

MIMIC-III Database: Heart Failure subtypes analysis

Capstone Project

Data Scientist Nanodegree

Author: Francisco J. Londono

Problem Introduction

Cardiovascular diseases are the leading cause of death worldwide affecting millions of each year. Among these, heart failure is becoming more prevalent, especially in an ageing population impacted by poor lifestyles (poor eating habits, sedentarism, smoking, pollution, etc.).

Heart failure is a complex, multifactorial disease process rather than a specific disease thus treatment has proven to be a challenging task as well. Even diagnosis is not a straightforward process, and differentiate between two defined subtypes of heart failure has proven to be a challenge as well. Indeed, it has been recognized two heart failure subtypes regarding affected cardiac function: systolic heart failure and diastolic heart failure. The fundamental measurement to diagnose one heart failure subtype is ejection fraction, i.e., the percentage of blood ejected after each systole. If the ejection fraction is lower than 50%, this is referred as heart failure with reduced ejection fraction, and it is related to a diminished systolic function. If ejection fraction is higher than 50%, this is referred as heart failure with preserved ejection fraction, and it is related to a diminished diastolic function of the heart. Even if this denomination is useful to support diagnosis and potential treatment, and in fact there is a prevalence of one of these two conditions, the complexity of the disease involves a deterioration of both functions: systolic and diastolic, impacting potential treatment as well.

The population more affected by heart failure is also, in general, differentiated by subtype. Systolic heart failure is more common in old men with hypertension, diabetes, and other comorbidities. Diastolic heart failure is more common in older women. In terms, of treatment, systolic heart failure has a proven route of action and drugs that help patients with symptoms and to improve the progression of the disease. In contrast, diastolic heart failure lacks a proven effective treatment, with the aggravating factor that is becoming more and more prevalent in the world's ageing population. Systolic failure treatment has been tested in diastolic failure patients but with less optimal results.

Strategy to solve the problem

In this context, the MIMIC-III database can provide a unique insight into the impact of heart failure and its subtypes in the daily routine of an emergency care unit, giving a more real idea of the type of population affected by the disease, the type of used treatment, and the final outcome of patients after an emergency related to the disease evolution. Thus,

following there is an analysis of the MIMIC-III database, starting with an overview of the most common diagnosis of patients, then the analysis focuses on cardiovascular diseases related directly with heart failure, highlighting the population with a systolic and a diastolic heart failure diagnosis, and comparing treatment followed at the ICU.

The second part of the analysis focuses on the systolic and diastolic heart failure patient's data, including some common lab exams and physiological measurements, such as hemodynamics, arterial blood gas (ABG), Hematology, Chemistry, Fluids - Other (Not In Use), Blood Products/Colloids, Labs, Cardiovascular (Pacer Data), Tandem Heart, non-invasive cardiac output monitor (NICOM). Hemodynamic measurements are the most relevant measurements to determine heart failure subtype but a deep analysis of the available data in the database showed that there were too few data points, or patients with a complete dataset, affecting the potential modeling of the data. Thus, other complementary measurements were included to build a more complete DataFrame. Finally, on the third part of the analysis, the built DataFrame was used to propose a model to classify the two heart failure subtypes based on the available physiological measurements.

Metrics

To perform this modelling, a pipeline was created, first imputing the data using a KNN imputer, then scaling the data and finally a Random Forest Classifier is implemented, and a cross validation scheme used to train the model, obtaining an accuracy and a F1 score of around 98% to classify heart failure sub-types. An additional analysis on the sample population shows that both categories (systolic and diastolic heart failure) are balanced in the DataFrame, and that imputed values are around only 3.3% of the total data. The balanced nature of the subsample justifies the use of accuracy of an adequate metric for the model. The F1 score will add value not only if data is not balanced but also to have an equilibrium between precision and recall.

