Prediction of Acute Kidney Injury in the ICU

Mia Kanzawa*†, Rohan Paul*, and Nielson Weng†‡

*Department of Biomedical Informatics, Stanford University; †Department of Medicine, Stanford University School of Medicine; ‡Department of Chemistry, Stanford University
Email: {mkanzawa, ropaul, nweng}@stanford.edu

INTRODUCTION

Acute Kidney Injury (AKI), is a clinicopathologic entity characterized by a sudden decrease in kidney function, leading to retention of metabolic waste products and the dysregulation of electrolyte homeostasis (1). Despite our progress in understanding the pathophysiology and a precise clinical definition and staging for diagnosis, AKI remains a global public health concern impacting approximately 13.3 million patients per year and resulting in 1.7 million deaths per year (2). Even though the economic burden of AKI is not well understood, AKI from hospitalization has been shown to be associated with an increase in hospitalization costs of $7933 and an increase in length of stay of 3.2 days. In addition, the corresponding results among patients hospitalized with AKI requiring dialysis were $42,077 and 11.5 days. AKI was associated with higher costs than myocardial infarction and comparable to those for stroke (3). Even if patients have survived, many go on to develop chronic kidney disease, end-stage renal disease, or exacerbate a pre-existing renal condition to accelerate toward ESRD, contributing to additional health care cost (4, 5). Furthermore, other than dialysis, no treatment reliably improves survival (6).

Basic science research has identified the pathogenesis of AKI and potential therapeutic approaches in animal models (7). Translational research and clinical trials have resulted in poor progress due to the lack of early biomarkers of AKI in humans (8-9).

In ICUs, the incidence of AKI is increased at 30-50% of patients from 5% of normal hospitalized patients, and there is substantial evidence that the incidence is rising (10).

From the diverse and complex etiology, the condition can be roughly divided into two major categories: hospital-acquired or community-acquired (6). In our study, we are interested in AKI acquired from the hospital setting because these cases are often caused by medical procedures and/or medication and may be prevented. Recent studies conclude that early nephrology consultation leading to preventative measure decreases both the incidence and severity of AKI (11–13). Furthermore, electronic medical records and e-Alerts offer the potential for identifying high-risk patients and warning the use of nephrotoxic medications (14-16). Therefore, a reliable predictive machine learning model could address this unmet need.

Previous attempts at AKI prediction have been built on logistic regression and cite low positive predictive values (as high as 22.7%) with AUROC as high as 0.79 (13, 14). These models have been built on features and cohorts of patients that are similar to ours. There is a precedent for using deep neural networks to provide improved prediction of mortality using similar healthcare data (20). We hypothesize that by using a deep learning model on well-engineered features, we will be able to achieve improved performance of prediction of AKI in the ICU.

MATERIALS AND METHODS

Data Source MIMIC-III (Medical Information Mart for Intensive Care III) Database is a freely accessible critical care database consisting of de-identified health data for 46,520 Intensive Care Unit (ICU) patients of the Beth Israel Deaconess Medical Center between 2001 and 2012 with information on demographics, lab tests, medications, etc.

Identifying onset of AKI:

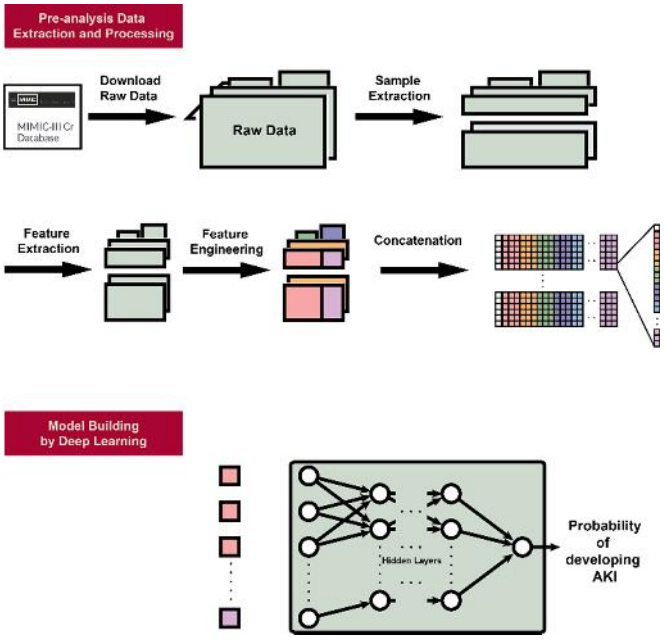

Figure 1. Schematic Conceptual Framework for the Workflow Used for Building the Deep Learning Model.

