

# Interpretable subgroup learning-based modeling framework: Study of diabetic kidney disease prediction

Health Informatics Journal  
1–18

© The Author(s) 2024

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/14604582241291379

[journals.sagepub.com/home/jhi](https://journals.sagepub.com/home/jhi)



**Bo Liu\*** 

Big Data Decision Institute, Jinan University, Guangzhou, China  
School of Management, Jinan University, Guangzhou, China

**Xiangzhou Zhang\***

Big Data Decision Institute, Jinan University, Guangzhou, China  
School of Medicine, Jinan University, Guangzhou, China

**Kang Liu\***

Big Data Decision Institute, Jinan University, Guangzhou, China  
School of Management, Jinan University, Guangzhou, China

**Xinhou Hu**

Big Data Decision Institute, Jinan University, Guangzhou, China

**Eric W. T. Ngai**

Faculty of Business, The Hong Kong Polytechnic University, Hong Kong, China

**WeiQi Chen**

School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China

**Ho Yin Chan**

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

**Yong Hu<sup>†</sup>**

Big Data Decision Institute, Jinan University, Guangzhou, China

**Mei Liu<sup>†</sup>**

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

\*These authors contributed equally to this study.

<sup>†</sup>Authors contributed equally and should be considered as co-corresponding authors.

## Corresponding authors:

Mei Liu, PhD Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, USA.

Email: [mei.liu@ufl.edu](mailto:mei.liu@ufl.edu)

Yong Hu, Big Data Decision Institute, Jinan University, Guangzhou, China.

Email: [yonghu@jnu.edu.cn](mailto:yonghu@jnu.edu.cn)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Abstract

**Objectives:** Complex diseases, like diabetic kidney disease (DKD), often exhibit heterogeneity, challenging accurate risk prediction with machine learning. Traditional global models ignore patient differences, and subgroup learning lacks interpretability and predictive efficiency. This study introduces the Interpretable Subgroup Learning-based Modeling (iSLIM) framework to address these issues.

**Methods:** iSLIM integrates expert knowledge with a tree-based recursive partitioning approach to identify DKD subgroups within an EHR dataset of 11,559 patients. It then constructs separate models for each subgroup, enhancing predictive accuracy while preserving interpretability.

**Results:** Five clinically relevant subgroups are identified, achieving an average sensitivity of 0.8074, outperforming a single global model by 0.1104. Post hoc analyses provide pathological and biological evidence supporting subgroup validity and potential DKD risk factors.

**Conclusion:** The iSLIM surpasses traditional global model in predictive performance and subgroup-specific risk factor interpretation, enhancing the understanding of DKD's heterogeneous mechanisms and potentially increasing the adoption of machine learning models in clinical decision-making.

## Keywords

clinical decision support system, diabetic kidney disease, electronic health record, predictive modeling, subgroup learning, clinical decision support system

## Introduction

The rapid digitization of the healthcare industry<sup>1</sup> has ushered in an era of unprecedented electronic health record (EHR) data, remarkably enhancing the entire digital health ecosystem and propelling the development of clinical decision support tools.<sup>2,3</sup> Within this transformative landscape, machine learning (ML), a subset of artificial intelligence (AI), has been set with high expectations, typically in application of clinical decision support systems (CDSS).<sup>4,5</sup> These systems significantly advanced disease risk prediction methodologies and augmented clinical decision-making.<sup>6,7</sup> The predominant predictive modeling strategy involves training a single model on data from a population at risk, known as the global modeling strategy. While effective in achieving superior generalization, it falls short when faced with the inherent heterogeneity of most complex diseases.<sup>8</sup> For example, as pointed out by Ford et al.,<sup>9</sup> the diagnosis of diabetes by a global model, typically reliant on glycosylated hemoglobin (HbA1c), exhibits intricate variations across ethnicities. This underscores the limitations of a global model, which may overlook delicate relationships and fail to be the optimal choice across diverse subpopulations.

In response, the subgroup learning strategy offers an alternative solution. This strategy requires the training of independent models using data from each subgroup and providing a refined approach to address the complexity of different patient populations.<sup>10</sup> Nevertheless, subgroup learning still faces certain problems in practical applications. First, the strategy relies on the accurate identification of subgroups. Given the incomplete understanding of many complex diseases, there is a growing emphasis on adopting data-driven subgrouping method.<sup>11,12</sup> Yet, a drawback arises from the lack of interpretability in most data-driven subgroup identification, a crucial factor for physicians

in understanding and making confident decisions.<sup>13</sup> For example, clustering, while capable of automatically identifying subgroups, is viewed with clinical skepticism due to its opacity. While some studies proposed hybrid approaches that incorporate knowledge,<sup>14</sup> the predominant focus remains on knowledge discovery rather than risk prediction.; second, subgrouping severely affects the predictive performance of the model, leading to insufficient training samples and a consequent decline in performance.<sup>15</sup> Although techniques such as multicenter modeling and oversampling, like the synthetic minority oversampling technique (SMOTE),<sup>16,17</sup> offer solutions, obtaining compatible data is difficult in practice and raises safety and cost concerns; third, predictive modeling performance measurement often ignores clinical needs, with physicians expressing a preference for metrics like sensitivity and precision over the area under the curve (AUC); finally, many subgroup models neglect to contribute to disease pathogenesis, depriving physicians of informative clues for further exploration. The above issues cannot be ignored when integrating subgroup strategies into broader digital health systems such as CDSS. Presently, there is a dearth of comprehensive studies addressing these multifaceted issues, and enhanced subgroup strategy approaches remain limited.

These issues are particularly relevant in the context of predicting complex diseases characterized by heterogeneity. Diabetic kidney disease (DKD) is a notable example, for which there is currently a lack of recognized understanding of pathogenesis and lacks international guidelines. Its traditional predictors, such as microalbuminuria, proteinuria, and estimated glomerular filtration rate (eGFR)<sup>18</sup> are considered inadequate and unreliable.<sup>19,20</sup> Variations in DKD incidence across groups is influenced by factors such as race,<sup>21</sup> gender,<sup>22</sup> and age.<sup>23</sup> While predictive models based on electronic medical record data, such as the logistic regression employed by Makino et al.<sup>24</sup> with an accuracy of 71%, have shown promise, it is essential to acknowledge the use of global learning strategies inevitably receives the benefit of potential heterogeneity, and the prevention and cure rates of DKD have so far been unsatisfactory. Improving the prediction and prevention of DKD risk requires addressing the multifaceted challenges associated with subgroup strategy. This is critical to advance understanding and prevention of all complex diseases.

This study aims to address the aforementioned issues of disease predictive modeling, especially for complex diseases with heterogeneity. We propose a subgroup learning-based framework that leverages DKD EHR data and machine learning with expert guidance. The main focus is on enhancing interpretability of subgroup identification, ensure the performance of predictive models, and provide valuable insights for disease exploration, thus enhance the predictive capabilities of CDSS and contribute to the advancement of personalized medicine and digital health solutions.

