Chronic Obstructive Pulmonary Disease (COPD) Detection, Audio/Sound Classification Using Deep Learning

Shubham Raj[1], Sandeep Vishwakarma[1]

[1] AI&ML Dept, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India

ABSTRACT

The advent and success of developers and researchers in the field of "Deep Learning", has led to the development and success of various algorithms used in various fields to achieve excellent results. This has led to the use of deep learning algorithms in the medical field to address real-world health-related problems thus providing a variety of exciting solutions with precision in medical approaches with sound analysis to diagnose various diseases For the detection of disorders in the health care industry, a quick and inexpensive method like audio analysis is required. A number of machine learning and deep learning algorithms were tested, examined, and implemented to see if they could help clinicians identify COPD by giving quicker and more accurate results in the face of so many limitations and the development of artificial intelligence. Airflow from the lungs becomes restricted due to the chronic inflammatory lung illness known as chronic obstructive pulmonary disease (COPD). However, there is very little support for voice features in the diagnosis of COPD. In this work, we propose a way to analyse breathing pattern, continuous wheezing, crackling sound, frequency, duration, initial deflection width, from speech (acoustics) data of patients to predict if the patient has COPD or not. In this project we aim to develop a modern system that uses CNN algorithm to detect COPD(Chronic obstructive pulmonary disease).

Keywords: COPD, CNN, mfcc, Mel-spectrogram, continuous sound, discontinuous sound, Emphysema, Chronic Bronchitis, Irreversible Asthma, Bronchiectasis.

INTRODUCTION

Airflow blockages are brought on by the lung illness known as chronic obstructive pulmonary disease (COPD)[1]. Wheezing, coughing up mucus (sputum), and breathing difficulties are all symptoms. Long-term exposure to irritating gases or particles, most often cigarette smoke, is the most common cause. Lung cancer, heart disease, and a variety of other illnesses are more common in COPD patients. The most common diseases that lead to COPD are emphysema and chronic bronchitis. These two conditions frequently coexist and can range in severity among COPD patients. Airflow blockages are brought on by the lung illness known as chronic obstructive pulmonary disease (COPD). Wheezing, coughing up mucus (sputum), and breathing difficulties are all symptoms. exposure to unpleasant gases or particles for a long time, frequently Two tubes in your windpipe (trachea) carry air into and out of your lungs (bronchi). These tubes, which resemble tree branches, split into a variety of smaller tubes (bronchioles) inside of your lungs, where they finally assemble as groups of tiny air sacs (alveoli). The air sacs have exceedingly thin walls that are dotted with tiny blood arteries (capillaries). Your bloodstream may receive oxygen from the air you breathe thanks to these blood veins. Carbon dioxide is exhaled at the same moment. To expel air from your body, your lungs rely on the bronchial tubes' and air sacs' innate elasticity. COPD makes them inflexible and overexpand, which causes some air to be held back in your lungs during exhalation. For many years, a complex, nonsensical vocabulary was used to describe respiratory noises. In order to create a more objective naming system, an ad hoc committee of the International Lung Sounds Association decided on a nomenclature that divided adventitious sounds into two primary classes in 1985: continuous wheezes and discontinuous sounds or crackling noises. These acoustically stated terms (such frequency, duration, initial deflection width, etc.) do not suggest a generating mechanism or location. The leading cause of COPD in developed countries is tobacco smoking. In developing countries, people are more prone to develop COPD if they are exposed to the fumes from burning fuel for cooking and heating in insufficiently ventilated dwellings. Even though many smokers with a long smoking history may have diminished lung function, very few chronic smokers acquire clinically evident COPD. Rarer pulmonary diseases can occur in some smokers. They might not actually have COPD until a more thorough examination is performed.

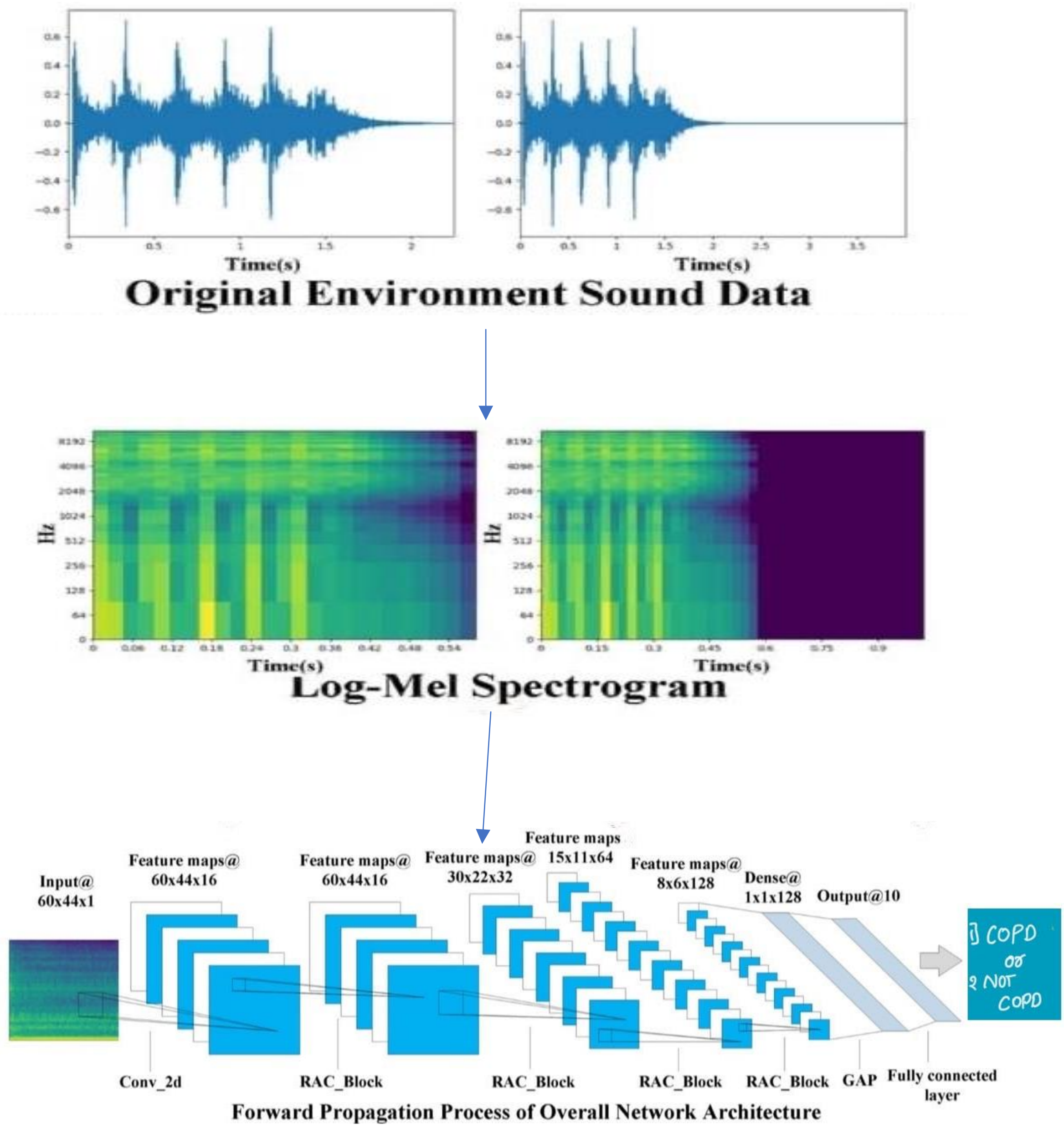
Biology Terms:

