Patient Survival Prediction

Jalaja Pothedar

2022-06-26

## Abstract

For patients who enter at the hospital as an emergency, Intensive Care Units (ICUs) usually lack solid medica (Raffa, et al., 2019) (Chi, et al., 2019)l histories. A troubled patient or one who is brought to emergency treatment in a disoriented or comatose state may be unable to provide any medical history information to the clinician.Due to the lack of medical records, the results of rapid laboratory tests performed on the patient upon arrival are the only means to determine the patient’s health and, more significantly, survival. This research project uncovers the characteristics that are important in predicting an ICU patient’s survival. These results may be useful for the medical practitioners to prioritize in obtaining the high-importance values first to improve the patient’s chances of survival.

The dataset contains clinical outcome information as well as patient survival rates. Predictive analysis was performed on the dataset with ‘hospital death’ (Whether the patient died during this hospitalization) as the variable of interest, and a prediction model with 91.5% efficiency(approximately) has been created using random forest algorithm. This study also identified key elements that aid in predicting the patient’s survival.

## Introduction

Intensive Care Units (ICUs) frequently lack reliable medical histories for patients who arrive at the hospital as an emergency. A distressed patient or one who is brought in a disoriented or unconscious state may be unable to communicate any medical history information to the doctor at emergency care. Transferring medical records may take many days, especially if the patient is coming from another medical provider or system. However, knowing the patient’s medical history can help doctors make better clinical decisions about the patient’s therapy. Because medical records are unavailable, the findings of fast laboratory tests done on the patient upon arrival are the only way to identify the patient’s health and, more importantly, survival.This research project uncovers the characteristics that are important in predicting an ICU patient’s survival. As a result, medical practitioners might try to prioritize in obtaining the high-importance values first to improve the patient’s chances of survival.

The dataset contains physical characteristics of the patient, laboratory test results, APACHE (‘Acute physiology and chronic health evaluation’) (Draper, 1984) values as well as information if the patient passed away during the hospitalization (0-No,1-Yes). The dataset contains information of 91,713 patients.

## Literature Review

Traces of 2 similar research works (Raffa, et al., 2019), (Chi, et al., 2019) of ICU patients are currently available on the internet. Although the dataset used and the complete research work is not available for free, a summary of the two studies is available on open source. Both the research papers mainly focused on the survival rate and disease severity and have focused on deducing conclusions for ICU patients battling with terminal diseases. Unlike the study that cited papers have put forward, the main objective of this research project is focused on emergency cases battling in the ICU. The dataset contains a column ‘ICU type’. The number of cases for each of the ICU type is depicted below as a pie chart. One of the studies (Raffa, et al., 2019) focused on calculating the severity of illness score at country and region level. For this study data was taken from the Australia New Zealand Intensive Care Society (ANZIS). During this study the missing data was filled using prediction models and logistic regression was used for analysis. The second study (Chi, et al., 2019) focused on palliative care (PC) for ICU patients. Palliative care is specialist medical care for those who are suffering from a terminal disease. A tool was created using predictive analysis to decide the unit where the patient needs to be admitted based on the survival value predicted by the model. The impact of PC consultation on outcomes was estimated using multivariate logistic regression analysis.

## Research Questions

The following points will be addressed in this research project: 1. Determine an efficient algorithm to predict the survival of the patient. 2. What are the most important elements in predicting the survival of a patient admitted to ICU?

## Theory

This paper is exploring best prediction algorithm, i.e., a high performing model to predict the variable of interest ‘hospital\_death’. The exploration includes finding variables that play key role in determeing the hospital\_death. This study also involves predetermined APACHE (“Acute physiology and chronic health evaluation”) values calculated by medical apparatus. No medical/clinical calculations are involved in this paper.

## Data

The data for this study is derived from open source: <https://www.kaggle.com/datasets/mitishaagarwal/patient/download> Variables available in the data set are mentioned below:

#loading the .csv file and creating a data frame.  
dataset <- read.csv("/Users/akshaymusuku/Downloads/R project\_JP/Dataset.csv")  
dictionary <- read.csv("/Users/akshaymusuku/Downloads/R project\_JP/Data Dictionary.csv")

str(dataset)

