Computers in Biology and Medicine, Volume 116, 103580, January 2020.
This is the accepted manuscript version of the article, the final published journal article is available at:
https://doi.org/10.1016/j.compbiomed.2019.103580

A Continual Prediction Model for Inpatient Acute Kidney Injury

Rohit J. Kate[a*], Noah Pearce[b], Debesh Mazumdar[c], Vani Nilakantan[d]
[a]Department of Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[b]Aurora Research Institute, Aurora Health Care, Milwaukee, WI, USA
[c]Milwaukee Kidney Associates, Milwaukee, WI, USA
[d]Allina Health System, Minneapolis, MN, USA

Abstract
Acute kidney injury (AKI) commonly occurs in hospitalized patients and can lead to serious medical complications. But it is preventable and potentially reversible with early diagnosis and management. Therefore, several machine learning based predictive models have been built to predict AKI in advance from electronic health records (EHR) data. These models to predict inpatient AKI were always built to make predictions at a particular time, for example, 24 or 48 hours from admission. However, hospital stays can be several days long and AKI can develop any time within a few hours. To optimally predict AKI before it develops at any time during a hospital stay, we present a novel framework in which AKI is continually predicted automatically from EHR data over the entire hospital stay. The continual model predicts AKI every time a patient's AKI-relevant variable changes in the EHR. Thus, the model not only is independent of a particular time for making predictions, it can also leverage the latest values of all the AKI-relevant patient variables for making predictions. A method to comprehensively evaluate the overall performance of a continual prediction model is also introduced, and we experimentally show using a large dataset of hospital stays that the continual prediction model out-performs all one-time prediction models in predicting AKI.

Keywords
Acute kidney injury; prediction; EHR; machine learning

1. Introduction

Acute kidney injury (AKI), formerly known as acute renal failure, is a sudden loss of kidney function. AKI affects 5-7% of hospitalized patients [1,2,3] and 22-57% of patients in intensive care units [4,5,6,7,8]. It can lead to serious medical complications and is potentially fatal. It also results in longer hospital stays and thus contributes to increasing healthcare costs [1]. Even after resolution, it can subsequently lead to severe kidney problems such as chronic kidney disease and progression to dialysis dependency. The incidence of AKI is highest in elderly patients [9,10], and its rate has been steadily increasing in this population due to an increasing number of comorbidities, aggressive medical treatments, and greater use of nephrotoxic drugs. Two factors complicate AKI diagnosis: it has a heterogeneous etiology, and it often develops stealthily in hospitalized patients being treated for other problems. However, up to 30% of hospital-acquired AKI is preventable [11] if predicted in time. AKI is also potentially reversible if diagnosed and managed in time.

The seriousness of AKI and its preventability make AKI a perfect candidate for predictive analytics. Hence many machine learning based predictive models have been built to predict inpatient AKI from electronic

* Corresponding author.
Correspondence to Rohit J. Kate, Department of Computer Science, University of Wisconsin-Milwaukee, 3200 N Cramer St, Milwaukee, WI 53211, USA; Phone: (414) 229-4216; Email: katerj@uwm.edu

health records (EHR) data; good reviews of these models can be found in [12,13]. All these models had a particular time when AKI prediction was made for the rest of the hospital stay, for example, at 24 hours after admission [14,15], or at 48 hours after admission [16,17], or around the time of a medical intervention [18,19,20]. However, there are two fundamental problems with such models that have a fixed time for predicting AKI. First, hospital stays can be several days long, during which a patient's medical condition can significantly change and after which AKI may develop within a few hours. For example, in our data, 39.4% of AKI incidences occurred after five days from admission and 15.7% occurred after 10 days from admission. Therefore, it is difficult to predict such incidences too far ahead in time. Second, if a prediction model has a fixed time of prediction, for example, 24 hours after admission, then it is bound to miss all the AKI incidences that occur during those 24 hours. For example, in our data, 12.8% of AKI incidences occurred within 24 hours from admission and 30.9% of AKI incidences occurred within 48 hours from admission. Therefore, the later the prediction time, the more incidences it will naturally miss.

Since a prediction time that is neither too early nor too late is desirable, AKI should be predicted continually during the entire inpatient stay so as to optimally predict it before it develops. This possibility was recently mentioned in a workgroup statement from the 15th Acute Dialysis Quality Initiative (ADQI) consensus conference [12] as a model that would generate a prediction score in real time as each new data value is received. However, to the best of our knowledge, no previous work has reported such a model to continually predict AKI. Although susceptible patients are continually monitored manually, no automated models have been reported that can continually monitor patients' variables in EHR to predict AKI. Automated continual prediction can not only reduce the manual work of caretakers, but is also particularly desirable given that AKI can stealthily develop in hospitalized patients being treated for other problems.

Our contributions in this paper are as follows. We introduce a novel framework for automatically predicting AKI continually over hospital stays. A trained machine learning model is used to predict AKI every time a patient's status changes in the EHR. For example, it will predict AKI every time a new medication is prescribed, or a new comorbidity is recorded, or a new laboratory test result becomes available. This new framework does not require constant monitoring and is designed to work automatically using EHR data and trigger an alarm when the potential for AKI increases. We also introduce an evaluation method that comprehensively measures the overall performance of such dynamic predictions. Furthermore, we retrospectively measure the performance of advance predictions, that is, how well the model could predict AKI, say, 6 hours in advance, or 12 hours in advance etc., which was not measured in previous studies. Our continual prediction framework also lends itself to discovery of the most important dynamic features in which change in values often trigger the prediction of AKI. To the best of our knowledge, researchers in previous work had not reported the dynamic importance of features in predicting AKI.

2. Related Work

The only work we know that is close to our work is by He et al. [21] in which they present a framework in which prediction is made after every 24 hours for AKI to occur within the next 24 hours. They set time for AKI incidence as the day on which AKI is diagnosed. Thus their framework has a granularity of 24 hours. In contrast, our framework has no granularity restriction and it continually predicts AKI to occur any time during the rest of the hospital stay. Since AKI can develop within a few hours following any change in medical condition, our framework that makes prediction every time a patient variable changes offers a significant advantage over their framework that makes prediction once every 24 hours. Another important difference is that He et al. [21] evaluated their dynamic prediction models separately for each day of prediction (i.e., they report one area under ROC curve (AUC) for prediction at 24 hours after admission, another AUC for prediction at 48 hours after admission, etc.). The authors note this limitation [21] and point out the need for a comprehensive index to evaluate the performance of models that predict dynamically. In this paper we introduce just such an evaluation method, which can also be used to directly compare with the performances of traditional one-time prediction models. Hence, the evaluation method

for the dynamic predictions that we present in this paper is a contribution of our work. Without such an evaluation method He et al. [21] could not directly compare performances of the traditional one-time prediction models with a dynamic prediction model, but using our evaluation method we are able to show in this paper that our dynamic prediction model significantly outperforms one-time prediction models.

We point out that our continual prediction framework is different from other well-known time-related data modeling frameworks. There are types of models that utilize repeated measures of variables from longitudinal data [22]. However, they still make predictions at only one time. In contrast, our continual prediction model uses only the latest values of the variables but makes predictions continually over time. Survival analysis models [23] predict the time when an event will occur (e.g. time of death) but our continual prediction model continually predicts whether an event (e.g. AKI) will occur or not as the variables change over time. Multilevel models for change [24] model how variables change over time (e.g. to model growth). In contrast, our model is applied continually over time to predict a particular event. In past, some researchers have employed time series analysis of medical data for various applications [25,26,27,28,29] using the aforementioned types of models, but they did not do continual prediction. A few researchers have also built sequential models to discover longitudinal patterns in patient data, for example, patterns of disease diagnoses to predict the diagnosis for the next hospital admission [30,31]. In contrast, ours is not a sequential model but is applied continually over an entire hospital stay in order to predict a particular disease as early as possible using the latest values of patient variables. Although our study focused on AKI, our proposed continual prediction framework and the method to evaluate it are general and can be applied to other diseases and disorders.

3. Materials and Methods

3.1. Data Collection

The data was collected from the EHRs of 15 hospitals that are part of Aurora Health Care system. These hospitals are located in southeastern region of Wisconsin state and use the same EHR system. Structured data along with their timestamps for entire hospital stays was collected for all adults older than 60 in 2013, 2014, and 2015 (number of patients=84,480). We focused on hospitalized older adults because the occurrence of AKI is especially common in this age group [9,10], otherwise our model itself is general and could be trained to apply to any age-group. Based on our exclusion criteria, we excluded patients who had chronic kidney disease stages III, IV and V, and those who were on dialysis. In addition, we included only those hospital stays that were longer than 24 hours and in which at least two serum creatinine measurements were taken. After applying the exclusion criteria, 36,614 patients remained with 44,691 total hospital stays among them, as shown in Figure 1. This study was approved by the Institutional Review Board of Aurora Health Care.

