Predicting Perioperative Outcomes from Surgical Data During One Lung Ventilation

by

Atif Ul Aftab

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

Department of Computer Science
The University of Manitoba
Winnipeg, Manitoba, Canada
August 2020

Thesis advisor Author

Dr. Michael Domaratzki **Atif Ul Aftab**

Predicting Perioperative Outcomes from Surgical Data During One Lung Ventilation

Abstract

Patients routinely undergo surgeries executed under general anesthesia that require ventilation. Sometimes this mechanical ventilation is not only harmful to health but also a significant cause of death. One lung ventilation (OLV) assisted lung surgery is not free from those complications. During OLV assisted lung surgery, ventilation is applied to the lung opposite the side of the operation, and the other lung (where the surgery will take place) is already damaged. For these reasons, mechanical ventilation procedures may cause acute ventilator-induced lung injuries. These lung injuries can be the reason for different pulmonary and respiratory complications, which may lead to patient death. So it is essential to predict OLV assisted surgery-related outcomes to reduce negative results. I have used machine learning to predict the perioperative outcome. I have used a real-world dataset of OLV assisted lung surgeries collected in Manitoba. The dataset is the real-time olv assisted lung surgery data from 80 patients surgery information. Due to the uneven class distribution, I have used SMOTE for oversampling the data. I divided the dataset into three parts: i) Preoperative Data, ii) Intraoperative Data, and iii) Combined Data. Three different classification algorithms, Random Forests (RF), Support Vector Machines (SVM), and Logistic Regression (LR), have been applied to the different combinations of datasets. Using intra-operative data oversampled by SMOTE using the SVM classification algorithm gives the best accuracy with an F1-score of 0.70 and AUC of 0.61 for the prediction of surgical complication.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my gratitude to Almighty God for granting me the opportunity to write this thesis.

I would like to thank my supervisor, Dr. Michael Domaratzki for giving me the opportunity to work under his supervision. I am very grateful for his guidance, suggestions, and feedback throughout the preparation of this thesis. I am thankful to the other members of my committee, Dr. Biniam Kidane, and Dr. Lorenzo Livi for the comment and advice they provided.

I am also grateful to all my collaborators for their guidance. I am also thankful to all my colleagues from Bioinformatics lab for their valuable consults and time. I am equally grateful to all the benevolent faculties of University of Manitoba Computer Science and Ahsanullah University of Science and Technology who shaped my research and academic standing.

Finally, I would like to thank *my wife and mother* for their countless sacrifices and inspirations throughout the years.

Dedicated to my father Late Aftabuddin Ahmed ...

Chapter 1

Introduction

Every day numerous surgeries take place around the world. Surgeries that are executed under general anesthesia requires artificial ventilation. This artificial mechanical ventilation may sometimes be injurious to health [1]. Most of the time, patients undergoing lung surgery need OLV. During OLV assisted lung surgery, ventilation is applied to the lung opposite the side of the surgery, and the other lung (where the surgery will take place) is already damaged. At the same time, different ventilation settings have different effects. Different OLV surgeries take different operating times also. Longer operating times can have different effects than shorter ones. So OLV may cause acute ventilator-induced lung injury, which can be the reason for different pulmonary and respiratory complications. In some cases, those complications can be a major reason for the death of the patients.

Different ventilation parameters include airway pressures, VT, respiratory rate, the fraction of inspired oxygen (FiO_2), PEEP, continuous positive airway pressure (CPAP), oxygenation and more [2; 3; 4; 5]. So the study of the assessment of the ventilation parameters and their relation with the perioperative complications is necessary. This study can help to reduce negative postoperative outcomes. Different ventilation parameter variations of those parameters have different effects on the postoperative complication.

I have applied machine learning to real-world surgical data of one-lung ventilation. This dataset has real-time surgery information of 80 patients during OLV assisted lung surgery. This dataset was collected by the Section of the Thoracic Surgery of the Department of Surgery, the University of Manitoba under the supervision of Professor Dr. Biniam Kidane. It has preoperative and intraoperative data. Preoperative data has FEV1, DLCO, Charlson Comorbidity Index (CCI), surgery type, etc. Intraoperative data is collected at regular intervals (1 sec, approx) during the surgery and has surgery duration, heart rate, FiO_2, CO_2-ET etc.

I have applied three different classification algorithms–RF, SVM, and LR–with different combinations of features for those patient records. For patients I have predicted complications, normal oxygenation at 4 hours, and normal oxygenation at 12 hours.

The dataset is imbalanced in terms of outcomes. Among 80 patients' records, only 20 records have surgical complications, and 60 records have no surgical complications. For the imbalanced proportion of the classes, I have applied an oversampling technique called Synthetic Minority Over-sampling Technique(SMOTE) [6].

I have also used the grid search cross validation (CV) for hyperparameter tuning [7; 8]. This technique finds the optimum value of hyperparameters for a given model. I have executed different 108 experiments and then compared the results.

Using machine learning with intra-operative data oversampled by SMOTE yielded the best results. Using SVMs, the F1-score for prediction of complications is 0.70.

In chapter 2, I have the background of this thesis. Chapter 3 contains the literature review of this thesis. Chapter 4 and 5 describes the results and conclusion of this thesis.

Chapter 2

Background and Methods

Every day thousands of lung surgeries are taking place around the world. Most of the time, these types of operations need OLV. Due to this mechanical ventilation, acute lung injury takes place. Prediction of postoperative complications due to OLV using machine learning can reduce the negative outcome of those surgeries.

2.1 Surgical Background

In the human body, the two lungs are two separate organs. But they work as a single functional unit of the body for ensuring the proper amount of oxygen and CO2 in the blood. During lung surgery, the function of one lung from the other needs to be separated, which is called lung isolation [9]. In this procedure, an individual lung operates as a single unit with the help of specific instrumentation and keeps breathing isolated in the other lung and not in the lung where the operation will take place. OLV is a technique that is used to complete this procedure [10]. During OLV assisted lung surgery, mechanical ventilation is applied to the lung opposite the side of the other lung where the operation will take place. That means the lung where the operation will take place is excluded from ventilation during the surgery, and all the bronchial functions are conducted by the other lung, which increases the chance of Acute Lung Injury (ALI).

The essential ventilatory variables during OLV are tidal volume (VT), respiratory rate, the fraction of inspired oxygen (FiO_2), positive end-expiratory pressure (PEEP), and continuous positive airway pressure (CPAP). Other variables that are also important during surgery are oxygenation, forced vital capacity (FVC), diffusing capacity of the lungs for carbon monoxide (DLCO), forced expiratory volume (FEV), surgery duration, surgery type, and charlson comorbidity index [2; 3; 4; 5]. The amount of air pushed into and out of the lungs throughout each ventilation period is called the VT [11]. Volumetric portion of the oxygen present in every breath is called the (FiO_2)[12]. After every breath, the pressure applied by the ventilator to ensure the functionality of the lung during the ventilation process is PEEP [13]. FEV is the amount of oxygen one can blow throughout a compelled breath and calculated throughout the first (FEV1), second (FEV2), and third (FEV3) seconds of that breath [14]. Throughout the FEV process, the cumulative volume of oxygen breathed is called FVC [14]. The DLCO indicates the extremity of parenchymal lung disease [15]. Charlson Comorbidity Index is a predictor of surgical mortality [16].

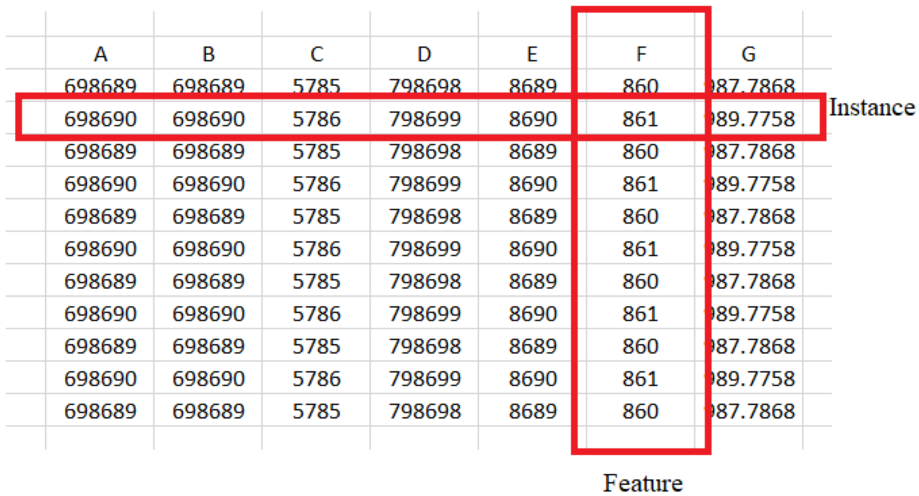
Many researchers have suggested different optimum levels of those variables and ventilatory management settings during OLV for reducing negative outcomes. But still, it's difficult to identify the optimum level for different patients. Sometimes applying different levels of ventilation variables damage the healthy lung even if the ventilation period is short [17]. Alveolar damage [18], postpneumonectomy pulmonary oedema [19], and various pulmonary complications [20] occur in different patients depending on different health conditions, ventilatory settings, and operational stress. All of those parameters vary over time and have different interactions between them. At the same time, different OLV surgeries take different operating times also. Longer operating times can have different effects than shorter ones.

2.2 Machine Learning Background

The main aim of this thesis is to build a machine learning predictor that will predict postoperative complications and delayed return to normal oxygenation from the ventilation parameters during OLV. The machine learning approaches are used to identify the relationship between ventilation parameters and perioperative outcomes. Different machine learning classification algorithms are applied to the different data sets to compare the results between them. These tools will be described in detail in chapter 4.

