Paper RE06

Applications of Machine Learning and Artificial Intelligence in Real-World Data in Personalised Medicine for Non-Small Cell Lung Cancer Patients

Samiul Haque, PhD, SAS Institute, Cary, NC, USA
Sherrine Eid, MPH, SAS Institute, Cary, NC, USA
Robert Collins, SAS Institute, Cary, NC, USA

ABSTRACT

In this work, we present a platform approach for analyzing Real World Data (RWD) that provided faster insight for personalized medicine for Non-Small Cell Lung Cancer (NSCLC) patients. NSCLC is the leading cause of cancer death. Personalized medicine is a viable option to improve patient outcomes when patient journeys are illustrated using Machine Learning (ML) and Real World Data (RWD).Chronic Obstructive Pulmonary Disease (COPD) persists during cancer treatment and requires interventions negating the effectiveness of the treatment. Patient journeys for 1.2M NSCLC (ICD-10 C34.*) patients from 2019-2020 using Symphony Claims (ICON Plc) were established and associations were examined using ML on SAS Viya 4.0 (SAS Institute, Inc.)Over 39% of patients had Metastatic Cancer (MC). The factors most related to MC were Dehydration, Weight Loss, Coagulopathy and Liver Disease, respectively. Over 30% of patients also had COPD and experienced an impact on their care. NSCLC patients should be assessed for comorbidities to better anticipate secondary treatment needs that may interfere with their cancer treatment providing opportunities for personalized medicine and better prognoses. Finally, we demonstrate that all aspects of the analysis, including cohort building, data visualization, machine learning, and governance, can be efficiently achieved using a data science platform approach to conduct real-world data (RWD) analysis. This approach may enable researchers to generate insights more rapidly than existing workflows.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a major public health problem. In 2020, COPD was projected to rank fifth worldwide in terms of disease burden and third in terms of mortality. Lung cancer is also one of the leading causes of death in many countries. Several clinical reports have shown that the proportion of deaths from lung cancer in patients with COPD ranges from 4 to 33%. The frequency of coexisting COPD has been reported to be 40–70% among lung cancer patients.

Both COPD and lung cancer are major worldwide health concerns owing to cigarette smoking, and represent a huge, worldwide, preventable disease burden. Whilst the majority of smokers will not develop either COPD or lung cancer, they are closely related diseases, occurring as co-morbidities at a higher rate than if they were independently triggered by smoking.

Chronic Obstructive Pulmonary Disease (COPD) is a progressive and ultimately fatal deterioration of lung function over time. COPD has a marked effect on a patient's quality of life affecting up to 50% of smokers. COPD was the third most common cause of death worldwide in 2010 and ranked fifth worldwide in terms of burden of disease. Damage to the lungs in COPD is caused by oxidative stress (both exogenous from smoking and endogenous), inflammatory cytokine release, protease activity (due to the protease: anti-protease imbalance) and autoantibody expression. These in turn can lead to airway destruction, air trapping and lung hyperinflation. COPD commonly persists during cancer treatment and requires interventions that could jeopardize the effectiveness of the treatment and their prognosis.

Lung cancer and COPD may be different aspects of the same disease, with the same underlying predispositions, whether this is an underlying genetic predisposition, telomere shortening, mitochondrial dysfunction or premature aging. In the majority of smokers, the burden of smoking may be dealt with by the body's defense mechanisms: anti-oxidants such as superoxide dismutases, anti-proteases and DNA repair mechanisms. However, in the case of both diseases these fail, leading to cancer if mutations occur or COPD if damage to the cell and proteins becomes too great.

Alternatively, COPD could be a driving factor in lung cancer, by increasing oxidative stress and the resulting DNA damage, chronic exposure to pro-inflammatory cytokines, repression of the DNA repair mechanisms and increased cellular proliferation.

Among patients diagnosed with NSCLC, those with pre-existing COPD have poorer median overall survival compared with patients without comorbid disease (192 days vs 206 days) and have an 11% higher risk of death. As of 2018, lung cancer was the most commonly diagnosed cancer and leading cause of cancer death. Most cases are Non-Small Cell Lung Cancer (NSCLC). Machine learning methods (ML) and Real World Data (RWD) make personalized medicine a viable possibility to save patients' lives. We explored patient journeys and novel opportunities to save lives.

