Respiratory Sound Dataset Based on Convolutional Neural Network

Koki Minami¹, Huimin Lu¹, Hyoungseop Kim¹
Shingo Mabu², Yasushi Hirano², Shoji Kido³

¹ Kyushu Institute of Technology, 1-1, Sensui, Tobata, Kitakyushu, Fukuoka 804-8550, Japan

² Yamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611

³ Osaka University, 2-2, Suita, Osaka 565-0871

Abstract: Auscultation of respiratory sounds is very important for discovering the respiratory disease. However, there

Disease Symptoms & Their Overlap

Symptom / Characteristic	Asthma	COPD	COVID-19
Cough type	Dry or intermittent	Chronic productive cough	Dry, persistent cough
Onset pattern	Sudden attacks	Slowly progressive	Acute infection
Fever	Rare	Rare	Very common
Fatigue / systemic symptoms	Usually absent	Mild	Common
Age group	Often younger individuals	Usually older adults/smokers	All age groups
Wheezing	Very common	Common	Rare
Upper respiratory symptoms	Rare	Rare	Common (sore throat, loss of smell)

Asthma, COPD, COVID-19

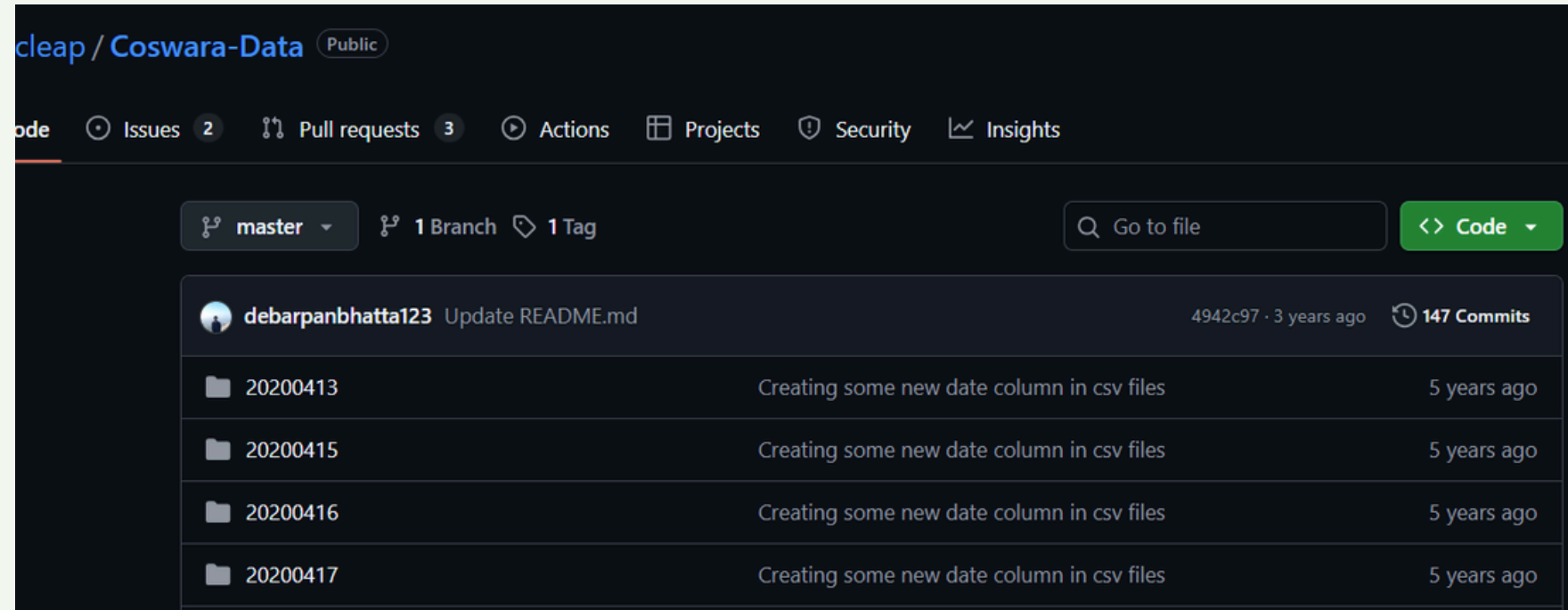
Existing solutions	Approach & Performance
<u>Swaasa AI platform</u>	<ul style="list-style-type: none"> • Model: CNN (MFCC spectrograms) + Feedforward ANN • Sound Type: Clinical cough recordings • Classification & Accuracy: Binary classification (COVID positive vs negative), ~75.54% accuracy
<u>Convolutional Neural Network</u>	<ul style="list-style-type: none"> • Model: CNN + VGG-16 transfer learning • Sound Type: Lung/respiratory sounds (spectrograms & scalograms) • Classification & Accuracy: Multi-class sound-type classification, ~54% accuracy
<u>Deep Neural Network</u>	<ul style="list-style-type: none"> • Model: BiLSTM-based Deep Neural Network • Sound Type: Smartphone-recorded cough sounds • Classification & Accuracy: Binary + multi-class disease classification, ~84% accuracy for binary pathological classification

GAPS

- **Binary or limited** classification focus
- **Limited multi-disease** coverage
- Reliance mainly on **cough/lung sounds only**
- **Insufficient feature extraction** (mostly MFCC/CNN-based features)
- **Lack of multimodal** analysis (e.g., vowel + cough + metadata)
- **No patient metadata**–driven classification framework
- Reduced accuracy/generalization in multi-class disease prediction

DATASET COLLECTION AND FEATURE PREPROCESSING

Dataset #1

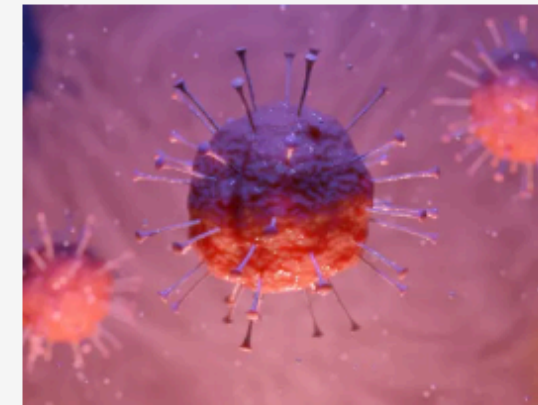


- **2,746 datapoints (65 hours)**
- Metadata: 36 features
- Classes: **COVID-19 (681), Healthy (2065)**
- Data taken: **two .wav files** (cough & o-sound) + **one .csv metadata file** for each datapoint

Source: Coswara Dataset (IISC Bangalore)

Dataset #2

The Corp Dataset



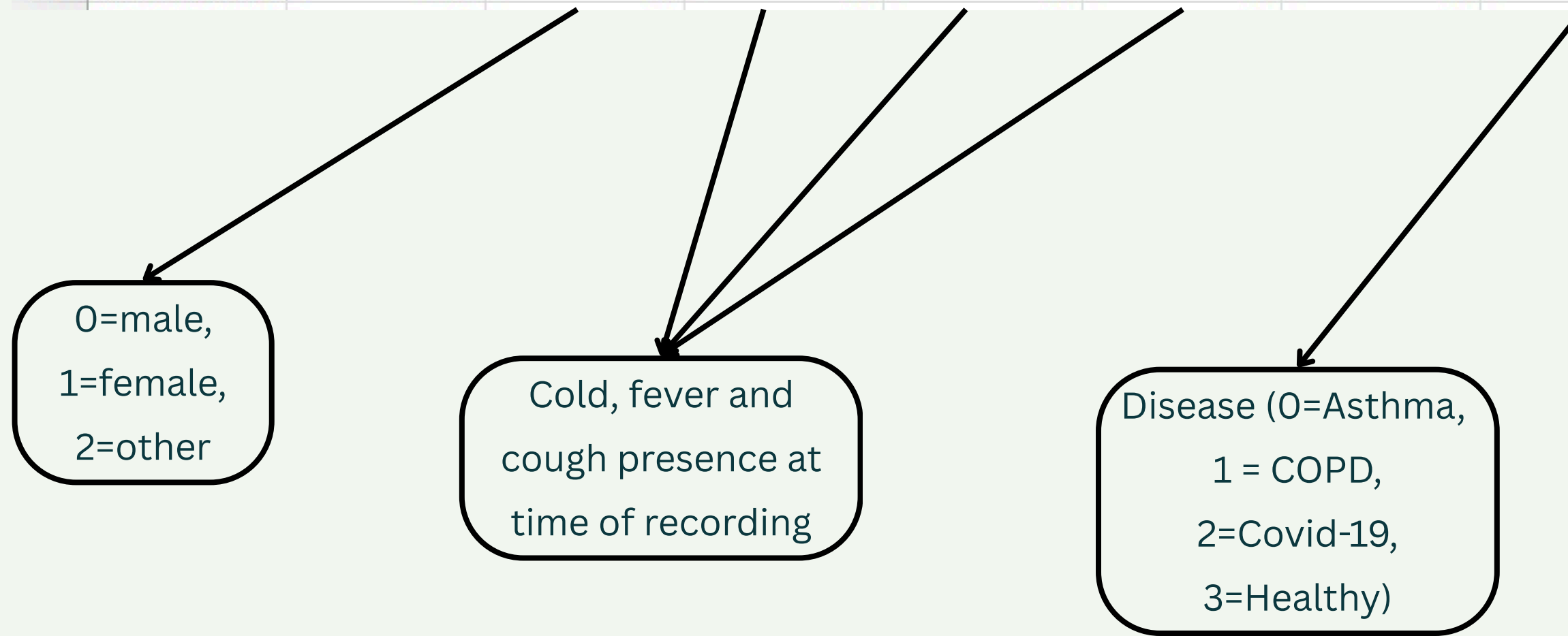
Citation Author(s): [Weihao Wang](#)
Submitted by: [Weihao Wang](#)
Last updated: [Mon, 05/17/2021 - 11:34](#)
DOI: [10.21227/q1cd-9n64](#)
Links: [THE CORP DATASET](#)

