Predicting Mortality, Length of Stay, and Pre Admission Conditions in Intensive Care Units

Luke Lorenz, Doug Rosin, and Jack Rudnick
Middlebury College

Abstract

Predicting outcomes in hospital Intensive Care Units (ICU) is an important and rapidly growing field. Many hospitals have begun to use an array of statistical learning methods to increase accuracy in these calculations. Anticipating patient outcomes is meaningful for hospitals' partitioning of scarce resources to deliver the best quality and mode of care for each patient. In this report, we employ linear, logistic, and quartile regression models, random forests, and K-nearest neighbor classification to predict mortality, length of stay, and liver disease in the ICU. In our best performing logistic regression model on mortality, we achieve 79.8% accuracy, observing 62.9% mortality within 2 years (731 days) for the high risk group and observed mortality of 13.9% in the low-risk group. We achieve 94.7% accuracy in predicting liver disease in patients, with 25.9% incidence of liver disease in patients predicted to have liver disease and 1.6% incidence of liver disease in patients predicted to be healthy.

I. Introduction

Intensive care units (ICU) provide complex and resource-intensive treatment for the sickest hospitalized patients. The need for critical care medicine has grown substantially over the past decade and has consumed a huge portion of the income in many countries worldwide (Vincent et al, 2014). In 2010, critical care units represented 13.2% of hospital costs, 4.1% of national health expenditures, and 0.72% of U.S. gross domestic product (Halpern et al 2010). Length of stay in the ICU is the primary predictor associated with ICU cost, as a small percentage of patients with a prolonged length of stay in the ICU could consume a large proportion (nearly 50%) of ICU resource use (Rapoport et al 1990, Rapoport et al 1994, Evans et al, 2018). Effective prediction tools are strongly needed by clinical staff to adequately manage patient and budget stress.

Intensive care unit (ICU) patient monitoring has the potential to reduce complications and morbidity, and to increase the quality of care by enabling hospitals to deliver higher-quality, cost-effective patient care, and improve the quality of medical services in the ICU (Alghantani et al 2021). Additionally, it is a patient-centered outcome and therefore of interest to families, ICU personnel, and managers (Peres et al 2021). The ICU is among the most resource intensive units in hospitals, so ICU surveillance has the potential to reduce morbidity and plan for improvement of care. There is therefore a demand for ICU monitoring that can provide guidance and report when adverse medical events are anticipated (Alghatani et al 2021). Currently, ICU length of stay is used retrospectively to evaluate ICU efficiency. Patient length of stay prediction at admission could help coordinate care, implement preventative measures, and better communicate with managers, families (Peres et al 2021).

With prolonged patient stays in the ICU, the burden of care for critically ill patients is massive (Young et al 2000). Given the role of the ICU in providing intensive and specialized nursing and medicinal care for critically ill patients and sustaining life during periods of life-threatening organ system insufficiency, it is unsurprising that ICU's contain the highest mortality rate of any unit of the hospital (Marshall et al 2017). Each year, about 500,000 out of 4 million patients entered into Intensive Care Units (ICU) pass away (Angus et al 1996, Wu et al 2002, Young et al 2000). In cases of severe illness, patients have an extraordinarily difficult choice to make. Patients can undergo intensive care, described above, to make every attempt to save their life, or they can request palliative care, which provides end of life care aimed at improving the quality of life for patients and their families (World Health Organization). Since early and frequent conversations over end of life treatment plans for patients suffering from serious conditions leads to increased satisfaction and quality of life stemming from care consistent to the patient's desire, many hospitals have begun implementing machine learning-informed algorithms derived from patient data to inform doctors of patients survival chances and spur these conversations (Wright et al 2008, Brinkman-Stoppelenburg et al 2014, Parikh et al 2021). Predicting mortality in patients hospitalized in ICU is crucial for assessing severity of illness and adjudicating the value of novel treatments mentioned above, interventions, and health care policies. Machine learning can utilize important baseline information that can be acquired in a timely manner and therefore serve as accurate predictors in our ICU length of stay calculation. When scaled to all patients that visit ICUs daily, massive amounts of data make it feasible to develop predictive models that can help with management.

