MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190245

Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements

Mengxin Sun[a], Jason Baron[b], Anand Dighe[b], Peter Szolovits[c], Richard G. Wunderink[d], Tamara Isakova[d], Yuan Luo[d]

[a] *Independent Researcher, Chicago, IL, USA*
[b] *Department of Pathology, Massachusetts General Hospital, Boston, MA, USA*
[c] *Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA*
[d] *Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*

Abstract

The onset of acute kidney injury (AKI) during an intensive care unit (ICU) admission is associated with increased morbidity and mortality. Developing novel methods to identify early AKI onset is of critical importance in preventing or reducing AKI complications. We built and applied multiple machine learning models to integrate clinical notes and structured physiological measurements and estimate the risk of new AKI onset using the MIMIC-III database. From the clinical notes, we generated clinically meaningful word representations and embeddings. Four supervised learning classifiers and mixed-feature deep learning architecture were used to construct prediction models. The best configurations consistently utilized both structured and unstructured clinical features and yielded competitive AUCs above 0.83. Our work suggests that integrating structured and unstructured clinical features can be effectively applied to assist clinicians in identifying the risk of incident AKI onset in critically-ill patients upon admission to the ICU.

Keywords:

Clinical Decision Support, Natural Language Processing, Acute Kidney Injury.

Introduction

Acute kidney injury (AKI) is commonly seen in adults in the intensive care unit (ICU). AKI is one of the major diagnoses among ICU patients and a leading factor associated with a prolonged hospital stay and subsequent morbidity or early mortality post discharge [1, 2]. Acute renal failure is a complex disorder that presents itself in a variety of settings with clinical manifestations, ranging from a minimal elevation in serum creatinine to anuric renal failure [3]. Unfortunately, the main biomarker of AKI, serum creatinine (SCr), is a late marker of injury, which delays diagnosis and treatment [4]. However, the efficacy of intervention greatly relies on the early identification of AKI [5]. Early recognition is critical in that AKI usually occurs over the course of a few hours to days and is potentially reversible if detected and managed early [5, 6].

In this study, we used the definition of AKI from the Kidney Disease Improving Global Outcomes (KDIGO) [7], in order to standardize the published diagnostic criteria. The diagnostic criteria are defined as an acute increase in the absolute level of serum creatinine of more than 0.3 mg/dl or 50% higher change in serum creatinine (SCr) from baseline within a 48-hour period or decreased glomerular filtration rate (GFR) to less than 0.5 ml/kg/hour for more than six hours [7, 8]. These criteria were based on accumulating evidence that even small alterations in SCr are associated with serious consequences. Therefore, an accurate creatinine forecast may enable prediction of AKI risk. In this study, we focus on predicting AKI using first-day measurements of a multivariate panel of physiologic variables, in order to elucidate early, subclinical deterioration of patient's physiologic baselines that are predictive of AKI.

Many factors including nephrotoxic medications, insufficient effective circulating fluid volume, and intrinsic renal disease can cause or contribute to AKI [9, 10]. In addition, epidemiology studies show multiple comorbidities, including diabetes mellitus, cardiovascular disease, chronic liver disease, cancer, and complex surgery have been associated with the development of AKI [5, 10, 11]. Thus, a comprehensive understanding of ICU patients is essential to predict the development of AKI.

The increasing use of electronic health records (EHRs) allows access a comprehensive and extensive amount clinical data to develop models to predict AKI in ICU [3]. There are several previous studies using EHR data to predict AKI [12-21]. Most studies achieved a modest performance with area under the receiver operation curve (AUC) close to 0.75. However, majority of these studies focus on specific patient population such as the elderly and cardic surgery patients. In addition, many prior studies rely on various static scoring algorithms and/or do not incorporate the temporal progression of the clinical, laboratory information that is shown to be effective for the prediction. Last but not least, some studies incoporate a limited set of predictors that are not inclusive enough to capture changes in clinical care that may impact AKI risk; similarly, some studies rely on non-routine biomarkers such as neutrophil gelatinase-associated lipocalin (NGAL) that are not always tested for general patients. Data-driven predictive modeling for AKI has recently gained traction, but is focused on using either structured EHR data [22] or unstructured clinical notes alone [23]. In summary, these models are not optimally suited for a relatively general patient population and are not set up as adaptive tools to assist the clinical decision-making process.

