Knowledge-Based Systems 245 (2022) 108655

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Temporal dynamics of clinical risk predictors for hospital-acquired acute kidney injury under different forecast time windows

Lijuan Wu [a,b,*], Yong Hu [a,b,*], Xiangzhou Zhang [a,b], Borong Yuan [a,b], Weiqi Chen [a,b], Kang Liu [a,b], Mei Liu [c,**]

[a] Big Data Decision Institute, Jinan University, Tianhe, Guangzhou, China
[b] Guangdong Engineering Technology Research Center for Big Data Precision Healthcare, Tianhe, Guangzhou, China
[c] Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, USA

ARTICLE INFO

Article history:
Received 11 November 2020
Received in revised form 21 March 2022
Accepted 23 March 2022
Available online 29 March 2022

Keywords:
Acute kidney injury
Knowledge mining approach
Risk prediction
Temporal fluctuation
Electronic medical records

ABSTRACT

Massive electronic medical records provide unique opportunities and possibilities for gaining medical knowledge and actionable insights to transform healthcare. Acute kidney injury (AKI) is considered one of the most common complications of acute illness, with a substantial impact on patient outcome and hospital costs. In this study, we developed a machine-learning-based knowledge mining approach, combining eXtreme Gradient Boosting (XGBoost) model and SHapley Additive exPlanations (SHAP) method, to explore the temporal dynamics of clinical risk predictors for hospital-acquired AKI under different forecast time windows. AKI occurred in 7,259 (9.43%) of 76,957 hospital admissions from November 2007 to December 2016 extracted from our institution's de-identified clinical database. Through qualitative visualization and quantitative analysis, we found and confirmed that relative importance of optimal risk factors fluctuates with the size of AKI prediction time window. For example, the contribution of most time-varying variables (such as medication and lab tests) to AKI risk increases as AKI onset approaches, while most non-time-varying variables (such as age and admission diagnosis) decrease. Besides, the optimal set of AKI risk predictors will also change under different time window. This study provides several practical implications, including recognizing the existence of feature weight volatility as it is desirable for model accuracy, identifying important AKI risk factor for further investigation, and facilitating early accurate prediction of AKI.

1. Introduction

Acute kidney injury (AKI) is prevalent in hospitalized patients and has been associated with poor short- and long-term outcomes in patients [1]. Delays in identification and intervention for AKI may lead to progression of injury, increasing the risk of death, the need for renal replacement therapy (RRT), and the risk of chronic kidney disease [2]. AKI is especially common among patients hospitalized with COVID-19 (46%) and is associated with extremely high mortality (only 30% survived with recovery of kidney function by the time of discharge) [3]. Therefore, the ability to predict patients at risk for AKI and managing them according to susceptibilities and exposures is more likely to result in better outcomes [4,5]. With the wide-adoption of electronic medical record systems (EMR) in modern healthcare, we are entering a new era where the abundant EMR data is available for data mining [6]. Data-driven approaches that incorporate "big" EMR data has presented a unique analytic opportunity for AKI.

Risk factor selection is an important step in constructing AKI prediction models. AKI risk factors can be either static (time-invariant) or dynamic (time-variant). Static risk factors (e.g., age [7]) are those predisposing factors of the patients that predict AKI risk but are not amenable to deliberate intervention. Admission diagnosis and medical history are also static information as doctors cannot change a patient history nor forgo certain life-saving procedures (such as RRT [8] and liver transplant [9]). In contrast, dynamic risk factors are potentially changeable factors during patient care, which is essential to physicians to prevent AKI. It is important that, if identified in a timely fashion, many risk factors for hospital-acquired AKI can be modified to reduce or prevent AKI from occurring. Thus, time before disease onset provides a window of opportunity to conduct surveillance and prompt intervention.

As AKI management is particularly time-sensitive, early identification of patients at high risk of developing AKI can improve

* Correspondence to: Jinan University, 601 Huangpu Road West, Guangzhou, 510632, China.
** Correspondence to: University of Kansas Medical Center, 3001B Student Center, Mail Stop 3065, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA.
E-mail addresses: wulj1989@163.com (L. Wu), henryhu200211@163.com (Y. Hu), meiliu@kumc.edu (M. Liu).

https://doi.org/10.1016/j.knosys.2022.108655
0950-7051/

patient outcome. For example, Palomba et al. developed the AKICS score system that incorporates five risk categories to predict AKI following cardiac surgery [10]. Lueangingkhasut et al. used five risk factors (congestive health failure, atherosclerotic coronary vascular disease, blood pH $\leq$ 7.3, sepsis and anemia) to develop a risk score model for hospital-acquired AKI in a tertiary care hospital in Thailand [11]. Similar to other existing AKI prediction studies [12,13] that screen known risk factors based on expert knowledge, they may not only miss potential unknown risk factors from the "big" EMR, but also fail to take into account the temporal fluctuation of risk factors.

Recent availability of EMR data and advances in artificial intelligence (AI) have sparked interest in machine learning-based healthcare risk prediction models. For example, Koyner et al. applied gradient boost machine (GBM) method to build an AKI prediction model [14]. Li et al. applied convolutional neural network to predict AKI within 24 h of admission to ICU (intensive care unit) [15]. Tomasev et al. used the deep learning method to capture the temporal short- and long-term variability of AKI predictors to improve prediction performance [16]. Although the sophisticated machine-learning methods provide good prediction accuracy, their application in actual clinical settings is limited because their predictions are difficult to interpret and therefore inoperable.

Improving human–AI collaboration is critical for applications where explaining machine learning model predictions can enhance human performance. One of the most common approaches is to use interpretable models directly. For example, Simonov et al. utilized a logistic regression approach to create a simple and implementable AKI prediction model [17]. Nematic et al. developed an interpretable model for sepsis prediction in ICU using a modified Weibull–Cox proportional hazards model [18]. Another approach is to develop explainable method to explain the "black-box" prediction model. For example, Ribeiro et al. proposed a novel explanation technique, namely LIME (Local Interpretable Model-agnostic Explanations), to explain the predictions of any classifier [19]. Lundberg et al. constructed an explainable AI for tree-models, namely SHAP (SHapley Additive exPlanations), which was used to explain GBM prediction model for hypoxemia during surgery [20,21].

To the best of our knowledge, no study has investigated and characterized the temporal dynamics of AKI risk factors, that is what risk factors are key to prediction at what time. In this study, we first explored an interpretable machine-learning-based knowledge mining method to delineate the risk contribution of each feature in each patient, and then captured fluctuation information of AKI risk factors under different prediction time windows, which may help improve the rationality and effectiveness of prediction models and decisions making.

2. Materials and methods

2.1. Study population

All adult patients (age $\geq$ 18) admitted to the University of Kansas Health System (a tertiary academic hospital) for two days or more from November 2007 to December 2016 were included in this retrospective observational cohort study. Given that a patient may have multiple eligible hospital admissions or encounters and develop AKI during one but not another, this study was conducted at the encounter level with a total of 179,370 encounters. From those encounters, we excluded the ones (a) missing necessary data elements for outcome determination, i.e., less than two serum creatinine measurements and (b) had evidence of moderate or severe kidney dysfunction at admission based on lab measurements, i.e., estimated Glomerular Filtration Rate (eGFR) less than 60 mL/min/1.73 m^2 or serum creatinine (SCr) level of >1.3 mg/dL within 24 h of hospital admission. As shown in Supplementary Figure S1, the final de-identified cohort contained 76,957 encounters.

Table 1
Clinical variables considered in the encounters.

Feature category	Number of variables	Details
Demographics (Demo)	3	Age, race, gender;
Vitals (Vitals)	5	BMI, diastolic BP, systolic BP, pulse, temperature;
Lab tests (Lab)	14	Albumin, ALT, AST, Ammonia, Calcium, BUN, Bilirubin, CK-MB, CK, Glucose, Lipase, Platelets, Troponin, WBC;
Admission diagnosis (DRG)	315	University Health System Consortium (UHC) APR-DRG;
Medical History (CCS)	280	ICD9 codes mapped to CCS major diagnoses.
Medications (MED)	1271	All medications are mapped to RxNorm ingredient;

Table 2
Comparison of 7 different machine learning methods in AKI prediction.