The fragile walls and elastic fibres of the alveolar walls are destroyed by the lung illness emphysema. Small airways constrict during exhalation, decreasing airflow from your lungs. Bronchitis is a chronic condition. In this sickness, your bronchial tubes swell and narrow, and your lungs produce more mucus, which can further clog the narrowed tubes. You begin to cough repeatedly in an effort to clear your airways. In the vast majority of COPD patients, long-term smoking causes the lung damage that results in the disease. However, since not all smokers get COPD, it is likely that additional factors, such as genetic susceptibility to the condition, are at play. Additionally, COPD sufferers are more likely to experience exacerbations, which are times when their symptoms increase and persist longer than usual. The vast majority of cases are directly related to smoking cigarettes, so the best way to prevent COPD is to never smoke or to stop doing so immediately away. These simple phrases might not seem so simple if you've been a smoker for a long time, especially if you've made past attempts to stop smoking - once, twice, or numerous times. But keep trying to give up. It's crucial to choose a programme that can help you stop smoking cigarettes permanently. Your best chance of avoiding lung injury is through it.

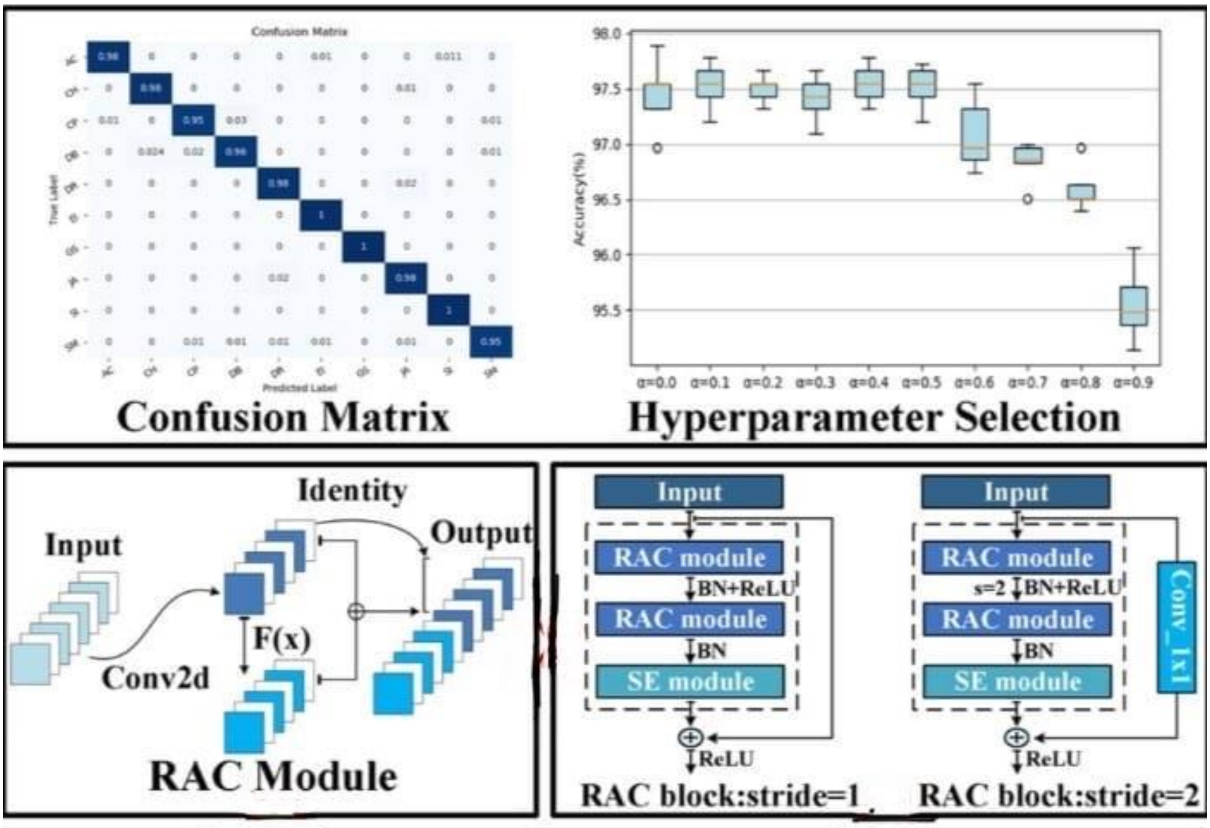
The work exposure to chemical dusts and fumes is another risk factor for COPD. Talk to your manager about the best ways to protect yourself if you work with these types of lung irritants, such as by donning respiratory protection.

RELATED WORK

Emphysema is a lung condition that damages the alveolar walls' brittle walls and elastic fibres. When you exhale, your small airways close off, reducing the amount of airflow from your lungs. Chronic bronchitis is a disease. Your bronchial tubes enlarge and narrow as a result of this illness, and your lungs also create extra mucus, which can further clog the already-narrowed airways. You start to cough continuously, trying to open up your airways. Long-term smoking is the primary cause of lung damage in the great majority of COPD patients. But since not all smokers develop COPD, it's likely that other factors, like a person's genetic vulnerability to the disease, are also at play. Furthermore, those who have COPD are more prone to go through exacerbations, which are periods when their symptoms worsen and remain. The precision and recall at the peak event level were 77.1% and 78.0%, compared to 83.9% and 83.2% for non-peak events. In the instance of COPD, it was suggested to create a complete system for chronic illnesses that would offer a platform for accurate diagnosis and continued patient health status review (Bellos et al., 2014). A machine or device is being designed to track a patient's condition in real time. A hybridised classifier that categorises a COPD series early and in real-time on a personal digital assistant was developed. It uses a variety of learning techniques, including SVM, random forest, and a predicate-based approach. The estimated quality of the categorization was 94%.A computer-based method to automatically analyse stethoscopically recorded breathing sounds has also been developed. This method has various potential applications, including telemedicine and self-screening (Liu et al., 2017). Three types of respiratory sounds are recorded by one specially designed test device from 60 participants. Convolutional neural networks were used to create the deep model, which was composed of six convolutional layers, three max-pooling layers, and three fully linked layers. The 60 bands of Log-scaled Mel-Frequency Spectral features that were present in the dataset were gathered frame by frame and segmented as model inputs in a span of 23 consecutive frames using time-frequency transformation.



The developed model was then put to the test using a brand-new dataset of 12 subjects, and its precision and recall were compared to the typical outcomes of 5 respiratory physicians. In order to record respiratory sounds on a monitor, Aykanat et al. (2017) proposed a simple and inexpensive digital stethoscope that can be used on any device. Using this instrument, 1,630 patients' 17,930 lung sounds were captured. The Convolutional Neural Network (CNN) and SVM, together with spectrogram images and two additional kinds of machine learning algorithms, were used in the study. Although the usage of MFCC functionality for an SVM algorithm is a frequently used technique for audio classification, the CNN algorithm's performance was assessed using user tests.



