USING EXPLAINABLE AI TO ASSESS MACHINE LEARNING MODELS ON THE FORECASTING OF HOSPITAL MORTALITY IN COVID-19 PATIENTS

Increasing clarity on the drivers of patient mortality due to COVID-19 hospitalization

Presentation and Code location

https://github.com/gregg2024/AI_in_HC_FinalProject

Video presentation location

https://utexas.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=f3922ae0-3ad1-42ae-af38-b16000d32cd8

Gregory Gipson
gg33252@my.utexas,edu

Using Explainable AI to Assess Machine Learning Models on the Forecasting of Hospital Mortality in COVID-19 Patients

Increasing clarity on the drivers of patient mortality due to COVID-19 hospitalization

Gregory Gipson

The University of Texas at Austin
e-mail: gg33252@my.utexas.edu

Abstract

Nearly 800 million cases of COVID-19 resulting in over 7 million deaths have been reported worldwide since the onset of the pandemic. This paper builds on prior research into mortality prediction for COVID-19 and uses synthetic patient data to examine factors that may lead to an enhanced ability to predict mortality, thus empowering health-care practitioners to focus efforts on those most at risk of death while hospitalized. The analysis constructs a gradient boosting model (XGBoost) then utilizes various approaches to explainable AI (XAI) including Shapley values, LIME, and Explain Like I'm 5 (ELI5) to evaluate the drivers of model decision making. Consistent across the various models was the explanatory power on mortality of the patient requiring ventilation, the length of their stay in the hospital, and their age. Other factors with varying degrees of explanatory power across models include gender, patient isolation, ICU admission, smoker status, and Body Mass Index (BMI). While lab results were initially included in the model, these were dropped from the analysis to only include those features that would be present and available early on in patient hospitalization. From these results, the analysis shows that simple, interpretable models for mortality prediction can be highly effective and can allow for rapid decision-making to prioritize patients and potentially aid in reducing patient mortality.

CCS CONCEPTS: Computing methodologies

Additional Keywords and Phrases: machine learning, COVID-19, mortality prediction, explainable AI (XAI), SHAP, LIME, ELI5, neural networks

1 Introduction

1.1 Background

Amidst the COVID-19 pandemic, mortality prediction for hospitalized patients with cardiovascular disease is essential for effective resource allocation and patient management. Machine learning models, such as XGBoost, offer significant potential to enhance prediction accuracy by interpreting complex patterns within patient data. However, these models often operate as black boxes, complicating their interpretation and trustworthiness in clinical settings. Addressing this challenge, Explainable AI (XAI) methodologies emerge as vital tools to shed light on the inner workings of these complex models. XAI can serve as an invaluable resource in unraveling the intricacies of mortality prediction models and can detail the influence of various risk factors on patient outcomes, fostering a deeper understanding of model predictions. The integration of machine learning models with XAI not only bolsters the predictive capabilities of mortality models but also enhances their interpretability. By uncovering the underlying mechanisms driving model decisions, XAI empowers clinicians to make informed decisions and tailor interventions to individual patient needs and fosters transparency and trust in the predictive process.

1.2 Related Work

A simple search on Google Scholar for the term "covid mortality prediction machine learning" returns approximately 157,000 papers. This paper was inspired in part by the work of Russell Yang [1], and Yi Lan et al [2] in developing gradient boosting models via XGBoost and implementing various XAI techniques to allow for greater transparency and interpretability of the model results. Yue Gao et al [3] inspired further analysis conducted outside of this paper by utilizing a neural-net based model for mortality prediction. Additional feature selection in model development was gleaned from research by Reddy et al [5] and Soeroto et al [6].