Positive Cases: We aimed to identify patients that had developed AKI within their hospital stay and only considered patients that had admissions of greater than or equal to 3 days. ICD-9 codes are assigned at the end of a hospital stay and are based on a physician's diagnosis. Therefore, an ICD-9 code will not capture time of AKI onset, nor patients for which the diagnosis of AKI was missed by the physician. It is imperative to specify a cutoff time at the time of AKI onset since features that are collected after the diagnosis may be caused by AKI and will not be useful for real-time prediction of AKI. The constraints for positive cases were decided based on the criteria recommended by Kidney Disease: Improving Global Outcomes (KDIGO) to be used by physicians to diagnose patients (17). The KDIGO definition and staging system is the most recent and preferred definition:

1. Increase in serum creatinine by >= 0.3 mg/dL within 48 hours, OR
2. Increase in serum creatinine by >= 1.5x baseline (known or presumed to have occurred within the prior seven days, OR
3. Urine volume < 0.5 mL/kg/hour for 6 hours

Onset of AKI was determined by the time point at which the patient met either criteria one or two at day 3 of their hospitalization or later. Criteria three was not used due to previously reported controversy over its significance as a diagnostic criterion given that it could be met by a healthy adult without pathology (18). For criterion two, baseline was determined as their first creatinine level taken upon admission.

Negative Cases: our negative samples were defined as patients who did not develop AKI at any time point during their hospitalization using the KDIGO criteria described above. In addition, patients that were admitted with AKI identified within the first day of admission were excluded from the positive cases, as well as patients with preexisting chronic end stage renal disease as defined by needing dialysis or with a creatinine level > 4 mg/dL.

After applying diagnostic criteria, we identified 18725 hospital admissions during which the patient developed AKI and 29221 non-AKI hospital admissions (39%). This is consistent with incidences cited from current literature with a range of 30-50% (10)

Feature Engineering

We included previous diagnostic codes, procedure codes, previous admissions, medications, vital signs, urine output, and labs as features for our model. Given that these features were often sparse and included both time series data and descriptors, we generally used two different strategies to engineer our features. For time series data such as labs, vital signs, and urine output, we used bins of 1 day, 1 week, and 1 month prior to AKI onset and used the average values over these time bins as our features. For descriptive data such as medications, previous diagnostic codes, and procedure codes we used a different approach. For medications, we identified presence or absence of each medication during each hospital admission. For previous diagnostic codes and procedure codes, we used frequency that these codes appeared in prior admissions, number of unique codes per patient, and a maximum number of codes assigned during any prior admission. Only diagnostic and procedural codes from prior admissions were used because they are assigned only at the end of the admission, so we are unable to extract actual time of diagnoses or procedures. This means that a code assigned during an admission in which AKI occurred may describe a procedure or diagnosis after the onset

