BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Transfer learning-enabled outcome prediction for guiding CRRT treatment of the pediatric patients with sepsis

Xiao-Qing Li[1], Rui-Quan Wang[1], Lian-Qiang Wu[1] and Dong-Mei Chen[2,3*]

Abstract

Continuous renal replacement therapy (CRRT) is a life-saving procedure for sepsis but the benefit of CRRT varies and prediction of clinical outcomes is valuable in efficient treatment planning. This study aimed to use machine learning (ML) models trained using MIMIC III data for identifying sepsis patients who would benefit from CRRT. We first selected patients with sepsis and CRRT in the ICU setting and their gender, and an array of routine lab results were included as features to train machine learning models using 30-day mortality as the primary outcome. A total of 4161 patients were included for analysis, among whom there were 1342 deaths within 30 days. Without data augmentation, extreme gradient boosting (XGBoost) showed an accuracy of 64.2% with AUC-ROC of 0.61. Data augmentation using a conditional generative adversarial neural network (c-GAN) resulted in a significantly improved accuracy (82%) and ROC-AUC (0.78%). To enable prediction on pediatric patients, we adopted transfer learning approaches, where the weights of all but the last hidden layer were fixed, followed by fine-tuning of the weights of the last hidden layer using pediatric data of 200 patients as the inputs. A significant improvement was observed using the transfer learning approach (AUCROC = 0.76) compared to direct training on the pediatric cohort (AUCROC = 0.62). Through this transfer-learning-facilitated patient outcome prediction, our study showed that ML can aid in clinical decision-making by predicting patient responses to CRRT for managing pediatric sepsis.

Keywords CRRT, MIMIC III, Newborn, Sepsis, Machine-learning, Clinical validation

Introduction

Sepsis represents a systemic inflammatory response to infection that can lead to significant morbidity and mortality, especially in critically ill patients in the Intensive Care Unit (ICU) [1]. Continuous renal replacement therapy (CRRT) is an established therapeutic procedure for patients with sepsis, aimed at managing renal dysfunction and fluid balance [2, 3], during which cytokines and endotoxins are also removed from septic patients, leading to improved patient outcomes [13]. Neonatal sepsis refers to sepsis in the newborn, a condition that can not only lead to the death of the infant but also long-term morbidities such as acute kidney injury (AKI). Until now, there are no effective prevention and treatment methods

*Correspondence:
Dong-Mei Chen
dongmei966231@163.com
[1]Quanzhou Women's and Children's Hospital, No. 700 Fengze Street, Quanzhou, Fujian 362800, China
[2]The Graduate School of Fujian Medical University, Fuzhou, Fujian 35000, China
[3]Teaching Hospital of Fujian Medical University, Quanzhou Children's Hospital, Quanzhou, Quanzhou, Fujian 362000, China

for neonatal sepsis combined with AKI. CRRT is traditionally used to treat critically ill adults but emerging studies showed the benefit of CRT in managing critically ill children and newborns. However, in managing newborns, such invasive procedure entails accurately determining the benefit of CRRT for each patient, which remains a challenge in clinics [4].

To develop such an outcome prediction tool, serum biomarkers have been the focus of research due to its cost-effectiveness and superior accessibility than imaging or pathological examination. For example, a study using blood samples of three children identified serum lysozyme C and leucine-rich alpha-2-glycoprotein as potential biomarkers correlated to CRRT therapy outcomes [4]. In another study, cell-free plasma DNA (cfDNA) was demonstrated as a prognostic marker of clinical outcome in septic patients who underwent CRRT [5]. Another retrospective study with 141 children suggested that higher plasma levels of IL-10, IL-6, IFN-γ and percentage of Treg are associated with severe neutropenia after treatment. Despite these promising results, a prognosis tool for neonatal sepsis patients has yet to be established, presumably due to insufficient sensitivity and specificity of using one or just a few lab tests for outcome prediction.

Machine learning (ML) has emerged as a powerful tool for the development of predictive models in healthcare, providing opportunities for improved patient stratification and treatment planning [6]. Particularly, the application of ML models allows the prediction of patient outcomes based on an array of patient characteristics, instead of just a few, which often demonstrates superior accuracy than traditional approaches [7, 8]. Previously, a number of studies have demonstrated capability of ML models in predicting outcomes of sepsis patients, and unveiled the tremendous potential of ML models in utilizing electronic medical record (EMR) data [9–11].

In this context, the present study aims to develop and validate ML models for identifying ICU patients with sepsis who will benefit CRRT treatment. Gender and routine laboratory results before CRRT treatment were used as features to train an ML model to predict patient's 30-day mortality as the clinical outcome. This approach leverages the Medical Information Mart for Intensive Care III (MIMIC III) dataset [12], an extensive collection of de-identified data that has been previously used in several critical care studies, utilizing data augmentation strategies to improve model performance. While previous studies that used MIMIC III are majorly focused on early detection of sepsis and sepsis-related mortality, our study aimed to train a model to identify sepsis patients who are more likely to respond to CRRT treatment before the treatment is performed. Unfortunately, as CRRT for pediatric patients is rarely adopted, MIMIC III did not contain pediatric patients who received CRRT. To address this, we leveraged transfer learning to fine-tune the model pre-trained on MIMIC III, a much larger dataset, on a new set of collected data of 200 pediatric patients with CRRT, to address the challenge of limited sample size of this cohort. We also assessed the utility of our ML approach in clinical decision-making for this vulnerable population. The results of the study could contribute to the integration of ML into routine clinical practice, potentially facilitating the early and targeted initiation of CRRT in sepsis patients who are most likely to benefit.

