**Machine Learning Survival Algorithms for Predicting Ventilation Duration after Cardiac Surgery**

**Abstract**

**1. Introduction**

Despite advances in surgical and anaesthetic practices, prolonged mechanical ventilation (PMV) remains a significant post-operative complication of cardiac surgery. PMV has been reported in up to 2-22% of cardiac surgery patients[1] and is associated with an in-hospital mortality of greater than 40%, compared to 2-3% in patients who do not require PMV [2,3]. PMV is also associated with increased economic and resource burdens on hospitals [4]. Accurate prediction of PMV risk in cardiac surgery patients is desirable for perioperative planning, resource allocation, and timely intervention. Early weaning protocols may also be improved by accounting for individual risks of PMV, which may decrease the rates of extubation failure [5].

Several PMV risk models have been proposed for cardiac surgery patients, all of which only predict a binary outcome of PMV. That is, the model can only return a probability that a patient is ventilated beyond a predefined duration. However, binary risk models for PMV are not able to provide an estimated duration of ventilation, a task in which clinicians tend to poorly perform [6]. Early and accurate prediction of ventilation duration could play a vital role in improving ICU resource allocation and post-operative decision making for cardiac surgery patients.

Time-to-event analysis, also known as survival analysis, is a group of statistical methods used to analyse the length of time until an outcome occurs [7]. Compared to linear regression models which predict a continuous target outcome, time-to-event models predict the Kaplan-Meier survival function or cumulative hazards function for each patient, providing greater granularity of information for clinicians. Time-to-event models can also account for censored patients and competing risks. The Cox proportional hazards model is a common example of a time-to-event model. In recent years, rapid developments in machine learning have led to new models such as Random Survival Forests [https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0863-0], Gradient Boosted Cox models [], and Deep Cox models [] which rely on neural networks. Machine learning models, for time-to-event analysis [8] or otherwise [x3], have consistently outperform traditional statistical methods in other clinical domains with similarly complex and multi-natured data. In this paper, we aim to develop and validate three time-to-event (i.e., time-to-extubation) models (penalised Cox proportional hazards, Random Survival Forests, and Gradient Boosted Cox) for the prediction of ventilation duration, using data available within the first four hours of post-operative ICU admission.

Machine learning models that can predict PMV risk in cardiac surgery patients are useful tools for improving perioperative planning and resource allocation.

**2. Methods**

*2.1. Data sources and study population*

We developed and validated all models on retrospective data from the MIMIC-III and MIMIC-IV databases. MIMIC-III and -IV are large, publicly available, single-centre critical care databases comprising deidentified data from a total of more than \_\_\_ patients admitted to intensive care units of the Beth Israel Deaconess Medical Centre (Boston, MA, USA) between 2001 and 2012 (MIMIC-III) and \_\_\_ (MIMIC-IV) [9]. The database contains multimodal data including patient demographics, hourly bedside vital sign measurements, ventilation settings, laboratory test results, medications, procedures, clinician notes, imaging reports, and diagnostic coding using International Classification of Diseases and Ninth Revision (ICD-9) codes [9]. We extracted and combined MIMIC-III and -IV data using Python 3.9.7 and BigQuery []. We included all patients aged ≥18 years who underwent coronary artery bypass grafting surgery, cardiac valve surgery, or combined valve and coronary artery bypass grafting surgery. As the database only includes deidentified data, patient consent has been waived for research on MIMIC data.

*2.2. Data preprocessing and exploration*

We removed any variables with greater than 80% missing data and performed multiple imputation by chained equations (MICE) on the remaining data using the miceforest package version 5.2.2 in Python []. MICE is a multiple imputation method which typically relies on a procedure of iteratively regressing on each variable to arrive at a stable, imputed dataset. The miceforest package uses a random forest algorithm to predict missing data instead of linear or logistic regression in traditional MICE. We created two copies of the original dataset to be imputed and ran the MICE algorithm for ten cycles on each dataset. We took the mean of each imputed value across all imputed datasets to arrive at a single complete dataset. In this work, we make the assumption that data were either Missing At Random (MAR) or Missing Completely At Random (MCAR).

Before building the time-to-event models, we explored the data by conducting mutual information analysis. Mutual information is general measure of the mutual dependence between two variables [https://github.com/AnotherSamWilson/miceforest]. As opposed to other measures of dependence like the correlation coefficient, mutual information does not rely on any assumptions about the underlying distribution of the data and thus can capture non-linear dependencies. We also presented the mean (standard deviation) and count (percentage) for continuous and categorical variables, respectively. All variables were included in the main analysis to maximise model performance.