Four data sets—safe versus unhealthy; rale, rhonchus, and standard classification of speech; singular classification of respiratory speech form; and classification of the audio form of all sound forms—were created for each CNN and SVM method. The results of the experimental accuracy tests were as follows: 86% CNN, 86% SVM, 76% CNN, 75% SVM, 80% CNN, 80% SVM, and 62% CNN, 62% SVM. As a result, it was discovered that both the CNN and the SVM algorithms perform well when used with spectrogram picture categorization. CNN and SVM algorithms may accurately diagnose and predict COPD using respiratory sounds given a sizable amount of information. In studies, ANFIS, MANFIS, and CANFIS models with various membership functions were used to classify spirometry results (Asaithambi, Manoharan & Subramanian, 2012). Comparing the ANFIS algorithm to the previous neural network method, the recognition accuracy is improved. This might be as a result of the fact that ANFIS combines the logic and learning capabilities of fuzzy inference systems with neural networks. As seen, CANFIS performs with a classification accuracy that is 97.5% higher than that of the traditional ANFIS and MANFIS. The two most common lung sounds that Chamberlain et al. (2016) concentrated on were wheezes and crackles, and their algorithm produced ROC curves with AUCs of 0.86 and 0.74 for wheezes and crackles, respectively. Another study included 155 samples from the COPD disease dataset, which was divided into two categories: Class 1: COPD, which contained 55 samples, and Class 2: Standard, which contained 100 samples (Er & Temurtas, 2008). Results were better with a two hidden layer MLNN system (95.33% accuracy) than with a single hidden layer MLNN system. A feed-forward NN design with a secret layer operating on the log sigmoid transfer feature was utilised in a study to categorise data samples (Asaithambi, Manoharan & Subramanian, 2012). There were several basic, layer-organized processing units in this network that resembled neurons. The error propagation backward method was used for network testing. Radial-base neural networks frequently concentrate on supervised learning. RBF networks may be trained in a single step as opposed to requiring a repeating process like the Multiple Layer Perceptron, making them an effective tool for modelling nonlinear data. Rapidly learning radial base networks. The network answer, which is interpreted with a desired solution to the output layer, is created by linearly combining the hidden layer outputs in response to an input vector. A controlled standard linear system is used to condition the weights. The ability of machine learning approaches to assist in the early diagnosis of COPD exacerbations is quite promising (Fernandez-Granero, Sanchez-Morillo & Leon-Jimenez, 2018). The inevitable negative consequences and the large costs associated with COPD patients may be reduced with the use of exacerbation forecasting. Chronic obstructive pulmonary disease (COPD) patients are more likely to be hospitalised as a result of acute exacerbations, which are also a major factor in declining quality of life. Early searing COPD exacerbations could be detected using the developed methodology 4.4 days before they started. To predict symptom-based exacerbations, an automated decision tree forest classifier was trained and validated using recorded data. Deep learning-based algorithms like CNN and LSTMS are receiving a lot of attention from practical, important learning systems (Bai et al., 2020).

Table 1 Existing work.

YEAR	METHODOLOGY	ADVANTAGE	DISADVANTAGE	ACCURACY
2008	Radial basis function Neural Network And Back Propagation	Individual Spirometry Test Features.	By including more input feature values (spirometric) and a sizable training database, the correctness can be improved even more.	96%
2008	MLNN with: one hidden Layer + Backpropagation and one hidden Layer + Levenberg– Marquardt algorithm	Individual Features were used.	The neural network with two hidden layers was shown to be preferred to the neural network with one hidden layer. They haven't looked for any secret layers deeper than two, which might produce even better results.	93.14%
2012	Worked on the selection of features for various classification algorithms such as linear bayesian , KNN, decision trees, ANN, and SVM	CT images from two datasets that include scans that were inspired by soft kernel	Does not classify the severity of the disease whether mild, moderate or severe.	87%
2020	Classification by colorful snapshot , Classification by gray	Features from Spirometry Tests	Includes only COPD and healthy control patients , Small dataset	88.2%

	snapshot,Classification by binary snapshot.			
2020	Classification of COPD using CT images and a 3D Convolutional Neural Network Without Transfer Learning	CT images from two datasets consisting of inspiration scans using soft kernel	Validating the model with a detailed comparison with other methods on bigger and balanced data sets, ideally using k-fold cross-validation.	58.8%
2016	Classification of lung sounds using a deep learning algorithm by Denoising Auto Encoder	30-second audio recordings from 11 different chest locations were used from the patients.	Not large enough training data and larger sample data needed to get more accurate results to determine the actual accuracy of the model. Longer training time on audio	86%
2018	Analysis of 3D-space quantization that originates with successive three data points in the signal and Deep Belief Networks	120 lung sounds from smokers (12 channels x 5 subjects) and COPD patients (12 channels x 5 subjects) were used.	Severity classification of respiratory abnormalities not done	95.84%,
2018	Identification of asthma and chronic obstructive pulmonary disease automatically using an Expert Diagnostic System with Artificial Neural Network	Seven features from symptom questionnaire + four features from the result of SPIR test	Evaluation of the EDS was conducted in a healthcare center in Bosnia and Herzegovina. However, it is recommended to further test in other healthcare establishments in order to further justify the EDS.	95.17%
2018	Deep Learning for COPD Detection using Lung Sounds with Sequential Forward Feature Selection applied to DBN	Features were extracted using a discrete wavelet transform. The input dataset contained 18 wavelet features	Results suffered from noise presence. Overall also, the accuracy, sensitivity and specificity are less without SFFS.	90.83%,
2021	CNN(mfcc) -based deep learning assistive model	Features were extracted using mfcc (librosa).	The Sensitivity and Specificity of mfcc were consistent, this time both being 0.92. None other feature worked well after augmentation	92.0%

LIMITATIONS OF EXISTING METHODS

Previous analytical techniques use more resources, particularly non-CNN neural networks that employ complicated analytical networks. As a result, it necessitates sophisticated mathematical abilities and can be expensive for infrastructure. Without infrastructure investment, training and disease forecasting can take a very long time. To assess whether a patient has dementia, current methods including a physical diagnosis by a doctor take longer and frequently require a trip to the hospital. The number of sound samples usually varies depending on the illness. Any network that is trained with uneven data aids in the prediction of the disease, which had a very high number of samples, so there is always a need to measure data. The sound samples typically have a lot of background noise. Deep networks of random forests combined with medical pictures are commonly used in current studies to detect COPD. According to numerous research, using language analyses that support speech and language variables to train machine learning models for AD can help diagnose COPD in its early stages. The predicted number of people enrolled in the study was approximated with 75% accuracy using mechanical learning techniques based on the description. Bias is introduced since the data used for each of the aforementioned activities is biassed and only includes 32 to 36 subjects on average.

OBJECTIVES

Based on the promotion of the proposed project as discussed in the previous section, the objectives of the project have been formally implemented. Do a thorough literature review to understand and determine project compatibility. Consult a physician for information on sound effects. or Identification and collection of relevant data for training and evaluation of the neural network. Pre-processing and cleaning the database to remove unwanted noise and waves if present. Calculate the mathematical model of supervised training with a feature-based distribution of classification parameters. Analyse the problem-solving success of the proposed algorithm by focusing on factors such as accuracy, computer complexity and algorithmic accuracy by comparing noted results with other popular and comparable metaheuristic metrics. Build an automated system that can detect dementia using audio input. Develop an effective and measurable architecture that we can use

METHODS FOR COPD DIAGNOSIS

One of the most popular uses of audio deep learning is sound classification. It entails developing the ability to categorise sounds and foretell their category. This kind of issue can be used in a variety of real-world situations, such as categorising music clips to determine the genre of the music or categorising brief utterances made by a group of speakers to determine the speaker based on voice.

