Interpretable Sub-phenotype Identification in Acute Kidney Injury

Ho Yin Chan, Ph.D.[1], Mei Liu, Ph.D.[1]
[1]Division of Medical Informatics, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS, USA

Abstract
Acute kidney injury (AKI) is a life-threatening and heterogeneous syndrome. Timely and etiology-based personalized treatment is crucial. AKI sub-phenotyping can lead to better understanding of the pathophysiology of AKI and help developing more targeted intervention. Current dimensionality reduction and similarity-based clustering for AKI sub-phenotyping suffer from limited interpretability and specificity. To address these limitations, we propose a pattern mining approach with multiobjective evolutionary algorithm (MOEA) for AKI sub-phenotyping. AKI sub-phenotypes are presented as explicit rules, so no post-hoc explanation is needed. Also, our method can search feature subspace efficiently for minor and highly specific sub-phenotypes. We aimed to discover sub-phenotypes for AKI patients against non-AKI patients (AKI vs non-AKI) and moderate-to-severe AKI patients against mild AKI patients (AKI-2/3 vs AKI-1). We identified 174(178) significant sub-phenotypes with average confidence of 0.33(0.33). Our method can assign patients to a sub-phenotype with higher confidence than k-means clustering, with average improvement of 0.20(0.23).

Introduction
AKI is a prevalent and life-threatening clinical syndrome in hospitalized patients, affecting ~15% of all hospitalizations and >50% of patients in intensive care units[1-3]. Currently, AKI patient management is based on clinical manifestation assessed by serum creatinine, which is a delayed biomarker of AKI and disregards the underlying pathophysiology[4]. In contrast, appropriate and immediate interventions should be etiology-based. For example, cardiac failure associated AKI would encourage cardiac failure management while hypovolaemic-AKI would encourage volume replacement.
Recent availability of electronic health record (EHR) and advancement in machine learning have fostered a variety of data-driven approaches to study AKI. Machine learning methods have been applied mostly for AKI risk prediction[5, 6], with the area under the receiver operating characteristics curve (AUROC) ranging from 0.66-0.80 in internal validation studies and 0.65-0.71 in external validation studies[7-9]. However, the traditional prediction models tend to bias toward high-ranking features in general population and produce unstable prediction for minority subgroups. A recent study[10] showed that localized prediction models trained on stratified subgroups can improve performance and discrepancies in feature ranking when compared to the global model, which highlights the importance of personalized prediction based on local/minority subgroup.
Sub-phenotyping is challenging due to the high dimensionality of EHR data. Several studies aimed to subgroup patients via data driven approach. Xu *et. al.* performed k-means clustering with t-distributed Stochastic Neighbor Embedding[11] (t-SNE). Chaudhary *et. al.* utilized autoencoder for dimensionality reduction and performed k-means clustering on the extracted features[12]. Baytas *et. al.* used a similar method but applied LSTM network to exploit the temporalities of EHR data[13]. While these approaches are promising, sub-phenotyping based on k-means clustering has several disadvantages. First, clustering depends on feature selection and minority clusters are only distinguishable in certain feature subspace. Second, similarity-based clusters often require post-hoc explanations, especially after feature space projection with autoencoder or t-SNE. Third, number of clusters must be pre-defined, and outliers are assigned to the nearest big cluster. Typically, only a small k (<10) is tested. As a result, many features are used and only a few major clusters are identified. Rare subphenotypes of order <10% are often hidden or grouped into neighboring big clusters, thus it is difficult to associate a cluster to one specific clinical origin. Fourth, the result of similarity-based clustering is sensitive to the similarity metric used, which are often chosen without justification.
To address the above limitation, we propose a rule mining approach for sub-phenotyping AKI patients. With rule mining, sub-phenotypes are determined by rules instead of patient similarities. Rule mining can effectively explore a large amount of feature subspace and identify minor subgroups that are highly specific. In addition, rules that define a cluster are precise and inherently interpretable, which is useful for subsequent clinical analysis. Rule Mining have been previously applied to discover the temporal relations in the EHR data for prediction analysis, I. Batal *et. al.*[14] and D. Patel *et. al.*[15] explored the temporal interval relations based on Allen's interval algebra[16], whereas Z. Huang *et. al.*[17] investigated patterns based on temporal separation of events. To applying rule mining for sub-phenotype discovery, we focus on mining non-temporal rule for easier interpretation and more tolerance to noise in EHR data.

Method

Rule Mining

Rule mining aims to describe a dataset in forms of rules. Given a binary dataset D, with features set I and binary labels $L = \{l_0, l_1\}$, a pattern p can be defined as a collection of features $p = \{i_1, i_2, i_3, \dots\} \subseteq I$ and a rule r is defined as a pattern associating with a label $l \in L$, denoted as $p \rightarrow l$. A rule can be interpreted as an if-then association. If a data point satisfies the antecedent (pattern p), then the data is predicted by the rule as having a label l. Rule mining is a procedure to discover rules that satisfy certain predefined measures. In the rest of the paper, a pattern p will be used as a representation of the antecedent of a rule r.

The difficulty of rule mining lies in the power law dependencies of the number of possible patterns. Traditional algorithms like Apriori[18] has been implemented for medical sub-phenotyping but it is highly computationally intensive due to the high dimensionality of EHR. The length of pattern can grow and produce a very large sets of rules that are difficult to analyze. Other algorithms like pattern tree[19] are less computationally intensive but still lack scalability for high dimensional data. Since EHR often consists of hundreds to thousands of features, it would be impractical to adopt classical rule mining algorithms to mine full EHR dataset without reducing the number of features to the order of tenth, or resort to distributed computing[18] which is hard to implement.

To address the above limitation, we employed multiobjective evolutionary algorithm[20] (MOEA) in this paper. MOEAs have been successfully applied to biomedical fields such as cervix lesion classification[21], medical speciality classification[22], gene selection[23] and biclustering[24]. MOEAs aim to discover a set of solutions, called Pareto set, through optimization of multiple criterions. When applied to rule mining, the solutions are the rules, and the criterions will be defined as the predefined measures[25]. Given the randomness nature of MOEA, the rule set is not complete and the quality is subject to the amount of computational time spent, but the rule set is often smaller in size, and the rules are more distinct. Also, MOEA requires less memory and more scalable to high dimensional data. The choice of criterions is independent of the algorithms and thus can be customized for finding more relevant rules.

