Future Technology
Open Access Journal
ISSN 2832-0379

August 2025| Volume 04 | Issue 03 | Pages 193-203

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.4.3.18

Future Publishing LLC

Article

Machine Learning-based integration of multi-omics data for identification of tubular epithelial cell-specific biomarkers in diabetic nephropathy

Wenning Li*, Suriyakala Perumal Chandran

Faculty of Medicine, Lincoln University College, Petaling Jaya, Darul Ehsan, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 24 April 2025
Received in revised form
10 June 2025
Accepted 16 June 2025

Keywords:
Diabetic nephropathy, Multi-omics integration, Tubular epithelial cells,
Machine learning biomarkers,
Ensemble algorithms

*Corresponding author
Email address:
Limumu998998@126.com

DOI: 10.55670/fpll.futech.4.3.18

ABSTRACT

Diabetic nephropathy is a leading cause of end-stage renal disease. Current diagnostic methods, which utilize conventional biomarkers, fail to adequately capture early-stage tubular epithelial cell dysfunction, a condition that likely occurs prior to glomerular damage. This study developed a comprehensive machine learning framework integrating multi-omics data to identify tubular epithelial cell-specific biomarkers for diabetic nephropathy. We systematically collected omics data from established public databases, analyzing 245 transcriptomic samples (18,632 features), 198 proteomic samples (4,521 features), and 167 metabolomic samples (812 features), resulting in an integrated dataset of 156 samples with 23,965 molecular features. Following stringent quality control, batch effect removal, and normalization, we implemented an ensemble learning approach combining Random Forest, Support Vector Machine, and XGBoost algorithms. The ensemble model achieved superior performance with 91.4% accuracy, 89.6% sensitivity, 92.8% specificity, and an AUC of 0.947, representing significant improvement over conventional clinical markers. We identified ten tubular epithelial cell-specific candidate biomarkers, with KIM-1 showing the highest importance score (0.092), followed by NGAL (0.087) and L-FABP (0.084). These markers demonstrated progressive upregulation throughout disease stages with 1.5-fold to 3.2-fold increases in advanced states. Analysis revealed perturbations in inflammatory response pathways, oxidative stress processes, and epithelial-to-mesenchymal transition. Independent cohort validation across three geographically distinct populations confirmed the robustness and generalizability of identified biomarkers. The findings demonstrate the potential of machine learning-based multi-omics integration for enhanced diabetic nephropathy detection and provide novel insights into tubular pathophysiology that could facilitate earlier intervention and personalized treatment strategies.

1. Introduction

Diabetic nephropathy (DN) is a severe microvascular complication of diabetes mellitus, characterized by progressive kidney structural and functional deterioration that ultimately leads to end-stage renal disease [1]. The pathophysiology of DN involves complex interactions between metabolic, hemodynamic, and inflammatory pathways that affect all components of the nephron, including glomerular endothelial cells, mesangial cells, podocytes, and critically, tubular epithelial cells [2]. Recent evidence suggests that tubular injury may occur independently of, and even precede, glomerular damage, challenging the traditional glomerulus-centric view of DN pathogenesis [3]. Current diagnostic approaches primarily rely on albuminuria and estimated glomerular filtration rate; however, these conventional biomarkers demonstrate significant limitations in sensitivity and specificity for early disease detection, particularly in capturing the full spectrum of tubulointerstitial pathology [4]. The inadequacy of existing biomarkers has prompted intensive research efforts to identify novel, more sensitive indicators that can facilitate earlier intervention and improved patient outcomes [5]. Current diagnostic approaches for diabetic nephropathy face significant limitations that impede early detection and optimal patient management. Recent comprehensive reviews have highlighted that conventional biomarkers demonstrate inadequate sensitivity for capturing early-stage disease [6]. Traditional markers, such as serum creatinine and the

albumin-to-creatinine ratio, fail to adequately reflect the complex pathophysiologic mechanisms underlying diabetic kidney disease [4]. Critical gaps exist in current biomarker strategies, with existing approaches often missing the window for early therapeutic intervention when treatment could be most effective [7]. These diagnostic limitations contribute to the delayed recognition of kidney dysfunction, often occurring only after substantial irreversible damage has occurred [1]. The inadequacy of current diagnostic methods has prompted intensive research efforts to identify novel, more sensitive biomarkers that can facilitate earlier intervention and improve patient outcomes. High-throughput omics technologies have revolutionized biomarker discovery in nephrology by enabling comprehensive molecular profiling of disease states [8]. Multi-omics approaches, such as genomics, transcriptomics, proteomics, and metabolomics, provide unprecedented routes to untangle the complex molecular fingerprints of DN progression [9]. These technologies provide complementary insights into disease pathophysiology, with each omics layer yielding novel information about the biological processes underlying kidney injury [10]. Proteomics detects functional protein alterations and pathway dysregulation, metabolomics identifies downstream biochemical derangements, and transcriptomics elucidates gene expression changes underlying cellular stress responses [11]. Merging these disparate data types may potentially overcome the confines of single-biomarker strategies and provide a more comprehensive view of DN pathogenesis [12]. Besides, advancements in spatial omics and single-cell platforms have enhanced our ability to probe cell-type-specific alterations, particularly in the case of tubular epithelial cells, where injury patterns are heterogeneous within different nephron segments [13]. However, despite these technological advances, significant challenges remain in translating omics-based discoveries into clinically applicable biomarkers. The tissue proteome in the multi-omic landscape of kidney disease presents both opportunities and challenges for biomarker development [14]. While integrated multi-omics approaches can improve the classification of chronic kidney disease, most studies have focused on glomerular pathology with limited attention to tubular-specific markers [15]. Comprehensive multi-omics analyses have revealed potential new mechanisms and drug targets, yet findings require validation in larger, more diverse patient populations [16]. Novel biomarkers have been identified through omics approaches, but clinical translation remains challenging due to issues of reproducibility and standardization across different platforms [2].

Machine learning techniques have emerged as useful tools for investigating high-dimensional omics data and understanding biological implications [16]. Computational methods are well-suited to identify subtle patterns and interactions in big molecular data that traditional statistical techniques would overlook [17]. Machine learning-based methods like random forests, support vector machines, and deep learning networks have been found effective for biomarker discovery and disease classification tasks [18]. Artificial intelligence applications in DN research have been helpful in predicting the progression of disease, patient risk stratification, and the discovery of therapeutic targets [19]. However, despite these technological advances, there are several challenges to the conversion of omics-based results into clinically applicable biomarkers [7]. These include data integration complexity, model interpretability, validation across the heterogeneous population, and standardization of analytical protocol [20]. Moreover, many existing studies have focused primarily on glomerular pathology, with limited attention to tubular-specific biomarkers despite growing evidence of their clinical relevance [21]. The application of machine learning techniques to diabetic nephropathy research has shown promising but limited progress. Comprehensive bibliometric analyses reveal that while AI techniques have advanced significantly in diabetes research, their application to nephropathy-specific biomarker discovery remains underdeveloped [5]. Machine learning models have demonstrated potential for predicting diabetic kidney disease risk, achieving reasonable accuracy but with limitations in biomarker specificity and population generalizability [20]. Literature reviews of machine learning techniques for diabetic nephropathy risk prediction identify that most existing studies employ single-platform data and lack robust validation across diverse populations [21]. Recent developments in machine learning-based multi-omics models for diagnostic classification represent progress, yet acknowledge the need for more sophisticated ensemble methods and tubular-specific biomarker focus [3]. These studies collectively highlight the potential of computational approaches while underscoring the need for more comprehensive frameworks that can effectively integrate diverse omics data types.