METHODS

Traditional RWD analysis often involves a significant amount of data transfer from the source to the compute environment, which we consider to be a major bottleneck in generating insights from RWD. Additionally, traditional workflows result in multiple copies of large datasets, complicating the governance of data and analytics. In this work, we adopt a modern approach by leveraging a cloud-native platform that enables efficient connections to data sources, cohort building, and visualization of results through interactive dashboards.

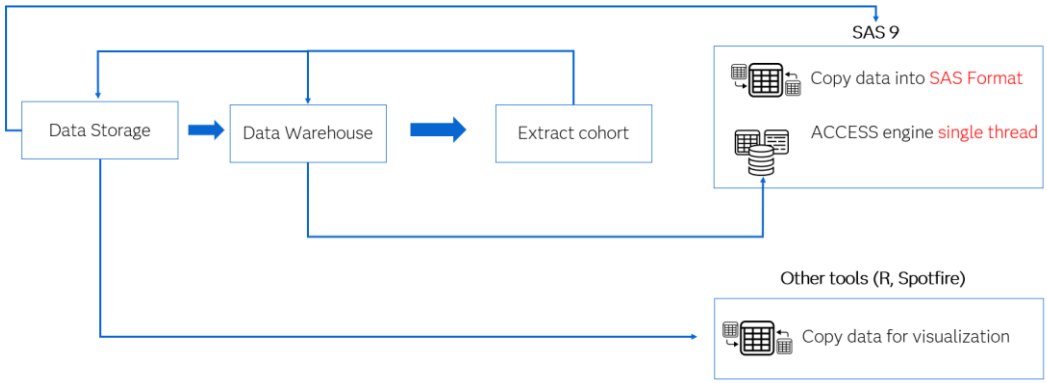

Figure 1: Typical Workflow for RWD Analysis

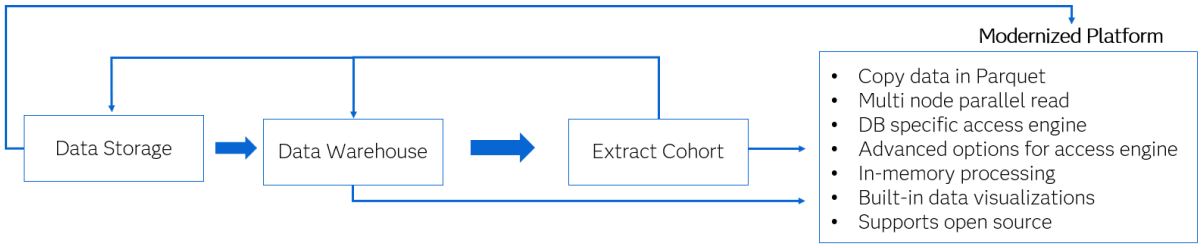

Figure 2: A Platform approach to RWD analysis

This study included 1.2 Million NSCLC (ICD-10 C34.*) patients from 2019-2020 using Symphony Health Claims (Symphony Health Solutions, ICON Plc). Metastatic cancer (MC) was the dependent variable. Independent variables included comorbidities as defined by Elixhauser Comorbidity Risk Factor categories, age, sex and drug product. Automated explanation explored the relationships between comorbidities and metastatic cancer. Patient journeys using diagnoses were established with Path Analysis. Likelihood of MC was determined using ML and artificial intelligence (AI) methods and compared using KS (Youden) and ROC. Machine learning algorithms used in this work included logistic regression, gradient boosting, decision tree, random forest and Bayesian networks. Network analysis explored the relationship between each comorbidity and MC. All analyses were done using SAS Viya 4.0 (SAS Institute, Inc.).

RESULTS

Network analysis illustrated the strong relationships between Solid Tumors and Metastatic Cancer with comorbidities such as Chronic Pulmonary Disease, and Hypertension (Figure 3). Over 39% of patients were diagnosed with MC (Figure 4). The factors most related to MC were Fluid and Electrolyte Disorders, Weight Loss, Coagulopathy and Liver Disease, respectively (Figure 4).

