Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Routine Biochemical and Metabolic Indicators Reveal Distinct Diabetes- and Hypertension-Associated Phenotypes in Chronic Renal Failure: A Four-Group Study With Interpretable Machine Learning

Baoye Ye
Fujian University of Traditional Chinese Medicine
Weikun Zhu
Fujian University of Traditional Chinese Medicine
Weirong Shi
The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine
Shuhui Huang
The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine
Zhongfu Xiao
The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine
Xikui Zhang
zhxk1963@163.com

Fujian University of Traditional Chinese Medicine

Article

Keywords: Chronic kidney disease, Metabolic dysregulation, Routine laboratory biomarkers, Machine learning, Model interpretability

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11094958/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Tables 1 to 4 are available in the Supplementary Files section.

Routine Biochemical and Metabolic Indicators Reveal Distinct Diabetes- and Hypertension-Associated Phenotypes in Chronic Renal Failure: A Four-Group Study With Interpretable Machine Learning

Baoye Ye[1],Weikun Zhu[2],Weirong Shi[1],Shuhui Huang[1], Zhongfu Xiao[1],Xikui Zhang[*1,2]

Corresponding author: Xikui Zhang

The Email Address of Xikui Zhang:zhxk1963@163.com

1.The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine, Fuzhou, China

2.Fujian University of Traditional Chinese Medicine, Fuzhou, China

Abstract

Diabetes and hypertension are major comorbidities in chronic renal failure, yet their biochemical phenotypes remain unclear, especially when they coexist. This retrospective cross-sectional study included 267 patients with chronic renal failure, classified as CKD-only, HTN-only, DM-only, and HTN+DM. Routine indicators of protein metabolism, lipid metabolism, renal function, mineral metabolism, and urinary findings were compared among groups. A 2 × 2 factorial model estimated diabetes, hypertension, and interaction effects after adjustment for age, sex, and disease duration. Exploratory machine learning distinguished DM-only from HTN-only patients, with interpretation based on permutation importance and SHAP analysis. The most robust differences involved albumin, albumin/globulin ratio, urea, creatinine, and urine protein, indicating distinct biochemical patterns across subgroups. Factorial analysis showed different

association patterns for diabetes and hypertension, while interaction effects were generally weak. In the DM-only versus HTN-only task, XGBoost performed best, with an AUC of 0.727, followed by CatBoost with an AUC of 0.720. Creatinine, urea, disease duration, phosphorus, and uric acid were the leading explanatory variables. These findings suggest that routine biochemical indicators may provide low-cost clues for differentiating diabetes- and hypertension-associated chronic renal failure phenotypes, although they should not be considered standalone diagnostic tools.

Keywords**:** Chronic kidney disease; Metabolic dysregulation; Routine laboratory biomarkers; Machine learning; Model interpretability

1. Introduction

Chronic renal failure represents an advanced stage of chronic kidney disease characterized by progressive and irreversible loss of kidney function, leading to disturbances in fluid balance, nitrogenous waste excretion, mineral metabolism, and systemic homeostasis. Diabetes and hypertension are among the most common comorbid conditions in patients with chronic renal failure and frequently coexist in clinical practice[1]. Although both conditions contribute to renal function decline, they are associated with different initiating mechanisms, pathological targets, and biochemical consequences. Understanding these differences is clinically important, because patients with similar degrees of renal dysfunction may have heterogeneous metabolic profiles and may require different monitoring and management strategies.

Diabetes-associated kidney injury is primarily driven by chronic hyperglycemia and its downstream metabolic and hemodynamic effects[2]. Hyperglycemia promotes glomerular hyperfiltration, oxidative stress, inflammation, activation of the renin–angiotensin–aldosterone system, and accumulation of advanced glycation end products[3]. These changes can affect glomerular endothelial cells, mesangial cells, podocytes, and the glomerular basement membrane, eventually leading to filtration barrier dysfunction and proteinuria. In addition, diabetes is commonly accompanied by lipid metabolic abnormalities, which may further aggravate endothelial injury, inflammation, and renal fibrosis[4]. Therefore, diabetes-related renal injury is often conceptualized as a glomerular–metabolic phenotype involving proteinuria, dyslipidemia, and progressive renal dysfunction.

In contrast, hypertension-associated kidney injury is more closely related to chronic vascular stress and renal microvascular remodeling[5]. Sustained elevation of systemic blood pressure can induce arteriolar wall thickening, hyaline change, luminal narrowing, and impaired renal autoregulation[6]. These vascular alterations may reduce downstream perfusion, promote glomerular ischemia and sclerosis, and contribute to tubulointerstitial injury. As nephron loss progresses, patients may develop increasing azotemic burden, altered protein status, and disturbances in mineral and uric acid metabolism. Thus, hypertension-related renal injury is often considered a vascular–ischemic phenotype, in which renal function impairment may be closely linked to chronic arteriolar and tubulointerstitial damage.

Despite these mechanistic distinctions, differentiating diabetes-associated and hypertension-associated renal injury in routine clinical settings remains challenging. First, diabetes and hypertension frequently coexist, making it difficult to attribute biochemical abnormalities to one condition alone. Second, many patients with chronic renal failure are evaluated without renal biopsy, detailed longitudinal metabolic data, or complete information on albuminuria, glycemic control, blood pressure burden, and medication use. Third, advanced biomarkers, omics technologies, and pathological confirmation are not always available, particularly in retrospective datasets or resource-limited settings. In contrast, routine biochemical and metabolic indicators, such as albumin, urea, creatinine, uric acid, lipid fractions, calcium, phosphorus, and urine test results, are inexpensive, widely available, and repeatedly measured during clinical care. However, whether these routine indicators can reveal distinct diabetes- and hypertension-associated phenotypes in patients with chronic renal failure has not been fully

clarified.

Most previous studies have focused on diabetic kidney disease or hypertensive nephrosclerosis as separate entities, while fewer studies have systematically examined their combined and independent associations with routine biochemical profiles within the same chronic renal failure population. A four-group framework based on diabetes and hypertension status provides an opportunity to address this issue. By comparing patients without diabetes or hypertension, those with hypertension alone, those with diabetes alone, and those with both conditions, it becomes possible to explore overall phenotypic differences, estimate the independent effects of diabetes and hypertension, and examine whether their coexistence produces interaction effects. Such a design may help move beyond simple disease-label comparisons toward a more structured biochemical phenotyping approach.

In addition to conventional statistical analysis, interpretable machine learning may provide complementary insight by evaluating whether routine indicators contain enough information to distinguish clinically relevant phenotypes. Importantly, in this context, machine learning should not be viewed as a standalone diagnostic tool, but rather as an exploratory method for assessing discriminative patterns and identifying variables that contribute most strongly to phenotype separation. This is particularly relevant for distinguishing diabetes-only from hypertension-only patients, because these two groups represent relatively “pure” clinical phenotypes without the confounding effect of coexistence.

Therefore, this study aimed to characterize routine biochemical and metabolic phenotypes associated with diabetes and hypertension in patients with chronic renal failure. We included 267

patients and classified them into four groups: CKD-only, HTN-only, DM-only, and HTN+DM. First, we compared routine indicators across the four groups to identify global and key biochemical differences. Second, we applied a 2 × 2 factorial model to estimate diabetes effects, hypertension effects, and diabetes–hypertension interaction effects after adjustment for age, sex, and disease duration. Third, we focused on the DM-only versus HTN-only comparison and used exploratory machine learning with permutation importance and SHAP analysis to evaluate discriminative ability and identify key explanatory variables. Through this integrated framework, we sought to determine whether routine biochemical indicators could provide low-cost clues for phenotypic differentiation of diabetes- and hypertension-associated chronic renal failure.

2. Materials and Methods

2.1 Study design and participants

This retrospective cross-sectional study was conducted to investigate whether routine biochemical and metabolic indicators could reveal distinct diabetes- and hypertension-associated phenotypes in patients with chronic renal failure. Clinical and laboratory data were collected from patients diagnosed with chronic renal failure who received medical care at the Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine between January 2020 and January 2025. A total of 267 eligible patients were included in the final analysis.

Patients were eligible for inclusion if they had a clinical diagnosis of chronic renal failure and had available records for diabetes status, hypertension status, demographic information, and routine biochemical or urinary indicators. Patients were excluded if they had incomplete key

grouping information, acute kidney injury, pregnancy, malignant tumor, severe acute infection, or other conditions that could substantially affect biochemical measurements and confound the interpretation of chronic renal failure-related metabolic profiles. For patients with repeated records during the study period, only the first eligible record was retained to avoid duplication.

The study population was subsequently classified according to the presence or absence of diabetes and hypertension. This grouping strategy allowed the comparison of patients without diabetes or hypertension, those with hypertension alone, those with diabetes alone, and those with both conditions. The purpose of this design was not to establish a pathological diagnosis of diabetic nephropathy or hypertensive nephrosclerosis, but rather to characterize diabetes- and hypertension-associated biochemical phenotypes within a chronic renal failure population.

This study was performed in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethics Committee of The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine (approval number: SPHFJP-K2024001-01). Because this was a retrospective study using anonymized clinical data, the requirement for written informed consent was waived by the ethics committee.

2.2 Grouping strategy and clinical data collection

Patients were classified according to the presence or absence of diabetes and hypertension recorded in the clinical medical records. Diabetes status was determined based on a documented clinical diagnosis of diabetes, history of antidiabetic treatment, or relevant diagnostic information recorded by the treating physicians. Hypertension status was determined based on a documented

clinical diagnosis of hypertension, history of antihypertensive treatment, or relevant clinical records. According to these two conditions, the study population was divided into four mutually exclusive groups: patients without diabetes or hypertension were assigned to the CKD-only group; patients with hypertension but without diabetes were assigned to the HTN-only group; patients with diabetes but without hypertension were assigned to the DM-only group; and patients with both hypertension and diabetes were assigned to the HTN+DM group. In the final cohort, 94 patients were classified as CKD-only, 85 as HTN-only, 40 as DM-only, and 48 as HTN+DM.

Demographic and clinical variables were extracted from the electronic medical records, including age, sex, disease duration, diabetes status, and hypertension status. Disease duration was recorded according to the available clinical history and converted into years for subsequent analysis. For patients with ambiguous or non-standard descriptions of disease duration, the recorded information was standardized as far as possible before analysis. Age, sex, and disease duration were considered potential confounding factors because they may influence biochemical indicators and renal function status. Therefore, these variables were included as adjustment covariates in the factorial regression models.

