Predictive Analysis of ICU Patient Survival Dataset

Weining Wu
Lehigh University

Professor Catherine Ridings

BIS 357 Artificial Intelligence for Business

December 11th, 2022

1. Introduction

Medical Data Growth

The big data industry is growing rapidly as the cost of data storage, network communication and computing decreases. People accumulate huge amounts of data everyday. Medical data is one such data. The chief executive of Talkspace, an online and mobile therapy company in New York, wrote about the development of medical data in an article in the New York Times. He mentioned that doctors used to handwrite medical records and store them in paper folders. Now, however, doctors use computers to enter medical records. With vast amounts of patient health data being collected and digitized every day, "if aggregated, our anonymized health records could become part of a large-scale data set to improve the diagnosis and treatment of illnesses across all medical fields using machine learning algorithms "(Frank, 2019). Artificial Intelligence(AI), which applies machine learning, has the ability to learn large numbers of cases quickly, and this learning efficiency is crucial to improving hospital operations. The core business of hospitals is to race against the clock and treat patients before their conditions get worse. With the help of AI, if doctors can quickly gain insights from similar cases, they will be able to tailor treatment plans for patients faster.

Doctors must read a large number of cases before they can summarize a few treatments that have a high probability of success. According to a 2021 paper written by Ron Winslow, the Deputy Bureau Chief of Health and Science at The Wall Street Journal, although people now use randomized clinical trials to study more effective treatments, yet "the elderly, children, women, minority groups and people who live far from medical research centers have long been underrepresented in such studies"(Winslow, 2021). “Researchers have believed for at least a halfcentury that data in patient medical records [outside clinical studies] could help fill the gaps"(Winslow, 2021). Therefore, it is necessary to analyze medical data extensively.

Intensive Care Unit Context

Of all the operations in a hospital, Intensive Care Unit (ICU) is the most expensive, both in terms of the hospital's investment in the ICU and in terms of the bills for patients staying in the ICU. The latest data suggested that even with insurance, the cost of an ICU stay for a COVID-19 patient can exceed $100,000 (Arradondo, 2021). Hospitals expect that their investment in the ICU will result in a decline in mortality rate in the unit. A high ICU survival rate of a hospital also convinces patients and their families more willing to pay the expensive bills for the hospital's ICU. One way to improve survival rate is to use data analysis to predict patients' mortality and to adjust their treatment plans based on the prediction. The medical data set selected by the analyst of this predictive analysis project was collected from ICU specifically. The aim of this project is to find the most appropriate predictive model to predict a patient's mortality in ICU based on his or her clinical data.

Project Dataset

The data set of this prediction analysis has a total of 91713 data points, each with 85 variables. Each data point represents an ICU patient in a U.S. hospital in 2021. The data set records the survival or death of these patients. It also includes patient data that were monitored during the patients' stay in the ICU, including their physical characteristics such as age, height, and weight, as well as vital signs such as blood pressure and heart rate. Massachusetts Institute of Technology (MIT) 's GOSSIS (Global Open Source Severity of Illness Score) initiated this statistical study and a Kaggle user posted the data on the Kaggle website (GOSSIS, 2021).

During the 2021 study, the GOSSIS researchers monitored and recorded patients' vital signs 1 hour after their admission to the ICU and again 24 hours after the admission. Some of the variables (columns) have names marked "apache" followed by different numbers. The researchers used Apache I, II, III and IV scoring systems during the study, which are commonly used in the ICU to show "an accurate measurement of patient severity and correlate strongly with outcome in critical patients" (Akavipat, Phuping, et al., 2019). The variables with the "apache" tags came from an Apache scoring process. Finally, the researchers recorded whether each patient died during their stay in the ICU.

The data set is very comprehensive because it includes a wide range of patient data. However, there are some limitations. First of all, the researchers only recorded whether the patients died in ICU in 2021, but they have not followed up the patients to record whether they died after leaving ICU. Therefore, this data set is limited to helping analysts study mortality within the ICU, not for studying whether certain clinical data can predict long-term health outcomes of patients. Secondly, the analyst of this prediction analysis detected 288,046 missing values in the data. Because the analyst could not determine which variables are important predictors of patient mortality, she assumed that any missing variables may lead to bias in the following model results. The analyst should clean all the missing values prior to modeling analysis. Finally, the publisher of the data has not given the information of the hospitals from which the data came, including the equipment condition and doctor qualification of these hospitals. Patient deaths could be caused by outdated hospital equipment or improper decision-making by doctors. These possibilities need to be tested with more studies.