Models (AUC, 95% CI)	72-hour forecast time window	48-hour forecast time window	24-hour forecast time window
XGBoost	**0.736 [0.723–0.745]**	**0.764 [0.750–0.775]**	**0.805 [0.791–0.824]**
GBM*	0.722 [0.710–0.732]	0.750 [0.738–0.760]	0.785 [0.774–0.800]
Random Forest[a]	**0.740 [0.728–0.747]**	**0.770 [0.759–0.781]**	**0.803 [0.788–0.814]**
LinearSVC*	0.728 [0.712–0.737]	0.743 [0.731–0.753]	0.753 [0.745–0.765]
Logistic Regression*	0.726 [0.709–0.738]	0.741 [0.728–0.753]	0.750 [0.741–0.764]
Naive Bayes*	0.670 [0.659–0.682]	0.697 [0.685–0.708]	0.699 [0.684–0.720]
MLP*	0.660 [0.630–0.680]	0.686 [0.674–0.697]	0.706 [0.686–0.714]

Abbreviations and notes: XGBoost: eXtreme Gradient Boosting; GBM: Gradient Boosting Machine; LinearSVC: Linear Support Vector Classification; MLP: Multilayer Perceptron neural network. AUC, the area under the receiver operating characteristic curve. CI, confidence interval.
[a]There is no significant difference in the classification results between XGBoost and Random Forest model, namely p value $>=$ 0.05 of DeLong test.
*The XGBoost model is significantly superior to this model, namely p value < 0.05 of DeLong test.

2.2. AKI definition

We defined AKI according to the SCr-based criteria described in the KDIGO (Kidney Disease Improving Global Outcomes) clinical practice guidelines [22] (see Supplementary Table S1). Baseline SCr level was defined as the most recent SCr value within two-day window prior to admission if available; otherwise, the first SCr value after admission was used as the baseline. Then all pairs of SCr levels measured between admission and discharge were evaluated on a rolling basis to determine the occurrence of any AKI.

2.3. Clinical variables

For each hospital encounter in the final analysis cohort, we extracted time stamped clinical data on demographics, vital signs, medications, laboratory values, past medical diagnoses, and admission diagnoses. Details of the 1888 clinical variables considered are available in Table 1. Patient vital signs were discretized into groups as shown in Supplementary Table S2. Laboratory

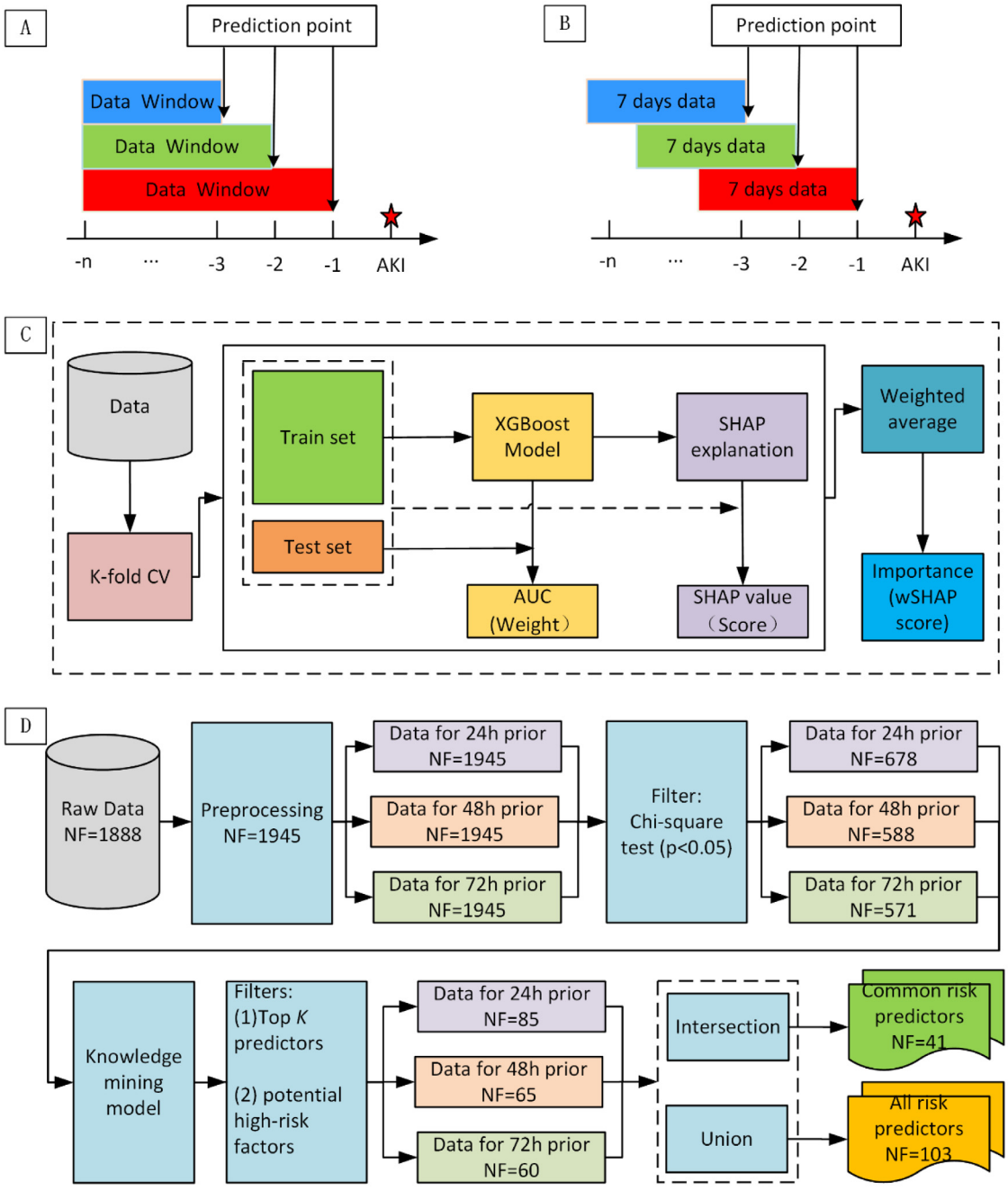

Fig. 1. Schematic diagram of data analysis and processing. (A) An illustration of data collection window to predict AKI event 1–3 days prior; (B) An illustration of medication data collection window to predict AKI event 1–3 days prior; (C) The knowledge mining model; (D) The flowchart of screening AKI risk factors. AKI, acute kidney injury; NF, the number of features.

values were categorized as unknown, less than reference normal range, within normal range, or greater than the reference normal range. Medication exposure included inpatient (i.e., dispensed during stay) and outpatient medications (i.e., medication reconciliation and prior outpatient prescriptions). All medication names were normalized by mapping to RxNorm ingredient. Admission diagnosis, i.e., all patient refined diagnosis related group (APR-DRG), were collected from the University Health System Consortium (UHC; http://www.vizientinc.com) data source in HERON (Health Enterprise Repository for Ontological Narration). Patient past medical history was captured as major diagnoses (ICD-9 codes grouped according to the Clinical Classifications Software (CCS) diagnosis categories by the Agency for Healthcare Research and Quality.

2.4. Data processing

Only the most recently recorded vitals and labs before the AKI prediction point (i.e., 24/48/72 hours prior to AKI event or last normal SCr for non-AKI cases, see Fig. 1(A)) were used for each encounter over a specified time-window. Meanwhile, instead of any imputation of missing numerical values, missing values were captured as a separate category because information may be contained in the choice to not perform a particular test [23]. To enhance interpretability, the multivalued variables are converted into multiple binary variables. Medical history was defined as true if it happened before the AKI prediction point. Hence, Medical history and admission diagnosis were all binary variables (i.e., presence or absence). Medication variables were

