Team Leader: Junwei(Alex) Huang Active Members: Sarayu, Noha, Jess, Oliver

Report

Project Description and Details:

Intensive care units (ICU) is something that all major hospitals have, as a way of providing round the clock care, specialised treatment, critical care and life support to ones who are acutely unwell and require critical medical care. In today’s world that is filled with data, healthcare providers collect data to evaluate and improve the quality-of-care in hospitals, especially in ICUs, to compare effectiveness of medications and treatments, and improve efficiencies of emergency departments. etc.

The dataset provided was a list of about 1500 ICU patients and their vital recordings across 48 hours, including whether they survived or not. The main aim of the project was to build a model that predicted whether a patient would survive their ICU stay. Using information on patient demographics and indicators of health, we determined which variables have a greater impact on the outcome. The model can be used as a base for hospitals to compare and rank their performance against as well as assess a patient's risk.

We approached this aim through the use of an outcome descriptor known as SAPS. SAPS is an adult physiology score calculated into a percentage based upon 17 variables derived from the APACHE score and the data is collected within the first 24 hours of ICU care to notify people of the severity of the disease (Bersten, 2019). The SAPS score allowed us to determine which factors impacted mortality rate the most by picking out the factors in the data which were most relevant instead of manually checking on python to see which factors correlated with each other to impact mortality.

The data provided had two versions, an unprocessed and preprocessed dataset. The preprocessed ICU dataset, which had a shape of 1474 rows and 232 columns, with 340940 instances, provided us with a dataset containing 6 general descriptor variables, 37 time series variables and 5 outcome related descriptors. The main reason for the two types was that the time series variables were observed at certain times for the duration of the patients stay, meaning that the values could have fluctuations, whereas the descriptors would remain constant. It is also important to note that ‘Weight (kg)’ can be a general descriptor which is recorded on admission to the ICU or a time series variable that is measured hourly for the estimation of fluid balance (Physionet, n.d.).

The 6 general descriptors included a unique ID of every patient, including their personal details such as their age (years), gender (0: female,1: male), height (cm), weight (kg) and the respective ICU quarters they reside in (1: Coronary Care Unit, 2: Cardiac Surgery Recovery Unit, 3: Medical ICU, 4: Surgical ICU) . The Coronary Care unit provides care for the more acutely ill cardiac patients (NHS Foundation Trust, n.d.), while the Cardiac Surgery Recovery unit supplies complex invasive hemodynamic monitoring and treatment of adult patients during the critical phase of recovery from cardiac and thoracic surgery (University of Florida, n.d.). The Medical ICU refers to the specialised treatment given to patients who are acutely unwell and require critical medical care (Health Direct, 2020) and the Surgical ICU provides for critically ill patients who require surgery or are recovering from it (UC San Diego, 2022).

The 37 time series variables were a combination of blood drawn indicators and general health measurements that were recorded at either regular or irregular intervals, and only either once or multiple times a day. Some of the main indicators that relate the outcome used in our model were the ID of the patients, SAPS-I score, SOFA score, length of stay in the ICU, the duration of patients being alive (days) and in-hospital deaths (0:survivor, 1:died).

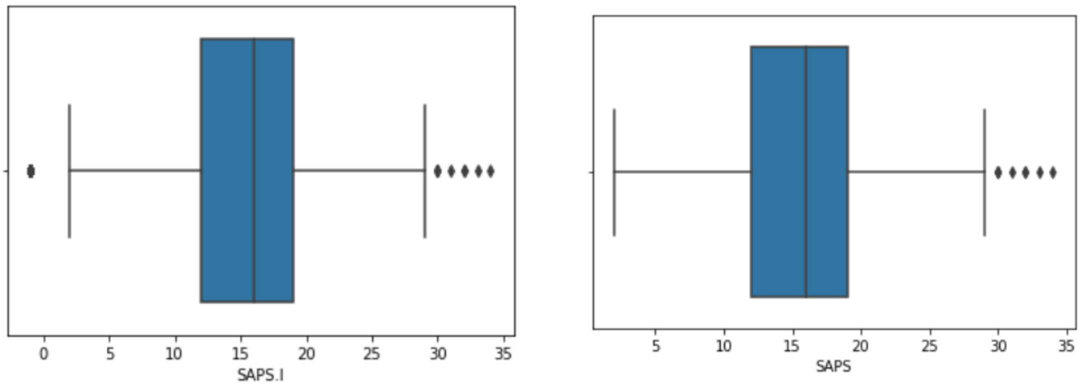
SOFA score stands for Sequential Organ Failure Assessment and is also a scoring system that assesses the performance of several organs based on six different criteria and is also used to estimate mortality in patients (ASPR TRACIE, 2015). The reason why we chose using the SAPS score over the SOFA score was because the SAPS score was more relevant and clinically in line with modern medicine standards as a predictor of mortality, so we decided to look at the correlation between the SAPS variables to determine which ones were the highest (Mungan et al., 2019).