3.2. AKI Definition

We used the AKIN criteria [32] as the definition of AKI, in which, AKI is determined by either a 1.5-fold increase or an absolute increase of 0.3 mg/dL in serum creatinine (SCr) within 48 hours. In our data, AKI developed in 3,786 hospital stays. In this study, hospital stays were classified as AKI or non-AKI and not patients, because a patient can have multiple hospital stays and may acquire AKI during one hospital stay and not during another. The time of the SCr measurement that satisfies the AKIN criteria is taken as the time of the AKI incidence. Figure 2 shows the number of AKI incidences that cumulatively occurred by different times from admission.

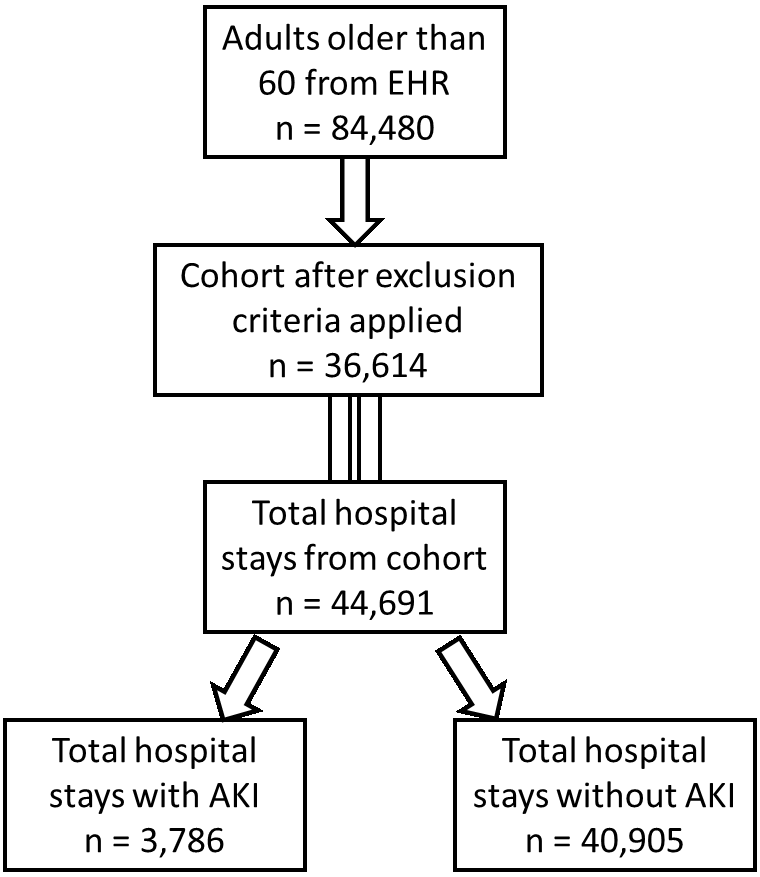

Figure 1. Flowchart depicting number of patients and hospital stays in the data.

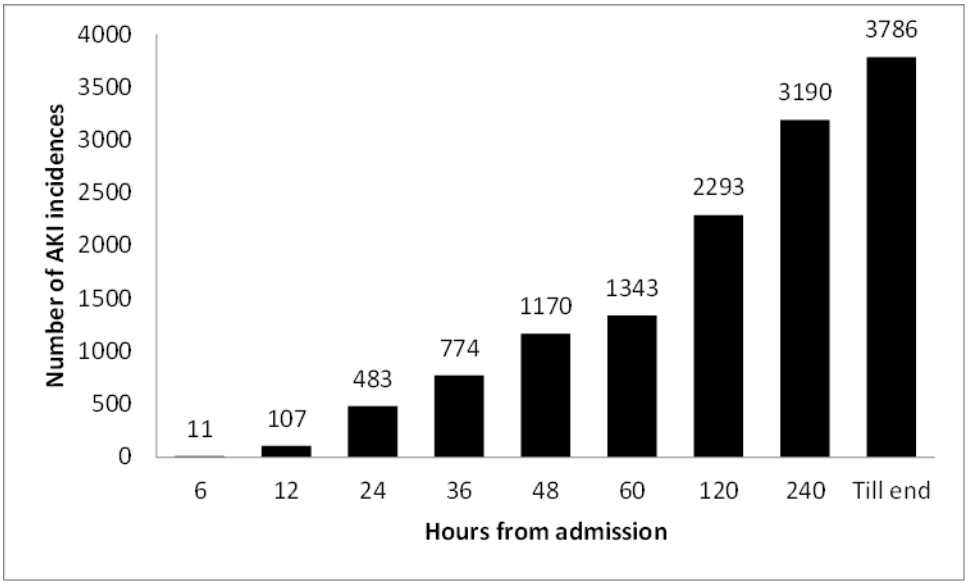

Figure 2. Cumulative number of AKI incidences by different hours from admission.

In the following subsections we first describe the traditional one-time prediction models for AKI and then describe our new continual prediction model.

3.3. Traditional One-Time Prediction Models for AKI

The task of an AKI prediction model is to predict in advance whether AKI will develop or not during a patient's hospital stay. Machine learning methods have been used in past to build AKI prediction models that classify hospital stays as AKI or non-AKI from various predictive variables, also called features. A machine learning method is given training examples of hospital stays in the form of their features (described in Section 3.3.2) and the correct class (AKI or non-AKI). From this training data, it learns to predict AKI on new hospital stays using only their features.

3.3.1. Time of Prediction

Most of the prior studies used a particular time for making predictions, for example, 24 hours from admission [14,15], or 48 hours from admission [16,17], or even at the time of admission [33], but some studies were not clear about it [34,35]. In the rest of this paper, a prediction model built to make predictions at 24 hours from admission is referred to as *one-time-at-24-hour* prediction model. When such a model is trained, it uses the latest values of the features as they were at 24 hours from admission for all its training examples. Then, when this model is applied to predict at 24 hours from admission whether AKI will develop during rest of the hospital stay, it uses the latest values of the features as they were at 24 hours from admission during that hospital stay. One-time prediction models to predict at other particular times are analogously built.

3.3.2. Features

Common types of features that have been used to predict AKI include demographic information, comorbidities, medications and laboratory values. In this work, for both one-time and continual models, we used the same features that had been used in our previous study [14]. These features were selected with the help of a nephrologist, who is a co-author. The only differences in the features used in this study from our previous study is that here we used separate pre-admission and post-admission features for each comorbidity and medication, and we also used prior AKI as a feature. Table 1 shows the distribution of the feature values in our data across AKI and non-AKI hospital stays for the features that do not change over hospital stays. These features include all the demographic features and prior AKI.

Table 1. Distribution of feature values across AKI and non-AKI hospital stays in the data for the features that do not change over hospital stays. The p-values for statistical significance were computed using "N-1" chi-squared test [36] for proportions and using *t*-test for means.

Variable	AKI (3786)	Non-AKI (40905)	*P*-value
Age	73.0 ± 9.0	73.0 ± 9.3	0.85
BMI	29.4 ± 8.3	28.5 ± 7.3	**< 0.01**
Gender = Male	2031 (53.7%)	20563 (50.3%)	**< 0.01**
Gender = Female	1755 (46.3%)	20342 (49.7%)	**< 0.01**
Race = White	3340 (88.2%)	37430 (91.5%)	**< 0.01**
Race = Black	331 (8.8%)	2440 (6.0%)	**< 0.01**
Race = Other	115 (3.0%)	1035 (2.5%)	0.06
Alcohol Use	1014 (26.8%)	13949 (34.1%)	**< 0.01**
Tobacco Use	408 (10.8%)	4771 (11.7%)	0.09
Drug Use	69 (1.8%)	472 (1.2%)	**< 0.01**
Prior AKI	618 (16.3%)	2838 (6.9%)	**< 0.01**

Table 2. Distribution of feature values across AKI and non-AKI hospital stays up to 24 hours and after 24 hours following admission for the features that can change over hospital stays. The hospital stays in which AKI developed within 24 hours of admission (483) were excluded for the purpose of this table. The differences shown in bold were found to be statistically significant ($p < 0.05$) using "N-1" chi-squared test [36] for proportions and using *t*-test for means.