Method

MIMIC III cohort identification

The Intensive Care Medical Information Mart III (MIMIC-III) is a publicly accessible database encompassing clinical data for over 60,000 anonymous patients treated in the Intensive Care Unit at Beth Israel Deaconess Medical Center between 2001 and 2012. Managed by the Laboratory of Computational Physiology at the Massachusetts Institute of Technology (MIT), the database stores a comprehensive array of patient records including demographics, vital signs, and medication details. This extensive collection of data has proven instrumental in various academic and clinical studies, such as epidemiology research, clinical decision-making support, and the advancement of healthcare technology tools. Our study identified sepsis patients using "sepsis", "septic shock", "septicemia" as keywords to identify ICD9 codes associated with sepsis (Table S1). Those who underwent CRRT, using the CPT code of 99,231, 99,232, 99,233 were included. The inclusion criteria were age>18 years old. The cohort contains 4161 patients, including 2978 survival and 1342 deaths identified using EXPIRE_FLAG=1 and death within 30 days of admission.

Machine learning using standard models

Extreme gradient boosting (XGBoost) [13], an ensemble model, was used as the model for classifying patients using 30-day mortality as the primary outcome. Features used in the study include gender, blood pH, lactate, potassium, sodium, BUN, C-reative protein (CRP), white blood cell (WBC), and Hemoglobin (Hb). Gender was one-hot encoded. Patient records with >25% missing values were excluded. Missing value was imputed using the most frequent value. Data was split into 70% training data and 30% testing data. Model was trained using 5-fold grid-search cross validation method with XGBoost. Logistical regression, and random forest were also trained for comparisons.

Conditional generative adversarial neural network (c-GAN)

To accommodate for the small size of the dataset, a conditional GAN was used to augment data, which consists

of 10 continuous features (lab results), and two categorical variables (gender and mortality). Mortality was used as the label for data augmentation. The GAN consists of two components: a generator network (G), which synthesizes new data instances, and a discriminator network (D), which evaluates the authenticity of instances from the generator. The GAN model is trained through a two-player minimax game where the generator tries to fool the discriminator and the discriminator tries to correctly classify instances as real or fake. For the generator, which was a four-layer MLP with hidden sizes of 512 and 128, it will take a latent space vector z and label y as input and pass it through a series of fully connected layers, with a final output layer of 12 nodes corresponding to the 12 continuous variables. The discriminator, which was also a four-layer MLP with hidden sizes of 256 and 512, will take a 12-dimensional vector (either real data or synthesized data from the generator) and the label y as input and pass it through a series of fully connected layers, with a final output of a single node with a sigmoid activation function representing the probability that the input is real data. Real data is fed directly to the discriminator to train the discriminator.

The quality of synthesized data was evaluated by distribution of variables of the fake data to real data. Synthesized data are added to the real data at 1:1 ratio. Classes were also balanced by generating more mortality patient samples. The new data were trained using XGboost, logistical regression and random forest.

Pediatric cohort

A retrospective analysis was conducted on 33 cases of sepsis in children admitted to our hospital from September 2015 to December 2022. The inclusion criteria were as follows: ① compliance with the standards related to "Chinese Guidelines for the Treatment of Severe Sepsis/Septic Shock"; ② satisfaction of diagnostic criteria: Scr > 26.5μmol/L (0.3 mg/dl) within 48 h, an increase of ≥0.5 times from baseline, and urine output < 0.5 ml/kg·h for 6 h, all in conjunction with AKI; ③ meeting the indications for CRRT treatment; ④ informed consent obtained from patients and their families. The exclusion criteria included: ① a history of chronic renal failure; ② AKI not caused by sepsis; ③ severe hemodynamic fluctuations requiring cessation of CRRT treatment; ④ missing clinical data. This study was reviewed and approved by the medical ethics committee. Patient blood tests were performed to record blood pH, Scr, BUN, K+, Na+, WBC, Hb, PLT, lactate, CD4, CD8+, IL6 and TNF-α. Sequential Organ Failure Assessment (SOFA) [14] score were also measured to assess the severity of organ dysfunction.