- **884 datapoints (13 hours)**
- Metadata: 11 features
- Classes: **COPD (168), Asthma (238), Healthy (140)**
- Data taken: **two .wav files** (cough & o-sound) + **one .csv metadata file** for each datapoint

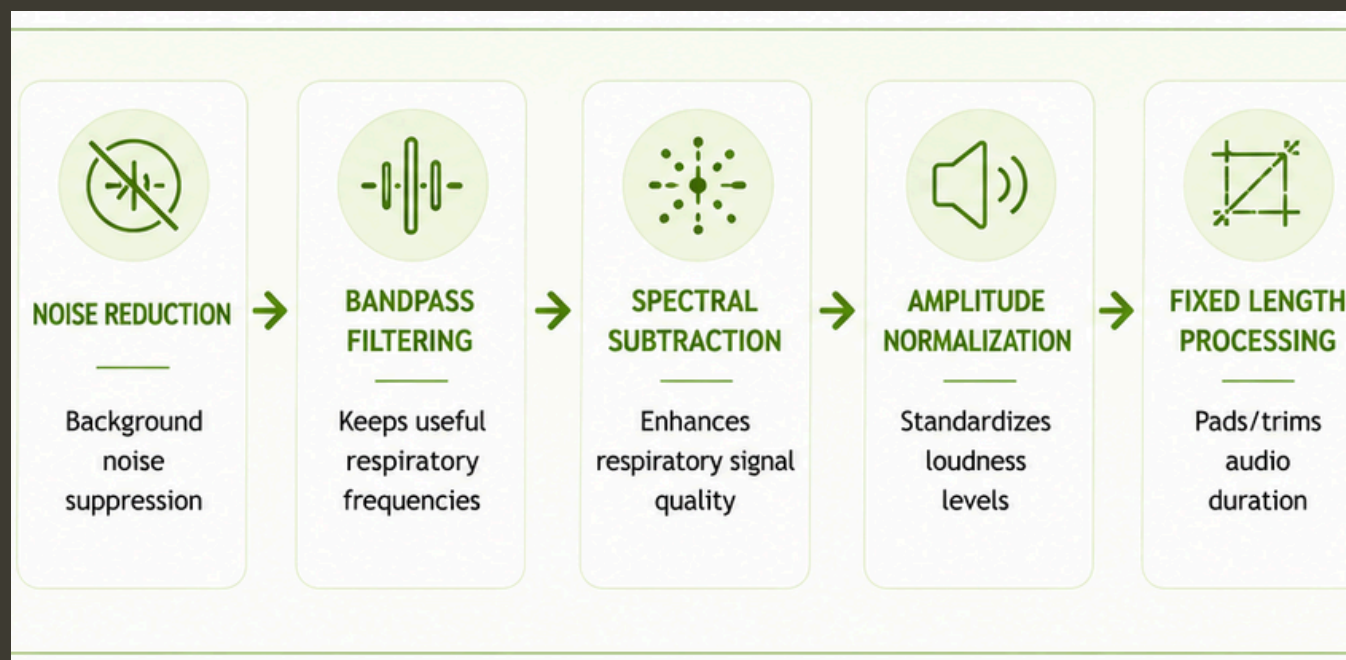
Source: Corp Dataset

Total instances: 2.7k
Total Features: 6

	A	B	C	D	E	F	G	H
1	id	age	gender	cough	fever	cold	smoker	disease
2	91fd57845	27	1	0	0	0	0	1
3	5414ff66ff	24	1	1	1	1	0	1
4	5093190b7	34	1	0	0	1	0	1
5	74603007f	40	0	0	0	0	0	1
6	80a6dd437	22	0	1	0	0	0	1
7	c914ff276	42	0	0	1	0	0	1
8	e4a2c4b88	42	1	0	0	0	0	1
9	1e0047a88	48	1	0	0	0	0	1



Preprocessing for vowel files



Preprocessing for cough audio

```

Total number of instances: 2783

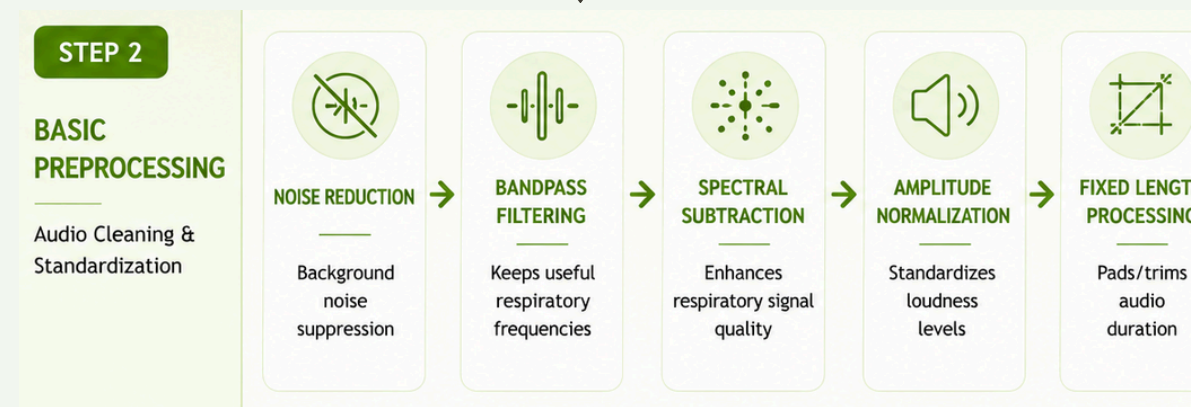
==== Disease ====
disease
healthy    1716
covid      545
asthma     354
copd       168
Name: count, dtype: int64
  
```



```

Rows      : 10940
Columns   : 73

Class distribution (disease_y):
disease_y
healthy   4040
asthma    2500
covid     2200
copd      2200
  
```

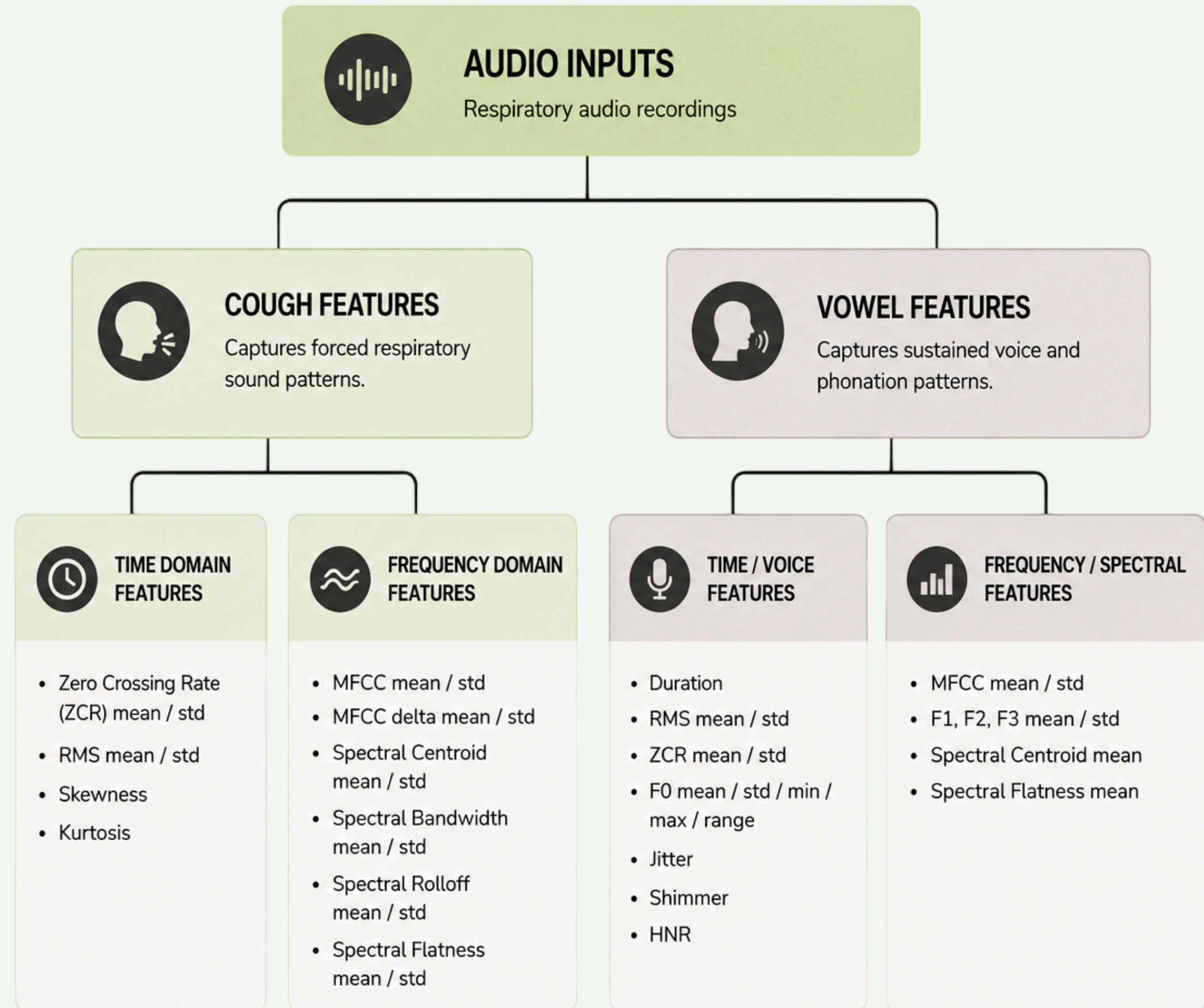


Feature Extraction

Audio recordings are transformed into meaningful numerical features that capture important temporal, spectral, and voice-based patterns in respiratory sounds.