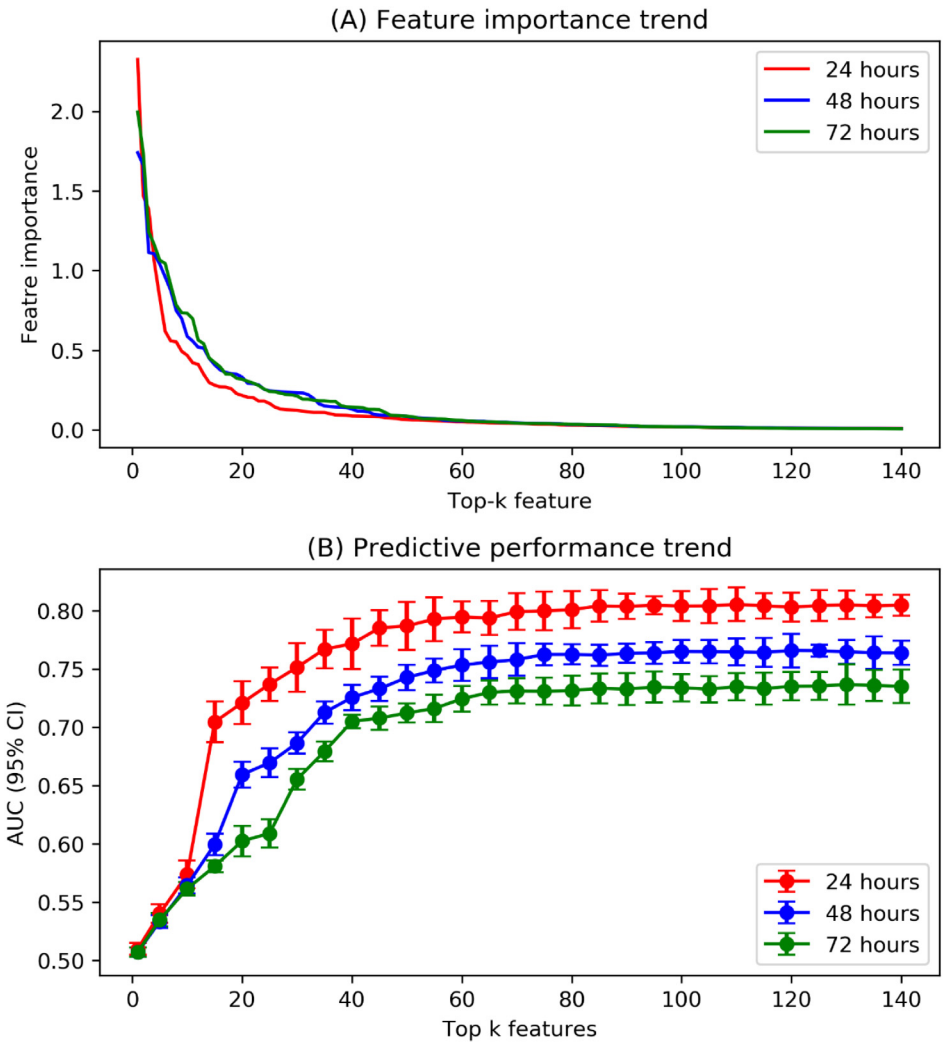

Fig. 2. The feature importance and predictive performance trend of top-ranking features for AKI events 24/48/72 hours prior. (A) The importance trend of the top-ranking features obtained by the knowledge mining approach; (B) The performance trend of the XGBoost prediction model with top K features. AUC, the area under the receiver operating characteristic curve; CI, confidence interval.

defined as the number of days taken within 7-days before the AKI prediction point, as shown in Fig. 1(B).

2.5. Knowledge mining model

As shown in Fig. 1(C), in the knowledge mining model, we applied two machine learning methods: eXtreme Gradient Boosting (XGBoost) [24] and SHapley Additive exPlanations (SHAP) method [25,26] (see Supplementary Methods S1-S2). XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the gradient boosting framework, and provides a parallel tree boosting that solve many data science problems in a fast and accurate way. Inspired by cooperative game theory, SHAP constructs an additive explanation model, with all the characteristics considered as "contributors". SHAP values can explain the output of any machine learning model, in this study, we applied the tree SHAP method, which is a high-speed exact explanation algorithm for tree ensemble methods.

Meanwhile, in order to reduce the influence of sample differentiation and enhance the stability and effectiveness of knowledge mining, cross-validation strategy was adopted to introduce this data drift [27] and obtained the weighted SHAP value ($wSHAP$), where the weight was the area under the receiver operating characteristic curve (AUC) of the XGBoost model in each fold and the risk score for each variable was derived from the SHAP interpretation using the entire dataset. In addition, to justify the choice of XGBoost method, we compared XGBoost against other 6 machine learning algorithms: GBM [14], Random Forest [28], LinearSVC (linear support vector classification) [29], Logistic Regression, Naive Bayes, MLP (multi-layer perceptron neural network) [30].

2.6. Risk predictor screening process and statistical analysis

To screen for risk predictor, as shown in Fig. 1(D), we explored three prediction models at different time-windows (namely, 24 h, 48 h and 72 h prior to onset) and implemented three filters in sequence: a) Chi-square test, only keeping variables that have a statistically significant ($p < 0.05$) univariate correlation with AKI label; b) top K predictors, obtaining the best top K according to the growth trend of AUC with top k; c) potentially high risk predictors, screening factors with the mean weighted SHAP greater than 0, which will filter out some risk-free but highly predictive factors, such as normal BMI (body mass index), the standard WBC (white blood cell) lab test, and *dexamethasone* for treating kidney

Table 3
The 41 common risk factors under AKI 24/48/72-hour time windows.

Name	Samples with the risk factor, n (%)		Median [IQR] of the weighted SHAP values		
	non-AKI	AKI	72 h prior	48 h prior	24 h prior
age_6['>64']	18596 (26.68)	2193 (30.21)	0.049 [0.038–0.062]	0.046 [0.035–0.053]	0.021 [0.017–0.028]
BMI_4['>30.0 obese']	25347 (36.37)	3095 (42.64)	0.073 [0.061–0.081]	0.068 [0.056–0.079]	0.043 [0.031–0.056]
Pulse_1['<50']	346 (0.50)	80 (1.10)	0.084 [0.062–0.176]	0.089 [0.07–0.134]	0.038 [0.032–0.065]
Temperature_1['<95.0 Hypothermia']	67 (0.10)	36 (0.50)	0.281 [0.229–0.37]	0.329 [0.251–0.438]	0.185 [0.14–0.234]
Temperature_4['99.5–104.0 Fever']	1217 (1.75)	252 (3.47)	0.119 [0.092–0.149]	0.126 [0.102–0.158]	0.095 [0.068–0.134]
Lab0_1['Albumin, less than standard value']	31637 (45.39)	4143 (57.07)	0.084 [0.069–0.108]	0.069 [0.057–0.087]	0.061 [0.046–0.078]
Lab5_3['BUN, more than the standard value']	6563 (9.42)	1290 (17.77)	0.174 [0.151–0.203]	0.236 [0.206–0.271]	0.261 [0.205–0.323]
Lab6_3['Bilirubin, more than the standard value']	2952 (4.24)	708 (9.75)	0.027 [0.022–0.052]	0.077 [0.056–0.114]	0.205 [0.168–0.252]
DRG0["liver transplant"]	147 (0.21)	157 (2.16)	1.518 [0.887–2.742]	1.380 [0.891–2.538]	1.099 [0.799–1.99]
DRG131["hip joint replacement"]	1565 (2.25)	116 (1.60)	0.116 [0.088–0.157]	0.249 [0.228–0.284]	0.082 [0.073–0.097]
DRG177["major bladder proc"]	242 (0.35)	42 (0.58)	0.211 [0.126–0.287]	0.202 [0.105–0.324]	0.047 [0.035–0.06]
DRG178["kidney/urinary tract malig"]	264 (0.38)	60 (0.83)	0.327 [0.242–0.482]	0.239 [0.182–0.392]	0.255 [0.198–0.355]
DRG2["bone marrow transplant"]	400 (0.57)	124 (1.71)	0.405 [0.292–0.542]	0.345 [0.261–0.491]	0.394 [0.278–0.514]
DRG256["acute leukemia"]	242 (0.35)	63 (0.87)	0.091 [0.07–0.129]	0.024 [0.02–0.058]	0.099 [0.061–0.119]
DRG261["infect & parasitic disease"]	322 (0.46)	147 (2.03)	0.863 [0.751–1.016]	0.720 [0.584–0.901]	0.487 [0.381–0.588]
DRG262["post-op/trauma infect proc"]	378 (0.54)	85 (1.17)	0.433 [0.332–0.494]	0.251 [0.197–0.309]	0.164 [0.107–0.214]
DRG263["septicemia & dissem infect"]	2420 (3.47)	425 (5.85)	0.163 [0.121–0.23]	0.058 [0.051–0.08]	0.011 [0.009–0.013]
DRG3["trach w/dmv w exten proc"]	80 (0.11)	112 (1.54)	2.069 [1.243–2.777]	1.821 [1.107–2.39]	1.408 [1.034–1.807]
DRG301["HIV w multiple major HIV"]	73 (0.10)	28 (0.39)	0.427 [0.308–0.505]	0.239 [0.185–0.279]	0.083 [0.058–0.126]
DRG309["extensive proc unrel pdx"]	291 (0.42)	69 (0.95)	0.235 [0.166–0.373]	0.122 [0.103–0.227]	0.096 [0.074–0.13]
DRG4["trach w/dmv w/o enten proc"]	50 (0.07)	30 (0.41)	0.592 [0.416–0.967]	0.306 [0.199–0.701]	0.189 [0.124–0.275]
DRG49["respiratory system diag"]	39 (0.06)	34 (0.47)	0.743 [0.544–0.926]	0.504 [0.403–0.73]	0.205 [0.138–0.307]
DRG66["cardiac valve proc w/cath"]	137 (0.20)	68 (0.94)	0.858 [0.715–1.931]	0.757 [0.645–1.576]	0.501 [0.479–0.651]
DRG67["cardiac valve proc w/0 cath"]	493 (0.71)	138 (1.90)	0.817 [0.752–1.175]	0.655 [0.603–0.961]	0.33 [0.326–0.336]
DRG68["coronary bypass w/card cath"]	437 (0.63)	163 (2.25)	0.649 [0.579–1.436]	0.768 [0.705–1.267]	0.674 [0.646–0.848]
DRG69["cor bypass w/o card cath"]	430 (0.62)	162 (2.23)	0.804 [0.733–1.362]	0.673 [0.589–1.095]	0.454 [0.445–0.551]
DRG84["heart failure"]	544 (0.78)	149 (2.05)	0.157 [0.131–0.21]	0.059 [0.047–0.08]	0.061 [0.051–0.082]
DRG97["maj small & large bowel proc"]	972 (1.39)	172 (2.37)	0.600 [0.526–0.706]	0.565 [0.509–0.682]	0.461 [0.406–0.572]
CCS118['Congestive heart failure; non-hypertensive']	3572 (5.12)	725 (9.99)	0.160 [0.135–0.198]	0.143 [0.111–0.181]	0.133 [0.106–0.168]
CCS135['Cancer of kidney and renal pelvis']	545 (0.78)	160 (2.20)	0.222 [0.171–0.324]	0.174 [0.141–0.272]	0.023 [0.013–0.037]

