CAPSTONE PROJECT PRESENTATION

DETECTION OF RESPIRATORY DISEASES USING DEEP LEARNING ALGORITHMS BASED ON RESPIRATORY SOUND ANALYSIS

Team Mentor: Pankaj Agarwal

Team Members:
Praneeth Devanabanda
Forum Joshi
Varnikha Sree A R
Manasa Shetty
B Shyam Sundar

PROBLEM STATEMENT

→ Traditional methods for diagnosing respiratory diseases are subjective and time-consuming, leading to potential misinterpretations and delays in treatment.

→ There is a need for more objective and efficient diagnostic tools to improve accuracy and streamline the detection process.

Importance of Problem:

→ Implementing deep learning models in healthcare settings can standardize diagnostic procedures, reduce subjectivity, and enhance the reliability of diagnoses.

→ Addressing the challenges associated with respiratory disease diagnosis can contribute to overall healthcare system efficiency and cost-effectiveness.

→ Improving the detection of respiratory diseases can lead to earlier treatment initiation, better disease management, and ultimately, improved quality of life for patients.

Approach:

→ Leveraging deep learning techniques, developing hybrid models that combine CNNs for feature extraction and LSTMs for sequential data analysis to improve accuracy and performance.

→ Training deep learning models on extensive datasets of respiratory sound recordings to learn complex patterns and associations with various respiratory conditions.

Technology Used:

→ Utilization of the Librosa machine learning library for feature extraction from audio files, including MFCC, Mel-Spectrogram, and Chroma features.

→ Application of optimization techniques and hyperparameter tuning to improve model performance and accuracy.

→ Implementation of deep learning algorithms, such as CNN-LSTM hybrids, to automate the analysis of respiratory sounds and enhance disease detection capabilities.

PROPOSED SOLUTION AND EDA

Overview of Final Process:

→ We collected audio files using the glob function and merged them with patient diagnosis data to create a dataframe linking audio files to respiratory disease diagnoses.

→ We extracted 189 features from each audio file, addressing class imbalance by using SMOTE to oversample minority classes and downsizing the majority class.

→ After encoding target classes and splitting the data, we applied both machine learning and deep learning models. Our final choice was the CNN-LSTM model, selected for its superior performance in meeting project objectives.

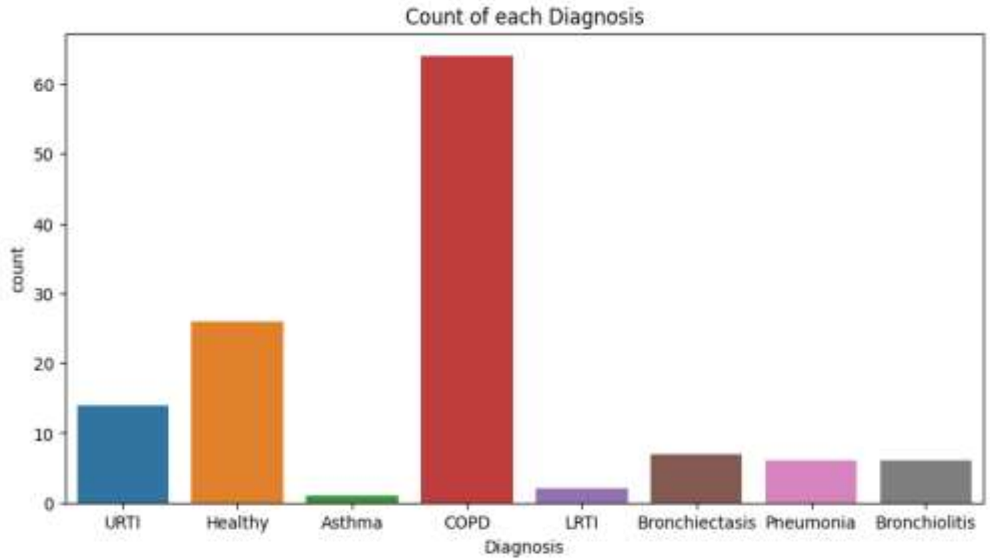
Data Description:

→ We have used the 2017 ICBHI dataset which is a large database of labeled respiratory sounds. It comprises 920 audio recordings totaling 5.5 hours in duration.