2. Detailed Description of Data

In this predictive analysis specifically, the analyst used only 15 variables in the data set, among which hospital_death was the target variable and the remaining 14 variables were potential predictors of hospital_death, including: age, elective_surgery, weight, arf_apache, d1_temp_max, d1_temp_min, d1_heartrate_max, d1_heartrate_min, d1_mbp_max, d1_mbp_min, d1_resprate_max, d1_resprate_min, d1_glucose_max, d1_glucose_min, and hospital_death.

The definitions in detail for these variables are in Appendix A. They are the data available to doctors within the first day of the patient's admission to the ICU, and they are also commonly used data that hospitals routinely record. These variables help analyze if hospitals can use the patients' data recorded on the first day of their admission to the ICU to predict patient mortality.

To optimize the following modeling process, the data were preprocessed. The predictor variables should be completely independent of each other. Based on the correlation table of current data (see Appendix B for details), d1_temp_max， d1_heartrate_max, d1_mbp_max, d1_resprate_min and d1_glucose_min have high correlations with other variables, indicating that they are strongly dependent on other variables. Therefore, the analyst removed these variables. Next, the analyst removed the missing values from the data. The cleaned data set used for modeling has a total of 77,616 data points, with 10 variables. This data set contains three categorical variables: elective_surgery, arf_apache and hospital_death， each with two categories. Their definitions are in Appendix C. The remaining seven variables are numerical. Variables in the data set after data preprocessing are relatively independent, according to the new correlation table (see Appendix E).

According to the data description tables (see Table 2 and Table 3), most of the cases in the data were over 50 years old and weighed more than 70 kg in 2021. Therefore, most of the ICU patients in this analysis are elderly. Data of the three categorical variables is severely tilted. Most of the ICU patients were admitted to the hospital not for an elective surgical operation, had no acute renal failure during the first 24 hours of their unit stay, or did not die in the ICU. From the point of view of vital signs, compared with the normal vital sign ranges (Vital signs, 2019), patients in the study had the following characteristics: the first day body temperatures were concentrated around 36 degrees Celsius, with rarely fever; there was a huge variation in heart rates among the patients, as well as rapid heart rates and cardiac arrest on the first day; blood pressure varied widely among patients, with hypotension but no hypertension; the respiratory rate was generally too high; and more than half of the patients had high glucose concentration.

Table 2. *Cleaned Data Description Table Part I*

	age	**elective_surgery**	**weight**	**arf_apache**	**d1_temp_min**
count	77616.00	77616.00	77616.00	77616.00	77616.00
mean	62.58	0.19	84.83	0.03	36.26
Standard deviation	16.59	0.40	25.13	0.17	0.76
min	16.00	0.00	38.60	0.00	31.89
25%	53.00	0.00	67.70	0.00	36.10
50%	65.00	0.00	81.30	0.00	36.40
75%	75.00	0.00	98.00	0.00	36.61
max	89.00	1.00	186.00	1.00	37.80

(Keep two decimal places)

Table 3. *Cleaned Data Description Table Part II*

	d1_heartbeat_ min	**d1_mbp_ min**	**d1_resperate_ min**	**d1_glucose_ max**	**hospital_death**

count	77616.00	77616.00	77616.00	77616.00	77616.00
mean	70.68	64.64	28.73	175.79	0.08
Standard deviation	16.88	15.53	10.39	87.25	0.28
min	0.00	22.00	14.00	73.00	0.00
25%	60.00	55.00	22.00	118.00	0.00
50%	70.00	64.00	26.00	151.00	0.00
75%	81.00	74.00	32.00	203.00	0.00
max	160.00	112.00	92.00	611.00	1.00

(Keep two decimal places)

3. Three AI Business Questions

People can use the data set of this project to solve the following but not limited to the following 3 business questions.

Question 1. Can hospitals use the patients' data recorded on the first day of their admission to the ICU to predict patient mortality?