One of the most frequently used measures in rule mining is support[25], which is defined as

$$sup_D(p) = \frac{count_D(p)}{|D|}$$

where $count_D(p)$ is the number of instances in $d \in D$ that p matches, i.e., $p \subseteq d$. But given the imbalanced nature of EMR, support can be an ineffective measure that would filter out useful rules in targeted labels, which is often a minority group. Instead, we use the following measures as the criterions for the rules:

True positive rate[25] is defined as the support of rule in the targeted label group D_l, i.e.

$$TPR(r) = sup_{D_l}(p)$$

We used true positive rate instead of support because we wanted to focus on discovering significant sub-phenotypes of the targeted label. True positive rate allows the discovery of rules that are populated in the targeted group instead of the general populations, which is more clinically useful in differentiating different groups of patients. In this paper, true positive rate is both set as a criterion and a constraint with a minimum value of 0.01.

Length of pattern is the cardinality $|p|$ of the pattern. Given the sparsity of EHR, the true positive rate drops rapidly with the length of the pattern. As such, the resulting rule set often consist of rules with a single feature pattern, which defeats the purpose of rule mining. To counter this trend, the length of pattern is set as a criterion to maximize so that longer patterns will be preferred in the Pareto set. In addition, the minimum length of patterns is constrained to two.

Growth rate[25] is defined as the ratio between the support of the targeted class against the non-targeted class, i.e.

$$growthrate(r) = \frac{sup_{D_l}(p)}{sup_{D_{\sim l}}(p)}$$

where $\sim l$ indicates the other label. Growth rate represents the statistical difference of a pattern between the associated group and the other group, which is set as a criterion to discover rules that are more representative in the targeted class.

Confidence[25] is defined as the conditional probability of target label l given an instance matching pattern p, i.e.

$$conf(r) = \frac{sup_D(p \cup \{l\})}{sup_D(p)}$$

Confidence is set as a criterion for identifying rules that better predict outcome label for patients.

Confidence difference is defined in this paper as

$$confdiff(r) = conf(r) - \max_{i \in p}(conf(\{i\}))$$

The idea of confidence difference is to focus on rules that have additional predictive power arising from the combination of features used in the pattern. Without this criterion, the discovered rules are often dominated by a certain feature, which made it non-distinctive to one feature patterns. Note that a stricter criterion $\min_{q \subset p}(growthrate(q)) > growthrate(p)$ is often used in traditional rule mining[26], which would require a complete knowledge of sub-patterns to compute. This criterion is computationally ineffective to be applied in MOEA, thus we used the confidence difference instead.

Study Design

AKI definition. Following the clinical practice guideline for AKI by KDIGO, we staged AKI severity based on the SCr-based criteria[27]. AKI Stage 1 is defined as an increase of 0.3 md/dL within 48h or 1.5 times of baseline SCr within 7 days. The baseline SCr level is chosen to be the most recent SCr after admission if past measurement is not available. Stage 2 AKI is defined as an increase of 2.0 to 2.9 times of the baseline SCr level in 7 days. Stage 3 AKI is defined as an increase of 4mg/dL after an acute increase of at least 0.3 mg/dL in 48h, an increase of more than 3.0 times of the baseline SCr level in 7 days or an initiation of renal replacement therapy.

Data Source. The clinical data for this study is collected from the University of Kansas Medical Center's data warehouse that has mapped EHR data to the common data model (CDM) developed by The National Patient-Centered Clinical Research Network[28] (PCORnet). The data is de-identified according to the Health Insurance Portability and Accountability Act of 1996 (HIPAA) 'Safe Harbour' criteria. This study was determined by the institutional review board as non-human subject research because it only involved the collection of existing and de-identified patient medical data.

Data Extraction. Patients aged between 18 and 90, admitted from the beginning of 2010 to the end of 2019, hospitalized for at least 2 days (n=286,723) and with at least two SCr records are extracted (n=247,457). Patients who had evidence of severe kidney dysfunction at or before admission are excluded using the criteria[29] (a) estimated glomerular filtration rate <15 mL/min/1.73 m^2 (n=11,778), or (b) has undergone any dialysis procedure or renal transplantation (RRT) prior to the visit (n=16,419), (c) required RRT within 48h of their SCr measurement record (n=1,702), or (d) has pre-existing end stage renal disease (n=15,800). Burn patients are also excluded since SCr is less reliable in accessing renal function during hypermetabolic phase[30] (n=653). The resulting cohort contained 218,365 records, with 183,235 non-AKI patients, 32,407 AKI stage 1 patients, 2,402 AKI stage 2 patients and 1,321 AKI stage 3 patients.

Data preprocessing. We collected clinical variables available in the PCORnet CDM schema, including demographic details (age, gender, race, Hispanic), diagnosis code (ICD-9 or ICD-10), procedure code (ICD and CPT code), lab test (LOINC code), medication (RXNORM and NDC code) and vital signs (weight, systolic, diastolic, BMI). Each record is timestamped relative to admission date. Patient data consist of most recent measures of vital signs and lab test values relative to onset day, historical record of diagnosis within a year and medication record between 30 days before admission and onset day. Multiple measurements of vitals and labs in the same day are averaged. Diagnosis codes are separated as before and after 6 months relative to the admission date, with ICD-10 code converted to ICD-09 code and all ICD-09 code are rolled up to the 3-digit category level. Medications are converted to the Anatomical Therapeutic Chemical (ATC) code and rolled up to 4th level. Then the dataset was converted to one-hot encoding with numerical features binned using quantiles (with a maximum of 5 bins). Features with less than 5% positive occurrence (not NaN for numerical features or True for binary/categorical features) are dropped. The label of patients is assigned to the most severe AKI stage in record. The final dataset consists of 1,020 features.