## 'data.frame': 91713 obs. of 186 variables:  
## $ encounter\_id : int 66154 114252 119783 79267 92056 33181 82208 120995 80471 42871 ...  
## $ patient\_id : int 25312 59342 50777 46918 34377 74489 49526 50129 10577 90749 ...  
## $ hospital\_id : int 118 81 118 118 33 83 83 33 118 118 ...  
## $ hospital\_death : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ age : int 68 77 25 81 19 67 59 70 45 50 ...  
## $ bmi : num 22.7 27.4 31.9 22.6 NA ...  
## $ elective\_surgery : int 0 0 0 1 0 0 0 0 0 0 ...  
## $ ethnicity : chr "Caucasian" "Caucasian" "Caucasian" "Caucasian" ...  
## $ gender : chr "M" "F" "F" "F" ...  
## $ height : num 180 160 173 165 188 ...  
## $ hospital\_admit\_source : chr "Floor" "Floor" "Emergency Department" "Operating Room" ...  
## $ icu\_admit\_source : chr "Floor" "Floor" "Accident & Emergency" "Operating Room / Recovery" ...  
## $ icu\_id : int 92 90 93 92 91 95 95 91 114 114 ...  
## $ icu\_stay\_type : chr "admit" "admit" "admit" "admit" ...  
## $ icu\_type : chr "CTICU" "Med-Surg ICU" "Med-Surg ICU" "CTICU" ...  
## $ pre\_icu\_los\_days : num 0.541667 0.927778 0.000694 0.000694 0.073611 ...  
## $ readmission\_status : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ weight : num 73.9 70.2 95.3 61.7 NA ...  
## $ albumin\_apache : num 2.3 NA NA NA NA NA NA NA 2.7 3.6 ...  
## $ apache\_2\_diagnosis : int 113 108 122 203 119 301 108 113 116 112 ...  
## $ apache\_3j\_diagnosis : num 502 203 703 1206 601 ...  
## $ apache\_post\_operative : int 0 0 0 1 0 0 0 0 0 0 ...  
## $ arf\_apache : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ bilirubin\_apache : num 0.4 NA NA NA NA NA NA NA 0.2 0.4 ...  
## $ bun\_apache : num 31 9 NA NA NA 13 18 48 15 10 ...  
## $ creatinine\_apache : num 2.51 0.56 NA NA NA 0.71 0.78 2.05 1.16 0.83 ...  
## $ fio2\_apache : num NA 1 NA 0.6 NA NA 1 NA 1 NA ...  
## $ gcs\_eyes\_apache : int 3 1 3 4 NA 4 4 4 4 4 ...  
## $ gcs\_motor\_apache : int 6 3 6 6 NA 6 6 6 6 6 ...  
## $ gcs\_unable\_apache : int 0 0 0 0 NA 0 0 0 0 0 ...  
## $ gcs\_verbal\_apache : int 4 1 5 5 NA 5 5 5 5 5 ...  
## $ glucose\_apache : num 168 145 NA 185 NA 156 197 164 380 134 ...  
## $ heart\_rate\_apache : int 118 120 102 114 60 113 133 120 82 94 ...  
## $ hematocrit\_apache : num 27.4 36.9 NA 25.9 NA 44.2 33.5 22.6 37.9 37.2 ...  
## $ intubated\_apache : int 0 0 0 1 0 0 1 0 0 0 ...  
## $ map\_apache : int 40 46 68 60 103 130 138 60 66 58 ...  
## $ paco2\_apache : num NA 37 NA 30 NA NA 43 NA 60 NA ...  
## $ paco2\_for\_ph\_apache : num NA 37 NA 30 NA NA 43 NA 60 NA ...  
## $ pao2\_apache : num NA 51 NA 142 NA NA 370 NA 92 NA ...  
## $ ph\_apache : num NA 7.45 NA 7.39 NA NA 7.42 NA 7.14 NA ...  
## $ resprate\_apache : num 36 33 37 4 16 35 53 28 14 46 ...  
## $ sodium\_apache : num 134 145 NA NA NA 137 135 140 142 139 ...  
## $ temp\_apache : num 39.3 35.1 36.7 34.8 36.7 36.6 35 36.6 36.9 36.3 ...  
## $ urineoutput\_apache : num NA NA NA NA NA NA NA NA NA NA ...  
## $ ventilated\_apache : int 0 1 0 1 0 0 1 1 1 0 ...  
## $ wbc\_apache : num 14.1 12.7 NA 8 NA 10.9 5.9 12.8 24.7 8.4 ...  
## $ d1\_diasbp\_invasive\_max : int 46 NA NA 62 NA NA 107 NA 64 74 ...  
## $ d1\_diasbp\_invasive\_min : int 32 NA NA 30 NA NA 65 NA 52 57 ...  
## $ d1\_diasbp\_max : int 68 95 88 48 99 100 76 84 65 83 ...  
## $ d1\_diasbp\_min : int 37 31 48 42 57 61 68 46 59 48 ...  
## $ d1\_diasbp\_noninvasive\_max : int 68 95 88 48 99 100 76 84 65 83 ...  
## $ d1\_diasbp\_noninvasive\_min : int 37 31 48 42 57 61 68 46 59 48 ...  
## $ d1\_heartrate\_max : int 119 118 96 116 89 113 112 118 82 96 ...  
## $ d1\_heartrate\_min : int 72 72 68 92 60 83 70 86 82 57 ...  
## $ d1\_mbp\_invasive\_max : int 66 NA NA 92 NA NA 138 NA 72 92 ...  
## $ d1\_mbp\_invasive\_min : int 40 NA NA 52 NA NA 84 NA 66 73 ...  
## $ d1\_mbp\_max : int 89 120 102 84 104 127 117 114 93 101 ...  
## $ d1\_mbp\_min : int 46 38 68 84 90 80 97 60 71 59 ...  
## $ d1\_mbp\_noninvasive\_max : int 89 120 102 84 104 127 117 114 93 101 ...  
## $ d1\_mbp\_noninvasive\_min : int 46 38 68 84 90 80 97 60 71 59 ...  
## $ d1\_resprate\_max : int 34 32 21 23 18 32 38 28 24 44 ...  
## $ d1\_resprate\_min : int 10 12 8 7 16 10 16 12 19 14 ...  
## $ d1\_spo2\_max : int 100 100 98 100 100 97 100 100 97 100 ...  
## $ d1\_spo2\_min : int 74 70 91 95 96 91 87 92 97 96 ...  
## $ d1\_sysbp\_invasive\_max : int 122 NA NA 164 NA NA 191 NA 94 126 ...  
## $ d1\_sysbp\_invasive\_min : int 64 NA NA 78 NA NA 116 NA 72 103 ...  
## $ d1\_sysbp\_max : int 131 159 148 158 147 173 151 147 104 135 ...  
## $ d1\_sysbp\_min : int 73 67 105 84 120 107 133 71 98 78 ...  
## $ d1\_sysbp\_noninvasive\_max : int 131 159 148 158 147 173 151 147 104 135 ...  
## $ d1\_sysbp\_noninvasive\_min : num 73 67 105 84 120 107 133 71 98 78 ...  
## $ d1\_temp\_max : num 39.9 36.3 37 38 37.2 36.8 37.2 38.5 36.9 37.1 ...  
## $ d1\_temp\_min : num 37.2 35.1 36.7 34.8 36.7 36.6 35 36.6 36.9 36.4 ...  
## $ h1\_diasbp\_invasive\_max : int NA NA NA 62 NA NA 107 NA 64 73 ...  
## $ h1\_diasbp\_invasive\_min : int NA NA NA 44 NA NA 79 NA 52 62 ...  
## $ h1\_diasbp\_max : int 68 61 88 62 99 89 107 74 65 83 ...  
## $ h1\_diasbp\_min : int 63 48 58 44 68 89 79 55 59 61 ...  
## $ h1\_diasbp\_noninvasive\_max : int 68 61 88 NA 99 89 NA 74 65 83 ...  
## $ h1\_diasbp\_noninvasive\_min : int 63 48 58 NA 68 89 NA 55 59 61 ...  
## $ h1\_heartrate\_max : int 119 114 96 100 89 83 79 118 82 96 ...  
## $ h1\_heartrate\_min : int 108 100 78 96 76 83 72 114 82 60 ...  
## $ h1\_mbp\_invasive\_max : num NA NA NA 92 NA NA 138 NA 72 92 ...  
## $ h1\_mbp\_invasive\_min : int NA NA NA 71 NA NA 115 NA 66 78 ...  
## $ h1\_mbp\_max : int 86 85 91 92 104 111 117 88 93 101 ...  
## $ h1\_mbp\_min : int 85 57 83 71 92 111 117 60 71 77 ...  
## $ h1\_mbp\_noninvasive\_max : int 86 85 91 NA 104 111 117 88 93 101 ...  
## $ h1\_mbp\_noninvasive\_min : int 85 57 83 NA 92 111 117 60 71 77 ...  
## $ h1\_resprate\_max : int 26 31 20 12 NA 12 18 28 24 29 ...  
## $ h1\_resprate\_min : int 18 28 16 11 NA 12 18 26 19 17 ...  
## $ h1\_spo2\_max : int 100 95 98 100 100 97 100 96 97 100 ...  
## $ h1\_spo2\_min : int 74 70 91 99 100 97 100 92 97 96 ...  
## $ h1\_sysbp\_invasive\_max : int NA NA NA 136 NA NA 191 NA 94 126 ...  
## $ h1\_sysbp\_invasive\_min : num NA NA NA 106 NA NA 163 NA 72 106 ...  
## $ h1\_sysbp\_max : int 131 95 148 136 130 143 191 119 104 135 ...  
## $ h1\_sysbp\_min : int 115 71 124 106 120 143 163 106 98 103 ...  
## $ h1\_sysbp\_noninvasive\_max : int 131 95 148 NA 130 143 NA 119 104 135 ...  
## $ h1\_sysbp\_noninvasive\_min : int 115 71 124 NA 120 143 NA 106 98 103 ...  
## $ h1\_temp\_max : num 39.5 36.3 36.7 35.6 NA 36.7 36.8 38.5 36.9 36.9 ...  
## $ h1\_temp\_min : num 37.5 36.3 36.7 34.8 NA 36.7 35 38.5 36.9 36.9 ...  
## $ d1\_albumin\_max : num 2.3 1.6 NA NA NA NA NA NA 2.7 3.6 ...  
## [list output truncated]