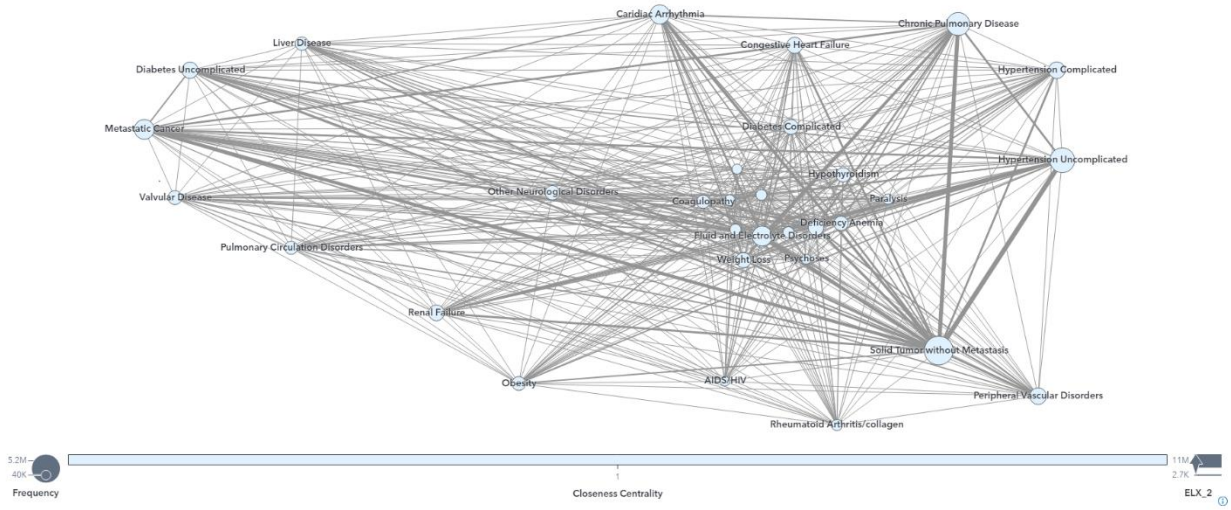

Figure 3: Network Analysis of Comorbidities in Patients with Non-Small Cell Lung Cancer

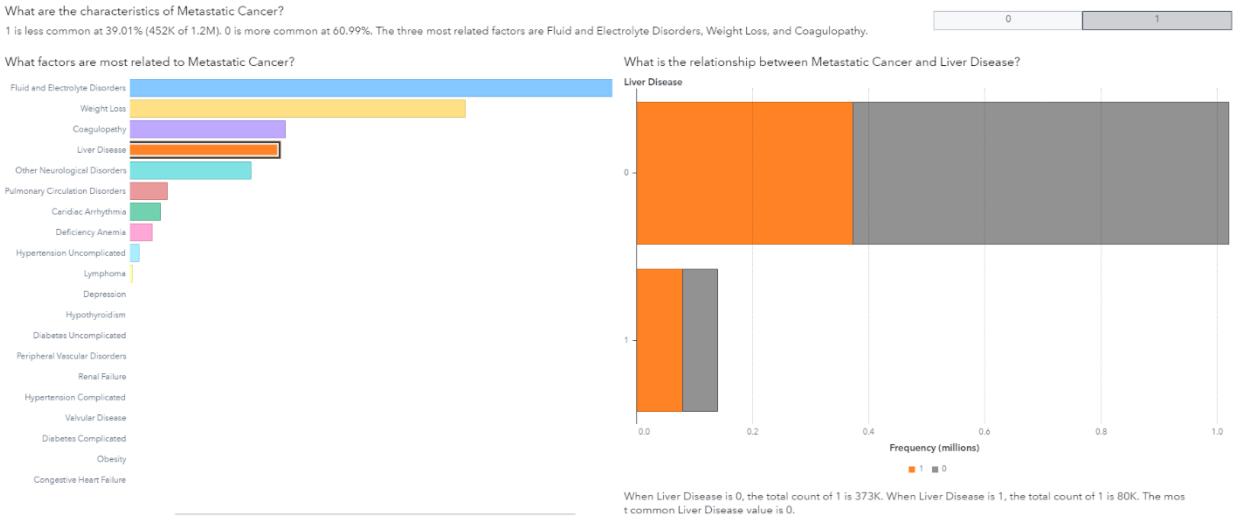

Figure 4: Automated Explanation of Likelihood of Metastatic Cancer in Patients with Non-Small Cell Lung Cancer

Figure 5 illustrates the correlations between comorbidities in patients with NSCLC. Some correlations are expected such as renal failure, congestive heart failure and hypertension. However, there seem to be a relationship between COPD and Solid Tumors.