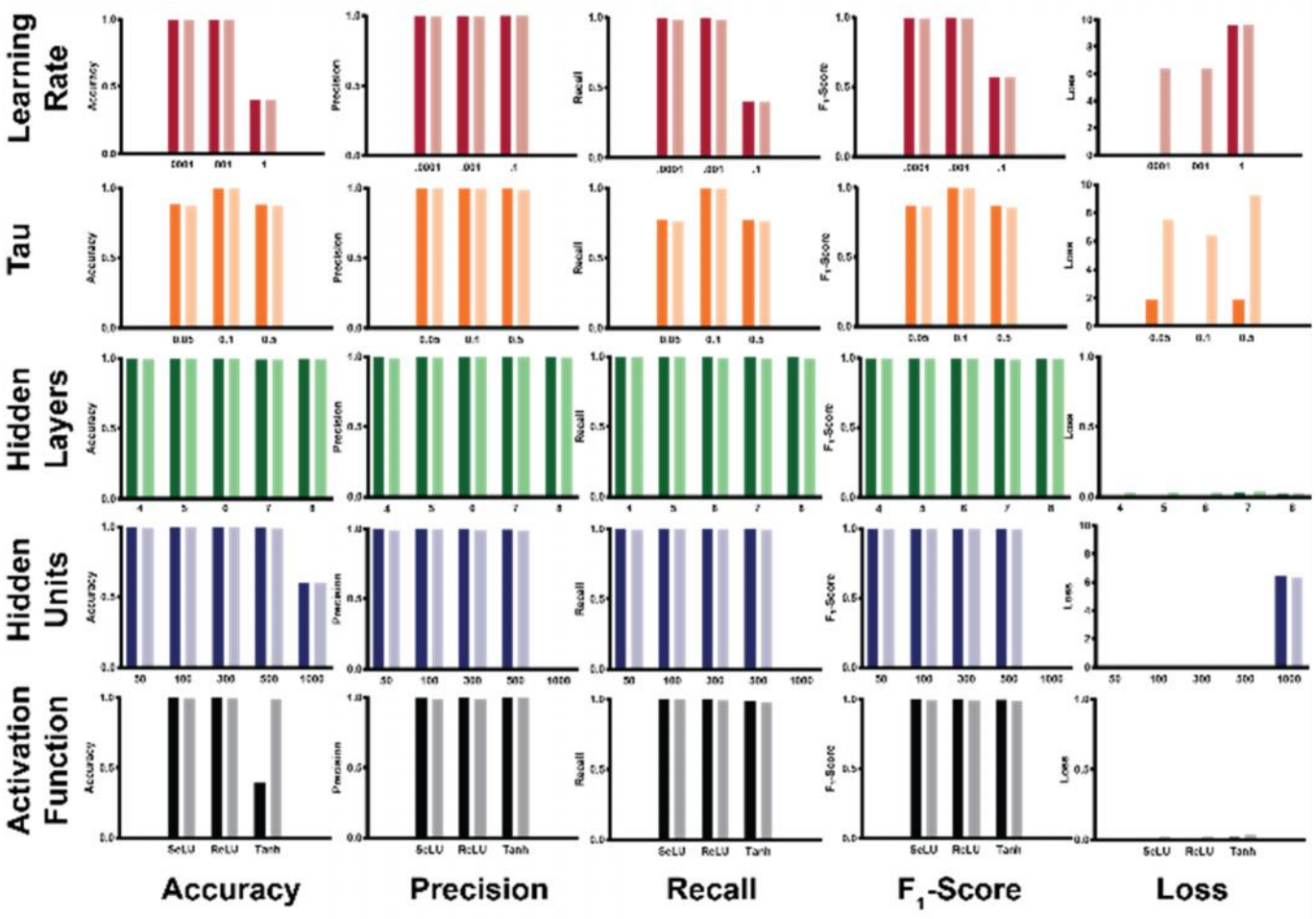

Figure 2. Results of experiments to determine parameters for neural network. Final model with 6 hidden layers, 300 units/layer, dropout rate of 0.1, learning rate of 0.001, tau of 0.1, mini batch size of 128, ran with 30 epochs. When optimizing a parameter, the rest of the parameters were held constant at the initial values. Darker bars represent the Training set while lighter bars represent the Dev Set.

of AKI. Given the sparseness of these codes, we also generalized the codes to Clinical Classification Software (CCS) categories as generated by the Healthcare Cost and Utilization Project (HCUP) (19). Using these more generalized categories we also counted frequencies of appearance in prior admissions, number of unique categories per patient, and a maximum number of categories assigned during any prior admission.

We then removed near zero variance features which reduced our number of features from over 4046 to 717 and normalized our features. The decrease in the feature space was due to the rare occurrence of certain ICD9 codes and medications.

