Predicting the Optimal Time for Weaning of Patients in Mechanical Ventilation

Nicholas Hale, Venkata Sai Manohar Gedela, Liangliang Cui, Yunduo Hu

May 19, 2021

Abstract

Despite relevant clinical criteria having been proposed and used to guide care physicians to execute the weaning process, there are still 25-40% of patients who cannot successfully wean from a mechanical ventilator. This is caused by individual differences between patients and differences between mechanical ventilators, Therefore, the existing criteria cannot solve the problem of how to wean from mechanical ventilators at the optimal time. One effective solution is to develop a machine learning model which can predict the optimal weaning time based on actual conditions of the patients. In this study, a machine learning model will be built to estimate the optimal weaning time. Different machine learning algorithms will be applied for model establishment, such as logistic regression, decision tree and other algorithms in supervised learning, as well as k-means clustering in unsupervised learning, to try as many different models as possible to explore what kind of algorithm can provide the best prediction according to the different conditions and situation of patient. There are two main stages in this study: data preparation and subsequent machine learning model development. The data preparation can be divided into 7 substages: Missing value analysis, exploratory data analysis, variable selection, identification and exclusion of outliers, data aggregation, inclusion criteria and feature construction. After the data preparation stage, the processed data will be applied in different algorithms to find out which can provide the best prediction. Algorithms are divided into two categories, supervised learning and unsupervised learning. After each model is trained and tested, the results obtained will be compared with the actual results to get the accuracy of the model. Compared with previous studies, in order to adapt to the complex and changeable situation of the patient's body, the model selects more types of input variables to cover different situations. Besides, the study also observed the prediction performance of different types of algorithms in different situations, in order to choose the best one from them to ensure that the final model can make accurate predictions under complex and changeable actual conditions, so as to aid clinicians to execute the weaning process at the optimal time.

Under the same training data set and test data set, we got the following results: the accuracy rate for logistic regression is 80.62%, 80.28% for

Principal Component Analysis, 80.28% for k-means clustering and 99.20% for decision tree. Therefore, the decision tree algorithm will be selected as the final selected model.

In conclusion, under complex conditions, the prediction accuracy obtained by the decision tree is the highest. Therefore, the model based on the decision tree algorithm can effectively assist the clinicians in determining the optimal time for the weaning procedure.

1 Introduction

Mechanical ventilators are essential components in Intensive Care Units, which are used to assist patients who are failing to spontaneous breath or undergoing surgery. In ICU, there are a high proportion of patients who are under the assistance of mechanical ventilators. At the same time, the cost of maintaining the ventilator is also very expensive, accounting for 12% of the total hospital expenditure.

The process of removing a patient from the ventilator can be defined as weaning, or extubation. Delaying or unnecessarily extending the weaning time will increase the length of stay in the intensive care units (ICU), which causes the increase of burden on the ICU and financial pressure on patients (Walsh et al, 2014).

Meanwhile, it is of vital importance to choose the appropriate weaning time. Incorrect weaning time, earlier or later, will both have a serious negative impact on the health of patients. Starting the weaning process early will cause the patient to have several uncomfortable symptoms, and if it is more serious, the patient will stay in the ICU for a longer time. If the weaning process starts too late, the patient will have symptoms such as post-extubation delirium, ventilator induced pneumonia, etc., and the patients will face a higher mortality rate.

Although relevant weaning criteria has been published by relevant organizations that critical care physicians can use to assess the optimal weaning time, 10-20% of patients who meet all the criteria will fail in the extubation (Walsh et al, 2014). This is because there are great individual differences between people, and existing criteria cannot be applied to everyone.

In order to solve this problem, one feasible method is to develop a machine learning algorithm based on the massive data recorded by ICU. The past data set will be used to train the model so that the model can predict the results based on actual conditions. In this condition, the actual condition is the medical condition of the patients.

The purpose of this study was to build a machine learning model based on the large amount of data recorded by ICU. The model will be trained by the ICU dataset, and then will be able to predict the optimal time to execute the weaning process by analyzing the patient's data, which can be used to assist the clinicians to execute the weaning process. The input variables are selected from previous peer-reviewed journals, in order to obtain a more accurate analysis, the number of variables selected in this study will be higher than in previous

similar studies. Different machine learning algorithms will be applied in this study, the outcomes of each model will be compared in order to find the model with the highest prediction accuracy. The algorithm with the highest accuracy rate will be adopted, perfected into a complete model, and packaged into the pipeline, which can generate practical application value.

2 Background & Literature Review

Weaning from mechanical ventilation is an important process in the ICU stay. Weaning refers to the entire process of removing patient from artificial assistance and from the endotracheal tube (Boles et al, 2007). Unsuccessful weaning process may cause serious consequence, for the patients who fail extubation in the weaning process will suffer a high mortality rate of 25-50%. Additionally, extubation failure and/or reintubation, regardless of the seriousness of the underlying condition, may have a significant negative impact on the patient's outcome (Arnaud et al, 2013). Although there are clear clinical criterial which indicates the time the patient requires to be performed the weaning trial, but there is no guarantee that the whole weaning process can be successfully performed (10-20% of the patients will fall at the stage of extubation).

In the early research, due to the lack of computer science, scientists usually use the controlled variable method directly on the patient to study the influence of different ventilation methods on the outcome of weaning from mechanical ventilation.

For example, in 1994, Brochard and his colleagues found that the ventilatory strategy can affect the outcome of weaning from mechanical ventilation. In their research, three different ventilation methods were proposed and randomly applied to the patients. Experiment result indicated that the pressure support ventilation (PSV) contributed to the other two methods (T-piece trials and synchronized intermittent mandatory ventilation (SIMV)) (Brochard, 1994).

Another example is the experiment conducted by Wesley and Albert in 1996. Wesley and Albert carried out a randomised, controlled experiment in surgical and coronary intensive care units, with 300 adult patients who were on mechanical ventilation. The patients were divided into two groups, the physical condition of the patients in the intervention group will be checked by medical staff every day to determine whether they can breathe spontaneously. The other group only checks the physical condition of the patients but does not allow the patients to breathe spontaneously.

Not only the research method, the research results of this research were also very valuable. The results showed that appropriate spontaneous breathing trials for patients with suitable conditions can effectively reduce the time of mechanical ventilation and reduce medical expenses, and there will be fewer complications compared with conventional care [4].

The key value of this research was it proved that studying the subject of weaning from mechanism ventilation can help patients from different dimensions. This provides research impetus for other scientists to conduct research in

this area.

In the 21st century, the focus of research has shifted from exploring the outcomes of weaning from mechanism ventilation to how to correctly predict the weaning success. Researchers predict the time of weaning from mechanical ventilation to reduce the damage caused to the human body during the weaning process.

This change can be reflected in Meade's research on predictors in 2001. In the research, Meade and his colleagues aimed to find which predictors can be used to determine the success of mechanical ventilation. In their research, by consulting a large number of documents, some predictive factors that meet the standards were determined, and these predictors were tested, and compared with the actual situation to judge whether the prediction is accurate [3].

However, the results of this research were not satisfactory, a significant number of predictors were found to be ineffective in predicting weaning outcomes. The most plausible reason for the poor performance of the research is that when clinicians choose patients for weaning trials, they already consider the outcome. In other words, by the time investigators formally measure the tests' predictive ability in patients that physicians have already determined are candidates for weaning, the tests' predictive power has been "used up." As a result, expecting physiologic assessments to be extremely predictive in patients with an intermediate likelihood of weaning success is impractical [3].

