Research article

Predictors of in-ICU length of stay among congenital heart defect patients using artificial intelligence model: A pilot study

João Chang Junior [a,b,c,*], Luiz Fernando Caneo [a], Aida Luiza Ribeiro Turquetto [a,d], Luciana Patrick Amato [a,d], Elisandra Cristina Trevisan Calvo Arita [a], Alfredo Manoel da Silva Fernandes [a], Evelinda Marramon Trindade [d,e,f], Fábio Biscegli Jatene [a], Paul-Eric Dossou [g], Marcelo Biscegli Jatene [a]

[a] *Hospital Das Clínicas HCFMUSP, Universidade de São Paulo, Instituto Do Coração - InCor, Av. Dr. Enéas Carvalho de Aguiar, 44, CEP 05403-000, São Paulo, Brazil*
[b] *Escola Superior de Engenharia e Gestão - ESEG, Rua Apeninos, 960, São Paulo, Brazil*
[c] *Centro Universitário Armando Alvares Penteado - FAAP, Rua Alagoas, 903, São Paulo, Brazil*
[d] *Núcleo de Avaliação de Tecnologias da Saúde - NATS-HCFMUSP, Brazil*
[e] *Laboratório de Ensino, Pesquisa e Inovação Em Saúde - LEPIC-HCFMUSP, Superintendência / Hospital Das Clínicas da FMUSP, Rua Dr. Ovidio Pires de Campos, 225, 5°. Andar – Superintendência, Sao Paulo, Brazil*
[f] *Sao Paulo State Health Secretariat–SES-SP, Sao Paulo, Brazil*
[g] *Institut Catholique des Arts et Metiers–Icam, Paris-Senart, France*

ARTICLE INFO

ABSTRACT

Objective: This study aims to develop a predictive model using artificial intelligence to estimate the ICU length of stay (LOS) for Congenital Heart Defects (CHD) patients after surgery, improving care planning and resource management.
Design: We analyze clinical data from 2240 CHD surgery patients to create and validate the predictive model. Twenty AI models are developed and evaluated for accuracy and reliability.
Setting: The study is conducted in a Brazilian hospital's Cardiovascular Surgery Department, focusing on transplants and cardiopulmonary surgeries.
Participants: Retrospective analysis is conducted on data from 2240 consecutive CHD patients undergoing surgery.
Interventions: Ninety-three pre and intraoperative variables are used as ICU LOS predictors.
Measurements and main results: Utilizing regression and clustering methodologies for ICU LOS (ICU Length of Stay) estimation, the Light Gradient Boosting Machine, using regression, achieved a Mean Squared Error (MSE) of 15.4, 11.8, and 15.2 days for training, testing, and unseen data. Key predictors included metrics such as "Mechanical Ventilation Duration", "Weight on Surgery Date", and "Vasoactive-Inotropic Score". Meanwhile, the clustering model, Cat Boost Classifier, attained an accuracy of 0.6917 and AUC of 0.8559 with similar key predictors.

* Corresponding author. Hospital Das Clínicas HCFMUSP, Universidade de São Paulo, Instituto Do Coração - InCor, Av. Dr. Enéas Carvalho de Aguiar, 44, CEP 05403-000, São Paulo, Brazil.
E-mail addresses: chang.joao@gmail.com (J. Chang Junior), caneo@me.com (L.F. Caneo), aidaturquetto@mac.com (A.L.R. Turquetto), lu.amato@gmail.com (L.P. Amato), elisandra.arita@incor.usp.br (E.C.T.C. Arita), alfredo.fernandes@incor.usp.br (A.M.S. Fernandes), evelinda.trindade@hc.fm.usp.br (E.M. Trindade), fabio.jatene@incor.usp.br (F.B. Jatene), paul-eric.dossou@icam.fr (P.-E. Dossou), marcelo.jatene@incor.usp.br (M.B. Jatene).

Conclusions: Patients with higher ventilation times, vasoactive-inotropic scores, anoxia time, cardiopulmonary bypass time, and lower weight, height, BMI, age, hematocrit, and presurgical oxygen saturation have longer ICU stays, aligning with existing literature.

1. Introduction

Congenital heart defects (CHD), with a global incidence of about 9 per 1000 live births, significantly impact morbidity, mortality, and healthcare costs. Predicting cardiac surgery outcomes and ICU stays for CHD patients is challenging due to the diversity of anomalies and surgical procedures. Registries like the Society of Thoracic Surgeons (STS), World Society for Pediatric and Congenital Heart Surgery (WSPCHS), and Brazil's ASSIST are crucial for understanding CHD and tailoring risk assessments and outcome evaluations to diverse populations. Machine learning methods have potential in predicting CHD patient outcomes, with studies employing neural networks and decision trees to develop risk classification systems for pediatric cardiac surgeries. This study aims to evaluate predictive models for ICU stays in CHD patients using machine learning, improving surgical scheduling and resource allocation in similar hospitals.

2. Methods

2.1. Study design

This retrospective, post-hoc analysis is based on artificial intelligence models constructed from data sourced from the Prospective Registry Multicenter CHD study conducted between 2014 and 2018. The analysis seeks to determine the most accurate machine learning (ML) model to forecast the duration of stay in the intensive care unit (ICU) for patients who have undergone surgery for congenital heart defects.

PyCaret, an open-source, low-code Python library designed for automating machine learning workflows, has been utilized in this research [1]. Notably, this library features a function named "compare_models", which serves to train and evaluate the performance of all available models within the library using the following statistical parameters: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (a measure of the proportion of the endogenous variable's variance explained by the exogenous variable(s)), Root Mean Squared Logarithmic Error (RMSLE), Mean Absolute Percentage Error (MAPE), for regression models. Additionally, it utilizes a confusion matrix for classification models.

In this study, the following models, available in the PyCaret library, were employed to predict the duration of ICU stay.