Building our Model

We used a deep neural network with 6 hidden layers and 300 hidden units/layer. On each of the hidden layers we used a Scaled Exponential Linear Unit (SeLU) activation function and finally used a sigmoid function for our output layer. We applied dropout with a probability of 0.01 prior to each layer and regularized using L2-regularization. . We used the Adam optimizer with a mini-batch size of 128 ran over 30 epochs. We determined our parameters through extensive experimentation (Figure 2).

RESULTS

Our model appeared robust to various parameters (Figure 2) and ultimately performed well with high accuracy (99.1%), precision (0.991), and recall (0.986) (Table 1). The AUROC was We also observed good calibration with a Brier score of 0.010.

DISCUSSION

	Training Set	Dev Set	Test Set
Loss	0.013	0.036	0.040
Accuracy	99.6%	99.2%	99.1%
Precision	0.998	0.991	0.986
Recall	0.998	0.989	0.992
F1-score	0.998	0.999	0.989

TABLE I FINAL MODEL RESULTS

AKI is a costly and dangerous ailment that affects a significant proportion of ICU patients. The recent application of machine learning and other computational tools to the prediction of AKI has improved our ability to preempt the disease and possibly start early treatment. However, novel methodologies like deep learning remain largely unexplored in this area.

Our work focuses on using deep learning to predict a probability of contracting AKI in the near future. We hypothesized that a neural network would be able to identify complex relationships in the high dimensional feature space. We have developed and validated such a model on a cohort of 50,000 patients including 20,000 unique hospital admissions who showed agreement with the clinical criteria for diagnosis with AKI. We show that our model performed very well with test set accuracy at 99%. Precision and recall were similarly high at 0.97 - 0.99 (Figure 4). High performance metrics indicate that our model has potential applications in clinical decision support. Specifically, a model with high recall allows physicians to identify patients for early intervention and prevention, thus opening the possibility of improving outcomes. Our choice to optimize recall over other metrics was made with this application in mind.

Identifying high risk patients will not only help reduce ICU mortality rate but also significantly reduce cost. A patient diagnosed with AKI in the ICU will cost the health system in excess of $8,000 on average. With approximately 75,000 patients afflicted each year, our model can lower expenditure by hundreds of millions of dollars annually.

Given our very high performance metrics, we investigated the possibility that some features could be very highly correlated with the outcome and therefore allow the model to easily make accurate predictions. After sequentially removing features to identify potential strong contributors, we

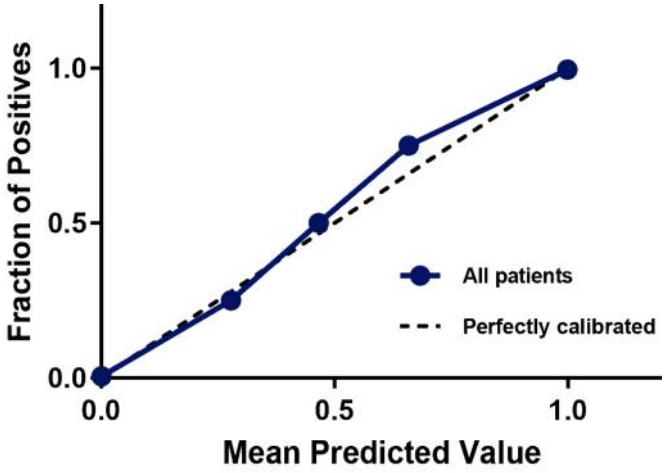

Figure 3. Reliability curve (calibration plot) of the model output probabilities.

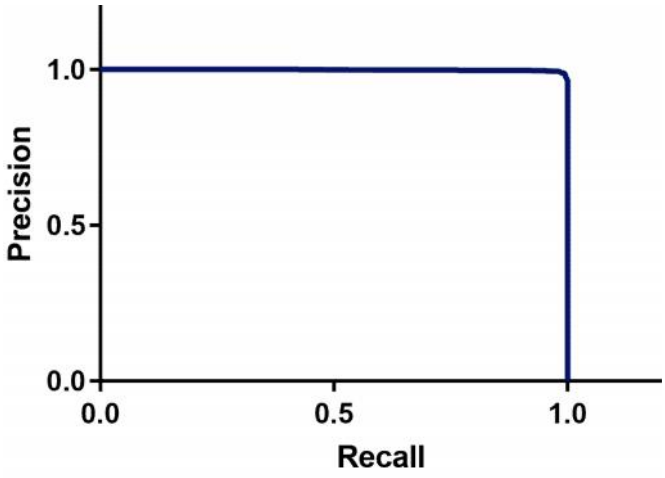

Figure 4. Receiver Operating Characteristic (ROC) of the model performance.