CRRT treatment

All validation cohort underwent routine treatment, including early resuscitation, anti-infection measures, fluid administration, and the use of vasoactive drugs. Subsequently, CRRT treatment was performed using the Plasauto iQ21 type blood purification machine (Asahi Kasei Medical (Hangzhou) Co., Ltd.), selecting the continuous veno-venous hemofiltration and dialysis (CVVHDF) mode. A double-lumen central venous catheter was placed in the femoral vein, and the filter and tubing were pre-filled with heparinized saline. After prefilling was completed, they were connected to the blood purification machine for circulation. Parameter adjustments included: an initial pump flow rate of 3 ml/kg·min, which could be discretionarily adjusted to 5 ml/kg·min, replacement fluid 2030 ml/kg·h (using Shijiazhuang Siyao Pharmaceuticals' replacement fluid), and dialysis fluid 1525 ml/min·m^2 (using Baxter Company's Baxter dialysis fluid). After treatment, regular heparin (if there was no risk of bleeding) or citrate (if there was a risk of bleeding) was used for anticoagulation. If clotting occurred in the filter membrane, it should be promptly replaced. Infections, hypothermia, obstruction, electrolyte imbalance, thrombocytopenia, hypotension, and hemorrhage are used as the basis for calculating morbidity counts.

Data from the validation cohort were fed into the train model, with missing features represented as NaN (not a number). Output from the model, i.e., morbidity or not, were compared to their 1-month follow-up results to acquire true positive and false-positive rates.

Transfer learning

For transfer learning, the weights of the first three layers of the pre-trained MLP were fixed and fine-tuning was performed on the last hidden layer to adapt the model using the pediatric dataset as inputs. The training process utilized a stochastic gradient descent optimizer, with a learning rate of 0.01 and a momentum of 0.9, to minimize the categorical cross-entropy loss function.

Statistical analysis

Data are presented as mean(SD) or percentage(95% CI). Comparison of numeric data and categorical data were performed using t-test and Chi-square test respectively. $P < 0.05$ was considered statistically significant.

Results

Dataset and preprocessing

A total of 4161 sepsis patients who received CRRT in the ICU setting were included in the analysis. The dataset was extracted from the MIMIC-III database, comprising both males and females with various routine lab results, such as platelet count and blood urea nitrogen (BUN). Table 1 describes the characteristics of the cohort.

Table 1 Cohort characteristics

	Mortality group	Control group	*P value
Gender male, n (%)	771(57.5%)	1634(54.9%)	0.111
CD3 (#/uL)	959.25(701.99-1216.51)	953.20(776.39-1130.01)	0.9710
CD4 (#/uL)	362.83(239.64-486.02)	304.43(248.84-360.02)	0.4226
CD8 (#/uL)	610.65(383.96-837.34)	646.28(495.24-797.31)	0.8081
Lymphocyte Count (#/uL)	1447.69(1013.17-1882.21)	1253.06(1049.35-1456.77)	0.4516
Albumin (g/dL)	2.9946(2.9587–3.0304)	3.11(3.08–3.13)	*P* < 0.0001
Albumin, Urine (g/dL)	23.99 (13.43–34.54)	30.29(20.27–40.30)	0.4214
Amylase (IU/L)	100.17(88.25–112.10)	99.52(90.69-108.34)	0.9342
Anion Gap (mEq/L)	16.46(16.23–16.70)	16.20(16.05–16.35)	0.0748
Bicarbonate (mEq/L)	23.61(23.33–23.89)	23.69(23.51–23.86)	0.6701
Bilirubin, Total (mg/dL)	2.70(2.41–2.99)	1.51(1.39–1.62)	*P* < 0.0001
CRP (mg/mL)	86.90(76.80–97.00)	75.50(70.19–80.80)	0.0632
Cholesterol Ratio (Total/HDL)	4.64(4.20–5.07)	4.61(4.38–4.85)	0.9282
Cholesterol, Total (mg/dL)	142.04(136.64-147.43)	156.63(152.50-160.76)	**0.0001**
Creatine Kinase (CK) (IU/L)	371.75(261.19-482.31)	488.68(316.84-660.52)	0.2872
Scr (mg/dL)	2.0145(1.9158–2.1133)	1.78(1.72–1.84)	**0.0003**
Ferritin (ng/mL)	841.43(606.61-1076.26)	625.31(568.168–682.46)	0.0964
Folate (ng/mL)	11.61(11.17–12.06)	11.59(11.35–11.83)	0.9255
Glucose (mg/dL)	150.06(145.31–154.80)	151.12(147.86-154.37)	0.7322
Hematocrit (%)	33.99(33.66–34.31)	34.66(34.45–34.88)	**0.0013**
Hemoglobin (g/dL)	11.28(11.16–11.39)	11.61(11.54–11.69)	*P* < 0.0001
Immunoglobulin G (mg/dL)	1218.27(1070.24-1366.30)	1205.64(1124.22-1287.06)	0.8894
Iron (ug/dL)	53.20(49.85–56.56)	45.49(43.65–47.33)	**0.0002**
Lactate (mmol/L)	2.84(2.73–2.95)	2.47(2.41–2.53)	*P* < 0.0001
PLT (K/uL)	231.86(224.53-239.18)	246.14(241.43-250.84)	**0.0023**
Protein, Total (g/dL)	5.87(5.76–5.99)	6.13(6.06–6.21)	**0.0005**
RBC (m/uL)	3.72(3.68–3.76)	3.86(3.83–3.88)	*P* < 0.0001
Na+ (mEq/L)	137.43(137.09-137.77)	137.82(137.63-138.01)	0.0599
Triglycerides (mg/dL)	148.04(134.41-161.68)	170.94(160.59-181.28)	**0.0129**
BUN (mg/dL)	38.64(37.18–40.10)	31.75(30.92–32.59)	*P* < 0.0001
Uric Acid	7.20(6.79–7.61)	6.14(5.90–6.39)	*P* < 0.0001
Vitamin B12 (pg/mL)	922.91(878.58-967.25)	843.87(818.50-869.24)	**0.0041**
WBC (K/uL)	7.83(6.35–9.31)	8.14(7.20–9.08)	0.7392
pCO2 (mm Hg)	41.14(40.34–41.93)	41.52(41.0-42.022)	0.4429
pH	7.35(7.34–7.35)	7.35(7.354–7.36)	0.1465
pO2(mm Hg)	136.41(130.77-142.05)	137.36(133.48-141.23)	0.7969
K+ (mEq/L)	4.39(4.33–4.45)	4.20(4.16–4.24)	*P* < 0.0001