(*continued on next page*)

Table 3 (*continued*).

Name	Samples with the risk factor, n (%)		Median [IQR] of the weighted SHAP values		
	non-AKI	AKI	72 h prior	48 h prior	24 h prior
CCS230['Acute and unspecified renal failure']	3803 (5.46)	683 (9.41)	0.120 [0.096–0.149]	0.103 [0.082–0.131]	0.108 [0.082–0.134]
CCS58['Cystic fibrosis']	979 (1.40)	223 (3.07)	0.314 [0.256–0.411]	0.307 [0.25–0.389]	0.452 [0.397–0.521]
MED1086['tazobactam']	11385 (16.33)	2276 (31.35)	0.196 [0.151–0.266]	0.302 [0.235–0.434]	0.285 [0.201–0.478]
MED1096['posaconazole']	767 (1.10)	183 (2.52)	0.072 [0.048–0.097]	0.071 [0.042–0.101]	0.081 [0.063–0.105]
MED134['benzoic acid']	17751 (25.47)	3075 (42.36)	0.164 [0.115–0.194]	0.215 [0.172–0.244]	0.256 [0.199–0.315]
MED256['fludarabine']	117 (0.17)	56 (0.77)	0.086 [0.019–0.11]	0.055 [0.039–0.075]	0.124 [0.094–0.149]
MED319['amphotericin b liposome']	104 (0.15)	90 (1.24)	0.746 [0.584–0.973]	1.061 [0.855–1.358]	1.074 [0.814–1.256]
MED321['vancomycin']	14266 (20.47)	2403 (33.10)	0.027 [0.018–0.05]	0.069 [0.045–0.127]	0.120 [0.077–0.224]
MED478['dornase alfa']	994 (1.43)	222 (3.06)	0.036 [0.016–0.06]	0.050 [0.027–0.108]	0.038 [0.024–0.103]
MED572['diazepam']	419 (0.60)	133 (1.83)	0.067 [0.048–0.086]	0.146 [0.086–0.192]	0.174 [0.122–0.232]
MED733['glipizide']	1295 (1.86)	216 (2.98)	0.049 [0.025–0.077]	0.037 [0.001–0.058]	0.012 [0.008–0.024]

Abbreviations: AKI = acute kidney injury; non-AKI = not acute kidney injury; Median [IQR] = median value [interquartile range].

disease. Finally, we took the intersection and union of the filtered AKI predictors from the three prediction time windows to explore temporal fluctuation of the importance/weight of AKI risk factors.

DeLong test [31] was used to compare the areas under two or more correlated receiver operating characteristic curves (ROC) and calculate the statistical significance of AUC improvement. Kolmogorov–Smirnov (KS) test [32] is a form of minimum distance estimation used as a nonparametric test (it does not assume any particular underlying distribution) to check if two independent distributions are similar or different. We applied KS test to compare whether the fluctuations in risk weight/importance are statistically significant. In addition, the network was applied to clarify the co-occurrence relationship of AKI risk factors in three prediction time-windows, where the phi (Φ) coefficient was used to measure the degree of association between two binary variables. Two-tailed $p < 0.05$ denoted statistical significance for all comparisons. Data extraction and all analyses were performed using Python 3.7 software.

3. Results

3.1. Screening for risk factors

Among the final analysis cohort of 76,957 hospital admissions, AKI occurred in 7259 (9.43%) encounters. Distribution of patient demographic variables among any AKI stage and non-AKI encounters is listed in Supplementary Table S3. The detailed risk factor screening process is shown in Fig. 1(D), from which we can see that the first univariable feature selection (i.e., Chi-square test) screened out 678, 588 and 571 factors under the 24 h, 48 h and 72 h prediction windows, respectively.

Table 2 compares results of 7 machine learning methods, among which the tree-based ensemble models performed best, and there is no significant difference between XGBoost and Random Forest models (i.e., $p \geq 0.05$ of DeLong test). The best cross-validated AUC achieved by the XGBoost method (parameters used are available in Supplementary Table S4) were 0.805 (95% confidence interval (CI) 0.791-0.824) for 24-h prior, 0.764 (95% CI, 0.750-0.775) for 48-h prior, and 0.736 (95% CI, 0.723-0.745) for 72-h prior. Supplementary Figure S2 shows the ROC curves of the 7 machine learning methods for AKI prediction 48 h prior.

There is an exponential decline trend in the importance (i.e., $|wSHAP|$) of clinical AKI predictors (see Fig. 2(A)). And Fig. 2(B) shows the performance trend of the XGBoost prediction model with top $K(1 \leq K \leq 140)$ features. We set 0.0001 as the threshold of ΔAUC, which means that the AUC growth tends to be stable, and the optimal top K for 24 h, 48 h and 72 h time windows were 102, 83 and 83, respectively. Further, we used the $mean(wSHAP) > 0$ as the restriction threshold to obtain potentially high-risk factors (resulted in 85, 65 and 60 features for 24/48/72-h prior respectively, see Supplementary Tables S5-S7). In Supplementary Figure S4, there are only 41 intersection factors among 103 union risk factors.

3.2. Temporal fluctuation of risk factors

Table 3 shows the fluctuations of the 41 intersection risk factors of AKI from the 24-h, 48-h and 72-h time windows. Fig. 3 illustrates the weighted SHAP summary plot of 41 intersection risk factors for AKI in the 48-h time window. Supplementary Figure S4 and Figure S5 show the importance fluctuations of $mean\ (wSHAP)$ of 41 intersection and 103 union risk factors, respectively. In addition, in order to show the temporal fluctuation of the importance of AKI risk factors more clearly, we normalized the $mean\ (wSHAP)$ of each feature to [0,1]. Fig. 4 shows the heat map of the normalized importance of the 103 union risk factors, where the default weight of unfiltered factors is 0.

