Comparing Predictive Machine Learning Models' Capacity to Objectively Predict Dyspnea

Hannah Liu[1], Hersh Sagreiya[2], Alireza Akhbardeh[3,4]

University of Pennsylvania[1], Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA[2],
Department of Diagnostic and Interventional Imaging, The University of Texas Health Science Center, Houston, TX, USA[3], Ambient Digital LLC, Daly City, CA, USA[4]

Introduction

- Dyspnea is a symptom characterized by shortness of breath and a major predictor of mortality for patients suffering from serious respiratory illnesses
- It must be treated quickly, as failure to address it can result in respiratory failure or death
- It is currently assessed via subjective numerical rating scales based on perceived level of breathing effort
- **Purpose:** to develop machine learning models that will automatically and accurately estimate a patient's breathing exertion levels; to provide doctors with real-time monitoring of dyspnea

Materials and Methods

- We obtained clinical data from a prospective cohort study that collected cerebral hemodynamic changes and vital signs from COPD patients while they performed treadmill walking tests; these became our two feature sets
- Scored based on the Borg Rating of Perceived Exertion Scale
- Utilized Decision Tree Regressors, Random Forest Regressors, XGBoost Regressors, CatBoost Regressors, and LightGBM regressors; total of 22 ML algorithms
- Both 5-fold cross-validation (k = 5) and train test split validation for each model (except for LightGBM)
- Prediction target was the max Borg score given by the patient within each minute interval
- Decision Tree Regressors were also tested with and without optimal leaves
- Accuracy of each model was the mean absolute error (MAE) between each model's prediction values and the subjective dyspnea scores given by patients

Please email hmliu@sas.upenn.edu about any questions.

Vital Signs Feature Set (CV = cross-validation, TTS = train test split)

Model Type	Mean Absolute Error (MAE)
Decision Tree w/ optimal leaves TTS	1.631
Decision Tree w/ optimal leaves CV	2.978
Decision Tree w/o optimal leaves TTS	1.505
Decision Tree w/o optimal leaves CV	3.001
Random Forest TTS	1.202
Random Forest CV	2.345
XGBoost TTS	1.210
XGBoost CV	2.563
CatBoost TTS	1.143
CatBoost CV	2.432
LightGBM TTS	1.395

Cerebral Hemodynamic Feature Set (CV = cross-validation, TTS = train test split)

Model Type	Mean Absolute Error (MAE)
Decision Tree w/ optimal leaves TTS	1.610
Decision Tree w/ optimal leaves CV	2.463
Decision Tree w/o optimal leaves TTS	1.407
Decision Tree w/o optimal leaves CV	2.921
Random Forest TTS	1.144
Random Forest CV	2.369
XGBoost TTS	1.126
XGBoost CV	2.575
CatBoost TTS	1.087
CatBoost CV	2.551
LightGBM TTS	1.379

Figure 1 (left): MAE values for each predictive model type trained on the vital signs feature set
Figure 2 (right): MAE values for each predictive model type trained on the cerebral hemodynamic feature set

Results

- For both cerebral hemodynamic and vital sign feature sets, CatBoost and Random Forest regressors performed the best
  - CatBoost yielded the lowest TTS validation MAE and second lowest CV MAE
  - Random Forest yielded the second lowest TTS validation MAE and lowest CV MAE
- Cross-validation MAE scores for both feature sets were much higher than train test validation MAE scores

Conclusions

- The performance of our machine learning models helps establish a foundation for more predictive model work focusing on objectively predicting dyspnea
- The subjective Borg rating school used utilizes values 6 to 20, so having MAE values ranging from 1 to 3 displays a promising start to our research
- Future research plans include implementing techniques such as applying dimensionality reduction on input variables, using correlation coefficients to measure model performance, further fine-tuning model parameters, and integrating leave-one-out cross validation

Acknowledgments

- Special thanks to the National Science Foundation Research Experience for Undergraduates for funding this opportunity