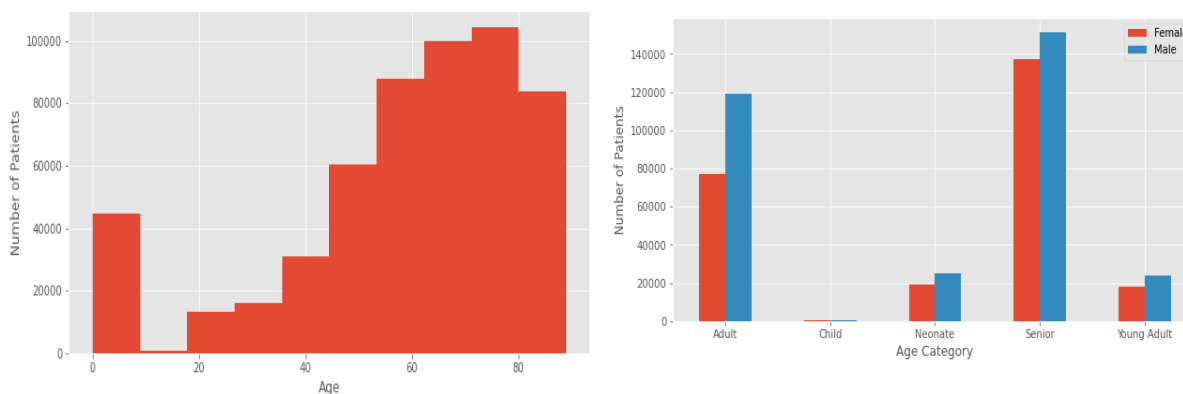
EDA

Part 1: MIMIC database EDA

Data exploration in an extensive database, such as MIMIC-III, constraints to perform an organized and focused analysis. Following this idea, a general demographic analysis is performed regarding age and gender to obtain some insights about the general population of the database. Then, diagnosis is selected as the main driven subject to explore the database, starting with the most common established diseases for patient's entry at the ICU to identify patterns and relevant concepts.

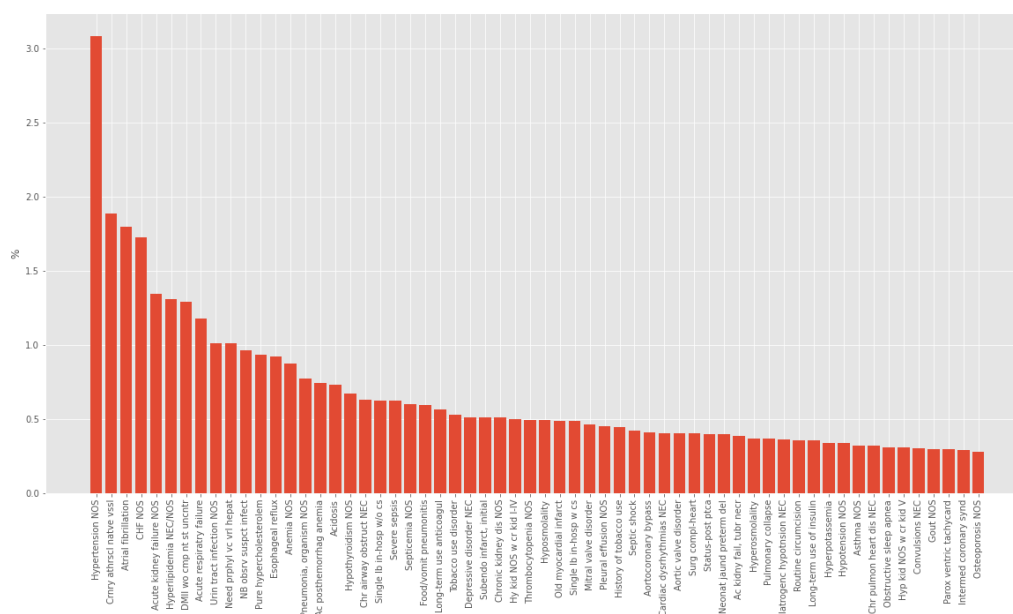
What are the demographic characteristics of the MIMIC-III database patients?

Age distribution shows a small peak of neonatal patients, and then a valley during childhood and teen years that grows from the 20's onward, peaking around ages 60-80. This trend is consistent with the fact that relevant pathological process develops through life and have major impact at certain threshold ages, due to poor lifestyle, genetic predisposition, or normal ageing processes.



Grouping by gender, group ages show a consistent trend in which female patients tend to be less prone to ICU entry. This could be due to social roles, a genetic or phenotypic difference, lifestyles, and many other reasons, that impact this specific sample population. To arrive to potential causal reasons, a more detailed analysis and more detailed information, such as epidemiological, demographic, economic and/or social situation, should be introduced to the analysis.

What are the most common diagnoses of patients included in the MIMIC database?