From the above results, almost all time-varying variables have higher $wSHAP$ at 24-hr and the non-time-varying variables (e.g., admission DRGs) actually have higher $wSHAP$ at 72-hr mark, suggesting that the contribution of those time-invariant predictors to AKI decreases as time progresses, which is consistent with human intuition. The significance of the importance distribution differences based on the Kolmogorov–Smirnov test is shown in Supplementary Figure S6, from which almost all the temporal fluctuations are statistically significant.

Fig. 5 presents the correlation network of AKI risk factors under 24-, 48- and 72-h time windows, where if two nodes

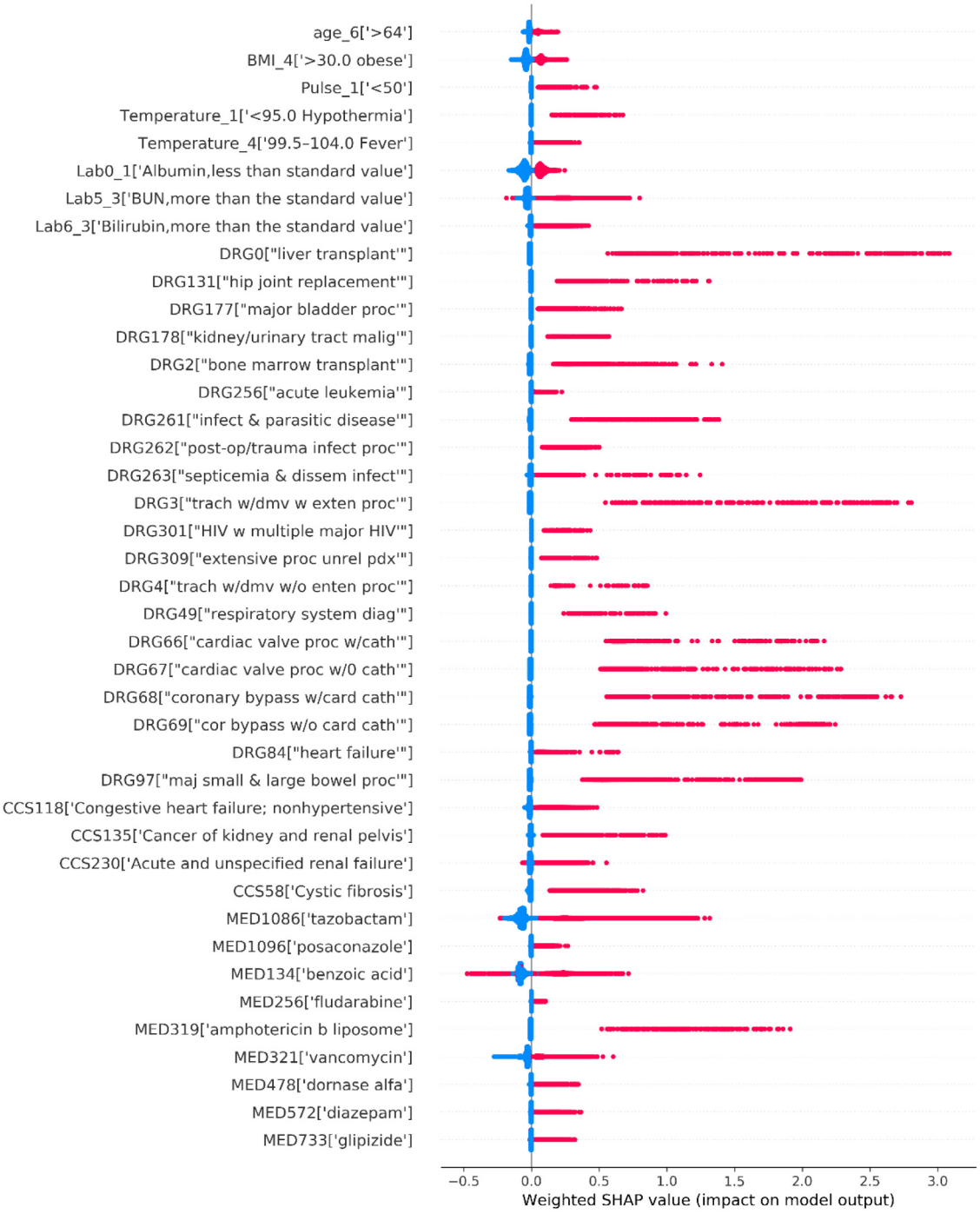

Fig. 3. The weighted SHAP summary plot of 41 common risk factors for AKI in the 48-h time window. (The higher the weighted SHAP value of a feature, the higher risk of AKI due to this feature. The point on each feature line represents a sample with that feature. Dots are colored by the feature value for that person and pile up vertically to show density. For the binary (0/1) variables, the red value is 1, that is, the variable exists. The blue value is 0, that is, this variable does not exist. For the drug variables, red means the drug has been taken within the past week, blue means not taking it.)

(factors) exhibit a certain degree of clinical Φ-correlation, they are connected by an edge. For better visualization, we made a tradeoff between the number of associations included in a network and the clarity with which these associations can be appreciated; for example, a cutoff value $\Phi^* = 0.03$ was specified so that only links satisfying $\Phi \geq \Phi^*$ were kept. As can be seen from Fig. 5, the correlation between risk factors increases as time approaches AKI onset. As an illustration, Supplementary Table S8 shows the top 5 positive and negative influencing factors of one AKI patient (male, randomly selected from the AKI sample) in the 24/48/72 h prediction time window, where the *wSHAP* values of *tazobactam* were 0.154, 0.307 and 0.440 for 72-, 48- and 24-h time windows, respectively. In addition, Fig. 6 illustrates the differences in the distributions of four AKI risk factors (namely, age_6['>64'], BMI_4['>30.0 obese'], Lab5_3['BUN, more than the standard value'], and MED321['vancomycin']) under 24, 48, and 72-h time windows.

4. Discussion

Currently international guidelines recommend risk assessment for AKI for the purpose of preventing kidney injury progression

