January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

Machine Learning Application to Predict the Length of Stay of type 2 Diabetes Patients in the Intensive Care Unit

[1]carol Anne Hargreaves, [2]chow An An Cherie
[1,2]Department of Statistics and Applied Probability, National University of Singapore, Singapore
Email: [1]stacah@nus.edu.sg, [2]e0005056@u.nus.edu

Article Info
Volume 82
Page Number: 6143 - 6163
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 30 January 2020*

Abstract:
The Intensive Care Units (ICU) are costly units in hospitals catered to a small group of critically ill patients who are monitored round-the-clock, and cost as much as six times that of normal wards. Shortage of beds is a prevalent issue in many countries, including Singapore. This problem is worsened by the aging population which has led to rising healthcare demands and costs. It is a serious problem as there may be delays in care due to the lack of beds. Reducing patients' length of stay (LOS), especially for the ICU, is one of the key priorities for hospitals in a bid to save cost and manage hospital resources more efficiently. One way to do this is to identify patients that are at risk of having prolonged hospital stay at
the start of their hospitalisation. Accurate identification of such patients allow early planning of treatment and provision of more intensive care to speed up their recovery. As a result, these patients may be discharged earlier, hence reaping cost saving benefits to hospitals and mitigate the bed shortage problem. This project is focused on Type 2 Diabetes Mellitus, which is recognised as a global epidemic due to increasing prevalence and the potentially serious complications resulting from this disease. This project first identifies risk factors for prolonged ICU LOS for patients with Type 2 Diabetes Mellitus, and then makes use of these factors to develop an accurate machine learning model to identify these patients. Finally, incentives to reduce LOS is explored, and hospital cost savings, specifically for Singapore, is calculated from reducing LOS in ICUs.

***Keywords**: Machine Learning, Type 2 Diabetes, Prolonged Length of Stay, Intensive Care Unit, High Risk Patients, Cost Savings, Logistic Regression, Random Forests, Predictor Variables.*

---

1. INTRODUCTION

1.1 PROJECT BACKGROUND

Type 2 Diabetes Mellitus is a chronic metabolic disorder in which the body is unable to respond to the insulin produced by the pancreas, leading to persistently high blood sugar levels. The rapid increase in the prevalence of Type 2 diabetes has led to it being recognized as a global epidemic, with obesity, poor diet, physical inactivity, smoking, alcohol use being a few of the most important drivers [1,2]. The American Diabetes Association (ADA) estimated that the total costs of diabetes increased 26% from 2012 to $327 million in 2017, mostly contributed by the high medical expenditures [3]. These include hospital inpatient care, prescription medications, anti-diabetic agents and physician visits.

Chronic complications of Type 2 Diabetes Mellitus include cardiovascular diseases, stroke, eye diseases, kidney diseases, neurological disorders and foot problems.

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

Acute conditions include diabetic ketoacidosis and hyperosmolar hyperglycaemic state, which are both life-threatening [4]. As a result of the serious complications that results from diabetes, patients with diabetes are more than twice as likely to be admitted to the hospital and are more likely to have prolonged stay as compared to the general population [5]. Older diabetic patients generally require even more care due to the increased risk and severity of acute and chronic complications [6]. These patients also face higher mortality rates [7].

Since complications resulting from type 2 diabetes can be potentially life-threatening, some patients may be transferred to the intensive care unit (ICU) which provides intensive care and supervision. Patients can be admitted to the ICU directly from the emergency room, their wards if their condition rapidly deteriorates, or after a surgery which requires intensive post-operative care. Typically, patients stay for approximately 1 to 3 days in the ICU. However, there is a small percentage of patients who stay way beyond the usual length of stay (LOS) and are categorised as patients with prolonged stay. These patients have a significant impact on ICU bed utilisation and ICU costs as they require a disproportionate amount of ICU resources [8,9]. This poses a concern for hospitals since the ICUs cost as much as six times that of regular wards and account for 8% to 30% of the hospital expenditures [10,11].

Furthermore, having a significant proportion of prolonged stay patients in the ICU means that there are less beds available for new admissions. This may mean that patients with severe or life-threatening conditions may not be able to get the immediate care that they need due to the lack of ICU beds and resources. Hence, there is a strong need for hospitals to find ways to reduce the LOS of patients with prolonged stay. The early identification of patients that require prolonged stay can enhance the planning and management of hospital and ICU resources. At the same time, it allows for these high-risk patients to be targeted for more intensive management of the disease to prevent or delay complications, hence speeding up their recovery [12]. This may allow hospitals to save a significant part of their budget, thus allowing them to channel the budget to other areas like Research and Development.

1.2 PROJECT OBJECTIVE

The objectives of this project are:

1. To identify factors associated with prolonged ICU stay for diabetic patients for early intervention and better patient care
2. To compare the characteristics of patients with prolonged and normal stay
3. To build prediction models using machine learning techniques to identify patients
4. with prolonged stay to compare the performance between several data mining models and select the one with the best performance. The main models used are naïve Bayes classification, logistic and penalised logistic regression, decision trees and random forest.
5. To calculate cost savings for reduced length of stay to help hospitals achieve greater economic efficiency.

2. Study of Related Papers

Many research papers have predicted the length of stay (LOS) for patients admitted to the hospital, but only a few identified patients with prolonged stay. Noticeably, there were only a few papers that focused on the risk factors associated with prolonged stay for diabetic patients but none assessed and compared the performance of different machine learning algorithms. Furthermore, there is no uniform definition of what constitutes a prolonged stay [13]. Several methods to identify patients with prolonged LOS were recognised and evaluated as follows:

1. A simplistic method is to use the empirical rule,where a prolonged LOS will be defined as a LOS that is more than 2 standard deviations above the mean LOS [14,15]. However, this

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

method is undesirable as ICU LOS does not follow normal distribution as it is heavily right skewed.
2. Designating a specific proportion of patients as those with prolonged stay was also used in some papers (eg. 25% with the prolonged stay) [16,17,18].
3. Some papers define prolonged LOS by designating a specific duration >7 days [19,20], ≥ 10 days [16], ≥ 14 days [21], ≥ 21 days [18] or ≥ 30 days [22].
4. Another method is to visually examine the LOS frequency distribution and identify a threshold for the "tail" of the distribution. Any LOS above that hreshold is considered prolonged LOS.

Despite differences in the definition of a prolonged stay, studies have repeatedly shown that a small percentage of patients with lengthy stay account for a disproportionate amount of resource use and cost [9, 19]. Hence early identification of patients with prolonged stay allows for targeted treatment which may reduce their stay, thus improving ICU efficiency.

The review of related literature served to extract useful and appropriate variables for this project. These papers included patient characteristics, presence of complications and laboratory test results as variables. A summary of the predictor variables in the papers reviewed are recorded in the Table 1.

Table 1: Common Predictor Variables in Related Papers

Category	Variables
Patient Information	Age, gender, smoking status, body mass index (BMI), alcohol consumption, insulin use, admission type
Presence of Chronic Conditions	Hypertension, diabetic retinopathy (eye disease), diabetic nephropathy (kidney disease), Diabetic Neuropathy (Nerve damage), heart diseases, hyperlipidaemia (high lipid levels), foot diseases, stroke
Laboratory Test Results	Blood glucose level, haemoglobin A1c level, glomerular filtration rate, serum potassium, serum sodium, serum chlorine, serum creatinine, albumin, blood urea nitrogen (BUN) level

The papers analysed also used different machine learning methods for prediction. The most common method used is the logistic regression. As there has yet to be any study done that identifies patients that are at risk of having prolonged stays and at the same time compare between different prediction models for the best performance, this project seeks to stand out by achieving both.