Data dictionary of the dataset is as below:

#displaying few columns of the data dictioanry to refer the description of variable names  
print(dictionary[, c(2, 5)])

## Variable.Name  
## 1 encounter\_id  
## 2 hospital\_id  
## 3 patient\_id  
## 4 hospital\_death  
## 5 age  
## 6 bmi  
## 7 elective\_surgery  
## 8 ethnicity  
## 9 gender  
## 10 height  
## 11 hospital\_admit\_source  
## 12 icu\_admit\_source  
## 13 icu\_admit\_type  
## 14 icu\_id  
## 15 icu\_stay\_type  
## 16 icu\_type  
## 17 pre\_icu\_los\_days  
## 18 readmission\_status  
## 19 weight  
## 20 albumin\_apache  
## 21 apache\_2\_diagnosis  
## 22 apache\_3j\_diagnosis  
## 23 apache\_post\_operative  
## 24 arf\_apache  
## 25 bilirubin\_apache  
## 26 bun\_apache  
## 27 creatinine\_apache  
## 28 fio2\_apache  
## 29 gcs\_eyes\_apache  
## 30 gcs\_motor\_apache  
## 31 gcs\_unable\_apache  
## 32 gcs\_verbal\_apache  
## 33 glucose\_apache  
## 34 heart\_rate\_apache  
## 35 hematocrit\_apache  
## 36 intubated\_apache  
## 37 map\_apache  
## 38 paco2\_apache  
## 39 paco2\_for\_ph\_apache  
## 40 pao2\_apache  
## 41 ph\_apache  
## 42 resprate\_apache  
## 43 sodium\_apache  
## 44 temp\_apache  
## 45 urineoutput\_apache  
## 46 ventilated\_apache  
## 47 wbc\_apache  
## 48 d1\_diasbp\_invasive\_max  
## 49 d1\_diasbp\_invasive\_min  
## 50 d1\_diasbp\_max  
## 51 d1\_diasbp\_min  
## 52 d1\_diasbp\_noninvasive\_max  
## 53 d1\_diasbp\_noninvasive\_min  
## 54 d1\_heartrate\_max  
## 55 d1\_heartrate\_min  
## 56 d1\_mbp\_invasive\_max  
## 57 d1\_mbp\_invasive\_min  
## 58 d1\_mbp\_max  
## 59 d1\_mbp\_min  
## 60 d1\_mbp\_noninvasive\_max  
## 61 d1\_mbp\_noninvasive\_min  
## 62 d1\_resprate\_max  
## 63 d1\_resprate\_min  
## 64 d1\_spo2\_max  
## 65 d1\_spo2\_min  
## 66 d1\_sysbp\_invasive\_max  
## 67 d1\_sysbp\_invasive\_min  
## 68 d1\_sysbp\_max  
## 69 d1\_sysbp\_min  
## 70 d1\_sysbp\_noninvasive\_max  
## 71 d1\_sysbp\_noninvasive\_min  
## 72 d1\_temp\_max  
## 73 d1\_temp\_min  
## 74 h1\_diasbp\_invasive\_max  
## 75 h1\_diasbp\_invasive\_min  
## 76 h1\_diasbp\_max  
## 77 h1\_diasbp\_min  
## 78 h1\_diasbp\_noninvasive\_max  
## 79 h1\_diasbp\_noninvasive\_min  
## 80 h1\_heartrate\_max  
## 81 h1\_heartrate\_min  
## 82 h1\_mbp\_invasive\_max  
## 83 h1\_mbp\_invasive\_min  
## 84 h1\_mbp\_max  
## 85 h1\_mbp\_min  
## 86 h1\_mbp\_noninvasive\_max  
## 87 h1\_mbp\_noninvasive\_min  
## 88 h1\_resprate\_max  
## 89 h1\_resprate\_min  
## 90 h1\_spo2\_max  
## 91 h1\_spo2\_min  
## 92 h1\_sysbp\_invasive\_max  
## 93 h1\_sysbp\_invasive\_min  
## 94 h1\_sysbp\_max  
## 95 h1\_sysbp\_min  
## 96 h1\_sysbp\_noninvasive\_max  
## 97 h1\_sysbp\_noninvasive\_min  
## 98 h1\_temp\_max  
## 99 h1\_temp\_min  
## 100 d1\_albumin\_max  
## 101 d1\_albumin\_min  
## 102 d1\_bilirubin\_max  
## 103 d1\_bilirubin\_min  
## 104 d1\_bun\_max  
## 105 d1\_bun\_min  
## 106 d1\_calcium\_max  
## 107 d1\_calcium\_min  
## 108 d1\_creatinine\_max  
## 109 d1\_creatinine\_min  
## 110 d1\_glucose\_max  
## 111 d1\_glucose\_min  
## 112 d1\_hco3\_max  
## 113 d1\_hco3\_min  
## 114 d1\_hemaglobin\_max  
## 115 d1\_hemaglobin\_min  
## 116 d1\_hematocrit\_max  
## 117 d1\_hematocrit\_min  
## 118 d1\_inr\_max  
## 119 d1\_inr\_min  
## 120 d1\_lactate\_max  
## 121 d1\_lactate\_min  
## 122 d1\_platelets\_max  
## 123 d1\_platelets\_min  
## 124 d1\_potassium\_max  
## 125 d1\_potassium\_min  
## 126 d1\_sodium\_max  
## 127 d1\_sodium\_min  
## 128 d1\_wbc\_max  
## 129 d1\_wbc\_min  
## 130 h1\_albumin\_max  
## 131 h1\_albumin\_min  
## 132 h1\_bilirubin\_max  
## 133 h1\_bilirubin\_min  
## 134 h1\_bun\_max  
## 135 h1\_bun\_min  
## 136 h1\_calcium\_max  
## 137 h1\_calcium\_min  
## 138 h1\_creatinine\_max  
## 139 h1\_creatinine\_min  
## 140 h1\_glucose\_max  
## 141 h1\_glucose\_min  
## 142 h1\_hco3\_max  
## 143 h1\_hco3\_min  
## 144 h1\_hemaglobin\_max  
## 145 h1\_hemaglobin\_min  
## 146 h1\_hematocrit\_max  
## 147 h1\_hematocrit\_min  
## 148 h1\_inr\_max  
## 149 h1\_inr\_min  
## 150 h1\_lactate\_max  
## 151 h1\_lactate\_min  
## 152 h1\_platelets\_max  
## 153 h1\_platelets\_min  
## 154 h1\_potassium\_max  
## 155 h1\_potassium\_min  
## 156 h1\_sodium\_max  
## 157 h1\_sodium\_min  
## 158 h1\_wbc\_max  
## 159 h1\_wbc\_min  
## 160 d1\_arterial\_pco2\_max  
## 161 d1\_arterial\_pco2\_min  
## 162 d1\_arterial\_ph\_max  
## 163 d1\_arterial\_ph\_min  
## 164 d1\_arterial\_po2\_max  
## 165 d1\_arterial\_po2\_min  
## 166 d1\_pao2fio2ratio\_max  
## 167 d1\_pao2fio2ratio\_min  
## 168 h1\_arterial\_pco2\_max  
## 169 h1\_arterial\_pco2\_min  
## 170 h1\_arterial\_ph\_max  
## 171 h1\_arterial\_ph\_min  
## 172 h1\_arterial\_po2\_max  
## 173 h1\_arterial\_po2\_min  
## 174 h1\_pao2fio2ratio\_max  
## 175 h1\_pao2fio2ratio\_min  
## 176 apache\_4a\_hospital\_death\_prob  
## 177 apache\_4a\_icu\_death\_prob  
## 178 aids  
## 179 cirrhosis  
## 180 diabetes\_mellitus  
## 181 hepatic\_failure  
## 182 immunosuppression  
## 183 leukemia  
## 184 lymphoma  
## 185 solid\_tumor\_with\_metastasis  
## 186 apache\_3j\_bodysystem  
## 187 apache\_2\_bodysystem  
## 188 pred  
## Description  
## 1 Unique identifier associated with a patient unit stay  
## 2 Unique identifier associated with a hospital  
## 3 Unique identifier associated with a patient  
## 4 Whether the patient died during this hospitalization  
## 5 The age of the patient on unit admission  
## 6 The body mass index of the person on unit admission  
## 7 Whether the patient was admitted to the hospital for an elective surgical operation  
## 8 The common national or cultural tradition which the person belongs to  
## 9 The genotypical sex of the patient  
## 10 The height of the person on unit admission  
## 11 The location of the patient prior to being admitted to the hospital  
## 12 The location of the patient prior to being admitted to the unit  
## 13 The type of unit admission for the patient  
## 14 A unique identifier for the unit to which the patient was admitted  
## 15   
## 16 A classification which indicates the type of care the unit is capable of providing  
## 17 The length of stay of the patient between hospital admission and unit admission  
## 18 Whether the current unit stay is the second (or greater) stay at an ICU within the same hospitalization  
## 19 The weight (body mass) of the person on unit admission  
## 20 The albumin concentration measured during the first 24 hours which results in the highest APACHE III score  
## 21 The APACHE II diagnosis for the ICU admission  
## 22 The APACHE III-J sub-diagnosis code which best describes the reason for the ICU admission  
## 23 The APACHE operative status; 1 for post-operative, 0 for non-operative  
## 24 Whether the patient had acute renal failure during the first 24 hours of their unit stay, defined as a 24 hour urine output <410ml, creatinine >=133 micromol/L and no chronic dialysis  
## 25 The bilirubin concentration measured during the first 24 hours which results in the highest APACHE III score  
## 26 The blood urea nitrogen concentration measured during the first 24 hours which results in the highest APACHE III score  
## 27 The creatinine concentration measured during the first 24 hours which results in the highest APACHE III score  
## 28 The fraction of inspired oxygen from the arterial blood gas taken during the first 24 hours of unit admission which produces the highest APACHE III score for oxygenation  
## 29 The eye opening component of the Glasgow Coma Scale measured during the first 24 hours which results in the highest APACHE III score  
## 30 The motor component of the Glasgow Coma Scale measured during the first 24 hours which results in the highest APACHE III score  
## 31 Whether the Glasgow Coma Scale was unable to be assessed due to patient sedation  
## 32 The verbal component of the Glasgow Coma Scale measured during the first 24 hours which results in the highest APACHE III score  
## 33 The glucose concentration measured during the first 24 hours which results in the highest APACHE III score  
## 34 The heart rate measured during the first 24 hours which results in the highest APACHE III score  
## 35 The hematocrit measured during the first 24 hours which results in the highest APACHE III score  
## 36 Whether the patient was intubated at the time