

Development and validation of a Multi-Causal investigation and discovery framework for knowledge harmonization (MINDMerge): A case study with acute kidney injury risk factor discovery using electronic medical records

Mingyang Zhang^{a,b,1}, Xiangzhou Zhang^{a,c,1}, Mingyang Dai^{a,d}, Lijuan Wu^{e,f}, Kang Liu^{a,b}, Hongnian Wang^{a,b}, Weiqi Chen^{g,*}, Mei Liu^{h,*}, Yong Hu^{a,c,*}

^a Big Data Decision Institute, Jinan University, Guangzhou 510632, PR China

^b School of Management, Jinan University, Guangzhou 510632, PR China

^c School of Medicine, Jinan University, Guangzhou 510632, PR China

^d College of Information Science and Technology, Jinan University, Guangzhou 510632, PR China

^e Institute of Sciences in Emergency Medicine, Department of Emergency Medicine, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 519041, PR China

^f Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 519041, PR China

^g School of Computer Science, Guangdong Polytechnic Normal University, 510632, PR China

^h Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, 32610, USA

ARTICLE INFO

Keywords:

Causal discovery
Causal network synthesis
Network cycle-breaking
Acute kidney injury
Risk factor identification

ABSTRACT

Objective: Accurate diagnoses and personalized treatments in medicine rely on identifying causality. However, existing causal discovery algorithms often yield inconsistent results due to distinct learning mechanisms. To address this challenge, we introduce MINDMerge, a multi-causal investigation and discovery framework designed to synthesize causal graphs from various algorithms.

Methods: MINDMerge integrates five causal models to reconcile inconsistencies arising from different algorithms. Employing credibility weighting and a novel cycle-breaking mechanism in causal networks, we initially developed and tested MINDMerge using three synthetic networks. Subsequently, we validated its effectiveness in discovering risk factors and predicting acute kidney injury (AKI) using two electronic medical records (EMR) datasets, eICU Collaborative Research Database and MIMIC-III Database. Causal reasoning was employed to analyze the relationships between risk factors and AKI. The identified causal risk factors of AKI were used in building a prediction model, and the prediction model was evaluated using the area under the receiver operating characteristics curve (AUC) and recall.

Results: Synthetic data experiments demonstrated that our model outperformed significantly in capturing ground-truth network structure compared to other causal models. Application of MINDMerge on real-world data revealed direct connections of pulmonary disease, hypertension, diabetes, x-ray assessment, and BUN with AKI. With the identified variables, AKI risk can be inferred at the individual level based on established BNs and prior information. Compared against existing benchmark models, MINDMerge maintained a higher AUC for AKI prediction in both internal (AUC: 0.832) and external network validations (AUC: 0.861).

Conclusion: MINDMerge can identify causal risk factors of AKI, serving as a valuable diagnostic tool for clinical decision-making and facilitating effective intervention.

1. Introduction

Epidemiologic studies aim to scrutinize disease patterns within

specific groups to infer potential causes and consequences. However, establishing definitive causality from an association between exposure risk factors and a disease is challenging as correlations do not imply

* Corresponding authors.

E-mail addresses: chenwqi@foxmail.com (W. Chen), mei.liu@ufl.edu (M. Liu), yonghu@jnu.edu.cn (Y. Hu).

¹ These authors contributed equally to this study.

causation. Solely relying on correlations identified through statistical or machine learning (ML) approaches is inadequate for informed decision making in healthcare. Causal discovery holds promise in uncovering crucial insights for medical diagnosis [1], and its growing relevance within healthcare presents a transformative pathway for advancing precision medicine and personalized diagnosis. Bayesian networks (BNs), providing a formal probabilistic framework to model and infer causal relationships among variables [2], offer a robust approach for classification. BNs have widespread utility across scientific domains, including biomedicine [3,4], and their application in healthcare could revolutionize medical decision-making by fostering a deeper understanding of causal relationships.

Causal discovery using electronic medical records (EMRs) is challenging due to the inherent complexity and limitations of observational data. Over the years, various causal discovery methods have shown promise and can be classified into three types: constraint-based [5,6], score-based [7], and hybrid algorithms [8,9]. Constraint-based algorithms assess network structures through conditional independence tests (e.g., chi-squared tests or mutual information tests). However, their computational complexity exponentially increases as the number of nodes, leading to reduced efficiency and reliability. Score-based algorithms employ scoring functions (e.g., BIC, BDeu, K2) and search methods (e.g., hill-climbing (HC) or tabu search algorithms) to identify underlying network structures. Yet, obtaining an optimal network structure within a large structural space remains challenging. Hybrid approaches, such as H2PC [9] and MMHC [10], combine the strengths of constraint-based and score-based algorithms. However, the selection of an appropriate causal algorithm remains an ongoing challenge. Comparative assessments of different BNs' accuracy and efficiency are presented by Scutari et al [11] and Hussung et al [12]. The general consensus of these studies is that no single causal method exhibits significantly superior performance in structure reconstruction. Moreover, individual causal algorithms often suffer from local optimality and limited generalization. Applying different causal algorithms to the same dataset frequently yields different results, contributing to inconsistent knowledge.

To improve the accuracy and efficiency of causal structure search, several scholars have adopted intelligent optimization and ensemble methods. Evolutionary algorithms (e.g., ant colony [13,14], particle swarm [15–17], genetic algorithm [18,19]) and combinatorial optimization algorithms (e.g., Information Flow (IF) Theory [20,21], Breeding Swarm [16], Glowworm swarm optimization algorithm [22], Greedy Equivalence Search [23]) have been proposed for learning BNs. However, limitations are also evident. Firstly, most of these methods rely on a single causal discovery method, potentially compromising the accuracy of discovered knowledge due to algorithmic bias. Secondly, optimization-based methods often suffer from reduced applicability owing to algorithm complexity and extensive search time. Some ensemble learning approaches for causal discovery (e.g., MIC (maximum information coefficient) [24], perturbed features [25], causal IF theory [26], causal strength scoring matrix [27], and weighted adjacency matrix of local data slices [28]) may overlook algorithm bias and the impact of data perturbation on causal algorithms. Most importantly, the potential existence of cyclic structures is often disregarded when synthesizing network knowledge acquired from different sources. This oversight may result in erroneous inferences.

In summary, data-driven causal discovery faces two critical challenges: 1) the diversity of causal discovery algorithms, and 2) the presence of cyclic structures in synthesized causal networks. To address these challenges, we propose MINDMerge, a multi-causal investigation and discovery framework aimed at harmonizing networks derived from multiple causal discovery algorithms to enhance accuracy, robustness, and generalization. Our evaluation of this framework spans synthetic and real-world datasets, investigating causal relationships among clinical variables and acute kidney injury (AKI). AKI is a prevalent complication among critically ill patients [29,30], associated with

significantly high mortality and morbidity rates [31]. Globally, an estimated 13.3 million individuals suffer from AKI annually, with 85 % in developing countries [32], and over 70 % of cases undiagnosed within 3 days [33]. Additionally, AKI is prevalent among patients hospitalized for COVID-19 [34,35]. However, etiology of AKI is complex, posing challenges in prediction and risk factor discovery. Some studies employ machine learning algorithms, such as [36–38], to analyze electronic health record (EHR) data and uncover patterns related to AKI occurrence. While the methodologies employed in these studies provide valuable insights into AKI risk prediction and factor discovery, they may be limited by their reliance on correlation-based analyses and not establish causal relationships. A robust causal network model, such as MINDMerge, can elucidate potential AKI risk factors, supporting clinical decisions and enabling effective interventions.

2. Material and methods

The proposed MINDMerge framework was initially tuned and evaluated using synthetic data before its application to EMR data for causal relationship discovery and AKI risk prediction. In this study, BN learning from observational data operated on the following assumptions:

- Strong causality between variables without symmetry.
- Data completeness – If the dataset is complete and the observations correctly reflect the true state of the variables, the causal network learned from the data will accurately reflect the true causal relationship.
- Causal sufficiency – It was assumed that all relevant variables were observed without any hidden variables.

These assumptions aimed to simplify the reasoning and interpretation of the BN model. Deviations from these assumptions could potentially complicate the establishment of causality.

2.1. Synthetic data

Synthetic data plays an important role in evaluating the efficacy of methods. We developed and evaluated the proposed MINDMerge framework using three widely used synthetic datasets – Cancer [39], Child [40], Alarm [41] – representing varying sizes and distinct structures for BN structure learning. Experiments were conducted on each dataset under different settings with sample sizes ranging from 500 to 5000. The Cancer dataset comprises 5 nodes, 4 arcs, and 10 parameters, representing a small network used in clinical diagnosis. The Child dataset, designed for referrals of newborn babies with congenital heart disease in London, comprises 20 nodes, 25 arcs, and 230 parameters, making it a medium-sized network. On the other hand, the Alarm dataset, based on an alarm message system for emergency patient monitoring, includes 37 nodes, 46 arcs, and 509 parameters, representing a large network. Synthetic data in the form of RDA files was downloaded from the *Bayesian network Repository* and the 'rbn' function was used to generate samples of desired sizes. Source data is available at <https://www.bnlearn.com/bnrepository>.

2.2. Real-world data

The proposed MINDMerge framework was also validated using two de-identified EMR datasets, both of which comply with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, a US federal law designed to protect patient health information. Access permissions were obtained for research purposes after the researchers completed the MIT CITI training modules, which are specifically designed for the secondary use of de-identified data.

- **The eICU Collaborative Research Database (eICU)** is a large multi-center critical care database made available by Philips Healthcare in

partnership with the MIT Laboratory for Computational Physiology [42]. It contains over 200,000 distinct ICU admissions between 2014 and 2015 across 208 US hospitals. We collected 92 variables from eICU dataset, encompassing demographic information, vital signs, laboratory tests, comorbidities and procedures, based on clinician recommendations.

- **The Medical Information Mart for Intensive Care III (MIMIC-III)** comprises of deidentified health-related data associated with over 40,000 patients admitted to critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. For our analysis, we collected 66 variables from MIMIC-III dataset, guided by clinician recommendations. The dataset served as an independent external validation set.