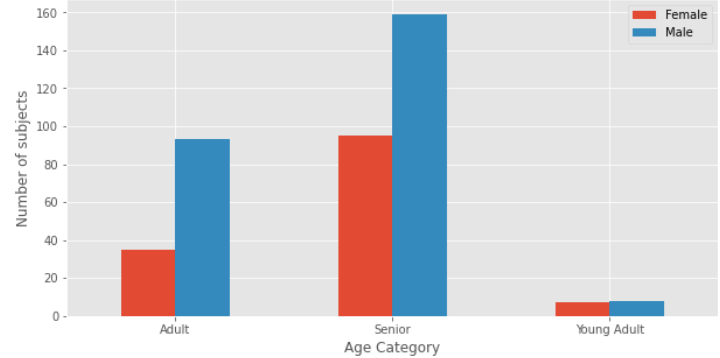
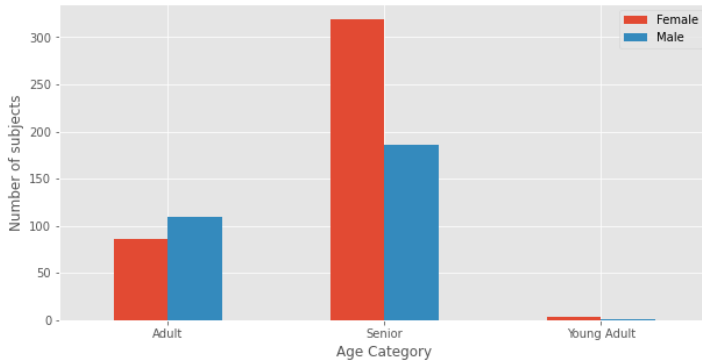
From the subsample population diagnosed with Heart Failure, what are the most common subtype diagnosed disease?

Indeed, Heart failure (HF) affects about 5.1 million people in the United States only and about 23 million people worldwide. HF is responsible for a large burden of morbidity and mortality, poor quality of life, and healthcare costs.

To complement the bar chart, this pie chart shows the graphical percentage distribution of diagnosis that include the expression "Heart Failure".



Among the diagnosis related with Heart Failure, there are many categories in which Systolic or Diastolic Heart Failure are included. These two categories represent Heart Failure subtypes, which are differentiated not only by the physiological cardiac function diminished in each case but also in their prognosis and potential treatment. Thus, even if interrelated by the complexity of the Heart Failure complex progression and multifactorial causality, the course of action could be completely diverse for each case, making a correct diagnosis a fundamental step for these patients.

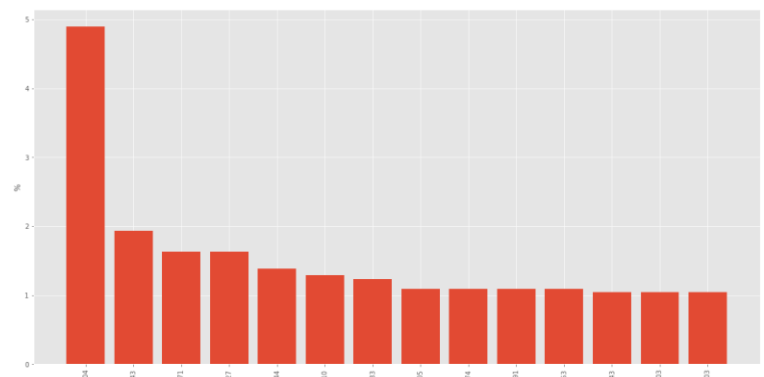
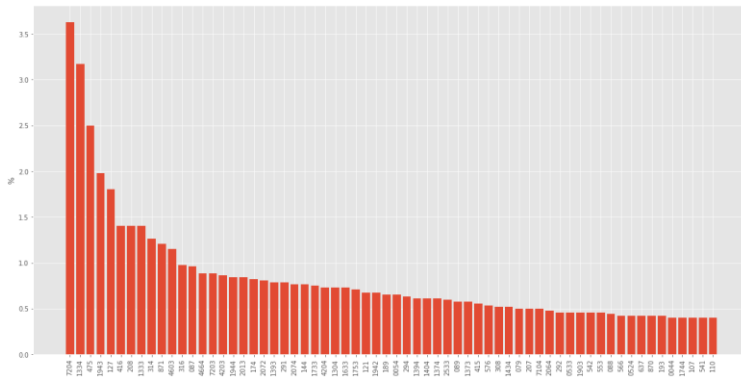


As expected, the selected subsample population show a higher percentage of older patients affected by these conditions, with most of the population belonging to the Senior category. An important difference between diastolic and systolic heart failure patients is the gender distribution. While systolic failure seems to be a predominantly male diagnosis, a diastolic failure diagnosis has a greater impact in older females.