These extracted cough and vowel biomarkers are used to train machine learning models for respiratory disease classification.

From Cough + 'O' Vowel Audio Files





Final CSV

id	cont_age	cat_gend	bin_smol	bin_coug	bin_fever	bin_cold	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc	cont_mfc
00ad476fb	0.974542	0	1	1	0	0	0.826649	-0.82187	-1.03687	-0.08263	-0.24516	0.509427	-1.71256	1.359902	1.205856	-0.59088	-0.27268	-1.00473	1.175966	-0.33
00ad476fb	0.974542	0	1	1	0	0	0.04519	-1.03483	-0.84983	-0.24239	-0.89607	0.625826	-3.17701	1.314504	1.851955	-0.69849	-0.08397	-0.44797	1.678209	-1.08
00ad476fb	0.974542	0	1	1	0	0	-1.67785	-1.11831	1.77779	2.071662	0.912859	-0.46815	-3.33439	1.398159	0.629826	-1.88995	1.197774	-0.82858	0.545545	-0.29
00ad476fb	0.974542	0	1	1	0	0	0.421969	-1.02733	-0.48231	-0.29476	-0.65408	0.472153	-2.91289	1.355185	1.883519	-0.25766	-0.21382	-0.6939	1.264993	-0.53
00ad476fb	0.974542	0	1	1	0	0	-0.68438	-1.16401	-0.90559	-0.96895	-0.20507	1.561799	-2.93804	0.770184	2.21927	-1.44446	-1.00562	1.653223	1.921355	0.140
00ad476fb	0.974542	0	1	1	0	0	-0.08974	-0.79218	-0.12023	0.315708	-0.1232	0.884445	-3.06461	1.032159	1.606327	-0.65728	-0.40115	-0.55382	1.059406	-0.21
00ad476fb	0.974542	0	1	1	0	0	0.256655	-0.98113	-0.86692	-0.42462	-0.68201	0.581362	-3.38361	1.039208	1.859728	-0.51267	-0.06141	-0.52011	1.260161	-0.9
00ad476fb	0.974542	0	1	1	0	0	-1.71094	-1.05384	0.058598	0.651973	-0.48929	1.335146	-2.45877	1.402267	3.137089	-0.89129	-1.6522	-2.14454	0.271704	0.972
00ad476fb	0.974542	0	1	1	0	0	-0.22046	-1.14304	-0.05931	0.125244	-0.25597	0.731105	-2.46854	0.85712	2.287704	0.17725	0.288117	-0.95429	0.485859	-0.59
00ad476fb	0.974542	0	1	1	0	0	0.434868	-0.78828	-0.16858	-0.15805	-0.30738	1.471145	-3.07683	0.865153	1.322094	-0.59276	-0.14933	-0.92309	0.556286	-0.16
00ad476fb	0.974542	0	1	1	0	0	0.890933	-0.66073	-0.78109	-0.39055	-0.17096	0.529749	-2.81251	0.838414	0.97335	-0.91772	-0.32748	-1.63401	0.969137	-1.5
00ad476fb	0.974542	0	1	1	0	0	-1.05143	-1.12567	-1.17847	0.001416	-0.98215	0.495207	-1.30833	1.672816	1.952918	-1.19541	-0.83331	-0.3112	1.558755	1.045
00xKcQMr	-0.86731	0	0	0	0	0	0.544569	0.780967	-1.10018	0.645666	0.04252	1.09867	0.818359	-0.2149	1.174762	0.390527	0.262718	0.00922	0.416967	0.960
00xKcQMr	-0.86731	0	0	0	0	0	0.726036	0.38775	-0.58024	-0.03805	-0.04908	0.325061	0.159706	-0.28143	0.227721	-0.2546	-0.80591	-0.64447	0.624989	0.701
00xKcQMr	-0.86731	0	0	0	0	0	1.183825	0.112327	-0.86497	0.839152	0.01864	0.742356	0.72163	-0.39812	0.834346	0.36016	-0.33669	0.753016	0.221768	0.221
00xKcQMr	-0.86731	0	0	0	0	0	0.4513	-0.20613	-1.47567	0.017497	-0.68321	0.4328	0.572571	-0.56067	0.396528	-0.06857	-0.29723	0.089531	-0.1137	0.202
00xKcQMr	-0.86731	0	0	0	0	0	1.126472	0.377119	-0.98957	0.650173	-0.10386	0.421845	0.488773	-0.15203	0.460907	0.426513	-0.21614	0.136293	0.354538	0.883
00xKcQMr	-0.86731	0	0	0	0	0	0.292964	0.503344	-0.96604	-0.29451	-0.16272	0.689309	0.348821	-0.63723	-0.49338	-0.50232	-0.6572	-0.95533	-0.192	0.975
01e409528	-1.06465	0	1	0	0	1	0.025741	-1.16254	-0.64587	1.053161	-1.02624	1.155166	-0.28384	-2.06161	0.783494	-0.2303	-0.5425	-0.13655	0.793601	0.600

Total instances: 11.3k

Total features: Patient ID + 120 (6 metadata, 66 cough, 47 vowel, 1 target)

Train-Validate-Test Split

Grouped & Stratified Data Splitting

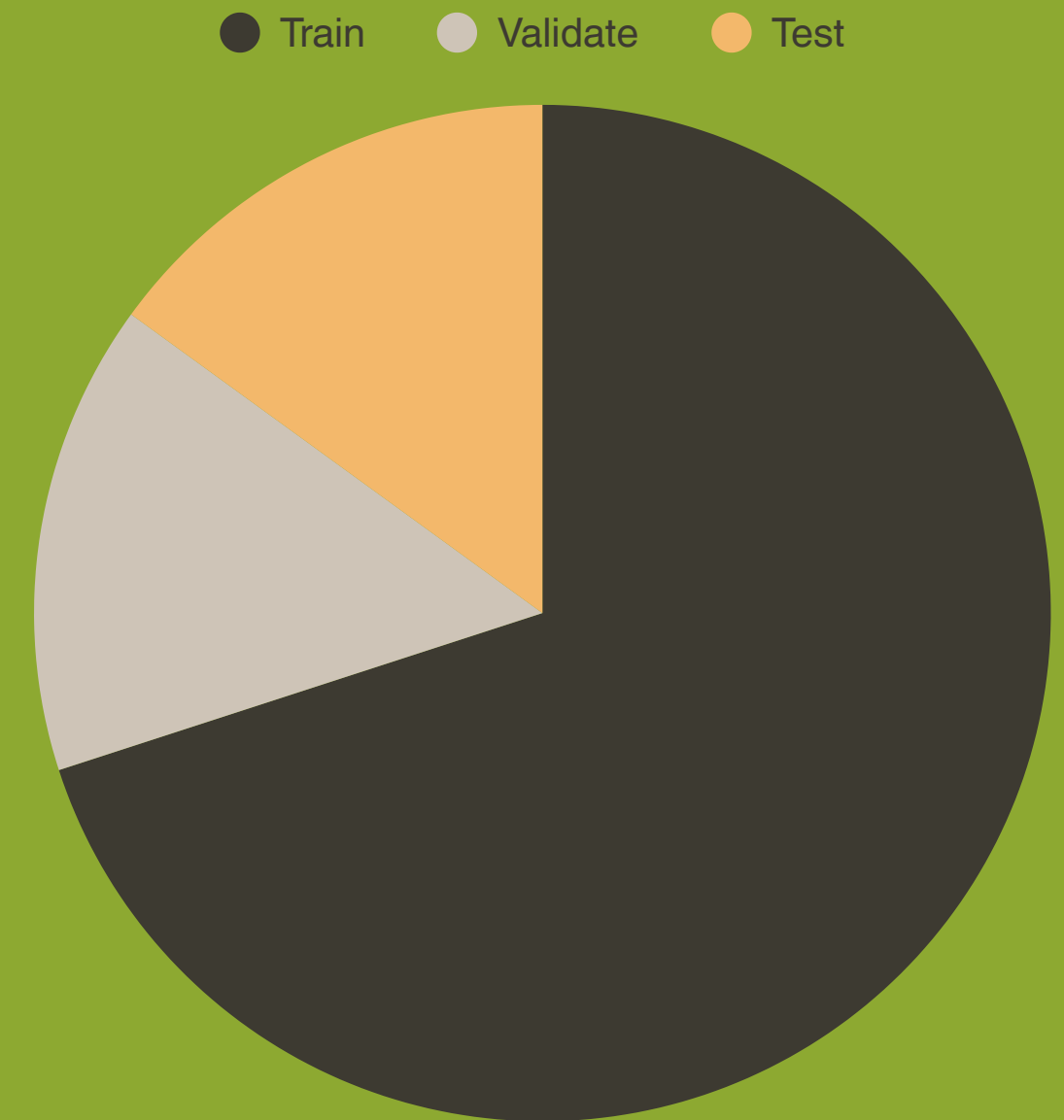
Why grouped splitting?