1. Regression Models: Extra Trees Regressor (et), Light Gradient Boosting Machine (lightgbm), K Neighbors Regressor (knn), Random Forest Regressor (rf), CatBoost Regressor (catboost), Gradient Boosting Regressor (gbr), Extreme Gradient Boosting (xgboost), Huber Regressor (huber), Dummy Regressor (dummy), Lasso Least Angle Regression (llar), Ridge Regression (ridge), Least Angle Regression (lar), Linear Regression (lr), Elastic Net (en), Lasso Regression (lasso), Bayesian Ridge (br), AdaBoost Regressor (ada), Decision Tree Regressor (dt), Orthogonal Matching Pursuit (omp), and Passive Aggressive Regressor (par).
2. Classification Models: Cat Boost Classifier (catboost), Random Forest Classifier (rf), Gradient Boosting Classifier (gbc), Extreme Gradient Boosting (xgboost), Light Gradient Boosting Machine (lightgbm), Ada Boost Classifier (ada), Extra Trees Classifier (et), Logistic Regression (lr), K Neighboors Classifier (knn), Decision Tree Classifier (dt), Naïve Bayes (nb), Linear Discrinant Analysis (lda), Ridge Classifier (ridge), SVM – Linear Kernel (svm), Quadratic Discriminant Analysis (qda), and Dummy Classifier (dummy).

2.2. Study population

Between January 2014 and December 2018, a total of 2240 patients with Congenital Heart Defects (CHD) were consecutively referred for surgical treatment to a specialized Brazilian hospital known for its expertise in transplants and cardiopulmonary surgeries. All the pertinent data were extracted from the ASSIST Registry dataset, adhering strictly to institutional governance rules pertaining to security and privacy.

The ASSIST database currently houses records of over 3000 patients [2]. However, data from other participating centers included in this collective dataset have not yet been subjected to an active audit. To ensure reliability, therefore, this analysis solely utilizes data from a single hospital.

Rigorous quality checks were performed periodically to uphold the accuracy of the data. In order to maintain integrity and consistency in the data pool, patients from the ASSIST database who had undergone transplantation, those who had passed away during their hospital stay, as well as those whose records lacked critical information (missing values) were excluded from this analysis. Consequently, the studied sample from the ASSIST database was narrowed down to include 1642 patients.

2.3. Predicting variables

Ninety-three preoperative and intraoperative variables from the ASSIST database were examined as potential predictors for patients' length of stay in the ICU. These variables were selected based on a comprehensive review of existing literature, practical

Table 1
Main variables pertaining to preoperative and intraoperative phases.

Variable	Description	n (%)	Mean	SD	Median	IQR
Provider	1 = SUS	73				
	2 = Particular	0.4				
	3 = Health insurance	7.3				
	4 = Ignored	19.3				
Sex	1 = Male	48.2				
	2 = Female	51.8				
Number of Previous Hospitalizations	From 0 to 8	100	0.6	1	0	1
Number of Previous Surgeries	From 0 to 6	100	0.33	0.74	0	0
Non-Cardiac Abnormality—Chromosomal Syndrome	0 = No	83.5				
	1 = Yes	16.5				
Down's Syndrome	0 = No	87				
	1 = Yes	13				
Preoperative Hematocrit	From 21% to 79%	100	38.9	6.45	38	7
Preoperative Arterial Oxygen Saturations	From 61% to 100%	100	93.85	6.38	96	3
Diagnosis Category	1 = cardiomyopathy	0.3				
	2 = Conduit failure	0				
	3 = cor triatriatum	0.12				
	4 = double outlet left ventricle	0				
	5 = double outlet right ventricle	2.07				
	6 = electrophysiological	0				
	7 = left heart lesions	7.92				
	8 = miscellaneous, other	0.3				
	9 = pericardial disease	0				
	10 = pulmonary venous anomalies	3				
	11 = pulmonary venous stenosis	0				
	12 = right heart lesions	20.9				
	13 = septal defects	50.19				
	14 = shunt failure	0				
	15 = single ventricle	5.4				
	16 = systemic venous anomalies	0				
	17 = thoracic and mediastinal disease	0				
	18 = thoracic arteries and veins	5.6				
	19 = transportation of the great arteries	4.2				
Atrial Septal Defect	0 = No	81				
	1 = Yes	19				
Tetralogy of Fallot	0 = No	85.5				
	1 = Yes	14.5				
Preoperative Resuscitation	0 = No	98.7				
	1 = Yes	1.3				
Preoperative Mechanical Ventilation	0 = No	96.5				
	1 = Yes	3.5				
Preoperative Pulmonary Hypertension	0 = No	95.9				
	1 = Yes	4.1				
Previous ICU Admission	0 = No	91.4				
	1 = Yes	18.6				
Mitral Valve Disease	0 = No	97.8				
	1 = Yes	2.2				
Patient Age on the Surgery Date	From 0 to 24,316 days	100	3303.42	4836.80	733.66	4379.17
Patient Weight on the Surgery Date	From 2 Kg to 116 Kg	100	23.78	24.83	10.2	34.3
Patient Height on the Surgery Date	From 40 cm to 189 cm	100	99.6	43.53	82	85
Body Mass Index on the Surgery Date	From 7.9 to 39.1	100	17.31	4.85	15.92	5.6
Body Surface on the Surgery Date	From 0.16 m2 to 3.93 m2	100	0.78	0.59	0.48	1
ICU Length of Stay	From 1 day to 290 days	100	11	18.56	2.81	9.15
ICU Length of Stay in 3 Clusters	From 1 to 3 days	33.4				
	From 4 to 7 days	30.6				
	More than 7 days	36				
Mechanical Ventilation Time	From 0 to 8767 h	100	79.66	262.89	15.21	85.79
Cardiopulmonary Bypass Time	From 0 to 417 min	100	108.8	49.7	101	66
Anoxia Time	From 2 min to 280 min	100	73.77	40.61	68	60
Vasoactive-Inotropic Score in Surgery	From 0 to 292	100	13.86	19.95	8	11.5
Surgical Procedure (1–92)	12 - Arterial Switch Operation	1.3				
	13 - Arterial Switch Operation and Ventricular Septal Defect	1.5				
	15 - Atrial Septal Defect	0.8				
	Others	96.4				