2.2.1 Machine Learning Methods and Data

Machine learning is the advanced utilization of Artificial Intelligence (AI). It learns dynamically from data with the help of different algorithms and improving performance from experience. So based on a dataset, the algorithm makes its decision. But the decision is more like a tool to be able to make predictions on new, unseen data. Datasets are a collection of data. Each datasets can be viewed as being in rows and columns. Figure 2.1 is an example of a dataset. Each row of a dataset is an instance, and each column of the dataset is a feature. Each feature can have different data types, such as numerical, categorical, textual values, etc. A portion



A	B	C	D	E	F	G
698689	698689	5785	798698	8689	860	87.7868
698690	698690	5786	798699	8690	861	89.7758
698689	698689	5785	798698	8689	860	87.7868
698690	698690	5786	798699	8690	861	89.7758
698689	698689	5785	798698	8689	860	87.7868
698690	698690	5786	798699	8690	861	89.7758
698689	698689	5785	798698	8689	860	87.7868
698690	698690	5786	798699	8690	861	89.7758
698689	698689	5785	798698	8689	860	87.7868
698690	698690	5786	798699	8690	861	89.7758
698689	698689	5785	798698	8689	860	87.7868

Figure 2.1: Example of a dataset

of features are used as input of the machine learning model, and there are some variables that are used as the output. Here the machine learning model is a module that is trained to identify different types of patterns. A portion of data from datasets is used as a training dataset used to train the machine learning model. The rest of the data are used as the test or training dataset for evaluating that model.

There are different types of machine learning techniques, such as supervised learning, unsupervised learning, reinforcement learning. In supervised learning, data is pre-categorized or has a numerical value and is a type of learning technique where both input and output data are provided [21].

2.2.2 Classification Algorithms

Random Forests

Random Forests (RF) are used regularly in both classification and regression [22] and are a popular supervised classification technique, which are a combination of decision tree predictors [23]. RFs are a collection of unpruned classification and regression decision trees, which combines the decision of individual decision trees using bagging and feature randomness. An individual decision tree makes a decision by going from the root to a leaf and the entire RF makes a decision by majority vote.

We have an example of decision tree in Figure 2.2. Internal nodes of a decision tree are a test and leaf nodes are the ultimate outcome of evaluating several of these tests. Testing starts at the root node. The uppermost node on the tree is called the root node of a tree. In figure 2.2, if a patient has a DLCO of 100 and we are the root then, it decision tree will follow the next internal node, which checks a threshold of ($\mathrm{FiO_2}$) less than equal to 0.49. Then it will make the decision based on the value of ($\mathrm{FiO_2}$). Thus it will move forward, and the final or leaf node will have the ultimate verdict.

In training, each tree in an RF relies on a subset of the features of an instance sampled independently. In simple terms, RFs connect randomized decision trees and provides an average prediction from them [22]. They construct n individual decision trees where n is a value that the user can adjust (hyperparameter), and each decision tree makes a class prediction. Then they considers all the results of decision trees, and the class with maximum number of correct prediction becomes the final prediction of the model. RFs work efficiently when the number of cases

Figure 2.2: Example of a random decision tree from RFs

is much less than the number of features, which match with our research traits. Random Forests require all individuals to have the same number of features.

Support Vector Machines

A popular useful technique for classification is SVMs [24]. It's popular for obtaining good results in medical diagnostics, optical character identification, forecasting, and other areas [25].

This algorithm is a supervised learning technique requiring a labeled training

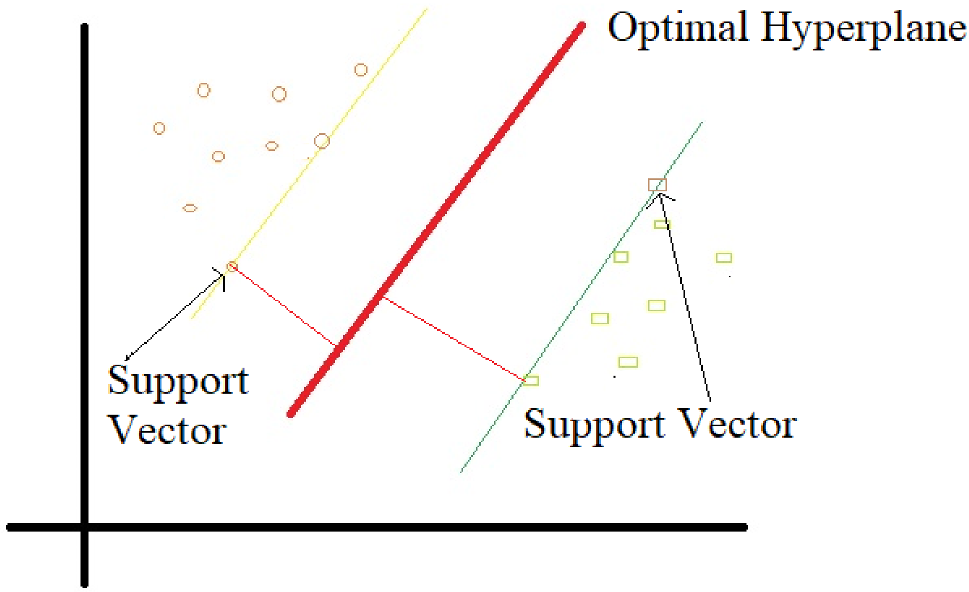


Figure 2.3: A Support Vector Machine

set [26; 27]. This algorithm is also efficient for large data sets and the broad diversity of biological data. SVMs require all data to have the same number of features.

For a given set of input, SVM tries to find a decision surface for separating two classes. In SVMs, data points which are closest to the decision surface are called support vectors. In the most basic case, the decision surface is a hyperplane. SVMs maximizes the boundary between the separating hyperplane and instances and finds an optimal solution among the infinite number of solutions. New instances are classified by the side of the decision boundary that the new instance lies on.

For example, let us consider a dataset with green rectangles and yellow circles illustrated in figure 2.3. This is a 2D case, so the instances have two features and plotted as points in the 2D plane. The objective of SVM is to find an optimal line

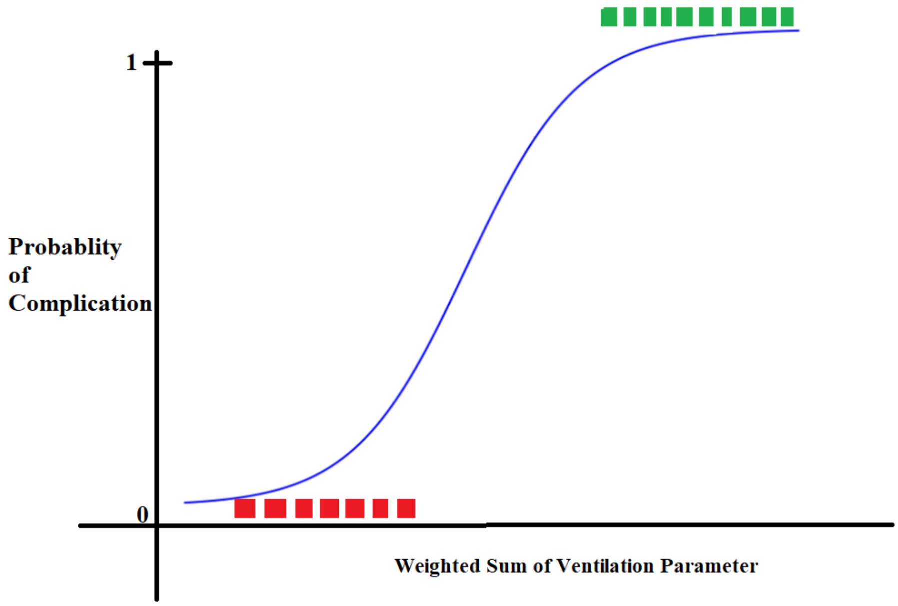


Figure 2.4: Example of Logistic Regression

that divides the dataset into two different classes.

After finding the decision boundary, the SVM finds the nearest points to the decision boundary. These points are considered as support vectors. Secondly, it measures the distance between the line and the support vectors. SVM tries to maximize the gap between them and find out the optimal hyperplane for which it can maximize the distance. We can see how its done in figure 2.3.

Logistic Regression

Logistic Regression is a predictive analysis and used when the outcome class is binary. It builds models based on the logistic function and describes the connection between the binary classification and feature variables. LR assesses the outcome based on previous knowledge or information by identifying the most critical relationship between the different circumstances and the consequences [28].

Let us consider a model where we need to identify the postoperative outcome of surgery for a given set of ventilation parameters as input features. Here the final outcome is a complication and no complication. LR will first calculate the weighted sum of the ventilation parameters. After that, it will input this into the sigmoid function. LR uses sigmoid function as the logistic function. By computing the sigmoid function, LR will get a probability. This probability is then converted to a binary outcome by use of a threshold. For a new instance, the probability of that instance is calculated and then assigned to one of the classes. The formula for the sigmoid function is given below:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

Let us consider an example of LR in figure 2.4, where the x-axis represents the weighted sum of the ventilation parameters, and the y-axis represents the probability of complication. If we plot the mentioned sigmoid function, the graph will be an S curve. Even if the logistic regression model meets an outlier, it will handle it with the help of this logistic function. Here outliers are data points that don't fit a particular data group.

2.2.3 Hyperparameter Tuning

In machine learning, a value which is employed to regulate the training process of a model is called the hyperparameter. Hyperparameter tuning is finding the optimized hyperparameters for a specific model as different machine learning models have different hyperparameters. For finding a machine learning model with the best accuracy, it is essential to supply the optimum value of hyperparameters to that model. I have used Grid Search Cross Validation for hyperparameter tuning [7; 8]. Grid Search CV finds those best hyperparameters by using cross-validation and exhaustively checking with all the possible values of those hyperparameters in that model.

