Computer Methods and Programs in Biomedicine 243 (2024) 107911

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases

Santosh Kumar [a,*], Vijesh Bhagat [a], Prakash Sahu [a], Mithliesh Kumar Chaube [b], Ajoy Kumar Behera [c], Mohsen Guizani [d], Raffaele Gravina [e], Michele Di Dio [e,f], Giancarlo Fortino [e], Edward Curry [g], Saeed Hamood Alsamhi [g,h]

[a] Department of Computer Science and Engineering, IIIT-Naya Raipur, Chhattisgarh, India
[b] Department of Mathematical Sciences, IIIT-Naya Raipur, Chhattisgarh, India
[c] Department of Pulmonary Medicine & TB, All India Institute of Medical Sciences (AIIMS), Raipur, Chhattisgarh, India
[d] Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, United Arab Emirates
[e] Department of Informatics, Modeling, Electronic, and System Engineering, University of Calabria, 87036 Rende, Italy
[f] Annunziata Hospital Cosenza, Italy
[g] Insight Centre for Data Analytics, University of Galway, Galway, Ireland
[h] Faculty of Engineering, IBB University, Ibb, Yemen

ARTICLE INFO

Keywords:
Deep learning
Chronic Obstructive Pulmonary Disease (COPD)
Computed Tomography (CT)
Texture
Lung sound
Classification

ABSTRACT

Background and objective: Chronic Obstructive Pulmonary Disease (COPD) is one of the world's worst diseases; its early diagnosis using existing methods like statistical machine learning techniques, medical diagnostic tools, conventional medical procedures, and other methods is challenging due to misclassification results of COPD diagnosis and takes a long time to perform accurate prediction. Due to the severe consequences of COPD, detection and accurate diagnosis of COPD at an early stage is essential. This paper aims to design and develop a multimodal framework for early diagnosis and accurate prediction of COPD patients based on prepared Computerized Tomography (CT) scan images and lung sound/cough (audio) samples using machine learning techniques, which are presented in this study.

Method: The proposed multimodal framework extracts texture, histogram intensity, chroma, Mel-Frequency Cepstral Coefficients (MFCCs), and Gaussian scale space from the prepared CT images and lung sound/cough samples. Accurate data from All India Institute Medical Sciences (AIIMS), Raipur, India, and the open respiratory CT images and lung sound/cough (audio) sample dataset validate the proposed framework. The discriminatory features are selected from the extracted feature sets using unsupervised ML techniques, and customized ensemble learning techniques are applied to perform early classification and assess the severity levels of COPD patients.

Results: The proposed framework provided 97.50%, 98%, and 95.30% accuracy for early diagnosis of COPD patients based on the fusion technique, CT diagnostic model, and cough sample model.

Conclusion: Finally, we compare the performance of the proposed framework with existing methods, current approaches, and conventional benchmark techniques for early diagnosis.

1. Introduction

The chronic obstructive pulmonary disease (COPD) is a significant public health concern worldwide, the world's third leading cause of mortality. Every year, more than five million people worldwide die from one of the five severe major pulmonary respiratory disorders: (1) COPD, (2) asthma, (3) tuberculosis (TB), (4) lung cancer, and (5) lower respiratory tract infection and other respiratory diseases [1]. Moreover, COPD

* Corresponding author.
E-mail addresses: santosh@iiitnr.edu.in (S. Kumar), vijesh20300@iiitnr.edu.in (V. Bhagat), prakash@iiitnr.edu.in (P. Sahu), mithilesh@iiitnr.edu.in (M.K. Chaube), drajoybeherakims@gmail.com (A.K. Behera), mguizani@ieee.org (M. Guizani), r.gravina@dimes.unical.it (R. Gravina), m.didio@aocs.it (M. Di Dio), giancarlo.fortino@unical.it (G. Fortino), edward.curry@insight-centre.org (E. Curry), saeed.alsamhi@insight-centre.org (S.H. Alsamhi).

https://doi.org/10.1016/j.cmpb.2023.107911
Received 10 August 2023; Received in revised form 23 October 2023; Accepted 1 November 2023
Available online 10 November 2023

is a non-communicable disease that takes a considerably long time to diagnose. The lungs' airways and alveoli are elastic or flexible in the human respiratory body. However, in the severity of COPD, in the human respiratory, the lung passes less air inside and out of the lung airways due to the following reasons: (1) the airway walls are thickened, and it becomes inflammatory during breathing in and out through the lung airways, and (2) the airways producing dense mucus around the airway walls and it might become choked during breathing [1]. Therefore, the diagnosis of COPD has recently increased due to wide applications in the medical field [1].

The traditional medical assessment method is complex for quantitatively analyzing COPD measures as some critical diseases exhibit common symptoms. Therefore, traditional medical diagnostic methods are leading to provide results for misdiagnosis of COPD patients, which is common in clinical practice for early diagnosis of COPD. For example, dyspnea and coughing are common side effects of asthma and COPD diseases. However, reversibility and persistency (respectively) are distinguishing features that in asthma, respiratory failure is reversible, and coughing is consistent. Heavy breathing and respiratory symptoms prompt an asthma diagnosis, even though sputum production leads to a COPD diagnosis [1] [2]. Therefore, there is a need to design the early diagnosis and prognosis framework of COPD for accurate classification of these non-communicable disease symptoms (i.e., COPD and asthma) using Machine Learning (ML) and Deep Learning (DL) techniques [2–4]. Asthmatic patients are often significantly younger, have a quick illness start, and are less likely to smoke [1] [5]. On the other hand, COPD patients typically present at an advanced age (older than 40 years), with slow development and long-standing cigarette smoking and pollution being the essential risk factors [6–12]. Although these distinguishing symptoms and traits aid in differentiation, there is also some overlap because a patient's clinical description only sometimes includes all belonging to one class and eliminates those belonging to the other. For example, wheezing and reversibility of airway obstruction can be found in COPD [3], and older age and smoking in asthma patients may be deceptive [6–12]. The diagnostic power of the above-mentioned discriminative symptoms and traits is mainly determined by how precisely the patient reports their lifestyle and complaints [6–12].

Lung function test is known as the spirometry test along with subjective lung auscultation which is defined as the standard diagnostic technique for severity analysis of COPD patients. Recently, spirometry-based diagnostic model has become easy to use due to its fast diagnosis and prevalent for regular monitoring of lung capacity with the emergence of portable spirometers. Under a spirometry test, a COPD patient needs do inhale deeply and do exhale forcefully into the attached mouthpiece to measure the different spirometry parameters while the nose is closed during exhale. To diagnose of COPD and perform statistical assessment-based evaluation method to measure severity levels of COPD patients, the values for spirometry parameters, including significant required variables such as, Forced Expiratory Volume in 1s (FEV1), Forced Vital Capacity (FVC), and calculated ratio of FEV1 and FVC, or measured ration of Forced Expiratory Volume (FEV), are utilize. The significant parameters to discriminate asthma and COPD are: 1) ratio of FEV1/FVC and ratio value ≤ 0.70 indicates COPD and the ratio value of FEV1/FVC less than or equal to 0.70 indicates asthma. (2) The percentage change of FEV1 with respect to the baseline after taking bronchodilators ($\leq$ 12% for COPD and $\leq$12% for asthma). On the contrary, the most significant symptom of COPD patient during lung auscultation is wheezing, which is caused by narrowed lung airways and blockages caused by highly dense layer of sputum (which is amalgamation of saliva and mucus) [6–12] [36].

The only objective assessment of the pulmonary respiratory condition of patients can be diagnosed through a pulmonary function test called spirometry. However, the diagnostic procedure on spirometry based test for COPD treatment takes a long time to diagnose and prognosis under supervision of pulmonary experts and professionals, even for experienced medical experts; clinical professionals and pulmonary specialists cannot provide accurate observation for early diagnosis of COPD patients. Moreover, experienced pulmonary professionals and specialists cannot provide correct classification of COPD by observing scanned lung images using Computerized Tomography (CT) scanning methods and listening to contaminated patient's cough audio or lung sound audio samples with stethoscope devices. For detail observation of COPD patients, they also required extensive human resources to conduct large-scale test scenarios and clinical diagnosis procedures of COPD patients using a medical procedure and enabled computing paradigm for the early diagnosis of patients [1] [35]. Therefore, it is necessary to design and develop efficient systems and method to analyze abnormalities and assessment of COPD severity in simple ways [1]. Exarchos et al. [1] proposed a CHRONIOUS system that caters to different types of health services, assists clinicians, and monitors activities for patients suffering from COPD. It has also been developed as an interactive platform to support intelligent services for remote people; however, interpretation of different subsystems could be better for monitoring other chronic diseases and predicting COPD cases.

Yang et al. [22] studied the classification of COPD subtypes, which includes a pulmonary emphysema-based classification model, on extracted spatially-informed lung texture from CT lung scan images for subtype classification. They prepared a lung CT scan image dataset based on the multi-ethnic study of atherosclerosis (MESA) for detailed analysis of each subtypes. Roy et al. [35] proposed a model using DL techniques for the classification of COPD subtypes based on lung sound samples for COPD severity classification. The major shortcoming of the proposed method is that the performance of the proposed model is not validated on massive amounts of lung audio samples.

The lung sound may take many different forms depending on the condition. It may also have audible and non-audible features, with the most frequent sick lung noises being crackle, stridor, and wheezing. Crackle is a blatantly annoying non-musical sound. It is typically associated with those who have lung fibrosis, pneumonia, asbestosis, or chronic bronchitis. Different stridor sounds include loud, highly melodic respiratory sounds and high-pitched sounds. It may be heard in those with laryngeal inflammation, airway inflammation, tumors, or tracheal constriction. Wheezing is a lung sound linked with chronic respiratory illnesses, including asthma and COPD. Constricted airways and sputum obstruction distinguish wheezing.

On the other hand, CT scan images are employed in the new computing paradigm to conduct an early prognosis of COPD. For early diagnosis, CT imaging techniques separate COPD into two key components: (1) chronic bronchitis and (2) emphysema. These computing paradigms provide efficient and quick diagnostic models and methods for visual representation and automated evaluation of. However, visualizing sickness severity and progression to create forecasts is very challenging [5]. As a consequence, the visual assessment technique for noncommunicable COPD diagnosis is subjective, time-consuming, and prone to various findings from expert to expert, referred to as intra-observer and inter-observer variability [6–12]. Several research studies have been undertaken to study intra-observer and inter-observer variances in high-resolution CT images to accurate COPD misclassification owing to intra-observer and inter-observer discrepancies in CT scan images. Author [13] presented a technique for evaluating inter-observer variability in high-resolution CT patterns of scanned images to help in early diagnosis. However, they could not detect the differences between particular patterns in inter-observer variability, which remains an open question for the CT scan-based technique of COPD diagnosis.

Deep learning (DL) is gaining proliferation due to massive applications in the early diagnosis of noncommunicable diseases. Discriminatory features are extracted from CT scan images and cough samples to classify non-communicable diseases using DL techniques and ML techniques [1–5]. By evaluating the acquired datasets using AI and DL algorithms that include "intelligence" to the prognosis system for classifying COPD and measuring severity level for individual life sustainability [5]. Recent noncommunicable disease research has focused on applying DL

methods, statistical machine learning, and data mining techniques. Chin et al. [5] created a strategy for diagnosing Rheumatoid Arthritis (RA) that combines association rule mining and classification algorithms for Rheumatoid Arthritis categorization (RA) [33]-[34] [35].

Deep learning (DL) is gaining high proliferation due to massive applications in the early diagnosis of non-communicable diseases. DL techniques are vital in accurately representing extracted features from CT scan images. Discriminatory features are chosen from the extracted CT scan images and cough sample features to classify non-communicable diseases using DL and ML techniques [1,5,7–10]. By evaluating the acquired datasets using AI and DL algorithms that include "intelligence" to the prognosis system for classifying COPD and measuring severity level for individual life sustainability [5]. Recent noncommunicable disease research has focused on applying DL methods, statistical machine learning, and data mining techniques. C.Y. Chin et al. [5] created a strategy for diagnosing Rheumatoid Arthritis (RA) that combines association rule mining and classification algorithms for Rheumatoid Arthritis categorization (RA).

Lung sounds may take many forms depending on the condition. It may also have audible and non-audible features, with the most frequent sick lung noises being crackle, stridor, and wheezing. Crackle is a blatantly annoying non-musical sound. It is typically associated with those who have lung fibrosis, pneumonia, asbestosis, or chronic bronchitis. Different stridor sounds include loud, highly melodic respiratory sounds and high-pitched sounds. It may be heard in that with laryngeal inflammation, airway inflammation, tumors, or tracheal constriction. Wheezing is a kind of lung sound that is linked with chronic respiratory illnesses, including asthma and COPD. Constricted airways and sputum obstruction distinguish wheezing. To diagnose the early stages and assess the severity levels of COPD, the estimated values for spirometry variables, forced expiratory volume in 1s (FEV1), forced vital capacity (FVC), and the measured ratio of FEV1/FVC and forced expiratory volume ratio (FEVR) play a significant role for fast diagnosis. Based on observed spirometric measurement, FEV ratio, and the prominent characteristics of wheezing from the human respiratory sound, the global initiative has segmented the degree of COPD severity and group into five classes for chronic obstructive lung disease (GOLD): COPD-0, COPD-1, COPD-2, COPD-3, and COPD-4. COPD patients belong to COPD-0, a low-risk condition for those who have been smoking for a few years, with a FEVR of more than 85% [9].

On the other hand, COPD-1 depicts moderate-level severity and exhibits prolonged symptoms with minimal wheezing. The group of patients has a ratio of FEV more than 80%. At the same time, COPD-2 signifies a group of patients with an intermediate degree of COPD severity, with ratio values of FEV that range from 50%-80%. During expiration and inspiration, it is essential to wheezing a precise observation to define class as COPD-3 and COPD-4. It is caused by blockage and constriction of the airways and usually coexisting heart disorders. People with an FEVR of 30%–50% with all chronic symptoms and lung infections are likely to have severe COPD and fall into the COPD-3 group. COPD patients with COPD-4 level indicate the severe stage of COPD. They have an FEVR of less than 30%, and they experience all chronic symptoms and respiratory issues [9].

