BMC Cardiovascular Disorders

RESEARCH

Open Access

Using artificial intelligence to predict post-operative outcomes in congenital heart surgeries: a systematic review

Ida Mohammadi[1], Shahryar Rajai Firouzabadi[1,2], Melika Hosseinpour[1], Mohammadhosein Akhlaghpasand[1,3*], Bardia Hajikarimloo[1], Sam Zeraatian-Nejad[1,3] and Peyman Sardari Nia[4,5]

Abstract

Introduction Congenital heart disease (CHD) represents the most common group of congenital anomalies, constitutes a significant contributor to the burden of non-communicable diseases, highlighting the critical need for improved risk assessment tools. Artificial intelligence (AI) holds promise in enhancing outcome predictions for congenital cardiac surgery. This study aims to systematically review the utilization of AI in predicting post-operative outcomes in this population.

Methods Following PRISMA guidelines, a comprehensive search of Pubmed, Scopus, and Web of Science databases was conducted. Two independent reviewers screened articles based on predefined criteria. Included studies focused on AI models predicting various post-operative outcomes in congenital heart surgery.

Results The review included 35 articles, primarily published within the last four years, indicating growing interest in AI applications. Models predominantly targeted mortality and survival ($n=16$), prolonged length of hospital or ICU stay ($n=7$), postoperative complications ($n=6$), prolonged mechanical ventilatory support time ($n=4$), with additional focus on specific outcomes such as peri-ventricular leucomalacia ($n=2$) and malnutrition ($n=1$). Performance metrics, such as area under the curve (AUC), ranged from 0.52 to 0.997. Notably, these AI models consistently outperformed traditional risk stratification categories. For instance, in assessing the risk of morbidity and mortality, the AI models demonstrated superior performance compared to conventional methods.

Conclusion AI-driven prediction models show significant promise in improving outcome predictions for congenital heart surgery. They surpass traditional risk prediction tools not only in immediate postoperative risks but also in long-term outcomes such as 1-year survival and malnutrition. Further studies with robust external validation are necessary to assess the practical applicability of these models in clinical settings.

The protocol of this review was prospectively registered on PROSPERO (CRD42024550942).

Keywords Artificial Intelligence, Machine Learning, Cardiac Surgery, Congenital Heart Defects

*Correspondence:
Mohammadhosein Akhlaghpasand
Akhlaghpasandm@yahoo.com
Full list of author information is available at the end of the article

Introduction

Congenital heart disease (CHD) is the most prevalent type of major congenital anomalies, constituting nearly one-third of such disorders [1]. Not long ago, only a small fraction of patients with moderate and severe CHDs reached adulthood. In the 1950s, however, the introduction of cardiopulmonary bypass significantly advanced the application of surgery as a treatment for congenital heart disease, resulting in marked improvement in long-term outcomes [2, 3]. Yet even with advancements in surgery, CHD plays a substantial role in the overall burden of non-communicable diseases (NCDs) [4], with postoperative complications in 46–74% of all operated patients [5, 6]. Prediction of these outcomes can be critical in operative and post-operative decision-making.

To predict postoperative outcomes in CHD patients, such as morbidity and mortality, several risk stratification categories have been introduced, including Risk Adjustment in Congenital Heart Surgery-1 (RACHS-1) [7] and Society of Thoracic Surgery-European Association for Cardiothoracic Surgery Congenital Heart Surgery Mortality Categories (STS-EACTS). However, it has been demonstrated that only the risk stratifications that depend predominantly on expert opinion and consensus, underperform in comparison to those that are more evidence-based [8]. Therefore, there is a need for new prediction models based on big data. [9]. Linear logistic regression models have shown promise in the past [10], yet the advent of artificial intelligence (AI) and the development of more complex AI models has shown great applicability [11].

In recent years, AI has emerged as a novel promising approach to data science [12]. With regard to AI prediction models, the most distinctive factor is that these models learn from examples rather than being programmed by rules as in traditional predicting models [13]. This makes these AI models non-linear complex tools well-suited for identifying and illustrating patterns that are either unknown to or too complex for traditional biostatistics [14, 15]. Moreover, AI models can distinguish variables most impactful in a particular process, allow for a greater number of predictive variables to be incorporated into a model, and learn important features by training on original datasets [16]. As the diverse array of AI predictive modeling techniques and the utility of each may seem confounding, Table 1 showcases models that are most frequently utilized by reports in the literature in order to provide better insight [17–21]. Numerous AI models have been designed to predict outcomes following various types of surgeries, including cardiac surgeries [22].

The aim of the current study was to systematically review the relevant literature and identify all studies that have utilized AI to predict post-operative outcomes in patients undergoing surgery for congenital heart defects.

Methods

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [23]. The protocol of this review was prospectively registered on PROSPERO (CRD42024550942).

Literature search strategy

A comprehensive search was conducted on May 4th, 2024 in 3 electronic databases, Pubmed, Web of Science, and Scopus by one author (IM). No language or publication limitations were applied. The search results were then passed on to S.R. and M.H. for further evaluation.

The search terms used included combinations of 3 separate search parameters, combined using Boolean operators:

Parameter a: ("artificial intelligence", "deep learning", and "machine learning").

Parameter b: ("congenital heart defect(s)", "tetralogy of Fallot", "atrial septal defect", "ventricular septal defect", "patent ductus arteriosus", and "patent foramen ovale").

Parameter c: ("surgery", "operation", "heart", and "cardiac").

Screening and eligibility criteria

Screening was carried out in 2 phases, with an initial title-abstract screening followed by a full-text evaluation, by two independent authors (S.R., M.H.), with discrepancies being resolved through discussion with a third reviewer (I.M.). Papers employing an AI model to predict a post-operative outcome after congenital heart surgery were included based on the PICOS (population, intervention, comparator, outcome, and study design (PICOS) criteria defined below. Artificial intelligence predictive models were identified according to the definitions provided by Jovel et al. [24].

Population: Patients undergoing congenital heart surgery.

Intervention: AI models and algorithms.

Comparator: Current algorithms used to predict outcomes, if available.

Outcome: Postoperative outcomes.

Study design: AI development and/or validation studies.

Papers that were conducted on animals were excluded. Study designs including reviews, meta-analyses, case reports, preclinical studies, editorials, book chapters, and conference abstracts were also excluded. In addition, due to ambiguity in defining what exactly constitutes a machine learning model, some consider LR a traditional

Table 1 A brief description of the most frequently used AI modeling techniques and advantages and disadvantages

Model Type	Description	Advantages	Disadvantages
Logistic Regression	A statistical model used for predicting binary outcomes based on one or more predictor variables	Easy to implement, interpretability, and computationally efficient	Assumes linearity, not suitable for complex relationships
Decision Trees and Ensemble Variants (Random Forest, Extra Trees, Optimal Classification Trees)	Models that use tree structures for decision-making, including individual trees and ensembles like Random Forest, Extra Trees, and Optimal Classification Trees	Easy to visualize and interpret, reduces overfitting with ensembles, handles high-dimensional data well	Prone to overfitting (single trees), less interpretable (ensembles), computationally intensive
Support Vector Machine (SVM)	A supervised learning model used for classification and regression that finds the optimal hyperplane	Effective in high-dimensional spaces, robust to overfitting	Requires careful parameter tuning, less effective on large datasets
Gradient Boosting (GBM, XGBoost, LightGBM, CatBoost)	An ensemble method that builds models sequentially, correcting previous errors; includes variants like XGBoost, LightGBM, and CatBoost	High accuracy, speed, scalability, automatic handling of categorical variables	Prone to overfitting, requires careful tuning, high computational demands
Neural Networks and Deep Learning (DNN, CNN)	Models inspired by the human brain, consisting of layers of interconnected nodes (neurons); includes DNNs and CNNs for complex representations	Capable of capturing complex patterns, highly flexible, exceptional at processing unstructured data	Requires large datasets, extensive computational resources, can be a "black box"
Naive Bayes	A probabilistic classifier based on Bayes' theorem, assuming independence among predictors	Simple, fast, performs well with small datasets	Assumes feature independence, less accurate with correlated features
K-Nearest Neighbors (KNN)	A non-parametric model that classifies based on the closest training examples in the feature space	Simple to implement, intuitive	Sensitive to irrelevant features and data scale
Adaptive Boosting (AdaBoost)	An ensemble technique that adjusts weights of misclassified instances in successive iterations	Improves weak learners, robust to noise	Sensitive to outliers, may overfit with noisy data
Linear Discriminant Analysis (LDA)	A statistical method for classifying samples based on linear combinations of features	Handles multiclass classification, interpretable	Assumes normal distribution of features, linear boundaries

statistical model, while others refer to it as a machine learning one [17–21]. For the purposes of this review, we considered LR to be an AI model and thus included predictive models employing it.

Outcomes

The intention of the reviewers was to review all postoperative outcomes examined in the literature. After the completion of full-text screening, we ended up reviewing the following outcomes: prediction of risk in congenital heart surgery based on RACHS (risk adjustment for congenital heart surgery) classification, critical events, morbidities, and complications such as prolonged length of hospital stay and prolonged mechanical ventilatory support time, in-hospital and 30-day mortality, periventricular leukomalacia (PVL), one-year transplant-free survival after Norwood procedure, pulmonary arterial hypertension, echocardiography variables related to the right atrium, postoperative pulmonary vein obstruction, interstage mortality between stage I and II of the Norwood procedure while at home, and malnutrition.