3. Methodology for Data Preparation

Data was obtained from the MIMIC-III (Medical Information Mart for Intensive Care III) Critical Care Database, a large database comprising of anonymized health-related data associated with patients who stayed in the ICUs at the Beth Israel Deaconess Medical Centre in Boston, Massachusetts, USA between 1 June 2001 and 10 October 2012 [23]. The database includes information such as patients' demographics, vital signs measurements, laboratory test results, procedures, medications, mortality and length of stay.

3.1 Derivation of Length of Stay Definition

There are 26 csv files in the MIMIC-III database. After careful consideration, the following 7 files are used in this project and a short description of the files is provided in Table 4. The full dataset includes 55323 hospital admissions across 44091 patients. In the 55323 hospital admissions, there are 58873 ICU admissions. This is possible because in each hospital admission, there may be more than one ICU admission.

There is a total of 20 ICD-9 codes that are Type 2 Diabetes coded. Only primary Type 2 Diabetes is included in this project as the

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

inclusion of secondary diabetes, defined as diabetes acquired as a consequence of other medical conditions, may not lead to accurate predictions. Table 5 shows the ICD-9 codes and frequency and description of each diagnosis. There is a total of 12225 hospital admissions and 13153 ICU admissions across 9444 Type 2 Diabetes patients.

ICU (length of stay) LOS is defined as the total number of fractional days that a patient stayed in the ICU for each hospital admission. Since there are patients who were admitted to the ICU multiple times throughout a single hospital admission, the LOS for each ICU admission is summed up to make up the ICU LOS for a single hospital admission.

The mean and median ICU LOS are 2.3 and 4.5 days respectively, hence extremely right skewed. This observation is consistent with the nature of hospital LOS stated in various literature [14,15,24] . A review of the ICU LOS frequency distribution showed that the "tail". began at about 7 days (Figure 1). Hence, a stay of seven days or longer is chosen as the threshold for prolonged stay. This threshold is reasonable as it demonstrated enough time to reflect any complications and the response to treatment if required [9]. Furthermore, an ICU LOS of seven or more days represents the 84th percentile, which is an appropriate LOS for a prolonged stay. Admissions with prolonged stay also took up 54% of the total ICU LOS across all Type 2 diabetes admissions, a significant percentage considering the proportion of prolonged stay patients was only about 16%. There is hence an incentive to reduce the length of stay of these prolonged stay patients. Admissions with a LOS more than seven days will be coded as longstay, while a LOS less than that will be coded as normal.

3.2 Class Imbalance Problem

It is worthwhile to note that there is class imbalance between normal and longstay, which happens when the class distributions are highly imbalanced. The normal class makes up 84% of all admissions, while the prolonged class only makes up 16%. (Figure 1) Predictions will be dominated by data from the majority class normal and there will be a high probability of misclassification of the minority class longstay compared to the majority class normal. In view of the class imbalance problem, there are methods employed to overcome it and they are explained in Section 3.6

3.3 Prior Variable Selection

After analysing the potential variables from related literature, the common variables illustrated in Table 2 were extracted from the MIMIC-III database. As laboratory test results are longitudinal with multiple measurements for each patient, the first measurement for each of the variables for each admission is taken and used as a potential predictor variable

General patient information included age, gender, insurance type, body mass index (BMI), smoking status, hospital admission type, insulin use, presence of uncontrolled diabetes. There is a strong association between age and prevalence of diabetes. Middle-aged and older adults have the highest risk of developing type 2 diabetes [27]. The categorical variable age comprises of 5 age groups: 54 and below, 55 – 64, 65 – 74, 75 – 84 and 85 and above.

Body Mass Index (BMI) is measured by the ratio of height and the square of weight. A BMI of 25 and below is considered within a normal range, a BMI that is between 25 and 30 is considered overweight, and a BMI of 30 and above is considered obese. People who are obese are at greater risk of developing diabetes [34]. A patient's BMI is represented as the categorical variable BMI.

Smoking is a known to increase insulin resistance and is major risk factor for diabetes [35]. Smokers are reported to be 30–40% more likely to develop type 2 diabetes than non-smokers [31]. Smoking is indicated by the dummy variable smoking.

There are 3 types of hospital admissions: emergency, urgent and elective. Emergency cases refer to life-threatening conditions that require immediate care, while urgent admission refers to non-emergency cases but still requiring medical care (e.g. fractures).

Elective admission can be in the form of direct admission from the clinic, planned admission for an elective surgery, or transfer from another hospital or health institution. The categorical variable admission type is used to indicate the type of hospital admission.
Diabetic patients may require the use of insulin to control their blood glucose levels. Patients that use insulin in this dataset are all on long-term current use (before admission). The dummy variable insulin is used to indicate insulin use.

Uncontrolled diabetes occurs when diabetic patients do not manage their condition well, leading to dangerously high levels of blood glucose. Patients with uncontrolled diabetes have an even higher risk of complications including kidney, heart and eye problems, and in serious cases even lead to a life-threatening condition called diabetic ketoacidosis [32]. The dummy variable uncontrolled is used to indicate the presence of uncontrolled diabetes.

3.4 Presence of Chronic Conditions

The papers reviewed in Section 1.3 included complications due to diabetes. It is worthwhile to extract some of these chronic conditions or complications from the MIMIC-III dataset for diabetic patients. These conditions are: hypertension, hyperlipidemia, heart disease, foot disease, stroke, kidney disease, diabetic neuropathy and diabetic retinopathy.

Hypertension (high blood pressure), is defined as having a blood pressure of more than 140/90 millimetres of mercury (mmHg) and is common in people with diabetes [33]. Presence of hypertension is indicated by the dummy variable hypertension.
Hyperlipidemia (high lipid levels) are also common in people with diabetes and is largely caused by unhealthy diet and sedentary lifestyles [34]. Presence of Hyperlipidemia is indicated by the dummy variable lipid.

Diabetes is strongly associated with heart diseases. The prevalence of heart disease is reported to be as high as 65% in patients in diabetes and is a major cause of death for diabetic patients [35,36]. Presence of heart disease is indicated by the dummy variable heart.

Foot problems are common amongst diabetic patients. The most common ones include gangrene, corns and callosities, hammertoe, bunion, foot and toe infections, foot ulcers, plantar wart, foot and toe blisters [37]. Patients with the presence of any of these foot problems was considered as having foot disease indicated by the dummy variable foot.

A stroke occurs when a blood vessel supplying the brain with oxygen becomes damaged or blocked off [38]. Diabetes greatly increases the risk of stroke by 1.5 times. The presence of kidney disease was indicated by the dummy variable stroke.

Kidney disease are commonly amongst diabetic patients. High blood glucose also damages blood vessels in the kidneys [39]. When the kidneys become impaired, they are unable to excrete waste materials and water out of the blood as well as before. The presence of kidney disease was indicated by the dummy variable kidney.

Diabetic Neuropathy is a type of nerve damage resulting from high blood glucose levels, causing pain in diabetic patients living with this condition [40]. The presence of diabetic neuropathy was indicated by the dummy variable neuropathy.