The grouping strategy was designed to support three levels of analysis. First, the four-group framework was used to compare overall biochemical and metabolic profiles among patients with different diabetes and hypertension backgrounds. Second, diabetes and hypertension were coded as two binary factors to estimate their independent and interactive associations with routine biochemical indicators. Third, the DM-only and HTN-only groups were selected for exploratory machine learning analysis, because these two groups represented relatively distinct clinical

phenotypes without overlap between diabetes and hypertension. This approach enabled the study to distinguish broad group differences from the more focused comparison between diabetes-associated and hypertension-associated chronic renal failure phenotypes. The overall study design, grouping strategy, and analytical workflow are summarized in Figure 1.

2.3 Biochemical and urinary measurement

Routine biochemical and urinary indicators were extracted from the laboratory records corresponding to the first eligible clinical encounter for each patient. All measurements were obtained as part of routine clinical care and were performed in the hospital clinical laboratory according to standard operating procedures. To ensure consistency, laboratory data collected during the same hospitalization or clinical evaluation period were used when available.

The included biochemical indicators were organized into five clinically relevant domains. Protein metabolism indicators included total protein (TP), albumin (ALB), globulin (GLB), and the albumin-to-globulin ratio (A/G). Lipid metabolism indicators included triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), apolipoprotein A1 (ApoA1), and apolipoprotein B (ApoB). Renal function-related indicators included urea, creatinine (CREA), and uric acid (UA). Mineral metabolism indicators included calcium (Ca) and phosphorus (P). Urinary findings included urine protein and occult blood.

Urine protein and occult blood were recorded according to routine semiquantitative dipstick results. For statistical analysis and machine learning, these urinary findings were converted into ordinal scores to preserve their graded clinical meaning. The scoring rule was defined as:

$$S = \begin{cases} 0, & negative \\ 0.5, & trace\ or\ \pm \\ 1, & 1+ \\ 2, & 2+ \\ 3, & 3+ \\ 4, & 4+ \end{cases} \quad (1)$$

where S represents the ordinal score assigned to urine protein or occult blood. This coding strategy allowed urinary abnormalities to be analyzed together with continuous biochemical indicators while retaining their ordinal nature.

All variables were checked for missing values and implausible entries before analysis. Continuous biochemical variables were treated as numeric variables, whereas sex, diabetes status, hypertension status, urine protein grade, and occult blood grade were treated as categorical or ordinal variables as appropriate. The selected indicators were intended to reflect major biochemical dimensions of chronic renal failure, including protein status, lipid metabolism, renal function burden, mineral metabolism disturbance, and urinary abnormalities.

2.4 Statistical analysis

Statistical analyses were performed using Python with the pandas, NumPy, SciPy, statsmodels, scikit-learn, XGBoost, LightGBM, CatBoost, matplotlib, and SHAP libraries. Continuous variables were assessed for distribution characteristics before analysis. Because most biochemical indicators showed non-normal distributions, continuous variables were summarized as median and interquartile range (IQR), whereas categorical variables were expressed as counts and percentages.

Comparisons among the four groups were performed using the Kruskal–Wallis test for continuous variables and the chi-square test or Fisher's exact test for categorical variables, as

appropriate. When overall group differences were statistically significant, pairwise post hoc comparisons were subsequently performed. To reduce the risk of false-positive findings caused by multiple testing, false discovery rate (FDR) correction based on the Benjamini-Hochberg method was applied where appropriate. For a set of ordered raw P values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$, the Benjamini-Hochberg adjusted value was calculated as:

$$p_{(i)}^{FDR} = \min_{j \ge i} \left(\frac{m}{j} p_{(j)} \right) \quad (2)$$

where m is the total number of tests. A two-sided P value < 0.05 was considered statistically significant.

To further evaluate the independent associations of diabetes and hypertension with routine biochemical indicators, a 2×2 factorial regression framework was constructed. Diabetes status and hypertension status were treated as two binary factors, and their interaction term was additionally included in the model. Age, sex, and disease duration were incorporated as adjustment covariates because of their potential influence on biochemical profiles and renal function status. For each biochemical indicator, the model was specified as:

$$Y_i = \beta_0 + \beta_1 DM_i + \beta_2 HTN_i + \beta_3 (DM_i \times HTN_i) + \beta_4 Age_i + \beta_5 Sex_i + \beta_6 Duration_i + \varepsilon_i \quad (3)$$

where Y_i represents the biochemical or urinary indicator of patient i, DM_i indicates diabetes status, HTN_i indicates hypertension status, and $DM_i \times HTN_i$ represents the interaction term between diabetes and hypertension. The coefficients β_1, β_2, and β_3 correspond to the diabetes effect, hypertension effect, and interaction effect, respectively.The dependent variable was standardized when presenting effect sizes to improve comparability across indicators. Robust standard errors were used to reduce the influence of heteroscedasticity.

Regression coefficients, 95% confidence intervals, and P values were estimated for diabetes effects, hypertension effects, and interaction effects. Forest plots were generated to visualize the direction and magnitude of associations across different biochemical indicators.

Missing-value patterns were evaluated before analysis. For descriptive statistical analyses, available-case analysis was used. For exploratory machine learning analyses, missing continuous variables were imputed using median values within the modeling pipeline to avoid data leakage. No feature selection based on outcome information was performed before model construction.

2.5 Machine learning analysis and model interpretation

Exploratory machine learning analysis was performed to evaluate whether routine biochemical and metabolic indicators could distinguish patients with diabetes-associated and hypertension-associated chronic renal failure phenotypes. To minimize the confounding influence of coexistence between diabetes and hypertension, the machine learning task was restricted to the DM-only and HTN-only groups. Patients in the HTN-only group were coded as 0, and patients in the DM-only group were coded as 1:

$$Z_i = \begin{cases} 0, & HTN\text{-}only \\ 1, & DM\text{-}only \end{cases} \quad (4)$$

where Z_i denotes the binary classification label of patient i. This task was designed to assess the discriminative information contained in routine clinical indicators between these two relatively distinct phenotypes, rather than to develop a standalone diagnostic model.

All available demographic, biochemical, and urinary variables were used as candidate predictors, including age, sex, disease duration, protein metabolism indicators, lipid metabolism indicators, renal function-related indicators, mineral metabolism indicators, and urinary findings.

Missing values were handled within the modeling pipeline using median imputation to avoid information leakage. Standardization was applied for logistic regression, whereas tree-based models were fitted without feature scaling. No feature selection based on outcome information was performed before model training.

Five commonly used machine learning algorithms were evaluated: logistic regression, random forest, XGBoost, LightGBM, and CatBoost. Model performance was assessed using stratified cross-validation. The evaluation metrics included accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). These metrics were defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{Precision + \text{Recall}} \tag{8}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. ROC curves were generated to compare the discriminative performance of different models, and the model with the highest AUC was selected for subsequent interpretation. To explore the contribution of individual variables to model prediction, permutation importance was calculated for the best-performing model. For each predictor X_j, permutation importance was defined as the decrease in model performance after randomly

permuting that predictor:

$$PI_j = \mathrm{AUC}(M, X, Z) - \frac{1}{B}\sum_{b=1}^{B} AUC\,(M, X^{n_{j,b}}, Z) \tag{9}$$

where PI_j is the permutation importance of feature j, M is the trained model, X is the original predictor matrix, Z is the outcome label, $X^{n_{j,b}}$ represents the dataset in which feature j is randomly permuted in the b-th repetition, and B is the total number of permutation repetitions. A larger PI_j indicates a greater dependence of model performance on that variable.

In addition, SHAP analysis was used to interpret the direction and magnitude of feature contributions at the individual-patient level. For each prediction, the SHAP additive explanation model can be expressed as:

$$f(x_i) = \phi_0 + \sum_{j=1}^{p} \phi_{ij} \tag{10}$$

where $f(x_i)$ is the model output for patient i, ϕ_0 is the baseline prediction, p is the number of predictors, and ϕ_{ij} is the SHAP value representing the contribution of feature j to the prediction for patient i. SHAP summary plots were generated to visualize the relative importance of predictors and their association with classification toward DM-only or HTN-only phenotypes. In this study, machine learning and SHAP analyses were interpreted as exploratory tools for biochemical phenotype differentiation. Model performance and interpretability outputs are presented in Table 4 and Figures 5–7.

3. Results

3.1 Baseline characteristics of the study population

A total of 267 patients with chronic renal failure were included in the final analysis. According to the presence or absence of diabetes and hypertension, 94 patients were classified into the CKD-only group, 85 into the HTN-only group, 40 into the DM-only group, and 48 into the HTN+DM group. The baseline characteristics of the four groups are summarized in Table 1.

The median age differed significantly among the four groups ($P < 0.001$). Patients in the HTN+DM group were the oldest, with a median age of 61.00 years, followed by the DM-only group, HTN-only group, and CKD-only group. Sex distribution also differed significantly across groups ($P = 0.011$). The proportion of male patients was highest in the HTN+DM group, followed by the HTN-only group, whereas the CKD-only and DM-only groups had relatively lower male proportions. In contrast, disease duration did not differ significantly among the four groups ($P = 0.101$).

By definition, patients in the CKD-only group had neither diabetes nor hypertension, those in the HTN-only group had hypertension without diabetes, those in the DM-only group had diabetes without hypertension, and those in the HTN+DM group had both conditions. Because age and sex showed significant differences across groups, they were considered potential confounding factors and were adjusted for, together with disease duration, in the subsequent factorial regression analyses.

3.2 Distinct biochemical and urinary profiles across the four groups

To characterize the overall biochemical phenotype across the four clinical subgroups, standardized group medians of routine biochemical and urinary indicators were visualized using a

heatmap. As shown in Figure 2, the CKD-only, HTN-only, DM-only, and HTN+DM groups displayed distinct biochemical patterns across protein metabolism, lipid metabolism, renal function, mineral metabolism, and urinary findings. These differences suggested that chronic renal failure patients with different diabetes and hypertension backgrounds had heterogeneous routine biochemical profiles rather than a uniform metabolic pattern.

The four-group comparison further identified several indicators with statistically significant differences (Table 2). The most robust differences were observed in albumin (ALB), albumin-to-globulin ratio (A/G), urea, creatinine (CREA), and urine protein, all of which remained significant after false discovery rate correction. These findings indicate that the major biochemical differences across the four groups were concentrated in protein status, renal function burden, and urinary protein abnormality. The distribution of key indicators is illustrated in Figure 3.