2.2.1. Description of study cohort

The study included adult patients (age ≥ 18) admitted to the hospital for a minimum of 2 days. AKI was defined using the KDIGO (Kidney Disease Improving Global Outcomes) guidelines, which was determined by:

- An increase in serum creatinine (SCr) levels by ≥ 0.3 mg/dL (26.5 μ mol/L) within 48 h OR
- An increase of > 1.5 times the baseline SCr level within seven days OR
- Urine output below 0.5 mL/kg/h for more than 6 h.

2.2.2. Data processing

For each patient, we collected information on clinical variables up to 48 h before the prediction point (Supplementary Fig. S1A). For patients with AKI, the prediction point was 48 h before onset. For non-AKI patients, we set the prediction point to 48 h before the last SCr measurement. For clinical variables with longitudinal measures, we only used their most recent value. SCr and eGFR were not included as variables because they were used to determine AKI occurrence. Data extraction flowchart can be found in Supplementary Fig. S1B.

We processed datasets from four hospitals in the eICU database, forming a final cohort of 9688 patients: eICU-1 (2,477 patients; AKI incidence of 27.6 %), eICU-2 (4,280 patients; AKI incidence of 29.06 %), eICU-3 (1,630 patients; AKI incidence of 33.37 %), and eICU-4 (1,301 patients; AKI incidence of 32.2 %). Additionally, a cohort from the MIMIC-III database was extracted and processed, comprising 15,298 patients with an AKI incidence rate of 30.76 %. For experiments, eICU-1 served as the derivation cohort, while eICU-2 served as the external validation cohort. The remaining cohorts – eICU-3, eICU-4, and MIMIC-III – were utilized for external geographical validation to assess generalizability.

We adopted Li et al.’s approach [43] for handling missing data. Among vital signs and laboratory tests, variables with missing rate exceeding 75 % were excluded, while the remaining variables were imputed using the predictive mean matching (pmm) method with multiple imputation by Chained Equations (MICE). Although other methods such as KNN and Random Forest were considered, the pmm method showed the least discrepancy between distributions of the original and the imputed data by the Kolmogorov-Smirnov test (Supplementary Table S1 and Fig. S2).

Continuous variables were discretized, and discretization thresholds were chosen based on existing knowledge (MSD MANUAL Professional Version). The continuous features were divided into different categories as shown in Supplementary Table S2. We took the interaction of three feature selection methods, including the Chi-square (Chi2) method with a significance level $p < 0.05$, Boruta, and Random Forest, were applied to exclude irrelevant variables, resulting in 26 variables (Supplementary Fig. S3 and Table S3). And generate a causal network structure based on these features. Baseline descriptive characteristics of other cohorts are shown in Supplementary Table S4-S7.

2.3. Mindmerge framework

2.3.1. Causal Bayesian network

The constraint-based methods have been discarded due to the presence of undirected edges within the structure. We employed five causal discovery algorithms: three score-based algorithms (BIC [44,45], BDeu [46,47] and BDla [48,49]) with tabu-search, and two hybrid algorithms (MMHC [8,10] and H2PC [9,50]). Detailed descriptions of each method can be found in Supplementary method 1, and their key properties are summarized in Table 1. These algorithms excel in identifying specific data pattern, ensuring efficiency and ease of implementation while achieving a balanced trade-off between model fitting and complexity. However, they ignore the diverse knowledge inherent in the data. By combining different causal algorithms, we can mitigate reliance on a single method and capture different information from data, resulting in more robust knowledge.

Algorithm ensemble can be approached in two ways. Homogeneous ensembles utilize similar models, but may not capture the diverse and complex relationships in data, thus limiting generalizability. Conversely, heterogeneous ensembles leverage multiple algorithms to capture patterns from diverse perspectives to enhance robustness. However, interpreting results from such a diverse set of models pose challenges due to their complexity. In this study, we sought to leverage the strengths of both homogeneous and heterogeneous ensembles within a hybrid framework for harmonizing causal networks. This approach allows synthesis of the best aspects of each method to attain a more comprehensive understanding of causal relationships.

2.3.2. Network fusion strategy

We designed a credibility-weighted fusion strategy, CW, to evaluate the final credibility score of each edge. A higher the score denotes higher the credibility, indicating a greater likelihood of the edge’s existence. Subsequently, edges with weak credibility were filtered out based on the threshold θ to derive a transition matrix. The credibility weights are defined as follows:

$$CW = \sum_{i=1}^n \sqrt{\frac{N_{node}}{N_i}} M_i \odot \frac{\sum_{i=1}^n M_i}{n}$$

Table 1
Description of the key properties of baseline model.

Structure Learning		Description
Score-based	BIC	Overfitting can be avoided by selecting models based on their log-likelihood and a penalty term for model complexity
	BDeu	It is assumed that the prior of the parameters obeying Dirichlet distribution, and it is used as the basis to find the optimal structure, which does not need to obtain the node order in advance. The key properties of BDeu arise from its uniform prior over the parameters of each local distribution in the network, which makes structure learning computationally efficient; it does not require the elicitation of prior knowledge from experts; and it satisfies score equivalence.
	BDla	BDla can avoid the need to set free parameters and produce better results when the parameter space of the model is complex (a mix of uniform and skewed parameter distributions).
Hybrid-based	MMHC	MMHC is capable of dealing with thousands of nodes in reasonable time. Firstly, the MMPC (max–min parents and children) algorithm was used to determine the parent and child node sets of each node, so as to construct the network structure framework. Then the frame of the obtained network structure is searched and scored according to the K2 search strategy to obtain the optimal network structure.
	H2PC	It first reconstructs the skeleton of a Bayesian network by integrating multiple PC algorithms to identify the parents and children set of each variable, and then performs a greedy hill-climbing search to filter and orient the edges.

where n is number of causal algorithms involved in the ensemble, N_{node} refers to the number of nodes in the dataset, N_i refers to the numbers of arcs of causal algorithm i , the larger the value of N_{node}/N_i , the larger the weight, and the more significant the causality of arcs, M_i is an 0–1 adjacency matrix of BN structure, \odot is Hadamard product.

When merging causal networks, a circular structure can emerge, causing causal conflicts. Since BNs are DAGs, it is crucial to derive an acyclic weighted adjacency matrix W that approximates the cyclic matrix W_0 . However, manual deletion in large-scale dataset is infeasible. To solve this problem, we referred to the NOTEARS algorithm introduced by Zheng et al [51]. NOTEARS works by converting the standard combinatorial optimization problem into a continuous regularization problem on matrices, and then solving it using a numerical optimization algorithm. Thus

$$\min_{W \in \mathbb{R}^{d \times d}} F(W)$$

subject to $G(W) \in \text{DAGs}$.

$$\Leftrightarrow \min_{W \in \mathbb{R}^{d \times d}} F(W)$$

$$\text{subject to } h(W) = 0 \quad (1)$$

where $F(W)$ is a score function on $\mathbb{R}^{d \times d}$, $G(W)$ is a directed acyclic graph generated by the weighted adjacency matrix W . Zheng et al confirmed the existence of such a smooth function h : A matrix $W \in \mathbb{R}^{d \times d}$ is a DAG if and only if $h(W) = \text{tr}(e^{W \odot W}) - d = 0$, where d is the dimension of matrix W , $\text{tr}(\cdot)$ is trace of matrix, \odot is Hadamard product and e^W is the matrix exponential of W .

Given that the loss function of NOTEARS is designed for linear structural equation models (SEM), it's essential to redefine the loss function appropriately in light of the distinct context of our study. So, we proposed a novel cycle-breaking algorithm termed CBAMN (Cycle-Breaking Algorithm based on Modified NOTEARS). CBAMN aims to identify acyclic solutions that closely resemble the cyclic solutions obtained through an iterative optimization process. we redefined a loss function:

$$F(W) = \frac{\lambda_1}{2} \|W_0 - W\|_F^2 + \lambda_2 \|W\|_1$$

$$\text{subject to } h(W) = 0, \quad (2)$$

where F is Frobenius norm, W_0 is matrix with a cycle, W is the final ideal acyclic matrix. The primary modification involves the incorporation of a cyclicity penalty term, which penalizes cyclic structures within the learned causal graph. This penalty incentivizes the algorithm to prioritize acyclic solutions while discouraging the formation of cyclic dependencies. Additionally, we introduce a regularization term to promote sparsity in the learned graph, facilitating simpler and more interpretable causal models. $\|W_0 - W\|_F^2$ corresponds to the least squares loss of matrix differences, adjusting the degree of approximation between two matrices via the parameter $\lambda_1 > 0$. The second term is the penalty term. To obtain an ideal matrix, the sum of all weights must be minimized. The sparsity of matrix can be adjusted by penalty parameter $\lambda_2 > 0$. Similar to NOTEARS, we utilized the augmented Lagrange algorithm [51,52] to solve equation (2) by converting the objective function into a dual problem, based on the given the objective function and constraint conditions:

$$L^\rho(W, \alpha) = \arg \min_{W \in \mathbb{R}^{d \times d}} \left\{ F(W) + \frac{\rho}{2} |h(W)|^2 + \alpha h(W) \right\} \quad (3)$$

where α is Lagrange multiplier, ρ is penalty parameter. Therefore, we can perform dual gradient ascent to optimize $\alpha = \alpha + \rho h(W^*)$, where W^*

is the local minimizer matrix of this iteration. The optimization process of CBAMN is shown in **Algorithm 1**.

Algorithm 1: CBAMN Algorithm

Input: degree of decay $c \in (0, 1)$, growth rate of penalty $r > 1$, constraint term coefficient α_0 , optimize error rate $\varepsilon > 0$, dual function: $L^\rho(W, \alpha)$
 Output: A directed acyclic graph W^*

```

for  $t = 0, 1, 2, \dots, \infty$  do:
   $W_{t+1} = \arg \min W L^\rho(W, \alpha_t)$  if  $h(W_{t+1}) \geq c \times h(W_t)$  then
     $\rho = r\rho$ 
  continue
  end if
   $\alpha_{t+1} = \alpha_t + \rho h(W_{t+1})$  if  $h(W_{t+1}) < \varepsilon$  then
    break
  end if
end for
Output  $W^*$ 

```

2.3.3. Experimental framework

To evaluate the proposed MINDMerge framework, we employed eleven baseline algorithms. These included 5 single-cause discovery algorithms (BDeu, BDla, BIC, H2PC, MMHC), their corresponding homogeneous ensemble versions (BDeu-en, BDla-en, BIC-en, H2PC-en, MMHC-en), and a heterogeneous ensemble algorithm combining all five causal algorithms (Het-en). Additionally, we compared MINDMerge against a voting-based network fusion strategy (Voting-N), which picks edges identified by majority of the algorithms. Fig. 1 provides an overview of the MINDMerge framework. Training data was extracted via bootstrapping from the original dataset. Phase 1 involved causal discovery on these bootstrapped samples to generate Phase 1 ensemble results. Phase 2 ensemble processing consolidated the Phase 1 results into a final output. Finally, given that the final output may be a cyclic graph, it is converted into an acyclic graph using CBAMN (Algorithm 1). Supplementary Fig.S4 provides an example shows the algorithm.