Data are represented using mean(95% confidence interval). *For numerical data, two-sample t-test was used to compare between two groups. For gender, chi-square test was used.CRP: C-reactive protein;

Gender of the two groups did not significantly differ and male patients constituted 57.5% and 54.9% for mortality and control groups. Albumin, bilirubin, total cholesterol, serum creatinine level (Scr), hematocrit, hemoglobin, iron, lactate, platelet count (PLT), total protein, red blood cell (RBC), triglyceride, blood urea nitrogen (BUN), uric acid, vitamin B12 and potassium (K+) levels showed significant differences (two-sample t test).

Machine learning models and data augmentation

We initially trained the machine learning models using XGBoost, which without data augmentation showed an accuracy of 64.2% and an AUCROC of 0.61, which is higher than the metrics acquired form logistical regression (52.3% and AUCROC of 0.53) and MLP (59.2% and AUCROC of 0.56). Hence, XGBoost was chosen as the model for further analysis.

We reasoned that the suboptimal accuracy was due to limited sample size, so we performed 2x data augmentation using c-GAN, meanwhile balancing the classes by synthesizing more samples in the mortality group. Figure 1A illustrates the basic architecture of c-GAN. The use of c-GAN led to the generation of realistic synthetic samples that augmented the dataset significantly and the synthesized data resembled the real data (Fig. 1B). Re-training using the augmented dataset resulted in a