Diabetic retinopathy is one of the most common complications of diabetes. People with this condition have damaged blood vessels in their retina, causing blurred vision and even blindness [41]. Presence of diabetic retinopathy is indicated by the dummy variable retinopathy.

3.5 Laboratory Data Preparation

There are several laboratory test results that were extracted from the MIMIC-III database as potential variables. These variables are: blood glucose level, hemoglobin A1C test, serum sodium, serum potassium, serum chloride and blood urea nitrogen (BUN) level.

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

The normal blood glucose level for diabetic patients is between 80 to 130 milligrams per decilitre (mg/dL) [42]. Blood glucose levels beyond that are considered high. The first blood glucose reading taken in each admission was used as the predictor variable glucose.

Hemoglobin A1C test measures the average blood sugar for the past 3 months. The normal range for HA1C level is between 4% and 5.7% [43]. The higher the HA1C level, the higher the risk of developing complications associated with diabetes.

The normal serum sodium level is between 135 to 145 milliequivalents per liter (mEq/L) [44]. Blood sodium levels below that are considered low. Decreased serum sodium levels are occasionally observed in patients with diabetes mellitus, especially those with kidney problems [45]. The first serum sodium reading taken in each admission was used as the predictor variable sodium.

Serum potassium, serum chloride and Blood Urea Nitrogen (BUN) are indicators of renal health. The normal values for these laboratory test results are 3.5 to 5.0 mEq/L [46], 97 to 107 mEq/L [47], and 7 to 20 mg/dL [48] respectively. One of the complications caused by diabetes is kidney disease. Patients with kidney diseases have diminished kidney capacity to excrete these electrolytes and waste products into urine. In such cases, these patients will have high levels of these products in the blood. The first readings for serum potassium, serum chloride, serum creatinine and BUN are taken in each admission is used as the predictor variables potassium, chloride, and BUN respectively.

3.6 Data Exploration

Before relationships between variables are analysed, distribution and patterns of individual variables should first be examined. Figure 1 shows the distribution of laboratory test results. The severe and moderate categories made up the majority for each of the laboratory test results.

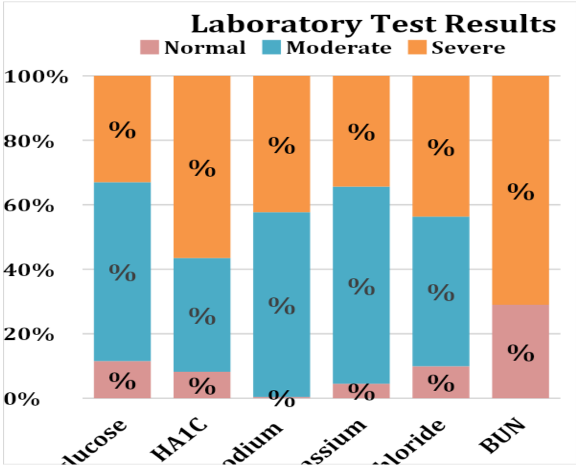

Figure 1: Distribution of laboratory test results

The majority of admissions are emergency cases (84.6%), while only a very small percentage of admissions belong to urgent cases (2.4%).

There was an imbalance in category levels. For example, Self-pay and Government constitute less than 2% of the total observations. These infrequent levels are unnecessary since they only make up a very small percentage, hence it makes sense to combine them to reduce the number of levels so that the variable will be more balanced. After combining levels, for example, Insurance was reduced from 5 to 2 levels.

The biggest age group was the 65 to 74 group, comprising of 27.8% of the total admissions. The total percentage of elderly patients (65 and above) adds up to 61.9%. This is consistent with the trend of increased risk of diabetes for older people. Older people also have higher risks of developing acute and chronic complications of diabetes and requiring hospitalisation [25].

Bivariate analysis was used to explore the relationship between two variables and understand the association or differences between variables. In the context of this project, it is of interest to analyse the relationship between predictor variables and outcome. (longstay or normal)

Stacked column charts were used to show the difference in the laboratory test variables between longstay and normal groups. (Figures 7 and 8) It was evident that the percentage of severe measurements for all laboratory test variables was higher for longstay compared to normal stay. This was not surprising as patients with longstay can be presumed to have more serious conditions than patients with normal stay

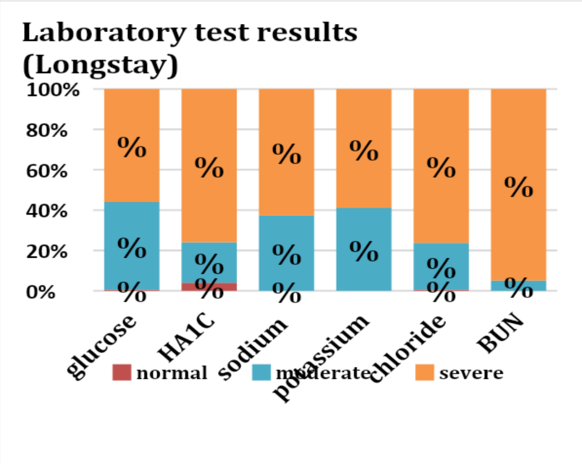

Figure 2: Stacked laboratory test results (longstay)

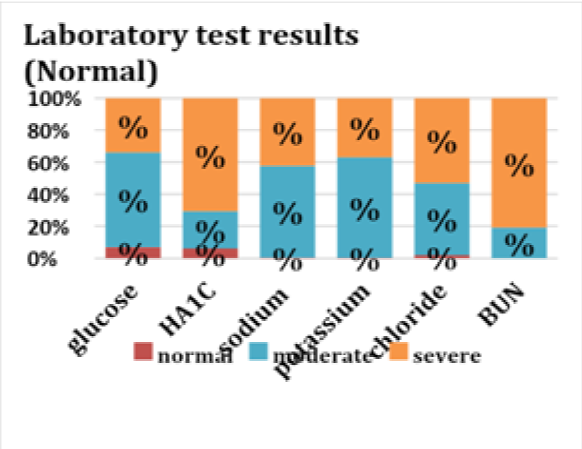

Figure 3: Stacked laboratory test results (Normal)

It is also interesting to compare the presence of chronic conditions between longstay and normal using stacked column charts. Figures 9 and 10 show that generally, conditions like hypertension, heart disease (heart) and kidney disease (kidney) are common for both longstay and normal stay. For longstay, there is a higher percentage of patients with heart disease (heart), kidney disease (kidney) and diabetic retinopathy (retinopathy) than that for normal. In fact, there is a much higher percentage of retinopathy for longstay (43%) than normal (12%).

For patient information variables, there is generally no significant difference between that of longstay and normal stay, especially for the variables insurance, gender, smoking, insulin use, and uncontrolled. According to the pie charts in Figure 11, only admission type and BMI seem to show some differences between longstay and normal. It is interesting to note that longstay admissions have a higher proportion of patients who were admitted via an emergency. There is also a higher percentage of patients that are obese.

3.7 Variable Selection

Since the subset of initial predictors is very large, there is a need to remove insignificant variables and only keep the important ones to reduce noise in the data. The Pearson's chi-squared test was performed on all potential variables and those that are independent of the outcome (p-value < 0.05) were removed.