evidence, applicability, and consensus among the participating researchers. The predictive variables used were: "Provider", "Sex", "Diagnosis of Congenital Heart Disease in Prenatal Care", "Prematurity", "Son of a Diabetic Mother", "Previous Surgery", "Number of Previous Hospitalizations", "Number of Previous Surgeries", "Non-Cardiac Abnormality-Chromosomal Syndrome", "Down's Syndrome", "DiGeorge Syndrome", "Turner Syndrome", "Williams Syndrome", "Edwards Syndrome", "Noonan Syndrome", "Other Syndromes", "Non-cardiac Abnormality-Malformations", "Esophageal Atresia", "Imperforated Anus", "Tracheoesophageal Fistula", "Diaphragmatic Hernia", "Omphalocele", "Cleft Palate", "Other Anomalies", "Preoperative Hematocrit", "Arterial Oxygen Saturations", "Diagnosis Category", "Truncus Arteriosus", "Cardiomyopathy", "Partial Anomalous Drainage of the Pulmonary Veins", "Total Anomalous Drainage of the Pulmonary Veins", "Aortic Aneurysm", "Aortic Valve Disease", "Mitral Valve Disease", "Tricuspid Valve Disease", "Ebstein Anomaly", "Aortopulmonary Window", "Persistence of the Arterial Canal", "Atrioventricular Septal Defect", "Atrioventricular Canal Defect", "Ventricular Septal Defect", "Aortic Arch Coarctation and Hypoplasia", "Miscellaneous", "Corrected Transposition of the Great Arteries", "Cor Triatriatum", "Coronary Artery Anomalies", "Double Outlet Right Ventricle", "Hypoplastic Left Heart Syndrome", "Tetralogy of Fallot", "Right Ventricular Outflow Tract Obstruction", "Aortic Arch Interruption", "Pulmonary Atresia", "Pulmonary Valve Disease", "Transposition of the Great Arteries", "Single Ventricle", "Heart Rhythm Changes", "Aorto-left Ventricular Tunnel", "Shone Syndrome", "Vascular Ring/Sling", "Preoperative Resuscitation", "Preoperative Arrhythmia", "Inotropic Use in the Preoperative Period", "Preoperative Mechanical Ventilation", "Preoperative Hypothyroidism", "Preoperative Diabetes", "Preoperative Diagnosis of Endocarditis", "Preoperative Sepsis", "Preoperative Seizure", "Preoperative Neurological Changes", "Preoperative Renal Dysfunction", "Preoperative Pulmonary Hypertension", "Preoperative Tracheostomy", "Preoperative Gastrostomy", "Preopeative ECMO Need", "Previous ICU Admission", "Patient Age on the Surgery Date", "Patient Weight on the Surgery Date", "Patient Height on the Surgery Date", "Body Mass Index on the Surgery Date", "Body Surface on the Surgery Date", "Reoperated", "ICU Length of Stay", "Mechanical Ventilation Time", "Cardiopulmonary Bypass", "Cardiopulmonary Bypass Time", "Anoxia Time", "Type of Surgery", "Vasoactive-Inotropic Score in Surgery", "Intraoperative Echocardiography", "Intraoperative Bleeding", "Intraoperative Arrhythmia", "Cardiopulmonary Bypass Return", "Extracorporeal Membrane Oxygenation in Surgery", "Surgical Procedure". A detailed description of the most significant variables explaining ICU length of stay and their parameterizations can be found in Table 1. These selected variables served as exogenous inputs for the algorithms used by the PyCaret library.

Fig. 1 provides a comparative illustration of the performance of the 20 regression models, as measured by the statistical parameters of adherence: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE). Fig. 2 provides a comparative illustration of the performance of the 16 classification models as per the indicators: "Accuracy", "AUC", "Recall", "Precision", "F1", "Kappa", and "MCC".

Chicco et al. [3] argue that in regression models, the R^2 adherence statistic, also known as the coefficient of determination, is more informative and reliable due to its lack of interpretability limitations inherent in the Mean Standard Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). In our research, the Light Gradient Boosting

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lightgbm	Light Gradient Boosting Machine	6.5667	256.3960	15.3773	0.2908	0.6482	0.7896	0.0940
rf	Random Forest Regressor	6.2327	261.2200	15.4963	0.2753	0.5766	0.7628	0.1380
gbr	Gradient Boosting Regressor	6.3729	262.7572	15.5827	0.2733	0.5940	0.8038	0.0450
knn	K Neighbors Regressor	6.4223	275.6775	15.9429	0.2558	0.6483	0.8069	0.0070
et	Extra Trees Regressor	6.2941	276.2756	16.0651	0.2156	0.5948	0.7702	0.1350
catboost	CatBoost Regressor	6.3927	298.8103	16.5613	0.1432	0.5946	0.8093	1.3920
huber	Huber Regressor	5.5041	451.6802	17.9369	-0.1599	0.5502	0.7247	0.0690
dt	Decision Tree Regressor	7.4284	353.4164	18.1480	0.0019	0.6949	0.7872	0.0060
xgboost	Extreme Gradient Boosting	7.0636	351.8369	18.2194	-0.0468	0.6319	0.8567	0.0770
llar	Lasso Least Angle Regression	9.7574	352.6005	18.3706	-0.0036	0.9970	2.0040	0.4410
dummy	Dummy Regressor	9.7574	352.6005	18.3706	-0.0036	0.9970	2.0040	0.0050
ridge	Ridge Regression	7.8872	425.9960	18.4496	-0.1230	0.7883	1.3294	0.0050
en	Elastic Net	7.5688	463.8348	18.6746	-0.2100	0.7523	1.3256	0.0050
omp	Orthogonal Matching Pursuit	7.4721	444.3789	18.7219	-0.1777	0.7282	1.2185	0.2090
br	Bayesian Ridge	7.5450	471.0667	18.7226	-0.2283	0.7416	1.2893	0.0090
lasso	Lasso Regression	7.5980	468.0804	18.7451	-0.2216	0.7542	1.3370	0.0050
lr	Linear Regression	8.1927	441.7997	19.0037	-0.1802	0.8147	1.3835	0.0060
ada	AdaBoost Regressor	18.0046	541.2532	22.7589	-0.7412	1.3505	4.1827	0.0320
par	Passive Aggressive Regressor	28.9948	7093.4975	50.5493	-16.0802	1.2181	8.2720	0.0060