Our approach, in contrast, involves the careful modeling of a wide array of predictor data including clinical treatments and the temporal aggregation of predictor data. Specifically, we include not only structured data such as laboratory test results, but also unstructured data, the ICU clinical notes (e.g., physician notes and nurse notes) to provide a comprehensive picture of the patients' current pathophysiologic condition and help develop a powerful model to predict AKI.

From a pratical standpoint, our approach focuses on the early prediction of AKI on patients who do not meet AKI criteria on admission to the ICU, thus targeting a population that could

benefit from early intervention which might reverse the development of AKI or minimize its clinical impact. This is especially important since prior studies on automated AKI detection (as opposed to prediction) show limited effectiveness of therapeutic interventions in patients already meeting AKI criteria [24]. We expect the early prediction of AKI in this study to have a wide range of clinical applications.

Methods

Dataset

Data for this study was acquired from Medical Information Mart for Intensive Care III (MIMIC-III). MIMIC-III captures de-identified health information for more than 46,000 patients admitted to the critical care units at Beth Israel Medical Center between 2001 and 2012. We developed a SQL script to extract data of patients for who had a creatinine measured at 72 hours following ICU admission from the MIMIC-III database [25]. Only patients 18 years of age or older were included, and patients with the pre-existing condition of Chronic Kidney Disease (CKD), who have an estimated glomerular filtration rate GFR (eGFR) < 60 mL/min/1.73 m^2 were excluded [26, 27]. Data extracted include patients' age, gender, ethnicity, 72-hour serum creatinine (with only the creatinine value from the first 24 hours used as predictor, the value from the 2nd and 3rd days are used to identify AKI), vital signs and lab values during the first day of ICU admission, whether the patient was mechanically ventilated during the first day of ICU admission, the hourly rate of urine output during the first day of ICU admission, and the clinical notes during the first 24 hours of ICU admission. Only clinical notes from the first 24 hours of the ICU stay are included given the commonly seen lag between the provide-patient encounter and the time the notes were performed.

Based on the criteria from KDIGO [7], AKI is defined as either of the following two conditions being met: (1) greater than or equal to 50% increase from the baseline creatinine value to the current creatinine value and (2) greater than or equal to 0.3 mg/dL change in creatinine from the baseline creatinine to the current creatinine value. Since our study focuses on AKI developed after the ICU admission, patients who got AKI on admission (day 1) were excluded. After this step, the remaining patients' AKI status were determined by comparing day 2 and day 3 maximum creatinine to the day 1 minimum creatinine level (the baseline).

A total of 16,558 ICU stays of 14,469 patients met the inclusion criteria for this study. The dataset was first split into training and test sets by an 7:3 ratio. Then, reassignment was conducted to ensure multiple ICU stays of the same patient stayed in the same set. The AKI prevalence rate overall and in both sets are approximately 17% as shown in Table I below. Table II presents the predictor variables used in this study, along with their statistical characteristics such as mean and standard deviation for continuous variables and count and percentage for categorical or discrete variables.

Table I – AKI Status Distribution Overall and in Training and Test Sets.

Set	All	AKI	Non-AKI	AKI Prevalence
Overall	16,558	2,785	13,773	16.82%
Training	11,558	1,927	9,631	16.67%
Test	5,000	858	4,142	17.16%

Natural language processing of the clinical notes

To better interpret clinical notes, some preprocessing steps were needed. Masked protected health information (PHI) in the notes were removed first. Then, in order to make the information usable in machine learning classifiers, clinical notes were converted to structured features. Our first set of features consists of unigram bag-of-words [1], which identified and normalized lexical variants from the unstructured text content. Those features with document-frequency under 10 were removed to reduce noise. A total of 313 stop words were applied according to NCBI guide. Term frequency-inverse document frequency (tf-idf) weighting adjustment was also applied [28].