The overall dataset was split into training, internal validation (20 % patients), and external validation (include network validation and geographic validation). Bootstrapping ($N = 50$) was performed on the training data. Following this, each causal discovery method was applied across all bootstrap samples. We investigated credibility-weighted ensemble methods, generating weighted matrices EBN1-EBN50 derived from five different causal discovery methods on each bootstrap sample.

Subsequently, multiple BN structures were integrated using credibility weighting, and edges with weak reliability was filtered based on a predetermined threshold, resulting in a transition matrix. Finally, a cycle-breaking algorithm was applied to the transition matrix to obtain the adjacency matrix of DAG.

For parameter optimization, we performed parameter tuning on synthetic data before deploying MINDMerge with the identified optimal parameters on real-world EMR data for AKI risk prediction. The final learned network structure was externally validated. Additionally, to assess the portability of MINDMerge, we performed external geographic validation.

2.3.4. Model evaluation

Using synthetic data, we directly compared the learned network with the ground-truth network using Recall and F1-score. TP refers to the number of edges that exist in both the ground-truth and the learned networks, FP refers to the number of edges that exist in the learned network but not in the ground-truth network, and FN refers to the

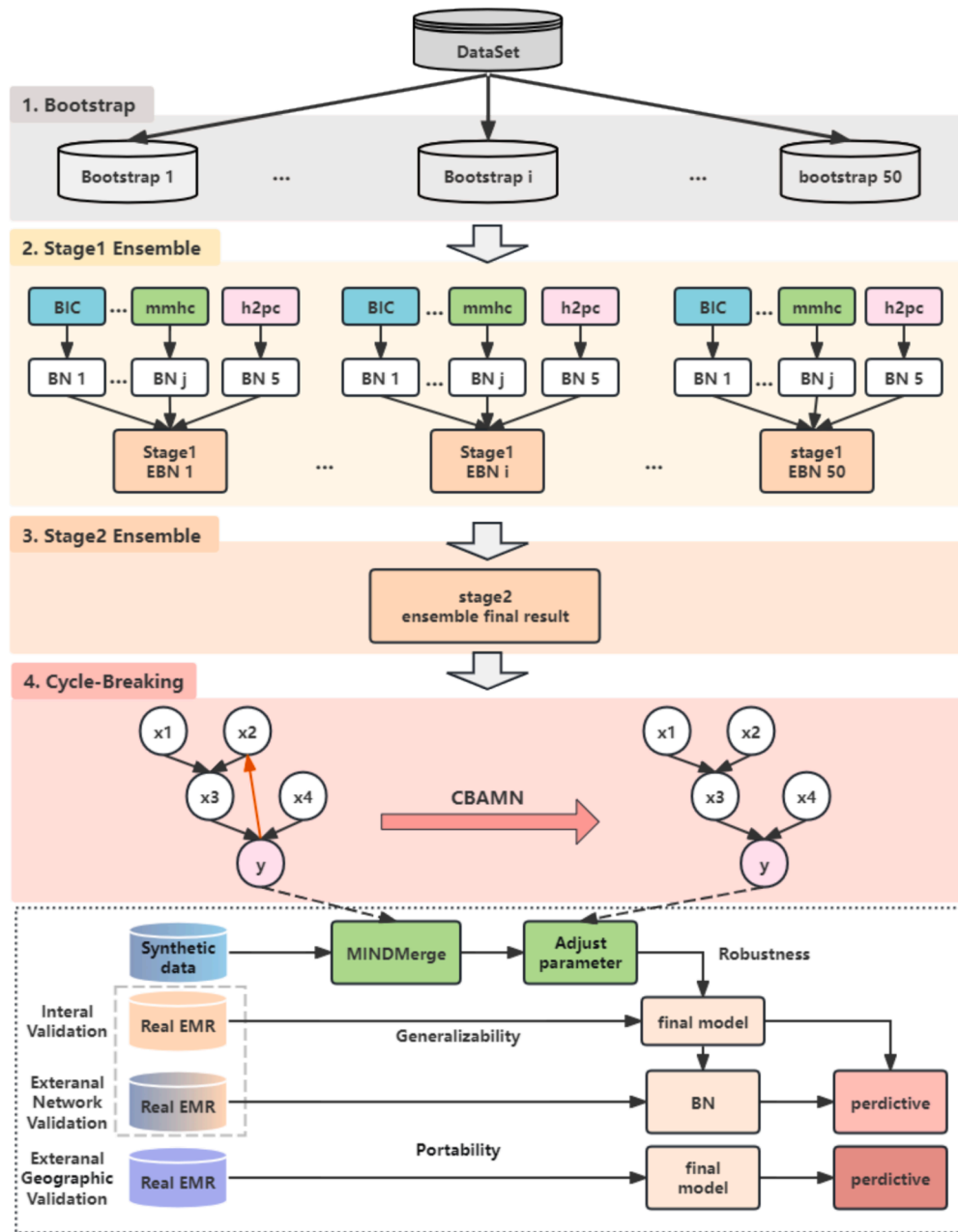


Fig. 1. Overview of the MINDMerge framework. MINDMerge (Multi-Causal Investigation and Discovery Framework for Knowledge Harmonization), CBAMN (novel cycle-breaking algorithm based on modify NOTEARS).

number of edges that exist in the ground-truth network but not in the learned network. We repeated the experiment 10 times and reported the average performance. Achieving comparable prediction accuracy indicates the model's ability to capture the causal relationships between variables.

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$F1 - \text{score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

In the real-world data application, we compared the prediction performance of various classification models – CART, C5.0, Random Forest (RF), XGBoost, Logistic Regression (LR), Support Vector Machine

(SVM), ANN, and Naïve Bayes (NB) – using the area under the receiver operating characteristic curve (AUC) and Recall metrics. The performance evaluation involved 5-fold cross-validation, and we determined the optimal threshold using the Youden index.

3. Results

3.1. Experiments on synthetic datasets

To strike a balance between achieving a network closer to the ground-truth (i.e., high Recall) and maintaining optimal sparsity, we set hyperparameters $\lambda_1 = 1$, $\lambda_2 = 0.5$, and the threshold $\theta = 0.85$. [Supplementary Fig. S5 and Table S8](#) provide an in-depth sensitivity analysis of these parameters.

Across various datasets – Cancer, Child, and Alarm – we conducted comparisons among different causal algorithms using MINDMerge with

Table 2

Structure learning results of MINDMerge ensemble algorithm and other algorithms under different dataset sizes (best performances are highlighted in bold).