	Up to 24 hours following admission				**After 24 hours following admission**		
	AKI (3,303)	**Non-AKI (40,905)**	**Difference**		**AKI (3,303)**	**Non-AKI (40,905)**	**Difference**
Laboratory Values (Means, Standard deviations)							
AST	30.0 ±18.1	26.4 ± 15.3	**3.58**		31.3 ± 17.9	29.1 ± 17.8	**2.2**
Blood Bilirubin	0.60 ± 0.37	0.57 ± 0.34	**0.03**		0.60 ± 0.37	0.58 ± 0.36	**0.02**
BP Diastolic	68.7 ± 10.5	70.3 ± 9.6	**-1.6**		65.6 ± 8.4	68.7 ± 8.3	**-3.1**
BP Systolic	131.2 ±19.8	133.9 ± 18.4	**-2.7**		125.9 ± 15.8	130.5 ± 15.8	**-4.6**
BUN	20.6 ± 9.7	18.6 ± 7.7	**2.0**		21.6 ± 9.5	17.9 ± 7.7	**3.7**
Heart Rate	82.3 ±14.6	79.8 ± 14.4	**2.5**		81.7 ± 12.8	77.8 ± 12.2	**3.9**
Platelet Count	220.5 ±87.8	219.9 ± 79.1	0.6		217.4 ± 88.0	212.4 ± 80.7	**5.0**
Temperature	98.2 ± 0.8	98.1 ± 0.6	0.1		98.2 ± 0.6	98.1 ± 0.5	**0.1**
Troponin	0.33 ± 0.72	0.29 ± 0.65	0.04		0.38 ± 0.82	0.38 ± 0.80	0
Comorbidities (Percentages)							
Coronary Artery Disease	0	0	0		0	0	0
Diabetes	5.21	4.36	**0.85**		6.96	5	**1.96**
Disorders of Lipoid Metabolism	1	1.68	**-0.68**		1.94	2.26	**-0.32**
Heart Failure	12.72	5.36	**7.36**		18.07	6.98	**11.09**
Hypercalcemia	0.36	0.19	0.17		0.67	0.26	**0.41**
Hyperlipidemia	2.27	3.6	**-1.33**		2.76	4.09	**-1.33**
Hypertension	26.73	30.4	**-3.67**		40.33	37.79	**2.54**
Pancreatitis	0.97	1.22	-0.25		1.27	1.35	-0.08
Respiratory Failure	1.94	1.05	**0.89**		3.39	1.35	**2.04**
Rhabdomyolysis	0.42	0.34	0.08		0.45	0.38	0.07
Sepsis	5.33	3.69	**1.64**		7.36	4.31	**3.05**
Thrombocytopenia	1.88	1.14	**0.74**		3.48	1.84	**1.64**
Medications (Percentages)							
ACE Inhibitors	17.62	16.06	**1.56**		26.1	19.05	**7.05**
ACE Inhibitors or NSAIDS or Diuretics	46.26	38.59	**7.67**		71.84	47.51	**24.33**
ACE Inhibitors or ARB or NSAIDS or Diuretics	49.35	43.38	**5.97**		74.72	51.88	**22.84**
Acylovir	1.18	0.79	**0.39**		2.36	1.11	**1.25**
Aminoglycosides	0.27	0.08	**0.19**		0.54	0.16	**0.38**
ARB	8.39	8.88	-0.49		12.44	9.96	**2.48**
Cisplatin	0.33	0.02	**0.31**		0.7	0.07	**0.63**
Diuretics	35.15	21.41	**13.74**		59.88	29.76	**30.12**
K Sparing	5.24	2.37	**2.87**		8.08	3.28	**4.8**
Lipid Lowering Drugs	37.63	39.03	-1.4		50.89	47.34	**3.55**
NSAIDS	5.24	9.14	**-3.9**		10.81	11.92	**-1.11**
Radiocontrast Dyes	16.65	18.15	**-1.5**		27.52	22.57	**4.95**

Table 2 shows the distribution of laboratory value, comorbidity and medication features across AKI and non-AKI hospital stays in our data. These features can change over hospital stays. Comorbidities and medications were binary features (true or false). The laboratory value features were the latest numeric values of the test results obtained since five days before admission till during the stay. To show that these features change over the course of hospital stays, Table 2 shows the same statistics up to 24 hours following admission and after 24 hours following admission till AKI incidence or till the end for non-AKI hospital stays. Averages are shown for the numeric features.

The hospital stays in which AKI developed within 24 hours from admission (483) were excluded from Table 2 because they are not useful for comparing feature values up to 24 hours and after 24 hours following admission. Also, only post-admission comorbidity and medication features are shown because they can change over hospital stays. Like previous studies on AKI prediction models, we did not use SCr as a feature because it is used as a gold-standard to determine AKI. In other words, the predictive model is expected to predict AKI in the absence of SCr measurement before the caregivers suspect the onset of AKI and order the SCr test. It can be observed from Table 2 that for most features, the differences between their values for AKI and non-AKI cases become more pronounced later during the hospital stays. For example, the difference in the use of diuretics between AKI and non-AKI cases was 13.74% up to 24 hours following admission, while it was 30.12% after 24 hours following admission. This shows that patient status changes significantly over hospital stays, and a one-time prediction model (e.g., predicting at 24 hours) will not be able to take this into account for predicting AKI.

3.3.3. Evaluation

The standard method for evaluating a machine learning model is through ten-fold cross-validation [37]. In this process, the data is first randomly divided into ten equal parts. Then, in each fold, nine parts are used for training the model and the remaining part (different in each fold) is used for testing it. The results of all the folds are combined and reported. Sensitivity (proportion of positive examples correctly identified) and specificity (proportion of negative examples correctly identified) are the most common evaluation measures for predictive models. Many machine learning methods give the probability (or confidence) of their prediction, and by varying the threshold value (we used a step size of 0.01) on a model's prediction confidence, sensitivity can be traded-off with specificity to generate an entire curve between them. Traditionally, a curve is plotted between sensitivity (true positive rate) and 1-specificity (false positive rate) and is called an ROC curve. The area under this curve (AUC) is one number that conveniently summarizes the entire curve. In our results, we reported AUC and used it for comparing different models.

3.4. Continual Prediction Model for AKI

We first define three terms *event*, *event-time* and *feature-snapshot* that will be useful in describing our new continual prediction model. We call *event* as any change in the value of any of the features during a hospital stay. For example, an event occurs any time a new medication is prescribed, or a new comorbidity is recorded, or a new laboratory test result becomes available. If multiple of these changes occur at the same time (to be more precise, if multiple of these get recorded in the EHR with the same timestamp) then all are considered as one event. Admission and discharge are also regarded as events. We call *event-time* as the timestamp of an event as recorded in the EHR. Lastly, we call the latest values of all the features at an event-time as its *feature-snapshot*. Figure 3 (a) shows a hypothetical illustrative timeline of a hospital stay with some events and their corresponding event-times and feature-snapshots. In our data, there were on average 21.5 (standard deviation=12.6) events per day, per hospital stay.