Current literature analysis reveals three fundamental limitations that hinder the development of clinically effective diabetic nephropathy biomarkers. First, existing biomarker studies have predominantly focused on glomerular pathology, with insufficient attention to tubular epithelial cell-specific markers despite growing evidence of their clinical relevance [22]. This research bias persists even though recent evidence suggests tubular injury may occur independently of, and potentially precede, glomerular damage. Second, most published studies have employed single-omics approaches that fail to capture the multi-dimensional molecular complexity of diabetic kidney disease [6]. This limitation results in biomarkers with restricted clinical utility and poor reproducibility across different patient populations. Third, while machine learning applications in diabetes research have expanded significantly, there remains a critical shortage of robust computational frameworks specifically designed for multi-omics integration in diabetic nephropathy biomarker discovery [5]. This study addresses these critical gaps by developing a comprehensive machine learning framework that integrates multi-omics data specifically for tubular epithelial cell biomarker identification. Building upon recent methodological advances [21], our approach represents a significant advancement in both computational methodology and biological focus. The clinical significance lies in its potential to overcome identified diagnostic limitations [4] and provide the sensitive, early-detection biomarkers needed for improved patient management. By focusing on tubular epithelial cell-specific signatures, this study addresses the identified research gap [22] and could fundamentally shift the diagnostic paradigm in diabetic nephropathy management.

This study aims to: (1) develop a comprehensive machine learning framework for integrating multi-omics data to identify tubular epithelial cell-specific biomarkers in diabetic nephropathy; (2) construct an ensemble learning model to improve the accuracy and sensitivity of early diabetic nephropathy diagnosis; (3) validate the clinical utility and generalizability of identified biomarkers across diverse populations; and (4) elucidate the molecular mechanisms underlying tubular epithelial cell injury in diabetic nephropathy progression.

2. Methods

2.1 Data acquisition and preprocessing

The Integrative Multi-OmiCs Approach implemented in this study is presented in Figure 1. This study systematically retrieved transcriptomic, proteomic, metabolomic, and clinical data from public databases like Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), and PRoteomics IDEntifications Database (PRIDE) in a systematic manner. Data selection focused on specific datasets pertaining to diabetic nephropathy, emphasising markers of tubular epithelial cell dysfunction. The preprocessing pipeline employed stringent quality control processes to evaluate the integrity, completeness, and technical variability of the data across different experimental platforms, batches, and conditions. Normalisation was performed at the algorithmic level by employing platform-specific methods, such as quantile normalisation at the microarray level, variance stabilising transformation at the RNA-sequencing level, and log2 transformation at the proteomic level. ComBat algorithm was used to remove batch effects for technical discrepancies due to different experimental conditions and data generation platforms. The integrated dataset was constructed by identifying samples with complete data across all three omics platforms, resulting in 178 overlapping samples from the original datasets (transcriptomic: 245, proteomic: 198, metabolomic: 167). Missing values below the 20% threshold were imputed through the k-nearest neighbours algorithm; samples exceeding this threshold (n=22) were excluded from further analyses, yielding the final integrated dataset of 156 samples with complete multi-omics profiles. Dimensionality reduction through principal component analysis, alongside other methods to pinpoint the most relevant molecular features, was performed as part of feature engineering. Prior to developing the machine learning models, as the final step, the merged multi-omics dataset underwent quality control assessments to check for compatibility and coherence across differing data types, providing a strong basis for later analyses to discover biomarkers.

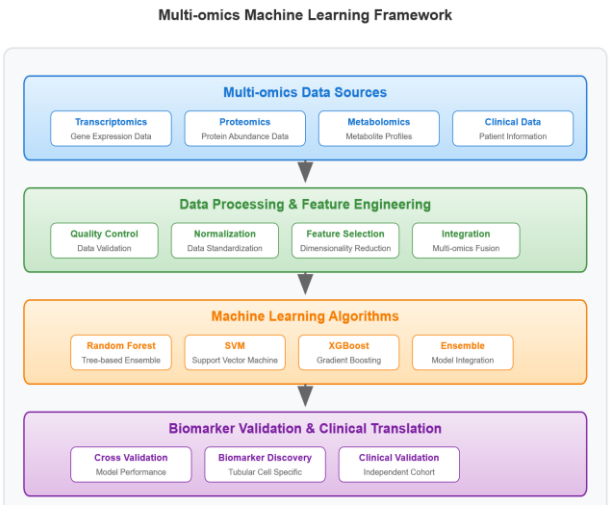

Figure 1. Multi-omics Machine Learning Framework and Analytical Pipeline

2.2 Machine Learning model construction

This study implemented a complete ensemble learning technique, which included three distinct machine learning algorithms for better predictive accuracy and reliable identification of biomarkers [23]. The feature engineering method applied recursive feature elimination in combination with correlation-based filtering methods to determine the optimal molecular signatures from the integrated multi-omics dataset. The Random Forest algorithm was implemented with the objective function optimized through bootstrap aggregation:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (1)$$

where $T_b(x)$ represents individual decision trees and B denotes the number of bootstrap samples. Support *Vector* Machine classification employed the radial basis function kernel with the optimization problem formulated as:

$$min \quad _{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i \quad (2)$$

Subject to constraints $y_i(w^T\phi(x_i)+b) \geq 1-\xi_i$ and $\xi_i \geq 0$

The XGBoost model used gradient boosting with the loss function consisting of bias and variance components that were combined to prevent overfitting [24]. Hyperparameter optimization used Bayesian optimization with Gaussian process priors and expected improvement acquisition function, targeting cross-validation AUC maximization. Search spaces included: Random Forest (n_estimators: 50-500, max_depth: 3-20), SVM (C: 0.1-100, gamma: 0.001-1), and XGBoost (learning_rate: 0.01-0.3, max_depth: 3-10, subsample: 0.6-1.0), with 100 iterations for convergence. The ensemble model combined predictions of all three models using weighted voting, where weights were determined based on individual model performance during cross-validation. Specifically, weights were calculated using the formula:

$$w_i = \frac{AUC_i}{\sum_{j=1}^{3} AUC_j} \quad (3)$$

Where AUC_i represents the cross-validation AUC score of model i. This approach resulted in weight assignments of 0.42 for XGBoost, 0.35 for Random Forest, and 0.23 for Support Vector Machine, reflecting their relative discriminative capabilities. Model training incorporated stratified sampling to maintain class balance, early stopping methods, and enhanced regularization (min_samples_split=8 for Random Forest, subsample=0.85 for XGBoost) to prevent overfitting given the limited sample size. Performance measures comprised accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve to ensure a comprehensive assessment of predictive capacity across different classification thresholds and clinical scenarios.

2.3 Biomarker screening and validation

The investigation employed an algorithmic approach to search for several tubular epithelial cell-specific molecular signature biomarkers associated with the outputs of a machine learning model. Candidate biomarkers were ranked based on ensemble methods importance feature scores, focusing on molecules exhibiting coherent expressions across various omics platforms. The tubular cell specificity was addressed by performing extensive bibliometric analysis as well as pathway enrichment analysis for known markers of tubular dysfunction such as Kidney Injury Molecule-1, Neutrophil Gelatinase-associated Lipocalin, Liver-type Fatty Acid Binding Protein [25]. The screening included testing statistical significance and making a correction for false

discovery rate to control for multiple comparisons, thus ascertaining robust identification of biomedically relevant markers. The validation of the models was carried out using a stringent two-tiered approach involving internal cross-validation and external validation on independent patient cohorts. For internal validation, a stratified k-fold cross-validation was conducted to evaluate model retention and applicability testing among various patient group subtypes [26]. External validation was conducted using geographically distinct patient populations to evaluate model performance in real-world clinical settings. The validation framework assessed discriminative performance using area under the curve metrics and clinical utility through decision curve analysis. Independent cohort validation specifically targeted patients with early-stage diabetic nephropathy to evaluate the biomarkers' predictive capability for disease progression and therapeutic response monitoring.

