Journal of the American Medical Informatics Association, 30(1), 2023, 94–102
https://doi.org/10.1093/jamia/ocac202
Advance Access Publication Date: 26 October 2022
Research and Applications

Research and Applications

A time-aware attention model for prediction of acute kidney injury after pediatric cardiac surgery

Xian Zeng[1,2], Shanshan Shi[3], Yuhan Sun[2], Yuqing Feng[1], Linhua Tan[3], Ru Lin[3,4], Jianhua Li[4], Huilong Duan[2], Qiang Shu[4], and Haomin Li[1]

[1]Clinical Data Center, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [2]The College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China, [3]CICU, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, and [4]Cardiac Surgery, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

Corresponding Author: Haomin Li, No. 3333 Binsheng Road, Hangzhou, Zhejiang 310052, PR China; hmli@zju.edu.cn

Received 12 July 2022; Revised 16 September 2022; Editorial Decision 10 October 2022; Accepted 12 October 2022

ABSTRACT

Objective: Acute kidney injury (AKI) is a common complication after pediatric cardiac surgery, and the early detection of AKI may allow for timely preventive or therapeutic measures. However, current AKI prediction researches pay less attention to time information among time-series clinical data and model building strategies that meet complex clinical application scenario. This study aims to develop and validate a model for predicting postoperative AKI that operates sequentially over individual time-series clinical data.

Materials and Methods: A retrospective cohort of 3386 pediatric patients extracted from PIC database was used for training, calibrating, and testing purposes. A time-aware deep learning model was developed and evaluated from 3 clinical perspectives that use different data collection windows and prediction windows to answer different AKI prediction questions encountered in clinical practice. We compared our model with existing state-of-the-art models from 3 clinical perspectives using the area under the receiver operating characteristic curve (ROC AUC) and the area under the precision-recall curve (PR AUC).

Results: Our proposed model significantly outperformed the existing state-of-the-art models with an improved average performance for any AKI prediction from the 3 evaluation perspectives. This model predicted 91% of all AKI episodes using data collected at 24 h after surgery, resulting in a ROC AUC of 0.908 and a PR AUC of 0.898. On average, our model predicted 83% of all AKI episodes that occurred within the different time windows in the 3 evaluation perspectives. The calibration performance of the proposed model was substantially higher than the existing state-of-the-art models.

Conclusions: This study showed that a deep learning model can accurately predict postoperative AKI using perioperative time-series data. It has the potential to be integrated into real-time clinical decision support systems to support postoperative care planning.

Key words: pediatric cardiac surgery, acute kidney injury, prediction model, deep machine learning, multi-perspectives evaluation

INTRODUCTION

Despite great progress achieved in the surgical management of congenital heart disease, there is still considerable associated risk of death and complications after surgery.[1] Acute kidney injury (AKI) is a common postoperative complication after pediatric cardiac surgery with a reported incidence of 15% to 64%, and is associated with increased early and long-term mortality, as well as prolonged hospital stay and higher medical costs.[2–6] Patients who survive an episode of AKI are also

© The Author(s) 2022. Published by Oxford University Press on behalf of the American Medical Informatics Association.
All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

at increased risk for developing long-term adverse outcomes, such as chronic kidney disease and end-stage renal disease.[7] It is widely accepted that glomerular filtration rate (GFR) is the most useful overall index of kidney function in health and disease, but direct GFR measurement is difficult.[8] In clinical practice, changes in serum creatinine (SCr) and urine output are used to assess changes in GFR.[9] Hence, there is a broad consensus that changes in SCr or urine output or estimated GFR (eGFR) form the basis of all diagnostic criteria for AKI such as the pediatric modified Risk, Injury, Failure, Loss, and End-stage renal disease (pRIFLE),[10] the International Kidney Disease Improving Global Outcomes (KDIGO),[9] and the pediatric reference change value optimized for AKI in children (pROCK).[11] However, the increases in SCr lag behind the actual renal injury by a considerable time period.[12] While there are often no specific treatments to reverse AKI after it has developed, an early identification of high-risk patients allows appropriate allocation of limited clinical resources to enhance clinical monitoring coupled with timely interventions.[13] Furthermore, pairing early detection and risk stratification of AKI has been shown to decrease AKI incidence and severity and drive improvements in AKI-related clinical outcomes in the inpatient setting.[14–17] Therefore, early prediction of patients at high risk of developing AKI is clinically significant.[13]

Promising recent studies[18,19] have suggested that the incorporation of machine learning may enable risk stratification, early prediction, and management of AKI. Tseng et al[20] developed and internally validated a AKI prediction model in patients undergoing cardiac surgery and achieved the greatest ROC AUC of 0.84. Dong et al[21] developed a machine learning model that was successful in predicting AKI in pediatric critical care patients prior to the detection of AKI by conventional criteria, with a median lead-time of 30 h. Recently, deep-learning methods have achieved improved performances over traditional models in the health care domain.[22] Recurrent neural networks (RNNs) are commonly used network architectures to model longitudinal patient data.[23,24] Rank et al[25] applied an RNN for the real-time prediction of AKI within the first 7 postoperative days following cardiothoracic surgery and showed that their model could predict 77.4% of all episodes of AKI. However, it is infeasible to interpret the outputs of models incorporating RNNs. RETAIN[26] and Dipole[27] introduced attention mechanisms that mimics cognitive attention to the healthcare domain and interpreted the models' output risks based on the learned attention weights for previous visits.