In detail, ALB and A/G showed visible differences across groups, with the HTN+DM group generally presenting lower levels, suggesting a more prominent disturbance in protein nutritional status or protein loss-related phenotype. Renal function-related indicators also differed among groups. Urea and CREA tended to be higher in the HTN-only and HTN+DM groups than in the DM-only group, indicating that hypertension-associated phenotypes were accompanied by a stronger azotemic burden in this cohort. Urine protein grades also differed across groups, with higher-grade proteinuria more frequently observed in patients with diabetes or combined diabetes and hypertension.

Lipid-related indicators showed more modest differences. TC and LDL-C tended to be higher

in the DM-only group, suggesting a possible diabetes-associated lipid metabolic pattern. However, these lipid indicators did not remain statistically significant after FDR correction in the overall four-group comparison. Therefore, lipid abnormalities were interpreted as exploratory phenotype signals rather than the most robust group-level differences. Full pairwise comparisons among the four groups are provided in Supplementary Table 2.

Overall, the four-group analysis demonstrated that routine biochemical and urinary indicators could capture clinically meaningful heterogeneity among chronic renal failure patients with different diabetes and hypertension backgrounds. The most stable differences were related to protein status, renal function burden, and urine protein abnormality, while lipid metabolism showed exploratory diabetes-associated trends.

3.3 Adjusted diabetes, hypertension, and interaction effects on biochemical indicators

To further distinguish the independent and joint associations of diabetes and hypertension with routine biochemical indicators, we performed adjusted 2 × 2 factorial regression analyses. Age, sex, and disease duration were included as covariates. The standardized regression coefficients for diabetes effects, hypertension effects, and diabetes × hypertension interaction effects are shown in Figure 4, and the corresponding model estimates are summarized in Table 3. Complete multivariable regression results for all included indicators are provided in Supplementary Table 3.

After adjustment, diabetes showed positive associations with several lipid-related and urinary indicators. The strongest exploratory diabetes-associated signals were observed for TC (β = 0.73, 95% CI: 0.14 to 1.31, P = 0.015), LDL-C (β = 0.59, 95% CI: 0.03 to 1.15, P = 0.040), and urine protein (β = 0.59, 95% CI: 0.07 to 1.12, P = 0.027). However, these associations did not remain

statistically significant after FDR correction. Diabetes was also negatively associated with Urea and CREA in the adjusted model, although these associations were borderline and did not reach statistical significance. These findings suggest that diabetes-related chronic renal failure may show exploratory trends toward lipid metabolic disturbance and urinary protein abnormality, but these signals should be interpreted cautiously.

Hypertension showed a different pattern of association. In the adjusted model, hypertension was negatively associated with ALB (β = −0.38, 95% CI: −0.72 to −0.04, P = 0.029) and showed a borderline negative association with A/G (β = −0.42, 95% CI: −0.84 to 0.00, P = 0.052). Hypertension also showed positive but non-significant associations with Urea, CREA, UA, and urine protein. These results suggest that hypertension-associated phenotypes may be more closely related to altered protein status and renal function burden, although most associations were attenuated after FDR correction.

The diabetes × hypertension interaction terms were generally weak. Exploratory interaction signals were observed for TC (β = −0.75, 95% CI: −1.49 to −0.02, P = 0.043) and LDL-C (β = −0.73, 95% CI: −1.42 to −0.03, P = 0.041), but these interaction effects were no longer significant after FDR correction. No consistent interaction effects were observed for renal function-related indicators or urinary protein. Overall, the factorial analysis indicated that diabetes and hypertension were associated with different biochemical directions across selected indicators, whereas strong interaction effects were not evident.

3.4 Exploratory machine learning for distinguishing DM-only from HTN-only phenotypes

To further evaluate whether routine biochemical and metabolic indicators contained

discriminative information for distinguishing diabetes-associated from hypertension-associated chronic renal failure phenotypes, we performed an exploratory machine learning analysis restricted to the DM-only and HTN-only groups. This task was designed to compare two relatively distinct clinical phenotypes without the confounding effect of diabetes and hypertension coexistence. Five machine learning models were evaluated, including logistic regression, random forest, XGBoost, LightGBM, and CatBoost. Model performance is summarized in Table 4, and the corresponding ROC curves are shown in Figure 5.

Among the five models, XGBoost achieved the highest AUC for distinguishing DM-only from HTN-only patients, with an AUC of 0.727, accuracy of 0.704, precision of 0.556, recall of 0.375, and F1-score of 0.448. CatBoost showed a similar discriminative performance, with an AUC of 0.720 and accuracy of 0.704. Random forest, LightGBM, and logistic regression showed lower AUC values of 0.683, 0.662, and 0.630, respectively. These findings suggest that routine biochemical and metabolic indicators had moderate discriminative ability for separating DM-only and HTN-only phenotypes, although the overall classification performance was not sufficient to support their use as a standalone diagnostic model.

To interpret the best-performing model, permutation importance was calculated for XGBoost. As shown in Figure 6, creatinine (CREA) was the most influential variable, followed by urea, disease duration, phosphorus (P), and uric acid (UA). These results indicate that the classification between DM-only and HTN-only patients was driven mainly by renal function burden, disease course, and mineral or metabolic disturbance, rather than by lipid indicators alone.

SHAP analysis further supported this interpretation (Figure 7). CREA showed the strongest

contribution to model prediction, followed by disease duration, P, urea, UA, age, and selected lipid-related indicators. The SHAP distribution suggested that different levels of renal function-related and metabolic variables contributed in opposite directions toward DM-only or HTN-only classification. Lipid-related variables, including TC, LDL-C, HDL-C, and ApoA1, also contributed to the model, but their overall importance was lower than that of CREA, urea, disease duration, P, and UA. Together, these results suggest that routine indicators can provide exploratory phenotypic clues for distinguishing diabetes-associated and hypertension-associated chronic renal failure, with the strongest signals arising from renal function and disease-burden-related variables.

4. Discussion

4.1 Principal findings

In this retrospective four-group study of 267 patients with chronic renal failure, we found that routinely available biochemical and urinary indicators captured clinically meaningful phenotypic heterogeneity associated with diabetes and hypertension. This finding is important because both diabetes and hypertension are central drivers of CKD progression and kidney failure, and contemporary KDIGO[7] and ADA/KDIGO recommendations[8]emphasize that CKD should be evaluated within the broader context of metabolic, cardiovascular, and kidney-risk profiles rather than by kidney function alone. By classifying patients into CKD-only, HTN-only, DM-only, and HTN+DM groups, our study showed that chronic renal failure patients with different comorbidity backgrounds do not share a completely uniform biochemical profile. The most robust group differences involved ALB, A/G, Urea, CREA, and urine protein, indicating that protein status,

renal excretory burden, and urinary protein abnormality are the most stable routine biochemical dimensions separating these subgroups(Figure 2, Figure 3, and Table 2).

A major observation was that diabetes- and hypertension-associated phenotypes appeared to involve partially different biochemical domains. Diabetes showed exploratory associations with lipid-related indicators and urine protein, whereas hypertension was more closely linked to altered protein status and renal function burden(Figure 4 and Table 3). This pattern is biologically plausible. Diabetic kidney disease is not only a disorder of reduced filtration, but also a metabolic and microvascular disease involving hyperglycemia-driven hemodynamic stress, glomerular hyperfiltration, oxidative stress, inflammation, podocyte injury, mesangial expansion, and filtration barrier dysfunction[9,10]. These processes provide a mechanistic basis for the association between diabetes, proteinuria, and lipid-metabolic abnormalities. In contrast, hypertension-related kidney injury is classically linked to chronic renal vascular stress, arteriolar hyalinosis, luminal narrowing, impaired autoregulation, ischemic glomerulosclerosis, and tubulointerstitial damage, which may explain why renal function burden-related indicators such as Urea and CREA were more prominent in hypertension-associated groups[11,12].

Another important finding was that the diabetes × hypertension interaction effects were generally weak(Figure 4 and Table 3).. This does not mean that coexistence of diabetes and hypertension is clinically unimportant[13]. On the contrary, diabetes and hypertension often coexist and jointly accelerate CKD progression[14]. However, our results suggest that their coexistence may not produce a simple additive or strongly synergistic pattern across routine biochemical indicators. Instead, the HTN+DM group may represent a more heterogeneous clinical

state shaped by age, disease duration, treatment exposure, renal reserve, nutritional status, and unmeasured factors such as glycemic control, blood pressure burden, albuminuria severity, and medication use. This interpretation is consistent with the increasingly recognized heterogeneity of diabetic kidney disease and hypertensive nephrosclerosis in clinical practice, where overlapping risk factors do not always map directly onto distinct pathological lesions.

The exploratory machine learning analysis further supported the presence of a measurable, but limited, biochemical distinction between DM-only and HTN-only phenotypes. XGBoost achieved the highest AUC of 0.727, followed by CatBoost with an AUC of 0.720, indicating moderate discriminative ability rather than strong diagnostic performance(Figure 5 and Table 4). Importantly, model interpretation showed that CREA, Urea, disease duration, P, and UA contributed most strongly to classification. This suggests that the model primarily captured differences in renal impairment burden, disease course, mineral metabolism, and nitrogenous or purine-metabolic disturbance, rather than relying solely on lipid metabolism. Therefore, our results should be interpreted as evidence that routine biochemical indicators provide low-cost phenotypic clues for differentiating diabetes- and hypertension-associated chronic renal failure, not as evidence that these indicators can independently diagnose diabetic nephropathy or hypertensive nephrosclerosis.

4.2 Biological interpretation of diabetes- and hypertension-associated biochemical phenotypes

The biochemical differences observed in this study can be interpreted within two partially overlapping but mechanistically distinct renal injury patterns: a glomerular–metabolic phenotype related to diabetes and a vascular–ischemic phenotype related to hypertension. This interpretation

is consistent with the conceptual framework we adopted at the beginning of the study, in which diabetes-related renal injury is considered to originate primarily from intraglomerular metabolic and filtration-barrier injury, whereas hypertension-related renal injury is considered to begin mainly from renal arteriolar remodeling and downstream ischemic damage. Routine biochemical indicators cannot directly replace kidney biopsy or disease-specific biomarkers, but they may reflect downstream consequences of these different injury pathways.

