Pay Attention to the cough: Early Diagnosis of COVID-19 using Interpretable Symptoms Embeddings with Cough Sound Signal Processing

Ankit Pal
Saama AI Research
Chennai, India
ankit.pal@saama.com

Malaikannan Sankarasubbu
Saama AI Research
Chennai, India
malaikannan.sankarasubbu@saama.com

ABSTRACT

COVID-19 (coronavirus disease 2019) pandemic caused by SARS-CoV-2 has led to a treacherous and devastating catastrophe for humanity. No specific antivirus drugs or vaccines are recommended to control infection transmission and spread at the time of writing. The current diagnosis of COVID-19 is done by Reverse-Transcription Polymer Chain Reaction (RT-PCR) testing. However, this method is expensive, time-consuming, and not easily available in straitened regions. An interpretable and COVID-19 diagnosis AI framework is devised and developed based on the cough sounds features and symptoms metadata to overcome these limitations. The proposed framework's performance was evaluated using a medical dataset containing Symptoms and Demographic data of 30000 audio segments, 328 cough sounds from 150 patients with four cough classes (COVID-19, Asthma, Bronchitis, and Healthy). Experiments' results show that the model captures the better and robust feature embedding to distinguish between COVID-19 patient coughs and several types of non-COVID-19 coughs with higher specificity and accuracy of $95.04 \pm 0.18\%$ and $96.83 \pm 0.18\%$ respectively, all the while maintaining interpretability.

CCS CONCEPTS

• Computing methodologies → Machine learning; • Applied computing → Life and medical sciences;

KEYWORDS

COVID-19, Audio Analysis, Deep Learning, Medical Data, Machine Learning

ACM Reference Format:

Ankit Pal and Malaikannan Sankarasubbu. 2021. Pay Attention to the cough; Early Diagnosis of COVID-19 using Interpretable Symptoms Embeddings with Cough Sound Signal Processing. In *The 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21)*, March 22–26, 2021, Virtual Event, Republic of Korea. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3412841.3441943>

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.
SAC '21, March 22–26, 2021, Virtual Event, Republic of Korea
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8104-8/21/03.
<https://doi.org/10.1145/3412841.3441943>

1 INTRODUCTION

The novel coronavirus (COVID-19) disease has affected over 31.2 million lives, claiming more than 1.02 million fatalities globally, representing an epoch-making global crisis in health care. At the time of writing, no specific antivirus drugs or vaccines are recommended to control infection transmission and spread. The current diagnosis of COVID-19 is made by Reverse-Transcription Polymer Chain Reaction (RT-PCR) testing, which utilizes several primer-probe sets depending on the assay utilized [10]. However, this method is time-consuming, expensive, and not easily available in straitened regions due to lack of adequate supplies, healthcare facilities, and medical professionals. A low-cost, rapid, and an easily accessible testing solution is needed to increase the diagnostic capability and devise a treatment plan. Computed Tomography(CT) helps clinicians perform complete patient assessments and describe the specific characteristic manifestations in the lungs associated with COVID-19 [19]. Hence, serving as an efficient tool for early screening and diagnosis of COVID-19. In analyzing medical images, AI-based methods have shown great success [9, 13, 14]. These methods are scalable, automatable, and easy to apply in clinical environments [1, 25]. Significant attempts have been made to use x-ray images for automatic diagnosis of COVID-19 [2, 21, 24, 28]. Studies dealing with the classification of COVID-19 show promising results in this task.

However, in [6] work, the classification limitations of x-ray images are examined since the network may learn more unique features to the dataset than those unique to the disease. Despite its success, the CT scan displays similar imaging characteristics, making it difficult to distinguish between COVID-19 and other pneumonia types. Moreover, CT-based methods can be integrated only with the Healthcare system to help clinical doctors, radiologists, and specialists detect COVID-19 patients using chest CT images. Unfortunately, an individual cannot utilize this method at home. To obtain the CT scan image and report, one must visit a well-equipped clinical facility or diagnostic center, which may increase the risk of exposure to the virus. According to the WHO and CDC official report, the four primary symptoms of the COVID are dry cough, fever, tiredness, and difficulty in breathing. [5]. However, cough is more common as it is one of the early symptoms of respiratory tract infections. Studies show that it occurs in 68% to 83% of the people showing up for the medical examination. Cough classification is usually carried out manually during a physical examination, and the clinician may listen to several episodes of voluntary or natural coughs to classify them. This information is crucial in diagnosis and treatment.