Most existing approaches[19,20,25] aim to develop an early prediction model of AKI for adult patients. In pediatric CHD patients, whose physiology differs greatly with age. Moreover, several examples of sequential AKI risk models[20,28] have focused on predictions across a short time horizon, which leaves little time for clinical assessment. The Acute Dialysis Quality Initiative (ADQI) group recommended developing models for the early prediction of AKI between 48 and 72 h before the diagnosis.[29] In addition, existing models relied on RNN-based models for longitudinal patient data modeling and implicitly assumed the stationary progression during each time period. This assumption does not leverage time information for risk prediction in a reasonable way. Furthermore, the most adopted model development and evaluation is by using the specific events (such as AKI onset) as the anchor point for data extraction and time window definition. However, the exact AKI onset time is unknown in real-world clinical practice, and AKI should be predicted continually during the entire inpatient stay.[29]

To tackle the above challenges, we proposed a time-aware attention-based RNN that allows early predictions of AKI following pediatric cardiac surgery. We designed a time-aware RNN to embed time information and a time-aware key-query attention mechanism to identify the key time steps among patients' historical events in their electronic health record data in the proposed model. Hence, the proposed model can not only operate sequentially over individual electronic health records but also can utilize time information among the patients' historical events. In an effort to conduct a more comprehensive assessment of our proposed model, we explored 3 evaluation perspectives to answer different AKI prediction questions: (1) predicting AKI before the onset; (2) continuously predicting after surgery; and (3) predicting AKI within various prediction windows at any random point after surgery. The 3 evaluation perspectives imply different data collection and prediction windows, and these can offer a comprehensive analysis of AKI prediction and provide useful guidance for future predictive modeling for clinical practice.

MATERIALS AND METHODS

Study design and population

This study used the Pediatric Intensive Care (PIC) database,[30] which contains deidentified clinical data of pediatric patients admitted to ICU of the Children's Hospital, Zhejiang University School of Medicine. In particular, the dataset collects over 15 000 admissions from 13 000 unique pediatric patients. Patients who underwent congenital heart surgery (5123 admissions/5091 patients) were included. We excluded any patients who did not have enough SCr measurements (<2 times) for determining AKI and patients who underwent surgery without cardiopulmonary bypass (CPB) or who died during the surgery (1737 admissions/1711 patients), as shown in Figure 1 and Supplementary Figure S1. Finally, a total of 3386 pediatric patients (no patients had end-stage renal disease before surgery) were included in this study. The patient characteristics were shown in Table 1.

Institutional Review Board approval from the Children's Hospital, Zhejiang University School of Medicine (2018_IRB_078, Title: Predictive and prognostic analysis of postoperative complications in pediatric cardiac surgery, approval date: 2018/09/19) was obtained prior to the commencement of this study. All procedures were followed in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975. In this study, the reporting of the development and validation of this prediction model widely follows the guidelines of the TRIPOD statement.[31]

Data collection and preprocessing

For each included patient, we extracted structured clinical variables, including demographic information, vital signs, and laboratory tests. Supplementary Table S1 gives an overview of all the considered input features. They can be grouped into static features (eg, patients and surgery characteristics) that do not change over the observation period and the frequently measured time-series features (eg, intraoperative vital signs, lab values, and blood gas values). The missing rate of raw input features is listed in Supplementary Table S2, and the missing rate for features not listed in this table is 0. The sampling frequency of raw time-series features and SCr recorded during hospitalization (except for intraoperative vital signs) is listed in Supplementary Table S3. In short, the missing rate is very low and the data density is generally consistent and tight. The average sampling frequency of SCr in all patients is 1.5 times per day, and more than 75% of patients sampled 1.05 times per day.

All numerical features were scaled as follows:

$$X_{\text{scaled}} = \frac{X - \mu(X_{\text{train}})}{\sigma(X_{\text{train}})}$$