2.3 Dataset

For this thesis, I have used a surgical time-series dataset, which contains data for each second, and general preoperative information for each of the patients. The sample size of the dataset is 80, which means it contains records for 80 patients. This dataset was collected by the Section of the Thoracic Surgery of the Department of Surgery, the University of Manitoba under the supervision of Professor Dr. Biniam Kidane. The dataset has two parts. Those are:

- Preoperative data
- Intraoperative time-series data

We have used four preoperative features: FEV1, DLCO, Charlson Comorbidity Index (CCI), and surgery type for analysis. We have also used four intraoperative features: surgery duration, heart rate, FiO_2, CO_2-ET for analysis.

We have four different outcome variables in the dataset. They are i) complication, ii) normal oxygenation at 4hour, iii) normal oxygenation at 12hour, and iv) normal oxygenation at 24hour. All of them have a binary outcome. Complication means the number of patients having any complications after surgery. There is data on sub-types of complication in the data set but these are very rare. Among 80 patients, 60 patients had no complications after surgery indicated by 0, and 20 patients had complications after surgery reported by 1 in the dataset.

Normal oxygenation at 4 hours means the number of patients are at normal oxygenation at 4 hours. Forty-three patients were in normal oxygenation at 4 hours indicated by 1, and thirty-seven patients were not in normal oxygenation at 4 hours indicated by 0 in the dataset. Similarly, we have data on normal oxygenation after 8, 12 and 24 hours. Sixty-three patients were in normal oxygenation at 8 hours indicated by 1, and seventeen patients were not in normal oxygenation at 12 hours indicated by 0 in the dataset. Seventy-one patients were in normal oxygenation at 24 hours indicated by 1, and nine patients are not in normal oxygenation at 12 hours indicated by 0 in the dataset. The main challenge of the data set is that the class distribution of the dataset is imbalanced. We have missing data in pre-operative records. Two features of pre-operative records have missing data: FEV1 and DLCO. FEV1 has 11, and DLCO has 17 missing records.

2.4 Oversampling and Imputation

Our dataset has uneven class distribution. For complication, only 25% of the data set has complications, and for oxygenation at 12 hrs, only 21.25% of the dataset has not returned to normal oxygenation after 12 hours. If we train a machine learning model with this data set and want to predict from it, then the outcome variable prediction can be poor due to the small minority class. So most of the machine learning models will fail to predict correctly. To resolve this problem, I have used SMOTE [6]. SMOTE offers better classification results by over-sampling the minority class and under-sampling the majority class.

SMOTE oversamples the minority class. SMOTE doesn't replicate the minority class instances. It creates synthetic minority class examples and balances the training dataset. For obtaining equal class distribution, an iterative process is executed [29]. Randomly minority class examples are taken from the training set. Then K-nearest neighbors are taken (by default, K=5). Then instances are randomly selected from the set of neighbours to compute new minority examples. The new minority examples are created by taking the average value between the sampled neighbor and the original sample. SMOTE is a successful technique in different field of operation in data science and considered as a benchmark for learning from imbalanced data [30].

The dataset also has missing data for two preoperative features. For that reason, I have taken the median value of all the values of a particular surgical feature and imputed the missing value with that median value for that particular feature. This technique imputes missing data of the dataset rather than losing the entire patient.

2.5 Metrics for Evaluating Machine Learning Models

Two different metrics are considered for all the experiments of this research. Those are f1-score and AUC. As I have mentioned earlier, our datasets are highly imbalanced. Accuracy only measures all correctly predicted cases, but it doesn't deal with the incorrectly classified cases. Accuracy is a good measure when the dataset has an equal distribution of classes. In our cases, our dataset has a highly imbalanced class distribution. So accuracy will not be enough to evaluate the model. That is why I have used f1-score and AUC. For imbalanced class, f1-score is more insightful than accuracy. F1-score ranges between 0 and 1. It measures the harmonic mean of precision and recall. Here precision is the ratio of correctly predicted positive instances and all predicted positive cases, and the recall is the ratio of correctly predicted positive examples and all positive of the dataset. The formula for precision is $\frac{TruePositives}{(TruePositives+FalsePositives)}$ and the formula for recall is $\frac{TruePositives}{(TruePositives+FalseNegatives)}$. It is better to evaluate a model with f1-score as this score uses both false negative and false positive.

Area Under Curve (AUC) is a metric for evaluating performance for classification problems under different conditions. It shows the performance of a model for distinguishing between classes. The value of AUC is ranges from 0 to 1. A higher value of AUC indicates a more accurate machine learning model.

2.6 Software Used for the Research

For data preprocessing and prediction, I have used python 3.0 scripting language and sckit-learn [7; 8]. I have used different libraries from sckit-learn such as simple imputer, RF, SVM, LR, feature importance from RF, Grid Search CV, pyplot. SMOTE has been used for the oversampling technique. Simple imputer imputes the missing value by taking the median of all the values. RF, SVM, LR are used for machine learning classification and prediction. Feature importance finds the importance of the features in RF classification. Grid Search CV is used for hyperparameter tuning of the model. SMOTE has been used from imbalanced-learn [31] library of python for oversampling process. Pyplot has been used from Matplotlib [32] library of python for the visualization of different graphs[1].

[1]https://github.com/atifulaftab/Predicting-Perioperative-Outcomes-from-Surgical-Data-During-OLV

Chapter 3

Literature Review

OLV assisted lung surgeries are dependent on mechanical ventilation which involves different ventilation parameters such as VT, respiratory rate, the fraction of inspired oxygen, PEEP, CPAP and more [2; 3; 4; 5]. The prediction of clinical outcomes based on these different parameters are very essential for reducing the negative outcomes.

In 2009, Pardos et al. [33] assessed the connection between the ventilatory mode implemented during OLV and intraoperative and early postoperative arterial oxygenation when patients are undergoing lung surgery. The study size was one hundred and ten thoracic surgeries with duration of at least one hour. They observed that arterial oxygenation during OLV or early postoperative oxygenation is not affected when using pressure-controlled ventilation (PCV) compared with volume-controlled ventilation (VCV) with the same VT. However the paper was not a predictive analysis.

Blank et al. [34] discussed the management of one-lung ventilation. They assessed the use of ventilator management during surgery involving OLV and measured the result on clinical consequences. They have used multivariate linear regression to assess the relationships between primary and secondary clinical outcomes and six ventilator parameters such as VT, PEEP, plateau ventilator pressures(Pplat), inspired oxygen fraction, expiratory carbon dioxide, and respiratory rate etc. for 1019 cases. Hospital electronic records and the Society of Thoracic Surgery database are the sources of dataset. They found that a high number of patients go through high VT during OLV. They also identified that significant postoperative symptoms and respiratory complications were inversely associated with this VT. After analyzing the situation they advised that maintaining low VT is not capable of blocking postoperative respiratory complications without a sufficient amount of PEEP, though physiologically low VT is an essential element throughout OLV. The research work is a a ventilator management study and not a predictive analysis.

Colquhoun et al. [35] also conducted a multicenter study and reviewed the various courses of action in the ventilation method for patients experiencing OLV. They have used 5609 patient's different ventilation parameter records which includes VT, driving pressure, PEEP, etc., during analysis. They used the Multicenter Perioperative Outcomes Group database [35; 36; 37] which helped to distinguish different patients undergoing one-lung ventilation. They observed that a substantial proportion of patients who were female, obese, or of short stature included in high-risk subgroups were ventilated with less than suggested VT. They found that most of the patients receive different VT and PEEP level than the recommended level. The research is also a ventilator management study and not a predictive analysis

Belda et al. [38] conducted a descriptive multicenter national study on lung surgery patients experiencing one-lung ventilation in 2018. This study was conducted with data of 690 patients who underwent thoracic surgery with OLV. They observed that OLV has a lower rate of postoperative pulmonary complications (PPCs) for the patients undergoing OLV with open lung approach. A personalized open-lung approach includes an alveolar recruitment maneuver- an enhanced transpulmonary pressure to excavate collapsed alveoli [39] accompanied by an adjustment of PEEP to the best respiratory system. But their goal was not the prediction of any postoperative pulmonary complications. The main objective of this paper was to give an idea of patients having PPCs who underwent through open-lung approach during OLV. They actually evaluated the performance of one lung approach in this research.

Shen et al. [40] investigated the relationship between one primary ventilation parameter, VT, and pulmonary complications. This research was conducted among 101 patients who experienced left-lung ventilation during thoracoscopic esophagectomy during June 2011 and July 2012. They found that intraoperative OLV causes lung injury in addition to pulmonary complications after minimally invasive esophagectomy (MIE). They also identified that low average VT reduces ventilation-associated lung complications.

Neto et al. [41] assessed the relationship of VT, the level of PEEP, and driving pressure during mechanical ventilation with postoperative complications. The sample size for this research was 2250. They conducted a meta-analysis on data from patients who experienced general anesthesia. Their main intention was to identify mechanical ventilator-associated postoperative pulmonary complications. They identified that when patients experience postoperative pulmonary complications, those complications are related to intraoperative high driving pressure and changes in the level of PEEP that increase driving pressure. But they did not authenticate the validity of those facts experimentally and suggested some approaches to confirm the validity of those facts.

Ball et al. [42] studied and examined intraoperative ventilation parameters and their relationships with PPCs in overweight patients. They performed a secondary analysis which was only limited to 2012 obese patients. They found that obese patients are often ventilated with high VT and low PEEP which causes postoperative pulmonary complications. Their study was limited to only obese patients and the association between ventilation settings and postoperative complications. But they failed to evaluate any specific casualty.