Learning Algorithm. The MOEA employed in this study is Nondominated Sorting Genetic Algorithm II[31] (NSGA2), implemented by the pymoo[32] package in Python. We chose NSGA2 because of its ability to handle binary data type. All one feature patterns are used as biased initialization of the algorithm. Hyper parameters include population size = 10000, number of offspring = 10, crossover = half uniform, mutation = bitflip, number of generation = 10000.

Experimental Design. Two sets of experiments were performed. First evaluation used AKI patients of all stages as the targeted label against the non-AKI patients as the other group (AKI vs non-AKI). The experiment aims to identify patient subgroups that have higher risk in developing AKI. Second evaluation treated AKI-2 and AKI-3 patients as targeted label while AKI-1 patients are treated as another group (AKI-2/3 vs AKI-1). This experiment would allow discovery of sub-phenotypes of AKI patients that are more likely to progress to more severe form of AKI.

Evaluation. The p-value of the rules discovered by the algorithm were calculated using fisher's exact test and rules with $p \geq 0.05$ were discarded to ensure statistical significance. The quality of a rule was characterized by measures (length of rule, growth rate, confidence, confidence difference, true positive rate) used in the algorithm and their 7-day mortality rate.

Result and Discussion

Validation of current method. To validate our method against clinical knowledge, we examined the one feature rules of AKI vs non-AKI, where clinical result is more readily available. Table 1 shows the one feature rules with top-3 confidence. Multiple studies have confirmed the association of both high base deficit and high B-type Natriuretic peptide with risk of AKI. Relation between Base Deficit as a predictive factor of AKI has been suggested by Gerent et. al.[33], while that of Natriuretic peptide B has been studied by Nowak et. al.[34] Our data and method confirms with the existing studies at base level.

Table 1. Statistics of single feature rules in AKI vs non-AKI.

	Rule	Support	TPR	Confidence
1	Base deficit in Arterial blood (>7 mmol/L)	0.011	0.029	0.457
2	Natriuretic peptide B [Mass/volume] in Blood (>770 pg/mL)	0.042	0.103	0.414
3	Base deficit in Blood (>5.1 mmol/L)	0.023	0.056	0.413

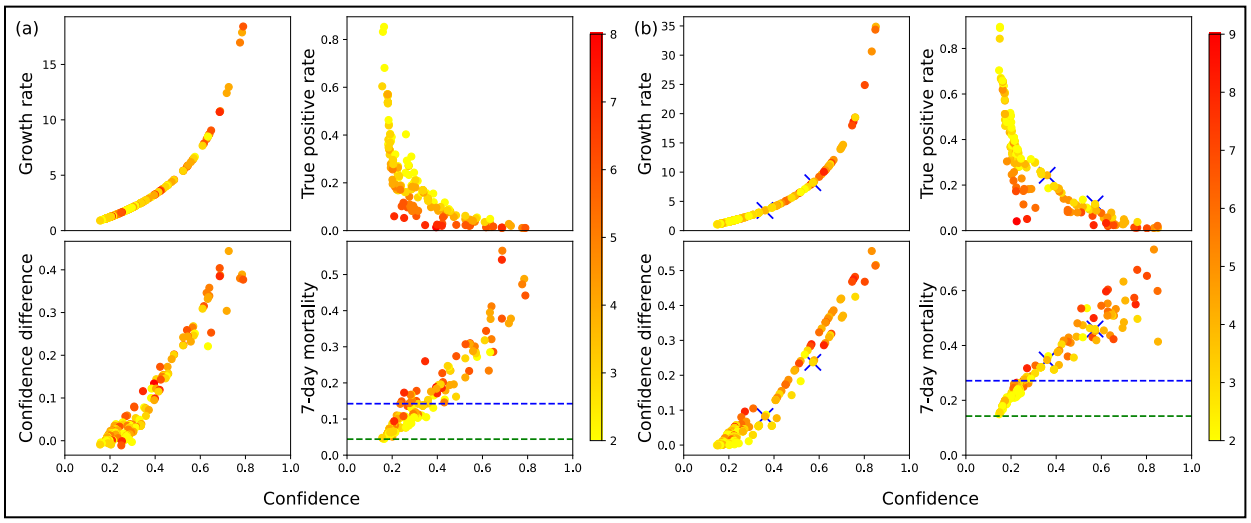

Figure 1. General trend of rule sets for (a) AKI vs non-AKI and (b) AKI-2/3 vs AKI-1. Each point represents a rule and the colormap represents the length of the rule. The blue crosses are the rules used as examples in the following section. The green dashed line represents the global 7-day mortality and the blue dashed line indicates that of the targeted label.

Rule statistics and trends. The advantage of rule mining lies in the ability of discovering complex rules. In the AKI vs non-AKI and AKI-2/3 vs AKI-1 experiments, total 174 and178 significant rules were discovered with average length of 3.51 and 3.65 and average confidence of 0.33 and 0.33, respectively. Compared to the class ratio of 0.17 and 0.14, the rule set shows significant distinction over general population. Figure 1 shows the trends of the measures of the discovered rules. The confidence difference, growth rate, 7-day mortality and the rule length are directly proportional to the confidence of the rule while the true positive rate are inversely proportional. The trend indicates that longer rules have higher predictability but have limited generalizability, thus more personalized. The higher quality rules are those with high true positive rate and high confidence. In the following section, two rules from the AKI-2/3 vs AKI-1 experiment are used as an illustration, the positions of those rules are indicated as blue crosses in Figure 1(b). Details of those rules are shown in Table 2.

Clinical interpretation of rules. While the individual features of the rule shown in Table 2 are associated with kidney problem, there are also other possible causes making their confidence low, especially for differentiating AKI-2/3 against AKI-1. For example, protein in urine can be caused by high blood pressure and high phosphate can be due to dietary habit. But in combination, bilirubin in urine is usually associated with liver problem where unitary tract infection, as suggested by nitrite presence in urine, is one of its complications[35]. Rule #169 is likely associated with patients having advanced liver cirrhosis, which is a cause of AKI[36, 37]. Protein in urine and high anion gap are both

associated with diabetes, and high phosphate can be a symptom of diabetic ketoacidosis. So, rule #87 is likely to be associated with diabetic patients, which is shown to be an independent risk factor of AKI[38, 39]. Both rules suggest that patients who develop AKI as a complication from diabetes and liver cirrhosis are more susceptible to advanced stage of AKI. Rule #119 is another example of significant rules that has high improvement in confident.