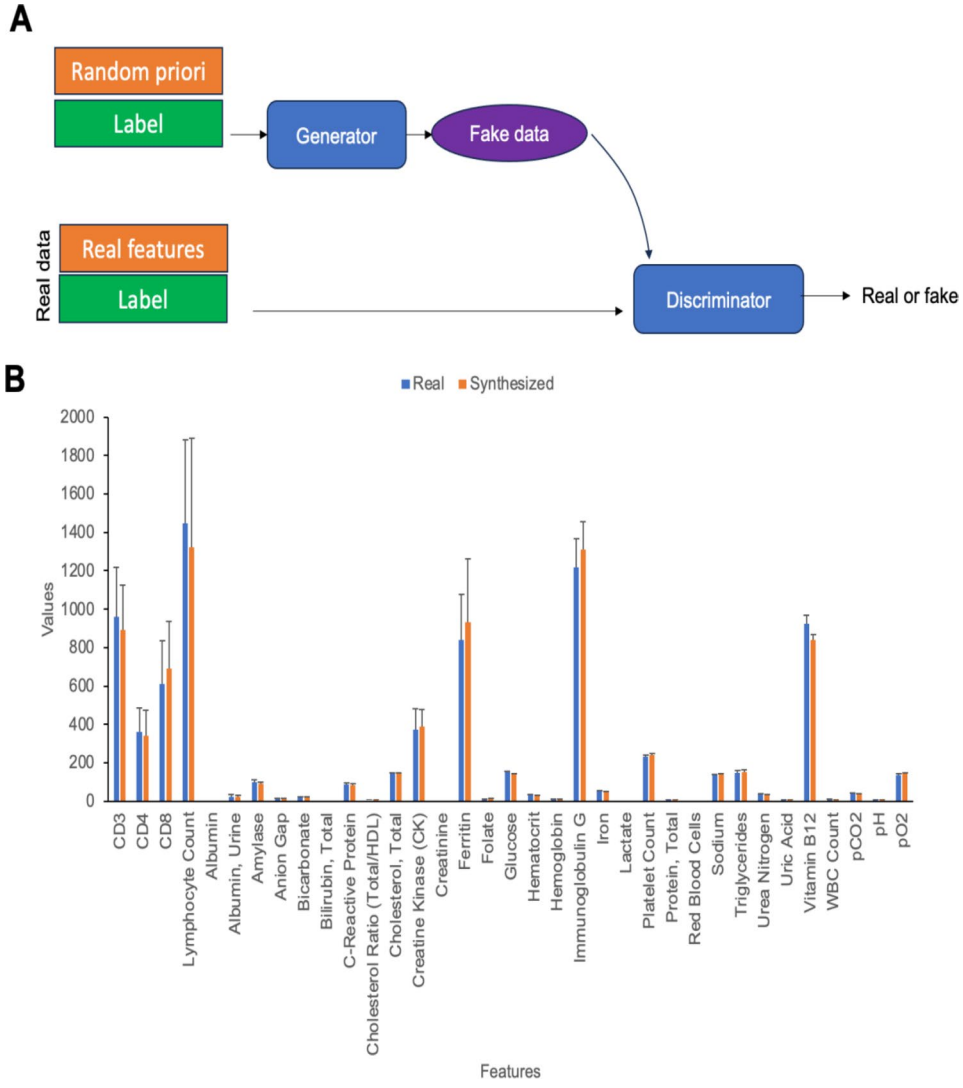

Fig. 1 (**A**) Schematic illustration of c-GAN for data augmentation. (**B**) Comparison of features of synthetic and real data. Data were represented as mean + 95%CI

substantial improvement in model performance using XGBoost, with accuracy, F1-score, and AUCROC reaching 82.4%, 0.82 and 0.78, respectively.

Transfer learning

We extended our study by adapting the model to classify 200 pediatric patients with sepsis, among whom 120 developed mortalities for 1 month follow-up. Finetuning was performed on the last hidden layer using pediatric data (Fig. 2) to facilitate domain transfer. Table 2 shows the characteristics of the pediatric cohort. The two groups showed significant differences in gender and lactate levels. Transfer learning was performed by fixing the layers prior to the last hidden layer of the MLP, followed by fine-tuning using the pediatric data. Our study indicated that the AUCROC of the model increased from 0.60 to 0.73 after 200 epochs. Whereas, training the 200 pediatric patients directly with the MLP can only achieve an AUCROC of 0.62 in the test dataset with a sign of overfitting, suggesting that the use of the pre-trained model offered a boost for model performance.

Feature importance

To understand features that predominantly contributed to the classification results of the model, we used the embedded F score in XGBoost to rank the importance of features (Fig. 3). It is unveiled that creatine kinase (CK), pO2, lactate, amylase, albumin, potassium levels, hemoglobin (Hb), platelet count (PLT), blood urea nitrogen (BUN), and pH are among the important contributors for outcome prediction. These features dictated the model's classification results and may hold significant clinical insights for understanding patients' response to CRRT.

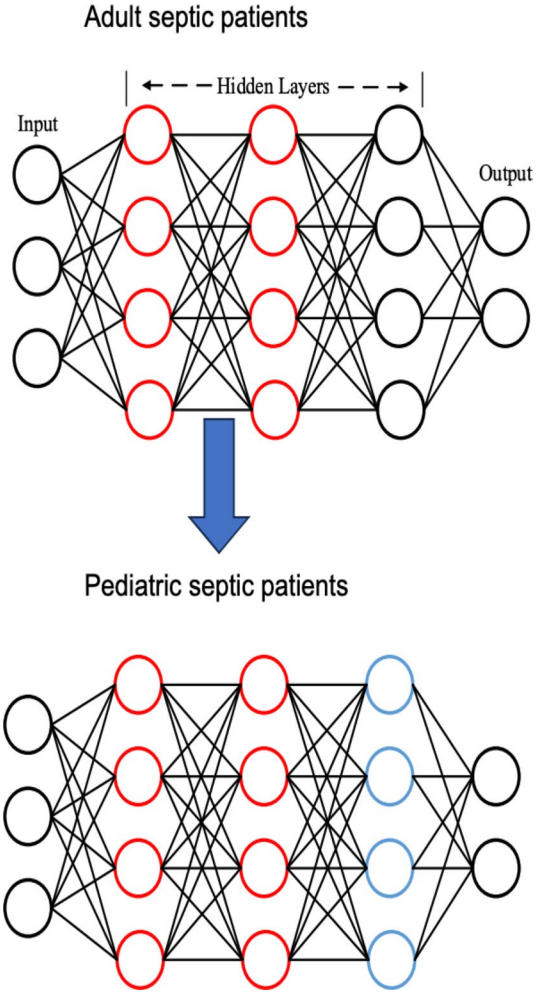

Fig. 2 Rationale for transfer learning. Model pre-trained on large and augmented adult dataset was applied to a different domain, i.e., pediatric patients, where the weights of the last hidden layer (labeled in blue) were fine-tuned

Clinical validation

We extended our study by deploying the model to classify 33 newborns with sepsis in a retrospective analysis, among whom 12 developed morbidities for 1 month follow-up. Table 2 shows the characteristic of the validation cohort. The two groups showed significant differences in gender and lactate levels. Twelve high-risk patients were identified using the trained XGBoost model, among whom 8 developed morbidities and 4 were in the control group, leading to a true-positive (TP) rate of 0.67, false positive (FP) rate of 0.16. This real-world validation confirmed the efficacy of our machine learning model in identifying high-risk patients.