Okahara et al. [43] investigated the relationship among intraoperative ventilator settings with PPCs after thoracic surgery. They worked with detailed surgery information, preoperative characteristics and information of patients and anesthesia information from 197 patients. They observed that higher $\mathrm{FiO_2}$ during one-lung ventilation might be correlated with the increased rate of PPCs. In their study, they failed to determine the causality of the complications.

After analyzing the above pieces of literature, we can identify that the researchers have identified some parameters including VT and PEEP that are linked to surgical outcomes. But there is no reliable and recognized way to predict postoperative pulmonary complications for ventilation parameters during OLV.

Chapter 4

Results and Discussion

4.1 Techniques

4.1.1 Data Processing

Preoperative and intraoperative information was extracted from OLV surgery information by the Section of the Thoracic Surgery in the Department of Surgery, the University of Manitoba, under the supervision of Professor Dr. Biniam Kidane. There are two raw datasets available for this research. Those are:

- Preoperative data
- Intraoperative time-series data

Though pre-operative varibales are widely used for predicting post-operative complications[44; 45; 46; 47], as a thoracic & foregut Surgeon, Professor Dr. Biniam Kidane suggested some critical both pre-operative and intra-operative features which they emphasize during OLV assisted lung surgery. Tables 4.1 and 4.3 denote those features or variables considered from the raw preoperative and intraoperative time-series datasets. Table 4.4 represents the outcome variable for these

Variable Name	Value Range
Surgery Type	0 to 5
FEV1	31 to 143
DLCO	48 to 118
Charlson Comorbidity Index (CCI)	0 to 9

Table 4.1: Preoperative Data

Surgery Type ID	**Surgery Type Name**
0	VATS Wedge Resection
1	VATS Lobectomy
2	Open Segmental Resection
3	Open Lobectomy
4	Right Extraplerual Pneumonectomy
5	Open LN biopsy

Table 4.2: Surgery Types

datasets, which contains the binary value (0 or 1). For complication outcome variables, 0 represents no complication, and 1 represents the complication. For normal oxygenation after 4, and 12 hours outcome variables, 0 represents abnormal oxygenation, and 1 represents normal oxygenation.

Each intraoperative feature was recorded as continuous time-series data during the entire OLV surgery. Each second of surgery records are kept as time series data in the dataset. Figure 4.1 shows the histogram for no. of patients for particular surgery duration in seconds. The mean surgery length was 7960.6875 seconds or 132.67 minutes. We have taken the median of all the values of each intraoperative feature for each patient. The primary motivation was representing the whole scenario of a particular feature during the surgery as a single numerical feature. We have also extracted the max value of all the intraoperative features for

Variable Name	Type
Heart Rate	Numerical
The Fraction of Inspired Oxygen (FiO_2)	Numerical
Carbon Dioxide Equivalent (CO_2 ET)	Numerical
Arterial Blood Pressure Mean (ABP-Mean)	Numerical
Arterial Blood Pressure Systolic (ABP-SYS)	Numerical
PEEP	Numerical
Ppeek	Numerical
Surgery Time	Numerical

Table 4.3: Intraoperative Data

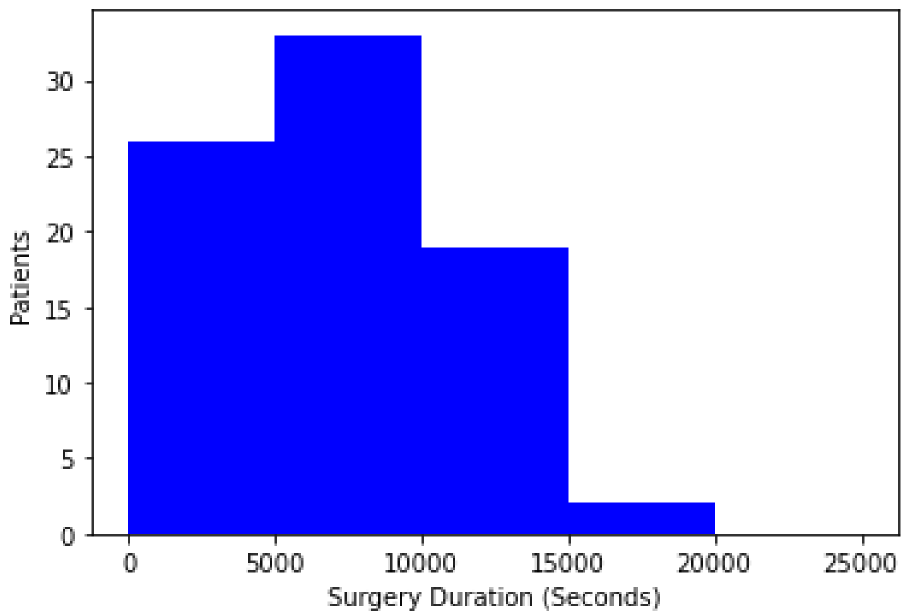
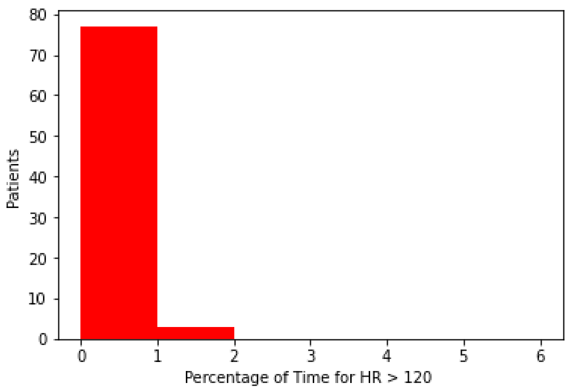
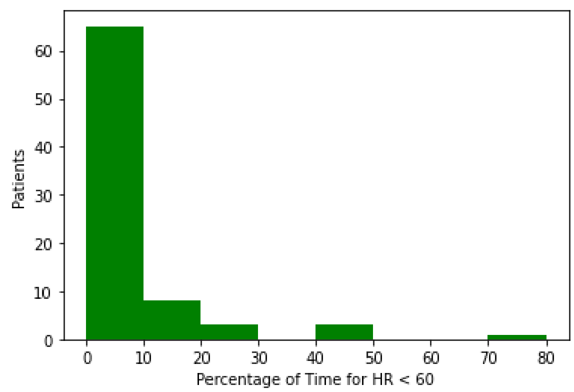


Figure 4.1: Histogram Visualization for Surgery Duration

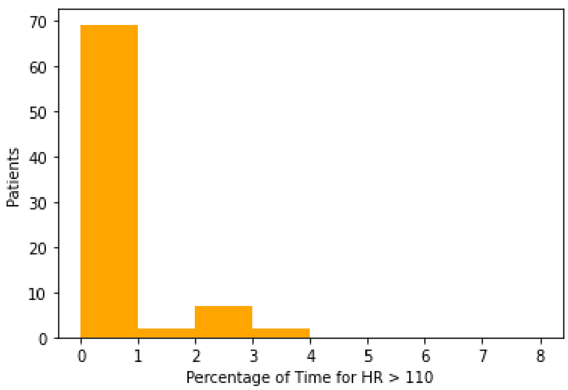
rendering the highest peak of the feature. We didn't consider the minimum value for any feature as most of the time, the minimum value of a particular feature was 0.