In contrast to Meade, Walsh (2004) successfully contributed a series of criteria which can be used to predict whether the weaning is successful or not. In the research, a checklist has been developed which contains a series of criteria that can be used to estimate the time to start the weaning process. These criteria were applied to the patients in one general ICU, and a daily based assessment was performed to monitor the patients' situation in order to assess whether it meet the criteria [7].

This experiment was a great success. Experimental results show that most patients who can breathe independently meet the set standards. This means that this criterion can be used to determine when to start the weaning process. Although not completely accurate, it can also provide a very valuable reference. At the same time, this also proves that it is feasible to predict the start time of the weaning procedure. The judgment criteria established by the research also provide reference for follow-up research [7].

With the development of computer science, models built by computers to simulate actual conditions and predict results are gradually adopted by researchers and used in research. In 2015, Kuo announced that he successfully developed an artificial neural network (ANN) model which can be used to accurate predict the extubation. Compared with previous researches, the ANN model established by Kuo achieved the highest accuracy rate. This can effectively assist the clinicians to choose the earliest time to perform the weaning process.

In previous studies on predicting weaning time, there were less categories of input variables used to make model predictions. For example, in the ANN model established by Kuo[6], only 8 input parameters were included in the simulation.

Although the accuracy rate obtained by this experiment is relatively excellent, it is limited by fewer types of input parameters, so the results obtained still have more room for improvement. This is also what needs to be improved in this experiment. We will select more input parameters based on the existing standards (such as those proposed by Wales in 2004), and use mechanical learning methods to build different models, such as linear regression, logistic regression, decision trees, etc., to explore which model can get the highest accuracy.

3 Research question and Hypothesis

Mechanical ventilation can provide an effective treatment method for patients with respiratory failure. Many studies have shown that with the development of technology, by using computers can provide clinicians with better help and guidance, because machine learning algorithms can be used to analyze large amounts of data that has been collected to provide personalized warning for patients (Hung-Ju et al. 2015) [9]. So far the developed ventilator intelligent management system and acute kidney injury (AKI) early warning system have been used to provide doctors with auxiliary warning and decision-making, and improved the survival rate and prognosis of patients, and also some of them can analyze clinical data in real time, display the most relevant parameters of the disease to the doctor according to the weight, and predict the development of the disease and the trend of death risk, assisting the doctor to use clinical experience to make rapid clinical diagnosis and treatment decisions (Hasan, 2014) [14]. However, there seems not to be a single prediction method or model to help intensive care physicians accurately and opportunely predict the weaning time.

The current research questions were primarily focusing on simple data pre-processing, and only few of them kept eyes on balancing data, dimension reduction or feature extraction. In this case, it often causes that the potential information of data is not mined as much as possible or improper handling with data leads to inaccurate prediction. The features applied to train a model in current research were generally containing demographic data, for example age, gender and height, regular check-up data such as mean breathing frequency and body temperature, and special items for ICU stays for example Level of consciousness. In addition, most of current research invest Artificial Neural Network (ANN), Support Vector Machine (SVM) or Logistic Regression (LR) (Walsh, 2004) [15].

In our project, it could use collected patient data from Mimic DataSet, and various machine learning algorithms can be used for model establishment, such as linear regression, logistic regression, decision tree and other algorithms in supervised learning, as well as k-means clustering in unsupervised learning, to try as many different models as possible to explore what kind of algorithm can provide the best prediction according to the different conditions and situation of patient. The regression model is probably the simplest of all models. The decision tree may only require a short time and a small amount of computation

5

to provide a good forecast for suitable weaning. K-means clustering can perform the task of classification in unknown data. For data pre-processing, under-sampling is used to balance the data and principal component analysis (PCA) is applied to reduce data dimension.

Furthermore, whether starting weaning protocols, especially without professional physician health-care staff implementing, would presumably have effect on weaning result (Joshua et al. 2018) [16]. Therefore, selecting proper criteria as an instruction for workers, on determining to complete the disconnection with mechanical ventilation for patients who are potentially achieved, is greatly essential before activating the weaning protocol. The criteria are generally required to be simple-data, easy to collect and cost-friendly. They should also have ability of prediction to weaning and extubation success. It is, in fact, hardly to find out the criteria like which could be suitable for each single individual, because of diversity and specificity among people. Hence, it could select respiratory rate, blood oxygen saturation, heart rate or PS_Above_PEEP, EPAP_PEEP, SET_FIO2, Spont_RR, DeadSpace, PaCO2, PaO2, SBI,P01; lab tests:pH and assessed their abilities to be potential criteria so that judging whether start a nurse-led weaning protocol.

4 Methodology

The stages in the data preparation and subsequent machine learning model development are as follows:

1. Data Preparation

    - Missing value analysis
    - Exploratory data analysis
    - Variable selection
    - Identification and exclusion of outliers
    - Data aggregation
    - Inclusion criteria
    - Feature construction

2. Model Developement

    - Patient stratification through unsupervised learning
        - K-mean clustering
    - Patient classification through supervised learning
        - Linear regression
        - Logistic regression
        - Decision tree

4.1 Missing Value Analysis

The data provided was incomplete with many missing values. This was partly due in some cases to the method in which the data was collected. Data features such as PH, PaCO2, Pa02, Lactate ABG and temperature had >95% of data missing because the process of extracting the measurements from the patients put the patients under stress and so therefore regular monitoring of these features is not possible without putting the patient under excessive discomfort. For the rest of the missing data values, this can be due to a variety of reasons, tests may have forgotten to be done or recorded, or pages simply lost or deleted [[8]]. Depending on the extent of this missing data, we are able to potentially impute values to greatly reduce the amount of lost data.

The first step in pre-processing was to investigate whether there were any patients with the majority of their data for the features that were not collected infrequently. For each of these features, the percent of missing data values was found for each patient to see if there were patients that were missing substantial information, table 1. There were 94 patients of this type who were removed from the training dataset. Following the removal of these patients, the percentage of missing values

Table 1: Missing Feature Data

Feature	0 to 20%	20 to 50%	50 to 80%	80 to 100%
Time	408	0	1	93
AdmissionDate	408	0	1	93
DischargeDate	407	0	1	94
CCOutcome	408	0	1	93
VentMode	408	0	1	93
EPAP_PEEP	408	0	1	93
SET_FI02	407	1	1	93
Spont_RR	407	1	1	93
Deadspace	408	0	1	93
ArtSystollic	395	8	5	94
ArtMAp	395	8	5	94

For the aforementioned features, PH, PaCO2, Pa02, Lactate ABG, no attempt to complete the data was made and these features were analysed as sparse features. For the remaining features, two methods were used to address the missing data: setting all the missing values to zero, and linearly interpolating a value from non-null values surrounding the missing data points.