*2.3. Model development and validation*

We modelled the survival functions of ventilation duration using a penalized Elastic Net-penalized Cox proportional hazards as a baseline model and two machine learning algorithms - random survival forest and gradient boosted Cox regression. The Elastic Net penalty combines the subset selection and regularization properties of the LASSO and Ridge penalties which results in improved model stability and better handling of highly dimensional data. Model hyperparameters in the machine learning algorithms were selected using a grid search strategy with 5-fold cross-validation, using the concordance index for internal validation. The full grid search strategy and final tuned hyperparameters are available in Supplementary Files.

We evaluated model performance using the cumulative/dynamic area under the receiver operating characteristic curve (AUCC,D(*t*)). The AUCC,D(*t*) is a time-dependent extension of the traditional area under the receiver operating characteristic curve (AUC). Incorporating time-dependency into the traditional AUC is a robust method that provides more granular performance evaluation of time-to-event models []. The AUCC,D(*t*) gives the probability that for a randomly chosen pair of patients, one who has been extubated before *t* and the other after *t*, the model correctly assigns a higher score to the patient who is extubated first. We trained each of the four tuned models on a 9:1 train-test split of the full dataset. We then obtain the predicted cumulative hazards functions for each patient in the test dataset to generate an AUCC,D(*t*) function for each model.

Model opacity is a common drawback of machine learning models as there are no easily interpretable variable coefficients, such as the hazard ratio in the Cox proportional hazards model. However, various statistical methods can be used to estimate feature importance and improve the transparency and interpretability of machine learning models. In this paper, we estimate feature importance using the concept of permutation importance, which measures the decrease in model performance when a feature is randomly shuffled and made unavailable.

*To facilitate comparison with existing risk models, we also obtained binary predictions of PMV >7d and numerical predictions of ventilation duration from each model to evaluate the area under the curve (AUC) and root-mean-squared-error (RMSE), respectively.*

*2.4 Sensitivity analysis*

Increasingly, there is an awareness that model fairness across different ethnicities and sexes can be influenced by inherent biases in the underlying dataset. We evaluated model fairness by conducting sensitivity analysis on cohorts of different genders and ethnicity. We evaluated the same models trained on the 9:1 train-test split and compared their performance on four cohorts within the test data, specifically male versus female cohorts and White versus non-White cohorts.

**3. Results**

This study included a total of 10,430 patients who underwent cardiac surgery from the MIMIC-III and -IV databases (mean age, 67.0 years; 7393 males [70.9%]; 8843 White [84.8%]). The three most common types of cardiac surgery were coronary artery bypass graft (CABG) surgery (55.4%), isolated aortic valve surgery (17.5%), and combined CABG and valvular surgery (15.3%). The median duration of ventilation was \_\_\_\_\_ hours (IQR: ). Summary statistics for all variables can be found in Supplementary Materials Table 1.

Univariate analysis using mutual information identified \_\_\_ as the most predictive variables.

The

**4. Discussion**

**Conclusions**

**References**

1. Rose, L.; McGinlay, M.; Amin, R.; Burns, K.E.A.; Connolly, B.; Hart, N.; Jouvet, P.; Katz, S.; Leasa, D.; Mawdsley, C., et al. Variation in Definition of Prolonged Mechanical Ventilation. *Respiratory Care* **2017**, *62*, 1324-1332, doi:10.4187/respcare.05485.

2. Fernandez-Zamora, M.D.; Gordillo-Brenes, A.; Banderas-Bravo, E.; Arboleda-Sánchez, J.A.; Hinojosa-Pérez, R.; Aguilar-Alonso, E.; Herruzo-Aviles, Á.; Curiel-Balsera, E.; Sánchez-Rodríguez, Á.; Rivera-Fernández, R. Prolonged Mechanical Ventilation as a Predictor of Mortality After Cardiac Surgery. *Respiratory Care* **2018**, *63*, 550-557, doi:10.4187/respcare.04915.

3. Pappalardo, F.; Franco, A.; Landoni, G.; Cardano, P.; Zangrillo, A.; Alfieri, O. Long-term outcome and quality of life of patients requiring prolonged mechanical ventilation after cardiac surgery. *European Journal of Cardio-Thoracic Surgery* **2004**, *25*, 548-552, doi:10.1016/j.ejcts.2003.11.034.

4. Rajakaruna, C.; Rogers, C.A.; Angelini, G.D.; Ascione, R. Risk factors for and economic implications of prolonged ventilation after cardiac surgery. *The Journal of Thoracic and Cardiovascular Surgery* **2005**, *130*, 1270-1277, doi:10.1016/j.jtcvs.2005.06.050.