On the other hand, CT-based diagnostic techniques are employed to improve the early diagnosis of COPD diseases. For early diagnosis, CT imaging techniques separate COPD disease analysis by considering it into two key components: (1) analysis of chronic bronchitis and (2) emphysema analysis. Chronic bronchitis is a class of COPD that shows a massive amount of the productive cough of more than three months occurring within two years. In which, the COPD patient presents with chronic productive cough, malaise, and symptoms of excessive coughing along with the chest or abdominal pain. On other hand, Emphysema has been morphologically depicted by the enlargement of airspaces with destruction of alveolar walls distal to the terminal bronchioles. I is a lung condition of the COPD that causes shortness of breath. There is a small airway disease, which at times results in emphysema. For classification of chronic bronchitis and emphysema, CT-based computing paradigms provide efficient and quick diagnostic models and methods for visual representation and automated evaluation of COPD patients based chest-X ray (CXR) images/CT scan images. However, visualizing sickness, severity and progression based on CT/CXR images to create forecasts takes a lot of work for proper analysis. Consequently, the visual assessment techniques for non-communicable COPD diagnosis is subjective, time-consuming, and obtained mis-classification results through various findings from the expert to expert for COPD analysis, referred to as intra-observer and inter-observer variability [5]. To solve these issues, several studies [5] [13] have analyzed intra-observer and inter-observer variability based on high-resolution CT/CXR images to correct COPD misclassification owing to intra-observer and inter-observer discrepancies in CT scan images. In similar direction, in [13], author proposed a technique for evaluating inter-observer variability in high-resolution CT patterns of scanned images to help in early diagnosis of COPD. However, the authors [13] could not detect the differences between particular patterns in inter-observer variability, which remains an open question for the CT scan-based technique of COPD diagnosis.

To solve these issues, the supervised machine learning (ML) techniques are applied to classify COPD and non-COPD based on extracted lung texture features from the CT scan images, where the built classification models are trained on labeled lung audio sample with manually annotated regions of interest (ROI) in the small melspectogram patches [5–13]. Compared to the densitometric measure approach, it requires many tagged CT image information for training for COPD diagnosis. The final output of pre-trained classification algorithms is used to assess the severity of COPD. However, supervised classification algorithms need massive labeled data for early detection of COPD, often provided by data manually marked by radiologists and other experts. Because manual annotation data has the same constraints as visual emphysema assessment in CT images [5], diagnosing COPD using a prepared manual annotation database is challenging. Emphysema zones (ROI) are difficult to define precisely because sickness patterns may be subtle and extensive, particularly in the early stages of COPD diagnosis. Furthermore, the investigation of emphysema zones is confined to human professionals' current experience and competence, even though there may be significant bias or error towards frequent situations in the annotated data [13].

As a result, previously unknown or even less common COPD-related patterns may be entirely overlooked by pre-trained classification algorithms, and critical discriminating information may be missing. The main research question of this work is ***how to identify COPD for early diagnosis and accurate prediction?*** To address this issue, we propose a novel multimodal framework based on CT scan/chest-X ray images and lung sounds (cough sample) database using machine learning algorithms.

The proposed multimodal framework consists of a CT scan-based diagnosis model and a lung sound-based diagnosis model. We prepared a database from the Department of Pulmonary Medicine and TB, All India Institute of Medical Sciences (AIIMS), Raipur, Chhattisgarh, India. The CT scan-based model extracts texture features from the CT image database. The lung sound-based technique builds a data-driven model for the classification of COPD based on extracted Mel Frequency Cepstral Coefficients (MFCCs) features of lung sounds using Incremental Decision Tree (IDT), Support Vector Machines (SVM), ensemble learning classifiers, deep learning-based Convolutional Neural Network (CNN) models and other texture feature descriptor techniques for early prediction of COPD. The significant contributions of the work are given below:

1. Gaze upon the vanguard of medical ingenuity, with unbridled passion, we proudly present our groundbreaking study—a symphony of innovation meticulously composed to forge a new path in COPD research. Our quest for excellence led us to unveil a revolutionary multimodal framework, a masterpiece born of the harmonious

union between cutting-edge deep learning techniques and the wonders of medical imaging. To the best of our knowledge, this is the first work that proposed a novel framework for early diagnosis and accurate prediction of COPD using deep learning techniques.
2. The multimodal framework integrates the CT image-based model and lung sound-based model using ensemble learning techniques for early prognosis and predicting COPD diseases.
3. The proposed framework performs the prepossessing steps to remove noises and specific artifacts from the collected lung sound samples of cough and CT scan database in order to get better quality data before feeding them into ML as well as DL models.
4. The lung sound sample-based technique detects the early and late phases of expiration, and expiration and inspiration features, which are reflected on the acoustic properties of pulmonary sounds using CNN architecture. The outcome of the lung-based technique is then utilized to enhance the performance using machine learning.
5. A comparative study is done with the proposed system and different existing methods, such as intensity and texture-based feature extraction methods for classifying COPD and non-COPD and analysis of different other diseases in CT scans and cough samples using machine learning techniques.

The rest of this paper is structured as follows. We outline the suggested structure with all steps in Section 2. Section 3 describes the implementation environment and model training processes. Section 4 describes the implementation and analysis of experimental results. Statistical analysis of extracted discriminatory features from the CT scan and cough/lung sample datasets is illustrated in Section 5. Finally, section 6 concludes the works and suggests further developments.

2. Proposed framework

This section describes the proposed framework for COPD diagnosis. The framework includes (1) a CT scan-based model and (2) a cough-based model. The framework employs a classification model using CT scan images and cough audio samples as training instances. Each model learns class-discriminating characteristics from CT scans and cough samples for early diagnosis.

2.1. Chest CT diagnostic model

The primary objective of the proposed framework is to integrate a chest diagnostic model that extracts information from the CT scan images for COPD early diagnosis. We used texture-based methods to extract features for classifying COPD and normal cases from segmented Regions of Interest (ROIs) of CT images for analyzing quantitative measurements for COPD using machine learning techniques. Each model learns the extracted features from CT images and performs the desirable results.

2.2. Preparation and description

The prepared database includes the CT scan image datasets from current and former chain smokers who participated in the Department of Pulmonary Medicine and TB, AIIMS, Raipur, Chhatisgarh. The Siemens/Denition AS+ was used to capture CT scan images, having a tube voltage capacity of 120 kV, the exposure of the connected device and model is 40 mAs (milli-ampere seconds), the slice thickness is 1 mm, and the in-plane resolution ranges from 0.7 to 0.78 mm. Chain smokers are exposed to a diagnostic of conjunctive years upon entrance and are scanned annually for early prognosis. PFTs and CT scans were taken of the participants, which included forced Expiratory Volume (FEV) in one second and Forced Vital Capacity (FVC).

Subjects (patients) were re-scanned after three months if non-calcified nodules with a diameter of 5-15 mm were discovered. We collected the CT scan images database from individuals who participated in two six-month sessions for early diagnosis. The prepared COPD dataset comprises derived features from 40,000 CT scan images of COPD patients (total number of patients: 2000) scanned at the AIIMS, Raipur, Chhattisgarh, India (i.e., 2000×10 images $\times$ 2 sessions). The Siemens/Denition scanner captured 20,500 CT pictures, while the Siemens/Denition AS+ scanner captured 19,500 images. 50 extracted features describe each picture, each specifying a volumetric ROI of size $41 \times 41 \times 41$ voxels extracted at arbitrary locations inside the lung mask.

The filter responses from rotation-invariant, multi-scale filters are used to build the histogram-based evaluation for the prognosis of COPD. We used the Gaussian filter banks to employ Gaussian filters to achieve smoothed images to compute features and statistically significant quantified measures of COPD severity. The severity level is measured based on estimated posterior probability by combining individual ROI's probability into one posterior probability. The phases for further processing of the chest CT scan are as follows, also shown in Fig. 1:

2.2.1. Pre-processing step:

- **Segmentation of the Lung Fields:** The proposed CT scan model segments collected CT scan images to establish unique RoI for extracting discriminating characteristics. It will also aid in the analysis and statistical significance of extracted characteristics for a more accurate disease prediction.
  A region-growing method segments the lung fields in a CT image. When the lung parenchyma is less than 400 Hounsfield Units (HU), a region-growing approach is employed to segment the lung fields in a CT image. The approach finds an RoI from the trachea by finding a black cylindrical object towards the top of the image, which is then utilized to segment the right and left main bronchi. After dividing the left and right primary bronchi, the left and right lung portions are separated. A post-processing method is applied to form the primary lung fields, which involves removing incorrectly inserted esophageal sections. This is achieved by identifying tube-like structures between the segmentation of the left and right lung fields.
- **Texture Feature Descriptors:** The extracted holistic characteristics of CT scan images may not accurately predict early COPD illness. As a result, we investigate the extraction of texture characteristics from CT images. Texture feature descriptors are employed to extract textural details from a CT image. This involves filtering various texture features in randomly selected Regions of Interest (ROIs) from the image. Based on the Gaussian function and combinations of Gaussian derivatives, an eight-filter bank is used at different sizes, resulting in multiple filtered copies of the CT images. Histograms of filter responses are then used to represent the RoIs in the lung images [13].
- **Gaussian Filters:** We applied the low pass filtering techniques to remove the noises from the CT scan image database. We employed an eight-filter bank based on the Gaussian function to measure the local image structure and similarity at the smoothed level of processed CT scan images that were used as base filters, including the low pass filter method, which also includes the Difference of Gaussian (DOG) and Laplacian of Gaussian (LoG) method, given as follows. The Gaussian function appears as in Eq: (1).

$$G(\mathbf{x},\sigma) = \frac{1}{\left(2\pi^{1/2}\sigma\right)^3} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right) \tag{1}$$

Where, (σ) is the σor scale. The voxel is represented as $\mathbf{X} = [X, Y, Z]^T$. We calculated λ_1, λ_2, & λ_3 as Eigenvalues ($\lambda_i \ \forall$ i = 1,2,3) from the defined Hessian matrix $H(X,\sigma)$ as follows (Eq. (2)).

$$\lambda_i(\mathbf{X},\sigma) = \quad |\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \tag{2}$$

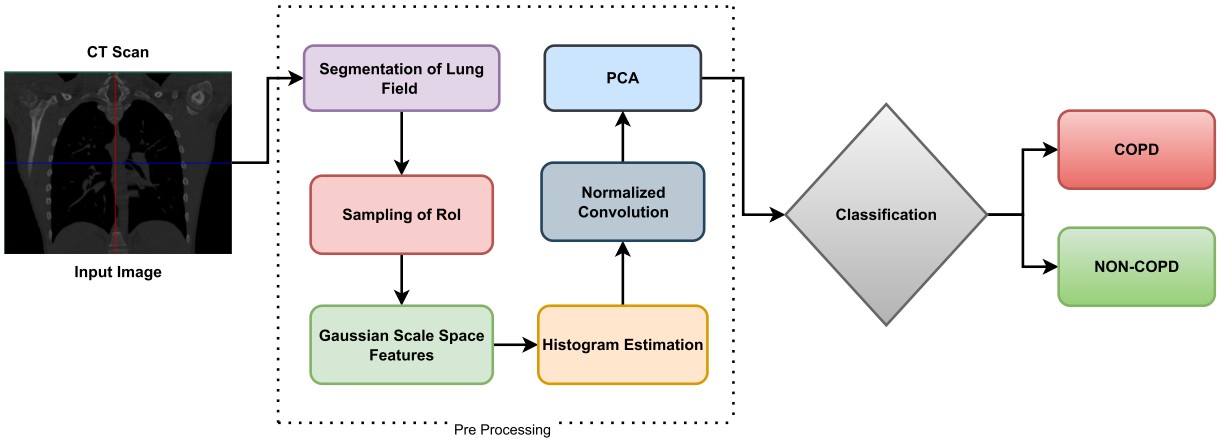

Fig. 1. Workflow of the proposed framework for classification of COPD and non-COPD based on CT scan images.

Where $[I_{X,\sigma}]$ depicts the first order derivative of given image [I] w.r.t. $[X,\sigma]$ as scale. Log method-based curvature and Hessian are defined below (see Eqs. (3)-(5)):

$$\nabla^2 G(x) = \sum_{i=1}^{3} \lambda_i(X,\sigma) \tag{3}$$

$$K(\mathbf{X},\sigma) = \Pi_{i=1}^{3}[\lambda_i(\mathbf{X};\sigma)] \tag{4}$$

$$\|H(\mathbf{X},\sigma)\|_F = \sqrt{\sum_{i=1}^{3} \lambda_i(\mathbf{X};\sigma)^2} \tag{5}$$

Because histograms are used, the voxel ordering is neglected, and a classifier cannot learn combinations of features such as the LoG filter method, from each eigenvalue. As a result, combinations of the eigenvalues, particularly (λ_3), (λ_4), and (λ_5), are explicitly used in the representation of extracted features from CT scan images.

- **Normalized Convolution:**
The lung field segmentation, obtained as a binary mask, is filtered using the normalized convolution approach [13]. The normalized convolution approach is described in Eq. (6):

$$Image_\sigma = \frac{(Seg(X)Image(X)) \otimes G(X,\sigma)}{Seg(X) \otimes G(X,\sigma)} \tag{6}$$

Where $(\otimes)$ shows convolution operation. Image segmentation is used as an indicator function to determine whether it is a lung parenchyma voxel. $Seg(X)$ is the segmented Image. Finite differences are used to calculate derivatives on Gaussian filtered images [23].

- **Histogram Estimation:**
Quantifying the filter responses yields histograms of the filter responses. The bin widths are determined through the use of adaptive binning [18]. This approach adjusts the histogram bin widths locally to the specific dataset, ensuring that each bin contains an equal amount of data. The histogram is constructed from all the data but adjusted to consider only lung segmentation regions of interest [18]. The number of histogram bins generated from the voxels is calculated using the following Eq. (7).

$$N_b = \sqrt[3]{N_s} \tag{7}$$

Where $[N_b]$ is an estimated bin width selection technique used in datasets described in [14], which asymptotically minimizes the true density histogram estimate's integral mean squared error. Adaptive binning can further reduce this error.