Table 2. Sample rules in AKI-2/3 vs AKI-1 experiment.

	Rule	**Confidence (single)**	**TPR**	**Confidence**	**Support**
169	Bilirubin.total [Presence] in Urine by Automated test strip	0.162	0.243	0.360	0.095
	Nitrite [Presence] in Urine by Automated test strip	0.162			
	Anion gap in Serum or Plasma (>10.0)	0.278			
	Salt solutions (Med)	0.151			
87	Protein [Presence] in Urine by Automated test strip	0.162	0.117	0.572	0.029
	Anion gap in Serum or Plasma (>10.0)	0.278			
	Phosphate [Mass/volume] in Serum or Plasma (>4.3mg/dL)	0.335			
119	Urobilinogen [Presence] in Urine by Automated test strip	0.162	0.123	0.478	0.143
	Phosphate [Mass/volume] in Serum or Plasma (>4.3mg/dL)	0.335			
	Infusion, normal saline solution, 250 cc	0.175			

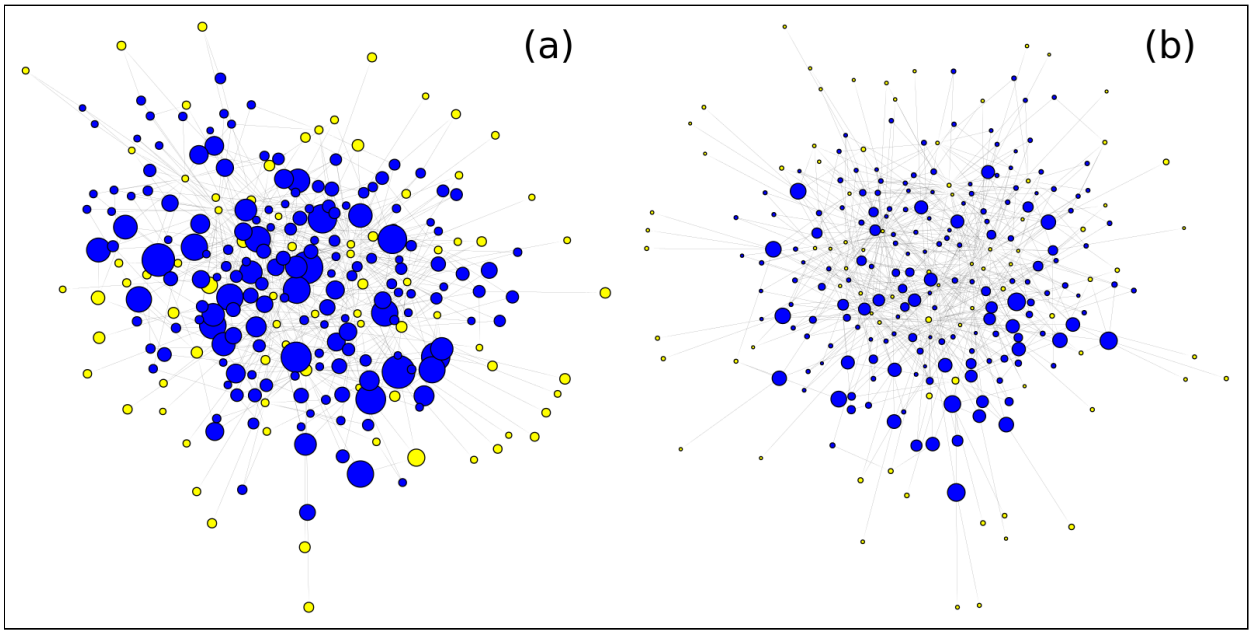

Figure 2. Bubble plots of the rule set of (a) AKI vs non-AKI and (b) AKI-2/3 vs AKI-1. Each blue circle represents a rule and each yellow circle represents a one feature rule. The edge indicates a feature belongs to a rule and the size scales with the confidence of the rule or feature.

Rule improvement. Table 2 showed that in general, combination of multiple features would increase the predictability over rules with a single feature. To quantify the improvement of the predictability, we compared confidence of the combinatorial rule with the maximum confidence of the single featured rules. In this case, rule #169 had a confidence difference of 0.082 and rule #87 had a confidence difference of 0.237. The improvement weakly depends on the length of rule, as shown in Figure 1. Figure 1 shows the confidence relation between the rules and their features. This improvement can be difficult to explore in clinical trials given the number of possible combinations to be tested and our algorithm can suggest potential combinations to be studied clinically. Also, these rules can give a more precise and personalize diagnose for the patients and suggests additional lab testing if a partial match is found. When compared to traditional predictive modeling, the criteria of rules are clearer and more understandable. Figure 2 illustrates the connection between the discovered rules and their relative confidence.

