Discovering predictive temporal patterns for Acute Kidney Injury from critical care data

Beatrice Amico, PhD[1], Carlo Combi, PhD[1], Giovanni Gambaro, MD, PhD[1]
[1]University of Verona, Verona, Italy

Abstract

Acute Kidney Injury is a severe clinical condition with a high risk of multi-organs complications and mortality. For this reason, early recognition is crucial. Our proposal based on a 3-window framework discovers all hidden regularities, called Approximate Predictive Functional Dependencies, with the aim to enable early recognition of high-risk patients during hospitalization in the Intensive Care Unit (ICU). We evaluated the different severity stages according to the Kidney Disease Improving Global Outcomes (KDIGO) guidelines, building different pathological state patterns, from admission to the discharge from ICU. According to the clinical practice, for each patient, we examined various characteristics expressed as a temporal history of events that may predict a pathological state pattern. We evaluated our proposal exploiting the MIMIC-IV dataset, a collection of Electronic Medical Records from ICU. The obtained results showed promising possibilities to use this type of dependency to support clinical practice.

Introduction

Current clinical database systems have the capacity to store vast amounts of data. However, extracting relevant knowledge from these large datasets requires sophisticated data mining techniques that can handle the temporal dimension of the data. Temporal data mining is a crucial research area that can help us gain a deeper understanding of complex systems and phenomena that evolve over time, revealing hidden temporal knowledge.
The analysis of these clinical data sources holds great potential for predicting medical events, understanding various disease mechanisms, and improving patient care quality. However, defining a pathological state can be challenging as it often depends on inherently temporal criteria. Therefore, temporal data mining plays a critical role in analyzing clinical data to extract valuable insights and improve patient outcomes.

One clinical example that illustrates this issue is Acute Kidney Injury (AKI), a critical event characterized by a sudden loss of kidney function. AKI is typically diagnosed based on increased serum creatinine levels (a marker of kidney excretory function) and reduced urinary output (oliguria) (1). Unfortunately, the management of AKI is often suboptimal, and there is currently no real cure for this syndrome. Therefore, the early prediction of deterioration could play a central role in improving patient outcomes. Identifying an increased risk of AKI sufficiently in advance to perform preventative actions before the event occurs could be crucial. To delineate more precisely the patient's healthy/pathological state, two main needs must be addressed: (i) representing the evolution of potentially changing pathological states, which must be periodically re-evaluated, and (ii) enabling the early prediction of such states, as they may be known after some (possibly varying) delay. This could help reduce the overall risk for patients and allow for the timely preparation of suitable therapies and interventions.

Based on the depicted scenario, our approach offers the opportunity to identify patients at risk within a time window that enables potential early treatments. Specifically, we adopt a 3-window framework and use Approximate Predictive Functional Dependencies (APFDs), as partially presented in (2), to discover hidden knowledge expressed as dependencies between the patients' history and the following evolution of AKI stages over time. We apply this original approach to the MIMIC-IV database. Apart from discussing the technical details, we also delve into the clinical implications and the most significant mined APFDs. By leveraging this approach, we aim to facilitate the early identification of patients at risk of AKI and help healthcare professionals initiate timely interventions to improve patient outcomes.

Background

Acute Kidney Injury

Acute Kidney Injury (AKI), previously known as acute renal failure (ARF), is a syndrome characterized by sudden kidney failure or kidney damage that occurs within a few hours or a few days and rarely has a sole and distinct

pathophysiology. The need to evaluate the adequacy and efficacy of different therapeutic protocols, in addition to the possibilities of prevention and/or limitation of the damage, has led to formulate a classification of AKI that also includes slight alterations in the renal function. AKI is not a single organ failure clinical event, but a syndrome where the kidney plays an active role in the progress of multi-organ dysfunction, with different critical clinical conditions ranging from a slight increase in creatinine to anuria, namely the complete cessation of urine flow (3).

The early detection, the prompt treatment, and the anticipated interventions are elements that likely provide benefits for the patient which has the possibility to avoid temporary support from a dialysis machine or death itself. AKI is often a quickly evolving clinical event with high morbidity that represents an important complication in patients admitted to the hospital (10-15% of all hospitalizations). The mortality rate can be very high, between 50% and 80%, especially for patients in the Intensive Care Unit (ICU), where it sometimes exceeds 50% (4). The major challenge to AKI diagnosis and treatment is that specific syndromes often coexist, without the immediate onset of alarming symptoms such as chest pain, dyspnea, palsy, or blindness; hence, diagnosis requires specific technical assessments.