- **Principal Component Analysis (PCA) Method:** The PCA approach is utilized to reduce the dimensionality of retrieved characteristics from COPD patients' chest CT scan images and lung images. The discriminating features are chosen from the reduced feature sets based on uncorrelated statistical variables ($[\mu]$, $[\sigma]$) that maximize the variance of chosen features in order to build the model for accurate prediction [15].
- **Labeling and Classification:** The label of each lung image is assigned based on segmented RoI using image threshold-based segmentation techniques. The label of each lung RoI is determined by the defined subject's diagnosis using COPD classification with label 1 and non-COPD classification with 0 labels using the proposed framework. We have used clinical characteristics for COPD diagnosis, which is measured according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria, where pulmonary function test (Forced Expiratory Volume in 1 second- FEV1), range (80% to 120%) and FVC values are measures from clinical data of the patients (FEV1/FVC $\leq$ 0.7). RoI posterior probabilities were measured & fused via the mean rule to get the overall subject posterior probability [19].

$$P(\omega_i \mid I) = \frac{1}{N_r} \sum_{j=1}^{N_r} P(\omega_i \mid \mathbf{x}_j, I) \tag{8}$$

Where $[N_r]$ is the number of RoI examined. The average sample posterior probability (Eq. (8)) offers an estimate of the likelihood that an individual has COPD based diagnosis on the CT image. The proposed system performs the RoI-based binary classification on CT scan images using a decision tree, random forest, linear discriminant analysis, logistic regression, Naive Bayes-based probabilistic methods, Support Vector Machine (SVM) classifiers, *k*-Nearest Neighbor (*k*-NN) techniques and its variants methods. The performance of the classifiers is evaluated in terms of accuracy, precision, recall, F1 Score, and specificity.

2.3. Cough samples based diagnostic model

Particularly in the absence of accurate spirometric testing, the pulmonary sound analysis may be a supplementary tool in clinical practice for the differential diagnosis of asthma and COPD. The proposed framework includes a lung sound model consisting of two steps: (1) the training phase and (2) the testing phase. The input cough samples are used to train the proposed model in the training phase. The lung sound model is composed of the following steps: (1) data collection, (2) preprocessing, (3) feature extraction and dimensionality reduction, and (4) classification of COPD and non-COPD using the CNN technique. In the testing phase, the query/unknown lung sound/cough samples are fed to the lung diagnostic model for testing purposes based on the extracted MFCCs feature and augmentation for the classification of COPD and non-COPD using the CNN technique.

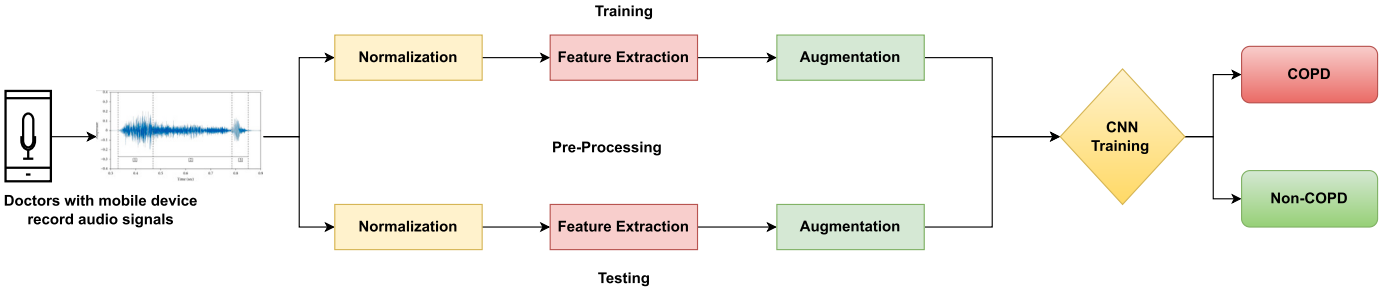

Fig. 2. Workflow of proposed Framework for early diagnosis of COPD and non-COPD based on pulmonary respiratory cough (audio) sample.

Table 1
Chest points location for the acquisition of audio dataset.

Chest location	AM	REU
Trachea (Tc)	X	A
Anterior left (Al)	Y	B
Anterior right (Ar)	NA	C
Posterior left (Pl)	NA	D
Lateral left (Ll)	NA	NA
Lateral right (Lr)	NA	NA

AM = Acquisition mode, REU = Recording equipment Used, A = AKG C417L B = 3M Littmann Classic. C = 3M Litmmann 3200 Litt3200X = sequential and single channel (sc), Y = simultaneous channel NA = Not Available.

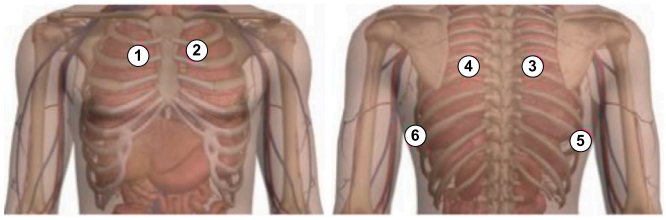

Fig. 3. Alternative chest locations of audio recording sensors.

Table 2
Augmentation methods of lung sound datasets.

Augmentation technique	Parameters
LA	0.5–1.7% of loudness
MA	DM
SA	Shift to left/right with second
SBA	Range: 0.6× and 2× speed

LA = Loudness Based augmentation, MA = Mask augmentation, SA = Shift augmentation, DM = Dynamically selected mask area, SA = Speed based augmentation.

2.4. Preparation of pulmonary respiratory cough/lung sound dataset

The pulmonary respiratory cough/lung sound database is partially prepared by the Department of Pulmonary Medicine and Tuberculosis (TB) at AIIMS, Raipur, Chhattisgarh, India; we collected standard lung sound data also from other open sources [17] [18] for COPD classification. The lung sound samples provided as input were recorded from normal and abnormal male and female patients suffering from different respiratory disorders such as COPD, Asthma, and Lower and Upper Respiratory Tract Infections (URTI) (LRTI, URTI). The prepared dataset contains 920 annotated recordings ranging from 10s to 90s.

These recordings have been made from 126 different patients, including adults and elderly patients to prepare datasets. In the dataset, we have a total of more than 5.5 hours of recordings with 6898 respiratory cycles, 1864 of which include crackles, 886 of which contain wheezes, and 506 of which have both crackles and wheezes. The dataset comprises clean respiratory cough sounds and noisy recordings that mimic real-world situations. There are several artifacts, such as noises, in most respiratory audio recordings in the prepared datasets. We used filtering methods to mitigate noises and other artefacts from our audio samples.

Table 1 shows the dataset description of cough audio samples. Fig. 2 shows the overall working of cough based diagnostic model based on audio cough samples. Fig. 3 depicts alternative body locations from where we captured the respiratory sound.

Non-COPD samples, the audio samples were enhanced using augmentation techniques to increase the number of non-COPD samples (see Table 2). The recordings, which were made with various tools, ranged from 10 to 90 seconds. Additionally, supplied are the chest locations where the recordings were made. Some respiration cycles have significant noise levels, simulating real-world settings.

2.5. Data pre-processing

In our recorded cough samples, the artifacts are reduced by using signal filtering techniques. The pre-processing steps involved in the first phase of the lung sound-based approach are as follows:

1. **Sampling rate:** The sampling rate is defined as the number of times the lung sound is sampled per second. In this work, the sampling rate defined for lung sound analysis is 5 kHz.
2. **Noise filtering:** Because it has minimal computational and spatial difficulties, we utilized the Moving Average (MA) based filtering technique to mitigate the irrelevant captured noisy data from the lung sound data. Let us consider that [x] is a sensory lung sound data vector with defined x[i] as the i^{th} lung sound value of the defined data vector [x]. The MA filter method measures x[N] using [M] prior [x] values, as shown in Eq (9). As depicted in Fig. 4, the MA filter successfully eliminates noisy data while preserving meaningful peaks.

$$ArrayX[N] = \frac{1}{M}\sum_{j=0}^{M} x[N-j] \tag{9}$$

3. **Augmentation:** The prepared cough/lung audio sample dataset of COPD patients consists of different majority classes, and the remaining audio classes belong to the minority class, as each of these classes having lesser subjects (patients). To solve this class imbalance problem, this work employed various min-max normalization and augmentation techniques for time-domain lung audio data augmentation, which are as follows.
   - **Normalization:** The min-max normalization technique is used to scale the lung sound signal to a specific range for analysis.

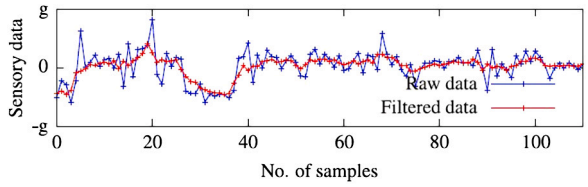

Fig. 4. Shows visualization of MA filter method on lung sound samples.

A typical normalization method scales the signal with a mean ($\mu = \sum_{i=0}^{N}[X_i] = 0$ with a standard deviation ($\sigma = 1$).

- **Triangular Filtering method**: We used the moving average filtering method and the triangular bank filter method to remove the specific artifacts and noises from the captured lung sound sample of the patient for an early diagnosis of COPD. A typical filter type for lung sound analysis is a band-pass filter with a pass-band of 100 Hz to 1000 Hz.
- **Time Stretching:** The prepared audio sample dataset can be expedited by speeding up or slowing down while maintaining lung sound pitch. In this work, we augmented the sample of the lung audio signals from each of the minority classes where each sample is stretched by two parameterized factors [0.4,0.17].
- **Pitch Shifting:**Pitch shifting mechanism is used to modify the lung sound pitch either by sped up or slowing down the sound pitch. In contrast, the duration of the lung audio signal is kept unaltered. In [31], the significance of the lung pitch shifting process is explored for the proposed framework using convolution neural network (CNN) based lung sound classification task. Here, we employed a similar method with two semitones or chosen factors of $-2, 1$ for pitch shifting to augment the minority class sample lung recordings for this work.
- **Additive noise** In this work, white noise addition is used as another audio data augmentation mechanism to increase the number of audio samples to analyze COPD diseases in the minority class. We used the Librosa library to mitigate the anomalies and unstructured audio data in the augmented lung sound sample dataset. Furthermore, to achieve segmented audio, we used the Python module Librosa to trim and pad the audio files to 20 seconds to standardize the cough sample/lung sample.

4. **Frame blocking:** During this stage, the pre-emphasized lung sound (audio) signal s(n) is segmented into N-sample frames, with M-sample gaps between them. If $M \leq N$, the whole frame is overlapped, and the resultant spectral estimates are correlated from framework to framework. There is no overlapping between neighboring frames if $[M \geq N]$.
5. **Windowing:** The next pre-processing step is to use the Hamming window method on each frame to reduce audio signal discontinuities at the beginning and end of the frames. The window approach is used to taper the audio signal to zero at the start and end of each frame. If we define the window as w(n), the signal is the result of windowing if $[0 \leq n \leq N-1]$ (see Eqs. (10)-(11)).

$$x'[n] = x[n]w(n) 0 \leq n \leq N-1 \tag{10}$$

$$W[n] = 0.54 - 0.46 \times cos[\frac{2\pi n}{N-1}], 0 \leq n \leq N-1 \tag{11}$$

2.6. Feature extraction

Feature extraction is one of the most critical phases in converting an audio signal data stream into a collection of statistical characteristics. Mel Frequency Cepstral Coefficients (MFCCs) and Mel-Spectrogram were among the characteristics of cough sound retrieved. The Chromagram feature contains the spectrogram of lung sound chroma-stft (chroma Short-Time Fourier Transform (STFT)), Constant-Q Chromagram (chroma cqt), and chroma-cens (Chroma Energy Normalized Variant (CENS)).

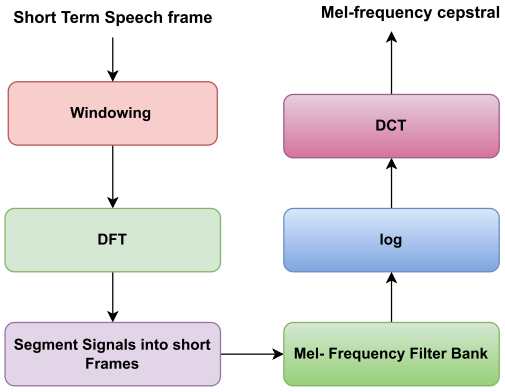

Fig. 5. Block diagram of feature extraction from lung sound.

Using the log filter bank and Discrete Cosine Transform (DCT) techniques, we estimated the first and second-order derivatives for Cepstrum analysis, done on the Mel filter bank energies, as shown in Fig. 6. The feature vector for assessment is the low-order coefficients of the cosine-converted signal. Because it is weighted with zero frequency cosine coefficients, the first coefficient $[C_0]$ refers to energy, while the second C1 shows the global energy balance between low and high frequencies. The extraction of MFCCs features includes the following steps shown in Fig. 5:

1. A pre-emphasis filter is utilized for noise elimination and signal enhancement, which is a high pass filter on the time-domain lung signal x[n] (see Eq (12)).

$$y[n] = x[n] \times ax[n] \tag{12}$$

2. The lung sound signal is regularly multiplied by a Hamming window (w(n)) with a set length to create a short temporal audio segment for analysis. Temporal smoothing of filter bank output vectors is done.
3. A Fast Fourier transform (FFT) technique is used to get the discrete Fourier spectrum, from which the magnitude squared spectrum is obtained.
4. The output is routed via a series of triangle filters. The filter bank simulates critical band filtering using triangle band-pass filters that act directly on the magnitude spectrum.
5. Take the logarithm of all filter-bank energies, then divide it by the number of filter-bank energies.
6. The next step is to decorrelate the filter-bank energies using DCT of the log filter-bank energies.

Adding temporal derivatives to the fundamental static MFCC parameters substantially improves lung sound-based COPD classification's efficacy. It is often referred to as delta (△) coefficients, and it is calculated using the following regression formula:

$$D_t = \frac{\sum_{n=1}^{N} n(c_{t+n}) - (c_{t-n})}{2 \times \sum_{n=1}^{N}(n^2)} \tag{13}$$

Where D_t is a delta coefficient calculated from frame [t], and $[c_t n]$ is a constant. A typical value for N is 2. The delta-delta (△△) or acceleration coefficients are generated in the same way from the cough samples, but from the deltas (△) rather than the static cough coefficients for COPD diagnosis. For the classification of lung sound of the COPD patient and characteristics analysis, we considered 19 MFCCs feature coefficients including $[C_0]$, computed their first order derivatives and second order derivatives for a total of $[C_q + C_0$+delta (△)+delta-detla (△△)] = 19+1+20+20 created 60 MFCCs features is typically used as acoustic cough features for classification.

The proposed framework on cough (audio) samples extracts the chroma features from the spectrogram of the cough sample dataset.