Variables that are independent of the outcome variable are: gender (p-value 0.2835), insurance (p-value 0.7061), smoking status (p-value 0.07544), insulin use (p-value 0.1521), uncontrolled (p- value 0.8591). This result is not surprising as there was not much difference in these variables between longstay and normal stay. These variables are hence removed.

For the remaining variables, the reported p-values are generally very small, indicating significant differences in each variable between longstay and normal stay groups.

3.8 Checking for Multi-collinearity

There may be the issue of multi-collinearity, in which predictor variables are strongly correlated to one another. Variables that show high degree of multi-collinearity provide little unique information in the model, leading to unstable estimates with inflated standard errors of the coefficients [49]. The coefficients of correlated variables in the model are highly sensitive to slight changes in the data set, leading to some variables potentially being statistically insignificant when they should be significant. Hence, multi-collinearity affects the importance of predictor variables, which is crucial in this project. Furthermore, some data mining algorithms like logistic regression and Naïve Bayes classification require the data to have little or no multi-collinearity. Therefore, it is crucial to check the data for the presence of multi-collinearity and remove variables that are high correlated.

The generalised variance inflation factor (GVIF) is computed for all predictor variables. As a rule of thumb, a GVIF value above 5 suggests that there is high correlation between two predictor variables. After checking through the GVIF values, it can be said that there is no presence of multi-collinearity in the data. There are a total of 16 predictor variables for predicting length of stay.

4. Predictive ModellingTechniques

Two different methods were used in the prediction of prolonged ICU stay: Logistic regression, random forest, and decision trees. The data is split randomly into training and test sets: 70% for training (n=8557) and 30% for testing (n=3668). The model is first trained on the training set, before being tested on the test set.

The performance of the different models is analysed using the confusion matrix which is a table to summarise the prediction results. The number of correct and incorrect predictions are summarized with count values and broken down by each class. (Figure 12) True positive (TP) is the outcome where the model correctly predicts the positive class (longstay). True negative (TN) is the outcome where the model correctly predicts the negative class (normal). False positive (FP), also known as Type 1 error, is the outcome where the model incorrectly predicts the positive class (longstay). False negative (FN), also known as Type 2 error, is the outcome where the model incorrectly predicts the negative class (normal).

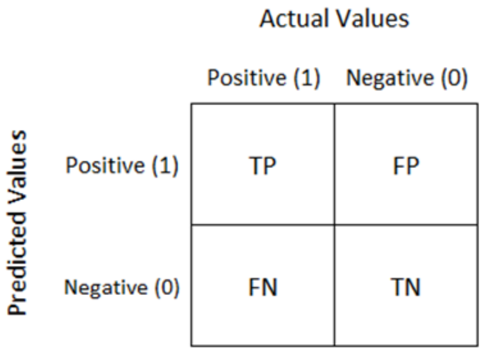

Figure 4: Confusion matrix

The performance of the machine learning models was compared using four evaluation metrics: overall accuracy, sensitivity, specificity and Area Under Curve (AUC). Overall accuracy is the ratio of correctly predicted classes to the total classes predicted. It is calculated using the following formula:

$$overall\ accuracy = \frac{TP + TN}{total}$$

Sensitivity is the proportion of actual positives that are corrected identified, which in this case represents the proportion of longstay admissions that are correctly identified as longstay. It is calculated using the following formula:

$$sensitivity = \frac{TP}{TP + FN}$$

Specificity is the proportion of actual negatives that are correctly identified, which in this case represents the proportion of normal admissions that are correctly identified as normal. It is calculated using the following formula:

$$specificity = \frac{TN}{TN + FP}$$

In this project, sensitivity is of more concern than specificity since one of the goals is to identify prolonged stay patients. Hence it makes sense to focus on improving the sensitivity more than specificity.

The Area Under Curve (AUC) measures the degree of separability between classes. It tells how well the model is capable of distinguishing between classes (longstay and normal). The higher the AUC, the better the model is at predicting the classes. The AUC arises from the area under the receiving operating characteristic (ROC) curve, which is a curve that illustrates the trade-off between sensitivity and specificity at different thresholds.

4.1.2 Overcoming Class Imbalance

As mentioned in Section 2.2.2, there is class imbalance present in the data, particularly in the outcome variable longstay. This poses a problem because the models will likely perform very poorly in terms of sensitivity (proportion of longstay admissions that are correctly identified as longstay) since predictions are dominated by contribution from the normal class. Since the aim is to predict the minority class longstay, it is crucial to increase the sensitivity to make the predictions meaningful.

There are many methods to overcome class imbalance, but the project will focus on four methods: up-sampling, down-sampling, Random Over-Sampling Examples (ROSE) and Synthetic Minority Oversampling Technique (SMOTE). All four methods involve the sampling of observations, allowing the classes to be more balanced. The best sampling method is used in the final model in each of the four algorithms.

4.2 Methodology

4.2.1 Logistic Regression Technique

Logistic regression uses the logistic function to return probability values based on predictor variables. The logistic function in this context is given by the equation.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$$

Where:

p is the probability of an admission being longstay

β_0 is the intercept of the regression line

β_1…β_n are the fit parameters

X_1…〖X〗_1 are the predictor variables.

The estimated probabilities can be mapped to two or more discrete classes by setting a threshold value, usually at 0.5. Probabilities above that will be classified as class longstay, and class normal otherwise.

p≥0.5,class=longstay
p<0.5,class=normal

Interpretation of Logistic Regression Model

The inverse of the logistic function gives the logit function, also known as log-odds, and is given by the equation below where p/(1-p) represents the odds of having longstay. The coefficients of the logistic regression model are represented by the log-odds.

$$logit(p) = log\ log\ \left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

The odds ratio is a useful result of logistic regression to measure the strength of association between each predictor variable and the outcome. The odds ratio for a categorical variable is the change in the odds of the outcome compared to that in the reference group of the categorical variable, holding all other variables constant. The greater the odds ratio, the better the predictor variable. The odds ratio can be obtained by taking the exponential of the coefficients of the logistic regression model.

4.2.2 Random Forest

Random forest involves creating many decision trees and aggregating the results from all the trees. The algorithm works by:

1. Draw ntree bootstrap samples, where ntree is the predefined number of trees to be grown. Bootstrapping involves repeatedly sampling samples of datasets with replacement from the original dataset.

2. For each bootstrap sample, a tree is grown. However, when forming each split, the random forest algorithm randomly selects mtry predictors within which the best split is selected. This differs from the decision tree algorithm in which all the predictors are searched when forming each split to find the best split. The predictor that results in the largest decrease in Gini impurity is selected as the best split.
3. The final outcome (longstay or normal) is predicted based on majority vote

The hyperparameter ntree is tuned by first examining the plot of error rate against number of trees, then finding an appropriate number of trees whereby the out-of-bag (OOB) error rate is stabilised. (Figure 14) Then, using this value of ntree, a grid search is performed to find the optimal value for mtry.

The random forest algorithm reduces correlation between trees by introducing randomness in two ways: 1) growing each tree to a bootstrapped sampled data set, and 2) randomly selecting the number of predictors to be considered at each split.