Method	Recall Cancer 500	F1	Recall Cancer 1 K	F1	Recall Cancer 2 K	F1	Recall Cancer 5 K	F1
BDeu	0.6(0.129)	0.732(0.09)	0.7(0.158)	0.804(0.112)	0.775(0.142)	0.814(0.115)	0.925(0.169)	0.925(0.169)
BDla	0.25(0.289)	0.276(0.311)	0.425(0.237)	0.485(0.285)	0.275(0.079)	0.297(0.099)	0.8(0.329)	0.797(0.334)
BIC	0.325(0.169)	0.508(0.131)	0.6(0.175)	0.729(0.168)	0.65(0.129)	0.744(0.136)	0.9(0.129)	0.932(0.093)
H2PC	0.35(0.129)	0.495(0.152)	0.625(0.177)	0.748(0.171)	0.65(0.129)	0.725(0.153)	0.925(0.121)	0.936(0.108)
MMHC	0.35(0.129)	0.5(0.145)	0.575(0.169)	0.729(0.168)	0.65(0.129)	0.735(0.145)	0.875(0.132)	0.896(0.116)
BDeu-en	0.75(0.204)	0.753(0.135)	0.8(0.197)	0.793(0.192)	0.8(0.197)	0.807(0.177)	0.95(0.158)	0.95(0.158)
BDla-en	0.425(0.265)	0.425(0.279)	0.6(0.269)	0.577(0.245)	0.4(0.175)	0.395(0.172)	0.75(0.264)	0.721(0.267)
BIC-en	0.6(0.175)	0.677(0.157)	0.75(0.167)	0.786(0.127)	0.825(0.121)	0.835(0.102)	0.975(0.079)	0.964(0.083)
H2PC-en	0.6(0.211)	0.656(0.183)	0.7(0.105)	0.758(0.101)	0.775(0.142)	0.814(0.115)	0.975(0.079)	0.975(0.079)
MMHC-en	0.6(0.175)	0.662(0.15)	0.675(0.121)	0.731(0.1)	0.8(0.158)	0.828(0.129)	0.975(0.079)	0.975(0.079)
Het-en	0.45(0.158)	0.596(0.152)	0.625(0.177)	0.748(0.171)	0.675(0.121)	0.721(0.13)	0.925(0.121)	0.946(0.091)
Voting-N	0.3(0.197)	0.533(0.143)	0.5(0.236)	0.701(0.146)	0.675(0.121)	0.78(0.108)	0.875(0.132)	0.929(0.075)
MINDMerge	0.825(0.169)	0.66(0.149)	0.825(0.121)	0.727(0.077)	0.85(0.129)	0.781(0.063)	1(0)	0.954(0.071)
Child 500			Child 1 K		Child 2 K		Child 5 K	
BDeu	0.676(0.072)	0.746(0.079)	0.74(0.047)	0.782(0.044)	0.748(0.092)	0.765(0.092)	0.784(0.087)	0.793(0.096)
BDla	0.6(0.105)	0.653(0.119)	0.708(0.057)	0.733(0.059)	0.82(0.087)	0.824(0.091)	0.808(0.045)	0.790(0.056)
BIC	0.66(0.069)	0.738(0.072)	0.736(0.054)	0.784(0.051)	0.748(0.073)	0.757(0.077)	0.788(0.09)	0.793(0.096)
H2PC	0.584(0.057)	0.679(0.06)	0.7(0.043)	0.764(0.043)	0.736(0.054)	0.784(0.044)	0.86(0.028)	0.869(0.021)
MMHC	0.496(0.051)	0.608(0.059)	0.592(0.037)	0.7(0.041)	0.628(0.038)	0.728(0.034)	0.6(0.044)	0.695(0.051)
BDeu-en	0.74(0.043)	0.754(0.051)	0.764(0.058)	0.769(0.063)	0.82(0.078)	0.805(0.086)	0.776(0.063)	0.761(0.068)
BDla-en	0.724(0.051)	0.689(0.051)	0.788(0.078)	0.774(0.069)	0.84(0.027)	0.806(0.028)	0.848(0.032)	0.824(0.038)
BIC-en	0.692(0.063)	0.729(0.065)	0.788(0.05)	0.797(0.061)	0.824(0.078)	0.803(0.079)	0.788(0.065)	0.778(0.07)
H2PC-en	0.608(0.082)	0.682(0.082)	0.692(0.046)	0.76(0.047)	0.732(0.042)	0.758(0.041)	0.824(0.021)	0.841(0.02)
MMHC-en	0.528(0.059)	0.634(0.065)	0.616(0.043)	0.713(0.051)	0.66(0.043)	0.75(0.047)	0.628(0.05)	0.702(0.053)
Het-en	0.64(0.068)	0.709(0.073)	0.692(0.038)	0.769(0.038)	0.792(0.067)	0.822(0.069)	0.836(0.048)	0.853(0.049)
Voting-N	0.652(0.053)	0.759(0.05)	0.736(0.034)	0.825(0.033)	0.876(0.035)	0.908(0.026)	0.928(0.017)	0.943(0.009)
MINDMerge	0.824(0.034)	0.79(0.038)	0.86(0.051)	0.83(0.061)	0.948(0.019)	0.917(0.033)	0.96(0)	0.958(0.006)
Alarm 500			Alarm 1 K		Alarm 2 K		Alarm 5 K	
BDeu	0.641(0.106)	0.602(0.111)	0.661(0.067)	0.639(0.065)	0.696(0.047)	0.659(0.047)	0.717(0.067)	0.686(0.077)
BDla	0.646(0.057)	0.587(0.063)	0.685(0.048)	0.65(0.057)	0.698(0.093)	0.656(0.093)	0.761(0.086)	0.728(0.088)
BIC	0.5(0.066)	0.543(0.072)	0.55(0.032)	0.574(0.038)	0.635(0.054)	0.643(0.064)	0.693(0.042)	0.698(0.046)
H2PC	0.393(0.046)	0.519(0.051)	0.459(0.075)	0.574(0.082)	0.554(0.043)	0.662(0.052)	0.663(0.06)	0.738(0.068)
MMHC	0.354(0.053)	0.484(0.066)	0.487(0.068)	0.606(0.077)	0.524(0.022)	0.644(0.029)	0.635(0.034)	0.726(0.04)
BDeu-en	0.65(0.11)	0.548(0.102)	0.722(0.061)	0.633(0.054)	0.713(0.06)	0.633(0.065)	0.752(0.059)	0.689(0.068)
BDla-en	0.62(0.076)	0.511(0.065)	0.652(0.072)	0.568(0.06)	0.661(0.045)	0.567(0.047)	0.693(0.076)	0.621(0.072)
BIC-en	0.502(0.045)	0.511(0.044)	0.602(0.047)	0.594(0.04)	0.643(0.031)	0.628(0.034)	0.665(0.034)	0.641(0.033)
H2PC-en	0.502(0.071)	0.615(0.076)	0.537(0.051)	0.631(0.056)	0.589(0.042)	0.68(0.044)	0.676(0.042)	0.737(0.05)
MMHC-en	0.459(0.068)	0.573(0.072)	0.515(0.045)	0.617(0.047)	0.574(0.055)	0.677(0.061)	0.641(0.029)	0.721(0.033)
Het-en	0.517(0.062)	0.619(0.066)	0.7(0.05)	0.698(0.052)	0.724(0.053)	0.712(0.06)	0.77(0.044)	0.807(0.048)
Voting-N	0.507(0.047)	0.625(0.051)	0.635(0.032)	0.712(0.031)	0.663(0.031)	0.731(0.04)	0.729(0.051)	0.799(0.053)
MINDMerge	0.691(0.059)	0.632(0.053)	0.735(0.025)	0.675(0.019)	0.746(0.041)	0.674(0.046)	0.811(0.051)	0.762(0.056)

varying sample sizes (500, 1 k, 2 k, 5 k). Table 2 presents the outcomes, highlighting that distinct causal methods yield varied learning accuracies for the same dataset. Ensemble algorithms exhibited superior performance compared to single causal algorithms, with MINDMerge consistently demonstrating better results than other ensemble approaches. Moreover, our observations indicate that learning efficiency is constrained with smaller sample sizes, and progressively improved as the sample size increases. The significance analysis is shown in Supplementary Fig. S6-S9. Fig. 2 illustrates the cosine similarity and Recall metrics for networks at a sample size of 5000. Notably, algorithmic bias is evident among different causal methods (Fig. 2A-C). MINDMerge exhibits the highest Recall (1, 0.96, 0.807 for Cancer, Child and Alarm datasets respectively) (Fig. 2D).

To evaluate the performance of the proposed cycle-breaking algorithm CBAMN in the MINDMerge framework, we compared it to other cycle removing algorithms like colored-DFS[15,53] and MIGGA [54] (Supplementary Table S9). As network size increases, CBAMN exhibited a more pronounced cycle-breaking effect, maintaining higher Recall and F1 compared to other methods.

3.2. Experiments on real-world datasets

Table 3 summarizes the characteristics of the real-world study cohort. In Fig. 3, the AKI incidence and crude odds ratio (cOR) for varied Age, Sex and BMI are illustrated in both the derivation and external network validation cohorts from the eICU database. When dividing age

into five bins, the AKI incidence increased with age and plateauing thereafter. Similarly, the division of BMI into four bins revealed a positive correlation with AKI, demonstrating a significant increase from 19.5 % to 31.8 % and 20.3 % to 34.2 % in the derivation and validation cohorts, respectively. Male patients exhibited a relatively higher AKI risk (28.5 % vs. 26.6 % and 31.1 % vs 26.2 %). Further details regarding AKI incidence and cOR in the geographical validation cohort can be found in Supplementary Fig. S10.

When applying MINDMerge to real-world EMR data, we imposed constraints based on domain knowledge and temporal considerations to ensure reliability of the derived causal network. For instance, no variables can be the cause of demographic variables like age, sex, and race. Consequently, all edges originating from any variable to “age” or “sex” or “race” were removed. Additionally, AKI as an outcome variable cannot be a cause of any variable. Following the experimental setup, causal network was learned using a training dataset (80 % of patients) and is presented in Fig. 4. Here, clinical variables serve as nodes, and directed edges represent causal relationships. The causal graph highlights pulmonary disease, hypertension, diabetes, and BUN as direct parent nodes contributing to AKI, representing direct relationships. Other variables such as age, sex, Paco2, Hco3, coronary artery disease, and stroke were identified as indirect risk factors associated with AKI, consistent with previous reports [29,33,55,56]. We have also explained the corresponding dependency (Supplementary Tables SA).

AKI risk can be inferred using the BNs (Fig. 4). The probabilistic dependencies between AKI and its risk factors are illustrated in

