Predicting the Onset of Delirium on Hourly Basis in an Intensive Care Unit Following Cardiac Surgery

Linda Lapp
Dept. of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
0000-0003-3743-434X

Marc Roper
D*ept.of Computer and Information Sciences*
University of Strathclyde
Glasgow, UK
0000-0001-6794-4637

Kimberley Kavanagh
Dept. of Mathematics and Statistics
University of Strathclyde
Glasgow, UK
0000-0002-2679-5409

Stefan Schraag
Dept. of Perioperative Medicine
Golden Jubilee National Hospital
Clydebank, UK
0000-0003-2375-9549

***Abstract*—Delirium, affecting up to 52% of cardiac surgery patients, can have serious long-term effects on patients by damaging cognitive ability and causing subsequent functional decline. This study reports on the development and evaluation of predictive models aimed at identifying the likely onset of delirium on an hourly basis in intensive care unit following cardiac surgery. Most models achieved a mean AUC > 0.900 across all lead times. A support vector machine achieved the highest performance across all lead times of AUC = 0.941 and Sensitivity = 0.907, and BARTm, where missing values were replaced with missForest imputation, achieved the highest Specificity of 0.892. Being able to predict delirium hours in advance gives clinicians the ability to intervene and optimize treatments for patients who are at risk and avert potentially serious and life-threatening consequences.**

Keywords—delirium, dynamic risk prediction, intensive care

I. INTRODUCTION

According to Oxford Dictionary, delirium is an "acutely disturbed state of mind characterized by restlessness, illusions, and incoherence, occurring in intoxication, fever and other disorders."[1] Delirium can affect up to 50% of hospital patients who are over the age of 65 years [2], and is also relatively common in cardiac surgery patients, with the incidence between 26% to 52% [3]. Delirium can have serious long-term effects on patients, such as permanent damage to cognitive ability, and due to potential functional decline, can also lead to complications such as infections or blood clots that weaken patients and increase the risk of mortality. [4] Hence, delirium also has consequences such as an extended length of hospital stay [5], increased healthcare costs [5], and higher re-admission rates [6].

To diagnose delirium, a formal cognitive assessment is required. In hospitals, delirium is normally assessed using the Confusion Assessment Method (CAM), initially developed by Inouye et al. in 1990 [7]. CAM, the most widely used tool for identifying delirium, has been validated and used in over 4,000 published studies and translated into at least 12 languages, and has been adapted for ICU usage. CAM-ICU, requiring little training, is usable for clinicians without psychiatric education to monitor whether critically ill patients have developed delirium. [8] In this study, CAM-ICU classification is used to formally diagnose delirium (similarly to many other delirium prediction studies [9]. The onset of delirium can occur within hours [10], and rapid diagnosis is critical if the condition is to be successfully reversed, and mortality avoided [11]. CAM-ICU assessments are usually administered fairly infrequently (approx. every 10 hours in the dataset used in this study), and so having a predictive model which is able to identify the likely onset of delirium based on standard ICU laboratory measures is an essential tool for preventing this dangerous condition [2].

II. RELATED WORK

Three recent systematic reviews found 26 unique prediction models for predicting delirium in ICU [9, 12, 13]. Of these models, 4 were identified by the review papers to be "dynamic" models: DYNAMIC-ICU [14], Auto-DelRAS [15], ABD-pm [16], and a model developed by Oh et al. [17]. A model is considered dynamic if its prediction is updated as the time changes, based on the renewed input information [18]. Predicting clinical outcomes in a dynamic manner helps clinicians to be informed of patient's risk for delirium on near real-time basis.

For DYNAMIC-ICU, it is unknown how often the model is calculated and how much time in advance delirium is predicted [14]. For Auto-DelRAS, it is known that the model is calculated once a day, however, again we do not know at what time point before delirium the prediction is made [15]. ABD-pm is calculated once a day to predict the next day probability for delirium (and other outcomes), however, it is unknown at what time in the next day delirium could happen [16]. As delirium can manifest itself in a matter of hours [10], a daily prediction is too infrequent.