identified a list of features that were highly correlated with AKI onset in our dataset. This list was primarily made up of medications, including epinephrine, diltiazem, diphenhydramine, and entacapone. The relationship between medications that regulate blood pressure such as epinephrine

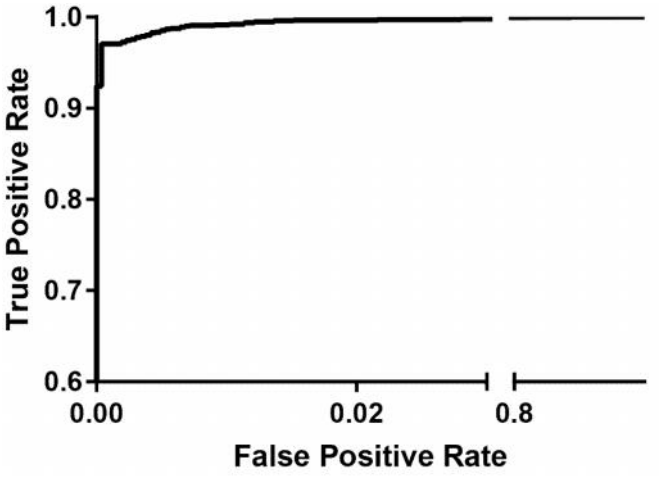

Figure 5. Interpolated Precision-Recall Curve.

and diltiazem and AKI onset is more readily interpretable given that renal hypoperfusion is a frequent cause of AKI; however, the strong correlation with diphenhydramine and entacapone warrants further investigation as they are not known to be nephrotoxins.

Additionally, while creatinine is the most frequently used biomarker for kidney function, it has been shown to be a delayed predictor of AKI onset (7). Because of this, it is possible that by using only serum creatinine levels to identify our patient cohort, we may mislabeled our time of AKI onset and used information from vital signs, medications and procedures after the actual onset of AKI that would bias our accuracy. Another explanation of the high accuracy includes systematic errors in data recording (medication administration recorded at an incorrect time). If medications were recorded at an earlier time than when they were actually administered, it is possible our model assigned high weights to such features and was therefore able to get a biased prediction.

FUTURE DIRECTION

In the future, we would like to quantify the model uncertainty as this is an important consideration for a physician before using valuable resources and time. If the model gives only a modest increase in risk but a disproportionately high uncertainty, it would be prudent to allocate only a minimal amount of extra resources to this patient. We propose using a bayesian neural network to accomplish this task (21).

Furthermore, due to time constraints, we were limited in the amount and complexity of feature engineering we could perform. We would like to improve prescription features by broadening the categories using the RxNorm ontologies, include demographic information, include diagnoses from notes using natural language processing techniques and adding more time bins like 48, 72 etc. hour bins.