Preprocessing:

The data was provided to us in two different versions, a pre-processed and unprocessed version. Logically, we preferred using the preprocessed dataset that was given to us rather than the unprocessed raw data as pulling out data for analysis would be easier compared to building a completely new dataframe from raw data. Upon comparing the two datasets, there was not much difference in the values nor in the information provided, so we thought it more practical to stick to the pre-processed version. In addition, the preprocessed excel file provided already included most of the NaN values removed so not much additional data cleansing had to be done.

While working and browsing through the preprocessed dataset, we noticed that there were several columns that had duplicated values and would take up a lot of space in python when trying to describe the data, thereby extending loading times for future modelling and calculations. That led us to the decision of removing the certain columns we believed were extra. For example the 6 columns of Mech Vent (mechanical ventilation) as all the values in the column were ones.

Utilising codes such as ICU.describe() and creating boxplots, multiple extreme outliers were spotted and were needed to be removed completely. In the beginning the preprocessed data originally started with 1474 rows, we decided that removing the rows with the extreme outliers altogether would not severely affect the dataset as the sample size was quite large and excluding extreme outliers would cause our results to be more statistically accurate and representative of the true outcome, as outliers increase the variability of the data, and hence decrease reliability.



Figures 1.1 and 1.2: Shows how preprocessing affected the data within the "SAPS" category. Data that contained a SAPS of -1 were removed as it is impossible to obtain such a score.

Some renaming of columns were required as it would benefit us from some miscommunications while coding. For instance, turning columns that have a "." into "_" to prevent problems in the modelling phase.

We chose to report upon 5 research questions, related to the questioning of the SAPS variables and how it would be useful in the prediction of the final outcome. The questions are listed down below.

1. Which variables contribute the most to a person's death?
2. How does the SAPS score determine which variables carry a greater risk of death?
3. What combination of factors would contribute to the increase in mortality the most?
4. How do the unknown variables contribute to the final outcome?
5. Is it possible to make a predictive model that would determine if a hypothetical patient would survive in the ICU?

We believe that these 5 questions covered the whole aspect of the ICU dataset to only achieve our aim, but also allow for further exploration and analysis to help broaden our understanding of the project as a whole.

Processing & Manipulation:

As SAPS is only an outcome indicator, we cannot fully conclude that it is the main reason for in-hospital deaths of patients, it has been proven to be valuable tool in providing a "net benefit" when predicting ICU mortality, especially if the probability threshold is low (Le Gall et al., 2005). This allowed us to determine more meaningful causes of death, narrowing down the 37 variables to only the 12 that were used in SAPS calculations.