Table II - Univariate Characteristics for Predictors of Interest, N= 16,558

Variable	Mean	SD
Age (year)	60.80	16.16
Gender - N, %		
Female	6,735	40.68%
Male	9,823	59.32%
Ethnicity - N, %		
African-American	1,048	6.33%
White	11,981	72.36%
Hispanic	532	3.21%
Other	2,997	18.10%
Heart rate maximum (bpm)	105.43	19.76
Heart rate mean (bpm)	87.10	15.03
Systolic BP minimum (mmHg)	92.45	17.04
Systolic BP mean (mmHg)	119.06	15.85
Diastolic BP minimum (mmHg)	44.60	10.95
Diastolic BP mean (mmHg)	61.33	10.00
Temperature maximum (Celsius)	37.65	0.74
SpO2 minimum (%)	92.20	7.24
SpO2 mean (%)	97.53	1.83
Glucose level maximum (mg/dL)	173.82	72.33
Bicarbonate level minimum (mg/dL)	23.86	4.26
Creatinine level minimum (mg/dL)	0.73	0.20
Creatinine level maximum (mg/dL)	0.80	0.22
Hemoglobin level minimum (g/dL)	10.34	2.07
Platelet count minimum (K/μL)	208.83	113.13
Potassium level maximum (mg/dL)	4.42	0.80
Partial thromboplastin time minimum (s)	32.49	11.49
Partial thromboplastin time maximum (s)	40.39	23.63
International normalized ratio minimum	1.33	0.43
International normalized ratio maximum	1.47	0.68
Prothrombin time minimum (s)	14.48	3.00
Prothrombin time maximum (s)	15.42	4.23
Blood urea nitrogen level maximum (mg/dL)	16.54	8.34
White blood cell count maximum (K/μL)	12.94	8.14
Calcium level minimum (mg/dL)	8.08	0.75
Average urine output (mL)	2,249.40	1419.51
Estimated glomerular filtration rate (eGFR)	110.33	50.20
Mechanical Ventilation - N, %		
No (0)	7,649	46.20%
Yes (1)	8,909	53.80%

Missing data imputation for structured data

There are no missing values among age, gender, and race. Other structured clinical and laboratory data obtained from clinical settings contained missing values, which indicates certain tests were not performed during the patient's ICU stay. For example, the variables having a large proportion of missing values include minimum albumin level (74.1%), maximum bilirubin level (67.2%), maximum lactate level (55.8%), maximum c-reactive protein level (99.0%), maximum asparate aminotransferase level (66.8%), maximum pH level (36.6%), and minimum base excess level (64.8%).

We employed a two-step process to handle missing values. First, we removed the variables with missing values greater

than 20%. We thenfilled in the values for predictors (e.g., labs not performed or recorded) using Multivariate Imputation by Chained Equations (MICE) [29] for those variables with less than 20% missing values. MICE estimates a conditional model for each variable to be imputed, with the other variables as possible predictors [30]. The term chained equation comes from the adoption of a Gibbs sampler, which is an iterative Markov Chain Monte Carlo algorithm for obtaining a sequence of observations that are approximated from a joint probability distribution. As MICE closely tracks the conditional interdependencies among variables, we expect MICE to produce more accurate imputation. Then, as the second step, we used the measured and imputed values for these predictors plus age and gender to predict maximum creatinine results during day 2 and day 3. In this step, we predicted both numerical results for creatinine (linear regression) and whether creatinine increase would be classified as AKI (logistic regression). Although no creatinine results were actually missing from our dataset per the inclusion criteria, we assessed model performance and creatinine predictability by masking creatinine results from a test fold during five-fold cross validation and then compared predicted creatinine results to the masked (measured) values. The masked-measured values were treated as the "ground truth" in assessing model performance. The imputation stage was required because the prediction algorithms used in the second stage of our procedure could not directly accommodate missing data in predictors.

Machine learning classifiers

Logistic regression (LR), random forest (RF), multinomial naïve Bayes (NB), and supported vector machine (SVM) classifiers were implemented in scikit-learn to find the best prediction model. As the class ratio is imbalanced, we set the class_weight parameter to "balanced" for logistic regression, random forest and supported vector machine classifiers to down weight the more popular class. In order to tune parameters and reduce over-fitting and instablity, a grid search with 3-fold cross-validation was performed on the training set. The best parameters for each classifier were then applied to the test set to assess the prediction performance of each model.