4.2 Exploratory data analysis

The below note tell us that the dataset has information regarding all the patients with physiological vital signs :PS_Above_PEEP, EPAP_PEEP, SET_FIO2,

Spont_RR, DeadSpace, PaCO2, PaO2,SBI,P01; lab tests: pH. Each observation/row has a timestamp (column 'TIMEDELTA') that indicates the number of hours after ICU entry where the observation was taken. There are several observations for the same variable/column in each icustay(ID).

```
Number of ICU stays: 97
Number of survivors: 74
Number of non-survivors: 23
Mortality: 23.7%

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1048575 entries, 45056 to 47502
Data columns (total 22 columns):
 #   Column         Non-Null Count      Dtype
---  ------         --------------      -----
 0   Time           1048575 non-null    object
 1   AdmissionDate  1048575 non-null    object
 2   DischargeDate  1048575 non-null    object
 3   CCOutcome      1048575 non-null    int32
 4   HospOutcome    799430 non-null     object
 5   VentMode       1048575 non-null    object
 6   PS_Above_PEEP  659608 non-null     float64
 7   EPAP_PEEP      1048466 non-null    float64
 8   SET_FIO2       1046527 non-null    float64
 9   Spont_RR       1045294 non-null    float64
 10  Deadspace      1047188 non-null    float64
 11  SBI            702194 non-null     float64
 12  P01            773384 non-null     float64
 13  PH             5329 non-null       float64
 14  PaCO2          5351 non-null       float64
 15  PaO2           5309 non-null       float64
 16  LactateABG     5341 non-null       float64
 17  HeartRate      1011677 non-null    float64
 18  ArtSystollic   875634 non-null     float64
 19  ArtMAP         880870 non-null     float64
 20  Temperature    50764 non-null      float64
 21  TIMEDELTA      1048575 non-null    float64
dtypes: float64(16), int32(1), object(5)
memory usage: 180.0+ MB
```

Figure 1: Summary of the dataset being used including information on the number of patients for each outcome.

The dataset consists of 97 unique ICU stays and 1,048,575 observations. All columns except for 'TIMEDELTA', 'CCOutcome','VentMode' have missing information. The existence of outliers can be detected by examining the maximum and minimum values. Both missing data and outliers are popular in ICU datasets and must be considered before implementing ML algorithms.