Out of the four models, to our knowledge, the model developed by Oh et al, is the only truly dynamic prediction model for delirium in the ICU. It reports CAM-ICU assessment in every 8 hours, and it uses heart rate variability to predict delirium every 3 hours. Even though this model shows some promising results, it was developed using 94 patient records only. [17]

Overall, none of the 26 studies report at what time specifically delirium occurs in their patient population. Pisani et al., however, state that 70.4% of their patient population had delirium within 48 hours of ICU admission [19]. This means that these models could attempt to predict delirium when it has already happened, which is a serious limitation to these models.

To improve upon delirium prediction in ICU, in this study various steps were taken:

- Experiments were undertaken to predict delirium on an hourly basis, using CAM-ICU status with its time stamps.

This is a peer-reviewed, accepted author manuscript of the following article: Lapp, L., Roper, M., Kavanagh, K., & Schraag, S. (2022). Predicting the onset of delirium on hourly basis in an intensive care unit following cardiac surgery. In 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS) IEEE. https://doi.org/10.1109/cbms55023.2022.00048

- As opposed to static variables used in models found in the review papers [9, 12, 13], dynamic laboratory variables were used to provide more up-to-date data for prediction.
- The analysis was undertaken using data recorded within the first 21 hours since ICU admission: this is to make sure that delirium is predicted as early as possible (see Section III.B).
- Experiments were undertaken using complete data, missing data, and imputation methods to understand how missing data affects the models' performance.

III. METHODS

A. Source of Data

This study has ethical approval from the UK NHS Health Research Authority (REC18/YH/0366).

This study is undertaken in Golden Jubilee National Hospital, Glasgow, which is the largest cardiac center in Scotland. The data is acquired from two databases: a preoperative database Cardiac, Cardiology and Thoracic Health Information (CaTHI) System and Centricity™ Critical Care (CCC) database. CaTHI is an audit database, including pre-operatively available variables about the patient's demographics, co-morbidities, information about the surgery, and outcome of surgery, such as total days in intensive care and hospital. The CCC database includes all information recorded in the cardiac intensive care unit, including laboratory test results and vital signs. The list of variables included in this study can be found from Table I.

All variables in the CaTHI database were measured once before surgery at the pre-operative clinic. From the CCC database, in this study, only laboratory values were used. Most variables were recorded every 2-3 hours, apart from creatinine, urea and c-reactive protein, which were measured approximately every 24 hours. Each variable in the ICU, including the CAM-ICU score, was recorded with a timestamp in the CCC database. The two databases were linked, using the Community Health Index number, and then anonymized, using a Patient ID.

B. Participants and Outcome

The outcome of this study was delirium, defined by the diagnosis of CAM-ICU (Yes/No), and was predicted on an hourly basis in an intensive care unit.

Since CAM-ICU has been recorded for ICU patients at this hospital since 2016, patients undergoing coronary artery bypass graft (CABG), valve, or combined CABG and valve surgery between 1st January 2016 and 31st December 2018 were included in the analysis: a total of 3367 patient records. Patients with CAM-ICU diagnosis within the first hour since ICU admission were excluded to make sure that only future CAM-ICU diagnoses were predicted. This resulted in the final overall population of 3322 (unique) patient records.

The prevalence of delirium in this patient population was 12.47% (95% CI 11.39%-13.63%) out of 3322 patients. The mean time of delirium occurrence was at 11.21 ± 2.84 hours since ICU admission (Fig. 1), the latest initial onset of delirium occurring at 20.60 hours. CAM-ICU score was measured in mean time of every 10.52 ± 4.02 hours (median = 11.90, IQR = 0.95 hours). This may explain why so many patients received a delirium diagnosis between 10 to 13 hours. However, for 7.30% patients CAM-ICU assessment was undertaken within the first hour since ICU admission. The patient demographics can be found in Table II.

TABLE I. VARIABLES EXTRACTED FROM THE DATABASES TO DEVELOP THE PREDICTION MODELS

Variable type	Variables
Static pre-operative variables from CaTHI database	
Patient characteristics	Age, sex, type II diabetes, BMI, smoking status
Cardiac pre-operative variables	Congestive cardiac failure, previous myocardial infarction, active endocarditis, hypertension history, NYHA grade, angina status, rhythm, left ventricular function, left main stem, extracardiac arteriopathy, logistic EuroSCORE
Non-cardiac pre-operative variables	Neurological dysfunction, pulmonary disease, pre-operative creatinine, renal impairment
Surgical variables	Surgical priority, critical pre-operative state, procedure, previous cardiac surgery, previous percutaneous coronary intervention
Dynamic ICU variables from Centricity Critical Care database	
laboratory variables (Minimum, maximum, first, last)	Arterial base excess, arterial haematocrit, bicarbonate, c-reactive protein, creatinine, daily fluid balance, haemoglobin, hydrogen ion, lactate, potassium, sodium, urea, urine
Binary variables (yes/no)	CAM-ICU; Medicines: Dobutamine, dopamine, noradrenaline, vasopressin