## Methods

### *Study cohort*

We utilized a cohort from the Healthcare Enterprise Repository for Ontological Narration (HERON) at the University of Kansas Medical Center (KUMC), which comprised EHR of 11,559 patients with Type 2 Diabetes Mellitus (T2DM) from January 2010 to December 2016. The surveillance, prevention, and management of diabetes mellitus (SUPREME-DM) definition of diabetes was adopted in this study to define diabetes status.<sup>25</sup> All patients had no prior history of renal disease, with diabetes onset occurring at least 30 days before the endpoint, and had at least one valid eGFR or albumin-to-creatinine (ACR) record from outpatient visits to determine DKD status. The case group included patients diagnosed with DKD, defined by the first occurrence of an abnormal eGFR or ACR. The control group consisted of diabetic patients with consistently normal eGFR values ( $\geq 60$  mL/min/1.73 m<sup>2</sup>) and no history of microalbuminuria, with the endpoint defined as the last

recorded normal eGFR or ACR. In this cohort, 5580 patients (44.73%) were identified with DKD, aligning with the typical range of 30% to 50% for T2DM patients developing DKD.<sup>26</sup>

This dataset received approval from the HERON Data Request Oversight Committee and complied with HIPAA de-identification criteria. The KUMC Institutional Review Board (IRB) reviewed and approved the consent procedures.

### *Expert knowledge for subgroup learning*

Leveraging domain knowledge can enhance the interpretability and reliability of disease prediction and patient classification.<sup>27</sup> Consequently, we selected well-established interpretable risk factors of DKD<sup>28–31</sup> as the expert knowledge to guide subgroup identification. Notably, these selected risk factors exclude those derived solely from data-driven methods lacking pathological validation. Table 1 presents seven DKD risk factors, including age, gender, and hypertension etc. The eGFR and ACR were excluded from the list, since they were already used as the traditional diagnostic criteria for the cohort selection.

### *Clinical variables for prediction modeling*

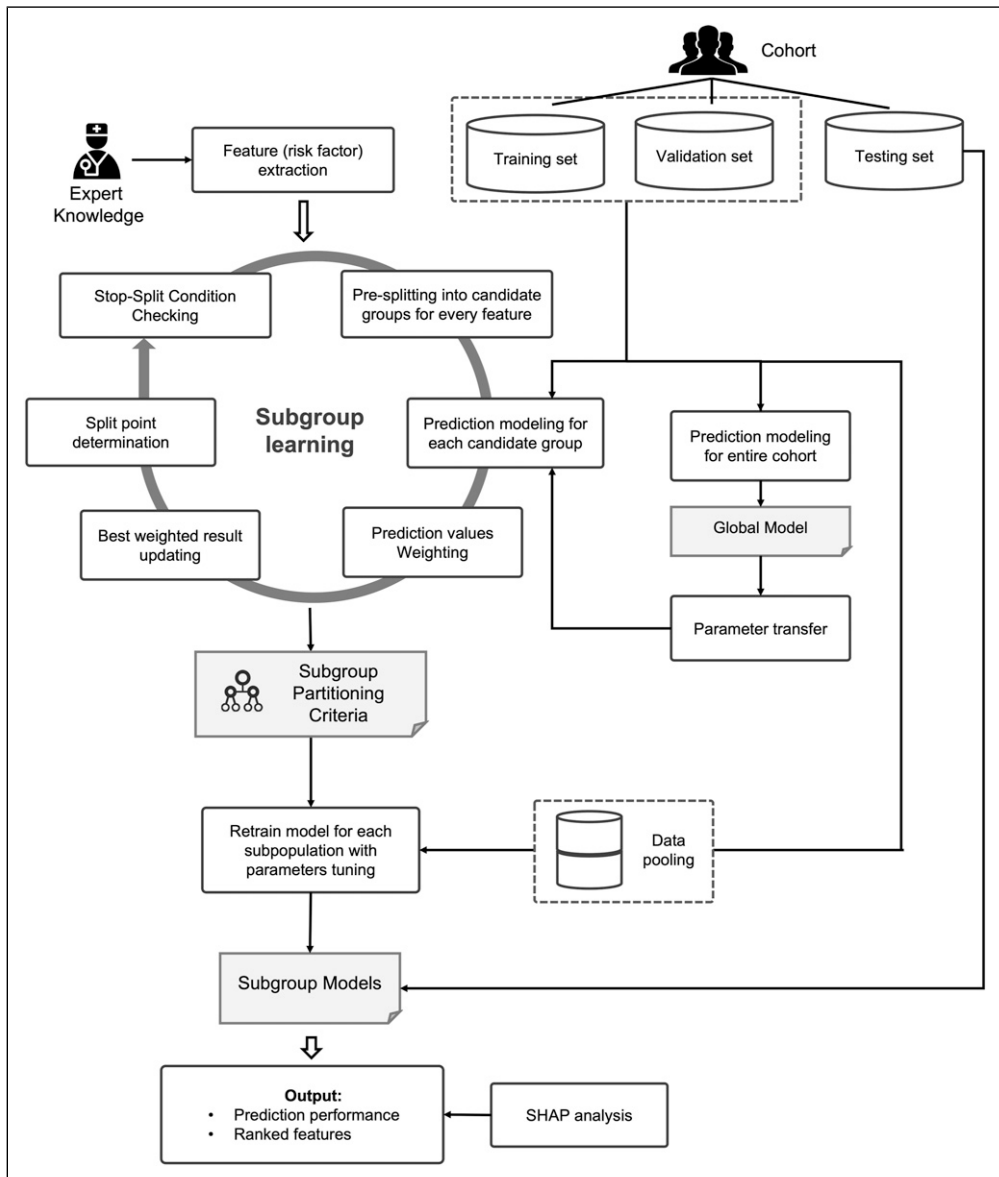
We utilized a set of 401 clinical variables (features) identified as important for predicting DKD.<sup>32</sup> The selected features included the subgroup learning factors, lab. tests, medications, and history etc. These features demonstrated significance in predicting DKD onset and were robust to data perturbations.