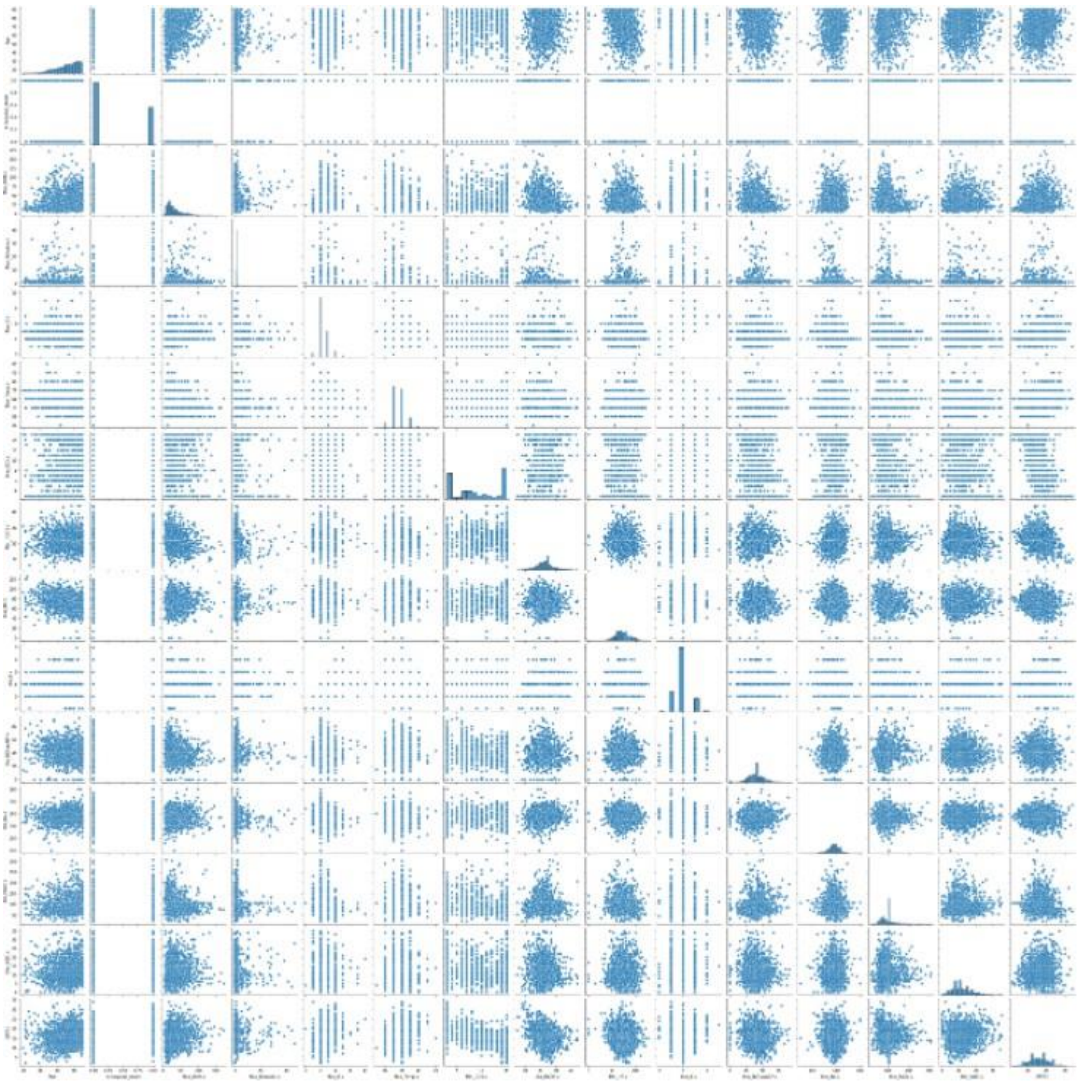


Figure 2.1: Pairplot of the 12 selected variables and SAPS and in hospital death outcomes.

When posing questions and hypotheses to analyse, we used variables suggested by the SAPS score to create a pivot table which would only include the outcome of in hospital death, and 12 of the most relevant variables to the SAPS score listed below.

- Maximum Bilirubin (mg/dL)
- Mean Bilirubin (mg/dL)
- Minimum Bilirubin (mg/dL)
- Maximum Partial Pressure of Oxygen (PaO2) if on mechanical ventilation (mmHg/%)
- Mean Partial Pressure of Oxygen (PaO2) if on mechanical ventilation (mmHg/%)
- Minimum Partial Pressure of Oxygen (PaO2) if on mechanical ventilation (mmHg/%)
- Maximum Blood Urea Nitrogen (BUN) (mg/dL)
- Minimum Sodium (Na) (mEq/L)
- Maximum Potassium (K) (mEq/L)
- Minimum Potassium (K) (mEq/L)
- Minimum White Blood Count (WBC) (x 10^3/mm^3)
- Minimum Bicarbonate (HCO3) (mEq/L)

Some of the variables appeared more than once as the maximum and minimum values, meaning they posed the most risk to mortality according to the SAPS score. Those variables include bilirubin which is an indicator of liver function (Healthline, n.d.), PaO2 which is a measurement of oxygen in arterial blood (Very Well Health, 2022), blood urea nitrogen (BUN) indicates the functionality of the liver (National Cancer Institute, n.d.), potassium (K) is used to maintain fluids in our cells whereas sodium (Na) maintains fluid levels outside our cells (Harvard School of Public Health, 2019), white blood cell count (WBC) as an indicator of infections, diseases or alike, and bicarbonate (HCO3) measuring the total amount of carbon dioxide in the blood (Pathology Tests Explained, n.d.)

