Detecting Respiratory Diseases

01

BY DANA NICOLAS

02

Outline

Project Background

Background, opportunity areas, potential impact, current status of healthcare, and previous studies.

Project Details

Origin of the dataset, methodologies used, deep learning models, and results.

Recommendations and future applications

Where do we go from here? What can be improved in the future, how can this be used in the present time.

Dana Nicolas

Experience

5 years of technical experience in Mobile Development and a year of experience in Product Management. I've worked on multiple kinds of products and industries such as E-commerce and IoT.

Passion

Health, Education, Conservation, and Transportation industries. Keen in modernising these industries as I believe these are the foundations of any society. This project you can say is my passion project.

Project Background

Healthcare

- Healthcare has lagged behind in terms of IoT integration compared to other industries.
- Using AI to improve and modernise how doctors do diagnosis.
- Much more accurate and uniform way of diagnosing. Minimising misdiagnosis.
- A starting point for IoT integration in the health industry.

2017-18 Distribution of GPs in Australia

Table 2. FSE* GPs per 100,000 people across each State/Territory									
	NSW	VIC	QLD	WA	SA	TAS	ACT	NT	AUS
Major cities	104.7	99.9	113.9	94.8	109	NA***	77.1	NA***	103.5
Inner regional	107.1	103.3	109.8	101.4	78.2	97.9	NP**	NA***	104.2
Outer regional, remote and very remote	NP**	NP**	96.8	77.9	85	84.2	NA***	54.9	89.9
*1 FSE is a 37.5 hour working week									
**NP not published									
***NA not available									

MISDIAGNOSIS

2000-4000 deaths due to misdiagnosis.

In 2013, Australia has reported about 2k-4k deaths per year due to misdiagnosis. That is 1 in 10 people being misdiagnosed.

90%-95% accuracy

On a good day, doctors can get it right 90%-95% of the time. However this is still variable due to a number of factors, one being fatigue.

Misdiagnosis compensation can be from $150k to $3.3M per patient.

Depending on the condition of the patient, the monetary compensation from a medical standpoint is costly. In 2018, a 61-yo was awarded $3.3M in compensation for a misdiagnosis in 2015.

DIAGNOSIS

Big 5

Asthma, COPD, LRTI, Tuberculosis, Lung Cancer are among most common causes of severe illness and death. With millions from all ages being affected yearly.

Spirometry and Chest X-Ray

Most common ways of diagnosing respiratory diseases.

Cost of Diagnosing Respiratory Diseases

The Global Impact of Respiratory Disease – Second Edition (published by European Respiratory Society in 2017)						
People affected / year (million)		25	16	5	10.4	14.1
	Average Cost	Asthma	COPD	LRTI	TB	Lung Cancer
Spirometry	$100	$2,500	$1,600	$500	$1,040	$1,410
Chest X-Ray	$370	$9,250	$5,920	$1,850	$3,848	$5,217
		$11,750	$7,520	$2,350	$4,888	$6,627
Total						$33,135
Total (M)						$33,135,000,000

Cost of Diagnosing Respiratory Diseases

Eliminate 20% of the cost by using AI as the main tool to identify diseases.

$6.63B

11

Potential Impact

Increase speed and accuracy of diagnosis

Improving the speed and accuracy WITHOUT sacrificing precision. And removing human error factor in diagnosing diseases.

Doctors can attend to more patients at a time

By having AI to aid doctors in diagnosing illnesses, they can now attend to more patients at a time. This improves the speed of treatment as well since doctors can start treatment just as soon as they know the diagnosis.

Wider access to Healthcare

By simplifying, making it cheaper and making it more accessible, countries that have limited capabilities and equipment can now have a way of diagnosing patients and thus treat them accordingly.

PREVIOUS STUDIES

From Italy

Diego and Andrea from the University of Calabria, Italy used the same dataset--ICBHI 2017 challenge. In Jul 2019 they came up with a RNN model and used MFCC to extract each sound files' features.

From Japan

A research group from Japan conducted a study with the same dataset. They used a pre-trained VGG-16 on a spectrogram images. Additionally they created scalogram images

From Norway

A research group from Norway used a pre-trained ResNet101 on ImageNet. They converted their respiratory sounds to spectrogram images with varied length. The study was published on April 2019.

13

Additional Information

Respiratory cycles are affected by sex.

According to a couple of studies even from 1985, male and females are different (obv) anatomically including respiratory muscles and organs.

Respiratory cycles are affected by age.

Normal respiratory rates for children are much more different with adults. The younger the person is, the higher their normal respiratory cycles are.

At Medius Health we are building a world where the right to good healthcare for everyone is realized by harnessing the potential of artificial intelligence.