### *Measurement for prediction performance*

Although AUC is commonly used in model evaluation, it is less likely to represent the outcomes of individual patients. Clinicians prioritize clinically relevant thresholds over AUC's overall performance, favoring sensitivity and precision, which inherently trade off, and the choice depends on

**Table 1.** Features to guide subgroup learning according to the DKD risk factors selected from the literature.

Features	DKD no. (%)	Non-DKD no. (%)	p Value
N (sample size)	5,171 (44.73)	6,388 (55.27)	
Age			
18~65	2,942 (25.45)	4,585 (39.66)	<0.001
65+	2,229 (19.28)	1,803 (15.60)	<0.001
Male	2,453 (21.22)	3,234 (27.98)	0.011
Hyperglycemia (in T2DM)	302 (2.61)	442 (3.82)	0.023
Hypertension	1,425 (12.33)	1,461 (12.64)	<0.001
Obesity (BMI ≥28)	2,192 (18.96)	2,914 (20.89)	<0.001
Smokeless Tobacco used	2,258 (19.76)	3,482 (0.3013)	<0.001
Race			
Black	1,031 (8.92)	1,316 (11.77)	0.108
Others	481 (4.16)	862 (7.45)	<0.001
White	3,659 (31.65)	4,210 (36.42)	0.002



**Figure 1.** Flowchart of the proposed iSLIM framework.

the disease management strategy. In the context of DKD, underdiagnosis is costlier than misdiagnosis, so we focused on identifying as many at-risk individuals as possible, emphasizing sensitivity both in subgrouping and modeling process, as the main criterion for model performance assessment.

### *Subgroup learning-based modeling*

The proposed Interpretable Subgroup Learning-based Modeling (iSLIM) framework, as illustrated in [Figure 1](#), begins by integrating expert knowledge with a transparent 'white-box' tree-based recursive partitioning method<sup>33,34</sup> to identify subgroups in the dataset, thereby ensuring interpretability. During the iterative tree construction, optimal sensitivity gain—derived from the prediction values of the LightGBM model—is used as the criterion for node splitting to maintain clinical relevance. This process continues until the tree is fully grown, with each leaf node representing an identified subgroup. The pooled data from the corresponding training and validation subsets are then merged to re-train the model and adjust hyperparameters, creating a unique LightGBM prediction model for each subgroup. This approach not only optimizes predictive performance but also captures subgroup-specific risk factors and relationships that may be obscured in a global model. Finally, the models are validated using their respective testing subsets.

To counter model performance degradation from sample partitioning in subgroup learning, we implement parameters-based transfer learning.<sup>35,36</sup> This approach initializes the LightGBM subgroup prediction model using the global model, with the residuals from the initial model serving as inputs for subsequent weak classifiers. Through iterative training, this method ensures effective inheritance of global knowledge while maintaining performance.

In evaluating the representativeness and influence of each subgroup's key features, we employ SHAP analysis.<sup>37</sup> SHAP values are utilized to interpret feature importance within each subgroup model, helping to identify subgroup-specific risk factors and their relative impacts. The iSLIM framework is a systematically developed approach, not just a combination of methods, and is aligned with real-world clinical scenarios.

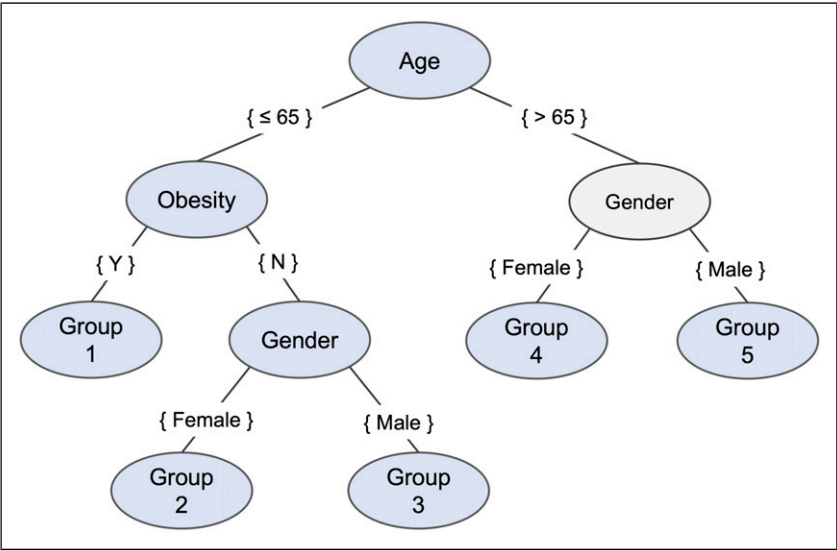
### *Experimental setting*

Stratified random sampling extracted 20% of the original dataset (2312) as the testing set, with the remaining samples (9247) divided into an 8:2 ratio for training (7398) and validation (1849). Key parameters for the tree-based recursive method included the minimum leaf node size (MLN) and minimum internal node size (MIN). A grid-search technique with a range of MLN = 500–2000 and MIN = 1000–2500 (step size = 50) prevented overfitting or underfitting. Given that the cohort dataset had undergone preliminary cleaning, features with more than 90% missing values and those not specific to DKD patients were already removed, effectively reducing data sparsity. Consequently, we opted to utilize LightGBM's default missing value handling mechanism without additional preprocessing. This approach leverages LightGBM's built-in capabilities for managing sparse data, streamlining our modeling process while maintaining data integrity. Sensitivity was the primary performance metric, with precision, AUC, and f1-score also calculated. Twenty-five bootstrap rounds on the testing subset were conducted for each group, applying the student's *t*-test for *p*-value calculation.

## **Results**

### *Identified subgroups*

With final parameters set at MLN = 650 and MIN = 1500, we obtained the optimal tree split by three risk factors (including root node): age, obesity, and gender, which constitute five subgroups (see [Figure 2](#) and [Table 2](#)). Other risk factors were eliminated in the optimal split node selection. DKD incidence rate was



**Figure 2.** Final learned subgroup tree.

**Table 2.** Learned subgroups details.

Subgroups	Characteristics	Samples No.(%) <sup>a</sup>	DKD No.(%) <sup>b</sup>
Group 1	Non-elderly, obesity	3491 (30.2)	1302 (37.3)
Group 2	Non-elderly, non-obesity, female	2023 (17.5)	874 (43.2)
Group 3	Non-elderly, non-obesity, male	2011 (17.4)	754 (37.5)
Group 4	Elderly, female	1907 (16.5)	1141 (59.8)
Group 5	Elderly, male	2127 (18.4)	1101 (51.7)

<sup>a</sup>The percentage of Sample No. refers to the proportion of the total cohort.

<sup>b</sup>The percentage of DKD No. represents the proportion of DKD samples in the subgroup, i.e., incidence rate.

above 50% in the elderly (groups 4, 5), which was considerably higher than that of the non-elderly men (groups 1, 3; 37%) and non-elderly, non-obesity women (Group 2; 43%) subgroups.

*Model performance*

The Youden index was used to determine the best cut-off point on the receiver operating characteristic curve for each prediction model.<sup>38</sup> In Figure 3, the subgroup models had an average sensitivity of 0.8074, outperforming the global model by 0.1104 on average, with Groups 2 having the largest gains of 0.1554. Although the average precision of subgroup models (0.666) was lower than the global model (0.735), their superior f1-score (average of 0.729 compared to 0.712 for the global model) demonstrates enhanced performance driven by heightened sensitivity. Despite a slightly lower average AUC for subgroup models (0.795) compared to the global model (0.814), it still aligns with expectations, ensuring effective prediction performance after sample partitioning.