In diabetes-associated renal injury, chronic hyperglycemia produces both metabolic and hemodynamic stress within the glomerulus. Hyperglycemia-induced glomerular hyperfiltration, activation of inflammatory and oxidative pathways, advanced glycation end-product accumulation, mesangial expansion, glomerular basement membrane thickening, and podocyte injury collectively impair the filtration barrier. This mechanism provides a biological explanation for why urine protein emerged as an important diabetes-related signal in our analysis(Figure 3 and Figure 4). Podocyte injury is particularly relevant because podocytes are terminally differentiated cells that maintain slit diaphragm integrity; once podocyte depletion or dysfunction occurs, albumin leakage and progressive glomerulosclerosis may follow. High-quality reviews of diabetic kidney disease similarly emphasize that DKD is not simply a decline in filtration, but a complex metabolic, inflammatory, hemodynamic, and microvascular disorder involving the glomerular endothelium, mesangium, podocytes, and tubulointerstitium[15,16].

The exploratory lipid-related pattern observed in the DM-only group is also biologically plausible[17]. Diabetes is frequently accompanied by insulin resistance, altered hepatic lipoprotein metabolism, increased triglyceride-rich particles, qualitative changes in LDL, and reduced HDL

function[18,19]. In the kidney, dysregulated lipid metabolism may promote lipotoxicity, mitochondrial dysfunction, oxidative stress, endothelial injury, and inflammation, thereby aggravating glomerular and tubulointerstitial damage[20,21,22,23]. This may explain why TC and LDL-C showed exploratory diabetes-associated signals in the factorial analysis(Figure 4 and Table 3), even though these signals were not the strongest or most stable findings after FDR correction. Therefore, the lipid results should be interpreted as a supportive metabolic pattern rather than definitive evidence that lipid abnormalities are the dominant discriminator between diabetes- and hypertension-associated renal injury.

By contrast, the hypertension-associated phenotype appears more consistent with chronic vascular and ischemic renal injury. Long-standing hypertension exposes renal arterioles to persistent mechanical stress, leading to arteriolar wall thickening, hyaline change, luminal narrowing, impaired autoregulation, and reduced downstream perfusion[24]. These vascular lesions can induce ischemic glomerulosclerosis, nephron loss, and tubulointerstitial fibrosis. This mechanism is compatible with the relatively higher Urea and CREA burden observed in the HTN-only and HTN+DM groups(Figure 3 and Table 2), because these markers primarily reflect reduced renal excretory capacity and accumulated nitrogenous waste. The association between hypertension and altered ALB or A/G may also reflect the combined effects of chronic inflammation, protein-energy wasting, vascular injury, and kidney disease severity rather than a single disease-specific pathway.

The prominence of CREA, Urea, UA, and P in both statistical and machine-learning analyses further suggests that the distinction between DM-only and HTN-only patients is not purely

etiological, but partly reflects differences in renal functional burden and systemic metabolic consequences of chronic renal failure. Phosphorus is closely linked to CKD-mineral and bone disorder, fibroblast growth factor 23 activation, secondary hyperparathyroidism, vascular calcification, and cardiovascular risk. Similarly, UA reflects purine metabolism and renal excretory function, and hyperuricemia often coexists with CKD, hypertension, insulin resistance, and vascular dysfunction, although whether UA is a causal driver or mainly a risk marker remains debated. Therefore, the contribution of P and UA should be interpreted as evidence of broader mineral-metabolic and waste-retention phenotypes rather than as disease-specific signatures of diabetes or hypertension alone.

Another important biological implication is that the HTN+DM group should not be viewed as a simple sum of the DM-only and HTN-only phenotypes. In clinical practice, diabetes and hypertension often interact through shared pathways, including endothelial dysfunction, renin–angiotensin–aldosterone system activation, oxidative stress, inflammation, arterial stiffness, sodium retention, and progressive nephron loss[25,26,27]. However, once chronic renal failure is established, routine biochemical indicators may become dominated by the severity of kidney dysfunction, nutritional status, treatment history, and residual renal reserve. This may partly explain why the interaction terms in our factorial analysis were generally weak. The coexistence of diabetes and hypertension is undoubtedly clinically important, but its biochemical expression in routine laboratory indicators may be heterogeneous and non-linear rather than strongly additive.

Taken together, these findings support a cautious but clinically meaningful interpretation. Diabetes-associated chronic renal failure in this cohort tended to show a glomerular–metabolic

signal, reflected by urine protein and exploratory lipid-related changes, whereas hypertension-associated chronic renal failure tended to show a vascular–ischemic and renal-burden signal, reflected by Urea, CREA, UA, and protein-status alterations. Nevertheless, these phenotypes are not mutually exclusive. In advanced chronic renal failure, multiple pathways converge, and routine indicators may reflect both disease mechanisms and downstream severity. Therefore, the value of these indicators lies not in making a definitive etiological diagnosis, but in providing accessible phenotypic clues that can help describe clinical heterogeneity and guide more detailed future evaluation using eGFR, UACR, glycemic control, blood pressure burden, medication exposure, and kidney pathology.

4.3 Interpretation of the factorial model and machine learning findings

The 2 × 2 factorial model provided a useful framework for moving beyond simple four-group comparisons. By treating diabetes and hypertension as two binary factors, the model separated three clinically meaningful components: the diabetes-associated effect, the hypertension-associated effect, and the diabetes × hypertension interaction effect. This approach is particularly relevant in chronic renal failure, where diabetes and hypertension frequently coexist and where biochemical abnormalities may reflect shared downstream kidney dysfunction rather than a single etiological pathway. Therefore, the factorial model should be interpreted as an association-based phenotyping tool rather than a causal model. The weak interaction effects observed in our study suggest that coexistence of diabetes and hypertension did not produce a consistent synergistic pattern across routine biochemical indicators(Figure 4 and Table 3); instead, the combined group likely reflects clinical heterogeneity arising from disease duration, age,

524 residual renal function, treatment exposure, and unmeasured metabolic control.

525 The machine learning findings complement, rather than replace, the factorial analysis. The
526 DM-only versus HTN-only task was intentionally restricted to two relatively “pure” phenotypes to
527 reduce the confounding effect of diabetes–hypertension coexistence. XGBoost achieved the
528 highest AUC of 0.727, followed by CatBoost with an AUC of 0.720, indicating moderate
529 discrimination(Figure 5 and Table 4). In clinical prediction research, however, model performance
530 should not be judged by AUC alone. Transparent reporting guidelines such as TRIPOD+AI
531 emphasize that prediction models, especially those using machine learning, should be interpreted
532 in relation to study design, sample size, validation strategy, intended use, and clinical
533 consequences rather than discrimination metrics alone[28]. In our study, the relatively modest
534 recall and F1-score indicate that the model is not suitable as a standalone classifier for clinical
535 diagnosis, but it is useful for exploring whether routine indicators contain phenotype-related
536 information.

537 The feature-importance and SHAP analyses provide the most informative biological
538 interpretation of the machine learning results. CREA and Urea dominated the model, followed by
539 disease duration, P, and UA(Figure 6 and Figure 7). This pattern suggests that the model primarily
540 captured differences in renal functional burden, disease course, mineral metabolism, and
541 nitrogenous or purine-metabolic retention. This is clinically important because it indicates that the
542 machine learning model may be distinguishing the severity and systemic consequences of chronic
543 renal failure as much as the underlying diabetes- or hypertension-associated phenotype. In other
544 words, the model’s discriminative signal should not be overinterpreted as direct evidence of

disease-specific etiology.

SHAP analysis was valuable because it allowed feature contributions to be examined at the individual-patient level rather than only at the global model level. SHAP is based on Shapley values and assigns each feature a contribution value for a given prediction, making complex models more transparent and clinically interpretable. Nevertheless, SHAP does not establish causality; it explains how a trained model used available variables to make predictions. Therefore, in this study, SHAP results should be interpreted as an explanatory map of model behavior. The fact that lipid-related variables contributed but were not dominant is consistent with our statistical results: lipid metabolism may provide auxiliary diabetes-associated information, whereas renal function and disease-burden variables are the principal discriminative signals.

Overall, the factorial model and interpretable machine learning analyses converge on the same cautious conclusion. Routine biochemical indicators can reveal meaningful phenotype-related patterns in chronic renal failure, but these patterns are not sufficiently specific to define diabetic nephropathy or hypertensive nephrosclerosis. This interpretation is also consistent with concerns raised in prediction-model methodology literature, where insufficient validation, small datasets, and unclear clinical use are common limitations of medical machine learning studies. Thus, the value of the present machine learning analysis lies in hypothesis generation and phenotype exploration. Future studies should integrate eGFR, UACR, HbA1c, longitudinal blood pressure burden, medication exposure, and kidney pathology to determine whether these routine biochemical signals can be translated into more robust clinical phenotyping models.

4.4 Clinical implications, strengths, and limitations

The present study has several clinical implications. First, it suggests that routine biochemical and urinary indicators, although non-specific, may provide accessible clues for describing phenotypic heterogeneity among patients with chronic renal failure. In daily clinical practice, patients with chronic renal failure are often evaluated using standard laboratory panels rather than kidney biopsy or advanced biomarkers[29]. Our findings indicate that these routine indicators may help clinicians recognize whether a patient's biochemical profile is more consistent with a diabetes-associated metabolic–glomerular pattern, a hypertension-associated vascular–renal burden pattern, or a mixed phenotype. Such information may be useful for guiding closer monitoring of protein status, renal function burden, mineral metabolism, and urinary abnormalities.

Second, the results support a phenotype-oriented interpretation of chronic renal failure rather than a purely diagnosis-label-based approach. Diabetes and hypertension are frequently treated as binary comorbidities, but their biochemical consequences are not always linear or disease-specific. The relatively weak interaction effects and the moderate machine learning performance suggest that routine indicators can support clinical phenotyping, but cannot replace disease-specific evaluation. Therefore, these findings should be used to generate clinical hypotheses and guide further assessment, rather than to make definitive etiological diagnoses. In particular, patients showing prominent renal function burden, mineral metabolism disturbance, or worsening proteinuria may require more detailed evaluation of eGFR, UACR, glycemic control, blood pressure burden, medication exposure, and, when clinically indicated, pathological information.