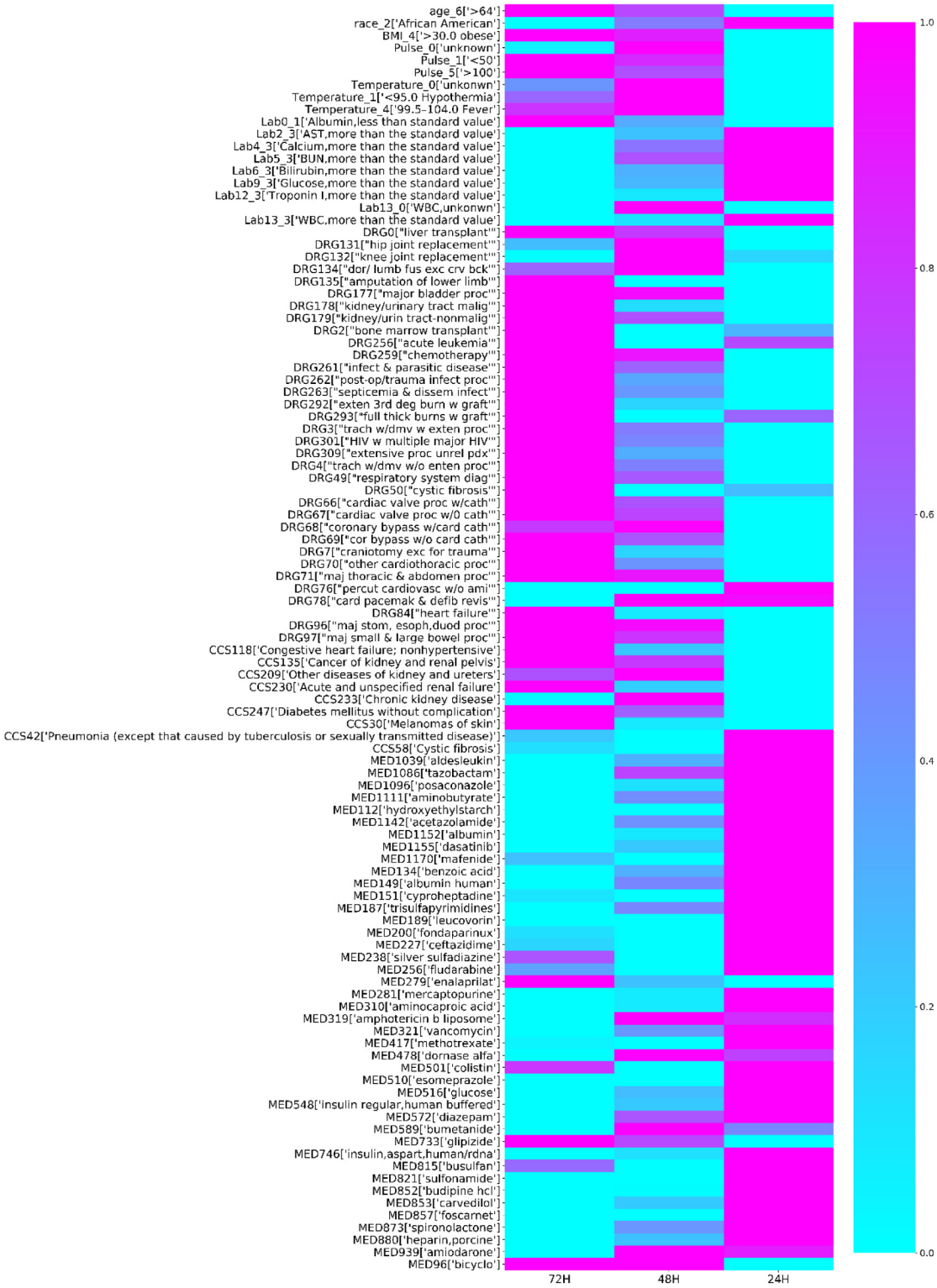

Fig. 4. Heatmap of the normalized importance of the 103 union risk factors for AKI events prior to 24, 48 and 72 h. (To display the temporal fluctuation of feature importance more clearly, we normalized the importance of each feature to [0,1], where the default weight of unfiltered factors is 0.)

and severity [33]. And the 15th Acute Dialysis Quality Initiative (ADQI) Consensus Group recommended AKI prediction with a lead time of 48 to 72 h to be the most useful in practice as the group believed that it would give practitioners adequate time to respond to mitigate potential injury without sacrificing prediction accuracy [34]. In this study, we developed a machine-learning-based knowledge mining approach, combining XGBoost model and SHAP explainer, to explore the temporal dynamics of clinical risk predictors for hospital-acquired AKI under different forecast time windows.

Model transparency is critical for many clinical applications and will accelerate the widespread adoption of such approaches in clinical practice. There are a large number of interpretable risk-scoring models. However, the construction of these transparent

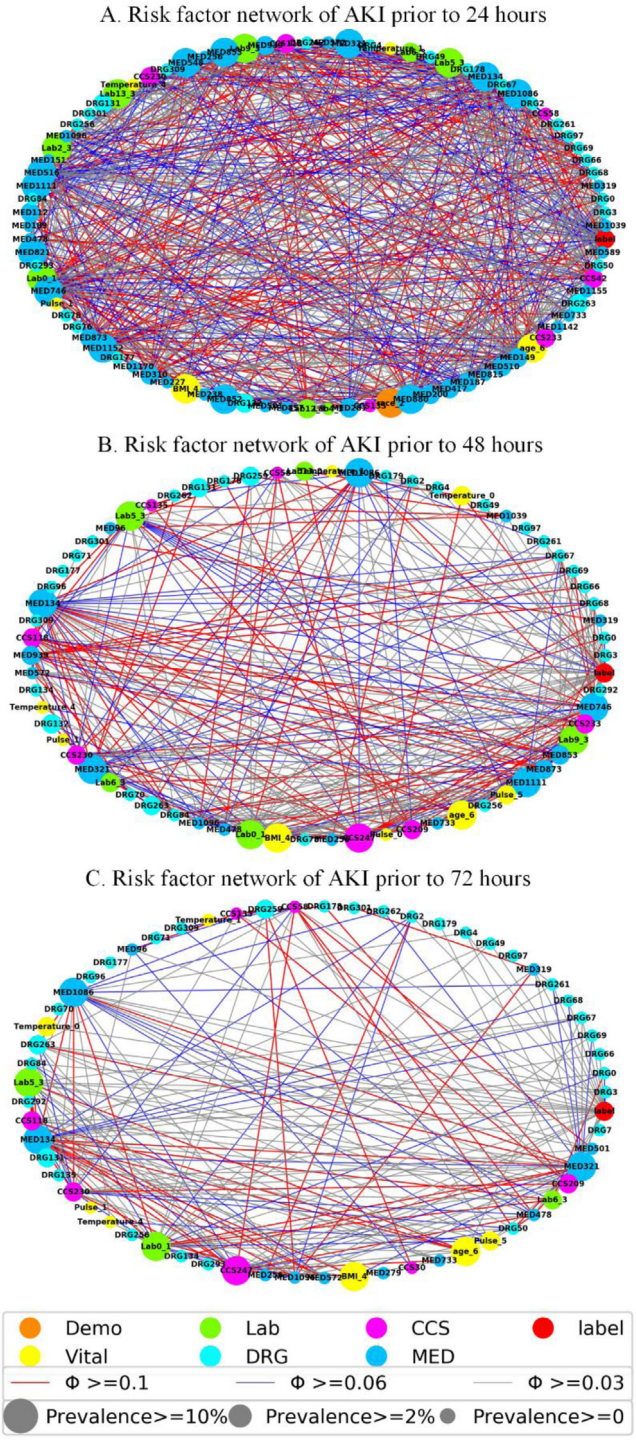

Fig. 5. Clinical risk feature networks of AKI prior to 24/48/72 hours. (Nodes are features; links are correlations; node color identifies feature categories; node size is proportional to the prevalence; link color indicates Φ correlation strength)

models is not only based on expert knowledge to select a small number of known risk factors, but is also limited to a relatively small specific scenario, such as AKI following elective cardiac surgery [10], after aortic surgery [35], sepsis associated AKI [36] and contrast-induced AKI [37]. More importantly, these traditional scoring models do not consider the temporal volatility of risk, which will reduce model accuracy. Some researchers [38,39] focused only on improving predictive performance at the expense of model interpretability (e.g., complex machine learning or deep learning methods).

Importantly, we implemented a trade-off between accuracy and interpretability [40] by introducing a SHAP interpreter to explain the XGBoost model. The SHAP method can be viewed as a personalized risk scoring system (see Supplementary Table S8). In addition, the explanations for the AKI risk are broadly consistent with the literature and with prior knowledge from nephrologists.