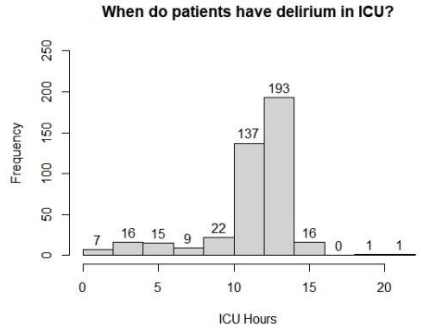

Fig. 1. Time of delirium onset in ICU based on CAM-ICU diagnosis.

TABLE II. PATIENT DEMOGRAPHICS

Demographic	Mean ± SD or %
Age	65.82 ± 11.19
Female	28.56%
CABG	51.97%
Valve	33.13%
CABG & Valve	14.89%
ICU hours	51.77 ± 109.90

C. Missing Data

If a numerical variable had many missing values in the CaTHI database, the variable was excluded from analysis (e.g., pre-operative hemoglobin). If a categorical variable had large number of missing values, a new category "Unknown" was

created. Hence, 25 preoperative variables were included in the analysis (Table I).

Since ICU is an incredibly data-rich environment, the problem of missing data is inevitable. Firstly, all data were checked for obvious incorrect values which are physiologically impossible. These values were marked as NA in the dataset. Secondly, if a patient had a missing value, but a timestamp recorded with it, the recorded value at previous timestamp was carried forward to replace the missing value in the next time stamp.

Since medicine doses were recorded for a very small number of patients, the medicine variables were changed into binary categorical variables (yes/no) based on whether they were given medication.

Overall, from the CCC database, 17 variables were used in the analysis (Table I).

In addition to the above, modelling experiments were undertaken to approach the problem of missing data (explained in Section D2.

D. Predictive Modelling Methods

1) Data Preparation

For the hourly prediction of delirium, time windows were created based on the time stamps recorded for each variable entry. To assign time windows for the recorded laboratory values, time differences (in hours) were calculated of when the laboratory values were measured for each patient. The cumulative time differences were created and rounded up to create time windows. The time window = 0 indicates the first measurement since admission to the ICU.

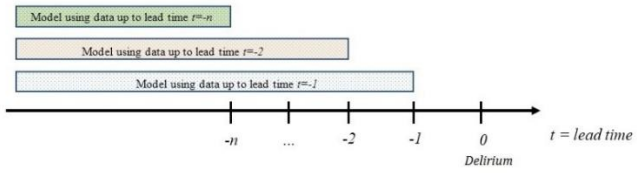

Fig. 2. How models are developed with lead times

Since the prediction of delirium was undertaken on an hourly basis, lead times were created to show how many hours in advance the outcome was predicted (Fig. 2). For example, if delirium happened at time window = 12, and the prediction was made 1 hour in advance, all data recorded for patients from time window = 0 to time window = 11 were included in the model.

Because most patients in this study had delirium in ICU within 13 hours (Fig. 1), meaning the percentage of patients developing delirium afterwards was less than 1%, the prediction was made on hourly basis, up to 13 hours in advance. This resulted in 13 models for each method, i.e., a model was developed for each lead time.

Due to working with a very large amount of data, to simplify the analysis, for each lead time, minimum, maximum, first and last measurement of a variable were calculated, as done elsewhere [20]. This means that if delirium happened in time window = 12, if predicting 1 hour in advance, for each variable, first, last, minimum, and maximum measurements that occurred in time windows 0 – 11 were calculated.

Since delirium predicted was a binary outcome (Yes/No), and only patients with delirium had a timestamp indicating when delirium was recorded for them, an arbitrary endpoint was chosen for patients who did not have delirium. Since the latest time of delirium onset in this patient population was 20.60 hours since ICU admission (Section III.B), in this study, delirium was predicted within 21 hours since ICU admission, and hence the endpoint for non-delirium patients was also 21 hours.

