The Alan Turing Institute

Data Study Group Final Report: Great Ormond Street Hospital

8 – 12 April 2019

Augmenting clinical decision-making in intensive care

https://doi.org/10.5281/zenodo.3670726

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1

Contents

1 Executive summary

Challenge overview

This report presents the outputs of a week-long collaboration between The Alan Turing Institute and Great Ormond Street Hospital. The purpose was to scope how vital signs monitoring data can be better used to inform the removal of a breathing tube (i.e. 'extubation') in intensive care units (ICUs).

Data overview

All children were admitted to an ICU at Great Ormond Street Hospital (GOSH) between 2016 and 2018. Data included physiological monitor outputs recorded every 5 seconds (including heart rate, respiratory rate, blood pressure, etc.), relevant demographic information, and flags for time of first extubation, including whether the attempt was successful.

Main objectives

Evaluating the effectiveness of different modelling methods for predicting whether an extubation attempt was successful (a failed attempt was defined as re-intubation within 48 hours of attempted extubation).

Methods

1. Automatic feature extraction using the 'tsfresh' library.
2. Clinical expertise driven manual feature engineering
3. State-of-the-art time series classification algorithms using the 'sktime' library

Conclusion

Automatic feature extraction and specialized state-of-the-art time series classification algorithms did not perform better than a naive (statistical) baseline, i.e. a model that makes a simple guess of the outcome. In contrast, domain expertise driven feature engineering, based on the duration and frequency of missing values in the end-tidal CO_2 measurements (hypothesized to represent episodes of coughing or secretion suctioning), improved the predictive accuracy for cases of failed extubation.

2 Introduction

Approximately 20,000 children in the UK are admitted to an intensive care unit (ICU) annually. Whilst recovery rates are quite high (just over 95% of admissions recover), this requires constant care and monitoring [1]. To aid recovery, many of these children require organ support in the form of drugs (e.g Inotropes & vasopressors) and machinery (e.g. Haemofiltration & ventillation). However, the decision to wean a child off organ support is not straightforward and requires clinicians to determine the exact point in time at which a child is 'sufficiently recovered', such that their organs can independently support them. Withdrawing support too early could lead to a deterioration in the child's health, whereas waiting too long could result in unnecessary treatment that exposes them to a greater risk of complications. Furthermore, once successfully weaned off organ support, the patient can be transferred from the ICU to a general ward, which is important both for the patient's recovery (ICU's are not known to be calm environments) and for the safe and effective functioning of the hospital, as ICU beds are always in high demand.

In this study we have looked at predicting the optimal timing for removing ventilatory support (i.e. removing the tube placed directly into the wind-pipe (i.e. trachea), to support breathing). Currently, the decision of when to remove the aforementioned tube, a process known as extubation, is determined by specialist physicians, using the patient's vitals (e.g. heart rate, breathing rate, the level of oxygen in the blood, etc), and several other sources of information. However, in 10% of extubation cases, the breathing tube has to be re-inserted (re-intubation) within 48 hours. Thus, there is clearly a role for a more accurate and precise methods of determining the ideal extubation period for a specific patient. However, attempts to apply predictive modelling to this research area have been limited due to a scarcity of data [2].

Great Ormond Street Hospital (GOSH) is currently in a unique position: (1) it is the only childrens hospital in the UK with the capability to record high resolution monitor data via the Etiometry T3 system (Etiometry Inc, MA, USA); and (2) it is one of the first hospitals within the NHS to have a digital research infrastructure, with a physical space (GOSH DRIVE Digital Research Informatics and Virtual Environment unit) and a Digital Research Environment (DRE). The DRE team operates at the interface between the clinical teams at GOSH and the virtual environment. The latter is used to extract relevant data from multiple sources. Then, these data are de-itentified and preprocessed, so that they can be used for analysis. This report describes the results of an initial collaboration between GOSH and The Alan Turing Institute, analyzing a series of high resolution datasets to address the aforementioned lack of robust research surrounding the use of advanced predictive modelling methods to address the issue of weaning organ support in paediatrics.