In previous studies, several methods with speech features have been proposed to automate different cough types classification. In the study published by [17], the sound of voluntary cough in patients with respiratory diseases was investigated. Later, in 2015, [12] published the study on the analysis of asthmatic breathing sounds. These studies utilized the wavelet transformation, which is a type of signal processing technique, generally used on non-stationary signals. In a study [27], a Logistic Regression model was utilized to classify the dry and wet cough from pediatric patients with different respiratory illnesses. For pertussis cough classification, three separate classifiers' performance was analyzed in [23]. Several AI-based approaches, motivated by prior work, have been presented to detect patients with COVID-19 using cough sound analysis. [8] gives an overview of Audio, Signal, Speech, NLP for COVID-19, [4, 22, 26] have collected a crowdsourced dataset of respiratory sounds and shared the findings over a subset of the dataset. [11, 16] performed similar analyses on cough data and achieved good accuracy. Most studies use short-term magnitude spectrograms transformed from cough sound data to the convolutional neural network (CNN). However, these methods have the following limitations :

- **Ignoring domain-specific sound information** Cough is a non-stationary acoustic event. CNN is based only on a spectrogram input; some domain-specific important characteristics (besides spectrogram) of cough sounds might be overlooked in the feature space.
- **Using cough features only** These methods exploit the cough features only, ignoring patient characteristics, medical conditions, and symptoms data. Both cough features and other symptoms accompanied by demographic data are responsible for COVID-19 infection. Wherein the prior carries vital information about the respiratory system and the pathologies involved, the latter encodes patient characteristics, signs, and health conditions (fever, chest pain, dyspnea). However, their existence alone is not a precise enough marker of the disease. Therefore, determining the symptoms (besides cough) presented by suspected cases, as best predictors of a positive diagnosis would be useful to make rapid decisions on treatment and isolation needs.
- **Lack of interpretability** In AI research, the model has not limited to accuracy and sensitivity reports; instead, it is expected to describe the predictions' underlying reasons and enhance medical understanding and knowledge. Clinical selection of an algorithm depends on two main factors, its clinical usefulness, and trustworthiness. When the prediction does not directly explain a particular clinical question, its use is limited.

To overcome the limitation of the existing methods, a novel interpretable COVID-19 diagnosis AI framework is proposed in this study, which uses symptoms and cough features to classify the COVID-19 cases from non-COVID-19 cases accurately. A three-layer Deep Neural Network model is used to generate cough embeddings from the handcrafted signal processing features, and symptoms embeddings are generated by a transformer-based self-attention network called TabNet. [3] Finally, the prediction score is obtained by concatenating the Symptoms Embeddings with Cough

Embeddings, followed by a Fully Connected layer. In a sensitive discipline such as healthcare, where any decision comes with an extended and long term responsibility, making wrong predictions can lead to critical judgments in life and death situations.

In this study, it is illustrated that this framework is not limited to accurate predictions or projections. Instead, it explains the underlying reasons for the same and answers the question as to why the model predicts it. The contributions of the paper can be summarized as follows:

- A novel explainable & interpretable COVID-19 diagnosis framework based on deep learning (AI) uses the information from symptoms and cough signal processing features. The proposed solution is a low-cost, rapid, and easily accessible testing solution to increase the diagnostic capability and devise a treatment plan in areas where adequate supplies, healthcare facilities, and medical professionals are not available.
- In this study, an interpretable diagnosis solution is presented, capable of explaining and establishing a dialogue with its end-users about the underlying process. Hence, resulting in transparent human interpretable outputs.
- Three binary and one multi-class classification tasks are developed in this study; Task 1 uses only cough features to classify between COVID-19 positive and COVID-19 negative. In Task 2, only demographic and symptoms data is used. In Task 3, both types of information are used, which helps the model learn deeper relationships between temporal acoustic characteristics of cough sounds and Symptoms' features and hence perform better, and in Task 4, multi-class classification is performed to explain the proposed model's effectiveness in classifying between four cough types, including Bronchitis, Asthma, COVID-19 Positive, and COVID-19 Negative.
- An in-depth clinical analysis is performed on different cough sounds. The observations and findings are presented, distinguishing COVID-19 cough from other types of cough.
- A python module¹ was developed to extract better and re-boost cough features from raw cough sounds. This module is open-sourced to help users, developers, and researchers. Those are not necessarily experts in domain-specific cough feature extraction, contributing to real-time cough based research application, and provide better mobile health solutions.
- This study hence provides a medically-vetted approach.

2 MODEL ARCHITECTURE

The model architecture consists of two subnetworks components, including the Symptoms Embeddings component and the Cough Embeddings component, that process the data from different modalities.

2.1 Symptoms Embeddings

Symptoms Embeddings capture the hidden features of patient characteristics, diagnosis, symptoms.

¹<https://github.com/coughresearch>

Figure 1: Illustration of the overall structure of Proposed Model, The model consists of two subnetworks that process the data from different modalities. The TabNet network at the top process features symptoms & Demographic data, while the network in the bottom processes the audio signal from cough. Attention Masks from each decision step are aggregated to provide model interpretability results.

2.1.1 Gated Linear Unit (GLU). GLU [7] layer consist of Fully connected layer (FC), Ghost batch normalization (GBN) [15] and GLU. FC layer map the input features to $2(n_d + n_a)$ where n_d and n_a are size of the decision layer and size of the attention bottleneck respectively. GBN splits each batch in chunks of virtual batch size(BS), standard Batch normalization(BN) applies separately to each of these, and concatenates the results back into the original batch.