Although most of the papers reviewed here highlight the power of AI and machine learning methods, no real translation of clinical practice has been achieved.

The aim of this study was to provide a new clinical study tool based on early detection of dementia. To find the correct partition model, we compared the algorithms for selecting different features with the dividing algorithms using the same data. We performed sensitivity analysis to test the validity of the results with our differentiation algorithms.

Algorithms are trained in voice files and learn to detect small changes in speech patterns that may indicate conflict.

METHODOLOGY

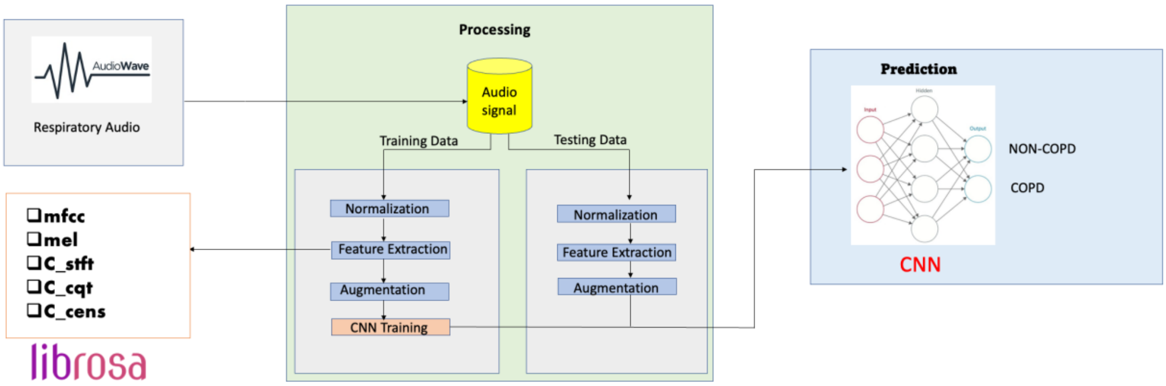


Figure 1. Workflow of the proposed system.

Data collectionOne of the most often utilised applications for audio deep learning is sound classification. Learning how to categorise sounds and foretell their type is a necessary skill. This kind of issue can be used in many real-world situations, such as classifying audio samples to determine the music's genre or short utterances by a group of speakers to determine who is speaking based on voice. 126 input recordings were collected from a database of respiratory sounds that Two Portuguese and Greek research teams produced

(ICBHI, 2017 Challenge). The breathing sounds of both healthy people and those with respiratory ailments were included in the data samples. Patients range in age from infants to elderly patients, and all age groups are represented. The collection consists of 920 annotated audio samples with 6898 breathing cycles spread over 5.5 hours from 126 persons. Of these, 1,864 respiratory cycles have crackles, 886 have wheezes, and 506 have both. Experts in respiratory biology annotated the cycles as having crackles, wheezes, a combination of them, or no accidental respiratory noises. The recordings were made with various tools, and they ranged in length from 10 to 90 seconds. It also includes the chest locations where the recordings were made. Some respiration cycles have high noise levels that mimic real-world settings. Data preprocessing: The dataset was very irregular and had a lot of unstructured data. Using the Python module Librosa, we chopped and padded the audio files to a length of 20 seconds in order to standardise the data (Librosa, 2020). Librosa [13] is a Python library for analyzing audio signals. It has utilities for reading audio from files, computing different audio representations, and working with audio data in general. Additionally, Librosa has numerous tools for modifying audio signals and dealing with music notation. The Librosa package, which may be used to create Python programmes for working with sound and music, is a useful tool for creating such programmes. The library contains a number of components related to the management and extraction of sound recordings, including the recording of various spectrogram depictions, the organisation of source detachments, standard decay, stacks, and translations of sound recordings. A beat tracker counts the number of casings associated with each beat as well as the rhythm (measured in beats per moment). The outline k * hop length, which is built on focused outlines, serves as the foundation for the kth outline in librosa. In this lesson, we'll go through a more complex syntax example that combines beat-synchronous feature aggregation, harmonics-percussive separation, and various spectral features.

Feature Extraction: for the feature extraction, we calculated five features.