3 Data Overview

The data was extracted and pre-processed by GOSH's Digital Research Environment (DRE) team, using their digital research environment (GOSH DRIVE), as described below:

1. ***Identification and anonymisation of ICU admissions***
   To compile a list of patients in the cohort under consideration for this study, the unique hospital IDs of all individuals that were admitted to a GOSH ICU ward between Jan 1st 2016 and December 31st 2018 were extracted from a pre-existing list of ICU admissions. De-identification was subsequently achieved by: 1) replacing the unique hospital IDs with random fixed IDs (so as to allow linkage across datasets); and 2) by shifting all recorded dates (e.g. admission, extubation, intubation, etc.) by a constant number of days that was determined by randomly sampling from the discrete Uniform (-6, 6) distribution.

2. ***Extraction of patient monitor data from the T3 system***
   A Python (v3.7) script was written to extract ICU monitor data from the T3 clinical system for each patient in the list compiled previously. Monitor data was then collated and written to a series of patient-specific .csv (comma-separated value) files; where each .csv included all the data from the different measurement methods (i.e. invasive vs. non-invasive) for each vital sign (e.g. Oxygen Saturation, Heart Rate, Blood pressure, etc.). A summary dataset was then produced for each individual in the patient cohort detailing the number of rows written for each group of monitor variables for that individual.

3. ***Identification of extubation times and labelling 'success' or 'failure'***
   First, structured and free-text data was extracted for each patient from the Carevue clinical system. Then, a combination of programmatic analysis and expert clinical opinion was applied to the structured ventilation data to determine whether an extubation had been attempted for a patient. A subjective categorical validity score was assigned to each identified extubation attempt, reflecting the experts' opinion on the validity of the associated timestamp for the extubation attempt. Unless evidence of a failed extubation was present for an extubation event, a successful extubation label was applied. Finally, to confirm the accuracy of the failed extubation labels, the string tuba was used to identify relevant entries in free-text fields that referenced the extubation process. Matching entries were then compared to a pre-determined list of 3-word-contiguous phrases identifying a failed extubation event. The date-time stamp of the first occurrence of a phrase that identified a failed extubation event in the patient record was taken as the indicative time of the event. The labels, anonymized IDs, extubation timestamps, and validity scores were then written out to a separate .csv file.

4 Predicting Successful Extubation

4.1 Problem Definition

Task To build a clinical decision support tool that provides a score illustrating the risk of failure for a given extubation attempt.

Clinical Context Physicians make the decision to extubate up to 6 hours prior to the event. Before the actual act of extubation (i.e. removal of the endotracheal tube), the patient is gradually weaned off the ventilation to encourage them to be more self-sufficient. Currently this decision is made based on an understanding of monitored physiological signals.

Problem Formulation We treated the problem as a classification task. Formally estimating the function,

$$\hat{g} : \mathcal{X} \rightarrow [0, 1] \tag{1}$$

where $\mathcal{X}$ is the set of features, i.e. different summaries and time windowing of partial time series data of sensor measurements observed up to six hours before extubation. The vital signs included in the subsequent analysis are: etCO2 (end-tidal CO2), SpO2 (Oxygen saturation), HR (Heart Rate), and RR (Respiratory Rate). The prediction is a probability of a failed extubation. We assume an extubation is either a failure or a success such that if we predict some value p for the probability of failure, then the probability of success is given by $1 - p$.

4.2 Data Overview

From the data described in Section 3, two partial time series datasets were generated:

1. The first dataset consisted of the 48 hours of vital sign measurements leading up-to the extubation-decision point (i.e. 6 hours prior to extubation attempt). This data was up-sampled to 1-minute frequency using the mean of the 5-second frequency of measurements for heart rate (HR), respiratory rate (RR), end tidal CO2 (etCO2), oxygen saturation (spO2)
2. The second dataset consisted of the last six hours of vital sign measurements leading up-to the extubation-decision point (i.e. 6 hours prior to extubation attempt), at the original 5-second frequency of measurements for heart rate (HR) and respiratory rate (RR).

Both datasets were enriched with time invariant features, including: age in days at admission to the ward, the patient's ward, and the number of days the patient had been in hospital prior to the attempted extubation.

Exclusion and Missingness Patients were excluded from the data if: 1) They had less than two days of measurements; or 2) more than 25% of measurements were missing. In all other cases, missing values were linearly interpolated by previous and subsequent values.