5. Blackwood, B.; Burns, K.E.A.; Cardwell, C.R.; O'Halloran, P. Protocolized versus non-protocolized weaning for reducing the duration of mechanical ventilation in critically ill adult patients. *Cochrane Database of Systematic Reviews* **2014**, 10.1002/14651858.CD006904.pub3, doi:10.1002/14651858.CD006904.pub3.

6. Figueroa-Casas, J.B.; Connery, S.M.; Montoya, R.; Dwivedi, A.K.; Lee, S. Accuracy of Early Prediction of Duration of Mechanical Ventilation by Intensivists. *Annals of the American Thoracic Society* **2014**, *11*, 182-185, doi:10.1513/AnnalsATS.201307-222OC.

7. Schober, P.; Vetter, T.R. Survival Analysis and Interpretation of Time-to-Event Data. *Anesthesia & Analgesia* **2018**, *127*, 792-798, doi:10.1213/ane.0000000000003653.

8. Spooner, A.; Chen, E.; Sowmya, A.; Sachdev, P.; Kochan, N.A.; Trollor, J.; Brodaty, H. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports* **2020**, *10*, doi:10.1038/s41598-020-77220-w.

9. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.-w.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Scientific Data* **2016**, *3*, doi:10.1038/sdata.2016.35.

DLS template: <https://jamanetwork.com/journals/jama/fullarticle/2665775>

Derivation and validation model: <https://reader.elsevier.com/reader/sd/pii/S002252231631039X?token=4EB6C36CF2B5FE3C504865473DD19C2409BBF826584F803EF01CA193EB3E5AA429C349FD0B1D0B82C1D8B26147D95295&originRegion=us-east-1&originCreation=20211010084455>

DLS original paper: <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0482-1>

DLS survival template: <https://link.springer.com/article/10.1186/s12859-020-3431-z>

Prediction of duration of mechanical ventilation by intensivist: <https://pubmed.ncbi.nlm.nih.gov/24069941/>

XGB model (C-index of 73.9 vs STS 72.6) for PMV >24h: <https://www-sciencedirect-com.ezproxy.lib.monash.edu.au/science/article/pii/S0003497520311565>

XBG predicting reintubation in PMV patients (C-index:0.908): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8104124/>