Data extraction and risk of bias assessment

A data extraction template was prepared beforehand which consisted of the following data: first author, year of publication, country, source of data, type of source, type of study, population size, population age, gender, whether the patients are adult or pediatric, type of defect, algorithm/model, NO. of variables, outcomes, mode of validation, calibration, missing data strategy, measures of performance of each outcome, the best algorithm, and main findings. The template was independently filled with the required qualitative and quantitative data by two authors (SH.R., M.H.), with a third author (I.M.) re-evaluating the extracted data and correcting any disagreements.

The risk of bias in the selected articles was assessed using the PROBAST tool [25] by two independent reviewers (SH.R., M.H.) with a third reviewer (I.M.) resolving any variations through discussion. This tool consists of 4 domains for bias detection, Participants, Predictors, Outcome, and Analysis, with a total of 20 criteria. Each domain can be graded with a low risk, a high risk, or an unclear risk of bias. Studies with a low risk of bias in all four domains were classified as low-risk while having a high risk of bias in even one domain would result in judging the study as high-risk. If one or more domains had an unclear risk of bias with the rest being low-risk, the risk of bias attributed to the study would then be declared unclear.

World map was designed using https://www.mapchart.net/world.html.

Results

Study selection

A total of 335 articles were identified in the literature search. There were 90 duplicates, the removal of which yielded 245 references. A further 188 references were excluded in the initial phase of screening based on the title and the abstract, and 57 articles were sought for full-text retrieval. During the second phase of screening 22 studies were excluded due to the targeted population not being that of patients undergoing congenital heart surgeries ($n=8$), the intervention not being AI-based ($n=4$), or the assessed endpoint not being a postoperative outcome ($n=10$), resulting in the inclusion of 35 articles in the database search. (Fig. 1).

Study characteristics

The characteristics of the included population in each study are outlined in Table 2. The included studies were published from 2013 to 2023 and mostly from United States (Fig. 2). The source of data in the majority of studies were the local electronic medical records (EMR). The included studies trained and tested their models in between 56 and 221,335 individuals, mostly pediatric patients with no particular focus on the type of congenital heart disease. Among the studies with a focus on specific types of defects, single ventricle physiology and tetralogy of Fallot (TOF) were the most frequently investigated (Fig. 2).

The details regarding the AI prediction models constructed in each study are displayed in Table 3. The most commonly utilized algorithm was logistic regression (LR, $n=16$) followed by boosting (B, $n=14$), random forests (RF, $n=9$), support vector machine (SVM, $n=8$), decision trees and artificial neural networks (DT and ANN, $n=7$ each), k-nearest neighbor and naïve bayes (KNN and NB, $n=4$ each) (Fig. 2). In addition, the most frequently assessed outcome was short-term mortality with 11 studies either reporting in-hospital or 30-day mortality. All of the included studies were retrospective in design. Moreover, most of the studies ($n=15$) utilized K-fold cross-validation while 14 used training and testing with or without bootstrapping as their mode of validation. Only 6 studies externally validated their models.

AI Models and Measures of Performance

Overall AUCs above 0.7 were achieved in at least one model in all studies which reported their respective AUC, with the exception of one study by Sunthankar et al. [52]. Short-term mortality was the most investigated outcome, and different AI models across four different articles were used to measure it (Fig. 3). The best performance among the models predicting this outcome was achieved by

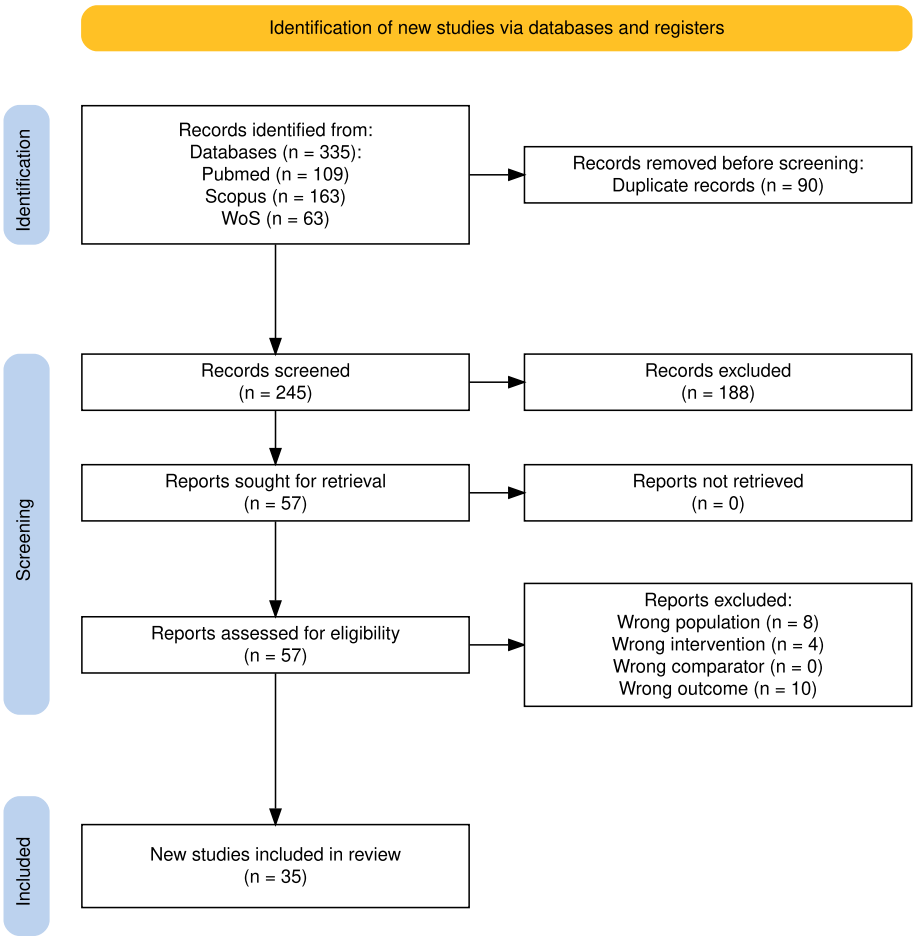

Fig. 1 PRISMA flowchart detailing the search and screening process, made using the Shinyapp by Haddaway et al. [26]

logistic regression (LR) with an AUC of 92.6% as reported by Zürn et al. [54], while the worst-performing model was also an LR model with an AUC of 72%, as reported by Cocomello et al. [36].

Risk of bias assessment

Overall, ten, fourteen, and eleven studies had high, low, and unclear risk of bias, respectively. The most common domain of bias among the studies was domain 4 (analysis), while the domain imposing the least amount of bias was domain 3 (outcome). Overall, 60% of the included studies had a high or unclear risk of bias, highlighting the need for more detailed quality control in this field (Fig. 4).

Discussion

This systematic review explores the application of AI (deep learning and machine learning) in the prediction of outcomes in congenital cardiac surgery. Overall 35 articles were included, most of which were published in the past four years, showing an increased interest in the use of AI. A meta-analysis to assess and compare the performance of the models was not possible due to missing data, yet we can still observe the robust predictive performance exhibited by different artificial intelligence models, regardless of the outcomes they were designed to predict.

The models included focused on the prediction of mortality and survival ($n=16$) [27, 29, 31, 33, 37, 40, 46–48, 52–55, 60], prolonged length of hospital or ICU stay ($n=7$) [29, 40, 43, 46, 47, 57, 58], postoperative complications ($n=6$) [34, 35, 41, 42, 44, 59] prolonged mechanical ventilatory support time ($n=4$) [40, 47, 57, 59], with additional focus on specific outcomes such as periventricular leucomalacia [32, 39], acute kidney injury [56], malnutrition [50]. The AUCs of the models ranged between 0.52 to 0.997, with most models achieving an AUC above 0.7, highlighting the predictive potential of artificial intelligence in congenital heart surgery. Details for each assessed outcome are laid out below:

Discrimination evaluated by area under the curve was above 0.8 in almost all of the models evaluating mortality,