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

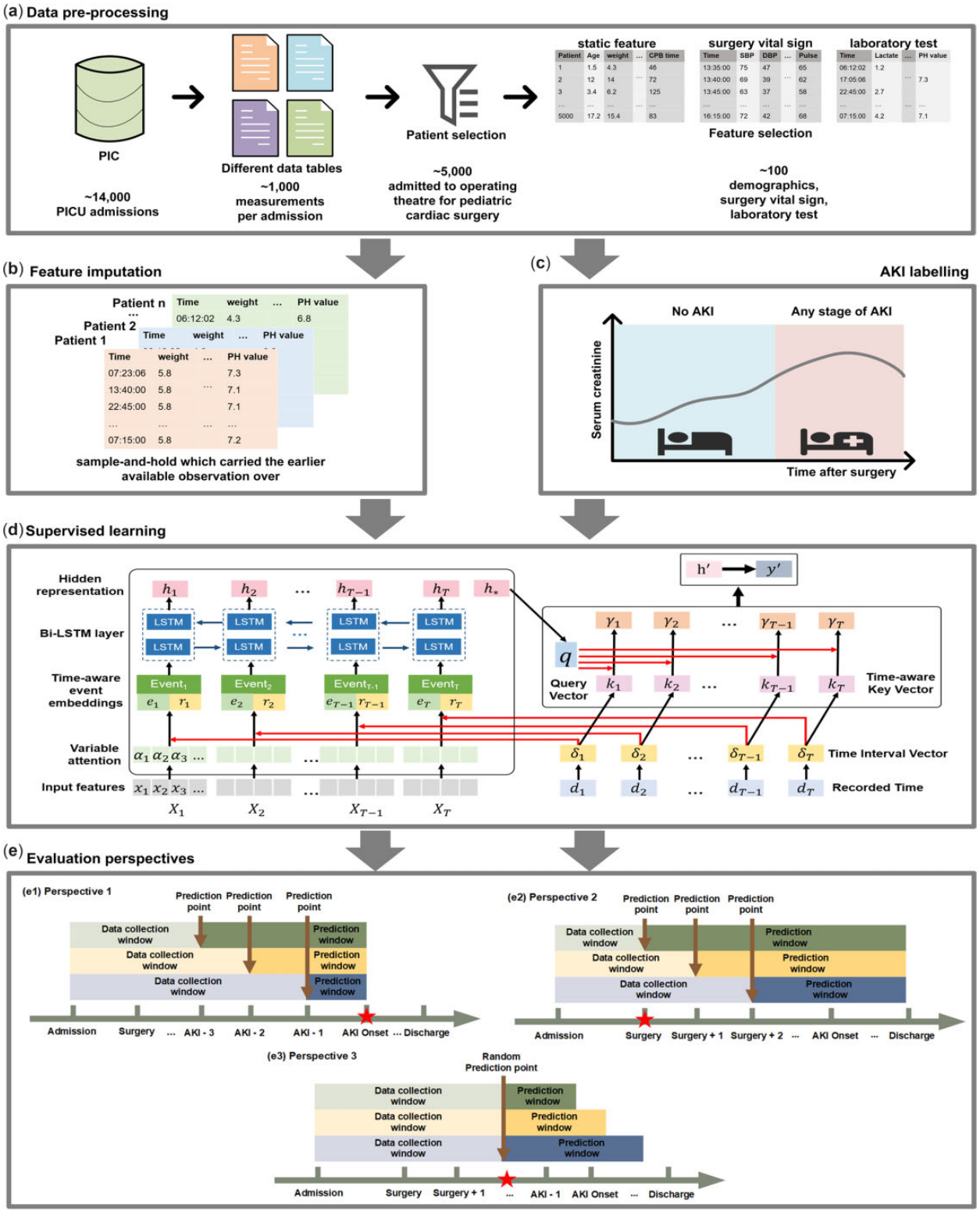

Figure 1. Overview of research. (a) Data preprocessing. (b) Feature imputation. (c) AKI labeling. (d) Supervised time-aware learning model. (e) Three evaluation perspectives designed to answer different AKI prediction questions encountered in clinical practice.

where $\mu(X_{\text{train}})$ denotes the mean and $\sigma(X_{\text{train}})$ is the standard deviation of feature X in the training set. We performed one-hot coding on the categorical variables (eg, diagnosis and procedure codes) to convert them into binary representations. The data were resampled to 5 min during surgery and 8 h during the hospital stay using sample-and-hold imputation which means when any measurements were missing for a certain time interval, we carried the earlier available observation over. AKI is determined by either a 1.5-fold

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

Table 1. Patient's characteristics

Variables	All	AKI	No-AKI	*P*
Patient population	3386	331	3055	
Demographics				
Age (months)	11.9 [4.5–28.9]	5.6 [1.9–15.3]	12.6 [5.1,30.6]	*
Male gender	1672 (49.4%)	173 (52.3%)	1499 (49.1%)	
Height (cm)	73.0 [62.0–91.0]	64.0 [55.0–78.0]	74.0 [63.0–92.0]	*
Weight (kg)	8.6 [6.8–12.5]	6.3 [4.3–9.6]	9.0 [6.0–12.9]	*
Surgery procedure				
VSD repair, patch	1530 (45.2%)	135 (40.8%)	1395 (45.7%)	
ASD repair, patch	935 (27.6%)	68 (20.5%)	867 (28.4%)	*
PDA closure	853 (25.2%)	140 (42.3%)	713 (23.3%)	*
PFO, primary closure	799 (23.6%)	80 (24.2%)	719 (23.5%)	
ASD repair, primary closure	736 (21.7%)	91 (27.5%)	645 (21.1%)	*
VSD repair, primary closure	375 (11.1%)	28 (8.5%)	347 (11.4%)	
Valvuloplasty, tricuspid	143 (4.2%)	18 (5.4%)	125 (4.1%)	
Valvuloplasty, mitral	137 (4.0%)	18 (5.4%)	119 (3.9%)	
TAPVC repair	89 (2.6%)	19 (5.7%)	70 (2.3%)	*
TOF repair, ventriculotomy, transannular patch	110 (3.2%)	23 (6.9%)	87 (2.8%)	*
Preoperative condition				
Preoperative length of stay (days)	4.0 [2.0–7.0]	7.0 [4.0–11.0]	4.0 [2.0–7.0]	*
Previous cardiac surgery	54 (1.6%)	12 (3.6%)	42 (1.4%)	*
Noncardiac malformation	207 (6.1%)	31 (9.4%)	176 (5.8%)	*
Other preoperative risk factors	569 (16.8%)	86 (26.0%)	483 (15.8%)	*
Intraoperative variables				
Elective	3341 (98.7%)	314 (94.9%)	3027 (99.1%)	*
Operation time (min)	125.0	148.0	123.0	*
Cardiopulmonary bypass time (min)	[108.0–156.0]	[119.0–207.0]	[107.0–150.5]	*
Cross clamp time (min)	60.0 [48.0–82.0]	85.0 [59.0–133.5]	59.0 [47.0–78.0]	*
Mechanical ventilation time (h)	40.0 [28.0–55.0]	53.0 [38.5–91.0]	39.0 [28.0–53.0]	*
pRBC transfusion during surgery (units)	7.0 [4.0–22.0]	24.0 [16.0–90.0]	6.0 [4.0–21.0]	*
FFP transfusion during surgery (mL)	1.0 [1.0–2.0]	2.0 [1.0–3.0]	1.0 [1.0–1.5]	
Autologous blood transfusion during surgery (mL)	150.0 [130.0–250.0] 125.0 [120.0–125.0]	260.0 [140.0–410.0] 120.0 [0.0–125.0]	150.0 [130.0–240.0] 125.0 [120.0–125.0]	*
Postoperative hospital stay	10.0 [7.8–13.9]	13.1 [9.7–19.9]	9.9 [7.7–13.8]	*
Requiring dialysis at the time of discharge	2 (0.1%)	2 (0.6%)	0 (0.0%)	*
Requiring temporary dialysis	15 (0.9%)	15 (4.5%)	0 (0.0%)	*
Mortality	45 (1.3%)	22 (6.6%)	23 (0.8%)	*

Abbreviations: AKI: acute kidney injury: ASD: atrial septal defect; FFP: fresh frozen plasma; PDA: patent ductus arteriosus; PFO: patent foramen ovale; pRBC: packed red blood cell; TAPVC: total anomalous pulmonary venous connection; TOF: tetralogy of Fallot; VSD: ventricular septal defect.

