ISSN: 2754-4761

Journal of Pulmonology Research & Reports

SCIENTIFIC Research and Community

Research Article

Open Access

Differentiating Productive and Nonproductive Cough using Cough Acoustics & Machine Learning

Rahul Pathri* and Shekhar Jha

Docturnal Private Limited, India

ABSTRACT

A wet cough by definition is when the cough brings about mucous in the phlegm due an irritant or an inflammation in the Lungs. Every cough regardless of it being wet or dry brings about phlegm and the degree of mucous varies and is much more in quantity in cases of a wet cough & can be used to determine the extent of SOB (shortness of breath), dyspnea due to excessive mucous. This is not always the case in that there are instances where the dry cough subjects may also experience dyspnea that can be attributed to other underlying & related conditions. From an acoustics perspective, a dry cough is usually something like a bland and hacking sound that is followed by an irritation in the Lungs while wet cough has a certain tonality to it whether it is crackling or a rattling sound. Tonality here can be measured using Spectral Flatness & other spectral variables that provide explainability are kurtosis, spectral centroid and spectral skewness. While this is not an exhaustive list, there are other variables that or of importance. Mucous is agnostic to whether it is a bacterial or viral infection but with Pneumonia, there are some indicators with sticky mucous with a bacterial infection Vs viral. However, an indicator of dyspnea induced by phlegm and whether it can differentiate between a bacterial and a viral Pneumonia is an important use case, the later in this context. Wet cough with wheezing caught at an early stage can determine an underlying Emphysema condition [1]. Emphysema: Types, Symptoms, Causes, Diagnosis, Treatment (health.com) or other forms of COPD such as chronic Bronchitis. Cystic Fibrosis has a certain type of mucous that if screened early can be priceless. The shortness of breath caused due to excessive mucus in the Phlegm, if it can rule out Lung Cancer as a screening modality are other areas where the applicability of wet cough can be useful. While the applications are many, in this paper we discuss the AI/ML methods on the acoustics of the "wetness or dryness" of a cough which provides an additional biomarker under the primary biomarker, "cough" in the field of acoustic epidemiology.

***Corresponding author**

Rahul Pathri, Docturnal Private Limited, India. Tel: 91 9676721717.

Received: April 22, 2023; **Accepted:** April 28, 2023, **Published:** May 05, 2023

Introduction

Cough and its associated wetness or dryness is an important measure (biomarker) for the Physician to further delve into the underlying disease diagnosis. Most importantly, many cough syrups are sold for wet cough and dry cough respectively. A very common precursor to the sale of these syrups in the Pharma chains of India is when the Pharmacist asks whether the subject intending to purchase the cough syrup has productive (wet) or nonproductive (dry) cough. Most of the times, the subject is unaware or sometimes has both wet and dry (hybrid) OR in some cases, provides the correct information. The key here is whether a Physician or Pharmacist needing this information without subjectivity is what we are presenting in this paper along with the various tools and techniques around Acoustic AI & Machine Learning. The results are encouraging and can be scaled up using Mobile phones for both diagnosis and treatment monitoring (prognosis). An objective assessment by a subject on their mobile phone that can record at different intervals can track and store this information which in turn can be used by a physician during the patient visit or teleconsultation instead of the subject providing an ambiguous response for the type of cough.

Materials and Methods

1. Chronic cough lasting for more than 8 weeks is surely a definitive path towards understanding the underlying condition and identifying whether the cough is chronic in nature as defined by Learn About Chronic Cough | American Lung Association [2] and provide the correct treatment to the subject. In addition, a more detailed assessment around the wet/dry cough texture shall be a valuable screening tool to rule out other conditions of interest from a differential diagnosis or screening perspective.
2. Here we consider Wet cough as a primary area of interest while still exploring dry coughs. GERD, PNDS, Viral infections infecting epithelial cells of the respiratory system producing mucous, Allergies or taking ACE inhibitors can still produce cough and mucous
3. Providing accurate information to the healthcare provider from the previous day recorded and analyzed coughs for the wet/dry status shall be of great value in accurate investigation and treatment plans thereby reducing time for the physician, patient and the providers is the goal of these findings
4. Data collection involving recording of the voluntary cough sounds using TimBre Android app that was connected to a Zoom H1 Microphone Array as described in Abstract | Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 Leveraging explainable AI/ML (onlinescientificresearch.com) [3]. A filter was used while recording the cough and subjects

were also equipped with a surgical mask to follow an infection control protocol. The subjects were asked to take a deep breath and cough for 5 seconds. The Model testing used cough data that was recorded on a Zoom H1n and Nokia 2.3 Mobile phone running the TimBre app. From the existing data set from [3], labelling of 253 WAV files was conducting using standard headphones by two individuals – 1) RMP (registered/ rural medical practitioner) 2) An individual with an ear for the cough acoustics. The Labels were finalized where there was concordance between the two.