Discussion

With the aim of training a ML model for predicting outcomes of CRRT treatment among patients with neonatal sepsis, we initiated our study by analyzing a large dataset

Table 2 Characteristics of validation cohort

	Morbidity($n=12$)	Control($n=21$)	*p value
Gender male n%	10(83%)	14(67.7%)	**0.0096**
pH	7.16(0.19)	7.156(0.19)	0.87855251
Scr (mg/dL)	1.3276(0.2802)	1.4266(0.5148)	0.47718778
BUN (mg/dL)	12.50(5.27)	13.39(6.29)	0.66546308
K+ (mEq/L)	5.54(1.30)	5.53(1.36)	0.97629452
Na+ (mEq/L)	137.23(7.63)	139.25(7.29)	0.46464174
WBC (K/uL)	14.45(4.89)	15.29(6.52)	0.67758494
Hb (g/dL)	10.59(1.98)	9.74(1.99)	0.24255486
PLT (K/uL)	109.81(37.45)	103.02(39.89)	0.62799096
Lactate (mmol/L)	3.54(1.43)	2.28(2.02)	**0.04430989**
CRP (mg/dL)	67.27(17.12)	68.3(14.82)	0.86302044
IL6 (pg/mL)	785.49(154.24)	728.52(146.97)	0.30742201
TNF (ng/L)	53.31(13.28)	48.90(12.05)	0.35103748
CD4 (#/uL)	369.0(85.7)	367.5(95.0)	0.96201768
CD8 (#/uL)	403.8(105.4)	39.58(9.75)	0.83142445
SOFA	6.38(1.24)	6.66(0.98)	0.50066978

Data were presented as mean(sd). *Two-tailed t-test was used for numerical data; chi-square test was used for gender. Scr: serum creatinine; BUN: blood urea nitrogen; K+: potassium; Na+: Sodium; WBC: white blood cell; Hb: hemoglobin; CRP: C-reactive protein; IL6: interleukin-6; TNF: tumor necrotic factor; PLT: platelet counts

of 4161 sepsis patients undergoing CRRT in the ICU setting within the MIMIC III dataset. Hence, our study differs from previous approaches that examined a single or a combination of a few lab tests for outcome prediction [15, 16], leveraging the big health data generated within EHR system and the capacities of ML models to consider different variables simultaneously. Using XGBoost, an ensemble model, we noted a considerably higher performance in both accuracy and AUCROC values compared to logistic regression and random forest. This could be attributed to (1) the intrinsic ability of XGBoost to handle missing values, which was prevalent in all features [17], (2) the robustness of the boosting approach against overfitting [18], which made XGBoost the focus of our approach.

To enhance model performance, we incorporated data augmentation with c-GAN, aiming to enlarge the sample size and especially balance the representation of the mortality group. c-GAN has been applied in various domains [19], but its application in augmenting EHR data is relatively novel [20, 21]. By generating additional data points that adhere to certain conditions or constraints, c-GAN allows for more robust model training, particularly when handling complex and high-dimensional medical data such as the MIMIC-III dataset. Our results, which demonstrated a significant improvement in both accuracy and AUCROC using XGBoost after the application of c-GAN, align with recent studies that have also highlighted the potential of c-GAN in enhancing predictive models [22, 23]. Our approach, however, is different in that we only synthesized lab records, which are an array of numerical values, which, in comparison to discrete

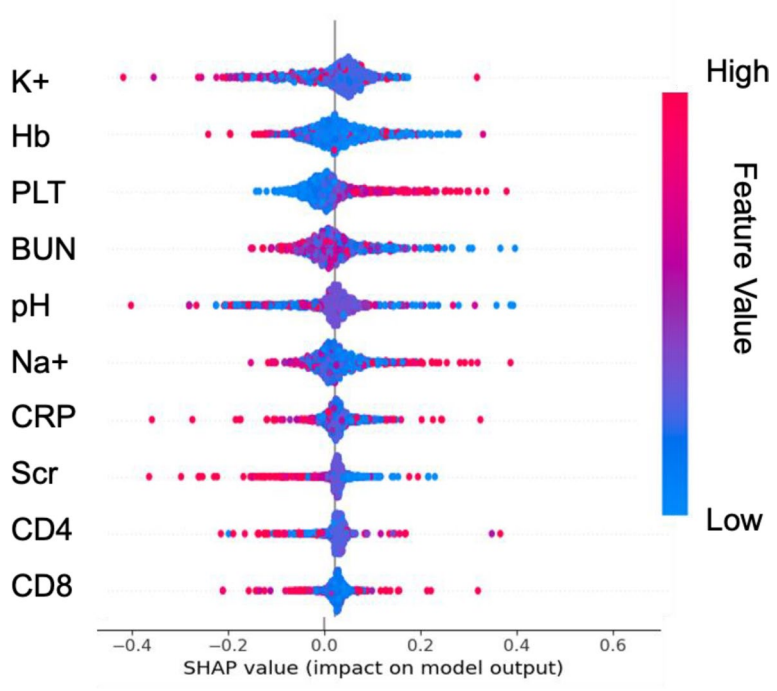

Fig. 3 Feature importance scores of top 10 features based on SHAP on augmented dataset. Scr: serum creatinine; BUN: blood urea nitrogen; K+: potassium; Na+: Sodium; WBC: white blood cell; Hb: hemoglobin; CRP: C-reactive protein; PLT: platelet count; pO2: oxygen pressure of blood