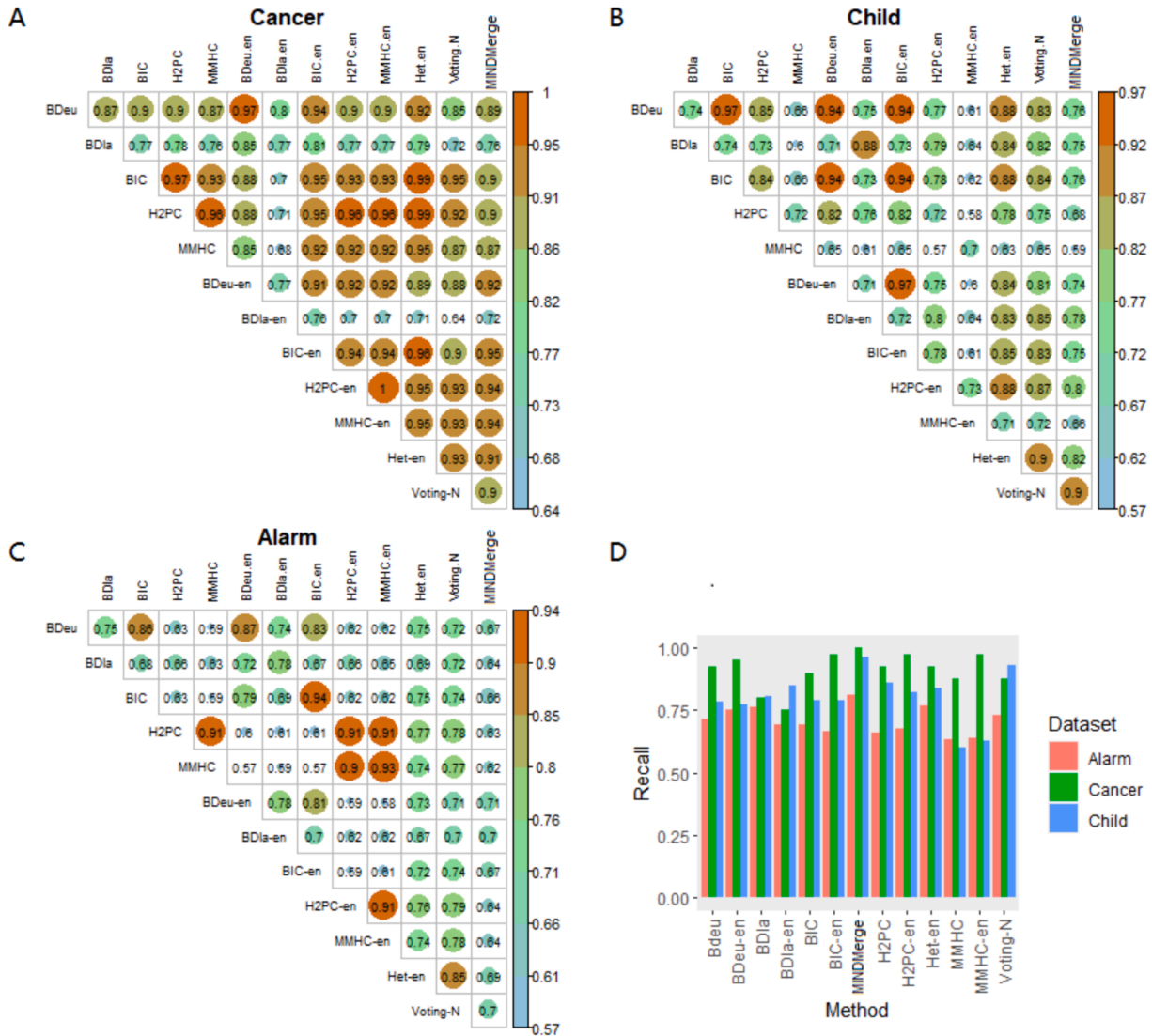


Fig. 2. The similarity and Recall of between different causal network. A, B, C shows the similarity of different causal discovery methods on the three datasets (i.e., Cancer, Child, and Alarm), D shows the Recall on the three datasets.

Supplementary Fig. S11A. Leveraging Bayes' theorem can allow the design of an intervention plan based on these inferred probabilities. Consider a hospitalized patient at age 68, without hypertension or diabetes, and has a normal BUN (blood urea nitrogen) level, the initial probability of this patient in developing AKI within 48 h stands at 33 %. However, upon monitoring, if the BUN level increases beyond the normal range, the patient's probability of developing AKI significantly increases from 33 % to 51 %. Moreover, before encountering abnormal BUN levels, the detection of urinary tract obstruction and the identification and treatment of potential AKI-associated factors through X-ray and CT scanners can significantly reduce the AKI risk from 33 % to 10 % as demonstrated in **Supplementary Fig. S11B-D**. This reasoning suggests that diagnosing AKI solely based on creatinine or urine volume might have a time lag.

BN not only enables causal inference but also facilitates accurate prediction of AKI risk. Leveraging MINDMerge, we constructed an AKI prediction model using Bayesian causal network classification (**Supplementary method 2**). **Fig. 5** illustrates the average performance and confidence intervals from 5-fold cross-validation on the test dataset. Notably, MINDMerge achieved the highest AUC compared to other

models. To assess statistical differences between performances, Scott-KnottESD [57,58] statistical rank test (SK-test) was used. As shown in **Supplementary Fig. S12**, MINDMerge exhibited significant difference in AUC compared to single causal algorithms, but not to other ensemble algorithms. Moreover, no statistically significant difference was observed in Recall. In contrast to other machine learning models, MINDMerge showed statistically significant improvements in both AUC and Recall. MINDMerge and XGBoost achieved the highest AUC and Recall, performing equally well. Furthermore, we selected the top-5 models to conduct a comprehensive analysis using the test set. **Supplementary Table S10** and **Fig. S13** present additional performance indicators such as Precision, Recall and AUPRC, decision curves, and calibration curves.

To assess the robustness of the causal network, **eICU-2** was used as an external validation dataset. **Supplementary Table S11** provides the AUCs and other performance indices for the top-5 models. In terms of AUC, the **eICU-1** (AUC is 0.861) model is 0.007 lower than the optimal XGBoost model, with a Recall of 0.744. When applying MINDMerge to **eICU-2**, it achieved an AUC of 0.848 and a Recall of 0.748. These results demonstrate that MINDMerge has excellent predictive performance on

Table 3

Patient characteristics of the derivation cohort and external validation cohort.

Characteristics	Derivation cohort	Externalnetwork validation	External geographical validation		
	eICU-1	eICU-2	eICU-3	eICU-4	MIMIC-III
Number of samples(%AKI)	2477(27.6 %)	4280 (29.06 %)	1630(33.37 %)	1301(32.2 %)	15298(30.76 %)
Age (AKI%)					
18–35	45 (6.6 %)	47 (3.8 %)	19 (3.5 %)	14 (3.3 %)	145 (3.1 %)
36–50	88 (12.9 %)	125 (10.0 %)	43 (7.9 %)	38 (9.1 %)	490 (10.4 %)
51–65	231 (33.8 %)	407 (32.7 %)	142 (26.1 %)	124 (29.6 %)	1294 (27.5 %)
66–80	228 (33.3 %)	461 (37.1 %)	212 (39.0 %)	178 (42.5 %)	1836 (39.0 %)
>80	92 (13.5 %)	204 (16.4 %)	128 (23.5 %)	65 (15.5 %)	941 (20.0 %)
Male (%)	376 (55.0 %)	773 (62.1 %)	326 (59.9 %)	248 (59.2 %)	2769 (58.8 %)
Race (%)					
Caucasian	453 (66.2 %)	1139 (91.6 %)	117 (21.5 %)	301 (71.8 %)	3839 (81.6 %)
African American	226 (33.0 %)	72 (5.8 %)	86 (15.8 %)	114 (27.2 %)	436 (9.3 %)
Hispanic	NA	2 (0.2 %)	335 (61.6 %)	2 (0.5 %)	152 (3.2 %)
Asian	5 (0.7 %)	28 (2.3 %)	6 (1.1 %)	2 (0.5 %)	279 (5.9 %)
Native American	NA	3 (0.2 %)	NA	NA	NA
BMI (%)					
<18.5	22 (3.2 %)	35 (2.8 %)	15 (2.8 %)	11 (2.6 %)	85 (1.8 %)
18.5–24.9	176 (25.7 %)	327 (26.3 %)	170 (31.3 %)	93 (22.2 %)	1318 (28.0 %)
25–29.9	187 (27.3 %)	348 (28.0 %)	169 (31.1 %)	119 (28.4 %)	1593 (33.9 %)
>29.9	299 (43.7 %)	534 (42.9 %)	190 (34.9 %)	196 (46.8 %)	1710 (36.3 %)

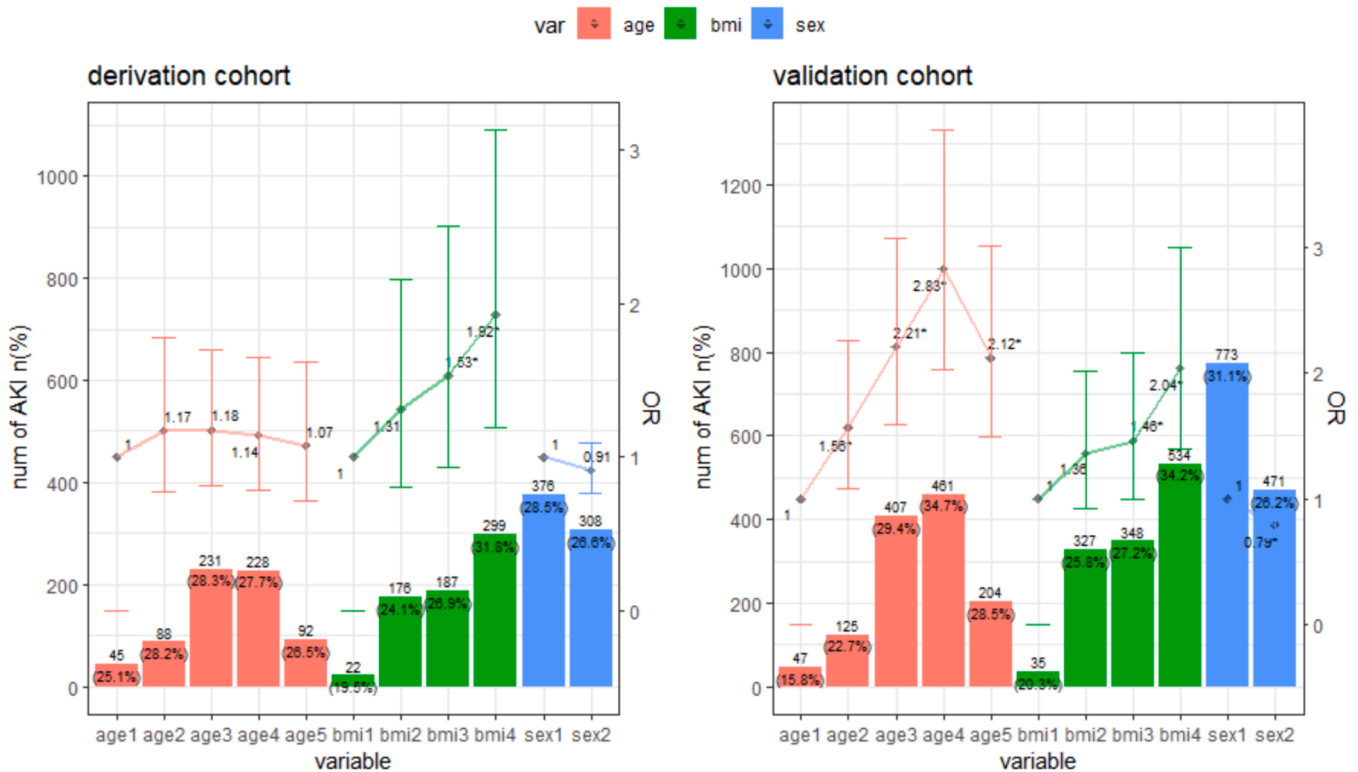


Fig. 3. AKI incidence and cOR with varied Age, Sex and BMI in derivation cohort and external network validation cohort. age1:18–35, age2:36–50, age3:51–65, age4:66–80, age5:>80; BMI: body mass index, bmi1:<18.5, bmi2:18.5–24.9, bmi3:25–29.9, bmi4:>29.9; sex1: male, sex2: female. * with statistical significance $p < 0.05$.

the external validation dataset. To further verify the portability of MINDMerge, external geographical validation was performed on datasets of different sizes (eICU-3:1630, eICU-4:1301, MIMIC-III:15298). [Supplementary Table S12](#), [Table S13](#) and [Fig. S14](#) showed that MINDMerge has superior AUC (0.842, 0.810, 0.730 respectively) compared to other causal and machine learning models. [Supplementary Table S14](#), [Fig. S15](#), [Fig. S16](#) and [Fig. S17](#) present additional performance indicators such as Precision, Recall and AUPRC as well as decision curves and calibration curves.