Table 2 Characteristics of the populations included in the studies

First author and year	Country	Source of data	Data range	Type of source	Population size	Train/Test Proportion	Gender (male)	Adult / Pediatric	Type of defect (procedure)
Crowe 2013 [27]	UK	Central Cardiac Audit Database	2000–2010	registry	37,044	2.5/1	training: 55% testing: NP	Pediatric	Not specified
Zapata-Impata 2015 [28]	Spain	NP	NP	NP	2432	3/1	NP	Pediatric	Not specified
Moein 2015 [29]	USA	Pediatric Health Information System (PHIS) database	2009–2014	registry	1036	NP	NP	Pediatric	Limited pulmonary blood flow (Systemic-to-Pulmonary Artery shunt)
Ruiz-Fernández 2015 [30]	Colombia	Cardiovascular Foundation of Colombia	NP	local EMR	2432	9/1	NP	Pediatric	Not specified
Rogers 2017 [31]	UK	National Congenital Heart Disease Audit	2009–2014	registry	21,838	4/1	NP	Pediatric	Not specified
Jalali 2018 [32]	USA	Children's Hospital of Philadelphia	NP	local EMR	71	1/6	47.89%	Pediatric	HLHS + non-HLHS
Luis Ahumadal 2018 [33]	USA	Pediatric Heart Networks Single Ventricle Reconstruction Trial	2005—2009	RCT	549	4/1	NP	Pediatric	HLHS (Norwood procedure)
Samad 2018 [34]	USA	Boston Children's Hospital	2005–2012	local EMR	153	4/1	49.67%	Adult	Tetralogy of Fallot
Ruiz 2019 [35]	USA	University of Pittsburgh Medical Center Children's Hospital of Pittsburgh	2014—2017	local EMR	93	9/1	NP	Pediatric	Single-ventricle physiology
Cocomello 2020 [36]	UK	National Congenital Heart Disease Audit (Bristol dataset)	2004–2009 2015–2019	registry	1st cohort: 1,352 2nd cohort: 1,197	NA	NP	Pediatric	Not specified
Chang Junior 2020 [37]	Brazil	ASSIST Registry (InCor's heart surgery program)	2014–2019	registry	2240	9/1	48.7%	Pediatric	Not specified
Huang 2020 [38]	China	Guangdong Provincial People's Hospital	2009–2017	local EMR	96	3/1	64.6%	Both	Single ventricle defects (a bidirectional Glenn procedure)
Bender 2021 [39]	USA	Children's Hospital of Philadelphia(CHOP)	NP	local EMR	56	2.12/1	NP	Pediatric	Not specified
Bertsimas 2021 [40]	USA	ECHSA Congenital Database	2000–2019	registry	221,335	3.8/1	55%	Both	Not specified
Faerber 2021 [41]	USA	Children's Hospital of Philadelphia	2012–2018	cohort	162	4/1	63%	Pediatric	Tetralogy of Fallot
Rusin 2021 [42]	USA	Texas Children's Hospital	2013–2019	local EMR	238	1/1	39.08%	Pediatric	Single-ventricle physiology
Ng 2022 [43]	China	Guangdong Provincial People's Hospital	NP	NP	58	8/2	NP	Pediatric	Complex CHD

Table 2 (continued)

First author and year	Country	Source of data	Data range	Type of source	Population size	Train/Test Proportion	Gender (male)	Adult / Pediatric	Type of defect (procedure)
Zeng 2021 [44]	China	Children's Hospital of Zhejiang University	2015–2018	local EMR	1964	7/3	48.77%	Pediatric	CHD
Thiriveedi 2021 [45]	USA	Johns Hopkins Hospital	NP	Cohort	162	NP	NP	Pediatric	Not specified
Jalali 2021 [46]	USA	Pediatric Heart Network Single Ventricle Reconstruction trial	2005–2009	Cohort	549	7/3	NP	Pediatric	Single ventricle disease (Norwood procedure)
Bertsimas 2022 [47]	USA	European Congenital Heart Surgeons Association (ECHSA) Congenital Database (ECCDB; 64 hospitals)	2010–2019	registry	31,792	2.8/1	NP	Both	Not specified
Du 2022 [48]	China	Shanghai Children's Medical Center	2006–2017	local EMR	24,685	7.5/2.5	57.59%	Pediatric	Not specified
Ekhomu 2022 [49]	USA	Retrospective cohort study	2012–2017	Cohort	153	7/3	62%	Pediatric	Tetralogy of Fallot
Shi 2022 [50]	China	Guangzhou Women and Children's Medical Center in China (registration number: NCT03626480)	2017–2020	Cohort	MDS: 536 EVS: 186	4/1	59.70%	Pediatric	Not specified
Pei 2022 [51]	China	Shanghai Children's Medical Center (SCMC) and Guangdong's Provincial People's Hospital (GPPH)	NP	local EMR	83	1.86/1	NP	Pediatric	Total anomalous pulmonary venous connection
Sunthankar 2023 [52]	USA	National Pediatric Cardiology Quality Improvement Collaborative (NPC-QIC) registry	2008–2019	registry	3267	4/1	62%	Pediatric	Single ventricle disease
Betts 2023 [53]	ANZ	Australia New Zealand Congenital Outcomes Registry for Surgery	2013–2022	registry	14,343	4/1	56.30%	Pediatric	Not specified
Zürn 2023 [54]	Germany	Departments of paediatric cardiology and cardiac surgery in Freiburg (train set) and Heidelberg (test set)	2014–2019	local EMR	MDS: 780 EVS: 985	1/1.26	NP	Pediatric	Not specified

Table 2 (continued)

First author and year	Country	Source of data	Data range	Type of source	Population size	Train/Test Proportion	Gender (male)	Adult / Pediatric	Type of defect (procedure)
Jiwani 2023 [55]	USA	NP	NP	NP	NP	NP	NP	NP	Not specified
Kong 2023 [56]	China	Children's Hospital of Chongqing Medical University	2002–2020	cohort	134	4/1	67.91%	Pediatric	Congenital malformations of the aorta
Sarris 2024 [57]	EU	ECCDB	1989–2022	registry	172,888	7/3	NP	NP	Not specified
Chang junior 2024 [58]	Brazil	ASSIST Registry	2014–2018	registry	1642	9/1	NP	NP	Not specified
Li 2024 [59]	China	Children's Hospital, Zhejiang University	2016–2021	local EMR	5030	18.6/1	training: 48.9% testing: 43.4%	Both	Not specified
Smith 2024 [60]	USA	Single Ventricle Reconstruction Extension Study	2005–2009	cohort	549	4/1	NP	Pediatric	Hypoplastic left heart syndrome
Tong 2024 [61]	China	Shanghai Children's Medical Center	2014–2021	cohort	23,000	3.1/2	training: 54.9% testing: 52.5%	Pediatric	Not specified

Abbreviations: EMR Electronic medical record, *PHIS* Pediatric Health Information System, *CHOP* Children's Hospital of Philadelphia, *ECHSA* European Congenital Heart Surgeons Association, *ECCDB* ECHSA Congenital Database, *SCMC* Shanghai Children's Medical Center, *GPPH* Guangdong's Provincial People's Hospital, *NP* not provided

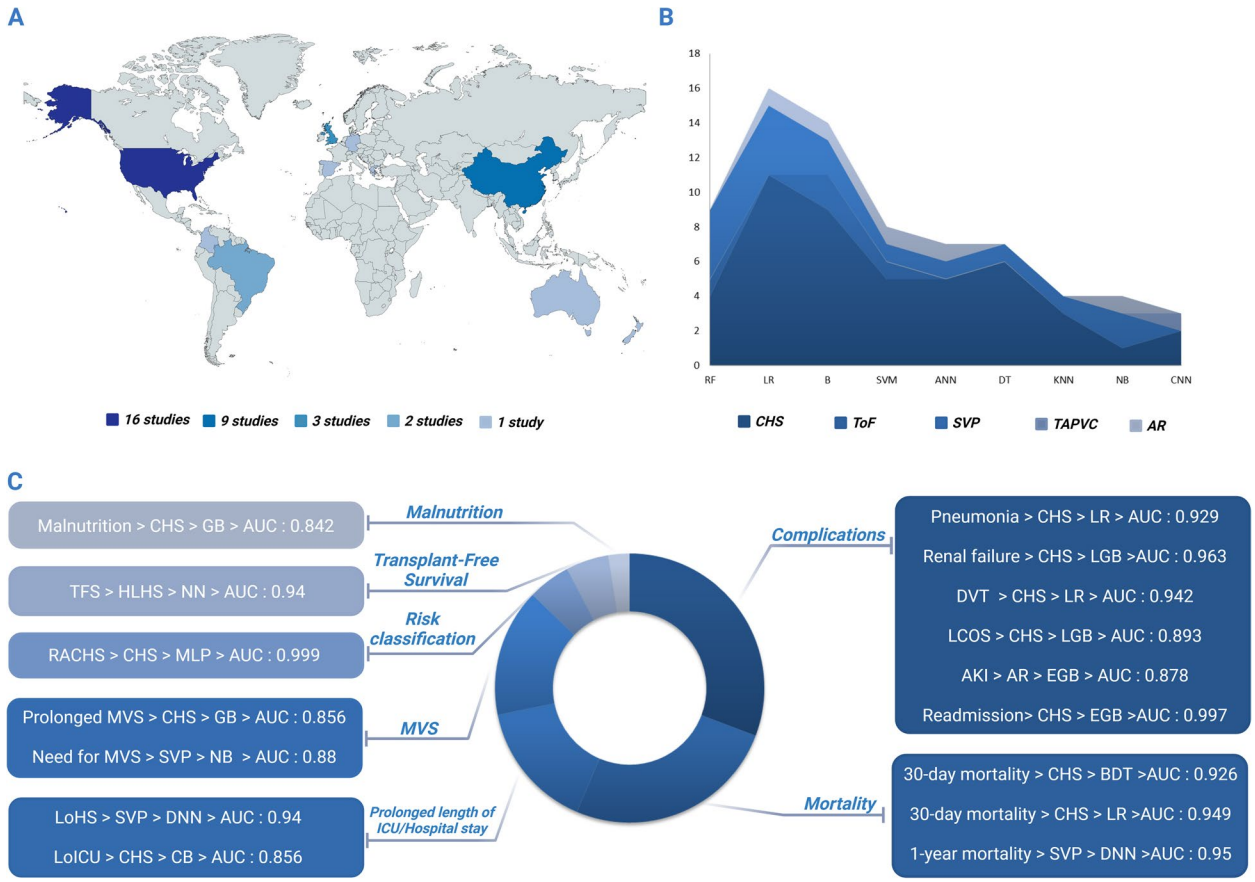

Fig. 2 Graphical abstract: **a** geographic distribution of the studies; **b** freaquancy of the machine learning algorthims for each congenital heart disease; **c** application and accuracy of machine learning algorithms in congenital heart surgery. Abbreviations: Boosting (B), Congential heart surgery (CHS), Single ventricle physiology (SVP), Aortic reconstruction (AR), Total anomalous pulmonary venous connection (TAPVC), Light gradient boosting (LGB), Extreme gradient boosting (EGB), CatBoost (CB), Transplant free survival (TFS), Mechanical ventilator support (MVS), Length of hospital stay (LoHS), Length of intensive care stay (LoICU), Acute kidney injury (AKI), artificial neural network (ANN), Neural network (NN), Deep neural network (DNN), Gradient boosting (GB), Multilayer Perceptron (MLP), Naive Bayes (NB), Bag Decision Trees (BDT), Deep venous thrombosis (DVT), Low cardiac output syndrome (LCOS), Support vector machine (SVM), Convolutional neural network (CNN), Random forest (RF), Decision tree (DT), K-nearest neighbor (KNN), Logistic regression (LR)