For this question, the independent variables involved in the analysis will include the patient's physical characteristics, such as age and weight, as well as data from the first day after their admission to the ICU, such as blood pressure, blood glucose, and heart rate in the first 24 hours. Dependent variables will be hospital_death indicating whether the patient died in the ICU. Necessary data transformations include clearing missing values and variables with excessive correlation with other variables.

Question 2. Does the stability of the patients' data on the first day of admission to the ICU correlate with the patient's eventual death?

For this question, the independent variables involved in the analysis will include vital signs at the first hour and the 24th hour after the patients were admitted to the ICU. Dependent variables will be hospital_death indicating whether the patient died in the ICU. Necessary data transformations include clearing missing values and variables with excessive correlation with other variables. In addition, the analyst will need to add columns to the data set to record the differences between the patient's physical signs in the first hour and 24 hours after admission to the ICU. The analyst can obtain the differences by subtracting the first hour's data from the data 24 hours later and then dividing the result by the first hour's data.

Question 3. Which of the APACHE I, II, III, and IV scoring systems has the highest accuracy in predicting patient mortality?

For this question, the independent variables covered by the analysis will include all 85 variables in the raw data set. Dependent variables will be hospital_death indicating whether the patient died in the ICU. Necessary data transformations include clearing missing values and variables with excessive correlation with other variables. In addition, the data set contains several categorical string variables. The analyst will need to create binary variables, substituting the string variables by 0 or 1. Before modeling, the analyst will need to divide independent variables into five categories: general variables (including basic patient information such as age, gender, and weight), variables recorded in Apache I, variables recorded in Apache II, variables recorded in Apache III, and variables recorded in Apache IV. The analyst will build four models, each of which uses general variables and variables in one of the four Apache systems as independent variables to train the model. Finally, the analyst should compare the accuracies of the four models to determine which Apache system can predict the mortality of ICU patients most effectively.

All the above three questions have the categorical variable hospital_death as their targets, so they are all classification problems. To solve classification problems, the analyst can use K-nearest Neighbors Modeling and Neural Network Modeling for prediction.

All three questions listed above are designed to predict mortality of ICU patients. In these cases, AI algorithms can help hospitals find methods to assess a patients' condition severities. Patients who are predicted to die have more severe conditions and should receive earlier treatments. Methods to assess or predict patients' conditions may include using a combination of individual variables as predictors, finding the relationships between variables as predictors, and assessing patients using the most effective scoring mechanism, corresponding to each of the three questions listed above.

From a business point of view, AI saves doctors the time and cost of reading and learning from plenty of medical records. At the same time, as mentioned in the introduction of this paper, the main business of a hospital is to treat patients. The ability to deliver treatment quickly and accurately to patients is the value a hospital brings to their clients and is the key to retain clients and to ensure their profitability. If AI can predict a patient's mortality within a short time, doctors will have the opportunity to determine the severity of a patient's condition early in the ICU, and plan treatment as early as possible to improve the chances of a cure. From a human health perspective, the prediction model could reduce ICU mortality. According to the Society of Critical Care Medicine, about 5 million people are admitted to intensive care units each year (Halpern). Furthermore, according to another statistic from the National Library of Medicine, 540,000 people die each year after being admitted to the ICU (Angus et al.). More than 10% of ICU patients die annually in the ICU, but AI intervention will potentially help doctors save time in evaluating patients' severity and improve treatment plans to increase the survival rate. If the AI can successfully select the most effective Apache scoring system, doctors can also apply the system directly to assess patients' conditions more accurately.

4. AI Model Analysis

---

a. Business Question and AI Technique:

This predictive analysis focused on the following question: can hospitals use the patients' data recorded on the first day of their admission to the ICU to predict patient mortality? To explore this question, the analyst tried to predict the categorical variable hospital_death. Therefore, she applied classification models in the analysis, including KNN Modeling and Neural Network Modeling. After the respective modeling, the analyst compared the accuracies of the two models and chose the Neural Network as the suitable one for this type of medical prediction because it had higher accuracy.