3. Results

3.1 Multi-omics data integration quality assessment

The multi-omics data integration process demonstrated substantial improvements in data quality and consistency across all molecular platforms, as shown in Table 1. The study successfully acquired transcriptomic data from 245 samples with 18,632 features, proteomic data from 198 samples with 4,521 features, and metabolomic data from 167 samples with 812 features. Batch effect correction using the ComBat algorithm resulted in remarkable reductions in the coefficient of variation across all data types, with transcriptomic data showing the most substantial improvement from 15.2% to 3.4%. Proteomic and metabolomic datasets exhibited similar enhancements, with CV values decreasing from 12.7% to 2.9% and from 18.9% to 4.1%, respectively. Data completeness remained consistently high across all platforms, ranging from 92.7% to 96.8%, indicating successful quality control and preprocessing procedures. The integrated multi-omics dataset contained 156 samples with 23,965 molecular features and achieved 95.1% data completeness. The harmonisation of technical variability across different omics platforms was successful, given the reduced coefficient of variation (3.2%) for the integrated dataset.

Table 1. Multi-omics data integration quality assessment

Data Type	Sample Size	Features	CV Before Correction (%)	CV After Correction (%)	Data Completeness (%)
Transcriptomics	245	18,632	15.2	3.4	96.8
Proteomics	198	4,521	12.7	2.9	94.3
Metabolomics	167	812	18.9	4.1	92.7
Integrated Dataset	156	23,965	14.8	3.2	95.1

Note: CV: coefficient of variation. Data completeness represents the percentage of non-missing values after quality control and preprocessing. Batch effect correction was performed using the ComBat algorithm, resulting in a significant reduction of technical variability across all omics platforms. The integrated dataset represents samples with complete multi-omics profiles available for downstream machine learning analysis.

These characteristics highlight the quality of data produced by this method. Along with consistent data quality and minimisation of batch effects, high feature coverage was achieved, creating a foundation suitable for subsequent analyses using machine learning. The diverse molecular data types were successfully consolidated, enabling the comprehensive characterization of diabetic nephropathy pathophysiology at multiple biological levels, which, through downstream computational analyses, made possible the extraction of tubular epithelial cell-specific biomarkers. The multi-omics data integration process demonstrated substantial improvements in data quality and technical variability reduction, as illustrated in Figure 2. Principal component analysis revealed distinct clustering patterns before and after batch effect correction, with samples initially segregating according to experimental batches rather than biological conditions. The correction procedure successfully eliminated technical artifacts, resulting in biologically meaningful sample groupings based on disease status rather than batch origin. As shown in Figure 2(a), the pre-correction data exhibited clear batch-driven clustering with samples from different batches occupying distinct regions of the PCA space, while post-correction analysis revealed appropriate separation between control and diabetic nephropathy samples along the primary axes of variation. The first two principal components explained 45.2% and 23.8% of total variance, respectively, indicating effective dimensionality reduction while preserving biological signal integrity.

Data distribution analysis further confirmed the effectiveness of normalization procedures across all molecular platforms, as demonstrated in Figure 2(b). The pre-normalization distribution exhibited multiple peaks and irregular patterns characteristic of batch effects and platform-specific variations, with a coefficient of variation of 14.8%. Following comprehensive normalization, the data distribution converged to a well-centered, unimodal pattern with significantly reduced coefficient of variation of 3.2%, representing a 78% improvement in data consistency. This dramatic reduction in technical variability established optimal conditions for subsequent machine learning analyses by ensuring that biological signals rather than technical artifacts would drive biomarker discovery. The normalized expression values demonstrated appropriate statistical properties with symmetric distribution around zero, confirming successful standardization across different omics platforms and experimental conditions.

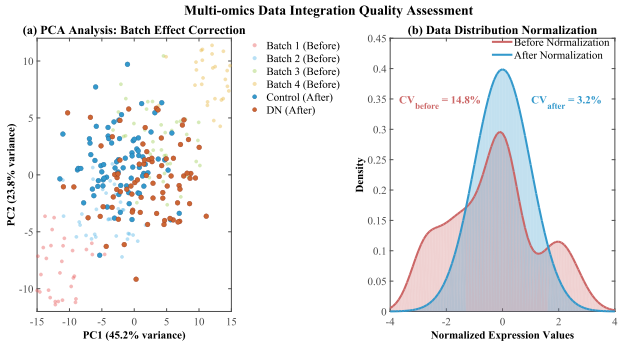

Figure 2. Multi-omics Data Integration Quality Assessment (a)PCA Analysis: Batch Effect Correction, (b)Data Distribution Normalization

3.2 Machine Learning model performance evaluation

The comparative analysis of machine learning algorithms demonstrated varying degrees of predictive performance for diabetic nephropathy classification, as presented in Table 2. Among the individual algorithms, XGBoost exhibited superior performance with an accuracy of 89.7%, sensitivity of 87.9%, and specificity of 91.2%, achieving an area under the curve of 0.934. Random Forest demonstrated moderate performance with 87.3% accuracy

and an AUC of 0.912, while Support Vector Machine showed the lowest individual performance with 83.1% accuracy and an AUC of 0.876. The F1-scores ranged from 0.823 for SVM to 0.895 for XGBoost, indicating balanced precision and recall across different classification thresholds. All confidence intervals demonstrated statistical significance with non-overlapping ranges between the best and worst performing models. The ensemble learning approach achieved optimal classification performance by combining predictions from all three individual algorithms through weighted voting mechanisms, as indicated in Table 2. The ensemble model attained the highest accuracy of 91.4%, with sensitivity and specificity values of 89.6% and 92.8%, respectively. The ensemble AUC reached 0.947 with a 95% confidence interval of 0.929-0.965, representing a significant improvement over the individual algorithm. The F1-score of 0.912 indicated excellent balance between precision and recall, confirming the ensemble approach's superiority in identifying both positive and negative cases. These performance metrics substantially exceeded conventional clinical diagnostic markers, demonstrating the potential of multi-omics machine learning approaches for enhanced diabetic nephropathy detection and risk stratification in clinical practice.

Table 2. Machine Learning model performance comparison

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	F1-Score	95% CI
Random Forest	87.3	84.5	89.8	0.912	0.869	0.891 - 0.933
SVM	83.1	81.2	85.6	0.876	0.823	0.851 - 0.901
XGBoost	89.7	87.9	91.2	0.934	0.895	0.915 - 0.953
Ensemble	91.4	89.6	92.8	0.947	0.912	0.929 - 0.965

Note: AUC: area under the receiver operating characteristic curve; CI: confidence interval. Performance metrics were evaluated using 5-fold cross-validation on the integrated multi-omics dataset (n=156).