Outcome Recording The binary target variable was defined as 0 if extubation was successful, or otherwise 1, when an attempt was recorded as a failed extubation (a failed attempt was defined as re-intubation within 48 hours of attempted extubation). This target variable was generated by using the 'failed extubation flag' provided by the clinicians.

4.3 Methods

Time series classification In these experiments, the algorithms were trained on the past-two-days and past-six-hours feature datasets using only the multivariate time series measurements without including the time invariant features. Trained algorithms were evaluated on an independent held-out sample of patients. A stratified train-test split using 25% of the patients for the testing was utilised.

Models were implemented using the 'sktime' Python toolbox for time series classification algorithms [3]. The algorithms utilised, were:

- time series forest [4]
- random interval spectral ensemble based on the time series forest algorithm [5]

Series as features An alternative approach was also trialled, where all of the time-points for all measurements were utilized as separate variables, ignoring their temporal ordering, in addition to the time-invariant features (i.e. sex, age on admission and hospital ward indicators).

Automated feature extraction The last set of automated methods attempted was that of using random forests to identify important features and subsequently using those features as the basis for prediction. For these experiments stratified sampling to split the data into 70% train, 30% test to train the models was used. These experiments were carried out using the Python library 'tsfresh' [6].

Expertise driven feature engineering Finally, in recognition of the limitations of automated feature extraction and feature engineering, i.e. advanced computational methods can only identify variables correlated with the outcome but are limited by the fact that they cannot understand important real-world findings, we consulted experts in the field to identify any other potential features. Discussions with medical staff generated the postulate that the breaks in CO2 measurements could serve as a proxy for coughing, i.e. evidence of in dependant respiratory effort. As such, a series of features were derived based on the frequency and duration of bouts of coughing. Using the 'etCO2' data a 'coughing indicator' (CI) was created for each patient such that,

$$CI = \begin{cases} 1, \text{if etCO2 data was missing for less than one hour} \\ 0, \text{otherwise} \end{cases} \tag{2}$$

This makes the assumption that if data is missing for a short period of time (less than one hour), this is likely due to the patient coughing and the data is not being recorded as the patient is helped through this period by removing the breathing tube. Data missing for longer than one hour is unlikely to be an extensive episode of coughing and much more likely to be a temporary malfunction in data collection or better explained by another intervention. An example of this data for a random patient is shown in Fig. 1.

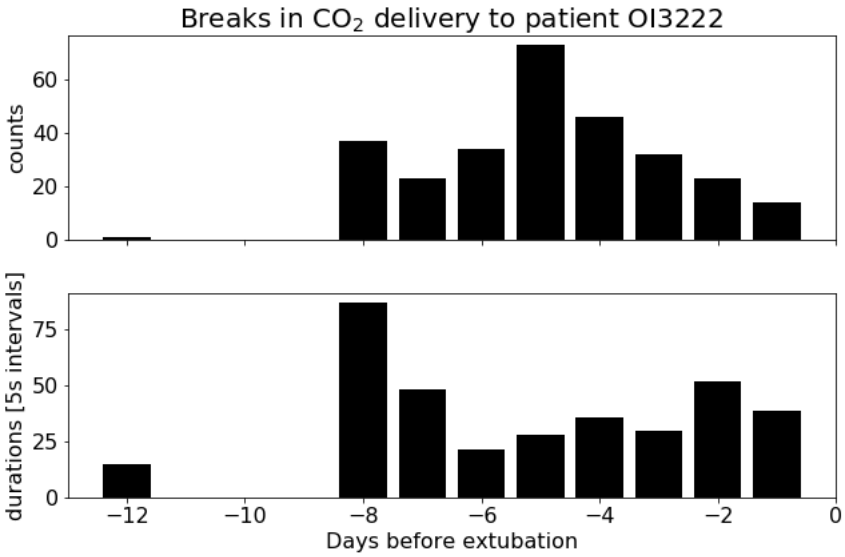

Figure 1: Demonstrating breaks in etCO2 data for a randomly-selected (anonymised) patient

A gap-finding algorithm was run on all patients excluding a sub-group of 400 from a single ICU (Flamingo Ward), to generate four features: (1) the number

of etCO2 gaps on the day of extubation; (2) the average gap duration on the day of extubation; (3) the average number of gaps per day throughout the entire ICU admission; and (4) the average gap duration across all gaps during the ICU admission. The subgroup of 400 patients from the Flamingo ward was split into train (75%) / test (25%) sets. A support vector classifier (SVC) was trained on these four features to predict the probability of a failed extubation.