→ The dataset encompasses recordings from 128 patients, each identified as either healthy or presenting one of several respiratory diseases or conditions, including COPD, Bronchiectasis, Asthma, upper and lower respiratory tract infections, Pneumonia, and Bronchiolitis. These respiratory condition labels are associated with the audio recording files.

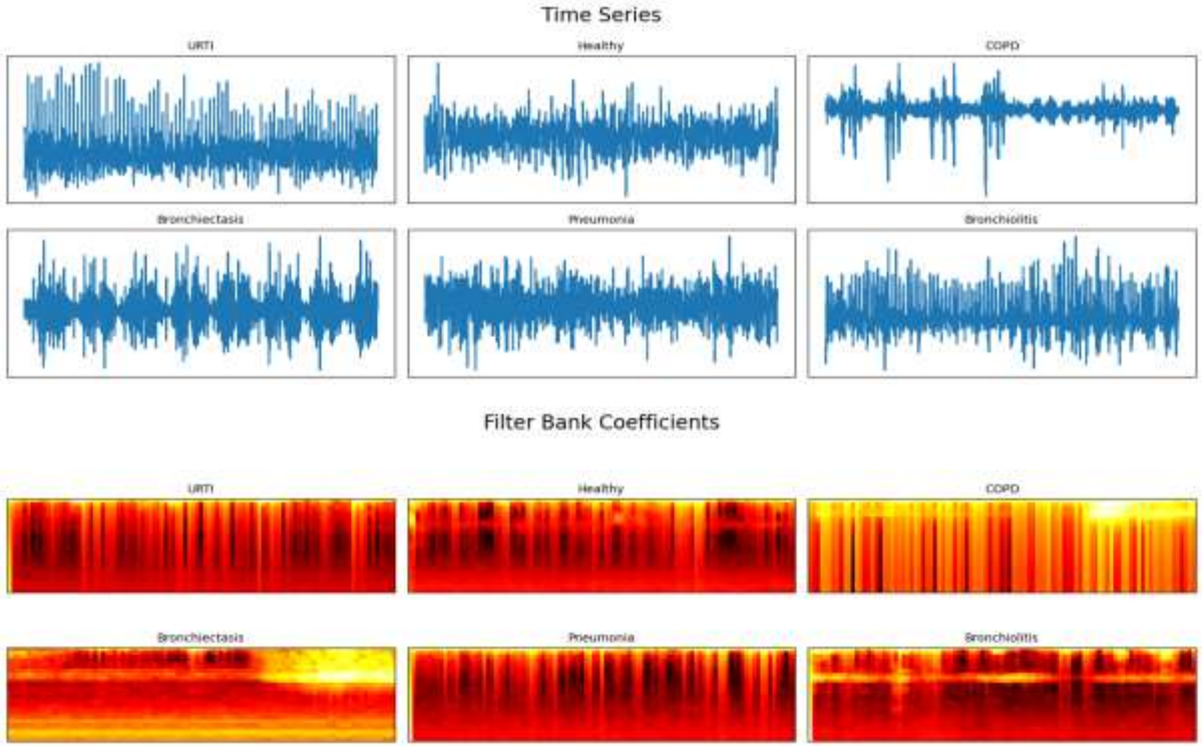
Exploratory Data Analysis:

→ The distributions of respiratory diseases among 126 patients is as follows. Each patient has multiple audio recordings and so target variable's distribution also varies. We have extracted a final data frame with audio files matched with the previous diagnosis information.