A Penn Medicine study examining the effect of an electronic nudge to doctors flagging patients who were at a high risk of death and would benefit from conversations surrounding end of life goals found a significant increase in the prevalence of these conversations between patients with doctors who received the nudge, in comparison to patients with doctors who were in the control group. Additionally, doctors in the test group still engaged in end of life conversations with patients who were flagged as low

mortality risk about 75% more regularly than doctors in the control group (Parikh et al 2019). Even in cases where patients are likely to survive, these conversations are important for doctors and patients to coordinate and plan the best care possible for the patient. In Section III of this paper, we build models which, similar to Penn Medicine, predict end of life outcomes for patients. We compare our models with Penn Medicine's models in the discussion.

While patients are admitted to the ICU with significant, typically life threatening injuries, many also enter with significant comorbidities. A patient admitted for a car crash injury may also unknowingly have kidney disease, which could result in complications in treatment or sub-par recovery. Adverse events and clinical complications are a major cause of mortality and poor outcomes in patients, and substantial effort has been made to improve their recognition (Tomašev 2018). Unfortunately, few predictors have found their way into routine clinical practice because they either lack effective sensitivity and specificity. For example, liver disease can only be positively diagnosed after extensive testing (ultrasound, CT scan, or MRI) or a liver biopsy (Mayo Clinic). These tests may not be performed rapidly upon a patient's arrival in the ICU, if the attending physician decides to have them performed at all, but as established above, the identification of these conditions can be critical. One medical review of car crash victims noted that conditions including other medical maladies could result in missed injuries (Garcia 2006).

The relevancy of ICU patients with various additional conditions remains after being released from the ICU. A 2010 study that followed 478 ICU patients found that preexisting conditions were associated with the majority of cases that experienced diminished quality of life after being released from the ICU (Orwelius et al.). Our hope is to use easily gathered patient information to predict which patients have pre existing medical conditions. This information could help ICU doctors identify patients that might have complicating factors, and more deliberately order tests for those conditions to better inform their treatment.

We aim to address these motivating questions of mortality, ICU stay, and pre-ICU conditions through rigorous application of statistical learning models. Due to the muli-faceted nature of our dataset, numerous models should be used in tandem to produce robust and accurate classification. These approaches allow us to thoroughly assess clinical features most commonly associated with our questions of interest.

II. Methods

We used a 1,776 patient subset of the 26,870 patient MIMIC-II dataset, which was constructed from the ICUs of Beth Israel Deaconess Medical Center between 2001 to 2008. The MIMIC-II ('Medical Information Mart for Intensive Care') dataset is a large, single-center database comprising information relating to patients admitted to critical care units at a large care hospital. Data includes vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more. Our subset of data contains 46 variables extracted from MIMIC-II, including demographics (e.g. age, weight), clinical observations collected during the first day of ICU stay (e.g. white blood cell count, heart rate), and outcomes (e.g. 28 day mortality and length of stay).

To clean our data, we imputed the median value for continuous numerical variables and the mode observation for factor variables in place of NA observations. We then iterated over the dataset 20 times, constructing a random forest and used the resulting proximity to repeatedly impute the NA values for a

more accurate estimate of the unknown data. Finally, we removed certain variables with obvious covariance, including hosp_exp_flg, icu_exp_flg, day_28_flg from the data, as in many cases incidences of mortality were indicated in multiple categories. We also removed mort_day_censored, since it was an extension of censor_flg, where observations under 730 indicate death and observations over 730 indicate survival.

