Machine Learning for Breathing Pathology Detection with Emphasis on Bronchiectasis

Lauren Everett, SBHSE

Advisors:
Sunil Rao, ECEE
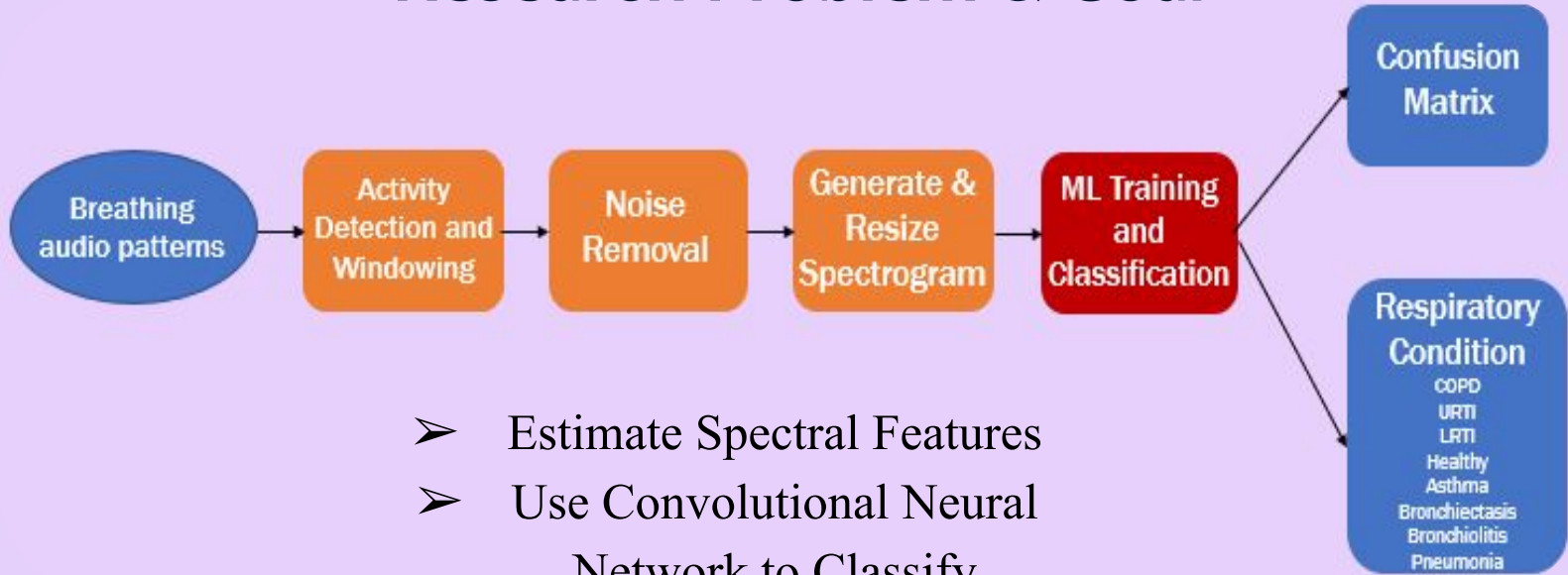**Dr. Andreas Spanias, ECEE**
Michael Esposito, Biological Sciences

July 15, 2021

SenSIP
INDUSTRY CONSORTIUM

1

Research Problem & Goal

Breathing audio patterns → Activity Detection and Windowing → Noise Removal → Generate & Resize Spectrogram → ML Training and Classification → Confusion Matrix / Respiratory Condition (COPD, URTI, LRTI, Healthy, Asthma, Bronchiectasis, Bronchiolitis, Pneumonia)

➢ Estimate Spectral Features
➢ Use Convolutional Neural Network to Classify
➢ Performance Analysis

Motivation

Health

Faster and non-invasive diagnostics

Economic

Average Cost of CT scan in US is $3,275

Technological

Gives rise to remote applications and telemedicine

Challenges

▷ Obtaining Databases

▷ Extracting Robust Features

▷ Real-time Operation

Working with Code

```
#%%
from sklearn.preprocessing import MinMaxScaler
#x_train, x_test = train_test_split(a, test_size=0.2)

x_train, x_test, y_train, y_test = train_test_split(a, Y, test_size=0.3)
#%%
model = Sequential()
model.add(Conv2D(32, kernel_size=(3, 3),
                 activation='relu',
                 input_shape=(128,64,1)))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(num_classes, activation='softmax'))

model.compile(loss=keras.losses.categorical_crossentropy,
              optimizer=keras.optimizers.Adadelta(),
              metrics=['accuracy'])

model.fit(x_train, y_train,
          batch_size=batch_size,
          epochs=epochs,
          verbose=1,
          validation_data=(x_test, y_test))
score = model.evaluate(x_test, y_test, verbose=0)
print('Test loss:', score[0])
print('Test accuracy:', score[1])
```

- Tensorflow Keras

- Simple Convolutional neural network with 3 layers

- Adadelta Optimizer

Figure: Screenshot of section of algorithm used on a covid-19 data set to predict healthy from covid-19 positive patients.