of the highest scoring arterial blood gas used in the oxygenation score  
## 37 The mean arterial pressure measured during the first 24 hours which results in the highest APACHE III score  
## 38 The partial pressure of carbon dioxide from the arterial blood gas taken during the first 24 hours of unit admission which produces the highest APACHE III score for oxygenation  
## 39 The partial pressure of carbon dioxide from the arterial blood gas taken during the first 24 hours of unit admission which produces the highest APACHE III score for acid-base disturbance  
## 40 The partial pressure of oxygen from the arterial blood gas taken during the first 24 hours of unit admission which produces the highest APACHE III score for oxygenation  
## 41 The pH from the arterial blood gas taken during the first 24 hours of unit admission which produces the highest APACHE III score for acid-base disturbance  
## 42 The respiratory rate measured during the first 24 hours which results in the highest APACHE III score  
## 43 The sodium concentration measured during the first 24 hours which results in the highest APACHE III score  
## 44 The temperature measured during the first 24 hours which results in the highest APACHE III score  
## 45 The total urine output for the first 24 hours  
## 46 Whether the patient was invasively ventilated at the time of the highest scoring arterial blood gas using the oxygenation scoring algorithm, including any mode of positive pressure ventilation delivered through a circuit attached to an endo-tracheal tube or tracheostomy  
## 47 The white blood cell count measured during the first 24 hours which results in the highest APACHE III score  
## 48 The patient's highest diastolic blood pressure during the first 24 hours of their unit stay, invasively measured  
## 49 The patient's lowest diastolic blood pressure during the first 24 hours of their unit stay, invasively measured  
## 50 The patient's highest diastolic blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 51 The patient's lowest diastolic blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 52 The patient's highest diastolic blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 53 The patient's lowest diastolic blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 54 The patient's highest heart rate during the first 24 hours of their unit stay  
## 55 The patient's lowest heart rate during the first 24 hours of their unit stay  
## 56 The patient's highest mean blood pressure during the first 24 hours of their unit stay, invasively measured  
## 57 The patient's lowest mean blood pressure during the first 24 hours of their unit stay, invasively measured  
## 58 The patient's highest mean blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 59 The patient's lowest mean blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 60 The patient's highest mean blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 61 The patient's lowest mean blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 62 The patient's highest respiratory rate during the first 24 hours of their unit stay  
## 63 The patient's lowest respiratory rate during the first 24 hours of their unit stay  
## 64 The patient's highest peripheral oxygen saturation during the first 24 hours of their unit stay  
## 65 The patient's lowest peripheral oxygen saturation during the first 24 hours of their unit stay  
## 66 The patient's highest systolic blood pressure during the first 24 hours of their unit stay, invasively measured  
## 67 The patient's lowest systolic blood pressure during the first 24 hours of their unit stay, invasively measured  
## 68 The patient's highest systolic blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 69 The patient's lowest systolic blood pressure during the first 24 hours of their unit stay, either non-invasively or invasively measured  
## 70 The patient's highest systolic blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 71 The patient's lowest systolic blood pressure during the first 24 hours of their unit stay, non-invasively measured  
## 72 The patient's highest core temperature during the first 24 hours of their unit stay, invasively measured  
## 73 The patient's lowest core temperature during the first 24 hours of their unit stay  
## 74 The patient's highest diastolic blood pressure during the first hour of their unit stay, invasively measured  
## 75 The patient's lowest diastolic blood pressure during the first hour of their unit stay, invasively measured  
## 76 The patient's highest diastolic blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 77 The patient's lowest diastolic blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 78 The patient's highest diastolic blood pressure during the first hour of their unit stay, non-invasively measured  
## 79 The patient's lowest diastolic blood pressure during the first hour of their unit stay, non-invasively measured  
## 80 The patient's highest heart rate during the first hour of their unit stay  
## 81 The patient's lowest heart rate during the first hour of their unit stay  
## 82 The patient's highest mean blood pressure during the first hour of their unit stay, invasively measured  
## 83 The patient's lowest mean blood pressure during the first hour of their unit stay, invasively measured  
## 84 The patient's highest mean blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 85 The patient's lowest mean blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 86 The patient's highest mean blood pressure during the first hour of their unit stay, non-invasively measured  
## 87 The patient's lowest mean blood pressure during the first hour of their unit stay, non-invasively measured  
## 88 The patient's highest respiratory rate during the first hour of their unit stay  
## 89 The patient's lowest respiratory rate during the first hour of their unit stay  
## 90 The patient's highest peripheral oxygen saturation during the first hour of their unit stay  
## 91 The patient's lowest peripheral oxygen saturation during the first hour of their unit stay  
## 92 The patient's highest systolic blood pressure during the first hour of their unit stay, invasively measured  
## 93 The patient's lowest systolic blood pressure during the first hour of their unit stay, invasively measured  
## 94 The patient's highest systolic blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 95 The patient's lowest systolic blood pressure during the first hour of their unit stay, either non-invasively or invasively measured  
## 96 The patient's highest systolic blood pressure during the first hour of their unit stay, non-invasively measured  
## 97 The patient's lowest systolic blood pressure during the first hour of their unit stay, non-invasively measured  
## 98 The patient's highest core temperature during the first hour of their unit stay, invasively measured  
## 99 The patient's lowest core temperature during the first hour of their unit stay  
## 100 The lowest albumin concentration of the patient in their serum during the first 24 hours of their unit stay  
## 101 The lowest albumin concentration of the patient in their serum during the first 24 hours of their unit stay  
## 102 The highest bilirubin concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 103 The lowest bilirubin concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 104 The highest blood urea nitrogen concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 105 The lowest blood urea nitrogen concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 106 The highest calcium concentration of the patient in their serum during the first 24 hours of their unit stay  
## 107 The lowest calcium concentration of the patient in their serum during the first 24 hours of their unit stay  
## 108 The highest creatinine concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 109 The lowest creatinine concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 110 The highest glucose concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 111 The lowest glucose concentration of the patient in their serum or plasma during the first 24 hours of their unit stay  
## 112 The highest bicarbonate concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 113 The lowest bicarbonate concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 114 The highest hemoglobin concentration for the patient during the first 24 hours of their unit stay  
## 115 The lowest hemoglobin concentration for the patient during the first 24 hours of their unit stay  
## 116 The highest volume proportion of red blood cells in a patient's blood during the first 24 hours of their unit stay, expressed as a fraction  
## 117 The lowest volume proportion of red blood cells in a patient's blood during the first 24 hours of their unit stay, expressed as a fraction  
## 118 The highest international normalized ratio for the patient during the first 24 hours of their unit stay  
## 119 The lowest international normalized ratio for the patient during the first 24 hours of their unit stay  
## 120 The highest lactate concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 121 The lowest lactate concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 122 The highest platelet count for the patient during the first 24 hours of their unit stay  
## 123 The lowest platelet count for the patient during the first 24 hours of their unit stay  
## 124 The highest potassium concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 125 The lowest potassium concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 126 The highest sodium concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 127 The lowest sodium concentration for the patient in their serum or plasma during the first 24 hours of their unit stay  
## 128 The highest white blood cell count for the patient during the first 24 hours of their unit stay  
## 129 The lowest white blood cell count for the patient during the first 24 hours of their unit stay  
## 130 The lowest albumin concentration of the patient in their serum during the first hour of their unit stay  
## 131 The lowest albumin concentration of the patient in their serum during the first hour of their unit stay  
## 132 The highest bilirubin concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 133 The lowest bilirubin concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 134 The highest blood urea nitrogen concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 135 The lowest blood urea nitrogen concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 136 The highest calcium concentration of the patient in their serum during the first hour of their unit stay  
## 137 The lowest calcium concentration of the patient in their serum during the first hour of their unit stay  
## 138 The highest creatinine concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 139 The lowest creatinine concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 140 The highest glucose concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 141 The lowest glucose concentration of the patient in their serum or plasma during the first hour of their unit stay  
## 142 The highest bicarbonate concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 143 The lowest bicarbonate concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 144 The highest hemoglobin concentration for the patient during the first hour of their unit stay  
## 145 The lowest hemoglobin concentration for the patient during the first hour of their unit stay  
## 146 The highest volume proportion of red blood cells in a patient's blood during the first hour of their unit stay, expressed as a fraction  
## 147 The lowest volume proportion of red blood cells in a patient's blood during the first hour of their unit stay, expressed as a fraction  
## 148 The highest international normalized ratio for the patient during the first hour of their unit stay  
## 149 The lowest international normalized ratio for the patient during the first hour of their unit stay  
## 150 The highest lactate concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 151 The lowest lactate concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 152 The highest platelet count for the patient during the first hour of their unit stay  
## 153 The lowest platelet count for the patient during the first hour of their unit stay  
## 154 The highest potassium concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 155 The lowest potassium concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 156 The highest sodium concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 157 The lowest sodium concentration for the patient in their serum or plasma during the first hour of their unit stay  
## 158 The highest white blood cell count for the patient during the first hour of their unit stay  
## 159 The lowest white blood cell count for the patient during the first hour of their unit stay  
## 160 The highest arterial partial pressure of carbon dioxide for the patient during the first 24 hours of their unit stay  
## 161 The lowest arterial partial pressure of carbon dioxide for the patient during the first 24 hours of their unit stay  
## 162 The highest arterial pH for the patient during the first 24 hours of their unit stay  
## 163 The lowest arterial pH for the patient during the first 24 hours of their unit stay  
## 164 The highest arterial partial pressure of oxygen for the patient during the first 24 hours of their unit stay  
## 165 The lowest arterial partial pressure of oxygen for the patient during the first 24 hours of their unit stay  
## 166 The highest fraction of inspired oxygen for the patient during the first 24 hours of their unit stay  
## 167 The lowest fraction of inspired oxygen for the patient during the first 24 hours of their unit stay  
## 168 The highest arterial partial pressure of carbon dioxide for the patient during the first hour of their unit stay  
## 169 The lowest arterial partial pressure of carbon dioxide for the patient during the first hour of their unit stay  
## 170 The highest arterial pH for the patient during the first hour of their unit stay  
## 171 The lowest arterial pH for the patient during the first hour of their unit stay  
## 172 The highest arterial partial pressure of oxygen for the patient during the first hour of their unit stay  
## 173 The lowest arterial partial pressure of oxygen for the patient during the first hour of their unit stay  
## 174 The highest fraction of inspired oxygen for the patient during the first hour of their unit stay  
## 175 The lowest fraction of inspired oxygen for the patient during the first hour of their unit stay  
## 176 The APACHE IVa probabilistic prediction of in-hospital mortality for the patient which utilizes the APACHE III score and other covariates, including diagnosis.  
## 177 The APACHE IVa probabilistic prediction of in ICU mortality for the patient which utilizes the APACHE III score and other covariates, including diagnosis  
## 178 Whether the patient has a definitive diagnosis of acquired immune deficiency syndrome (AIDS) (not HIV positive alone)  
## 179 Whether the patient has a history of heavy alcohol use with portal hypertension and varices, other causes of cirrhosis with evidence of portal hypertension and varices, or biopsy proven cirrhosis. This comorbidity does not apply to patients with a functioning liver transplant.  
## 180 Whether the patient has been diagnosed with diabetes, either juvenile or adult onset, which requires medication.  
## 181 Whether the patient has cirrhosis and additional complications including jaundice and ascites, upper GI bleeding, hepatic encephalopathy, or coma.  
## 182 Whether the patient has their immune system suppressed within six months prior to ICU admission for any of the following reasons; radiation therapy, chemotherapy, use of non-cytotoxic immunosuppressive drugs, high dose steroids (at least 0.3 mg/kg/day of methylprednisolone or equivalent for at least 6 months).  
## 183 Whether the patient has been diagnosed with acute or chronic myelogenous leukemia, acute or chronic lymphocytic leukemia, or multiple myeloma.  
## 184 Whether the patient has been diagnosed with non-Hodgkin lymphoma.  
## 185 Whether the patient has been diagnosed with any solid tumor carcinoma (including malignant melanoma) which has evidence of metastasis.  
## 186 Admission diagnosis group for APACHE III  
## 187 Admission diagnosis group for APACHE II  
## 188 Example mortality prediction, shared as a 'baseline' based on one of the GOSSIS algorithm development models.