2) Classification Experiments and Methods

The dataset was randomly divided into a training set (2/3 of the data) and testing set (1/3 of the data). All models were developed, using the training set at 10-fold cross-validation, and evaluated, using the testing set.

Three experiments were undertaken to predict delirium, employing different strategies to handle the missing data:

1. Using complete data only, removing all records with missing data from analysis. Classification methods, appropriate for this type of data were used: logistic regression (LR), random forest (RF), AdaBoost (AB), Gradient Boosting Model (GBM), Support Vector Machine (SVM), C5.0 and Bayesian Additive Regression Trees Machine (BARTm).
2. Using complete training data and testing data with missing values. Records with >40% of missing values were excluded from analysis, as done elsewhere [21]. Here, two classification methods which are robust at handling missing data were used: C5.0 and BARTm.
3. Using complete training data and the missForest imputation method [22] on the testing data to replace missing values. Again, records with >40% of missing values were excluded. The same classification methods were used as in experiment 1.

The ‘caret’ R package version 6.0.90 was used to develop the LR (‘glm’ method) and RF (‘rf’, n=200 trees) models.

The AB model was developed using the package ‘fastAdaboost’ version 1.0.0, which implements Freund and Schapire’s Adaboost.M1 algorithm (n=40 iterations) [23]. For the GBM, the package ‘gbm’ version 2.1.5 was used, which uses the Friedman’s gradient boosting algorithm (number of trees n=1000, shrinkage = 0.01) [24]. For SVM [25], package ‘e17071’ version 1.7.9. was used.

These models are widely known approaches to predict clinical outcomes. BARTm and C5.0 applied to missing data, however, are less commonly used. In this study, BARTm was experimented with by using the ‘bartMachine’ R package [26] version 1.2.6. For the C5.0 model development, R package ‘C50’ [27] was used, together with the default of including missing values as the model can accommodate these.

For missing data, the missForest imputation method was experimented with, which has been shown to outperform other sophisticated imputation methods like k-nearest neighbors imputation or multivariate imputation using chained equations [22]. For this, the ‘missForest’ R package [28] version 1.4. was used.

All analysis was undertaken, using R version 4.1.1 and code is available in [29].

3) Model Evaluation and Performance Measures

The models were evaluated on testing data, using receiver operating characteristic curve (ROC), sensitivity, specificity, positive (PPV) and negative predictive values (NPV).

The focus, in terms of performance measures in this paper is on sensitivity and specificity. This is to ensure that the models can recognize patients with and without delirium well, and in practical terms, the introduction of an intervention does not have serious consequences (in health or cost terms) whereas a failure to intervene does.

IV. RESULTS

A. Model Development and Performance

The mean number of patients across all lead times was similar in all three experiments (Experiment 1: training = 1519 ± 27.28, testing = 747 ± 13.47; Experiments 2 and 3: training = 1526 ± 25, testing = 762 ± 12.63). When using missForest (experiment 3), only 3% of missing data had to be replaced. As expected, based on Fig. 1, the proportion of patients with delirium increased as the lead times got closer to 0 (Fig. 3), hbut was similar in all three experiments.

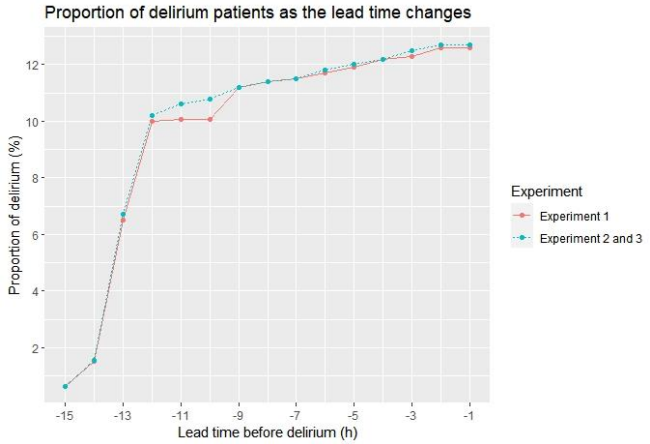

Fig. 3. Proportion of patients with delirium as the lead time changes.