b. Data Visualization

The analyst first focused on the impact of each single independent variable on patient mortality (see Appendix D). According to the figures, Patients who died in the ICU had lower body weight (Figure 3 in Appendix D), day 1 minimum heart rate of 40-120 beats per minute (Figure 6 in Appendix D), day 1 minimum mean blood pressure concentrated below 80 mmHg (Figure 7 in Appendix D), day 1 maximum respiration rate concentrated below 50 breaths per minute (Figure 8 in Appendix D), and day 1 maximum glucose rate concentrated blow 300 per kilogram body weight per minute (Figure 9 in Appendix D). However, patients with these characteristics do not necessarily die in the ICU. The other variables were not significantly indicative of patients' mortality.

c. AI Model Results

(1) KNN Modeling

Based on KNN modeling, the analyst also used K-fold Cross Validation to avoid model overfitting to the data used for training. Meanwhile, the analyst tested different k-values and compared their impacts on the accuracy of the model (see Table 4). The conclusion was that when k equals 15,17, or 19, the model had the highest accuracy of 91.84%. The accuracy represented that 91.84% of the predicted results of the data predicted by the KNN model were consistent with the actual results. This conclusion suggested that if hospitals decide to use KNN modeling to predict patient mortality, they should check 15, 17, or 19 neighbors to determine the classification of a specific query point. For example, if the person presiding over the prediction specifically chooses 15 as the k-value, and if most of the 15 neighbors of the instance are survivors, this instance will also be predicted as a survivor.

Table 4. *Comparing Model Accuracies with Different K-values*

K-value	**Model Accuracy**
1	87.13%
3	90.62%
5	91.43%
7	91.66%

9	91.74%
11	91.78%
13	91.81%
15	91.84%
17	91.84%
19	91.84%

(2) Neural Network Modeling

In Neural Network Modeling, the analyst used 75% of the data to train the model and 25% of the data to test the accuracy of the model. In addition, the analyst added three dense layers to the neural network, with 12, 8 and 1 nodes respectively. Because the neural network has a certain randomness, the analyst repeated the modeling 5 times to obtain an average model accuracy (see Table 5). The accuracies represented the percentages of the predicted results of the data predicted by the Neural Network model that were consistent with the actual results.

Table 5. *Model Accuracies of Neural Network Model*

Run Time	**Accuracy**
1	92.06%
2	92.04%
3	92.00%
4	92.07%
5	92.06%
Average	**92.05%**

(3) Model Comparison

From the point of view of the classifier efficiency, the computation time of KNN model and Neural Network model was similar due to the small data set used to train the model. From the classifier principal point of view, both classifiers are data-driven, so they are suitable for analyzing a wide range of data. Since the two classifiers perform similarly in both respects, the analyst determined the better model by comparing their accuracies. Because the accuracy of the Neural Network (92.05%) was higher than that of the KNN model (91.84%), Neural Network was the more suitable classifier for this analysis.

5. Conclusion

This predictive analysis was a preliminary attempt. The description of the data showed that the actual death cases in the data account for a small proportion. If analysts want to get more accurate conclusions through data analysis, collecting more cases in the future will be helpful. In addition, although the Neural Network is the best model based on this analysis, the analysts can keep searching for a more ideal Neural Network model by adding or removing layers or changing the number of nodes in each layer and conducting repeated tests. If this predictive model proves to be accurate after repeated testing, it could make a huge contribution to medicine. Although medical judgments made by machines may not be fully trusted today, such predictive models can be used as a tool for doctors to assess their usefulness over time.

References

Akavipat, Phuping, et al. (2019, March) Acute Physiology and Chronic Health Evaluation (Apache) II Score - the Clinical Predictor in Neurosurgical Intensive Care Unit. *Acta Clinica Croatica*, U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6629196/#:~:text=A%20widely%20used%20ICU%20prognostic,patients%20(4%2C%205)

Angus, Derek C., et al. Use of Intensive Care at the End of Life in the United States: An Epidemiologic Study*. *Critical Care Medicine*, vol. 32, no. 3, Mar. 2004, pp. 638–643., https://doi.org/10.1097/01.ccm.0000114816.62331.08

Arradondo, Briona. (2021, August 7). Stay in ICU for COVID Patients Can Cost More than $100k, Even with Insurance, Data Shows. *FOX 13 Tampa Bay*, FOX 13 Tampa Bay. https://www.fox13news.com/news/financial-burden-of-getting-covid-19

Frank, Oren. (2019, October 2). Donate Your Health Care Data Today. *The New York Times*, The New York Times. https://www.nytimes.com/2019/10/02/opinion/health-care-data-privacy.html