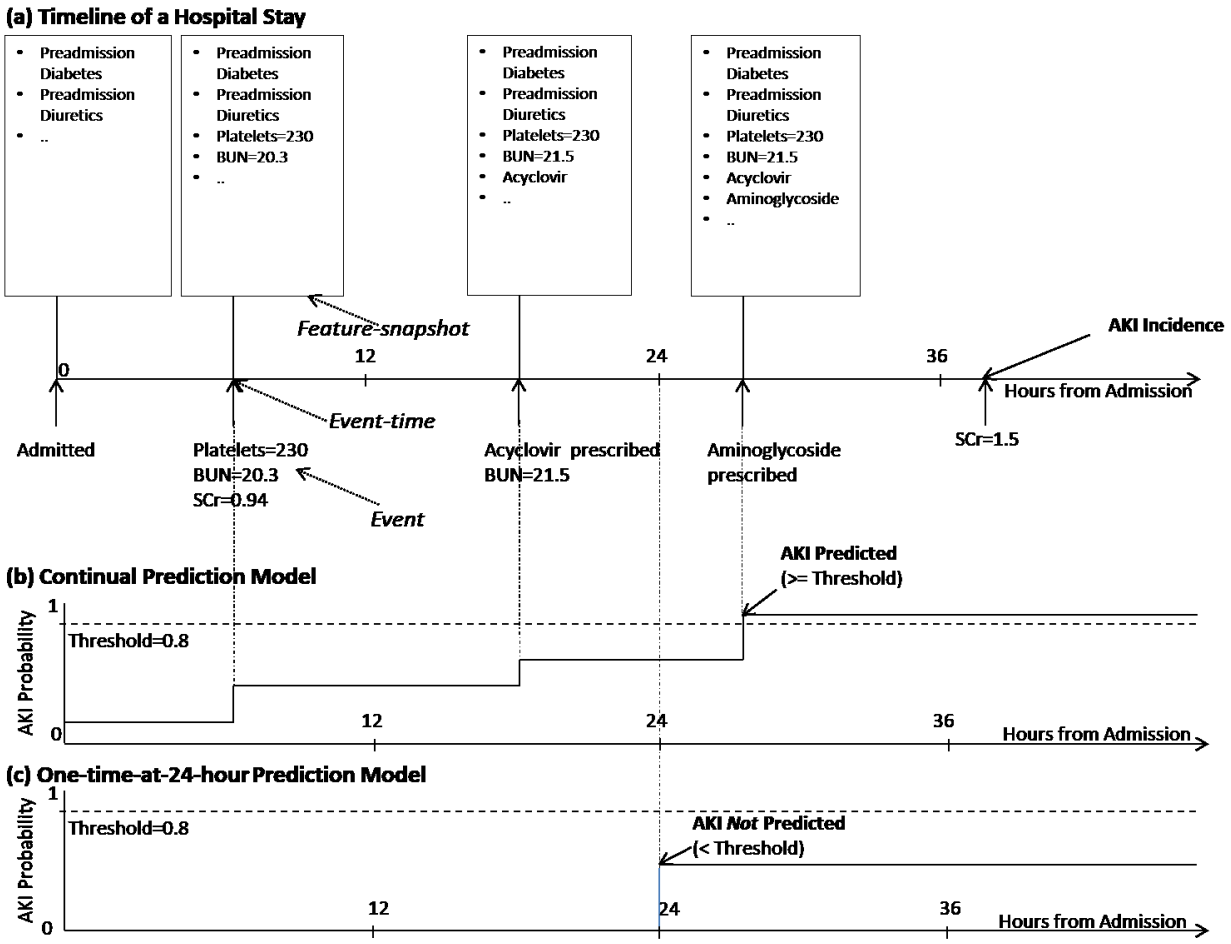

Figure 3. An illustrative timeline of a hospital stay along with the AKI prediction probabilities of the continual and the one-time-at-24-hour prediction models. The continual prediction model predicts AKI once aminoglycoside is prescribed sometime after 24 hours from admission but the one-time-at-24-hour prediction model fails to predict AKI at 24 hours from admission. AKI is determined sometime after 36 hours by the AKIN criteria. This is a hypothetical example for illustration purpose.

3.4.1. Applying the Model

Once trained (we describe in Section 3.4.3 how it is trained), our continual prediction model is applied to predict AKI at every event-time of a hospital stay starting from the admission time. There is no need to apply the model in between two successive events because, based on our definition of event, no feature can change in between two successive events (otherwise the two events will not be successive); therefore, the model's prediction will not change. To apply the model at an event-time, the values of the features are taken from that event-time's feature-snapshot.

Many machine learning methods give the probability of their prediction. For example, in our task, it is the probability that AKI will develop. Therefore, our model's AKI prediction probability dynamically changes during the course of a hospital stay as the feature values change. Figure 3 (b) illustrates this for the example hospital stay shown above it. To contrast this, Figure 3 (c) shows the static nature of the one-time-at-24-hour prediction model's prediction. Unlike one-time prediction models, the continual prediction model is able to take into account the latest medical status of the patient as it changes dynamically over the hospital stay. This helps the model to more accurately predict the AKI incidences that develop later in the hospital stays. Additionally, one-time prediction models often miss early AKI incidents because they are applied

too late in the hospital stay, but the continual prediction model is applied from the start of a patient's admission, and thus cannot miss early AKI incidents. Therefore, the continual prediction model circumvents the two problems inherent in one-time prediction models: being applied too early to accurately predict AKI or being applied too late to prevent AKI.

Although the continual prediction model gives the probability of AKI at every event-time, it will be highly impractical to alert healthcare providers about the probability of AKI every time a feature changes. This raises an important question: When is it appropriate to trigger an alarm for AKI as predicted by the model? Our method uses a threshold (a parameter) and whenever the AKI prediction probability exceeds this threshold it is deemed that the model has predicted AKI which should be alerted to the healthcare providers. Figure 3 (b) shows an example of AKI being predicted by the continual prediction model using a particular threshold value of 0.8 (this threshold value was arbitrarily chosen for illustration; for evaluation, threshold values between 0 and 1 in step-size of 0.01 were used to obtain different points on the ROC curve). The first event-time at which this happens is recorded as the time the model predicted AKI for that hospital stay. If the probability never exceeds the threshold over a hospital stay, then it is deemed that the model did not predict AKI for that hospital stay. When applying the model, a practitioner will choose the threshold that corresponds to the desired level of sensitivity and specificity from those that the model can deliver, as indicated in its ROC curve.

In the hypothetical illustrative example shown in Figure 3, AKI develops after aminoglycoside is prescribed sometime after 24 hours from admission. The continual prediction model is able to predict AKI when aminoglycoside is prescribed (the AKI probability increases and exceeds the threshold). On the other hand, the one-time-at-24-hour prediction model fails to predict AKI.

3.4.2. Evaluation

The continual prediction model we just introduced makes predictions multiple times over a hospital stay. A naïve way to evaluate such a dynamic prediction model will be to evaluate each of its predictions exactly in the way a one-time prediction model is evaluated for each of its predictions. This way, for example, in the example of Figure 3 (b), it would be deemed that the model missed predicting AKI the first three times it made predictions corresponding to the first three events (including at admission time), although it eventually correctly predicted AKI before its onset. It does not seem fair that the model should be penalized for the first three times it did not predict AKI when it later predicts it correctly before time. On the other hand, the continual prediction model may correctly predict non-AKI (or AKI) multiple times just because the hospital stay had many events and it made predictions at each of those events. In this case, the evaluation will unfairly boost the model's performance by counting correct predictions multiple times. In addition to these problems, this approach to evaluating the performance of dynamic prediction models does not lead itself to be directly comparable to the performance of one-time prediction models, which are evaluated for only one prediction per hospital stay. Hence we introduce the following method of evaluating the performance of dynamic prediction models. Instead of evaluating individual predictions over time, we evaluate the continual model's overall prediction for a hospital stay in terms of whether the model was able to correctly predict AKI in time or not and also whether it correctly predicted non-AKI. This evaluation also directly reflects the prediction model's usefulness in the clinical setting.

In our evaluation, for a hospital stay during which AKI had developed, the continual prediction model must predict AKI (based on the threshold) by the time of AKI incidence otherwise it will be considered that the model missed it (predicting it later will be too late). In other words, in our evaluation, to correctly predict AKI for a hospital stay, the continual prediction model's prediction probability for AKI must be above the chosen threshold *at least at one event-time before the time of AKI incidence*. The first event-time at which the prediction probability exceeds the threshold is regarded as the time the model predicted AKI. On the other hand, for a hospital stay during which AKI did not develop, the continual prediction model must not predict AKI (based on the threshold) at any event-time. In other words, in our evaluation, to correctly

predict non-AKI for a hospital stay, the continual prediction model's prediction probability for AKI should not be above the chosen threshold at *any* event-time over the entire hospital stay.

It is important to note that this is a very strict criterion for correct non-AKI prediction. In addition, the number of non-AKI hospital stays are more than 10 times the number of AKI hospital stays in our data, this makes the task for the continual prediction model especially difficult. The threshold offers a trade-off between sensitivity and specificity. A higher threshold will increase the model's specificity but decrease its sensitivity, and vice-versa. The rest of the evaluation proceeds in the same way as for the one-time prediction models described earlier. As noted in the Introduction section, this evaluation method comprehensively measures the overall performance of a dynamically predicting model, and it can also be used to directly compare with the performance of the traditional one-time prediction models.

3.4.3. Training the Model

For training the one-time-at-24-hour prediction model (or any one-time prediction model), it is clear that the feature values for positive (AKI) and negative (non-AKI) training examples must be collected at 24 hours from admission. However, for the continual prediction model, which has no specific time for prediction, an important question arises: At what time should the feature values be collected for positive and negative training examples? Ideally, for each hospital stay in the training data in which AKI developed, feature-snapshot of its each event-time before AKI should be used to create a positive example. This is because when the model is applied it is expected to predict AKI during those events (sooner the better). Similarly, for each hospital stay in the training data in which AKI did not develop, feature-snapshot of its each event-time till the end of hospital stay should be used to create a negative example. However, this will result in an explosion of the number of training examples (0.48 million positive and 3.2 million negative examples in our data) which will require unreasonable amounts of computational time and memory for training machine learning methods. Also, any two successive feature-snapshots will minimally differ from each other, hence most of the examples thus created will be largely repetitious.