Functional dependencies

Our proposal is based on a specific type of functional dependency. Before the introduction to the method, we recall some basic information proper to delineate our technique. We briefly describe the definition of functional dependency (FD), and then introduce its extensions: Approximate Functional Dependency (AFD) and Temporal Functional Dependency (TFD). Such concepts will lead to the definition of Approximate Temporal Functional Dependency (ATFD), where ATFD inherits the properties both from AFD and from TFD.
To extract knowledge from clinical databases, there are different methodologies. Functional dependencies can be seen as hidden regularities that are present in the database. For example, suppose we are studying the different drug administrations, in order to discover recurrent adverse events. We may discover that given a symptom, we always have the same administered drug, or given an administrated drug we always have the same adverse reaction. We can formally define the concept of Functional Dependency (FD) (5) as follows:
Let r be a relation over the relational schema R(U): let $X, Y \subseteq R$ be sets of attributes of U. We assert that r fulfils the functional dependency $X \to Y$ (written as $r \vDash X \to Y$) if the following condition holds: $\forall t, t' \in r(t[X] = t'[X] \Rightarrow t[Y] = t'[Y])$. The X represents the antecedent, while Y is the consequent.
When we consider temporal aspects, so considering to extend the concept of FDs to deal with data temporalities, we talk about Temporal functional dependency (TFD) (6). Using the previous motivating example, a TFD may express a concept as: for each symptom and an adverse reaction, the administered drug does not change over a time span of 6 months.
In real-world data, noise is often present, and as a result, some errors are expected when discovering dependencies. Therefore, it is necessary to accept a certain degree of error in the discovered dependencies. So, from the plain concept of FD derives the concept of approximation (7), first introduced in (5). Given a relation r where an FD holds for most of the tuples in r, we may identify some tuples for which that FD does not hold.

Methods

MIMIC-IV

MIMIC-IV (8) is a publicly available database, the result of a collaboration between Beth Israel Deaconess Medical Center (BIDMC) and the Massachusetts Institute of Technology (MIT). They collected electronic health records regarding the decade from 2008 to 2019. Data contains seventy thousand admissions of patients hospitalized in ICU. The recorded information mainly contains demographic data, vital signs, laboratory results, procedures, and medications.

Different criteria have been used to gather accurate conclusions on the epidemiology of this syndrome. The first one was the International consensus criteria introduced by the Acute Dialysis Quality Initiative, and afterwards modified by the AKI Network (9) until 2012 when KDIGO (10) provided the new guidelines. In this study, we employ the KDIGO criteria based on AKI stages:

Stage 1: Serum creatinine is 1.5–1.9 times baseline, or an increasing equal or greater than 0.3 mg/dl (26.5 mmol/l), or a urine output less than 0.5 ml/kg/h for at least 6 hours;

Stage 2: Serum creatinine is 2.0–2.9 times baseline, or a urine output less than 0.5 ml/kg/h for at least 12 hours;

Stage 3: Serum creatinine is ≥ 3.0 times baseline, or an increase in serum creatinine equal or greater than 0.4 mg/dl (greater or equal to 353.6 mmol/l), or the initiation of renal replacement therapy, or a urine output less than 0.3 ml/kg/h for at least 24 hours, or anuria for 12 hours.