2 Methodology

2.1 Workflow

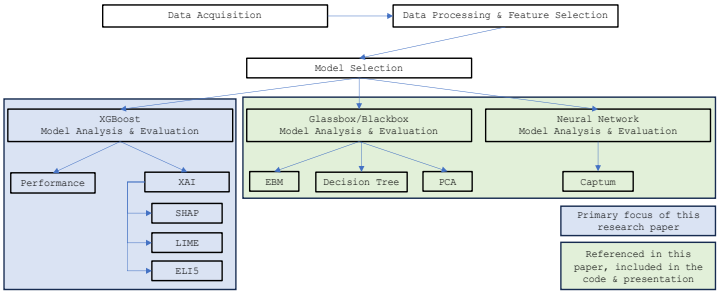

Figure 1: Workflow associated with this research paper

2.2 Data Acquisition

Data for this analysis was generated by Synthea™, an open-source patient population simulation made available by The MITRE Corporation for public use along with certain data pre-processing code from research conducted by Walonoski et al [4]. Simulated, synthetic data can serve as a secure alternative to deidentifying real-world health records. It eliminates the risk of reidentification since there are no actual patient records underlying the simulated data. The specific dataset used for analysis included 100,000 synthetic patient records with COVID-19 and includes a wealth of information on electronic health records that are consistent with what would be available for actual patients. Files were downloaded from the MITRE website and stored on Google drive for analysis.

2.3 Data Processing & Feature Selection

Synthetic patient data used for the analysis was screened to include only those patients admitted to the hospital with a COVID-19 diagnosis and further screened to identify death as only resulting from COVID-19, excluding patients who died for other reasons. Six specific lab results were incorporated in prior research on this dataset [4] which included levels of lymphocytes, D-dimer, serum ferritin, high sensitivity cardiac troponin I, lactate dehydrogenase, and IL-6 while certain of these measures were used more broadly by other researchers [2] [3]. From a patient characteristic perspective, additional features were either drawn from or calculated using the synthetic data and included ICU admission, age, need for isolation, need for ventilation, length of stay at the hospital, gender, smoking status, and Body Mass Index (BMI). The latter two features were included based on research surrounding the impact of smoking and obesity on the development and severity of COVID-19 [5] [6].

Two datasets were created for use in model development. One dataset included all the variables outlined above, lab-related and patient characteristics, while the second dataset excluded the lab results. Analyzing the feature correlations, the lab results are generally highly correlated with each other and with mortality while the non-lab features generally see lower correlation with each other and with mortality. The one exception is lymphocytes which are a type of white blood cell that plays a crucial role in the body's immune response. Thus, low lymphocytes would factor into mortality prediction so the negative correlations are sensible.

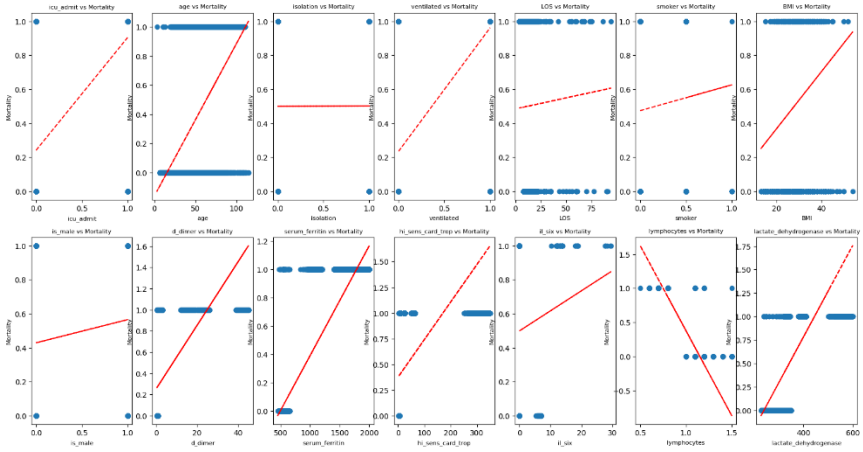

Figure 2: Charts showing the regressions between model features and patient mortality

2.4 Model Selection

Both datasets were run through various classification-based machine learning models including XGBoost, CatBoost, Support Vector (SVC), Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, and Stochastic Gradient Descent. The models were trained on default parameters using 70% of the respective datasets with the remaining 30% split evenly between test and validation sets. Results were generally consistent across the various methods.