5. A binary classifier was created using Classification Learner app from Matlab R2022a, with wet=1 and dry=0. Given limitations with the data size, "resubstitution validation" was used as the Validation mode and the hyperparameters for the KNN model are represented in the figure A. This was the best model when compared with others and the ROC curve is depicted in the Figure B that used Kruskal Wallis as a feature selection algorithm that gave the best curve with limited labelled data. The AUC was at 0.74 with a sensitivity of 72% and specificity of 73% which was reasonable considering the smaller data set.

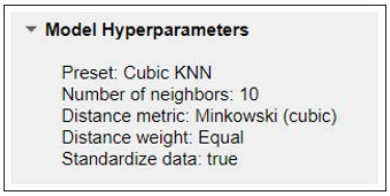

Figure A

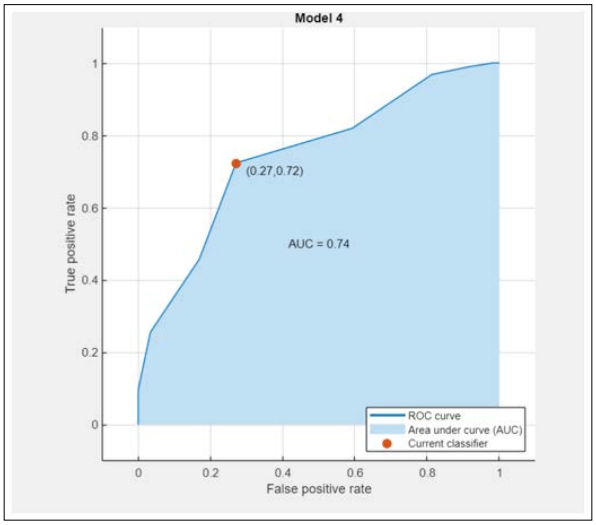

Figure B

Results & Discussion

1. The testing data used a clinical site that was earlier used for Covid19 validation as described in [3]. As per original observation and early reports for Covid19, the cough was perceived as dry cough and there was a mix of both wet and dry in the testing set
2. Where every there were discordant results when compared between the KNN model results and labelled data, the SME's (subject matter experts) reviewed their label and corrected their decision. This is probably an advantage with the "resubstitution validation" technique to identify gaps in data as described at [4] Is cross-validation better than resubstitution for ranking genes? - PubMed (nih.gov)
3. The SME's relabeled the data after a thorough review of their earlier decision post which the testing set yielded a 90% sensitivity and 92% specificity
4. There were cases where the SME's corrected labels that were earlier labelled as wet coughs which however were dry coughs and the initial cough phase was clearing the phlegm with subsequent plosive phases were dry in nature
5. With a small data set, the results look promising and if it can be scaled up, the various healthcare apps can use this as a feature for the subjects to assess whether their cough is dry or wet before going to a pharmacy or physician regardless of the cough condition being Acute or Chronic
6. "People that have a typical bacterial pneumonia will more frequently have a phlegmy kind of a cough," said Dr. Glatt. The mucus they cough up is "dirtier, thicker, and ugly," Dr. Glatt said [5]. By contrast, viral pneumonia often (but not always) produces less phlegm, and people whose immune systems are not working well may not produce any at all, Dr. Glatt said.

Limitations

1. The study used data from an existing clinical trial study published via [5] Abstract | Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 Leveraging explainable AI/ML (onlinescientificresearch.com). This study extensively used Zoom H1 and Zoom H1n Microphone Arrays, Nokia 2.3 & Xiomi Redmi Android mobile phones. Needs to be explored for other mobile phone models and operating systems
2. The supervised learning approach involved two individuals. One being a healthcare worker (RMP) and an individual that has an ear for the acoustics. Physicians OR Pulmonologists need to be involved for data labelling aka golden truth
3. Size of the data for the Machine Learning model needs to be increased for more findings that are circumvented using Machine Learning techniques such as Resubstitution Validation
4. Texture (color) of the sputum is not covered while that could be of great importance
5. Heart diseases causing Pulmonary Oedema & mucus patterns are not discussed which could be of great importance given the cardiac conditions reported post covid19.

References

1. Emphysema: Types, Symptoms, Causes, Diagnosis, Treatment (health.com)
2. Learn About Chronic Cough | American Lung Association
3. Rahul Pathri, Shekhar Jha, Samarth Tandon, Suryakanth Gangi Shetty (2022) Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 Leveraging Explainable AI/ ML. J Pulmonol Res Rep 4: 2-6.
4. Ulisses Braga-Neto, Ronaldo Hashimoto, Edward R Dougherty, Danh V Nguyen, Raymond J Carroll (2004) Is cross-validation better than resubstitution for ranking genes? Bioinformatics 20: 253-258.
5. 8 Symptoms of Pneumonia (health.com)