DALHOUSIE UNIVERSITY
FACULTY OF ENGINEERING

Team 16 : Jeffrey Lyn, Mark Makramalla, and Gavin Barter

Department of Electrical Engineering

Methods to Improve Signal Quality in Oscillometry and Improve Machine Learning Methods for Disease Classification

Introduction

Chronic Obstructive Pulmonary Disease (COPD) causes obstruction in airways making it more difficult to breath. Oscillometry and Spirometry are two methods of COPD diagnosis for clinicians however, they also rely on Patient-Reported Outcomes (PROs) and assessment tests completed by patients to determine if the patient is at risk of having COPD

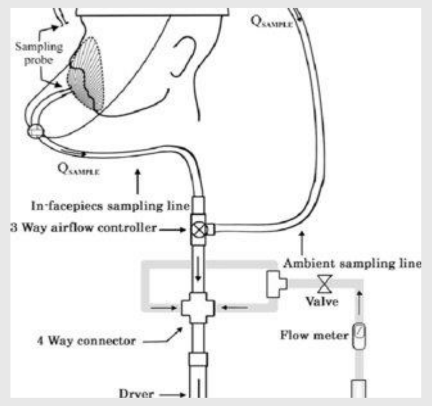

Figure 1. Oscillometry design and setup

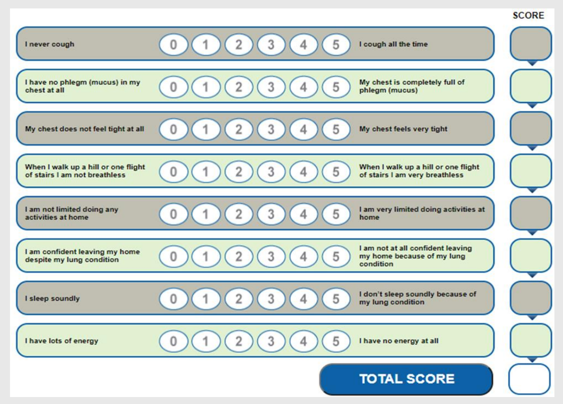

Figure 2. PRO: COPD Assessment Test (CAT)

The figures above are both means of measuring COPD within a patient, Figure 1 is an objective measurement, whereas Figure 2 is the CAT, a patient self-assessment test. The modified British Medical Research Council (mMRC) breathlessness questionnaire is another self-assessment test but is only scored out of 4, whereas CAT is out of 40.

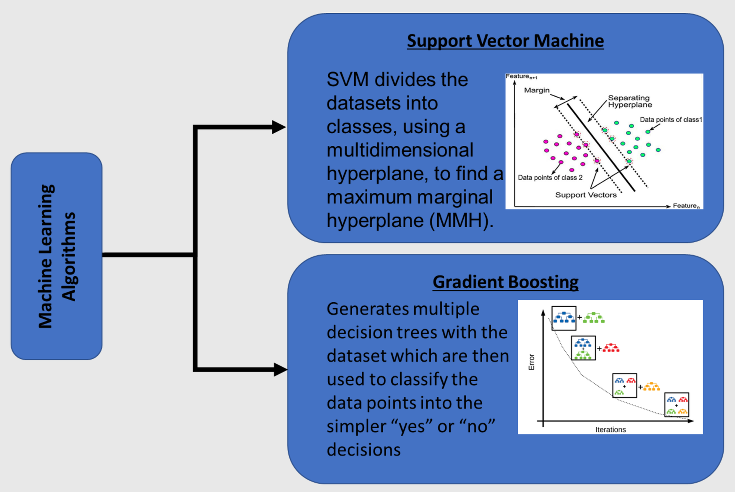

Figure 3. Overview of the two Machine Learning algorithms used

Machine learning is a method used to classify results of oscillometry to provide an objective, patient independent measurement of COPD. The figure to the left provides an overview of the two algorithms used.

Goal: To predict the presence of lung disease from patient reported outcomes using machine learning

Design Process

- A machine learning algorithm will use both PROs and objective measurements from clinician diagnoses to determine a correlation between the two and access the accuracy of PROs when diagnosing COPD
- The CAT system was the chosen PRO to generate our models
  - This is because the CAT is scored out of 40, whereas the mMRC is only scored out of 4 making more difficult to determine a trend based on the scoring
- We imported the SKLEARN library which contains Machine Learning and ROC curve algorithms
  - We then tweaked the parameters of these algorithms for optimization to obtain the highest accuracy score

Details of Design

- **CAT assessments scores, Spirometry, and Oscillometry data from 321 patients (21 Healthy, 300 COPD) were provided at the initial stages of the project. This data was split into 80% for training and 20% for testing**
- **As shown in Figure 4, patient CAT scores were utilized to determine presence of COPD, using GBT and SVM algorithms. This is confirmed by Spirometry and Oscillometry data.**