GOSSIS. (2021). Patient Survival Prediction [Data set]. Kaggle. https://www.kaggle.com/datasets/mitishaagarwal/patient?select=dataset.csv

Halpern, Neil A. SCCM: Critical Care Statistics. *Society of Critical Care Medicine (SCCM)*, https://www.sccm.org/Communications/Critical-Care-Statistics

Vital Signs. (2019, January 23). *Cleveland Clinic*, https://my.clevelandclinic.org/health/articles/10881-vital-signs

Winslow, Ron. (2021, December 3). Medical Records Data Offers Doctors Hope of Better Patient Care. *The Wall Street Journal*, Dow Jones & Company. https://www.wsj.com/articles/medical-records-data-offers-doctors-hope-of-better-patient-care-11638543613

Appendix

Appendix A. Selected Variables for Modeling and Their Definitions

Variable	**Definition**
age	The age of the patient on unit admission
elective_surgery	Whether the patient was admitted to the hospital for an elective surgical operation. Notes: Elective surgery: surgery that is scheduled in advance because it does not involve a medical emergency. Semi-elective surgery: surgery that must be done to preserve the patient's life but does not need to be performed immediately.
weight	The weight (body mass) of the person on unit admission
arf_apache	Whether the patient had acute renal failure during the first 24 hours of their unit stay. Notes: Acute renal failure is defined as a 24-hour urine output <410ml, creatinine >=133 micromol/L and no chronic dialysis
d1_temp_max	The patient's highest core temperature during the first 24 hours of their unit stay, invasively measured

d1_temp_min	The patient's lowest core temperature during the first 24 hours of their unit stays
d1_heartrate_max	The patient's highest heart rate during the first 24 hours of their unit stays
d1_heartrate_min	The patient's lowest heart rate during the first 24 hours of their unit stays
d1_mbp_max	The patient's highest mean blood pressure during the first 24 hours of their unit stays, either non-invasively or invasively measured
d1_mbp_min	The patient's lowest mean blood pressure during the first 24 hours of their unit stays, either non-invasively or invasively measured
d1_resprate_max	The patient's highest respiratory rate during the first 24 hours of their unit stays
d1_resprate_min	The patient's lowest respiratory rate during the first 24 hours of their unit stays
d1_glucose_max	The highest glucose concentration of the patient in their serum or plasma during the first 24 hours of their unit stay
d1_glucose_min	The lowest glucose concentration of the patient in their serum or plasma during the first 24 hours of their unit stay
hospital_death	Whether the patient died during this hospitalization