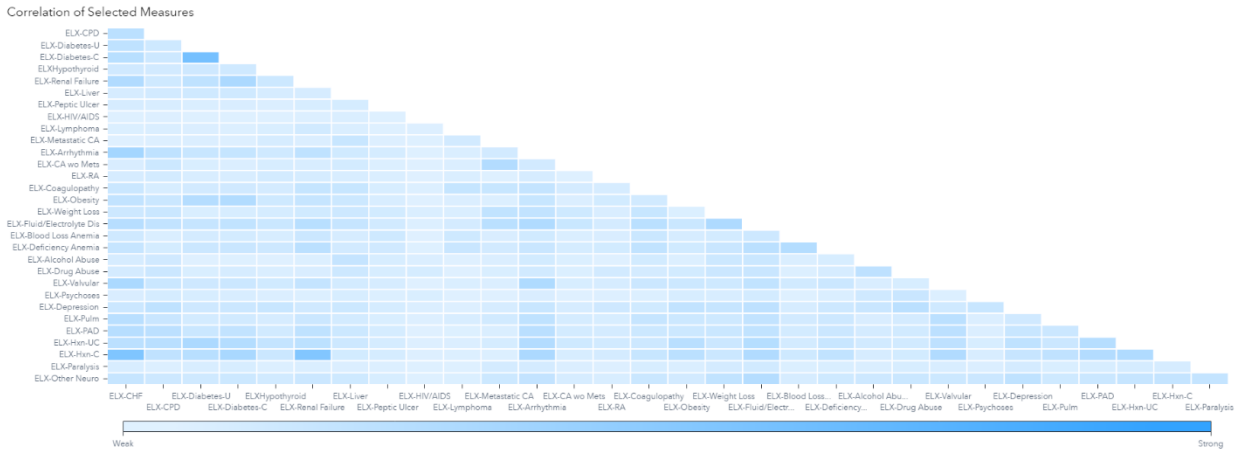

Figure 5: Correlation Matrix of Comorbidities in Patients with Non-Small Cell Lung Cancer

Furthermore, path analyses (Figure 6 and Figure 7) illustrated the impact of COPD on these patients and their care. This work illustrates that a major segment of this patient population had a comorbid diagnosis of COPD in addition to their NSCLC diagnosis and continued to require treatment for both, COPD, as well as NSCLC.

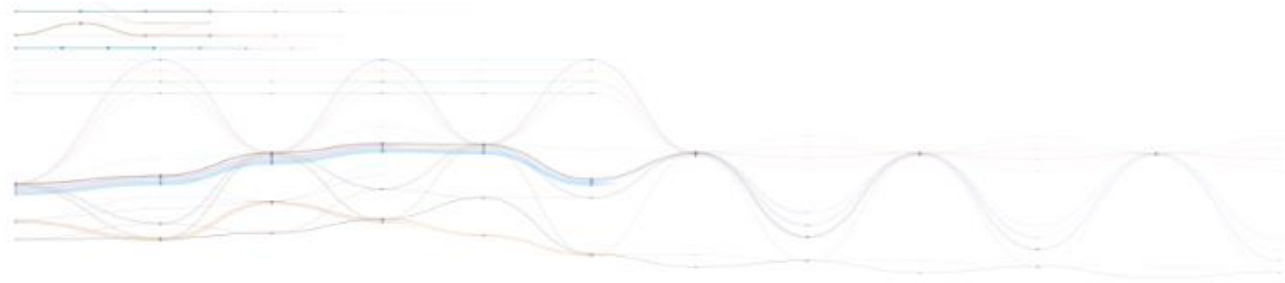

Figure 6: Patient Journey Based on Diagnosis in Patients with Non-Small Cell Lung Cancer