Fig. 1. Comparison of regressor model performance.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
catboost	CatBoost Classifier	0.6917	0.8559	0.6834	0.6980	0.6914	0.5365	0.5393	3.1550
rf	Random Forest Classifier	0.6910	0.8498	0.6810	0.6889	0.6848	0.5345	0.5389	0.0530
gbc	Gradient Boosting Classifier	0.6910	0.8548	0.6829	0.6985	0.6911	0.5356	0.5387	0.1640
xgboost	Extreme Gradient Boosting	0.6872	0.8500	0.6803	0.6948	0.6890	0.5303	0.5318	0.1830
lightgbm	Light Gradient Boosting Machine	0.6789	0.8507	0.6708	0.6843	0.6796	0.5175	0.5192	0.0440
ada	Ada Boost Classifier	0.6744	0.7841	0.6644	0.6723	0.6686	0.5096	0.5137	0.0250
et	Extra Trees Classifier	0.6391	0.8180	0.6274	0.6297	0.6297	0.4558	0.4595	0.0440
lr	Logistic Regression	0.6338	0.8070	0.6238	0.6311	0.6292	0.4484	0.4510	0.1280
knn	K Neighbors Classifier	0.6226	0.7853	0.6130	0.6464	0.6199	0.4321	0.4438	0.0080
dt	Decision Tree Classifier	0.5932	0.6978	0.5845	0.6015	0.5953	0.3892	0.3903	0.0060
nb	Naive Bayes	0.5917	0.7779	0.5854	0.6177	0.5931	0.3878	0.3949	0.0030
lda	Linear Discriminant Analysis	0.5887	0.7799	0.5816	0.6008	0.5919	0.3830	0.3848	0.0080
ridge	Ridge Classifier	0.5880	0.0000	0.5739	0.5676	0.5712	0.3771	0.3822	0.0030
svm	SVM - Linear Kernel	0.5594	0.0000	0.5402	0.5084	0.4747	0.3257	0.3586	0.0060
qda	Quadratic Discriminant Analysis	0.3586	0.5215	0.3602	0.4513	0.2916	0.0420	0.0590	0.0060
dummy	Dummy Classifier	0.3564	0.5000	0.3333	0.1270	0.1873	0.0000	0.0000	0.0050

Fig. 2. Comparison of classifier model performance.

Machine model demonstrated the highest R^2 and the lowest RMSE, the latter of which is commonly used as an indicator to compare the performance of regression models.

Based on these findings, the Light Gradient Boosting Machine was selected as the machine learning model for this study, aimed at estimating the length of stay in the ICU.

The Light Gradient Boosting Machine (Lightgbm) is an advanced version of the gradient learning framework based on decision trees. The term "Boosting" refers to a family of algorithms that are able to transform weak learners into strong ones. Owing to its high prediction accuracy, rapid computational speed, and commendable ability to minimize overfitting issues, Lightgbm has been extensively employed in various fields [4].

3. Results

The processing for this study was performed on an Intel Core™ i7-10700 CPU@ 2.90 GHz desktop with 16.0 GB of RAM, operating on the Windows 11 Pro platform. For data manipulation, analysis, and algorithm training, Jupyter Notebook was utilized, an interactive web-based development environment for creating notebook documents. The codes were written in Python 3.

All the models were acquired through the PyCaret library, a framework that employs models from the Sklearn library, the most widely used in data science studies, as well as other libraries [1]. This library supports basic machine learning tasks such as regression, classification, clustering, and anomaly detection, simplifying the workflow and making it an excellent choice for both beginners and experts wanting to prototype machine learning models quickly [5].

The development of the predictive model incorporated the following stages: 1. Dataset preparation: elimination of cases with missing values, normalization or standardization of variables, partitioning data into subsets for training, validation, and testing; 2. Training and tuning of algorithm hyperparameters; 3. Measurement and comparison of model performances. The PyCaret library facilitated all of these steps.

Each stage is described in more detail below.

For the purpose of this study, the hospital provided the dataset and its technical expertise. The dataset, extracted from the ASSIST database, contains the history of 2240 consecutive cardiac surgeries performed on patients with congenital heart diseases from 2014 to 2018. The patient information is structured into 145 pre-, intra-, and postoperative variables, 93 of which were selected as predictors of the length of stay in the ICU. Most of these variables were derived from international risk checklists such as Rachs-1 and Aristotle.

In the data pre-processing phase, variables were standardized and cases with missing values were eliminated, resulting in a sample of 1642 patients for this post-hoc analysis.

These 1642 patients were randomly partitioned into 1478 patients (90%) for modeling and 164 patients (10%) for testing the models' accuracy. The modeling dataset (1478 patients) was further randomly divided into a training set (1330 patients) and a validation set (148 patients). The training set was used to train the models and the validation set to adjust the model's hyperparameters for performance enhancement.

To ensure better performance of the machine learning models and minimize variance [6,7], the K-fold cross-validation technique was utilized. This method splits the dataset into K equally sized parts, also known as folds. The training process is then applied to all but one of these folds, which is reserved for validation. The average performance across all tests serves as a measure of the model's

performance. The benefit of this method is that it allows for the entire dataset to be both trained and tested, reducing the variance of the chosen estimator. A 10-fold cross-validation method (k = 10) was used, which has demonstrated good adequacy in numerous health-related studies, with low variability between the training and test samples [8,9–13].