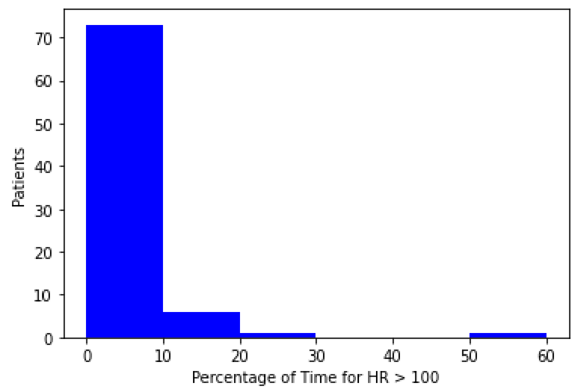
(a) $HeartRate > 120$



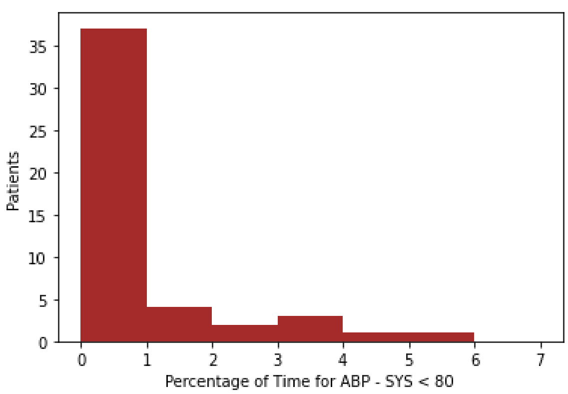
(b) $HeartRate < 60$



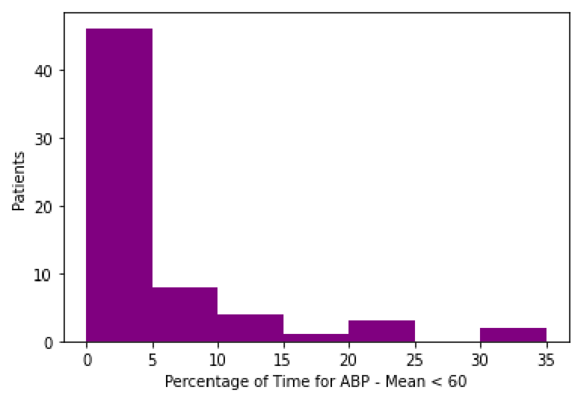
(c) $HeartRate > 110$



(d) $HeartRate > 100$

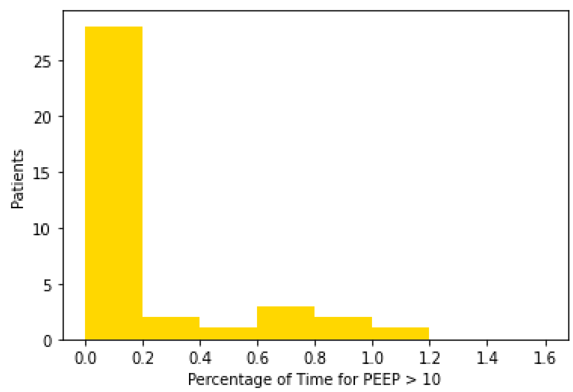


(e) $ABP - SYS < 80$

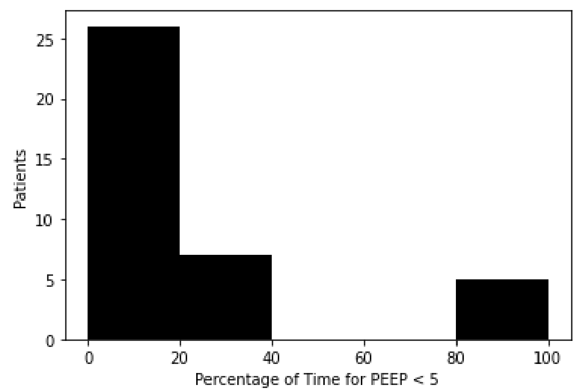


(f) $ABP - MEAN < 60$

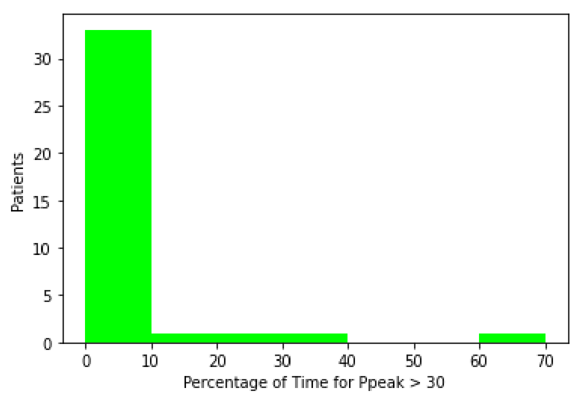
Figure 4.2: Histogram Visualization for Percentage of time of different ventilation parameters



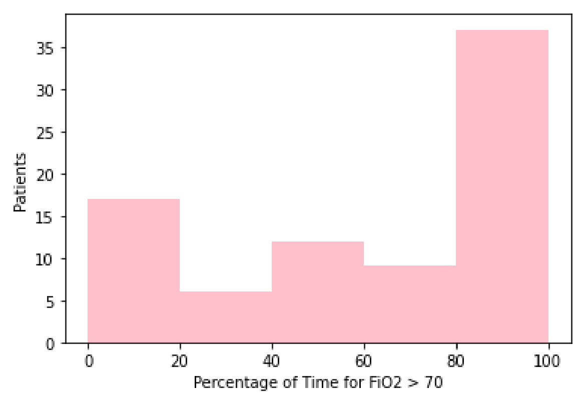
(a) $PEEP > 10$



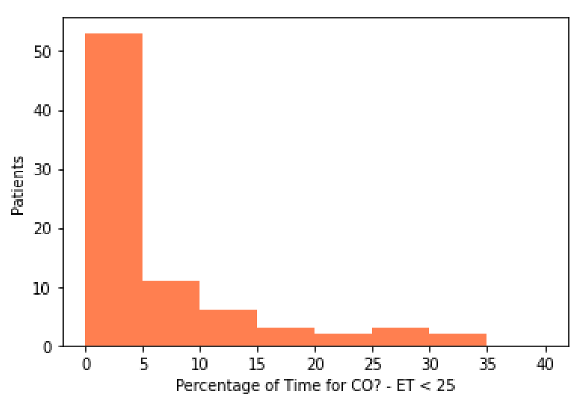
(b) $PEEP < 5$



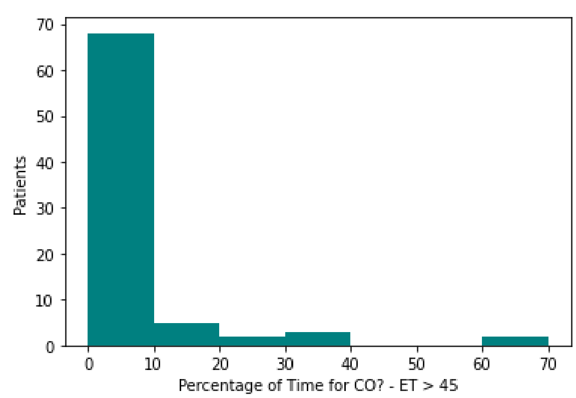
(c) $Ppeak > 30$



(d) $FiO_2 > 70$



(e) $CO_2 - ET < 25$



(f) $CO_2 - ET > 45$

Figure 4.3: Histogram Visualization for Percentage of time of different ventilation parameters

4.1.2 Summarizing the Time Series Data

We have also adopted an alternate way to summarize the time series data. For each of the patients, we have calculated the percentage of time of the surgery duration above a given threshold for different intraoperative features. We have– heart rate, FiO_2, ABP-Mean, ABP-Sys, PEEP, Ppeak, $CO_2 - ET$. Given a threshold T, we calculated the percentage of time that a feature was above or below T. The threshold for heart rate was greater than 120, 110, & 100 and less than 60. For FiO_2, the threshold was greater than 70 and for $CO_2 - ET$ was greater than 45 and less than 25. The threshold for ABP-SYS was less than 80 and for ABP-MEAN the threshold was less than 60. The threshold for PEEP was greater than 10 and less than 5. The threshold for Ppeak was greater than 30. All the thresholds are suggested by Professor Dr. Biniam Kidane. Due to a large amount of missing data, we removed ABP-Mean, ABP-Sys, PEEP, Ppeak from further analysis.

We have represented the processed data ranging from 0 to 1. By 1 we represented meeting the threshold value for the entire time and by 0 we represented the absence of threshold value for the entire time for each patient. There is also fraction value between 1 and 0 for the representation of the percentage of presence of threshold value during the surgery. Figure 4.2 & 4.3 shows the histogram visualization for percentage of time of different ventilation parameters. From the intraoperative time-series dataset, we have processed and extracted 12 features. Table 4.5 denotes the prepared intraoperative features used for this thesis.

I have used six different combinations of dataset. Initially, I divided the dataset into the main two parts. Those are Preoperative and Intraoperative information for each patient. Intraoperative datasets have been divided into two distinct parts. One of them consists of maximum and the average value of different ventilatory

Variable Name	Positive Cases	Negative Cases
Complication	20	60
Normal Oxygenation after 4 Hours	43	37
Normal Oxygenation after 12 Hours	63	17
Normal Oxygenation after 24 Hours	71	9

Table 4.4: Outcome Variables

Variable Name	Value Range
Heart Rate Mean	47.27 to 100.15
Heart Rate Max	56 to 208
FiO_2 Mean	44.58 to 93.59
FiO_2 Max	68 to 100
Percentage of time for $HeartRate > 120$	0 to 1
Percentage of time for $HeartRate < 60$	0 to 1
Percentage of time for $HeartRate > 110$	0 to 1
Percentage of time for $HeartRate > 100$	0 to 1
Percentage of time for $FiO_2 > 70$	0 to 1
Percentage of time for $C0_2 - ET < 25$	0 to 1
Percentage of time for $C0_2 - ET > 45$	0 to 1
Surgery Duration (Minutes)	20 to 377

Table 4.5: Processed Intraoperative Data

records. Another one is the percentage of time of various ventilatory records for the whole period of surgery. These three distinct records are used to generate 18 different combinations for the experiments.

4.2 Oversampling, Imputation, and Standardization

Our dataset has records for 80 patients. Among them, only 20 have the postoperative OLV associated complication, and 60 have no complications. Forty-three patients had normal oxygenation at 4 hours, and thirty-seven patients did not have normal oxygenation at 4 hours. These scenarios shows that our dataset is highly imbalanced. If we train a machine learning model with this data set and want to use it to make predictions, then complication or abnormal oxygenation prediction can be less likely to occur due to the small minority class. So most of the machine learning models will fail to predict correctly. To resolve this problem, I have used Synthetic Minority Over-sampling Technique (SMOTE) [6]. I have applied it to my training set. It oversamples the minority class by creating synthetic minority class examples and balances the training dataset.

I have two preoperative featurea which have missing values. One of the features is FEV, which has 11 missing values, and another one is DLCO, which has 17 missing values. For imputing the missing value, I have used simple imputer from scikit-learn [7; 8]. It fills up the missing value with the mean value of all the rows of that particular feature.

I have also standardized my dataset using the Standard Scaler from scikit-learn [7; 8]. Standardization modifies the values of the dataset to a standard scale, but it keeps the variations in the ranges of values the same.

4.2.1 Feature Extraction

Feature extraction or selection is choosing a subset of related features to apply in the machine learning model. For feature selection, we have used feature importance, which is the process of allotting scores to input features, which shows the significance of each feature. I have used the feature importance of RFs from sckit learn [7; 8] for identifying the critical features.

As a thoracic & foregut Surgeon, Professor Dr. Biniam Kidane also suggested some crucial features which they carefully observe in OLV assisted lung surgery. Some of the features are also considered based on the medical importance and previous sucessful medical research.

