In Hospital Mortality Predictions

Indigo Jones

Fisk University

Nashville, TN

IIJones02@my.fisk.edu

Abstract— In this project, results from predicting hospital mortality are analyzed. They were produced using several different models such as QDA, SVC, and many more. This is a mere evaluation of an approach of finding a more so unbiased way of predicting such casualties. The goal is to provide further insight into ways to rectify such issues on top of the many other solutions and approaches mentioned in this project. The dataset used is from the Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC-III) database.

I. INTRODUCTION

The impact of caring for critically ill patients within intensive care units (ICUs) has been increasing steadily over the years. Many factors such as resources, space (e.g., hospital beds, etc.), and even the gross domestic product (GDP) have been increasing and exhausting over the years. For example, in the United States alone, 22% of the hospital's cost is due to care for patients in ICUs, in addition, this even leads to those costs contributing to about more than 1% to the GDP. Also due to the pandemic that is currently taking place the admission rate of ICU was up by 32% and the mortality rate of ICU patients is now at a range from 42% to 93% as of 2021 while the mortality rate from 2009-2012 ranged from 11% to 12%. With these numbers, hospitals and even some parts of the healthcare system have been bearing the burden that ICU's patients have indirectly placed on it. However, to alleviate this problem, predicting the mortality of critically ill patients has been one of the many solutions to alleviate the effects that come along with the problem of having resources and money exhausted.

The issue at hand has been addressed in several diverse ways, one main way using different scoring techniques which use statistical modelling such as *Acute Physiology and Chronic Health Evaluation* (APACHE) as well as *Simple Acute Physiology score* (SAPs). However, this system of scoring is biased against patience that leaves alive or dead. Even with the updated state of both systems it is concluded that it still is not accurately predicting whether a patient will die or not. In addition, it seems to be biased towards regions such as France, southern and central parts of Europe, as well as the southern parts of Mediterranean.

Such a faulty system within a high-risk system is the main reason as to why it is important for this issue to be addressed. There have been many solutions and approaches to rectifying such a problem ranging from expanding and using different parameters within models to training and testing a more inclusive and expansive dataset. My goal with this project is to not to provide a solution to the problem, but to evaluate and provide a different approach to the issue at hand for scoring systems such as SAPs and APACHE are still widely used despite their faults meaning there is yet for a solution to be provided. By exploring and evaluating my approach, I hope to gain insight or spark influence on a solution to this detrimental problem.

II. PROBLEM & HYPOTHESIS

The problem at hand is that today's tools for predicting mortality of a patient are biased due to underlying discrimantion. I am hypothesizing that by evaluating and running multiple models, more specifically, classifiers, in addition to using a more extensive dataset, this could identify and/or help to begin to mitigate such biases.

III. RELATED WORKS

When it comes to related works, this project was heavily inspired by Jingmin Zhou from Zhongshan Hospital, Fudan University and his works on predicting in hospitality mortality. His approach was to use a more extensive dataset and used XGBoost and LASSO regression models to test to his hypothesis, *"to develop and validate a prediction model for all-cause in-hospital mortality among ICU-admitted HF patients",* referenced by Zhou and his team's abstract. With their use of linear regression analysis, their algorithm was able to produce *"good predictions".*

In addition to their works, this project was also inspired by and referenced to the works of Lancet Respir Med, where their works more so focused on the parameters being used within the models. By using a scoring algorithm, *Super Learner* (which ranks the algorithms based on its *"prediction performance, and then builds the aggregate algorithm given by the optimal weighted combination of all candidate algorithms"*), and comparing it to the most widely used ones of today such as SAPs and APACHE, they were able to produce better results compared to the widely used systems of scoring within in the health care systems.

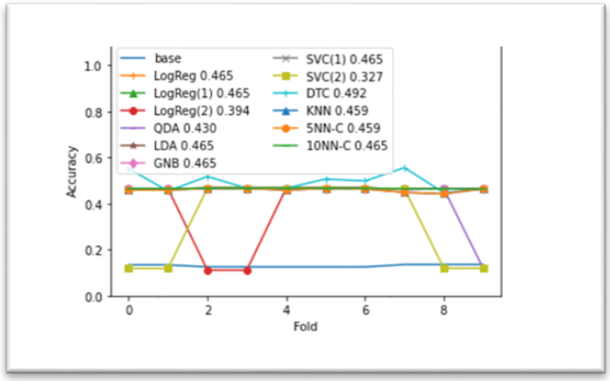
Both related works have contributed a bountiful amount of research, and with this project being on a smaller scale compared to those two, there was much inspiration taken from the two in order to continue with this approach. Further improvement on this project will be continued.