Data pre-processing: AKI definition and patient features

To effectively predict the risk of AKI, the first crucial step is to gather all relevant information about each patient, so that their medical history can be accurately represented. Using data from the MIMIC-IV database, we extract information for patients between the ages of 18 and 90, resulting in a cohort of approximately 70,000 individuals. To ensure that our predictions are as accurate as possible, we exclude patients who are admitted to the ICU with an already established AKI, analysing the first two occurrences of creatinine and treating these patients as unstable.
From a clinical perspective, from the entire database we identify several potentially relevant features for predicting AKI. Specifically, we focus on medications administered during the patient's ICU stay, including diuretics, antihypertensives, and nephrotoxic drugs. Additionally, we consider whether a patient was admitted to a specific post-surgery ICU (e.g., Surgical Intensive Care Unit (SICU), Trauma SICU (TSICU), Medical/Surgical Intensive Care Unit (MICU/SICU), …), whether the patient was diagnosed with sepsis and, if so, which type of sepsis (e.g., puerperal sepsis, salmonella sepsis, listeria sepsis, …) was present, whether the patient received a procedure involving contrast medium and the systolic and central venous pressures. To make the analysis more manageable, we discretize continuous variables into low, medium, and high categories based on clinical literature.
To evaluate each patient's risk of AKI, we use the KDIGO criteria and assign a label of 0 (if AKI is not present), 1, 2, or 3 at specific time points, from the time of the ICU admission until the discharge. By using different moving windows following the criteria specifications, we continuously monitor the patient history and track the evolution of the AKI stages. If a patient satisfies concurrent multiple criteria, serum creatinine and urinary output variations, for different AKI stages, we consider the worst one. This is because, by the end of the observation period, the characteristic is already underway, so we assume that the predicted event, such as an AKI diagnosis, started at the beginning of the observation period. In Figure 1, we illustrate how we evaluate the urine criteria. By considering the patient's entire history, we assess all stages of the criteria every 6 hours, using a different temporal window depending on the severity stage. The light blue section refers to the first criterion, which considers an observation period of 6 hours for urine; The green one refers to the second criterion, which considers an observation period of 12 hours for urine; The orange one refers to the third criterion, which considers an observation period of 24 hours for urine. Ultimately, we evaluate all pathological state patterns and choose the one with the worst severity.

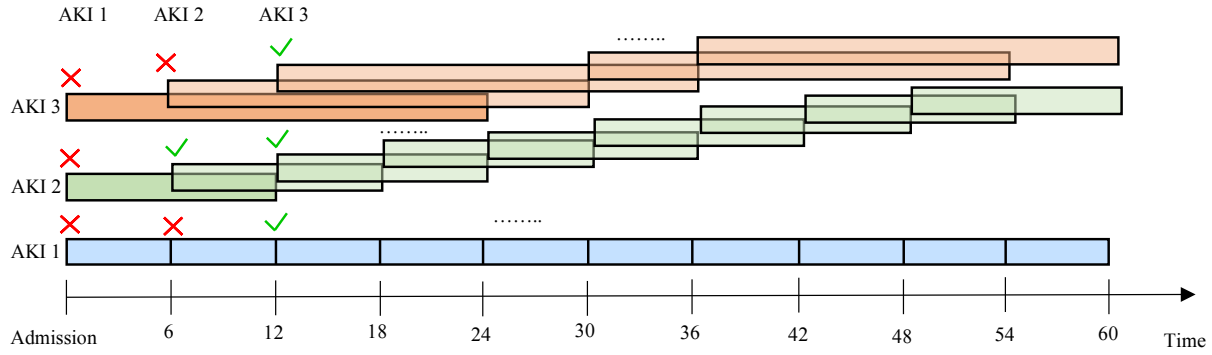


Figure 1. Temporal valuation of different AKI stage criteria: the worst satisfied stage is assigned every 6 hours. Red crosses refer to unsatisfied criteria, while green check marks refer to satisfied criteria.

The proposed model: Approximate predictive functional dependencies and the 3-window framework

In this section, we first introduce a 3-window model for the interpretation of predictive temporal data, and then we illustrate the definition of Approximate Predictive Functional dependencies (APFDs).
Typically, prediction models rely on two-time windows: the data collection or observation window, and the prediction window. We investigate the possibility to predict a future event using three temporal windows as follows: we collect information that could be relevant to a future event over a certain time span in the observation window (OW); we introduce a temporal gap before the event occurs, which we call the waiting window (WW); finally, we observe how the event evolves over time span in the prediction window (PW). It is worth noting that the length of the temporal

windows may vary according to the problem at hand. We may choose to collapse the waiting window to zero in case of observations with an instantaneous observable effect. Figure 2 provides a visual representation of this 3-window framework.

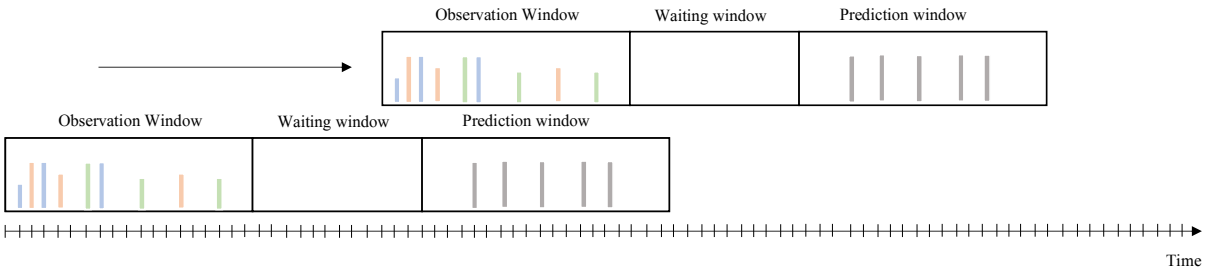


Figure 2. The 3-window framework. The temporal windows can be either fixed to a specific time point or movable along the time axis. The bar charts in the observation window represent various clinical parameters that have been observed, while the bar charts in the prediction window represent the observed clinical outcomes.