*Statistically significant difference $P < .05$.

increase from baseline SCr within 7 days or an absolute increase of 26.5 μmol/L in SCr within 48 h following the guideline of KDIGO. Baseline SCr was defined as the last SCr value before surgery. If there were no values for SCr available during hospitalization before surgery, we used the first postoperative value.

Models for predicting AKI

At each prediction time point, input features were provided to our proposed model, the output of which is a probability of AKI occurring within the prediction windows. Figure 1d gives a schematic view of our model, which makes predictions by first utilizing long short-term memory (LSTM) to learn a hidden state by considering both the input feature values and the time intervals and then uses key-query attention to model the input values with the time changes by embedding the hidden state into a "query" vector and embedding the time interval into a "key" vector, similar to position encodings in the transformer structure.[32] The source code of this model is available at https://github.com/Healthink/AKIPrediction. Hyperparameters of model training were show in Supplementary Table S4.

To understand how single features relate to the model output, we used Shapley additive explanation (SHAP) values,[33] which estimate the contribution of each feature to the overall model predictions. To learn the significance of each time interval during the risk prediction, we put the key and query vectors together and calculated the attention scores. The entire model can be trained end-to-end, that is, the parameters can be learned jointly without pretraining any parts of the model.

Experiment design

We evaluated our proposed model by comparing it with the following models.

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

- *The traditional methods:* support vector machine (SVM) and logistic regression (LR).
- *Plain RNNs*: LSTM and gate recurrent units (GRUs).
  Attention-based models: Dipole[27] and RETAIN.[26]

In an effort to conduct a more comprehensive assessment of our proposed AKI prediction model, we explored the following 3 clinical prediction perspectives as shown in Figure 1e (detailed description of data collection, prediction point and prediction windows of 3 perspectives is shown in Supplementary Tables S5).

- Perspective 1: Can we predict AKI before its onset using data before the onset time?
- Perspective 2: Can we predict whether AKI will occur in postoperative patients during their hospital stay?
- Perspective 3: Can we predict, at any random point after surgery, if a patient will develop AKI within a various number of days?

We randomized the study cases across the training (60%), calibration (15%), and internal test (25%) sets. The AKI prediction model was trained and optimized using the training set and was evaluated on the test set. For all of the variables, the differences between the training set, the calibration set, and the test set were nonsignificant.

Different experimental designs would result in different sample sizes for the model building and evaluation. In perspectives 1 and 2, the prevalence of AKI in our cohort was 9.8%, and the main difference was the data collection window and prediction point. In perspective 3, the prevalence of AKI increased with the increasing length of the prediction window (2.4%–9.5%). Given that the data are imbalanced, we applied the synthetic minority oversampling technique (SMOTE)[34] to achieve a 1:1 ratio for AKI cases and no AKI cases, and such an oversampling strategy is also used in many studies.[25,35] For comparative purposes, we also reported the models without any oversampling to present the possibility of operating this algorithm in a normal clinical environment where the prevalence of AKI is relatively low.

We used 5-fold cross validation on the training set to train the proposed models and to iteratively improve the models by selecting the best model architectures and hyperparameters. The models selected for cross validation were recalibrated on the calibration set to further improve the quality of the risk predictions. The recalibration ensures that consistent probabilistic interpretations of the model predictions can be made.[36] To compare uncalibrated predictions to recalibrated ones, we used the Brier score and mean absolute error.[37] The best model was evaluated on the independent test set that was retained during the model development.

The main metrics used in evaluation are the accuracy, recall, the area under the precision-recall curve (ROC AUC), and the area under the receiver operating curve (PR AUC). To gauge the uncertainty of the performance of a trained model, we calculated 95% confidence intervals with the pivot bootstrap estimator. This was done by sampling the entire calibration and test dataset with replacement 100 times.

RESULTS

Model performance

The performance of our proposed model from different perspectives within the different prediction windows was shown in Figure 2. For evaluation perspective 1, our model predicted AKI well, with the ROC AUC increasing from 0.862 [95% CI, 0.861–0.863] to 0.902 [95% CI, 0.901–0.903] within the training prediction point of 168 h to 6 h before the AKI onset (Figure 2a and b). The performance of the model is improved when the prediction point is close to the AKI onset. Figure 2c and d shows the performance of our proposed model used to predict AKI events from the end of surgery until 3 days after surgery in perspective 2. The model demonstrated improved performance over time and was steadily accurate 24 h after surgery (ROC AUC: 0.908 [95% CI, 0.907–0.909]); that is, using data collected 24 h after surgery to predict AKI during the patient's hospital stay resulted in a better accuracy than using data collected at other time points. The evaluation results of the models built from perspective 3 are shown in Figure 2e and f. The best length of the prediction window was 12 h, with a ROC AUC of 0.853 [95% CI, 0.852–0.854] and a PR AUC of 0.853 [95% CI, 0.852–0.855]. The model from the different perspectives of the different prediction windows showed calibration with an average Brier score of 0.137 for the test set, while it was significantly improved to 0.061 or 0.062 after simple recalibration using Platt scaling or isotonic regression (Supplementary Tables S6).

Comparing with other models

The corresponding receiver operating characteristic and precision recall curves of AKI prediction of different models including LR, RF, LSTM, GRU, Dipole, and RETAIN in the different time windows on 3 evaluation perspectives are shown in Supplementary Figures S2–S4. The average performance of the proposed model and these baseline models in the different time windows on 3 evaluation perspectives are shown in Table 2 (detailed performance metrics of these models are shown in Supplementary Tables S7–S9).