data, e.g. ICD code, require less sophisticated c-GAN architectures. To our best knowledge, this is the first time that such method has been used to address limitation of insufficient sample size within MIMIC III. The reason we chose gradient boosting as the baseline is that tree-based algorithms, like gradient boosting, are in general considered powerful in processing categorical data prevalent in the electronic health records (EHR), which are also present in our data, i.e., gender. But we do recognize that deep learning networks could be strong at dealing with lab test results, which are continuous data, and hence we also explored a typical deep learning network, MLP, to compare the classification performance.

To cope with overfitting caused by insufficient sample size, as is the case for the pediatric cohort, transfer learning was used. This approach leverages knowledge in a similar domain (adult sepsis patients in our study) to help modeling in the target domain (i.e. pediatric sepsis patients in this study). The similarity of the two domains is dependent on the hypothesis that pediatric sepsis patients share similar physiology to adult patients. Indeed, the enhanced model performance using the transfer learning approach on pediatric patients corroborated the benefit of pre-training on the larger MIMIC-III dataset.

Despite that MIMIC-III is a publicly available dataset that has been widely utilized for the development and validation of various clinical predictive models, the translation of these models from research settings to real-world clinical practice often encounters challenges. In our study, the use of a new clinical cohort of pediatric patients with sepsis attests the translatability of models developed using MIMIC-III. A salient feature of this study was the exploration of feature importance, facilitated by XGBoost's embedded F score. Our results indicate that certain routine lab results such as creatine kinase (CK), lactate, and platelet counts, BUN, among others, play a crucial role in predicting the patient's response to CRRT [24, 25]. These findings not only reinforce the clinical significance of these parameters in the management of sepsis patients but also spotlight areas where future clinical research could delve deeper.

Despite that MIMIC-III is a publicly available dataset that has been widely utilized for the development and validation of various clinical predictive models, the translation of these models from research settings to real-world clinical practice often encounters challenges. In our study, we not only developed the models but also carried out a retrospective validation on a new clinical cohort of newborns with sepsis. Our retrospective analysis of 33 newborns with sepsis further confirmed the utility of our machine learning model. The true positive rate of 0.67 and a false positive rate of 0.16 highlight the model's ability to successfully identify high-risk patients, though a larger external validation cohort would be valuable in further substantiating these findings.

In the present study, we only leveraged data from one time point, i.e., the time immediately before CRRT

treatment. However, it is possible that data from multiple time points could be valuable in training a more powerful model. In this regard, novel techniques in tensor factorization, e.g., tensor learning from mutual information [26] and a correlation matrix-based tensor decomposition method [27], could be useful in extracting valuable information and preprocessing of the high-dimensional data for machine learning. Other feature extraction techniques that utilize higher-order statistic relations of the features, such as ratio and power-based feature transformation [28], are also of significant value in enhancing model accuracy by leveraging temporal data. As opposed to the method of using the most frequent value for missing value imputation, as employed in our study, the feature extraction techniques are also powerful in imputing missing values [26–28], and should be explored if high-dimensional temporal data are to be exploited in future studies.

It is worth noting that our cohort in training ML models did not include newborn patients as no newborn patients in MIMIC III underwent CRRT, and training was performed using adult patients. Hence, our features did not include age to circumvent this issue. We also used 30-day mortality as the label for model training, but our validation dataset did not contain mortality so the model was used to predict morbidities. Despite these disparities, the inferencing on our validation dataset showed good performance in identifying patients that would still develop morbidities, suggesting that the capability of generalization of the trained model. It should also be noted that while our study is not aimed to detect mortality events early among newborns receiving CRRT treatment, hence making utility score not suitable, future studies could benefit from utility score as a metric to develop ML models that detect mortality early among newborn patients that receive CRRT treatment. Further studies are warranted to validate the trained model in adult sepsis patients who underwent CRRT. Longer follow-up period is also valuable. Moreover, although the model showed good performance after data augmentation, we reason that gathering more real-world data could yield more robust ML models to facilitate clinical translation.

Conclusion

In conclusion, our study provided another route for more personalized and effective interventions for sepsis patients, particularly in the context of CRRT treatment, demonstrating that ML models have the potential to facilitate decision making to achieved personalized medicine.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02623-y.

Supplementary Material 1

Acknowledgements
None.

Author contributions
CDM contributed to the conception of the study. LXQ wrote the manuscript. WRQ and WLQ performed the data analyses and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding
None.