4.3 Variable selection

The future value of the attribute in the model and the amount of data available should be balanced. We do know how much data is missing in each column, but we don't know how much information is missing at the patient stage. To do so, we'll aggregate data by ID and count the number of non-null values with the 'group by' function and the 'mean' operator. This would indicate how many ICU stays have at least one observation for each attribute.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 97 entries, 45056 to 47502
Data columns (total 17 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   CCOutcome      97 non-null     int32
 1   PS_Above_PEEP  84 non-null     float64
 2   EPAP_PEEP      97 non-null     float64
 3   SET_FIO2       97 non-null     float64
 4   Spont_RR       97 non-null     float64
 5   Deadspace      97 non-null     float64
 6   SBI            88 non-null     float64
 7   P01            97 non-null     float64
 8   PH             97 non-null     float64
 9   PaCO2          97 non-null     float64
 10  PaO2           97 non-null     float64
 11  LactateABG     97 non-null     float64
 12  HeartRate      97 non-null     float64
 13  ArtSystollic   96 non-null     float64
 14  ArtMAP         96 non-null     float64
 15  Temperature    96 non-null     float64
 16  TIMEDELTA      97 non-null     float64
dtypes: float64(16), int32(1)
memory usage: 13.3 KB
```

Figure 2: Table of the selected variables used for the investigation.

4.4 Identification and exclusion of outliers

4.4.1 Outliers

We already saw that outliers exist in the dataset, but we need to look at the data more closely before determining how to treat them. We can easily generate one boxplot for each vector by using the 'seaborn' library and the 'boxplot' feature. Seaborn is a matplotlib-based visualisation library that offers a high-level interface for drawing statistical graphics (Fig 3).
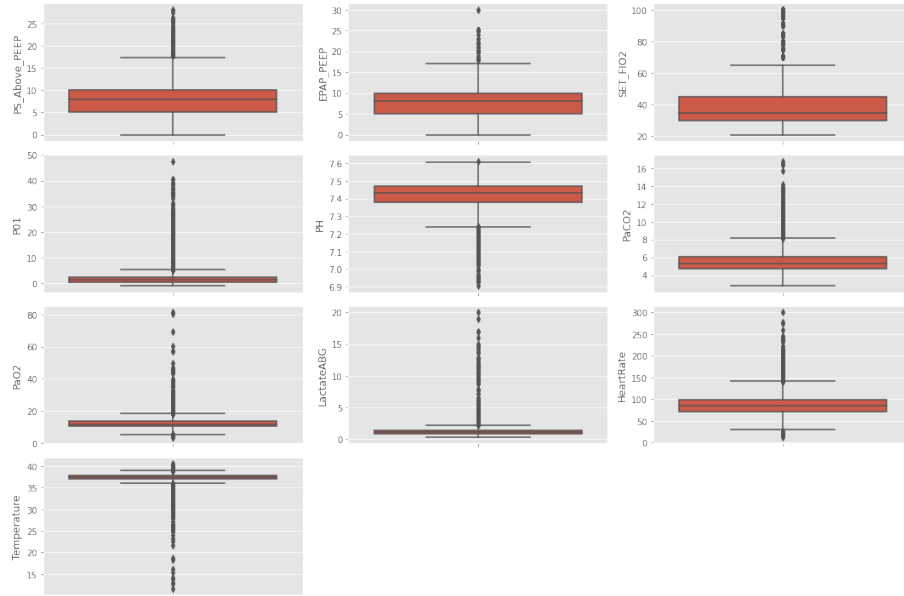
9

Figure 3: Box plot to detect the outliers

The same code will be used to validate the data distribution after outliers have been removed. The 'stripplot' function visualises the underlying distribution as well as the number of observations. When x = 'CCOutcome,' the boxplots are partitioned by result.

4.4.2 Exclusion

Ideally, we should hold drastic values associated with the patients' ill health and remove unlikely values (such as negative temperature) and likely outliers (such as heart rate above 100). Values that fall outside the limits established by expert expertise are omitted to do so. We stop removing extreme (but correct/possible) values this way.

The same code will be used to validate the data distribution after outliers have been removed (Fig 4). The 'stripplot' function visualises the underlying distribution as well as the number of observations. When x = 'CCOutcome,' the boxplots are partitioned by result.
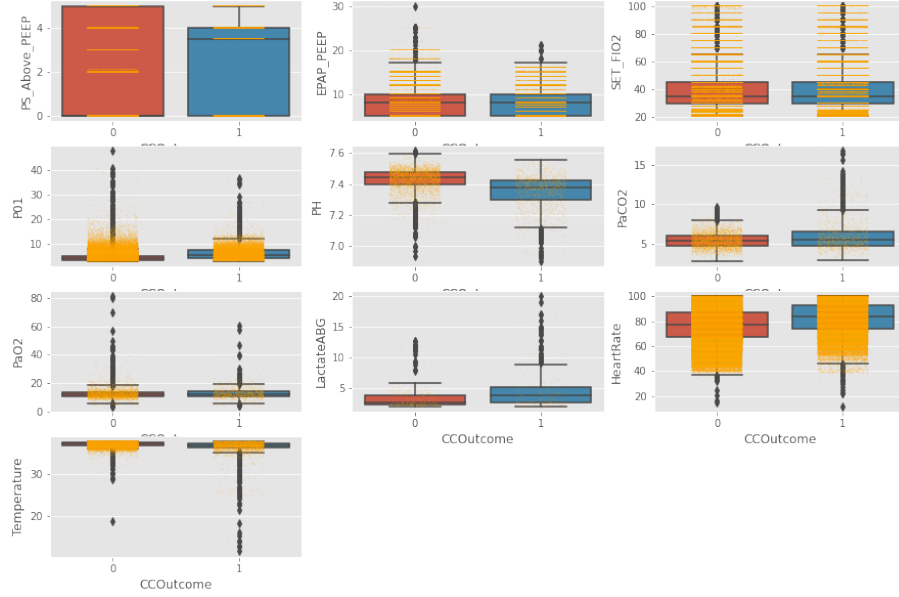
Figure 4: Boxplot and strip plot after excluding outliers

4.5 Data aggregation

4.5.1 Aggregate by hour

As previously said, the dataset includes details about the first two days in the ICU. Any observation has a time stamp that indicates the number of hours between ICU entry and the time the observation was obtained (e.g., 0.63 hours). Person, data is distilled into hourly observations for ease of analysis by choosing the median value of the available observations within each hour. First, the 'floor' operator is used to divide the hours into 48 bins. The 'group by' function is then used with the 'median' operator to obtain the average heart rate for each hour of ICU stay (ID). Fig 5 shows the vital signs for a specific ICU stay (ID=45056). Consecutive hourly observations are connected by line.
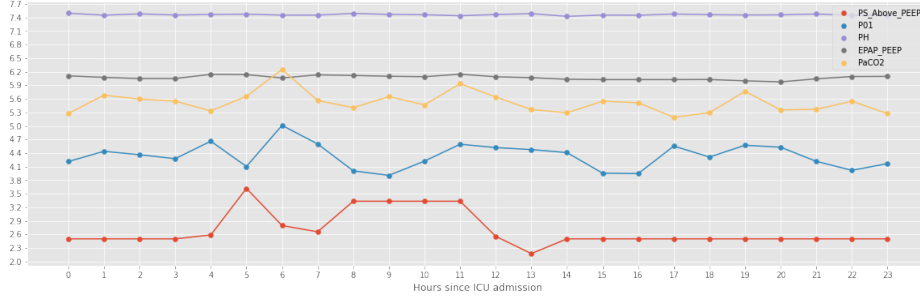
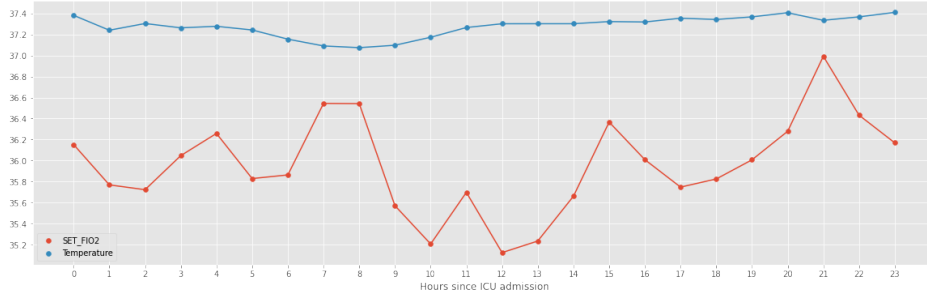Figure 5: Hourly observations of Vital signs for ID=45056

Figure 6: Hourly observations of Vital signs for ID=45056

4.5.2 Select minimum number of observations

We made the decision to leave all time-varying variables open. However, as seen in the preceding case, since not all variables have an hourly sampling rate, a lot of information is lost (coded as Nan). It is essential to determine how to treat missed information before training ML algorithms. There are two options: substitute the missing information with something valuable or exclude the missing information. We can prevent adding bias in this work by replacing missing values with expected worth. Rather, we will concentrate on a comprehensive case study. Complete knowledge may have different interpretations depending on how we construct the feature collection. For e.g., if we want to use one observation per hour, full information means no missed data from t=0 to t=23, which will result in the omission of most data. One typical method for reducing the size of the function space is to use only a portion of the time series. This is the approach that will be used in this project. To derive useful information from the time series, summary statistics such as the mean, limit, minimum, and standard deviation would be used. It is critical in this case to determine the minimum duration of the time series before beginning to pick portions of it. One solution will be to include all patients who have at least one observation about each component. Since summary figures have no significance if only one

observation is visible, a two-observation threshold would be used. Setting 'min num meas = 4' means we're looking for ICU stays (ID) where each vector was registered at least once at four different times. Again, we're aggregating data by ICU stay (ID) and counting the number of observations for each vector with the 'group by' function. Then, we omitted ICU stays (ID) in which any variable was registered less than twice.

4.6 Feature construction

The dilemma of finding the transformation of variables that contains the most valuable details is addressed by feature construction. Easy operations will be used in this project to construct/extract essential features from time series:

- Maximum

- Minimum

- Standard deviation

- Mean

From t=0 to t=23h, these characteristics summarise the worst, greatest, difference, and normal patient's status. In the suggested exercises, you will do this for each day individually, which will maximise the dimensionality of the dataset but, ideally, allow for the retrieval of more valuable information.

These features can be conveniently extracted by using the 'group by' function to combine data by ICU stay (ID), along with the 'max', 'min', 'std', and 'mean' operators.

4.7 Data partitioning

Data may be separated into preparation, evaluation, and validation sets to determine model results.