- Variables of interest:
    - aline_flg: IAC used (categorical, 1 = year, 0 = no)
    - icu_los_day: length of stay in ICU (days, quantitative)
    - hospital_los_day: length of stay in hospital (days, quantitative)
    - age: age at baseline (years, quantitative)
    - gender_num: patient gender (1 = male; 0=female)
    - weight_first: first weight, (kg, quantitative)
    - bmi: patient BMI, (quantitative)
    - sapsi_first: first SAPS I score (quantitative)
    - sofa_first: first SOFA score (quantitative)
    - service_unit: type of service unit (character: FICU, MICU, SICU)
    - service_num: service as a quantitative (categorical: 0 = MICU or FICU, 1 = SICU)
    - day_icu_intime: day of week of ICU admission (character)
    - day_icu_intime_num: day of week of ICU admission (quantitative, corresponds with day_icu_intime)
    - hour_icu_intime: hour of ICU admission (quantitative, hour of admission using 24hr clock)
    - hosp_exp_flg: death in hospital (categorical: 1 = yes, 0 = no)
    - icu_exp_flg: death in ICU (categorical: 1 = yes, 0 = no)
    - day_28_flg: death within 28 days (categorical: 1 = yes, 0 = no)
    - mort_day_censored: day post ICU admission of censoring or death (days, quantitative)
    - censor_flg: censored or death (categorical: 0 = death, 1 = censored)
    - sepsis_flg: sepsis present (categorical: 0 = no, 1 = yes -- absent (0) for all)
    - chf_flg: Congestive heart failure (categorical: 0 = no, 1 = yes)
    - afib_flg: Atrial fibrillation (categorical: 0 = no, 1 = yes)
    - renal_flg: Chronic renal disease (categorical: 0 = no, 1 = yes)
    - liver_flg: Liver Disease (categorical: 0 = no, 1 = yes)
    - copd_flg: Chronic obstructive pulmonary disease (categorical: 0 = no, 1 = yes)
    - cad_flg: Coronary artery disease (categorical: 0 = no, 1 = yes)
    - stroke_flg: Stroke (categorical: 0 = no, 1 = yes)
    - mal_flg: Malignancy (categorical: 0 = no, 1 = yes)
    - resp_flg: Respiratory disease (non-COPD) (categorical: 0 = no, 1 = yes)
    - map_1st: Mean arterial pressure (mmHg, quantitative)
    - hr_1st: Heart Rate (quantitative)
    - temp_1st: Temperature (F, quantitative)
    - spo2_1st: S_pO_2 (%, quantitative)
    - abg_count: arterial blood gas count (number of tests, quantitative)
    - wbc_first: first White blood cell count (K/uL, quantitative)

- hgb_first: first Hemoglobin (g/dL, quantitative)
- platelet_first: first Platelets (K/u, quantitativeL)
- sodium_first: first Sodium (mEq/L, quantitative)
- potassium_first: first Potassium (mEq/L, quantitative)
- tco2_first: first Bicarbonate (mEq/L, quantitative)
- chloride_first: first Chloride (mEq/L, quantitative)
- bun_first: first Blood urea nitrogen (mg/dL, quantitative)
- creatinine_first: first Creatinine (mg/dL, quantitative)
- po2_first: first PaO_2 (mmHg, quantitative)
- pco2_first: first PaCO_2 (mmHg, quantitative)
- iv_day_1: input fluids by IV on day 1 (mL, quantitative)

(SEE CODE FOR MORE METHOD EXPLANATION)

To predict mortality, we first test and compare the accuracy of models built using KNN-Classification, Random Forest, and Logistic Regression on the entirety of the data set. We chose censor_flg(, a binary variable where 1 indicates censoring, or survival, of the patient and 0 indicates death within 730 days (2 years), as our target predictor.

For KNN-Classification, we filter out all factor variables, including gender and incidences of stroke, congestive heart failure, and chronic obstructive pulmonary disease, since this method of modeling is based on the "distance" between two observations and it is impossible to find the distance between factor observations. We then split the data into train and test sets, and then performed 10-fold cross validation on the training set in order to obtain the optimal nearest neighbors, which was determined to be 12. We then fit a 12NN-Classification model on the entire training set, which was then used to make predictions on the test set. To assess the accuracy of our predictions, we create a confusion matrix to observe the overall accuracy, sensitivity, and specificity of the model.

For Logistic Regression, we omitted studying sepsis (sepsis_flg) from the model, as no patients in the data set contracted sepsis . We then split the data into train and test sets, with 70% of data in the training set and 30% in the test set. We fit a general linear model using logistic regression on the training set, which was then used to make predictions on the test set. The predictions take the form of a vector of probabilities, between 0 and 1, based on the determined likelihood of the observation being predicted as Censored. To determine which observations to classify as "Censored" and which to classify as "Death", we graph a receiver operating characteristic (ROC) curve, which maps the sensitivity vs specificity of the model. We use the ROC curve to determine the Youden index, or the ideal threshold to maximize sensitivity and specificity in the model. Observations above the Youden index are then classified as "Censored," and below are classified as "Death." Again, we assess the accuracy of our predictions through a confusion matrix.

Predicting mortality with a random forest follows a similar process, where we split the data into a training and test set, with 70% of data in the training set. We then fit a random forest model on the training set, which is then used to make predictions on the testing set. We compute accuracy, sensitivity, and specificity to assess our model's performance.