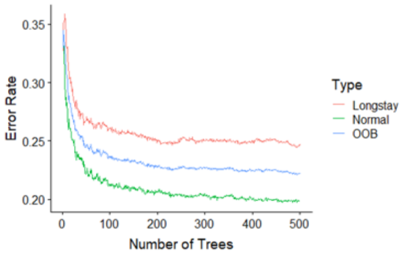

Figure 5: Plot of error rates against number of trees. The OOB error is stabilised after about 400 trees

Random forest helps in feature selection by weighing the importance of each variable according to how much they contribute to the model. This is done by calculating the variable importance for each predictor variable. Variable importance is measured by ranking the mean decrease in Gini. For each variable, the sum of GINI decrease across every tree in the forest is accumulated every time this particular variable is chosen to split a node. The sum will then be divided by the total number of trees in the forest to get the average. This makes the mean decrease in Gini. The larger the mean decrease in Gini, the more important the variable is. Hence, less important variables can be removed to increase the performance of the model.

4.3 RESULTS AND DISCUSSION

4.3.1 Implementation of the Logistic Regression

Using the glm function in the glmnet library, the logistic regression model was fitted. The original model had very poor sensitivity, which is not surprising due to class imbalance. ROSE achieved the best performance results among the 4 methods to overcome class imbalance. (Table 2)

Table 2: Summary of the Prediction Results

Models	Overall	Sensitivity	Specificity	AUC
Original	0.877	0.383	0.965	0.867
ROSE	**0.766**	**0.819**	**0.757**	**0.867**

Table 3 summarises the coefficients, odds ratio (OR) and corresponding 95% confidence interval, and p-value from likelihood ratio test.

Table 3: Analysis of logistic regression model with ROSE

Variables	Coefficients	OR (95%)	p-value
Age			

16 – 54*	0	1	
55 – 64	0.223	1.25 (1.09 - 1.41)	0.0052
65 – 74	0.278	1.32 (1.17 - 1.47)	< 0.001
75 – 84	0.253	1.29 (1.13 - 1.45)	< 0.001
Above 85	0.655	1.93 (1.73 - 2.13)	< 0.001
BMI			
Normal*	0	1	
Overweight	0.235	1.26 (1.06 - 1.46)	0.022
Obese	1.24	3.46 (3.26 - 3.66)	< 0.001
Admission Type			
Elective*	0	1	
Emergency	0.301	1.35 (1.20 - 1.50)	< 0.001
Urgent	0.334	1.40 (1.07 - 1.73)	0.047
Hypertension			
No*	0	1	
Yes	-0.216	0.81(0.71 - 0.91)	< 0.001
Hyperlipidemia			
No*	0	1	
Yes	-0.286	0.75(0.64 - 0.86)	< 0.001
Heart Disease			
No*	0	1	
Yes	0.265	1.30 (1.20 - 1.40)	< 0.001
Kidney Disease			
No*	0	1	
Yes	1.44	4.22 (4.11 - 4.33)	< 0.001
Foot Disease			
No*	0	1	
Yes	-0.365	0.69 (0.46 - 0.92)	< 0.001
Stroke			
No*	0	1	

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

Yes	0.853	2.35 (2.11 – 2.59)	< 0.001
Diabetic retinopathy			
No*	0	1	
Yes	-0.755	0.47 (0.23 – 0.71)	< 0.001
Glucose			
Below 130*	0	1	
130 to 250	1.58	4.85 (4.50 – 5.20)	< 0.001
Above 250	2.39	10.9 (10.5 – 11.3)	< 0.001
Sodium			
Below 135	2.17	8.76 (8.12 – 9.38)	< 0.001
135 to 145	1.23	3.42 (3.24 – 3.95)	< 0.001
Above 145*	0	1	
Potassium			
Below 3.5*	0	1	
3.5 to 5.0	1.33	3.78 (3.19 – 4.20)	0.01
Above 5.0	1.69	5.42 (4.90 – 5.82)	< 0.001
Chloride			
Below 97*	0	1	
97 to 107	1.28	3.60 (3.13 – 4.07)	< 0.001
Above 107	1.77	5.87 (5.40 – 6.34)	< 0.001
Hemoglobin A1C			
Below 5.7*	0	1	
5.7 to 6.4	-0.073	0.93 (0.87- 0.98)	0.04
Above 6.5	0.004	1.04 (1.01- 1.07)	0.04
Blood Urea Nitrogen			
Below 20*	0	1	
Above 20	1.04	2.83 (2.66 – 3.00)	< 0.001
*denotes reference group			

It is evident that the odds of having prolonged stay increases with age. The oldest age group (age above 85) had an OR of 1.93 (95% CI 1.73 - 2.13), which implies that a patient who

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

is 85 and above will be 1.93 times as likely to have a prolonged stay compared to a patient in the reference group (age 16 to 54). Patients who are admitted via emergency (1.35 [1.20 - 1.50]) and urgent (1.40 [1.07 - 1.73]), overweight (1.26 [1.06 - 1.46]), obese (3.46 [3.26 - 3.66]), have heart disease (1.30 [1.20 - 1.40]), kidney disease (4.22 [4.11 - 4.33]) and stroke (2.35 [2.11 – 2.59]) have higher odds of having prolonged stays.

For laboratory test results variables, the normal level is used as the reference group to allow for comparisons. For most of the laboratory test results, the moderate and severe groups have higher odds of having a prolonged stay compared to the normal group and severe groups have higher odds than moderate groups. This is especially so for variables like glucose above 250 (10.9 [10.5 – 11.3]), sodium below 135 (8.76 [8.12 – 9.38]), potassium above 5.0 (5.42 [4.90 – 5.82]), chloride above 107 (5.87 [5.40 – 6.34]) and BUN above 20 (2.83 [2.66 – 3.00]).

The variables age above 85, obese, kidney, stroke, glucose, sodium, potassium, chloride, BUN seem to be the most important predictors in this logistic regression model as they have larger absolute coefficients compared to the other predictors.

4.3.2 Implementation of the Random Forest

A Random Forest technique was implemented using the randomforest package in R. The down-sampling method is the best method to overcome class imbalance amongst the four methods.

The variable importance table (Table 4) shows that the least important variables are retinopathy,
hypertension and foot. After removing these unimportant variables, the model is tested again, but the performance did not improve. The prediction results are summarized in Table 4.

Table 4: Mean Decrease GINI for Random Forest

Variable	Importance (out of 100)
Chloride	52.4
Kidney	51.0
BUN	45.6
Glucose	41.0
Sodium	33.7
Potassium	33.6
HA1C	29.8
BMI	26.9
Age	24.6
Admission type	23.7
Stroke	23.3
Heart	11.9
Hyperlipidemia	14.1
*Retinopathy**	3.3
*Hypertension**	1.7
*Foot**	1.5
**denotes removed variables*	

Table 5: Summary of Prediction Results using Random Forest

Models	Overall Accuracy	Sensitivity	Specificity	AUC

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

Original	**0.878**	**0.376**	**0.967**	**0.830**
Down-sampling	**0.750**	**0.809**	**0.737**	**0.853**
Down-sampling (only significant variables)	**0.738**	**0.800**	**0.727**	**0.848**

4.4 Summary of Prediction Results

Table 6 Summary of the Best Performance

Model		Overall Accuracy	Sensitivity	Specificity	AUC
Logistic Regression	**ROSE**	**0.766**	**0.819**	**0.757**	**0.867**
Random Forest	Down-sampling	0.750	0.809	0.737	0.853

Models

The Logistic Regression has the best performance in terms of overall accuracy and area under curve. Therefore, the logistic regression model performed the best overall.