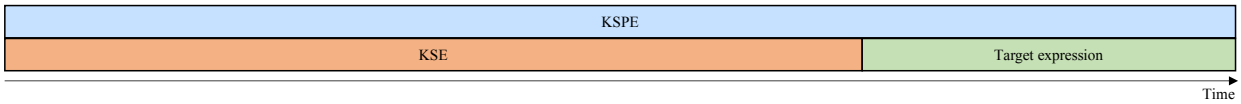
Within this 3-window framework, our goal is to define a general framework that captures a specific type of functional dependencies, with the antecedent composed of a set of characteristics called *predictive attributes*, and the consequent composed of the *predicted event*. Our proposal consists of three main elements.
Firstly, we propose that the relation containing the values of the predicted attribute, the target expression, is interval-based, i.e., its valid time represents the interval during which the considered value is true (11). A general target expression has schema $R^{TE}\left(ZB \cup \{VT_{start,}VT_{end}\}\right)$, where Z is the patient ID, B represent the predicted event (i.e. the diagnosis pattern), and VT_{start}, VT_{end} are the attributes representing the temporal dimension of the tuples, namely the start and the end of each diagnosis pattern. The second element, the K-State Evolution Expression (KSE), contains the predictive attributes. We define a relation K-State Evolution Expression as:

$$R^{KSE}(Z\overline{U}^{1}\overline{U}^{2}..\overline{U}^{k} \cup \{\overline{VT}^{1},\overline{VT}^{2},..,\overline{VT}^{k}\})$$

where attributes $\overline{U}^{i}$ represent properties holding at a valid time $\overline{VT}^{i}$ for $1 \leq i \leq k$; and for each tuple $t \in R^{KSE}$ it holds $t[\overline{VT}^{i}] < t[\overline{VT}^{(i+1)}]$ for $1 \leq i \leq k-1$. Therefore, the predictive attributes are temporally ordered.
Finally, we can define the third element, the K-State Prediction Expression (KSPE). We define a relation as K-State Prediction Expression as:

$$R^{KSPE} \equiv \{t | R^{KSE}(t[Z\overline{U}^{1}\overline{U}^{2}..\overline{U}^{k} \cup \{\overline{VT}^{1},\overline{VT}^{2},..,\overline{VT}^{k}\}]) \wedge R^{TE}\left(t\left[ZB \cup \{VT_{start,}VT_{end}\}\right]\right) \wedge t[\overline{VT}^{k}] - t[\overline{VT}^{1}] \leq OW \wedge t[VT_{start}] - t[\overline{VT}^{1}] > OW + WW \wedge t[VT_{end}] - t[\overline{VT}^{1}] < OW + WW + PW\}$$

In Figure 3, we report the general idea of a K-State Prediction Expression (KSPE), underlying which are the different involved parts, namely the K-State Evolution Expression (KSE) and the Target Expression.



Patient	$\overline{HR}^{1}$	$\overline{VT}^{1}$	$\overline{SpO_2}^{2}$	$\overline{VT}^{2}$	$\overline{Drug}^{3}$	$\overline{VT}^{3}$	AKI	VT_{start}	VT_{end}
Daisy	High	9	Low	11	Aspirin	13	333	17	20
Daisy	Low	2	High	4	Aspirin	6	322	9	12
Daisy	Low	2	High	4	Aspirin	6	221	13	16
Luke	Medium	7	High	8	Ibuprofen	12	222	14	17
Luke	Medium	7	High	8	Ibuprofen	12	221	18	21
Luke	Low	9	High	13	Sulindac	14	211	22	25
Stevie	High	1	Low	13	Aspirin	5	111	14	17

Figure 3. Graphical representation of an example of KSPE.

Once we have constructed these expressions, which allow us to associate ordered state values with a final target attribute value for a given patient, representing the event we wish to predict according to the specified temporal windows, we can then introduce the concept of Approximate Predictive Functional Dependency (APFDs).
In this framework, the consequent corresponds to the target attribute, while the antecedent is composed of a (sub)set of attributes that represent the various evolving states of a patient.