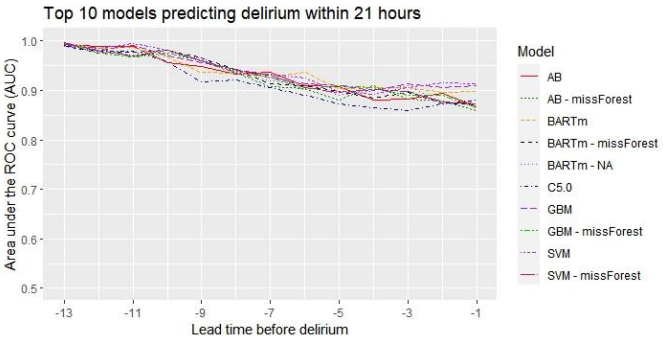

Fig. 4. Area under the receiver operating characteristic curve for each lead time for top 10 models based on mean AUC.

As seen from Fig. 4, all models had a notably high performance, mostly staying above AUC = 0.850. BARTm with complete data and GBM with missForest had equally the highest AUC of 0.997 when predicting delirium 13 hours in advance. At 8 hours in advance, BARTm with missForest had the highest AUC of 0.943 and AB with missForest imputation had the highest AUC at 4 hours before the onset of delirium (AUC = 0.910).

In terms of sensitivity (Fig. 5), the models had either high or moderately high performance, most staying above 0.700 at all lead times. C5.0, however, had a considerably large drop in performance when predicting delirium 4 hours or less in advance. At 13 hours in advance, GBM, SVM and RF had equally high sensitivity of 0.999, all three developed with missForest. When predicting delirium 8 hours in advance, RF with complete data had the highest sensitivity of 0.936. BARTm with missing values in test set had the highest sensitivity when predicting delirium 4 hours in advance (Sens = 0.895).

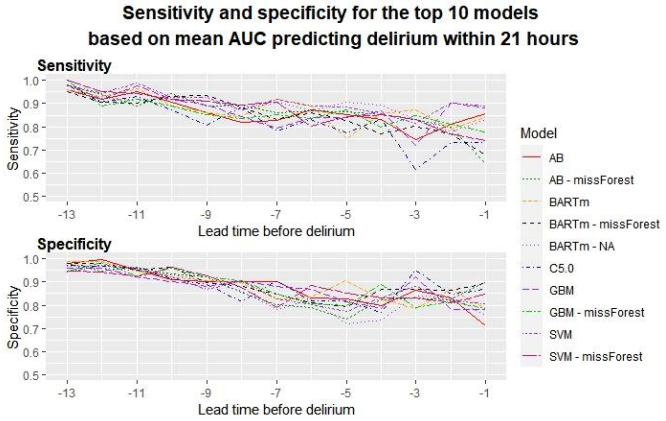

Fig. 5. Sensitivity and specificity for each lead time for the top 10 models based on mean AUC predicting delirium within 21 hours of ICU stay, 13 hours in advance.

TABLE III. MEAN PERFORMANCE MEASURES FOR EACH MODEL ACROSS ALL LEAD TIMES

Model	AUC (mean ± SD)	Sens (mean ± SD)	Spec (mean ± SD)
SVM	0.941 ± 0.038	0.907 ± 0.048	0.870 ± 0.066
GBM	0.939 ± 0.035	0.875 ± 0.070	0.875 ± 0.064
BARTm	0.937 ± 0.036	0.875 ± 0.062	0.884 ± 0.069
GBM - missForest	0.931 ± 0.041	0.861 ± 0.057	0.875 ± 0.071
SVM - missForest	0.931 ± 0.043	0.874 ± 0.075	0.880 ± 0.057
BARTm - NA	0.930 ± 0.041	0.885 ± 0.047	0.851 ± 0.087
AB	0.929 ± 0.043	0.863 ± 0.060	0.877 ± 0.078
BARTm - missForest	0.929 ± 0.044	0.850 ± 0.080	0.892 ± 0.061
AB - missForest	0.926 ± 0.046	0.865 ± 0.083	0.870 ± 0.076
C5.0	0.915 ± 0.047	0.822 ± 0.097	0.885 ± 0.067
RF - missForest	0.910 ± 0.049	0.852 ± 0.076	0.844 ± 0.076
RF	0.907 ± 0.050	0.850 ± 0.084	0.832 ± 0.083
C5.0 - missForest	0.903 ± 0.054	0.803 ± 0.099	0.890 ± 0.059
LR	0.901 ± 0.033	0.848 ± 0.054	0.884 ± 0.050
C5.0 - NA	0.898 ± 0.056	0.822 ± 0.089	0.854 ± 0.098
LR - missForest	0.879 ± 0.044	0.814 ± 0.066	0.878 ± 0.056