4.2.2 Machine Learning and Outcome Prediction

I executed different 108 experiments for predicting perioperative outcomes for operational data during OLV. Figure 4.4 shows the overall process executed during those experiments. I have used six different combinations of patient information for conducting those experiments. Those six combinations are:

- Combined Data
- Intraoperative Data
- Preoperative Data
- Intraoperative data - max and mean only
- Intraoperative data - percentage of time only
- Combined Data without percentage of time records

We have divided intraoperative data into two parts: Maximum and mean value for the features and percentage of time records for those features. Intraoperative data is basically the combined data of maximum and mean value for the features and percentage of time records for those features. From above list we can observe that we have used different combination of those intraoperative data.

For different 108 experiments, I have used three different classification algorithms. Those are RFs, SVMs, and LR. These three classification algorithm

Hyperparameter	Range
Number of Tress for RF	1 to 200
Max Depth for RF	1 to 25
Regularization parameter C for SVM	0.001 to 10000
Gamma for SVM	0.0001 to 0.01
Kernel for SVM	rbf / linear
Regularization Strength C for LR	0.01 to 1000
Penalty for LR	l1 or l2

Table 4.6: Hyperparameters

predicted complication, normal oxygenation at 4 hours, and normal oxygenation at 12 hours.

Each combination of the dataset was randomly split into two sets: the training set (80%) and the test set (20%). Models were trained on the training set only. All models used the same split of training and test data. SMOTE was only applied to training set. There is no synthetic data in the test set. Results were obtained using ten repeated 5-fold cross-validation and grid search over hyperparameters for hyperparameter tuning. The hyperparameters are: number of trees and max tree depth for RFs, regularization parameter C (used for harmonizing the model complexity for better prediction outcomes), kernel (defines the kernel type to be used in the algorithm) and gamma (defines the kernel coefficient) for SVMs, and the inverse of regularization strength C and penalty (applied to define the standard practiced in the penalization) for LRs. Table 4.6 illustrates the different values for hyperparameters.

For verification of results with grid search CV, I have used another technique called nested cross-validation. In this technique, data was split using stratified k-fold into five different folds, and grid search cv is applied to each fold. Each of

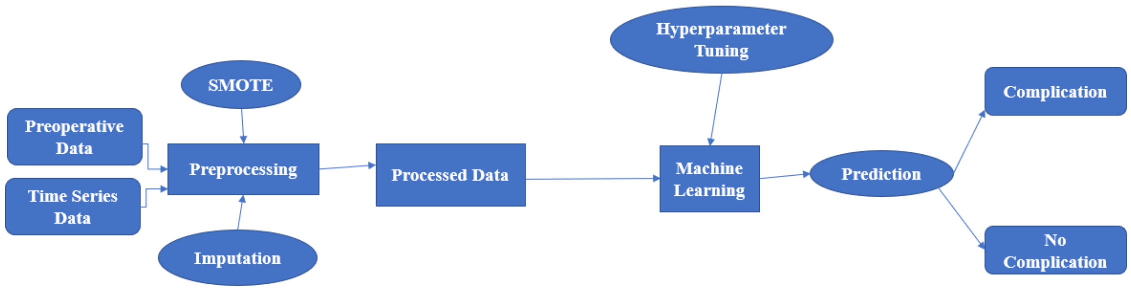


Figure 4.4: Flow chart for predicting perioperative outcomes

the fold returns average optimum hyperparameters for grid search cv. In each grid search, it used 5-fold cross-validation. Then those hyperparameters are applied to each fold of the stratified k-fold, and the final result was calculated by taking the average of all five results.

I have also used Principal Component Analysis (PCA), a dimensionality reduction algorithm. This algorithm is applied to project high dimensional data into a lower dimension as many features or dimensions in a dataset produces more noise. It can be used to reduce the number of features and to eliminate redundancy in the features. Since we had few instances, we wanted to investigate whether reducing the number of features as well would improve results. We have executed the PCA analysis for different components: 3, 4, 5, 6, and 7. We have used 3-7 components since it represents a range of values which includes up to half the number of features of the original set.

4.3 Evaluation

As I have mentioned earlier, I have measured two different metrics for each of the experiments. Those are f1-score, and AUC. Our datasets are highly imbalanced. For complication, only 25% of the data set has complications, and for oxygenation at 12 hrs, only 21.25% of the dataset has not returned to normal oxygenation after 12 hours. For imbalanced classes, f1-score & AUC is more insightful than accuracy score.

Table 4.7 shows the f1-score for complication prediction without SMOTE. The highest score we have is with combined preoperative data and intra-operative data classified by RFs. Preoperative values are usually recorded before OLV assisted thoracic surgery.

Now we move to table 4.8, which shows the f1-score for complication prediction with SMOTE. The highest value here we have with intra-operative data classified by SVMs, which is 0.70. We got 0.61 AUC for this prediction. These were obtained with an RBF kernel and with a avg. value of C=5311.0 and gamma= 0.0091. This f1-score with AUC is the best prediction among all the experiments.

The result for normal oxygenation prediction using the same experimental settings are mentioned in Table 4.9, 4.10, 4.11 and, 4.12 which is not acceptable due to poor score.

Using machine learning with combined preoperative data and intra-operative data yielded the best results for complication prediction. Using Rfs, the F1-score for prediction of complications is 0.59. Intra-operative data alone had the best prediction after applying SMOTE. It has 0.70 f1-score with SVM classification algorithm. Figure 4.5 shows an illustration of the decision boundary of an SVM.

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.59±0.07	0.57±0.09	0.57±0.09
Intra-operative	0.56±0.10	0.57±0.09	0.57±0.09
Pre-operative	0.58±0.09	0.57±0.09	0.57±0.09
Intra-operative without Percentage of Time Records	0.59±0.10	0.57±0.09	0.57±0.09
Percentage of Time Records	0.57±0.07	0.57±0.09	0.57±0.09
Combined without Percentage of Time Records	0.59±0.10	0.56±0.08	0.57±0.09

Table 4.7: Complication Prediction without SMOTE

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.67±0.11	0.68±0.11	0.64±0.09
Intra-operative	0.68±0.08	**0.70±0.08**	0.59±0.13
Pre-operative	0.56±0.08	0.57±0.13	0.59±0.13
Intra-operative without Percentage of Time Records	0.65±0.12	0.67±0.11	0.63±0.11
Percentage of Time Records	0.67±0.10	0.67±0.13	0.47±0.12
Combined without Percentage of Time Records	0.63±0.09	0.60±0.12	0.63±0.07

Table 4.8: Complication Prediction with SMOTE

The points exhibit instances from the data set, and the colored regions indicate predictions. A new data point on the graph will be classified according to the background color where it is placed. The areas and data points are represented by a 2-dimensional PCA plot to visualize high-dimensional data. The value of AUC for

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.51±0.09	0.36±0.09	0.35±0.10
Intra-operative	0.48±0.08	0.33±0.07	0.34±0.08
Pre-operative	0.47±0.16	0.46±0.12	0.34±0.08
Intra-operative without Percentage of Time Records	0.57±0.08	0.54±0.11	0.35±0.10
Percentage of Time Records	0.37±0.12	0.32±0.07	0.33±0.08
Combined without Percentage of Time Records	0.47±0.09	0.39±0.09	0.35±0.10

Table 4.9: Normal Oxygenation at 4 Hours Prediction without SMOTE

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.50±0.08	0.47±0.09	0.43±0.08
Intra-operative	0.51±0.08	0.46±0.11	0.55±0.09
Pre-operative	0.40±0.10	0.55±0.18	0.55±0.09
Intra-operative without Percentage of Time Records	0.58±0.10	0.59±0.10	0.45±0.11
Percentage of Time Records	0.48±0.09	0.38±0.13	0.36±0.11
Combined without Percentage of Time Records	0.53±0.11	0.46±0.08	0.45±0.09

Table 4.10: Normal Oxygenation at 4 Hours Prediction with SMOTE

this prediction was 0.61.

Without SMOTE, intra-operative data alone had an f1-score of 0.57. This shows that SMOTE improved the result drastically with the intra-operative data alone.

Pre-operative data alone yielded weaker results. Using LRs, the F1-score for

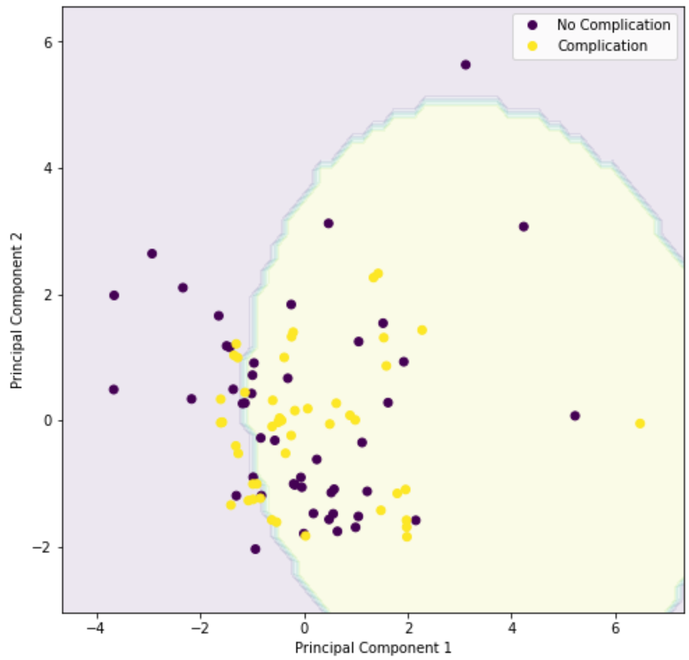


Figure 4.5: Illustration for SVM Analysis on Intraoperative data with SMOTE

the preoperative data alone is 0.59 with SMOTE & using RFs, 0.58 without SMOTE.

I have used the feature importance of RFs from sckit learn [7; 8] for identifying the critical feature. From that experiment, it is found that maximum heart rate, maximum F_iO_2, and surgery times are very crucial intraoperative features for the prediction in the case of RFs.