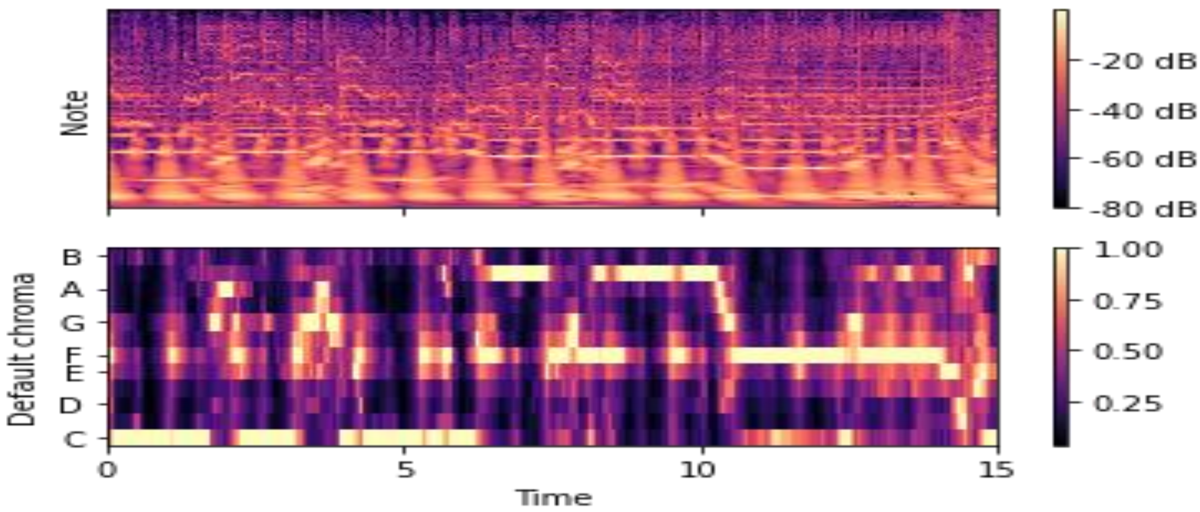


Figure 2 Chroma constant-Q time. X-axis: pitch class; y-axis: time

Mel-Frequency Cepstral Coefficients (mfcc), Mel-Spectrogram, ConstantQ Chromagram, Chroma Energy Normalized Variant (CENS), and Chroma stft (Chromagram created from the waveform/power spectrogram) were among the characteristics. A set of coefficients known as a mel-frequency cepstrum is an MFCC (MFC). An MFC is a picture of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrogram on a non-linear mel scale of frequency. These features serve as representations of phonemes, the discrete units of sound, due to the vocal tract's shape, which is crucial for sound development. As a result, MFCC is a great factor to consider when conducting a respiratory audio analysis. The Mel-Spectrogram is produced by taking samples of the air pressure over time, quickly transforming it from the time domain to the frequency domain using the fast Fourier transform, and then converting the frequency to a mel scale and the colour dimension to the amplitude. It displays the transient power spectrum of the sound. Chroma-based characteristics, which are also referred to as "pitch class profiles," are an effective group of features for analysing music whose pitches can be categorised. They have the characteristics we talked about earlier. Due to their resistance to dynamics, timbre, and articulation, the CENS characteristics are extensively used in audio matching and retrieval applications. We gave each feature a "n" value of 40 to maintain consistency throughout the features (equivalent to n mfccs in mfcc features and n chroma bins in chroma features). Since there were approximately four times as many COPD samples as there were non-COPD samples, we increased the quantity of non-COPD samples using a variety of audio augmentation techniques. As shown in Table 2, we used the following methods for audio augmentation. CNN architecture: Keras and a Tensorflow back-end were used to create the CNN that makes up our model. It is a sequential model made up of an input layer, convolutional layers in two dimensions, dropout layers, layers that maximise pooling, and a dense layer. The primary objective of a convolution layer is feature detection. It operates by adjusting a filter window over the input and multiplying matrices; the result is recorded as a feature map. Convolution is the name of the feature map creation procedure. The final layer has the GlobalAveragePooling2D type, with each convolutional layer having a related MaxPooling2D pooling layer. The role of the pooling layer is to reduce the dimensionality of the model, which is done by optimising the parameters and lowering the amount of computation required. By doing this, training time and overfitting concerns are reduced. For our dense output layer, Maxpooling and the Global Average Pooling type use the maximum and average window sizes for each window. The number of possible classifications is determined using the two nodes (num labels of the output layer). The softmax is turned on in the output layer. In order to convert the output number into a probability, Softmax raises it up to a maximum of one. The number of nodes in each layer is determined by the filter parameter. As the kernel size parameter controls the kernel window size, which is two 2 2 filter matrices, the size of each layer gradually increases from 16, 32, 64, and finally 128. The first layer will be given an input shape, as seen in Table 3. When padding is taken into consideration, the values of the input shape would be (40, 862, 1), showing the mono audio structure as 1, the number of MFCCs as 40, and the number of frames as 862. The data is then sent to the MaxPooling2D layer of the Convolution2D layer (16 filters, kernel size: 2, relu) after the Convolution2D layer (Pooling Size: 2). Secondly, to stop the data from being being overfit, we specified a 20% dropout rate. Following the dropout, the data is sent to a Convolution layer (32 filters, kernel size: 2, relu), which once more provides the input to a MaxPooling2D layer of (Pooling size: 2). A MaxPooling2D layer with a pooling size of two is followed by a Convolution2D layer with 64 filters and a kernel size of two, relu, and a final 20% dropout. The data is then transmitted via a second Convolution2D layer, a MaxPooling2D layer, and a dropout layer (20%). (32 filters, kernel size: 2, relu). An activation map is produced from a convolutional layer using the activation function ReLu. The GlobalAveragePooling2D layer (128), which flattens the output before passing it to a dense layer, which separates it into COPD and non-COPD outputs, receives the remaining data. It has been discovered that audio inputs frequently contain more noise than visual inputs do. We used the Adam optimization technique to optimise. Our method is more effective when dealing with ambiguous input values and is computationally efficient, needing less memory. Adam is a stochastic gradient descent and RMSprop method combination that offers superior network weight optimization logic and, as a result, more effective hyper parameter tweaking (Kingma & Ba, 2017). In comparison to other optimizers like RMSprop, SGD, ADAGrad, and NAdam, this approach delivers a faster convergence together with optimised performance parameters. Three benchmarked databases were used to compare various optimizers, and the results are shown in Table 4 below: The model then makes a forecast based on the option with the highest likelihood. Training and testing sets of data were created using the audio sample that is provided. Before being utilised to train the model, the training data is put through a number of processes, including normalisation, feature extraction, and augmentation. The training data are then saved alongside the modified model features. The testing data was processed using a similar set of steps. It makes use of the CNN-developed trained COPD categorization and prediction model.

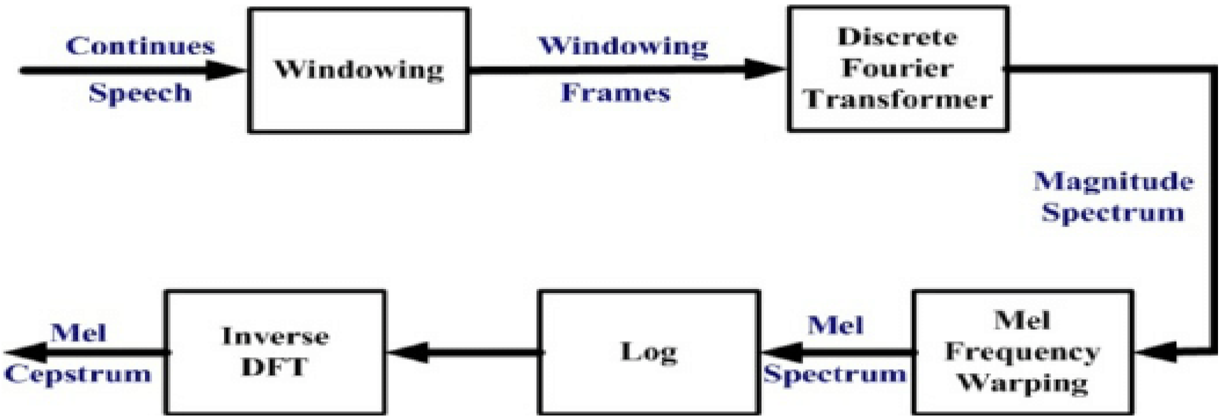


Figure 3 mfcc flowchart

When describing a signal's timing, frequency, and magnitude, a spectrogram is helpful. A spectrogram is a 2D representation of audio in which time and magnitude are the two dimensions and colours represent the third dimension. The presence of respiratory disorders also results in a distinct spectrogram since they are musical anomalies that can be detected by appropriate auscultation methods. Mel-Spectrograms, MFCC[Figure 3], and Chromagrams are the three major categories into which our retrieved features can be separated. The vocal tract’s shape manifests itself in the short-time power spectrum envelope, and the MFCC accurately represents this envelope.

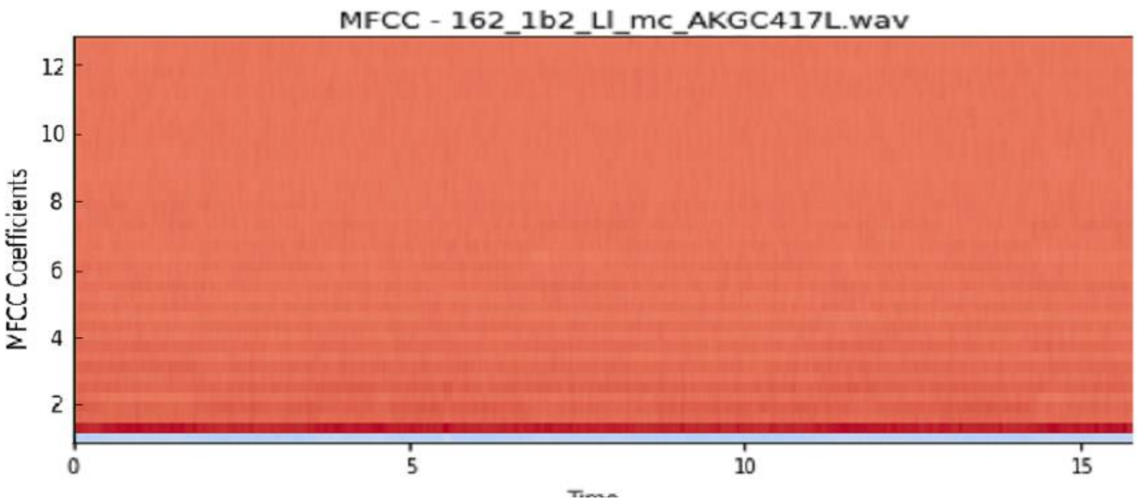


Figure 4 MFCC representation. X-axis: MFCC coefficients; y-axis: time [1]