2.1.2 Feature Transformer (FT). The feature transformer is one of the main components of TabNet; it consists of 4 GLU layers, two are shared across the entire network, and two are step-dependent across each decision step allowing for more modeling flexibility. GLU layers are concatenated with each other after being multiplied by a constant scaling factor($\sqrt{0.5}$). Feature Transformer process the filtered features by looking at all the symptoms features assessed and deciding which ones indicate which class.

2.1.3 Attentive Transformer (AT). Attention Transformer is another main component of TabNet architecture. It utilizes sparse intense wise features selection based on learned symptoms dataset and directs the model's attention by forcing the sparsity into the feature set, focusing on specific symptoms features only. It is a powerful way of prioritizing which features to look at for each decision step. The FC handles the learning part of this block. TabNet uses the Sparsemax, [20] an alternative of softmax function for soft feature selection. Sparsemax activation function is differentiable and has forward and backward propagation. Due to projection and thresholding, sparsemax process sparse probabilities lead to a selective and more compact attention focus on symptoms features.

2.1.4 Decision Step (DS). TabNet stacks the subsequent DS one after the other. Decision steps are composed of a Feature Transformer(FT), Attentive Transformer(AT), and feature masking. Symptoms features are mapped into a D-dimensional trainable embeddings $q \in \mathbb{R}^{B \times D}$, where B is the batch size, and D is the feature dimension. A batch normalization (BN) is performed across the whole batch. For the selection of specific soft features and explain the feature importance TabNet uses a learnable mask $M[i] \in \mathbb{R}^{B \times D}$. Each decision step has a specific mask and selects its own features; steps are sequential, so the second step needs the first to be finished.

We obtain the mask output at each decision step by multiplying the mask with Normalised symptoms features $q_i(M[i] \cdot q)$. Normalized domain features are passed to FT, and a split block divides the processed representation into two chunks for the next decision step.

$$[d[i], a[i]] = q_i(M[i] \cdot q) \quad (1)$$

where $d[i] \in \mathbb{R}^{B \times n_d}$ and $a[i] \in \mathbb{R}^{B \times n_a}$. After n^{th} step two outputs are produced.

- Mask outputs are aggregated from all the decision steps to provide model interpretability result Figure 7 and Figure 8 shows the interpretability result.
- The final output is a linear combination of all the summed decision steps, similar to the decision tree result.

$$d_{out} = \sum_{i=1}^{N_{steps}} \text{ReLU}(d[i]) \quad (2)$$

Pairwise dot product was computed between output d_{out} and FC layer to obtain the Symptoms Embeddings $S_e \in \mathbb{R}^{B \times F}$ where B is the batch size, and F is the output dimension.

2.1.5 Symptoms Embedding Loss. TabNet uses regularized sparse entropy loss to control the sparsity of attentive features. The regularization factor is a mathematical aggregation of the attention mask.

$$\text{Loss}_{se} = \sum_{i=1}^{N_{steps}} \sum_{b=1}^B \sum_{j=1}^D \frac{-M_{b,j}[i]}{N_{steps} B} \log(M_{b,j}[i] + \epsilon) \quad (3)$$

Where ϵ is a small positive value.

2.2 Cough Embeddings

Cough Embeddigs learn and capture deeper features in temporal acoustic characteristics of cough sounds.

2.2.1 Signal Preprocessing. Before extracting cough features and feeding it to Deep Neural Networks(DNN), some pre-processing of raw audio data is needed. Each cough recording was downsampled to 16 kHz; normalization was applied to the cough signal level with a target amplitude of -28.0 dBFS to keep the future features as close as possible to the same level.

Normalized features were split into cough segments based on the silence threshold. Let $s[t]$ be the discrete-time cough sound recording. The expression of signal $s[t]$ can be written as:

$$s[t] = y[t] + b[t] \quad (4)$$

Where $y[t]$ denotes the cough signal $b[t]$ denotes the noise in the signal. To reduce the noise and get $y[t]$ a High Pass Filter(HPF) was applied on $s[t]$. Cough segments $y[t]$ were divided into sub segments of non-overlapping Hamming-windowed frames. Let $y_i[t]$, $i = 1, 2, 3, \dots, n$ denotes the i^{th} cough sub-segment from signal $y[t]$ with length N .

2.3 Mel Frequency Cepstral coefficients (MFCCs)

Mel Frequency Cepstral coefficients represent the short-term power spectrum of a signal on the Mel-scale of the frequency. The hearing mechanism of human beings inspires MFCC. The coefficients of MEL-frequency that represent this transformation are called MFCCs. First, we apply the Discrete Fourier Transform(DFT) on each cough sub-segment

$$Y(k) = \sum_{t=0}^{N-1} y_i[t] w(t) \exp(-2\pi i k t / N), \quad k = 0, 1, \dots, N-1 \quad (5)$$

Where N denotes the number of samples in frame, $y_i[t]$ is the discrete time domain cough signal, obtained in Section 2.2.1, $w(t)$ is the window function in time domain and $Y(k)$ is the k^{th} harmonic corresponding to the frequency $f(k) = kF_s/N$ where F_s is the sampling frequency. MFCCs use Mel filter bank or triangular band-pass filter on each cough signal's DFT output, equally spaced on the Mel-scale. At last, we apply the Discrete Cosine Transform(DCT) on the output of the log filter bank in order to get the MFCCs :

$$c(i) = \sqrt{\frac{2}{M}} \sum_{m=1}^M \log(E(m)) \cos\left(\frac{\pi i}{M}(m-0.5)\right) \quad (6)$$

Where $i = 1, 2, \dots, l$ denotes the cepstrum order, $E(m)$ and M are the filter bank energies and the total number of mel-filters, respectively.