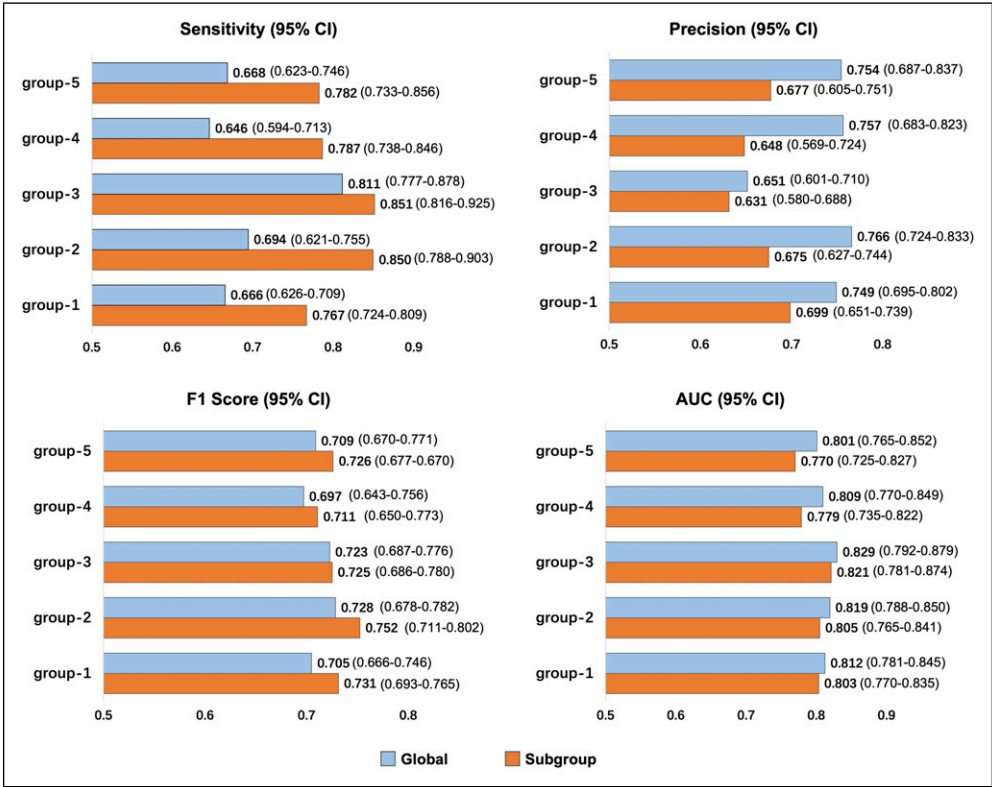


Figure 3. Comparison of predictive measures in the subgroup and global models.

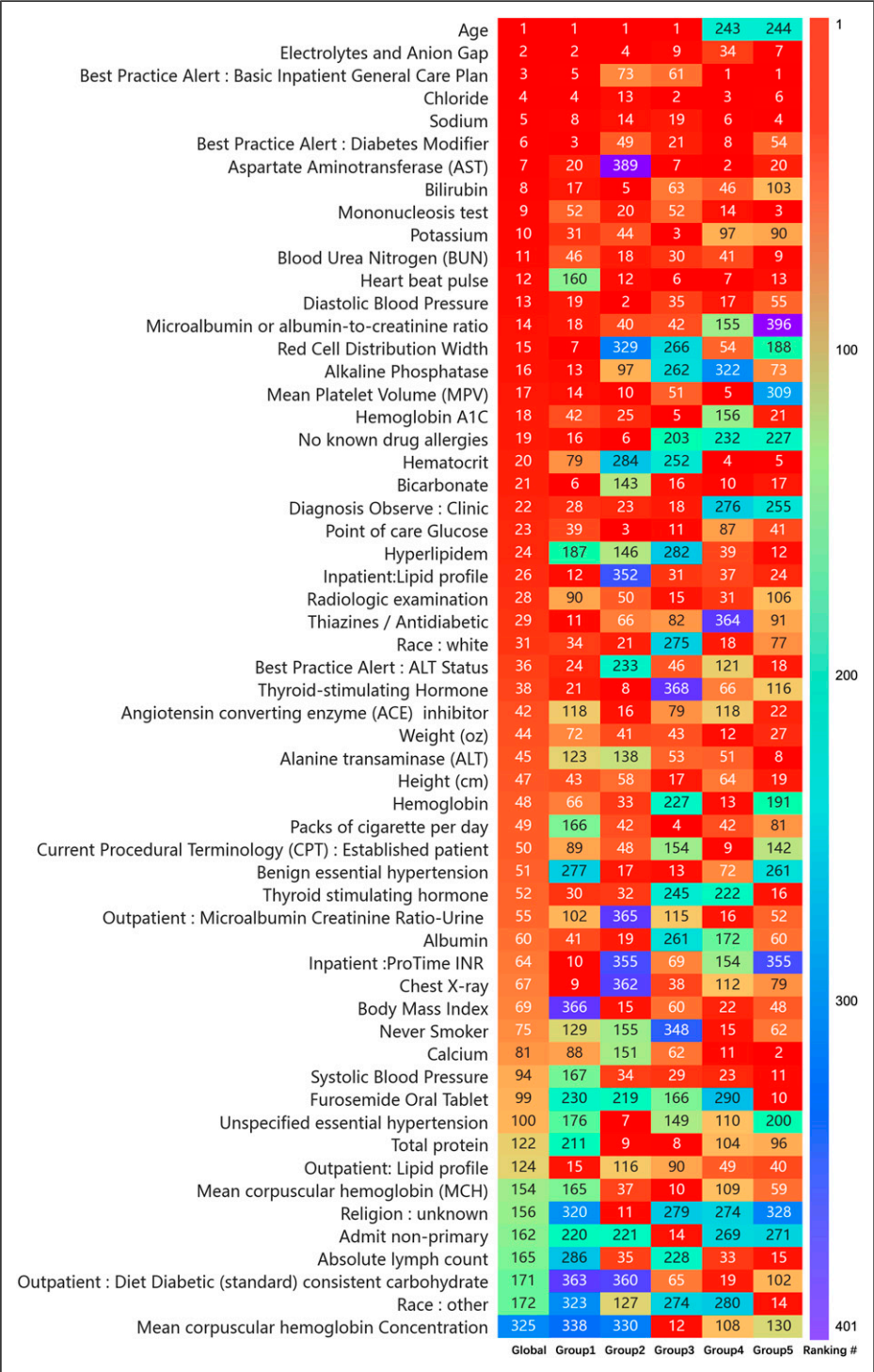
Feature comparison among learned subgroups

In analyzing feature importance output by LightGBM, we assessed the top-ranked features. Figure 4 shows the top 20 features identified in each subgroup model. Deduplication was applied to the top 20 features of each subgroup, resulting in a total of 58 unique features, ordered by their importance in the global model. The value in each color cell (feature) was specific ranking number among all 401 features of its subpopulation, the redder the color, the higher the ranking. Feature rankings were only partially consistent, and some vary greatly, which implies that some important features vital to the general population may have lost dominance in some individuals, while certain features that are more competent in describing the particularity emerged.