Data availability
The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Declarations

Ethical approval and consent to participate
This study was reviewed and approved by the ethics committee of Quanzhou women's and children's hospital. The patients/participants provided their written informed consent to participate in this study.

Consent to publish
We had obtained consent to publish from the participant.

Conflict of interest
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 October 2023 / Accepted: 25 July 2024
Published online: 27 September 2024

References

1. Romagnoli S, Ricci Z, Ronco C. CRRT for sepsis-induced acute kidney injury. Curr Opin Crit Care. 2018;24(6):483–92.
2. Lin J, et al. Impact of cumulative fluid balance during continuous renal replacement therapy on mortality in patients with septic acute kidney injury: a retrospective cohort study. Front Med. 2021;8:762112.
3. Uusalo P, et al. Early restrictive fluid balance is associated with lower hospital mortality independent of acute disease severity in critically ill patients on CRRT. Sci Rep. 2021;11(1):18216.
4. Cui Y, et al. The novel biomarkers for assessing clinical benefits of continuous renal replacement therapy in pediatric sepsis: a pilot study. Clin Proteomics. 2023;20(1):1–13.
5. Clementi A, et al. The role of cell-free plasma DNA in critically ill patients with sepsis. Blood Purif. 2016;41(1–3):34–40.
6. Dahiwade D, Patle G, Meshram E. *Designing disease prediction model using machine learning approach*. in 2019 *3rd International Conference on Computing Methodologies and Communication (ICCMC)*. 2019. IEEE.
7. Ratner B. Statistical and machine-learning data mining:: techniques for better predictive modeling and analysis of big data. CRC; 2017.
8. Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. IEEE Rev Biomed Eng. 2020;14:116–26.
9. Nemati S, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2018;46(4):547.
10. Fleuren LM, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46:383–400.
11. Desautels T, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inf. 2016;4(3):e5909.
12. Johnson AE, et al. MIMIC-III, a freely accessible critical care database. Sci data. 2016;3(1):1–9.

13. Chen T, et al. Xgboost: extreme gradient boosting. R Package Version 0 4-2. 2015;1(4):1–4.
14. Lambden S, et al. The SOFA score—development, utility and challenges of accurate assessment in clinical trials. Crit Care. 2019;23(1):1–9.
15. Kim CS, et al. A prospective observational study on the predictive value of serum cystatin C for successful weaning from continuous renal replacement therapy. Kidney Blood Press Res. 2018;43(3):872–81.
16. Jung S-Y, et al. Phosphate is a potential biomarker of disease severity and predicts adverse outcomes in acute kidney injury patients undergoing continuous renal replacement therapy. PLoS ONE. 2018;13(2):e0191290.
17. Latief MA, Bustamam A, Siswantining T. *Performance evaluation xgboost in handling missing value on classification of hepatocellular carcinoma gene expression data*. in 2020 *4th International Conference on Informatics and Computational Sciences (ICICoS)*. 2020. IEEE.
18. Liao X et al. *Research on short-term load forecasting using XGBoost based on similar days*. in *2019 International conference on intelligent transportation, big data & smart city (ICITBS)*. 2019. IEEE.
19. Wang J, et al. Data augment method for machine fault diagnosis using conditional generative adversarial networks. Proc Institution Mech Eng Part D: J Automobile Eng. 2020;234(12):2719–27.
20. Bernardini M et al. A novel missing data imputation approach based on clinical conditional generative adversarial networks applied to EHR datasets. Comput Biol Med, 2023: p. 107188.
21. Ghosheh G, Li J, Zhu T. *A review of Generative Adversarial Networks for Electronic Health Records: applications, evaluation measures and data sources* arXiv preprint arXiv:2203.07018, 2022.
22. Yang W et al. Mortality prediction among ICU inpatients based on MIMIC-III database results from the conditional medical generative adversarial network. Heliyon, 2023. 9(2).
23. Nik AHZ et al. *Generation of Synthetic Tabular Healthcare Data Using Generative Adversarial Networks*. in *International Conference on Multimedia Modeling*. 2023. Springer.
24. Griffin BR, et al. Platelet decreases following continuous renal replacement therapy initiation as a novel risk factor for renal nonrecovery. Blood Purif. 2022;51(7):559–66.
25. Hayes LW, et al. Outcomes of critically ill children requiring continuous renal replacement therapy. J Crit Care. 2009;24(3):394–400.
26. Nesaragi N, Patidar S, Aggarwal V. Tensor learning of pointwise mutual information from EHR data for early prediction of sepsis. Comput Biol Med. 2021;134:104430.
27. Nesaragi N, Patidar S, Thangaraj V. A correlation matrix-based tensor decomposition method for early prediction of sepsis from clinical data. Biocybernetics Biomedical Eng. 2021;41(3):1013–24.
28. Nesaragi N, Patidar S. Early Prediction of Sepsis from Clinical Data using ratio and power-based features. Crit Care Med. 2020;48(12):e1343–9.

Publisher's Note