To put some of these variables in context of the dataset, an example would be about 350 patients who had a minimum PaO2 count below 75, of which 45.7% died. This alludes that there is a high chance that when patients have a low volume of oxygen in their arterial blood, there is a connection to mortality rates. Another example from the dataset would be that 9% of the sample patients had a PaO2 count outside of the normal range (75-100 mmHg), resulting in low oxygen levels in the blood which suggest that further complications will arise. All patients in ICUs were under mechanical ventilation, meaning that their PaO2 levels had a contribution to their final SAPS score.

Another variable would be the blood urea nitrogen levels, with 67 people having a maximum BUN greater than or equal to 84 and 55.2% of them dying. This indicates severe problems with their liver and/or kidneys, while also gatewaying into several other problems such as heart failure.

Looking at biases in the dataset, age is a factor that influenced the outcome as a whole, as the average age of the patients is about 71 years old. This means that the predictive modelling would not be an accurate indicator for someone who is not of that range, but fits to the demographic of hospitals, as most patients tend to be elderly. In particular, there were 500 people who were over 80, and 40% of them had died in hospital,with age being the key factor. There were only 47 people less than 40 years, but had a higher death rate at 59.6%. This did not align with the SAPS calculator nor medical records in general, but could be caused from previous chronic conditions, which were not provided in the dataset.

Using the 12 variables, we created a heatmap to find the correlation between them and therefore, see which ones had the biggest contribution to SAPS. It would also give us an indication of which ones would have a bigger outcome on the patient. The heatmap suggests that there were 4 main variables (which did not include SAPS and in hospital death) that gave us the highest correlated values which were the Maximum Blood Urea Nitrogen (Max_BUN.x), Minimum and Maximum Potassium (Min_K.x & Max_K.x) and the Minimum Bicarbonate (Min_HCO3.x). Even though Minimum Glasgow Coma Scale (Min_GCS.x) and Maximum Temperature (Max_Temp.x) have a correlation of -0.27, which is a greater correlation than 2 of our combinations, problems arouse in further data analysis and the modelling portion, so these variables were eventually disregarded from the final model.