In this paper, we consider the observations of patients for which our models predicted mortality to be "high-risk," and observations for patients for which our model predicted survival to be "low-risk." As such, sensitivity, which is generally the accuracy in predicting mortality, can be interpreted as the percentage of high-risk patients who passed away, which we describe as the observed mortality for the

high risk group. Similarly, specificity can be interpreted as the percentage of low-risk patients who are deceased, which we describe as the observed mortality for the low-risk group. To assess the overall performance of our models, we will compare their accuracy, sensitivity, and specificity, both amongst each other and against the literature.

Finally, we split the data set in half, with observations of patients above the median age placed in one data set, and observations of patients below the median age placed in a separate data set. We then follow similar processes as above to fit models using logistic regression and random forests on each data set.

III. Results (SEE CODE FOR FULL RESULTS + EXPLANATION)

In our data set, there was an observed 72% survival rate and a 28% mortality rate (n = 1776, see Table 1). In the KNN Classification model (Table 2), overall accuracy was 78.4%, with observed 2 year mortality in the high-risk group of 50.5% and observed mortality in the low risk group of 11.2%. In the logistic regression model, overall accuracy was 80.3%, with an observed 2 year mortality of 70.5% in the high risk group and 15.9% in the low risk group. The random forest model performed with the same overall accuracy as the logistic regression model, but the random forest outperformed the logistic regression model in the low risk group, with only 9.9% incidence of mortality within 2 years, but performed slightly than logistic regression in predicting the high-risk group, observing 55% 2 year mortality. All three models significantly outperformed random assignment to “Survival” and “Death” based on data means.

The random forest (Figures 1-3) and logistic regression models (Table 3) also grant insight into the most important variables in predicting mortality. We find that age, stroke, malignant cancer presence, sapsi score, and blood urea nitrogen levels are most important to the accuracy of the random forest model, while blood urea nitrogen, sapsi score, and white blood cell count are most important for the homogeneity of the nodes and leaves. In the logistic regression model, some notable statistically significant variables include age, gender, incidence of stroke, cancer, and chronic obstructive pulmonary disease, and blood urea nitrogen, albeit with a small coefficient. It is notable that MeanDecreaseAccuracy and the logistic regression models tend to hold factor variables such as incidence of stroke in higher regard, while MeanDecreaseGini tends to prioritize continuous variables.

In all models, however, the importance of age in predicting mortality is undeniable. Due to the strong response of predicting mortality as a result of old age, we split the data into two age groups, “young” and “old,” based on the median age (56.3 years) in the data set, in order to make predictions more accurately in different age brackets. For both logistic and random forest models, we achieve moderately better mortality prediction for the old age group (mean age = 72.3, mortality rate = 46.3%, N=888) but suffer losses in overall accuracy and prediction of survival. In the young age group (mean age= 36.4, mortality rate = 9.7%, N = 888), we achieve higher accuracy and superior prediction of survival, but the observed 2 year mortality for the high risk group falls significantly, to 22.7% in the linear regression model and just 9.1% in the random forest model.

Table 1: Selected Means

Statistic	N	Mean	St. Dev.	Min	Max
censor_flg	1,776				
...Death	497	.28			
...Censored	1279	.72			
icu_los_day	1,776	3.346	3.356	0.500	28.240
age	1,776	54.380	21.063	15.180	99.111
weight_first	1,776	79.885	21.795	30.000	257.600
bmi	1,776	27.433	7.081	12.785	98.797
sapsi_first	1,776	14.130	4.015	3	32
sofa_first	1,776	5.818	2.331	0	17
day_icu_intime_num	1,776	4.054	1.994	1	7
hour_icu_intime	1,776	10.586	7.925	0	23
stroke_flg	1,776				
...Yes	222	.125			
...No	1554	.875			
mal_flg	1,776				
...Yes	265	.144			
...No	1520	.856			
copd_flg	1,776				
...Yes	157	.088			
...No	1619	.912			
liver_flg	1,776				
...Yes	99	.056			
...No	1677	.944			
chf_flg	1,776				
...Yes	213	.12			
...No	1563	.88			
map_1st	1,776	88.247	17.596	5.000	195.000
hr_1st	1,776	87.915	18.759	30	158
temp_1st	1,776	97.793	4.537	32.000	104.800
spo2_1st	1,776	98.433	5.512	4	100
abg_count	1,776	5.985	8.684	0	115
wbc_first	1,776	12.316	6.585	0.170	109.800
hgb_first	1,776	12.552	2.197	2.000	19.000
platelet_first	1,776	246.083	99.640	7	988
sodium_first	1,776	139.557	4.720	105	165
potassium_first	1,776	4.107	0.794	1.900	9.800
tco2_first	1,776	24.415	4.985	2.000	62.000
chloride_first	1,776	103.842	5.727	78	133
bun_first	1,776	19.266	14.348	2	139
creatinine_first	1,776	1.095	1.082	0.000	18.300
po2_first	1,776	211.220	145.216	22	634
pco2_first	1,776	43.266	13.235	8	158
iv_day_1	1,776	1,579.317	1,614.889	0.000	13,910.000