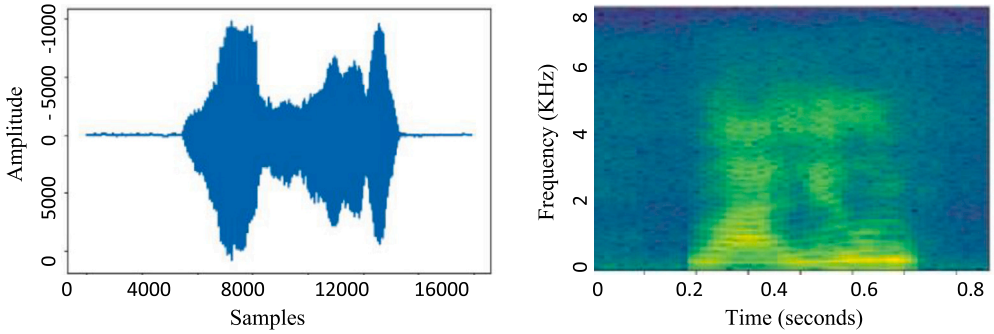

Fig. 6. Cough and power spectrum.

Algorithm 1: Feature extraction.

1. Initialization: Let y(n) be an input cough (voice signal): y[n] = x[n]-ax[n].
2. Processing and filtering noise of the input signal y[n] (see Eq. (14)):

$$y[n] = \frac{1}{n}\sum_{i=0}^{n-1} y[n-i] \tag{14}$$

3. Segmentation using Hamming Window is used to divide cough (audio) sample into distinct window
4. **Mel Frequency Cepstral Coefficients (MFCCs) features:** MFCCs feature are extracted from segmented window.
5. **Segment the signal into short frames:** Each frame's period gram was determined, and the power spectrum's period gram estimate was calculated.
6. **Computation of Discrete Cosine Transformation (DCT):** We used DCT to calculate the log filter bank energies.
7. The MeI scale is given as follows:

$$M(f) = 1125ln(1 + f/700) \tag{15}$$

These features are discriminatory for lung sound analysis for classifying COPD or non-COPD. Furthermore, the framework extracts the chroma Energy Normalized Statistics (CENS) of chroma features from the lung sound (see Fig. 7 (Upper) and Fig. 7 (Lower), respectively), because the chroma features are resistant to changes in dynamics, timbre, and articulation, these features are frequently employed in early COPD diagnosis based on lung sound in medical field. We used the FFT technique to transform the frequency domain to another domain for further analysis. The representation of lung sound samples using the FFT technique is given as follows:

Let us consider [m] as an integer number. The value of [N] can be represented as $N = 2^m$, $x = [x_0, \cdots, x_{N-1}]$. We applied the Discreet Fourier Transform (DFT) technique to measure the cepstral coefficients $V = F_N(x)$ is given as follows (see Eqs. (16)-(17)):

$$V(k) = \frac{1}{N}\sum_{j=0}^{N-1} x_j(n)\omega_N^{jk} \tag{16}$$

Where

$$\omega_N = \exp\frac{-j2\pi}{N} \tag{17}$$

To achieve the discriminatory features from lung sound, we used the FFT technique to transform the frequency to other domains for further processing of cough samples. It requires O ($N \times \log_2 N$) operations for implementing the DFT or unitary DFT, where the operation is a real multiplication and a real addition. Table 3 shows the types of cough (voice)/lung sound features and types of data available for training of convolution neural network (CNN) model for classification of COPD and non-COPD.

2.6.1. Parameter selection

The parameter selection for lung sound pre-processing is based on the characteristics of the lung sound signal and the specific tasks being performed. For example, the sampling rate of 5 kHz is used to capture the high-frequency components of the lung sound signal, which are essential for diagnosing COPD. The window size of 20 ms is used to balance the need for temporal and spectral resolutions. The overlap of 50% is used to ensure no information is lost during the analysis process. The bandpass filter with a passband of 100 Hz to 1000 Hz (filter bank method) removes artifacts or specific noises from the lung sound signal while preserving the essential frequency components for generating spectrograms for feature extraction using a convolutional neural network. The normalization technique scales the lung sound signal to a specific range, essential for machine-learning applications.

Table 3
Types of features and datatypes for CNN model.

Type of Features	Datatype
audio_series	Object
file_name	Object
mfcc	Object
melspectrogram	Object
chroma_stft	Object
chroma_cqt	Object
chroma_cens	Object
patient_id	Int64
rec_instrument	Object

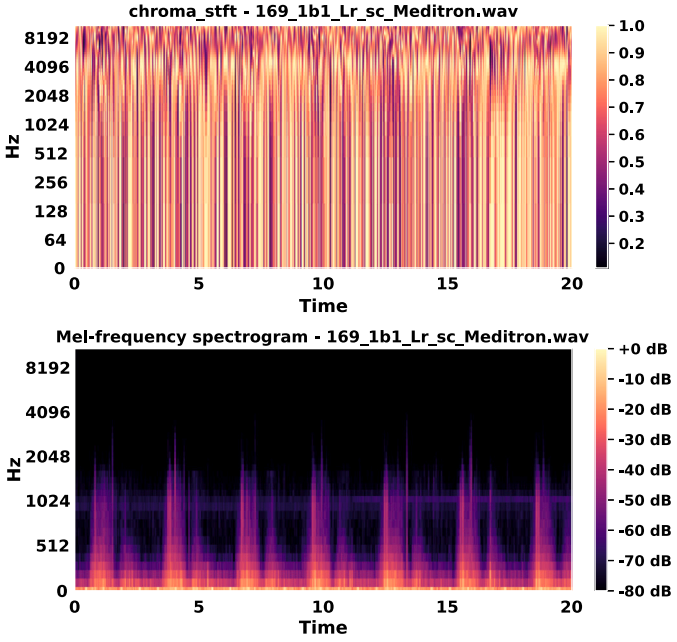

Fig. 7. (Upper) seven Mel frequency spectrogram of the lung sound and (Lower) Chroma constant-Q time-X-axis: pitch class; y-axis: time and Full size.

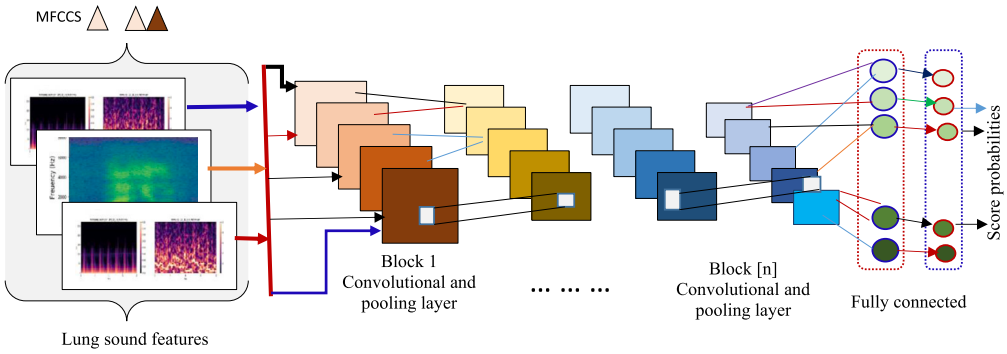

Fig. 8. Illustration of cough audio samples classification using CNN.

2.7. Classification

This stage creates a classification model using a CNN based on extracted Mel-frequency cepstral coefficients (MFCCs) and spectrogram generation feature vectors prepared from lung sound dataset by applying the deltas (△) and Delta-deltas (△ – △).

To identify the lung sound signal for early diagnosis of COPD, the proposed framework leverages CNN architecture, as shown in Fig. 8). The architecture comprises convolution (Conv2D (16, K2, ReLu), Conv2D (32, K2, ReLu), Conv2D (64, K2, ReLu), max pooling (MaxPooling2D (P2)-dropout (0.2), and Conv2D (128, K2, ReLu)), and fully connected layers. The proposed CNN framework begins with an input layer that contains a spectrogram of the collected spectrogram of lung sound features (i.e., a feature map) and forwards these feature maps to the convolution layers. To identify the features from the generated feature maps of lung sound spectrograms, the convolution layer employed the ReLU(0,1) non-linear activation function to classify whether patients have COPD or non-COPD. The max-pooling layer is then used to reduce the dimension of feature maps by recapitulating the features using the down-sampling approach. Each convolution layer contains a max-pooling-2D, pooling layer, and a final layer of the type global-average-pooling2D with ReLU and sigmoid non-linear activation functions, as illustrated in Eq. (18) non-activation functions to the classification of COPD patients. Finally, the fully connected layer computes the score probabilities of the lung sound sample output classes. These levels may be recounted as needed by the system. By evaluating the loss in the scores, CNN employs the back-propagation approach to learn the appropriate connection weights.

The cross-entropy loss function is combined with the softmax ($h\theta(x)$ activation function in the proposed framework. The fully connected layer is made up of a single neuron with a sigmoid activation function (see Eq. (18)), which creates a value ($\theta \in [0,1]$). If the value $\in$ is between 0 and 0.5, the patient is non-COPD (healthy); otherwise, the patient is COPD.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{18}$$

$$ReLU(x) = Max(0, x) = \begin{cases} 0, & \text{if } x \le 0 \\ x, & x \ge 0 \end{cases}$$

2.8. Gradient-weighted class activation mapping (Grad-CAM) method

For deep representation of visual information of a given object, several researchers included a representation mechanism in CNN architecture to preserve discriminatory information between high-level semantics and detailed spatial information. The authors of [32] proposed Grad-CAM technique enhances the transparency and explainability of CNN-based models, making them more interpretable and useful in various applications such as image classification, image captioning, and visual question answering. The neurons in these convolutional layers observe for semantic class-specific information in the image (parts of the objects). The proposed multimodal framework employs Grad-CAM techniques and uses the gradients of any target COPD or NON-COPD classification tasks flowing into the final convolutional layer to generate a coarse localization map to emphasize significant regions in the chest X-ray/spectrogram image for predicting and early detection of COPD. To achieve the class-discriminative localization feature map Grad-CAM $L^c{}_{Grad-CAM} \in R^{u \times v}$ of width u and height v for any class c, the gradient of the score is calculated for class c, y^c (before the softmax), concerning feature map activation A^k of a convolutional layer, i.e. $\frac{\partial y^c}{\partial A^k}$. These gradients flow back are global-average-pooled over the width and height dimensions (indexed by i and j, respectively) to obtain the neuron importance weights α_k^c.

During the estimation of α_k^c while backpropagating gradients concerning activations, the exact computation amounts to successive matrix products of the weight matrices and the gradient concerning activation functions till the final convolution layer that the gradients are being propagated to. Hence, this weight α_k^c represents a partial linearization of the deep network downstream from A and captures the 'importance' of feature map k for a target class c (COPD and Non-COPD).

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}} \tag{19}$$

The proposed framework performs a weighted combination of forward activation maps followed by a non-linear activation ReLU function to obtain the computed values by following mathematical expression:

$$L^c_{\text{Grad-CAM}} = \text{ReLU} \underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}} \tag{20}$$

Furthermore, we used the Adaptive Moment Estimation (Adam) optimizer technique to modify the network weights. In Adam, the exponential moving averages m_t and v_t of the gradient g_t, and also the square of the gradient g_t^2, have been used. Like m_t and v_t begin out as zero vectors. Thus, in every iteration, the values of m_t and v_t have been updated (t). Each iteration includes the same training data as the batch size is eight. For every iteration [t], Adam's parameter upgrade seems to be as follows (see Eqs. (21)-(26)):

$$g_t = \frac{\partial L}{\partial w} | w_t \tag{21}$$

$$m_t = \alpha \times m_{t-1} + (1 - \alpha) g_t \tag{22}$$

$$v_t = \beta \times v_{t-1} + (1 - \beta) g t^2 \tag{23}$$

$$m_t' = \frac{m_t}{(1 - \alpha)^t} \tag{24}$$

$$v_t' = \frac{v_t}{1 - \beta^t} \tag{25}$$

$$w_t = w_{t-1} - \eta\left(\frac{m'_t}{\sqrt{(v'_t + \epsilon)}}\right) \tag{26}$$

Where L represents the binary cross-entropy loss function, and (ϵ) is a small number being used to eliminate division by zero. To regulate the decay rates of m_t and v_t, (α) and (β) have been used. m_t and v_t have values nearer to one. The model learning rate is depicted by (η), and the objective function value of (η) is equivalent to 10^{-5}.

2.9. Classification techniques

This section demonstrates how to develop an ensemble classifier for reliable prediction of COPD and (non-COPD) healthy patients to make an early diagnosis. We employed classifiers such as support vector machines, quadratic discriminant classifiers, decision tree classifiers, *k*-NN, logistic regression, ensemble learning-based classifiers, discriminant classifiers, and others. These classifiers are used for the early classification of COPD based on lung sound/cough samples and CT image databases. These classifiers are used in the developed models to translate the input variables to the right class label for accurate classification and early diagnosis of COPD patients. The following are detailed discussions of classification algorithms:

1. **Support Vector Machine (SVM):** SVM algorithm proposes a technique to find the optimal separating hyperplane for binary classification problems. Let us consider the data vectors $q_t \in X$, $\forall$ t = 1, 2, $\cdots$,$\| X \|$, where X is the labeled dataset and $\| X \|$ is the size of the dataset, and the class labels are denoted by z_t, $z_t = +1$ if $[q_t] \in G_1$ and $z_t = -1$ if $[q_t[\in G_2$. Then, the SVM classifier finds the optimal separating hyperplane as follows:

$$r_t(W^T \alpha_t + W_0) \geq +1 \forall [t] \tag{27}$$

Where, W defines the hyperplane's average vector. The least error-prone one is chosen if there is no hyperplane to divide the cough sample data into linearly separable components; it means that the hyperplane exists in the feature space and it makes a proper separation for cough sample data and one of the hyperplanes depicts the least error which is searched in the feature space. The relationship changes when slack variables $\alpha_t \geq 0$ are included to display departures from the margin (shown in Eqn. (28)-Eqn. (30)).

$$z_t(W^T \alpha_t + W_0) \geq (1 - \alpha_t) \forall [t] \tag{28}$$

To solve for w, one should minimize

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_t \alpha_t + O \tag{29}$$

where O is defined as follows

$$O = \sum_t \alpha_t \left[z_t\left(\mathbf{w}^T \alpha_t + w_0\right) - 1 + \alpha_t\right] - \sum_t \mu_t - \alpha_t \tag{30}$$

μ_t and α_t are the Lagrange multipliers for the two primary constraints. The term G shows the notation for penalty factor to trade-off complexity and data misfitting. To classify extracted features from lung sound samples and CT scan images, polynomial kernel L^n SVM. We used kernel SVM with customization such as quadratic SVM (L^2), cubic SVM L^3 classifiers with polynomial with degree n = 1, 2, 3), and Gaussian-SVM kernel with $\sigma = 2$. We also applied a customized medium Gaussian SVM method with $\sigma = 0.7$ and a fine Gaussian-SVM kernel with $\sigma = 0.25$ for the classification of COPD cases.