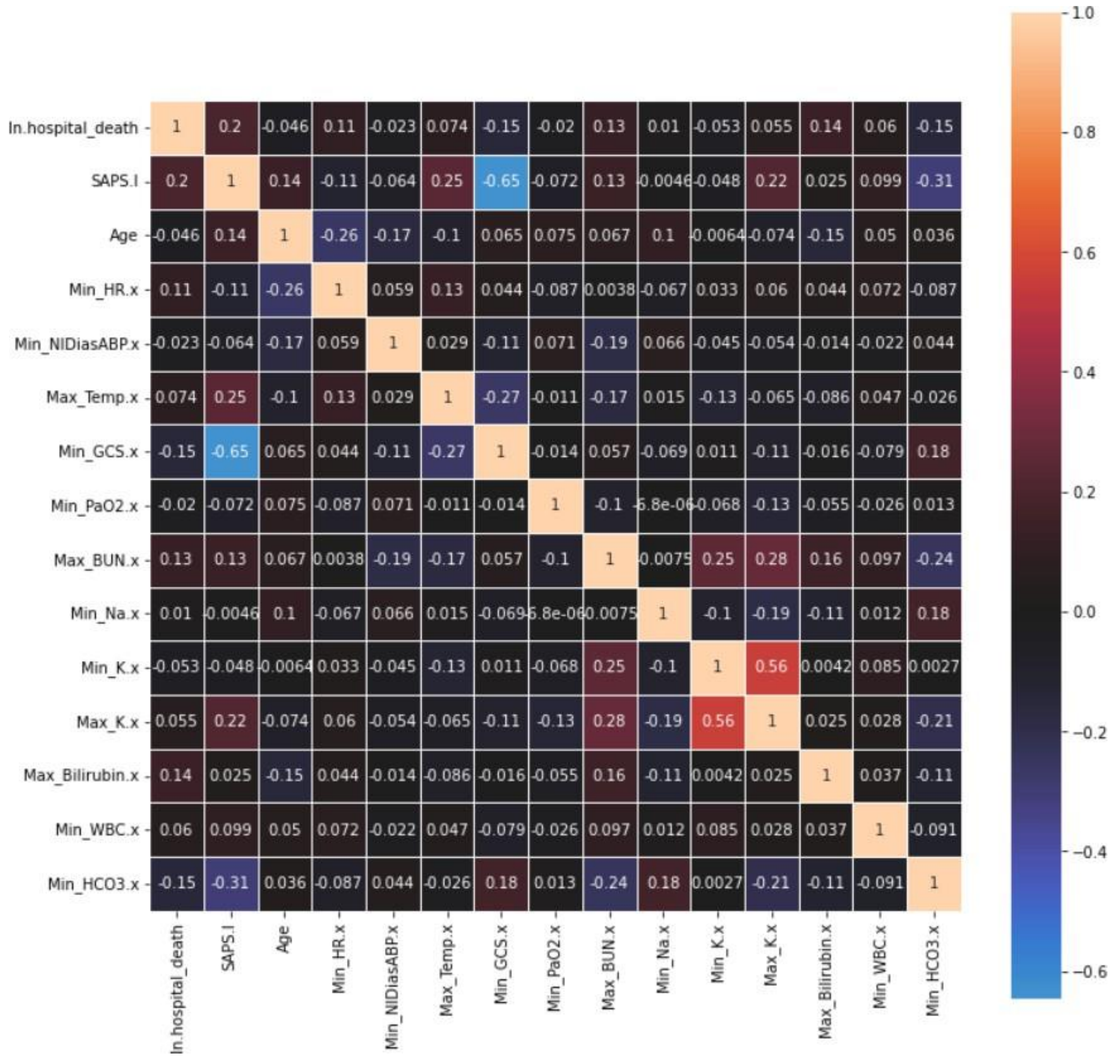


Figure 2.2: Heatmap of the selected variables and SAPS and in hospital death outcomes.

In addition to a heatmap and pairplot, histograms were created as a way to visualise the spread of the data across intervals. While we did not end up using them in our presentation, they served as a point of reference for how to model based on where the data aligned and a way to look for any possible biases present.

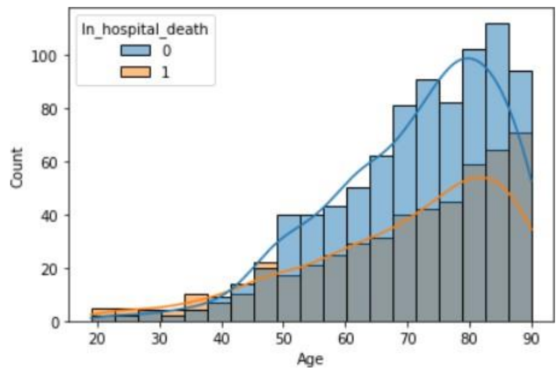


Figure 2.3: Histogram of the spread of patients' ages. The average age of the patients in this dataset was 70.74 years old. In addition, based on this histogram, the proportion of elderly patients (≥70 years old) greatly outweighed patients younger than 70 years old.

Exploratory Data Analysis:

To determine the odds of a patient's death based on the combination of factors, a logistic regression model was used to compare the four variables of Max_BUN.x, Min_HCO3.x, Min_K.x & Max_K.x.

Despite the Min_GCS.x and Max_Temp.x variables having a correlation greater than two of the combinations we used, the intercept value shown was at a very low value of 0.00922 when other variables are 0. When put in the context of the variables, flaws began to appear with this model. When the GCS is at 0, the patient is in a deep comatose state which increases the chance of death. In addition, when the maximum temperature is at 0, the patient should not be alive as the normal body temperature is 37 degrees celsius. For the intercept to be at 0.00922, the chances of the patient dying would be highly unlikely. However, once put in perspective of how the variables work, this would conclude that the model has its own flaws. Therefore, continuation of the usage of the model would be unwise. The intercepts and coefficients for the other variables were more significant and measurable, hence why Min_GCS.x and Max_Temp.x were not used in the final model.

These were the values given by the regression:

- Max BUN, Max K:
  - Intercept of 0.42
  - Coefficients - (1.00842, 1.02182)
- Max BUN, Min K:
  - Intercept of 1.16233
  - Coefficients - (1.01, 0.78)
- Max BUN, Min HCO3:
  - Intercept of 1.94854
  - Coefficients - (1.01, 0.94)

The predictions of the patient's death are given by the intercepts, where the levels of the given variables are at 0, and the figures represent percentages in decimal form whether the patient dies or not. For example, the patients who have the combination of Max BUN and Max K receive a probability of death of 0.42, where 0 is lowest probability and1 is highest probability. However, as there is an increase in the unit of Max BUN or Max K, the odds of the patient dying also increases.