Looking at the specificity of the models (Fig. 5), again all models stayed above 0.700 at all lead times. BARTm had the

highest performance when predicting delirium 13 hours in advance (Spec = 0.989). RF with missForest had the highest specificity of 0.908 at 8 hours before delirium, and GBM with missForest had the highest specificity of 0.889 at 4 hours before delirium.

According to Table III, SVM with complete data had the highest mean AUC of 0.941 and mean sensitivity of 0.907 across all lead times. This is a very high sensitivity, meaning that the model can recognize patients with delirium 90.7% of the time. Based on mean specificity, BARTm with missForest imputation had the highest performance (Spec = 0.892), which means that the model can recognize patients without delirium 89.2% of the time.

Finally, the positive predictive values for models were very low (highest mean = 0.030 ± 0.098 for C5.0 with NA) and so were negative predictive values. The highest mean NPV was 0.563 ± 0.133 across all lead times for SVM with complete data.

As shown in Fig. 6, the most used variables for the models to predict delirium were lactate, urine output, potassium, and hydrogen ion. Every model used lactate and potassium, and most models also used urine output, hydrogen ion, arterial hematocrit, arterial base excess, hemoglobin, bicarbonate, and sodium.

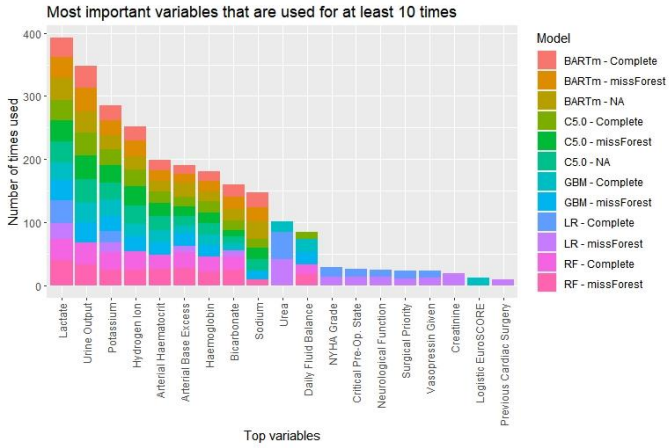

Fig. 6. The number of times the top-used variables were used in each model.

V. Discussion

Overall, SVM, developed with complete data, achieved the highest mean AUC of 0.941 (mean Sens = 0.907 and mean Spec = 0.870).

These results should be compared to other published "dynamic" models predicting delirium with caution. In terms of the "real-time" aspect, the model by Oh et al. works most similarly to the models developed in this study. Oh et al.'s best model reached a sensitivity of 0.880 and specificity of 0.278 [17]. Our SVM model had a slightly higher mean sensitivity, recognizing patients with delirium 90.7% of the time (vs Oh et al.'s 88.0%), and far higher specificity than the Oh et al.'s model. The model presented in this study was capable of recognizing patients without delirium 87.0% of the time (vs Oh et al.'s 27.8%).

Models developed with complete data performed slightly better than models with missing values or missForest imputation. However, the BARTm model with missing values in test set was the 6th best performing model based on AUC (0.930 vs SVM's 0.941) and 2nd best model based on sensitivity (0.885 vs SVM's 0.907). This is a very promising result as missing data is a big problem in healthcare databases [30]. Ideally, more effort should be directed towards developing higher quality databases and robust and reliable variable recording mechanisms, however, in the meantime, being able to use methods that handle missing data, is a great solution. This means that if a patient has some missing data, the clinician can still be informed whether a patient is likely to develop delirium due to the high-performing prediction model that is robust to missing values.