We employed texture feature descriptor approaches to extract discriminatory features. We classified them by polynomial kernel, linear kernel, Gaussian kernel, and Radial Basis Function (RBF) techniques to find separable feature subsets from feature spaces.

Table 4 shows different classifiers, such as linear SVM with two classes by separating hyperplanes between COPD (Class 1) and Non-COPD (Class 2). The cubic SVM classifier that employs the kernel function $\phi(K_1, K_2)$ as cubic kernel function which is defined as $k(x_i, x_j) = (x_i \top x_j)^3$. The fine Gaussian SVM and Medium Gaussian SVM classifier with kernel factor (K) is defined as $(\frac{\sqrt{(P)}}{4})$ for separating nonlinear values.

Table 4
Different types of kernels of SVM classifier.

Type	Description
SVM	A
Quadratic kernel-SVM	B
Cubic polynomial kernel-SVM	C
Fine Gaussian Kernel-SVM	D
Medium Gaussian Kernels-SVM	E
Coarse Gaussian Kernel-SVM	F
k-NN	G

Table 5
Different Ensemble Classifiers applied.

Model Type	EM	MF
Boosted Trees	A	P
Bagged Trees	B	Q
Subspace Discriminant	C	R
Subspace KNN	D	S
RUSBoost Trees	E	T

Abbreviations: EM = Ensemble Method, MF = Model Flexibility A = AdaBoost, with Decision Tree learners, B = Random forest Bag, with Decision Tree learners, C = Subspace, with Discriminant learners, P = increases with Number of Classifiers with the maximum number of split settings, T = Max number of splits, Q = used more learner D = Subspace, with Nearest Neighbor learners, E = RUSBoost, with Decision Tree learners.

2. **Decision Tree Classifiers:** Decision trees are supervised algorithms used in classification and regression problems. It is made up of nodes and branches. There are two kinds of nodes: root nodes and leaf nodes. An unknown sample is classified based on the following criteria: the requirement for the split in the root node, and finally, the leaf, a node that links branches. The present research investigates many decision tree types, including simple, medium, difficult, boosted, and bagged decision trees. The split criteria are Gini's diversity index, and the surrogate decision splits are off, as shown in Table 7.
3. **Ensemble Classifiers:** To increase the overall accuracy of the proposed framework, we applied the ensemble learning-based strategy. The framework combines weak and strong learners to learn the extracted features from the chest CT images and lung sound/cough samples. The voting system is utilized to bring together the weak and robust learners, as summarized in Table 5. Different Neural Network (NN) classifiers are applied, shown in Table 6.
4. **k-Nearest Neighbor Classifiers**: We used K-nearest neighbor *k*-NN classifiers to classify the extracted features from CT scan images. In this work, we used the customized *k*-NN classification with selected values of *k*. The primary objective is to find the optimal features from the scattered features matrix from feature spaces. We used fine *k*-NN ($k = 1$, Euclidean distance), medium *k*-NN ($k = 9$, Euclidean distance measure), coarse *k*-NN ($k = 99$, Euclidean distance measure), cosine *k*-NN ($k = 9$, cosine distance measure), cubic *k*-NN ($k = 9$, Minkowski (cubic) distance measure), and weighted

Table 6
Different Neural Network (NN) Classifiers applied.

Model Name	Features in model
Narrow NN	A: 1 B: 10, C: ReLU, D: 1000, E: 0, F: Yes
Medium NN	A : 1, B: 25, C: ReLU, D: 1000, E: 0, F: Yes
Wide NN	A: 1, B: 100, C: ReLU, D: 1000 E: 0, F: Yes
Bilayered NN	A: 2, B: 10, G: 10, C: ReLU, D: 1000, E: 0, F: Yes
Trilayered NN	A: 3, B: 10, G: 10, H: 10, C: ReLU, D: 1000, E: 0, F: Yes

Abbreviations: Several fully connected layers, B First layer size, C = Activation, D = Iteration limit, E = Regularization strength (Lambda), F = Standardize data, G = Second layer size, H = Third layer size.

Table 7
Different types of decision tree classifiers.

Classifier Type	Description
Coarse Tree	A
Medium Tree	B
Fine Tree	C

Abbreviation A = Few leaves make coarse distinctions between classes maximum number of splits is 4, B = Medium number of leaves makes finer distinctions maximum number of splits is 20, C = Many leaves to make many fine distinctions maximum number of splits is 100.

Table 8
Training parameters used in proposed based on *k*-NN classifier.

(*k*-NN, EX.)	(Size of RoI, EX.)	Histogram dissimilarity
25,5	21,0	L = L_1,3
35,3	31,8	L = L_2,0
45,2	41,2	L = L_{EMD},7

Abbreviation: EX. = Number of experiments.

k-NN (k = 9, Euclidean distance measure) where k signifies the nearest neighbors available in feature space (Table 6).

5. **Neural Network (NN) Classifiers:** Although NN models offer high predicted accuracy and may be utilized for multi-class classification, they are challenging to comprehend.

3. Experimental results

This section presents the experimental results of the proposed framework. All experiments were carried out on the CPU computer with a Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz (2 processors) and 128 GigaByte of random access memory (RAM). The deep learning algorithm was implemented using the Colab deep learning library.

3.1. Training and parameter selection

In the proposed framework, the parameters are selected to evaluate experimental results for *k*-NN classifiers.

1. Each region of interest (ROI) is defined with size $R_O \times R_O \times R_O$ where, the size of R_O is defined as $\{21, 31, 41\}$ voxels. The R_O sizes are at the resolution of the secondary pulmonary lobule, which is around 1–2.5 cm in diameter, and we assume that the lung texture is homogeneous at this level. In the *k*-NN classifier, k = $\{25, 35, 45\}$ represents the closest neighbors. A k closest neighbor query in the approximate nearest neighbor library has a computational cost of $\mathcal{O}(kd \log n)$, where n is the number of prototypes and is the dimension. Furthermore, the results did not change much within the specified range, which is between 1-NN and the standard square root formula $\sqrt{n} = \sqrt{200 \times 50} = 100$NN. The histogram dissimilarity metric $L = \{L_1, L_2, L_{\{EMD\}}\}$ (see Table 8).
2. The various base Gaussian filters (F1), (F2), (F3) (F4), and (F5) at scales as defined below:

$$\sigma = \left\{ 0.6(\sqrt{2})^i \right\}_{i=0,\ldots,6} \text{ mm} \tag{31}$$

3. The optimal combination of r, L, and sequential forward feature selection (SFS) [15] method is to select discriminatory features from CT scan images to combine to achieve optimal characteristics. Based on the obtained optimal features histogram subset from a total of $N_f = 56$ histogram for each combination [16]. The objective function of SFS is given in [15]. The area under the receiver operating characteristic (ROC) curve (AUC) is generated from the calculated R_O region of interest of segmented chest CT scan images. This work defined the number of R_O samples for each CT scan image as $N_r = 50$. The number of histogram bins is calculated as $N_b = r$ with the adaptive histogram binning method computed using a new set of 10 randomly selected RoIs from each image.
4. We used filters:

$$\left\| \nabla^2(\mathbf{x}, 0.8), 10 \right\|, \left\{ \|\nabla(\mathbf{x}, 4.8)\|_2, 5 \right\}, \left\{ \lambda_1(\mathbf{x}, 4.8), 6 \right\}, \{G(\mathbf{x}, 2.4), 4\},$$
$$\left\{ \lambda_2(\mathbf{x}, 0.85), 6 \right\}, \left\{ \lambda_3(\mathbf{x}, 4.8) \right\}, 6, \{K(\mathbf{x}, 4.8), 5\}$$

It should be noted that a filter can only be chosen ten times, once for each repeated experiment. SFS technique [15] chooses between 5 and 11 filters from a set of 56 potential filters, with seven being the most common.

4. Performance evaluation and discussion

We conducted experiments to assess the effectiveness of the proposed system and provide answers to the following research questions:

1. What is the performance of the proposed system for early diagnosis of COPD classification models based on the Chest CT Scans and Cough audio samples?
2. What is the evaluation system of the suggested method for early diagnosis of COPD illnesses based on different sizes of selected characteristics?
3. How does the suggested model with holistic features, Gaussian features, local feature extraction, and the incremental learning approach perform in classification accuracy?

4.1. Performance measure

We employed techniques based on conventional benchmark settings to calculate the performance of the proposed multi-modal system. We employed confusion matrix-based metrics to assess the system's performance, which is as follows:

Table 9
Classification accuracy (%) on CT scan images using the proposed method with customized Decision Tree techniques.

MN	Acc	Pr	Re	FS
A	70.8	0.75	0.43	0.54
B	56.7	0.73	0.18	0.29
C	52.6	0.56	0.22	0.32

Abbreviation: MN = Model Name, A = Fine Tree, B = Medium Tree, C = Coarse Tree.

Table 10
Classification results (%) on CT scan images using the proposed method with SVM and its variants + PCA technique.

Model	Accuracy	Precision	Recall	F1-Score
SVM [15]	49.9	0.49	1	0.66
QSVM	49.9	0.49	0.54	0.51
CSVM	51.3	0.50	0.94	0.65
FGSVM	51.3	0.52	0.28	0.36
MGSVM	51.1	0.51	0.46	0.48

Abbreviations: SVM = Linear, QSVM = Quadratic SVM, CSVM = Cubic SVM, FGSVM = Fine Gaussian-SVM, MGSVM = Medium Gaussian SVM, Acc = Accuracy.

1. **Sensitivity (Sn):** It is a way to evaluate a model's ability to forecast true positives in each current category properly.
2. **Accuracy (Acc):** The metric used to assess which model is better at recognizing correlations and patterns between variables in a dataset based on the input (or training) data.
3. **Precision (Pr):** Precision is the proportion of accurately anticipated instances that are classed as positive.

$$\frac{True - positive}{True - positive + False - positive} \tag{32}$$

4. **Recall (Re):** the proportion of times the model properly predicts a certain set of real positive instances.
5. **Specificity(Sp):** The proportion of actual negatives that were expected to be negative (or true negative). This implies that a proportion of true negatives were predicted to be positive, which may be called false positives.

$$\frac{True - positive}{True - positive + False - Negative} \tag{33}$$

6. **F1-score (FS):** It gives an integrated score using precision and recall, which is computed as

$$F1 - score = \frac{2Pr * Re}{Pr + Re} \tag{34}$$

4.2. Results of CT image and lung sound based modalities

The proposed framework includes (1) CT scan-based diagnosis and (2) lung sound-based diagnosis. The performance of each modality for COPD diagnosis is measured and validated on CT images and a lung sound cough sample database using various standard procedures and setting. We utilized assessment criteria to assess the performance of each model. Individual measures such as accuracy, specificity, precision, recall, and F1-Score are utilized.

4.3. Performance analysis of CT scan modality

The system's performance is assessed using a 5-fold cross-validation setting, with 60% of the total images used for training the CT scan model and the remaining 40% images used to test the performance of

Table 11
Classification Results on CT scans using *k*-NN classifier.

Model Name	Accuracy	Precision	Recall	F1-score
Fine *k*-NN	99	0.99	0.49	0.66
Medium *k*-NN	98.4	0.97	0.50	0.66
Coarse *k*-NN	70.7	0.70	0.50	0.58
Cubic *k*-NN	98.4	0.97	0.50	0.66
Weighted *k*-NN	98.9	0.98	0.49	0.66

Table 12
Classification Results on CT scans using Ensemble classifiers.

MN	Accuracy	Precision	Recall	F1-score
A	58.7	0.64	0.32	0.42
B	99	0.99	0.49	0.66
C	51.2	0.51	0.47	0.49
D	99	0.99	0.49	0.66
E	56.7	0.73	0.18	0.29

Abbreviations: MN = Model Name, A = Boosted Trees, B = Bagged Trees, C = Subspace Discriminant, D = Subspace KNN, E = RUSBoost Trees.

Table 13
Classification (%) on CT scans using proposed method with NN model and its variants + PCA technique.

Model	Accuracy	Precision	Recall	F1-Score
Narrow NN	52.1	0.52	0.47	0.49
Medium NN	54.2	0.54	0.50	0.52
Wide NN	56.5	0.55	0.56	0.55
Bilayered NN	54.1	0.53	0.53	0.53
Trilayered NN	57.7	0.56	0.58	0.57

the proposed CT model using a 5-fold cross-validation configuration. Table 9 illustrates the average accuracy, precision, recall, and F1-Score of the proposed CT image model utilizing 5-fold cross-validation on CT scans employing different classifiers. Table 10, Table 11, and Table 12 show accuracy respectively using *k*-NN, ensemble methods, and SVM. Based on overall observations, *k*-NN and Ensemble classifiers provided higher accuracy than other classifiers. We used to calculate the performance measures based on the customization of the *k*-NN and ensemble learning techniques. Based on these customizations, the measurement accuracy is shown in Table 10, Table 11, and we can observe that the AdaBoosted with decision tree classifiers along with subspace *k*-NN methods perform accurate prediction of COPD as compared to other classification methods. In order to achieve improved accuracy, neural networks are used on prepared datasets for the classification of COPD cases with customized turning parameters and learning rates. We used the backpropagation method to measure the difference between the target and predicted accuracy via using the gradient descent learning method (see Eqs. (21)-(26)). However, the proposed method with a customized NN model does not provide higher accuracy for the classification of COPD with precision (56%), accuracy (57.70%), recall (58%), and F1-measure (57%) to perform well on it as shown in Table 13.

4.3.1. Lung sound analysis with cough based modality

We also built a classification model using CNN for the classification of lung sound with cough samples. We fed each feature from cough samples. We fed input into our cough/sound model in the form of generated spectrogram images of cough samples to extract MFCCs, and Chroma features for accurate prediction of COPD.