	Patient number	audio_file_name	Diagnosis
0	101	101_1b1_Al_sc_Meditron.wav	URTI
1	101	101_1b1_Pr_sc_Meditron.wav	URTI
2	102	102_1b1_Ar_sc_Meditron.wav	Healthy
3	103	103_2b2_Ar_mc_LittC2SE.wav	Asthma
4	104	104_1b1_Al_sc_Litt3200.wav	COPD
5	104	104_1b1_Ll_sc_Litt3200.wav	COPD
6	104	104_1b1_Ar_sc_Litt3200.wav	COPD
7	104	104_1b1_Lr_sc_Litt3200.wav	COPD
8	104	104_1b1_Pl_sc_Litt3200.wav	COPD
9	104	104_1b1_Pr_sc_Litt3200.wav	COPD
10	105	105_1b1_Tc_sc_Meditron.wav	URTI
11	106	106_2b1_Pl_mc_LittC2SE.wav	COPD
12	106	106_2b1_Pr_mc_LittC2SE.wav	COPD
13	107	107_2b3_Pl_mc_AKGC417L.wav	COPD
14	107	107_2b3_Al_mc_AKGC417L.wav	COPD
15	107	107_2b4_Ar_mc_AKGC417L.wav	COPD
16	107	107_2b3_Ar_mc_AKGC417L.wav	COPD

→ Coming to audio files, we looked into some of them, played with sampling rates to observe any differences. Additionally, we created charts using random audio files from each target class. Some of the charts included:



Feature Extraction & Addressing Class Imbalance

Feature Extraction and Engineering:

→ We extract 189 features from each audio file, including Mel-frequency cepstrum coefficients (MFCCs) for spectral characteristics, chromagrams for pitch mapping, Mel-scaled spectrograms for visual representation, spectral contrast for frequency band insights, and tonal centroids for tonal characteristics. These features are crucial for capturing the nuances of respiratory sounds and are calculated using the librosa library.

→ Utilizing a user-defined parser function, we process the entire dataset to extract the 189 features and associate them with the corresponding target labels. This comprehensive approach ensures that each audio file's relevant information is captured effectively, laying the groundwork for subsequent model training and evaluation.

Dealing with Class Imbalance:

→ Addressing class imbalance is an important step as you don't want ML and DL models to get biased towards a single target class. For our data, Approximately 80% of the records belong to the 'COPD' class. To address this imbalance, we utilized the SMOTE function from the Imbalance Learn library. We also reduced the number of records in the 'COPD' class by half.

Target Labels	Bronchiectasis	Bronchiolitis	COPD	Healthy	Pneumonia	URTI
Before	16	13	793	35	37	23
After	100	100	397	150	100	150

→ After addressing the class imbalance, we used the LabelEncoder() to encode the target variable, making it compatible with DL models. Then, we split the data into training and testing sets in an 80:20 ratio.

Models Implemented & Final Model

Ensemble Classifiers:

→ Prior to deep learning models, we investigated ensemble learning techniques based on our mentor's recommendation. Surprisingly, these models showcased remarkable performance in classifying target classes, even in multiclass classification scenarios.

→ All models underwent hyperparameter tuning using GridSearchCV, highlighting the superiority of boosting algorithms over Random Forest. Notably, XGBoost exhibited superior accuracy, with Gradient Boosting slightly surpassing XGBoost in 5-fold cross-validation accuracy.

Model Name	Precision	Recall	F1-Score	Accuracy	CV Scores
Random Forest	95.76	95.5	95.5	95.5	95.29
Gradient Boosting	97.69	97.5	97.53	97.5	96.79
XG Boosting	98.16	98	98.01	98	96.69

Deep Learning Models:

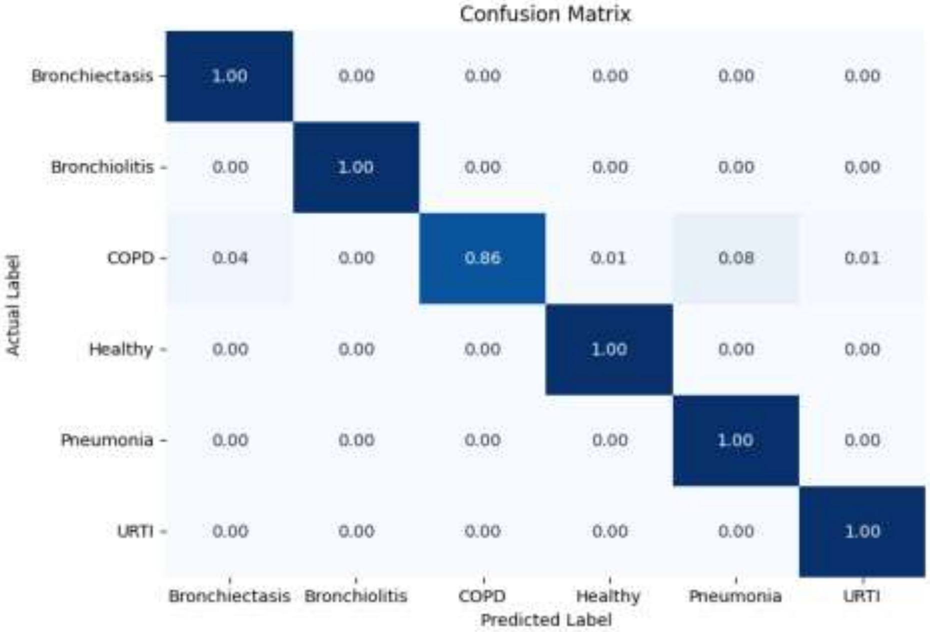
→ Transitioning to Deep Learning models was our primary objective from the inception of the project. While we initially explored simpler neural networks, our ultimate goal was to develop a deep learning model, rather than relying solely on traditional machine learning or ensemble classifiers.

→ Before diving into the final CNN-LSTM architecture, we experimented with two other deep learning models: a pure LSTM model and a pure 1D CNN model. Both models underwent hyperparameter tuning to optimize their architectures. Performance on test is as follows:

Model Name	Precision	Recall	F1-Score	Accuracy
LSTM	0.74	0.77	0.74	0.75
1D CNN	0.925	0.93	0.925	0.9275

Final Model i.e., CNN-LSTM Hybrid:

```
Model: "sequential_4"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_1 (Conv1D)           (None, 187, 128)          512

 conv1d_2 (Conv1D)           (None, 185, 64)           24640

 dropout_7 (Dropout)         (None, 185, 64)           0

 lstm_6 (LSTM)               (None, 185, 128)          98816

 dropout_8 (Dropout)         (None, 185, 128)          0

 lstm_7 (LSTM)               (None, 185, 64)           49408

 dropout_9 (Dropout)         (None, 185, 64)           0

 max_pooling1d_1 (MaxPoolin  (None, 92, 64)            0
 g1D)

 flatten_3 (Flatten)         (None, 5888)              0

 dense_6 (Dense)             (None, 100)               588900

 dense_7 (Dense)             (None, 6)                 606

=================================================================
Total params: 762882 (2.91 MB)
Trainable params: 762882 (2.91 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```



→ The accuracy of the model is 94.5%, with an F-score of 0.948. we can observe that the model achieved a perfect recall score for all target classes except for the 'COPD' class, which is the majority class among the target classes.

LIMITATIONS & CONCLUSIONS

Limitations:

→ While several classifiers could have been employed for this task, We opted for the CNN-LSTM architecture due to its potential to effectively capture temporal and spatial features from audio spectrograms, enhancing respiratory disease classification accuracy.

→ While balancing the dataset using SMOTE addresses class imbalance, it may inadvertently introduce biases and sensitivity to noisy samples, potentially impacting model performance.

→ More hyperparameter tuning often means more complex models. These models might be harder to interpret and explain, which is crucial in many applications, especially in domains like healthcare or finance where decisions have high stakes.

Takeaways & Conclusions:

→ The application of deep learning models, shows promise in the accurate and early detection of various respiratory diseases. These models have the potential to assist healthcare professionals in timely diagnosis and intervention.

→ We learned the intricacies of handling non-tabular data, such as audio files, including feature extraction techniques. Additionally, we gained hands-on experience in implementing and optimizing deep learning models, understanding their inner workings, and fine-tuning hyperparameters to enhance model performance. Overall, this project provided valuable insights into the complexities of DL models and their practical applications.