4.4 Results

fig:TSR illustrates the precision over all potential thresholds when the time series for each individual features is used to predict the outcome. The results clearly illustrate that in isolation, each vital sign time series is a reasonably poor predictor of failed extubation.

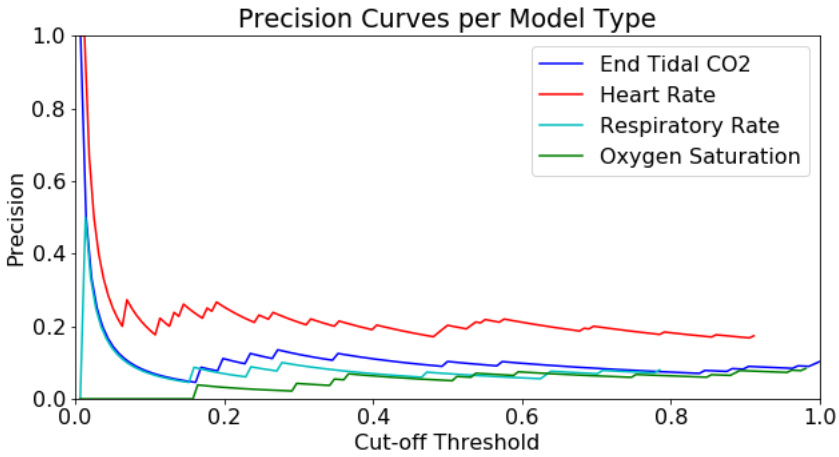

Figure 2: Using each individual vital sign time series as a feature

The results in Table 1 illustrate that the different time series classification methods (i.e. Forest & Random Interval Spectral) are no better at predicting failed extubation than the naive baseline. Moreover, when the series' (illustrated in 2) are utilized as features to train a Random Forest algorithm, the resulting model is also no better than the naive baseline. In effect, the results demonstrate that state-of-the-art off-the-shelf methods are not capable of identifying sufficiently insightful relationships to produce clinically (or even statistically) relevant improvements in decision-making relative to guessing.

Given the failure of the automated methods, we sought to leverage the experience of the ITU clinicians and experts available during the Data Study Group. The Support Vector Machine (SVM) trained on the end-tidal CO2

Algorithm	Log Loss	Standard Error
Naive Baseline	0.312	0.063
Time Series Forest	0.341	0.075
Random Interval Spectral Ensemble	0.315	0.062
Series as Features using Random Forest	0.318	0.064

Table 1: Results from last two days feature data. No sophisticated model outperforms the nave baseline.

coughing intervals achieved 85% precision and 60% recall, out-performing the majority baseline (uninformed) classifier (see table:confusion).

		Predicted	
		Failure	Success
True	Failure	92	1
	Success	4	6

Table 2: Confusion matrix for support vector classifier, threshold =

There are a number of reasons for why this could be the case, some of which are medical, and others due to hidden variables (i.e. the breaks could indicate not only coughing, but also interventions). For future research, it is would be interesting to look at the gap frequency, duration and counts filtered by minimum & maximum duration (e.g. gaps shorter than 5 minutes versus those over 30 minutes) – different gaps might prove to be more robust classification features (see fig:c02).

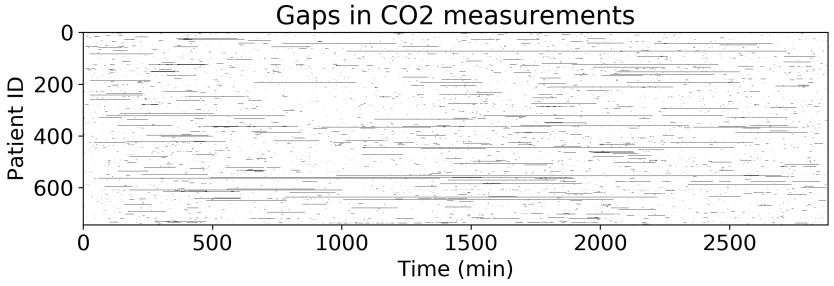

Figure 3: Breaks in CO2 measurements for all patients

5 Limitations & Challenges

Taking into consideration the clinical context of ICUs, the data extraction setup in place, and the clinical decision-making process which we are trying to inform, the following limitations and challenges have been highlighted to be inform future research.