In this work, we decided to take feature-snapshot at the time AKI was determined (i.e., when the SCr measurement satisfied the AKIN criteria) for creating positive training examples, and at the time last SCr measurement was taken for creating negative training examples. This way we create only one training example per hospital stay thus keeping under control the computational time and memory requirement for training machine learning methods. We decided to use the time of AKI for creating positive training examples because it best represents the status of patients who have acquired AKI, with the AKI-relevant features taking their AKI-indicative values. We also tried adding feature-snapshots from random event-times to create positive examples, but it did not improve results. This shows that there is nothing more to learn from earlier feature-snapshots. We decided to use the time of last SCr for creating negative training examples because it shows that the patient was still considered to be prone to AKI. As a result, the model will learn to better distinguish between AKI and non-AKI cases. If we had used, for example, feature-snapshot of the discharge time for creating negative examples then the model may simply learn to predict whether the patient is about to be discharged or not. We also tried random event-times to create negative examples, but it did not lead to any noticeable difference in the results. We point out that training a prediction model with the examples as described above is similar to the "AKI detection" model we had built in our past work [14]. However, that model was not applied continually and its purpose was to determine if AKI had already developed in order to prevent it from going undiagnosed.

We want to point out that the one-time-at-24-hour prediction model will do worse if it is trained with feature values collected at the same times as for the continual prediction model, instead of collected at 24 hours from admission as we did. This is because in order to obtain the best performance, a machine learning model must be trained with the same type of examples that it will encounter during testing. Instead, if the one-time-at-24-hour prediction model is trained with feature values collected at times close to AKI incidences then it will learn to make predictions based on those features that change over hospital stays.

But when applied at 24 hours from admission, which is too early for many hospital stays, these features may not have yet changed and thus the model will perform poorly. To make prediction at 24 hours from admission, a model has to rely primarily on features that do not change over hospital stays, such as pre-admission comorbidities and pre-admission medications (we later show this to be the case in the Appendix). The model will learn to make predictions using these features only when it is trained with feature values collected at 24 hours from admission, which is the time when the model will be applied during testing.

In our previous work [14] we compared several machine learning methods for building AKI prediction models and found that logistic regression [38] works best. Since the goal of the current study was to compare the proposed continual prediction model with the traditional one-time prediction models, we only used logistic regression to build both types of models. Unlike our previous work, comparing different machine learning methods was not a goal of this study. In this work, we used the Weka machine learning software [39] to build logistic regression models. Missing feature values were handled by Weka's default mechanism for logistic regression, which replaces missing values with the modes for nominal features and with the means for numeric features as computed from the training data. Given the unbalanced nature of our data, with more than 10 times the number of non-AKI examples than AKI examples, we used Weka's cost-sensitive classifier whose weight for the minority class was decided out of 1, 2, 4, 6, …, 18, 20 through internal cross-validation within the training data.

Table 3. Comparison of continual prediction model and one-time-at-24-hour predcition model for AKI. An AUC number in bold was found to be statistically significant compared to the AUC number in the same row ($p < 0.05$; two-tailed paired *t*-test).

	AKI Hospital Stays	**Non-AKI Hospital Stays**	**Continual Prediction Model AUC (95% CI)**	**One-time-at-24-hour Prediction Model AUC (95% CI)**
Excluding hospital stays in which AKI developed within 24 hours from admission	3303	40905	**0.724** (0.705, 741)	0.653 (0.641, 665)
All hospital stays	3786	40905	**0.709** (0.690, 0.728)	0.57 (0.555, 0.584)

4. Results and Discussion

4.1. Continual vs. One-time Prediction Models

We first compare our proposed continual prediction model with the one-time-at-24-hour prediction model and then with many other one-time prediction models. Table 3 shows a comparison between AUC obtained by the continual prediction model and by the one-time-at-24-hour prediction model through 10-fold cross-validation using exactly the same folds. The first row shows results in which we exclude hospital stays in which AKI developed within 24 hours from admission (483 such hospital stays), because it is beyond the capacity of the one-time-at-24-hour prediction model to predict these AKI incidences. The continual prediction model obtained statistically significantly better AUC than the one-time prediction model (0.724 vs. 0.653; $p < 0.05$; two-tailed paired *t*-test).

In the second row we show results for all hospital stays including the 483 hospital stays in which AKI developed within 24 hours after admission. The one-time-at-24-hour prediction model cannot predict these 483 AKI incidences; as a result, its AUC dropped to 0.57 while the AUC of the continual prediction model was 0.709. Figure 4 shows the corresponding ROC curves. Overall, this is a more realistic comparison because it does not exclude the cases which are beyond the capacity of the one-time prediction model. However, in prior studies, the results for one-time prediction models were always reported after excluding these cases, and thus those results looked better than they really were. It should be noted that the ROC curve

for the one-time-at-24-hour prediction model in Figure 4 does not end at the top-right corner. The reason for this is that even with zero threshold value for confidence, which usually yields maximum sensitivity of 1, the one-time-at-24-hour model simply cannot predict AKI for those 483 hospital stays because the prediction time of 24 hours is already too late for them.

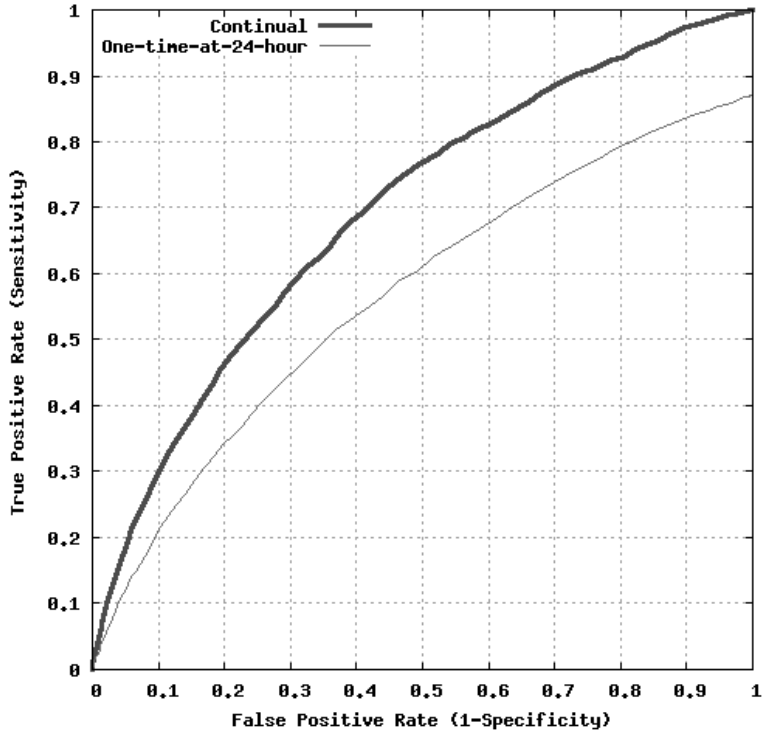

Figure 4. ROC curves of the continual prediction model and the one-time-at-24-hour prediction model when evaluated on all hospital stays.

To show that the continual prediction model is also better than other one-time prediction models besides the one-time-at-24-hour prediction model, we generated results for different one-time prediction models that predict from 0 hours (i.e., at the time of admission) to 120 hours from admission. Each of these models requires separate training and testing for its time of prediction. The second column of Table 4 shows the results for all hospital stays. The next column shows results of evaluation after excluding hospital stays in which AKI developed before the time of prediction as well as after excluding hospital stays which were shorter than the time of prediction. Finally, the last column shows a direct comparison of all the models on the same set of hospital stays on which none of them was applied too late. These results exclude hospital stays in which AKI developed within 120 hours as well as exclude hospital stays that were less than 120 hours. For comparison, the last row shows the corresponding results of the continual prediction model.

As Table 4 shows, none of the one-time prediction models performed better than the continual prediction model. As pointed out earlier, if the one-time prediction is made too late then it will miss AKI cases that developed earlier; the drop in AUC in the second column clearly shows this. It is interesting to note that the last column of Table 3 changes very little over different times of one-time prediction models, indicating that a later one-time prediction model is not necessarily better at predicting AKI incidences that develop later. This may seem counter-intuitive, but it has an explanation. Using information gain statistic [40] we show in the Appendix that the pre-admission features are the most important features for one-time prediction models. This is because these models are supposed to predict far ahead in time and hence the other features that later change over hospital stays are not helpful in making such early predictions. Hence

all the one-time prediction models predominantly use pre-admission features to make predictions. And given that the pre-admission features do not change over hospital stays, a later one-time prediction model is not necessarily better at predicting AKI that develops later. We also show in the Appendix that, in contrast, the features that change over hospital stays are the most important features for the continual prediction model.