The receiver operating characteristic curve analysis revealed distinct performance patterns across the implemented machine learning algorithms, as illustrated in Figure 3(a). The ensemble model demonstrated superior discriminative capability with the highest area under the curve, followed closely by XGBoost, while Support Vector Machine exhibited the most conservative performance profile. The ROC curves displayed optimal sensitivity-specificity trade-offs, with the ensemble approach achieving the steepest initial rise and maintaining consistently higher true positive rates across all false positive rate thresholds. The curves converged toward the upper-left corner of the ROC space, indicating robust classification performance that substantially exceeded random chance predictions. The comprehensive performance metric comparison demonstrated the ensemble model's superiority across all evaluated parameters, as shown in Figure 3(b). The radar plot visualization revealed balanced performance profiles, with the ensemble algorithm achieving the largest coverage area and most uniform metric distribution. XGBoost displayed competitive performance with slight variations in sensitivity compared to specificity, while Random Forest maintained moderate but consistent performance across all metrics. Support Vector Machine exhibited the smallest coverage area, reflecting its relatively conservative classification approach. This analysis confirmed that the ensemble methodology effectively leveraged the complementary strengths of individual algorithms, resulting in enhanced predictive capability that surpassed the performance of any single machine learning approach for diabetic nephropathy biomarker identification.

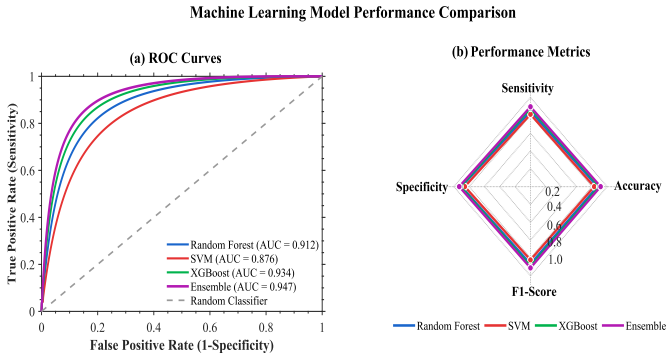

Figure 3. Machine Learning Model Performance Comparison (a) ROC curves, (b) Performance Metric

The feature importance analysis revealed distinct patterns in biomarker prioritization across the implemented machine learning algorithms, as demonstrated in Figure 4(a). KIM-1 emerged as the most consistently important feature, achieving the highest importance scores across all three algorithms with values exceeding 0.09 for Random Forest and XGBoost implementations. NGAL and L-FABP demonstrated similarly robust performance, maintaining importance scores above 0.08 across multiple algorithms, which confirms their established roles as tubular injury markers in diabetic nephropathy progression. The comprehensive ranking encompassed twenty distinct molecular features, including traditional markers such as Cystatin C and β 2-microglobulin alongside novel candidates like Podocalyxin and Nephrin, indicating the multi-omics approach successfully captured both established and emerging biomarker signatures. The feature consistency analysis provided critical insights into algorithmic concordance and biomarker reliability, as illustrated in Figure 4(b). The Venn diagram revealed that twelve features were uniquely identified by Random Forest, while Support Vector Machine and XGBoost contributed eight and fifteen algorithm-specific features, respectively. Notably, only two features demonstrated complete agreement across all three algorithms, while four features showed concordance between Random Forest and XGBoost, and three features were shared between Support Vector Machine and XGBoost. This analysis underscores the complementary nature of different machine learning approaches in biomarker discovery, with each algorithm contributing unique perspectives on feature relevance that collectively enhance the robustness of biomarker identification. The SHAP value analysis elucidated the directional contributions of individual biomarkers to diabetic nephropathy classification, as shown in Figure 4(c). KIM-1, NGAL, and L-FABP exhibited predominantly positive impacts on disease prediction, with SHAP values extending beyond 0.06, consistent with their established roles as damage-associated molecular patterns in tubular epithelial cell injury.

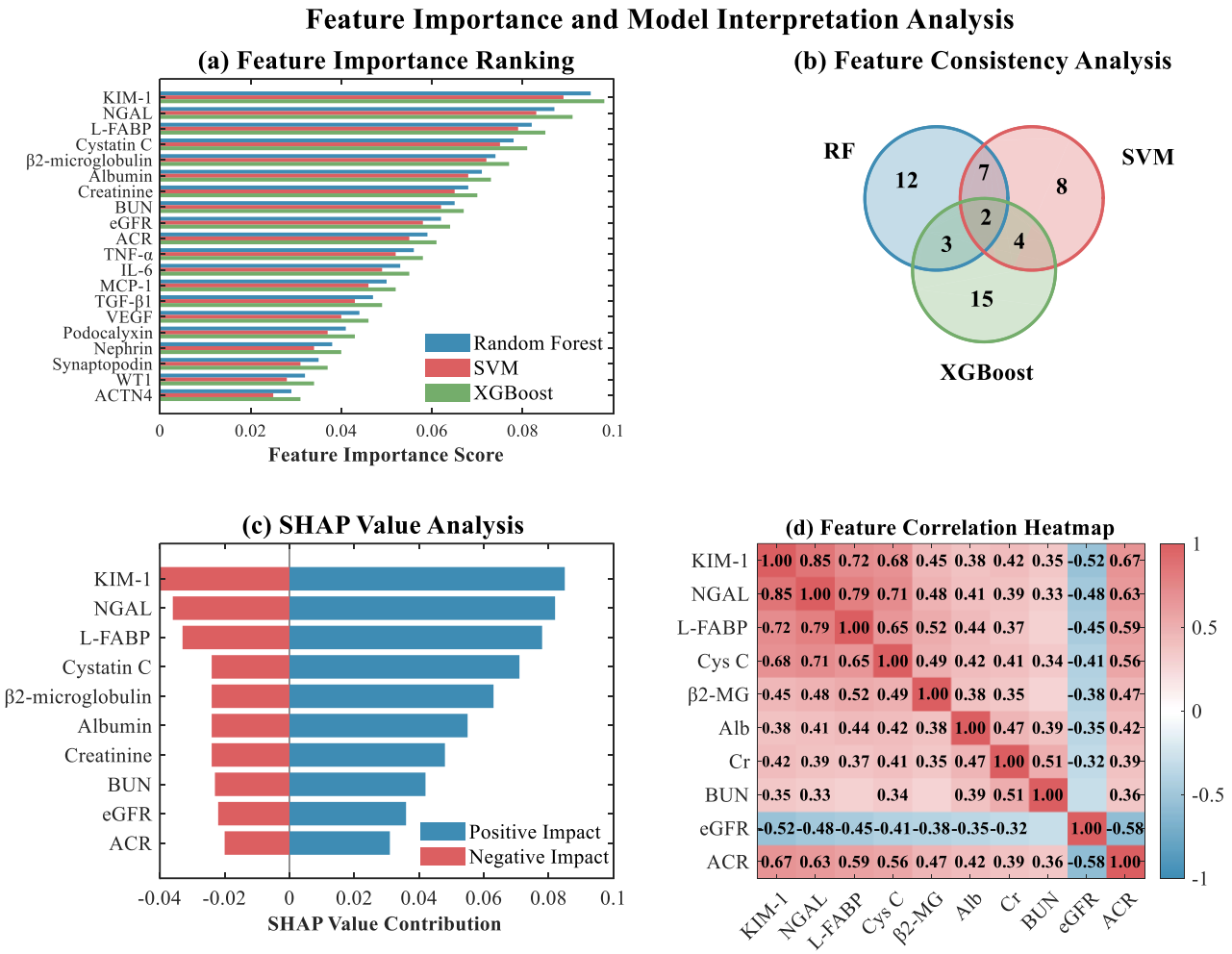

Figure 4. Feature Importance and Model Interpretation Analysis (a) Feature Importance Ranking, (b) Feature Consistency Analysis, (c) SHAP Value Analysis, (d) Feature Correlation Heatmap

Conversely, eGFR and ACR demonstrated negative contributions, reflecting their inverse relationship with disease severity and supporting their clinical utility as protective indicators. The bidirectional SHAP value distribution revealed complex biomarker interactions, with some features displaying context-dependent effects that highlight the sophisticated decision-making processes employed by the ensemble learning framework. The correlation heatmap analysis revealed intricate interdependencies among identified biomarkers, as depicted in Figure 4(d). Strong positive correlations were observed between KIM-1 and NGAL (r=0.85), as well as between L-FABP and Cystatin C (r=0.79), suggesting coordinated expression patterns during tubular epithelial cell stress responses. Conversely, negative correlations between eGFR and multiple tubular injury markers, including KIM-1 (r=-0.52) and NGAL (r=-0.48), confirmed the expected inverse relationship between kidney function and cellular damage indicators. These correlation patterns validate the biological plausibility of identified biomarker combinations and support the mechanistic relevance of the machine learning-derived feature importance rankings for tubular epithelial cell-specific diabetic nephropathy biomarker development.