Subjectivity of Extubation There are two methods by which subjectivity is introduced into this dataset. The first occurs as a consequence of imperfect decision making. Although in most cases there will be clear signs that the patient requires breathing support again, post-extubation, other situations exist when this is much harder to determine accurately. This means that it is possible a patient was re-intubated when they didn't necessarily need to be, or was re-intubated after 48 hours (which is not currently captured in our 'failure' label). Whilst this decision is made by clinical experts, there is still subjectivity being introduced as these experts do not necessarily make decisions in a perfectly uniform manner. Handling this inter-clinician variability is difficult without adequate labelling (a challenge we touch on later). The second layer of subjectivity occurred at the point of determining the exact extubation timestamp, which as discussed in Section 3 was not straightforward to label. The uncertainty introduced by the aforementioned methods for identifying the extubation and re-intubation is not known at present, and therefore somewhat undermine the models' ability to make predictions. In the next iteration of these experiments it would be useful to explore the intra-individual variation in the time-stamps produced by this automated method.

Censoring and Missingness In the context of medical data, high levels of missingness and censoring are always expected. Censoring occurs when a patient drops out of the study, for a reason that may be dependent or independent of the outcome. In this case, if a patient died after admission and before extubation then they would be censored. In the experiments described above, these instances were case-wise excluded from the analysis. A more subtle case occurs if a patient dies within the 48 hours following extubation. It may be clear their death is due to extubation and therefore the extubation should have been labelled as a 'failure', however if they die for any other reason then they should be censored. The current work does not make use of state-of-the-art methods for including censoring in the analysis (for a variety of reason), and thus, this is a limitation of the work.

Labels, and the lack thereof An important challenge in the utilization of high resolution data in this context is that distinct clinical events can present with highly similar morphological time series patterns. This issue is exacerbated

by the fact that these periods of significant deviation from the baseline are unlabelled, and therefore it is difficult to discern their relevance for the task at hand, and even when logical assumptions can be made, the naive methods employed to capture these episodes inevitably captures unrelated events. For example a 5 minute long absence of end tidal CO2 measurements can be produced by a number of events, from clinical interventions (e.g. removal of the end tidal co2 monitor to allow for suctioning of secretions), to a change in the physiological state of the patient (e.g. cardiac arrest), however a simple algorithm (especially in the absence of labels), will interpret them as being equivalent. Removal or labelling of these instances is an important element of data pre-processing that should be taken into consideration, however, due to time limitations it was not implemented in this analysis. In order to address this challenge, we have identified the following approaches:

1. At the Data Preprocessing Stage: Integrate medical records with the existing database to include time-stamps of clinical interventions versus other physiological events.
2. At the Data Analysis Stage: Produce metrics using multivariate analysis techniques in order to capture information that is stored in the physiological state of the patient as opposed to metrics produced through univariate analysis.
3. At the Algorithmic Design Stage: Utilize frameworks that take into consideration the physiological state of the patient and the interactions between the organ systems represented by the recorded signals. The production of state tracking algorithms could provide clinically significant insights to medical personnel.

6 Future Work

6.1 Reinforcement Learning

Problem Formulation From a framework of model-free and value-based Reinforcement Learning (RL), we sought to create an agent capable of imitating an ICU physician by learning a policy function which represents the decision making process, to extubate or not, given the current state of the patient [7]. Formally:

- We define the states, $s_t = (x_t, z_t)$, as the physical data x_t observed in the ICU, and the amount of time since the patient entered the ICU, z_t
- The doctor's decision, σ_t given by,

$$\sigma_t = \begin{cases} (1,0) & \text{if doctor extubates at time } t \\ (0,1) & \text{otherwise} \end{cases} \tag{3}$$

- There are two possible actions, a_t, given by,

$$a_t = \begin{cases} 1 & \text{extubation should start} \\ 0 & \text{extubation should not start} \end{cases} \tag{4}$$