Our model shows outstanding performance and achieves best scores in most of the metrics. Compared with deep learning methods, the overall performance of the classic methods (SVM and LR) is lower. In addition, the improvements in all the attention-based models, including Dipole and RETAIN, are significant in all perspectives compared with the basic RNN methods. All these results prove that the modeling time information of each event is meaningful and that the attention mechanism can make the models focus on the events that contain risk factors. In our model, the time parameter is generated by a complicated self-learned nonlinear layer that ensures learning the best time-level attention. We also tested the calibration in these baseline models and found that our proposed model is well calibrated in different time windows on 3 evaluation perspectives (Supplementary Tables S7–S9).

Clinical interpretation of the AKI prediction model

In Figure 3a, the most impactful 15 features of our prediction model using data collected at 6 h before AKI onset in order to predict AKI during this hospitalization were shown. Based on the SHAP values, the autologous blood transfusion during CPB can significantly reduce the risk of AKI and prolonged mechanical ventilation poses a higher risk of AKI. The contributions of different features in model at different time points are different as shown in SHAP summary plots for different prediction models of 3 prediction perspectives in Supplementary Figures S5–S7. We also looked inside the AKI prediction model by evaluating the attention weights of the different timestamps identified by the time-aware key-query attention mechanism. Figure 3b shows a heatmap of the average attention weights of the last 10 events for all 1473 patients in the oversampled test set. On

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

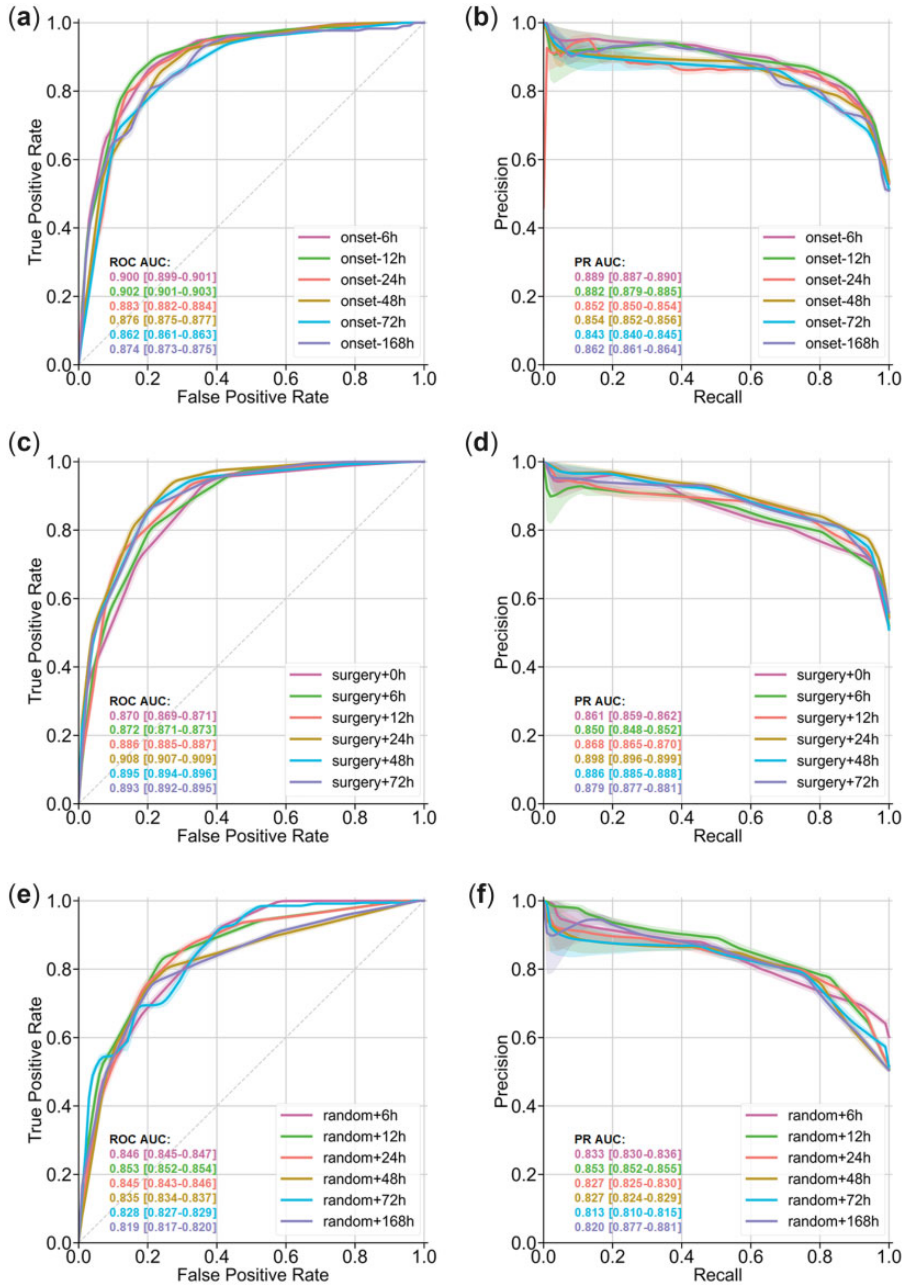

Figure 2. Model performance from the different evaluation perspectives. The model performance of our proposed model from different evaluation perspectives illustrated by the receiver operating characteristic curve, precision-recall curve for predicting any AKI events. (a and b) Perspective 1. (c and d) Perspective 2. (e and f) Perspective 3.

the heatmap, it can be seen that for different patients, the attention weights learned by the attention mechanism are different.

DISCUSSION

We developed a time-aware attention RNN for prediction of postoperative AKI after pediatric cardiac surgery based on routinely collected features during the hospital stay and then retrospectively validated the model on an independent test set. In an effort to conduct a more comprehensive assessment of our AKI prediction model and to test the clinical significance, we explored 3 evaluation perspectives that were designed to investigate the prediction performance in more realistic clinical application scenarios. The 3 evaluation perspectives answer different AKI prediction questions faced in clinical practice and offer a more comprehensive analysis of the current state of AKI prediction, which may provide useful guidance for future predictive modeling for clinical practice. Our proposed model significantly outperformed the existing state-of-the-art models with an improved average performance for any AKI prediction from the 3 evaluation perspectives.