Interestingly, the models' performance decreased as the lead time got closer to the onset of delirium. This is most likely due to the proportion of patients with delirium increasing (Fig. 3), which increases the variance of the observations, and reducing the certainty of the models' predictive ability at later lead times.

A. Explainability

Because the models were all performing quite well, based on the results shown in Table III, it is difficult to say which model should be ultimately chosen to further develop a clinical decision support tool. On one hand, the model with the highest AUC, sensitivity and specificity should be chosen, but on the other hand, the explainability and usability of the models should be considered.

In medicine, it is especially important to understand why a model predicts a patient to have high probability for the outcome. If the clinician knows which factors have contributed to the probability to be high, then the clinicians can investigate which factors to pay closer attention to and improve with medical interventions.

For example, even though SVM had an overall best performance, it is known to be a "black-box" classifier, meaning the decisions the model makes are not easily explainable [31]. The same issue was found with AB due to its complexity. This means that the model does not always give useful intuitive reasons why a particular point is classified in one class rather than another.

LR is a highly explainable classification method due to its built-in estimates that can be converted into odds ratios. The mean performance for LR, developed with complete data, was AUC of 0.901 (Sens = 0.848, Spec = 0.884). Even though these values are at the lower end of performance when compared to the other models, LR performance is still comparable with other models' performance and is still remarkably high. LR's competitiveness with other machine learning algorithms has also been shown in other studies [20, 32].

RF, C5.0 and BARTm, being tree-based models, and GBM having tree-based base learners, are easily explainable through looking at variable importance measures and at how the trees were formed. Based on the findings of this study, the most important variables that were associated with delirium were lactate, urine output, potassium, and hydrogen ion.

B. Limitations

Because delirium assessment with CAM-ICU is undertaken by clinical staff, the regularity of when the assessment is done can vary from patient to patient, which is a limitation to this study. On average, CAM-ICU assessments were undertaken in every 10.52 hours, only 7.30% of all patients receiving the assessment within the first hour since ICU admission. To improve delirium prediction, CAM-ICU assessment should be undertaken within the first hour of ICU admission for all patients and followed up with more regular assessments. As delirium can develop rapidly [33], the actual occurrence of it can be earlier than when delirium is assessed. This can be the reason why the performance of the prediction models in this study was decreasing as the lead times got closer to the recorded positive CAM-ICU result.

Another limitation, when analyzing CAM-ICU diagnoses is the possibility for observer bias. The score is measured by different clinical staff, which means that there can be variation in training and the results can be subjective [33]. Oh et al. minimize this problem by having the CAM-ICU scoring done both by psychiatrists and nurses [17]. However, as the study in hand is a retrospective study, it is unknown who exactly undertook the assessment and what their level of training was.

Even though the CAM-ICU could be recorded more often, being able to clearly state when exactly CAM-ICU was diagnosed for each patient is a strength to this study by offering transparency on how the models were developed. None of the studies discussed in "Related Work" reported at what time delirium occurred in their patient population.

C. Conclusion

Most of the models currently predicting delirium in the ICU are static models. This study demonstrated that it is possible to predict delirium on an hourly basis 13 hours in advance, with the average AUC of 0.941 (SVM) using complete data, and mean AUC of 0.930 (BARTm) with missing data. The models developed in this study could help clinicians optimize treatments for patients who are at risk of developing delirium hours in advance.

In addition to internal validation to ensure reproducibility, external validation with a larger patient population to support generalizability should be undertaken. Furthermore, for the models developed in this study to be successful in practice, a system needs to be developed that is integrated with the electronic health records.

ACKNOWLEDGMENT

This research was funded by the University of Strathclyde and Golden Jubilee National Hospital.