- The training set is what is used to train/build the learning algorithm.

- Validation (or development) set: used to tune parameters, pick functions, and make other learning algorithm decisions.

- Test set: used to measure the algorithm's output but not to make any assumptions about the learning algorithm design or parameters.

Scikit-learn is Python's basic machine learning bundle. It offers easy and effective data processing and data analysis methods. The following example demonstrates how to use the 'train test split' function from the 'sklearn' library to assign observations to each collection at random. The 'test size' parameter, which specifies the size of the test set and is set to 30% in this case, can be used to control the size of the sets. In situations where the data is heavily imbalanced, forcing an oversampling of the minority class or an under sampling

of the majority class may be a safe solution to ensure that the model is not skewed against the majority class. This should be done on the training set, while the test set should retain the class imbalance discovered on the original data, so that when testing the final model, a true representation of the data is used. Under sampling is used to improve the interpretability of clustering. However, as a general rule, and whether the dataset includes a large number of observations, oversampling is favoured over under sampling because it requires all of the information in the training set to be retained. In any case, using learning algorithms that consider class imbalance might be a safer option. The following example demonstrates how to under sample the majority class given a desired size of the minority class, which is governed by the parameter 'perc class1'. If 'perc class1' is greater than zero, under sampling is used to create a balanced training package. If 'perc class1' is set to 0, no balancing is done.

4.8 Patient classification through supervised learning

4.8.1 Data Preparation

We'll now plan the data so that it can be used by machine learning algorithms. In this notebook, we consider the following processes for the dataset: For categorical attributes, one-hot encoding is used, and function scaling is used for numerical attributes. Scikit-learn has utilities to help with these tasks:

1. OneHotEncoder() allows to transform a categorical variable to a one-hot encoding representation.

2. StandardScaler() performs feature scaling by standardisation.

3. Normalizer() Normalization, also known as scaling, is used to ensure that all functions fall between a given minimum and maximum value, which is usually between zero and one. During the training process, the maximum and minimum values of each element should be calculated, and the same values should be used during the testing phase.

Data transformations include OneHotEncoder() and StandardScaler(). These are known as transformers in scikit-learn, and they map data from one format to another. Transformers are classes in a programming context. They are available in the following methods:

- Fit that is used to learn the transformation from data.

- Transform that is used to transform the data once the transformer has been fitted.

- Fit transform that applies first fit and then transform to the data.

Typically, we use either a fit or fit transform for the training data and transform to the validation or test data.

Since the one-hot-encoding and standardisation transformations may be extended to separate columns in the dataset, having a function that applies both is useful. It is especially useful since such transformations must be applied to the train, validation, and test sets. We might write such a function from scratch, but in this case, we use ColumnTransformer(), a scikit-learn estimator that helps us to combine several transformations into a single process. ColumnTransformer is an example of a scikit-learn estimator. An estimator is a type of object that predicts new data. In addition to the match, transform, and fit transform methods, they often provide a predict method for making projections on test results.

4.9 Logistic Regression

When starting a machine learning project, it is usually a good idea to start with a basic model because it will give you an idea of how difficult the question is. Logistic regression (LR) is regarded as a simplistic model since the underlying math is simple to grasp, making its parameters and outcomes understandable. It also takes longer to compute than other ML models.

4.10 Principal Component Analysis (PCA)

The first discussion of this technique was made in 1933 by Harold Hotelling in a paper on principal component analysis. Hotelling was inspired to establish a statistical basis for factor analysis approaches, which were commonly used in psychology and the social sciences at the time. His model was a factor analysis model, but he found the model's noiseless 'limit.'

4.10.1 Explained Variance

The explained variance tells us how much information can be attributed to each other of the principal components. This is important, because we convert 15 dimension space to 4 dimension space, we may lose some of the information

4.10.2 Clustering

Clustering is a learning task that attempts to divide a given collection of observations into subgroups (clusters) based on data similarities, so that observations in the same cluster are more closely related than observations in different clusters. It is an unsupervised learning task because it detects structures in unlabelled datasets, and a classification task because it can classify findings based on the cluster to which they are assigned.

4.10.3 Pipeline

A pipeline enables the combination of several estimators into one. It is useful in a machine learning project since the various stages (e.g., data pre-processing,

function collection, prediction) can be placed sequentially one after the other in a single instruction.

Let us return to the dataset and implement the transformations we discussed earlier. We begin by creating lists of the names of the attributes, one for categorical attributes and one for numerical attributes.

4.11 Decision Tree classifier

Most ICU severity scores are constructed using logistic regression, which imposes stringent constraints on the relationship between explanatory variables and outcome. Logistic regression is based on the concept of a linear and additive relationship between the result and its predictors. This prediction may be unrealistic given the complexity of the processes causing death in ICU patients.