Given a KSPE $R^{KSPE}\left(Z\overline{U}^{1}\overline{U}^{2}..\overline{U}^{k}B \cup \left\{\overline{VT}^{1},\overline{VT}^{2},..,\overline{VT}^{k},VT_{start,}VT_{end}\right\}\right)$, we can extract the APFDs expressed as:

$$\bar{X}^{h}\bar{S}^{i}..\overline{W}^{j} \overset{\varepsilon}{\rightarrow} B \; with \; 0 \leq h < i..< j \leq k$$

$$where \; \varepsilon =< \varepsilon_g, \varepsilon_h, \varepsilon_j > and \; \bar{X}^{h} \subseteq \overline{U}^{h}, \bar{S}^{i} \subseteq \overline{U}^{i}, \overline{W}^{j} \subseteq \overline{U}^{j}$$

ε considers different kinds of approximation, i.e., different kinds of error thresholds: (i) ε_g represents a threshold for the number of tuples we have to delete for having the dependency satisfied; (ii) ε_h specifies the maximum number of patients we admit, having the complete deletion of their data, to satisfy the dependency; and (iii) ε_j control the maximum number of tuples we accept to delete per patient.
We compute all the APFDs, considering the three approximations, using an extended version of TANE algorithm (12). Our approach is mainly focused on the following steps: (i) Derive a set of tuples that satisfy the given APFD, according to error threshold ε_g (ii) From this set, we check whether the threshold ε_h is satisfied (iii) If ε_h is satisfied, we check ε_j.
As a result, our method allows us to investigate patterns in the past that involve specific temporally ordered stages, subsequently linking them to a specific pattern of pathological states occurring in the future.

It is worth observing that our approach deals with discovering Approximate Temporal Functional Dependencies from data, allowing some kind of errors. Thus, it can be considered complementary to Machine Learning (ML) techniques, which learn from data and try to predict pathological states from them. Indeed, while ML techniques propose predictive, often black-box, techniques, our proposal deals with the discovery of temporal features hidden in the data, with an explainable approach, which could be associated with the previously ML-mentioned techniques. Thus, we will see that our experimental results will be given in terms of error thresholds, while confusion matrices and the related rates and scores are not suitable for APFDs.

Results

From the initial court of 73729 patients admitted to the ICU, we select the "stable" patients considering the first two creatinine measures. Specifically, we exclude patients when the difference between the two values was greater than 0.3, according to the KDIGO guidelines. Following this preliminary phase, we end up with a cohort of 30,915 subjects. In Table 1, we provide a detailed report of all the specific features that were used during the pre-processing part.

For our analysis, we consider five different KSPEs using two different 3-window frameworks. Each framework consists of an observation window of 72 hours, during which we collect all the measures related to each patient, followed by a waiting window of either 6 or 12 hours where we do not consider any events. Finally, we have a prediction window of 24 hours. We build pathological state patterns of length 3.
The five KSPEs are composed as follows:

1. For the first KSPE, we use sepsis, nephrotoxicity, and diuretics to obtain 386 patients with this history in the database from the first framework (KSPE 1.a) and 371 from the second one (KSPE 1.b);
2. The second KSPE consists of central venous pressure, contrast medium, and another central venous pressure, which resulted in 4 patients within both frameworks (KSPE 2.a) and (KSPE 2.b);
3. The third KSPE uses sepsis, surgical operation, central venous pressure, diuretics, and nephrotoxicity to obtain 277 patients from the first framework (KSPE 3.a) and 254 from the second one (KSPE 3.b);
4. For the fourth KSPE, we use sepsis, surgical operation, and central venous pressure, resulting in 1705 patients from the first framework (KSPE 4.a) and 1669 from the second one (KSPE 4.b);
5. In the last one, we include systolic pressure, antihypertensives, and another systolic pressure obtaining 941 patients from the first framework (KSPE 5.a), and 548 from the second one (KSPE 5.b).

Table 1. Detailed features used during the pre-processing part.