Table 4. Comparison of one-time prediction models with different times of predictions. All evaluation numbers are AUC.

One-time prediction at X hours **X =**	**All hospital stays (AKI developed within X hours will be considered missed)** (AKI: 3786; Non-AKI: 40905)	**Excluding hospital stays shorter than X hours or in which AKI developed within X hours**	**Excluding hospital stays shorter than 120 hours or in which AKI developed within 120 hours** (AKI: 1493; Non-AKI: 12274)
0	0.653	0.653	0.597
12	0.637	0.655	0.594
24	0.570	0.653	0.592
36	0.560	0.650	0.593
48	0.525	0.641	0.595
60	0.543	0.629	0.595
72	0.535	0.634	0.607
84	0.559	0.628	0.607
96	0.553	0.624	0.609
108	0.566	0.611	0.607
120	0.560	0.607	0.607
Continual	0.709	-	0.651

4.2. Advance Prediction

One can also evaluate retrospectively how well a model could predict AKI certain time in advance, for example, how well a model could predict AKI at least 6 hours in advance, or at least 12 hours in advance, etc. This may be desirable to know in order to have sufficient time to respond to prevent AKI. To the best of our knowledge, no previous study has retrospectively evaluated such advance AKI prediction. Figure 6 shows how the AUC varied with the number of hours in advance a model must predict AKI before it developed. For this evaluation, the model may predict earlier than the number of hours required in advance, but if it predicted any later then it will be considered too late and the prediction will not count as correct. For example, if the model must predict AKI at least 12 hours in advance, then its time of prediction must be 12 hours before AKI developed.

For the one-time-at-24-hour prediction model the time of prediction is fixed at 24 hours, but for the continual prediction model the time of prediction is the first event-time when the AKI probability exceeds the threshold (see Figure 1 illustration). For a fair comparison across all time lengths from 0 to 24 hours, for this plot, we considered only those AKI incidences that developed at least 24 hours after admission (for example, if AKI developed in less than 24 hours, then one could not have possibly predicted it 24 hours in advance). We plotted the curves with 2-hour granularity which was sufficient to obtain smooth curves; they can be otherwise plotted at any level of granularity. It is important to note that the continual prediction model still makes predictions continually, here we are only evaluating it in 2 hour intervals (i.e., AUC for

prediction by 6 hours in advance, AUC for prediction by 8 hours in advance, etc.). Figure 6 shows, as one would expect, that for both types of models it is harder to predict AKI farther ahead in time. The AUC seemed to slowly drop, almost linearly, with the number of hours in advance by when the prediction is required. However, the performance of the one-time-at-24-hour model drops more precipitously than the continual prediction model.

We want to contrast our evaluation with a recent study [41] in which in order to evaluate the performance of an AKI prediction model to make predictions at, for example 24 hours in advance, the model was actually applied 24 hours before each AKI incidence. As was pointed out by He et al. [21], the framework of advance prediction presented in [41] is unrealistic and not possible to deploy in practice because one cannot know in advance the time of an AKI incidence in order to determine when the model should be applied (in fact, that defeats the very purpose behind the prediction model). In contrast, in our evaluation, a model makes prediction when it makes prediction (as depicted in the illustrative example in Figure 1) and it is then evaluated whether that prediction was made, say, by 24 hours in advance of the AKI incidence. Hence our evaluation does not require knowing the time of an AKI incidence beforehand in order to determine when to predict it. Therefore, our framework is deployable in practice for advance prediction, and our evaluation also realistically measures the advance prediction performance.

One may notice in Figure 6 that the AUC of the one-time-at-24-hour prediction model for making prediction by 24 hours in advance is nearly 0.5. A random prediction model also gives an AUC of 0.5 with a diagonal ROC curve. However, we want to point out that the one-time-at-24-hour prediction model still performs better than a random prediction model that makes prediction at 24 hours after admission. The reason is as follows. When the one-time-at-24-hour prediction model is required to predict AKI by 24 hours in advance, it simply cannot predict AKI incidences that occur between 24 hours and 48 hours of admission. As a result, those incidences become out of reach for the model to predict and thus lowers the maximum sensitivity it can obtain. Consequently, its ROC curve does not reach the top right corner. A similar situation was shown in Figure 4 for a curve with an AUC of 0.57. In a similar way, a random prediction model that predicts at 24 hours after admission and which needs to predict AKI by 24 hours in advance also cannot predict AKI incidences that occur between 24 hours and 48 hours after admission. As a result, its ROC curve will also not reach the top right corner and consequently its AUC will be lower than 0.5. Therefore, the one-time-at-24-hour learned prediction model still performs better than a random prediction model. Of note, because the continual prediction model has no fixed time of prediction, there are no AKI incidences which are beyond its reach to predict in advance. For instance, when the continual prediction model is required to predict AKI by 24 hours in advance, it can predict AKI incidences that occur between 24 hours and 48 hours of admission by making those predictions any time between 0 and 24 hours of admission.

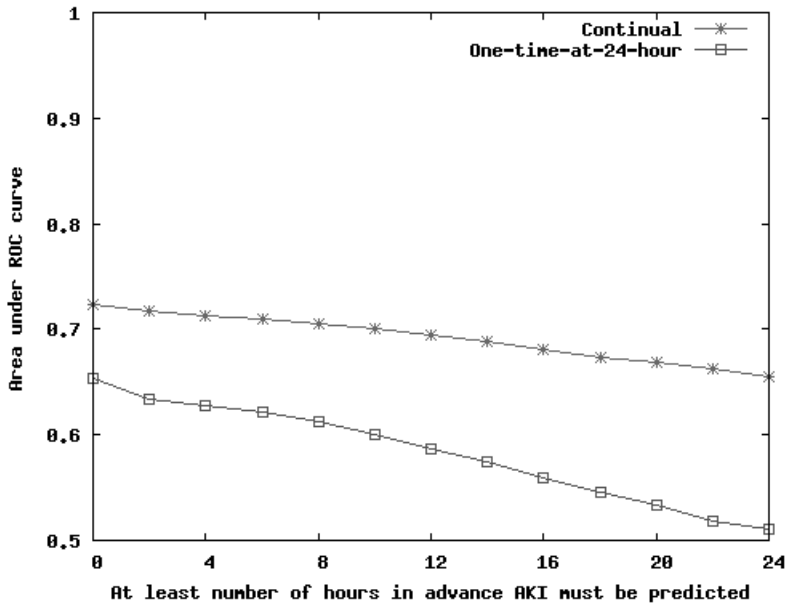

Figure 6. A graph showing how the prediction performance changed with the number of hours in advance by when a model must predict AKI.

Table 5. Top ten medication and comorbidity features whose prescription or diagnosis during a hospital stay were associated with a change in the model's prediction from non-AKI to AKI. The numbers in the last column are average values obtained over all confidence thresholds.

Rank	Medication or Comorbidity Feature	Percentage of hospital stays over which the feature changed (%)	How often change in feature over hospital stay changed prediction from non-AKI to AKI (%)
1	Cisplatin	0.123	25.6
2	Aminoglycosides	0.186	15.4
3	Hypercalcemia	0.311	5.81
4	Diuretics	36.4	3.84
5	Acyclovir	1.27	3.6
6	ACE inhibitors or NSAIDS or Diuretics	56.4	2.53
7	ARB or ACE inhibitors or NSAIDS or Diuretics	61.5	2.29
8	Respiratory failure	1.59	2.17
9	Rhabdomyolysis	0.403	2
10	K sparing	4.19	1.86

4.3. Dynamic Predictive Features

Another advantage of our continual prediction framework is that we can directly observe how the model's prediction dynamically changes as the values of the features change over the course of hospital stays. Table 5 shows this for the top ten medication and comorbidity features whose prescription or diagnosis during a hospital stay were associated with a change in the model's prediction from non-AKI to AKI. The third column shows the percentage of hospital stays over which a particular medication was prescribed or a particular comorbidity was diagnosed. The fourth (last) column shows how often the model's prediction changed from non-AKI to AKI whenever the medication was prescribed, or the comorbidity was diagnosed over the hospital stay (it is possible that other features could have also changed at the same time). These

numbers can be computed for every confidence threshold, but the table shows the combined results for all confidence thresholds instead of a particular threshold. Laboratory values are numeric features that can change multiple times during a hospital stay in different magnitudes and thus they are not included in this analysis.