In lung sound signal based diagnostic model, we fed all the extracted features (MFCCs, Chroma, and its variant features of the lung sound signal samples (shown in Fig. 7 (Upper) and Fig. 7 (Lower) respectively) one by one and we computed the experimental results in the form of final class labels as COPD and non-COPD (healthy). Based on Table 14

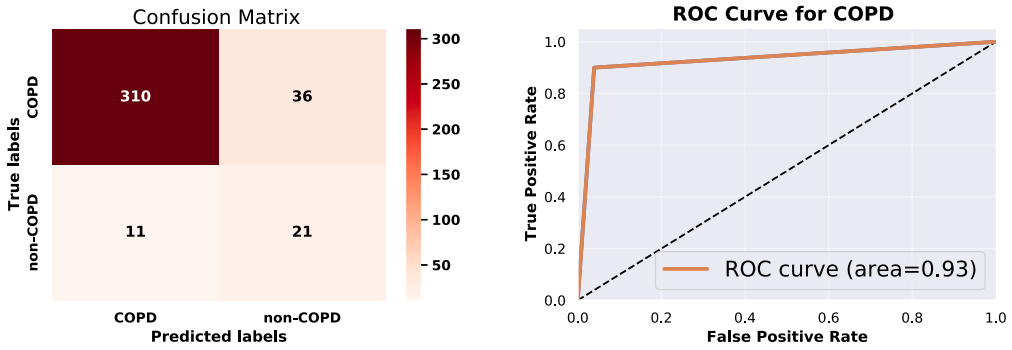

Fig. 9. (Left) confusion matrix and (Right) ROC curve obtained after feeding MFCC feature into CNN model.

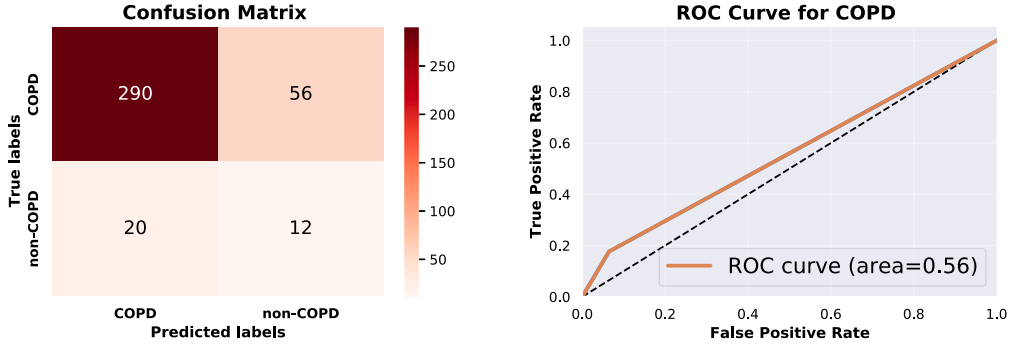

Fig. 10. (Left) confusion matrix and (Right) ROC curve obtained after feeding chroma_*cens* feature into CNN model.

Table 14
Classification metrics (%) of the proposed framework based on different features.

Feature	Accuracy	Loss	Sensitivity	Specificity
MFCCs	96.92	14.33	1	0.09
Melspectrogram	83.63	36.67	0.85	0.65
Chroma_stft	86.49	32.01	0.94	0.56
Chroma_cqt	77.26	46.0	0.962	0.43
Chroma_cens	80.34	41.25	0.79	0.48

Table 15
Classification metrics (%) of the proposed framework based on different sound features.

Feature	Precision	Recall	F1-score
MFCCs	0.94	0.99	0.97
Melspectrogram	0.94	0.96	0.95
Chroma_stft	0.977	0.893	0.933
Chroma_cqt	0.92	0.88	0.90
Chroma_cens	0.92	0.97	0.94

and Table 15, we observe that the sensitivity of the proposed cough model on MFCCs features and *chroma_cqt* features for analysis of lung sound with different cough (audio) is relatively high as compared to other extracted features, but on the other hand, their specificity is extremely poor as shown in Fig. 7, because the model performs a test to correctly identify people without the disease. When we fed all our features one by one to the proposed framework using CNN framework, our model provided a good accuracy of 93.50% when the proposed model fed MFCCs features, while other features allowed us to achieve less accuracy as compared to MFCC along with its loss comparison shown in Table 14. The Receiver Curve (ROC) are shown in Fig. 9, Fig. 10, Fig. 11, Fig. 12, and Fig. 13 [29]-[30].

Table 16
Lung regions based on intensity value.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.546	0.0589	-	-	-	-
TB	0.532	0.043	A	-	-	-
BI	0.558	0.042	-	C	-	-
Pn	0.509	0.047	B	-	C	-
COPD	0.504	0.051	C	B	C	-

TB = Tuberculosis, BI = Bronchiectasis infection, Pn = Pneumonia.

5. Statistical analysis

The discriminatory characteristics of clinical data are extracted from the chest X-ray/CT scan images to evaluate statistical significance measures in this study, as shown in Table 16, Table 17, and Table 18. Moreover, we used the standard bio-marker metrics based on extracted discriminatory features from CT images for early diagnosis and accurate prediction. The following are the standard bio-marker metrics measures from CT scan image analysis that are emphasized:

1. **Measuring the morphology structure:** Morphology structure analysis technique has been deployed segment CT scan images to find the infected lung areas, as shown in Fig. 14 (left) and Fig. 14 (center), it was evaluated on different classes.
2. **Computation of pixel intensity:** To validate relevant statistical measures of the segmented area of lung values, the statistical second-order ($\mu = \frac{1}{N}\sum_{i=1}^{N} CT_i$: mean value, ($\sigma = \sqrt{\mu_2}$, and variance) is analyzed (see Fig. 14(right)), where (N) is total number of CT image samples, μ defines the mean value computed from given CT sample.

$$Variance = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(CT_i - \mu\right)^2} \tag{35}$$

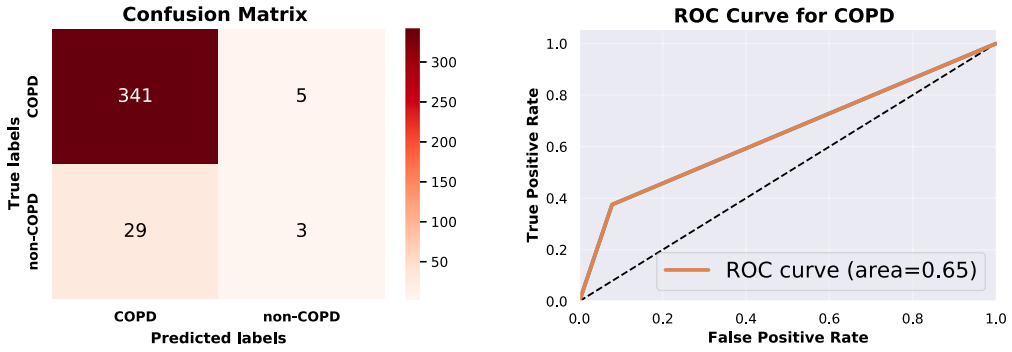

Fig. 11. (Left) confusion matrix and (Right) ROC curve obtained after feeding chroma_*cqt* feature into CNN model.

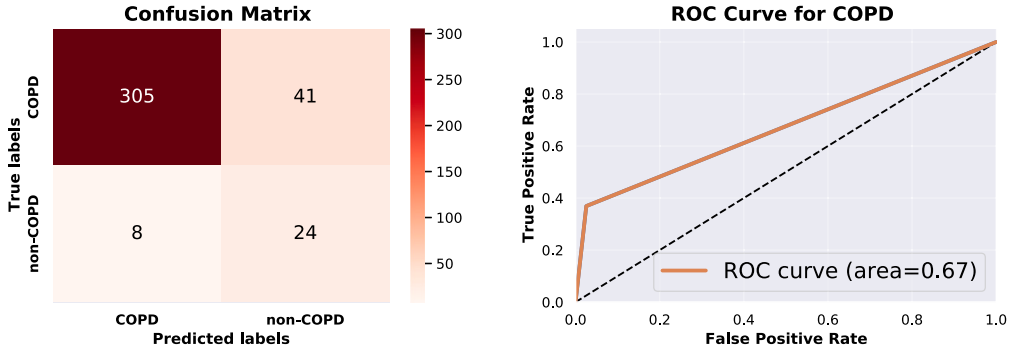

Fig. 12. (Left) confusion matrix and (Right) ROC curve obtained after feeding $chroma\_stft$ feature into CNN model.

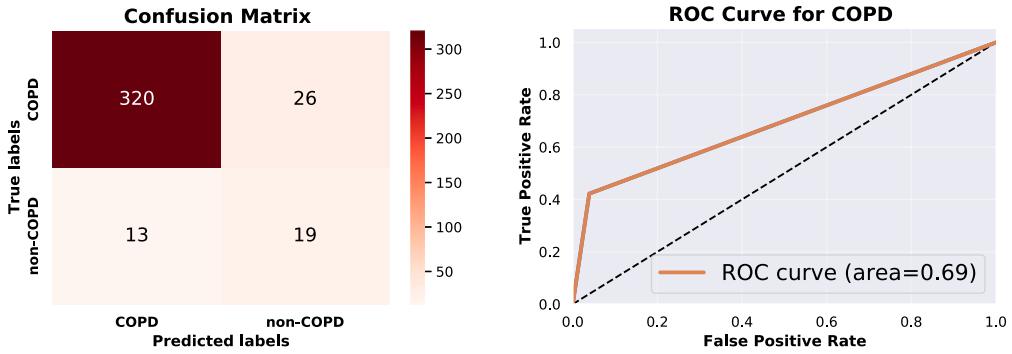

Fig. 13. (Left) Confusion Matrix curve based on melspectogram feature into CNN model, and (Right) ROC-AUC Curve obtained.

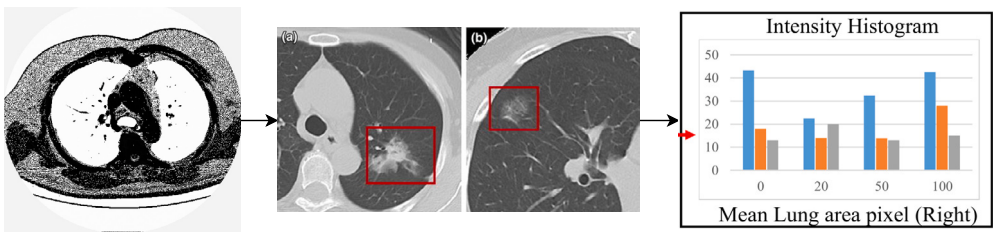

Fig. 14. (Left) Lung segmentation result, (center) contour detection in segmented CT images, and (right) Intensity histogram description of pixel from measured mean lung area values.

3. **σ-based similarity Analysis:** From the segmented images, we calculated the second-order statistic measure (σ). The segmented lung regions are considered for evaluation of the grey level intensity using weighted (χ^2)-square dissimilarity measure method based on histograms. The segmented images are used to consider the specified regions. Fig. 14(right) shows the black color-headed double arrow. In order to evaluate the histogram values of CT sample image features, Chi-square (χ^2) statistics as the dissimilarity measure is applied to find the match score values from each segmented regions smooth level of Gaussian pyramid. It is observed that CT sample image features consist of highly rich texture information. This information mainly lies in some regions of segmented CT images, infectious regions, and other regions. These texture patterns cater to more discriminatory information for accurately classifying COPD. Therefore, a weight can be set for each region of lung images based on the extracted discriminatory information it contains (see right of Fig. 14 (right) for an illustration). The weighted Chi-square X^2 dissimilarity measure is defined as follows (shown in Eq. (36))

$$\chi^2_w(s_1, s_2) = \sum_{i,j} w_{(i,j)} \frac{(s_{1(i,j)} - s_{2(i,j)})}{(s_{1(i,j)} + s_{2(i,j)})} \tag{36}$$

   Where s_1 and s_2 are two histogram values of the local binary pattern on Gaussian smooth level $[L_1]$ and $[L_2]$ of Gaussian filtering pyramid and W_j is the weight for lung regions $[j]$ of CT images.
4. **CTR:** We used the Cardiothoracic ratio (CTR) as a biomarker to identify cardiomegaly. It is defined as the ratio of measured values between the calculated maximal transverse cardiac diameter and the maximal internal thoracic diameter, as observed in annotated CT scan images. An off-average CTR value is used as an anomaly alert when the cardiothoracic boundary in COPD CXR is obscured due to rounded opacity or consolidation.
5. **Kolmogorov Smirnov test & Wilcoxon performance measure:** We assessed segmented CT images with obscured cardiothoracic borders due to opacities or consolidation to normalize possible biomarkers. Employing the Wilcoxon rank test, we examined the significance of chest pictures, comparing COPD candidates to other labels with varying data sizes [28]. Additionally, the rank-sum method evaluated segmentation performance using the Wilcoxon rank test, this time with equal data sizes [28].
6. **Statistical significance p-value thresholds:** We applied the statistical significance (SS) thresholds for computed values that are recommended for the p-test method to offer significant statistical measures for computed characteristics from segmented areas, i.e., p(A) $\leq$ 0.05, p(B) $\leq$0.01, and p(C) for p $\leq$ 0.001 and F1-measures.

Table 16 provides the related results utilizing CT scan pictures based on the statistical analysis approach. It displays SS values for several segmented photos of COPD patients' chests. For the scatter plot of the histogram, we computed the SS measures, as shown in Fig. 14 (c). The significant overlap regions between tagged CT chest image classes are depicted.

The lung area intensity variance analysis based on the pneumonia dataset and COPD (CT/CXR images) is summarized below. The pixel intensity of each divided lung region is used to calculate the standard deviation (σ). Table 16 represents the calculated values. We found that COPD patients had greater variance values than other labeled classes with SS (p $\leq$ 0.001), whereas viral cases have higher variance values than other labeled classes with SS (p$\leq$ 0.001). To evaluate the effect of scanning methodology on statistics, we studied the whole chest CT image collection with documented patient information, omitting chest Anterior-Posterior (AP) Supine radiographs. Because the AP Supine procedure is not commonly employed in typical cases, it is used as a substitute for conventional chest Anterior-Posterior (AP) or Posterior Anterior (PA) radiographs, depending on the patient's condition. We conducted an analysis based on the data from the Table 17 and Table 18. These

Table 17
Lung regions based on mean intensity value.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.159	0.028	-	-	-	-
TB	0.136	0.019	-	-	-	-
BI	0.144	0.025	-	-	-	-
Pn	0.163	0.026	C	C	B	-
COPD	0.162	0.024	C	C	C	A

Table 18
Lung intensity-based variance statistics.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.139	0.027	-	-	-	-
TB	0.136	0.017	-	-	-	-
BI	0.145	0.018	A	B	-	-
Pn	0.165	0.023	C	C	C	-
COPD	0.163	0.022	C	C	C	-

Table 19
Variance based on inter- and intra-class.