Features	
Nephrotoxic drugs	Gentamicin, Vancomycin, Tobramycin, Amikacin, Penicillamine, Auranofin, Sulfamethoxazole, Trimethoprim, Sulfametrole, Sulfamazone, Streptomycin, Netilmicin, Zoledronate, Colistin, Acyclovir, Foscavir, Ganciclovir, Adefovir, Tenofovir, Indinavir, Cidofovir, Cyclosporine, Tacrolimus, Carmustine, Mutamycin, Prevacid, Pamidronate
Diuretic drugs	Furosemide, Triamterene, Hydrochlorothiazide, Indapamide Spironolactone, Tolvaptan, Chlorothiazide, Bumetanide, Amiloride, Metolazone, Eplerenone, Chlorthalidone, Torsemide, Aldactone, Ethacrynic acid, Acetazolamide
Antihypertensive drugs	Nebivolol, Moexipril, Sotalol, Lisinopril, Carvedilol, Methyldopa, Propranolol, Benazepril, Aliskiren, Ambrisentan, Clonidine, Pindolol, Bosentan, Minoxidil, Irbesartan, Prazosin, Quinapril, Doxazosin, Atenolol, Diazoxide, Metoprolol, Esmolol, Candesartan, Nadolol, Losartan, Captopril, Valsartan, Trandolapril, Acebutolol, Ramipril, Macitentan, Guanfacine
Sepsis (ICD code)	67020, 67022, 67024, 99592, A021, A227, A267, A327, A40, A408, A409, A41, A4150, A4159, A418, A4189, A427, A5486, B377, O85, R652, R6520, R6521
Contrast medium (ICD code)	8702, 8811, 8840, 8860, 8861, 8862, 8863, 8864, 8865, 8866, 8867

Table 2. Discretization of Systolic pressure and Central venous pressure.

	Low	Medium	High
Systolic pressure	< 90	90 -120	>120
Central venous pressure	<0	0-12	>12

In Tables 3 and 4 we report all the mined APFDs with the related epsilon values and the related 3-window framework. We select some APFDs to show which are the values under the dependencies. For example, from KSPE 3.a, we analyse $\overline{\text{SEPSIS}}^1$,$\overline{\text{DIURETICS}}^4$,$\overline{\text{NEPHROTOXIC}}^5$ →AKI, reporting in Table 5 the most common value combinations. Another remarkable example is given by the dependency $\overline{\text{SEPSIS}}^1$,$\overline{\text{OPERATION}}^2$,$\overline{\text{CVP}}^3$,$\overline{\text{DIURETICS}}^4$,$\overline{\text{NEPHROTOXIC}}^5$ →AKI from the KSPE 3.b. In Table 6, we report all the combinations which characterize only one pathological state, the '333' pattern.

Table 3. APFDs mined using the 3-window framework OW 72 - WW 6 - PW 24 (continued)

	3-window framework 72-6-24	ε_g	ε_h	ε_j
KSPE 1.a	$\overline{SEPSIS}^1, \overline{NEPHROTOXIC}^2, \overline{DIURETICS}^3 \longrightarrow AKI$	36.2%	0%	50%
	$\overline{SEPSIS}^1, \overline{DIURETICS}^3 \longrightarrow AKI$ $\overline{SEPSIS}^1, \overline{NEPHROTOXIC}^2 \longrightarrow AKI$	36.5%	0%	50%
	$\overline{SEPSIS}^1 \longrightarrow AKI$	36.6%	0%	50%
	$\overline{SEPSIS}^1 \longrightarrow AKI$ $\overline{NEPHROTOXIC}^2 \longrightarrow AKI$ $\overline{DIURETICS}^3 \longrightarrow AKI$	37%	0%	50%
KSPE 2.a	$\overline{CVP}^1 \longrightarrow AKI$ $\overline{CONTRAST}^2 \longrightarrow AKI$ $\overline{CVP}^3 \longrightarrow AKI$	18.5%	0%	0%
KSPE 3.a	$\overline{SEPSIS}^1, \overline{CVP}^3, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$ $\overline{SEPSIS}^1, \overline{DIURETICS}^4, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$ $\overline{SEPSIS}^1, \overline{OPERATION}^2, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$	32.8%	0%	50%
	$\overline{OPERATION}^2, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$ $\overline{CVP}^3, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$ $\overline{SEPSIS}^1, \overline{NEPHROTOXIC}^5 \longrightarrow AKI$	33%	0%	50%
KSPE 4.a	$\overline{SEPSIS}^1 \longrightarrow AKI$	33.24%	0%	50%
	$\overline{SEPSIS}^1 \longrightarrow AKI$ $\overline{OPERATION}^2 \longrightarrow AKI$ $\overline{CVP}^3 \longrightarrow AKI$	34%	0%	50%
	$\overline{SYSTOLIC}^1, \overline{ANTIHYPERTENSIVE}^2 \longrightarrow AKI$	47%	0%	50%