Table 2. Accuracy in Mortality Models

	KNN Classification	Logistic Regression	Random Forest
Model Accuracy	78.0%	80.3%	80.3%
Specificity (Predicting Survival)	89.6%	84.1%	90.1%
Sensitivity (Predicting Mortality)	47.4%	70.5%	55.1%

Table 3. Accuracy in Age Delimited Data

	Logistic Regression		Random Forest	
	Old	Young	Old	Young
Model Accuracy	70.4%	89.5%	68.5%	92.1%
Specificity (Predicting Survival)	67.9%	95.5%	70.1%	99.6%
Sensitivity (Predicting Morality)	72.9%	22.7%	66.9%	9.1%

Figure 1. Variable importance plot for random forest predicting mortality

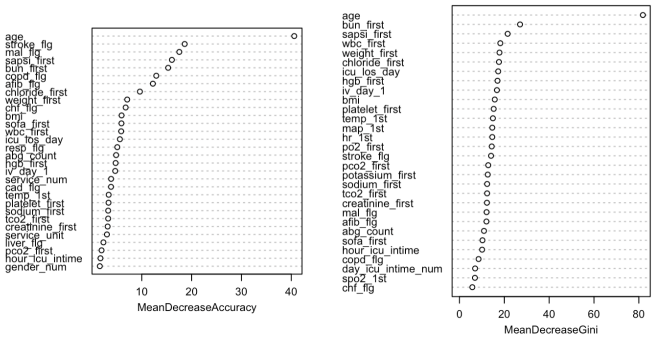

Figure 2. Partial dependence plot (PDP) for age on death

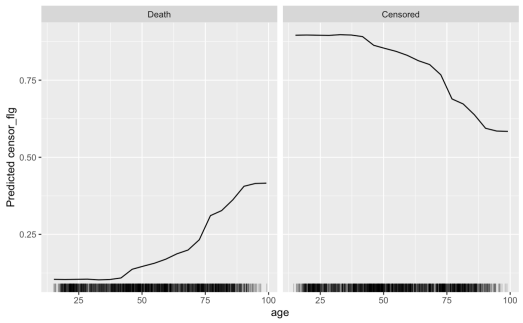

Figure 3. PDP for blood urea nitrogen on death

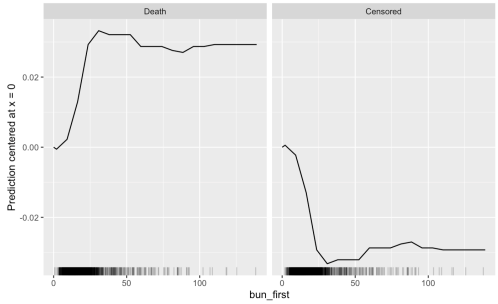

Table 4: Logistic Regression Results

	Dependent variable:		
	censor_flg		
	(1)	(2)	(3)
icu_los_day	−0.041	−0.047	−0.182**
	(0.032)	(0.041)	(0.074)
age	−0.049***	−0.053***	−0.044**
	(0.006)	(0.011)	(0.021)
gender_numMale	−0.519***	−0.517**	0.155
	(0.196)	(0.226)	(0.420)
weight_first	0.013*	0.015*	−0.023**
	(0.007)	(0.009)	(0.012)
bmi	0.008	0.008	0.126***
	(0.020)	(0.025)	(0.045)
sapsi_first	−0.039	−0.076**	−0.070
	(0.027)	(0.035)	(0.057)
service_unitMICU	0.237	−0.336	1.601*
	(0.536)	(0.673)	(0.862)
service_unitSICU	0.345	−0.334	1.805**
	(0.516)	(0.655)	(0.817)
service_numSICU			
chf_flgYes	−0.061	−0.219	−1.337*
	(0.259)	(0.284)	(0.752)
afib_flgYes	−0.540**	−0.488*	−0.892
	(0.238)	(0.259)	(0.929)
renal_flgYes	0.681	0.971**	2.622
	(0.417)	(0.453)	(1.978)
liver_flgYes	−0.062	−0.341	0.458
	(0.372)	(0.441)	(0.635)
copd_flgYes	−0.920***	−0.725**	−0.621
	(0.269)	(0.288)	(0.755)
cad_flgYes	0.444	0.121	−0.911
	(0.305)	(0.303)	(0.923)
stroke_flgYes	−1.618***	−1.763***	−2.264***
	(0.244)	(0.290)	(0.652)
mal_flgYes	−1.350***	−1.211***	−2.276***
	(0.222)	(0.277)	(0.510)
resp_flgYes	−0.530***	−0.518**	−0.926**
	(0.202)	(0.238)	(0.405)
wbc_first	0.004	−0.035	0.114***
	(0.013)	(0.021)	(0.044)
hgb_first	0.146***	0.169***	0.172**
	(0.046)	(0.060)	(0.084)
platelet_first	0.001	0.001	0.001
	(0.001)	(0.001)	(0.002)
chloride_first	0.085***	0.080**	0.154***
	(0.024)	(0.033)	(0.044)
bun_first	−0.015**	−0.005	−0.046**
	(0.007)	(0.008)	(0.022)
Constant	−3.111	2.352	0.098
	(3.675)	(4.322)	(12.751)
Observations	1,243	621	621
Akaike Inf. Crit.	1,028.112	726.821	333.155
Note:			*p<0.1; **p<0.05; ***p<0.01

Model (1) is the full logistic regression . Model (2) is the logistic regression model on the old-age group and model (3) is the logistic regression model on the young group.

IV. Discussion