This study also has several strengths. First, the four-group design allowed us to compare CKD-only, HTN-only, DM-only, and HTN+DM patients within the same chronic renal failure population, reducing the limitations of simple binary comparisons. Second, the 2 × 2 factorial model provided a structured framework to separate diabetes-associated effects, hypertension-associated effects, and their interaction effects after adjustment for age, sex, and disease duration. Third, the combination of conventional statistics and interpretable machine learning improved the analytical depth of the study. Instead of reporting only model performance, permutation importance and SHAP analysis were used to identify the variables that contributed most strongly to phenotype discrimination, thereby improving the clinical interpretability of the findings.

Several limitations should be acknowledged. First, this was a retrospective cross-sectional study, and therefore causal relationships between diabetes, hypertension, biochemical indicators, and renal injury phenotypes cannot be inferred. Second, the study population was relatively small, especially the DM-only group, and the findings require validation in larger, multicenter cohorts. Third, several clinically important variables were unavailable or incomplete, including eGFR, UACR, HbA1c, longitudinal blood pressure levels, medication use, disease duration precision, and kidney biopsy findings. The absence of these variables limited our ability to distinguish true disease-specific mechanisms from differences in renal impairment severity. Fourth, diabetes-associated and hypertension-associated phenotypes were defined according to clinical comorbidity status, not by pathological confirmation of diabetic nephropathy or hypertensive nephrosclerosis. Fifth, the machine learning analysis showed only moderate discrimination, and

the model was not externally validated. Thus, it should be interpreted as an exploratory phenotyping tool rather than a clinical diagnostic model. Finally, the strong contribution of CREA and Urea suggests that part of the model signal may reflect renal function burden rather than etiology-specific injury mechanisms.

Future studies should incorporate prospective data, larger sample sizes, longitudinal follow-up, standardized kidney function measures, albuminuria quantification, glycemic and blood pressure control indicators, treatment information, and pathological validation where possible. Such studies may clarify whether routine biochemical indicators can be integrated with clinical and pathological variables to build more robust phenotyping models for diabetes- and hypertension-associated chronic renal failure.

5. Conclusions

This retrospective four-group study suggests that routine biochemical and urinary indicators can reflect distinct phenotypic patterns associated with diabetes and hypertension in patients with chronic renal failure. The observed differences were most consistently related to protein status, renal function burden, and urinary protein abnormality. These findings are biologically compatible with the concept that diabetes-associated renal injury is more closely linked to glomerular–metabolic dysfunction and filtration barrier injury, whereas hypertension-associated renal injury is more closely related to vascular–ischemic damage and renal excretory burden.

Exploratory machine learning further showed moderate ability to distinguish DM-only from HTN-only patients, with CREA, Urea, disease duration, P, and UA contributing most strongly to classification. These results indicate that routine indicators may provide low-cost clues for

differentiating diabetes- and hypertension-associated chronic renal failure phenotypes. However, they should be interpreted as supportive phenotypic signals rather than standalone diagnostic tools for diabetic nephropathy or hypertensive nephrosclerosis. Further prospective and multicenter studies incorporating eGFR, UACR, glycemic control, blood pressure burden, medication exposure, and pathological evidence are needed to validate these findings.

Data availability statement

All data supporting the findings of this study are available within the paper and its Supplementary Information.

Ethics statement

The study protocol was reviewed and approved by the Ethics Committee of [The Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine] (approval number: SPHFJP-K2024001-01). Because this was a retrospective study using anonymized clinical data, the requirement for written informed consent was waived by the ethics committee.

Author contributions

Xikui Zhang is the corresponding author.He completed the project design and all patients came from his outpatient clinic.All authors contributed to the article and approved the submitted version.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This project is supported by Special Funds for Education and Scientific Research of Directly Affiliated Units of Fujian Province (Project No.: X2023009 and Project No.:X2024001). The research plan has been approved by the Ethics Committee of the Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine (Approval No.: SPHFJP-K2024001-01), and informed consent exemption has been obtained.

References

[1]Qasim H, Ktaifan M, Awawdeh A, Khattab K, Leoni MLG, Varrassi G. Hypertensive Nephrosclerosis: Pathological Changes and Overlap with Diabetic Nephropathy. Cureus. 2025 Sep 22;17(9):e92949.

[2]Martinez Leon V, Hilburg R, Susztak K. Mechanisms of diabetic kidney disease and established and emerging treatments. Nat Rev Endocrinol. 2026 Jan;22(1):21-35.

[3]Zatz R, Brenner BM. Pathogenesis of diabetic microangiopathy. The hemodynamic view. Am J Med. 1986 Mar;80(3):443-53.

[4]Moorhead JF, Chan MK, El-Nahas M, Varghese Z. Lipid nephrotoxicity in chronic progressive glomerular and tubulo-interstitial disease. Lancet. 1982 Dec 11;2(8311):1309-11.

[5]Kohagura K, Zamami R, Oshiro N, Shinzato Y, Uesugi N. Heterogeneous afferent

667 arteriolopathy: a key concept for understanding blood pressure-dependent renal damage.
668 Hypertens Res. 2024 Dec;47(12):3383-3396.
669 [6]Tylicki L, Rutkowski B. Nefropatia nadciśnieniowa--patogeneza, diagnostyka i leczenie
670 [Hypertensive nephropathy: pathogenesis, diagnosis and treatment]. Pol Merkur Lekarski. 2003
671 Feb;14(80):168-73.
672 [7]Group, K.D.I.G., KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management
673 of Chronic Kidney Disease. Kidney Int, 2024. 105(4S): p. S117-S314.
674 [8]de Boer IH, R.P.C.M., KDIGO 2022 Clinical Practice Guideline for Diabetes Management in
675 Chronic Kidney Disease. Kidney Int, 2022. 102(5S): p. S1-S127.
676 [9]Garoufis M, Sakkou SF, Kostara CE, Bairaktari E, Tsimihodimos V. Metabolomics for
677 Preclinical Detection of Diabetic Kidney Disease: A Comprehensive Review. Int J Mol Sci. 2026
678 Jan 19;27(2):998.
679 [10]Patel HA, Wang J, Zinn CJ, Learmonth M, Bian X, Jurk D, Lerman LO, Wolfram J, Hickson
680 LJ. Targeting cellular senescence in diabetic kidney disease: potential of regenerative, cell-based
681 therapies and other senotherapeutic approaches. Kidney Int. 2026 Apr;109(4):684-703.
682 [11]Al Rashdi F, Tabche C, Atwan Z, Al-Qanubi H, Al Khaldi S, Al-Zadjali N, Rawaf S. Chronic
683 kidney disease in hypertensive patients: the urgent need for targeted interventions in Arab
684 countries: a systematic review. Front Nephrol. 2026 Feb 9;6:1735217.
685 [12]Mahfoud F, Götzinger F, Kramann R, Marx N, Schwenger V. The Synergistic Treatment of
686 Heart and Kidney Disease. Dtsch Arztebl Int. 2026 Jan 1;123(1):19-26.
687 [13]Pascat V, Zudina L, Maurin L, Ulrich A, Maina JG, Demirkan A, Balkhiyarova Z, Pupko I,

Sharhorodska Y, Pattou F, Staels B, Kaakinen M, Khamis A, Bonnefond A, Munroe P, Froguel P, Prokopenko I. Partitioned polygenic scores show mechanistic heterogeneity in type 2 diabetes and hypertension comorbidity. Nat Commun. 2026 Feb 9;17(1):1446.

[14]Agrawal M, Kumar A, Singh AK, Singh H, Kumar P, Kumar P. Understanding the Molecular Mechanisms of Diabetes mellitus, Hypertension, Cardiovascular Disease, and Chronic Kidney Disease: Clinical Insights and Future Perspectives. Curr Hypertens Rev. 2026 Jan 6.

[15]Martinez Leon V, Hilburg R, Susztak K. Mechanisms of diabetic kidney disease and established and emerging treatments. Nat Rev Endocrinol. 2026 Jan;22(1):21-35.

[16]Yu S, Han Z, Li C, Lu X, Li Y, Yuan X, Guo D. Cross Talk Between Macrophages and Podocytes in Diabetic Nephropathy: Potential Mechanisms and Novel Therapeutics. Mediators Inflamm. 2025 May 2;2025:8140479.

[17]Tu QM, Jin HM, Yang XH. Lipid abnormality in diabetic kidney disease and potential treatment advancements. Front Endocrinol (Lausanne). 2025 Mar 18;16:1503711.

[18]Hirano T. Clinical significance of small dense low-density lipoprotein cholesterol measurement in type 2 diabetes. J Diabetes Investig. 2025 Mar;16(3):370-383.

[19]Khan AW, Jandeleit-Dahm KAM. Atherosclerosis in diabetes mellitus: novel mechanisms and mechanism-based therapeutic approaches. Nat Rev Cardiol. 2025 Jul;22(7):482-496. =-14

[20]Hughes E, Wang XX, Sabol L, Barton K, Hegde S, Myakala K, Krawczyk E, Rosenberg A, Levi M. Role of nuclear receptors, lipid metabolism, and mitochondrial function in the pathogenesis of diabetic kidney disease. Am J Physiol Renal Physiol. 2025 Oct 1;329(4):F510-F547.

[21]Giardini E, Moore D, Sadlier D, Godson C, Brennan E. The dual role of lipids in chronic kidney disease: Pathogenic culprits and therapeutic allies. Atherosclerosis. 2024 Nov;398:118615.

[22]Apte M, Kumthekar G, Tupe RS. Connecting glycation and lipotoxicity to mitochondrial dysfunction in diabetic kidney disease: A 'tubulocentric' perspective. Biochem Biophys Res Commun. 2026 Apr 16;809:153479.

[23]Al-Masri S, Coelho JN, Thomas L. Molecular pathways and emerging therapeutic targets in the pathogenesis of diabetic kidney disease. Front Physiol. 2026 Feb 19;17:1747053.

[24]Kitada K. Hypertension research 2024 update and perspectives: basic research. Hypertens Res. 2024 Dec;47(12):3304-3309.

[25]Verma S, Ansari IS, Gangwar V. The Vicious Synergy: Intertwined Pathophysiology of Hypertension and Type 2 Diabetes Mellitus. Indian Journal of Cardiovascular Diseases in Diabetes and Metabolic Disorders. 2025;1(1):31-35.

[26]Wang N, Kong J, Lu Z, Cao F, Tian E, Li J, Li S, Liu S, Yue B. Effects of SGLT-2 inhibitors on renin-angiotensin-aldosterone system and their correlation with glucose metabolism in type 2 diabetes mellitus patients with hypertension: A prospective study. PLoS One. 2025 Nov 18;20(11):e0336158.