Table 3 outlines the top 10 ranked features for each subgroup, revealing universal influences such as age and electrolyte levels. However, subgroup-specific factors like calcium and furosemide oral tablet were critical only in specific demographics, highlighting subgroup-specific importance. The importance of each feature, normalized and visualized in Figure 5, emphasizes that “laboratory test,” “procedure,” and “diagnosis” feature types play crucial roles in predicting DKD risk across all subgroups. Subgroup variations are evident, with certain features being more impactful in specific populations.

Further investigating feature impact, SHAP analysis (Figure 6) reveals diverse influences. For instance, total bilirubin significantly decreases DKD risk for non-elderly women, while calcium has a substantial impact on elderly individuals. Mean corpuscular hemoglobin (MCH) affects various

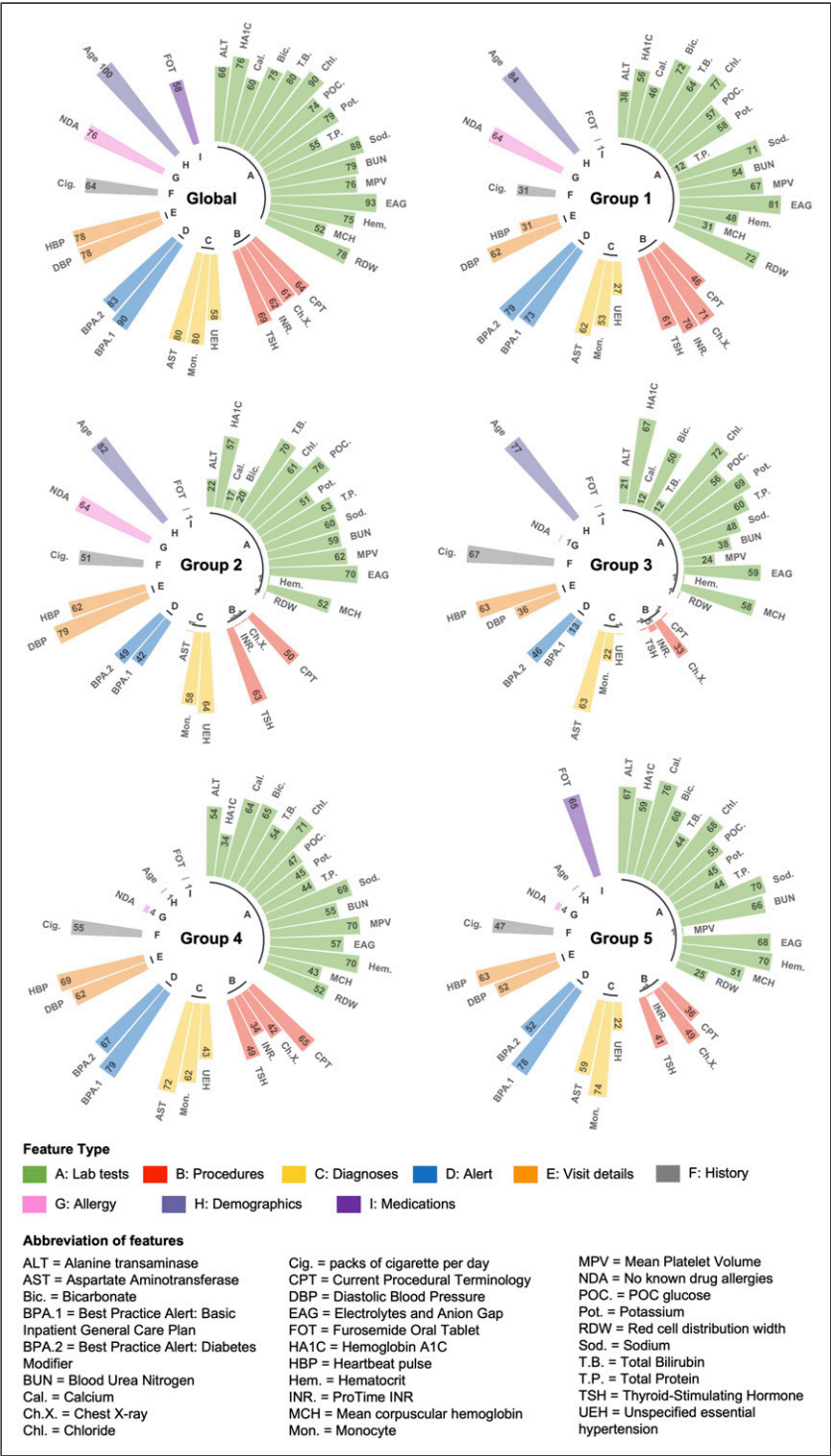




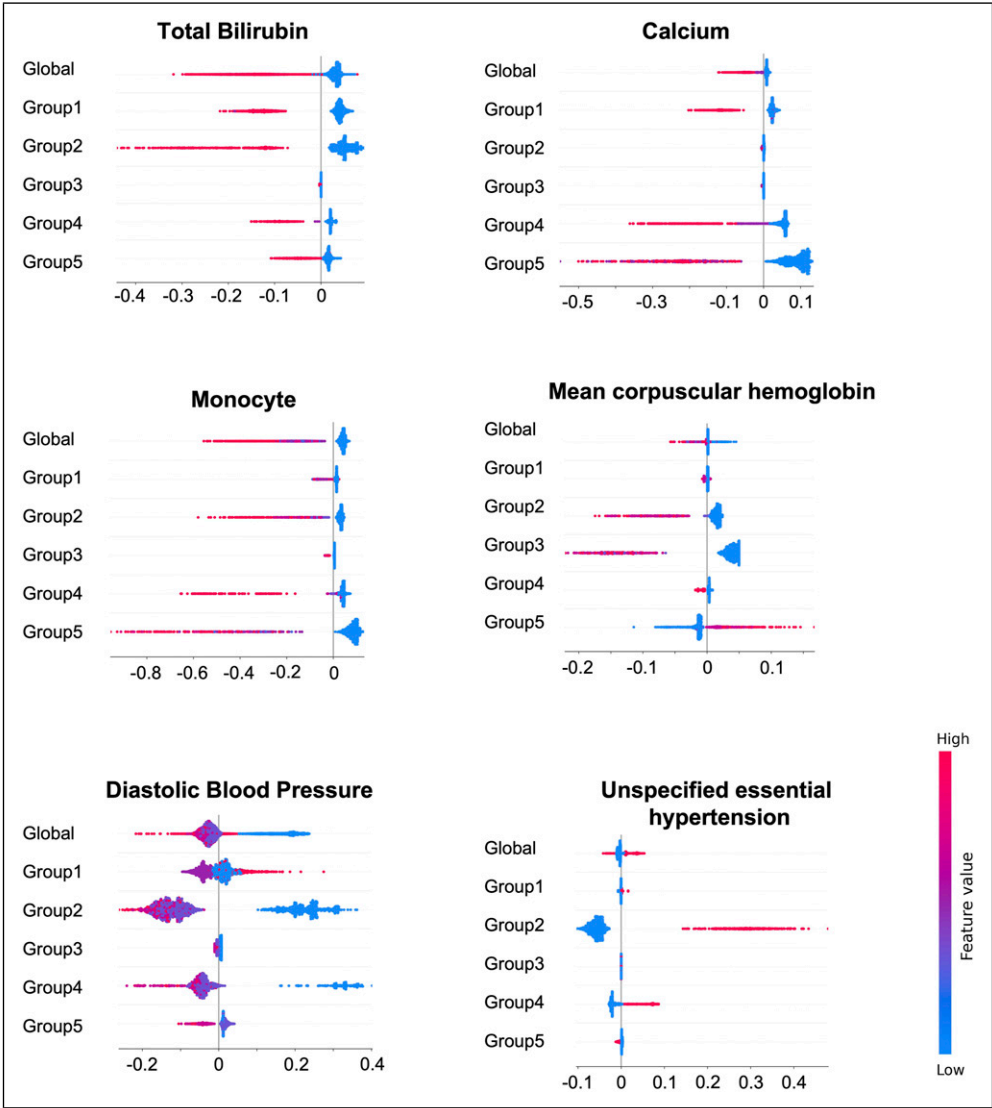
**Figure 4.** Feature ranking heatmap of all subpopulations and cohort. Excerpts from the top 20 features of each subpopulation and global model.

**Table 3.** Top 10 features in the subgroup and global models.

Rank	Group 1	Group 2	Group 3	Group 4	Group 5	Global
1	Age	Age	Age	Best practice alert: Inpatient general care plan	Best practice alert: Basic Inpatient general care plan	Age
2	Electrolytes and Anion Gap	Diastolic blood pressure	Chloride	Aspartate Aminotransferase	Calcium	Electrolytes and Anion Gap
3	Best practice alert: Diabetes Modifier	POC glucose	Potassium	Chloride	Monocyte	Best practice alert: Basic Inpatient general care plan
4	Chloride	Electrolytes and Anion Gap	Packs of cigarette per day	Hematocrit	Sodium	Chloride
5	Best practice alert: Basic Inpatient general care plan	Total bilirubin	Hemoglobin A1C	Mean Platelet Volume	Hematocrit	Sodium
6	Bicarbonate (total CO <sub>2</sub> )	No known drug allergies	Heartbeat pulse	Sodium	Chloride	Best practice alert: Diabetes Modifier
7	Red cell distribution width	Unspecified essential hypertension	Aspartate Aminotransferase	Heartbeat pulse	Electrolytes and Anion Gap	Aspartate Aminotransferase
8	Sodium	Thyroid-Stimulating Hormone	Total Protein	Best practice alert: Diabetes Modifier	Alanine transaminase	Total bilirubin
9	Chest X-ray	Total Protein	Electrolytes and Anion Gap	Current Procedural Terminology	Blood Urea Nitrogen	Monocyte
10	Inpatient: ProTime INR	Mean Platelet Volume	Mean corpuscular hemoglobin	Bicarbonate (total CO <sub>2</sub> )	Furosemide oral tablet	Potassium



**Figure 5.** Top 10 important features comparison of all models. Importance of each feature from the lightGBM model was normalized by  $\log_{10} \times 100$  after, and the length of each bar represents the importance degree of the feature.



**Figure 6.** Comparison of the single feature SHAP values in multiple groups. The horizontal axis is the SHAP value. Each point represents a sample, and the sample size is stacked vertically.

subgroups, particularly men, but not elderly women. DBP and unspecified essential hypertension have greater impacts on female subgroups, particularly non-elderly women. These findings underscore the nuanced influence of features across diverse subpopulations, highlighting potential risk factors for specific subgroups that may not appear equally important in global models.