The reported incidence of AKI after pediatric cardiac surgery has been shown to be approximately 9.6% to 64%.[4,5,38] In this study, we oversampled the AKI events to balance the samples in dataset as previous studies.[25,35] In addition, we simulate how changes in the

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

Table 2. Average performance of the different models on the test set for predicting AKI within different prediction windows

Evaluation	Model	Accuracy	PR AUC	ROC AUC
Perspective 1	LR	0.743	0.806	0.800
	SVM	0.762	0.808	0.817
	LSTM	0.758	0.823	0.824
	GRU	0.769	0.821	0.826
	Dipole	0.766	0.829	0.832
	RETAIN	0.780	0.832	0.843
	Ours	0.814	0.864	0.883
Perspective 2	LR	0.742	0.800	0.798
	SVM	0.735	0.798	0.805
	LSTM	0.745	0.804	0.815
	GRU	0.751	0.802	0.809
	Dipole	0.751	0.814	0.812
	RETAIN	0.754	0.816	0.817
	Ours	0.808	0.874	0.887
Perspective 3	LR	0.739	0.790	0.794
	SVM	0.737	0.787	0.790
	LSTM	0.742	0.802	0.805
	GRU	0.741	0.797	0.802
	Dipole	0.753	0.813	0.820
	RETAIN	0.754	0.808	0.816
	Ours	0.776	0.829	0.838

Abbreviations: GRU: gate recurrent unit; LR: logistic regression; LSTM: long short-term memory; SVM: support vector machine.

prevalence of AKI impact the PR AUC and ROC AUC, the estimated PR AUC increases and ROC AUC remains stable with the increasing of AKI prevalence (Supplementary Tables S10 and S11). When AKI incidence is low, the ratio of positive to negative is unbalanced, we adjusted the weight parameter in the model training to make the model pay more attention to positive samples. One possible interpretation of PR AUC decline may be that the ratio of positive and negative is too unbalanced and adjusting the weight parameter to improve recall will result in a significant reduction in precision. The simulated results illustrate the application of the proposed model in natural clinical settings where the prevalence of AKI varies depending on the type of surgery, age or location of the institutions.

Many existing AKI prediction models have been developed and evaluated using data collected when using the AKI onset as an anchor point.[20,21] Our model predicted AKI in a time frame up to 7 days in Perspective 1. This is a much longer time period when compared to the observation windows of other studies.[19,39] Events in the near future are usually easier to predict than those in the more distant future. To intervene early when the kidneys are merely at risk of injury, a longer prediction window might be necessary. It has been shown that early intervention can prevent AKI or its progression to higher stages.[17,40] However, the testing data corresponding to these results do not conform to the real-life clinical application because the AKI onset is unknown. To investigate the prediction performance in more realistic clinical application scenarios, we made continuous AKI predictions at the end of surgery for forecasting the patient risk of developing AKI after surgery and found that our model achieved the best ROC AUC of 0.911 [95% CI, 0.909–0.913] when choosing 24 h after surgery as the prediction point. To further analyze the performance of the model for AKI prediction at any time after surgery in clinical environment, we evaluated whether the performance of the model varies with the prediction window size. The PR AUC for shorter prediction windows in perspective 3 was relatively lower. The major factor for this phenomenon may be related to the positive sample size. The positive sample proportion is only 2.4% in the 6-h prediction window.

When deploying predictive models to real world settings, the discrimination alone is insufficient to assess the prediction capability.[41] In this study, our proposed model performed better in both calibration and discrimination when compared to the baseline models. In addition, the clinical interpretability of machine-learning models is important because it may be difficult or impossible to detect subtle shortcomings of accuracy-driven black-box models. Using the SHAP values, we can see that the proposed model identified many already established risk factors such as age, operation time, cardiopulmonary bypass time, and blood product transfusion during surgery.[21,42] In particular, autologous blood transfusion was at the forefront of influencing factors in many scenario models, inspiring investigators to further explore the mechanisms involved. In addition, the attention weights of the different timestamps identified by the time-aware key-query attention mechanism can provide evidence when the patient's changes have a greater impact on the AKI risk. All the computer code used for in this study is available at https://github.com/Healthink/AKIPrediction to allow independent replication.

Generalization ability of the prediction model is also important when it is applied in clinical scenarios. Our model shows outstanding performance and achieves best scores in most of the metrics in different AKI prediction scenarios when compared with other existing models. In addition, we calculated 95% confidence intervals with the pivot bootstrap estimator (100 times) to gauge the uncertainty of the performance of our model. The prediction performance of our proposed model has little fluctuation according to 95% confidence intervals. Besides, we compare the performance of our proposed model with different AKI definitions, and the model achieves outstanding performance with different AKI definitions (Supplementary Table S12). The above 3 results can prove the robustness and generalization of our proposed model. We believe that this model can also achieve outstanding performance in other related clinical settings.

There are several limitations to our study. First, our analysis used only single-center data. The performance of the machine learning algorithm might differ for larger datasets with different institutions. As such, external validation is required in the future. Second, due to data availability, the model only used creatinine alone to define AKI. Some of the AKI patients could be diagnosed based on the urine output criteria, and the lack of urine criteria may have resulted in lower AKI rates in this study than previously described.[14] This may impact the model performances in a few ways, which includes the consideration that the model is likely to have lower sensitivity with regard to the urine-staged AKI patients. Third, our proposed model has more parameters when compared with simple models such as LR, and although we used Shapley values or attention weights to improve the model interpretability, it still takes some time for clinicians to understand how single features relate to the model output. Furthermore, we do not know whether factors such as data density play any unknown role in the hidden layer of such black-box algorithms. Although our model may create a larger computation burden in the training process due to the complex structure. We will deploy the trained model using historical patient data in the electronic medical record system in the real-world application. For new patients, the proposed model does not need to train again, but directly gives the prediction results according to the input features of new patients, which will not bring additional computational burden.