Class	Mean	STD	Normal	TB	BI	VI
Normal	0.446	0.051	-	-	-	-
TB	0.476	0.078	A	-	-	-
BI	0.472	0.074	-	-	-	-
Pn	0.502	0.064	C	A	-	-
COPD	0.499	0.068	C	C	A	-

C: infection of COPD.

Table 20
Fusion technique-based multimodal accuracy.

Model name	Weight	Accuracy	WS
CXR model (M1)	0.54	98%	52.92
Lung model (M2) M2	0.46	96.92%	44.58
Fused Accuracy (%)	-	-	**97.50**

M1: CT/CXR model, M2: Lung sound/ Cough (audio), ((W_i), A = Accuracy (S_i), WS = Weighted sum fusion score.

results are compared by different classification methods, as shown in Table 16, Table 17, and Table 19, respectively. Statistical measures technique reveals variations in mean values $\mu = [mu_1 \cdots, \mu_n$ and standard deviation (σ) values from calculated histogram responses of segmented lung images based on overall observations. Based on the measured metrics, the proposed framework distinguishes COPD patients, tuberculosis (TB), bacterial infection, and other Respiratory Tract Infections.

The computed accuracy shows that the statistical mean and STD values of COPD and viral classes demonstrate significant differences based on highly intensity-variable features in segmented lung regions, according to significant measurements. We did not account for dynamic changes in the scanning methodology. In addition, the categorization of COPD and other viral cases demonstrates a substantial difference based on overall observation. Despite statistical disparities between COPD patients and other disease classes, the chosen parameters (confidence intervals) are $p \leq 0.001$ for Normal, and TB, $p \leq 0.05$ for bronchitis, and there are significant overlaps across numerous disease groups.

Based on characteristics collected from lung areas, the approach classifies bronchitis infection, TB, COPD, and other Respiratory Tract Infections (RTI). A statistical model is used to assess the statistical significance of extracted characteristics, as described in Table 20. Statistical analysis was conducted on potentially derived characteristics from chest CT images, aimed at identifying significant alterations in the pixel intensity distribution of segmented lung images to enable accurate categorization. The study revealed that the precise distribution within the segmented lung region displayed distinct patterns that could be mod-

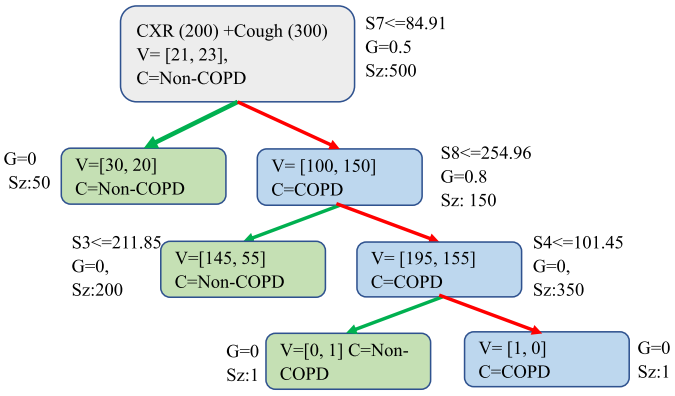

Fig. 15. RF classifier model.

ified. Consequently, these identified characteristics prove to be highly valuable for early diagnosis and precise prognosis based on CXR examinations.

5.1. Analysis of inter- and intra-class lung images

In our study, we analyzed chest CT images to calculate local and global pixel intensity distributions using features from segmented images. Our goal was to predict early diagnosis of COPD patients, identifying the best discriminant between inter-class and intra-class lung image databases. We computed the mean intensity for each lung imaging patch and the standard deviation (σ) for early COPD detection, defining these as intra-patch and inter-patch intensity distributions, respectively.

The lower intensity values ($p \leq 0.001$ for all) of other classes have extremely intensity-variant features, as shown by a significant error bar in the distribution of inter-patch intensity of the unified COPD and other Respiratory Tract Infection class. For the study of inter-patch and intra-patch lung pictures, the accuracy is tested using distinct Regions of Interest (RoI) of lung images and calculated histogram responses of each RoI. We found no changes between the normal and intra-patch intensity-based pixel distribution while studying intra-patch intensity-based pixel distribution ($p \geq 0.05$). Therefore, based on pixel variance, the intra-class image patches contain local textural information for categorizing chest CT images (COPD and non-COPD) based on overall observations. We calculated global features and multifocal intensity from segmented pictures, and the differences can be used as discriminating characteristics for COPD diagnosis in individuals who have had post-COPD infections. It has an excellent link to CT image analysis and is statistically significant as shown in Table 17.

5.2. Analysis of ensemble learning classification models

The proposed ensemble approach consists of two levels: the first level utilizes decision trees, random forests, logistic regression, and boosted and bagged learning models with the majority voting to make predictions. These models serve as the base-level structures in the ensemble. In the second level, the predictions from the base-level models are used as feature inputs to accurately predict COPD patients based on CT scans and cough samples. To classify participants as positive or negative for COPD, a 350-tree ensemble classifier is employed. Additionally, a bagging model with 350 decision trees, each with a maximum depth of 4, is utilized (see Fig. 15). For adaptive boosting, the AdaBoost M1 method with 350 decision stumps (one-level trees with two leaf nodes) is applied. The ERLX model's second level utilizes an extreme gradient boosting (XGBoost) classifier, which is a decision tree-based ensemble method with a gradient-boosting framework.

It is observed that the performance of the proposed fusion model is lower because of the availability of inter-class and intra-class samples of chest-X ray/CT scan images and cough sample. These images and cough sample have not provided discriminatory features for training the individual model for COPD diagnosis (shown in Table 20. For all samples of all classes (COPD and non-COPD), we measure the inter-class scatter matrix (SB) and the intra-class scatter matrix (SW) to improve the overall accuracy of the proposed fusion model. Therefore, we used the linear discriminant analysis (LD) method to maximize inter-class scatter matrix(SB) (i.e., ratio $det|SB|/det|SW|$) while minimizing matrix (SW) for selecting prominent features for fusion of individual model. The measured ratio has been maximized when the column feature vectors of the projection matrix are the eigenvectors of ($SW^{-}1\not\supseteq SB$) for the classification of COPD and non-COPD.

5.2.1. Impact on defined clinical problem and motivation

After post-COVID-19, the level of infection severity of pulmonary respiratory diseases is spreading rapidly worldwide. India, one of the developing countries, has more than 8.60% of the tribal population. The tribal population is highly dependent on minor products (i.e., honey bees, fruits, leaf products, and others) from the forests, and they live in forested areas and earn lower income from buying and selling these minor products in the local markets in the tribal community. There is an unavailability of efficient solutions for early diagnosis and accurate prediction of critical diseases of tribal patients. Moreover, the population of the tribal people lags in availing emerging healthcare facilities for pulmonary respiratory disorders, which includes several health indicators, with an enormous number of women and children being the most vulnerable to severe infection of pulmonary diseases. In the tribal population, it is reported that the total fertility rate includes more than 2.5%, and only 15% complete their hospital visits. Instead, what is more disheartening is the lack of availability of maternal mortality ratio (MMR) data containing the tribal population in India. It is reported that 10 points have declined the ratio of MMR data (dropped from 113 (2016-18) to 103 (2017-2019). Moreover, in tribal communities, children (10-35 years) suffer from pulmonary diseases, including full immunization medical ratio (IMR) coverage that is reported at only 56%, which is reflected in IMR as 44.40%, and the Under-five mortality rate is reflected by 57.20%. The tribal population suffers from the triple burden of critical diseases, namely, communicable diseases, non-communicable diseases, mental health, Bronchiectasis infection, malnutrition, and addictions. More than 8.60% of the tribal population in India is composed of 30%-40% of all cases of pulmonary diseases, including malaria. The estimated prevalence of communicable diseases (TB) (per 100,000) reported 703 points against 256 in the non-tribal population. The overall percentage of the children population reported as underweight, which is depicted as more than 56%, and about 89% of the total population under-five children are reported as suffering from COPD and anemia. It has been shown that severe deaths from malnutrition and children in spurts are depicted mostly during different seasons. Moreover, more than 50%-80% tribal girl populations counted as lower body mass index measured population less than 18.5, and about 60%-75% of the tribal women population (from years 15–50) reported severity of COPD population from anemia (non-tribal 50%) in each period cycle. Hence, there is a need to design and develop efficient and fast diagnostic solutions to bridge this gap between tribal and non-tribal populations for healthcare. In this work, the proposed multimodal framework provides efficient solutions for early diagnosis of multipulmonary diseases such as Tuberculosis (TB), Bronchiectasis infection (BI), Pneumonia (Pn), and COPD diseases using ensemble learning techniques (shown in Table 16). The primary motivation of proposed solutions is to leverage the diagnostic of clinical data analysis of different diseases based on extracted discriminatory characteristics from chest-X-ray/CT and lung/cough (audio) samples using statistical analysis.

5.2.2. Clinical relevance

The efficient solutions of the proposed multimodal framework focus on the detection of Tuberculosis (TB), Bronchiectasis infection (BI), Pneumonia (Pn), and COPD patients at an early stage based

on accurate observation of collected chest-X-ray/CT scan images, cough(voice)/Lung sound sample dataset. We collected patient clinical data from tribal people based on taken consents. We proposed a multimodal framework that performs different statistical measures, including FEV FVC values with stage 0 (at risk) (FEV1/FVC≤70), where stage 1 defines the mild stages that depict evaluation of the ratio where FEV1≤80% of the predicted value. The measured percentage (FEV1/FVC≤ 70%) signifies that the patients may have severe symptoms. In the early diagnosis of patients at stage 2, patients are declared moderate- that is, lies between 50% ≤FEV1≤80% of predicted value and ratio depicts that the ratio FEV1/FVC≤70% and 30%≤FEV1≤50% of the expected value of COPD diseases. Moreover, patients with stage 3 depict that patients are in the severe stage, where the measured ratio of FEV1/FVC≤70% of the predicted values declared that patients are at high risk for COPD and stage 4 (very severe), where FEV1≤30% of the expected value or FEV1≤50% of the predicated value+ severe chronic symptoms for early detection of COPD patients. The clinical relevance of our proposed solutions is in enhancing early COPD diagnosis and subtype disease classification, improving patient outcomes, and reducing the burden of this debilitating disease. We understand that clinical viability is critical and not appropriately defined, and we are committed to conducting further validation studies, collaborating with clinicians and pulmonary experts, and ensuring that our proposed solutions meet the rigorous standards and trustworthiness expected by healthcare professionals.

5.3. Comparison with existing work

Finally, we compare the proposed framework's performance to other COPD diagnostic and prognosis systems based on CT scan and cough audio sample databases. Most COPD health monitoring systems and IoT enabled devices or models diagnose COPD using standard processes and protocols, as illustrated in Table 21. According to our findings, no study based on the multimodal fusion-based technique for early detection of COPD using CT scan and cough samples has been described in the literature (see Table 21 and Table 22). This study found the proposed framework paradigm for early COPD detection on various datasets. It is, nevertheless, comparable enough to establish reasonable conclusions about early COPD diagnosis in different benchmark settings (see Table 21 and Table 22).

5.4. Application of proposed framework

Based on various available open datasets of COPD patients' CT and cough/lung sounds, we conducted the experiments to compute the overall accuracy of the proposed system separately for early diagnosis. However, there is no availability of complete CT scan images, clinical characteristics datasets, or cough/lung sound datasets of COPD patients in the public domain to perform computation for early prediction and prognosis of COPD. There may be severe problems in preparing the complete CT scan as pulmonary respiratory databases due to changes in the patient's physiological and human respiratory system with different standard settings. In selecting the study population, criteria may be changed for different countries (e.g., Turkey and the USA are among the countries with the highest prevalence of COPD. In the BOLD-Adana Turkey COPD study, they studied the prepared CT scan, cough samples along with the prevalence of COPD; it was found 19.10% (28.50% in men and 10.30% in women) with a fixed ratio (FEV1/FVC≤70%) in adults over 40 years of age.

To solve the early COPD diagnosis problem of the tribal population, we considered the age of diagnosis ≥ 35 years to ≤ 80 years. Diagnosis of COPD within the previous 3 months or after the initiation visit. Diagnosis of COPD was based on a post-bronchodilator forced expiratory volume (FEV) in 1s (FEV1)/forced vital capacity (FVC) ratio of ≤ 0.7. We considered the case where in the absence of a primary diagnosis of bronchiectasis, asthma, or any other critical respiratory disease. The patients also had a smoking history of at least 10 packs/year, family disease history, lifestyle, behavioral changes, or occupational exposure to irritants, toxic gases, or biomass. The patients (or their legal representatives) who accepted to sign the informed consent form.

Table 21
Comparison of proposed cough based model.

Ref.	Technique	Acc	F	R
[24]	A	86%	30 Sounds	Low Acc
[25]	DBN	93.84%,	120	High error
[26]	DL	70.28%	DB	Low
[28]	A+FL	91%	C	Less
[27]	DT	87.40%	60%	Low accuracy
Prop.	**CNN [MFCC]**	**95.30**	**Y**	**Z**

Table 22
Comparison of existing techniques and the proposed framework based on CT scan images.

Ref.	Technique	Acc(%)	F-1
[20]	ML	91%	0.78
[21]	SVM	0.742	NA
[22] [16]	*k*-NN & SVM	81.7.	NA
Prop.	***k*-NN+PCA**	**98.4**	**Y**

Abbreviation: R = Remark, F = Features, TU = Technique Used, P = multiple instance learning (MIL), O = one classifier, Q = Larger datasets are not used Y = Medium distinctions between classes with k = 9, NA = Not Available.

The prepared COPD dataset was captured from the tribal people living in forest areas (e.g., 33% of the population in Chhattisgarh (India) is living in tribal areas without the availability of a preliminary level of required medical facilities for tribal people. We prepared the CT scan, cough/lung sound, and clinical characteristics datasets from AIIMS, Raipur, Chhattisgarh, India. The proposed framework is evaluated based on captured datasets using CT scans as well as cough diagnostic models to perform early diagnosis and prediction of COPD patients. The CT scan and cough model reported accuracy of 95.30% and 97.50% for early diagnosis. The proposed novel framework integrates the accuracy of both models with 98% accuracy using ensemble learning techniques on CT and cough sample datasets. We can find that the audio-based model shows similar significance compared to the CT scan-based model for early COPD prediction. Based on the literature work, the CT scanning method is a well-known method for early diagnosis of COPD and other critical diseases in the medical field in comparison to other diagnostic models. However, CT scanning models are costly and require substantial computational resources to process the highly dimensional CT scan dataset for early prediction, which may not be available for fast diagnosis of tribal people. Therefore, cough (audio) samples are taken as input for validating the working of the proposed model. The cough samples are collected through user smart devices and stored in the database to process further collected audio datasets for early COPD diagnosis.