2.4 Log energy

To calculate the log energy of each sub-segment, the following formula was used:

$$L_t = 10 \log 10 \left(\epsilon + \frac{1}{N} \sum_{t=1}^N y_i(t)^2 \right) \quad (7)$$

where ϵ is a minimal positive value.

2.5 Zero crossing rate(ZCR)

ZCR is used to calculate the number of times a signal crosses the zero axis. To detect the cough signal's periodic nature, we compute the number of zero crossings for each sub-segment.

$$Z_t = \frac{1}{N-1} \sum_{t=1}^{N-1} \Pi[y_i(t)y_i(t-1)] \quad (8)$$

Where $\Pi[A]$ is a indicator function and is defined as

$$\Pi[A] = \begin{cases} 1, & \text{if } A < 0 \\ 0, & \text{otherwise} \end{cases}$$

2.6 Skewness

Skewness is the third order moment of a signal, Which measures the symmetry in a probability distribution.

$$S_t = \frac{E(y_i[t] - \mu)^3}{\sigma^3} \quad (9)$$

Where μ and σ is mean and stand deviation of the sub-segment $y_i[t]$ respectively.

2.7 Entropy

We compute the Entropy for each sub-segment of the cough signal to capture the difference between signal energy distributions.

$$E_t = - \sum_{t=1}^{N-1} y_i(t)^2 \ln(y_i(t)^2), \quad 1 \leq t \leq N-1 \quad (10)$$

2.8 Formant frequencies

In the analysis of speech signals, Formant frequencies capture a human vocal tract resonance's characteristics. We compute the Formant frequencies by peak picking the Linear Predictive Coding(LPC) spectrum. The first four Formant frequencies(F1-F4) are enough to discriminate various acoustic features of airways.

2.9 Kurtosis

Kurtosis can be defined as the fourth-order moment of a signal, which measures the peakiness or heaviness associated with the cough sub-segment probability distribution.

$$k_t = \frac{E(y_i[t] - \mu)^4}{\sigma^4} \quad (11)$$

Where μ and σ is mean and stand deviation of the sub-segment $y_i[t]$ respectively.

	F1-score	Precision	Sensitivity	Specificity	Accuracy
Covid-19 Positive	86.38 ± 0.03%	81.88 ± 0.01%	91.39 ± 0.04%	97.49 ± 0.03%	96.81 ± 0.05%
Covid-19 Negative	92.16 ± 0.01%	95.09 ± 0.02%	89.41 ± 0.08%	98.64 ± 0.05%	96.55 ± 0.11%
Bronchitis	92.85 ± 0.04%	97.70 ± 0.05%	88.45 ± 0.05%	98.08 ± 0.12%	93.46 ± 0.02%
Asthma	83.88 ± 0.13%	75.46 ± 0.03%	94.41 ± 0.04%	93.10 ± 0.01%	93.34 ± 0.05%
Overall	90.09 ± 0.17%	90.92 ± 0.09%	90.41 ± 0.14%	96.83 ± 0.06%	95.04 ± 0.18%

Table 1: Model performance metrics across four different diseases

2.10 Fundamental frequency(F0)

To estimate the cough sub-segment’s fundamental frequency (F0), we used the center-clipped auto-correlation method by removing the formant structure from the auto-correlation of the cough signal.

A total of 22 cough features were extracted from each cough recording segment $y[t]$, including 12 MFCC features, 4 Formant frequencies, Zcr, Kurtosis feature, Log energy, Skewness feature, Entropy, and Fundamental frequency(F0). The final feature matrix was grouped by chunks of n Consecutive feature matrix, and a total of 44 cough features were extracted by taking the mean and standard deviation for all the cough features in each chunked matrix.

Later we feed the final feature matrix to 3 layers Deep Neural Network(DNN) with ReLu activation function to get the final Cough

2.10.1 Cough Embeddings Loss. In Multi-class classification setting, we use Categorical Crossentropy loss function to calculate the loss of Cough Embeddings

$$\text{Loss}_{ce} = - \sum_{i=1}^N y_i \cdot \log \hat{y}_i \quad (12)$$

where N is the number of classes in dataset, $\hat{y}_i$ denotes the i -th predicted class in the model output and y_i is the corresponding target value. In a binary setting, we use the Binary Cross-Entropy loss function.

$$\text{Loss}_{ce} = - \frac{1}{N} \sum_{n=1}^N [y_n \log \hat{y}_n + (1 - y_n) \log (1 - \hat{y}_n)] \quad (13)$$

2.11 Classification layer

We get the prediction score by concatenating the Symptoms Embeddings with Cough Embeddings followed by a FC layer.

$$\hat{y} = \underbrace{[S_e, C_e]}_{\text{Concatenate}} \cdot FC \quad (14)$$

Figure 1 shows the overall structure of the proposed architecture. After this, the Total loss was calculated as follows.

$$\text{Loss}_{total} = \underbrace{(1 - \alpha) \text{Loss}_{ce}}_{\text{Cough Embeddings Loss}} + \underbrace{\alpha \text{Loss}_{se}}_{\text{Symptoms Embeddings Loss}} \quad (15)$$

Where α is a small constant value to balance the contribution of the different losses.