Since its introduction, MFCCs have been frequently employed in audio classification. Chroma features, commonly referred to as pitch-class features, are an effective way to represent audio samples. The usual Chroma variation (chroma stft) typically displays the energy levels of each pitch class in the signal wave. Additionally, the Chromagram has been used with a "Constant Q" transform (chroma cqt). The constant-Q transform is used to convert a data series from the time domain to the frequency domain. The alternative variation, Chroma Energy Normalized Statistics (CENS) (chroma cens), is founded on the idea that statistics applied over broad windows are necessary for smoothing local changes in various audio qualities (Librosa, 2019). Classification of audio: We will begin with sound data, transform them into spectrograms, add them to a CNN plus Linear Classifier model, and then generate predictions about the class—either COPD or non-COPD class—to which the sound belongs. FEATURE EXTRACTION After Data Processing we will use the Librosa library in python to extract Spectral Features from Feature extraction.

We will extract 5 features:-

chroma_stft (*[, Y, Sr, S, norm, n _fft,...)	Compute a chromagram from a waveform or power spectrogram.
chroma_cat (*L Y, sr, C, hop_length, fmin, ...)	Constant-Q chromagram
chroma_cens(*L, y, sr, C, hop_length, fmin, ..)	Computes the chroma variant "Chroma Energy Normalized" (CENS)
melspectrogram(*L, Y, sr, S, n_fft,...)	Compute a mel-scaled spectrogram.
mfce(*L, Y, sr, S, n _mfcc, det_type, norm, ...)	Mel-frequency cepstral coefficients (MFCCs)

Figure 5 Feature extraction

An MFC is made up of a number of coefficients known as mel-frequency cepstral coefficients (MFCCs). They were created using an audio clip's cepstral representation (a nonlinear spectrum-of-a-spectrum). The actions that must be taken to calculate the MFCC features are as follows: The method calculates the power spectrum periodogram estimate for each frame of the signal by framing the signal into brief frames. Summarizing the filter energies after applying the Mel-space filter banks on the power spectra. Calculate the DCT (Discrete Cosine Transform) of the logarithms and the logarithms of the filter banks' energies. The method will omit the remaining coefficients because the DCT coefficients must range from 2 to 13. Each feature vector may occasionally have the frame energy, Delta, and Delta-Delta features added. The frequencies in a spectrogram are translated to the mel scale to create a mel spectrogram. Mel scale: A unit of pitch that allows for the perception of equal distances in pitch by the listener. It is known as the mel scale. To accurately represent an audio signal digitally, we collected air pressure samples throughout time. On overlapping windowed portions of the audio signal, we applied the fast Fourier transform to translate the audio signal from the time domain to the frequency domain. To create the spectrogram, we transformed the colour dimension (amplitude) to decibels and the y-axis (frequency) to a log scale. To create the mel spectrogram, we translated the y-axis (frequency) onto the mel scale.

Metrics for Evaluation of the Model:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

The suggested model is assessed using the metrics listed below.

Accuracy: how many of the classifier's predictions were correct when compared to the label value during testing. It can also be expressed as the ratio of accurate reviews to all reviews. To assess accuracy, apply the formula below: Where TP, TN, FP, and FN stand for "true positives," "false positives," and "false negatives," respectively. True positives occur when a dataset's records have positive class labels and the classifier correctly predicted that the records would have positive class labels. When the classifier predicts that a dataset record's class label will be negative and the record already has a negative class label, False negatives occur when a class label for a record in a dataset is positive but the classifier predicts a negative class label for that record. False positives occur when the class label of a record in a dataset is negative but the classifier predicts that record's class label will be positive. Sensitivity: This is the proportion of true positives during testing that the classifier correctly identifies. It is calculated using the following Equation. Sensitivity:It is the percentage of true positives that are correctly identified by the classifier during testing. It is calculated using the following Equation. Specificity: It is the percentage of true negatives that are correctly identified by the classifier during testing. This equation is used to calculate it. Precision: Precision is an important metric for assessing exactness; it expresses the proportion of cases that the classifier classified as positive relative to all of the Equation's total forecasted positive occurrences. Recall: Recall establishes completeness, i.e., the proportion of cases that the classifier correctly classified as positive. When there is a large cost associated with False Negative, as illustrated in Equation, the recall is a performance parameter that is used to choose the optimum model. F1-measure: (F1 or F1-score) represents the harmonic mean of precision and recall as shown in Equation. F1 Score is required to find a balance between Precision and Recall. Accuracy is mainly contributed by a large number of True Negatives whereas False Negative and False Positive usually have business costs (tangible & intangible). Thus F1 Score might be a better measure when a balance between Precision and Recall is needed with an uneven class distribution (large number of Actual Negatives).

Device Configuration: all this was run on a machine with the configuration as listed below: Processor: GPU: Nvidia K80s, (24GB DDR5 Type) RAM: 36GB HDD: 250GB SSD, max sequential read speed: upto 550 MB/s, max sequential write speed: upto 520 MB/s Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz

DATASET

Respiratory Sound Database (KAGGLE) - There are 920 annotated recordings in it, with lengths ranging from 10 to 90 seconds. 126 different patients were used to create these recordings. There are 6898 respiratory cycles in 5.5 hours of recordings, 1864 of which include crackles, 886 of which have wheezes, and 506 of which have both. The data collection includes recordings of both clear respiratory sounds and chaotic ones that reflect actual environmental conditions. Patients include young people, adults, and the elderly.

Available at: https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database

Hack Respiratory Sound Dataset- Respiratory Sound files of affected users and Users does not have corona is available in Hack-Respiratory-Sound-Dataset

Respiratory sound files with dates and user IDs are available in test and train folders. Numerous categories of respiratory sound files include vowels A, E, and O, counting swiftly, counting normally, coughing strongly, and coughing shallowly.

Available at: https://www.kaggle.com/datasets/praveengovi/coronahack-respiratory-sound-dataset

CHALLENGES IN THIS AREA

The performance of most ML models, and deep learning models in particular, depends on the quality, quantity, and relevance of the training data. Lack of data, however, is one of the most typical issues with machine learning implementation in the workplace. This is due to the fact that it can often be costly and time-consuming to obtain such data. Businesses can lessen their dependency on the preparation and acquisition of training data by using data augmentation to develop machine learning models more quickly and accurately. As our data is split into training and testing in the ratio of 8:2. We have a very less amount of test data So, it is necessary to implement By creating additional and distinct instances for training datasets, data augmentation helps machine learning models perform better and produce better results. A machine learning model performs better and is more accurate if the dataset is large and sufficient. Data collection and labelling for machine learning models can be time-consuming and expensive activities. Using data augmentation approaches, companies can transform datasets to save these operational expenses. One of the steps in building a data model, which is necessary for high-accuracy models, is cleaning the data. However, if cleaning reduces the data's representability, the model cannot produce reliable predictions for inputs from the real world. By producing variables that the model could encounter in the real world, data augmentation approaches might help machine learning models become more resilient. Additionally, because CNN is being used, our execution time is rather long and requires more energy to implement. To decrease this, we can consider using different NLP models and device used to calculate respiratory sound is expensive and large in size for that we can develop Arduino or raspberry pi device at very low cost.