## Discussion

### *Balancing interpretability and performance*

In clinical settings, data-driven models can achieve impressive performance but often lack transparency and medical interpretability, compromising their clinical utility. Utilizing known medical knowledge can enhance the interpretability of subgroup identification and support clinical acceptance. However, overreliance on existing knowledge can result in biased perspectives, hinder the identification of new trends, and obscure certain relationships within the underlying mechanisms.

To achieve a balance between interpretability and performance, iSLIM offers a structured framework rooted in established DKD risk factors while also allowing for the discovery of novel patterns and risk factors within and across subgroups. It initially employs a white-box method to identify subgroups based on well-known factors like age, hypertension, and BMI, ensuring clinicians can easily understand the subgrouping process and logic. Subsequently, LightGBM is used for predictive modeling, with SHAP analysis enhancing interpretability by revealing feature importance and uncovering complex relationships. This approach not only validates and extends existing medical knowledge but also provides fresh insights into the heterogeneity of DKD.

### *Rational interpretation of the identified subgroups*

The five DKD subgroups were identified based on the factors of age, gender, and obesity. These factors served as the optimal split nodes, as determined by tree-based methods, to classify the heterogeneous cohort into subgroups that achieve the best predictive performance within our framework. From a medical perspective, the heterogeneity of patient subgroups defined by these three factors, particularly in combination, has been reported in many clinical studies.<sup>23,26,39</sup> This provides a rationale for our subgroup learning results. For example, one retrospective study has shown that older age, male, and being overweight are associated with increased creatinine (i.e., decreased kidney function).<sup>40</sup> Tanabe et al. found that patients with severe insulin resistant diabetes who are young and have high BMI is a subgroup with the highest risk of developing DKD,<sup>41</sup> thereby corresponding to our subgroup 1. Another study on renal function prediction has shown that the variability of prediction performance depends on gender, age, and BMI.<sup>42</sup> These findings suggest that the three factors are representative but also reflect the instability of a single risk equation or global model when facing heterogeneous diseases. However, previous research has failed to emphasize the importance of combinatorial impact of the three features, as well as how they may be used to distinguish DKD patients. Our study identified the five most representative subgroups and delivered a rank list of important features, thereby addressing the preceding issues and potentially inspiring further exploration of the DKD mechanisms.