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

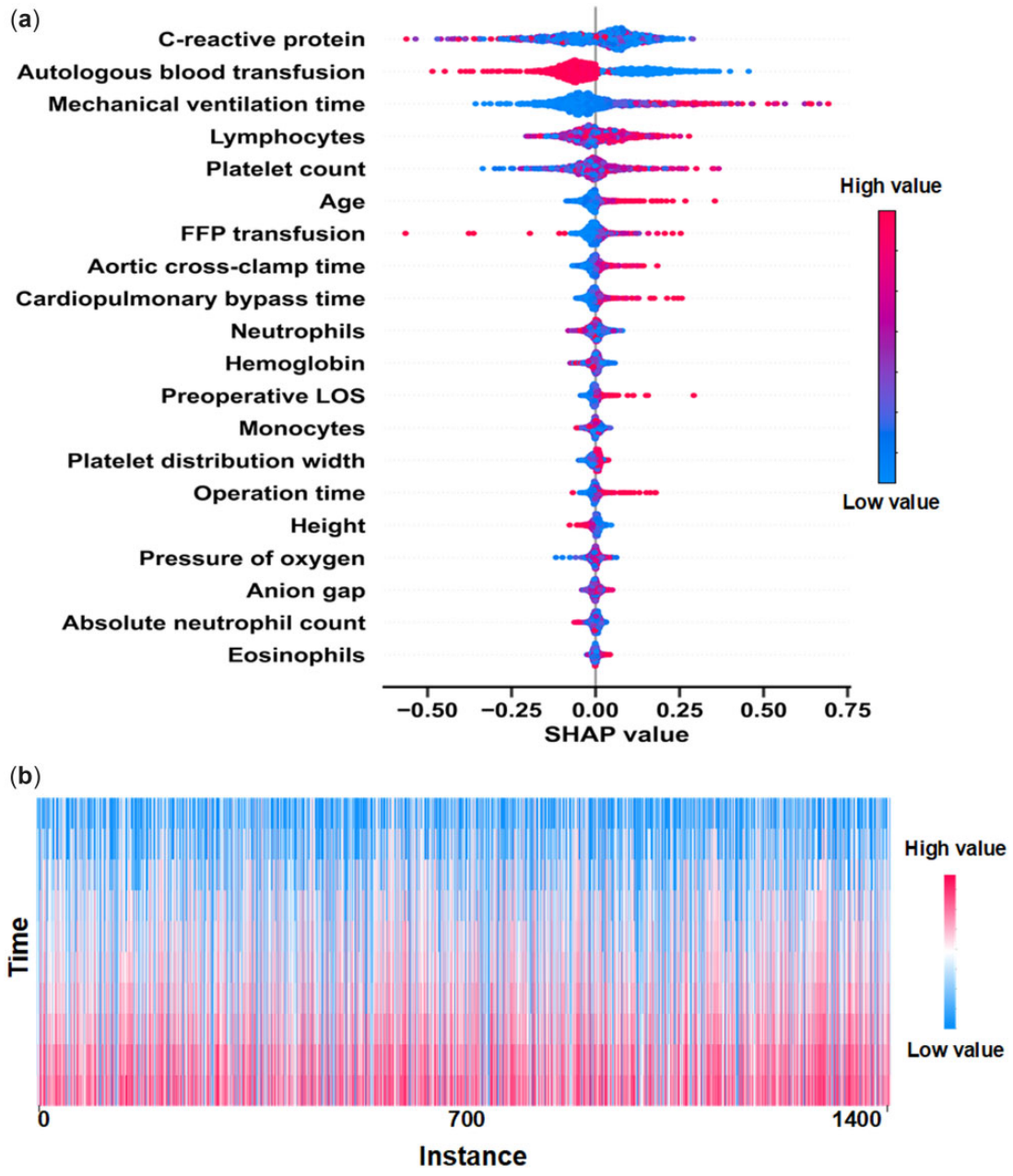

Figure 3. Feature inspection and model interpretation. (a) The 15 features with the highest mean absolute SHAP values in one model. (b) Heatmap plot showing the time attention weights of the different patients in the last 10 timestamps before the prediction point.

CONCLUSIONS

In conclusion, we developed a highly accurate model for the prediction of AKI after pediatric cardiac surgery, and the model significantly outperformed the state-of-the-art prediction models from different evaluation perspectives. The model could potentially be integrated into EHR systems to predict postoperative AKI through real-time patient surveillance.

AUTHOR CONTRIBUTIONS

HL, XZ, QS, and HD designed the experiments; SS, TL, RL, and JL defined and developed the labeling of AKI events and collected the data; XZ, YS, and YF preprocessed and cleaned the data; XZ developed code for proposed model; XZ and YS developed the baseline models; XZ, SS, LT, RL, JL, HL, and QS contributed to various analyses of the data; XZ and HL wrote the manuscript with the assistance and feedback of all the other co-authors.

FUNDING

This work was supported by the National Natural Science Foundation of China (81871456).

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

PIC database is available on http://pic.nbscn.org, and the full clinical dataset can be obtained from the corresponding author on reasonable request. The computer code used in this research is available at https://github.com/Healthink/AKIPrediction to allow independent replication.