Mixed-feature Convolutional Neural Networks

We also explored deep learning models for AKI prediction. We used mixed-feature Convolutional Neural Networks (CNN) to combine word features and structured clinical features. The architecture of mixed-feature CNN model is shown in Figure 1. It used pre-trained word embeddings and structured clinical data as the input. A one-dimensional convolution layer was built on the input embeddings. We used max pooling to select the most important feature with the highest value in the convolutional feature map and then concatenated the max pooling results of word embeddings and structured clinical features. The concatenated hidden features were fed into a fully connected layer, followed by a dropout and ReLU activation layer. Finally, a fully connected layer was fed to a softmax output layer, whose output is the probability distribution over labels (AKI or Non-AKI).

We adopted the following parameter settings for the mixed-feature CNN after further separating 10% from training data as validation set for parameter tuning, the number of convolution filters: 32, the convolution kernel size: 5, the dimension of hidden layer in the fully connected layer: 64, dropout keep probability: 0.8, learning rate: 0.001, batch size: 64. We used weight blancing in mixed-feature CNN training to address the class imbalance issue. In particular, we used the labels of classes to automatically adjust weights inversely proportional to class frequencies using the training data. The weights are then used in calcualting the weighted cross entropy from logits in the softmax layer.

Evaluation

We adopted area under the receiver operating characteristic (AUC) to evaluate the performance of imbalanced binary classifiers. Precision, recall, F-measure of positive AKI status were also calculated for reference.

Results

In this section, we report the evaluation results on the held-out test set using four supervised learning classifiers (LR, RF, NB, SVM) over three different feature configurations (structured feature only, unstructured clinical notes only, and structured features combined with unstructured clinical notes), as well as results by mixed-feature CNN. Table III presents the results from the above model configurations.

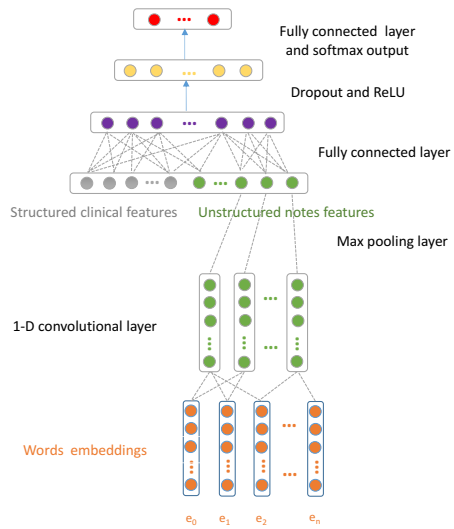

Figure 1. Mixed-feature Convolutional Neural Networks.

We first investigated the combinations of three feature representations and four supervised learning classifiers. The baseline predictor of each representation used Naive Bayes algorithm yielded AUC in the range from 0.65 to 0.75, F-measure from 0.17 to 0.37. In particular, combining structured and unstructured features in NB does not lead to better AUC. Logistic regression with L2-regularization over the mixed features yielded the best F-measure of 0.5423 on AKI status prediction, while calibrated SVM with L1-regularization gave the best AUC of 0.8352. Overall, all classifiers, except for NB as the baseline, yielded a competitive AUC score over 0.79, given that most previous models had modest AUCs around 0.75 (see Related Work section). In addition, the best LR model favored high recall (i.e., sensitivity: 0.7284) over precision (i.e., positive predictive value: 0.4319). This serves well for the clinical application of AKI onset alarm as we want to capture as many future AKI onsets as possible, while tolerating modest false alarms.

For the mixed-feature CNN, using ratio balancing class weights, yielded the best performance of 0.8167 AUC and 0.3559 F-measure. Despite heavy parameter tuning, all configurations of parameters of mixed-feature CNN did not outperform the best non-CNN classifiers. Based on our previous experiments on general domain text corpora [31], CNN-based architectures generally work well for datasets with

short texts, but may not outperform bag-of-word on corpus with long texts such as the AKI clinical note corpus. In addition, due to the fact that training CNN models is usually time-consuming, in our case, well-calibrated non-CNN classifiers seem to be a more suitable choice.