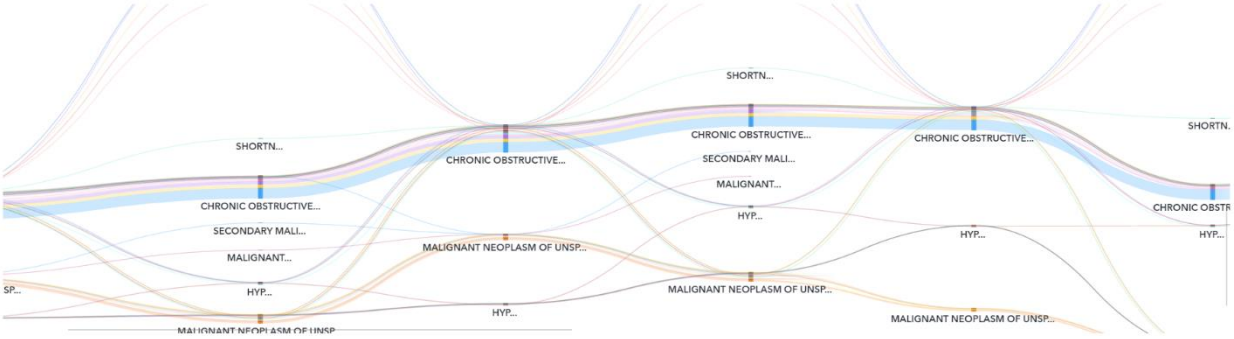

Figure 7: Patient Journey in Patients Who Are Diagnosed with Chronic Obstructive Pulmonary Disease and Non-Small Cell Lung Cancer

Machine learning algorithms were utilized to assess adjusted likelihood of and associations with MC and comorbidities and demographic variables. Conventional logistic regression demonstrated an overfit model with compared to the other models (Figure 8). These parameters were duplicated in a Gradient Boosting model, Decision Tree, Random Forest, and Bayesian Network, respectively (Figure 9, Figure 10, Figure 11, and Figure 12). The majority of the models showed the same comorbidities as most strongly associated with MC -namely, Fluid and Electrolyte Disorders, Weight Loss, Coagulopathy and Liver Disease, respectively.

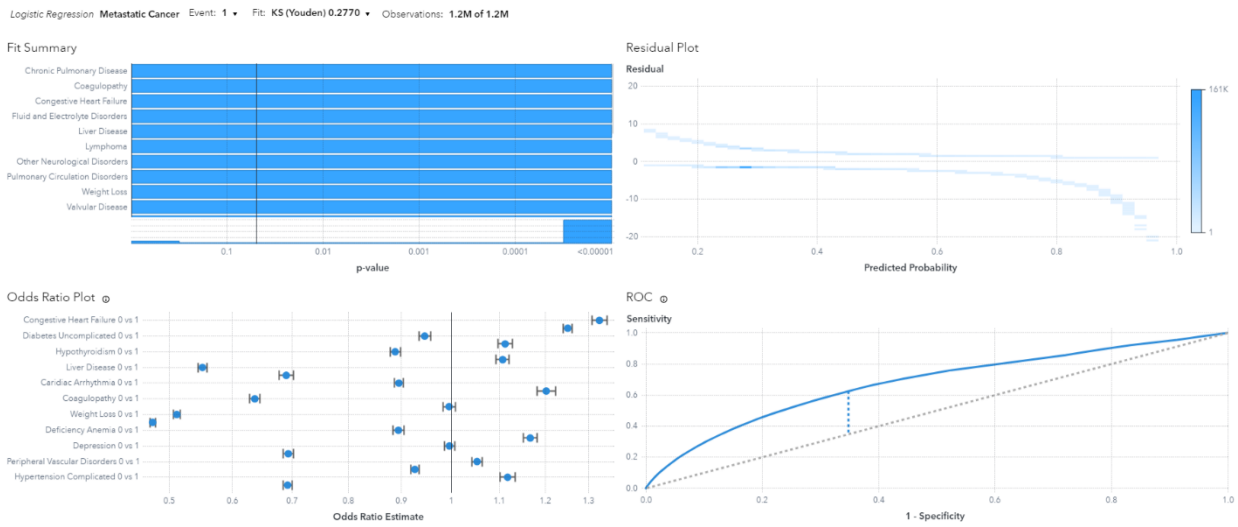

Figure 8: Likelihood of Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer Adjusting for Comorbidities