with the exception of one study by Sunthankar et al. [52] which predicted mortality between stages I and II of single ventricle surgery. Other mortality-predicting models focused on survival during the hospital stay and up to 1-year post-operation. The sensitivity of the models ranged from 0.1 in the LR model by Jalali et al. [46] to 0.92 in the random forest model by Chang Junior et al. [37]. The specificity of the predictions ranged from 0.542 in the gradient-boosted tree model by Sunthankar et al. [52] to 0.985 from the multilayer perception model by Chang Junior et al. [37]. The number of predictors used in each prediction model varied greatly yet some were of high predictive value throughout the different models; the type of procedures [27, 31, 40, 47, 52], age [27, 31, 33, 40, 48, 52], weight at the time of surgery [27, 31, 33, 40, 47, 53], arterial oxygen saturation [37, 48], and days since the previous hospitalization [40, 47] were all among these, with the type of procedure contributing more than 45% of predictive value in two instances [40, 47].

The type of operative procedure itself was a significant predictor in some models, with the Norwood procedure being the most predictive of mortality in some studies [27, 31, 47, 52]. This is in accordance with relevant literature stating the higher mortality rate of this procedure in comparison to most other procedures [9, 62]. Crowe et al.'s analysis clearly demonstrated that the Norwood procedure has the highest 30-day mortality rate among congenital heart surgeries, with a statistically significant difference compared to the second-highest rate, observed in interrupted aortic arch repair [27]. Sunthankar et al. developed a model aimed at predicting mortality after the Norwood procedure in single ventricle physiology and found the Norwood/Sano conduit as a significant predictor of mortality, yet deemed the hybrid Norwood method

Table 3 The characteristics of AI prediction models

First author and year	Algorithm/model	NO. of variables	Outcome(s)	Mode of validation	Missing data strategy	Best performing algorithm (Corresponding AUC)
Crowe 2013 [27]	LR	8	30-day mortality	Train/test	Record exclusion	LR (0.81)
Zapata-Impata 2015 [28]	PSO + KNN	Combination A: 33 Combination A: 83	RACHS classification	Train / Test	NP	PSO + KNN (NP)
Moein 2015 [29]	RF	1089	Post-operative poor outcomes: Morbidity (need for ECMO, pLOS, etc.) and Mortality	Train /Test / Validation (SwB)	NP	RF with 400 trees + clinical features (0.743)
Ruiz-Fernández 2015 [30]	MLP, SOM, RBF, DT	87	RACHS classification	10-CV	Single imputation	MLP (0.999)
Rogers 2017 [31]	LR	11	30-day mortality	5-CV + EV	Record exclusion	LR (0.86)
Jalali 2018 [32]	SVM using 3 ranking methods: (1) mutual information (2) mutual information modified with reliability index (3) mutual information with reliability index and considering mutual information of a set	HLHS: 14 non-HLHS: 11	Periventricular leukomalacia	K-fold cross validation	NP	SVM + 3rd ranking system (NP)
Luis Ahumadal 2018 [33]	NN, RF	26	One-Year Transplant-Free Survival (Norwood Procedure)	5-CV	Multiple imputation	NN (0.94)
Samad 2018 [34]	LSVM	24 (total): a: 6 b: 6 c: 5 d: 10	a) Major vs No DVSF b) Major or Minor vs No DVSF c) Major vs Minor vs No DVSF d) Major vs Minor or No DVSF	5-CV	NP	LSVM: a: 0.87 b: 0.82 c: 0.7 d: 0.77
Ruiz 2019 [35]	C-WIN + NB	34	Critical events: CPR, UETI, and ECMO in infants with SVP before second-stage surgery	10-CV	Single imputation	C-WIN + NB (0.88)
Cocomello 2020 [36]	LR	11	30-day mortality	EV	Record exclusion	LR (1st cohort = 0.72; 2nd cohort = 0.88)
Chang Junior 2020 [37]	MLP, RF, ET, SGB, ABC, BDT	MLP, BDT: 84; RF, SGB, ET, ABC: 42	In-hospital /30-day mortality	10-CV	NP	BDT (0. 926)
Huang 2020 [38]	LR, NB, RF, LDA, SVM, KNN	7	Mean pulmonary arterial pressure > 15 mmHg	Train / Test	NP	RF (0.79)

Table 3 (continued)

First author and year	Algorithm/model	NO. of variables	Outcome(s)	Mode of validation	Missing data strategy	Best performing algorithm (Corresponding AUC)
Bender 2021 [39]	SVM + Genetic Algorithm (the optimization technique)	53	Periventricular leukomalacia	Train / Test	NP	SVM (1)
Bertsimas 2021 [40]	LR, OCT, RF, GB	13	In-hospital /30-day mortality pMVST pLOS	Train / Test	NP	Mortality: GB (0.874) pMVST: GB (0.856) pLOS: RF (0. 821)
Faerber 2021 [41]	GB: two-stage with and without a MARS	without MARS: 56 with MARS: 21	Postoperative cardiac complications	10-CV	Multiple imputations	GB (0.71)
Rusin 2021 [42]	LR	7	Cardiorespiratory deterioration events[1]	Train / Test	NP	LR (0.958)
Ng 2022 [43]	A deep learning based perioperative parameter classifier composed of CNN + RBF + fusion strategy	NA	- Length of ICU stay (LICUS) - Perioperative complications (PC)	Train /Test / Validation(SwB)	NP	LICUS: All components (0.73) PC: All components (0.72)
Zeng 2021 [44]	XGBoost, LR	45	- Prediction of postoperative complications - Classification of postoperative complications	5-CV	Multivariate imputation	Prediction: XGBoost (0.839) Classification: XGBoost (0.85)
Thiriveedi 2021 [45]	XGBoost: STS model -Biomarker model -Clinical model	-STS model: NP -Biomarker model: 4 -Clinical model: NP	Readmission following 30 days post operation	NP	NP	XGBoost: clinical model (0.997)
Jalali 2021 [46]	LR, RF, DT, GB, DNN	25 (1-year mortality) 49 (pLOS)	- 1-year mortality/need for cardiac transplant - Prolonged length of hospital stay (pLOS)	5-CV	Multiple imputation	1-year mortality: DNN (0.95) pLOS: DNN (0.94)
Bertsimas 2022 [47]	OCT	12	- Mortality: In-hospital /30-day mortality - pMVST - pLOS	Train / Test	NP	OCT: Mortality: 0.872 OCT: pMVST: 0.814 OCT: pLOS: 0.813
Du 2022 [48]	XGBoost	59	In-hospital mortality	Train /Test / Validation(SwB)	NP	XGBoost (0.874)

Table 3 (continued)

First author and year	Algorithm/model	NO. of variables	Outcome(s)	Mode of validation	Missing data strategy	Best performing algorithm (Corresponding AUC)
Ekhomu 2022 [49]	GB	NP	- Postoperative peak RA strain - Postoperative systolic RA strain rate - Postoperative early diastolic RA strain rate - Postoperative RV global longitudinal strain	3-CV	Multiple imputations	GB (NP)
Shi 2022 [50]	LR, SVM, MLP, XGBoost, AB	15	Malnutrition, defined as underweight: weight below −2 z-scores	K-fold CV + EV	Categorical: mode Continuous: multiple imputations	XGBoost (0.842)
Pei 2022 [51]	Deep learning framework: 1. Segmentation and 3D modeling of LA and PV using V-net (CNN) 2. Computation of morphological features from LA and PV 3. Determination of Risk Factors 4. Risk Prediction Model by Morphological Features of LA and PV	29	Postoperative Pulmonary Vein Obstruction	3-CV + EV	NP	CNN (0.870)
Sunthankar 2023 [52]	LR, RF, XGBoost, GBDT, LightGBM	180	Interstage mortality between stage I and II surgery while at home	5-CV	Single imputation	Light gradient boosting machine (0.642)
Betts 2023 [53]	GBDT, ANN, LR	NP	30-day mortality	5-CV + EV	No missing data	Gradient boosting trees (0.87)
Zürn 2023 [54]	LR	5	30-day mortality	Leave-one-out CV + EV	Simple imputation	LR (0.9486)
Jiwani 2023 [55]	CNN	NP	In-hospital mortality	K-fold CV	NP	CNN (NP)
Kong 2023 [56]	LR, NB, XGBoost, SVM, LightGBM, MLP	16	Acute kidney injury	Train / Test / Validation (SwB)	NP	XGBoost (0.878)
Sarris 2024 [57]	Decision trees	12	In-hospital mortality, pMVST, pLOS	Train / Test	NP	DT (Mortality: 0.866, pMVST: 0.851, pLOS: 0.818)
Chang junior 2024 [58]	Catboost, RF, GB, NB, XGBoost, SVM, LightGBM. LR, DT, KNN, AB, LDA, ET, ridge, QDA	93	ICU length of stay	10-CV	No missing data	Catboost (0.8559)