[27]Adeshara K, Lithovius R, Mutter S, Harjutsalo V, Lehto M, Groop PH, Sandholm N; FinnDiane Study Group. Soluble RAGE further stratifies risk of coronary artery and end-stage kidney disease in high-risk individuals with type 1 diabetes and treatment-resistant hypertension. Cardiovasc Diabetol. 2025 Dec 12;25(1):13.

[28]Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for

730 reporting clinical prediction models that use regression or machine learning methods. BMJ.
731 2024;385:e078378.
732 [29]Wu Y, Ying Q, He Y, Xie X, Yuan X, Wang M, Fei X, Yang X. Effect of external therapies of
733 traditional Chinese medicine on constipation in patients with CRF: A meta-analysis. PLoS One.
734 2023 Oct 5;18(10):e0291968.

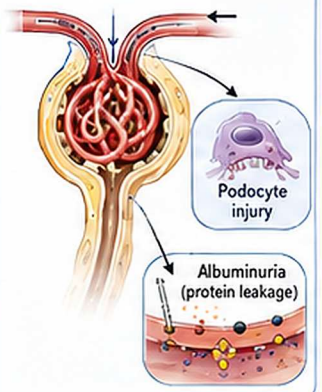
Routine Biochemical and Metabolic Indicators Reveal Distinct Diabetes- and Hypertension-Associated Phenotypes in Chronic Renal Failure

Four-group analysis, 2 × 2 factorial modeling, and interpretable machine learning

A Pathophysiological mechanisms

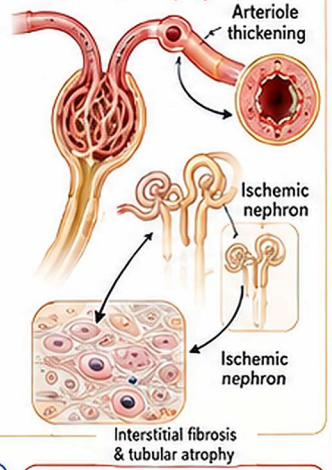
1) Diabetes-associated renal injury

- Hyperglycemia
- Glomerular hyperfiltration
- Mesangial expansion
- Advanced glycation end products (AGEs)
- Podocyte injury
- Albuminuria (protein leakage)
- Dyslipidemia tendency



2) Hypertension-associated renal injury

- Blood pressure elevation
- Arteriolar thickening / hyaline change
- Narrowed lumen
- Renal ischemia
- Glomerulosclerosis
- Tubulointerstitial injury / fibrosis
- Impaired concentrating function
- Renal function burden



Diabetes-related phenotype:
Albuminuria + lipid-metabolic disturbance tendency

Hypertension-related phenotype:
Renal ischemic injury + stronger azotemic burden

B Study design and analytical framework

267 patients with chronic renal failure

CKD-only	HTN-only	DM-only	HTN+DM
n = 94	n = 85	n = 40	n = 48

Routine biochemical and metabolic indicators

Protein metabolism	TP, ALB, GLB, A/G
Lipid metabolism	TG, TC, HDL-C, LDL-C, ApoA1, ApoB
Renal function	Urea, CREA, UA
Mineral metabolism	Ca, P
Urinary findings	Urine protein, occult blood

1 Four-group comparison
(ANOVA/Kruskal–Wallis with post hoc tests)

2 Adjusted 2 × 2 factorial model
(Diabetes, Hypertension, Diabetes × Hypertension)
Adjusted for age, sex, disease duration

Diabetes \ Hypertension	No	Yes
No	●	●
Yes	●	●

3 Exploratory machine learning
DM-only vs HTN-only
(Model training and evaluation)

C Main biochemical phenotypes

Distinct biochemical phenotypes exist across four CRF subgroups.

1 Robust group differences

	CKD-only	HTN-only	DM-only	HTN+DM
ALB	↓	↓	↓	↓↓
A/G	↓	↓	↓	↓↓
Urea	↑	↑↑	↑	↑↑
CREA	↑	↑↑	↑	↑↑
Urine protein	↑	↑	↑↑	↑↑

↑↑ Markedly higher ↑ Higher ↓ Lower ↓↓ Markedly lower

2 Diabetes-associated pattern
(exploratory trend)

- TC ↑
- LDL-C ↑
- Urine protein ↑



3 Hypertension-associated pattern

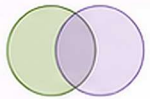
- Urea ↑↑
- CREA ↑↑
- UA ↑↑
- ALB ↓
- A/G ↓



4 Interaction effects (Diabetes × Hypertension)

Most indicators: Weak / not significant

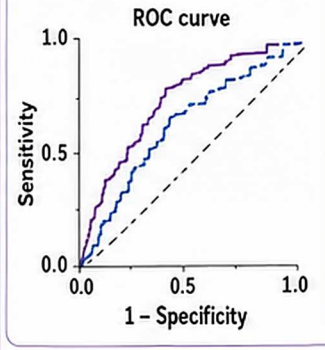
Lipids (e.g., TC, LDL-C, ApoB): Exploratory interaction signals



Phenotypic differentiation is primarily driven by renal function burden and proteinuria, with lipid alterations mainly linked to diabetes.

D Interpretable machine learning

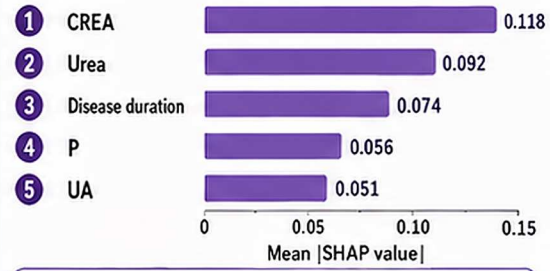
Task: DM-only vs HTN-only



Best model: **XGBoost**
AUC = 0.727
CatBoost
AUC = 0.720

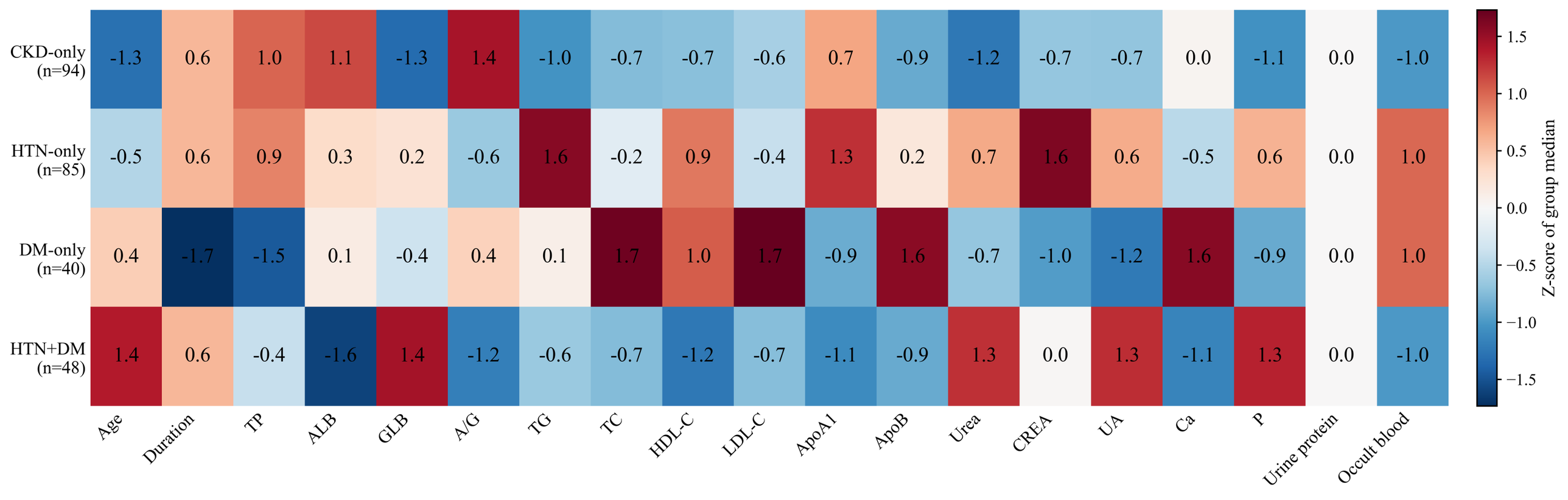
Interpretation: moderate discriminative ability