- The policy function, $\pi(s_t)$, is the conditional distribution of extubating at time t, given by

$$\pi(s_t) = (P(a_t = 1|s_t), P(a_t = 0|s_t)) \tag{5}$$

- We denote the (first) extubation time as D

Reinforcement learning models are not assessed by comparing a prediction to a ground truth observation (as is the case in supervised learning). Instead a reward function is established and the agent tries to maximise the expected reward. We define the reward, r, by

$$r(s_t, a_t, s_{t+1}) = \begin{cases} -KL(\sigma_t||\pi_t) - (z_t - D)^2, & a_t = 1 \\ -KL(\sigma_t||\pi_t), & a_t = 0 \end{cases} \tag{6}$$

where $KL(\sigma_t||\pi_t)$ is the Kullback-Leibler divergence, which measures the deviation between the doctor's decision and the agent's decision

We use the weighting $(z_t - D)^2$ to attach more importance to the accuracy of the model at the early time. The expected return is given by

$$R(\pi) = E\left[\sum_{t=1}^{T} r(s_t, a_t, s_{t+1})\right] \tag{7}$$

where T is the time when first-time extubation happened following the strategy π

Data Overview The data is utilized in two ways: 1) to represent the patient's state with monitored autocorrelated variables at time t; and 2) represent doctor's extubation decision with a sparse vector ending with 1. We have been formatted the data into a matrix with 3 dimensions (patient, time, monitor).

Methods The reinforcement learning framework *gym* (OpenAI) was used to build the environment for this task, and the deep learning library *TensorFlow* (Google) was used to train the policy function.

The policy function is modeled by by a 2-layer neural network with 32 hidden units and $\tanh$ activation function.

We adopted the simplest policy optimization method: vanilla policy gradient and leave more complicated algorithms for future work.

Unfortunately, due to the time constraints, we did not finish training the policy but we established the environment with which the agent can interact.

6.2 Deep Learning

There are a number of deep learning methods suited to time-series data. During the Data Study Group we considered RNNs and 1-dimensional CNNs (implemented in TensorFlow using code available on Aymeric Damien's GitHub), in the context of predicting successful extubation of a patient from observations made within the two days prior to extubation. Initial work during the DSG focused on using the hourly averages of the patients' observations (for 40 patients) to develop the necessary code-base and format the data so as to run a simple LSTM model. Deep learning models typically require thousands of patients, and the RNN was not able to make good predictions when trained on 30 patients, achieving baseline performance on the remaining 10 test patients. Subsequently 500 patients with observations averaged, instead of hourly, over each minute was sampled from the dataset. Due to the computational complexity of linking the patient observations and summary data, it was not possible to do so for more patients within the given time frame. Moreover, we were unable to run this experiment to completion and thus this remains an issue for future research.

6.3 Auto-regressive (AR) models

It was hypothesized that the ICU vital sign monitor data likely demonstrates a high degree of auto-correlation, and as such we sought to determine the value of AR models in this setting. We extracted a 6 hour window pre- and post-extubation dataset, with values up-sampled by taking the hourly mean of the data. Our aim was to use this data to fit an auto-regressive model (AR) to each individual time-series, and then examine the predicted variance to see if differences in variance parameters corresponded to different clinical or demographic features. All of this analysis was to be performed using the tseries and forecast packages within R. Unfortunately, due to time constraints we were unable to complete this analysis. Future work in this domain would also include a robust assessment of how more complex AR models, such as GARCH models (which can capture changing variances), would perform in this setting.

6.4 Survival Analysis

In the field of survival analysis, the task is to predict the time to an event (usually death). In the context of extubation, we could create a model that predicts the time to successful extubation. It is unclear at this stage whether this would be a better strategy than others discussed previously, however it would provide a mechanism by which to investigate for factors associated with a positive outcome. For example by stratifying patients, we could identify how estimated survival curves differ between groups, which could be useful for feature engineering. Given more time we would have carried out an in-depth analysis using survival analysis techniques to predict time-to-extubation, however to illustrate the potential useful of these methods, below is a brief example of how survival analysis could be applied to predicting mortality from the data provided.