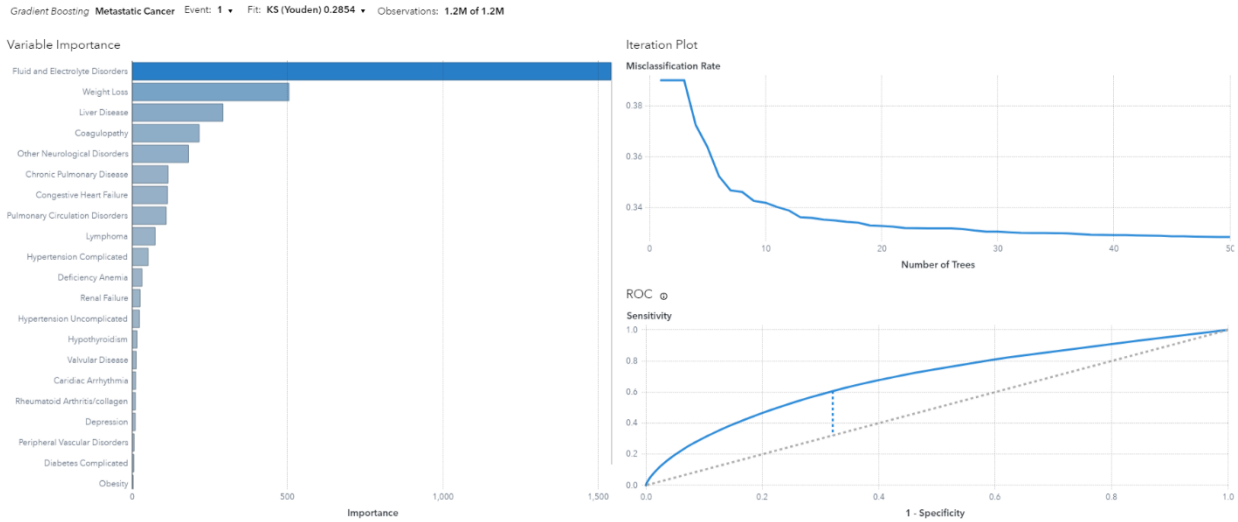

Figure 9: Variables of Importance As Related to the Likelihood of Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer Using a Gradient Boosting Model

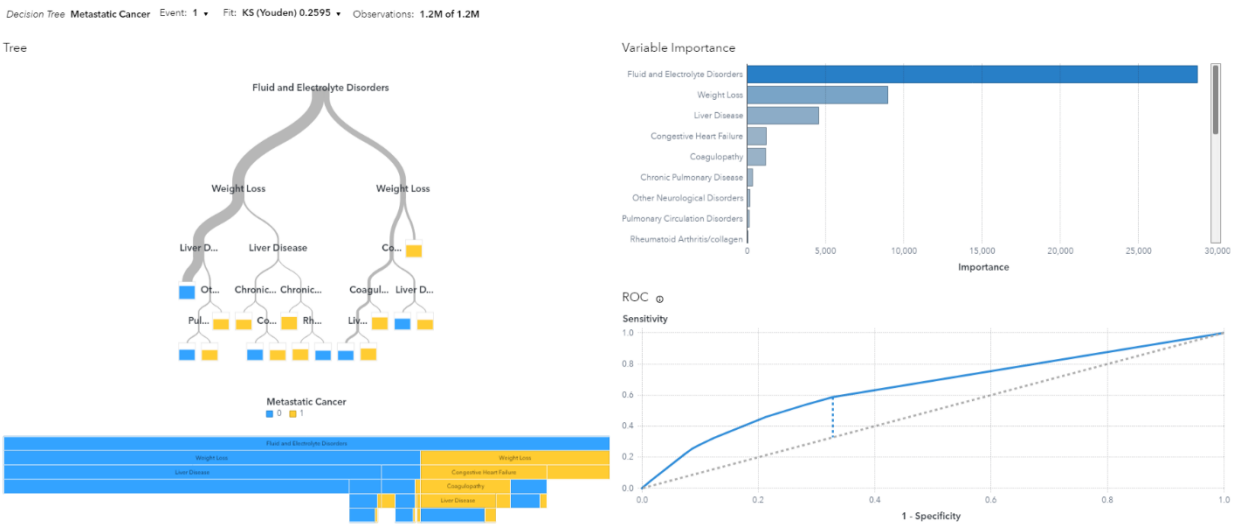

Figure 10: Association of Comorbidities Associated with Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer

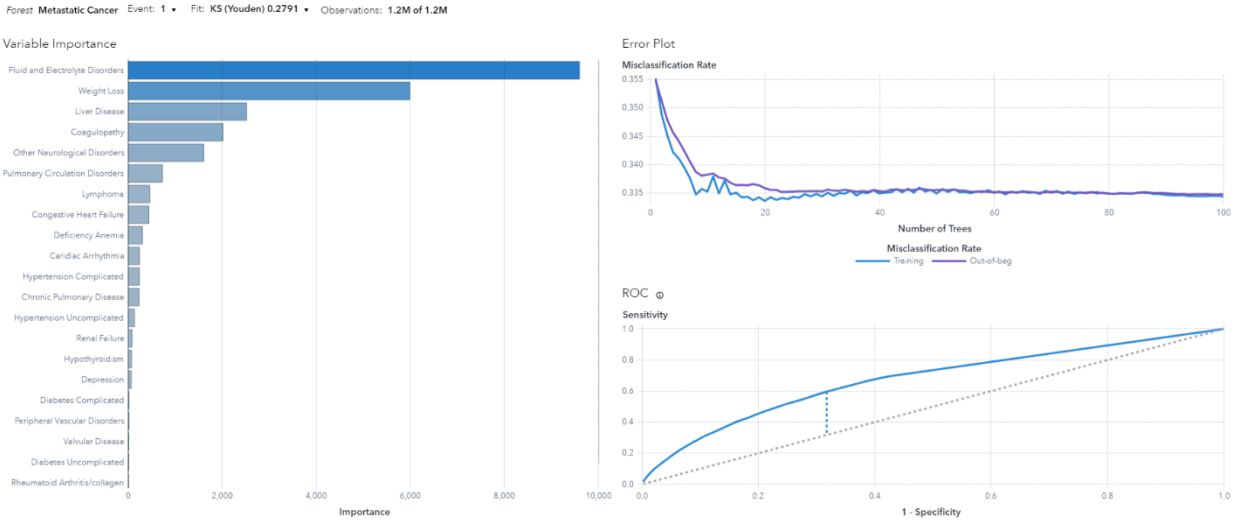

Figure 11: Variables of Importance Using a Random Forest Plot in Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer

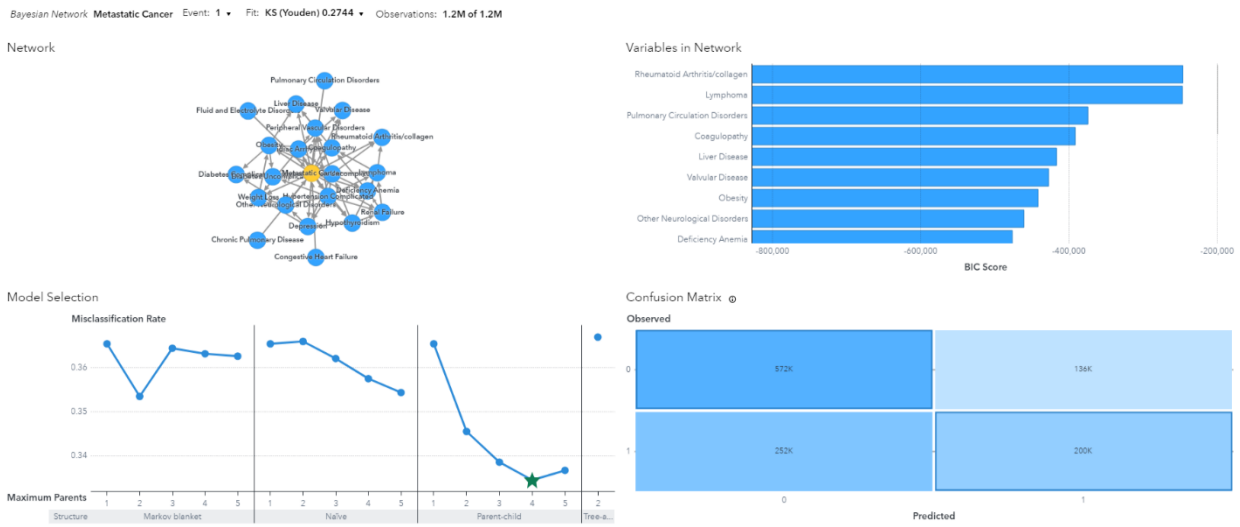

Figure 12: Associations of Comorbidities with Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer

Gradient boosting was the best fit model among the ML methods (KS (Youden)=0.286) followed by Forest, Logistic Regression and Bayesian Network (0.279, 0.277, and 0.274, respectively). ML demonstrated greater accuracy in assessing the likelihood of MC (Figure 13).