How treatment compares for each Heart Failure subtype?

Even if there are several similarities between Heart Failure subtypes, there are relevant differences that raise the importance of a differentiated treatment. Systolic Heart Failure has been treated successfully since decades ago, and evolution in terms of drug development and follow up has been important to improve prognosis and quality of life of patients. Even if these medical advances have become a starting point to diastolic failure treatment, a successful approach is still a topic of research, adding to the complexity of the disease.

These two bar charts and additional information from the dataframe show that medicaments used for the treatment of both Heart Failure subtypes at the ICU are consistent, and that differences are not as relevant as could be expected since diastolic heart failure treatment is still a pending task in cardiovascular research.



The following information indicates the percentage of deaths among the patients at the ICU with diastolic and systolic heart failure diagnosis. Diastolic failure has almost a 67% rate of deaths while systolic failure counts for almost 60%. Even if similar, this gap raises several questions, including if the mortality rate of diastolic failure is, in general, higher or if treatment is more adequate for systolic failure patients, and thus more attention is needed for diastolic failure.

Modelling

Part 2: Subsample Data Analysis (Heart Failure)

Hemodynamic measurements are fundamental in terms of Heart Failure diagnosis and subtype determination. However, the main reference values are extracted from medical imaging, which are not available in this case. Thus, a first approach is to use the available hemodynamic data to model patient conditions with commonly used physiological measurements at the ICU.

Which physiological measurements are available for patients with a Heart Failure subtype?

A subsample population of patients with Systolic Heart Failure is selected, and hemodynamic measurements are analyzed to obtain a DataFrame to model the data.

The number of empty values, in this case NaNs, is high for this subsample. Only a threshold of 15 NaNs values removes all rows from the created DataFrame, making the modeling step difficult and especially inaccurate. Thus, it is necessary to include more information, considering not only hemodynamic data but further physiological measurements and lab exams data.

Which measurements can be used for classifying Heart Failure subtypes?

From the MIMIC database, several additional physiological measurements and lab exams data is available. To improve the dataframe selection, relevant items were selected in addition to hemodynamics. Specifically, the categories arterial blood gas (ABG) data, hematology, chemistry, fluids, blood products/colloids, labs, cardiovascular, tandem heart and non-invasive cardiac output monitor (NICOM).

Again, the subsample population of patients with Systolic Heart Failure is selected, and all selected physiological measurement data are analyzed in order to obtain a DataFrame to model the data.

Complementary to the obtained dataframe, a subsample population of patients with Diastolic Heart Failure is selected, and all selected physiological measurement data are analyzed in order to obtain a DataFrame to model the data.

After processing and obtaining the dataframe for each subsample (Systolic and Diastolic Heart Failure), a unique DataFrame is created for modeling the data, experimenting if there is a model that can use the available data to separate patients with Systolic Heart Failure from patients with Diastolic Heart Failure.

Part 3: Data Modelling

The previously created DataFrame is uploaded, and the feature matrix and target vector are selected from DataFrame columns. A classifier model is also selected, as well as an imputer to fill the NaNs values of the feature matrix, and a pipeline is defined to perform the data imputation, scalation and modeling. Then, a cross validation is performed assessing the accuracy and F1 score of the selected model.

