

ORIGINAL RESEARCH ARTICLE

Temporal correlation rough set-based three-way clustering model for chronic kidney disease diagnosis

Yan Zhang^{1*}, Wanzhen Long², Ruihui Chen³, Zihui Lin³, Haole Huang³, Lincong Li³, and Xiaoli Chu^{4,5*}

¹Intelligent Computing and Big Data Technology Research Center, Economics and Management Experimental Teaching Center, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

²Department of Computer Application, School of Big Data and Artificial Intelligence, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

³Department of Data Science, School of Statistics and Data Science, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

⁴State Key Laboratory of Traditional Chinese Medicine Syndrome, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China

⁵Department of Traditional Chinese Medicine Big Data Research, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China

*Corresponding authors:

Yan Zhang
(zhangyan@gdufe.edu.cn);
Xiaoli Chu
(xlchu85@126.com)

Citation: Zhang Y, Long W, Chen R, *et al.* Temporal correlation rough set-based three-way clustering model for chronic kidney disease diagnosis. *Artif Intell Health*. 2026;3(3):025470107.
doi: 10.36922/AIH025470107

Received: November 23, 2025

Revised: January 19, 2026

Accepted: January 28, 2026

Published online: April 3, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Chronic non-communicable diseases (CNCDs) are the leading cause of morbidity and mortality worldwide. Currently, diagnostic information for CNCDs remains incomplete and is subject to ongoing change, leading to uncertainty in diagnostic conclusions and severely impacting disease control. Therefore, it is necessary to study decision-making theory and methods for assisting clinical diagnosis. Given the characteristics of clinical diagnostic decision problems, this paper discusses the temporally correlated rough set theory and its three-way clustering model. Firstly, the temporal correlation rough set model based on the gradient and cosine similarity is defined. Then, the clinical diagnostic decision process for CNCDs is transformed into a three-way clustering problem with temporal correlation attributes. A three-way clustering model based on a temporal correlation rough set is constructed to support clinical diagnosis decisions for CNCDs. In this model, the common features of temporal association attributes are identified using a positive domain-based three-way clustering algorithm. Finally, the validity and applicability of the theoretical model are verified using real clinical data from 80,139 visit time points and 83 laboratory examination indicators from 3,094 chronic kidney disease (CKD) patients. Based on the algorithm's calculations and a retrospective literature review, β_2 -microglobulin is identified as a candidate marker for CKD that warrants further validation and could provide auxiliary decision support for clinical practice. The main contribution of this paper is twofold. One is to make a new theoretical contribution to the data-driven research paradigm of clinical diagnosis decision-making. Another is to provide quantitative decision support for clinical diagnosis.

Keywords: Rough sets; Three-way clustering; Clinical decision-making; Temporal correlation features

1. Introduction

Chronic non-communicable diseases (CNCDs) are persistent health-related diseases that impair quality of life and are a major cause of morbidity and mortality worldwide.^{1,2} The accuracy of clinical diagnosis decisions is crucial for CNCD patients to receive effective and timely treatment, improving their quality of life and prolonging survival. The clinical diagnosis of CNCDs is a dynamic decision-making process that integrates patient symptoms with progressively accumulating diagnostic indicators to arrive at the most accurate possible conclusion. In principle, clinical assessment should integrate physiological indicators (laboratory tests and imaging), symptoms, medical history, and other clinical diagnostic indicators, and, through comprehensive analyses, provide the disease name, disease progression, and other diagnostic conclusions. However, given a variety of complex factors, such as clinical load, testing technology, out-of-pocket costs, and medical insurance burdens, clinicians aim to provide diagnostic conclusions quickly with as few primary indicators as possible.

Chronic kidney disease (CKD) is one of the most common CNCDs.³ According to a previous study, 11–13% of the world population suffers from CKD.⁴ Because the etiology of CKD is unknown, the glomerular filtration rate is a key indicator of CKD progression. However, as the spectrum of human diseases continues to change, chronic disease comorbidities are becoming increasingly diverse. Retrospective studies have shown the limitations of a single laboratory test⁵ and glomerular filtration rate^{6,7} in diagnosing CKD. Starting with the most comprehensive laboratory test data, exploring the disease's key influencing factors, and identifying the key markers for clinical diagnosis are urgent needs for clinical decision-making.^{8,9}

In our study, a class of clinical diagnosis decision problems with temporal correlation characteristics was extracted from the dataset of 80,139 visit time points and 83 laboratory examination indicators of 3,094 patients with CKD, and the common characteristics of correlated objects driven by massive attributes were studied to identify primary laboratory indicators (or indicator sets) for clinical diagnosis of CKD. Our study helps establish a data-driven clinical diagnostic decision-making paradigm and provides auxiliary decision support for clinical diagnosis by analyzing large-scale clinical data.

The clinical diagnostic process generates a large number of attributes and data (each CKD evaluation covers approximately 80 indicators) that require special theoretical models or techniques for analysis and meaningful interpretation. Attempts to employ data discovery and machine learning models in order to unearth

complex data characteristics and relationships have a long track record in clinical decision-making.^{10–13} Additionally, in parallel with the advances of Pawlak's Rough Sets (RS) theory¹⁴, numerous studies emerged to address problems of using RS in clinical decision-making applications.^{15–17} RS offers a powerful mathematical framework to acquire information about data that cannot be categorized with traditional set theory.¹⁸ The basic idea of classical Pawlak RS¹⁴ is to describe imprecise or uncertain concepts by using the definable set obtained from the equivalence relation, so as to give the upper and lower approximate sets of the target concept, and to describe the uncertainty of the target concept through the boundary field between the upper and lower approximate sets. RS properties and ability to produce strong decision rules from incomplete and noisy data¹⁹ deliver flexibility and interpretability, which are lacking in black-box models based on machine learning methods. RS is widely used in numerous fields, such as credit evaluation²⁰, industrial processes²¹, design²², financial forecasting²³, and environmental forecasting.^{24,25}

Using RS to address temporal correlation in feature information without artificial labels is an important problem in the study of RS extended models.^{26–28} For temporal attribute information, Yao²⁶ demonstrated an RS data analysis model for the discovery of decision rules from time series data. Podsiadlo and Rybinski²⁷ investigated the feasibility of RS for building profitable trend-prediction models for financial time series. Temporal information is essentially data whose attribute values and attributes are constantly updated over time. Dong *et al.*²⁹ proposed a relative discernibility relation to represent the samples and features of fuzzy RS in a unified manner. Ni *et al.*³⁰ devised incremental mechanisms of positive region and dependency function using concepts from fuzzy RS on incremental datasets. For attribute correlation, a common approach is to count the co-occurrence frequency of attributes in the output space. To address the lack of correlation attributes classification methods, Che *et al.*³¹ demonstrated the RS attribute reduction algorithm for aggregation-related local attribute correlations to obtain global label correlation. Lee *et al.*³² described a relative value trading system based on a correlation and RS analysis for the foreign exchange futures market.

However, existing research on the temporal correlation RS (TCRS) model primarily focuses on its application to processing the temporal correlation hybrid data, such as attribute reduction and feature selection. Theoretical research on the TCRS model, such as the temporal correlation hybrid binary relation, requires further in-depth study, given the prevalence of practical temporal correlation decision-making problems.

Furthermore, to extract the common features of the objects driven by interrelated massive attributes—based on the construction of the TCRS model to achieve the pre-classification of data without artificial labels—the clustering thinking is introduced, so that the objects with the same attribute characteristics are aggregated in a cluster, and the primary correlation features are found in this cluster. Traditional clustering methods group individuals (outliers) who deviate from the overall pattern. Individuals (edge points) may exist and hover between two or more classes. In addition, traditional clustering methods are ineffective when dealing with clusters with the nearest regions.³³ Both of these conditions can lead to unstable and irrational clustering results. Three-way clustering (TWC) is a fundamental problem in three-way decision theory³⁴, which describes the three-way relationship between an object and a cluster, whether it may/may not/may or may not belong to the cluster. Sun *et al.*³⁵ extended density peaks clustering to TWC nearest neighbors to deal with higher density samples. There are also studies that combine evidence theory with TWC to develop a TWC expansion model that addresses specific data characteristics.³⁶ Ali *et al.*³⁷ defined TWC to improve the label allocation strategy of the nearest sample to distinguish the intension and extension of a cluster to improve the clustering accuracy. However, theoretical research on TWC needs further development, particularly the meaningful construction of two sets and three regions.³⁷

The current paper summarizes a class of clinical diagnosis decision-making problems characterized by temporally correlated attribute information characteristics from the decision-making process and clinical data of clinical diagnosis of CKD, applies RS and TWC theory to construct a TWC model based on TCRS, and applies it to the decision-making problems of clinical diagnosis of CKD. First, a TCRS model was constructed to perform pre-classification of temporal correlation information without artificial labels. Second, the unsupervised TWC model was used to construct the two-set and three-region RS model of time series, and a TWC model based on a positive domain was designed. Finally, using the unmarked dataset of 3,094 CKD patients (including 80,139 visit records and 83 laboratory indicators), based on the proposed theoretical model, unsupervised clustering of CKD patients was realized, and the characteristics of patients with a clinical diagnosis of CKD were identified, and the initially undetectable trends in CKD diagnosis were identified. The primary correlation indexes for CKD diagnosis were determined, and the effect of clustering was evaluated using the silhouette coefficient (SC). The finding provides additional support and guidance for clinical decision-making based on the experimental results.