4. Discussion

Causal learning is crucial in healthcare for unveiling the underlying relationships and mechanisms that drive medical conditions and treatment outcomes. However, singular causal models may not fully capture the complexity of relevant factors. The integration of multiple models through BN fusion offers a more comprehensive approach to causal inference, which is essential for informed and effective decision-making. Despite the efficiency of existing BN fusion methods (e.g., Puerta's [23]

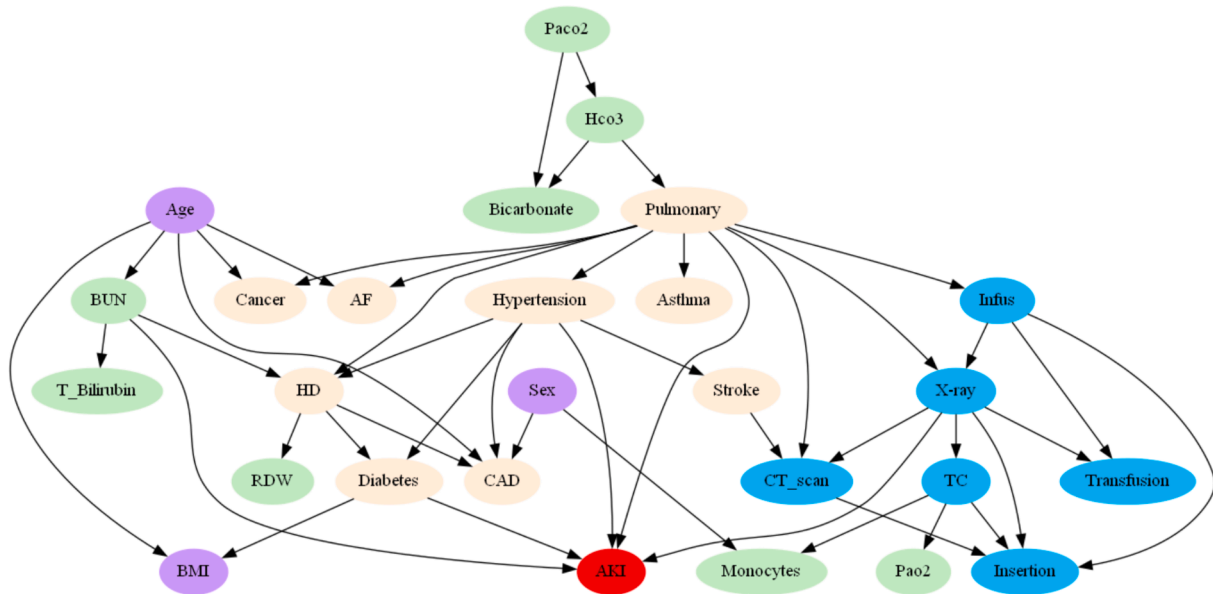


Fig. 4. The causal structure of AKI using MINDMerge ensemble learning. Purple nodes refer to demographic information and vital signs, red node refer to AKI onset (No/Yes), cyan, yellow, blue nodes refer to laboratory tests, comorbidities and procedures, respectively. BMI: body mass index, CAD: coronary artery disease, BUN: Blood Urea Nitrogen, AF: Atrial Fibrillation, HD: Heart Disease, TC: Tissue culture, RDW: Red cell Distribution Width, T_Bilirubin: Total Bilirubin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

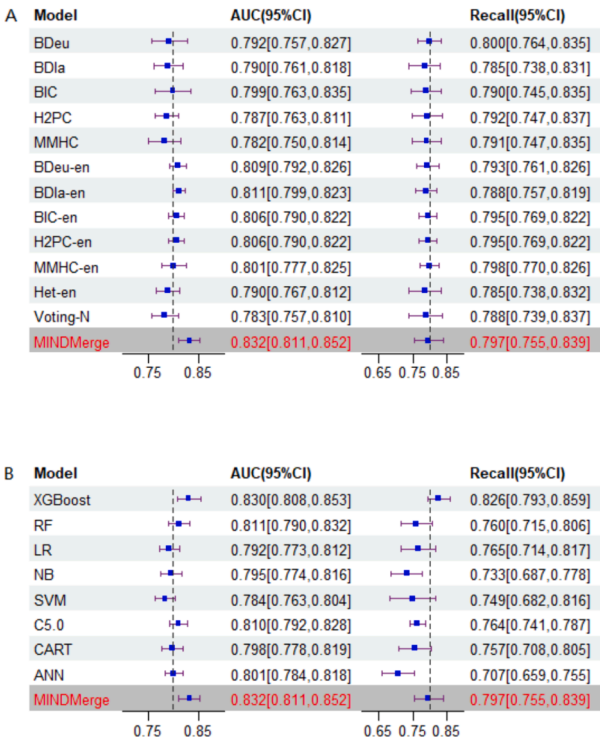


Fig. 5. Forest plot of the prediction performance of different models.

and others [24–28]), they face challenges such as algorithm diversity, local optima, and cyclic structures. To mitigate these issues, we developed MINDMerge, a novel causal network fusion framework that harmonizes knowledge across algorithms. MINDMerge integrates learning results from different causal discovery algorithms, facilitating the construction of more accurate network structures across varying sample sizes. Experimental results on both synthetic and real-world data demonstrate that MINDMerge outperforms other causal algorithms and

effectively captures true causal relationships between variables. It also has significant advantages over conventional machine learning by revealing the causality underlying complex diseases and facilitating new risk factor discoveries.

Although knowledge mining methods are adept at revealing factors with high predictive power, their identification does not necessarily imply causal influence [36]. In our work, five critical risk factors have been identified that are directly related to AKI, namely pulmonary disease, hypertension, diabetes, X-ray, and BUN. Significant interactions between the kidneys and lungs exist in both physiological and pathological conditions. Clinical observations have revealed kidney involvement in chronic respiratory disease. Studies have reported that the incidence of renal failure is higher in patients with concomitant pulmonary disease, especially in chronic obstructive pulmonary disease (COPD) [59]. Additionally, pulmonary infections, inflammatory responses, and certain treatment modalities such as mechanical ventilation can also contribute to AKI [60]. Lun et al [61] and Cai et al [55] confirmed hypertension as an independent risk factor of AKI through meta-analysis. Persistent hypertension can increase glomerular capsule pressure, leading to glomerular fibrosis and renal arteriosclerosis, which can cause ischemia of the renal parenchyma and renal failure [62]. Similarly, patients with diabetes are more likely to develop AKI than those without diabetes, demonstrated in multiple studies that diabetes was an independent risk factor of AKI [55,63,64]. In the diabetic states, glucose metabolism mainly takes place in the kidneys, which increases the glycemic load on the kidneys that may cause kidney damage. Elevated BUN levels can signal early AKI, preceding increases in serum creatinine, offering an initial alert for renal impairment [65]. Therefore, monitoring BUN aids in timely AKI detection and intervention, potentially enhancing patient outcomes. However, BUN has limitations as a diagnostic tool due to influences from other factors, leading to potential false positives or negatives. Thus, integrating additional biomarkers can refine and expedite AKI diagnosis [66]. Moreover, the use of X-ray [67] and CT imaging offers unique insights into intrarenal hemodynamics and function [68]. These imaging modalities, often employed for assessing abdominal region issues and lung infections, provide crucial diagnostic information potentially associated with kidney damage.

While machine learning models like Extreme Gradient Boosting

(XGB) excel in outcome prediction, they lack inherent insights into causal mechanisms. Our BN fusion approach, in contrast, uncovers causal relationships between variables, enabling the identification of modifiable risk factors and potential intervention targets. This provides invaluable clinical interpretability and enhances the efficacy of clinical interventions by tailoring them to specific patient populations. The risk factors identified in our study can provide actionable strategies for the prevention and management of AKI. Recognizing these risk factors allows clinicians to identify high-risk patients, monitor them more closely, and implement preventive strategies earlier. For instance, patients with pulmonary disease, hypertension, or diabetes could benefit from more rigorous control of their underlying conditions to potentially reduce the risk of AKI. Additionally, early detection and prompt intervention are critical for patients exposed to X-ray or with elevated BUN levels to prevent further kidney damage. Finally, educating patients about the signs and symptoms of AKI empowers them to seek medical attention promptly.

Fig. 4 illustrates BUN’s direct and indirect effect on AKI with the chain “Age → BUN→AKI”. AKI risk increases with age [37], Urea is the metabolic product of protein decomposition in human body, the increase of BUN levels can be caused by various factors, including age. Generally, BUN levels tend to increase with age, which is often attributed to a gradual decline in kidney function with age, leading to abnormal renal function [29], subsequently elevating the risk of AKI. The causal chain “PaCO2 → HCO3 → Pulmonary Disease → AKI” involves PaCO2 and Hco3, which are mainly used to assess the acid-base balance in the body and respiratory function, and they are closely associated with pulmonary disease. Avoiding acid-base disturbances may be beneficial for the management of lung disease and kidney function [69], which may help reduce the risk of AKI. We also observed that pulmonary disease may cause multiple complications that may indirectly or directly contribute to AKI. Studies have indicated a higher prevalence of comorbidities such as diabetes (12.2 % vs.4.6 %), heart disease (15.0 % vs.7.7 %) and hypertension (38.8 % vs. 22.8 %) in individuals with abnormal lung conditions [70]. Participants with restrictive or obstructive lung function had 1.49 and 1.42-fold higher atrial fibrillation risks, respectively [71]. Conditions like asthma, characterized by impaired lung function [72], have emerged as potential biomarkers for lung cancer development [73]. Therefore, good cardio-pulmonary function stands as a cornerstone of human health, supporting the normal functioning of all bodily organs and, in some instances, reducing the risk of AKI.