Table 3 (continued)

First author and year	Algorithm/model	NO. of variables	Outcome(s)	Mode of validation	Missing data strategy	Best performing algorithm (Corresponding AUC)
Li 2024 [59]	LR, KNN	NP	Postoperative complications Mechanical ventilation duration	Train/ test	NP	LR + KNN (0.810)
Smith 2024 [60]	NP	45 to 195 (in different models)	Transplant-free survival	5-CV	Missing forest	NP
Tong 2024 [61]	LightGBM, LR, SVM, RF, CatBoost	39	LCOS, Pneumonia, Renal failure, Deep venous thrombosis	Train/ Test	Record exclusion	LCOS: LightGBM (0.893) Pneumonia: LR (0.929) Renal failure: LightGBM (0.963) DVT: LR (0.942)

Abbreviations: 1: defined as either a cardiac arrest requiring CPR (cardiac deterioration) or an unplanned intubation (respiratory deterioration)

ECMO Extracorporeal membrane oxygenation, PSO Particle swarm optimization, *KNN* k-nearest neighbors, *RF* Random Forest, *MLP* Multilayer Perceptron, *SOM* Self-organizing map, *RBF* Radial Basis Function, *DT* Decision Tree, *3-CV* threefold cross-validation, *5-CV* fivefold cross-validation, *10-CV* tenfold cross-validation, *EV* External validation, *LightGBM* Light gradient boosting machine, *GBDT* Gradient Boosting Decision Trees, *SVM* Support vector machine, *CNN* Convolutional neural network, *NN* Neural network, *LSVM* Linear support vector machine, *DVSF* Deterioration of ventricular size and function, *C-WIN* Cardiac intensive-care Warning INdex, *NB* Naive Bayes, *CPR* Cardiopulmonary resuscitation, *UETI* Unplanned Endotracheal Intubation, *SVP* Single-ventricle physiology, *ET* Extra Trees, *SGB* Stochastic Gradient Boosting, *ABC* Ada Boost Classification, *BDT* Bag Decision Trees, *LR* Logistic regression, *LDA* Linear discriminant analysis, *OCT* Optimal classification trees, *GB* Gradient boosting, *MARS* Multivariate adaptive regression spline model, *XGBoost* Extreme gradient boosting, *STS* Society of Thoracic Surgeons, *DNN* Deep neural networ, *AB* Adaptive boosting, *SwB* Split with bootstrapping, *QDA* Quadratic discriminant analysis, *DVT* Deep venous thrombosis, *LCOS* Low cardiac output syndrome, *pMVST* prolonged mechanical ventilatory support time, *pLOS* prolonged length of hospital stay, *RACHS* risk adjustment for congenital heart surgery

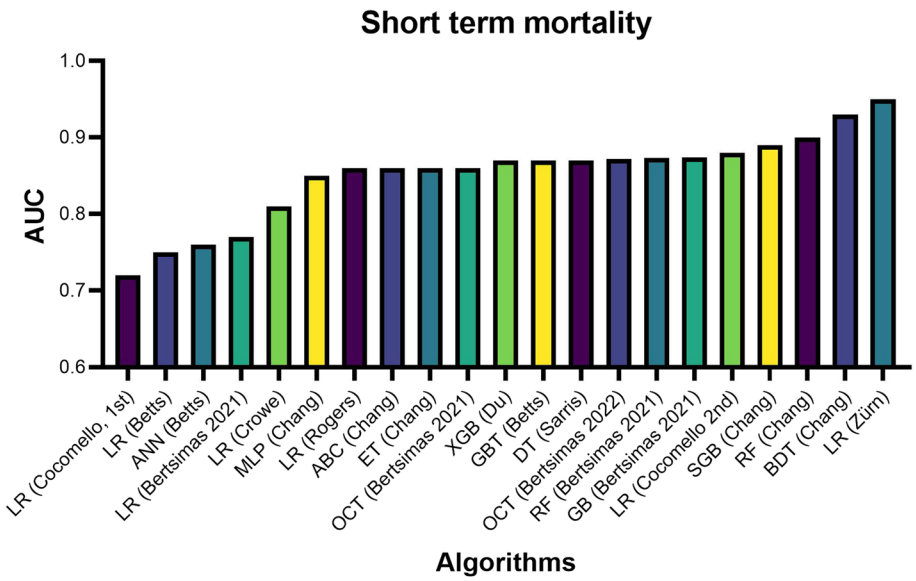

Fig. 3 AUCs of different models investigating short term mortality. Algorithm abbreviations: random forest (RF), Multilayer perception (MLP), Logistic regression (LR), Bagged decision trees (BDT), Extra trees (ET), Ada boost classification (ABC), Stochastic gradient boosting (SGB), Extreme gradient boosting (XGBoost), Optimal classification trees (OCT), Gradient boosting (GB), Adaptive boosting (ADA)

of less importance [52]. Another study also aimed to predict mortality after the Norwood procedure; however, it did not use the type of Norwood procedure as a predictor, yet still managed to obtain an accuracy of 88% for the prediction of mortality [33]. It should also be noted the extreme gradient boosting machine model developed by Du et al. _which included a large variety of procedures_ did not attribute a significant predictive value to the type of procedure [48]. Additionally, STS-EACTS scores _ estimates of the mortality risk associated with each type of congenital cardiac procedure_were also found to be highly predictive of mortality by some studies [53, 54]. Whether the type of procedure can be used to correctly predict mortality after congenital heart surgery requires more extensive investigations, yet it seems to be promising in the few instances where it has predictive value.

The application of age in the predictive models was rather heterogeneous, with some finding gestational age [52] and age at first extubation [52] as a significant predictor, while others found the age at the time of surgery to be an important predictor [27, 31, 40, 48, 53]. One study by Chang Junior et al., however, found a maximum of 2.7% predictive importance for age at the time of the procedure [37]. This factor similar to the previous one is also in accordance with the literature as a large study found increasing levels of age lower the likelihood of mortality [63]. The utilization of this predictor also shows promise.

Weight at the time of surgery was highly predictive in some instances [27, 31, 40, 47, 53], while in a model by Chang Junior et al. [37] showed rather diminished predictive value. This model however found height and BMI to be of high importance in their algorithms, suggesting that weight on its own may not be a reliable predictor and lacks generalizability. The value of weight as a predictor, much like the previous factors, is also evident in the literature [64].

The diagnostic group was also identified to be an important predictor of mortality by some studies [27, 31, 37]. Among the anomalies, univentricular heart status [27, 31] and by extension, hypoplastic left heart syndrome (HLHS) were found to be important predictors of mortality. Crowe et al. [27] classified patients with HLHS and pulmonary atresia with an intact ventricular septum as high risk groups for mortality, which was also corroborated by Rogers et al. [31]. Both studies also found that univentricular heart status is a separate indicator of 30-day mortality. In addition, Chang Jr. et. al. not only found that diagnostic groups are a significant predictor, they identified HLHS status as a separate significat predictor as well [37]. This finding is in alignment with the previous literature, which has often found univentricular heart anomalies to be the most fatal among congenital heart anomalis [65, 66].

Patient height at surgery [37], previous ICU admission [37], gender [33, 53], usage of cardiopulmonary bypass [27, 31], Down syndrome status [27], presense of congenital or acquired comorbidities [31], severity of illness [31], digoxin use at discharge [52], atrial shunt [48], and cardiopulmonary bypass time [52] were also notable