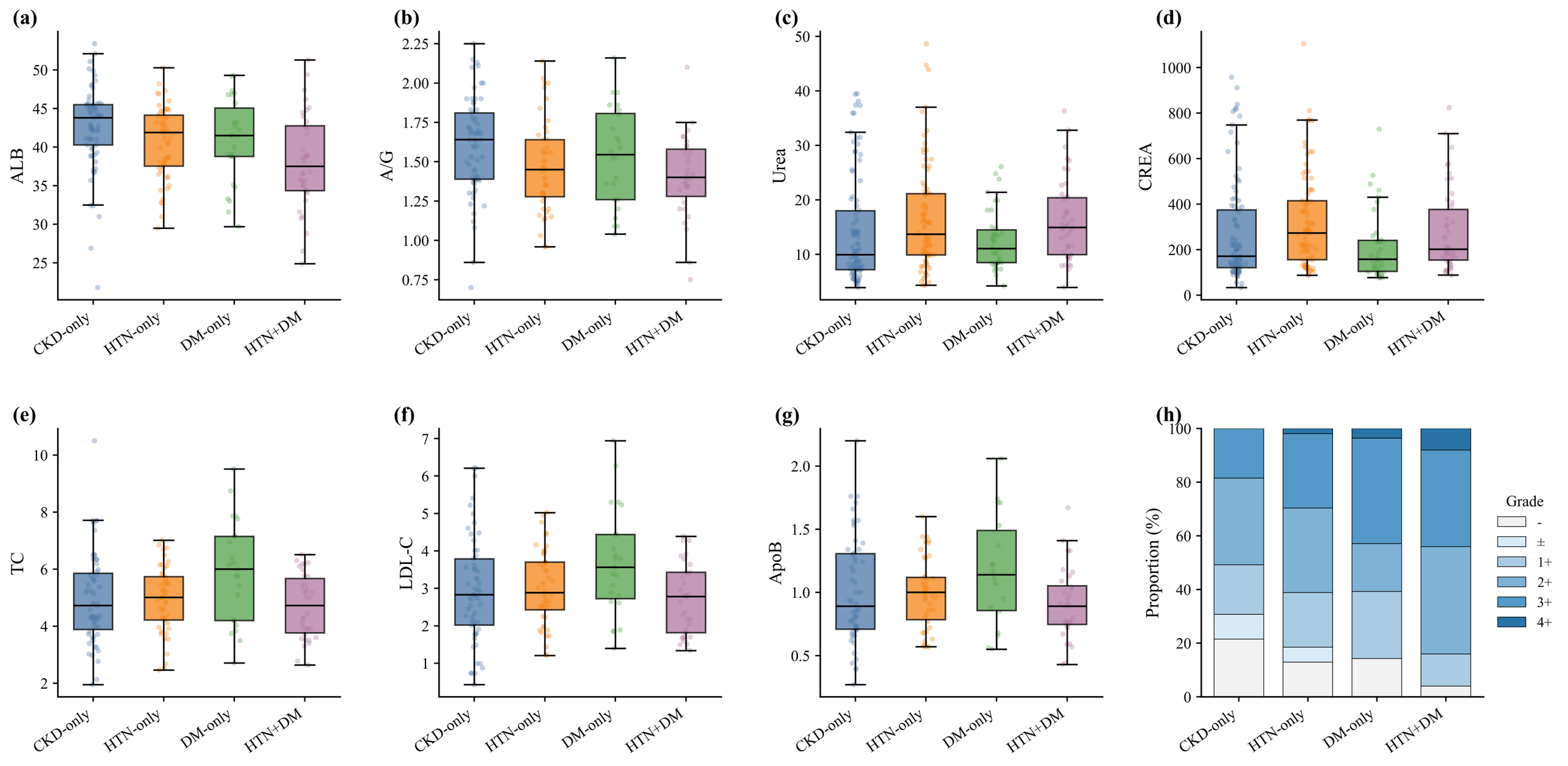
Top features (SHAP mean |value|)

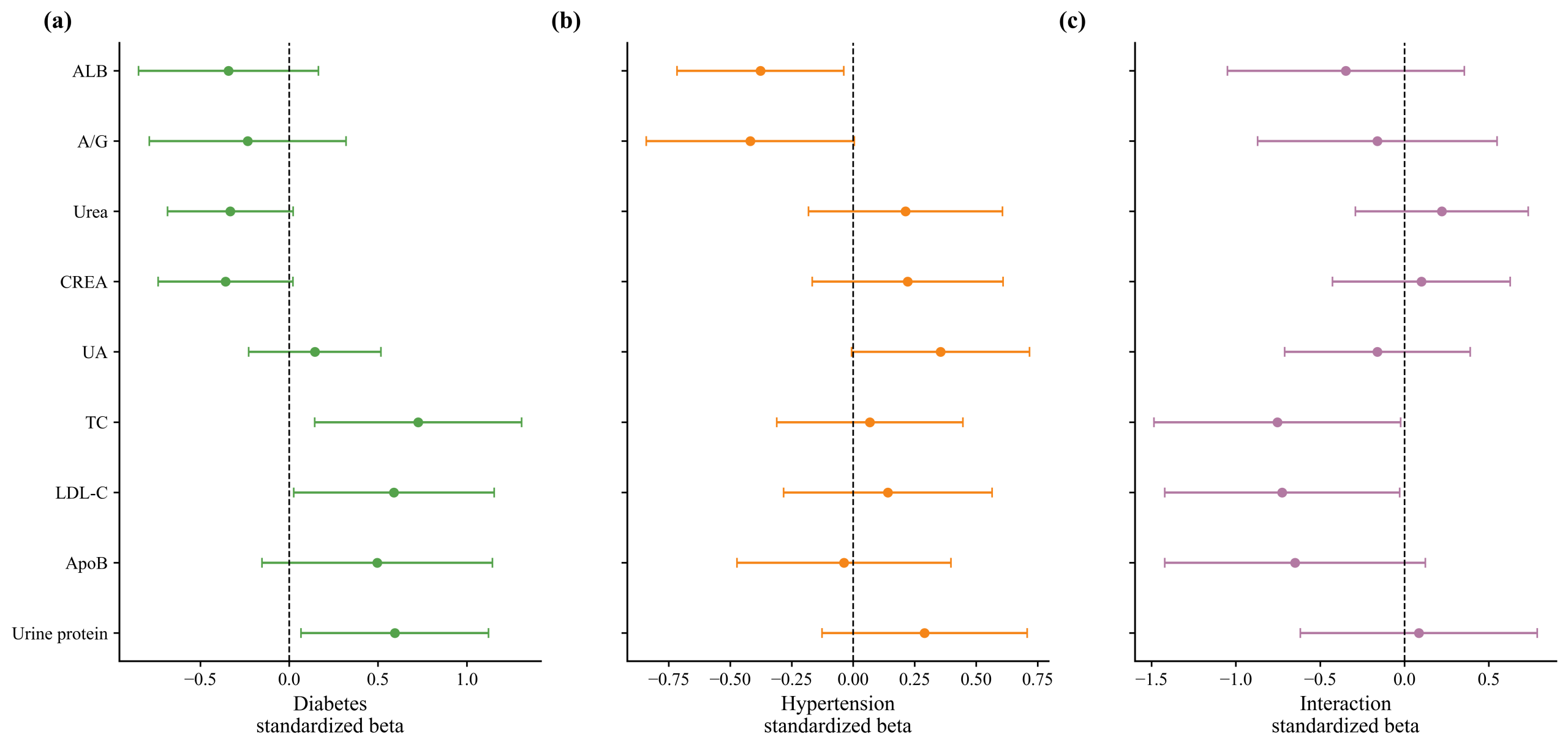


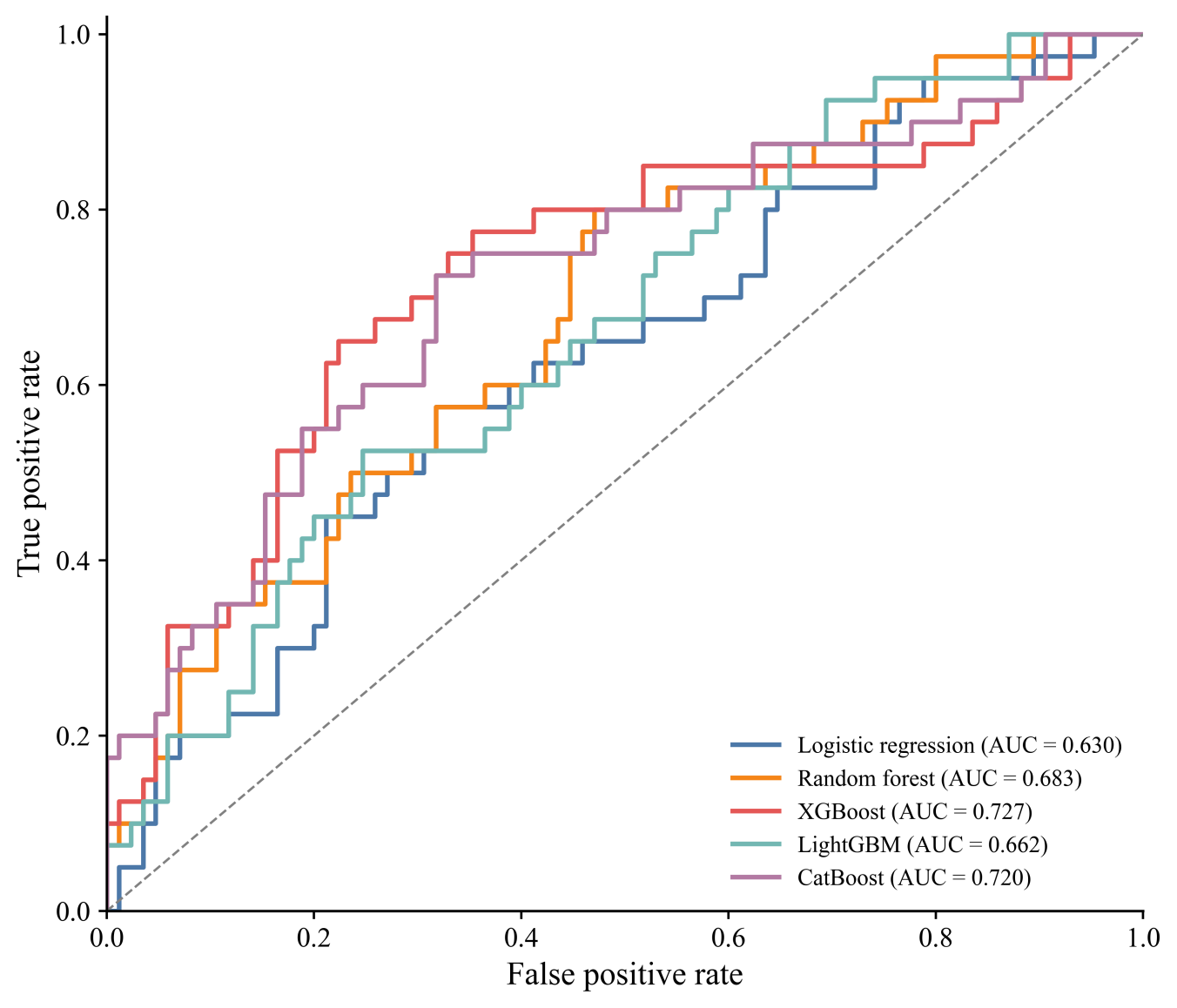
Classification was driven mainly by renal function burden, disease course, and mineral-metabolic disturbance; lipid variables contributed auxiliary signals.