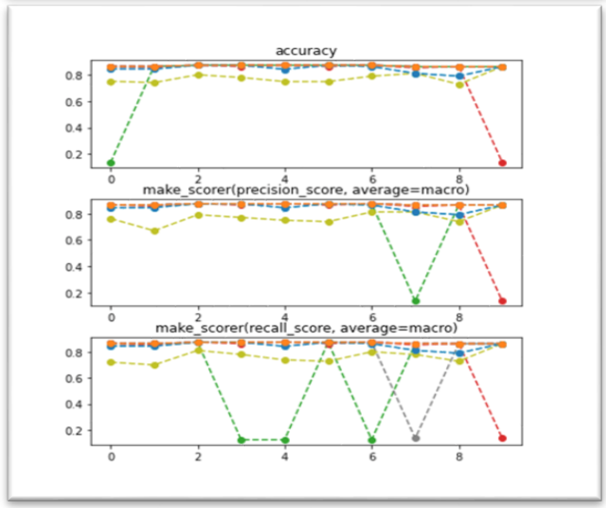
IV. DATA & RESULTS

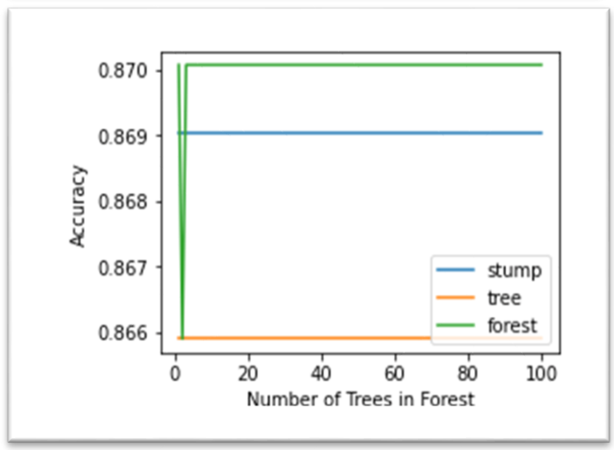
The data that was used for this project was the *Multiparameter Intelligent Monitoring in Intensive Care III or MIMIC-III* which consisted of a variety of different factors and conditions that was associated with each patient. However, due to the small scale of this project few features were omitted, but results of all the features are included in the Bibliography below.

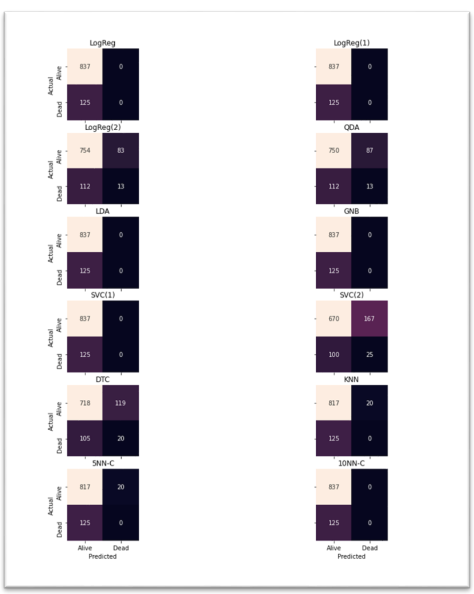
When all 12 models were ran, the LogReg, LDA, GNB, SVC(1), and 10NN-C were consistently accurate. As shown in Figure-A.0. Those models predictions were 100% accurate, however, their accuracy scores were a little lower than expected with them all having a score of 0.84905 which is shown in Figure-A.1.

When it came to DTC graph shown in Figure-B.0, the forest was set to go through maximum of 5 features with a depth of 3. Because of this it performed well on accuracy alone, however, being compared to other models, its score was off by more than 10%. As for the other models, they performed not as well when being compared to the top performing ones, with their accuracy score ranging form 0.74 to 0.80