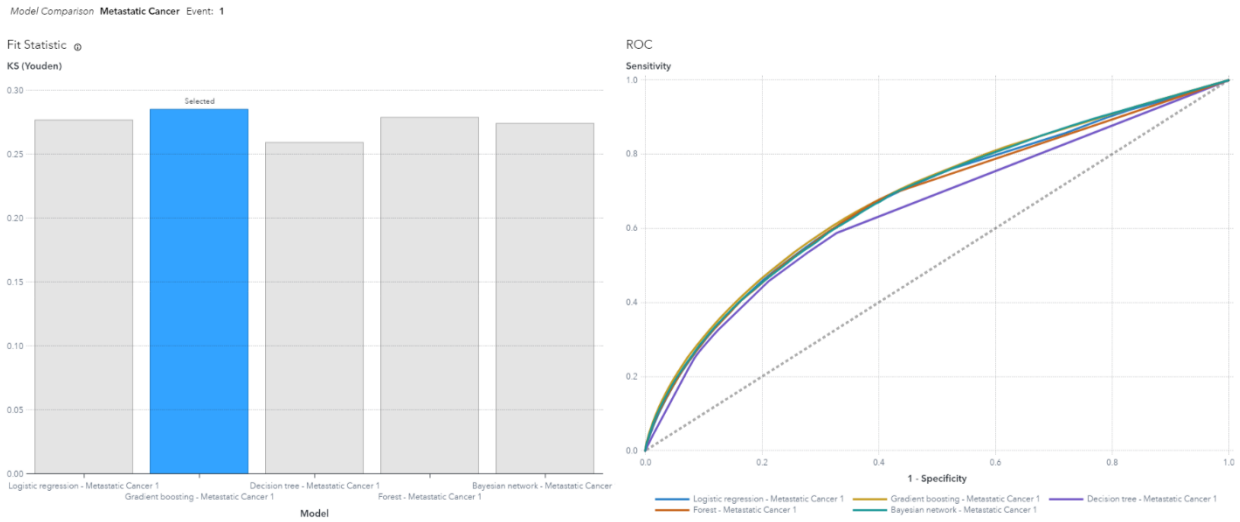

Figure 13: Model Comparison in the Likelihood of Metastatic Cancer in Patients Diagnosed with Non-Small Cell Lung Cancer

CONCLUSION

In conclusion, this work demonstrates the importance of more comprehensive analyses to better the outcomes and care that our patients deserve. Knowing that COPD is a contributor to mortality in patients diagnosed with NSCLC, in addition to the existing burden of COPD for a patient that will absolutely affect their quality of life, it is imperative that the findings of this work are taken into consideration. At the point of care, NSCLC patients should be assessed for comorbidities to better anticipate secondary treatment needs that may interfere with their cancer treatment and/or contribute to their overall morbidity and mortality. Furthermore, future studies examining more specific risk factors that affect outcomes and quality of life in this patient population are necessary to optimize treatment and positive outcomes. These approaches to personalized medicine could contribute to better prognoses for these patients. With greater compute power, scalable environments and ease of access to advanced analytics, our patients are owed more comprehensive approaches to evidence generation to assist in optimizing their outcomes.

REFERENCES

Durham AL, Adcock IM. The relationship between COPD and lung cancer. *Lung Cancer*. 2015 Nov;90(2):121-7. doi: 10.1016/j.lungcan.2015.08.017. Epub 2015 Aug 29. PMID: 26363803; PMCID: PMC4718929.

Vestbo, J., Hurd, S. S., Agustí, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., ... & Rodriguez-Roisin, R. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American journal of respiratory and critical care medicine*, 187(4), 347-365.

Trends in Tobacco Use. American Lung Association, Research and ProgramServices, Epidemiology and Statistics Unit, 2011.

Lundbäck, B., Lindberg, A., Lindström, M., Rönmark, E., Jonsson, A. C., Jönsson, E., ... & Larsson, K. (2003). Not 15 but 50% of smokers develop COPD?—report from the obstructive lung disease in Northern Sweden studies. *Respiratory medicine*, 97(2), 115-122.

Brusselle, G. G., Joos, G. F., & Bracke, K. R. (2011). New insights into the immunology of chronic obstructive pulmonary disease. *The Lancet*, 378(9795), 1015-1026.

Shah S, Blanchette C, Coyle J, Kowalkowski M, Arthur S. Howden R. Survival associated with chronic obstructive pulmonary disease among SEER-Medicare beneficiaries with non-small-cell lung cancer [published online April 29, 2019]. *Int J Chron Obstruct Pulmon Dis*. doi: 10.2147/COPD.S185837.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Samiul Haque
Company: SAS Institute Inc.
Email: samiul.haque@sas.com