The rest of this paper is organized as follows: We briefly introduce the preliminaries related to our work in Section 2. The details of the TCRS-TWC are introduced in Section 3. The CKD diagnosis model based on TCRS-TWC is introduced in Section 4. Section 5 presents the experimental results on the datasets and the sensitivity analysis. Finally, we summarize our work and future research in Section 6.

2. Preliminaries

2.1. Rough sets

Rough sets are well-suited to managing vague and uncertain information. By reducing knowledge while preserving classification capability, one can derive decision or classification rules. The extension of the RS model fundamentally depends on how the approximation operator is defined.

Definition 2.1.¹⁴

Let (U, R) be an approximation space and $[u_i]_R = \{up \in U | (u_i, u_p) \in R\}$ be an equivalence class of u_i . For any $X \subseteq U$, the lower and upper approximations of X are defined as in **Equation (1)**:

$$\begin{aligned} appr_R(X) &= \{ui \in U | [ui]_R \subseteq X\} \\ \overline{appr_R}(X) &= \{u_i \in U | [u_i]_R \cap X \neq \emptyset\} \end{aligned} \quad (1)$$

Definition 2.2.¹⁴

Given a set-pair $(appr_R(X), \overline{appr_R}(X))$, the three regions of X are defined as in **Equation (2)**:

$$\begin{aligned} POS(X) &= appr_R(X) \\ BND(X) &= \overline{appr_R}(X) - appr_R(X) \\ NEG(X) &= U - \overline{appr_R}(X) \end{aligned} \quad (2)$$

2.2. Three-way clustering representation

Unlike classical clustering methods, TWC utilizes interval sets to represent clustering outcomes, which differs distinctly from the representation of $C = \{C_1, C_2, \dots, C_k\}$ with k clusters as a result.

Definition 2.3.^{38,39}

Given a universe set $U = \{x_1, x_2, \dots, x_M\}$, which has M objects and each object has K attributes, represented as $xm = (mm_1, mm_2, \dots, mm_k)$. The three-way clustering is defined as **Equation (3)**:

$$C_i = [\underline{C}_i, \overline{C}_i] = [POS(C_i), POS(C_i) \cup BND(C_i)] \quad (3)$$

where $\underline{C}_i$ is called the lower bound (i.e. the core region of a cluster) and $\overline{C}_i = POS(C_i)$. $\overline{C}_i$ denotes the largest

possible region of a cluster, which is named the upper bound, $\overline{C}_i = POS(C_i) \cup BND(C_i)$.

Let $NEG(C_i) = U - POS(C_i) \cup BND(C_i)$ indicates all objects in this region that do not belong to the cluster C_i . A cluster result is divided into three regions: $POS(C_i)$, $BND(C_i)$ and $NEG(C_i)$. $POS(C_i)$ contain points in C_i and belong to the cluster. $BND(C_i)$ contains fringe points in C_i , though it also belongs to the cluster; it is distributed in edge regions and may belong to other clusters. $NEG(C_i)$, formed by the remaining objects in $U - C_i$, does not belong to the cluster. For the family of clusters C , its TWC representation is shown in **Equation (4)**:

$$C = \{[\underline{C}_1, \overline{C}_1], [\underline{C}_2, \overline{C}_2], \dots, [\underline{C}_k, \overline{C}_k]\} \quad (4)$$

3. Three-way clustering based on temporal correlation rough sets

We constructed a TCRS model and a TWC model to classify temporal association attributes information. First, for the time-series and correlation features, change-measure functions based on two methods (gradient and cosine similarity) and the distance function based on residual were constructed. On this basis, a TCRS model was constructed. Secondly, a TWC algorithm based on a temporal association RS model was developed to classify temporal-association hybrid attribute information. The research idea of this section is shown in **Figure 1**.

3.1. Temporal rough set model based on gradient

In this part, we introduce the temporal RS. We illustrate the temporal feature decision-making problem by introducing an example from clinical diagnosis and treatment.

Example 1. The decision object x_i (i.e., the patient) is associated with the following examination indicators $A = \{a1, a2, a3, a4, a5\}$: urinary microalbumin, creatinine (Cr), urea nitrogen, carbon dioxide (CO_2), and hemoglobin (HE). Within the selected treatment period, patient x_i was examined on four occasions $T = t1, t2, t3, t4$. The corresponding clinical and treatment records for this patient are provided in **Table 1**.

Table 1. The dataset on the examination and treatment of patient

T	Urinary microalbumin (mg/L)	Creatinine (mg/g)	Urea nitrogen (mg/dL)	Carbon dioxide (mmol/L)	Hemoglobin (g/L)
t1	1,959.00	220.00	18.70	22.00	112.00
t2	2,093.50	250.00	12.96	25.30	103.00
t3	2,228.00	465.00	23.42	20.80	101.00
t4	2,160.75	577.00	22.57	25.80	141.00

Definition 3.1. Let $TS = \{(x_1, T, A, f), (x_2, T, A, f), \dots, (x_M, T, A, f)\}$ be a decision-making information system with temporal correlated hybrid attribute feature, where $U = \{x_1, x_2, \dots, x_m, \dots, x_M\}$ be a set of decision-making objects, $T = \{t_1, t_2, \dots, t_p, \dots, t_N\}$ be a decision time point set, $A = \{a_1, a_2, \dots, a_k, \dots, a_K\}$ be a set of attributes. The decision object x_i corresponds to the decision scheme s_i at each decision time point t_j . For each object $x \in U$, attribute $a \in A$, and time point $t \in T$, there is a corresponding information function $a: U \times T \rightarrow f_x(a, t)$. Let $\nabla F_x(a, t)$ be the gradient set of the decision object x on the attribute a and the time point t . For $x_i \in U$, the metric formula between temporal attributes, i.e., the gradient, is shown in **Equation (5)**:

$$\nabla F_{x_i}(a, t) = (f_a(a_k, t_j), f_t(a_k, t_j)) \cos \theta \quad (5)$$

Example 2. The gradient values of **Table 1** derived from **Equation (5)** are shown in **Table 2**.

Proposition 1. The universe: U is a non-empty finite set, R is a serial relation on U .

- (i) serial, if $\forall x \in U$, there $\exists y \in U$ such that xR_y .
- (ii) reflexive, $\forall x \in U$, there xR_x .
- (iii) symmetric, $\forall x, y \in U$, there $xR_y \Rightarrow yR_x$.

3.2. Temporal correlation rough sets based on cosine similarity and distance measurement

In this section, two RS operators with temporal correlation features, namely directional binary relation R_s and distance binary relation R_d , were constructed. An RS is one where the upper approximation is larger than the lower approximation. The lower approximation of a set $X(XU)$ uses equivalence classes containing period information.²⁶

Definition 3.2. Consider a decision-making information system with temporally correlated hybrid attributes, denoted as $TS = \{(x_1, T, A, f), (x_2, T, A, f), \dots, (x_M, T, A, f)\}$. Here, for any object $x \in U$, attribute $a \in A$, and time point $t \in T$, the corresponding information is given by the function $a: U \times T \rightarrow f_x(a, t)$. Let $\nabla F_x(a, t)$ represent the gradient set of object x regarding attribute a at time t , and let ρ denote the attribute correlation degree. The