3 EXPERIMENTS

3.1 Evaluation

In this section, a comprehensive evaluation is carried out to investigate the results of four clinical classification tasks. Based on the dataset collected, the model was trained on the following combination of features.

- **Task 1, Using cough data only** In this experiment setup, only cough features were utilized from the collected dataset to train the Model and distinguish between COVID-19 positive and negative cases. Cough features were extracted using the signal processing pipeline, as described in Section 2.2.1

Figure 2: Healthy Cough

Figure 3: Asthma Cough

Figure 4: Bronchitis Cough

Figure 5: COVID-19 Cough

Figure 6: Four types of cough with their original sound, FFT output, and 1D image representation in Figure 2: Healthy cough, Figure 3: Asthma Cough, Figure 4: Bronchitis cough, and Figure 5: Covid-19 positive cough

- **Task 2, Using Demographic & Symptoms Data Only** In this setup, experiments were conducted on Demographic & Symptoms data. The symptoms (Fever, Headache, aches, sore throat, etc.) were used to train the model and classify between covid-19 positive and negative cases.
- **Task 3, Using Both** when using both types of data, including cough features from section 3 with Demographic & Symptoms data, Model learn the hidden patterns and relationship between both types of features and classify between COVID-19 positive and negative cases.
- **Task 4, Using Both with different Cough Types**, to demonstrate the effectiveness of the Model, the Model is trained on four different types of cough, including COVID-19, Bronchitis, Asthma, and Healthy using both types of data.

		F1-score	Precision	Sensitivity	Specificity	Accuracy
Cough data	Covid-19 Positive	90.6 $\pm$ 0.2%	89.1 $\pm$ 0.4%	86.2 $\pm$ 0.3%	92.4 $\pm$ 0.2%	89.3 $\pm$ 0.1%
	Covid-19 Negative	90.6 $\pm$ 0.1%	91.7 $\pm$ 0.1%	94.3 $\pm$ 0.3%	89.3 $\pm$ 0.1%	92.4 $\pm$ 0.1%
	Overall	90.6 $\pm$ 0.3%	90.4 $\pm$ 0.5%	90.1 $\pm$ 0.6%	90.3 $\pm$ 0.3%	90.8 $\pm$ 0.2%
Symptoms data	Covid-19 Positive	91.5 $\pm$ 0.2%	86.9 $\pm$ 0.5%	87.8 $\pm$ 0.2%	86.0 $\pm$ 0.3%	94.1 $\pm$ 0.6%
	Covid-19 Negative	91.5 $\pm$ 0.1%	93.7 $\pm$ 0.3%	93.3 $\pm$ 0.2%	94.1 $\pm$ 0.1%	86.0 $\pm$ 0.2%
	Overall	91.5 $\pm$ 0.3%	90.3 $\pm$ 0.8%	90.5 $\pm$ 0.4%	90.8 $\pm$ 0.3%	91.1 $\pm$ 0.8%
Both	Covid-19 Positive	96.8 $\pm$ 0.4%	95.1 $\pm$ 0.1%	94.6 $\pm$ 0.3%	95.6 $\pm$ 0.1%	97.3 $\pm$ 0.2%
	Covid-19 Negative	96.8 $\pm$ 0.1%	97.6 $\pm$ 0.3%	97.8 $\pm$ 0.4%	97.3 $\pm$ 0.2%	95.6 $\pm$ 0.3%
	Overall	96.8 $\pm$ 0.5%	96.3 $\pm$ 0.4%	96.2 $\pm$ 0.7%	96.5 $\pm$ 0.3%	96.5 $\pm$ 0.5%

Table 2: Model performance metrics across different models on Covid-19 data

Figure 7: Attention distribution over the Symptoms of a Healthy(COVID-19 Negative) person. The color depth expresses the seriousness of a symptom.

Figure 8: Attention distribution over the Symptoms of a COVID-19 infected person. Fever, Cough, Dizziness, or Confusion, Chest pain is with high color depth, showing that the model has learned the symptoms embedding based on demographic & cough features.