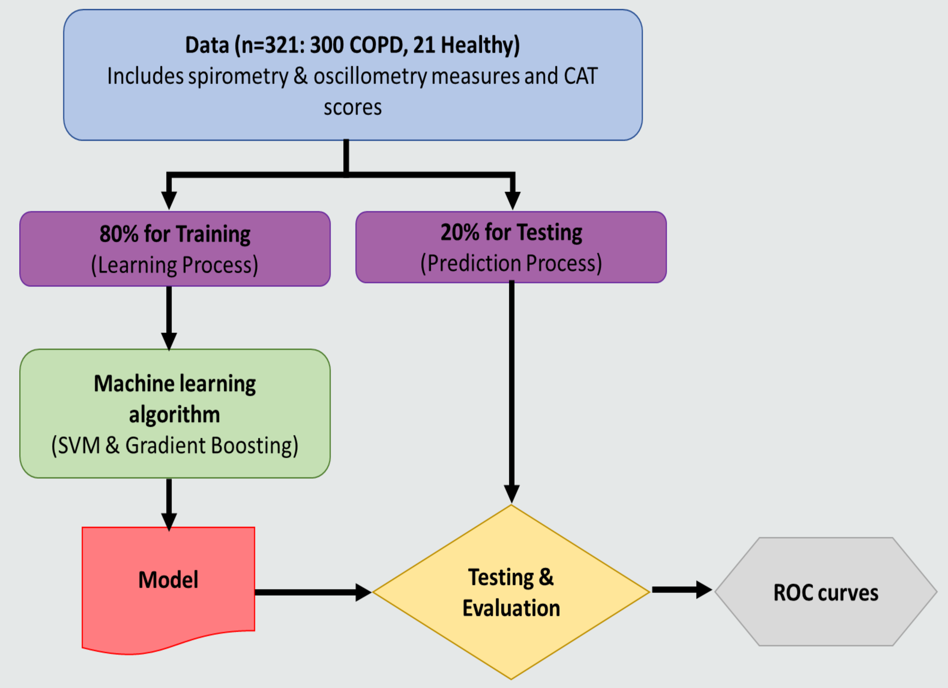

Figure 4. Flow chart of method used to train and evaluate the different machine learning models

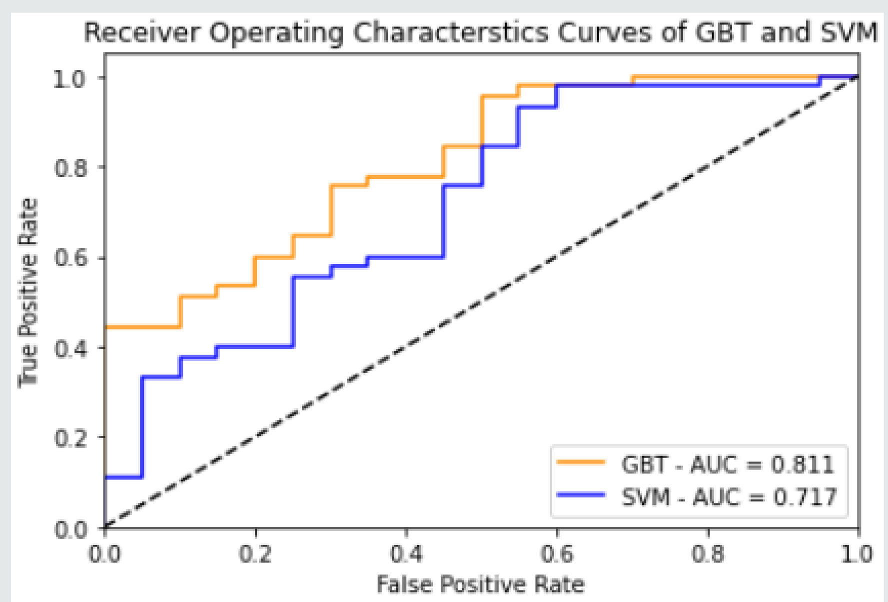

Figure 5. Plot of Receiver Operating Characteristics (ROC) curve utilizing Gradient Boosting model (Orange), and SVM model (Blue). NOTE: Multiple trials of algorithms generate the same ROC curves.

- **The performance of each algorithm was analyzed and compared at different thresholds to determine result predication accuracy. This was accomplished by utilizing the entire dataset and holding all randomization parameters.**
- **As shown in Figure 5, the area under the curve (AUC) of GBT (AUC = 0.811) is greater than that of SVM (AUC = 0.717), indicating GBT model outperforming the SVM model**
- **Running multiple trials of the same algorithms generated the same ROC plot shown in Figure 4, indicating stability of code.**

- **Specifying randomized and equal portions of healthy and COPD patients using the 80:20 split (80% healthy + 80% COPD for training ; 20% healthy + 20% COPD for testing), the variability of using the GBT model was assessed**
- **Performing 20 randomized trails, the ROC curves, and an average ROC curve were plotted, as shown in Figure 6.**
- **From Figure 6, there exists noticeable variability in multiple ROC plots, indicating more optimization and analysis of model are still available**