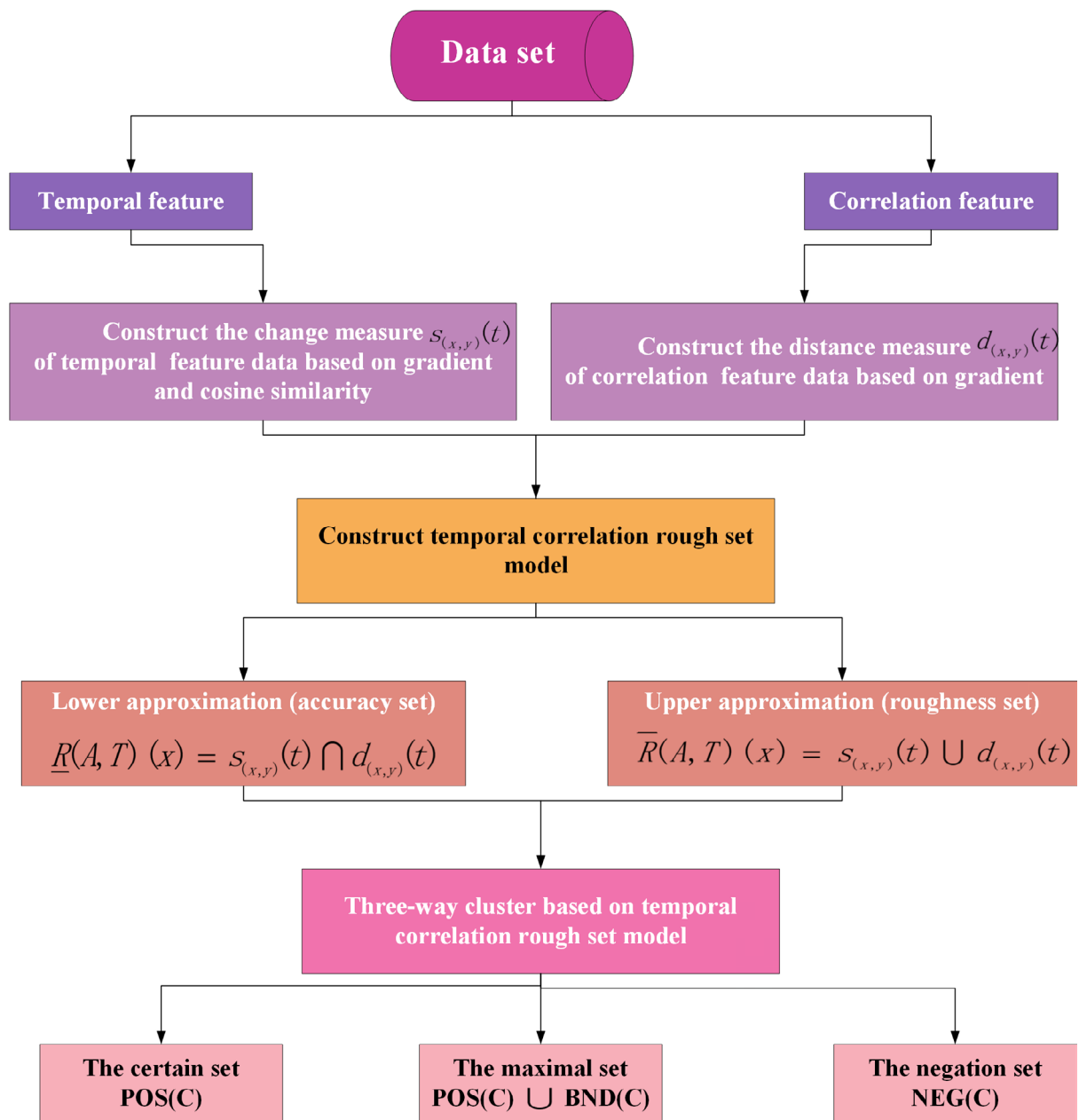


Figure 1. The flowchart for three-way clustering based on temporal correlation rough sets

Table 2. The gradient values of x_i

T	$\nabla Fx_i(a_1, t)$	$\nabla Fx_i(a_2, t)$	$\nabla Fx_i(a_3, t)$	$\nabla Fx_i(a_4, t)$	$\nabla Fx_i(a_5, t)$
t_1	134.50	30.00	-6.26	3.30	-9.00
t_2	134.50	122.50	2.36	-0.60	-5.50
t_3	33.63	163.50	4.80	0.25	19.00
t_4	-67.25	112.00	-0.85	5.00	40.00

similarity between two decision objects $x, y \in U$ is defined in Equation (6).

$$s_{(x,y)}(t) = \cos(\nabla F_x(a, t), \nabla F_y(\rho_a, t)) \quad (6)$$

Definition 3.3. Let R_d be a binary relationship based on attribute values changing over time (Equation (7)):

$$R_s = \{(y, t) \in U \times T \mid (x, y) \in S_{(x,y)}(t)\} \quad (7)$$

Definition 3.4. Within the same information system TS described in Definition 3.2, the distance $d(x, y)(t)$ between decision objects x and y ($x, y \in U$), computed from residuals, is defined in Equations (8)–(12). This definition utilizes the previously introduced gradient set $\nabla F_x(a, t)$ and the attribute correlation degree ρ .

$$f(a_k, t_N) = \frac{1}{|N|} \sum_{i \in N} f(a_k, t_i) \quad (8)$$

$$f(a_K, t_i) = \frac{1}{|K|} \sum_{i \in K} f(a_k, t_i) \quad (9)$$

$$f(a_K, t_N) = \frac{1}{|K||N|} \sum_{i \in K, i \in N} f(a_k, t_i) \quad (10)$$

$$r(a_k, t_i) = (f(a_k, t_i) - f(a_k, t_N) - f(a_K, t_i) - f(a_K, t_N)) \quad (11)$$

$$d_{(x,y)}(t) = (\sum_{k \in K, i \in N} (r_x(a_k, t_i) - \rho r_y(a_k, t_i))^2)^{\frac{1}{2}} \quad (12)$$

where $r(a_k, t_i)$ is the residual error of the domain value of the function $f_x(a, t)$. ρ is the correlation coefficient between attributes. The relationship between attributes can be divided into the following three cases (Figure 2): (i) $\rho > 0$, indicating a positive correlation between attributes; (ii) $\rho = 0$, indicating that attributes are independent; and (iii) $\rho < 0$, indicating a negative correlation between attributes. In clinical diagnostic decision-making, ρ is used to represent the same or opposite trend between two known physical examination indexes.

Definition 3.5. Let R_d be a binary relationship based on attribute distance (Equation (13)):

$$R_d = \{(x, y) \in U \times U \mid (x, y) \in d_{(x,y)}(t)\} \quad (13)$$

Definition 3.6. Let R_s and R_d be two binary relations on U , where R_s is a similarity binary relation, and R_d is a differential binary relation. $R(A, T)(x) = \{R_s, R_d\}$ be a temporal correlation binary relation on U , the

approximation operators of TCRS are shown in Equation (14):

$$\begin{aligned} \bar{R}(A, T)(x) &= R_s \cup R_d = \{(x, y) \in U \times U \mid s_{(x,y)}(t) \cup d_{(x,y)}(t)\} \\ \underline{R}(A, T)(x) &= R_s \cap R_d = \{(x, y) \in U \times U \mid s_{(x,y)}(t) \cap d_{(x,y)}(t)\} \end{aligned} \quad (14)$$

The two operators describe the direction and distance of objects x and y ($x, y \in U$) in the time series. $R(A, T)(x)$ indicates that the closer the direction and distance are, the higher the similarity between the objects. Furthermore, the two basic RS operators can correspond to risk preferring $R(A, T)(x)$ decision making and risk-averse $R(A, T)(x)$ decision making of reality. $R(A, T)(x)$ indicates the decision maker prefers either direction or distance. $R(A, T)(x)$ indicates the decision maker prefers both direction and distance.

3.3. Three-way clustering based on the temporal correlation rough sets model

From a decision-making perspective, there are two types of relationships between an object and a class cluster: the object is determined to belong to this class cluster, and the object is determined not to belong to this class cluster. This is the clustering result of two-way decision-making. This representation cannot capture the object's uncertainty during data partitioning. The TWC method introduces a new division that this object may or may not belong to this class cluster, and is more in line with the practical management and decision-making problems.

Definition 3.7. Let $s(x, y)(t)$ and $d(x, y)(t)$ be direction measure and distance measure, respectively. $R(A, T)(x) = \{R_s, R_d\}$ be a temporal correlation binary relation on U , where R_s is a directional binary relation, and R_d is a distance binary relation. The decision objects are classified by $R(A, T)(x)$ into four sets according to risk preference. Given a similarity threshold θ ($\theta \in [0, 1]$) and a distance threshold μ ($\mu \geq 0$). The family of clusters C , its TWC $C = \{POS(C), POS(C) \cup NEG(C), NEG(C)\}$, is shown in Equation (15).

$$\begin{aligned} POS(C) &= \{(x, y) \in (s_{(x,y)}(t)) \cap d_{(x,y)}(t) \leq \mu\} \\ POS(C) \cup BND(C) &= \{0 \leq |s_{(x,y)}(t)| < \theta \cap (d_{(x,y)}(t) > \mu)\} \\ NEG(C) &= \{\theta \leq |s_{(x,y)}(t)| < 1 \cap (d_{(x,y)}(t) > \mu)\} \end{aligned} \quad (15)$$

where a certain positive cluster $POS(C)$ contains a collection of objects with both similar directions and close distances.

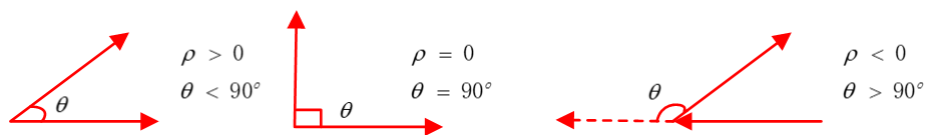


Figure 2. The relationships between attributes

The uncertain cluster $POS(C) \cup BND(C)$ contains a collection of objects with both similar directions and far distances. The certain negation set $NEG(C)$ contains other objects.