The goal of our project was to assess whether a troupe of machine learning techniques offer any advantages to predicting length of stay, mortality, and pre-ICU conditions. Health professionals can benefit from advanced accurate predictions of ICU patient monitoring to adequately make stronger decisions on challenges in managerial health care. Poor management can result in patient rejection due to bed scarcity, loss of resources and manpower, and reduction in hospital revenue and quality of service (Schmidt et al 2013). Integrating machine learning models into intelligent ICU patient monitoring can have several beneficial implications. These implications were explored in three main questions: ICU length of stay, patient mortality, and pre-ICU conditions. Receiving more informed decision making in these areas can result in reductions in unwanted medical situations.

As we assumed, there was not a clear linear relationship between ICU length of stay and our variables of interest. Support vector regression worked well with our unstructured data, as did a random forest that fit our entire dataset. The models allowed us to observe variables that contributed differences to patients with increased length of stay. Variables such as the abg_count, age, and the service unit were statistically significant predictors of increased ICU stay. All models produced rmse values that indicated increases in prediction accuracy (in the scope of the range of our data, compared to random classification). Gaining full access to the dataset and filtering for patients that stayed beyond 5 days could allow exclusion of patients who passed away in the ICU and therefore more precise data points. A larger sample size would additionally assist in more accurate prediction. When split, our testing dataset for long length of stay contained 46 observations; to truly analyze the performance of a model we would need more observations among long stay patients to see why they remain in the ICU for a longer period of time.

In a study similar to ours, researchers at the University of Pennsylvania predicted mortality for cancer patients within the Penn Medicine oncology practices. Using random forest algorithms, this study observed 500 day mortality of 64.4% in the high risk group and 7.6% mortality in the low risk group (logarithmic regression and gradient boosting algorithms were also employed, but only results for the random forest model are published) (Parikh et al 2019). Our best mortality prediction model, using logistic regression, observed 2 year mortality of 70.5% in the high risk group and 15.9% in the low risk group. These models both aim to predict mortality in order to assist in providing the end of life care that the patient desires. Since classifying someone as in the high risk group increases the likelihood of introducing conversations surrounding end of life care, and therefore a choice to choose palliative care, we consider it more important to classify the high-risk group correctly. As such, we consider our model to have outperformed Penn Medicine's.