KSPE 5.a	$\overline{SYSTOLIC^1} \longrightarrow AKI$ $\overline{ANTIHYPERTENSIVE^2} \longrightarrow AKI$ $\overline{SYSTOLIC^3} \longrightarrow AKI$	50%	0%	50%

Table 4. APFDs mined using the 3-window framework OW 72 - WW 12 - PW 24

	3-window framework 72-12-24	ε_g	ε_h	ε_j
KSPE 1.b	$\overline{SEPSIS^1}, \overline{NEPHROTOXIC^2} \longrightarrow AKI$ $\overline{DIURETICS^3} \longrightarrow AKI$	36.5%	0%	50%
	$\overline{SEPSIS^1} \longrightarrow AKI$ $\overline{NEPHROTOXIC^2} \longrightarrow AKI$ $\overline{DIURETICS^3} \longrightarrow AKI$	37 %	0%	50%
KSPE 2.b	$\overline{CONTRAST^2} \longrightarrow AKI$	11%	0%	0%
	$\overline{CVP^1} \longrightarrow AKI$ $\overline{CONTRAST^2} \longrightarrow AKI$ $\overline{CVP^3} \longrightarrow AKI$	13%	0 %	0%
KSPE 3.b	$\overline{SEPSIS^1}, \overline{OPERATION^2}, \overline{CVP^3}, \overline{DIURETICS^4}, \overline{NEPHROTOXIC^5} \longrightarrow AK$I	33%	0%	50%
	$\overline{SEPSIS^1}, \overline{CVP^3}, \overline{NEPHROTOXIC^5} \longrightarrow AKI$ $\overline{SEPSIS^1}, \overline{OPERATION^2}, \overline{NEPHROTOXIC^5} \longrightarrow AKI$	33.2%	0%	50%
	$\overline{SEPSIS^1}, \overline{NEPHROTOXIC^5} \longrightarrow AKI$	33.5%	0%	50%
	$\overline{OPERATION^2}, \overline{NEPHROTOXIC^5} \longrightarrow AKI$ $\overline{CVP^3}, \overline{NEPHROTOXIC^5} \longrightarrow AKI$ $\overline{\text{SEPSIS}^1}, \overline{\text{NEPHROTOXIC}^5} \longrightarrow$AKI	34%	0%	50%
KSPE 4.b	$\overline{SEPSIS^1} \longrightarrow AKI$	33%	0%	50%
	$\overline{SEPSIS^1} \longrightarrow AKI$ $\overline{OPERATION^2} \longrightarrow AKI$ $\overline{CVP^3} \longrightarrow AKI$	33.24%	0%	50%
KSPE 5.b	$\overline{SYSTOLIC^1} \longrightarrow AKI$ $\overline{ANTIHYPERTENSIVE^2} \longrightarrow AKI$ $\overline{SYSTOLIC^3} \longrightarrow AKI$	50%	0%	50%

It is easy to see how the pattern '333' is the most popular one, and hence, the algorithm selects it. However, our proposal uncovers diverse diagnostic patterns within the data. In Table 7, we reported some value combinations under the APFD $\overline{SEPSIS^1}, \overline{DIURETICS^3} \longrightarrow AKI$, which highlight pathological states patterns leading to deterioration of health conditions.

Table 5. Value combinations of $\overline{\text{SEPSIS}^1}, \overline{\text{DIURETICS}^4}, \overline{\text{NEPHROTOXIC}^5} \longrightarrow$AKI

Sepsis ICD code	**Diuretic**	**Nephrotoxic**	**AKI pattern**
99592	Bumetanide	Vancomycin	300
99592	Chlorothiazide	vancomycin	333
A4189	Furosemide	Cidofovir	333
A4189	Furosemide	Vancomycin	033
R6521	Bumetanide	Vancomycin	333
R6521	Metolazone	Vancomycin	003

Table 6. Value combinations of $\overline{\text{SEPSIS}^1}, \overline{\text{OPERATION}^2}, \overline{\text{CVP}^3}, \overline{\text{DIURETICS}^4}, \overline{\text{NEPHROTOXIC}^5} \longrightarrow$AKI (continued)