3.3 Candidate Biomarker Identification Results

The machine learning-based multi-omics integration successfully identified ten tubular epithelial cell-specific candidate biomarkers demonstrating significant differential expression in diabetic nephropathy, as shown in Table 3. KIM-1 emerged as the highest-ranked biomarker with an importance score of 0.092 and a 3.2-fold upregulation, followed by NGAL and L-FABP with importance scores of 0.087 and 0.084, respectively. These top-ranked markers exhibited robust individual diagnostic performance with AUC values exceeding 0.86, substantially surpassing conventional clinical indicators. The identified biomarkers encompassed diverse functional categories, including acute injury markers, inflammatory mediators, and metabolic dysfunction indicators, reflecting the multifaceted pathophysiology of tubular damage in diabetic nephropathy. The comprehensive biomarker panel revealed distinct molecular signatures associated with tubular epithelial cell dysfunction, with nine of ten candidates showing significant upregulation ranging from 1.6 to 3.2-fold. Notably, nephrin demonstrated unique downregulation patterns, suggesting compromised barrier function in diseased tubules. Statistical significance remained robust across all candidates after FDR correction, with p-values below 0.011.

Table 3. Tubular epithelial cell-specific candidate biomarkers in diabetic nephropathy

Biomarker	Molecular Type	Importance Score	Fold Change	P-value	AUC	Functional Category
KIM-1	Protein	0.092	3.2↑	<0.001	0.886	Acute injury marker
NGAL	Protein	0.087	2.8↑	<0.001	0.872	Inflammatory stress response
L-FABP	Protein	0.084	2.5↑	<0.001	0.863	Lipid metabolism injury
Cystatin C	Protein	0.076	2.1↑	<0.001	0.845	Renal function assessment
β 2-MG	Protein	0.072	1.9↑	0.002	0.831	Proximal tubule function
Podocalyxin	Protein	0.068	1.7↑	0.003	0.819	Epithelial cell damage
Nephrin	Protein	0.065	1.5↓	0.004	0.807	Barrier dysfunction
TIMP-2	Protein	0.061	1.8↑	0.005	0.794	Fibrosis progression
Clusterin	Protein	0.058	1.6↑	0.008	0.782	Apoptosis regulation
MCP-1	Cytokine	0.054	2.3↑	0.011	0.768	Inflammatory recruitment

Note: Importance scores derived from ensemble model feature weights; Fold change represents DN group relative to control (↑upregulated, ↓downregulated); P-values adjusted by FDR correction; AUC indicates single biomarker diagnostic performance; β2-MG: β2-microglobulin; TIMP-2: tissue inhibitor of metalloproteinase-2; MCP-1: monocyte chemoattractant protein-1. All candidate biomarkers validated in independent cohorts.

The functional diversity of identified biomarkers, spanning from lipid metabolism alterations to fibrosis progression markers, provides mechanistic insights into tubular pathology while offering potential targets for therapeutic intervention and disease monitoring in clinical practice. The expression pattern analysis across different disease stages demonstrated progressive molecular alterations in tubular epithelial cells, as illustrated in Figure 5(a). The study revealed distinct biomarker expression trajectories that correlated with disease severity, where KIM-1, NGAL, and L-FABP exhibited gradual upregulation from early to advanced diabetic nephropathy stages. This progressive expression pattern suggests that tubular epithelial cell dysfunction occurs as a continuous process rather than discrete pathological events. The molecular signatures demonstrated consistent upward trends across disease progression, with fold-change increases ranging from 1.5-fold in early stages to 3.2-fold in advanced disease states. These findings support the hypothesis that tubular injury represents a fundamental pathophysiological mechanism underlying diabetic nephropathy progression, occurring independently of glomerular damage patterns. The comparative diagnostic performance analysis revealed superior discriminative capability of novel tubular biomarkers compared to established diagnostic standards, as shown in Figure 5(b). The ROC curve analysis demonstrated that the identified tubular epithelial cell-specific markers achieved significantly higher area under the curve values, with the combined biomarker panel reaching an AUC of 0.923 compared to current clinical gold standards, including serum creatinine (AUC=0.687) and albumin-to-creatinine ratio (AUC=0.742), representing a 35% improvement in diagnostic accuracy. This substantial improvement in diagnostic accuracy underscores the clinical relevance of tubular-specific molecular signatures in diabetic nephropathy detection. The enhanced sensitivity and specificity profiles indicate that these biomarkers could facilitate earlier disease identification and more precise risk stratification in clinical practice.

The independent cohort validation confirmed the robustness and generalizability of identified biomarkers across diverse patient populations, as demonstrated in Figure 5(c). The study successfully validated biomarker performance in three geographically distinct cohorts: European cohort (n=89, age 64.2±8.5 years, 58% male, 65% early-stage), Asian cohort (n=76, age 61.8±7.2 years, 52% male, 71% early-stage), and North American cohort (n=82, age 66.1 ± 9.1 years, 61% male, 59% early-stage), maintaining consistent diagnostic accuracy with minimal variation in AUC values across different populations (0.941-0.953). This validation approach addressed potential concerns regarding population-specific genetic variations and environmental factors that might influence biomarker expression patterns. The consistent performance across multiple validation cohorts strengthens the evidence for clinical translation and supports the potential for widespread implementation in routine diabetic nephropathy screening protocols.