Feature analysis

We further examined important features by ranking coefficients in the L2-regularized logistic regression (best F-measure) over bag-of-words and structured clinical features. Table IV presents top 10 positive structured clinical features that contribute to AKI onset. The maximal creatinine level in day 1 carries a weight almost 10 times higher than the predictor with the second highest weight. The information critical to predict AKI is concentrated in the top three predictors by weight. These predictors are consistent with the known pathophysiology of AKI that older patients have higher incidence of AKI. Mechanical ventilation and coagulopathy (prolonged prothrombin times) are also known risk factors of AKI and also might represent patients with higher severity of illness and/or sepsis. The elevated potassium level likely represents early electrolyte disturbances in the setting of injured kidneys likely to meet AKI definition in the subsequent days. Finally, the elevated creatinine and BUN levels, while not meeting AKI criteria in these patients given the exclusion criteria of the study, likely represents an early elevation indicative of injured kidneys which has not peaked yet.

Table III – Machine Learning Model Results. Str: structured features. Unstr: unstructured features. ALG: algorithm. MCNN: Mixed-feature CNN. Best AUC and F-measure are in bold.

Features	ALG	AUC	Precision	Recall	F-meausre
Str	LR	0.8336	0.4319	0.7284	**0.5423**
+	RF	0.7914	0.6774	0.0490	0.0913
Unstr	NB	0.6728	0.2681	0.6166	0.3737
	SVM	**0.8352**	0.7274	0.2340	0.3541
	MCNN	0.8167	0.7292	0.2354	0.3559
Str	LR	0.8117	0.3782	0.7401	0.5006
	RF	0.8132	0.8125	0.1364	0.2335
	NB	0.6574	0.2560	0.6247	0.3631
	SVM	0.8097	0.7439	0.2133	0.3315
Unstr	LR	0.7735	0.4009	0.6457	0.4946
	RF	0.7582	0.5769	0.0350	0.0659
	NB	0.7495	0.5658	0.1002	0.1703
	SVM	0.7727	0.5306	0.0606	0.1088

Figure 2 presents top 50 positive bag-of-words features with theirs font sizes proportional to their coefficients in the model. For the selected features, in most cases, these features appear to be clinically meaningful. For example, 'lasix' with highest coefficient in bag-of-words is one of the diuretics that can treat fluid retention and edema that might be caused by kidney dysfunction. Also, 'co' and 'ci' in bag-of-words is short for 'cardiac output and cardiac index', which assess whether a patient's heart is pumping enough blood and delivering sufficient oxygen to cells. Patients with abnormal cardiac output and even heart failure often have a higher risk of AKI [36]. The words 'insulin' and 'incisional' indicate diabetes mellitus comorbidity and procedural risks that may predispose AKI onset. Feature examination confirms that clinically meaningful key words in clinical notes can be used to predict AKI onset and the models we built do capture those words.

Discussions, Limitations and Future Work

Due to data accuracy concern, this study only uses the increase in creatinine to determine AKI. Should data that accurately record the decrease of GFR for a prolonged period exist, future studies will include both criteria to better capture AKI.

Table IV – Top physiologic variables that are associated with increased AKI risk by L2-regularized logistic regression.

Physiologic Variable	Weight
Creatinine level maximum (mg/dL)	4.8374
Mechanical Ventilation	0.4578
International normalized ratio maximum	0.2315
Potassium level maximum (mg/dL)	0.0589
Prothrombin time minimum (s)	0.0267
Estimated glomerular filtration rate (eGFR)	0.0199
Age (year)	0.0192
Diastolic BP mean (mmHg)	0.0101
Partial thromboplastin time minimum (s)	0.0095
Blood urea nitrogen level maximum (mg/dL)	0.0063

Figure 2 Ranked top 50 positive features in Bag-of-words with its coefficients as font size in AKI onset prediction by L2-regularized logistic regression.

When working with structured clinical variables, we have used MICE imputation to fill in missing entries [30]. In principle, MICE assumes the missing-at-random pattern, an assumption that almost certainly will not hold in real clinical practice, since clinicians often order tests under certain expectations about the likely results. However, our previous study shows that even when the missing-at-random assumption may not hold, in practice, MICE may still be used as an effective way and baseline for comparing other multiple imputation methods due to its simple implementation [32, 33]. On the other hand, missingness may represent no indication for having the test performed, hence may offer clinical information. Thus, we plan to investigate missingness patterns as additional predictors for AKI prediction in future studies.