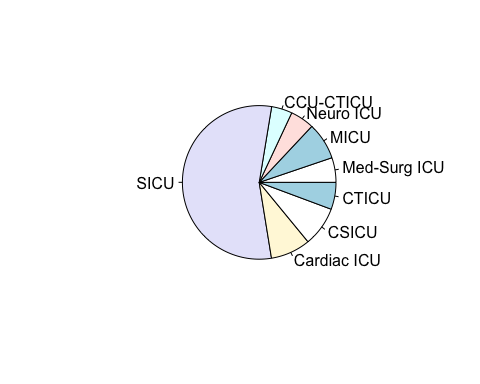
Below is the indication of total number of hospital deaths (whether the patient died during hospitalization or not)- 1 did not survive and 0 as survived.

#using table function to create a categorical representation of the data  
table(dataset['hospital\_death'])

## hospital\_death  
## 0 1   
## 83798 7915

Below pie chart depicts the types of ICUs that the study includes.

#counting labels in icu\_type column  
maj\_icu = table(dataset['icu\_type'])  
#pie chart for icu\_type column  
labels <- c('Med-Surg ICU','MICU','Neuro ICU','CCU-CTICU','SICU','Cardiac ICU','CSICU','CTICU')  
pie(maj\_icu,labels)



As part of the data cleaning process following steps have been done. Considering the fact that undersampling of non-events as one of the most important steps in data cleaning unnecessary variables i.e., non-medical variables like patient ID, encounter ID have been eliminated and required variables have been captured.