The performance of the models developed by the PyCaret library for predicting ICU length of stay was evaluated and compared using statistical adherence parameters shown in Figs. 1 and 2. According to MSE, RMSE, and R^2 criteria, the Light Gradient Boosting Machine regression model performed the best. The optimal hyperparameters for this model were: boosting_type = 'gbdt', class_weight = None, colsample_bytree = 1.0, importance_type = 'split', learning_rate = 0.1, max_depth = −1, min_child_samples = 20, min_-child_weight = 0.001, min_split_gain = 0.0, n_estimators = 100, n_jobs = −1, num_leaves = 31, objective = None, random_state = 24, reg_alpha = 0. Fig. 3 presents the most influential variables impacting a patient's ICU length of stay, as per the regression method.

The model yielded a root mean square error (RMSE) of 15.4 days in the training dataset, 11.8 days in the validation dataset, and 15.2 days in the unseen dataset, representing the discrepancies between the observed values of ICU stay and those predicted by the model.

For this current analysis, despite an RMSE of approximately 15 days, outliers were included to enhance the model's ability to predict such exceptional cases. However, another option to tackle the issue of outliers would be to classify the ICU stay into categories, that is, patients who stayed in the ICU for 1–3 days (short duration), patients who stayed in the ICU between 4 and 7 days (moderate duration), and patients who stayed in the ICU beyond 7 days (long duration). According to Accuracy, AUC, Recall, F1, Kappa, and MCC criteria, the Cat Boost Classifier classification model performed the best. Figs. 4–6 present the most influential variables impacting a patient's ICU length of stay, as per the classification method, for each of the 3 categories (short, moderate, and long ICU stay). Finally, Figs. 7 and 8 display the ROC Curve and the Confusion Matrix for the CatBoost Classifier classification model.

4. Discussion

This study aligns with recent research on the use of individualized artificial intelligence models to predict the length of stay in the ICU for cardiac patients who are potential surgical candidates. Such research has shown that machine learning techniques can be an effective tool to support medical decision-making [14].

For instance, Bertsimas et al. [14] posited that numerous risk factors affecting outcomes in congenital heart surgery may not interact in a linear and additive manner, as assumed in a logistic regression model. This justifies the application of machine learning models. These authors developed machine learning models to predict mortality, duration of postoperative mechanical ventilatory support, and hospital stay length for patients undergoing congenital heart surgery, based on data from over 235,000 patients and 295,000 operations provided by the European Congenital Heart Surgeons Association Congenital Database. Optimal Classification Trees, Random Forest, and Gradient Boosting models demonstrated superior predictive accuracy compared to the logistic regression model for all three outcomes.

Fig. 3. The most influential variables in predicting the length of stay in the ICU by regression method.
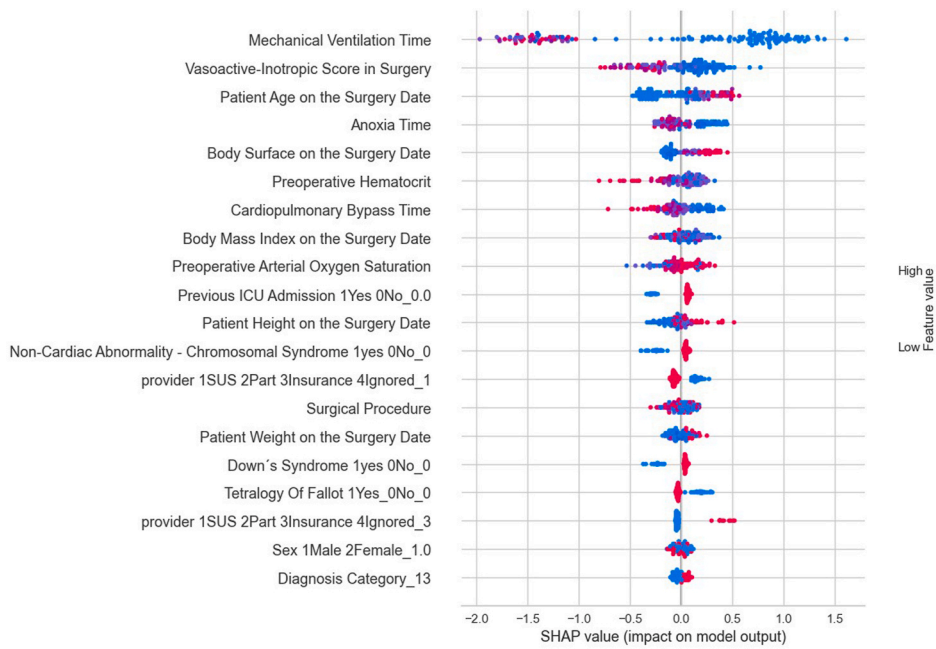
Fig. 4. The most influential variables in predicting from 1 to 3 Days in the ICU by classifier method.
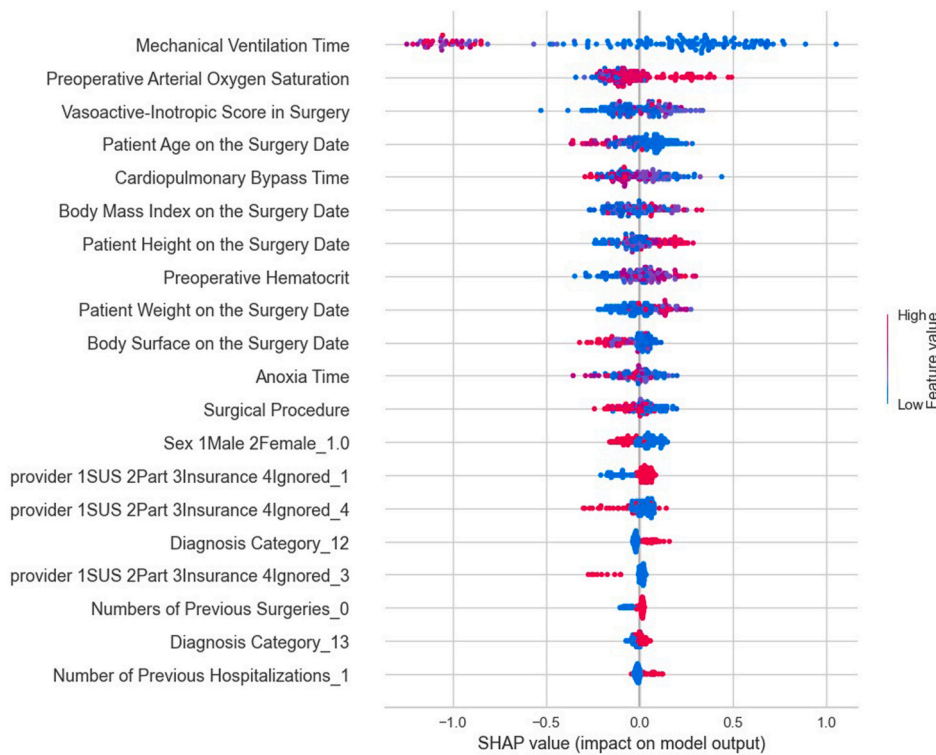