It is important to recognize the main drivers for prolonged stay for these models to be used in the clinical setting. Even though the four models had different prediction results, they are relatively consistent in terms of the most important risk factors. The most important variables in these four models are: age, BMI, kidney, chloride, glucose, sodium, potassium, BUN (blood urea nitrogen). These observations are consistent with previous research done. [17,25,26,30].

5. DISCUSSION

5.1 Clinical Application of Prediction Model

Clinicians can make use of the best performing model, the logistic regression model, to identify patients at risk of having a prolonged stay.

Suppose there are three diabetic patients that go into the ICU with different risk factors. (Figure 6) Patient A is relatively young, was admitted via urgent care, is overweight, and has some non life-threatening chronic conditions. By inputting the weights from table 5 into the logistic function, patient A is predicted to have a 54% chance of having a prolonged stay and is hence classified as longstay.

Patient B is old, admitted via an emergency, and has more serious chronic complications such as kidney disease, heart disease and stroke. The laboratory test results also seem to suggest that patient B is in a more serious state than patient A. Patient B is predicted to have a 99% chance of having a prolonged stay.

Lastly, patient C is in an evidently less serious condition compared to patient A and B as suggested by the less serious chronic conditions and laboratory test results. Patient C has a 33% chance of having a prolonged stay and is hence classified as normal.

The logistic regression model can not only classify patients into the different outcomes but also allows clinicians to see the estimated probability of these patients having a prolonged stay. Even though patients A and B have similar outcomes (longstay), the probability of them having a prolonged stay is very different due to the difference in the severity of their conditions. Hence, this allows clinicians to possibly devise a more intensive treatment plan for patients B compared to patient A.

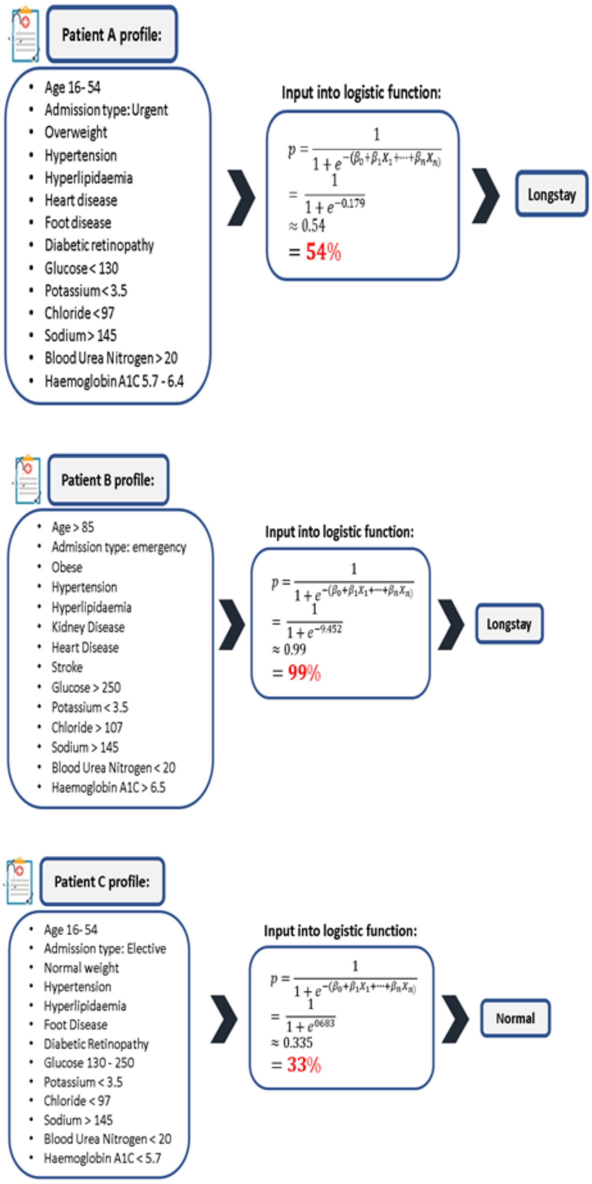

Figure 6: Patient Profiles for three different patients

5.2 Impact of Early Identification on High-risk Patients

5.2.1 Better Care and Management of Disease

Since patients who are at risk of having prolonged stay generally have more serious conditions, early and accurate identification is crucial in allowing them to have better care and management of their disease. This is especially important for patients with diabetes since complications due to diabetes are serious and potentially life-threatening, like heart attack or stroke.

In the context of Singapore, which has the second highest proportion of diabetics among developed nations according to a report in 2015 by the International Diabetes Federation (IDF), the early identification of high-risk patients is extremely crucial. Currently, there are about 450000 Singaporeans with diabetes [50]. This figure is expected to rise due to the aging population and increasing sedentary lifestyles, and the Ministry of Health (MOH) has estimated that this chronic disease could potentially affect one million Singaporeans by 2050 [51]. In fact, MOH has already declared a war on diabetes and is setting up a national taskforce to address prevention, screening and management of this disease.

Hence, early identification of patients at risk of prolonged stay can help in the management of diabetes significantly. Appropriate risk management plans including close follow-up and monitoring may be an effective way of preventing or delaying complications, which will alleviate patients' pain and suffering. Reduction of these diabetes-related complications would reduce direct health cost by decreasing the frequency of hospital stays.

5.2.2 Mitigate Bed Shortage Problem

Reducing LOS can also mitigate the bed shortage problem in Singapore. With the aging population in Singapore, the demand for healthcare is rapidly growing and more admissions can be expected. Even with the Singapore government increasing its health expenditure yearly [52] and investing in more healthcare infrastructure such as hospitals to provide more beds, bed crunch still exists in public hospitals. Singapore's bed density ratio, or the number of beds per 1000 people, is currently at 2.4, which is significantly lower than neighbouring countries like Korea and Japan that is similarly facing aging population [53].

By identifying patients that are expected to have prolonged stay, hospitals can plan ahead and have more efficient resource management. They can target these high-risk patients for more intensive treatment and change or devise new treatment plans if necessary, to cater to the needs of these patients. It is hoped that by doing this, these prolonged stay patients can recover faster and get discharged from the ICUs to the general wards or other healthcare facilities, freeing up beds in the ICUs for other critically-ill patients. This allows higher bed turnover rates in the ICUs and mitigate the bed shortage problem.

5.2.3 Better Planning and Management of ICU Resources

Identification of patients with prolonged stay can allow hospitals to have better management of limited resources. These patients who are likely to have prolonged stays can be assessed for transfer to a long-term acute care facility [54], referred for early mobility therapy to restore physical functionality [55], early discharge planning [56] and palliative care consultation [57]. These interventions allow for early planning and management of resources. They also allow for a more seamless transfer of patients from hospitals to relevant health institutions and at the same time reduce their LOS.

5.3 Cost Saving for Hospitals and Government

Critical care is a costly component of the hospital budget, comprising of about 20% of hospital costs. These costs are largely explained by the need for complex equipment, and round-the-clock care by highly-trained medical professionals.

It is in the interest of hospitals to reduce cost, especially in the costly ICUs. The following shows some calculation on the potential cost savings that a reduction in LOS can bring, in the context of Singapore.