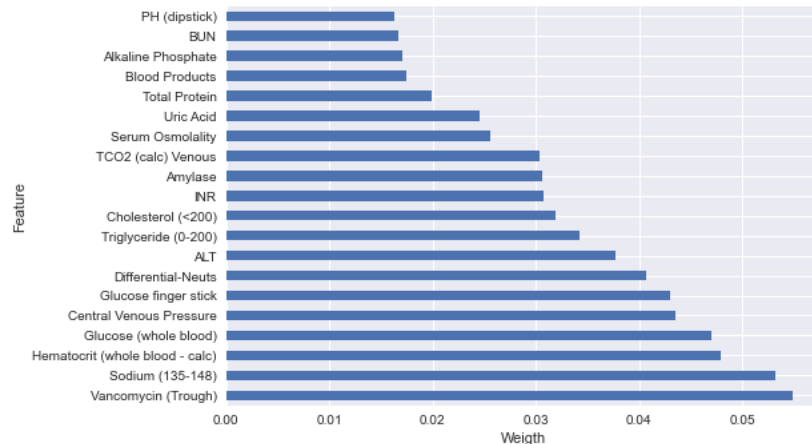
Model Evaluation and Validation

The distribution and balance of both classification categories (Systolic and Diastolic Heart Failure) is checked, including during the cross-validation splits. The values show that the number of patients from each category is similar and, thus, the population is balanced, making accuracy an insightful metric to evaluate the model. The first model assessment calculates a cross-validation score from a Repeated Stratified K-Fold strategy. To have an idea of how the model evaluates, a loop is created, and each iteration delivers a final accuracy and F1 score. All scores are higher than 98%. The validation performance is stable and doesn't fluctuate much, reflecting that the model is robust against small perturbations in the training data.

Then, feature importance is analyzed delivering interesting results regarding which measurements seem to have relevant difference between Systolic and Diastolic Failure patients.

Which features are the most relevant for achieving the proposed classification task?

The graph shows which features, in this case physiological measurements, are the most relevant regarding the classification model to differentiate from both Heart Failure subtypes. These results could give an interesting idea of which features should be evaluated and could complement standard measurements.



How imputation impacts the used data?

The impact of the performed imputation is taken into account as it could have an important influence in the modeling results, showing that only around 3% of values in the feature matrix are imputed.

Hyperparameter tuning

After results were in general adequate, improvement of the model is proposed through hyperparameter tuning using GridSearchCV, specifically for the number of trees in the forest (n_estimators) and for the number of features to consider when looking for the best split (max_features). This search delivers accuracy and F1 scores higher than 98%, although previous models without optimization processes show similar scores and the steps involve were less computational consuming.

Results

The initial step of this database assessment includes a general demographic data analysis, in which age and gender show interesting patterns, for example the fact that the highest percentage of ICU patients in this MIMIC database are older males. Causality for this distribution needs a deeper analysis. Then, diagnosis is selected for analysis, building a graphical description of the most common diseases for ICU entry. Cardiovascular diseases are among the top causes for ICU entry in this population, which is consistent with the impact of CVDs worldwide. Thus, an analysis in CVD patients is performed, specifically those affected by Heart Failure subtypes: Systolic and Diastolic Heart Failure. A demographic description of patients diagnosed with either one of both subtypes show that even if older individuals are more impacted by these pathological conditions, there is a gender differentiation between them. Previous studies are consistent with these differences and in the fact that treatment should be specific. However, the database shows that the treatment for each diagnosis is almost the same at ICU level. Another differentiation level is the death rates that show around a 7% higher chance of dying at the ICU for patients with a Diastolic Heart Failure diagnosis. Further analysis of the subsample population from the MIMIC database containing patients with a diagnosis of Diastolic and Systolic Heart Failure is performed in the following Part 2.

A fundamental step to analyze the MIMIC database in the context of Heart Failure patients is to know the related available data. Making use of the lab exams and physiological measurements done at the ICU, it could be interesting to determine if there is relevant information about Heart Failure patients and specific differences between Heart Failure subtypes. Heart Failure diagnosis is done by hemodynamic analysis of medical imaging and extracted parameters. Thus, the first measurements taken for the analysis are the ones indicated with the tag "Hemodynamics". A DataFrame of Systolic Failure patients is prepared but the limited number of patients with values forces to include other measurements such as arterial blood gas (ABG) data, hematology, chemistry, fluids, blood products/colloids, labs, cardiovascular, tandem heart and non-invasive cardiac output monitor (NICOM). Then, a subsample of patients with Systolic and Diastolic Failure is selected, prepared and cleaned to build a DataFrame, including a label with the correspondent diagnosis, for modeling the data, experimenting if there is a model that can use the available data to identify and differentiate patients with Systolic Heart Failure and with Diastolic Heart Failure. The DataFrame is saved as a pickle file to be used in Part 3: Data Modeling.