Theorem. Let the certain set $POS(C)$, the maximal set $POS(C) \cup BND(C)$ and the negation set $NEG(C)$ be TWC sets on the low approximation set $R(A, T)(x)$ of TCRS (Equation (16)).

$$\begin{aligned} POS(C) \cup POS(C) \cup BND(C) \cup NEG(C) &= R(A, T)(x) \\ POS(C) \cap (POS(C) \cap BND(C)) \cap NEG(C) &= \emptyset \end{aligned} \quad (16)$$

Proof. $POS(C) \cup POS(C) \cup BND(C) \cup NEG(C)$
 $= \{\theta \leq |s_{(x,y)}(t)| \leq 1 \cap d_{(x,y)}(t) \leq \mu\} \cup \{\theta \leq$
 $|s_{(x,y)}(t)| \leq 1 \cap d_{(x,y)}(t) \geq \mu\} \cup \{0 \leq |s_{(x,y)}(t)| \leq$
 $\theta \cap ((x, y) \in d_{(x,y)}(t))\} = \{\theta \leq |s_{(x,y)}(t)| \leq 1 \cap ((x, y)$
 $\in d_{(x,y)}(t))\} \cup \{0 \leq |s_{(x,y)}(t)| \leq \theta \cap ((x, y)$
 $\in d_{(x,y)}(t))\} = \{((x, y) \in s_{(x,y)}(t)) \cap ((x, y) \in d_{(x,y)}(t))\}$
 $= \{((x, y) \in s_{(x,y)}(t)) \cap ((x, y) \in d_{(x,y)}(t))\} = R(A, T)(x)$

4. Diagnosis model of chronic kidney disease based on temporal correlation rough set-based three-way clustering

4.1. Problem description

The clinical diagnosis of CNCDs is highly incomplete and time-series dynamic. Clinical decision-makers comprehensively analyze the characteristic indicators of abnormalities from laboratory examination results, patient symptoms, and disease history, and draw definitive conclusions about the type of disease or the extent of disease progression. To improve the quality of clinical diagnostic decision-making, common characteristic indicators that help determine disease types can be extracted from historical disease-diagnosis data, and the quantitative representation and characterization of diagnostic decision-making results can be achieved using these indicators. In addition, an accurate diagnosis can be made possible by quantitative analysis and calculation.

Chronic kidney disease is a kind of chronic disease with an unknown cause, rapid progression, and requires regular follow-up and treatment once diagnosed. CKD is also associated with various diseases, such as anemia and heart disease. The evaluation of each visit should be based on the perspectives of nutrition, metabolism, fluid and electrolyte balance, and complications, using indicators to assess disease progression and test results that cover a wide range of diseases. In particular, as complications increase, the glomerular filtration rate, a primary indicator often

used to assess CKD progression, has gradually lost clinical significance. It is an urgent need for clinical diagnosis to analyze the trends of various correlated indicators and explore the historical time-series dynamic laboratory indicators of CKD progression as comprehensively as possible.

The proposed TCRS-TWC model can support clinicians in diagnosing and monitoring CKD through several practical applications. Firstly, a decision support tool: by identifying key temporal patterns in laboratory indicators, the model can highlight patients showing abnormal progression trends, enabling early intervention. Secondly, guideline complement: while existing guidelines (e.g., Kidney Disease: Improving Global Outcomes [KDIGO]) focus on established markers, such as estimated glomerular filtration rate, our model can identify additional correlated indicators that may provide complementary information, especially in complex cases. Thirdly, workflow integration: the model could be embedded within Electronic Health Record (EHR) systems to provide real-time alerts when patients' laboratory trends deviate from expected patterns, supporting personalized monitoring plans. Fourthly, biomarker discovery: the unsupervised approach can reveal novel associations between laboratory markers, generating hypotheses for further clinical validation.

Therefore, this paper addresses how to assist clinicians in rapidly generating scientifically sound, accurate diagnostic conclusions under time- and resource-constrained conditions by integrating incomplete, temporally evolving, and interrelated laboratory data with historical diagnostic evidence, thereby improving the accuracy and efficiency of clinical decision-making.

4.2. Model construction

Based on the above description of the characteristics of CKD clinical diagnosis decision problems, the following uses the TCRS-TWC model defined in Section 3 to present a decision model of CKD clinical diagnosis decision problems with the characteristics described in Section 4.1.

Let $U = \{x_1, x_2, \dots, x_m, \dots, x_M\}$ be a set of CKD patients, $T = \{t_1, t_2, \dots, t_p, \dots, t_N\}$ be a decision time point set, $A = \{a_1, a_2, \dots, a_k, \dots, a_K\}$ be a set of laboratory test indicators for patients with CKD, such as Cr, and urinary microalbumin. Each patient x_i has a set of laboratory test indicator values $set_{f_x}(a, t)$ at the time point t_j .

Through the statistical analysis of the diagnostic data of CKD in a certain continuous period, the time series laboratory index A and the inspection index value f of the disease at different time points T can be obtained, as well as the correlation measure $R(A, T)(x)$ between the indicators. Affected by objective factors such as time and

cost, the dataset is randomly sampled over time. From the perspective of the whole disease course cycle, A is the incomplete index set on the patient (decision object).

The closer the direction $s_{x,y}(t)$ and distance $d_{x,y}(t)$ indicate a closer relationship $R(A, T)(x)$ between the two laboratory indexes. Therefore, a temporal correlation decision information system TS for chronic disease clinical diagnosis decision is constructed. The upper and lower approximations of $R(A, T)(x)$ for the correlation of laboratory test indicators with respect to TS can be obtained from **Definition 3.5**.

From the discussion in Section 3, the lower approximation $\underline{R}(A, T)(x)$ is the minimum set considering the correlation of indicators in both direction and distance, and the upper approximation $\overline{R}(A, T)(x)$ is the maximum set considering the similar direction or distance. Direction and distance thresholds are set so that the indicators most relevant to the direction of disease progression can be obtained within the minimum set (indicator set).

4.3. Algorithm

The algorithm for the TCRS-TWC model is as follows:

Input: Temporal correlation decision information for clinical diagnosis of chronic diseases without artificial markers.

Output: Optimal decision result.

Step 1: The direction similarity relation R_s is calculated for the temporal attribute information according to Equations (6) and (7).

Step 2: The distance similarity relationship R_d is calculated for the associated attribute information according to Equations (8)–(12).

Step 3: The minimum index set of temporal correlation attribute information is calculated according to Equation (14).

Step 4: The indicators most associated with disease progression (indicator sets) are calculated according to Equation (15).

Step 5: The optimal decision conclusion is obtained.

5. Experiment

5.1. Dataset

With the widespread adoption of her system and the explosive growth of medical data, it has become a primary research direction to mine valuable information from EHRs to assist clinical decision-making without artificial markers. In this section, we used EHRs from 3,094 patients and 80,139 treatment points without artificial markers,

from the Nephrology Department of Zhongshan Hospital of Traditional Chinese Medicine from 2019 to 2022, to verify the validity and applicability of the theoretical model proposed in this paper. Furthermore, compared to the performance of clustering algorithms such as K-means, the experiments were conducted. The data distribution with 80 laboratory indicators is shown in [Figure 3](#).

5.2. Data preprocessing and feature engineering

To ensure reproducibility and methodological rigor, our data preprocessing pipeline is as follows:

Missing data: For patients with missing values in their time series, we employed linear interpolation for gaps of ≤ 3 consecutive time points. Patients with $>30\%$ missing data across all time points were excluded from the analysis.

Normalization: All laboratory indicators were normalized using Min-Max scaling to the range $[0,1]$ to ensure comparability across different measurement units:

$$x_{norm} = \frac{x - x_{\max}}{x_{\max} - x_{\min}}$$

Outlier detection: Values beyond ± 3 standard deviations from the mean for each indicator were treated as potential outliers and reviewed by a clinical expert.

Feature selection: Initial feature selection was performed using variance thresholding (removing features with variance < 0.01) and correlation analysis (removing features with pairwise correlation > 0.95).

Computational implementation: All analyses were implemented in Python 3.9 (Python Software Foundation, Netherlands) using NumPy (v1.24.0), SciPy (v1.10.0), and scikit-learn (v1.2.0) libraries. Experiments were conducted on a server with 32 GB RAM and an Intel Xeon E5-2690 processor.

We used EHRs from 3,094 patients and 80,139 treatment points, without artificial markers, from the Nephrology Department at Zhongshan Hospital of Traditional Chinese Medicine, covering 2019 to 2022, to assess the validity and applicability of the theoretical model proposed in this paper.

5.3. Experimental results

It is assumed that the entire sample space is the lower approximation space of the sample. Given the change measure $\theta = 0.1$ and distance $d = 3,300$, the result of TCRS-TWC is shown in [Figure 4](#). The green dot ($NEG(C)$) indicates the Cr. The blue dots ($POS(C) \cup (BND)$) indicate uric acid, platelet (plt), microalbuminuria, parathyroid hormone and ferritin (Fe). The red dots ($POS(C)$) indicate other indicators.



Figure 3. The data distribution with 83 laboratory indicators and 3,094 samples ($n = 3,094$, $a = 83$)
 Abbreviations: AST/ALT: Aspartate aminotransferase/alanine aminotransferase; CV: Coefficient of variation; HGB: Hemoglobin; PLT: Platelet; RBC: Red blood cell; SD: Standard deviation.