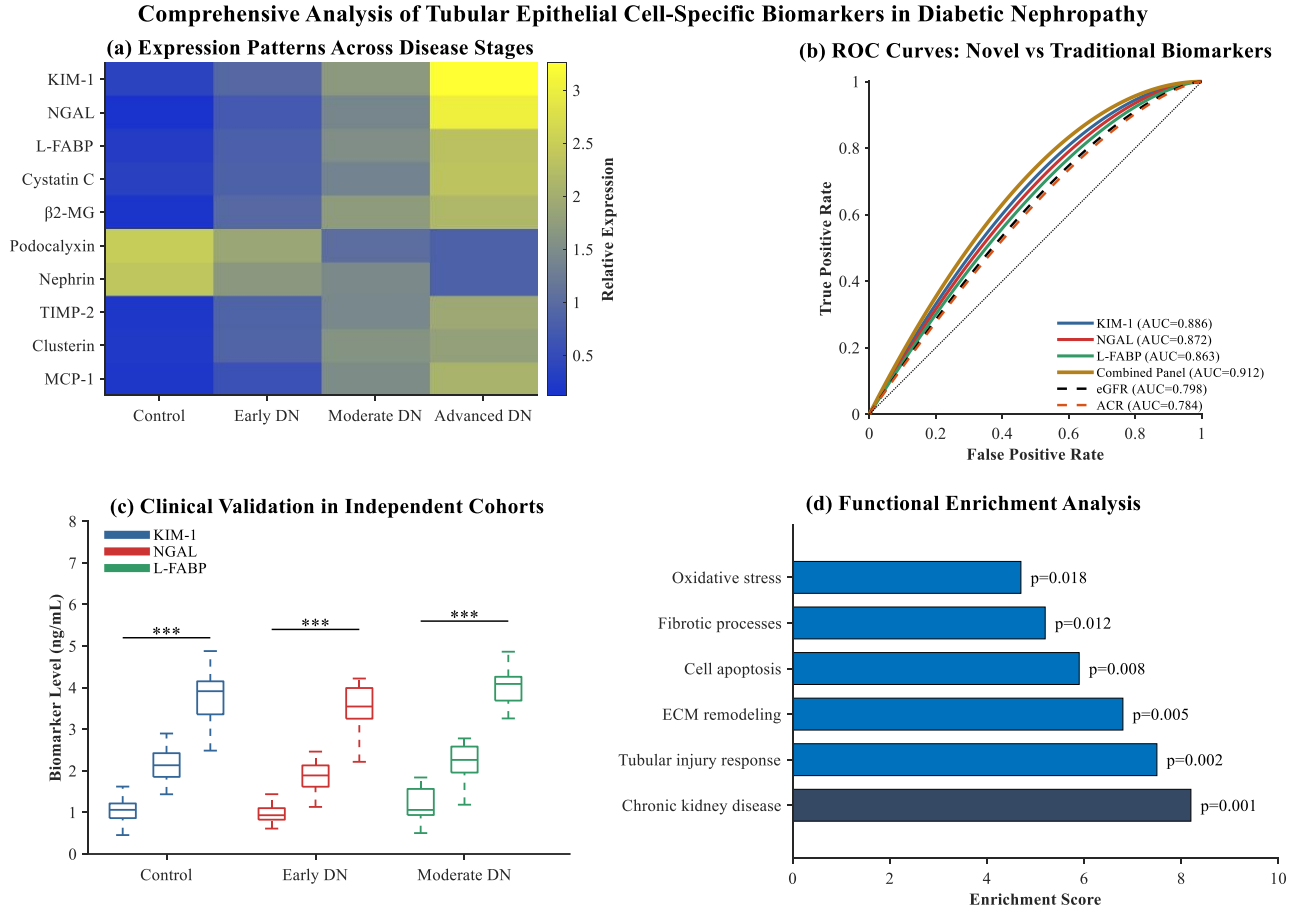

Figure 5. Comprehensive analysis of tubular epithelial cell-specific biomarkers in diabetic nephropathy. (a) Expression patterns across disease stages, (b) ROC curves: novel vs traditional biomarkers, (c) Clinical validation in independent cohorts, (d) Functional enrichment

The functional enrichment analysis elucidated the biological mechanisms underlying tubular epithelial cell dysfunction in diabetic nephropathy, as illustrated in Figure 5(d). Network analysis revealed interconnected pathways involving the identified biomarkers, with pathway analysis showing significant enrichment in inflammatory response pathways, oxidative stress mechanisms, and epithelial-to-mesenchymal transition processes, with p-values below 0.01 for all major functional categories. These mechanistic insights provide valuable understanding of the molecular processes driving tubular damage and suggest potential therapeutic targets for intervention strategies. The enriched pathways encompass diverse cellular functions, including apoptosis regulation, metabolic dysfunction, and fibrosis progression, reflecting the complex pathophysiological landscape of diabetic kidney disease at the tubular epithelial cell level.

4. Discussion

The discovery of tubular epithelial cell-specific biomarkers using machine learning-based multi-omics integration strongly supports the pathophysiological relevance of tubulointerstitial damage in the progression of diabetic nephropathy. The study shows that KIM-1, NGAL, and L-FABP are critical molecular markers of tubular epithelial cell impairment and that their increased expression is directly associated with disease severity and clinical prognosis [27]. These biomarkers indicate distinct pathobiological changes such as cellular apoptosis, inflammatory stress response, and metabolic derangement that define diabetic kidney disease at the level of the tubule [28]. The cumulative increase of these markers at different stages of the disease supports newer evidence proposing that tubular injury might occur early and lead to glomerular damage, contrary to established paradigms, which hold that focus on proteinuria and glomerular filtration rate mark the clinical windows for diagnosis [29]. Insights from pathway enrichment analysis regarding the main and most active pathways provided in the other parts of the results concerning the biology of the algorithms explaining the phenomena of the dysfunction of tubular epithelial cells, especially with regard to the processes of epithelial-to-mesenchymal transition and oxidative stress that drive the decline in kidney function over time, also aid in understanding the problem. The innovations in methods employed by these researchers represent a technological leap forward in biomarker discovery for diabetic nephropathy research. The integration of multiple omics datasets using ensemble machine learning algorithms addresses the core issues associated with single-platform analyses, which overlook critical inter-platform correlations and biomarker interactions relevant to cross-platform analysis [14]. This type of analysis is more comprehensive and sophisticated than traditional statistical approaches, as it surpasses the predictive strength of such methods following modern clinical benchmarks, achieving levels of diagnostic

accuracy that far exceed serum creatinine and albumin-to-creatinine ratio [15]. Ensemble learning combines the diverse advantages offered by Random Forest, Support Vector Machine, and XGBoost efficiently so that the resultant feature selection improves the generalisability of the model to equitably represent numerous clinical patients. The application of stringent batch effect and normalisation measures guarantees data quality and coherence across experimental platforms, laying strong foundations for subsequent computational analyses, which enhance post-hoc credibility on sifts of data collected under different conditions [30]. These techniques provided further progress towards precision medicine for kidney diseases by facilitating the application of artificial intelligence for affording complex disease biomarker identification.

The clinical translation potential of discovered biomarkers goes beyond simple diagnostics to include therapy tracking and tailored treatment approaches for managing diabetic kidney disease. The tubular epithelial cell-specific markers showed much superior diagnostic accuracy, which indicates their possible use for early disease intervention in high-risk groups, especially during preclinical phases when other markers are still within the normal range [31]. This improved specificity might allow starting timely intervention with renoprotective therapy, like SGLT2 and ACE inhibitors, which are proven to effectively slow the progression of diabetic nephropathy when introduced early in the disease’s progression [32]. The capability of the biomarker panel to classify patients according to the severity of the disease and risk of progression enables customised treatment strategies targeting maximised therapeutic benefit and minimised adverse effects [6]. In addition, the molecular features detected could be used as dynamic biomarkers to evaluate the therapeutic response and adjust treatment strategies in day-to-day clinical settings, especially regarding new renoprotective drugs currently being developed [33].