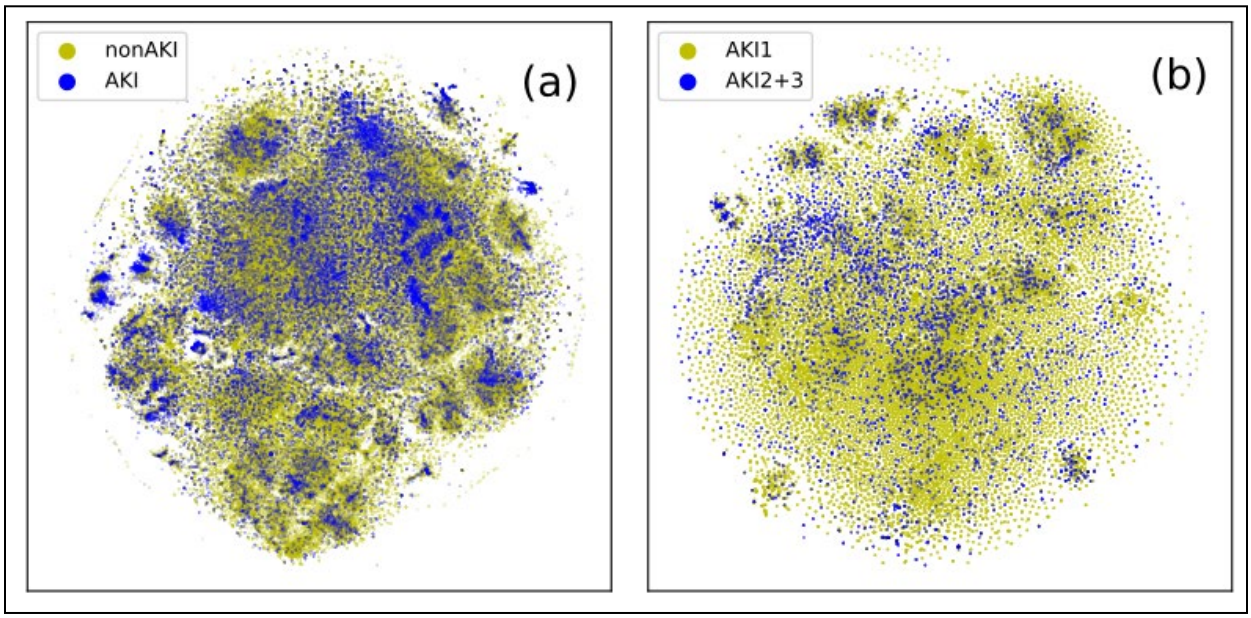

Figure 3. t-SNE plot of (a) AKI vs non-AKI and (b) AKI-2/3 vs AKI-1. Yellow dots represent the targeted class and blue dots represent the other class.

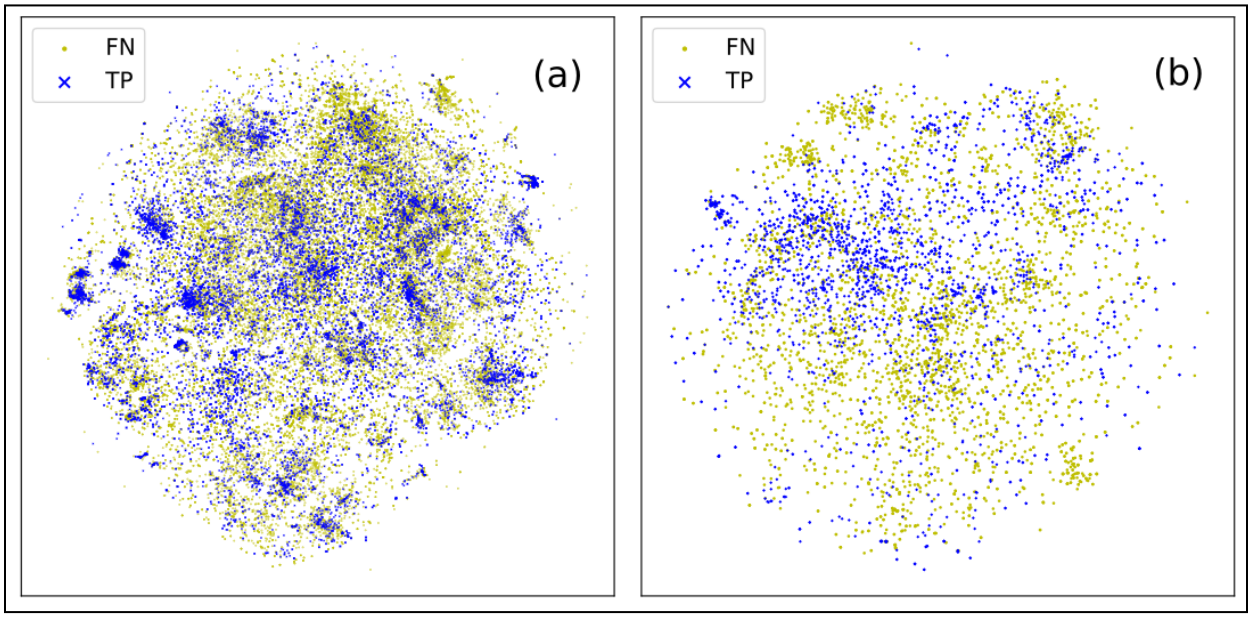

Figure 4. t-SNE plot of a rule coverage in (a) AKI vs non-AKI and (b) AKI-2/3 vs AKI-1. Yellow dot represents the false negative and blue cross represents the true positive.

Coverage of rules. To compare the coverage of the rule set and the similarity, we showed data distribution using t-SNE[40]. Figure 3 shows the global distribution for both datasets. From Figure 3, there is no clear cluster of the targeted labels, which suggests that clustering technique will not perform well on this data. In contrast, our method is not restricted by similarity. Two examples are shown in Figure 4 where the true positive rules are not confined to a specific region of the plot. To quantify this behavior, we calculate the root-mean-square deviation from mean of the t-SNE coordinates of the true positives for each rule. The average deviation is 5.91±0.38 for AKI vs non-AKI and 14.37±0.74 for AKI-2/3 vs AKI-1. The above deviation for the targeted class is 6.12 for AKI vs non-AKI and 13.53 for AKI-2/3 vs AKI-1, which suggests that most rules have approximately the same level of spread as the underlying class and MOEA does not depend on similarity. In addition, we compared MOEA, single length rules, and k-means clustering[41] based on t-SNE ($k = 2$ to 10). Figure 5 demonstrates the resulting performance in which k-means clustering performed worse in general. While some single feature rules have the same performance as MOEA at low confidence, more specific rules can only be obtained by MOEA.