A series of survival curves were generated based on the Cox proportional hazard model, utilizing the following covariates: patient's age, gender, and the type of intensive care unit (ICU). Three types of ICU were considered: Neonatal Intensive Care Unit (NICU), Paediatric Intensive Care Unit (PICU), and Cardiac Intensive Care Unit (FLAMI). Age was grouped into four categories: 0 - 1 month old, 1 month - 1 year old, 1 - 5 years old, and $>$ 5 years old. The analysis was limited to children for whom the discharge date from the Intensive Care Unit (ICU) was consistent with the recorded date of death (i.e. 60% of total deaths/240 deaths). Time to death was defined as time in days from the ICU admission to recorded time of death.

Survival curves for gender, age categories, and ICU type are shown in fig:surv. Notably, the hazard of death was significantly lower for the 1 month - 1 year old, 1 - 5 years old, and $>$ 5 years old age categories (reference category: 0

- 1 month old), with HR = 0.39 (0.27, 0.57), HR = 0.34 (0.22, 0.54), and HR = 0.58 (0.39, 0.86) respectively.

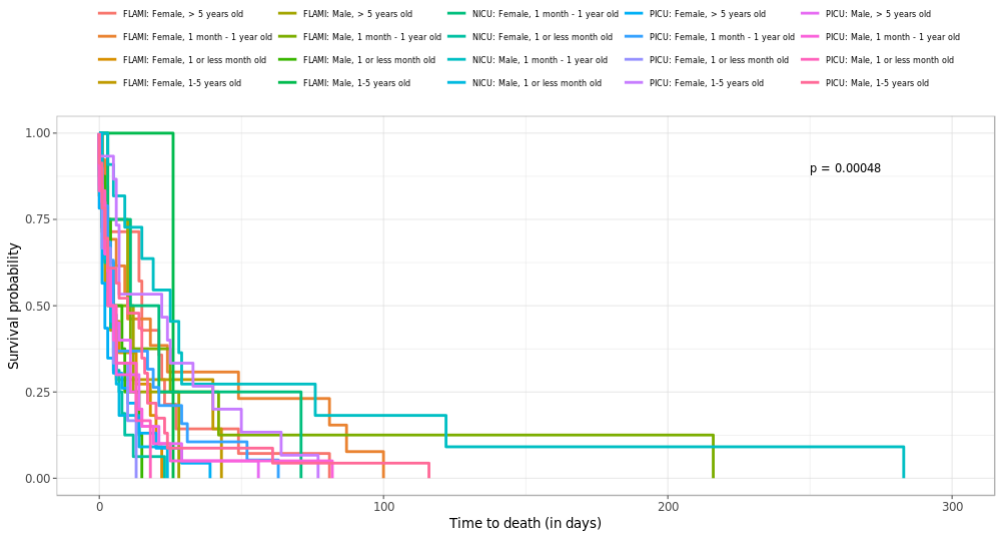

Figure 4: Survival curves for gender, several age categories, and ICU type with respect to time to death.

7 References

[1] Wood D, Goodwin S, Pappachan J, Davis P, Parslow R, Harrison D, Ramnarayan P. Characteristics of adolescents requiring intensive care in the United Kingdom: a retrospective cohort study. Journal of the Intensive Care Society. 2018 Aug;19(3):209-13.

[2] Eytan D, Goodwin AJ, Greer R, Guerguerian AM, Mazwi M, Laussen PC. Distributions and Behavior of Vital Signs in Critically Ill Children by Admission Diagnosis. Pediatr Crit Care Med. 2018 Feb;19(2):115-124.

[3] Lning M, Bagnall A, Ganesh S, Kazakov V, Lines J, Kirly FJ. sktime: A Unified Interface for Machine Learning with Time Series. arXiv preprint arXiv:1909.07872. 2019 Sep 17.

[4] Deng H, Runger G, Tuv E, Vladimir M. A time series forest for classification and feature extraction. Information Sciences. 2013 Aug 1;239:142-53.

[5] Bagnall A, Lines J, Bostrom A, Large J, Keogh E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery. 2017 May 1;31(3):606-60.

[6] Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time series feature extraction on basis of scalable hypothesis tests (tsfresha python package). Neurocomputing. 2018 Sep 13;307:72-7.

[7] Prasad et al 2017. A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units. https://arxiv.org/pdf/1704.06300.pdf

8 Team members

8.1 Participants

Evangelos Kafantaris PhD student at the University of Edinburgh working on signal processing and machine learning algorithms for decision support and outcome prediction in intensive care units. He contributed to this report by defining a framework for further multivariate analysis aimed at the identification of the physiological state of the patients.