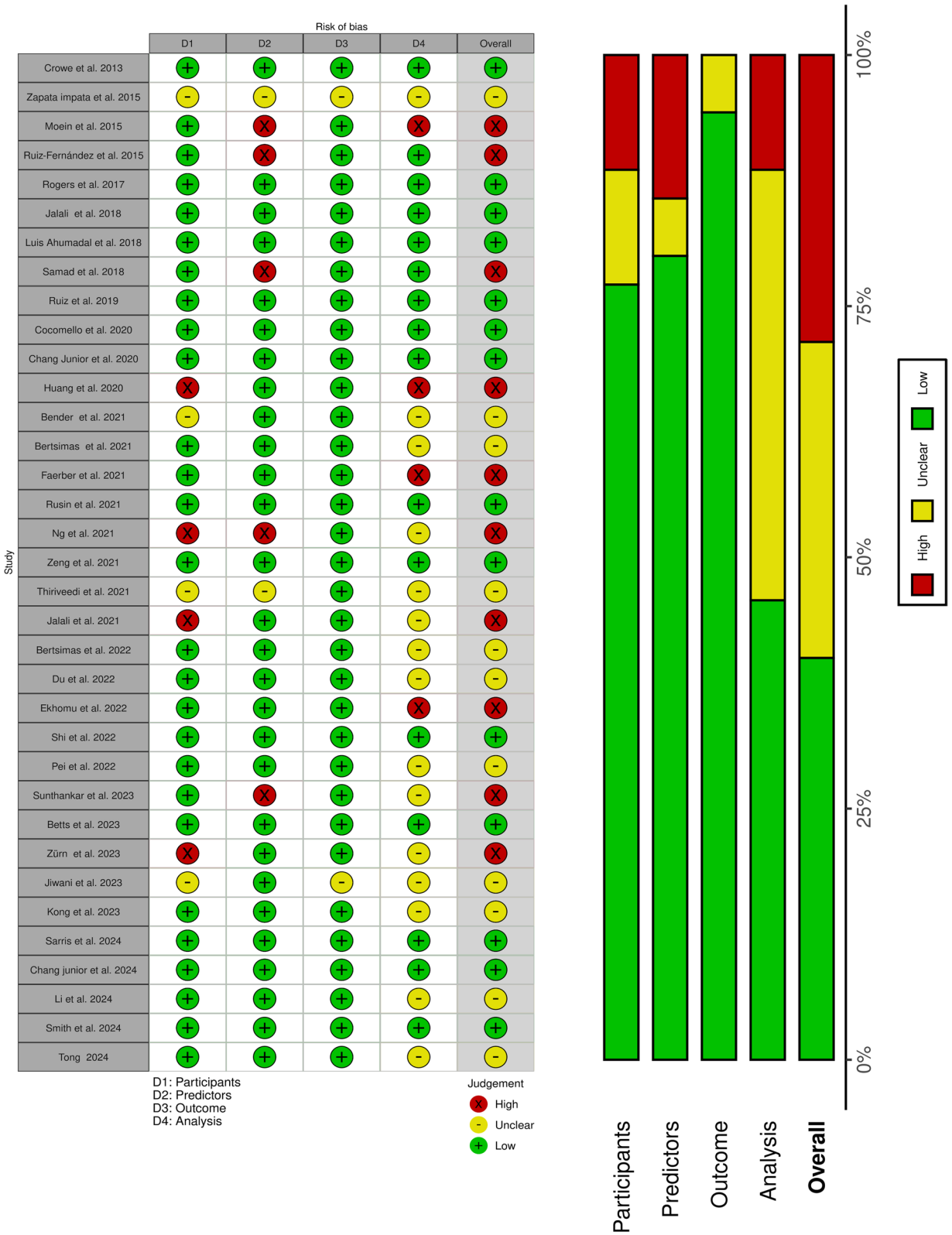

Fig. 4 Quality assessment of the studies using PROBAST. Domains: D1: Participants/ D2: Predictors/ D3: Outcome/ D4: analysis

predictors among the models for mortality prediction, yet these findings were only exclusive to one model or two models and were not found of considerable value in the other included models.

All in all, we observed an exceptional performance among the AI models predicting mortality in patients undergoing congenital heart surgeries; however, more studies need to be conducted in this area in order to perform quantitative synthesis and objectively measure their performance.

The area under the curve for studies predicting cardiac complications was more heterogeneous than mortality, with a range of 0.71 to 0.95. We observed that models tasked with predicting a broader spectrum of complications rather than a specific cardiac complication had a lower AUC. Furthermore, the sensitivity of predictions ranged from 0.34 in an LR model to 0.791 in an extreme gradient boosting model, both by Zeng et al. [44] Specificity of the models predicting complications was not reported. As for the predictors used in cardiac complications, birth weight [49], weight at surgery [44], gestational age [49], age at surgery [34, 49], cardio-pulmonary bypass time [41, 44], ECG variables [42], heart rate [35, 42], echocardiographic variables [34, 41, 49] and post-surgery oxygen pressure [41, 44, 49] were all important predictors. Echocardiographic variables, such as left ventricular ejection fraction [34], right ventricular global strain [41], and right pulmonary artery Z score [49], have been shown to correctly foretell cardiac complications [67] and due to the ubiquitous and simple nature of echocardiography, these variables can be of immense value in predicting cardiac complications.

Another intriguing aspect with regard to cardiac morbidity and mortality is the notable superiority of AI to traditional prediction models. Zeng et al. demonstrated that the model they produced using extreme gradient boosting achieved a higher AUC compared to univariable logistic regression models of Aristotle's Basic Complexity (ABC) score, STS-EACTS mortality and morbidity score, and RACHS-1 [44], indicating the potential superiority of an AI model to the tools currently available to clinicians. In addition, a systematic review assessing the ability of these risk stratification systems to predict morbidity and mortality found that the AUC of ABC ranges from 0.59 to 0.743, the AUC of RACHS-1 ranges from 0.68 to 0.782, and the AUC of STS-EACTS score from 0.732 to 0.8 [68]. Given the fact that AI models consistently demonstrate AUCs higher than the numbers stated, adopting them would likely assist surgeons in better risk stratification and decision-making. This is notable as risk stratification tools play a crucial part in reducing the morbidity and mortality of patients undergoing congenital heart surgery [68]. More investigations are needed to compare and validate AI prediction models against traditional ones. Furthermore, to implement these models clinically, traditional machine learning models should be optimized with as few features as possible in order to require lower computational power [69], and models should be trained and externally validated on larger sets of data to maximize the predictive ability before commercialization.

The area under the curve for length of stay predictive models was even more heterogeneous than the previous outcomes, ranging between 0.54 to 0.95. Sensitivity also ranged from 0.32 in an LR model to 0.91 in a deep neural network, both developed by Jalali et al. [46]. The specificity of the models predicting length of stay was not reported. Variables associated with increased length of stay were unfortunately not provided in detail by the included models, yet Bertsimas et al. [40] found that the type of procedure, days since previous admission, weight, and age were all significant predictors of prolonged length of hospital stay. Moreover, Chang Jr. et al. found that mechanical ventilation time, weight, previous ICU admission, vasoactive-inotropic score, and height were significant predictors of prolonged length of IC stay [58].

This pattern of AUC appears to show artificial intelligence predictive models are more geared toward predicting singular binary outcomes, yet newer machine learning algorithms perform better than logistic regression models in this regard. This is evident in the few instances they were compared in the included studies [40, 44, 46, 52].

While most of the included articles did not focus on a specific CHD, outcomes in patients with single ventricle physiology ($n=8$) and TOF ($n=4$) were the most investigated. The management of patients with single-ventricle physiology is an intricate process, as not only do single-ventricle defects vary greatly in anatomical subtypes, but surgical palliation is undertaken in multiple stages [70]. As such, precise prediction models for post-operative outcomes in these patients are invaluable. AI models can predict postoperative complications [35, 42] and 1-year mortality [33, 46] with great accuracy. These results could not be replicated when predicting interstage mortality_i.e. mortality occurring between the first and second stages of the Norwood procedure_ as the AUC of neither the mortality-predicting models of Sunthankar et al. nor Smith et al.'s survival-predicting model surpassed 70% [52, 60]. Among the studies in patients with TOF, the reports were more divergent with respect to the measured outcome, yet complications, mortality, and indices of heart function have been successfully predicted by AI models [29, 34, 41, 49].

Our systematic review has several limitations. First, by the nature of congenital heart defects, the patient populations used to develop the models are quite heterogeneous.

As a result, unless the models are specialized in one specific type of defect, very large sample sizes are needed to develop reliable AI models, a condition that was not met by some of the included studies. In addition, mainly due to the heterogeneous and incomplete reporting of the analyses conducted to develop the AI models, the quality of most of the included studies was either low or unclear. We suggest that future studies report the steps of model development thoroughly to avoid this issue. External validation _ i.e. evaluating the developed model on a separate original data set different than the one used to train and test the model_ was not used in most of the included studies in our review. Doing so is important to assess the realistic applicability of predictive models [71] and without this process, we cannot say how practical these models can be in the healthcare system, therefore more studies are needed to validate our findings. Another important point that was ignored by most of the included reports was first ensuring that the model was calibrated and second, reporting quantitative measures of calibration of the model in addition to plots. In addition, the high rate of heterogeneity among methods, outcomes, and participants did not allow us to perform a meta-analysis on any of the outcomes.

Conclusion

The implementation of artificial intelligence in congenital heart surgery has resulted in the creation of many predictive models that have the potential to change the landscape of clinical outcome prediction. When assessing the risk of postoperative morbidity and mortality, these AI-generated predictive models demonstrate superior performance compared to traditional risk prediction models. Their applicability, however, is not limited to the immediate risk of complications after surgery and encompasses long-term outcomes such as 1-year mortality and malnutrition. This systematic review provides a comprehensive report on such models and can serve as a bedrock to conduct future studies with more insight into the field.