The five standard evaluation metrics (Accuracy, specificity, sensitivity/recall, precision, and F1-score) are adopted to evaluate the Model on the test dataset. Several iterations were performed, and results were reported. Table 1 and Table 2 represents the classification of results along with the following observations:

- The Model achieves Accuracy, Specificity, and Sensitivity of 90.8 $\pm$ 0.2%, 90.3 $\pm$ 0.3%, 90.1 $\pm$ 0.6% respectively. Task 1, using only the cough features. It demonstrates that Cough provides requisite data about the respiratory system and the pathogenies involved. Signal processing characteristics enable the Model to capture the hidden cough sound signatures and diagnose COVID-19 with sufficient sensitivity and specificity.
- When using symptoms and demographic data, the Model's performance represents a slight increase in Accuracy, Specificity, and Sensitivity 91.1 $\pm$ 0.8%, 90.8 $\pm$ 0.3%, and 90.5 $\pm$ 0.4%. Compared to Task 1. This increase is attributed to the rich categorical information present in Demographic and Symptoms data about suspected infection cases and Transformer based network; TabNet exploits attention mechanism to learn the characteristics, diagnosis automatically, and symptoms based on information in their dataset.
- In Task 3, the Model makes better use of available data by combining both types of representation, which complement each other and significantly improve the classification's performance with the Accuracy, Specificity, and Sensitivity of 96.5 $\pm$ 0.5%, 96.5 $\pm$ 0.3% and 96.2 $\pm$ 0.7%. It is observed that one data type, either cough features or symptoms' features,

Figure 9: t-SNE visualization of four types of Cough features

is insufficient to capture the features predicting COVID-19 disease effectively.

- To further demonstrate the proposed Model's flexibility, the Model was trained with a multi-class classification setting to distinguish between four types of cough classes, including Bronchitis, Asthma, COVID-19 Positive, and COVID-19 Negative. The experimental results of Multi-class classification are presented in Table 1. Results show that the Model can certainly classify between four types of cough classes with Accuracy, Specificity, and Sensitivity of $95.04 \pm 96.83 \pm 0.06\%$, $90.41 \pm 0.14\%$ respectively. Both types of data enable the Model to learn deeper relationships between temporal acoustic characteristics of cough sounds and Symptoms' features and hence perform better.

3.2 Interpretability

It is demonstrated that the proposed framework benefits from the high accuracy and generality of deep neural networks and TabNet's interpretability, which is crucial for AI-empowered healthcare. TabNet aggregate the feature masks from all decision steps to provide the importance of the overall symptoms for a single sample. It used a weight coefficient to decide the contribution of features at each decision step. A feature that has been masked a lot has low importance for the model and vice-versa. Figure 7 and Figure 8 visualizes the symptoms of a healthy and COVID-19 infected individual. The color depth indicates the seriousness of a particular symptom or medical condition.

It shows that the model comprehends the hidden pattern in symptoms data and its relationship with cough sound characteristics. To intuitively show the representation's quality, the cough features are visualized in Figure 9.

3.3 In Depth Clinical Analysis

It has been suggested that the progression of COVID-19 symptoms is distinct, in that it most likely starts with fever, followed by a cough due to inflammation of respiratory epithelia [18]. This leads to irritation of the lung tissue that manifests into one of the earliest symptoms of COVID-19, dry cough, which is quite persistent. Compared to non-COVID-19 respiratory illnesses like bacterial and viral

pneumonia, the manifestation of COVID-19 from early-stage involves inflammation confined to the subpleural or peribronchovascular regions of the lungs, showing patchy Ground-Glass Opacities (GGOs) commonly associated with vascular dilation.

The progression of COVID-19 in later stages in the lungs can resemble acute bronchitis as SARS-CoV-2 can spread to the lower respiratory tract and inflame the alveoli thereby paving the way to secondary conditions such as pneumonia and Acute Respiratory Distress Syndrome (ARDS). This shows up as another commonly observed symptom of shortness of breath which happens due to the accumulation of mucus and fluid in the air sacs which hinders the exchange of carbon dioxide with oxygen. The distinct nature of the COVID-19 disease progression in the respiratory tract itself could be indicative that the associated symptoms such as cough-type could also be distinctive in this condition. Different types of cough samples are visualized in Figure 6 and peaks in Figure 14 Based on the analyzed data, the findings are as follows.

The coughing sound consists of three phases- Phase 1- Initial burst, Phase 2- Noisy airflow, and Phase 3- Glottal closure. It is observed that in the cough sample of healthy individuals, phase 3 finished with vocal folds activity. Figure 2 shows that after Phase 1, i.e., initial burst, the energy levels are high at higher frequencies. The peaks in the energy envelope of a healthy cough are also analyzed. Figure 10 shows that the healthy cough sample had the minimum number of peaks compared to the other three cough types. It starts with a peak followed by a float region with two small peaks and falls following a large peak.

Asthma & Bronchitis come under the wet cough (carries mucus and sputum caused by bacteria or viruses, secretion in the lower airways) category. In particular, vocal fold activity looks random, and the energy is expanded over a broader frequency band. Asthma & Bronchitis Figure 3 and Figure 4 shows these characteristics and the energy levels. Figure 11 & Figure 12 shows the peaks in energy envelope of Asthma and bronchitis cough respectively. Cough samples of Bronchitis exhibit the maximum number of peaks. It starts with small peaks and ends with high peaks; multiple small peaks were observed in between in phase 2. Whereas, cough samples of Asthma exhibit fewer peaks than bronchitis' but more peaks than healthy and COVID-19 ones, starting with multiple large peaks in phase 1, ending with small peaks in phase 3.