As data for ICUs in Singapore is not published, only estimations from reputable academic sources can be made. A study by Dr Shahla Siddiqui, a consultant at Khoo Teck Puat Hospital (KTPH) [58], published by the Indian Journal of Critical Care Medicine, involved approximately 2880 annual ICU admissions in a 500 bedded public hospital in Singapore. Using the number of beds as a measure of hospital size [59], the total number of annual ICU admissions can be estimated by the formula below. This amounts to approximately 49300 annual ICU admissions in Singapore.

Annual ICU admissions in Singapore (estimated)

$$= \frac{total\ number\ of\ beds\ in\ Singapore\ *}{550}\ X\ 2880$$

$$= \frac{9418}{550}\ X\ 2880 \approx 49300$$

$*obtained\ from\ latest\ data\ from\ Department\ of\ Statistics\ Singapore$ [65]

Since there is no data published on the proportion of the diabetic patients in the ICUs in Singapore, only estimations on the general population can be made (for illustration purposes). According to the MIMIC-III database, the number of prolonged stay patients in the general population makes up about 15% of the total population. In the context of Singapore, this amounts to approximately 7395 ICU admissions with prolonged stay per year.

Annual number of patients with prolonged stay in Singapore (estimated)

$$= \frac{15}{100}\ X\ 49300 \approx \mathbf{7395}$$

The estimated cost savings from a one-day reduction in the LOS of a single patient is equivalent to taking the difference between the cost per day in the ICU and that in the general ward. According to the inpatient charges of the National University Hospital, a reputable hospital in Singapore, the charges per day in the ICU is $834.60, while that in the general ward is $240.75. The difference is $593.85. The total cost saving for a one-day reduction

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

in the LOS for all prolonged stay patients is be estimated to be $4,391,520.75 (calculations summarised in Table 27) This shows that hospitals will be able to save millions of dollars if they are to focus on reducing LOS for patients with prolonged stay.

Table 7: Annual Cost savings calculation for a 1-day reduction in LOS

Annual Cost Savings Calculation	
Median Days in Prolonged Stay in ICU (≥ 7 days)	12
Estimated total number of ICU admissions per year	49300
Estimated total number of prolonged ICU stays per year	7395
Difference in charges between ICU and general ward/ day	$593.85*
Total Cost Saving of 1-day Reduction in LOS in ICU	**$4,391,520**
**From Inpatient Charges of National University Hospital in Singapore [66]*	

At the same time, reducing LOS also allows patients to save on their medical expenses and also for the government to save on the heavy subsidies provided to these patients. All in all, reducing the length of stay of prolonged stay patients not only help hospitals to save cost but also the patients and government.

6. CONCLUSION AND FUTURE ENHANCEMENT

This project aims to identify Type 2 Diabetes Mellitus patients who are at risk of having prolonged ICU stays. Risk factors were first examined from previous research papers. A set of appropriate factors including patient information, presence of chronic conditions and laboratory test results was taken from the MIMIC-III dataset and analysed. Variables with too many missing values as well as insignificant variables were removed and a final set of variables were used as predictor variables in four different machine learning methods. Four methods were used to overcome class imbalance. The logistic regression model with ROSE had the best performance with: AUC 0.867, overall accuracy 0.766, sensitivity 0.819 and specificity 0.757.

It has been found that the risk factors age, BMI, kidney, chloride, glucose, sodium, potassium, BUN (blood urea nitrogen) are crucial in stratifying the patients into longstay and normal groups.

It is hoped that this project has clinical utility in helping hospital identify patients with prolonged stay at admission. Doctors can possibly use this prediction model on top of their medical knowledge and experience to identify these high-risk patients and devise the best treatment plan for them. The timely identification of these patients provides the opportunity for intervention and improvement in clinical outcome, thus reducing the length of stay. Furthermore, it also helps hospitals and ICU in the planning and management of resources.

Other machine learning algorithms like neural networks and support vector machine can be used to improve the performance of the prediction model. In addition, a possible future improvement will be to predict prolonged stay patients not only at admission to the ICU, but during their stay. This helps hospital to evaluate the improvement in the condition of the patients, and devise new treatment plans if necessary in a bid to reduce the length of stay of patients.

REFERENCES

1. Tabish SA. Is Diabetes Becoming the Biggest Epidemic of the Twenty-first Century? International Journal of Health Sciences 2007 Jul; 1(2): V–VIII.
2. Frank B. Hu. Globalization of Diabetes: The role of diet, lifestyle, and genes. Diabetes Care 2011 Jun; 34(6): 1249-1257
3. American Diabetes Association. The Cost of Diabetes. (2018 March) Retrieved from

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

http://www.diabetes.org/advocacy/news-events/cost-of-diabetes.html
4. World Health Organisation. About Diabetes. Retrieved from https://www.who.int/diabetes/action_online/basics/en/index3.html
5. Rosenthal MJ, Fajardo M, Gilmore S, Morley JE & Naliboff BD. Hospitalization and mortality of diabetes in older adults. A 3-year prospective study. Diabetes Care 1998; 21(2):231-5.
6. Naliboff BD, Rosenthal M. Effects of age on complications in adult-onset diabetes, Journal of the American Geriatrics Society 1989 Sep; 37 (9):838-42.
7. Holman N, Hillson R, Young RJ. Excess mortality during hospital stays among patients with recorded diabetes compared with those without diabetes. Diabetes Medicine 2013; 30: 1393-1402.
8. Weissman C. Analyzing the impact of long-term patients on ICU bed utilization. Intensive Care Medicine 2000; 26(9):1319-25.
9. Kramer AA, Zimmerman JE. A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. BMC Medical Informatics and Decision Making. 2010; 10:27
10. Norris C, Jacobs P, Rapoport J. ICU and non-ICU cost per day. Canadian Journal of Anesthesia 1995; 42:192–196
11. Chalfin DB, Cohen IL, Lambrinos J. The economics and cost-effectiveness of critical care medicine. Intensive Care Medicine 1995; 21:952– 961
12. Bo S, Ciccone G, Grassi G, Gancia R, Rosato R, Merletti F, Pagano GF. Patients with type 2 diabetes had higher rates of hospitalisation than general population. Journal of Clinical Epidemiology 2004; 57(11):1196-201.
13. Rapoport J1, Teres D, Zhao Y, Lemeshow S. Length of Stay Data as a Guide to Hospital Economic Performance for ICU Patients. Medical Care 2003 Mar; 41 (3):386-97.
14. Weissman C: Analyzing intensive care unit length of stay data: Problems and possible solutions. Critical Care Medicine 1997; 25: 1594-1600.
15. Ilona W. M. Verburg, Nicolette F. de Keizer, Evert de Jonge, Niels Peek Comparison of Regression Methods for Modeling Intensive Care Length of Stay. PLoS One. 2014; 9(10): e109684
16. Luiz Felipe de Campos Lobato (Risk factors for prolonged length of stay after colorectal surgery. Journal of Coloproctology 2013; 33(1): 22- 27
17. Nirantharakumar K, Hemming K, Narendran P, Marshall T, Coleman JJ. A Prediction Model for Adverse Outcome in Hospitalized Patients With Diabetes. Diabetes Care 2013; 36(11):3566-72
18. Martin CM, Hill AD, Burns K, Chen LM. Characteristics and outcomes for critically ill patients with prolonged intensive care unit stays. Critical Care Medicine 2005; 33: 1922-1927
19. Stricker K, Rothen HU, Takala J. Resource use in the ICU: Short- vs. long-term patients. Acta Anaesthesiologica Scandinavica 2003; 47:508–515
20. Desse TA, Eshetie TC. Determinants of Long Hospital Stay among Diabetic Patients Admitted with Diabetic Ketoacidosis at Jimma University Specialized Hospital. Journal of Traumatic Stress Disorders & Treatment 2016; 6:1
21. Becker GJ, Strauch GO, Saranchak HJ. Outcome and cost of prolonged stay in the surgical intensive care unit. Achieves of Surgery 1984; 119:1338–1342
22. Trottier V, McKenney MG, Beninati M, Manning R, Schulman CI. Survival after prolonged length of stay in a trauma intensive care unit. Journal of Trauma and Acute Care Surgery. 2007, 62:147-150.
23. MIMIC-III Critical Care Database. Retrieved from