Appendix B. Correlation Table for First Selected Variables

	age	elective_surgery	weight	arf_apache	d1_temp_max	d1_temp_min	d1_heartrate_max	d1_heartrate_min	d1_mbp_max	d1_mbp_min	d1_resprate_max	d1_resprate_min	d1_glucose_max	d1_glucose_min	hospital_death
age	1.000000	0.067320	-0.127252	-0.001684	-0.082764	-0.070002	-0.135417	-0.143705	0.006123	-0.131755	0.032206	0.034439	0.012538	0.067118	0.111017
elective_surgery	0.067320	1.000000	0.026900	-0.027357	0.077195	-0.033104	-0.073569	-0.025685	-0.141750	0.012453	-0.063515	-0.172378	-0.014436	-0.004182	-0.093574
weight	-0.127252	0.026900	1.000000	-0.010785	0.032039	0.036138	-0.038491	0.000898	0.070165	0.055879	-0.006198	-0.035975	0.090293	0.134588	-0.038362
arf_apache	-0.001684	-0.027357	-0.010785	1.000000	-0.019952	-0.027585	-0.014301	-0.004353	0.019230	-0.028282	0.012278	-0.022140	0.030502	-0.058829	0.027309
d1_temp_max	-0.082764	0.077195	0.032039	-0.019952	1.000000	0.256638	0.273183	0.181212	-0.006388	-0.121309	0.121657	-0.001549	0.001565	-0.025125	0.006293
d1_temp_min	-0.070002	-0.033104	0.036138	-0.027585	0.256638	1.000000	0.002657	0.160161	-0.013963	0.124983	-0.037062	0.109429	-0.105660	0.034895	-0.207239
d1_heartrate_max	-0.135417	-0.073569	-0.038491	-0.014301	0.273183	0.002657	1.000000	0.465056	0.108394	-0.176183	0.249341	0.028852	0.110566	0.053704	0.162934
d1_heartrate_min	-0.143705	-0.025685	0.000898	-0.004353	0.181212	0.160161	0.465056	1.000000	-0.072071	0.084784	0.039423	0.259548	0.096901	0.056860	-0.003587
d1_mbp_max	0.006123	-0.141750	0.070165	0.019230	-0.006388	-0.013963	0.108394	-0.072071	1.000000	0.247351	0.167384	-0.063986	0.021628	0.036721	-0.016752
d1_mbp_min	-0.131755	0.012453	0.055879	-0.028282	-0.121309	0.124983	-0.176183	0.084784	0.247351	1.000000	-0.065137	0.149848	-0.046753	0.054510	-0.195262
d1_resprate_max	0.032206	-0.063515	-0.006198	0.012278	0.121657	-0.037062	0.249341	0.039423	0.167384	-0.065137	1.000000	0.014941	0.034723	0.018707	0.103093
d1_resprate_min	0.034439	-0.172378	-0.035975	-0.022140	-0.001549	0.109429	0.028852	0.259548	-0.063986	0.149848	0.014941	1.000000	0.008490	0.054075	0.025667
d1_glucose_max	0.012538	-0.014436	0.090293	0.030502	0.001565	-0.105660	0.110566	0.096901	0.021628	-0.046753	0.034723	0.008490	1.000000	0.383948	0.081568
d1_glucose_min	0.067118	-0.004182	0.134588	-0.058829	-0.025125	0.034895	0.053704	0.056860	0.036721	0.054510	0.018707	0.054075	0.383948	1.000000	0.029884
hospital_death	0.111017	-0.093574	-0.038362	0.027309	0.006293	-0.207239	0.162934	-0.003587	-0.016752	-0.195262	0.103093	0.025667	0.081568	0.029884	1.000000

Appendix C. Categorical Variables and Their Definitions

	`Value: 0`	`Value: 1`
elective_surgery	The patient was admitted to the hospital not for an elective surgical operation	The patient was admitted to the hospital for an elective surgical operation
arf_apache	The patient had no acute renal failure during the first 24 hours of their unit stay	The patient had acute renal failure during the first 24 hours of their unit stay
Hospital death	The patient survived during this hospitalization	The patient died during this hospitalization

Appendix D. Scatter Plots of Patient Mortality by Different Individual Variables

(Note: The y value hosptal_death in the figures indicates the mortality of the patients.)

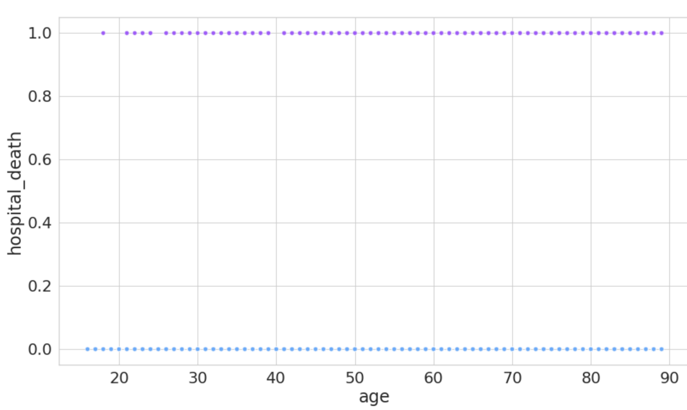

Figure 1: Scatter Plot of Patient Mortality by Age

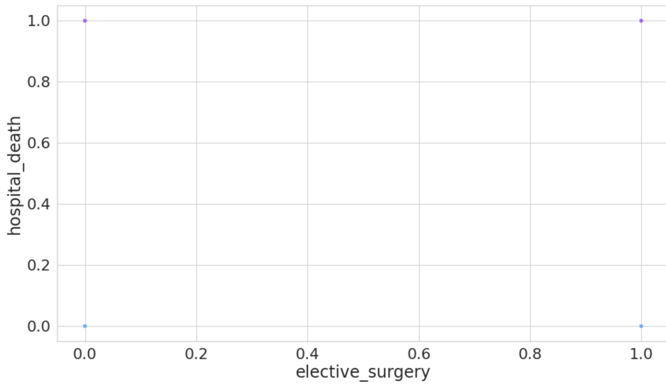

Figure 2: Scatter Plot of Patient Mortality by Whether Patients in ICU for Elective Surgery