3 Results

3.1 Model Analysis & Evaluation

The CatBoost Classifier model generated the highest AUC score on the full dataset with a score of 99.96% and generated the highest AUC score on non-lab dataset with a score of 98.23%. The XGBoost Classifier model generated an AUC score of 99.85% on the full dataset and an AUC score of 97.73% on the non-lab dataset. For consistency with the research papers [1] [2], we will use the XGBoost Classifier model, noting that its AUC is consistent with that of the optimal CatBoost models.

Table 1: Classification Report for the XGBoost Model as applied to the full dataset and the non-lab dataset

	Precision		Recall		F1-Score		Support	
Metric	**Full Data**	**Non-lab Data**	**Full Data**	**Non-lab Data**	**Full Data**	**Non-lab Data**	**Full Data**	**Non-lab Data**
Survival	0.99	0.91	1.00	0.94	1.00	0.93	498	498
Death	1.00	0.94	0.99	0.91	1.00	0.93	511	511
accuracy					1.00	0.93	1009	1009
macro avg	1.00	0.93	1.00	0.93	1.00	0.93	1009	1009
weighted avg	1.00	0.93	1.00	0.93	1.00	0.93	1009	1009

Both datasets demonstrate strong performance across various metrics, indicating their effectiveness in classification tasks. The Full Dataset (FDS) model outperforms the Non-lab Dataset (NLD) model in most metrics, showcasing its superior discriminative ability and overall predictive power. Specifically, the FDS model achieves higher scores in AUC, AUPRC, F1 Score, Cohen Kappa, Matthews Correlation Coefficient, and Precision, indicating better classification performance across different evaluation criteria. However, it's worth noting that the NLD model still performs well, and achieves high accuracy, recall, and specificity, demonstrating its ability to effectively discriminate between classes.

The model on the full dataset has near-perfect forecasting ability which could raise some concern. Noting that the lab results are highly correlated with each other and with mortality, one could presume that dropping any single lab item would not materially impact the forecasting ability. As such, when we look at various implementations of Explainable AI (XAI), we will use the model based on the non-lab items. One could look at this model as one which has the most data available upon hospital admission as the lab results may take some time to complete. The non-lab feature model is thus a good snapshot at inception of patient admission.

3.2 Explainable AI (XAI)

3.2.1 SHAP (SHapley Additive exPlanations)

SHAP is a popular library for explaining individual predictions of machine learning models. It computes Shapley values, a method from cooperative game theory, to explain the contribution of each feature to the model's output.

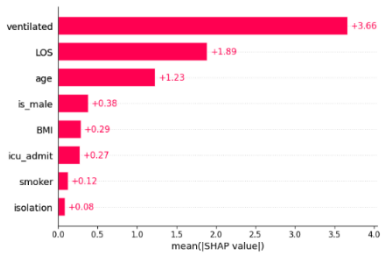

Figure 3: SHAP values for the non-lab features

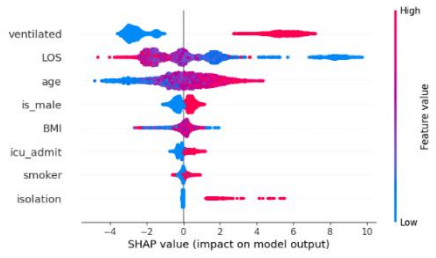

Figure 4: SHAP bee-swarm plot for the non-lab features

In Figure 3, the SHAP explainer process shows that need for ventilation, length of stay (LOS), and age are the most important features with the highest explanatory power for mortality. Figure 4 depicts a bee-swarm plot which is another visualization tool that shows the SHAP values for each feature across all patients. In analyzing the plot, we see that need for ventilation, length of stay (LOS), ICU admission, and isolation are all skewed to the right and may be viewed as relatively more predictive than the other features. Age and gender are centered right though have long left tails and could be viewed as less predictive.

The following figures show two patient cases, one for survival and one for death. In Figure 5 although the patient was elderly, their short length of stay (LOS), lack of ventilation, non-ICU admission, and lack of isolation pushed the prediction to survival. In Figure 6, ICU admission, age, and need for ventilation were the primary drivers of mortality prediction.