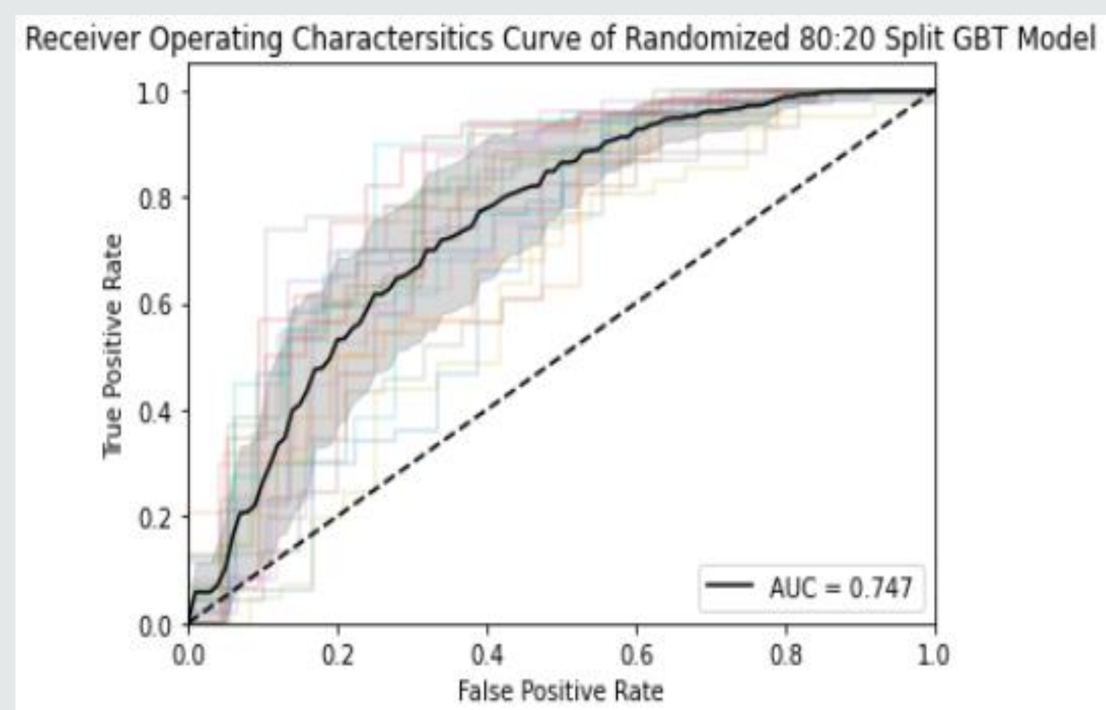

Figure 6. Plot of an Average Receiver Operating Characteristics (ROC) curve of 20 trials of randomized 80:20 dataset, utilizing Gradient Boosting model

Conclusion and Recommendations

- Based on the dataset and the parameters specified in the Machine Learning mode, Gradient Boosting algorithm outperforms Support-Vector Machine algorithm
- Multiple trials of utilizing the same parameters and dataset show no changes in accuracy scores or ROC curve properties. This shows that coding the algorithms do not add error or noise to results, indicating variability in data as only variable to changes in accuracy scores.
- Randomization of equal portions of healthy and COPD patients in the specified 80:20 split with the Gradient Boosting algorithm show noticeable variability, indicated by the ROC curves and the average ROC curve. This indicates more optimization or better modelling may still be available.
- Recommendations:
  - Utilize Receiver Operator Characteristic Curves to determine optimal CAT score threshold that provides greatest accuracy in prediction.
  - Obtain and utilize different datasets of CAT scores, Spirometry data, and Oscillometry data to compare and analyze the effects on accurate predictions of the Machine Learning model.

Acknowledgements

The authors are thankful to our supervisors Dr. Geoff Maksym, Anas Abufardeh, and Dr. Sergey Ponomarenko for their support

References

- *Handheld portable impulse oscillometer*. USPTO.report. (n.d.). Retrieved March 27, 2022, from https://uspto.report/patent/grant/10,492,711
- Oostveen, E., MacLeod, D., Lorino, H., Farré, R., Hantos, Z., Desager, K., & Marchal, F. (2003, December 1). *The forced oscillation technique in clinical practice: Methodology, recommendations and future developments*. European Respiratory Society. Retrieved March 27, 2022, from https://erj.ersjournals.com/content/22/6/1026
- Reii NakanoReii Nakano 1, MonaMona 32122 silver badges33 bronze badges, Cherry WuCherry Wu 3, Yohann L.Yohann L. 1, MaxMax 1, & Brian ChanBrian Chan 4111 bronze badge. (1962, May 1). *How to plot ROC curve in python*. Stack Overflow. Retrieved March 27, 2022, from https://stackoverflow.com/questions/25009284/how-to-plot-roc-curve-in-python
- SABHARWAL, A. (2017, August 10). *Forced oscillation technique based lung function testing*. WO2017136639 FORCED OSCILLATION TECHNIQUE BASED LUNG FUNCTION TESTING. Retrieved March 27, 2022, from https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2017136639