#Input required value adding columns for analysis. Ignored columns that contained non-medical data.  
input\_cols <- c('hospital\_id', 'age', 'bmi', 'elective\_surgery', 'ethnicity', 'gender',  
 'height', 'icu\_admit\_source', 'icu\_id', 'icu\_stay\_type', 'icu\_type',  
 'pre\_icu\_los\_days', 'weight', 'apache\_2\_diagnosis',  
 'apache\_3j\_diagnosis', 'apache\_post\_operative', 'arf\_apache',  
 'gcs\_eyes\_apache', 'gcs\_motor\_apache', 'gcs\_unable\_apache',  
 'gcs\_verbal\_apache', 'heart\_rate\_apache', 'intubated\_apache',  
 'map\_apache', 'resprate\_apache', 'temp\_apache', 'ventilated\_apache',  
 'd1\_diasbp\_max', 'd1\_diasbp\_min', 'd1\_diasbp\_noninvasive\_max',  
 'd1\_diasbp\_noninvasive\_min', 'd1\_heartrate\_max', 'd1\_heartrate\_min',  
 'd1\_mbp\_max', 'd1\_mbp\_min', 'd1\_mbp\_noninvasive\_max',  
 'd1\_mbp\_noninvasive\_min', 'd1\_resprate\_max', 'd1\_resprate\_min',  
 'd1\_spo2\_max', 'd1\_spo2\_min', 'd1\_sysbp\_max', 'd1\_sysbp\_min',  
 'd1\_sysbp\_noninvasive\_max', 'd1\_sysbp\_noninvasive\_min', 'd1\_temp\_max',  
 'd1\_temp\_min', 'h1\_diasbp\_max', 'h1\_diasbp\_min',  
 'h1\_diasbp\_noninvasive\_max', 'h1\_diasbp\_noninvasive\_min',  
 'h1\_heartrate\_max', 'h1\_heartrate\_min', 'h1\_mbp\_max', 'h1\_mbp\_min',  
 'h1\_mbp\_noninvasive\_max', 'h1\_mbp\_noninvasive\_min', 'h1\_resprate\_max',  
 'h1\_resprate\_min', 'h1\_spo2\_max', 'h1\_spo2\_min', 'h1\_sysbp\_max',  
 'h1\_sysbp\_min', 'h1\_sysbp\_noninvasive\_max', 'h1\_sysbp\_noninvasive\_min',  
 'd1\_glucose\_max', 'd1\_glucose\_min', 'd1\_potassium\_max',  
 'd1\_potassium\_min', 'apache\_4a\_hospital\_death\_prob',  
 'apache\_4a\_icu\_death\_prob', 'aids', 'cirrhosis', 'diabetes\_mellitus',  
 'hepatic\_failure', 'immunosuppression', 'leukemia', 'lymphoma',  
 'solid\_tumor\_with\_metastasis', 'apache\_3j\_bodysystem',  
 'apache\_2\_bodysystem')

All the missing values have been replaced with median values. In one of the studies put forward by (Raffa, et al., 2019), missing values were filled with predicted values. Median imputation has been chosen in this study as the data is skewed.

#Replacing missing data with median values  
dataset <- dataset %>%   
 dplyr::mutate\_if(is.numeric, function(x) ifelse(is.na(x), median(x, na.rm = T), x))

The dataset is then split into 2 parts with split ratio 0.8 for traning and testing.

#splitting the dataset into 2 sets (80% and 20%) for training and testing  
split <- sample.split(dataset, SplitRatio = 0.8)  
train <- subset(dataset, split == "TRUE")  
test <- subset(dataset, split == "FALSE")

## Methodology

Once the data is cleaned and prepared for analysis, the randomForest function is used to train the model.To improve the quality of the model, random ‘ntree’ values (like 3, 4, 5, 20, 40, 60, 80,..) have been picked up, the model has been tested and the OOB error (out-of-box error) has been recorded for each case. The number of trees=5 has been chosen as it has the least error.

#Using set.seed to ensure the results are same for randomization  
set.seed(120)  
#Training the random Forest model.  
classifier\_RF = randomForest(x = subset(train, select = c(input\_cols)),  
 y = train$hospital\_death,  
 ntree = 5, importance=TRUE)

Then the prediction of test set results has been carried out in the below chunk.

#Predicting the Test set results  
y\_pred = predict(classifier\_RF, newdata = subset(test, select = c(input\_cols)))

Confusion matrix has been generated for each of the nodes of 5 trees.

#creating confusion matrix for each node of random forest  
confusion\_mtx = table(test$hospital\_death, y\_pred)  
#confusion\_mtx

The plot of the model looks as depicted below. The error of each of the trees is ranging between 9.5% to 11.5%.

#Plotting modelChart

Description automatically generated  
plot(classifier\_RF)

Below is the importance plot of the variables. The Mean Decrease Accuracy (%IncMSE) shows the value of how much the model accuracy reduces if we leave out the respective variable and Mean Decrease Gini (IncNodePurity) is a measure of the importance of the variable based on the Gini Impurity Index that is used for calculating the randomness in the trees. Higher the value of %IncMSE or IncNotePurity , higher is the importance of the variable to the model that has been built.

Hence, the following variables can be considered as the most important ones in predicting the hospital\_dealth with the highest accuracy possible.

#Importance plot to show "importance" of variables: higher value indicates higher importance  
importance(classifier\_RF)

## %IncMSE IncNodePurity  
## hospital\_id 2.0446394 75.5126218  
## age 3.7544660 85.1121925  
## bmi 1.9989024 91.0193894  
## elective\_surgery 1.3030091 3.9939077  
## ethnicity 0.3932532 23.8753384  
## gender 0.3178573 7.8195496  
## height 2.9632224 80.8754394  
## icu\_admit\_source 1.7152603 30.8762027  
## icu\_id 5.7701742 82.6836187  
## icu\_stay\_type -0.7846047 4.0109228  
## icu\_type 0.3673275 40.0679856  
## pre\_icu\_los\_days 0.9515222 108.3852490  
## weight 1.9174882 88.9219169  
## apache\_2\_diagnosis 2.0017653 43.5001061  
## apache\_3j\_diagnosis 2.7200288 104.4963073  
## apache\_post\_operative 0.8111780 3.6784846  
## arf\_apache 1.4075611 6.4698719  
## gcs\_eyes\_apache 5.1673523 17.7205629  
## gcs\_motor\_apache 3.1798582 83.2444763  
## gcs\_unable\_apache -0.4000742 6.0474092  
## gcs\_verbal\_apache 3.0906135 21.1876655  
## heart\_rate\_apache 2.6834325 80.1857299  
## intubated\_apache 1.7568731 7.4969567  
## map\_apache 2.5616699 74.8205552  
## resprate\_apache 1.7275529 67.1940828  
## temp\_apache 4.1911199 78.8761388  
## ventilated\_apache 1.9545494 17.2910374  
## d1\_diasbp\_max 4.3654887 51.2314255  
## d1\_diasbp\_min 5.1745333 49.3093156  
## d1\_diasbp\_noninvasive\_max 2.5170346 51.5470406  
## d1\_diasbp\_noninvasive\_min 3.4719198 50.2333790  
## d1\_heartrate\_max 6.9438781 76.9240393  
## d1\_heartrate\_min 2.7420433 163.6484332  
## d1\_mbp\_max 3.9089407 49.9463820  
## d1\_mbp\_min 6.5606748 76.2865604  
## d1\_mbp\_noninvasive\_max 2.6267270 59.0927366  
## d1\_mbp\_noninvasive\_min 6.6622346 58.7072278  
## d1\_resprate\_max 2.2677701 80.7637327  
## d1\_resprate\_min 5.7229079 82.3781913  
## d1\_spo2\_max 2.4327993 33.3149112  
## d1\_spo2\_min 8.5811459 157.4225156  
## d1\_sysbp\_max 2.9270771 56.2152029  
## d1\_sysbp\_min 3.4686178 129.5550871  
## d1\_sysbp\_noninvasive\_max 2.4217814 66.6728502  
## d1\_sysbp\_noninvasive\_min 2.6231620 76.4319348  
## d1\_temp\_max 3.6622711 112.3104214  
## d1\_temp\_min 3.5593877 110.7668565  
## h1\_diasbp\_max 5.2544922 52.7575441  
## h1\_diasbp\_min 2.7682736 46.5680539  
## h1\_diasbp\_noninvasive\_max 7.9084240 54.9643525  
## h1\_diasbp\_noninvasive\_min 4.4873880 41.9562580  
## h1\_heartrate\_max 3.5932936 72.2855544  
## h1\_heartrate\_min 6.6398651 74.5293740  
## h1\_mbp\_max 4.1220171 42.7219022  
## h1\_mbp\_min 4.4969646 38.8076271  
## h1\_mbp\_noninvasive\_max 5.8853097 45.6512327  
## h1\_mbp\_noninvasive\_min 2.7318580 50.9722387  
## h1\_resprate\_max 7.4647613 57.6599933  
## h1\_resprate\_min 3.1424218 72.0828952  
## h1\_spo2\_max 6.5350693 38.4901077  
## h1\_spo2\_min 2.8632441 46.5478493  
## h1\_sysbp\_max 3.2972983 51.1578464  
## h1\_sysbp\_min 4.1172969 54.5094236  
## h1\_sysbp\_noninvasive\_max 3.9078800 41.2845934  
## h1\_sysbp\_noninvasive\_min 4.1793313 46.5458175  
## d1\_glucose\_max 2.3277239 83.5398238  
## d1\_glucose\_min 1.6615883 98.3618980  
## d1\_potassium\_max 0.3850400 66.8800707  
## d1\_potassium\_min 1.1492357 80.9214278  
## apache\_4a\_hospital\_death\_prob 6.9411931 781.6522793  
## apache\_4a\_icu\_death\_prob 11.5002856 334.1043519  
## aids 0.0000000 0.3820776  
## cirrhosis -1.2425196 4.6705271  
## diabetes\_mellitus -0.8423194 9.9217376  
## hepatic\_failure -0.6321519 5.2847896  
## immunosuppression -0.3246643 12.4916825  
## leukemia -0.3490094 3.8790860  
## lymphoma -2.1362714 2.4167093  
## solid\_tumor\_with\_metastasis -0.5291495 7.2740839  
## apache\_3j\_bodysystem 0.3304217 18.7398662  
## apache\_2\_bodysystem 0.6855845 23.7546281