All recordings from the same patient were kept within a single dataset split to:

- Prevent data leakage
- Avoid the model learning patient-specific patterns

Why stratified splitting?

Disease proportions were preserved to maintain balanced class representation.



TRAIN
70%
7993 instances
1657 patients

VALIDATE
15%
1772 instances
355 patients

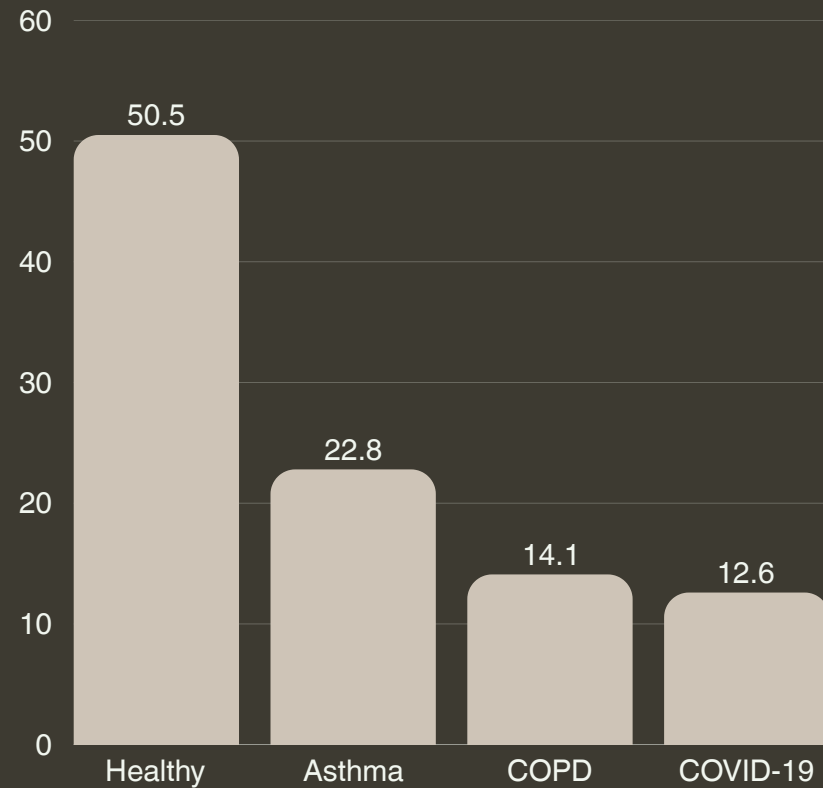
TEST
15%
1626 instances
356 patients

Patient overlap across splits: 0

Disease classes: Healthy, Asthma, COPD, COVID-19

Why balancing was needed:

The dataset showed unequal representation across disease classes, with healthy samples dominating the training data.



To improve minority disease classification

Class Imbalance Handling

1. Random Over Sampling (ROS)

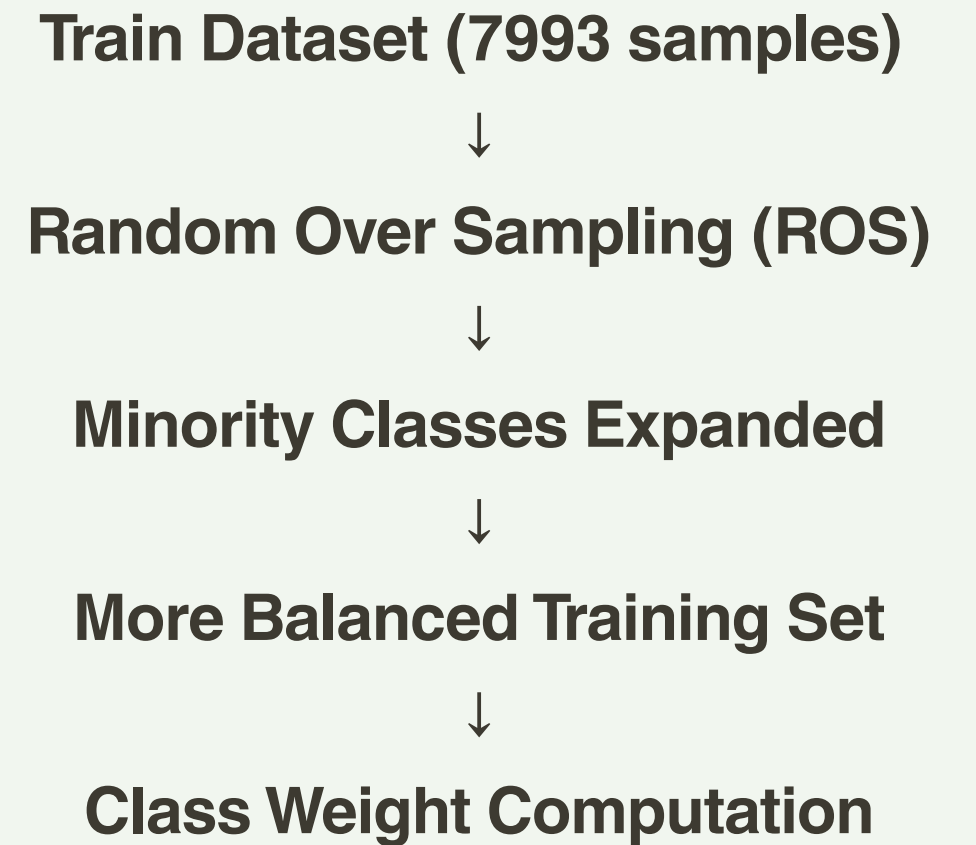
- Applied only on training data
- Minority classes duplicated randomly
- Validation and test sets left untouched

Class	Target Samples
Healthy	8000
Asthma	4000
COPD	2500
COVID	2500

2. Additional Balancing

Class weights were computed from the resampled training set and supplied to ML models during training.

Pipeline



METHODOLOGY

Approach

#1 Cough-only Model

Input:

Extracted cough features only

Features extracted:

we have around 11.3k instance with 74 features

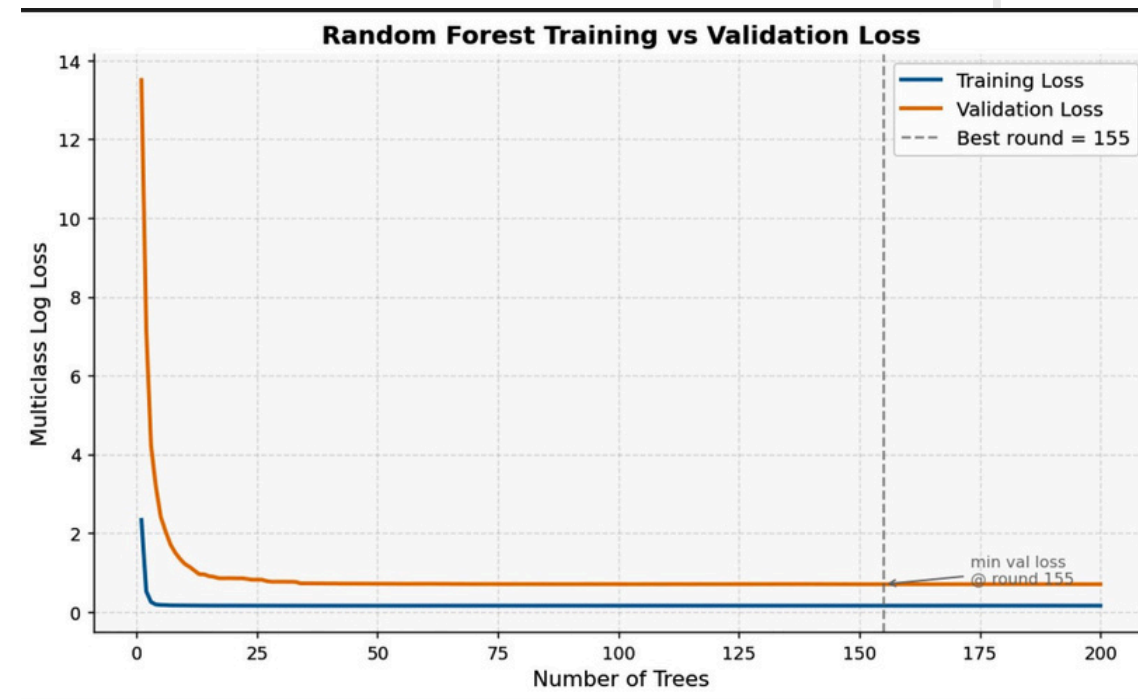
Dataset size:

~13 k instances (post ROS), 74 features

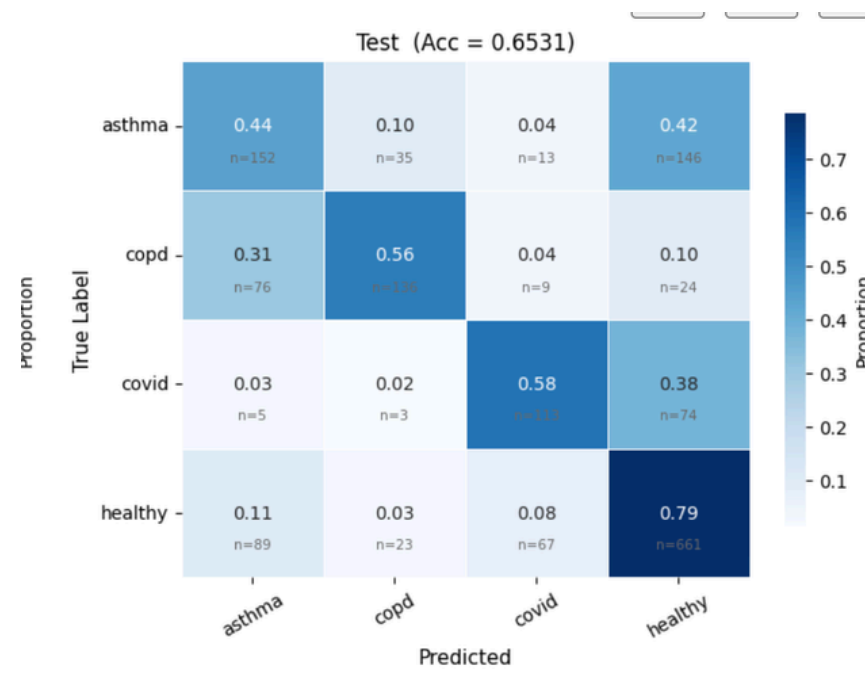
Models tested:

Random Forest, SVM, Logistic Regression, XGBoost, HistGrad Boosting

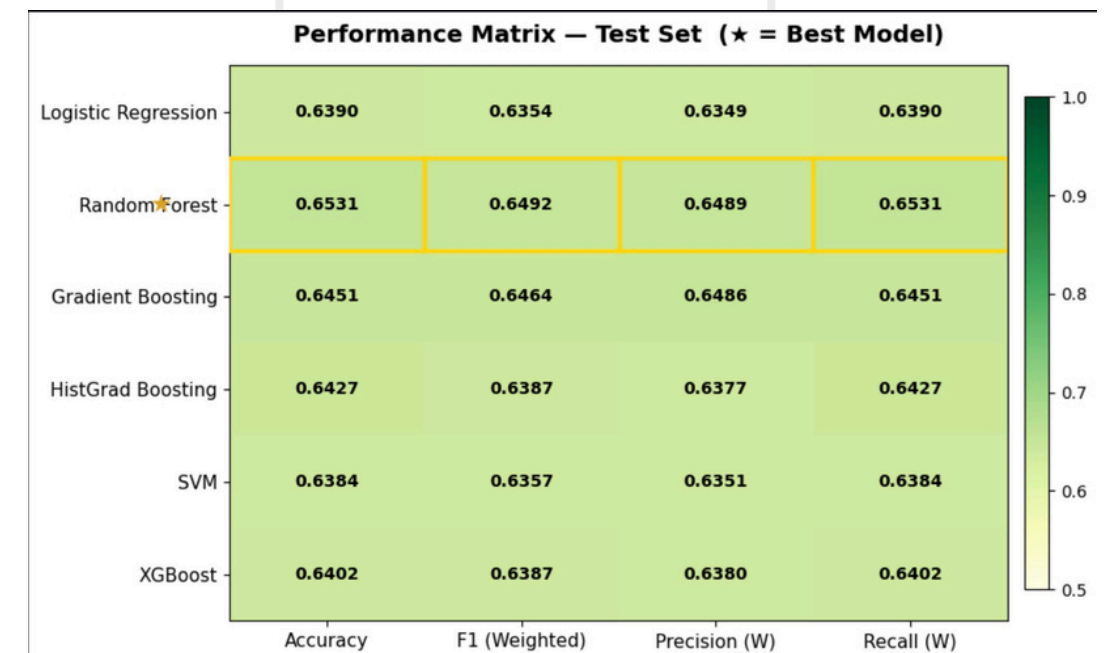
Training vs. Validation Loss



- Random Forest achieved the best overall performance among all tested ML models.
- **~65.3% accuracy**
- Training and validation loss curves show stable learning with minimal overfitting.
- Cough-only audio features were sufficient for meaningful respiratory disease classification.
- The model performed best on healthy and COVID respiratory samples.
- Feature extraction and preprocessing improved model consistency and generalization.



Confusion Matrix



Performance Metrics

Effective for general respiratory health-status detection, but limited in capturing fine-grained disease-specific respiratory variations.

Approach

#2 *Vowel-only Model*

Input:

Extracted vowel features only

Features extracted:

47 vowel-based features

Dataset size:

~2.7k instances, ~3.8k after ROS

55 features

Models tested:

Random Forest, SVM-RBF, Logistic Regression, XGBoost

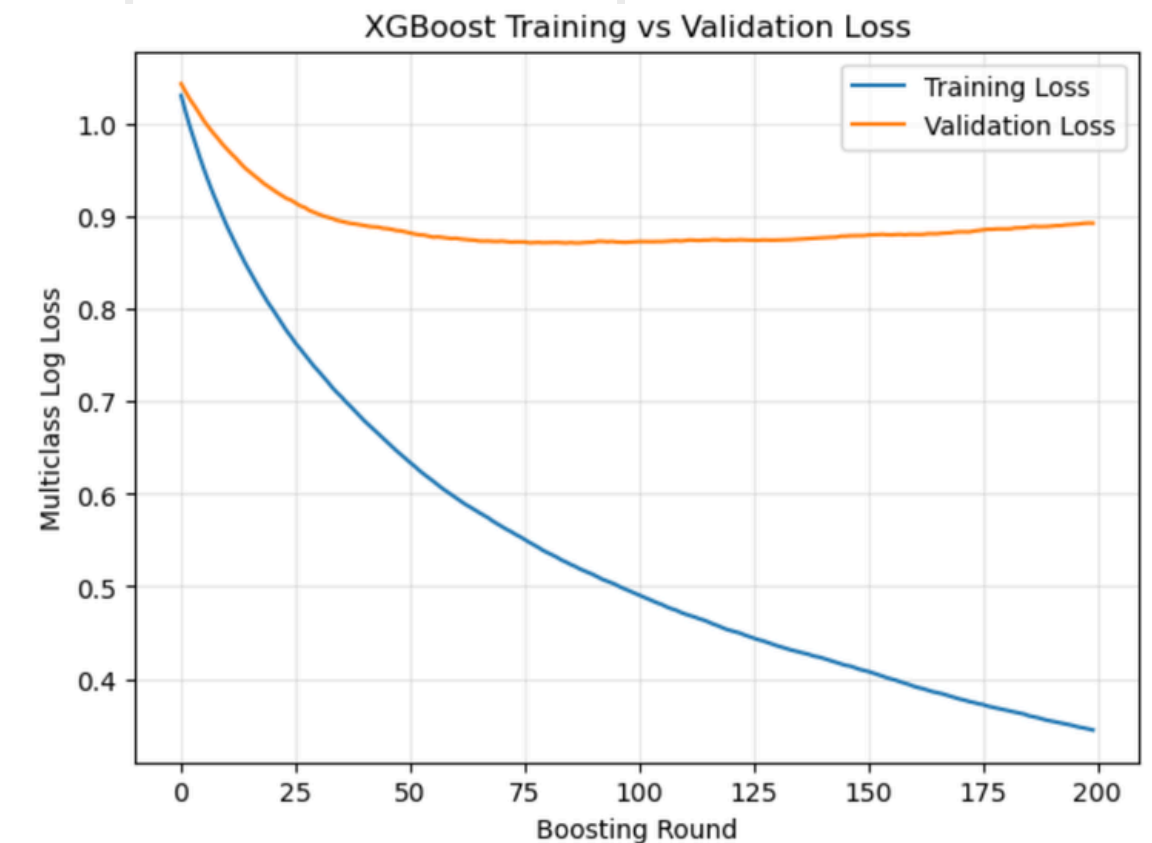
- XGBoost achieved ~**68.6%** accuracy using vowel-only features
- Healthy class showed **strongest** classification performance
- Asthma, COPD, and COVID classes showed significant **overlap**
- Validation loss plateauing indicates **overfitting**

	asthma	copd	covid	healthy
asthma	12	9	2	30
copd	8	9	0	8
covid	2	0	26	53
healthy	5	1	11	235

Confusion Matrix

```
XGBoost Test Accuracy: 0.6861313868613139
      precision    recall  f1-score
asthma      0.44      0.23      0.30
copd        0.47      0.36      0.41
covid       0.67      0.32      0.43
healthy     0.72      0.93      0.81
```

Performance Metrics



Training vs. Validation Loss

Useful for health-status detection, limited for disease-specific classification 20

Approach #3 *Cough+Vowel Model for 2 classes*

Input:

Extracted cough & vowel features

Features extracted:

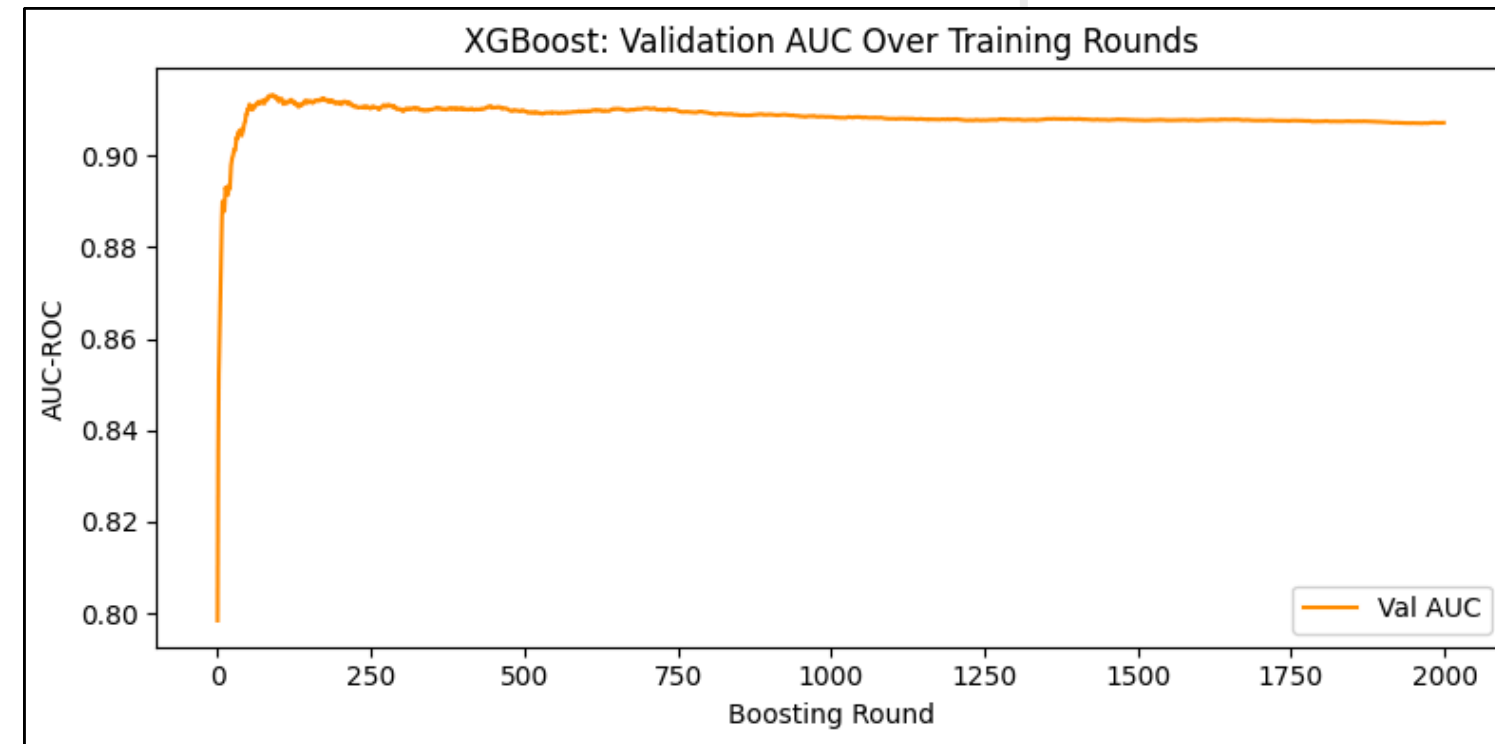
121 features

Dataset size:

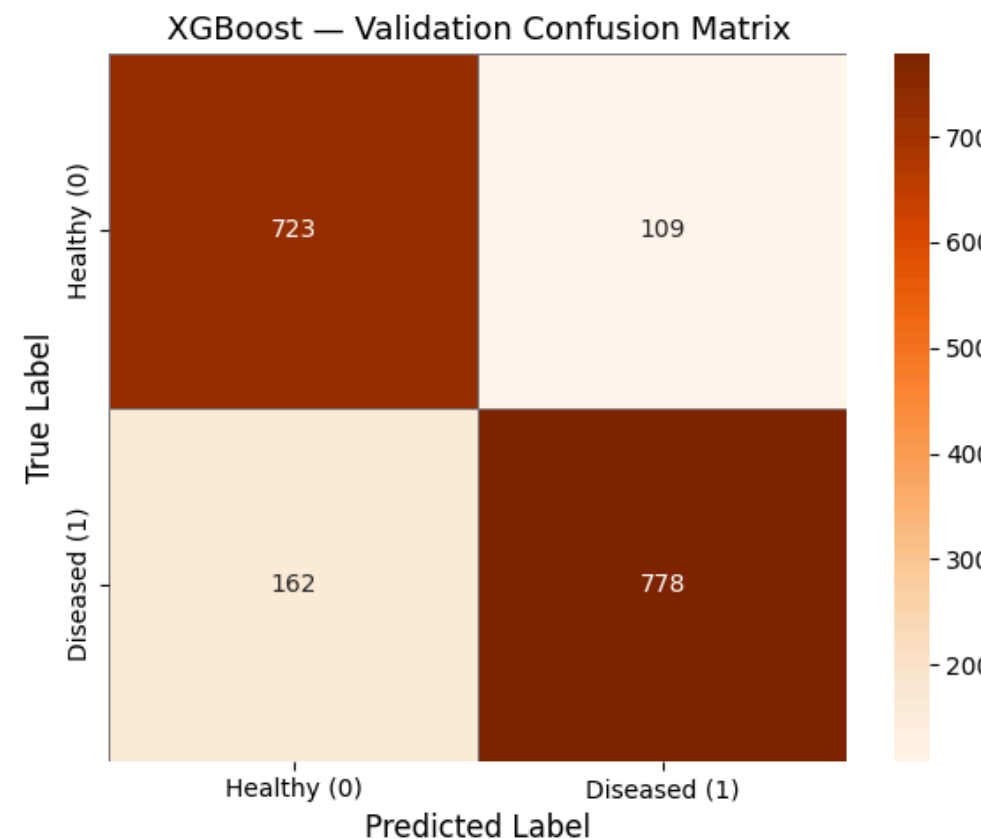
~11.3k instance

Models tested:

Random Forest, SVM-RBF, Logistic Regression, XGBoost, Gradient Boost, HistGradBoost



- XGBoost achieved the best overall performance with the least FNR for **healthy vs diseased**.
- **~75.5% binary accuracy**
- AUC loss curves show stable learning with minimal overfitting.
- Cough + vowel features are excellent for differentiating between healthy and non-healthy classes.
- Thus, we proceed with cough+vowel features used with XGBoost.



```
=====
FINAL SUMMARY - XGBoost Layer 1
=====
Val Accuracy : 0.8471
Val AUC-ROC  : 0.9071
Val F1 Macro : 0.8469
Val Recall M : 0.8483
Val FN Rate  : 0.1723
---
Test Accuracy : 0.7558
Test AUC-ROC  : 0.8392
Test F1 Macro : 0.7557
Test Recall M : 0.7560
Test FN Rate  : 0.2405
```

Approach

#4 *Cough+Vowel Model for all 4 classes*

Input:

Extracted cough & vowel features

Features extracted:

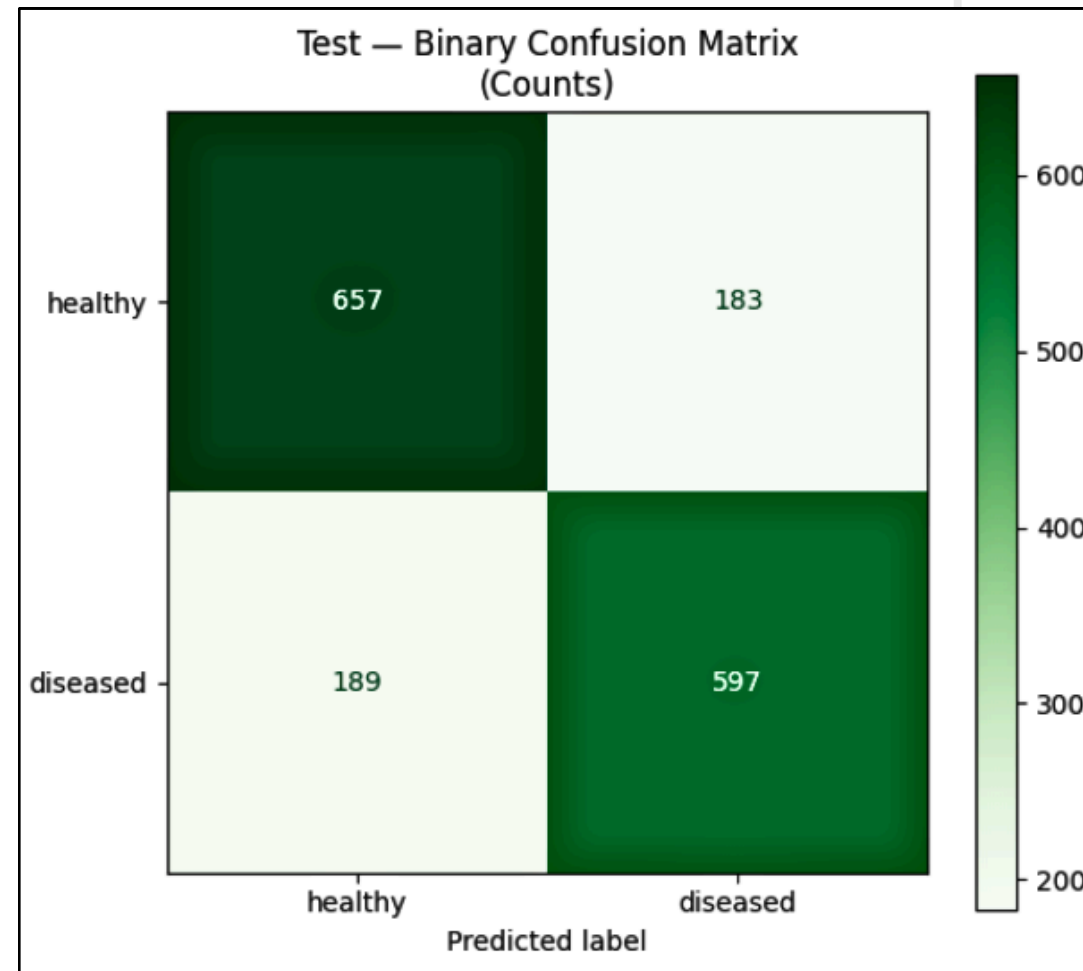
121 features

Dataset size:

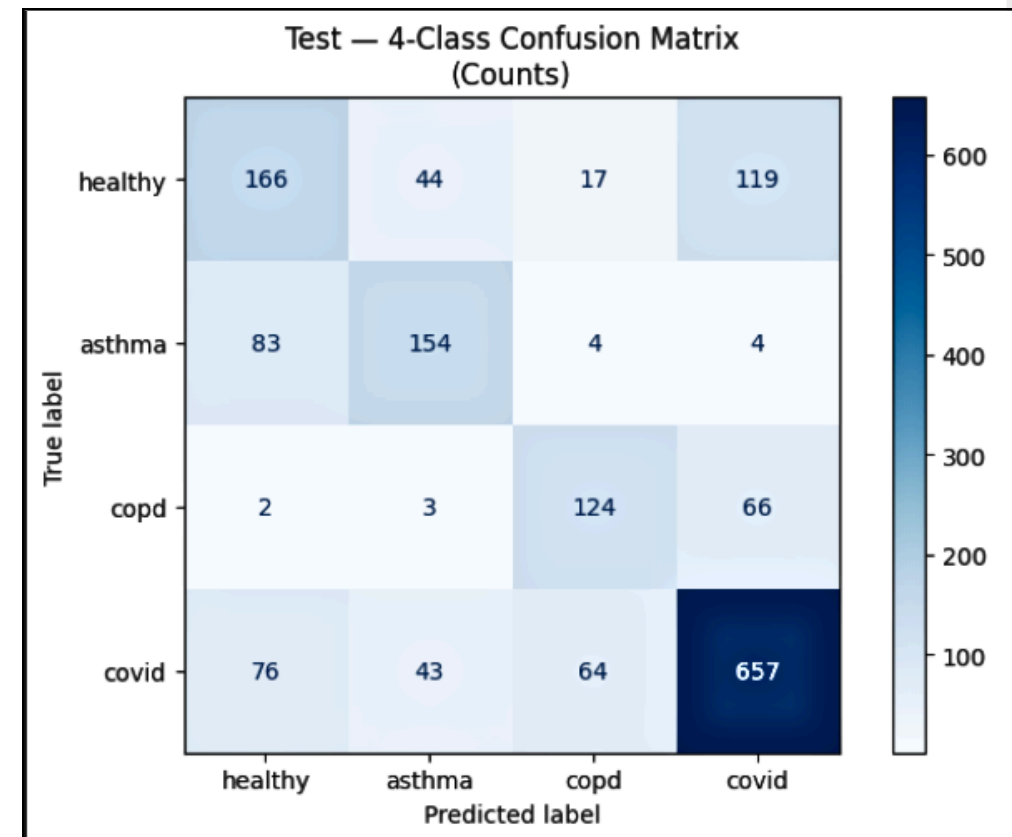
~11.3k instance

Models tested:

Random Forest, SVM-RBF, Logistic Regression, XGBoost, Gradient Boost, HistGradBoost



- HistGradBoost is our
- **~67.1% accuracy for multi-disease, 77.1% for binary**
- Cough + vowel features are not enough for differentiating between all four classes.



Classification Report:

	precision	recall	f1-score	support
healthy	0.5076	0.4798	0.4933	346
asthma	0.6311	0.6286	0.6299	245
copd	0.5933	0.6359	0.6139	195
covid	0.7766	0.7821	0.7794	840
accuracy			0.6771	1626
macro avg	0.6272	0.6316	0.6291	1626
...				
Overall Accuracy :	0.6771			

Approach

#5 *CNN trained on audio files*

Input:

Cough + vowel audio files

Features extracted:

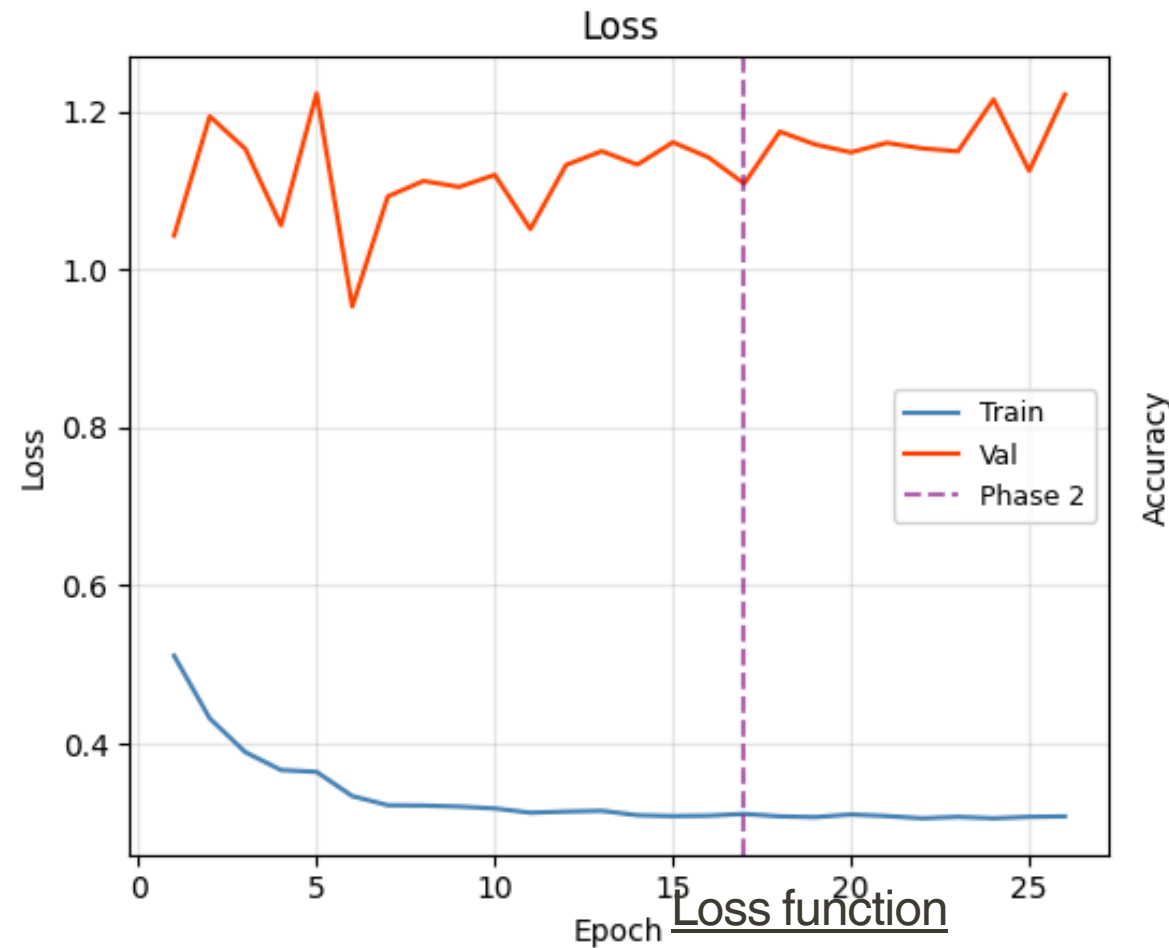
2.3 Million params from transfer learning

Dataset size:

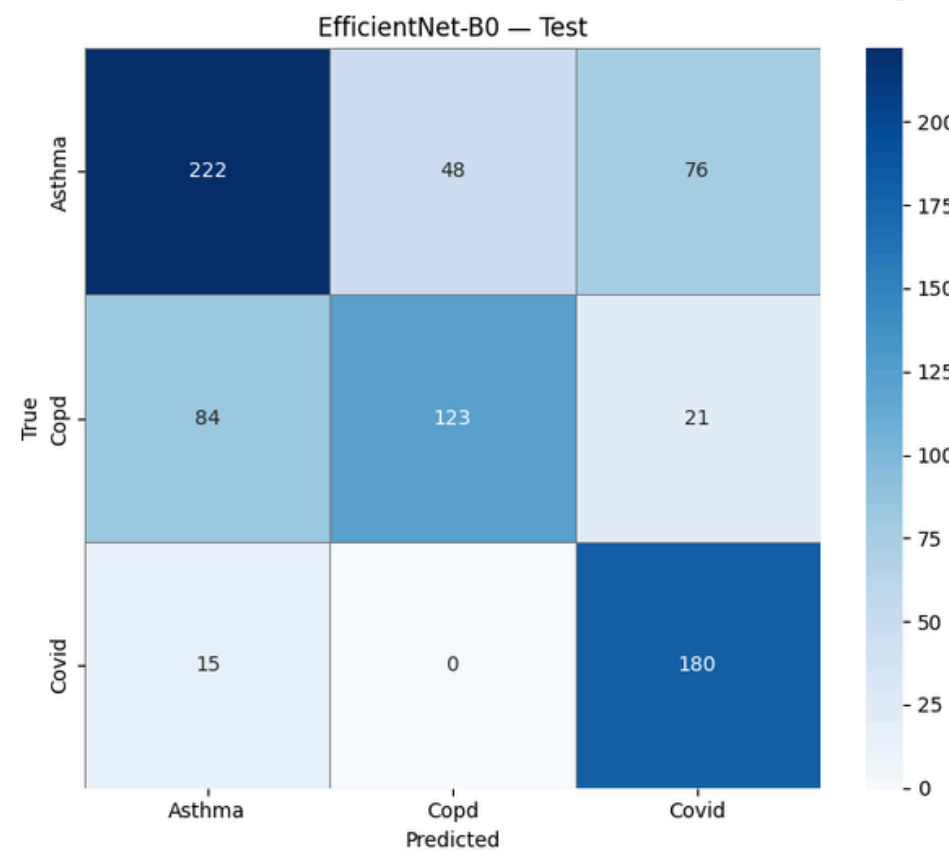
11.3k instances, 121 features

Models tested:

Used a pre-trained CNN EfficientNet-B0 (light weight)



- CNN performed with decent accuracy and recall for differentiating between diseased classes
- **~68.3% accuracy**
- Loss curves show decent decrease in loss.
- CNN is overall the best at differentiating between diseased classes out of all the other models we tested.



Confusion matrix

```

=====
TEST METRICS SUMMARY
=====
Accuracy           : 0.6827
Precision (macro) : 0.6869
Recall (macro)    : 0.7014
F1 Score (macro)  : 0.6816
    
```

Performance Metrics

FINAL APPROACH

CNN + XGboost

Input:

extracted acoustic features from cough and vowel

Features extracted:

121 features

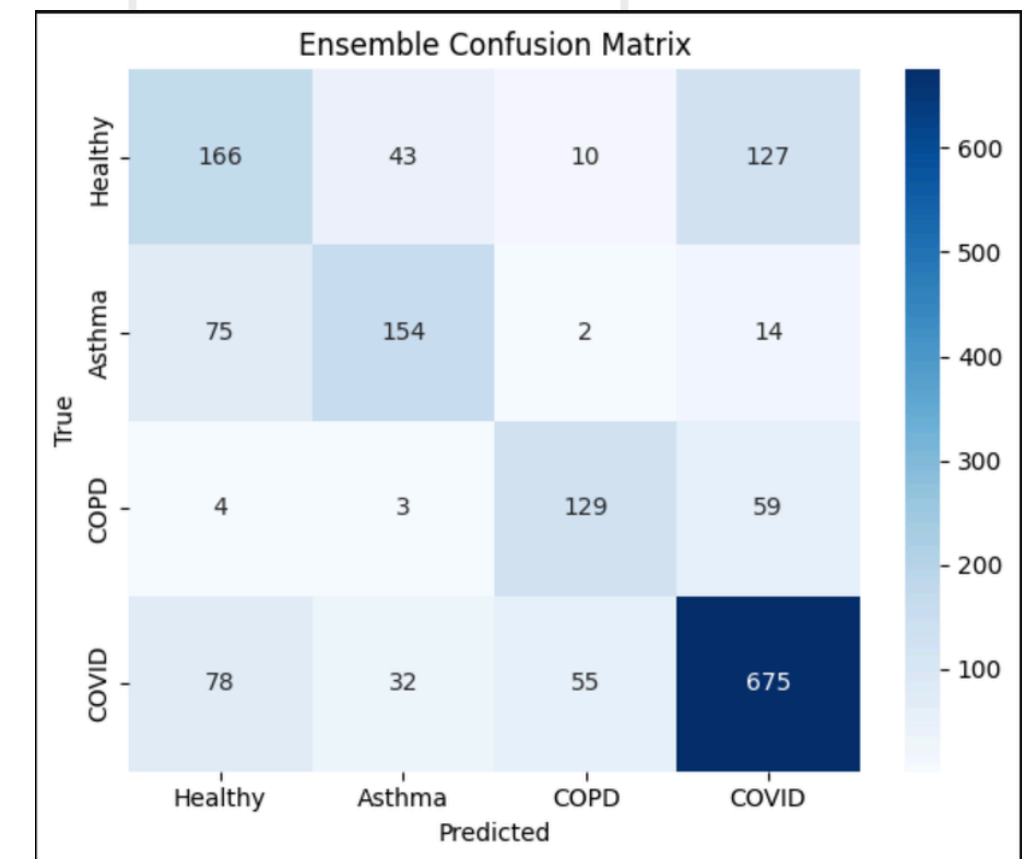
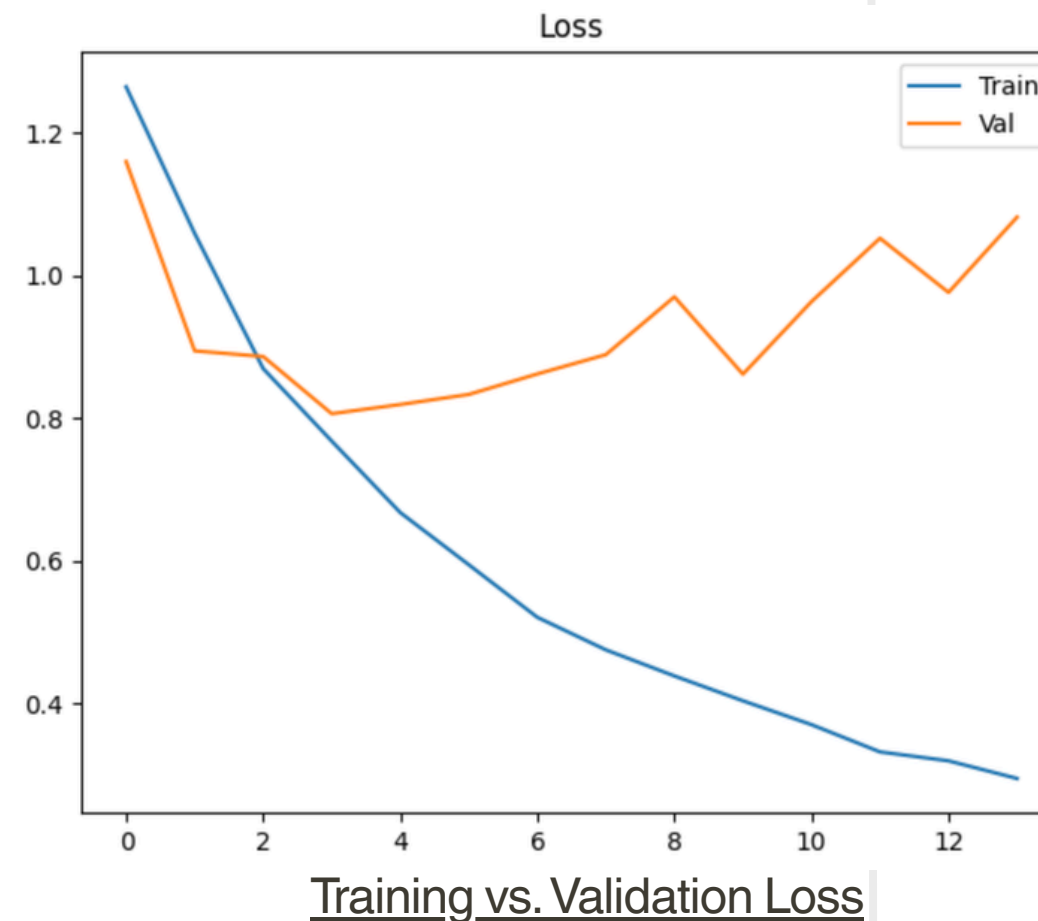
Dataset size:

~11.3k instance and 121 features

Models tested:

used cnn + xgboost

- **CNN** learned deep acoustic patterns from cough + vowel features
- 121 audio features used as model input
- Achieved ~69% test accuracy
- Strong disease-specific classification performance
- Healthy vs non-healthy differentiation remained challenging



Confusion Matrix

Has the best accuracy, recall and F1-score among all for multi-disease classification.

This is our final model.

PERFORMANCE METRICS AND DEPLOYABILITY

Performance Metrics of Final Model

- **Accuracy(69.13%)** = prioritised because overall prediction correctness
- **Recall(69.13%)** = (prioritised because minimises false negatives (missed disease cases))
- **F1-score(68.87%)**= (prioritised because balances precision + recall for imbalanced data)

FUTURE SCOPE

- Refinement to **2-layer ML model**
- Expansion to **additional respiratory diseases**
- Use of high-performance **GPUs** to train deeper and more accurate CNN architectures
- **Multimodal** data fusion using cough, vowel, breathing sounds, and patient metadata
- Model **compression and optimization for efficient real-time deployment**
- **Explainable AI** for interpretable clinical decision support
- **Real-time** mobile and edge-device deployment
- **Noise-robust** respiratory audio classification in real-world environments

DEPLOYABILITY

STRENGTHS:

- Hybrid system combining XGBoost (features) + CNN (audio learning) for better coverage
- Two-stage pipeline enables fast screening (healthy vs diseased) followed by detailed disease classification
- Lightweight inference suitable for low-resource clinical settings

LIMITATIONS:

- Limited dataset size and variability
- CNN is currently lightweight (transfer learning used) due to compute constraints, needs GPU

Thank
You.