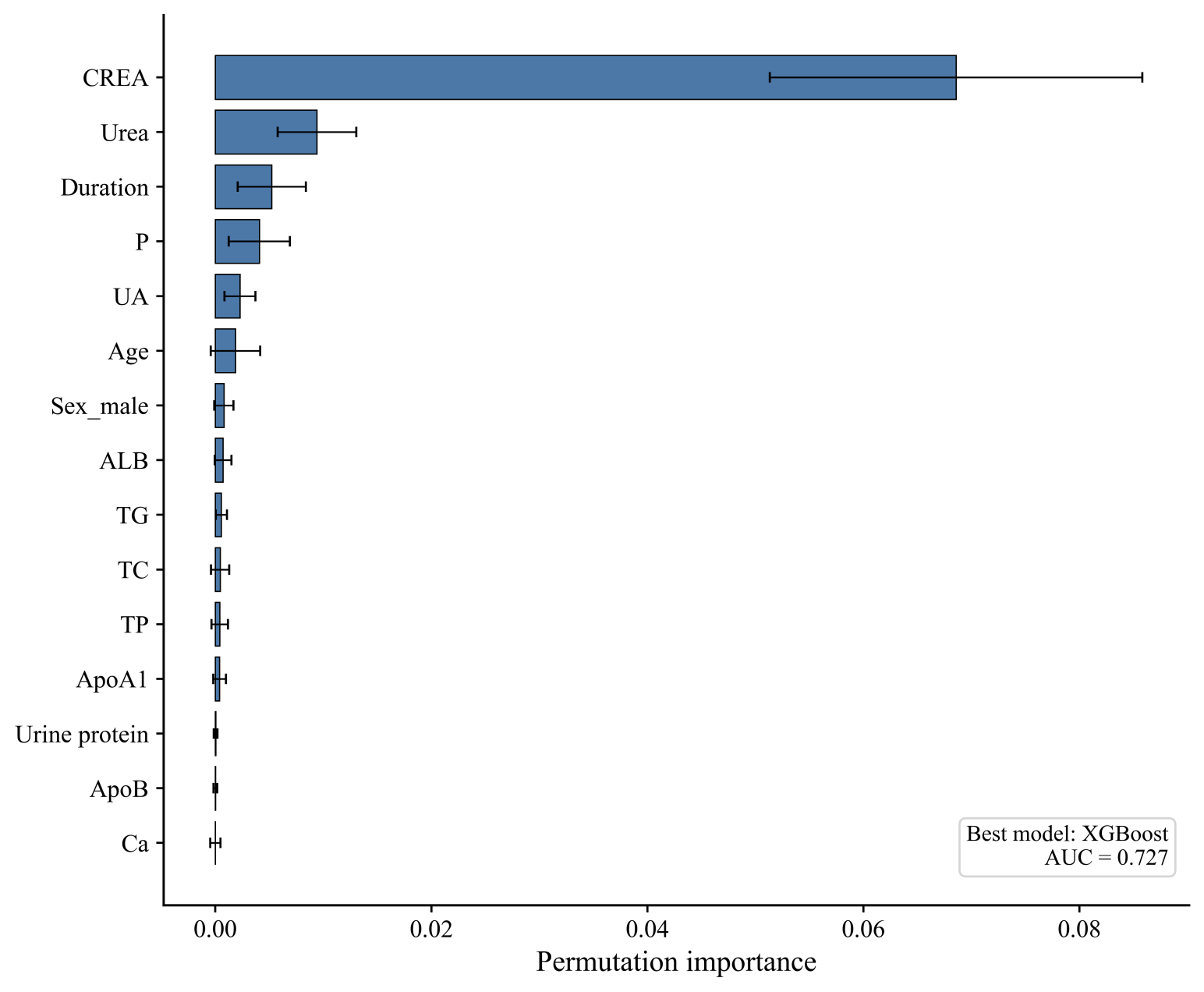
Routine biochemical indicators provide low-cost clues for phenotypic differentiation of diabetes- and hypertension-associated chronic renal failure, but are not standalone diagnostic tools.

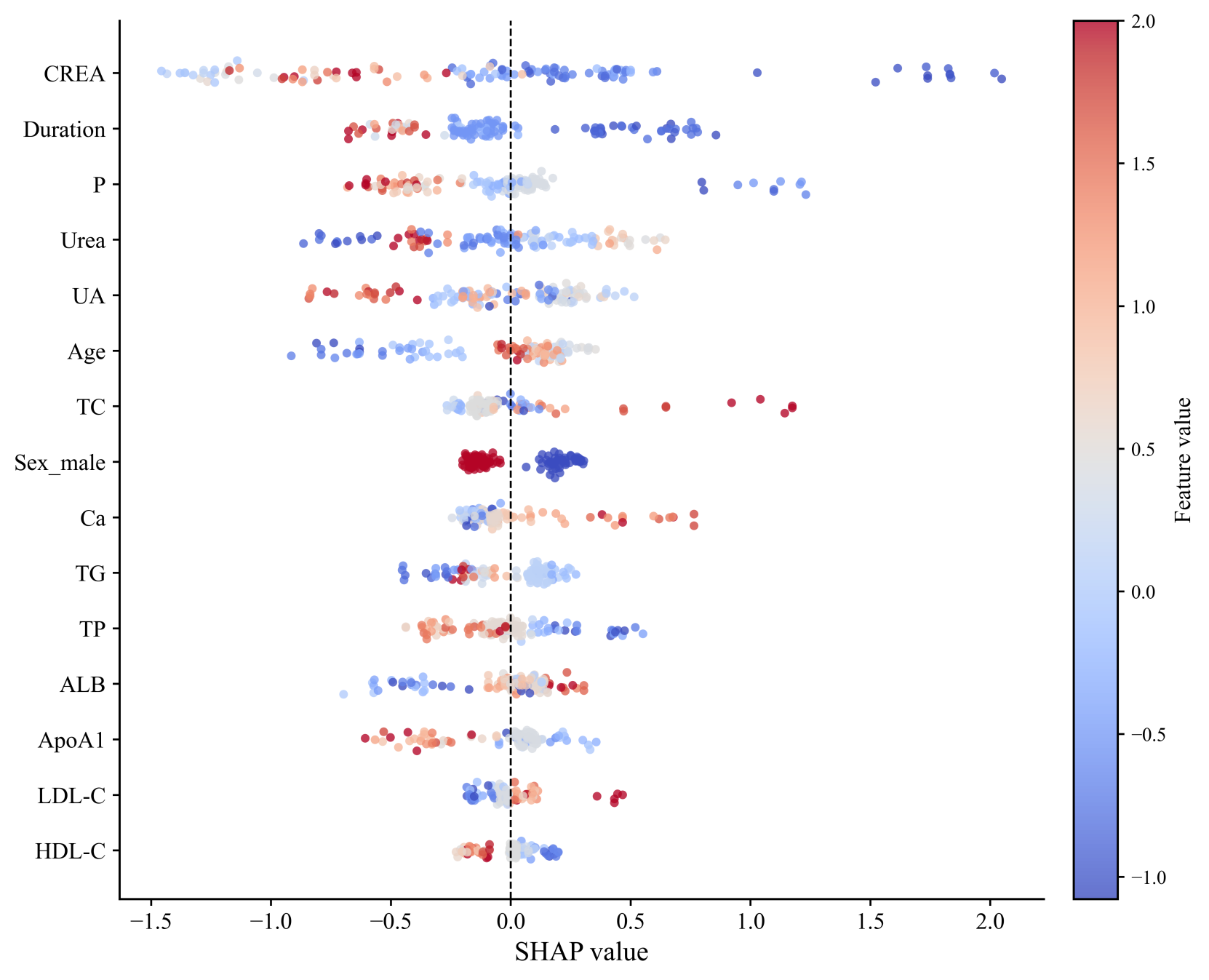

CKD-only (n=94)
HTN-only (n=85)
DM-only (n=40)
HTN+DM (n=48)
-1.3 0.6 1.0 1.1 -1.3 1.4 -1.0 -0.7 -0.7 -0.6 0.7 -0.9 -1.2 -0.7 -0.7 0.0 -1.1 0.0 -1.0
-0.5 0.6 0.9 0.3 0.2 -0.6 1.6 -0.2 0.9 -0.4 1.3 0.2 0.7 1.6 0.6 -0.5 0.6 0.0 1.0
0.4 -1.7 -1.5 0.1 -0.4 0.4 0.1 1.7 1.0 1.7 -0.9 1.6 -0.7 -1.0 -1.2 1.6 -0.9 0.0 1.0
1.4 0.6 -0.4 -1.6 1.4 -1.2 -0.6 -0.7 -1.2 -0.7 -1.1 -0.9 1.3 0.0 1.3 -1.1 1.3 0.0 -1.0
Age
Duration
TP
ALB
GLB
A/G
TG
TC
HDL-C
LDL-C
ApoA1
ApoB
Urea
CREA
UA
Ca
P
Urine protein
Occult blood
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Z-score of group median


(a)
ALB
50
45
40
35
30
25
CKD-only
HTN-only
DM-only
HTN+DM
(b)
A/G
2.25
2.00
1.75
1.50
1.25
1.00
0.75
CKD-only
HTN-only
DM-only
HTN+DM
(c)
Urea
50
40
30
20
10
CKD-only
HTN-only
DM-only
HTN+DM
(d)
CREA
1000
800
600
400
200
0
CKD-only
HTN-only
DM-only
HTN+DM
(e)
TC
10
8
6
4
2
CKD-only
HTN-only
DM-only
HTN+DM
(f)
LDL-C
7
6
5
4
3
2
1
CKD-only
HTN-only
DM-only
HTN+DM
(g)
ApoB
2.0
1.5
1.0
0.5
CKD-only
HTN-only
DM-only
HTN+DM
(h)
Proportion (%)
100
80
60
40
20
0
CKD-only
HTN-only
DM-only
HTN+DM
Grade
-
±
1+
2+
3+
4+


(a)
(b)
(c)
ALB
A/G
Urea
CREA
UA
TC
LDL-C
ApoB
Urine protein
−0.5
0.0
0.5
1.0
Diabetes
standardized beta
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
Hypertension
standardized beta
−1.5
−1.0
−0.5
0.0
0.5
Interaction
standardized beta


True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic regression (AUC = 0.630)
Random forest (AUC = 0.683)
XGBoost (AUC = 0.727)
LightGBM (AUC = 0.662)
CatBoost (AUC = 0.720)


CREA
Urea
Duration
P
UA
Age
Sex_male
ALB
TG
TC
TP
ApoA1
Urine protein
ApoB
Ca
Best model: XGBoost
AUC = 0.727
0.00
0.02
0.04
0.06
0.08
Permutation importance


CREA
Duration
P
Urea
UA
Age
TC
Sex_male
Ca
TG
TP
ALB
ApoA1
LDL-C
HDL-C
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
SHAP value
Feature value
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.xlsx
- Table2.xlsx
- Table3.xlsx
- Table4.xlsx
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryFigure2.png
- SupplementaryFigure3.png