REFERENCES

1. Kumar V, Abbas A, Fausto N, Robbins S, Cotran R (2010) *Pathologic Basis of Disease*.
2. Lewington AJP, Cerdá J, Mehta RL (2013) Raising awareness of acute kidney injury: a global perspective of a silent killer. *Kidney Int* 84(3):457–467.
3. Silver SA, Long J, Zheng Y, Chertow GM (2017) Cost of acute kidney injury in hospitalized patients. *J Hosp Med* 12(2):70–76.
4. Ingelfinger JR, Chawla LS, Eggers PW, Star RA, Kimmel PL (2014) Acute Kidney Injury and Chronic Kidney Disease as Interconnected Syndromes. *N Engl J Med* 371:58–66.
5. Ishani A, et al. (2009) Acute kidney injury increases risk of ESRD among elderly. *J Am Soc Nephrol* 20(1):223–8.
6. Zuk A, Bonventre J V. (2016) Acute Kidney Injury. *Annu Rev Med* 67(1):293–307.
7. Devarajan P (2006) Update on mechanisms of ischemic acute kidney injury. J Am Soc Nephrol 17:1503–1520
8. Allgren RL, Marbury TC, Rahman SN, Weisberg LS, Fenves AZ, Lafayette RA, Sweet RM, Genter FC, Kurnik BR, Conger JD, Sayegh MH (1997) Anaritide in acute tubule necrosis. Auriculin Anaritide Acute Renal Failure Study Group. N Engl J Med 336:828–834
9. Hirschberg R, Kopple J, Lipsett P, Benjamin E, Minei J, Albertson T, Munger M, Metzler M, Zaloga G, Murray M, Lowry S, Conger J, McKeown W, O'Shea M, Baughman R, Wood K, Haupt M, Kaiser R, Simms H, Warnock D, Summer W, Hintz R, Myers B, Haenftling K, Capra W, Pike M, Guler H-P (1999) Multicenter clinical trial of recombinant human insulin-like growth factor 1 in patients with acute renal failure. Kidney Int 55:2423–2432

10. Xue JL, Daniels F, Star RA, Kimmel PL, Eggers PW, Molitoris BA, Himmelfarb J, Collins AJ (2006) Incidence and mortality of acute renal failure in medicare beneficiaries, 1992 to 2001. J Am Soc Nephrol 17:1135–1142
11. Mehta RL, et al. (2002) Nephrology consultation in acute renal failure. *Am J Med* 113(6):456–461.
12. Nephrology B, et al. (2007) Prognosis and serum creatinine levels in acute renal failure at the time of nephrology consultation: an observational cohort study. *BMC Nephrol* 8(8). doi:10.1186/1471-2369-8-14.
13. Malhotra, R. et al. (2017) A risk prediction score for acute kidney injury in the intensive care unit. *Nephrol Dial Transplant* **32,** 814–822 .
14. Basu, R. K. *et al.* (2014) Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children. *Kidney Int* **85,** 659–667.
13. Balasubramanian G, et al. (2011) Early Nephrologist Involvement in Hospital-Acquired Acute Kidney Injury: A Pilot Study. *Am J Kidney Dis* 57(2):228–234.
14. McCoy AB, et al. (2012) Real-time pharmacy surveillance and clinical decision support to reduce adverse drug events in acute kidney injury. *Appl Clin Inform* 3(2):221–238.
15. Colpaert K, et al. (2012) Impact of real-time electronic alerting of acute kidney injury on therapeutic intervention and progression of RIFLE class*. *Crit Care Med* 40(4):1164–1170.
16. Thomas M, Sitch A, Dowswell G (2011) The initial development and assessment of an automatic alert warning of acute kidney injury. *Nephrol Dial Transplant* 26(7):2161–2168.
17. KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney Int Suppl 2012; 2:8.
18. Erdbruegger, U, Okusa, M. Etiology and Diagnosis of Prerenal Disease and Acute Tubular Necrosis in Acute Kidney Injury in Adults Post TW, ed. UpToDate. Waltham, MA: UpToDate Inc. http://www.uptodate.com (Accessed on December 02, 2017.
19. HCUP-US Home Page. Available at: https://hcup-us.ahrq.gov/. (Accessed: 13th December 2017)
20. Avati, A. *et al.* (2017) Improving Palliative Care with Deep Learning. *arXiv:1711.06402 [cs, stat]*.
21. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 2951–2959 (Curran Associates, Inc., 2012).

Contributions:

Mia contributed to background research and with the feature engineering including extracting and generalizing ICD codes. She built the framework for the model, ran a large portion of the experiments, and led the writing of the Methods section.

Rohan contributed to background research and did the majority of the feature engineering including binning the time series data and extracting medication features. He helped experiment with the model and led the writing of the Results and Discussions section.

Nielson did the bulk of the background research, contributed to feature engineering and experimenting with the model, put our poster together, and led the writing for the Introduction section and led the making of the figures.