FUTURE SCOPE

Several lines of future research have been identified for further work. In the future, this project could be expanded to: Instead of one use at a time, can be used in various approaches and combinations to determine if that produces better results. Use of smartphone microphones to capture noise so that suspects can detect the disease in the comfort of their homes. The plan can also define the severity of the identified disease, such as mild, moderate or severe.

After comparing all the models, we have come up with a Transformer model for our project. According to Vaswani et al. (2017), there were a number of reasons why they decided against using RNN and CNN. For shorter sequence lengths, self-attention layers were found to be faster than recurrent layers, and for extremely long sequence lengths, they can be constrained to only consider a neighbourhood in the input sequence. A self-attention layer requires a fixed number of sequential operations, whereas a recurrent layer depends on the length of the series. The kernel width in convolutional neural networks has a direct impact on the potential for long-term relationships to develop between pairs of input and output positions**.** Large kernels or stacks of convolutional layers would be needed to track long-term dependencies, which could raise the computing cost.

CONCLUSION

Convolutional neural networks (CNNs) have been widely adopted as the main building block for end-to-end audio classification models in the past decade, with the intention of learning direct label mapping from audio spectrograms. To better capture long-range global context, a recent trend is to add a self-attention mechanism on top of the CNN, forming a CNN-attention hybrid model. However, it is uncertain whether using a CNN is necessary, and whether using attention-only neural networks alone is sufficient to get good performance in audio categorization. In our paper, we got yhe answer of questions by introducing the Audio Spectrogram Transformer (AST), the first convolution-free, purely attention-based model for audio classification. paper evaluate AST on various audio classification benchmarks, where its accuracy on Speech Commands V2achieves new state-of-the-art results. Thought, we can apply an Audio Spectrogram transformer to classify COPD and non-COPD patients. And also we can develop Arduino or raspberry pi based devices at a very low cost.

REFERENCES

[1] Srivastava A, Jain S, Miranda R, Patil S, Pandya S, Kotecha K. 2021. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Computer Science* 7:e369 https://doi.org/10.7717/peerj-cs.369

[2] (Google Researcher) **Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones,** Attention Is All You Need (2017). arXiv:1706.03762v5 [cs.CL] 6 Dec 2017

[3]Ahmed J, Vesal S, Durlak F, Kaergel R, Ravikumar N, Rémy-Jardin M, Maier A, Tolxdorff T, Deserno T, Handels H, Maier A. 2020. COPD classification in CT images using a 3D convolutional neural network. In: Maier-Hein K, Palm C, eds. Bildverarbeitung für Die Medizin 2020—Informatik Aktuell. Wiesbaden: Springer Vieweg,

[4] Aykanat M, Kılıç Ö, Kurt B, Saryal S. 2017. Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing 2017(1):65 Classification of lung sounds using convolutional neural networks | EURASIP Journal on Image and Video Processing | Full Text (springeropen.com)

[5] Yuan Gong, Yu-An Chung, James Glass(MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA**).** AST: Audio Spectrogram Transformer. arXiv:2104.01778v3 [cs.SD] 8 Jul 2021

[6] Er O, Temurtas F. 2008. A study on chronic obstructive pulmonary disease diagnosis using multilayer neural networks. Journal of Medical Systems 32(5):429–432 DOI 10.1007/s10916-008-9148-6.

[7] Shaikh A, Patil S. 2018. A survey on privacy enhanced role based data aggregation via differential privacy. In: 2018 International Conference On Advances in Communication and Computing Technology. Piscataway: IEEE, 285–290.

[8] Bellos CC, Papadopoulos A, Rosso R, Fotiadis DI. 2014. Identification of COPD patients' health status using an intelligent system in the CHRONIOUS wearable platform. IEEE Journal of Biomedical and Health Informatics 18(3):731–738 DOI 10.1109/JBHI.2013.2293172.

[9] Perna D, Tagarelli A. 2019. Deep auscultation: predicting respiratory anomalies and diseases via recurrent neural networks. arXiv. Available at http://arxiv.org/abs/1907.05708

[10] Pham L, McLoughlin I, Phan H, Tran M, Nguyen T, Palaniappan R. 2020. Robust deep learning framework for predicting respiratory anomalies and diseases. arXiv. Available at http://arxiv.org/ abs/2002.03894.

[11] Khatri KL, Tamil LS. 2018. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. IEEE Journal of Biomedical and Health Informatics 22(1):285–290 DOI 10.1109/JBHI.2017.2698418.

[12] Weese J, Lorenz C. **2016**. Four challenges in medical image analysis from an industrial perspective. Medical Image Analysis 33:44–49 DOI 10.1016/j.media.2016.06.023.

[13] Librosa. 2022. Feature extraction—Librosa 0.8.0 documentation available at https://librosa.org/ doc/latest/feature.html

[15] Transformers 4.24.0 documentation. Available at
https://pypi.org/project/transformers/

[16] Chronic obstructive pulmonary disease (**COPD**): documentation available at
https://en.wikipedia.org/wiki/Chronic_obstructive_pulmonary_disease

[17] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3(11):e442. doi: 10.1371/journal.pmed.0030442. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[18]. Zhong N, Wang C, Yao W, et al. Prevalence of chronic obstructive pulmonary disease in China. *Am J Respir Crit Care Med.* 2007;176(8):753–760. doi: 10.1164/rccm.200612-1749OC. [PubMed] [CrossRef] [Google Scholar]

[19] Mapel DW, Dalal AA, Blanchette CM, et al. Severity of COPD at initial spirometry-confirmed diagnosis: data from medical charts and administrative claims. *Int J Chron Obstruct Pulmon Dis.* 2011;6:573–581. doi: 10.2147/COPD.S16975. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[20] Bellamy D, Smith J. Role of primary care in early diagnosis and effective management of COPD. *Int J Clin Pract.* 2007;61:1380–1389. doi: 10.1111/j.1742-1241.2007.01447.x. [PubMed] [CrossRef] [Google Scholar]

[21] Gurney JW, Jones KK, Robbins RA, et al. Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. *Radiology.* 1992;183(2):457–463. doi: 10.1148/radiology.183.2.1561350. [PubMed] [CrossRef] [Google Scholar]

[22] Lynch DA, Austin JH, Hogg JC, et al. CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the Fleischner Society. *Radiology.* 2015;277(1):192–205. doi: 10.1148/radiol.2015141579. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[23] Kauczor HU, Wielpütz MO, Jobst BJ, et al. Computed tomography imaging for novel therapies of chronic obstructive pulmonary disease. *J Thorac Imaging.* 2019;34(3):202–213. doi: 10.1097/RTI.0000000000000378. [PubMed] [CrossRef] [Google Scholar]