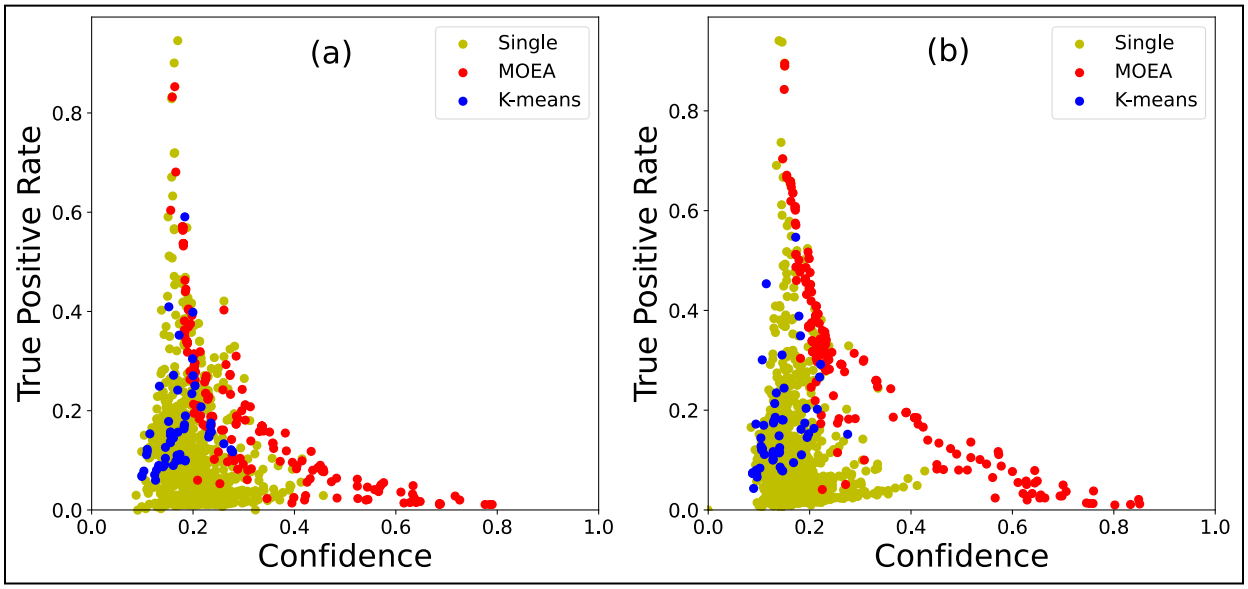

Figure 5. True positive rate vs Confidence of (a) AKI vs non-AKI and (b) AKI-2/3 vs AKI-1 for three models.

Rule Mining as a mean of subphenotype discovery and feature selection. We have shown previously the global t-SNE plot does not show obvious subgroup of patients that is statistically useful, due to the curse of dimensionality. Traditionally, to gather useful information, clinically important features have often been selected by expert. While expert selection produces promising result, complex feature correlations is often hard to be identified and established. Rule mining can provide a mean for selecting useful features and identifying useful subphenotypes from it. For illustration, we used rule #169 from Table 2 for feature selection. The resulting t-SNE plot is shown in Figure 6 (a). Under this projection the data is divided into one big cluster and several small clusters, and one of the small clusters in blue is the sub-phenotype represented by rule #169 for its distinctive statistical property.

Scalability Evaluation. To demonstrate the scalability of our method, we generated subsamples of our dataset into 1/1000, 1/100, 1/10, 3/4, 1/2, 1/4, 1 of the original size and performed the algorithm under the same hyperparameter. The result is shown in Figure 6 (b). The scaling is weakly linear to the size of the dataset. Given that the current working dataset already consists of all AKI patients within 10 years of KUMC record, our model should be practical in real world scenario.

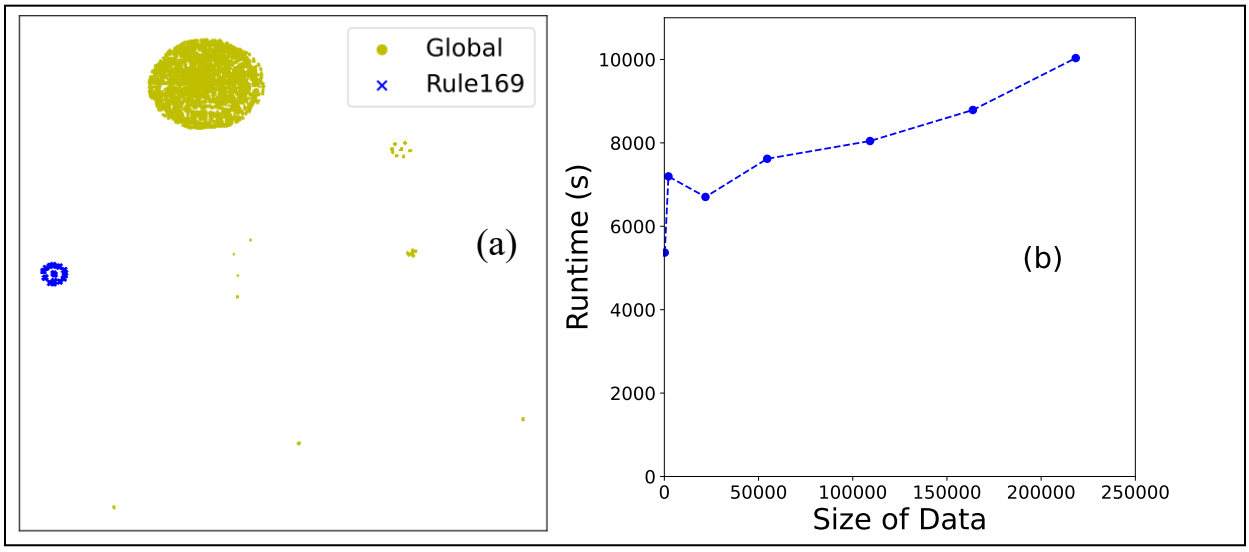

Figure 6. (a) t-SNE plot after using feature selection by rule #169. Blue crosses represent data point matching rule #169 and yellow dots otherwise. (b) The scalability of current method.