The SC and Calinski–Harabasz index (CH) are used to evaluate the effect of TCRS-TWC. Figure 5 shows the SC in terms of cosine similarity and distance, respectively. Figure 6 illustrates the effect evaluation of TCRS-TWC by CH. The value increases gradually from blue through cyan and green to yellow and red.

After clustering into three categories, each category selected a benchmark index and other indicators of the class to calculate the T value, which was then used to generate a box plot, as shown in Figure 7. Two-sided unpaired t -tests were performed. In boxplot graphs, the center line indicates the median, the bounds of the box indicate the 25th and 75th percentiles, and whiskers indicate the minimum and maximum ($p < 0.0001$). The sample distribution after the t -test (Figure 7) is consistent with Figure 4. The clustering results show meaningful alignment with clinical staging systems and expert judgment, as further detailed in Section 5.7.

5.4. Experimental comparison

Traditional clustering algorithms, including K-means, hierarchical clustering, agglomerative clustering, and Density-Based Spatial Clustering of Applications with Noise, were used to process the dataset in this study. SC and CH were also used to evaluate the clustering effect.

Experimental comparison results are shown in Figure 8.

The proposed TCRS-TWC, with a good classification effect, divided the raw data into three categories. However, increasing the number of clusters in baseline methods tended to reduce cluster stability. In addition, traditional unsupervised clustering methods are primarily used to identify and create data groups or classifications, but cannot be used to evaluate relationships between data features. Therefore, the proposed TCRS-TWC, compared to traditional unsupervised clustering methods, realizes:

- (i) Stability of clustering effect.
- (ii) The relationship between attributes could be used for evaluation.
- (iii) New clinical findings were calculated from large clinical sample data.

5.5. Analytical studies

Clinical experience shows that five critical indicators reflect changes in CKD: urinary microalbumin, Cr, urea nitrogen, CO_2 , and HE. Using the proposed theoretical model, a comprehensive analysis of large-sample data, the top 30 indicators can be identified from 80 indicators. Figure 9 shows the correlations among these 30 indicators; the first row shows cosine similarity, and the second row shows the distance value. The following five sets of strongly correlated indicators are also listed in Figure 9.

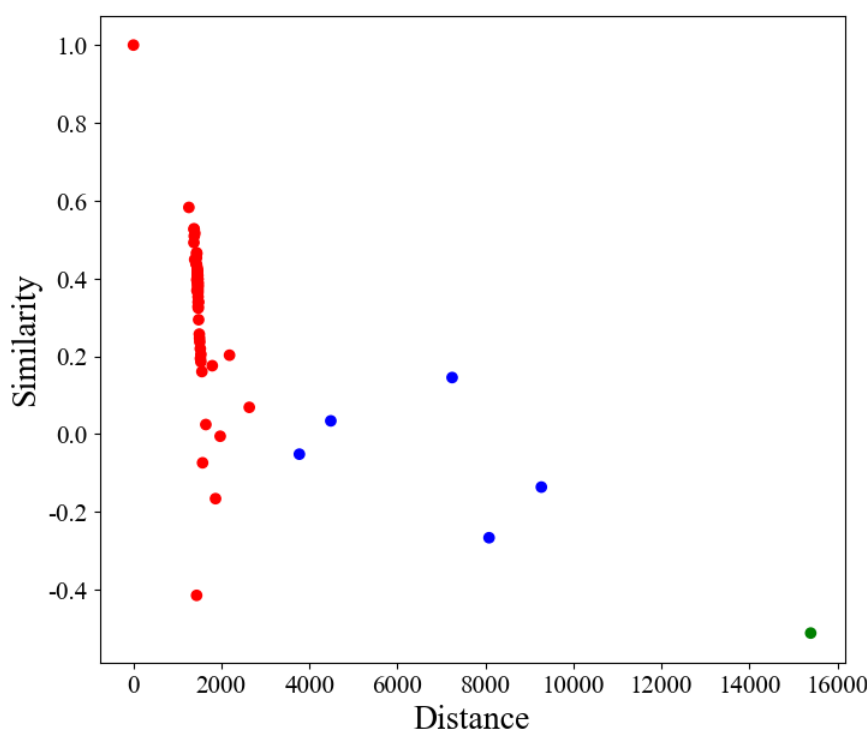


Figure 4. The result of the temporal correlation rough set-based three-way clustering

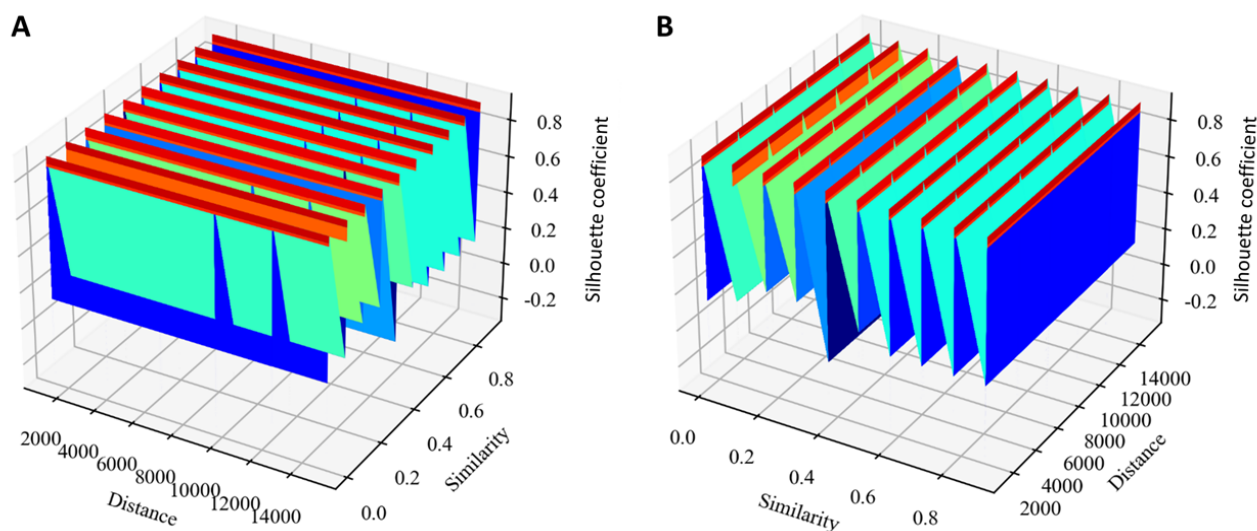


Figure 5. Effect evaluation of the temporal correlation rough set-based three-way clustering by silhouette coefficient from different angles. (A) In terms of cosine similarity. (B) In terms of distance.

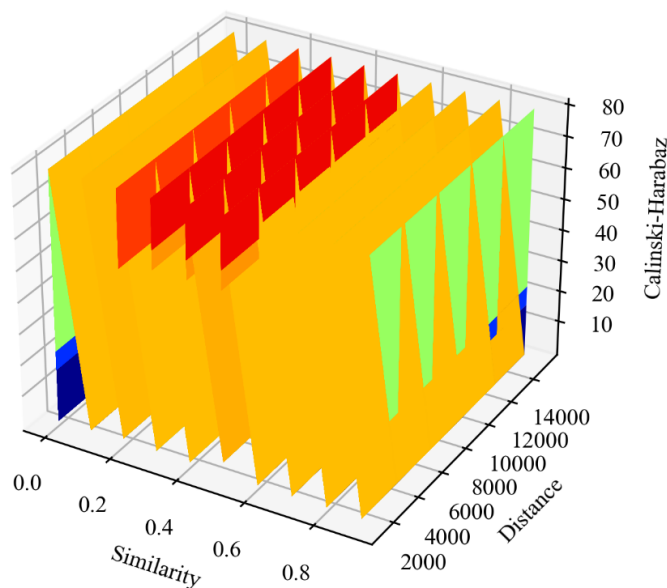


Figure 6. Calinski-Harabasz index evaluation of the temporal correlation rough set-based three-way clustering

Numerous studies and clinical studies have confirmed the relationship between CO_2 , Cr, HE and CKD. Next, we present pioneering studies on the association among $\beta 2$ -microglobulin ($\beta 2$ -MG), Fe, iron, and kidney disease (Table 3).

The results of our study, as well as those from previous studies, suggest that $\beta 2$ -MG may be a biomarker of kidney disease progression. A data-driven clinical decision-making research paradigm can provide clinical cues for diagnosing diseases of unknown etiology using large clinical datasets

and offer a method and technology for marker screening in laboratory research. Our model suggests associations rather than confirming causal or diagnostic relevance. Noting that prospective studies and clinical validation are needed before any clinical application.