It can be seen from the table that prescriptions of medications cisplatin and aminoglycosides were most prominently associated with the change in prediction from non-AKI to AKI. The third column shows that these were not prescribed frequently though. The next two medications to follow were diuretics and acyclovir. All these medications are, in fact, known for their nephrotoxicity. Hypercalcemia and respiratory failure were found to be the most prominent among comorbidities which, when diagnosed during a hospital stay, were most often associated with change in prediction from non-AKI to AKI. Such dynamic feature information was never reported in past work and could not have been obtained without a continual prediction model.

4.4. Limitations and Future Work

Our data was limited to one system of hospitals, which limits generalization. The data was obtained only from the structured part of EHR. In the future, natural language processing techniques [42] could be used to extract AKI-relevant features and their values from the text part of EHR. Our continual prediction model always used the latest values of all the features; in the future, the model could be improved by also taking into account the past values of the features and how they vary over time, for example, by mining temporal clinical event patterns [43] or by using temporal models such as recurrent neural networks [44]. Surgery is a major factor that leads to AKI, and a future study that specifically looks at surgery patients and their heterogeneity is another possibility.

Furthermore, our study was retrospective. It remains to be seen how our continual prediction framework will perform when deployed in a clinical decision support built into an EHR, and how much that will improve AKI predictions compared with predictions made by humans alone. However, it should not be difficult to integrate our continual prediction framework in an EHR system, because the trained logistic regression model is only a mathematical equation (this is also true for many other machine learning methods) that requires only the latest values of patient variables to make predictions, which are already present in the EHR. The system can then be made to trigger an alarm whenever the AKI prediction probability exceeds the threshold set corresponding to the desired level of sensitivity and specificity.

5. Conclusions

A new framework of continual prediction from EHR data was introduced and applied for predicting AKI. Instead of applying the trained model at a particular chosen time, as has been done in the past, the continual prediction model was applied continually over entire hospital stay whenever any patient variable changed. A method to evaluate the overall performance of a continual prediction model was also presented. Our experiments on a large dataset showed that the model out-performed one-time prediction models. Unlike one-time prediction models, the continual prediction model can take into account the latest values of variables as they dynamically change over a hospital stay. The continual prediction model also circumvents the shortcomings faced by one-time prediction models from either being applied too early to accurately predict AKI or too late to prevent AKI. It also performed well when evaluated for predicting AKI at desired number of hours in advance.

Conflict of Interest

None declared.

References

1. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology. 2005;16(11):3365-70.

2. Hou SH, Bushinsky DA, Wish JB, Cohen JJ, Harrington JT. Hospital-acquired renal insufficiency: a prospective study. The American journal of medicine. 1983;74(2):243-8.

3. Nash K, Hafeez A, Hou S. Hospital-acquired renal insufficiency. American Journal of Kidney Diseases. 2002;39(5):930-6.

4. Thakar CV, Christianson A, Freyberg R, Almenoff P, Render ML. Incidence and outcomes of acute kidney injury in intensive care units: a Veterans Administration study. Critical Care Medicine. 2009;37(9):2552-8.

5. Kes P, Jukić NB. Acute kidney injury in the intensive care unit. Bosnian journal of basic medical sciences. 2010;10(Suppl 1):S8.

6. Ostermann M, Chang RW. Acute kidney injury in the intensive care unit according to RIFLE. Critical care medicine. 2007;35(8):1837-43.

7. Sileanu FE, Murugan R, Lucko N, Clermont G, Kane-Gill SL, Handler SM, Kellum JA. AKI in low-risk versus high-risk patients in intensive care. Clinical Journal of the American Society of Nephrology. 2014:CJN-03200314.

8. Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, Edipidis K, Forni LG, Gomersall CD, Govil D, Honoré PM. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive care medicine. 2015;41(8):1411-23.

9. Coca SG. Acute kidney injury in elderly persons. American Journal of Kidney Diseases. 2010;56(1):122-31.

10. Anderson S, Eldadah B, Halter JB, Hazzard WR, Himmelfarb J, Horne FM, Kimmel PL, Molitoris BA, Murthy M, O'Hare AM, Schmader KE. Acute kidney injury in older adults. Journal of the American Society of Nephrology. 2011;22(1):28-38.

11. Yamout H, Levin ML, Rosa RM, Myrie K, Westergaard S. Physician prevention of acute kidney injury. The American Journal of Medicine. 2015;128(9):1001-6.

12. Sutherland SM, Chawla LS, Kane-Gill SL, Hsu RK, Kramer AA, Goldstein SL, Kellum JA, Ronco C, Bagshaw SM. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15 th ADQI Consensus Conference. Canadian Journal of Kidney Health and Disease. 2016;3(1):11.

13. Hodgson LE, Sarnowski A, Roderick PJ, Dimitrov BD, Venn RM, Forni LG. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. BMJ open. 2017;7(9):e016591.

14. Kate RJ, Perez RM, Mazumdar D, Pasupathy KS, Nilakantan V. Prediction and detection models for acute kidney injury in hospitalized older adults. BMC Medical Informatics and Decision Making. 2016;16(1):39.

15. Bedford M, Stevens P, Coulton S, Billings J, Farr M, Wheeler T, Kalli M, Mottishaw T, Farmer C. Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study. Health Service Delivery Research. 2016;4(6).

16. Cronin RM, VanHouten JP, Siew ED, Eden SK, Fihn SD, Nielson CD, Peterson JF, Baker CR, Ikizler TA, Speroff T, Matheny ME. National Veterans Health Administration inpatient risk stratification models for hospital-acquired acute kidney injury. Journal of the American Medical Informatics Association. 2015;22(5):1054-71.

17. Schneider DF, Dobrowolsky A, Shakir IA, Sinacore JM, Mosier MJ, Gamelli RL. Predicting acute kidney injury among burn patients in the 21st century: a CART analysis. Journal of Burn Care & Research: Official Publication of the American Burn Association. 2012;33(2):242.

18. Tsai TT, Patel UD, Chang TI, Kennedy KF, Masoudi FA, Matheny ME, Kosiborod M, Amin AP, Weintraub WS, Curtis JP, Messenger JC. Validated contemporary risk model of acute kidney injury in patients undergoing percutaneous coronary interventions: insights from the National Cardiovascular Data Registry Cath-PCI Registry. Journal of the American Heart Association. 2014;3(6):e001380.

19. Slankamenac K, Beck-Schimmer B, Breitenstein S, Puhan MA, Clavien PA. Novel prediction score including pre-and intraoperative parameters best predicts acute kidney injury after liver surgery. World Journal of Surgery. 2013;37(11):2618-28.

20. Demirjian S, Schold JD, Navia J, Mastracci TM, Paganini EP, Yared JP, Bashour CA. Predictive models for acute kidney injury following cardiac surgery. American Journal of Kidney Diseases. 2012;59(3):382-9.

21. He J, Hu Y, Zhang X, Wu L, Waitman LR, Liu M. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. JAMIA Open. 2018;2(1):115-22.

22. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986:121-30.

23. Miller Jr RG. Survival Analysis. 2nd ed. John Wiley & Sons; 1998.

24. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford: University Press; 2003.

25. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, Levin A, Levey AS. A predictive model for progression of chronic kidney disease to kidney failure. JAMA. 2011;305(15):1553-9.

26. Zhou J, Liu J, Narayan VA, Ye J. Modeling disease progression via multi-task learning. NeuroImage. 2013;78:233-48.

27. Liu YY, Ishikawa H, Chen M, Wollstein G, Schuman JS, Rehg JM. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden markov model. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2013:444-451.

28. Sukkar R, Katz E, Zhang Y, Raunig D, Wyman BT. Disease progression modeling using hidden Markov models. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE 2012 (pp. 2845-2848). IEEE.

29. Moskovitch R, Shahar Y. Classification-driven temporal discretization of multivariate time series. Data Mining and Knowledge Discovery. 2015;29(4):871-91.

30. Arandjelović O. Discovering hospital admission patterns using models learnt from electronic hospital records. Bioinformatics. 2015;31(24):3970-6.

31. Vasiljeva I, Arandjelović O. Diagnosis Prediction from Electronic Health Records Using the Binary Diagnosis History Vector Representation. Journal of Computational Biology. 2017;24(8):767-86.

32. Palevsky PM, Liu KD, Brophy PD, Chawla LS, Parikh CR, Thakar CV, Tolwani AJ, Waikar SS, Weisbord SD. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for acute kidney injury. American Journal of Kidney Diseases. 2013;61(5):649-72.