https://mimic.physionet.org/about/mimic/
24. Gamsjäger T. Hospital length of stay: Analysis across a nationwide healthcare system. 2015
25. Khalid JM, Raluy-Callado M, Curtis BH, Boye KS, Maguire A, Reaney M. Rates and risk of hospitalisation among patients with type 2 diabetes: retrospective cohort study using the UK General Practice Research Database linked to English Hospital Episode Statistics International Journal of Clinical Practice. 2014 Jan; 68(1):40-8.
26. Moss SE, Klein R, Klein BE. Risk Factors for Hospitalization in People with Diabetes. Achieves of Internal Medicine. 1999 Sep;159(17):2053-7.
27. Kirkman MS, Briscoe VJ. Diabetes in Older Adults. Diabetes Care 2012; 35(12): 2650-2664
28. Al-Goblan AS, Al-Alfi MA, Khan MZ. Mechanism Linking Diabetes Mellitus and Obesity. Diabetes, Metabolic Syndrome and Obesity 2014; 7:587-5
29. Attvall S, Fowelin J, Lager I, Von Schenck H, Smith U. Smoking induces insulin resistance: a potential link with the insulin resistance syndrome. Journal of Internal Medicine 1993; 233:327–332
30. Wijayaratna SM, Cundy T, Drury PL, Sehgal S, Wijayaratna SA, Wu F. Association of type 2 diabetes with prolonged hospital stay and increased rate of readmission in patients with lower limb cellulitis Internal Medical Journal 2017 Jan; 47(1):82-88.
31. Center for Disease Control and Prevention. Smoking and Diabetes. Retrieved from https://www.cdc.gov/tobacco/campaign/tips/diseases/diabetes.html
32. Mayo Clinic. Diabetes Ketoacidosis. Retrieved from https://www.mayoclinic.org/diseasesconditions/diabetic-ketoacidosis/symptoms-causes/syc-20371551
33. Sowers JR, Khoury S, Standley P, Zemel P, Zemel M. Mechanisms of Hypertension in Diabetes. American Journal of Hypertension 1991 Feb;4:177-82.
34. Goldberg RB. Lipid Disorders in Diabetes. Diabetes Care 1981; Sep-Oct;4(5):561-72
35. American Heart Association. Cardiovascular Disease and Diabetes. Retrieved from https://www.heart.org/en/health-topics/diabetes/why-diabetes-matters/cardiovascular-disease--diabetes
36. Giuseppe MC Rosano, Cristiana Vitale, Petar Seferovic. Heart Failure in Patients with Diabetes Mellitus. Cardiac Failure Review 2017 Apr; 3(1): 52–55.
37. National Kidney Foundation. Diabetes- A Major Risk Factor for Kidney Disease. Retrieved from https://www.kidney.org/atoz/content/diabetes
38. Mayo Clinic. Diabetic Retinopathy. Retrieved from https://www.mayoclinic.org/diseasesconditions/diabetic-retinopathy/symptoms-causes/syc-20371611
39. American Diabetes Association. Checking Your Blood Glucose. Retrieved from http://www.diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/checkingyour-blood-glucose.html

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

40. What are the risk of the sodium blood test? https://www.healthline.com/health/sodiumblood#risk-factors
41. Liamis G, Tsimihodimos V, Elisaf M. Hyponatremia in Diabetes Mellitus: Clues to Diagnosis and Treatment. 2015
42. Hyperkalaemia and Kidney Dysfunction. Retrieved from https://www.medicinenet.com/hyperkalemia/article.htm#hyperkalemia_and_kidney_dysfunction
43. Hypochloraemia (Low Chloride). Retrieved from http://chemocare.com/chemotherapy/sideeffects/hypochloremia-low-chloride.aspx
44. Creatinine (Low, High, Blood test results explained). Retrieved from https://www.medicinenet.com/creatinine_blood_test/article.htm
45. Blood and Urine tests. Retrieved from https://lifeoptions.org/learn-about-kidney-disease/bloodand-urine-tests/
46. Baek HY, Cho M,Kim S, Hee H, Song M, Yoo,S. Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. PLoS One 2018;13(4):e0195901
47. Exploratory regression and model selection. Retrieved from http://www.researchtraining.net/addedfiles/2011aManchester/2011ManchesterModelSelection.pdf
48. Ministerial Conference on Diabetes to Tackle Global Diabetes Epidemic. 2018, Nov. Retrieved from https://www.moh.gov.sg/news-highlights/details/ministerial-conference-ondiabetes-to-tackle-global-diabetes-epidemic
49. New protocol helps tell when patients are ready for oral feeding. 2017 April 19. Retrieved from https://www.nuh.com.sg/wbn/slot/u3007/Patients%20and%20Visitors/Newsroom/Media%20Articles/2017/APR_2017/TODAY_NewProtocolOralFeeding_190417-Page_20-21.pdf
50. Eskildsen MA: Long-term acute care: A review of the literature. J Am Geriatr Soc. 2007, 55:775-779
51. Needham DM. Mobilizing patients in the intensive care unit: Improving neuromuscular weakness and physical function. JAMA. 2008, 300: 1685-1690
52. Kleinpell RM: Randomized trial of an intensive care unit based early discharge planning intervention for critically ill elderly patients. American Journal of Critical Care. 2004, 13: 335-345.
53. Norton SA, Hogan LA, Holloway RG, Temkin-Greener H, Buckley MJ, Quill T: Proactive palliative care in the medical intensive care unit: Effects on length of stay for selected high risk patients. Crit Care Med. 2007, 35: 1530-1535
54. Mallick R, Strosberg M, Lambrinos J, Groeger J S. The intensive care unit director as manager: Impact on performance. Med Care 1995; 33:611-24.
55. Shahla Siddiqui Mortality profile across our Intensive Care Units: A 5-year database report from a Singapore restructured hospital. Indian J Crit Care Med. 2015 Dec; 19(12): 726–727
56. Giancotti M, Guglielmo A, Mauro M. Efficiency and optimal size of hospitals: Results of a systematic search PLoS One. 2017 Mar 29;12(3):e0174533
57. https://www.tablebuilder.singstat.gov.sg/publicfacing/createDtaTable.action?refId=15276

January - February 2020
ISSN: 0193 - 4120 Page No. 6143 - 6163

58. https://www.nuh.com.sg/patients-and-visitors/appointments/hospital-charges/inpatientcharges.html
59. https://data.gov.sg/dataset/government-health -expenditure?view_id=cdc03adc-b1b0-4eaa-99e2 269b174d1ef4&resource_id=cf7b1696-9b0e-425d-a96a-e61c41629623
60. https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=15276