This is shown by the coefficients given by the model, which can be interpreted as; for every one unit increase in the Max BUN, the odds of a patient dying increases by a factor of 1.01 (roughly), and for a one unit increase in the Max K, the odds of a patient dying roughly increases by a factor of 1.02. Those with intercepts higher than one, Max BUN, Min K/HCO3, this figure shows that the patient has a very likely chance of death when both values are at 0. Unlike the first combination with Max BUN and Max K, because the combinations of the other two have minimum values, the odds of a patient's death decreases for a one unit change in the variables. This also shows that it is crucial for a patient to have certain levels of potassium and bicarbonate in the body, otherwise it would lead to certain death.

The main conclusion which was made during the analysis was the chance of a patient's death given the combination of the selected factors. The more fatal combinations were determined and confirmed using this model, as well as the chance of death being dependent on the variation of factors. Other discoveries made were which factors are necessary for patients to live, and which were required in low amounts. Using the findings from this model, a model was then created to predict the outcome of a hypothetical patient when given the combinations. These findings from the logistic regression model were essential to the predictive model, as it determined and provided the relevant and necessary factors to work with.

Modelling:

To determine the odds of a patient's death based on the combination of factors, a logistic regression model was used, as a prediction of “true or false” was to be made. Furthermore, this model also calculated the increase or decrease in chance of a patient’s death depending on their levels in each respective factor. For example, the model would calculate the increase or decrease in chance of a patient’s death if their BUN levels increased. Using these scores, a prediction of a patient’s death if they had one of these combinations was determined. 80% of the data was used as training data for the model itself, and 20% was used for the testing data.

Logistic regression works with categorical data, which is data not continuous such as age, weight, or height. It is a statement whether something is one thing or not, or “true or false”. As in hospital death was used in the model (the patient dies or doesn’t), logistic regression must be used. In this model, either a patient survives or doesn't. Using the given data the model places all the data points into each respective category and then finds a relationship between the two factors (in hospital death and the combination of selected factors) and makes a prediction. This prediction is based on the curve the logistic model uses, much like the fitted line in linear regression.

	0	1	Death	true
0	0.653106	0.346894	0	0
1	0.714336	0.285664	0	0
2	0.750892	0.249108	0	0
3	0.539376	0.460624	0	1
4	0.746877	0.253123	0	0
...	...	...	...	...
1087	0.608625	0.391375	0	0
1088	0.662850	0.337150	0	1
1089	0.642483	0.357517	0	0
1090	0.661501	0.338499	0	0
1091	0.670888	0.329112	0	0

Figure 3.1: Example predictive model, showing the relationship between Maximum BUN and Maximum Potassium

For the predictive models, their accuracies were determined through training and testing scores, where 80% of the data was allocated to training data, and the rest was allocated to the testing data. In all instances of the combination pairs, both training and testing scores were above 60%, which is high enough for the predictive models to not be labelled as a coin toss. The training scores are expected to be slightly higher than the testing scores, as there is a larger sample size allocated to the training data. This causes a bias towards the training data, as a larger sample size allows for a more defined trend compared to the testing data, where data points may be quite sparse. However, for the combination of Max BUN and Min K, it is apparent that the testing score is slightly higher than the training score.

The testing score should not be higher than the training score, as the model is optimised for the training data. Some possible reasons as to why the testing score was higher in that instance are:

- The test/train data split might not have been suitable
- The data used varied a lot, not all outliers were removed as they were medically possible, and it might just have been that the split happened to favour the testing data side.

Dataset Limitations/Improvements:

While the dataset contained a range of information that we could not utilise due to time constraints, there were limitations and improvements that could have been made to the values. One such limitation would be the entire dataset being outdated, especially when it comes to the SAPS I scores. The most up to date SAPS model is SAPS III, which has a plethora of information and data to explain and help understand what each variable does and how it contributes to the score. It has only recently been adapted into the medical scene, prevailing over its long-time predecessor, SAPS-II. SAPS-I, however, has not been in use for almost 20 years, which is the value our dataset gave us. Almost no relevant information was to be found in terms of how to interpret the SAPS-I score or what variables can be removed/substituted.