[24] Ostridge K, Wilkinson TM. Present and future utility of computed tomography scanning in the assessment and management of COPD. *Eur Respir J.* 2016;48(1):216–228. doi: 10.1183/13993003.00041-2016. [PubMed] [CrossRef] [Google Scholar]

[25] Feragen A, Petersen J, Grimm D, et al. Geometric tree kernels: classification of COPD from airway tree geometry. *Inf Process Med Imaging.* 2013;23:171–183. doi: 10.1007/978-3-642-38868-2_15. [PubMed] [CrossRef] [Google Scholar]

[26] Bodduluri S, Newell JD, Jr, Hoffman EA, et al. Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. *Acad Radiol.* 2013;20(5):527–536. doi: 10.1016/j.acra.2013.01.019. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[27] Cheplygina V, Sorensen L, et al (2014) Classification of COPD with multiple instance learning in 2014 22nd International Conference on Pattern Recognition. 10.1109/ICPR.2014.268

[28] Cheplygina V, Pena IP, Pedersen JH, et al. Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE J Biomed Health Inform.* 2018;22(5):1486–1496. doi: 10.1109/JBHI.2017.2769800. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[29] González G, Ash SY, Vegas-Sánchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med.* 2018;197(2):193–203. doi: 10.1164/rccm.201705-0860OC. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[30] Hatt C, Galban C, Labaki W, Kazerooni E, Lynch D, Han M (2018) Convolutional neural network based COPD and emphysema classifications are predictive of lung cancer diagnosis. In: Stoyanov D. et al. (eds) Image analysis for moving organ, breast, and thoracic images. RAMBO 2018, BIA 2018, TIA 2018. Lecture Notes in Computer Science, vol 11040. Springer, Cham. 10.1007/978-3-030-00946-5_30

[31] Tang L, Coxson HO, Lam S, et al. Towards large-scale case-finding: training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. *Lancet Digit Health.* 2020;2(5):e259–e267. doi: 10.1016/S2589-7500(20)30064-9. [PubMed] [CrossRef] [Google Scholar]

[32] Ju J, Li R, Gu S, et al. Impact of emphysema heterogeneity on pulmonary function. *PLoS ONE.* 2014;9(11):e113320. doi: 10.1371/journal.pone.0113320. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[33] Ahmed J, Vesal S, Durlak F, et al (2020) COPD Classification in CT images using a 3D convolutional neural network. arXiv:2001.01100

[34] Ho TT, Kim T, Kim WJ, et al. A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects. *Sci Rep.* 2021;11(1):34. doi: 10.1038/s41598-020-79336-5. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[35] Liu J, Tan G, Lan W, et al. Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. *BMC Bioinformatics.* 2020;21(Suppl 6):123. doi: 10.1186/s12859-020-3437-6. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[36] Zhang X, He L, Chen K, et al. Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease. *AMIA Annu Symp Proc.* 2018;5(2018):1147–1156. [PMC free article] [PubMed] [Google Scholar]

[37] TA Song S Roy Chowdhury F Yang et al 2019 Graph convolutional neural networks for Alzheimer's disease classification Proc IEEE IntSymp Biomed Imaging 414–41710.1109/ISBI.2019.8759531 [PMC free article] [PubMed]

[38] Jiang H, Cao P, Xu M, et al. Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Comput Biol Med.* 2020;127(1):104096. doi: 10.1016/j.compbiomed.2020.104096. [PubMed] [CrossRef] [Google Scholar]

[39] Liang X, Zhang Y, Wang J, Ye Q, Liu Y, Tong J. Diagnosis of COVID-19 pneumonia based on graph convolutional network. *Front Med.* 2021;7:612962. doi: 10.3389/fmed.2020.612962. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[40] Wang SH, Govindaraj VV, Górriz JM, et al. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf Fusion.* 2021;67:208–229. doi: 10.1016/j.inffus.2020.10.004. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[41] Li Y, Chen J, Xue P, et al. Computer-aided cervical cancer diagnosis using time-lapsed colposcopic images. *IEEE Trans Med Imaging.* 2020;39(11):3403–3415. doi: 10.1109/TMI.2020.2994778. [PubMed] [CrossRef] [Google Scholar]

[42] Ye H, Wang DH, Li J, et al (2019) Improving histopathological image segmentation and classification using graph convolution network. in ICCPR '19: 2019 8th International Conference on Computing and Pattern Recognition. 10.1145/3373509.3373579

[43] Zhou Y, Graham S, Koohbanani NA, et al (2019) CGC-net: cell graph convolutional network for grading of colorectal cancer histology images. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), pp. 388–398. 10.1109/ICCVW.2019.0050

[44] Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer ct screening trial—overall design and results of the prevalence round. *J Thorac Oncol.* 2019;4(5):608–614. doi: 10.1097/JTO.0b013e3181a0d98f. [PubMed] [CrossRef] [Google Scholar]

[45] Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. *GOLD executive summary.* 2017;53(3):128–149. doi: 10.1016/j.arbr.2017.02.001. [PubMed] [CrossRef] [Google Scholar]

[46] Bruna J, Zaremba W, Szlam A, et al (2013) Spectral networks and locally connected networks on graphs. Computer Science. arXiv:1312.6203

[47] Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. arXiv:1606.09375

[48] Wang Z, Zheng L, Li Y, et al (2019) Linkage based face clustering via graph convolution network. arXiv:1903.11306

[49] Niepert, M, Ahmed M and Kutzkov K (2016) Learning convolutional neural networks for graphs. arXiv:1605.05273

[50] Li Q, Han Z, Wu XM (2018) Deeper insights into graph convolutional networks for semi-supervised learning. arXiv:1801.07606

[51] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* 2020;42(2):318–327. doi: 10.1109/TPAMI.2018.2858826. [PubMed] [CrossRef] [Google Scholar]

[52] Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* 2016;35(5):1285–1298. doi: 10.1109/TMI.2016.2528162. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[53] Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging.* 2016;35(5):1299–1312. doi: 10.1109/TMI.2016.2535302. [PubMed] [CrossRef] [Google Scholar]

[54] Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv:1609.02907

[56] Sorensen L, Nielsen M, Lo P, et al. Texture-based analysis of COPD: a data-driven approach. *IEEE Trans Med Imaging.* 2012;31(1):70–78. doi: 10.1109/tmi.2011.2164931. [PubMed] [CrossRef] [Google Scholar]

[57] Shuman DI, Narang SK, Frossard P, et al. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process Magazine.* 2013;30(3):83–98. doi: 10.1109/MSP.2012.2235192. [CrossRef] [Google Scholar]

[58] Li, RY, Yao JW, Zhu XL, et al (2018) Graph CNN for survival analysis on whole slide pathological images. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, 11071. Springer, Cham. 10.1007/978-3-030-00934-2_20

[59] Chen, ZM, Wei XS, Wang P, et al (2019) Multi-label image recognition with graph convolutional networks. in 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). 10.1109/CVPR.2019.00532

[60] Xu C, Qi S, Feng J, et al. DCT-MIL: Deep CNN transferred multiple instance learning for COPD identification using CT images. *Phys Med Biol.* 2020;65(14):145011. doi: 10.1088/1361-6560/ab857d. [PubMed] [CrossRef] [Google Scholar]