Literature Review

- ▷ Problems with diagnosing Bronchiectasis
  - ▸ Bronchography is invasive
  - ▸ HRCT scans produce false negatives in children
- ▷ AI in healthcare and diagnostics
  - ▸ COVID-19, Parkinson's, TBI, Psychiatric conditions
- ▷ Limited research for use of audio signatures for detecting bronchiectasis

Methods

- ▷ Kaggle Respiratory Audio Database with 920 samples
- ▷ Spectral Estimation + Convolutional Neural Nets
- ▷ VGG-13 deep learning model
- ▷ Challenge: data is heavily biased with COPD
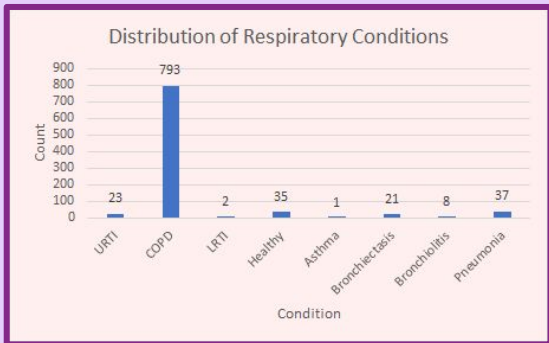
Bronchiectasis
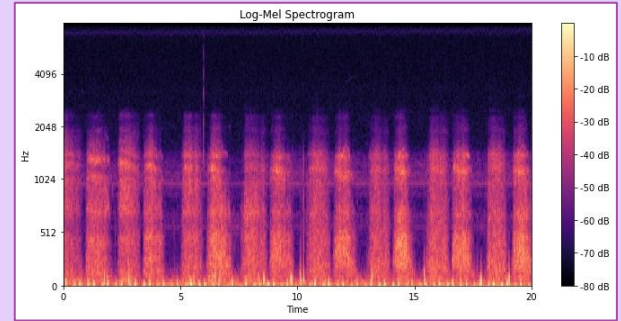
Healthy
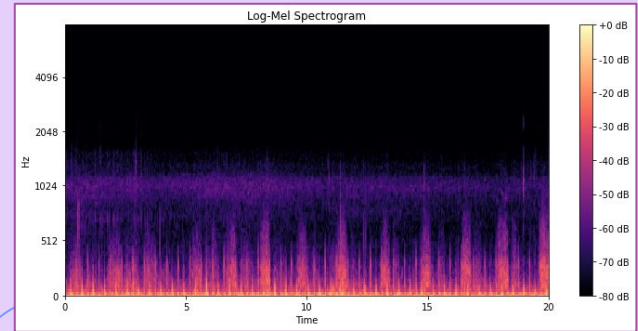
Figure: Top image displays sample spectrogram of an audio signature from a patient with bronchiectasis. Bottom image displays sample spectrogram of audio signature from a healthy patient

Preliminary Results
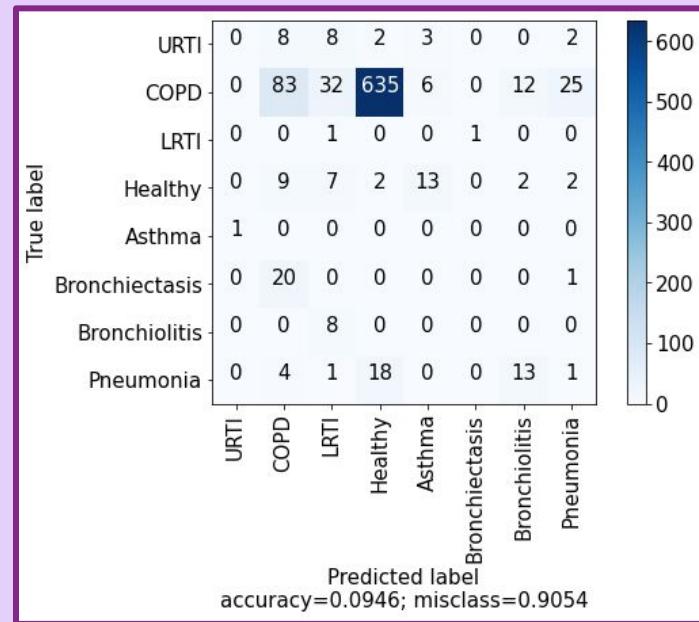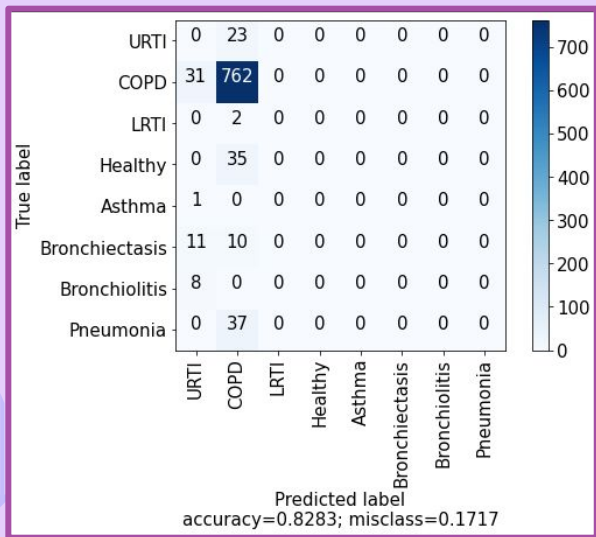