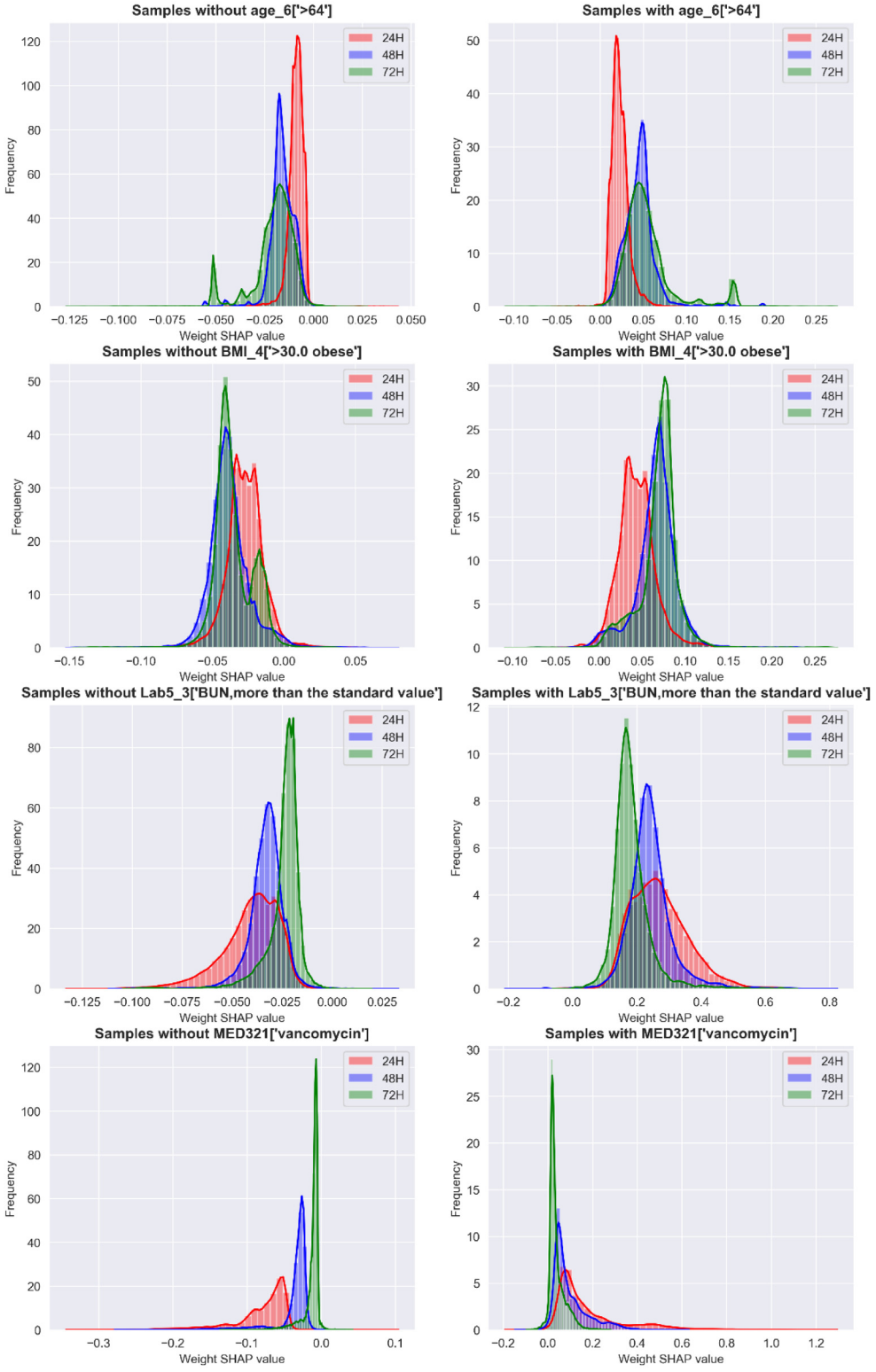

Fig. 6. Distribution charts of the weighted SHAP values for samples without/with some AKI risk factors (age_6['>64'], BMI_4['>30.0 obese'], Lab5_3['BUN, more than the standard value'], and MED321['vancomycin']) under 24, 48, and 72-h time windows. (BMI, body mass index; BUN, blood urea nitrogen.).

For example, as shown in Fig. 3, procedures related to heart, trachea, and liver diseases are associated with higher AKI risks, that is because they are associated with reduced kidney reserve or failure of other organs with known cross-talk with the kidneys (e.g., heart, liver, and respiratory system) [41]. Meanwhile, this study's main focus is to mine the temporal fluctuation of AKI risk predictors, rather than simply build an AKI prediction model. Furthermore, what we analyze was the difference in the importance of the same feature in different prediction time windows, rather than the temporal change of feature values.

It is worth noting that optimal risk factors used to construct AKI prediction models will differ under different prediction time

windows (see Supplementary Figure S3 and Tables S5–S7). Fig. 4 demonstrated that the contribution of most time-varying variables (such as medication and lab tests) to AKI risk increases as AKI onset approaches, while most non-time-varying variables (such as age and admission diagnosis) decrease, suggesting that static risk factors would be more advantageous in the longer time window prediction. From the clinical perspective, the static predictor is more likely to provide a baseline risk estimation and the dynamic predictor is more likely to be a superimposed risk on the baseline risk. Static predictors may be equally important in each time window, but what may differ is their relative importance in each forecast time window.

Temporal information in longitudinal EMR is valuable, however, one of the challenges in improving AKI prediction performance is how to consider temporality when using time-stamped clinical events in EMR. Previous research [42] on this topic suggested that representing EMR data as a bag of temporally weighted clinical events is promising; however, how to assign weights in an optimal manner remains unexplored. In this paper, our results showed that it is not appropriate to consider only one weighting strategy, because the fluctuation trends of time-invariant and time-variant characteristics are different (see Table 2, Figs. 4 and 6). Besides, we hypothesized that topological analysis of clinical variables in high-dimensional clinical EMR feature space may identify meaningful knowledge discovery of AKI patients. As shown in Fig. 5, with the narrowing of the prediction time window, that is, the closer AKI onset is, the stronger the correlation between risk predictors is. In the future, knowledge graph may help us further analyze the risk factor network of AKI to improve prevention and intervention levels.

Several limitations in the present research must be considered. First, although we utilized a large cohort observed for up to a decade, they only reflect the population of one academic medical center. Replicating this study in other institutions would generalize conclusions. Second, we limited the analysis to patients with a minimum eGFR (estimated glomerular filtration rate) of 60 ml/min/1.73 m^2 and normal serum creatinine on the day of admission at hospital admission. We acknowledge that patients with reduced eGFR have an increased risk of developing AKI; however, we made the decision to focus on hospital-acquired AKI. Third, this study explored the entirety of the above mentioned EMR data types except for laboratory tests where we only selected certain lab tests based on previous literature for AKI prediction [23]. Finally, because urine output can be influenced by factors other than kidney health, and it is not frequently collected among the general inpatient population, we did not include urine output criteria as a predictor nor using it to define AKI.

5. Conclusions

Large-scale EMR offer unique opportunities and possibilities for generating real-world clinical evidence and actionable insights to transform healthcare. In this study, we used 9 years of "big" EMR data, including 76,957 encounters, and established a machine-learning-based knowledge mining approach to explore the temporal dynamics of clinical risk predictors for hospital-acquired AKI under different forecast time windows. Results of this study illustrated the existence of fluctuations in the risk importance of features, which need to be considered for accurate modeling, and identify important AKI risk factors under different time windows for further research to improve the accuracy of early AKI prediction.

CRediT authorship contribution statement

Lijuan Wu: Conceptualization, Methodology, Software, Writing – original draft. **Yong Hu:** Supervision, Resources, Writing – review & editing. **Xiangzhou Zhang:** Formal analysis, Project administration. **Borong Yuan:** Visualization. **Weiqi Chen:** Data curation, Software. **Kang Liu:** Data curation. **Mei Liu:** Supervision, Resources, Writing – review & editing.

Declaration of competing interest

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

Acknowledgments

This research was partially supported by the Youth Science Fund of the National Natural Science Foundation of China (Grant No. 61802149), the Major Research Plan of the National Natural Science Foundation of China (Key Program, Grant No. 91746204), the Guangzhou Basic and Applied Basic Research Project (Grant No. 202102020863), the Science and Technology Development in Guangdong Province (Major Projects of Advanced and Key Techniques Innovation, Grant No. 2017B030308008), Guangdong Engineering Technology Research Center for Big Data Precision Healthcare (Grant No. 603141789047), the Fundamental Research Funds for the Central Universities (Grant No. 21618315), the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (Grant No. R01DK116986), and the National Science Foundation (Grant No. 2014554). The content is solely the responsibility of the authors and does not represent the official views of the funders. The clinical dataset used for analysis described in this study was obtained from KUMC's HERON clinical data repository which is supported by institutional funding and by the KUMC CTSA grant UL1TR002366 from National Institutes of Health.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.knosys.2022.108655.

References

[1] J.A. Kellum, N. Lameire, Diagnosis, evaluation, and management of acute kidney injury : a KDIGO summary (part 1), Crit. Care 17 (2013) 204.
[2] J. Himmelfarb, M. Joannidis, B. Molitoris, M. Schietz, M.D. Okusa, D. Warnock, F. Laghi, S.L. Goldstein, R. Prielipp, C.R. Parikh, N. Pannu, S.M. Lobo, S. Shah, V. D'Intini, J.A. Kellum, Evaluation and initial management of acute kidney injury, Clin. J. Am. Soc. Nephrol. 3 (2008) 962–967, http://dx.doi.org/10.2215/CJN.04971107.
[3] L. Chan, K. Chaudhary, A. Saha, K. Chauhan, A. Vaid, S. Zhao, I. Paranjpe, S. Somani, F. Richter, R. Miotto, et al., AKI in hospitalized patients with COVID-19, J. Am. Soc. Nephrol. 32 (2021) 151–160.
[4] J.L. Koyner, K.A. Carey, D.P. Edelson, M.M. Churpek, The development of a machine learning inpatient acute kidney injury prediction model, Crit. Care Med. 46 (2018) 1070–1077.
[5] R.A. Taylor, J.R. Pare, A.K. Venkatesh, H. Mowafi, E.R. Melnick, W. Fleischman, M.K. Hall, Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data-driven, machine learning approach, Acad. Emerg. Med. 23 (2016) 269–278, http://dx.doi.org/10.1111/acem.12876.
[6] B.A. Goldstein, A.M. Navar, M.J. Pencina, J.P.A. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, J. Am. Med. Inf. Assoc. 24 (2017) 198–208, http://dx.doi.org/10.1093/jamia/ocw042.