When working with clinical notes, we only explored the bag-of-words model in conventional machine learning models and the word embeddings in mixed-feature CNN model. Though this yields promising performance, there are other options for using medical concepts as features such as through NLP pipelines including MetaMap [34], clinical Text Analysis and Knowledge Extraction System (cTAKES) [35]. Further investigation is necessary in order to identify which pipeline is the most suitable tool to generate medical concepts in disease prediction with clinical notes. In addition, we only explored mixed-feature CNN as one deep learning framework. Other deep learning architecture that is suitable for clinical dataset with mixed feature types and long text content are of future interest.

The improved model performance implies the potential clinical application in the near future. However, it requires further study and discussion to determine the model performance that is needed before it could be intergreted in clincal decision-making process.

Conclusions

Our study demonstrates that a supervised learning method that integrates both structured physiologic variables and unstructured clinical notes can be effective in early prediction of AKI onset in the first 72 hours following ICU admission for general adult patient population. We showed that carefully selected physiologic variables and well-represented clinical notes as features can predict new AKI onset with an AUC greater than 0.83, competitive with previous studies focused on specific patient groups or on novel biomarkers. Our work suggests that prospective trials with independent model training and validation cohorts are needed to further evaluate the clinical utility of this approach for identifying at risk patients early in their hospital course and potentially instituting interventions to decrease the likelihood of developing AKI.

Acknowledgements

This study was supported in part by grant R21LM012618.

References

[1] T. Ali *et al.*, "Incidence and outcomes in acute kidney injury: a comprehensive population-based study," (in eng), *J Am Soc Nephrol,* vol. 18, no. 4, pp. 1292-8, Apr 2007.

[2] C. V. Thakar, A. Christianson, R. Freyberg, P. Almenoff, and M. L. Render, "Incidence and outcomes of acute kidney injury in intensive care units: a Veterans Administration study," (in eng), *Crit Care Med,* vol. 37, no. 9, pp. 2552-8, Sep 2009.

[3] H. E. Wang, P. Muntner, G. M. Chertow, and D. G. Warnock, "Acute kidney injury and mortality in hospitalized patients," (in eng), *Am J Nephrol,* vol. 35, no. 4, pp. 349-55, 2012.

[4] H. R. de Geus, M. G. Betjes, and J. Bakker, "Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges," (in eng), *Clin Kidney J,* vol. 5, no. 2, pp. 102-108, Apr 2012.

[5] J. A. Kellum, N. Lameire, and K. A. G. W. Group, "Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1)," (in eng), *Crit Care,* vol. 17, no. 1, p. 204, Feb 2013.

[6] P. K. T. Li, E. A. Burdmann, R. L. J. A. j. o. n. Mehta, and transplantation, "Acute kidney injury: global health alert," vol. 6, no. 2, pp. 75-81, 2013.

[7] K. D. I. G. O. C.-M. W. Group, "KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD)," *Kidney international. Supplement,* no. 113, p. S1, 2009.

[8] R. L. Mehta *et al.*, "Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury," (in eng), *Crit Care,* vol. 11, no. 2, p. R31, 2007.

[9] R. Bellomo, J. A. Kellum, and C. Ronco, "Acute kidney injury," (in eng), *Lancet,* vol. 380, no. 9843, pp. 756-66, Aug 2012.

[10] G. M. Chertow, E. Burdick, M. Honour, J. V. Bonventre, and D. W. Bates, "Acute kidney injury, mortality, length of stay, and costs in hospitalized patients," (in eng), *J Am Soc Nephrol,* vol. 16, no. 11, pp. 3365-70, Nov 2005.

[11] R. Bellomo, J. A. Kellum, and C. J. T. L. Ronco, "Acute kidney injury," vol. 380, no. 9843, pp. 756-766, 2012.

[12] R. J. Kate, R. M. Perez, D. Mazumdar, K. S. Pasupathy, and V. Nilakantan, "Prediction and detection models for acute kidney injury in hospitalized older adults," (in eng), *BMC Med Inform Decis Mak,* vol. 16, p. 39, Mar 2016.