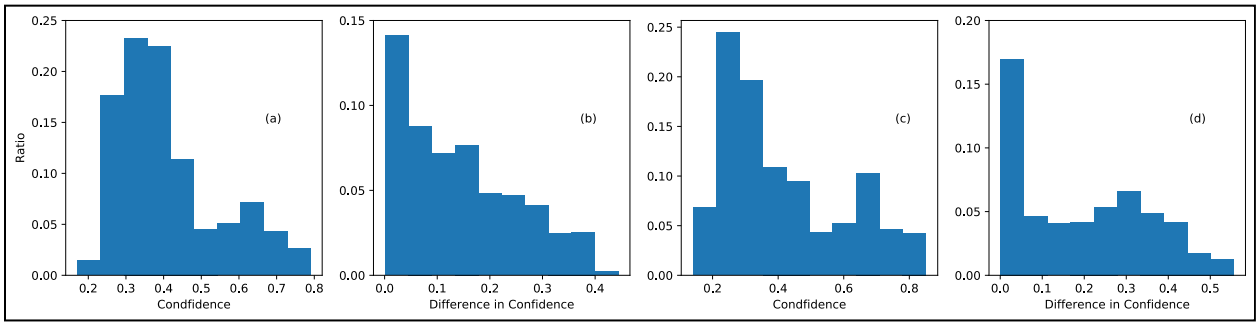

Figure 7. The histogram for AKI vs non-AKI of (a) confidence of data point using the highest matching rules of the rule set combined with the single length rules and (b) the difference in confidence when compared to single length rules only (excluding data with no difference). (c) and (d) are the corresponding histogram for AKI-2/3 vs AKI-1.

Global performance of rule set. To evaluate the predictive power of the rule set as a whole, we tested a simple scheme in which we assigned each patient in the targeted class the highest confidence of the matching rule in the rule set combined with all single feature rules. To evaluate the additional predictive power resulted from the rule set, we repeated the same process but with the single length rules only. Then the difference of confidence for each data is calculated. The result is shown in Figure 7. Every patient in the targeted label is covered by at least one rule and only 0.01% of the patients has confidence less than the class ratio for AKI-2/3 vs AKI-1 (0% for AKI vs non-AKI), 65% of patients in AKI vs non-AKI has confidence more than double of the class ratio (70% for AKI-2/3 vs AKI-1). On the other hand, the benefit of complex rules is also significant. 56% of patients have increased confidence for AKI vs non-AKI and 54% for AKI-2/3 vs AKI-1, while 32% of patients has an increase > 0.1 for AKI vs non-AKI and 33% for AKI-2/3 vs AKI-1. Figure 7(b) and (d) show the distribution of patients with increased confidence only. Average improvement was 0.14 for AKI vs non-AKI and 0.19 for AKI-2/3 vs AKI-1. When compared to k-means clustering, 99% of the patients for AKI vs non-AKI (99% for AKI-2/3 vs AKI-1) has increased confidence assigned to a sub-phenotype using the rule set and one length rules vs assigned to a sub-phenotype defined by k-means clustering, as indicated in Figure 5, and 75% for AKI vs non-AKI (71% for AKI-2/3 vs AKI-1) had improvement > 0.1, with an average improvement of 0.20 for AKI vs non-AKI (0.23 for AKI-2/3 vs AKI-1).

Clinical Application. Clinical interpretability of machine-learning models is utmost important in clinical practice. In contrast to accuracy-driven black-box models, where shortcomings are difficult to detect and prevent[42], all the rules discovered in our method can be independently verified and interpretable by clinicians. In practice, a system can quickly identified the rules that a patient match. In addition to provide the risk of a patients in developing AKI, the system can also provide which factors are contributing to the prediction. Clinician can determine whether a certain

rule match is applicable based on their own experience and circumstances. Clinician can also provide more personalized treatment based on the group of rules that the patient matches. The system can potentially identify cases where occurrence is rare but with high confidence. A partial match of a rule could suggest a potential clinical test to perform for narrowing down the possibility. Results from different hospital sites can also be shared and aggregated without compromising patient privacy. Furthermore, clinical trials can be expensive and time consuming, as such limited to simple and common correlations. Our method can serve as an exploratory research where complex and unanticipated relations can be discovered and verified by more rigorous clinical trials.

Conclusion

In this paper we proposed to use rule mining algorithm to identify sub-phenotypes of AKI patients. We have successfully identified subphenotypes of major and minor subgroups in a human readable format. We have generated both major and minor sub-phenotypes (in terms of true positive rate) with increased confidence. We also demonstrated the interpretability of our method by associating two sample sub-phenotypes to specific clinical etiology. Our approach showed improvement in confidence when compared to k-means clustering, and some specific sub-phenotypes also showed increased confidence when compared to baseline sub-phenotypes.

References

1 Cheng P, Waitman LR, Hu Y, Liu M. Predicting inpatient acute kidney injury over different time horizons: How early and accurate? AMIA Annu Symp Proc. 2018;2017:565-74.

2. Hoste EAJ, Bagshaw SM, Bellomo R, et al. Epidemiology of acute kidney injury in critically ill patients: The multinational aki-epi study. Intensive Care Medicine. 2015;41(8):1411-23.

3. Zeng X, McMahon GM, Brunelli SM, Bates DW, Waikar SS. Incidence, outcomes, and comparisons across definitions of aki in hospitalized individuals. Clinical Journal of the American Society of Nephrology. 2014;9(1):12.

4. Endre ZH, Mehta RL. Identification of acute kidney injury subphenotypes. Curr Opin Crit Care. 2020;26(6):519-24.

5. Le S, Allen A, Calvert J, et al. Convolutional neural network model for intensive care unit acute kidney injury prediction. Kidney International Reports. 2021;6(5):1289-98.

6. Roy S, Mincu D, Loreaux E, et al. Multitask prediction of organ dysfunction in the intensive care unit using sequential subnetwork routing. Journal of the American Medical Informatics Association. 2021;28(9):1936-46.

7. Churpek MM, Carey KA, Edelson DP, et al. Internal and external validation of a machine learning risk score for acute kidney injury. JAMA Network Open. 2020;3(8):e2012892-e.