Fig. 5. The most influential variables in predicting from 4 to 7 Days in the ICU by classifier method.

Additionally, Daghistani et al. [15] proposed a Random Forest model to predict the length of hospital stay (LOS) for cardiac patients at three levels (low – less than 3 days, medium – between 3 and 5 days, and high – more than 5 days). They noted that patients' LOS varied widely, depending on their personal characteristics and the specific features of the hospital.

Furthermore, Alsinglawi et al. [16] concluded that the Gradient Boosting Regressor machine learning model had the best
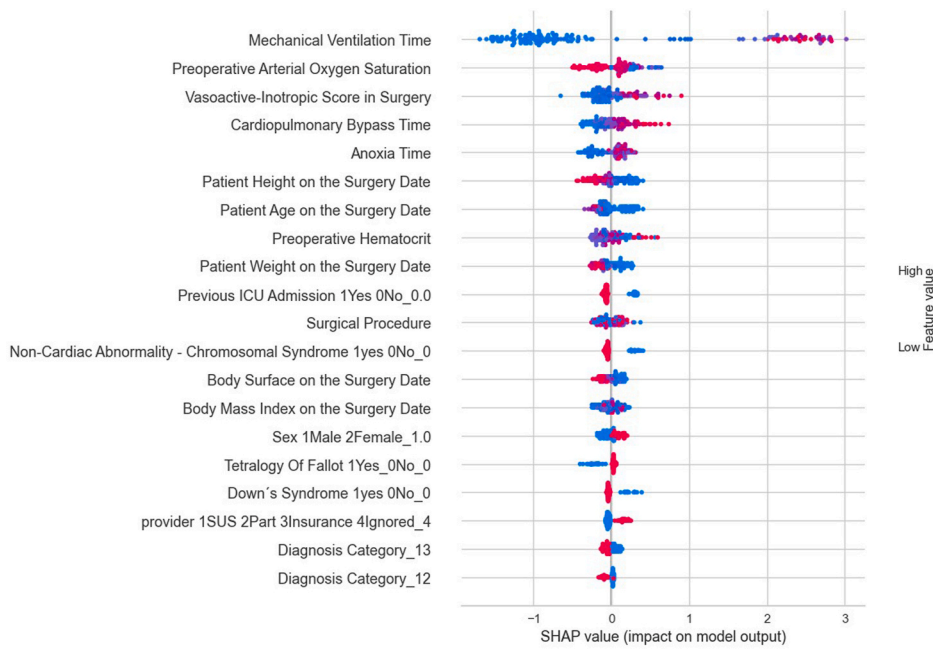
Fig. 6. The most influential variables in predicting above 7 Days in the ICU by classifier method.
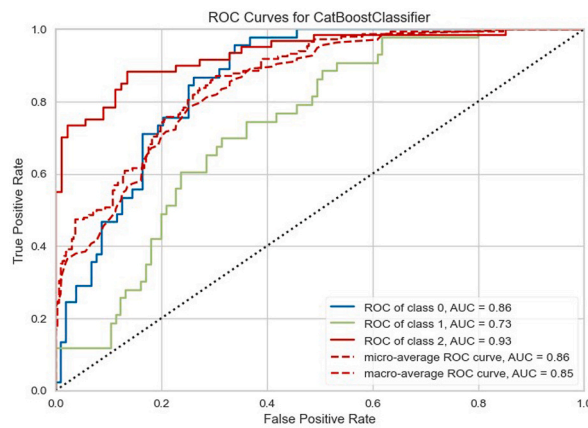
Fig. 7. Roc curves for CatBosst classifier.

performance in predicting the length of hospital stay for coronary patients.

Our results support these findings in the published international scientific literature. The variables identified as the most significant in predicting the ICU length of stay align with aggregated data published by the STS.

Another consideration is the extent to which parameter measures influence the outcomes. The international RACHS-1 risk strata were primarily designed to reflect the complexity of surgical procedures. Boethig et al. [17] applied the risk-adjusted classification for congenital heart surgery – RACHS-1 – to 2223 patients under 18 years of age in Bad Oeynhausen, Germany, to analyze its relationship with mortality and length of hospital stay. They found that the length of hospital stay increases exponentially with the RACHS-1 category, but the RACHS-1 classification only explained 13.5% of the hospital stay times and 16.8% of the individual postoperative times for surviving patients. This result was attributed to the substantial variability in the length of hospital stay and post-surgery times within each RACHS-1 group. This intra-group variability, as observed in the Boethig et al. [17] study, may indicate the influence of hospital or individual characteristics.