We also view logistic regression as advantageous over other models due to the ease of modifying the threshold between "Censored" and "Death." In this report, we use the Youden index for this threshold in order to maximize both sensitivity and specificity, which resulted in a threshold of 0.5866, which means that observations had to have at least a 58.6% chance of survival to be classified by the model into the low-risk group. Admittedly, this is a pretty high threshold for this classification; in the future, hospitals could choose a lower threshold in order to increase certainty for classification into the high risk group.

One important variable that we included in our model is the Simplified Acute Physiology Score (I) predictor, which is formed from the collection of 14 physiological variables collected within 24 hours of admittance to the ICU to indicate the risk of death (Le Gall et al. 1984). While many of these variables are also included in the general dataset that we use in this paper, we keep SAPSI and view it as an

interaction term between the variables that form the score. Using the formula by Le Gall et al. 1993, we mutate a SAPSI mortality prediction into our dataset. Since the maximum predicted mortality derived through this formula was under 20%, we claim that our models outperform this score calculator. However, due to the relative importance of SAPSI in predicting in random forests, we still support the collection of this metric for use in machine learning algorithms to predict mortality.

Our efforts to predict ICU pre-conditions with easily gathered medical metrics yield disappointing results. Both our logistic regression model and our random forest model had high accuracy (94.7% and 98.4% respectively) and specificity (98.4% and 99.8% respectively), but underwhelmed with low sensitivity scores (25.9% and 7.4% respectively). Catching positive cases of liver disease is important in an ICU setting where it can cause complications, so we hoped for a higher sensitivity.

Fortunately, our results appear in line with similar research. Previous efforts to predict liver disease in patients achieved a positive predictive rate between 20% and 30%, similar to our logistic regression model (McLernon 2014). It is possible that liver disease is simply too difficult to predict with great certainty. However, the optimal decision threshold for our logistic regression model of 0.07 yielded a sensitivity of 70%. This result is exciting given the significant improvement in sensitivity compared to our other models, but the low decision threshold and lower specificity of 85% does make us wonder if the higher performance is reflective of a good model or merely an incredibly low decision threshold. Still, our results do suggest that there might be hope for doctors and researchers attempting to identify patients that might have liver disease or other comorbidities.

Additionally, the limited MIMIC-II data that we were using only had 99 cases of liver disease. This may constrain our ability to properly train and test the model. Getting access to MIMIC-IV would greatly increase the amount of data we had to work with and increase the number of observations we could train and test our model on. Additionally, we only used two predictive methods for this research question. Focused study on predicting liver disease would allow for more predictive methods to be used. One of these other methods may prove superior to the ones we used. Finally, we selected live disease in part because it is a difficult comorbidity to diagnose. Applying our same work to a different comorbidity contained in the dataset might produce different results. Testing our process on different diseases would be an interesting extension of this research. In sum, we are unable to confidently predict comorbidities that ICU admittance may have with our current models; however, our work does suggest that this line of research could be fruitful if further pursued.

As is with many real world applications, the best result from a model will depend on the data used. When wrangling data, it was crucial to extract covariates, but extraction of any variables must be done with caution because excluding relevant features in advance may generate sub optimal results. Furthermore, many variables interact with cloud data. For example, more severe patients tend to have longer lengths of stay, but the sickest patients are also those at higher risk of death which will decrease the LOS. Having access to a full dataset, we could filter for patients who stayed beyond 5 days in the ICU, thus eliminating patients who passed shortly after admission, and examine what predicts longer stays in these patients. We additionally predicted mortality after two years; in many cases, hospitals are interested primarily in 6 month mortality. We also used the MIMIC database which represents a subset of a patient population from a singular hospital in Boston, and does not generalize to other populations or hospital systems in other areas across the United States and rest of the world. Additionally, machine learning algorithms to predict patient mortality typically utilize much larger patient databases with a wider array of variables. Finally, our data contained no observations of sepsis, which is a primary target to either prevent or predict using machine learning since sepsis management is highly time sensitive in

order to prevent mortality (Goh et al. 2021). It is unrealistic that there are incidences of sepsis at Beth Israel Hospital, so obtaining and training models on data containing sepsis is critical for future research to understand how sepsis affects patients in hospitals and the ICU.