REFERENCES

1. Triedman JK, Newburger JW. Trends in congenital heart disease. *Circulation* 2016; 133 (25): 2716–33.
2. Taylor ML, Carmona F, Thiagarajan R, Ferguson M, del Nido P, Rajagopal S. Early postoperative acute kidney injury and outcomes following surgery for congenital heart disease. *J Am Coll Cardiol* 2012; 59 (13): E754.
3. Tóth R, Breuer T, Cserép Z, *et al.* Acute kidney injury is associated with higher morbidity and resource utilization in pediatric patients undergoing heart surgery. *Ann Thorac Surg* 2012; 93 (6): 1984–90.
4. Blinder JJ, Goldstein SL, Lee VV, *et al.* Congenital heart surgery in infants: effects of acute kidney injury on outcomes. *J Thorac Cardiovasc Surg* 2012; 143 (2): 368–74.
5. Taylor ML, Carmona F, Thiagarajan RR, *et al.* Mild postoperative acute kidney injury and outcomes after surgery for congenital heart disease. *J Thorac Cardiovasc Surg* 2013; 146 (1): 146–52.
6. Morgan CJ, Zappitelli M, Robertson CMT, *et al.*; Western Canadian Complex Pediatric Therapies Follow-Up Group. Risk factors for and outcomes of acute kidney injury in neonates undergoing complex cardiac surgery. *J. Pediatr* 2013; 162 (1): 120–7.e1.
7. Chawla LS, Eggers PW, Star RA, Kimmel PL. Acute kidney injury and chronic kidney disease as interconnected syndromes. *N Engl J Med* 2014; 371 (1): 58–66.
8. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders HJ. Acute kidney injury. *Nat Rev Dis Prim* 2021; 7: 1–17.
9. Walther CP, Podoll AS, Finkel KW. KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl* 2012; 2: 1–138.
10. Akcan-Arikan A, Zappitelli M, Loftis LL, Washburn KK, Jefferson LS, Goldstein SL. Modified RIFLE criteria in critically ill children with acute kidney injury. *Kidney Int* 2007; 71 (10): 1028–35.
11. Xu X, Nie S, Zhang A, *et al.* A new criterion for pediatric AKI based on the reference change value of serum creatinine. *J Am Soc Nephrol* 2018; 29 (9): 2432–42.
12. Lachance P, Villeneuve PM, Rewa OG, *et al.* Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review. *Nephrol Dial Transplant* 2017; 32 (2): 265–72.
13. Hodgson LE, Selby N, Huang TM, Forni LG. The role of risk prediction models in prevention and management of AKI. *Semin Nephrol* 2019; 39 (5): 421–30.
14. Kaddourah A, Basu RK, Bagshaw SM, *et al.*; AWARE Investigators. Epidemiology of acute kidney injury in critically ill children and young adults. *N Engl J Med* 2017; 376 (1): 11–20.
15. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-hospital AKI. *J Am Soc Nephrol* 2018; 29 (2): 654–60.
16. Goldstein SL, Mottes T, Simpson K, *et al.* A sustained quality improvement program reduces nephrotoxic medication-associated acute kidney injury. *Kidney Int* 2016; 90 (1): 212–21.
17. Meersch M, Schmidt C, Hoffmeier A, *et al.* Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive Care Med* 2017; 43 (11): 1551–61.
18. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med* 2018; 46 (7): 1070–7.
19. Tomašev N, Glorot X, Rae JW, *et al.* A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019; 572 (7767): 116–9.
20. Tseng PY, Chen YT, Wang CH, *et al.* Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* 2020; 24: 1–13.
21. Dong J, Feng T, Thapa-Chhetry B, *et al.* Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. *Crit Care* 2021; 25 (1): 1–8.
22. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018; 19 (6): 1236–46.
23. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* 2017; 24 (2): 361–70.
24. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep* 2018; 8: 1–12.
25. Rank N, Pfahringer B, Kempfert J, *et al.* Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *NPJ Digit Med* 2020; 3 (1): 1–12.
26. Choi E, Taha Bahadori M, Kulas JA, *et al.* RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv Neural Inf Process Syst* 2016; 29: 3504–12.
27. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; 2017: 1903–1911; Halifax, NS.
28. Sato N, Uchino E, Kojima R, Hiragi S, Yanagita M, Okuno Y. Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. *Comput Methods Programs Biomed* 2021; 206: 106–29.
29. Sutherland SM, Chawla LS, Kane-Gill SL, *et al.* Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15(th) ADQI Consensus Conference. *Can J Kidney Heal Dis* 2016; 3: 11.
30. Zeng X, Yu G, Lu Y, *et al.* PIC, a paediatric-specific intensive care database. *Sci Data* 2020; 7: 1–8.
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015; 13: 1–10.
32. Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. In: *31st conference on Neural Information Processing Systems (NIPS 2017)*; 2017, Long Beach, CA.
33. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions, In: *NIPS 2017*; 2017: 4765–74; Long Beach, CA.
34. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *JAIR* 2002; 16: 321–57.
35. Goh KH, Wang L, Yeow AYK, *et al.* Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat. Commun* 2021; 12: 1–10.
36. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *Proceedings of the International Conference on Machine Learning*; 2017: 1321–30; Sydney, Australia.
37. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc* 2020; 27 (4): 621–33.
38. Li S, Krawczeski CD, Zappitelli M, *et al.*, TRIBE-AKI Consortium. Incidence, risk factors, and outcomes of acute kidney injury after pediatric cardiac surgery: a prospective multicenter study. *Crit Care Med* 2011; 39 (6): 1493–9.
39. Mohamadlou H, Lynn-Palevsky A, Barton C, *et al.* Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. *Can J Kidney Heal Dis* 2018; 5: 1–9.
40. Balasubramanian G, Al-Aly Z, Moiz A, *et al.* Early nephrologist involvement in hospital-acquired acute kidney injury: a pilot study. *Am J Kidney Dis* 2011; 57 (2): 228–34.
41. Alba AC, Agoritsas T, Walsh M, *et al.* Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017; 318 (14): 1377–84.
42. Huen SC, Parikh CR. Predicting acute kidney injury after cardiac surgery: a systematic review. *Ann Thorac Surg* 2012; 93 (1): 337–47.

Downloaded from https://academic.oup.com/jamia/article/30/1/94/6775039 by McGill University Libraries user on 20 August 2023