This meant that the SAPS-I data had to be moulded to fit into what should have been SAPS-II data, leading to missing or incomplete variables, which would have provided an incorrect SAPS score. If the dataset at least had the variables that were relevant to SAPS-II calculations or even provided the SAPS-II scores itself, there would have been more validity in our model, leaving less room for errors.

Further improvements to the data could have been to include a section about previous or current ailments/diseases. This would allow for an acute understanding of their stay in the different ICUs as well as allow for monitoring of the variables related to previous conditions, which might influence their death. This means that all readings are consistent and are taken at certain time intervals, not having irregular values taken whenever, especially if the measurement is only taken once for the patient's overall stay.

Conclusion:

In conclusion, the predictive model to determine the patient's outcome based on their SAPS score consisted of 4 main variables, all derived from the 17 that are used in SAPS calculations. The correlation between Max_BUN.x, Min_HCO3.x, Min_K.x & Max_K.x allowed for us to combine and compare the intercepts and coefficients, which allowed for the building of the model. Through the training and testing scores, we found that the final predicted outcomes were only about 15% better than a coin toss, which isn’t considered super reliable nor accurate due to missing variables, but serves as a base for allowing the integration of more variables into the model, allowing for more validity and accuracy.

In terms of the testing and training scores themselves, there are many possible factors as to why the testing scores were higher, including issues with outliers. For future research and expansion upon this project, reassessing the data and eliminating more outliers instead of only the extreme outliers may improve the models’ validity and accuracy.

Further analysis of the dataset without technical or time limitations would allow for further analysis of links between all variables and especially their impact on the outcome descriptors. While this model predicted based on the given SAPS scores, advanced and more intuitive models could predict with a range of factors and use SAPS as a variable, not a final solution.

References:

- Healthy WA. Intensive care units (ICUs). Healthy WA, from https://www.healthywa.wa.gov.au/Articles/F_I/Intensive-care-units-ICUs.
- Le Gall, J. *Simplified Acute Physiology Score (SAPS) II*. MDCalc, from https://www.mdcalc.com/simplified-acute-physiology-score-saps-ii.
- Le Gall, J. (1993). A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA: The Journal Of The American Medical Association*, *270*(24), 2957-2963. https://doi.org/10.1001/jama.270.24.2957
- Leader, D. (2022). *Understanding the Partial Pressure of Oxygen (PaO2) Test*. Verywell Health, from https://www.verywellhealth.com/partial-pressure-of-oyxgen-pa02-914920#:~:text=The%20partial%20pressure%20of%20oxygen%2C%20also%20known%20as%20PaO2%2C%20is,oxygen%20pressure%20in%20arterial%20blood.
- Mayr, V., Dünser, M., Greil, V., Jochberger, S., Luckner, G., & Ulmer, H. et al. (2006). Causes of death and determinants of outcome in critically ill patients. *Critical Care*, *10*(6), R154. https://doi.org/10.1186/cc5086
- Mungan, İ., Bektaş, Ş., Altınkaya Çavuş, M., Sarı, S., & Turan, S. (2019). The predictive power of SAPS-3 and SOFA scores and their relations with patient outcomes in the Surgical Intensive Care Unit. *Turkish Journal Of Surgery*, *35*(2), 124-130. https://doi.org/10.5578/turkjsurg.4223
- National Cancer Institute. *NCI Dictionary of Cancer Terms*. National Cancer Institute, from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/blood-urea-nitrogen.
- NHS Foundation Trust. *Coronary Care Unit (CCU) and Cardiology Ward*. Hampshirehospitals.nhs.uk, from https://www.hampshirehospitals.nhs.uk/our-services/az-departments-and-specialties/coronary-care-unit-ccu-and-cardiology-ward.
- Rossiaky, D. (2022). *High Bilirubin Levels: Symptoms, Causes, and Treatment*. Healthline, from https://www.healthline.com/health/high-bilirubin.
- Silva, I., Moody, G., Mark, R., & Celi, L. (2012). *Predicting Mortality of ICU Patients: The PhysioNet/Computing in Cardiology Challenge 2012*. PhysioNet, from https://physionet.org/content/challenge-2012/1.0.0/.
- UC San Diego. (2022). *Critical Care*. UC San Diego Health, from https://health.ucsd.edu/specialties/critical-care/Pages/default.aspx.