Project Details

ABOUT THE DATASET

- The Respiratory Sound Database was created by two research teams in Portugal and Greece collected independently from 2 different countries over several years.
- It includes **920 annotated recordings** of varying length - 10s to 90s. There are a total of 5.5 hours of recordings containing **6898 respiratory cycles** - 1864 contain crackles, 886 contain wheezes and 506 contain both crackles and wheezes.
- These recordings were taken from **126 patients including diagnosis.**
- The data includes both clean respiratory sounds as well as noisy recordings that simulate real life conditions.
- The patients span all age groups - **children, adults and the elderly.**
- The Respiratory Sound database was originally compiled to support the **scientific challenge organized at Int. Conf. on Biomedical Health Informatics - ICBHI 2017.**

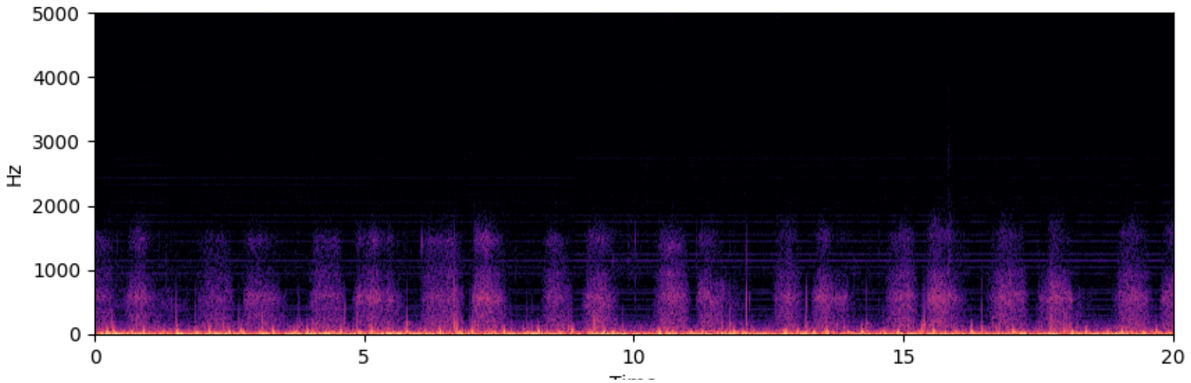
About the dataset

The demographic info file has 6 columns:
- Patient number
- Age
- Sex
- Adult BMI (kg/m2)
- Child Weight (kg)
- Child Height (cm)

COPD	51%
Healthy	21%
URTI	10%
Bronchiectasis	6%
Other (4)	12%

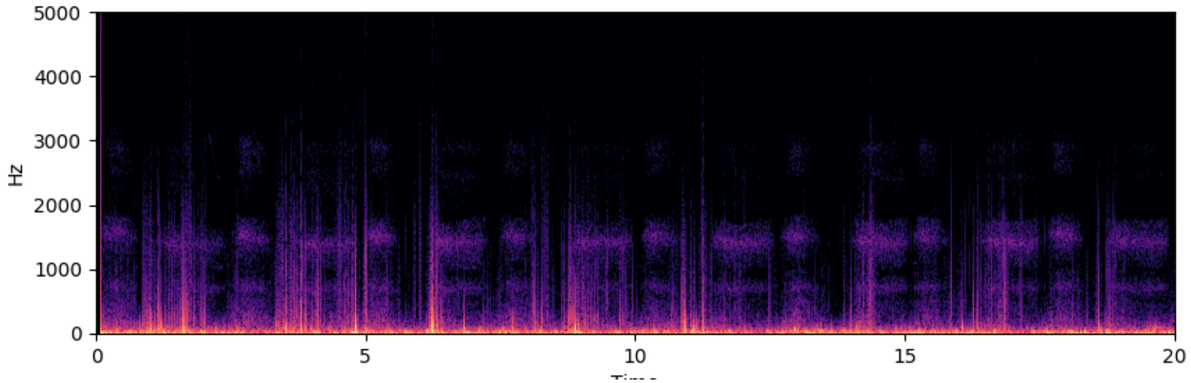
The annotation text files have four columns:
- Beginning of respiratory cycle(s)
- End of respiratory cycle(s)
- Presence/absence of crackles (presence=1, absence=0)
- Presence/absence of wheezes (presence=1, absence=0)

About the dataset



HEALTHY

COPD



Data Preprocessing

Split

Given that we have multiple respiratory cycles in each audio, let's split them to get a 1:1 ratio of a respiratory cycle and its own sound file.

Limit to 6 sec

Setting a uniform limit per sound file.

Spectrogram Images

Convert the files to spectrogram images which will be used for training the model.