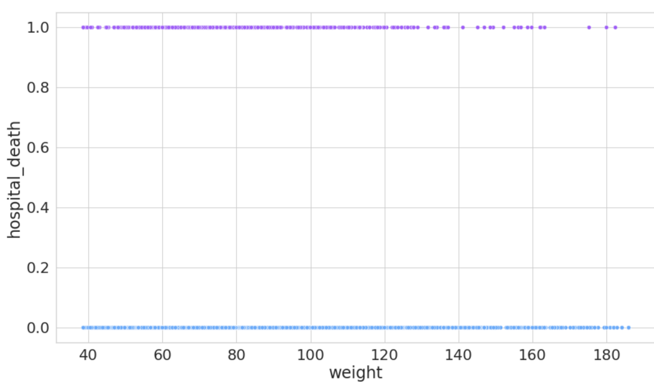

Figure 3: Scatter Plot of Patient Mortality by Weight

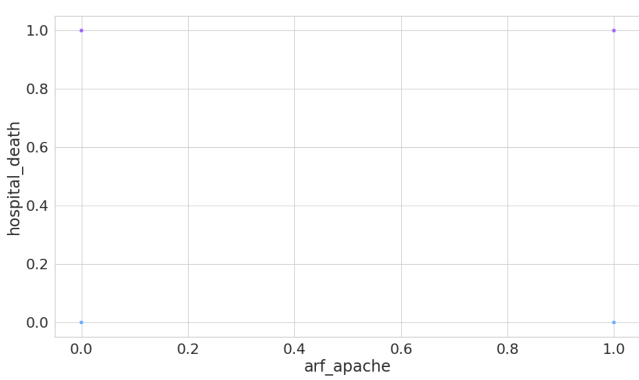

Figure 4: Scatter Plot of Patient Mortality by Whether Patient Had ARF

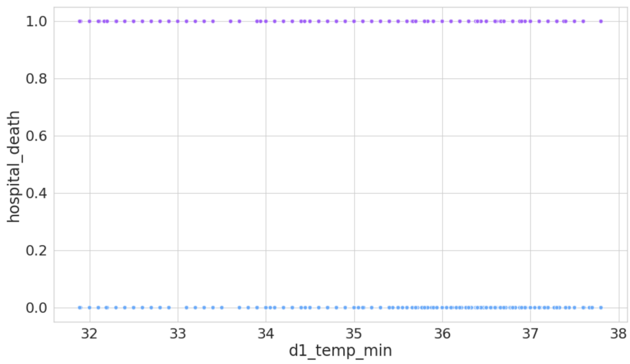

Figure 5: Scatter Plot of Patient Mortality by Day 1 Minimum Temperature

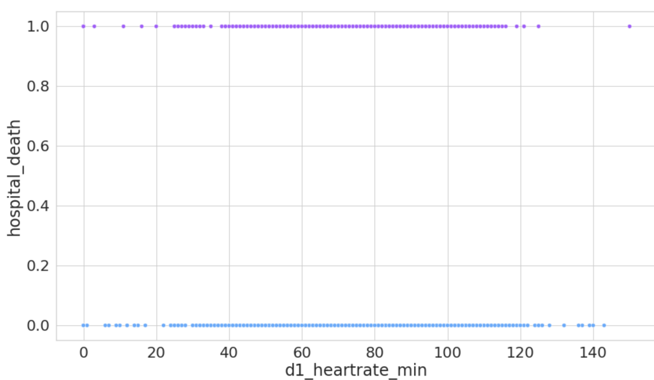

Figure 6: Scatter Plot of Patient Mortality by Day 1 Minimum Heart Rate

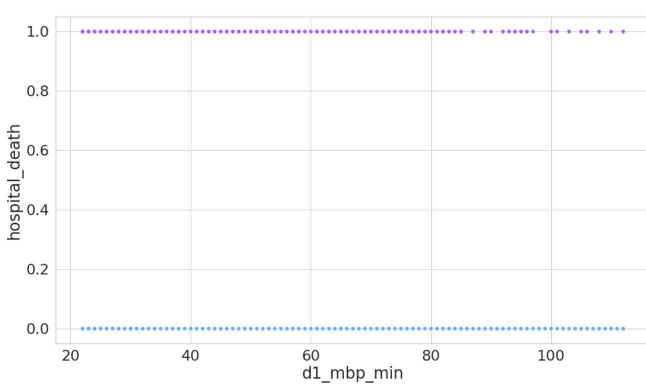

Figure 7: Scatter Plot of Patient Mortality by Day 1 Minimum MBP

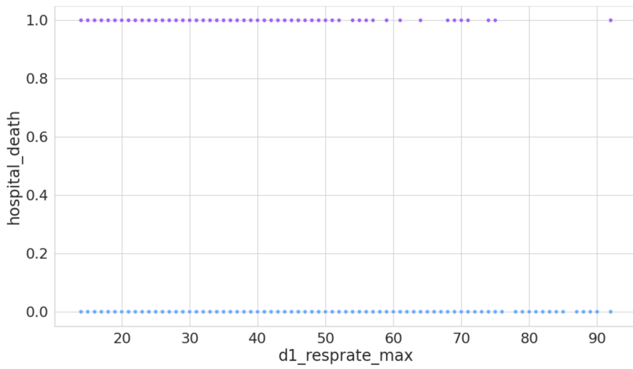

Figure 8: Scatter Plot of Patient Mortality by Day 1 Maximum Respiration Rate

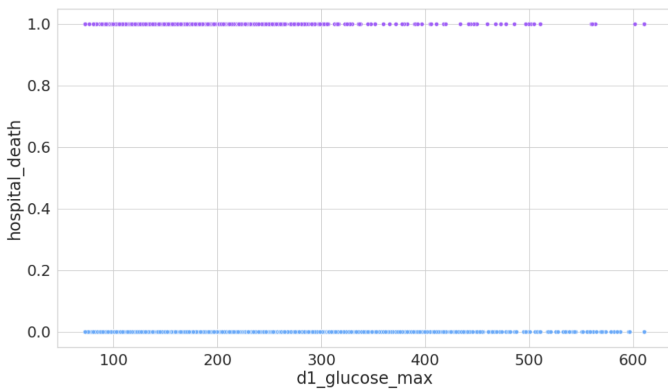

Figure 9: Scatter Plot of Patient Mortality by Day 1 Maximum Glucose Rate

Appendix E. Correlation Table of Cleaned Data

	age	elective_surgery	weight	arf_apache	d1_temp_min	d1_heartrate_min	d1_mbp_min	d1_resprate_max	d1_glucose_max	hospital_death
age	1.000000	0.067320	-0.127252	-0.001684	-0.070002	-0.143705	-0.131755	0.032206	0.012538	0.111017