Applying Focal Loss Function to reduce bias

Figure: Top left image displays distribution of data for Kaggle Respiratory dataset. Bottom left displays initial confusion matrix from VGG-13 model. Right image displays confusion matrix after applying loss function.

Final Results & Discussion

▷ Focus on Bronchiectasis

▷ Current Accuracy: 75%

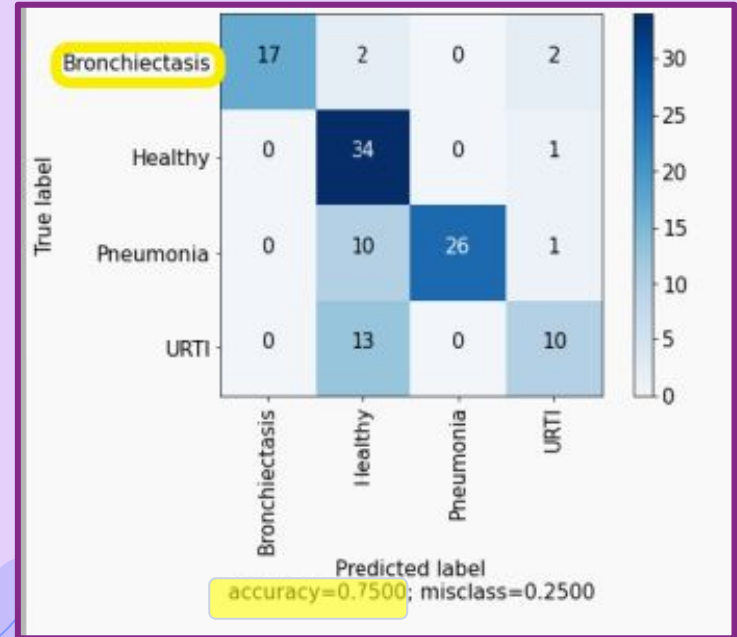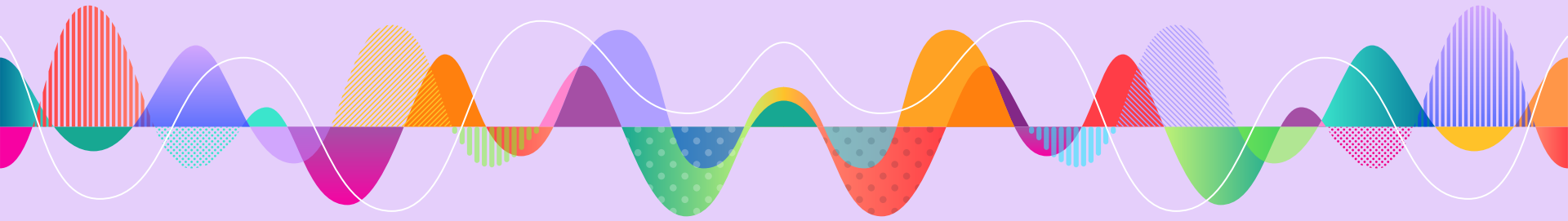▷ Ongoing work to continue tuning hyperparameters to increase accuracy

Figure: Image above displays confusion matrix once COPD data has been removed.

Next Steps and Future Work

▷ Report findings and publish

▷ Expand use of model to other respiratory ailments and datasets

▷ Consider implementation on a smartphone

▷ Develop software for use in clinical setting

▷ Investigate difference of accuracy compared to HRCT scans in children

Thank You!

Any questions?