[13] M. Bedford *et al.*, *Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study*. Southampton (UK): NIHR Journals Library, 2016.

[14] M. A. Perazella, "The urine sediment as a biomarker of kidney disease," (in eng), *Am J Kidney Dis,* vol. 66, no. 5, pp. 748-55, Nov 2015.

[15] K. F. Kerr, A. Meisner, H. Thiessen-Philbrook, S. G. Coca, and C. R. Parikh, "Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers," (in eng), *Clin J Am Soc Nephrol,* vol. 9, no. 8, pp. 1488-96, Aug 2014.

[16] W. H. Kim *et al.*, "Simplified clinical risk score to predict acute kidney injury after aortic surgery," (in eng), *J Cardiothorac Vasc Anesth,* vol. 27, no. 6, pp. 1158-66, Dec 2013.

[17] H. D. Kiers *et al.*, "Comparison and clinical suitability of eight prediction models for cardiac surgery-related acute kidney injury," (in eng), *Nephrol Dial Transplant,* vol. 28, no. 2, pp. 345-51, Feb 2013.

[18] T. Ploug and S. Holm, "Informed consent and routinisation," (in eng), *J Med Ethics,* vol. 39, no. 4, pp. 214-8, Apr 2013.

[19] L. Z. Zhou *et al.*, "Development and Validation of a Risk Score for Prediction of Acute Kidney Injury in Patients With Acute Decompensated Heart Failure: A Prospective Cohort Study in China," (in eng), *J Am Heart Assoc,* vol. 5, no. 11, Nov 2016.

[20] T. Wilson *et al.*, "Risk prediction models for acute kidney injury following major noncardiac surgery: systematic review," (in eng), *Nephrol Dial Transplant,* vol. 31, no. 2, pp. 231-40, Feb 2016.

[21] L. N. Sanchez-Pinto and R. G. Khemani, "Development of a Prediction Model of Early Acute Kidney Injury in Critically Ill Children Using Electronic Health Record Data," (in eng), *Pediatr Crit Care Med,* vol. 17, no. 6, pp. 508-15, Jun 2016.

[22] L. Zimmerman *et al.*, "Early Prediction of Acute Kidney Injury Following ICU Admission," *BMC Medical Informatics and Decision Making (accepted),* no. S1, pp. 5-16, 2019.

[23] Y. Li, L. Yao, C. Mao, A. Srivastava, X. Jiang, and Y. Luo, "Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018.

[24] K. Colpaert *et al.*, "Impact of real-time electronic alerting of acute kidney injury on therapeutic intervention and progression of RIFLE class," (in eng), *Crit Care Med,* vol. 40, no. 4, pp. 1164-70, Apr 2012.

[25] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," (in eng), *Sci Data,* vol. 3, p. 160035, May 2016.

[26] A. S. Levey *et al.*, "A new equation to estimate glomerular filtration rate," (in eng), *Ann Intern Med,* vol. 150, no. 9, pp. 604-12, May 2009.

[27] L. A. Stevens, J. Coresh, T. Greene, and A. S. Levey, "Assessing kidney function--measured and estimated glomerular filtration rate," (in eng), *N Engl J Med,* vol. 354, no. 23, pp. 2473-83, Jun 2006.

[28] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," Carnegie-mellon univ pittsburgh pa dept of computer science1996.

[29] S. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of statistical software,* vol. 45, no. 3, 2011.

[30] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *2011,* vol. 45, no. 3, p. 67, 2011-12-12 2011.

[31] L. Yao, C. Mao, and Y. Luo, "Graph Convolutional Networks for Text Classification," in *AAAI*, 2019.

[32] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "Using Machine Learning to Predict Laboratory Test Results," (in eng), *Am J Clin Pathol,* vol. 145, no. 6, pp. 778-88, Jun 2016.

[33] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data," *Journal of the American Medical Informatics Association (JAMIA),* vol. 25, no. 6, pp. 645-653, Nov 30 2017.

[34] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program," in *Proceedings of the AMIA Symposium*, 2001, p. 17: American Medical Informatics Association.

[35] G. K. Savova *et al.*, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association,* vol. 17, no. 5, pp. 507-513, 2010.

Address for Correspondence

Yuan Luo, email: yuan.luo@northwestern.edu