33. Drawz PE, Miller RT, Sehgal AR. Predicting hospital-acquired acute kidney injury—a case-controlled study. Renal Failure. 2008;30(9):848-55.

34. Matheny ME, Miller RA, Ikizler TA, Waitman LR, Denny JC, Schildcrout JS, Dittus RS, Peterson JF. Development of inpatient risk stratification models of acute kidney injury for use in electronic health records. Medical Decision Making. 2010;30(6):639-50.

35. Kane-Gill SL, Sileanu FE, Murugan R, Trietley GS, Handler SM, Kellum JA. Risk factors for acute kidney injury in older adults with critical illness: a retrospective cohort study. American Journal of Kidney Diseases. 2015;65(6):860-9.

36. Campbell I. Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. Statistics in medicine. 2007;26(19):3661-75.

37. Japkowicz N, Shah M. Evaluating Learning Algorithms: A Classification Perspective. Cambridge: University Press; 2011.

38. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.

39. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter. 2009;11(1):10-8.

40. Kent JT. Information gain and a general measure of correlation. Biometrika 1983;70(1):163-173.

41. Cheng P, Waitman LR, Hu Y, Liu M. Predicting Inpatient Acute Kidney Injury over Different Time Horizons: How Early and Accurate? In Proceedings of the American Medical Informatics Association (AMIA) Annual Symposium, pp. 565-574, Washington, DC, November 2017.

42. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of Biomedical Informatics. 2009;42(5):760-72.

43. Li J, Tan X, Xu X, Wang F. Efficient mining template of predictive temporal clinical event patterns from Patient Electronic Medical Records. IEEE journal of biomedical and health informatics. 2019;23(5):2138-2147.

44. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting clinical events via recurrent neural networks. In Machine Learning for Healthcare Conference 2016 (pp. 301-318).

Appendix

To see the difference between continual and one-time prediction frameworks, the following tables show the information gain statistic [40] for all the features. The first table shows for the data that is used to train the one-time-at-24-hour prediction model; the second table shows for the data that is used to train the continual prediction model (information gain is a property of the data and is independent of any model). The Weka machine learning software was used to obtain these numbers.

A higher value of the information gain statistic for a feature means that the feature is more helpful in distinguishing between AKI and non-AKI examples. The features are shown in the decreasing order of information gain. It can be observed that for the one-time prediction, the most important features are pre-admission comorbidities and pre-admission medications. The features that change over hospital stays are not important for a one-time prediction because such a model is supposed to make predictions far ahead in time (for example, incidence of AKI may occur even after 120 hours from admission) and at 24 hours from admission these features are not helpful in distinguishing between AKI and non-AKI hospital stays (as indicated by their low information gain statistic). We obtained similar trend of information gain statistic for other one-time prediction models as well. This also shows why a later one-time prediction model is not necessarily better at predicting AKI that develop later (as shown by the last column of Table 4 in the article that changes very little). On the other hand, for continual prediction, the features that change during hospital stays are the most important features because they are helpful in distinguishing between AKI and non-AKI hospital stays.

1. Features given to the one-time prediction model that predicts at 24 hours from admission ranked by the information gain statistic. The top features are preadmission features.

Information Gain	Feature
0.006935	Preadmission Diuretics
0.006867	Preadmission Heart Failure
0.006867	Preadmission Congestive Heart Failure
0.005041	Prior AKI
0.004813	Preadmission Respiratory Failure
0.003634	Preadmission K Sparing
0.002644	Preadmission Radiocontrast Dyes
0.002591	Preadmission Diabetes

0.002451	Preadmission Sepsis
0.002281	Preadmission ACE Inhibitors or NSAIDS or Diuretics
0.002082	Preadmission ARB or ACE Inhibitors or NSAIDS or Diuretics
0.001865	Preadmission ACE Inhibitors
0.001715	Preadmission Coronary Artery Disease
0.001107	Preadmission Lipid Lowering Drugs
0.000888	BP Systolic
0.000826	Preadmission Hypertension
0.000821	Preadmission ARB
0.000737	Race
0.000735	Preadmission Thrombocytopenia
0.000704	BP Diastolic
0.00041	Alcohol
0.000345	Tobacco
0.000327	Gender
0.000326	Heart Rate
0.000313	BMI
0.000281	Preadmission Aminoglycosides
0.00025	BUN
0.000223	Hypertension
0.000153	Drug
0.000147	Preadmission Hypercalcemia
0.000123	Preadmission Acyclovir
8.24E-05	Preadmission Cisplatin
7.97E-05	Preadmission Disorders of Lipoid Metabolism
5.26E-05	Heart Failure
5.26E-05	Congestive Heart Failure
4.46E-05	AST
4.11E-05	Preadmission Rhabdomyolysis
3.79E-05	NSAIDS
3.41E-05	Diuretics
2.63E-05	Respiratory Failure
1.89E-05	Thrombocytopenia
1.85E-05	Sepsis
1.7E-05	Preadmission NSAIDS
7.72E-06	Troponin
7.6E-06	Pancreatitis
6.18E-06	Disorders of Lipoid Metabolism
5.07E-06	Hypercalcemia
4.03E-06	Preadmission Hyperlipidemia
2.53E-06	Rhabdomyolysis

2.53E-06	ACE Inhibitors
1.19E-06	Diabetes
1.08E-06	ARB or ACE Inhibitors or NSAIDS or Diuretics
1.08E-06	ACE Inhibitors or NSAIDS or Diuretics
9.58E-07	Radiocontrast Dyes
2.63E-09	Hyperlipidemia
0	Blood Bilirubin
0	Acyclovir
0	Platelets
0	Preadmission Pancreatitis
0	Coronary Artery Disease
0	Temperature
0	Cisplatin
0	Aminoglycosides
0	K Sparing
0	Lipid Lowering Drugs
0	ARB
0	Age

2. Features given to the continual prediction model ranked by the information gain statistic. The top features are those that change during hospital stays.

Information Gain	**Feature**
0.015545	Diuretics
0.014529	BP Systolic
0.012971	BP Diastolic
0.008671	ACE Inhibitors or NSAIDS or Diuretics
0.008189	Heart Rate
0.00816	Preadmission Diuretics
0.00748	ARB or ACE Inhibitors or NSAIDS or Diuretics
0.007039	Preadmission Heart Failure
0.007039	Preadmission Congestive Heart Failure
0.006287	Congestive Heart Failure
0.006287	Heart Failure
0.00597	Temperature
0.005515	Prior AKI
0.004994	Preadmission Respiratory Failure
0.004883	BUN
0.003912	Preadmission K Sparing
0.003288	Platelets
0.00309	Preadmission Diabetes
0.002884	Preadmission Radiocontrast Dyes

0.002803	Preadmission ACE Inhibitors or NSAIDS or Diuretics
0.002744	Preadmission Sepsis
0.002601	Preadmission ARB or ACE Inhibitors or NSAIDS or Diuretics
0.00222	Preadmission ACE Inhibitors
0.00206	K Sparing
0.001759	Preadmission Coronary Artery Disease
0.001608	AST
0.001371	Preadmission Lipid Lowering Drugs
0.001154	Respiratory Failure
0.001153	Preadmission Hypertension
0.001094	BMI
0.000995	Preadmission ARB
0.000986	Sepsis
0.000881	Preadmission Thrombocytopenia
0.000803	Race
0.000802	Cisplatin
0.000802	ACE Inhibitors
0.000416	Acyclovir
0.000403	Alcohol
0.000394	Thrombocytopenia
0.000377	Tobacco
0.000356	Hyperlipidemia
0.000339	Radiocontrast Dyes
0.000327	Diabetes
0.000295	Blood Bilirubin
0.000255	Gender
0.000244	Preadmission Aminoglycosides
0.000225	Aminoglycosides
0.000188	Hypercalcemia
0.000185	Drug
0.000139	ARB
0.000132	NSAIDS
0.00013	Preadmission Acyclovir
0.000128	Preadmission Hypercalcemia
0.000116	Troponin
7.82E-05	Preadmission Disorders of Lipoid Metabolism
6.82E-05	Preadmission Rhabdomyolysis
4.65E-05	Preadmission Cisplatin
2.88E-05	Disorders of Lipoid Metabolism
2.01E-05	Rhabdomyolysis
1.25E-05	Hypertension

7.27E-06	Lipid Lowering Drugs
6.95E-06	Preadmission NSAIDS
2.86E-06	Coronary Artery Disease
1.86E-06	Preadmission Hyperlipidemia
2.95E-07	Pancreatitis
0	Preadmission Pancreatitis
0	Age