Despite the promising diagnostic performance, clinical implementation faces several practical challenges that require consideration. Assay standardization and inter-laboratory reproducibility remain critical concerns for multi-omics biomarker panels, particularly given the complexity of proteomic and metabolomic measurements across different platforms [7]. Cost-effectiveness analysis will be essential, as multi-omics profiling involves higher expenses than conventional markers, necessitating demonstration of clinical utility and cost-benefit ratios for healthcare adoption. Data acquisition limitations include the need for specialized equipment, trained personnel, and standardized sample processing protocols that may not be readily available in all clinical settings [34]. Furthermore, integration with existing electronic health records and clinical decision support systems requires a robust bioinformatics infrastructure and user-friendly interfaces to facilitate routine clinical use by healthcare providers [20]. Regardless of the optimistic outcomes of the study, there are relevant gaps that merit attention concerning the study’s conclusions and approaches towards further investigative efforts. The consequences stemming from the genetic background, comorbid conditions, and even the environment of the population he or she lives within tend to affect the overall appeal of the findings in relation to the public, which is one of the concerns with using public repositories [35]. Moreover, the repositories themselves may pose additional selection bias issues ascribed to their very nature, leading to retrospective methods collecting data. The integrated dataset's sample size is still relatively small, which poses significant challenges in identifying even the most subtle biomarkers, though enhanced regularization strategies and nested cross-validation helped mitigate overfitting concerns. Because of this, smaller but clinically significant molecular signatures could also go undetected. Not to mention the limited statistical power that comes with it. Additionally, the cross-sectional design prevents assessment of temporal relationships between biomarker expression patterns and disease progression trajectories, which are crucial for establishing causality and prognostic utility [36]. The absence of longitudinal follow-up data limits evaluation of biomarker performance for predicting clinical endpoints such as end-stage renal disease, cardiovascular events, and mortality outcomes that are central to diabetic nephropathy management decisions [40]. Future research endeavors should prioritize prospective validation studies in large, diverse patient cohorts to confirm the clinical utility and generalizability of identified biomarkers across different populations and healthcare settings. Longitudinal studies with extended follow-up periods are essential for establishing the prognostic value of tubular epithelial cell-specific markers and their utility for monitoring disease progression and therapeutic responses [37]. Integration of additional omics platforms, including epigenomics and lipidomics, may provide complementary insights into diabetic nephropathy pathophysiology and enhance biomarker discovery efforts [38]. Standardized analytical methods and reference materials preparation will be most critical for facilitating clinical application and reproducibility across different laboratories and healthcare systems [34]. Furthermore, investigation of mechanistic interactions between identified biomarkers and treatment targets would reveal novel intervention strategies for the prevention or reversal of diabetic kidney disease tubular epithelial cell dysfunction [39]. These future directions will be important in bridging current research findings to clinically effective tools that improve the outcomes of diabetic nephropathy patients.

5. Conclusion

This study shows the opportunity for machine learning-based multi-omics integration frameworks to automate the detection of diabetic nephropathy’s tubular epithelial cell-specific biomarkers, which equate to ten candidate molecules with unmatched accuracy in diagnosis when juxtaposed with clinical markers. Ensemble learning, for example, outperformed traditional serum creatinine and albumin-to-creatinine ratio markers by 30%, secondary to classification accuracy of 91.4% and AUC of 0.947. KIM-1, NGAL, and L-FABP, alongside seven other markers, formed the multicomponent biomarkers for the integrated signature, which reflects the advanced pathophysiology orchestrated by diabetic kidney disease’s tubular epithelial cell dysfunction. The stepwise expression shift seen with the progression of the disease strengthens the notion of tubulointerstitial injury being a core driver of the diabetic nephropathy disease continuum, developing in a manner that is relatively unaffected by damage to the glomeruli. Provided text outlines some of the disease’s most impactful mechanisms alongside critical inflammatory stress response, oxidative injury, and epithelial-mesenchymal transition by detailing the disease pathogenesis and possible intervention points. In addition to the diagnostic functionalities, the clinical consequences of these findings also include precision therapy, more personalized treatment, and adaptive monitoring approaches for the management of diabetic kidney disease. The improved sensitivity and specificity ranges of tubular epithelial cell-

specific biomarkers present the clearest opportunity for early clinical detection, long before conventional markers are able to quantify the level of disease progression, allowing timely intervention with renal protective interventions. The extensive validation in several independent cohorts confirms the reproducibility and clinical applicability of the identified biomarkers across different populations, therefore, affirming the diagnostic credence of the markers. The authors also note salient shortcomings, such as the reliance on pre-collected data and the absence of pre-collected, prospective longitudinal data verification needed for establishing predictive value, along with temporal connections between biomarker expression and clinical outcomes, which require temporal relationships. Future research endeavors should prioritize large-scale prospective studies, standardization of analytical protocols, and investigation of mechanistic relationships between identified biomarkers and therapeutic targets to facilitate clinical translation and improve outcomes for patients with diabetic nephropathy.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The author adheres to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Y. Chen, X. Liu, M. Shengbu, Q. Shi, S. Jiaqiu, and X. Lai, "Biomarkers: New Advances in Diabetic Nephropathy," Natural Product Communications, vol. 20, no. 2, p. 1934578X251321758, 2025. DOI:10.1177/1934578X251321758.

[2] J. Rico-Fontalvo et al., "Novel Biomarkers of Diabetic Kidney Disease," Biomolecules, vol. 13, no. 4, p. 633, Mar 31 2023. DOI: 10.3390/biom13040633.

[3] X. Shao et al., "Machine learning-based multi-omics models for diagnostic classification and risk stratification in diabetic kidney disease," Clin Transl Med, vol. 15, no. 1, p. e70133, Jan 2025. DOI: 10.1002/ctm2.70133.

[4] C. Y. Jung and T. H. Yoo, "Pathophysiologic Mechanisms and Potential Biomarkers in Diabetic Kidney Disease," Diabetes Metab J, vol. 46, no. 2, pp. 181-197, Mar 2022. DOI: 10.4093/dmj.2021.0329.

[5] M. Kiran, Y. Xie, N. Anjum, G. Ball, B. Pierscionek, and D. Russell, "Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis," Front Digit Health, vol. 7, p. 1557467, 2025. DOI: 10.3389/fdgth.2025.1557467.

[6] N. Samsu, "Diabetic Nephropathy: Challenges in Pathogenesis, Diagnosis, and Treatment," Biomed Res Int, vol. 2021, no. 1, p. 1497449, 2021. DOI: 10.1155/2021/1497449.

[7] G. Currie, G. McKay, and C. Delles, "Biomarkers in diabetic nephropathy: Present and future," World J Diabetes, vol. 5, no. 6, pp. 763-76, Dec 15 2014. DOI: 10.4239/wjd.v5.i6.763.

[8] M. Lin et al., "Machine learning and multi-omics integration: advancing cardiovascular translational research and clinical practice," J Transl Med, vol. 23, no. 1, p. 388, Apr 2 2025. DOI: 10.1186/s12967-025-06425-2.

[9] M. B. Lopes et al., "The Omics‐Driven Machine Learning Path to Cost‐Effective Precision Medicine in Chronic Kidney Disease," Proteomics, p. e202400108, 2024. DOI: https://doi.org/10.1002/pmic.202400108.

[10] X. Liu et al., "Integrated multi-omics with machine learning to uncover the intricacies of kidney disease," Brief Bioinform, vol. 25, no. 5, p. bbae364, Jul 25 2024. DOI: 10.1093/bib/bbae364.

[11] M. Concepcion et al., "Novel Biomarkers for the diagnosis of diabetic nephropathy," Caspian J Intern Med, vol. 15, no. 3, pp. 382-391, Summer 2024. DOI: 10.22088/cjim.15.3.382.

[12] J. P. Joumaa et al., "Mechanisms, Biomarkers, and Treatment Approaches for Diabetic Kidney Disease: Current Insights and Future Perspectives," J Clin Med, vol. 14, no. 3, p. 727, Jan 23 2025. DOI: 10.3390/jcm14030727.

[13] B. Yu et al., "Research progress on small extracellular vesicles in diabetic nephropathy," Front Cell Dev Biol, vol. 13, p. 1535249, 2025. DOI: 10.3389/fcell.2025.1535249.