**Supplementary Material**

**Table 1:** General characteristics of the patients

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Total cohort (n=10430)** | **Mutual Information** | **R2 value** |
| Age, mean (SD) | 67.0 (11.7) | 0 | 0.001 |
| Male gender, n (%) | 7393 (70.9) | 0 | 0.002 |
| Body mass index | 29.4 (5.9) | 0.008 | 0.001 |
| Height | 170.8 (10.2) | 0.163 | 0.001 |
| Weight | 85.9 (19.0) | 0 | 0 |
| **Race, n (%)** |  |  |  |
| White | 8843 (84.8) | 0.001 | 0.003 |
| Black | 393 (3.8) | 0 | 0.001 |
| Hispanic | 358 (3.4) | 0 | 0.001 |
| Native | 53 (0.5) | 0.007 | 0.001 |
| Other | 499 (4.8) | 0.014 | 0 |
| **Type of Cardiac Surgery, n (%)** |  |  | 0.001 |
| Aortic, isolated | 1823 (17.5) | 0 | 0 |
| Mitral, isolated | 851 (8.2) | 0 | 0 |
| Tricuspid, isolated | 44 (0.4) | 0 | 0.004 |
| Pulmonary, isolated | 9 (0.1) | 0.002 | 0 |
| Coronary artery bypass surgery (CABG), isolated | 5779 (55.4) | 0.005 | 0.007 |
| Multiple valve surgery | 326 (3.1) | 0.005 | 0.005 |
| Combined valve and CABG | 1598 (15.3) | 0.005 | 0.005 |
| **Medical history, n (%)** |  |  |  |
| Arrhythmia | 5013 (48.1) | 0.008 | 0.01 |
| Congestive cardiac failure | 2883 (27.6) | 0.036 | 0.021 |
| Acute myocardial infarction | 2836 (27.2) | 0 | 0.002 |
| Cerebrovascular disease | 829 (7.9) | 0 | 0.007 |
| Type 2 diabetes | 2061 (19.8) | 0.037 | 0 |
| Type 1 diabetes | 85 (0.8) | 0.016 | 0 |
| Diabetes, uncomplicated | 2746 (26.3) | 0 | 0 |
| Diabetes, complicated | 833 (8.0) | 0.003 | 0 |
| Peripheral vascular disease | 1556 (14.9) | 0.005 | 0.009 |
| Chronic obstructive pulmonary disease | 1967 (18.9) | 0.008 | 0.001 |
| Chronic kidney disease | 1269 (12.2) | 0.015 | 0.003 |
| Liver disease, severe | 16 (0.2) | 0 | 0.001 |
| Liver disease, mild | 328 (3.1) | 0.006 | 0.007 |
| Peptic ulcer disease | 65 (0.6) | 0 | 0.001 |
| Rheumatological disease | 263 (2.5) | 0.003 | 0 |
| Malignancy | 227 (2.2) | 0.002 | 0 |
| Metastatic cancer | 21 (0.2) | 0.002 | 0 |
| Paraplegia | 67 (0.6) | 0 | 0.002 |
| Dementia | 39 (0.4) | 0.001 | 0 |
| AIDS | 13 (0.1) | 0.009 | 0 |
| Smoking, current or previous | 2836 (27.2) | 0.034 | 0.002 |
| **Ventilation settings, first measurement within 4h after intubation, mean (SD)** |  |  |  |
| Pulmonary artery diastolic pressure (mmHg) | 18.7 (5.7) | 0.062 | 0.037 |
| Tidal volume, observed, mL | 569.8 (127.8) | 0.074 | 0 |
| Tidal volume, set, mL | 505.5 (60.1) | 0.021 | 0 |
| Minute volume, L | 8.2 (2.2) | 0.073 | 0.002 |
| Plateau pressure, mmHg | 18.7 (3.8) | 0.021 | 0.025 |
| Peek end expiratory pressure, mmHg | 5.4 (1.3) | 0.034 | 0.028 |
| Respiratory rate, set, bpm | 15.4 (2.0) | 0.01 | 0.018 |
| Respiratory rate, total, bpm | 16.3 (3.6) | 0.022 | 0.003 |
| Respiratory rate, spontaneous, bpm | 4.9 (7.4) | 0.098 | 0 |
| **Vital sign predictors, mean (SD)** |  |  |  |
| Heart rate, bpm | 83.9 (12.3) | 0.024 | 0.005 |
| Respiratory rate, bpm | 17.3 (4.7) | 0.037 | 0 |
| Peripheral oxygen saturation, % | 98.6 (2.3) | 0.021 | 0.003 |
| Systolic blood pressure, mmHg | 111.5 (16.2) | 0.022 | 0.007 |
| Diastolic blood pressure, mmHg | 57.7 (9.6) | 0.006 | 0.002 |
| Mean arterial pressure, mmHg | 75.0 (11.1) | 0.02 | 0.003 |
| Temperature, °C | 36.8 (0.6) | 0.094 | 0 |
| Cardiac index | 2.7 (0.7) | 0.018 | 0.001 |
| SOFA score | 5.3 (2.7) | 0.081 | 0.071 |
| **Laboratory values, mean (SD)** |  |  |  |
| pt\_first | 13.9 (2.4) | 0.017 | 0.019 |
| ptt\_first | 32.0 (10.1) | 0.039 | 0.031 |
| inr\_first | 1.3 (0.2) | 0.008 | 0.017 |
| hb\_first | 10.3 (1.4) | 0.011 | 0 |
| wcc\_first | 13.0 (4.9) | 0.003 | 0 |
| chloride\_first | 106.9 (3.4) | 0.014 | 0.002 |
| magnesium\_first | 2.3 (0.5) | 0.064 | 0.007 |
| potassium\_first | 4.3 (1.3) | 0.005 | 0 |
| creatinine\_first | 1.1 (0.8) | 0.026 | 0.011 |
| free\_calcium\_first | 1.1 (0.1) | 0.015 | 0.006 |
| sodium\_first | 138.1 (2.9) | 0.017 | 0.009 |
| bicarb\_first | 23.3 (2.4) | 0.022 | 0.01 |
| bun\_first | 18.0 (9.2) | 0.024 | 0.026 |
| glucose\_first | 138.1 (35.4) | 0.037 | 0.013 |
| lactate\_first | 2.2 (1.4) | 0.033 | 0.05 |
| aki\_first | 0.1 (0.4) | 0.016 | 0.012 |
| po2\_first | 76.9 (70.8) | 0.144 | 0.002 |
| pco2\_first | 40.8 (5.9) | 0.013 | 0.001 |
| baseexcess\_first | -1.3 (2.5) | 0.024 | 0.003 |
| ph\_first | 7.4 (0.1) | 0.024 | 0 |
| fio2\_first | 54.6 (45.0) | 0.225 | 0.001 |
| insulin\_first | 41.8 (100.9) | 0.162 | 0.032 |