8. 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.

9. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model*. Critical Care Medicine. 2018;46(7).

10. Shou X, Mavroudeas G, New A, et al., editors. Supervised mixture models for population health. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2019 18-21 Nov. 2019.

11. Xu Z, Chou J, Zhang XS, et al. Identifying sub-phenotypes of acute kidney injury using structured and unstructured electronic health record data with memory networks. Journal of Biomedical Informatics. 2020;102:103361.

12. Chaudhary K, Vaid A, Duffy Á, et al. Utilization of deep learning for subphenotype identification in sepsis-associated acute kidney injury. Clinical Journal of the American Society of Nephrology. 2020;15(11):1557.

13. Baytas IM, Xiao C, Zhang X, Wang F, Jain AK, Zhou J. Patient subtyping via time-aware lstm networks. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Halifax, NS, Canada: Association for Computing Machinery; 2017. p. 65–74.

14. Batal I, Valizadegan H, Cooper GF, Hauskrecht M. A temporal pattern mining approach for classifying electronic health record data. ACM Trans Intell Syst Technol. 2013;4(4).

15. Patel D, Hsu W, Lee ML. Mining relationships among interval-based events for classification. Proceedings of the 2008 ACM SIGMOD international conference on Management of data; Vancouver, Canada: Association for Computing Machinery; 2008. p. 393–404.

16. Allen JF. Towards a general theory of action and time. Artificial Intelligence. 1984;23(2):123-54.

17. Huang Z, Lu X, Duan H. On mining clinical pathway patterns from medical behaviors. Artif Intell Med. 2012;56(1):35-50.

18. Phinney MA, Zhuang Y, Lander S, Sheets L, Parker JC, Shyu C-R. Contrast mining for pattern discovery and descriptive analytics to tailor sub-groups of patients using big data solutions. Medinfo 2017: Precision healthcare through informatics: IOS Press; 2017. p. 544-8.
19. García-Vico AM, Carmona CJ, Martín D, García-Borroto M, del Jesus MJ. An overview of emerging pattern mining in supervised descriptive rule discovery: Taxonomy, empirical study, trends, and prospects. WIREs Data Mining and Knowledge Discovery. 2018;8(1):e1231.
20. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation. 2002;6(2):182-97.
21. Sahoo A, Chandra S. Multi-objective grey wolf optimizer for improved cervix lesion classification. Applied Soft Computing. 2017;52:64-80.
22. Faris H, Habib M, Faris M, Alomari M, Alomari A. Medical speciality classification system based on binary particle swarms and ensemble of one vs. Rest support vector machines. Journal of Biomedical Informatics. 2020;109:103525.
23. Shahbeig S, Rahideh A, Helfroush MS, Kazemi K. Gene selection from large-scale gene expression data based on fuzzy interactive multi-objective binary optimization for medical diagnosis. Biocybernetics and Biomedical Engineering. 2018;38(2):313-28.
24. Mitra S, Banka H. Multi-objective evolutionary biclustering of gene expression data. Pattern Recognition. 2006;39(12):2464-77.
25. García-Vico ÁM, Carmona CJ, González P, Jesus MJd. Moea-efep: Multi-objective evolutionary algorithm for extracting fuzzy emerging patterns. IEEE Transactions on Fuzzy Systems. 2018;26(5):2861-72.
26. Fan H, Ramamohanarao K, editors. Efficiently mining interesting emerging patterns. International Conference on Web-Age Information Management; 2003: Springer.
27. Khwaja A. Kdigo clinical practice guidelines for acute kidney injury. Nephron Clinical Practice. 2012;120(4):c179-c84.
28. Qualls LG, Phillips TA, Hammill BG, et al. Evaluating foundational data quality in the national patient-centered clinical research network (pcornet®). EGEMS (Wash DC). 2018;6(1):3-.
29. Song X, Yu ASL, Kellum JA, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nature Communications. 2020;11(1):5668.
30. Ibrahim AE, Sarhane KA, Fagan SP, Goverman J. Renal dysfunction in burns: A review. Ann Burns Fire Disasters. 2013;26(1):16-25.
31. Deb K, Agrawal S, Pratap A, Meyarivan T, editors. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. Parallel Problem Solving from Nature PPSN VI; 2000 2000//; Berlin, Heidelberg: Springer Berlin Heidelberg.
32. Blank J, Deb K. Pymoo: Multi-objective optimization in python. IEEE Access. 2020;8:89497-509.
33. Gerent A, Almeida J, Almeida E, et al. Base deficit and sofa score are predictive factors of early acute kidney injury in oncologic surgical patients. Crit Care. 2015;19(Suppl 1):P297-P.
34. Nowak A, Breidthardt T, Dejung S, et al. Natriuretic peptides for early prediction of acute kidney injury in community-acquired pneumonia. Clin Chim Acta. 2013;419:67-72.
35. Pleguezuelo M, Benitez JM, Jurado J, Montero JL, De la Mata M. Diagnosis and management of bacterial infections in decompensated cirrhosis. World J Hepatol. 2013;5(1):16-25.
36. Francoz C. Acute kidney injury in cirrhosis: An immediate threat but also a ticking time bomb. Journal of Hepatology. 2020;72(6):1043-5.
37. Chancharoenthana W, Leelahavanichkul A. Acute kidney injury spectrum in patients with chronic liver disease: Where do we stand? World J Gastroenterol. 2019;25(28):3684-703.
38. Yu SM-W, Bonventre JV. Acute kidney injury and progression of diabetic kidney disease. Advances in Chronic Kidney Disease. 2018;25(2):166-80.
39. Chen J, Zeng H, Ouyang X, et al. The incidence, risk factors, and long-term outcomes of acute kidney injury in hospitalized diabetic ketoacidosis patients. BMC Nephrology. 2020;21(1):48.
40. Van der Maaten L, Hinton G. Visualizing data using t-sne. Journal of machine learning research. 2008;9(11).
41. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python. the Journal of machine Learning research. 2011;12:2825-30.
42. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-8.