We aim to build on LR's estimation by using a nonparametric algorithm such as a decision tree. A decision tree is a model that uses a tree–like graph of rules that give probabilities of outcome. It can be used for classification and regression, performs feature selection automatically, is simple to understand and translate (if the tree has a limited depth and a small number of features), and needs no data planning. Since this type of algorithm has no clear decisions about the structure of the mapping function, it is a good candidate when you have a large amount of data and no previous experience, and you don't want to think too hard about selecting the right features.

However, decision trees learners are associated with several disadvantages. They are vulnerable to overfitting because they produce overly complex trees that do not generalise well, and they can be unreliable because minor changes in the data can result in entirely different trees being produced. To solve these problems, methods such as bagging and boosting, such as random woods, are often used.

5 Results

5.0.1 Pairwise plotting

The pairs plot is a common technique in exploratory data analysis (also called scatterplot). This approach helps you to see the distribution of individual variables as well as the relationships between each pair of variables. It is simple to implement in Python with the 'seaborn' library. The following example demonstrates how to map pairwise relationships between variables as well as histograms of single variables partitioned by result (Alive vs Dead). The parameter 'vars' specifies the collection of variables to plot, and the parameter 'hue' specifies the use of separate markers for each stage of the hue variable.
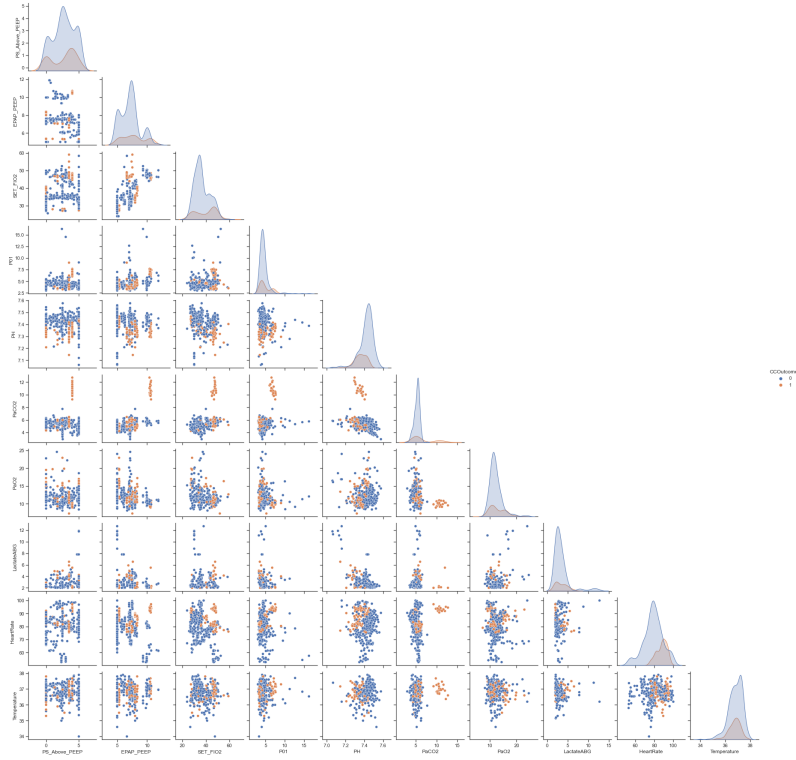
Figure 7: Pairwise plotting for the selected variables

5.0.2 Time series plotting

It is useful to visualise the mean HR partitioned by result from t=0 to t=23 to investigate time patterns. The DataFrame must be restructured to execute this process easily. The following function takes a panda DataFrame and the variable name as input and transposes/pivots the DataFrame to have columns corresponding to time and rows corresponding to ICU stays. If 'filldata' is set to 1, the feature will use the forward fill method to fill missing information, replacing Nans with the value preceding it. We can conveniently execute these two acts by using the function 'fillna' with the method parameter set to 'ffill' and 'bfill'. When 'filldata' is set to 0, no missing data imputation occurs.

Figure 8: Time series plotting for the selected variables

Physiological decline or recovery over time differs greatly between survivors and non-survivors. We couldn't see any distinctions between the groups when we used the pairwise plot, but this sort of plot shows very strong differences.

- P01
    * higher in the dead group
    * rapidly decreasing during the first 5 hours

- SET_FIO2
    * higher in the dead group, increasing over time
    * different between both groups at admission, and converging at one point but diverging thereafter

- Heart rate
    * lower in the alive group

- Temperature
    * low variation from t=0 to t=23h
    * slightly increasing during the first 10 hours

- PH
    * Increasing over time in both groups
    * pH < 7.35 from t=4 to t=23h in the dead group

18

5.1 Clustering

Clustering is a learning task that attempts to divide a given collection of observations into subgroups (clusters) based on data similarities, so that observations in the same cluster are more closely related than observations in different clusters. It is an unsupervised learning task because it detects structures in unlabelled datasets, and a classification task because it can classify findings based on the cluster to which they are assigned.

This work is concerned with the following issues:

- Will we distinguish distinct trends in the absence of class labels?

- How are the various trends expressed through various outcomes?

To answer these questions, we will first provide a summary of the fundamental principles underlying k-means clustering, the most well-known and straightforward clustering algorithm. We will demonstrate how the algorithm operates by using 2D data as an example, perform time series clustering, and use the knowledge obtained from clustering to train predictive models.

5.1.1 Exemplification with 2D data

While pairwise plots revealed no interesting trends, some clusters may have appeared after the data was transformed. You should re-run the code for pairwise plots of transformed features, but due to the high dimensionality of the dataset, this would be time consuming. For illustration purposes, the features 'max PH' and 'mean PaCO2' were selected.
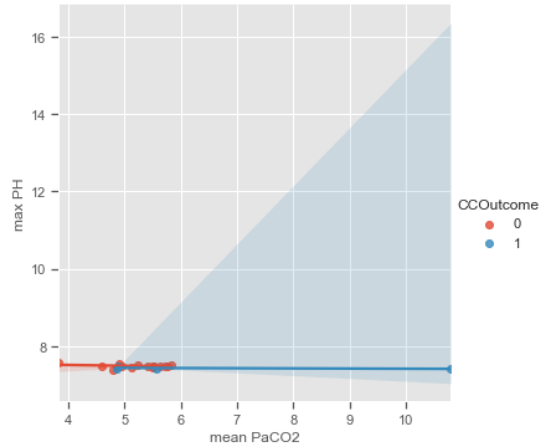

Figure 9: lmplot for max PH and mean PaCO2

The attribute 'labels_' contains the cluster labels that indicate which cluster each observation belongs to, and the attribute 'cluster centres' contains the

coordinates of cluster centres that represent the mean of all observations in the cluster. It is possible to map the clusters' centres and data in each cluster using various colours to differentiate the clusters using these two attributes:
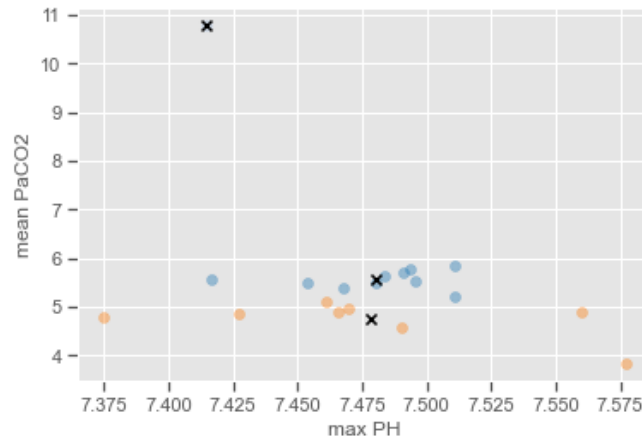

Figure 10: Scatter plot for mean PaCO2 and max PH

The algorithm is simple enough that it can be applied with only a few lines of code. If you want to see how the centres converge after many iterations, you can use the code below, which is a step-by-step implementation of the k-means clustering algorithm.
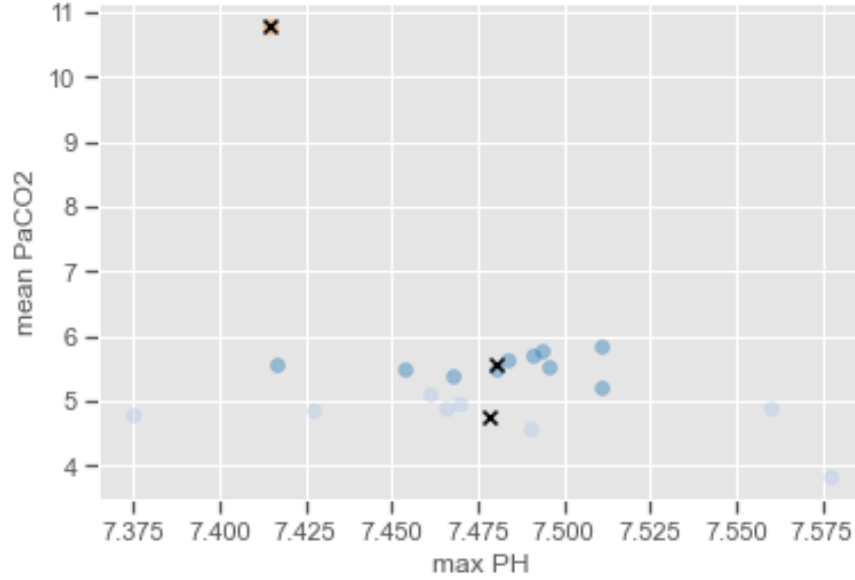
Figure 11: Scatter plot for mean PaCO2 and max PH after many iterations

5.1.2 Time series clustering

Time series research showed distinct and intriguing trends of both the living and the deceased. Following that, K-means clustering can be used to look for correlations in time series. The aim is to stratify patients based on their progression in the ICU, from admission to discharge, with each

component separately. It is worth noting that we are now dealing with time series data rather than built functions. It is important for this task and style of algorithm to normalise data for each patient separately. This allows for a study of time averages rather than magnitudes of measurements. In fact, if the data is normalised separately for each patient, clustering would appear to group together patients who (for example) began with the lowest values and ended up with the highest values, while if the data is not normalised, the same patients may end up in separate clusters despite a similar pattern.

5.1.3 Examining the right number of clusters for each vector visually

Temperature is clustered using k-means with a variable number of clusters (K). To classify the clusters, only the training data is used. The figures indicate, in descending order of appearance:
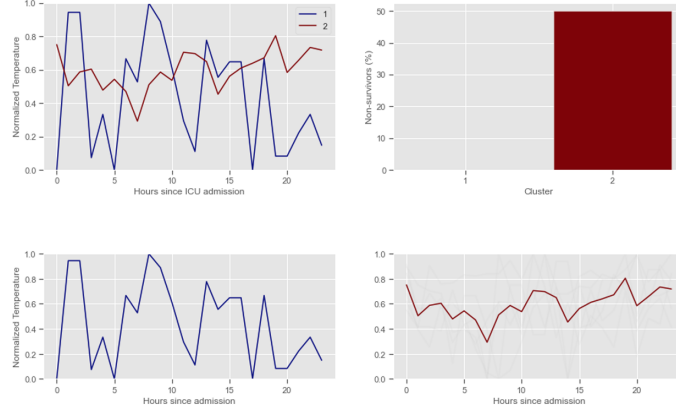
Figure 12: Using 7 ICU stays for creating the clusters; 2 clusters. Cluster 1: 1 observation. Cluster 2: 6 observations. Top left: shows cluster centres. Top right: shows the number of dead in each cluster. Bottom: shows cluster centres and preparation results in each cluster
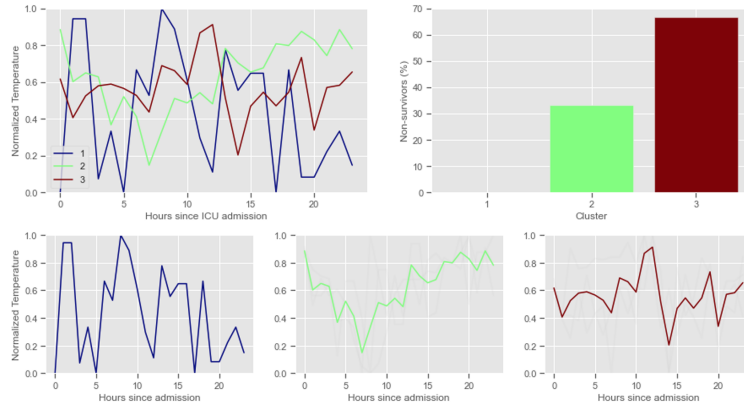

Figure 13: Using 7 ICU stays for creating the clusters; 3 clusters. Cluster 1: 1 observation. Cluster 2: 3 observations. Cluster 3: 3 observations. Top left: shows cluster centres. Top right: shows the number of dead in each cluster. Bottom: shows cluster centres and preparation results in each cluster
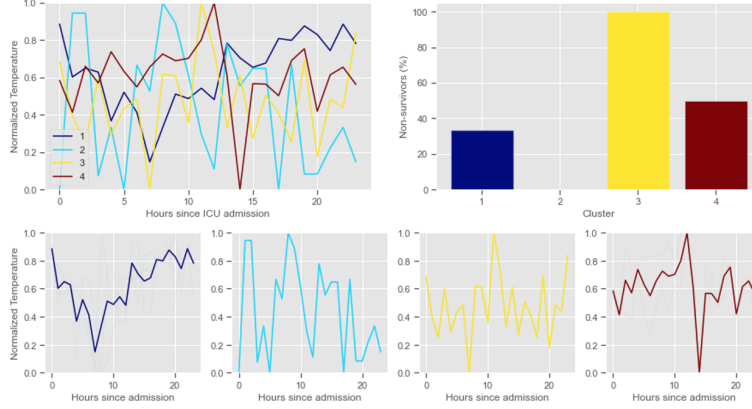
Figure 14: Using 7 ICU stays for creating the clusters; 4 clusters. Cluster 1: 1 observation. Cluster 2: 1 observations. Cluster 3: 1 observations. 4: 2 observations. Top left: shows cluster centres. Top right: shows the number of dead in each cluster. Bottom: shows cluster centres and preparation results in each cluster

We were able to stratify patients using clustering based on their physiological evolution within the first 24 hours in the ICU. Since cluster centres represent cluster tendencies, it is possible to examine the relationship between distinct physiological trends and mortality and determine whether or not the relationship is predicted. In terms of heart rate, for example, clusters 2 and 4 are more or less symmetric: in cluster 4, patients begin with a low heart rate that rises over time until it decreases again; in cluster 2, patients begin with a high heart rate that decreases over time until it increases again. While this is clearly insufficient to estimate mortality, it does suggest a potential connection between the evolution of heart rate and mortality (Alive or Dead). When a patient has a heart rate rhythm comparable to cluster 2, there could be more cause for alarm than if the pattern is similar to cluster 4.

Such characteristics of the method of normalisation conducted can now be identified:

- It impairs interpretability

- It helps the algorithm to group together patients that did not exhibit major improvements in their physiological condition over time, regardless of the absolute significance of the findings.
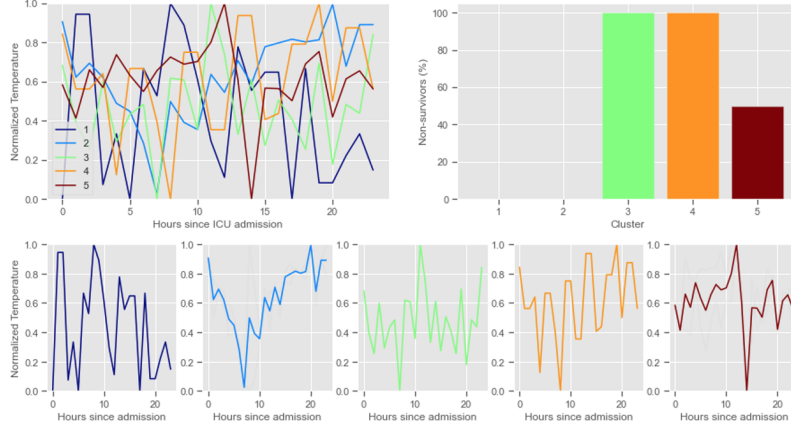
Figure 15: Using 7 ICU stays for creating the clusters; 5 clusters. Cluster 1: 1 observation. Cluster 2: 2 observations. Cluster 3: 1 observations. 4: 1 observations. 5 : 2 observations. Top left: shows cluster centres. Top right: shows the number of dead in each cluster. Bottom: shows cluster centres and preparation results in each cluster

5.2   Results for Supervised learning

With all the hyperparameters defined, the model is built. For Linear Regression,Logistic Regression and pipeline based model using K-means and Logistic regression. The model is built, and the model is used on test data, 3 models are trained, Linear regression method gets 2.543 error rate on test data, and Logistic regression gets 80.06% accuracy on test data. For the purpose of comparison, we provide accuracy of 2 models and the average model on test data,model 1 is with PCA which has an accuracy of 80.02%, and the 2nd model is with k-means which has an accuract of aprrox 80%.

6   Discussion

Although we tried many models in this experiment, there are still many limitations. Firstly, linear regression is one of ML models applied in our project to predict the weaning protocol from mechanical ventilation. Linear regression is a technique used to model and analyse the linear relationship between variables and analyse how the variables affect the results. Linear regression refers to a regression model composed entirely of linear variables. The general situation is multi-variable linear regression, which embodies the relationship between multiple independent input variables (feature variables) and output dependent

variables. The model remains linear because the output is a linear combination of input variables. The linear regression invests gradient descent or normal equation method to achieve the fitting problem, as a result that the training speed and prediction speed are faster, it performs well on small data sets, the results are interpretable and easy to explain and also when new data is added, it is easy to update the model. (Tài, 2017) [10] However, this model is not applicable to non-linear data and leads to overfitting problems. In addition, its prediction accuracy is generally low. The endurance to separating signal and noise is not ideal, and unrelated features need to be removed before using.

Logistic regression can be regarded as a generalized linear regression that invests maximum likelihood estimation (MLE) to fit the data and has the similar advantages and limitations with linear regression, for example, the training speed is also faster. While classifying, the computation is only related to the number of features. Both linear regression and logistic regression are 'white box' models, which are simple and easy to understand, and the interpretability of the model is very good (Behnood, 2018) [11] We can learn the attribution of different features to the final prediction from the weight of features. Whereas logistic regression cannot be used to solve nonlinear problems and is sensitive to multicollinearity data, as well as difficult to deal with the problem of imbalance data. Furthermore, the accuracy is not very high, because the form is very simple (very similar to a linear model), it is difficult to fit the real distribution of data. Logistic regression itself cannot filter features so that having to depend on other feature selection techniques to filter features.

Then the k-means clustering, as a simple model with a fast-fitting speed, is also introduced in our project. Similarly, its principle is relatively simple, and the algorithm is easy to implement because the main parameter that needs to be adjusted is only the value of clusters K. If denser the resulting clusters are, more obvious the difference between clusters and clusters is. In this case, k-means can work better. However, the value of K needs to be determined in advance, and it is difficult to determine a proper value of K in unsupervised clustering tasks because it is not known how many categories there are in the data set. K-means is easily affected by outliers, since k-means obtains the average value of samples in the cluster. For example, a sample with large value appears in a cluster on a certain dimension, which will make the cluster centre toward the outlier, resulting in a poor clustering performance (Ewan C, 2016) [12] Since k-means uses Euclidean distance to measure the similarity between samples, the obtained clusters are all convex, which cannot solve the clustering of "S-shaped" data distribution, as a result it is difficult to handle with non-convex fitting problems.