5.6. Parameter sensitivity analysis

To address concerns about parameter selection and ensure robustness, we conducted a systematic sensitivity analysis (Table 4). We varied the similarity threshold θ from 0.05

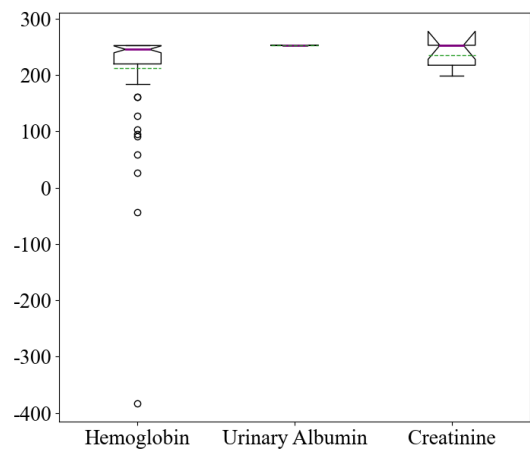


Figure 7. Indicator distribution after the temporal correlation rough set-based three-way clustering

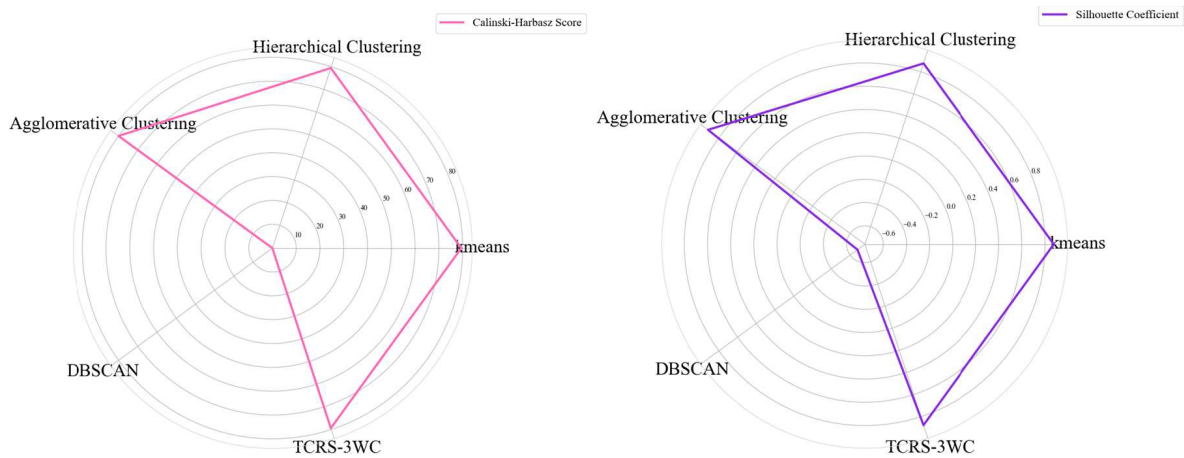


Figure 8. The clustering effects of different factors evaluated by SC and CH based on the chronic kidney disease dataset. (A) SC evaluation of algorithms. (B) CH evaluation of algorithms. Abbreviations: DBSCAN: Density-Based Spatial Clustering of Applications with Noise; TCRS-TWC: Temporal correlation rough set-based three-way clustering.

Table 3. Pioneering studies of the association between β 2-MG, Fe, iron and kidney disease

Indicators	Associations
β 2-MG	• β2-MG as a tool for evaluating donor kidney status for transplantation.⁴⁰ • β2-MG levels associate with the severity of acute kidney injury.⁴¹ • β2-MG is associated with survival in patients with nephropathy.⁴² • β2-MG associated with chronic hemodialysis.⁴³
Fe and Iron	• Clinical diagnosis of CKD and iron treatment.⁴⁴ • Iron homeostasis and Fe in kidney disease.⁴⁵ • The relationship between the expression of Fe and tumor differentiation is discussed.⁴⁶ • Temporal changes in binding of anti-laminin IgG and cationized Fe in developing kidney glomeruli.⁴⁷

Abbreviations: CKD: Chronic kidney disease; Fe: Ferritin; IgG: Immunoglobulin G; β 2-MG: β 2-microglobulin.

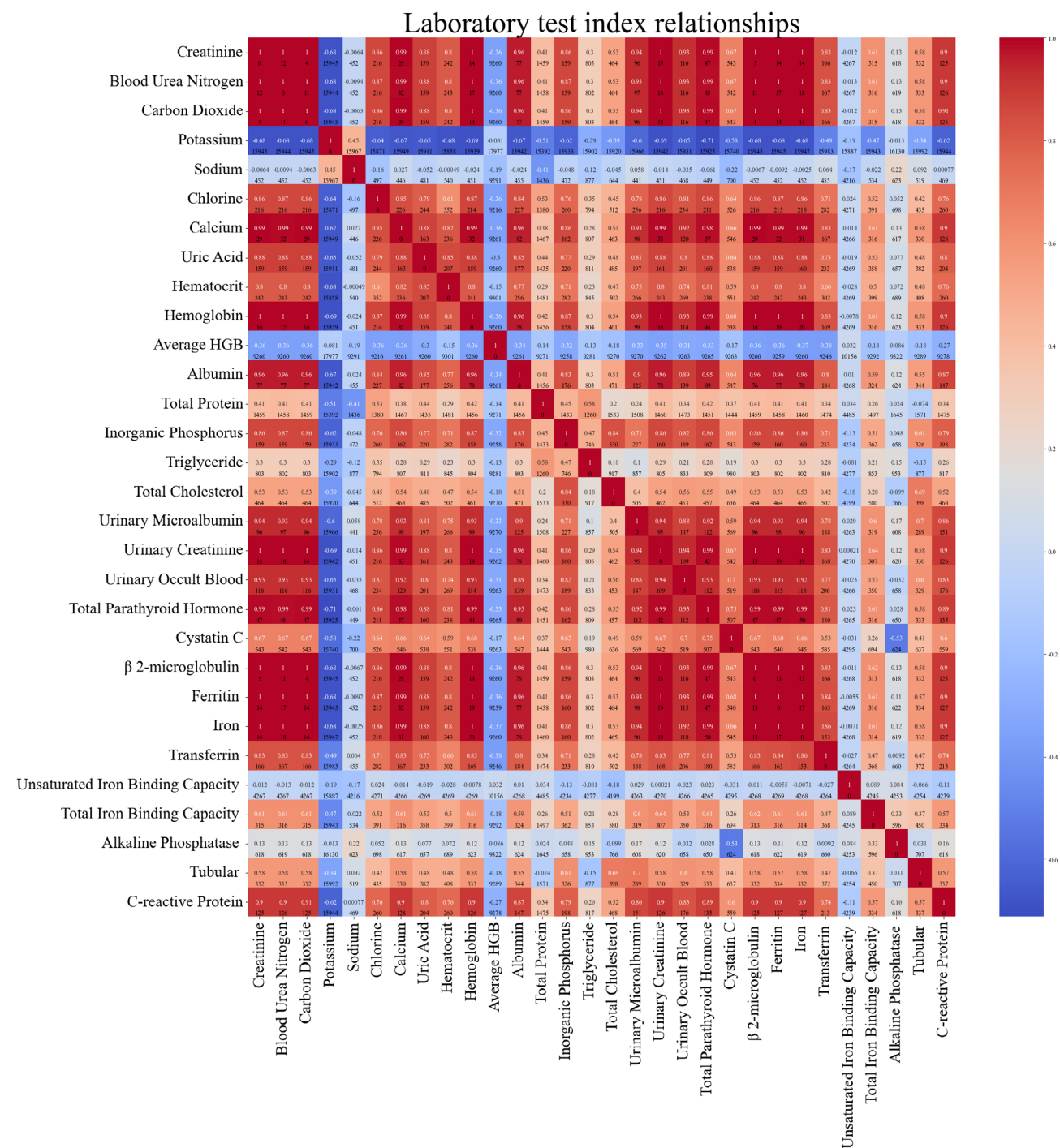


Figure 9. Correlation between the top 30 indicators

to 0.3 (in increments of 0.05) and the distance threshold μ from 2,500 to 4,000 (in increments of 500). For each parameter combination, we computed SC, CH index, and cluster stability (measured as the Jaccard similarity between cluster assignments across different initializations).

Table 4. Sensitivity analysis of parameters θ and μ

θ	μ	SC	CH index	Stability
0.05	2,500	0.68	1,250	0.85
0.05	3,000	0.71	1,320	0.88
0.05	3,500	0.69	1,280	0.86
0.10	2,500	0.72	1,350	0.89
0.10	3,000	0.75	1,420	0.92
0.10	3,500	0.73	1,380	0.90
0.15	3,000	0.70	1,300	0.87
0.15	3,500	0.68	1,260	0.85
0.20	3,000	0.65	1,200	0.82

Abbreviations: CH index: Calinski–Harabasz index; SC: Silhouette coefficient.

The selected values ($\theta = 0.1$, $\mu = 3,300$) correspond to the maximum SC (0.75), indicating well-separated clusters. High CH index (1,420) suggests dense, well-separated clusters. Excellent stability (0.92) ensures reproducible results. Clinical interpretability of the resulting clusters.