When using nested CV, the average result was almost identical as the actual result prepared with the grid search cv. The variation with the actual result was 0.01-0.08 at most.

I have also used Principal Component Analysis (PCA), a dimensionality reduction algorithm. I have executed the PCA analysis for different components: 3, 4, 5, 6, and 7. Results were obtained using same ten repeated 5-fold cross-validation and grid search over hyperparameters for hyperparameter tuning. From Table

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.73±0.10	0.74±0.09	0.74±0.09
Intra-operative	0.69±0.12	0.74±0.09	0.74±0.09
Pre-operative	0.74±0.09	0.74±0.09	0.74±0.09
Intra-operative without Percentage of Time Records	0.74±0.08	0.73±0.10	0.74±0.09
Percentage of Time Records	0.71±0.09	0.74±0.09	0.74±0.09
Combined without Percentage of Time Records	0.72±0.09	0.74±0.09	0.74±0.09

Table 4.11: Normal Oxygenation at 12 Hours Prediction without SMOTE

Dataset	F1-Score for RF	F1-Score for SVM	F1-Score for LR
Combined	0.66±0.10	0.69±0.08	0.65±0.16
Intra-operative	0.69±0.12	0.72±0.11	0.58±0.13
Pre-operative	0.65±0.14	0.63±0.15	0.58±0.13
Intra-operative without Percentage of Time Records	0.69±0.09	0.70±0.09	0.58±0.10
Percentage of Time Records	0.64±0.08	0.61±0.08	0.54±0.11
Combined without Percentage of Time Records	0.65±0.05	0.64±0.09	0.65±0.15

Table 4.12: Normal Oxygenation at 12 Hours Prediction with SMOTE

4.13 and 4.14, we can see the f1-score for different PCA analysis for complication prediction with and without SMOTE. It does not make the results significantly worse, but it does not improve on the results. They cannot beat the result of intra-operative data with SMOTE classified by the SVM classification algorithm.

PCA Components	RF	SVM	LR
3	0.58±0.15	0.58±0.17	0.60±0.15
4	0.65±0.13	0.63±0.09	0.62±0.09
5	0.65±0.13	0.62±0.14	0.62±0.14
6	0.65±0.07	0.62±0.08	0.64±0.10
7	0.62±0.09	0.62±0.11	0.62±0.11

Table 4.13: F1 Score for Different PCA Analysis for Complication Prediction without SMOTE

PCA Components	RF	SVM	LR
3	0.64±0.11	0.63±0.10	0.58±0.08
4	0.63±0.08	0.62±0.08	0.60±0.07
5	0.63±0.10	0.62±0.11	0.56±0.07
6	0.65±0.11	0.61±0.07	0.54±0.09
7	0.65±0.08	0.55±0.11	0.49±0.05

Table 4.14: F1 Score for Different PCA Analysis for Complication Prediction with SMOTE

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.50±0.04	0.50±0.00	0.50±0.01
Intra-operative	0.47±0.14	0.50±0.02	0.50±0.03
Pre-operative	0.49±0.03	0.50±0.00	0.50±0.02
Intra-operative without Percentage of Time Records	0.53±0.16	0.50±0.11	0.49±0.04
Percentage of Time Records	0.47±0.05	0.50±0.00	0.49±0.03
Combined without Percentage of Time Records	0.51±0.08	0.48±0.00	0.50±0.01

Table 4.15: AUC for Complication Prediction without SMOTE

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.50±0.07	0.57±0.12	0.54±0.09
Intra-operative	0.56±0.08	0.61±0.07	0.50±0.06
Pre-operative	0.39±0.10	0.46±0.16	0.48±0.15
Intra-operative without Percentage of Time Records	0.51±0.13	0.65±0.11	0.50±0.08
Percentage of Time Records	0.55±0.07	0.54±0.11	0.48±0.09
Combined without Percentage of Time Records	0.46±0.11	0.53±0.12	0.52±0.12

Table 4.16: AUC for Complication Prediction with SMOTE

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.51±0.08	0.45±0.07	0.44±0.06
Intra-operative	0.48±0.11	0.42±0.10	0.45±0.08
Pre-operative	0.49±0.11	0.48±0.06	0.46±0.07
Intra-operative without Percentage of Time Records	0.58±0.08	0.57±0.15	0.49±0.04
Percentage of Time Records	0.41±0.05	0.46±0.02	0.46±0.03
Combined without Percentage of Time Records	0.48±0.15	0.48±0.09	0.46±0.07

Table 4.17: AUC for Normal Oxygenation at 4 Hours Prediction without SMOTE

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.50±0.10	0.47±0.08	0.45±0.08
Intra-operative	0.52±0.11	0.49±0.13	0.42±0.11
Pre-operative	0.43±0.13	0.55±0.08	0.55±0.08
Intra-operative without Percentage of Time Records	0.58±0.15	0.60±0.11	0.46±0.06
Percentage of Time Records	0.48±0.06	0.45±0.10	0.39±0.10
Combined without Percentage of Time Records	0.53±0.07	0.47±0.06	0.47±0.07

Table 4.18: AUC for Normal Oxygenation at 4 Hours Prediction with SMOTE

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.48±0.04	0.50±0.00	0.50±0.00
Intra-operative	0.44±0.03	0.50±0.00	0.50±0.00
Pre-operative	0.49±0.04	0.50±0.00	0.50±0.00
Intra-operative without Percentage of Time Records	0.51±0.07	0.48±0.05	0.49±0.00
Percentage of Time Records	0.46±0.02	0.50±0.00	0.50±0.00
Combined without Percentage of Time Records	0.47±0.06	0.50±0.02	0.50±0.00

Table 4.19: AUC for Normal Oxygenation at 12 Hours Prediction without SMOTE

4.4 Discussion

The 108 different experiments with the various classification algorithms give us some insightful observations. First of all, the highly imbalanced dataset is the

Dataset	AUC for RF	AUC for SVM	AUC for LR
Combined	0.45±0.14	0.55±0.14	0.59±0.14
Intra-operative	0.54±0.08	0.63±0.13	0.50±0.13
Pre-operative	0.50±0.11	0.56±0.12	0.45±0.11
Intra-operative without Percentage of Time Records	0.59±0.16	0.59±0.15	0.56±0.17
Percentage of Time Records	0.49±0.11	0.52±0.11	0.53±0.14
Combined without Percentage of Time Records	0.46±0.18	0.46±0.12	0.61±0.12

Table 4.20: AUC for Normal Oxygenation at 12 Hours Prediction with SMOTE

main reason for not getting higher accuracy. Class distribution for the complication, normal oxygen saturation at 4 hours, and normal oxygen saturation at 12 hours highly uneven. Though we overcome this obstacle using SMOTE. It is clear from the experiments that SMOTE gives a improvement over the imbalanced dataset.

I observed that normal oxygenation at 12 hours gives us higher accuracy without smote. But after analyzing, it is clear that it's always predicting patients are in normal oxygenation as the class distribution is highly unequal here. The class ratio is for predicting normal oxygenation at 12 hours is highly uneven. That means it is not predicting adverse outcomes at all.

SMOTE is only applied to the training set but not in the test set for fair prediction. All the comparisons are fair, as I have used the same test and training split for all the experiments. I observed that intraoperative data has improved prediction accuracy and has a higher AUC score than preoperative data. Combining them is also not capable of bringing improvement in accuracy and AUC.

From this research, it is visible from the result that intraoperative data are more crucial than preoperative in terms of F1-score and AUC. So this research suggests using intraoperative data as input to the machine learning model. As of now, it is not possible to identify the outcome in real-time as we have used the data for the whole surgery. But it is possible to use real-time data and send them instantly to a deep learning model for immediately predicting complications of surgery for that particular time of surgery.

In our research, Professor Dr. Biniam Kidane suggested using all types of complications as outcome variables. But if we only use pulmonary complications or some specific complications, then the dataset will be more imbalanced. Using only one specific complication will make the dataset more imbalanced, and the prediction of complications likely be more difficult with a higher F1-score.

Chapter 5

Conclusion

5.1 Future Work

The main obstacle to this thesis was the dataset. Class distribution was unequal for complication prediction, normal oxygenation at 4 hours prediction, and normal oxygenation at 12 hours prediction. The sample size for the experiment was 80, which is very small for predicting complications or normal oxygenation with higher accuracy. Many information was incomplete or had missing values, which has been sorted out by taking the average value for that specific patient. Shen et al. [40], Ball et al. [42], Neto et al. [41], Blank et al. [34] used VT, PEEP etc. in their research and considered them as critical ventilation parameters. But those essential features can not be used due to high volume of missing data.

As our dataset's main challenges are the small sample size and uneven class distribution, we can work further if we get a balanced dataset with bigger sample size.

I am not able to explore other machine learning algorithms such as Bayesian

Network, ADA Boost, etc. From one undergraduate research under Professor Dr. Michael Domaratzki, I came to know that Naive Bayes, and Gradient Boosting, doesn't improve the result when tested with similar experimental setup. In the near future, I want to explore those with a bigger dataset, if available.

5.2 Conclusion

All lung surgeries are highly dependent on OLV. At the same time, it is also true that this type of mechanical ventilation can cause acute ventilator-induced lung injuries which can be the reason for different pulmonary complications and, respiratory complications which may lead to the death of a patient. So the evaluation of ventilation parameter and their relationship with the postoperative outcome is highly required. In this thesis I designed a predictor using machine learning which can predict postoperative complications and delayed return to normal oxygenation from ventilation parameter during OLV for reducing the negative outcomes. After executing different 108 experiments, we find that the machine learning model with SVM applied to the intra-operative dataset oversampled with SMOTE gets us the best accuracy with the f1-score of 0.70 and AUC 0f 0.61. This is a drastic improvement over all other experiments. Further research with a balanced dataset can give us a more optimized model for the prediction of perioperative complications of OLV with more accuracy.