Sepsis ICD code	**Operation**	**CVP**	**Diuretics**	**Nephrotoxic**	**AKI pattern**
99592	Yes	high	Chlorothiazide	Vancomycin	333
99592	Yes	high	Furosemide	Acyclovir	333
99592	Yes	high	Furosemide	Sulfamethoxazole-trimethoprim	333
R6521	Yes	high	Furosemide	Sulfamethoxazole-trimethoprim	333

R6521	Yes	normal	Furosemide	Acyclovir	333

Table 7. Value combinations of $\overline{SEPSIS}^1, \overline{DIURETICS}^3 \longrightarrow AKI$

Sepsis ICD code	Diuretics	AKI pattern
99592	furosemide	203
A408	furosemide	303
B377	furosemide	030
R6521	bumetanide	302
R6521	furosemide	031
R6521	metolazone	303

Discussion and Conclusion

In this study, we proposed the application of a new temporal data mining technique to support the discovery of temporal properties of data in clinical practice, to support the prediction of the evolution of pathological states of Acute Kidney Injury during hospitalizations in the Intensive Care Unit. We considered different temporal frameworks. Analyzing the results, it is easy to see there is no significant difference between the mined APFDs from the two different frameworks. Temporal windows were selected based on the clinical practice, and waiting windows of 6 and 12 hours, respectively, are not showing differences, concerning the APFDs obtained. Indeed, the dependencies were similar.

Based on the data we analyzed, we discovered that the most prevalent pattern of pathological states within MIMIC-IV is designated as 333, corresponding to three consequent states of AKI stage 3. As a result of this pattern prevalence, our technique highlights the predictive attribute value combinations associated with this pattern, even though it allows the discovery of the presence of many other patterns in the database.

A limitation of this technique is given by the user-dependent selection of the 3-window framework. The choice of length is not automatically set according to the data or to the problem at hand. but is always related to a choice made through clinical practice. To address this limitation, future research will focus on leveraging the waiting window nature to generate a model that is always tailored to the specific clinical problem to prevent future events.

A significant strength of our approach is the absence of learning. We do not use any learning methods, to predict/classify patient data. Indeed, we allow the discovery of temporal data dependencies hidden inside the data. This feature enables us to overcome the "black box problem," which often generates a layer between the user and the selected model.

References

1. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders HJ. Acute kidney injury. Nat Rev Dis Prim [Internet]. Springer US; 2021;7(1). doi: http://dx.doi.org/10.1038/s41572-021-00284-z

2. Amico B, Combi C. A 3-Window Framework for the Discovery and Interpretation of Predictive Temporal Functional Dependencies. Vol. 13263 LNAI, Lecture Notes in Computer Science. Springer International Publishing; 2022. 299-309 p. doi: http://dx.doi.org/10.1007/978-3-031-09342-5_29

3. Makris K, Spanou L. Acute Kidney Injury: Definition, Pathophysiology and Clinical Phenotypes. Clin Biochem Rev. 2016.

4. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. Lancet [Internet]. Elsevier Ltd; 2019;394(10212):1949–64. doi: http://dx.doi.org/10.1016/S0140-6736(19)32563-2

5. Codd EF. Normalized data base structure: A brief tutorial. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. 1971. doi: 10.1145/1734714.1734716

6. Wijsen J. Temporal Dependencies. In: Encyclopedia of Database Systems. 2018. doi:10.1007/978-1-4614-8265-9_396

7. Kivinen J, Mannila H. Approximate inference of functional dependencies from relations. Theor Comput Sci. 1995. doi:10.1016/0304-3975(95)00028-U

8. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. Nature Research; 2023 Dec 1;10(1). doi: 10.1038/s41597-022-01899-x

9. Mehta RL, Kellum JA, Shah S V., Molitoris BA, Ronco C, Warnock DG, et al. Acute Kidney Injury Network: Report of an initiative to improve outcomes in acute kidney injury. Crit Care. 2007. doi:10.1186/cc5713. PMID: 17331245; PMCID: PMC2206446

10. KDIGO Group. "KDIGO clinical practice guideline for acute kidney injury." *Kidney Int. Suppl.* 2 (2012): 1.

11. Jensen CS, Snodgrass RT. Valid Time. In: Encyclopedia of Database Systems. 2017. doi:10.1007/978-1-4614-8265-9_1066

12. Huhtala Y, Kärkkäinen J, Porkka P, Toivonen H. TANE: An efficient algorithm for discovering functional and approximate dependencies. Comput J. 1999. doi: 10.1093/comjnl/42.2.100