REFERENCES

[1] Delirium. [https://www.lexico.com/definition/delirium].
[2] Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014; 383(9920):911–922.
[3] Brown CH. Delirium in the Cardiac Surgical Intensive Care Unit. Curr. Opin. Anaesthesiol. 2014; 27(2):117–122.
[4] Collier R. Hospital-induced delirium hits hard. CMAJ 2012; 184(1):23–24.
[5] Schubert M, Schürch R, Boettger S et al. A hospital-wide evaluation of delirium prevalence and outcomes in acute care patients - a cohort study. BMC Health Serv. Res. 2018.
[6] Gleason LJ, Schmitt EM, Kosar CM et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. JAMA Surg. 2015; 150(12):1134–1140.
[7] Inouye SK, van Dyck CH, Alessi CA et al. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann. Intern. Med. 1990; 113(12):941–948.
[8] Ely EW, Margolin R, Francis J et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit. Care Med. 2001; 29(7):1370–1379.
[9] Ruppert MM, Lipori J, Patel S et al. ICU Delirium-Prediction Models: A Systematic Review. Crit. Care Explor. 2020.
[10] Cole MG, Mccusker J. Delirium in older adults: a chronic cognitive disorder? Int. Psychogeriatrics 2016; 28(8):1229–1233.
[11] Regal PJ. Delirium, in 405 articles of medical (non-surgical or ICU) inpatients: unproven speed of onset and recovery. Clin. Interv. Aging 2017; 12:377–380.
[12] Chen J, Yu J, Zhang A. Delirium risk prediction models for intensive care unit patients: A systematic review. Intensive Crit. Care Nurs. 2020.
[13] Chen X, Lao Y, Zhang Y et al. Risk predictive models for delirium in the intensive care unit: a systematic review and meta-analysis. Ann. Palliat. Med. 2021; 10(2):1467–1479.
[14] Fan H, Ji M, Huagn J et al. Development and validation of a dynamic delirium prediction rule in patients admitted to the Intensive Care Units (DYNAMIC-ICU): A prospective cohort study. Int. J. Nurs. Stud. 2019; 93:64–73.
[15] Moon KJ, Jin Y, Jin T, Lee S-M. Development and validation of an automated delirium risk assessment system (Auto-DelRAS) implemented in the electronic health record system. Int. J. Nurs. Stud. 2018; 77:46–53.
[16] Marra A, Pandharipande PP, Shotwell MS et al. Acute Brain Dysfunction. Development and Validation of a Daily Prediction Model. Chest 2018; 154(2):293–301.
[17] Oh J, Cho D, Park J et al. Prediction and early detection of delirium in the intensive care unit by using heart rate variability and machine learning. Physiol. Meas. 2018.
[18] Ellner SP, Guckenheimer J. What are dynamic models? Dyn. Model. Biol., Princeton University Press, 2011:1–30.
[19] Pisani MA, Murphy TE, van Ness PH et al. Characteristics associated with delirium in older patients in a medical intensive care unit. Arch. Intern. Med. Intern. Med. 2007; 167(15):1629–1634.
[20] Johnson AE., Mark R. Real-time mortality prediction in the Intensive Care Unit. AMIA Annu. Symp. Proc., 2017:994–1003.
[21] Ho JC, Lee CH, Ghosh J. Septic Shock Prediction for Patients with Missing Data. ACM Trans. Manag. Inf. Syst. 2014; 5(1):1–15.
[22] Stekhoven DJ, Bühlmann P. MissForest - non-parametric missing value imputation for mixed-type data. Bioinformatics 2012; 28(1):112–118.
[23] Freund Y, Schapire RE. Experiments with a new boosting algorithm. Thirteen. Int. Conf. Mach. Learn., 1996:148–156.
[24] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 2001; 29(5):1189–1232.
[25] Meyer D. Support Vector Machines. 2021.
[26] Kapelner A, Bleich J. Package "bartMachine." 2020.
[27] Kuhn M, Weston S, Culp M et al. Package "C50." 2021.
[28] Stekhoven DJ. Package "missForest." 2016.
[29] Lapp L. Code for Hourly Delirium Preparation and Prediction. 2022. doi:10.15129/1ab360f7-0779-4cf3-8a9a-dae621892a51.
[30] Mazzali C, Duca P. Use of administrative data in healthcare research. Intern. Emerg. Med. 2015; 10:517–524.
[31] Barbella D, Benzaid S, Christensen JM et al. Understanding Support Vector Machine Classifications via a Recommender System-Like Approach. Proc. 2009 Int. Conf. Data Min. (DMIN 2009), 2009.
[32] Christodoulou E, Ma J, Collins GS et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J. Clin. Epidemiol. 2019; 110:12–22.
[33] Mistraletti G, Pelosi P, Mantovani ES et al. Delirium: Clinical approach and prevention. Best Pract. Res. Clin. Anaesthesiol. 2012; 26:311–326.