5.7. Clinical correlation analysis

To validate the clinical meaningfulness of our clustering results beyond internal metrics, we conducted two additional analyses:

(i) Comparison with KDIGO Staging

For the 2,145 patients with available estimated glomerular filtration rates, we compared our cluster assignments with KDIGO stages (G1–G5). We observed moderate agreement (Cohen's $\kappa = 0.42$), with cluster 1 (red) containing predominantly G1–G2 patients, cluster 2 (blue) containing G3 patients, and cluster 3 (green) containing G4–G5 patients.

(ii) Expert validation

Two board-certified nephrologists independently reviewed the indicator clusters identified by our model. Both experts agreed that:

Cluster 1 (red) represented “stable renal function” indicators,

Cluster 2 (blue) represented “inflammatory/metabolic” indicators,

Cluster 3 (green) represented “advanced renal dysfunction” indicators.

The experts noted that the inclusion of $\beta 2$ -MG in cluster 3 was clinically plausible given its known association with advanced CKD, though they emphasized this requires prospective validation.

5.8. Discussion

From a methodological perspective, the model's performance, as measured by SC (0.75) and CH (1,420), indicates well-separated and internally coherent clusters (Figures 5 and 6). This suggests that the constructed TCRS effectively captures meaningful patient subgroups based on the evolving trajectories of laboratory indicators, rather than static snapshots. Importantly, the model identified a distinct cluster (green/NEG(C)) centered on Cr, consistent with its established clinical role as a cornerstone marker of renal function. The separation of this cluster from others containing inflammatory/metabolic markers (e.g., uric acid, parathyroid hormone, Fe) in the blue region (POS(C)UBND(C)) suggests the model's ability to distinguish between pathways of renal dysfunction from associated systemic complications. The most significant output, however, is the model's hypothesis-generating capability. By analyzing correlation patterns among the top 30 indicators (Figure 9), the model consistently associated $\beta 2$ -MG with key markers of CKD progression (CO_2 , HE, Fe, and Iron). While $\beta 2$ -MG is well known in nephrology, its strong embedding in the derived correlation network suggests it may have underutilized or composite prognostic value when combined with other indicators. This data-driven insight is supported by the retrospective literature analysis (Table 3), which highlights $\beta 2$ -MG's relevance in kidney injury severity, transplantation evaluation, and patient survival. Thus, the model not only rediscovers known clinical relationships (validating its method) but also nominates $\beta 2$ -MG as a high-priority candidate biomarker warranting targeted prospective validation in CKD staging and monitoring.

6. Conclusion

Medical decision-making is an important area of research in decision theory. Data science has changed the traditional decision-making paradigm in healthcare. Data-driven methods are being applied clinically to improve diagnosis accuracy and treatment for individual patients.^{48–50} Based on the interdisciplinary theories, models, methods and technologies of data science, management science, decision science and intelligent computing, the construction of clinical diagnosis and decision theory models and methods driven by massive attribute data can bring new opportunities and effective ideas for the construction of scientific and accurate chronic disease clinical diagnosis and decision-making technology.

This study summarizes a class of clinical diagnostic decision problems with temporal correlation and attribute information characteristics, drawn from the decision-making process and clinical data for diagnosing CKD. By

applying the RS and TWC theory, a TCRS-TWC model was constructed and applied to clinical decision-making problems in CKD. Firstly, a TCRS model based on the similarity of gradient and cosine was constructed to realize the pre-classification of temporal correlation information without artificial labels. Secondly, the unsupervised TWC model was used to construct the two-set and three-region RS model of temporal collation, and a TWC model based on a positive domain was designed. Finally, using the unmarked dataset of 3,094 CKD patients, based on the proposed theoretical model, unsupervised clustering of CKD patients was realized, the characteristics of the patient population in clinical diagnosis of CKD were found, the trends in CKD diagnosis that could not be detected initially were identified, and the key correlation indicators of CKD diagnosis were identified, providing supplementary support and suggestions for clinical decision-making based on experimental results.

Our study has several limitations that warrant consideration. The analysis relies on single-center, retrospective data, which may limit generalizability, and the model's unsupervised nature limits causal inference. Additionally, the absence of external validation and clinical endpoint assessment underscores the need for further investigation. Despite these constraints, the TCRS-TWC model demonstrates potential as a data-driven tool for biomarker discovery and pattern recognition in chronic disease management.

Future research should focus on multi-center validation to enhance generalizability, prospective studies to establish predictive value, and integration of multimodal data, such as imaging and genomic information. Extending the model to other chronic diseases, developing real-time clinical dashboards, and conducting interventional studies to assess impact on patient outcomes represent important directions for advancing this line of research. While not yet suitable for direct clinical application, the TCRS-TWC framework offers a meaningful contribution to the evolving paradigm of data-informed clinical decision support.

Acknowledgments

We sincerely thank Guangdong Provincial Hospital of Chinese Medicine for providing the valuable clinical data and imaging materials that enabled this multimodal CKD analysis. We also extend our appreciation to all medical staff and research assistants involved in data collection and preliminary processing for their diligent work. Furthermore, we are grateful to our research team members for their insightful suggestions during model construction and experimental validation, and to our peers for the stimulating academic exchanges. We also

acknowledge the general research support and resources provided by our institution that facilitated this study.

Funding

The work was supported by the National Natural Science Foundation of China (72301082), the China Post doctoral Science Foundation (2025M773863), the Guangdong Basic and Applied Basic Research Foundation (2022A1515110703), the Guangdong Provincial Hospital of Chinese Medicine Science and Technology Research Project (YN2022QN33, YN2024GZRPY077), the Guangzhou Key Research and Development Program (202206010101), the National Key Laboratory of Chinese Medicine Syndrome (QZ2023ZZ07), the Postdoctoral Fellowship Program of CPSF under Grant (GZC20252561), the Special Project of State Key Laboratory of Dampness Syndrome of Chinese Medicine (SZ2021ZZ36, SZ2021ZZ09), the Guangzhou Science and Technology Plan Project (2024A03J0117, 2025A03J4062) and the 2020 Guangdong Provincial Science and Technology Innovation Strategy Special Fund (Guangdong-Hong Kong-Macau Joint Lab) (2020B1212030006).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Yan Zhang

Formal analysis: Wanzhen Long

Investigation: Ruihui Chen, Zihui Lin

Methodology: Lincong Li

Project administration: Haole Huang

Resources: Xiaoli Chu

Writing—original draft: Ruihui Chen

Writing—review & editing: Ruihui Chen

Ethics approval and consent to participate

Guangdong Provincial Hospital of Traditional Chinese Medicine Ethics Committee.BE2025-098-01.

Consent for publication

All patients provided informed oral consent during their clinical visits for the anonymous use of their data for scientific publication.

Availability of data

The dataset used and analyzed in this study is non-public data, and its ownership belongs to Guangdong Provincial Hospital of Traditional Chinese Medicine. Due to patient privacy concerns and restrictions under the hospital's data-sharing policy, the original data is not publicly

available. If there is a legitimate research need, please contact the corresponding author of this paper or the Data Management Committee of Guangdong Provincial Hospital of Traditional Chinese Medicine. The relevant data may be applied for use under the condition that the ethical review of the institution is passed and a data use agreement is signed.