Bibliography

[1] Arthur S Slutsky and Lorraine N Tremblay. Multiple system organ failure: is mechanical ventilation a contributing factor? *American journal of respiratory and critical care medicine*, 157(6):1721–1725, 1998.

[2] Katherine P Grichnik and Jeffrey A Clark. Pathophysiology and management of one-lung ventilation. *Thoracic surgery clinics*, 15(1):85–103, 2005.

[3] Giorgio Della Rocca and Cecilia Coccia. Acute lung injury in thoracic surgery. *Current Opinion in Anesthesiology*, 26(1):40–46, 2013.

[4] G Rocca Della and C Coccia. Ventilatory management of one-lung ventilation. *Minerva anestesiologica*, 77(5):534–536, 2011.

[5] Marcelo Gama de Abreu, Manuel Heintz, Axel Heller, Roswitha Széchényi, Detlev Michael Albrecht, and Thea Koch. One-lung ventilation with high tidal volumes and zero positive end-expiratory pressure is injurious in the isolated rabbit lung model. *Anesthesia & Analgesia*, 96(1):220–228, 2003.

[6] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[8] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[9] Atul Purohit, Suresh Bhargava, Vandana Mangal, and Vinod Kumar Parashar. Lung isolation, one-lung ventilation and hypoxaemia during lung isolation. *Indian journal of anaesthesia*, 59(9):606, 2015.

[10] Mayank Mehrotra and Ankit Jain. Single lung ventilation. In *StatPearls [Internet]*. StatPearls Publishing, 2019.

[11] Soraya Abbasi, Emidio M Sivieri, and Vinod K Bhutani. Evaluation of pulmonary function in the neonate. In *Fetal and Neonatal Physiology*, pages 919–926. Elsevier, 2004.

[12] Jérôme Allardet-Servent, Jean-Marie Forel, Antoine Roch, Christophe Guervilly, Laurent Chiche, Matthias Castanier, Nathalie Embriaco, Marc Gainnier, and Laurent Papazian. Fio2 and acute respiratory distress syn-

drome definition during lung protective ventilation. *Critical care medicine*, 37(1):202–e6, 2009.

[13] ”mechanical ventilation- peep positive end expiratory pressure.”. https://www.criticalcarepractitioner.co.uk/mechanical-ventilation-peep-positive-end-expiratory-pressure/. [Online; accessed 2020-06-02].

[14] Forced Expiratory Volume and Forced Vital Capacity. https://www.healthlinkbc.ca/health-topics/aa73564. [Online; accessed 2020-06-02].

[15] Victor Chernick. *Kendig's disorders of the respiratory tract in children*. Saunders Philadelphia, 1990.

[16] Alessandro Brunelli and Pieter E Postmus. Preoperative functional evaluation of the surgical candidate. In *IASLC Thoracic Oncology*, pages 265–273. Elsevier, 2018.

[17] Peter Slinger and Bruce Kilpatrick. Perioperative lung protection strategies in cardiothoracic anesthesia: are they useful? *Anesthesiology clinics*, 30(4):607–628, 2012.

[18] Alf Kozian, Thomas Schilling, Christoph Röcken, Christian Breitling, Thomas Hachenberg, and Göran Hedenstierna. Increased alveolar damage after mechanical ventilation in a porcine model of thoracic surgery. *Journal of cardiothoracic and vascular anesthesia*, 24(4):617–623, 2010.

[19] John M Alvarez, Ranjit K Panda, Mark AJ Newman, Peter Slinger, Jean Deslauriers, and Mark Ferguson. Postpneumonectomy pulmonary edema. *Journal of cardiothoracic and vascular anesthesia*, 17(3):388–395, 2003.

[20] Thomas Schilling, Alf Kozian, Christof Huth, Frank Bühling, Moritz Kretzschmar, Tobias Welte, and Thomas Hachenberg. The pulmonary immune effects of mechanical ventilation in patients undergoing thoracic surgery. *Anesthesia & Analgesia*, 101(4):957–965, 2005.

[21] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach.* Malaysia; Pearson Education Limited,, 2016.

[22] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016.

[23] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[24] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[25] Laura Auria and Rouslan A Moro. Support vector machines (SVM) as a technique for solvency analysis. 2008.

[26] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[27] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565, 2006.

[28] Juliana Tolles and William J Meurer. Logistic regression: relating patient characteristics to outcomes. *Jama*, 316(5):533–534, 2016.

[29] Nitesh V Chawla, David A Cieslak, Lawrence O Hall, and Ajay Joshi. Auto-

matically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery*, 17(2):225–252, 2008.

[30] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905, 2018.

[31] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[32] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[33] Patricia Cruz Pardos, Ignacio Garutti, Patricia Piñeiro, Luis Olmedilla, and Francisco de la Gala. Effects of ventilatory mode during one-lung ventilation on intraoperative and postoperative arterial oxygenation in thoracic surgery. *Journal of cardiothoracic and vascular anesthesia*, 23(6):770–774, 2009.

[34] Randal S Blank, Douglas A Colquhoun, Marcel E Durieux, Benjamin D Kozower, Timothy L McMurry, S Patrick Bender, and Bhiken I Naik. Management of one-lung ventilationimpact of tidal volume on complications after thoracic surgery. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 124(6):1286–1295, 2016.

[35] Douglas A Colquhoun, Bhiken I Naik, Marcel E Durieux, Amy M Shanks, Sachin Kheterpal, S Patrick Bender, and Randal S Blank. Management

of 1-lung ventilation—variation and trends in clinical practice: A report from the multicenter perioperative outcomes group. *Anesthesia & Analgesia*, 126(2):495–502, 2018.

[36] Robert E Freundlich and Sachin Kheterpal. Perioperative effectiveness research using large databases. *Best Practice & Research Clinical Anaesthesiology*, 25(4):489–498, 2011.

[37] Sachin Kheterpal. Clinical research using an information system: the multicenter perioperative outcomes group. *Anesthesiology clinics*, 29(3):377–388, 2011.

[38] Javier Belda, Carlos Ferrando, Ignacio Garutti, et al. The effects of an open-lung approach during one-lung ventilation on postoperative pulmonary complications and driving pressure: a descriptive, multicenter national study. *Journal of cardiothoracic and vascular anesthesia*, 32(6):2665–2672, 2018.

[39] Dr Chris Nickson Chris. Lung recruitment manoeuvres in ards, Apr 2020.

[40] Yaxing Shen, Ming Zhong, Wei Wu, Hao Wang, Mingxiang Feng, Lijie Tan, and Qun Wang. The impact of tidal volume on pulmonary complications following minimally invasive esophagectomy: a randomized and controlled study. *The Journal of thoracic and cardiovascular surgery*, 146(5):1267–1274, 2013.

[41] Ary Serpa Neto, Sabrine NT Hemmes, Carmen SV Barbas, Martin Beiderlinden, Ana Fernandez-Bustamante, Emmanuel Futier, Ognjen Gajic, Mohamed R El-Tahan, Abdulmohsin A Al Ghamdi, Ersin Günay, et al. Association between driving pressure and development of postoperative pulmonary complications in patients undergoing mechanical ventilation for gen-

eral anaesthesia: a meta-analysis of individual patient data. *The Lancet Respiratory Medicine*, 4(4):272–280, 2016.

[42] L Ball, SNT Hemmes, A Serpa Neto, T Bluth, J Canet, M Hiesmayr, MW Hollmann, GH Mills, MF Vidal Melo, C Putensen, et al. Intraoperative ventilation settings and their associations with postoperative pulmonary complications in obese patients. *British journal of anaesthesia*, 121(4):899–908, 2018.

[43] Shuji Okahara, Kazuyoshi Shimizu, Satoshi Suzuki, Kenzo Ishii, and Hiroshi Morimatsu. Associations between intraoperative ventilator settings during one-lung ventilation and postoperative pulmonary complications: a prospective observational study. *BMC anesthesiology*, 18(1):13, 2018.

[44] Alessandro Brunelli, Michele Salati, Gaetano Rocco, Gonzalo Varela, Dirk Van Raemdonck, Herbert Decaluwe, and Pierre Emmanuel Falcoz. European risk models for morbidity (eurolung1) and mortality (eurolung2) to predict outcome following anatomic lung resections: an analysis from the european society of thoracic surgeons database. *European Journal of Cardio-Thoracic Surgery*, 51(3):490–497, 2017.

[45] Alessandro Brunelli, Silvia Cicconi, Herbert Decaluwe, Zalan Szanto, and Pierre Emmanuel Falcoz. Parsimonious eurolung risk models to predict cardiopulmonary morbidity and mortality following anatomic lung resections: an updated analysis from the european society of thoracic surgeons database. *European Journal of Cardio-Thoracic Surgery*, 57(3):455–461, 2020.

[46] Felix G Fernandez, Andrzej S Kosinski, William Burfeind, Bernard Park, Malcolm M DeCamp, Christopher Seder, Blair Marshall, Mitchell J Magee,

Cameron D Wright, and Benjamin D Kozower. The society of thoracic surgeons lung cancer resection risk model: higher quality data and superior outcomes. *The Annals of thoracic surgery*, 102(2):370–377, 2016.

[47] Benjamin D Kozower, Shubin Sheng, Sean M O'Brien, Michael J Liptay, Christine L Lau, David R Jones, David M Shahian, and Cameron D Wright. Sts database risk models: predictors of mortality and major morbidity for lung cancer resection. *The Annals of thoracic surgery*, 90(3):875–883, 2010.