### *Potential risk factors of subgroups*

Improvement in sensitivity within subgroups is driven by unique features that, combined with pathology, guide physicians in exploring mechanisms and identifying risk factors for early alert and prevention. Despite residual confounders, potential risk factors are discussed below.

Lowering *total bilirubin* concentration clinically heightens the risk of renal insufficiency,<sup>43</sup> aligning with our findings that high total bilirubin reduces DKD risk. Our results also showed that Total bilirubin's impact is pronounced in non-elderly females (subgroup 2) but minimal in males of

the same age range (subgroup 3). In prior studies, total bilirubin emerges as a predictor of metabolic syndrome in females,<sup>44</sup> indicating a novel DKD risk factor with strong gender heterogeneity.

*Calcium* has a significant impact on the elderly, especially older men (subgroup 5). Decreased calcium metabolism associated with aging will lead to increased renal burden.<sup>45</sup> In our findings, *furosemide* was one of the top predictors of DKD in older men, and it increased renal calcium excretion.<sup>46,47</sup> The association between calcium and furosemide needs to be further explored in studies of DKD risk in men.

Low *mean corpuscular hemoglobin* (MCH) generally increases the risk of DKD, but an exception exists in our result for older men (subgroup 5), where high MCH becomes a risk factor. Macrocytic anemia may be represented by high MCH.<sup>48</sup> Some studies have explained that the causes of anemia may be gradually complicated by aging, and the number of cases of macrocytic anemia increases in older people.<sup>49</sup> MCH warrants further investigation as a potential risk predictor for DKD in older males.

High-value *unspecified essential hypertension* shows a stronger association with DKD risk in non-elderly women (subgroup 2) than others, underscoring the importance of studying specific types of hypertension in understanding renal disease risk in women, such as pre-eclampsia, which can persist for years after delivery.<sup>50</sup>

Decreased *diastolic blood pressure* (DBP) generally correlates with increased DKD risk,<sup>51</sup> particularly for women (subgroups 2 and 4). An exception is observed for obese individuals (subgroup 1), where increasing DBP renders them likely to develop DKD. In-depth analyses of DBP and obesity are warranted to elucidate their relationship or mechanism with DKD. As many studies have repeatedly confirmed the positive correlation between body weight and blood pressure,<sup>52</sup> and only a few studies have discussed that there are increased intracardiac filling pressures and impaired relaxation of the heart during diastole.<sup>53</sup> Accordingly, we advocate for in-depth analyses of DBP and obesity to elucidate their relationship or mechanism with DKD.

While the above factors may be non-specific and not directly causal, studying them can provide valuable insights into disease pathogenesis, playing significant roles in the etiological network. Our research identifies potential DKD risk factors as informative clues for physicians to decide on further investigation.

## Limitation and future work

Several limitations should be addressed in future work. First, the dataset is derived from a single organization, which may lead clinical doctors to question the model's generalizability and performance across different institutions. Expanding to multi-center data sources is essential to enhance the model's applicability and robustness. Second, ongoing model optimization is necessary, and clinical doctors may anticipate the integration of new technologies, such as advanced deep learning models like explainable neural networks or large-scale language models (LLMs), to further improve prediction accuracy and interpretability. However, implementing these innovations requires fine-tuning and the use of techniques like reinforcement learning from human feedback (RLHF) and direct preference optimization (DPO), despite the associated learning costs. Finally, iSLIM's dependence on expert knowledge underscores the need for continuous updates to incorporate the latest clinical insights. Without timely updates and rigorous clinical validation, the model's practical applicability and reliability could be questioned by clinicians.

## Conclusions

This study introduces the iSLIM framework, which focuses on often-overlooked patient populations and combines knowledge-based and data-driven technologies. Validation with the DKD cohort demonstrates its effectiveness in addressing heterogeneous disease prediction challenges through subgroup strategies. The framework can be integrated with EHR systems and CDSS to enhance machine learning tools in clinical settings, aid doctors in creating personalized prevention plans, and advance digital health systems for managing complex diseases.

## Acknowledgements

Thanks to the Healthcare Enterprise Repository for Ontological Narration (HERON) at the University of Kansas Medical Center for providing the dataset used in this study. We appreciate the HERON Data Request Oversight Committee for their approval and adherence to Health Insurance Portability and Accountability Act (HIPAA) privacy rules.

## Author Contributions

ML, YH, and BL designed and conceptualized the overall study. BL and XH performed model development and conducted the experiments. BL prepared the manuscript. ML, Eric Ngai, KL, XZ, WC, and HYC contributed to the evaluation of experimental results. ML and YH advised regarding motivation of the clinical design and domain for the study. All authors reviewed the manuscript critically for scientific content, and all authors gave final approval of the manuscript for publication.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

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.

## Ethical statement

### *Ethical approval*

The retrospective cohort dataset utilized in this study was derived from the Healthcare Enterprise Repository for Ontology Narratives (HERON) at the University of Kansas Medical Center (KUMC), ensuring adherence to the stringent de-identification standards mandated by the Health Insurance Portability and Accountability Act (HIPAA). The KUMC Institutional Review Board (IRB) diligently reviewed and sanctioned the written consent procedures employed in this study. Given the approval received from the HERON Data Request Oversight Committee and the dataset's HIPAA compliance, IRB approval is not required.