Abbreviations

AI	Artificial intelligence
AUC	Area under the curve
CHD	Congenital heart disease
RACHS-1	Risk Adjustment in Congenital Heart Surgery-1
STS-EACTS	Society of Thoracic Surgery-European Association for Cardiothoracic Surgery Congenital Heart Surgery Mortality Categories
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PICOS	PICOS (population, intervention, comparator, outcome, and study design
PVL	Periventricular leukomalacia
EMR	Electronic medical records
TOF	Tetralogy of Fallot
LR	Logistic regression
STAT	The Society of Thoracic Surgeons-European Association for Cardio-Thoracic Surgery
ABC	Aristotle's Basic Complexity
ECMO	Extracorporeal membrane oxygenation
PSO	Particle swarm optimization
KNN	K-nearest neighbors
RF	Random Forest
MLP	Multilayer Perceptron
SOM	Self-organizing map
RBF	Radial Basis Function
DT	Decision Tree
3-CV	3-Fold cross-validation
5-CV	5-Fold cross-validation
10-CV	10-Fold cross-validation
EV	External validation
LightGBM	Light gradient boosting machine
GBDT	Gradient Boosting Decision Trees
XGBoost	Extreme gradient boosting
SVM	Support vector machine
CNN	Convolutional neural network
NN	Neural network
LSVM	Linear support vector machine
DVSF	Deterioration of ventricular size and function
C-WIN	Cardiac intensive-care Warning INdex
NB	Naive Bayes
CPR	Cardiopulmonary resuscitation
UETI	Unplanned Endotracheal Intubation
SVP	Single-ventricle physiology
ET	Extra Trees
SGB	Stochastic Gradient Boosting
ABC	Ada Boost Classification
BDT	Bag Decision Trees
LDA	Linear discriminant analysis
OCT	Optimal classification trees
GB	Gradient boosting
MARS	Multivariate adaptive regression spline model
XGBoost	Extreme gradient boosting
STS	Society of Thoracic Surgeons
DNN	Deep neural network
AB	Adaptive boosting
SwB	Split with bootstrapping
QDA	Quadratic discriminant analysis
DVT	Deep venous thrombosis
LCOS	Low cardiac output syndrome
B	Boosting
CHS	Congential heart surgery
SVP	Single ventricle physiology
AR	Aortic reconstruction
TAPVC	Total anomalous pulmonary venous connection
LGB	Light gradient boosting
EGB	Extreme gradient boosting
CB	CatBoost
TFS	Transplant free survival
MVS	Mechanical ventilator support
LoHS	Length of hospital stay
LoICU	Length of intensive care stay
AKI	Acute kidney injury
ANN	Artificial neural network
NN	Neural network
DNN	Deep neural network
GB	Gradient boosting
MLP	Multilayer Perceptron
NB	Naive Bayes
BDT	Bag Decision Trees
DVT	Deep venous thrombosis
LCOS	Low cardiac output syndrome
SVM	Support vector machine
CNN	Convolutional neural network
RF	Random forest
DT	Decision tree
KNN	K-nearest neighbor
LR	Logistic regression
pMVST	Prolonged mechanical ventilatory support time
pLOS	Prolonged length of hospital stay
HLHS	Hypoplastic left heart syndrome

Acknowledgements
None

Authors' contributions
IM: Conceptualization, investigation, data curation, methodology, and writing original draft; SRF: Investigation, data curation, formal analysis, and writing original draft; MH: Investigation, formal analysis, and writing original draft; MA: Conceptualization, investigation, project administration, writing-review and editing, and supervision; BH: Writing-review and editing; SZN: Supervision, writing-review and editing; PSN: Writing-review and editing.

Funding
None.

Data availability
The data utilized and/or analyzed in the present study are accessible from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Cardiovascular Surgery Research and Development Committee, Iran University of Medical Sciences (IUMS), PO box 14665-354, Tehran, Iran. [2]Student Research Committee, School of Medicine, Shahid Beheshti University of Medical Sciences, Teheran, Iran. [3]Department of Surgery, Surgery Research Center, School of Medicine, Rasool-E Akram Hospital, Iran University of Medical Sciences, Tehran, Iran. [4]Department of Cardiothoracic Surgery, Maastricht University Medical Centre, Maastricht, Netherlands. [5]Foundation Heart Team Academy, Maastricht, the Netherlands.

Received: 1 July 2024 Accepted: 11 November 2024
Published: 20 December 2024

References

1. van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J Am Coll Cardiol. 2011;58(21):2241–7.
2. Bouma BJ, Mulder BJ. Changing Landscape of Congenital Heart Disease. Circ Res. 2017;120(6):908–22.
3. Takach TJ, Ott DA. Congenital heart surgery in Houston. The early years. Tex Heart Inst J. 1997;24(3):233–7.
4. Su Z, Zou Z, Hay SI, Liu Y, Li S, Chen H, et al. Global, regional, and national time trends in mortality for congenital heart disease, 1990–2019: An age-period-cohort analysis for the Global Burden of Disease 2019 study. eClinicalMedicine. 2022;43:101249.
5. Pick E, Quirk J, Birkett J, Cullington D, Oliver J, Congiu S, et al. 85 Post operative complications in adult congenital heart disease. Heart. 2017;103(Suppl 5):A62–3.
6. Javed F, Aleysae NA, Al-Mahbosh AY, Zubani AA, Atash AM, Salem HB, et al. Complications After Surgical Repair of Congenital Heart Disease in Infants. An Experience From Tertiary Care Center. J Saudi Heart Assoc. 2021;33(4):271–8.
7. Jenkins KJ, Gauvreau K. Center-specific differences in mortality: preliminary analyses using the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. J Thorac Cardiovasc Surg. 2002;124(1):97–104.
8. Hörer J, Kasnar-Samprec J, Cleuziou J, Strbad M, Wottke M, Kaemmerer H, et al. Mortality Following Congenital Heart Surgery in Adults Can Be Predicted Accurately by Combining Expert-Based and Evidence-Based Pediatric Risk Scores. World J Pediatr Congenit Heart Surg. 2016;7(4):425–35.
9. O'Brien SM, Clarke DR, Jacobs JP, Jacobs ML, Lacour-Gayet FG, Pizarro C, et al. An empirically based tool for analyzing mortality associated with congenital heart surgery. J Thorac Cardiovasc Surg. 2009;138(5):1139–53.
10. Martínez-Alario J, Tuesta ID, Plasencia E, Santana M, Mora ML. Mortality Prediction in Cardiac Surgery Patients. Circulation. 1999;99(18):2378–82.
11. Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: Systematic review and meta-analysis. J Thorac Cardiovasc Surg. 2022;163(6):2075–87.e9.
12. Hunter DJ, Holmes C. Where Medical Statistics Meets Artificial Intelligence. N Engl J Med. 2023;389(13):1211–9.
13. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019;380(14):1347–58.
14. Volkova A, Ruggles KV. Predictive Metagenomic Analysis of Autoimmune Disease Identifies Robust Autoimmunity and Disease Specific Microbial Signatures. Front Microbiol. 2021;12:621310.
15. Ripley BD. Pattern recognition and neural networks: Cambridge: Cambridge university press; 2007.
16. Naruka V, Arjomandi Rad A, Subbiah Ponniah H, Francis J, Vardanyan R, Tasoudis P, et al. Machine learning and artificial intelligence in cardiac transplantation: A systematic review. Artif Organs. 2022;46(9):1741–53.
17. Dangeti P. Statistics for machine learning. Birmingham, UK: Packt Publishing Ltd; 2017.
18. Lantz B. Machine learning with R: expert techniques for predictive modeling. Birmingham, UK: Packt Publishing Ltd; 2019.
19. Amornsamankul S, Pimpunchat B, Triampo W, Charoenpong J, Nuttavut N. A Comparison of Machine Learning Algorithms and Their Applications. Int J Simul Syst Sci Technol. 2019:n. pag.
20. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021;2(3):160.
21. Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research (IJSR)[Internet]. 2020;9(1):381–6.
22. Penny-Dimri JC, Bergmeir C, Perry L, Hayes L, Bellomo R, Smith JA. Machine learning to predict adverse outcomes after cardiac surgery: A systematic review and meta-analysis. J Card Surg. 2022;37(11):3838–45.
23. Moher D, Liberati A, Tetzlaff J, Altman DG, Group* P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264–9.
24. Jovel J, Greiner R. An introduction to machine learning approaches for biomedical research. Front Med. 2021;8: 771607.
25. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51–8.
26. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev. 2022;18(2):e1230.
27. Crowe S, Brown KL, Pagel C, Muthialu N, Cunningham D, Gibbs J, et al. Development of a diagnosis- and procedure-based risk model for 30-day outcome after pediatric cardiac surgery. J Thorac Cardiovasc Surg. 2013;145(5):1270–8.
28. Zapata-Impata BS, Ruiz-Fernandez D, Monsalve-Torra A. Swarm intelligence applied to the risk evaluation for congenital heart surgery. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:214–7.
29. Moein S, Yan H, Das S, Hall M, Eghtesady P. Prediction of Systemic-to-Pulmonary Artery shunt surgery outcomes using administrative data. IEEE International Conference on Bioinformaticsand Biomedicine (BIBM). 2015:737–41.
30. Ruiz-Fernández D, et al. Aid decision algorithms to estimate the risk in congenital heart surgery. Comput Methods Programs Biomed. 2016;126:118–27.
31. Rogers L, Brown KL, Franklin RC, Ambler G, Anderson D, Barron DJ, et al. Improving Risk Adjustment for Mortality After Pediatric Cardiac Surgery: The UK PRAiS2 Model. Ann Thorac Surg. 2017;104(1):211–9.
32. Jalali A, Simpao AF, Gálvez JA, Licht DJ, Nataraj C. Prediction of Periventricular Leukomalacia in Neonates after Cardiac Surgery Using Machine Learning Algorithms. J Med Syst. 2018;42(10):177.
33. Luis Ahumadal M, Peck J, Guerra J, Do N, Gupta M, Ghazarian S, et al. Prediction of One-Year Transplant-Free Survival after Norwood Procedure