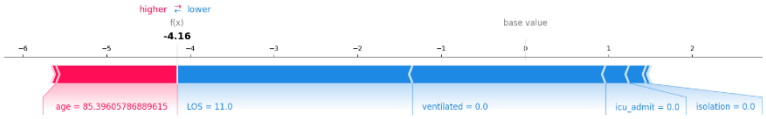

Figure 5: SHAP force plot for an example of patient survival

Figure 6: SHAP force plot for an example of patient death

3.2.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME is another library for explaining individual predictions of machine learning models. It provides locally faithful explanations by approximating complex models with interpretable surrogate models.

Figure 7: LIME interpretability for an example of patient survival

The left-most bar plot in Figure 7 shows the prediction probabilities, which can be treated as the model's confidence level in making the prediction, in this case a 63% change of survival. The second visualization is probably the most important one which provides maximum explainability for each of the features. The third visualization shows the features and their respective values. Here, the features highlighted in orange are contributing toward death, while features highlighted in blue are contributing toward survival. Lack of both ventilation and isolation and non-ICU admission coupled with a short length of stay (LOS) are the drivers for prediction survival despite the patient being elderly and a former smoker.

LIME is also capable of analyzing interpretability using bar charts similar to SHAP. In Figure 8 we see that patient survival is driven by a relatively young age and lack of both ventilation and isolation while Figure 9 depicts patient death as driven primarily by a need for ventilation and to a lesser extent gender, ICU admission, and being a former smoker.

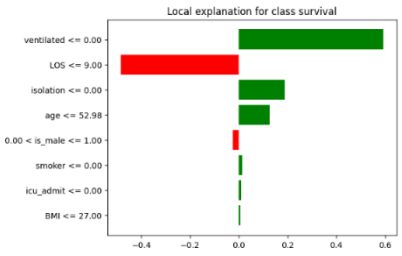

Figure 8: LIME explanation plot for patient survival

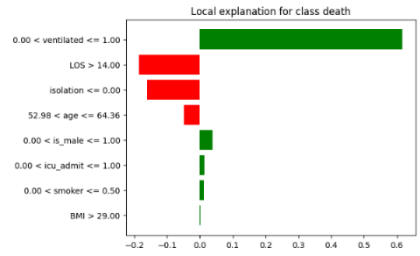

Figure 9: LIME explanation plot for patient death

3.2.3 ELI5 (Explain Like I'm 5)

ELI5 is a library that provides explanations for machine learning models at various levels of detail, including feature importances, permutation importances, and visualization of decision trees.

Feature importance is a popular XAI technique that helps us understand which features are most important in a machine learning model. ELI5 also provides tools for generating textual explanations for machine learning models. Textual explanations can help in understanding how a model arrived at a particular decision or prediction.

Table 2: ELI5 Feature Importance of non-lab features

Feature Importance	
Weight	**Feature**
0.8423	ventilated
0.0747	LOS
0.036	isolation
0.0118	age
0.0102	is_male
0.0091	icu_admit
0.0086	smoker
0.0071	BMI

Table 3: ELI5 Textual explanation example

y=died (probability 0.986, score 4.224)	
Contribution	**Feature**
+3.855	ventilated
+1.220	LOS
+0.607	**icu_admit**
+0.376	**age**
+0.069	**is_male**
+0.008	**<BIAS>**
-0.034	**isolation**
-0.203	**smoker**
-1.674	**BMI**

Feature importance confirms what was seen in the other XAI techniques as need for ventilation and length of stay (LOS) are the features with the highest explanatory power. In the example above, the textual explanation shows that the model predicts a 98.6% probability of death with need for ventilation, length of stay (LOS), and ICU admission being the primary contributors to this prediction. This chart is helpful as the contributions sum to the score generated whereby higher scores reflect mortality.