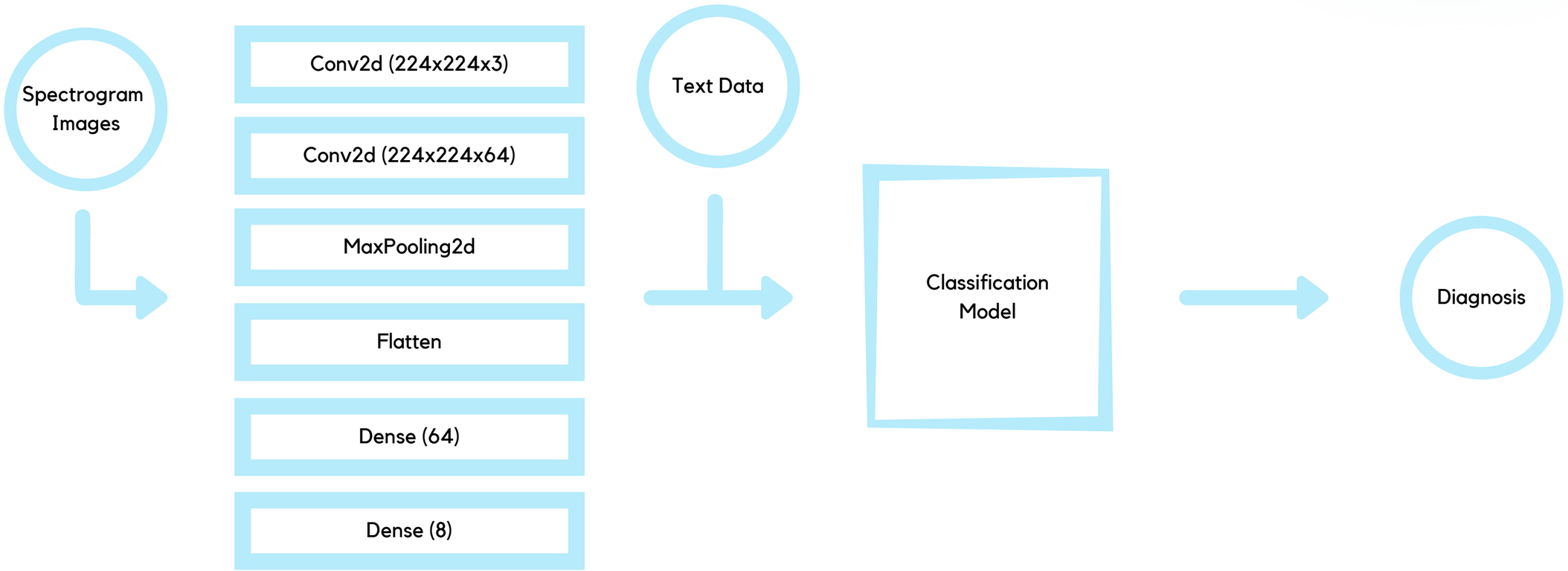
Split to Train-Test

Split the converted images to a 80-20 split.

Prepare Text Data

Text files include patient id, diagnosis, age, and sex.

Deep Learning Models



Result

	Using probabilities as new feature			Using prediction as new feature		
	Accuracy	Recall	F1 Score	Accuracy	Recall	F1 Score
Logistic Regression	45.83%	45.83%	56.24%	44.34%	54.86%	44.34%
Support Vector Classifier	67.07%	72.46%	67.07%	47.12%	57.82	47.12%
Decision Tree Classifier	90.91%	90.91%	91.24%	83.40%	86.45%	83.40%
Random Forest Classifier	90.54%	90.54%	89.48%	87.20%	88.96%	87.20%
Multi-layer Perceptron	90.82%	90.82%	88.58%	91.47%	89.38%	91.47%

Analysis

WHAT DOES THIS MEAN?

90% of the time we can predict the diagnosis accurately. And we are 90% confident of the correctness of each diagnosis predicted.

~90%

Recommendations and Future work

Additional Data

- **Patient Lifestyle**

  Respiratory cycles are also influenced by the lifestyle of the patient. This may include if the patient regularly exercises or if the patient smokes.

- **Health Factors**

  Health Factors such as BMI, and body temperature can also affect respiratory rate/cycles.

- **More Data for other diseases**

  Asthma, Pneumonia, and RTI diseases have too few data (asthma only has 1 patient). It would be a great help if we have more samples for these diseases.

SARS

A contagious and sometimes fatal respiratory illness caused by a coronavirus from 2002-04

MERS

Middle East respiratory syndrome (MERS) is a contagious, sometimes fatal respiratory illness. It's often spread through close contact with an infected person.

COVID-19

A new outbreak of respiratory disease caused by a novel (new) coronavirus that was first detected in China in the late 2019 and which has now been detected in more than 100 locations internationally.

Modern Epidemics

SAMPLE DATA

Detecting Respiratory Diseases | March 2020

MORE EXPLORATIONS

MFCC / STFT

Extracting features using MFCC or STFT or both which can be used to feed AI models.

RNN

Explore using RNN instead of CNN. This includes LSTM, GRUs and different variants of these two.

Filters

Explore the use of filters so that features will be more prominent. Examples of these are Chebyshev, Butterworth, and Ellipsis filters.

References

Dataset

- ICBHI 2017 Challenge

Respiratory Rate

- Factors that affect Respiratory Rate

State of Healthcare

- Distribution of GPs across Australia
- Cost of Misdiagnosis

Other studies

- Predicting respiratory anomalies and diseases via recurrent neural networks
- Convolutional Neural Network for Breathing Phase Detection in Lung Sounds
- A Novel Method for Automatic Identification of Breathing State

References

Diagnosing Respiratory Health

- Most common ways of diagnosing respiratory diseases
- Spirometry
- Chest X-ray

Respiratory Diseases

- Global Impact of Respiratory Diseases

Modern Epidemics

- 2019 Novel Coronavirus
- SARS
- MERS

Thank you

QUESTIONS?