Moreover, Nina et al. [18] used the same RACHS-1 classifier on 145 patients in a public hospital in Northeast Brazil to evaluate the applicability of RACHS-1 as a mortality predictor. They concluded that, "despite the ease of application of the RACHS-1, it cannot be applied in our environment because it does not include other variables present in our reality that can interfere with the surgical result," such as nutritional status, among other factors.

Furthermore, upon excluding patients over 18 years of age from the ASSIST database, Fig. 9a and b exhibit significant variability in
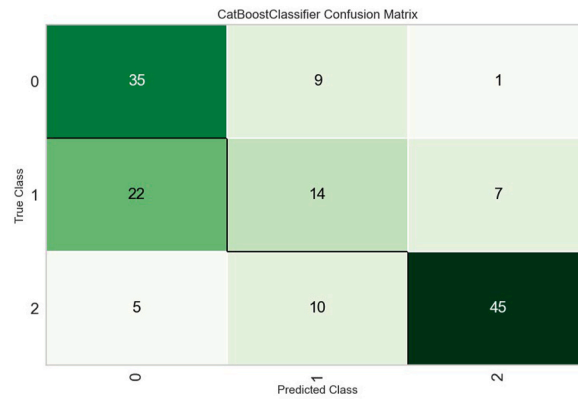
Fig. 8. CatBoost classifier confusion matrix.

ICU length of stay within each RACHS-1 class, mirroring the observations made by Boethig et al.

Given the reasons stated above, it was essential to propose a model to estimate the ICU length of stay on an individual basis. Understanding which diagnoses and variables affect the duration of ICU stay for a CHD patient undergoing cardiac surgical intervention enables patients, their families, doctors, surgeons, and healthcare professionals to critically evaluate associated risks, plan, comprehend, and communicate such information. This knowledge serves to educate and support their decision-making process.

Upon examining the variables identified by the models, through their respective SHAP values, as most impactful, it was observed that the risk of a prolonged ICU stay increases with the duration of mechanical ventilation, a higher vasoactive-inotropic score (VIS) during surgery, a longer duration of extracorporeal circulation (ECC), extended anoxia time, lower body mass index (BMI), lower height and weight, younger age at the time of surgery, lower pre-surgical hematocrit, and lower pre-surgical oxygen saturation. Therefore, it is crucial to direct specific studies towards these groups and variables, which could guide actions to mitigate ICU length of stay.

Additionally, Cordeiro et al. [19] identified factors that prolong mechanical ventilation time (VT) and consequently extend the ICU stay: advanced age, female gender, longer cardiopulmonary bypass duration, dysfunction, and low cardiac output. The same authors [20] also concluded that an extended duration of invasive VT negatively impacts peripheral muscle strength in patients undergoing cardiac surgery, and suggested strategies for early VT weaning. One of these strategies was tested by Chiang et al. [21] who implemented a six-week physical training program in a specialized respiratory care unit. Following the training, patients gained peripheral muscle strength in their upper and lower limbs, reduced their ICU stay length, and achieved higher measures of functional independence and Barthel Score. Chiang et al. noted that among the procedures performed by postoperative physical therapists, early ambulation generates hemodynamic impact safely and without risk for this patient profile.

Tabib et al.'s research [22] also concluded that factors such as younger age, lower weight, heart failure, higher doses of vasoactive-inotropic, pulmonary hypertension, respiratory infections, and late sternal closure were predictors of mechanical
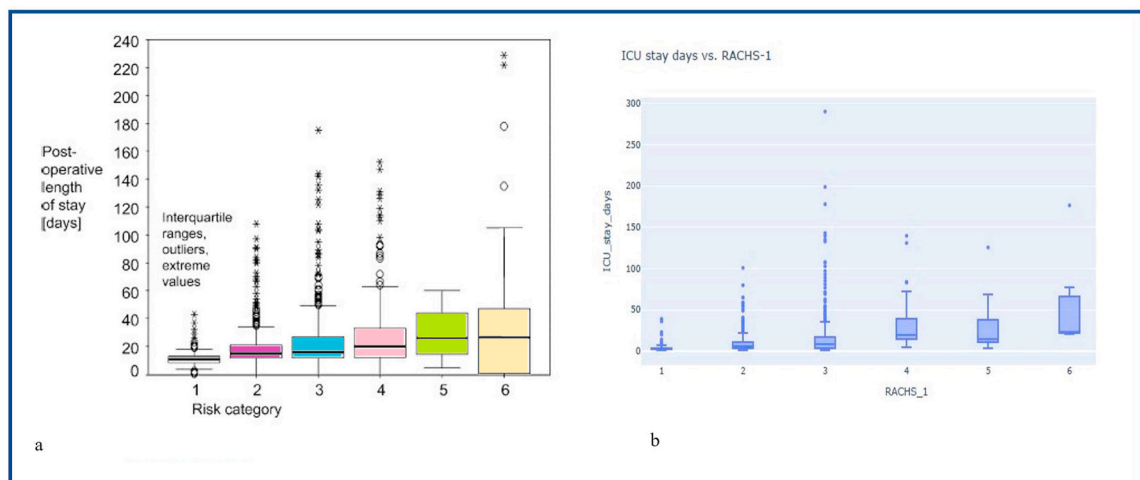

Fig. 9. a - Postoperative permanence time in Germany ICU.
b - Length of stay in the ICU of the researched hospital.
Source: Boethig et al. [17].

ventilation time in a multivariate analysis. Tabib et al.'s results suggested that predictors of mechanical ventilation time can be specific to each operating room. Therefore, it is crucial to have an efficient management program for each pediatric heart surgery center, emphasizing the value of preoperative management for patients undergoing congenital cardiac surgery and the planning of intensive care facilities.

5. Perspectives

Future research is necessary to evaluate new machine learning algorithms, test novel variables and diagnoses, and conduct an in-depth analysis of the effects of these variables on the explanation of the ICU length of stay for patients with congenital heart disease undergoing cardiac surgery.