4 Conclusion

This study underscores the critical importance of enhancing mortality prediction models for COVID-19 patients, particularly those hospitalized, through the lens of Explainable AI (XAI). With the global impact of the pandemic resulting in millions of deaths, there is an urgent need to identify the factors driving mortality to facilitate targeted interventions and resource allocation.

The use of machine learning models, such as XGBoost, coupled with XAI methodologies like SHAP, LIME, and ELI5, offers promising avenues for advancing mortality prediction accuracy while ensuring interpretability and transparency in model decision-making. Through comprehensive analyses, this research details key predictors of mortality, including ventilation requirement, hospital length of stay, and patient age, underscoring their consistent impact across various models. The incorporation of non-lab features, such as gender, patient isolation, ICU admission, smoking status, and BMI, expands the scope of predictive variables and provides valuable insights into mortality risk prediction. Notably, the non-lab feature model demonstrates robust performance, offering a snapshot of patient prognosis upon hospital admission.

The findings emphasize the need for interpretable and actionable mortality prediction models that can inform clinical decision-making and improve patient outcomes. By leveraging XAI techniques, clinicians may gain deeper insights into model predictions, enabling personalized treatment based on individual patient needs. Future research efforts should focus on refining the model by incorporating real-world data and validating model performance across diverse patient populations. Ongoing advancements in XAI methodologies hold promise for further improvement in model interpretability and integration into clinical practice.

Additional research into other models was conducted outside of the scope of this paper. Analysis via InterpretML was leveraged in a few specific cases: the Glassbox approaches using the Explainable Boosting Machine (EBM) model and Decision Trees model along with the Blackbox Principal Component Analysis (PCA) model. All three were run on the non-lab dataset and confirmed the importance of ventilation and length of stay in mortality prediction.

Subsequent to this analysis, the original XGBoost model was converted to a neural net model via Captum which is a PyTorch-based library for model interpretability in machine learning. It offers a suite of algorithms and tools designed to help users understand and interpret the decisions made by deep learning models. Analysis was run across five separate algorithms and confirmed that age, ventilation, ICU admissions, and BMI had positive feature importance with mortality prediction across the various algorithms.

Acknowledgements

None

References

1. Russell Yang. 2020. Who dies from COVID-19? Post-hoc explanations of mortality prediction models using coalitional game theory, surrogate trees, and partial dependence plots. Retrieved April 4, 2024 from https://www.medrxiv.org/content/10.1101/2020.06.07.20124933v1.full.pdf

2. Li Yan, Hai-Tao Zhang, Jorge Goncalves et al. An interpretable mortality prediction model for COVID-19 patients. Nature Machine Intelligence 2, 283–288 (2020). https://doi.org/10.1038/s42256-020-0180-7

3. Yue Gao, Guang-Yao Cai et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nature Communications 11, 5033 (2020). https://doi.org/10.1038/s41467-020-18684-2

4. Jason Walonoski, Sybil Klaus, Eldesia Granger, Dylan Hall, Andrew Gregorowicz, George Neyarapally, Abigail Watson, Jeff Eastman, Synthea™ Novel coronavirus (COVID-19) model and synthetic data set, Intelligence-Based Medicine, Volumes 1–2,2020,100007,ISSN 2666-5212, https://doi.org/10.1016/j.ibmed.2020.100007

5. Rohan K. Reddy, Walton N. Charles, Alexandros Sklavounos, Atul Dutt, Paul T. Seed, Ankur Khajuria. The effect of smoking on COVID-19 severity: A systematic review and meta-analysis. J Med Virol. 2021; 93: 1045–1056. https://doi.org/10.1002/jmv.26389

6. Arto Yuwono Soeroto, Nanny Natalia Soetedjo, Aga Purwiga, Prayudi Santoso, Iceu Dimas Kulsum, Hendarsyah Suryadinata, Ferdy Ferdian. Effect of increased BMI and obesity on the outcome of COVID-19 adult patients: A systematic review and meta-analysis. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, Volume 14, Issue 6, 2020, Pages 1897-1904, ISSN 1871-4021, https://doi.org/10.1016/j.dsx.2020.09.029