Meanwhile, another popular ML model called decision tree is employed, which is also easy to understand and implement and does not need users to understand a lot of background knowledge about the project. Data preparation is often simple or unnecessary with a decision tree algorithm, and it can handle various data types at the same time and provide feasible and effective results for a large data set. On the other hand, it is easy to lead to overfitting problems and ignore the correlation between features (Chinedu, 2021)[13] Also, decision tree

cannot process the unbalanced dataset, if dataset is unbalance, it commonly will have a prediction of belonging to the class whose data with more information. Finally, cross validation (CV) such as holdout, k-fold CV or leave-one-out, is ignored to determine parameters or assess all results in our project. Besides lacking hypothesis testing, t-testing or other testing methods, the results are provided only with the form of accuracy, which cannot guarantee the results are reliable.

There are a number of approaches that we can attempt to improve this project. 1) Feature selection algorithms can be introduced to implement data processing, such as forward feature selection, backward feature selection or two-stage steps, and that can depend on different criteria to analyse the relationship between features and classes for example correlation, information gain (IG), information entropy and so on. In this case, uncorrelated information would be eliminated to reduce disturbance to train models beforehand. 2) The regularization should be added to some linear regression methods, for instance lasso or elastic net. They cannot only achieve feature selection, but also reduce the level of overfitting. 3) Cross validation (CV) and hypothesis testing, t-testing or other testing methods should be used to determine parameters or assess all results and enrich the forms of results to guarantee the reliability. 4) Considering that the project is medical field, the prediction is a kind of advice or sign to assist a staff with diagnosis. Hence, the model should provide more information, not only the probability or cluster, but also provide the explanation.

7   Conclusion

The devised model, developed with the Logistic Regression algorithm, trained by data collected of patients in ICU , can predict the optimal time for weaning of patients in mechanical ventilators with high accuracy. Under the assistance of this model, the clinicians can implement a weaning process for patients who meet the conditions at the optimal time. This can effectively reduce the various complications that patients face due to improper weaning time, as well as the financial burden for both hospital and patients.

References

[1] Brochard, L., Rauss, A., Benito, S., Conti, G., Mancebo, J., Rekik, N., Gasparetto, A. and Lemaire, F., 1994. Comparison of three methods of gradual withdrawal from ventilatory support during weaning from mechanical ventilation. American Journal of Respiratory and Critical Care Medicine, 150(4), pp.896-903.Weaning from the ventilator and extubation in ICU.

[2] Boles, J., Bion, J., Connors, A., Herridge, M., Marsh, B., Melot, C., Pearl, R., Silverman, H., Stanchina, M., Vieillard-Baron, A. and Welte, T., 2007. Weaning from mechanical ventilation. European Respiratory Journal, 29(5), pp.1033-1056.

[3] Meade, M., Guyatt, G., Cook, D., Griffith, L., Sinuff, T., Kergl, C., Mancebo, J., Esteban, A. and Epstein, S., 2001. Predicting Success in Weaning From Mechanical Ventilation. Chest, 120(6), pp.400S-424S.

[4] Ely, E., Baker, A., Dunagan, D., Burke, H., Smith, A., Kelly, P., Johnson, M., Browder, R., Bowton, D. and Haponik, E., 1996. Effect on the Duration of Mechanical Ventilation of Identifying Patients Capable of Breathing Spontaneously. New England Journal of Medicine, 335(25), pp.1864-1869.

[5] Thille, A., Cortés-Puch, I. and Esteban, A., 2013. Weaning from the ventilator and extubation in ICU. Current Opinion in Critical Care, 19(1), pp.57-64.

[6] Kuo, H., Chiu, H., Lee, C., Chen, T., Chang, C. and Bien, M., 2015. Improvement in the Prediction of Ventilator Weaning Outcomes by an Artificial Neural Network in a Medical ICU. Respiratory Care, 60(11), pp.1560-1569.

[7] Walsh, T., Dodds, S. and McArdle, F., 2004. Evaluation of simple criteria to predict successful weaning from mechanical ventilation in intensive care patients. British Journal of Anaesthesia, 92(6), pp.793-799.

[8] Cismondia, F., Fialho, A,S., Vieirab S, M., Retic, S,R., Finkelstein S,R. and Sousab, J,M,C.Missing data in medical databases: Impute, delete or classify? Artificial Intelligence in Medicine, pp63

[9] Kuo, Hung-Ju et al. (2015) 'Improvement in the prediction of ventilator weaning outcomes by an artificial neural network in a medical ICU', Respiratory care, 60(11), pp. 1560–1569. doi: 10.4187/respcare.03648.

[10] Pham, Tài, Brochard, Laurent J and Slutsky, Arthur S (2017) 'Mechanical Ventilation: State of the Art', Mayo Clinic proceedings, 92(9), pp. 1382–1400. doi: 10.1016/j.mayocp.2017.05.004.

[11] Gholami, Behnood, Haddad, Wassim M and Bailey, James M (2018) 'AI in the ICU: In the intensive care unit, artificial intelligence can keep watch', IEEE spectrum, 55(10), pp. 31–35. doi: 10.1109/MSPEC.2018.8482421.

[12] Goligher, Ewan C, MD, Ferguson, Niall D, MD and Brochard, Laurent J, Dr (2016) 'Clinical challenges in mechanical ventilation', The Lancet (British edition), 387(10030), pp. 1856–1866. doi: 10.1016/S0140-6736(16)30176-3.

[13] Ossai, Chinedu I and Wickramasinghe, Nilmini (2021) 'Intelligent decision support with machine learning for efficient management of mechanical ventilation in the intensive care unit – A critical overview', International journal of medical informatics (Shannon, Ireland), 150, pp. 104469–104469. doi: 10.1016/j.ijmedinf.2021.104469.

[14] Guler, Hasan, Turkoglu, Ibrahim and Ata, Fikret (2014) 'Designing Intelligent Mechanical Ventilator and User Interface Using LabVIEW', Arabian Journal for Science and Engineering, 39(6), pp. 4805–4813. doi: 10.1007/s13369-014-1090-y.

[15] Walsh, T. S, Dodds, S and McArdle, F (2004) 'Evaluation of simple criteria to predict successful weaning from mechanical ventilation in intensive care patients', British journal of anaesthesia: BJA, 92(6), pp. 793–799. doi: 10.1093/bja/aeh139.

[16] Parreco, Joshua et al. (2018) 'Using artificial intelligence to predict prolonged mechanical ventilation and tracheostomy placement', The Journal of surgical research, 228, pp. 179–187. doi: 10.1016/j.jss.2018.03.028.