[7] S.L. Kane-Gill, F.E. Sileanu, R. Murugan, G.S. Trietley, S.M. Handler, J.A. Kellum, Risk factors for acute kidney injury in older adults with critical illness: A retrospective cohort study, Am. J. Kidney Dis. 65 (2015) 860–869.
[8] G.M. McMahon, X. Zeng, S.S. Waikar, A risk prediction score for kidney failure or mortality in rhabdomyolysis, JAMA Int. Med. 173 (2013) 1821–1827.
[9] Y. Zongyi, L. Baifeng, Z. Funian, L. Hao, W. Xin, Risk factors of acute kidney injury after orthotopic liver transplantation in China, Sci. Rep. 7 (2017) 1–11, http://dx.doi.org/10.1038/srep41555.
[10] H. Palomba, I. De Castro, A.L.C. Neto, S. Lage, L. Yu, Acute kidney injury prediction following elective cardiac surgery: AKICS score, Kidney Int. 72 (2007) 624–631.
[11] N. Lueangingkhasut, P. Khumjai, C. Soontornpas, Development of risk score model for hospital-acquired acute kidney injury in a tertiary care hospital in Thailand, Indian J. Pharm. Sci. 16 (2020) 37–48.
[12] M. Flechet, F. Güiza, M. Schetz, P. Wouters, I. Vanhorebeek, I. Derese, J. Gunst, I. Spriet, M. Casaer, G. Van den Berghe, G. Meyfroidt, Akipredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin, Intens. Care Med. 43 (2017) 764–773, http://dx.doi.org/10.1007/s00134-017-4678-3.
[13] M.H. Park, H.S. Shim, W.H. Kim, H.-J. Kim, D.J. Kim, S.-H. Lee, C.S. Kim, M.S. Gwak, G.S. Kim, Clinical risk scoring models for prediction of acute kidney injury after living donor liver transplantation: A retrospective observational study, PLoS One 10 (2015) e0136230, http://dx.doi.org/10.1371/journal.pone.0136230.
[14] J.L. Koyner, K.A. Carey, D.P. Edelson, M.M. Churpek, The development of a machine learning inpatient acute kidney injury prediction model, Crit. Care Med. 46 (2018) 1070–1077.
[15] Y. Li, L. Yao, C. Mao, A. Srivastava, X. Jiang, Y. Luo, Early prediction of acute kidney injury in critical care setting using clinical notes, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2018, pp. 683–686.
[16] N. Tomašev, X. Glorot, J.W. Rae, M. Zielinski, H. Askham, A. Saraiva, A. Mottram, C. Meyer, S. Ravuri, I. Protsyuk, et al., A clinically applicable approach to continuous prediction of future acute kidney injury, Nature 572 (2019) 116–119.
[17] M. Simonov, U. Ugwuowo, E. Moreira, Y. Yamamoto, A. Biswas, M. Martin, J. Testani, F.P. Wilson, A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: a descriptive modeling study, PloS Med. 16 (2019) e1002861.
[18] S. Nemati, A. Holder, F. Razmi, M.D. Stanley, G.D. Clifford, T.G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the ICU, Crit. Care Med. 46 (2018) 547.
[19] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?: Explaining the predictions of any classifier, in: Proceedings Ofthe 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[20] S. Lundberg, G.G. Erion, H. Chen, A.J. Degrave, J.M. Prutkin, B.G. Nair, R. Katz, J. Himmelfarb, N. Bansal, S. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (2020) 56–67.
[21] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D.E. Liston, D.K.-W. Low, S.-F. Newman, J. Kim, et al., Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, Nat. Biomed. Eng. 2 (2018) 749–760.
[22] D. Fliser, M. Laville, A. Covic, D. Fouque, R. Vanholder, L. Juillard, W. Van Biesen, A European renal best practice (ERBP) position statement on the kidney disease improving global outcomes (KDIGO) clinical practice guidelines on acute kidney injury: Part 1: Definitions, conservative management and contrast-induced nephropathy, Nephrol. Dial. Transplant. 27 (2012) 4263–4272.
[23] M.E. Matheny, R.A. Miller, T.A. Ikizler, L.R. Waitman, J.C. Denny, J.S. Schildcrout, R.S. Dittus, J.F. Peterson, Development of inpatient risk stratification models of acute kidney injury for use in electronic health records., Med. Decis. Mak. 30 (2010) 639–650.
[24] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[25] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, 2018, arXiv preprint arXiv:180203888.
[26] S.M. Lundberg, S. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.
[27] J.G. Moreno-Torres, J.A. Saez, F. Herrera, Study on the impact of partition-induced dataset shift on k-fold cross-validation, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1304–1312, http://dx.doi.org/10.1109/TNNLS.2012.2199516.
[28] K. Lin, Y. Hu, G. Kong, Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model, Int. J. Med. Inf. 125 (2019) 55–61.
[29] R.J. Kate, R.M. Perez, D. Mazumdar, K.S. Pasupathy, V. Nilakantan, Prediction and detection models for acute kidney injury in hospitalized older adults, BMC Med. Inf. Decis. Mak. 16 (2016) 39.
[30] H.H. Rashidi, S. Sen, T.L. Palmieri, T. Blackmon, J. Wajda, N.K. Tran, Early recognition of burn-and trauma-related acute kidney injury: a pilot comparison of machine learning techniques, Sci. Rep. 10 (2020) 205.
[31] X. Sun, W. Xu, Fast implementation of Delong's algorithm for comparing the areas under correlated receiver operating characteristic curves, IEEE Signal Process. Lett. 21 (2014) 1389–1393.
[32] F.J. Massey Jr., The Kolmogorov–Smirnov test for goodness of fit, J. Am. Stat. Assoc. 46 (1951) 68–78.
[33] A. Khwaja, KDIGO clinical practice guidelines for acute kidney injury, Nephron Clin. Pract. 120 (2012) c179–c184.
[34] S.M. Sutherland, L.S. Chawla, S.L. Kane-Gill, R.K. Hsu, A.A. Kramer, S.L. Goldstein, J.A. Kellum, C. Ronco, S.M. Bagshaw, Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15 th ADQI consensus conference, Canadian J. Kidney Health Dis. 3 (2016) 11.
[35] W.H. Kim, S.M. Lee, J.W. Choi, E.H. Kim, J.H. Lee, J.W. Jung, J.H. Ahn, K.I. Sung, C.S. Kim, H.S. Cho, Simplified clinical risk score to predict acute kidney injury after aortic surgery, J. Cardiothoracic Vasc. Anesth. 27 (2013) 1158–1166.
[36] J. Zhou, Y. Bai, X. Wang, J. Yang, P. Fu, D. Cai, L. Yang, A simple risk score for prediction of sepsis associated-acute kidney injury in critically ill patients, J. Nephrol. 32 (2019) 947–956.
[37] C. Duan, Y. Cao, Y. Liu, L. Zhou, K. Ping, M.T. Tan, N. Tan, J. Chen, P. Chen, A new preprocedure risk score for predicting contrast-induced acute kidney injury, Canadian J. Cardiol. 33 (2017) 714–723.
[38] Y. Cheng, F. Wang, P. Zhang, J. Hu, Risk prediction with electronic health records : A deep learning approach, in: Siam International Conference on Data Mining, 2016, pp. 432–440.
[39] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, IEEE J. Biomed. Health Inf. 22 (2017) 1589–1604.
[40] D. Gunning, D. Aha, DARPA's explainable artificial intelligence (XAI) program, AI Mag. 40 (2019) 44–58.
[41] M. Leblanc, J.A. Kellum, R.T.N. Gibney, W. Lieberthal, J. Tumlin, R. Mehta, Risk factors for acute renal failure: inherent and modifiable risks, Curr. Opin. Crit. Care 11 (2005) 533–536.
[42] J. Zhao, Temporal weighting of clinical events in electronic health records for pharmacovigilance, in: Proceedings - 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015, 2015, pp. 375–381.