5.5. Complexity analysis

The proposed multimodal framework builds two classification models ((1) CT-scan-based model, and (2) cough (voice)/lung sound based model) using random forest based incremental decision tree and Convolutional Neural Network (CNN) which build major fraction of the space and time complexity. The proposed lung sound recognition model uses incremental decision tree algorithm that takes order of $O(MN^2)$ time to form a ensemble decision tree of chest and cough/lung sound signal features using C4.5 algorithm [37], where M and N depict number of discriminatory features of chest-X-ray and lung sound signal, re-

spectively. The constructed decision tree has been updated in O(h×M) time as soon as new combined features instance arrives. The term h shows the height of the constructed tree (see Fig. 15). Moreover, the time complexity of proposed multimodal using CNN is illustrated as $O(\sum_{i=0}^{D} P_{i-1} \cdot F_i^2 \cdot f_i \cdot X_i^2)$ [22], where terms D, and P show the total number of convolution layers, input channels. Moreover, F and f are defined as the spatial size of filter, and required total number of filters. The term X defines size of output feature map, at i^{th} layer. As incremental learning algorithm and CNN architecture have been implemented on various health assistive devices for early diagnosis, these diagnostic models can be implemented and properly executed in parallel. Hence, overall time complexity of implemented algorithms $O(P \times F^2.f.X^2)$ as time complexity of customized CNN is higher than the incremental learning method.

6. Conclusion and future work

In this paper, we proposed a novel multimodal framework for early diagnosis and classification of COPD and multivariate pulmonary respiratory diseases such as, Tuberculosis (TB), bronchitis, COPD, LRTI, and URTI. The proposed framework provided an accuracy of 98% (CT scanning model) and 95.30% (cough-based model), respectively. To improve the overall accuracy, the integrated novel multimodal framework fused CT image features and cough feature to provide 97.50% using the weighted sum-rule fusion method for early diagnosis of COPD patients. The multimodal extracted features such as, texture Gaussian space features, chroma, morphological features, and histogram features, structural features of the prepared CT images and cough sample datasets. Moreover, we evaluated the statistical significance measures based on extracted pulmonary morphological structure, mean intensity, and the cardiothoracic ratio of segmented parts of CT scan image features from the prepared database using statistical machine learning techniques for analysis of different multivariate pulmonary respiratory diseases for tribal people in Chhattisgarh, India. We considered the patient's age for diagnosis, which is $\geq$ 35 years and $\leq$ 80 years. It was observed that the age of early and fast diagnosis is between 35-40 years, and more than 40 years, a patient. Based on overall observations, a group of patient with age $\geq$ 80 years, the severity of COPD is higher due to smoking and other health parameters. Furthermore, it is challenging to diagnose COPD early for accurate prediction based on the collected CT scan and cough/lung sound sample datasets from the previous 3 months or after the initiation visit to the Department of Pulmonary Medicine and TB, AIIMS Raipur. To solve these issues, we collected the clinical characteristics of COPD patients for early diagnosis of COPD based on a post-bronchodilator forced expiratory volume in one second and (FEV1)/Forced Vital Capacity (FVC) ratio-based computations. In addition, we considered the ratio of FEV1/FVC $\leq$ 0.70 for the analysis of experimental results. The experimental results of the proposed framework are compared with the existing methods on different open datasets of CT scan images and cough samples on standard benchmark settings for early diagnosis and classification. The benefits of the proposed multimodal framework are fast, cost-effective, and remotely accessible for early diagnosis of critical diseases (i.e., TB, bronchitis, COPD, LRTI, and URTI), i.e., a user in testing phase needs only to record cough (audio) sample and transfer the multivariate pulmonary respiratory cough (voice) sample using microphone of their smart devices (smartphone, smartwatch, smart tablet) to the health monitoring server for the detection and accurate predication of respiratory diseases. The efficient solutions can be utilized to produce holistic health diagnostic report based on transferred cough sample datasets from patients those are living in the remote and tribal regions. Moreover, the multimodal framework integrates individual cough and CT diagnostic models that may take more time to process the collected respiratory cough (voice)/CT scan sample at server. The concluding remarks about the findings in this study is that the no work is reported in literature for development of a multimodal framework for early diagnosis of COPD using deep learning techniques. In addition, we believe that our proposed multimodal solutions will make it possible to build a fast diagnostic system that can automatically detect COPD severity from cough (voice)/lung sound signal (auscultations) in actual clinical settings for upliftment of society. In the future, we plan to extend the proposed novel framework for deployment in tribal regions of India for early diagnosis of COPD patients in real time. Furthermore, we will develop a multiclass framework for solving multivariate pulmonary sound analysis and other respiratory disease classifications and their prognosis using deep learning techniques. It may help build a platform to motivate interdisciplinary researchers, scientists, engineers, and medical practitioners to analyze critical diseases. Moreover, the research outcomes may help the Indian medical departments to perform early diagnosis and accurate prediction of non-communicable and communicable diseases. The early diagnosis solution can be deployed even in tribal regions at the early stages of various critical disease detections in dense tribal areas and take immediate action to stop the spread of infections to save the lives of tribal people and other Indian citizens.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

There will no availability of data for research work.

Acknowledgements

This publication has emanated from research conducted with the financial support of the Science and Engineering Research Board (SERB) India under award numbers (Sanction Order No. EEQ/2022/000617), (SERB Finance No: SERB/F/11403/2022-2023). Furthermore, this publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number SFI/12/RC/2289_P2. Partially supported by the POS RADIOAMICA project funded by the Italian Minister of Health (CUP: H53C22000650006).

The authors would like to thank experts and staff of Department of Pulmonary Medicine & TB, All India Institute of Medical Sciences (AIIMS), Raipur, Chhattishgarh, India, for their guidance and collaboration, and for coordination of the pulmonary specialists and patients for this research work.

References

[1] K.P. Exarchos, A. Aggelopoulou, A. Oikonomou, T. Biniskou, V. Beli, E. Antoniadou, K. Kostikas, Review of artificial intelligence techniques in chronic obstructive lung disease, IEEE J. Biomed. Health Inform. 26 (5) (2021) 2331–2338, https://doi.org/10.1109/JBHI.2021.3135838.

[2] Himanshu KumarSahu, Santosh Kumar, Saeed Hamood Alsamhi, Mithilesh Kumar Chaube, Edward Curry, Novel framework for Alzheimer early diagnosis using inductive transfer learning techniques, in: 2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA), IEEE, 2022, pp. 1–7.

[3] Q. Zhao, J. Li, L. Zhao, Z. Zhu, Knowledge guided feature aggregation for the prediction of chronic obstructive pulmonary disease with Chinese EMRs, IEEE/ACM Trans. Comput. Biol. Bioinform. (2022), https://doi.org/10.1109/TCBB.2022.3198798.

[4] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, Giancarlo Fortino, A survey on deep learning in medicine: why, how and when?, Inf. Fusion 66 (2021) 111–137.

[5] C.Y. Chin, M.Y. Weng, T.C. Lin, S.Y. Cheng, Y.H.K. Yang, V.S. Tseng, Mining disease risk patterns from nationwide clinical databases for the assessment of early rheumatoid arthritis risk, PLoS ONE 10 (4) (2015) e0122508.

[6] A. Badnjevic, L. Gurbeta, E. Custovic, An expert diagnostic system to automatically identify asthma and chronic obstructive pulmonary disease in clinical settings, Sci. Rep. 8 (1) (2018) 1–9.

[7] Lidia Fotia, Flávia Delicato, Giancarlo Fortino, Trust in edge-based internet of things architectures: state of the art and research challenges, ACM Comput. Surv. 55 (9) (2023) 1–34.

[8] Abdu Gumaei, Mohammad Mehedi Hassan, Md Rafiul Hassan, Abdulhameed Alelaiwi, Giancarlo Fortino, A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification, IEEE Access 7 (2019) 36266–36273.
[9] A. Alvar, M. Decramer, P. Frith, Global initiative for chronic obstructive lung a guide for health care professionals global initiative for chronic obstructive disease, Global Initiat. Chronic Obstr. Lung Dis. 22 (4) (2010) 1–30.
[10] Amir Amini, Wei Chen, Giancarlo Fortino, Ye Li, Yi Pan, May Dongmei Wang, Editorial special issue on "AI-driven informatics, sensing, imaging and big data analytics for fighting the COVID-19 pandemic", IEEE J. Biomed. Health Inform. 24 (10) (2020) 2731–2732.
[11] Konstantinos P. Exarchos, Agapi Aggelopoulou, Aikaterini Oikonomou, Theodora Biniskou, Vasiliki Beli, Eirini Antoniadou, Konstantinos Kostikas, Review of artificial intelligence techniques in chronic obstructive lung disease, IEEE J. Biomed. Health Inform. (2021), https://doi.org/10.1109/JBHI.2021.3135838.
[12] H.J. Davies, P. Bachtiger, I. Williams, P.L. Molyneaux, N.S. Peters, D. Mandic, Wearable in-ear PPG: detailed respiratory variations enable classification of COPD, in: IEEE Transactions on Biomedical Engineering, 2022.
[13] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1) (1996) 51–59.
[14] A. Abid, R.J. Mieloszyk, G.C. Verghese, B.S. Krauss, T. Heldt, Model-based estimation of respiratory parameters from capnography, with application to diagnosing obstructive lung disease, IEEE Trans. Biomed. Eng. 64 (12) (Dec. 2017) 2957–2967.
[15] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (July 1997) 711–720.
[16] Santosh Kumar, Mithilesh Kumar Chaube, Saeed Hamood Alsamhi, Sachin Kumar Gupta, Mohsen Guizani, Raffaele Gravina, Giancarlo Fortino, A novel multimodal fusion framework for early diagnosis and accurate classification of COVID-19 patients using X-ray images and speech signal processing techniques, Comput. Methods Programs Biomed. 226 (2022) 107109.
[17] R.J. Mieloszyk, et al., Automated quantitative analysis of capnogram shape for COPD–normal and COPD–CHF classification, IEEE Trans. Biomed. Eng. 61 (12) (Dec. 2014) 2882–2890.
[18] Gokhan Altan, Yakup Kutlu, RespiratoryDatabase@TR (COPD Severity Analysis), in: Mendeley Data V1, 2020.
[19] V. Cheplygina, I.P. Peña, J.H. Pedersen, D.A. Lynch, L. Sørensen, M. de Bruijne, Transfer learning for multicenter classification of chronic obstructive pulmonary disease, IEEE J. Biomed. Health Inform. 22 (5) (Sept. 2018) 1486–1496.
[20] I. Mohamed, M.M. Fouda, K.M. Hosny, Machine learning algorithms for COPD patients readmission prediction: a data analytics approach, IEEE Access 10 (2022) 15279–15287.
[21] Y. Nagaraj, H.J. Wisselink, M. Rook, J. Cai, S.B. Nagaraj, G. Sidorenkov, R. Veldhuis, M. Oudkerk, R. Vliegenthart, P. van Ooijen, AI-driven model for automatic emphysema detection in low-dose computed tomography using disease-specific augmentation, J. Digit. Imag. (2022) 1–13.
[22] J. Yang, et al., Novel subtypes of pulmonary emphysema based on spatially-informed lung texture learning: the multi-ethnic study of atherosclerosis (MESA) COPD study, IEEE Trans. Med. Imaging 40 (12) (Dec. 2021) 3652–3662.
[23] L. Sørensen, P. Lo, H. Ashraf, J. Sporring, M. Nielsen, M.D. Bruijne, Learning COPD sensitive filters in pulmonary CT, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2009, pp. 699–706.
[24] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, R.R. Fletcher, Application of semi-supervised deep learning to lung sound analysis, in: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 804–807.
[25] G. Altan, Y. Kutlu, A.Ö. Pekmezci, S. Nural, Deep learning with 3D-second order difference plot on respiratory sounds, Biomed. Signal Process. Control 45 (2018) 58–69.
[26] G. Altan, Y. Kutlu, N. Allahverdi, Deep learning on computerized analysis of chronic obstructive pulmonary disease, IEEE J. Biomed. Health Inform. 24 (5) (May 2020) 1344–1350.
[27] M.A. Fernandez-Granero, D. Sanchez-Morillo, A. Leon-Jimenez, An artificial intelligence approach to early predict symptom-based exacerbations of COPD, Biotechnol. Biotechnol. Equip. 32 (3) (2018) 778–784.
[28] R. Vilarinho, L. Serra, A. Águas, C. Alves, P.M. Silva, C. Caneiras, A.M. Montes, Validity and reliability of a new incremental step test for people with chronic obstructive pulmonary disease, BMJ Open Respir. Res. 9 (1) (2022) e001158.
[29] G. Altan, Deep learning-based mammogram classification for breast cancer, Int. J. Intell. Syst. Appl. Eng. 8 (4) (2020) 171–176.
[30] G. Altan, DeepOCT: an explainable deep learning architecture to analyze macular edema on OCT images, Eng. Sci. Technol. Int. J. 34 (2022) 101091.
[31] J. Salamon, J.P. Bello, Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Signal Process. Lett. 24 (3) (2017) 279–283.
[32] R.R. Selvaraju, M. Cogswell, A. Das, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359, https://doi.org/10.1007/s11263-019-01228-7.
[33] G. Altan, Y. Kutlu, A. Gökçen, Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds, Turk. J. Electr. Eng. Comput. Sci. 28 (5) (2020) 2979–2996.
[34] G. Altan, Computer-aided diagnosis system for chronic obstructive pulmonary disease using empirical wavelet transform on auscultation sounds, Comput. J. 64 (11) (2021) 1775–1783.
[35] A. Roy, U. Satija, A novel melspectrogram snippet representation learning framework for severity detection of chronic obstructive pulmonary diseases, IEEE Trans. Instrum. Meas. 72 (2023) 1–11.
[36] D.M. Huang, J. Huang, K. Qiao, N.S. Zhong, H.Z. Lu, W.J. Wang, Deep learning-based lung sound analysis for intelligent stethoscope, Mil. Med. Res. 10 (1) (2023) 1–23.
[37] J. Su, H. Zhang, A fast decision tree learning algorithm, in: Proc. AAAI Conf. Artif. Intell., 2006, pp. 500–505.