In addition, given the observed variation in ICU length of stay within the ASSIST sample patients, with an average of 10.93 days and a standard deviation of 18.56 days (a 10–12 day interval with 95% confidence, but with extremes ranging from 0.63 days to 290.09 days), it will be possible to perform clustering before applying regression methods once the ASSIST Registry has accrued a larger patient sample.

This work is an additive study producing supporting data for the entire AI prediction modeling enterprise.

Ethical approval

This work was approved by the Ethics Committee of the Heart Institute of University of São Paulo Medical School – InCor - HCFMUSP, São Paulo, Brasil, protocol CAEE: 29385320.1.0000.0068.

Data availability statement

The data are not publicly available due to their containing information that could compromise the privacy of research participants.

CRediT authorship contribution statement

João Chang Junior: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Luiz Fernando Caneo:** Formal analysis, Investigation, Resources, Supervision, Validation, Visualization, Writing – review & editing. **Aida Luiza Ribeiro Turquetto:** Data curation, Formal analysis, Investigation, Resources, Supervision, Validation, Visualization. **Luciana Patrick Amato:** Data curation, Formal analysis, Investigation, Resources. **Elisandra Cristina Trevisan Calvo Arita:** Data curation, Resources, Visualization. **Alfredo Manoel da Silva Fernandes:** Data curation, Formal analysis, Project administration, Resources, Visualization, Writing – review & editing. **Evelinda Marramon Trindade:** Investigation, Resources, Validation, Visualization, Writing – review & editing. **Fábio Biscegli Jatene:** Formal analysis, Supervision, Validation, Visualization. **Paul-Eric Dossou:** Formal analysis, Visualization, Writing – review & editing. **Marcelo Biscegli Jatene:** Conceptualization, Formal analysis, Investigation, Project administration, Resources, Supervision, Validation, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Available at: https://pycaret.readthedocs.io/en/stable/, PyCaret, 2020.
[2] F. Carmona, et al., Collaborative quality improvement in the congenital heart defects: development of the ASSIST consortium and a preliminary surgical outcomes report, Brazilian Journal of Cardiovascular Surgery 32 (2017) 260–269.
[3] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) e623, https://doi.org/10.7717/peerj-cs.623.
[4] F. Junliang, M. Xin, W. Lifeng, et al., Light Gradient Boosting Machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data, Agricultural Water Management 225 (2019), https://doi.org/10.1016/j.agwat.2019.105758.
[5] G. Tolios, Simplifying Machine Learning with PyCaret - A Low-Code Approach for Beginners and Experts! Leanpub, 2022.
[6] M.W. Browne, Cross-validation methods, Journal of Mathematical Psychology 44 (4) (2001) 108–132.

[7] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: International Joint Conference on Artificial Intelligence, 1995, 0, 0–6, https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf.

[8] D. Ruiz-Fernández, et al., Aid decision algorithms to estimate the risk in congenital heart surgery, Computer Methods and Programs in Biomedicine (2016).

[9] A. Awad, et al., Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach, International journal of medical informatics 108 (2017) 185–195, https://doi.org/10.1016/j.ijmedinf.2017.10.002.

[10] M.M. Churpek, et al., Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards, Critical Care Medicine 44 (2) (2017) 298–305. https://pubmed.ncbi.nlm.nih.gov/26771782/.

[11] T. Desautels, et al., Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach, BMJ Open 7 (9) (2017) 1–9.

[12] K.B. Cohen, et al., Methodological issues in predicting pediatric epilepsy surgery candidates through natural language processing and machine learning, Biomedical Informatics Insights 8 (BII) (2016) S38308.

[13] R. Pirracchio, et al., Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study, The Lancet Respiratory Medicine 3 (1) (2015) 42–52, https://doi.org/10.1016/S2213-2600(14)70239-5.

[14] D. Bertsimas, D. Zhuo, J. Dunn, et al., Adverse outcomes prediction for congenital heart surgery: a machine learning approach, World Journal for Pediatric and Congenital Heart Surgery 12 (4) (2021) 453–460, https://doi.org/10.1177/21501351211007106.

[15] T.A. Daghistani, R. Elshawi, S. Sakr, et al., Predictors of in-hospital length of stay among cardiac patients: a machine learning approach, International Journal of Cardiology (2019).

[16] B. Alsinglawi, F. Alnajjar, O. Mubin, et al., Predicting length of stay for cardiovascular Hospitalizations in the intensive care unit: machine learning approach, in: Annual Int Conf IEEE Eng Med Biol Soc, vol. 2020, 2020, pp. 5442–5445, https://doi.org/10.1109/EMBC44109.2020.9175889.

[17] D. Boethig, K.J. Jenkins, H. Hecker, et al., The RACHS-1 categories reflect mortality and length of hospital stay in a large German pediatric cardiac surgery population, European Journal of Cardio-thoracic Surgery 26 (2004) 12–17.

[18] R.V. Nina, M.E.A. Gama, A.M. Santos, et al., Is the RACHS-1 (Risk adjustment in congenital heart surgery) a useful tool in our scenario? Rev Bras Cir Cardiovasc 22 (4) (2007) 425–431.

[19] L.L. Cordeiro, et al., Analysis of Mechanical Ventilation Time and Hospitalization of Patients Undergoing Cardiac Surgery, ABCS Health Sciences, 2016, https://doi.org/10.7322/abcshs.v42i1.942.

[20] L.L. Cordeiro, et al., Mechanical ventilation time and peripheral muscle strength in post-heart surgery, Int J Cardiovasc Sci 29 (2) (2016) 134–138, https://doi.org/10.5935/2359-4802.20160021.

[21] L.L. Chiang, L.Y. Wang, C.P. Wu, et al., Effects of physical training on functional status in patients with prolonged mechanical ventilation, Physical Therapy 86 (9) (2006) 1271–1281.

[22] A. Tabib, et al., Predictors of prolonged mechanical ventilation in pediatric patients after cardiac surgery for congenital heart disease, Res Cardiovasc Med 5 (3) (2016) e30391, https://doi.org/10.5812/cardiovascmed.30391.