elective_surgery	0.067320	1.000000	0.026900	-0.027357	-0.033104	-0.025685	0.012453	-0.063515	-0.014436	-0.093574
weight	-0.127252	0.026900	1.000000	-0.010785	0.036138	0.000898	0.055879	-0.006198	0.090293	-0.038362
arf_apache	-0.001684	-0.027357	-0.010785	1.000000	-0.027585	-0.004353	-0.028282	0.012278	0.030502	0.027309
d1_temp_min	-0.070002	-0.033104	0.036138	-0.027585	1.000000	0.160161	0.124983	-0.037062	-0.105660	-0.207239
d1_heartrate_min	-0.143705	-0.025685	0.000898	-0.004353	0.160161	1.000000	0.084784	0.039423	0.096901	-0.003587
d1_mbp_min	-0.131755	0.012453	0.055879	-0.028282	0.124983	0.084784	1.000000	-0.065137	-0.046753	-0.195262
d1_resprate_max	0.032206	-0.063515	-0.006198	0.012278	-0.037062	0.039423	-0.065137	1.000000	0.034723	0.103093
d1_glucose_max	0.012538	-0.014436	0.090293	0.030502	-0.105660	0.096901	-0.046753	0.034723	1.000000	0.081568
hospital_death	0.111017	-0.093574	-0.038362	0.027309	-0.207239	-0.003587	-0.195262	0.103093	0.081568	1.000000