It is observed that COVID-19 cough is continuous; energy distribution is spread across frequencies preceded by a short catch. By analyzing the energy distribution of multiple COVID-19 cough sounds, it is found that energy distribution was high in Phase 2 and Phase 3. The abnormal oscillatory motion in the vocal folds may be produced by altered aerodynamics over the glottis due to respiratory irritation, Figure 5 shows these findings. Whereas, Figure 13 shows the peaks in the normalized energy envelope of COVID-19 cough. Compared to asthma and bronchitis, it consists of fewer peaks, starting with multiple small peaks and ending with large peaks in phase 3, which shows a significant difference between COVID-19 and other types of cough.

4 DATA ACQUISITION

All the COVID-19 data utilized in this study were obtained from 200 subjects in Dr. RamManohar Lohia Hospital, New Delhi, India.

Figure 10: Peaks in Healthy Cough

Figure 11: Peaks in Asthma Cough

Figure 12: Peaks in Bronchitis Cough

Figure 13: Peaks in COVID-19 Cough

Figure 14: Peaks analysis in four types of coughs, including Healthy, Asthma, Bronchitis, and Covid-19 positive cough. Where y-axis represents Normalized Energy envelope $E(t)$, obtained from applying band-pass filter followed by second-order Butterworth low pass filter, and the x-axis represents the time in seconds.

Out of 100 were confirmed positive from COVID-19 RT-PCR results. The Clinical Trials Registry, India (CTRI) had approved the study protocols and the patient recruitment procedure. After data preprocessing and Out of 200 samples, 50 samples were discarded due to low data quality. Aside from COVID-19 and healthy data, We also collected Bronchitis and Asthma cough from different online and offline sources. The data collection person followed all the clinical safety measures and inclusion-exclusion criteria While collecting the COVID-19 data, and the cough sounds, breathing sounds, counting 1 to 10 (natural voice samples), sustained phonation of 'a', 'e', 'o' vowel, demographic, symptoms data such as fever, headache, sore Thorat, or any other medical conditions were also collected at the same time. The average interaction time with the subject was 10–12 mins.

5 CONCLUSION

Mass COVID-19 monitoring has proved essential for governments to successfully track the disease's spread, isolate infected individuals, and effectively "flatten the curve" of the infection over time. In the wake of the COVID-19 pandemic, many countries cannot conduct rapid enough tests; hence an alternative could prove very useful. This study brings forth a Low cost, accurate and interpretable AI-based diagnostic tool for COVID-19 screening by incorporating the demographic, symptoms, and cough features and achieving mean accuracy, precision, and precision in the mentioned tasks. This significant achievement supports large-scale COVID-19 disease screening and areas where healthcare facilities are not easily accessible. Data collection is being performed daily. Experiments will be carried out in the future by incorporating different voice data features such as breathing sound, counting sound (natural voice samples), and sustained vowel phonation. The results prove to be transparent, interpretable, and multi-model learning in cough classification research.