A Classification model is proposed to differentiate Heart Failure subtypes. Specifically, a Random Forest Classifier implement within a pipeline that includes imputation and scaler steps. The general accuracy and F1 score of the model show values above 98%. Thus, an analysis of the feature importance for the model is also included, as an analysis of the imputation data process to assess the final results in context. Then, a Grid Search step is added to the pipeline for optimization of the model through hyperparameter tuning, specifically the number of trees in the forest (200, 700) and for the number of features to consider when looking for the best split ('auto', 'sqrt', 'log2'). Even if the best parameter option were 700 and log2, this step is computationally consuming, and the accuracy and F1 scores do not have a high improvement, reflecting how robust this type of model in standard conditions is for the subsample population and the classification tasks involve.

Justification

Heart Failure sub-type classification is a fundamental step for the treatment and follow up of patients, thus supporting tools are not only valuable but also necessary to enhance this step. Indeed, the detection of Heart Failure, and the posterior classification as Systolic or Diastolic Heart Failure is still a difficult task, even through the use of advanced imaging techniques, such as echocardiography or cardiac MRI sequences. Then, complementary tools that can add value and improve this diagnostic task. Thus, the proposed classification model has a high accuracy and F1 scores, making use of common physiological measurements and lab exams, potentially less complex than medical imaging techniques, for differentiating Heart Failure sub-types and potentially adding value to more establish diagnostic tools, and fundamentally enhancing the analysis of the complexity of the pathological condition known as Heart Failure and its sub-types.

Conclusions

The MIMIC-III database is a unique contribution for the analysis of healthcare set ups, providing insights into ICU patients through many conditions, a subsample that even if it could be biased due to geographical and availability of data, gives the opportunity of a vision of a real healthcare

setting and its necessities. The EDA shows the age and gender distribution of the patients, and the most common diagnosis in admitted ICU patients. From this, it is clear that age has a role to play, and in some cases also gender, and it confirms as well that cardiovascular diseases are on the top reasons for emergencies. A deeper analysis on the cardiovascular disease patients show that Heart Failure has a high impact in the older population, and that there are two sub-types with specific characteristics and necessities, even if it seems treatment remains similar for both cases and death rates are high for these patients. Then, this subsample population lacks standardization physiological measurements that could support treatment decision and, potentially, patient outcome. Consequently, a model using the available data is proposed to support more establish diagnostic tools with encouraging results. Indeed, the final implemented model accomplishes the classification task proposed. However, a more bridge subsample population must be included in order to validate the model in more complex datasets, and with a more complete set of values and physiological variables to adapt better the model to the primary problem.

Limitations and Difficulties

Due to the intrinsic characteristics of the database, the lack of complete data from each patient and the lack of certain important physiological measurements in the context of Heart Failure are issues to take into account, but also opportunities to perform an analysis on non-conventional measurements for Heart Failure analysis. In fact, the proposed model shows that the most relevant features are variables that in general are not considered in Heart Failure diagnosis and classification. Interestingly, the model hyperparameter tuning does not add much value and the default values show equally high metric score, highlighting how well the model adapted to the analyzed data. The size of the subsample population and the lack of consistency in some of the diagnostic labels have also an impact on the final result, and they should be considered in the data preparation stage.

Improvements

The main strategy to improvement is to have more patients in the selected subsample population, that is to say patients diagnosed with Systolic and Diastolic Heart Failure, and increase the number of physiological measurements and lab exams for each patient to ensure a more complete database and general state of the patients, which will improve the understanding of the potential differences between Heart Failure sub-types. Ultimately, a more refine process to select patients and the standardization of data acquisition will impact positively the analysis, results and conclusions of the project.

Adding more variables and physiological measurements, it could be very useful to compare standard measurements for Heart Failure diagnosis against other values, exams and measurements since these relationships could give a better idea which variables are more suitable to support sub-type classification, thus adding a correlation map and visualization tools from echocardiographic measurements and less complex physiological measurements could contribute with interesting insights.

In more practical terms, the assessment of other models can deliver better metric scores, even if for this strategy to success data quality and quantity remains a critical issue. GridSearchcv could include the assessment of different models in order to find a better result for the classification task. For example, boosting algorithms, such as the XGBoost, could be useful in the regime of limited training data, computational resources, and expertise for parameter tuning. Ensemble models are another potential way to manage the task and to improve results. Then, with a more bride and complex dataset, metrics could be wider reflecting and evaluating a higher complexity problem than the one is presented in this limited subsample population.