Limitations of the study exist. Firstly, to reduce data complexity, mitigate influence of outliers, and enhance model interpretability, we discretized the continuous variables. Employing different data imputation methods and discretization methods for continuous data might potentially impact the obtained results. Secondly, setting the threshold for the causal discovery phase is difficult; an inappropriate threshold can significantly impact the accuracy of structure learning. Achieving a balance between network complexity and accuracy required parameter selection based on experiential knowledge. Thirdly, our study was based on cross-sectional data, whereas in realistic healthcare setting, data collection is time-series in nature. Incorporating time-series data necessitates alternative causal analysis methods, such as Granger causality. Fourthly, our BN fusion approach, aimed at enhancing model accuracy and robustness, lacks a direct comparison with existing BN fusion methods due to the varied assumptions and applicability of different techniques. Lastly, the limited scope of clinical data variables, particularly the lack of medication factors, represents a significant limitation of this study. Medications, including nephrotoxic agents, play a crucial role in the development of AKI and their omission could influence the results of the model, potentially underestimating the true complexity of the causal pathways involved.

Nonetheless, this study holds significant implications. Firstly, the developed causal network harmonization framework presents a novel approach to structural learning, offering potential adaptation in

knowledge fusion across diverse datasets like in a federated learning setting. Secondly, MINDMerge has stronger generalizability and can maintain higher accuracy for different datasets regardless of the sample size compared to individual or ensemble causal discovery algorithms. Thirdly, distinct from conventional machine learning algorithms, MINDMerge not only achieves high performance but also provides inference and interpretation.

5. Conclusion

This study introduces a novel causal network harmonization framework that effectively integrates causal graphs from different causal discovery algorithms. To address the cyclic structure problem in causal network fusion, we developed a novel cycle breaking algorithm. Experimental results showed that our model significantly outperformed baseline models on both synthetic data and real-world data. The model has the potential to guide the development of personalized treatment strategies.

6. Summary Table

Problem or Issue	What is Already Known	What this Paper Adds
The diversity of causal discovery algorithms and the presence of cyclic structures in synthesized causal networks.	No single causal method exhibits significantly superior performance in structure reconstruction. The existence of cyclic structures is often disregarded when synthesizing network knowledge from different sources, which may result in erroneous inferences.	We proposed a novel network harmonization framework that effectively integrates causal graphs from different causal algorithms. To address the cyclic structure in network fusion, we developed a novel cycle-breaking algorithm. Experimental results showed that our model significantly outperformed baseline models on both synthetic data and real-world data. The model has the potential to guide the development of personalized treatment strategies.

CRedit authorship contribution statement

Mingyang Zhang: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Xiangzhou Zhang:** Writing – review & editing, Validation, Supervision, Resources. **Mingyang Dai:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Lijuan Wu:** Writing – review & editing. **Kang Liu:** Writing – review & editing. **Hongnian Wang:** Resources, Data curation. **WeiQi Chen:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition. **Mei Liu:** Writing – review & editing, Validation, Project administration, Investigation, Funding acquisition. **Yong Hu:** Writing – review & editing, Validation, Project administration, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Major Research Plan of the National Natural Science Foundation of China (Key Program, Grant No. 91746204), the Science and Technology Development in Guangdong Province (Major Projects of Advanced and Key Techniques Innovation, Grant No. 2017B030308008), Guangdong Engineering Technology Research Center for Big Data Precision Healthcare (Grant No. 603141789047), the National Natural Science Foundation of China (Grant No. 72371116). WC is supported by the Guangzhou Science and Technology Plan Project (Grant No. 2023A04J0360). ML is supported by the NIH/NIDDK under award number R01DK116986 and NSF Smart and Connected Health award 2014554.

Appendix A. . Supplementary material

Code for the main experiments is presented in <https://github.com/zmy-web/CD-code>.

Supplementary material is available at the end of this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmedinf.2024.105588>.

References

- J.G. Richens, C.M. Lee, S. Johri, Improving the accuracy of medical diagnosis with causal machine learning, *Nature Communications*. 11 (2020) 3923.
- J. Wang, J. Zhou, X. Chen, Bayesian Causal Network for Discrete Variables. Data-Driven Fault Detection and Reasoning for Industrial Monitoring, Springer Singapore, Singapore, 2022, pp. 233–249.
- K.X. Yu, Z.H. Cui, X. Sui, X. Qiu, J.F. Zhang, Biological Network Inference With GRASP: A Bayesian Network Structure Learning Method Using Adaptive Sequential Monte Carlo, *Frontiers in Genetics*. (2021) 12.
- E.Z. Liu, J. Li, G.H. Kinnebrew, P.Y. Zhang, Y. Zhang, L.J. Cheng, et al., A Fast and Furious Bayesian Network and Its Application of Identifying Colon Cancer to Liver Metastasis Gene Regulatory Networks, *Ieee-Acm Transactions on Computational Biology and Bioinformatics*. 18 (2021) 1325–1335.
- Cheng J, Bell DA, Liu W. Learning belief networks from data: An information theory based approach. *Proceedings of the sixth international conference on Information and knowledge management* 1997. p. 325–31.
- M. Tsagris, Bayesian Network Learning with the PC Algorithm: An Improved and Correct Variation, *Applied Artificial Intelligence*. 33 (2019) 101–123.
- S. Behjati, H. Beigy, Improved K2 algorithm for Bayesian network structure learning, *Engineering Applications of Artificial Intelligence*. 91 (2020).
- I. Tsamardinos, L.E. Brown, C.F. Aliferis, The max-min hill-climbing Bayesian network structure learning algorithm, *Machine Learning*. 65 (2006) 31–78.
- M. Gasse, A. Aussem, H. Elghazel, A hybrid algorithm for Bayesian network structure learning with application to multi-label learning, *Expert Systems with Applications*. 41 (2014) 6755–6772.
- W. Song, L. Qiu, J. Qing, W. Zhi, Z. Zha, X. Hu, et al., Using Bayesian network model with MMHC algorithm to detect risk factors for stroke, *Mathematical Biosciences and Engineering*. 19 (2022) 13660–13674.
- M. Scutari, C.E. Graafland, J.M. Gutiérrez, Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms, *International Journal of Approximate Reasoning*. 115 (2019) 235–253.
- S. Hussung, S. Mahmud, A. Sampath, M. Wu, P. Guo, J. Wang, Evaluation of data-driven causality discovery approaches among dominant climate modes, *UMBC Faculty Collection*. (2019).
- T. Liu, G. Duan, Task allocation optimization model in mechanical product development based on Bayesian network and ant colony algorithm, *Journal of Supercomputing*. 77 (2021) 13963–13991.
- K. Asghari, M. Masdari, F.S. Gharehchopogh, R. Saneifard, A fixed structure learning automata-based optimization algorithm for structure learning of Bayesian networks, *Expert Systems*. 38 (2021).
- S. Gheisari, M.R. Meybodi, BNC-PSO: structure learning of Bayesian networks by Particle Swarm Optimization, *Information Sciences*. 348 (2016) 272–289.
- A.R. Khanteymoori, M.H. Olyae, O. Abbaszadeh, M. Valian, A novel method for Bayesian networks structure learning based on Breeding Swarm algorithm, *Soft Computing*. 22 (2018) 3049–3060.
- J.Y. Wang, S.Y. Liu, A novel discrete particle swarm optimization algorithm for solving bayesian network structures learning problem, *International Journal of Computer Mathematics*. 96 (2019) 2423–2440.
- Y. Xiao, D.Y. Wang, Y. Gao, A Mobile Bayesian Network Structure Learning Method Using Genetic Incremental K2 Algorithm and Random Attribute Order Technology, *Scientific Programming*. 2021 (2021).
- L.W. Zhang, Predictive Analysis of Machine Learning Error Classification Based on Bayesian Network, *Wireless Personal Communications*. 127 (2022) 615–634.
- M. Li, K. Liu, Causality-based attribute weighting via information flow and genetic algorithm for naive Bayes classifier, *IEEE Access*. 7 (2019) 150630–150641.
- M. Li, R. Zhang, M. Hong, C. Bai, Improved structure learning algorithm of Bayesian network based on information flow, *Systems Engineering & Electronics*. 40 (2018) 1385–1390.
- X. Wang, H. Ren, X. Guo, A novel discrete firefly algorithm for Bayesian network structure learning, *Knowledge-Based Systems*. 242 (2022).
- J.M. Puerta, J.A. Aledo, J.A. Gámez, J.D. Laborda, Efficient and accurate structural fusion of Bayesian networks, *Information Fusion*. 66 (2021) 155–169.
- S. Wang, B. Qin, Bayesian Network Structure Learning by Ensemble Learning and Feedback Strategy, *Chinese Journal of Computers*. 44 (2021) 1051–1063.
- M. Sinha, P. Tadepalli, S.A. Ramsey, Voting-based integration algorithm improves causal network learning from interventional and observational data: An application to cell signaling network inference, *Plos One*. 16 (2021).
- M. Li, R. Zhang, K. Liu, A new ensemble learning algorithm combined with causal analysis for bayesian network structural learning, *Symmetry*. 12 (2020) 2054.
- Q. Cai, X. Chen, Bayesian Network Structure Merging Algorithm Based on Scoring Function, *Computer Engineering and Application*. 55 (2019) 147–152.
- Y. Tang, J. Wang, N. Mai, I. Altintas, PEnBayes: A Multi-Layered Ensemble Approach for Learning Bayesian Network Structure from Big Data, *Sensors*. (2019) 19.
- J.A. Kellum, N. Lameire, P. Aspelin, R.S. Barsoum, E.A. Burdman, S.L. Goldstein, et al., Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. (2012).
- X. Song, A.S.L. Yu, J.A. Kellum, L.R. Waitman, M.E. Matheny, S.Q. Simpson, et al., Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction, *Nature Communications*. 11 (2020).
- F.P. Wilson, M. Martin, Y. Yamamoto, C. Partridge, E. Moreira, T. Arora, et al., Electronic health record alerts for acute kidney injury: multicenter, randomized clinical trial. *Bmj-British Medical Journal*. (2021) 372.
- J.G. Raimann, M.C. Riella, N.W. Levin, International Society of Nephrology's Oby25 initiative (zero preventable deaths from acute kidney injury by 2025): focus on diagnosis of acute kidney injury in low-income countries, *Clin Kidney J*. 11 (2018) 12–19.
- L. Yang, G. Xing, L. Wang, Y. Wu, S. Li, G. Xu, et al., Acute kidney injury in China: a cross-sectional survey, *Lancet*. 386 (2015) 1465–1471.
- Y.C. Cheng, R. Luo, K. Wang, M. Zhang, Z.X. Wang, L. Dong, et al., Kidney disease is associated with in-hospital death of patients with COVID-19, *Kidney International*. 97 (2020) 829–838.
- M. Adamczak, S. Surma, A. Wiecek, Acute kidney injury in patients with COVID-19: Epidemiology, pathogenesis and treatment, *Advances in Clinical and Experimental Medicine*. 31 (2022) 317–326.
- L. Wu, Y. Hu, B. Yuan, X. Zhang, W. Chen, K. Liu, et al., Which risk predictors are more likely to indicate severe AKI in hospitalized patients? *Int J Med Inform*. 143 (2020) 104270.
- L. Wu, Y. Hu, X. Zhang, J. Zhang, M. Liu, Development of a knowledge mining approach to uncover heterogeneous risk predictors of acute kidney injury across age groups, *Int J Med Inform*. 158 (2021) 104661.
- K. Liu, B. Yuan, X. Zhang, W. Chen, L.P. Patel, Y. Hu, et al., Characterizing the temporal changes in association between modifiable risk factors and acute kidney injury with multi-view analysis, *Int J Med Inform*. 163 (2022) 104785.
- Kevin B. Korb AEN. *Bayesian Artificial Intelligence*. 2010: 2nd edition, Section ... CRC Press.
- D.J. Spiegelhalter, A.P. Dawid, S.L. Lauritzen, R.G. Cowell, Bayesian Analysis in Expert Systems, *Statistical Science*. 8 (2019–47) (1993) 29.
- E.H. Shortliffe, I.A. Beinlich, H.J. Suermondt, R.M. Chavez, G.F. Cooper, The ALARM Monitoring System: A Case Study with two Probabilistic Inference Techniques for Belief Networks, Springer, Berlin Heidelberg, 1989.
- T.J. Pollard, A.E.W. Johnson, J.D. Raffa, L.A. Celi, R.G. Mark, O. Badawi, The eICU Collaborative Research Database, a freely available multi-center database for critical care research, *Sci Data*. 5 (2018) 180178.
- J. Li, X.S. Yan, D. Chaudhary, V. Avula, S. Mudiganti, H. Husby, et al., Imputation of missing values for electronic health record laboratory data, *Npj Digital Medicine*. 4 (2021) 147.
- G. Schwarz, Estimating the Dimension of a Model, *The Annals of Statistics*. 6 (1978) 461–4.
- H.B. Ajmal, M.G. Madden, Dynamic Bayesian Network Learning to Infer Sparse Models From Time Series Gene Expression Data, *Ieee-Acm Transactions on Computational Biology and Bioinformatics*. 19 (2022) 2794–2805.
- D. Heckerman, D. Geiger, D.M. Chickering, Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*. 20 (1995) 197–243.
- L. Azzimonti, G. Corani, M. Scutari, A Bayesian hierarchical score for structure learning from related data sets, *International Journal of Approximate Reasoning*. 142 (2022) 248–265.
- A. Cano, M. Gómez-Olmedo, A.R. Masegosa, S. Moral, Locally averaged Bayesian Dirichlet metrics for learning the structure and the parameters of Bayesian networks, *International Journal of Approximate Reasoning*. 54 (2013) 526–540.
- J. Rusek, K. Firek, L. Slowik, Extracting structure of Bayesian network from data in predicting the damage of prefabricated reinforced concrete buildings in mining areas, *Eksploatacja i Niezawodność-Maintenance and Reliability*. 22 (2020) 658–666.
- Gasse M, Aussem A, Elghazel H. An experimental comparison of hybrid algorithms for Bayesian network structure learning. *Machine Learning and Knowledge*

- Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012 Proceedings, Part I 23: Springer; 2012. p. 58-73.
- [51] X. Zheng, B. Aragam, P. Ravikumar, Xing EP, Continuous Optimization for Structure Learning. *Adv Neur In, DAGs with NO TEARS*, 2018, p. 31.
 - [52] H. Wang, J. Li, G. Zhu, A Data Feature Extraction Method Based on the NOTEARS Causal Inference Algorithm, *Applied Sciences*. 13 (2023) 8438.
 - [53] S. Gheisari, M.R. Meybodi, A new reasoning and learning model for Cognitive Wireless Sensor Networks based on Bayesian networks and learning automata cooperation, *Computer Networks*. 124 (2017) 11–26.
 - [54] K. Yan, W. Fang, H. Lu, X. Zhang, J. Sun, X. Wu, Mutual Information-Guided GA for Bayesian Network Structure Learning, *IEEE Transactions on Knowledge and Data Engineering*. 1–16 (2022).
 - [55] X.Y. Cai, G.M. Wu, J. Zhang, L.C. Yang, Risk Factors for Acute Kidney Injury in Adult Patients With COVID-19: A Systematic Review and Meta-Analysis, *Frontiers in Medicine*. 8 (2021).
 - [56] Y.F. Liu, Z. Zhang, X.L. Pan, G.L. Xing, Y. Zhang, Z.S. Liu, et al., The chronic kidney disease and acute kidney injury involvement in COVID-19 pandemic: A systematic review and meta-analysis, *Plos One*. 16 (2021).
 - [57] C. Tantithamthavorn, S. McIntosh, A.E. Hassan, K. Matsumoto, The impact of automated parameter optimization on defect prediction models, *IEEE Transactions on Software Engineering*. 45 (2018) 683–711.
 - [58] A.O. Balogun, S. Basri, S. Mahamad, S.J. Abdulkadir, L.F. Capretz, A.A. Imam, et al., Empirical analysis of rank aggregation-based multi-filter feature selection methods in software defect prediction, *Electronics*. 10 (2021) 179.
 - [59] C. Sorino, N. Scichilone, C. Pedone, S. Negri, D. Visca, A. Spanevello, When kidneys and lungs suffer together, *J Nephrol*. 32 (2019) 699–707.
 - [60] J. Alge, K. Dolan, J. Angelo, S. Thadani, M. Virk, A.A. Acan, Two to Tango: Kidney-Lung Interaction in Acute Kidney Injury and Acute Respiratory Distress Syndrome, *Front Pediatr*. 9 (2021) 744110.
 - [61] Z. Lun, Z. Mai, L. Liu, G. Chen, H. Li, M. Ying, et al., Hypertension as a Risk Factor for Contrast-Associated Acute Kidney Injury: A Meta-Analysis Including 2,830,338 Patients, *Kidney Blood Press Res*. 46 (2021) 670–692.
 - [62] S. Nie, Z. Feng, L. Tang, X. Wang, Y. He, J. Fang, et al., Risk Factor Analysis for AKI Including Laboratory Indicators: a Nationwide Multicenter Study of Hospitalized Patients, *Kidney Blood Press Res*. 42 (2017) 761–773.
 - [63] Y.J. Wang, J.H. Li, Y. Guan, Q.H. Xie, C.M. Hao, Z.X. Wang, Diabetes mellitus is a risk factor of acute kidney injury in liver transplantation patients, *Hepatobiliary & Pancreatic Diseases International*. 20 (2021) 215–221.
 - [64] X. Shen, K.M. Lv, B.C. Hou, Q.G. Ao, J.H. Zhao, G. Yang, et al., Impact of Diabetes on the Recurrence and Prognosis of Acute Kidney Injury in Older Male Patients: A 10-Year Retrospective Cohort Study, *Diabetes Therapy*. 13 (2022) 1907–1920.
 - [65] S. Zachariah, K. Kumar, S.W.H. Lee, W.Y. Choon, S. Naeem, C. Leong, Chapter 7 - Interpretation of Laboratory Data and General Physical Examination by Pharmacists, in: D. Thomas (Ed.), *Clinical Pharmacy Education*, Elsevier, Practice and Research, 2019, pp. 91–108.
 - [66] Edelstein CL. Chapter Six - Biomarkers in Acute Kidney Injury. In: Edelstein CL, editor. *Biomarkers of Kidney Disease (Second Edition)*: Academic Press; 2017. p. 241-315.
 - [67] H. Boyacioglu, B.B. Yola, C. Karaman, O. Karaman, N. Atar, M.L. Yola, A novel electrochemical kidney injury molecule-1 (KIM-1) immunosensor based covalent organic frameworks-gold nanoparticles composite and porous NiCo2S4@CeO2 microspheres: The monitoring of acute kidney injury, *Applied Surface Science*. 578 (2022).
 - [68] L.O. Lerman, M. Rodriguez-Porcel, J.C. Romero, The development of x-ray imaging to study renal function, *Kidney International*. 55 (1999) 400–416.
 - [69] K. Yagi, T. Fujii, Management of acute metabolic acidosis in the ICU: sodium bicarbonate and renal replacement therapy, *Critical Care*. 25 (2021).
 - [70] D.M. Mannino, M.A. McBurnie, W. Tan, A. Kocabas, J. Anto, W.M. Vollmer, et al., Restricted spirometry in the Burden of Lung Disease Study, *International Journal of Tuberculosis and Lung Disease*. 16 (2012) 1405–1411.
 - [71] S.N. Lee, S.H. Ko, S.H. Her, K. Han, D. Moon, S.K. Kim, et al., Association between lung function and the risk of atrial fibrillation in a nationwide population cohort study, *Scientific Reports*. 12 (2022).
 - [72] M.A. Kim, S.W. Shin, J.S. Park, S.T. Uh, H.S. Chang, D.J. Bae, et al., Clinical Characteristics of Exacerbation-Prone Adult Asthmatics Identified by Cluster Analysis, *Allergy Asthma & Immunology Research*. 9 (2017) 483–490.
 - [73] H.W. Lee, H.J. Lee, J.K. Lee, T.Y. Park, E.Y. Heo, D.K. Kim, Rapid FEV1 Decline and Lung Cancer Incidence in South Korea, *Chest*. 162 (2022) 466–474.