REFERENCES

- [1] Zeeshan Ahmed, Khalid Mohamed, Saman Zeeshan, and Xinqi Dong. 2020. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database* 2020 (1 2020). <https://doi.org/10.1093/database/baaa010>
- [2] Ioannis D. Apostolopoulos and Tzani A. Mpesiana. 2020. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 43 (6 2020), 635–640. Issue 2. <https://doi.org/10.1007/s13246-020-00865-4>
- [3] Sercan O. Arık and Tomas Pfister. 2020. TabNet: Attentive Interpretable Tabular Learning. *arXiv:cs.LG/1908.07442*
- [4] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. 2020. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data.
- [5] CDC. 2019. centers for disease control and prevention. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
- [6] Joseph Paul Cohen, Mohammad Hashir, Rupert Brooks, and Hadrien Bertrand. 2020. On the limits of cross-domain generalization in automated X-ray prediction.
- [7] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. 2017. Language Modeling with Gated Convolutional Networks. *arXiv:cs.CL/1612.08083*
- [8] Gauri Deshpande and Björn Schuller. 2020. *An Overview on Audio, Signal, Speech, & Language Processing for COVID-19*.
- [9] Yue Du, Roy Zhang, Abolfazl Zargari, Theresa C. Thai, Camille C. Gunderson, Katherine M. Moxley, Hong Liu, Bin Zheng, and Yuchen Qiu. 2018. Classification of Tumor Epithelium and Stroma by Exploiting Image Features Learned by Deep Convolutional Neural Networks. *Annals of Biomedical Engineering* 46 (12 2018), 1988–1999. Issue 12. <https://doi.org/10.1007/s10439-018-2095-6>
- [10] Shannon L. Emery, Dean D. Erdman, Michael D. Bowen, Bruce R. Newton, Jonas M. Winchell, Richard F. Meyer, Suxiang Tong, Byron T. Cook, Brian P. Holloway, Karen A. McCaustland, Paul A. Rota, Bettina Bankamp, Luis E. Lowe, Tom G. Ksiazek, William J. Bellini, and Larry J. Anderson. 2004. Real-Time Reverse Transcription–Polymerase Chain Reaction Assay for SARS-associated Coronavirus. *Emerging Infectious Diseases* 10 (2 2004), 311–316. Issue 2. <https://doi.org/10.3201/eid1002.030759>
- [11] É G Furman, A Charushin, E Eirikh, S Malinin, V Shelud'ko, V Sokolovsky, and G Furman. 2020. THE REMOTE ANALYSIS OF BREATH SOUND IN COVID-19 PATIENTS: A SERIES OF CLINICAL CASES. *medRxiv*.
- [12] Gunes Guclu, Fatma Göğüş, and Bekir Karlik. 2015. Classification of Asthmatic Breath Sounds by Using Wavelet Transforms and Neural Networks. *International Journal of Signal Processing Systems* 3 (10 2015).
- [13] Morteza Heidari, Abolfazl Zargari Khuzani, Alan B Hollingsworth, Gopichandh Danala, Seyedehnafiseh Mirniaharikandehei, Yuchen Qiu, Hong Liu, and Bin Zheng. 2018. Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. *Physics in Medicine & Biology* 63 (1 2018), 035020. Issue 3. <https://doi.org/10.1088/1361-6560/aa11ca>
- [14] Morteza Heidari, Seyedehnafiseh Mirniaharikandehei, Wei Liu, Alan B. Hollingsworth, Hong Liu, and Bin Zheng. 2020. Development and Assessment of a New Global Mammographic Image Feature Analysis Scheme to Predict Likelihood of Malignant Cases. *IEEE Transactions on Medical Imaging* 39 (4 2020), 1235–1244. Issue 4. <https://doi.org/10.1109/TMI.2019.2946490>
- [15] Elad Hoffer, Itay Hubara, and Daniel Soudry. 2018. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. *arXiv:stat.ML/1705.08741*
- [16] Ali Imran, Iryna Posokhova, Haneya N. Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N. John, MD Iftikhar Hussain, and Muhammad Nabeel. 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked* 20 (2020), 100378. <https://doi.org/10.1016/j.imu.2020.100378>
- [17] Juliana Knocikova, J Korpas, M Vrabec, and Michal Javorka. 2008. Wavelet analysis of voluntary cough sound in patients with respiratory diseases. *Journal of physiology and pharmacology* 59 Suppl 6 (10 2008), 331–340.
- [18] Joseph R. Larsen, Margaret R. Martin, John D. Martin, Peter Kuhn, and James B. Hicks. 2020. Modeling the Onset of Symptoms of COVID-19. *Frontiers in Public Health* 8 (8 2020). <https://doi.org/10.3389/fpubh.2020.00473>
- [19] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, Kunlin Cao, Daliang Liu, Guisheng Wang, Qizhong Xu, Xisheng Fang, Shiqin Zhang, Juan Xia, and Jun Xia. 2020. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 296 (8 2020), E65–E71. Issue 2. <https://doi.org/10.1148/radiol.2020200905>
- [20] André F. T. Martins and Ramón Fernandez Astudillo. 2016. From Softmax to Sparsemax: A Sparse Model of Attention and Multi-Label Classification. *arXiv:cs.CL/1602.02068*
- [21] Ali Narin, Ceren Kaya, and Ziyne Pamuk. 2020. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks.

- [22] Lara Orlandic, Tomás Teijeiro, and D Atienza. 2020. The COUGHVID crowd-sourcing dataset: A corpus for the study of large-scale cough analysis algorithms. *ArXiv abs/2009.11644* (2020).
- [23] Danny Parker, Joseph Picone, Amir Harati, Shuang Lu, Marion H. Jenkyns, and Philip M. Polgreen. 2013. Detecting Paroxysmal Coughing from Pertussis Cases Using Voice Recognition Technology. *PLoS ONE* 8 (12 2013), e82971. Issue 12. <https://doi.org/10.1371/journal.pone.0082971>
- [24] Rodolfo M. Pereira, Diego Bertolini, Lucas O. Teixeira, Carlos N. Silla, and Yandre M.G. Costa. 2020. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. *Computer Methods and Programs in Biomedicine* 194 (10 2020), 105532. <https://doi.org/10.1016/j.cmpb.2020.105532>
- [25] Pratik Shah, Francis Kendall, Sean Khozin, Ryan Goosen, Jianying Hu, Jason Laramie, Michael Ringel, and Nicholas Schork. 2019. Artificial intelligence and machine learning in clinical development: a translational perspective. *npj Digital Medicine* 2 (12 2019), 69. Issue 1. <https://doi.org/10.1038/s41746-019-0148-3>
- [26] Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Nirmala R., Prasanta Kumar Ghosh, and Sriram Ganapathy. 2020. Coswara – A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis.
- [27] V. Swarnkar, U. R. Abeyratne, Y. A. Amrulloh, and A. Chang. 2012. Automated algorithm for Wet/Dry cough sounds classification. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3147–3150. <https://doi.org/10.1109/EMBC.2012.6346632>
- [28] Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxin Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, and Yong Xia. 2020. Viral Pneumonia Screening on Chest X-ray Images Using Confidence-Aware Anomaly Detection.