References

1. Quam L, Smith R, Yach D. Rising to the global challenge of the chronic disease epidemic. *Lancet*. 2006;368(9543):1221–1223.
doi: 10.1016/S0140-6736(06)69422-1
2. Furman D, Campisi J, Verdin E, *et al*. Chronic inflammation in the etiology of disease across the life span. *Nat Med*. 2019;25(12):1822–1832.
doi: 10.1038/s41591-019-0675-0
3. Kalantar-Zadeh K, Li P. Strategies to prevent kidney disease and its progression. *Nat Rev Nephrol*. 2020;16(3):129–130.
doi: 10.1038/s41581-020-0253-1
4. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2020;395(10225):709–733.
doi: 10.1016/s0140-6736(20)30045-3
5. Feinstein A. The pre-therapeutic classification of co-morbidity in chronic disease. *J Chronic Dis*. 1970;23(7):455–468.
doi: 10.1016/0021-9681(70)90054-8
6. Stockley R, Halpin D, Celli B, *et al*. Chronic obstructive pulmonary disease biomarkers and their interpretation. *Am J Respir Crit Care Med*. 2019;199(9):1195–1204.
doi: 10.1164/rccm.201810-1860SO
7. Pinto-Plata V. Use of Proteomic Patterns of Serum Biomarkers in Patients with Chronic Obstructive Pulmonary Disease: Correlation with Clinical Parameters. *Proc Am Thorac Soc*. 2006;3(6):465–466.
doi: 10.1513/pats.200603-030ms
8. Sinha K, Uddin Z, Kawsar H, *et al*. Analyzing chronic disease biomarkers using electrochemical sensors and artificial neural networks. *Trends Analyt Chem*. 2023;158:116861.
doi: 10.1016/j.trac.2022.116861
9. Topalovic M, Laval S, Aerts JM, Troosters T, Decramer M, Janssens W. Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints. *Respiration*. 2017;93(3):170–178.
doi: 10.1159/000454956
10. Liu Z, Liu L, Weng S, *et al*. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13(1):816.
doi: 10.1038/s41467-022-28421-6
11. Li J, Huang Q, Ren S, *et al*. A novel medical text classification model with Kalman filter for clinical decision making. *Biomed Signal Process Control*. 2023;82:104503.
doi: 10.1016/j.bspc.2022.104503
12. Choi MY, Chen I, Clarke AE, *et al*. Machine learning identifies clusters of longitudinal autoantibody profiles predictive of systemic lupus erythematosus disease outcomes. *Ann Rheum Dis*. 2023.
doi: 10.1136/ard-2022-223808
13. Wang X, Chen Y, Gao Y, *et al*. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat Commun*. 2021;12(1):1637.
doi: 10.1038/s41467-021-21674-7
14. Pawlak Z. Rough sets. *Int J Comput Inf Sci*. 1982;11(5):341–356.
doi: 10.1007/BF01001956
15. Xing J, Gao C, Zhou J. Weighted fuzzy rough sets-based tri-training and its application to medical diagnosis. *Appl Soft Comput*. 2022;124:109025.
doi: 10.1016/j.asoc.2022.109025
16. Srivastava S, Meena YK, Singh G. Forecasting on COVID-19 infection waves using a rough set filter-driven moving average model. *Appl Soft Comput*. 2022;131:109750.
doi: 10.1016/j.asoc.2022.109750
17. Tsumoto S. Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model. *Inf Sci*. 2004;162(2):65–80.
doi: 10.1016/j.ins.2004.03.002
18. Yao Y. Two views of the theory of rough sets in finite universes. *Int J Approx Reason*. 1996;15(4):291–317.
doi: 10.1016/S0888-613X(96)00071-0
19. Grzymala-Busse JW. LERS-A System for Learning from Examples Based on Rough Sets. In: *Intelligent Decision Support*. Springer; 1992:3–18.
doi: 10.1007/978-94-015-7975-9_1
20. Li W, Yang B. Three-way decisions with fuzzy probabilistic covering-based rough sets and their applications in credit evaluation. *Appl Soft Comput*. 2023;136:110144.
doi: 10.1016/j.asoc.2023.110144
21. Qahtan S, Alsattar HA, Zaidan AA, Deveci M, Pamucar D, Delen D. Performance assessment of sustainable transportation in the shipping industry using a q-rung orthopair fuzzy rough sets-based decision making

- methodology. *Expert Syst Appl.* 2023;223:119958.
doi: 10.1016/j.eswa.2023.119958
22. Zheng Q, Chang H, Liu Z, *et al.* Multi-stage design space reduction technology based on SOM and rough sets, and its application to hull form optimization. *Expert Syst Appl.* 2023;213:119229.
doi: 10.1016/j.eswa.2022.119229
23. Bai J, Guo J, Sun B, *et al.* Intelligent forecasting model of stock price using neighborhood rough set and multivariate empirical mode decomposition. *Eng Appl Artif Intell.* 2023;122:106106.
doi: 10.1016/j.engappai.2023.106106
24. Tishya M, Anitha A. Precipitation prediction by integrating rough set on fuzzy approximation space with deep learning techniques. *Appl Soft Comput.* 2023;139:110253.
doi: 10.1016/j.asoc.2023.110253
25. Yan HY, Zhang XR, Dong JH, Shang MS, Shan K, Wu D, *et al.* Spatial and temporal relation rule acquisition of eutrophication in Da'ning River based on rough set theory. *Ecol Indic.* 2016;66:180-189.
doi: 10.1016/j.ecolind.2016.01.032
26. Yao Y. The superiority of three-way decisions in probabilistic rough set models. *Inf Sci.* 2011;181(6):1080-1096.
doi: 10.1016/j.ins.2010.11.019.
27. Podsiadlo M, Rybinski H. Financial Time Series Forecasting using Rough Sets with Time-Weighted Rule Voting. *Expert Systems with Applications*, 2016;66:219-233.
doi: 10.1016/j.eswa.2016.08.066.
28. Selvakumar K, Karuppiiah M, SaiRamesh L, Islam SH, Hassan MM, Fortino G, *et al.* Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. *Inf Sci.* 2019;497:77-90.
doi: 10.1016/j.ins.2019.05.040
29. Dong L, Wang R, Chen D. Incremental feature selection with fuzzy rough sets for dynamic data sets. *Fuzzy Sets Syst.* 2023;467:108503.
doi: 10.1016/j.fss.2023.03.006
30. Ni P, Zhao S, Wang X. Incremental feature selection based on fuzzy rough sets. *Inf Sci.* 2020;536:185-204.
doi: 10.1016/j.ins.2020.04.038
31. Che X, Chen D, Mi J. Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets. *Fuzzy Sets Syst.* 2022;426:121-144.
doi: 10.1016/j.fss.2021.03.016
32. Lee S, Enke D, Kim Y. A relative value trading system based on a correlation and rough set analysis for the foreign exchange futures market. *Eng Appl Artif Intell.* 2017;61:47-56.
doi: 10.1016/j.engappai.2017.02.014
33. Afridi MK, Azam N, Yao J. Variance based three-way clustering approaches for handling overlapping clustering. *Int J Approx Reason.* 2020;118:47-63.
doi: 10.1016/j.ijar.2019.11.011
34. Yu H, Chen L, Yao J. A three-way clustering method based on an improved DBSCAN algorithm. *Phys A.* 2019;535:122289.
doi: 10.1016/j.physa.2019.122289
35. Sun C, Du M, Sun J, Li K, Dong Y. A three-way clustering method based on improved density peaks algorithm and boundary detection graph. *Int J Approx Reason.* 2023;153:239-257.
doi: 10.1016/j.ijar.2022.12.002
36. Yu H, Chen L, Yao J. A three-way density peak clustering method based on evidence theory. *Knowl Based Syst.* 2021;211:106532.
doi: 10.1016/j.knosys.2020.106532
37. Ali B, Azam N, Yao J. A three-way clustering approach using image enhancement operations. *Int J Approx Reason.* 2022;149:1-38.
doi: 10.1016/j.ijar.2022.07.001
38. Yu H, Zhang C, Wang G. A tree-based incremental overlapping clustering method using three-way decision theory. *Knowl Based Syst.* 2016;91:189-203.
doi: 10.1016/j.knosys.2015.05.028
39. Yao Y. Three-way decision and granular computing. *Int J Approx Reason.* 2018;103:107-123.
doi: 10.1016/j.ijar.2018.09.005
40. Cha S, Shin I, Kim D, *et al.* Effectiveness of serum β 2-microglobulin for evaluating donor kidney status. *Sci Rep.* 2020;10:8109.
doi: 10.1038/s41598-020-65134-6
41. Barton KT, Kakajiwala A, Dietzen DJ, Goss CW, Gu H, Dharnidharka VR. Using the newer Kidney Disease: Improving Global Outcomes criteria, beta-2-microglobulin levels associate with severity of acute kidney injury. *Clin Kidney J.* 2018;11(6):797-802.
doi: 10.1093/ckj/sfy056
42. Kanda E, Muenz D, Bieber B, Cases A, Locatelli F, Port FK, *et al.* Beta-2 microglobulin and all-cause mortality in the era of high-flux hemodialysis: results from the Dialysis Outcomes and Practice Patterns Study. *Clin Kidney J.* 2020;14(5):1436-1442.
doi: 10.1093/ckj/sfaa155
43. Gejyo F, Yamada T, Odani S, Nakagawa Y, Arakawa M, Kunitomo T, *et al.* A new form of amyloid protein

- associated with chronic hemodialysis was identified as β 2-microglobulin. *Biochem Biophys Res Commun*. 1985;129(3):701-706.
doi: 10.1016/0006-291x(85)91948-5
44. Fishbane S, Kalantar-Zadeh K, Nissenson AR. Serum Ferritin in Chronic Kidney Disease: Reconsidering the Upper Limit for Iron Treatment. *Semin Dial*. 2004;17(5):336-341.
doi: 10.1111/j.0894-0959.2004.17359.x
 45. McCullough K, Bolisetty S. Iron homeostasis and ferritin in sepsis-associated kidney injury. *Nephron*. 2020;144(12):616–620.
doi: 10.1159/000509073
 46. Fleming S. Immunocytochemical Localization of Ferritin in the Kidney and Renal Tumours. *Eur Urol*. 1987;13(6):407-411.
doi: 10.1159/000472835
 47. Abrahamson DR, John PLSt. Ultrastructure of developing kidney glomerular basement membranes: Temporal changes in binding of anti-laminin IgG and cationized ferritin. *Microsc Res Tech*. 1994;28(2):81-94.
doi: 10.1002/jemt.1070280202
 48. Operskalski J, Barbey A. Risk literacy in medical decision-making. *Science*. 2016;352(6284):413–414.
doi: 10.1126/science.aaf7966
 49. Athey S. Beyond prediction: Using big data for policy problems. *Science*. 2017;355(6324):483–485.
doi: 10.1126/science.aal4321
 50. Meara S. China's data-driven dream to overhaul health care. *Nature*. 2021;598:S1–S3.
doi: 10.1038/d41586-021-02694-1