Farhad Hatami Post-doctoral fellow at Lancaster University working on applications of Machine Learning to Medical Sciences. He contributed to this report by applying a joint mixture model to the time-series data and finding markers to predict the time of first extubation.

Hareem Naveed Senior Data Scientist with MunichRe's Integrated Analytics team where she focuses on applying predictive analytics to support underwriting analysis, and claims management.

Jialin Yi PhD student in Statistics at London School of Economics working on the statistical foundations of reinforcement learning and sequential decision making. He contributed to the work on reinforcement learning and implementation in Gym and TensorFlow.

Jo French Data scientist at Health Data Insight, working with Public Health England's National Cancer Registration and Analysis Service. She contributed the deep learning approach and the reinforcement learning approach by preprocessing the data and mathematically conceptualizing the clinical problem.

Leigh Shlomovich Statistics PhD student at Imperial College London researching time-frequency methods for cyber security.

Dr Magda Bucholc is a Lecturer in Data Analytics at the Intelligent Systems Research Centre at Ulster University. She works on the development and implementation of big data approaches to clinical decision making. She contributed to this report by performing the survival analysis given the patient-specific characteristics from the Neonatal Intensive Care Unit, Paediatric Intensive Care Unit, and Cardiac Intensive Care Unit of the Great Ormond Street Hospital.

Markus Loning PhD student at UCL and an Enrichment Scheme student at The Alan Turing Institute, working on supervised learning with time series data, and a co-developer of the Python toolbox 'sktime'. He contributed to this report by mapping out the different time series classification approaches and

implementing the specialised time series classification algorithms.

Ming Li Senior Data Scientist in industry. He contributed to the work on augmenting clinical decision making through applying reinforcement learning (RL) and building customised a RL environment using Gym.

Oliver Crook PhD student at the University of Cambridge working on Bayesian approaches to spatial proteomics and transcriptomics & facilitator for the GOSH project. He contributed to the sections on linear models, AR models and exploratory data analysis.

Pablo Leon Villagra PhD student at Edinburgh University working on understanding how people make generalizations and what kind of representations underpin these inferences. Pablo helped with data preprocessing, descriptive analysis, and auto-regressive models.

Piotr Oleskiewicz PhD student at the Institute for Computational Cosmology in Durham, working on galaxy formation, modified gravity cosmology and sensitivity analysis of models. He contributed to this report by computing time series correlations, developing an algorithm for finding gaps in CO2 delivery, extracting discrete features from the gaps, and feeding them to a support vector machine classifier.

Turing Principal Investigator: Dr.Bilal A. Mateen is a clinical-academic at Kings College Hospital (KCH), and a fellow at The Alan Turing Institute. At the Turing, Bilal is also the Clinical Data Science Liaison to the Data Study Groups Programme, helping academic health challenges (e.g. NHS Scotlands SPARRA challenge, and PLORAS), take full advantage of this unique opportunity.

8.2 Other Contributors

Ben Margetts Data scientist in the DRE team at Great Ormond Street Hospital.

Christina Pagel Director of UCL's Clinical Operational Research Unit (CORU), Professor of Operational Research, and a "researcher in residence" within the critical care units at Great Ormond Street Hospital.

John Booth Senior data steward in the DRE team at Great Ormond Street Hospital.

Samiran Ray Consultant in Paediatric Intensive Care at Great Ormond Street Hospital NHS Foundation Trust. Challenge owner, with an interest improving decision making in the intensive care unit environment using high resolution patient data.

In addition to work done by the GOSH DRE team, the challenge owners also collaborated with an Imperial College London Data Sciences team (Niall Adams, Nick Heard, & Leigh Shlomovich) who had been previous DSG challenge owners. They also helped prepare the data for DSG use, creating data dictionaries and Python functions to support data filtering and visualization from the created data structure. Moreover, Thalita Grossman & Yael Feinstein (Intensive Care Unit Fellows at Great Ormond Street Hospital), helped verify much of the data, manually going through patient records to ensure accuracy.

8.3 Editors

Dr. Bilal A. Mateen (Clinical Data Science Fellow, Turing) and Raphael Sonabend (PhD Student, UCL) edited the report for clarity, and appropriateness of scientific content.

turing.ac.uk
@turinginst