## Results

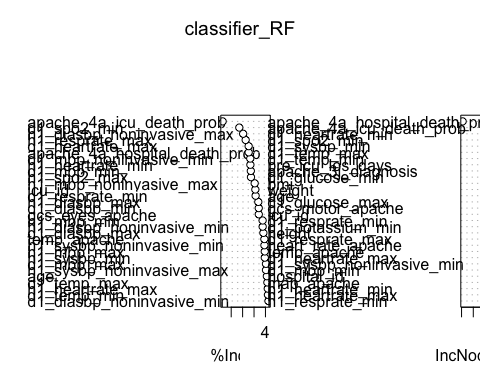
The importance plot below clearly depicts variables that play a key role in achieving an efficiency of ~91.5% in predicting whether a patient survives the hospitalisation or not. Important variables:

On the basis of %IncMSE: d1\_heartrate\_min (9.675), temp\_apache, bmi, d1\_glucose\_max, h1\_resprate\_min apache\_2\_diagnosis, d1\_temp\_min, d1\_sysbp\_noninvasive\_min, d1\_diasbp\_max, d1\_spo2\_max (5.98) On the basis of IncNodePurity: apache\_4a\_hospital\_death\_prob (716.2), apache\_4a\_icu\_death\_prob, d1\_heartrate\_min, d1\_sysbp\_noninvasive\_min, d1\_spo2\_min, pre\_icu\_los\_days, d1\_sysbp\_min, d1\_glucose\_min, d1\_temp\_min (108.22).

The factors that are of high importance in both scenarios may be considered as the top priority parameters that are essential in predicting the survival of a patient efficiently. This study may be taken as reference by medical professionals to create a checklist of the laboratory tests with importance/priority on each of the tests.

Below is the importance plot for each of the values (mean decrease accuracy & mean decrease gini).

#Variable importance plot  
varImpPlot(classifier\_RF)



## Conclusion

From this research project, random forest model with t=number of trees=5 shows high perfromnce on the dataset. It can also be deduced that variables at the top of the importance plot chart can be considered by the medical practitioner at the ICU as key parameters in deciding the distressed pateint’s condition. Laboratory tests related to these values may be listed as basic and high priority tests and may be ensured that they are performed as soon as the patient arrives at the facility. This would increase the survival chances of the patient. Extra care may be extended to patients who have already been to the ICU before as it is one of the most important variables with IncNodepurity value greater than 100, also placed among the top 10 positions of importance plot.

## Implications

This study may be used as a reference for anybody conducting medical-related analysis on predicting patient survival rates in an emergency. Medical expertise may be applied to improve the outcomes and the model’s performance. As an IT student with no medical knowledge, numeric variables were prioritized during data cleaning and preparation. As a medical practitioner, analyzing this dataset or a comparable one will undoubtedly provide higher efficiency and can be approved for real-time usage.

## References

Dataset: <https://www.kaggle.com/datasets/sadiaanzum/patient-survival-prediction-dataset/download> (Links to an external site.)

<https://journals.lww.com/ccmjournal/Citation/2019/01001/33> (Links to an external site.)

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/123-required-r-packages-for-principal-component-methods/>

Citation1: Raffa, Jesse1; Johnson, Alistair1; Celi, Leo Anthony2,,1; Pollard, Tom1; Pilcher, David3; Badawi, Omar4 33: THE GLOBAL OPEN SOURCE SEVERITY OF ILLNESS SCORE (GOSSIS), Critical Care Medicine: January 2019 - Volume 47 - Issue 1 - p 17 doi: 10.1097/01.ccm.0000550825.30295.dd

Citation2: Chi, Stephen; Buettner, Benjamin; Ma, Jessica; Pollard, Katherine; Muir, Monica; Kolekar, Charu; Al-Hammadi, Noor; Kollef, Marin; Dans, Maria 34: EARLY PALLIATIVE CARE CONSULTATION IN THE MEDICAL ICU: A CLUSTER RANDOMIZED CROSSOVER TRIAL Critical Care Medicine:January, 2019

<https://journals.lww.com/ccmjournal/toc/2019/01001>

Breiman, L. (2001), Random Forests, Machine Learning 45(1), 5-32.

Breiman, L (2002), ``Manual On Setting Up, Using, And Understanding Random Forests V3.1’’,

<https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest>

<https://www.listendata.com/2014/11/random-forest-with-r.html>