Based on the Pre-Operative Data. Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:3995–8.
34. Samad MD, Wehner GJ, Arbabshirani MR, Jing L, Powell AJ, Geva T, et al. Predicting deterioration of ventricular function in patients with repaired tetralogy of Fallot using machine learning. Eur Heart J Cardiovasc Imaging. 2018;19(7):730–8.
35. Ruiz VM, Saenz L, Lopez-Magallon A, Shields A, Ogoe HA, Suresh S, et al. Early prediction of critical events for infants with single-ventricle physiology in critical care using routinely collected data. J Thorac Cardiovasc Surg. 2019;158(1):234–43.e3.
36. Cocomello L, Caputo M, Cornish R, Lawlor D. External validation of the improving partial risk adjustment in surgery (PRAIS-2) model for 30-day mortality after paediatric cardiac surgery. BMJ Open. 2020;10(11):e039236.
37. Chang Junior J, Binuesa F, Caneo LF, Turquetto ALR, Arita E, Barbosa AC, et al. Improving preoperative risk-of-death prediction in surgery congenital heart defects using artificial intelligence model: A pilot study. PLoS ONE. 2020;15(9):e0238199.
38. Huang L, et al. Prediction of pulmonary pressure after Glenn shunts by computed tomography–based machine learning models. Eur Radiol. 2020;30(3):1369–1377.
39. Bender D, Licht DJ, Nataraj C. A Novel Embedded Feature Selection and Dimensionality Reduction Method for an SVM Type Classifier to Predict Periventricular Leukomalacia (PVL) in Neonates. Appl Sci (Basel). 2021;11(23):11156.
40. Bertsimas D, Zhuo D, Dunn J, Levine J, Zuccarelli E, Smyrnakis N, et al. Adverse Outcomes Prediction for Congenital Heart Surgery: A Machine Learning Approach. World J Pediatr Congenit Heart Surg. 2021;12(4):453–60.
41. Faerber JA, Huang J, Zhang X, Song L, DeCost G, Mascio CE, Ravishankar C, O'Byrne ML, Naim MY, Kawut SM, Goldmuntz E, Mercer-Rosa L. Identifying Risk Factors for Complicated Post-operative Course in Tetralogy of Fallot Using a Machine Learning Approach. Front Cardiovasc Med. 2021;8:685855.
42. Rusin CG, Acosta SI, Vu EL, Ahmed M, Brady KM, Penny DJ. Automated Prediction of Cardiorespiratory Deterioration in Patients With Single Ventricle. J Am Coll Cardiol. 2021;77(25):3184–92.
43. Ng WWY, Liang H, Peng Q, Zhong C, Dong X, Huang Z, et al. An automatic framework for perioperative risks classification from retinal images of complex congenital heart disease patients. Intl J Mach Learn Cybern. 2022;13(2):471–83.
44. Zeng X, Hu Y, Shu L, Li J, Duan H, Shu Q, et al. Explainable machine-learning predictions for complications after pediatric congenital heart surgery. Sci Rep. 2021;11(1):17244.
45. Thiriveedi LS, Everett AD, Parker D, Stabler M, Brown JR. Abstract 10834: Novel Biomarkers and Machine Learning Models for the Prediction of 30-Day Readmission Following Pediatric Congenital Heart Surgery. Circulation. 2021;144(Suppl_1):A10834-A.
46. Jalali A, Lonsdale H, Do N, Peck J, Gupta M, Kutty S, Ghazarian SR, Jacobs JP, Rehman M, Ahumada LM. Deep Learning for Improved Risk Prediction in Surgical Outcomes. Sci Rep. 2020;10(1):9289.
47. Bertsimas D, Zhuo D, Levine J, Dunn J, Tobota Z, Maruszewski B, et al. Benchmarking in Congenital Heart Surgery Using Machine Learning-Derived Optimal Classification Trees. World J Pediatr Congenit Heart Surg. 2022;13(1):23–35.
48. Du X, Wang H, Wang S, He Y, Zheng J, Zhang H, Hao Z, Chen Y, Xu Z, Lu Z. Machine Learning Model for Predicting Risk of In-Hospital Mortality after Surgery in Congenital Heart Disease Patients. Rev Cardiovasc Med. 2022;23(11):376.
49. Ekhomu O, Faerber JA, Wang Y, Huang J, Mai AD, DiLorenzo MP, et al. Right atrial function early after tetralogy of Fallot repair. Int J Card Imaging. 2022;38(9):1961–72.
50. Shi H, Yang D, Tang K, Hu C, Li L, Zhang L, et al. Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease. Clin Nutr. 2022;41(1):202–10.
51. Pei Y, et al. Building a Risk Prediction Model for Postoperative Pulmonary Vein Obstruction via Quantitative Analysis of CTA Images IEEE J Biomed Health Inform. 2022;26(7):3127–3138.
52. Sunthankar SD, Zhao J, Wei WQ, Hill GD, Parra DA, Kohl K, et al. Machine Learning to Predict Interstage Mortality Following Single Ventricle Palliation: A NPC-QIC Database Analysis. Pediatr Cardiol. 2023;44(6):1242–50.
53. Betts KS, Marathe SP, Chai K, Konstantinov I, Iyengar A, Suna J; ANZCORS Collaborative; Venugopal P, Alphonso N. A machine learning approach to predicting 30-day mortality following paediatric cardiac surgery: findings from the Australia New Zealand Congenital Outcomes Registry for Surgery (ANZCORS). Eur J Cardiothorac Surg. 2023;64(2):ezad160.
54. Zürn C, Hübner D, Ziesenitz VC, Höhn R, Schuler L, Schlange T, Gorenflo M, Kari FA, Kroll J, Loukanov T, Klemm R, Stiller B. Model-driven survival prediction after congenital heart surgery. Interdiscip Cardiovasc Thorac Surg. 2023;37(3):ivad089.
55. Jiwani N, Gupta K, Sarveshwaran V, Ravi V. Deep Learning Model to Predict In-hospital Mortality ofNewborns during Congenital Heart Disease Surgery. Open Bioinform J. 2023;16:e187503622306220
56. Kong X, Zhao L, Pan Z, Li H, Wei G, Wang Q. Acute renal injury after aortic arch reconstruction with cardiopulmonary bypass for children: prediction models by machine learning of a retrospective cohort study. Eur J Med Res. 2023;28(1):499.
57. Sarris GE, Zhuo D, Mingardi L, Dunn J, Levine J, Tobota Z, Maruszewski B, Fragata J, Bertsimas D. Congenital Heart Surgery Machine Learning-Derived In-Depth Benchmarking Tool. Ann Thorac Surg. 2024;118(1):199–206.
58. Chang Junior J, Caneo LF, Turquetto ALR, Amato LP, Arita ECTC, Fernandes AMDS, Trindade EM, Jatene FB, Dossou PE, Jatene MB. Predictors of in-ICU length of stay among congenital heart defect patients using artificial intelligence model: A pilot study. Heliyon. 2024;10(4):e25406.
59. Li H, Zhou M, Sun Y, Yang J, Zeng X, Qiu Y, Xia Y, Zheng Z, Yu J, Feng Y, Shi Z, Huang T, Tan L, Lin R, Li J, Fan X, Ye J, Duan H, Shi S, Shu Q. A Patient Similarity Network (CHDmap) to Predict Outcomes After Congenital Heart Surgery: Development and Validation Study. JMIR Med Inform. 2024;12:e49138.
60. Smith AH, Gray GM, Ashfaq A, Asante-Korang A, Rehman MA, Ahumada LM. Using machine learning to predict five-year transplant-free survival among infants with hypoplastic left heart syndrome. Sci Rep. 2024;14(1):4512.
61. Tong C, et al. Machine learning prediction model of major adverse outcomes after pediatric congenital heart surgery: a retrospective cohort study. Int J Surg 2024;110(4):2207–2216.
62. Mascio CE, Irons ML, Ittenbach RF, Gaynor JW, Fuller SM, Kaplinski M, et al. Thirty years and 1663 consecutive Norwood procedures: Has survival plateaued? J Thorac Cardiovasc Surg. 2019;158(1):220–9.
63. Yeh S-J, Chen H-C, Lu C-W, Wang J-K, Huang L-M, Huang S-C, et al. National database study of survival of pediatric congenital heart disease patients in Taiwan. J Formos Med Assoc. 2015;114(2):159–63.
64. Anderson BR, Blancha Eckels VL, Crook S, Duchon JM, Kalfa D, Bacha EA, et al. The Risks of Being Tiny: The Added Risk of Low Weight for Neonates Undergoing Congenital Heart Surgery. Pediatr Cardiol. 2020;41(8):1623–31.
65. Lynn MM, Salemi JL, Kostelyna SP, Morris SA, Tejtel SKS, Lopez KN. Lesion-Specific Congenital Heart Disease Mortality Trends in Children: 1999 to 2017. Pediatrics. 2022;150(4):e2022056294.
66. Çaylan N, Yalçin SS, Tezel B, Üner O, Aydin Ş, Kara F. Investigation of infant deaths associated with critical congenital heart diseases; 2018–2021, Türkiye. BMC Public Health. 2024;24(1):441.
67. Egbe AC, Miranda WR, Pellikka PA, DeSimone CV, Connolly HM. Prevalence and Prognostic Implications of Left Ventricular Systolic Dysfunction in Adults With Congenital Heart Disease. J Am Coll Cardiol. 2022;79(14):1356–65.
68. Fakhri D, Damayanti NMAS, Nurhanif M. Comparison of risk stratification scoring system as a predictorof mortality and morbidity in congenital heart disease patients requiring surgery. Ann Pediatr Cardiol.2023;16(5):349–53.
69. Staartjes VE, Kernbach JM, Stumpo V, van Niftrik CHB, Serra C, Regli L. Foundations of Feature Selection in Clinical Prediction Modeling. Acta Neurochir Suppl. 2022;134:51–7.
70. Walker SG, Stuth EA. Single-ventricle physiology: perioperative implications. Semin Pediatr Surg. 2004;13(3):188–202.
71. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2020;14(1):49–58.

Publisher's Note