Future directions and extensions of this research will involve extending the scope beyond Beth Israel Medical center, using holistic data to make broader optimal clinical intervention decisions. This research project thus lays a foundation for the future application of 3 patient related prediction models in ICU clinical practice.

V. References

Angus DC, Linde-Zwirble WT, Sirio CA, et al. The effect of managed care on ICU length of stay: implications for medicare. JAMA. Oct 2 1996;276(13):1075-1082

Alghatani, K., Ammar, N., Rezgui, A., & Shaban-Nejad, A. (2021). Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation. JMIR medical informatics, 9(5), e21347.

Brinkman-Stoppelenburg A, Rietjens JAC, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. Palliat Med. 2014;28(8):1000-1025.

Evans, J.; Kobewka, D.; Thavorn, K.; D'Egidio, G.; Rosenberg, E.; Kyeremanteng, K. The impact of reducing intensive care unit length of stay on hospital costs: Evidence from a tertiary care hospital in Canada. Can. J. Anesth. J. Can. d'anesthésie 2018, 65, 627–635

Garcia, Alberto. "Critical care issues in the early management of severe trauma." The Surgical clinics of North America vol. 86,6 (2006): 1359-87. doi:10.1016/j.suc.2006.07.004

Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med. Jan 2010;38(1):65-71

Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993;270(24):2957-63.

Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. Critical Care Medicine. 1984 Nov;12(11):975-977. DOI: 10.1097/00003246-198411000-00012. PMID: 6499483.

Marshall JC, Bosco L, Adhikari NK, Connolly B, Diaz JV, Dorman T, Fowler RA, Meyfroidt G, Nakagawa S, Pelosi P, Vincent JL, Vollman K, Zimmerman J. "What is an intensive care unit? A

report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine." *J Crit Care*. 2017 Feb;37:270-276. doi: 10.1016/j.jcrc.2016.07.015. Epub 2016 Jul 25. PMID: 27612678.

Mayo Clinic. "Liver Disease." Mayo Clinic.

Orwelius, Lotti et al. "Pre-existing disease: the most important factor for health-related quality of life long-term after critical illness: a prospective, longitudinal, multicentre trial." Critical care (London, England) vol. 14,2 (2010): R67. doi:10.1186/cc8967

Parikh, R.B., Liu, M., Li, E. et al. Trajectories of mortality risk among patients with cancer and associated end-of-life utilization. npj Digit. Med. 4, 104 (2021).

Parikh RB, Manz C, Chivers C, et al. Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer. JAMA Netw Open. 2019; 2(10).

Peres, I. T., Hamacher, S., Oliveira, F., Bozza, F. A., and Salluh, J. (2021). Prediction of intensive care units length of stay: a concise review. Predição do tempo de permanência em unidades de terapia intensiva: uma revisão concisa. Revista Brasileira de terapia intensiva, 33(2), 183–187. https://doi.org/10.5935/0103-507X.20210025

Rapoport J, Teres D, Lemeshow S, Avrunin JS, Haber R. Explaining variability of cost using a severity-of-illness measure for ICU patients. Med Care. Apr 1990;28(4):338-348.

Rapoport J, Teres D, Lemeshow S, Gehlbach S. A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. Crit Care Med. Sep 1994;22(9):1385-1391.

Shaban-Nejad A, Michalowski M, Buckeridge DL. (2018) Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med,1:53

Schmidt R, Geisler S, Spreckelsen C. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. BMC Med Inform Decis Mak. 2013 Jan 07;13:3

Tomašev, N., Glorot, X., Rae, J.W. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572, 116–119 (2019)

Vincent, J.L.; Marshall, J.C.; Namendys-Silva, S.A.; Francois, B.; Martin-Loeches, I.; Lipman, J.; Reinhart, K.; Antonelli, M.; Pickkers, P.; Njimi, H.; et al. Assessment of the worldwide burden of critical illness: The Intensive Care Over Nations (ICON) audit. Lancet Resp. Med. 2014, 2, 380–386

World Health Organization: WHO Definition of Palliative Care. Published 2017. Accessed August 15th, 2017.

Wu AW, Pronovost P, Morlock L. ICU incident reporting systems. J Crit Care. Jun 2002;17(2):86-94

Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA. 2008;300(14):1665-1673

Young MP, Birkmeyer JD. Potential reduction in mortality rates using an intensivist model to manage intensive care units. Eff Clin Pract. Nov-Dec 2000;3(6):284-289