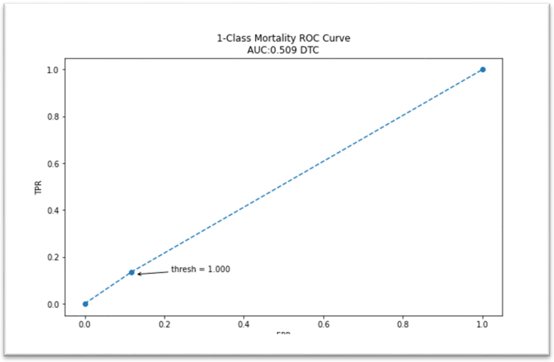


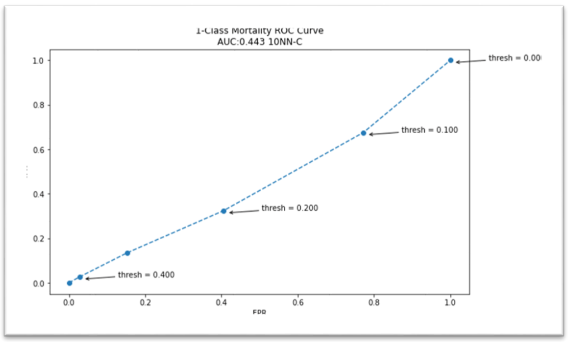


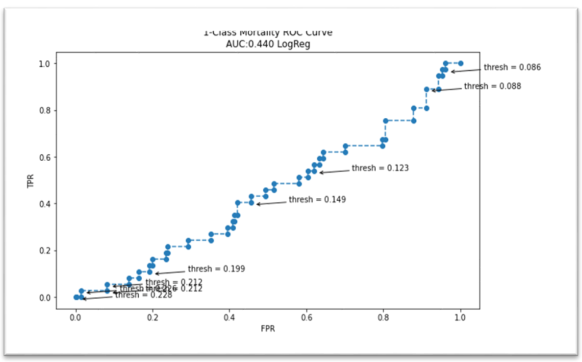


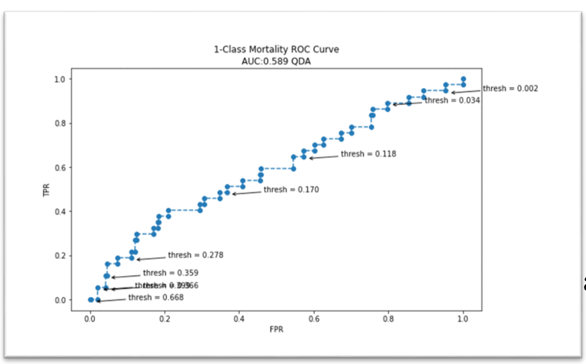


In the following three graphs shows the ROC Curve for the following models, LogReg, 10NN-C, DTC, and QDA. The ROC curve displays the different thresh-holds that the data goes through and the accuracy that is computed when data is at a certain threshold within each model. The best performing models mainly looked like Figure-C.0 and Figure-C.1, where when the threshold was nearly so close to 0 it performed well. However, what is interesting, was the Linearity of DTC in Figure-D.1. This was also shown back in Figure-B.0, and is assumedto be cause by the limited number of features it was traversing through. In addition, it was shown that the least performing models were more so linear, however, the 10NN-C proved to seems somewhat linear and still to perform well. Compared to 5NN-C and KNN, it is concluded the increasing number of neighbors set as a parameter in the KNNeighbor model proves to become more so accurate than anything lower than that number.









V. CONCLUSION

After training and testing out all 12 models I have concluded the following, that LDA or QDA is the best choice when it comes to predicting the the mortality of a patient simply because when looking at the ROC Curve it was curving more so closer to the top left corner making it more accurate compared to the LogReg or the DTC which was more linear. That being said, there is much work that is needed to be done to accurately approach this problem. One being using all of the features within the table, in addition to using GridSearchCV to generate and test out different parameters for each model.

These efforts will be included and improved on because, as stated before, SAPs and APACHE are still being widely used even though they possess biases with their scoring system. Though there are small outlets such as one of the related works I have mentioned, there should be a larger scale for not just clinicians to use but hospitals. This is my next goal to further my project and research and make it on a larger scale for such entities to use.

In conclusion, I was able to produce results of how all 12 models would work in these instances and what models were deemed to produce the best results.

VI. BIBLIOGRAPHY

Armstrong, R. A., Kane, A. D., Kursumovic, E., Oglesby, F. C., & Cook, T. M. (2021, February 1). *Mortality in patients admitted to intensive care with Covid‐19: An updated systematic review and meta‐analysis of observational studies*. Association of Anaesthetists. Retrieved November 20, 2021, from https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/10.1111/anae.15425.

Outcomes from intensive care in COVID-19 patients. American College of Cardiology. (2020, July 21). Retrieved November 20, 2021, from https://www.acc.org/latest-in-cardiology/journal-scans/2020/07/21/13/29/outcomes-from-intensive-care-in-patients.

Pirracchio, R., Petersen, M. L., Carone, M., Rigon, M. R., Chevret, S., & van der Laan, M. J. (2015, January). *Mortality prediction in intensive care units with the Super ICU Learner Algorithm (Sicula): A population-based study*. The Lancet. Respiratory medicine. Retrieved November 20, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4321691/.

Zhou, Jingmin et al. (2021), *Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database*, Dryad, Dataset, https://doi.org/10.5061/dryad.0p2ngf1zd