## ORCID iD

Bo Liu  <https://orcid.org/0000-0001-9709-3546>

## References

1. Jha AK, DesRoches CM, Kralovec PD, et al. A progress report on electronic health records in US hospitals. *Health Aff* 2010; 29: 1951–1957.
2. Piri S, Delen D, Liu T, et al. A data analytics approach to building a clinical decision support system for diabetic retinopathy: developing and deploying a model ensemble. *Decis Support Syst* 2017; 101: 12–27.
3. Zolbanin HM and Delen D. Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions. *Decis Support Syst* 2018; 112: 98–110.
4. Berner ES. *Clinical decision support systems*. Springer, 2007.
5. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020; 3: 17–20.
6. Yang C-S, Wei C-P, Yuan C-C, et al. Predicting the length of hospital stay of burn patients: Comparisons of prediction accuracy among different clinical stages. *Decis Support Syst* 2010; 50: 325–335.
7. Ozaydin B, Hardin JM and Chhieng DC. Data mining and clinical decision support systems. In *Clinical decision support systems*. Springer, 2016, pp. 45–68.
8. Chen IY, Szolovits P and Ghassemi M. Can AI help reduce disparities in general medical and mental health care? *AMA J Ethics* 2019; 21: 167–179.
9. Ford CN, Leet RW, Kipling L, et al. Racial differences in performance of HbA1c for the classification of diabetes and prediabetes among US adults of non-Hispanic black and white race. *Diabet Med* 2019; 36: 1234–1242.
10. Atzmueller M. Subgroup discovery. *WIREs Data Min Knowl* 2015; 5: 35–49.
11. Faghri F, Brunn F, Dadu A, et al. Identifying and predicting amyotrophic lateral sclerosis clinical subgroups: a population-based machine-learning study. *Lancet Digit Health* 2022; 4: e359–e369.
12. Dennis JM, Shields BM, Henley WE, et al. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019; 7: 442–451.
13. Ahmad MA, Eckert C and Teredesai A. Interpretable machine learning in healthcare. In: Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics, 2018, pp. 559–560.
14. Gamberger D and Lavrac N. Expert-guided subgroup discovery: methodology and application. *J Artif Intell Res* 2002; 17: 501–527.
15. Gao Y and Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nat Commun* 2020; 11: 5131–5138.
16. Bach M, Werner A, Żywiec J, et al. The study of under-and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis. *Inf Sci* 2017; 384: 174–190.
17. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–357.
18. Radbill B, Murphy B and LeRoith D. Rationale and strategies for early detection and management of diabetic kidney disease. In: Mayo Clinic Proceedings. Elsevier, 2008, pp. 1373–1381.
19. Elise S, Gansevoort RT, Jacqueline B, et al. Will the future lie in multitude? A critical appraisal of biomarker panel studies on prediction of diabetic kidney disease progression. *Nephrol Dial Transplant* 2015; iv96.
20. Krolewski AS, Niewczas MA, Skupien J, et al. Early progressive renal decline precedes the onset of microalbuminuria and its progression to macroalbuminuria. *Diabetes Care* 2014; 37: 226–234.



21. Bhalla V, Zhao B, Azar KM, et al. Racial/ethnic differences in the prevalence of proteinuric and non-proteinuric diabetic kidney disease. *Diabetes Care* 2013; 36: 1215–1221.
22. Guo M, Niu J-Y, Li S-R, et al. Gender differences in the association between hyperuricemia and diabetic kidney disease in community elderly patients. *J Diabetes Complications* 2015; 29: 1042–1049.
23. Russo GT, De Cosmo S, Viazzi F, et al. Diabetic kidney disease in the elderly: prevalence and clinical correlates. *BMC Geriatr* 2018; 18: 38–41.
24. Makino M, Yoshimoto R, Ono M, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci Rep* 2019; 9: 11862–11869.
25. Castillo RS and Kelemen A. Considerations for a successful clinical decision support system. *Comput Inform Nurs* 2013; 31: 319–326.
26. Pavkov M, Knowler W, Bennett P, et al. Increasing incidence of proteinuria and declining incidence of end-stage renal disease in diabetic Pima Indians. *Kidney Int* 2006; 70: 1840–1846.
27. Alonso F, Martínez L, Pérez A, et al. Cooperation between expert knowledge and data mining discovered knowledge: Lessons learned. *Expert Syst Appl* 2012; 39: 7524–7535.
28. MacIsaac RJ, Ekinci EI and Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. *Am J Kidney Dis* 2014; 63: S39–S62.
29. Gheith O, Farouk N, Nampoory N, et al. Diabetic kidney disease: world wide difference of prevalence and risk factors. *J Nephropharmacol* 2016; 5: 49–56.
30. Radcliffe NJ, Seah J, Clarke M, et al. Clinical predictive factors in diabetic kidney disease progression. *J Diabetes Investig* 2017; 8: 6–18.
31. Harjutsalo V and Groop P-H. Epidemiology and risk factors for diabetic kidney disease. *Adv Chronic Kidney Dis* 2014; 21: 260–266.
32. Song X, Waitman LR, Hu Y, et al. Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *J Am Med Inform Assoc* 2019; 26: 242–253.
33. Myles AJ, Feudale RN, Liu Y, et al. An introduction to decision tree modeling. *J Chemom* 2004; 18: 275–285.
34. Safavian SR and Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 1991; 21: 660–674.
35. Pan SJ and Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2009; 22: 1345–1359.
36. Yao Y and Doretto G. Boosting for transfer learning with multiple sources. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010, pp. 1855–1862.
37. Lundberg SM and Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*; 30: 32.
38. Fluss R, Faraggi D and Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J* 2005; 47: 458–472.
39. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. *Clin J Am Soc Nephrol* 2012; 7: 401–408.
40. Mjøen G, Oyen O, Midtvedt K, et al. Age, gender, and body mass index are associated with renal function after kidney donation. *Clin Transplant* 2011; 25: E579–E583.
41. Tanabe H, Saito H, Kudo A, et al. Factors associated with risk of diabetic complications in novel cluster-based diabetes subgroups: a Japanese retrospective cohort study. *J Clin Med* 2020; 9: 2083.
42. Cirillo M, Anastasio P and De Santo NG. Relationship of gender, age, and body mass index to errors in predicted kidney function. *Nephrol Dial Transplant* 2005; 20: 1791–1798.
43. Lee AT, Wang YY, Lin SY, et al. Higher serum total bilirubin concentration is associated with lower risk of renal insufficiency in an adult population. *Int J Clin Exp Med* 2015; 8: 19212–19222.

44. Kwon KM, Kam JH, Kim MY, et al. Inverse association between total bilirubin and metabolic syndrome in rural Korean women. *J Womens Health* 2011; 20: 963–969.
45. Bhattarai HK, Shrestha S, Rokka K, et al. Vitamin D, calcium, Parathyroid Hormone, and Sex Steroids in Bone health and Effects of aging. *J Osteoporos* 2020; 2020: 9324505.
46. Vasco RF, Moyses RM, Zatz R, et al. Furosemide increases the risk of hyperparathyroidism in chronic kidney disease. *Am J Nephrol* 2016; 43: 421–430.
47. Lee C-T, Chen H-C, Lai L-W, et al. Effects of furosemide on renal calcium handling. *Am J Physiol Renal Physiol* 2007; 293: F1231–F1237.
48. Jolobe O. Prevalence of hypochromia (without microcytosis) vs microcytosis (without hypochromia) in iron deficiency. *Clin Lab Haematol* 2000; 22: 79–80.
49. Nagao T and Hirokawa M. Diagnosis and treatment of macrocytic anemias in adults. *J Gen Fam Med* 2017; 18: 200–204.
50. Reddy S and Jim B. Hypertension and Pregnancy: management and future risks. *Adv Chronic Kidney Dis* 2019; 26: 137–145.
51. Rossing P, Hommel E, Smidt UM, et al. Impact of arterial blood pressure and albuminuria on the progression of diabetic nephropathy in IDDM patients. *Diabetes* 1993; 42: 715–719.
52. Rocchini AP. The influence of obesity in hypertension. *Physiology* 1990; 5: 245–249.
53. Kotsis V, Stabouli S, Papakatsika S, et al. Mechanisms of obesity-induced hypertension. *Hypertens Res* 2010; 33: 386–393.