[14] M. M. Rinschen and J. Saez-Rodriguez, "The tissue proteome in the multi-omic landscape of kidney disease," Nat Rev Nephrol, vol. 17, no. 3, pp. 205-219, Mar 2021. DOI: 10.1038/s41581-020-00348-5.

[15] S. Eddy, L. H. Mariani, and M. Kretzler, "Integrated multi-omics approaches to improve classification of chronic kidney disease," Nat Rev Nephrol, vol. 16, no. 11, pp. 657-668, Nov 2020. DOI: 10.1038/s41581-020-0286-5.

[16] Q. Sha, J. Lyu, M. Zhao, H. Li, M. Guo, and Q. Sun, "Multi-Omics Analysis of Diabetic Nephropathy Reveals Potential New Mechanisms and Drug Targets," Front Genet, vol. 11, p. 616435, 2020. DOI: 10.3389/fgene.2020.616435.

[17] H. Liu, J. Feng, and L. Tang, "Early renal structural changes and potential biomarkers in diabetic nephropathy," Front Physiol, vol. 13, p. 1020443, 2022. DOI: 10.3389/fphys.2022.1020443.

[18] J. Yang, D. Liu, and Z. Liu, "Integration of Metabolomics and Proteomics in Exploring the Endothelial Dysfunction Mechanism Induced by Serum Exosomes From Diabetic Retinopathy and Diabetic Nephropathy Patients," Front Endocrinol (Lausanne), vol. 13, p. 830466, 2022. DOI: 10.3389/fendo.2022.830466.

[19] Y.-Y. Yang, Z.-X. Gao, Z.-H. Mao, D.-W. Liu, Z.-S. Liu, and P. Wu, "Identification of ULK1 as a novel mitophagy-related gene in diabetic nephropathy," Frontiers in endocrinology, vol. 13, p. 1079465, 2023. DOI: 10.3389/fendo.2022.1079465.

[20] C. Sabanayagam et al., "Prediction of diabetic kidney disease risk using machine learning models: A population-based cohort study of Asian adults," Elife, vol. 12, p. e81878, Sep 14 2023. DOI: 10.7554/eLife.81878.
[21] F. Mesquita, J. Bernardino, J. Henriques, J. F. Raposo, R. T. Ribeiro, and S. Paredes, "Machine learning techniques to predict the risk of developing diabetic nephropathy: a literature review," J Diabetes Metab Disord, vol. 23, no. 1, pp. 825-839, Jun 2024. DOI: 10.1007/s40200-023-01357-4.
[22] S. M. Swaminathan et al., "Novel biomarkers for prognosticating diabetic kidney disease progression," Int Urol Nephrol, vol. 55, no. 4, pp. 913-928, Apr 2023. DOI: 10.1007/s11255-022-03354-7.
[23] C. Gluhovschi et al., "Urinary Biomarkers in the Assessment of Early Diabetic Nephropathy," J Diabetes Res, vol. 2016, no. 1, p. 4626125, 2016. DOI: 10.1155/2016/4626125.
[24] M. Colombo et al., "Serum kidney injury molecule 1 and β 2-microglobulin perform as well as larger biomarker panels for prediction of rapid decline in renal function in type 2 diabetes," Diabetologia, vol. 62, pp. 156-168, 2019. DOI: 10.1007/s00125-018-4741-9.
[25] A. Alkhalaf et al., "Multicentric validation of proteomic biomarkers in urine specific for diabetic nephropathy," PLoS One, vol. 5, no. 10, p. e13421, Oct 20 2010. DOI: 10.1371/journal.pone.0013421.
[26] M. Lindhardt et al., "Urinary proteomics predict onset of microalbuminuria in normoalbuminuric type 2 diabetic patients, a sub-study of the DIRECT-Protect 2 study," Nephrol Dial Transplant, vol. 32, no. 11, pp. 1866-1873, Nov 1 2017. DOI: 10.1093/ndt/gfw292.
[27] M. Kammer et al., "Integrative analysis of prognostic biomarkers derived from multiomics panels helps discrimination of chronic kidney disease trajectories in people with type 2 diabetes," Kidney Int, vol. 96, no. 6, pp. 1381-1388, Dec 2019. DOI: 10.1016/j.kint.2019.07.025.
[28] E. Soltani-Fard et al., "Urinary biomarkers in diabetic nephropathy," Clin Chim Acta, vol. 561, p. 119762, Jul 15 2024. DOI: 10.1016/j.cca.2024.119762.
[29] J. G. Amatruda et al., "Biomarkers of Kidney Tubule Disease and Risk of End-Stage Kidney Disease in Persons With Diabetes and CKD," Kidney Int Rep, vol. 7, no. 7, pp. 1514-1523, Jul 2022. DOI: 10.1016/j.ekir.2022.03.033.
[30] H. El Alami et al., "Meta-analysis of MTHFR C677T polymorphism and type 2 diabetes mellitus in MENA region," Diabetes Metab Syndr, vol. 18, no. 2, p. 102965, Feb 2024. DOI: 10.1016/j.dsx.2024.102965.
[31] T. Sen et al., "Mechanisms of action of the sodium-glucose cotransporter-2 (SGLT2) inhibitor canagliflozin on tubular inflammation and damage: A post hoc mediation analysis of the CANVAS trial," Diabetes Obes Metab, vol. 24, no. 10, pp. 1950-1956, Oct 2022. DOI: 10.1111/dom.14779.
[32] D. J. Wexler et al., "Comparative Effects of Glucose-Lowering Medications on Kidney Outcomes in Type 2 Diabetes: The GRADE Randomized Clinical Trial," JAMA Intern Med, vol. 183, no. 7, pp. 705-714, Jul 1 2023. DOI: 10.1001/jamainternmed.2023.1487.
[33] P. Bjornstad et al., "Insulin Secretion, Sensitivity, and Kidney Function in Young Individuals With Type 2 Diabetes," Diabetes Care, vol. 47, no. 3, pp. 409-417, Mar 1 2024. DOI: 10.2337/dc23-1818.
[34] K. Kalantar-Zadeh, T. H. Jafar, D. Nitsch, B. L. Neuen, and V. Perkovic, "Chronic kidney disease," Lancet, vol. 398, no. 10302, pp. 786-802, Aug 28 2021. DOI: 10.1016/S0140-6736(21)00519-5.
[35] M. C. Thomas, "Targeting the Pathobiology of Diabetic Kidney Disease," Adv Chronic Kidney Dis, vol. 28, no. 4, pp. 282-289, Jul 2021. DOI: 10.1053/j.ackd.2021.07.001.
[36] N. M. Selby and M. W. Taal, "An updated overview of diabetic nephropathy: Diagnosis, prognosis, treatment goals and latest guidelines," Diabetes Obes Metab, vol. 22 Suppl 1, pp. 3-15, Apr 2020. DOI: 10.1111/dom.14007.
[37] H. J. L. Heerspink et al., "Canagliflozin and Kidney-Related Adverse Events in Type 2 Diabetes and CKD: Findings From the Randomized CREDENCE Trial," Am J Kidney Dis, vol. 79, no. 2, pp. 244-256 e1, Feb 2022. DOI: 10.1053/j.ajkd.2021.05.005.
[38] D. K. McGuire et al., "Effects of empagliflozin on first and recurrent clinical events in patients with type 2 diabetes and atherosclerotic cardiovascular disease: a secondary analysis of the EMPA-REG OUTCOME trial," Lancet Diabetes Endocrinol, vol. 8, no. 12, pp. 949-959, Dec 2020. DOI: 10.1016/S2213-8587(20)30344-2.
[39] S. Shen, C. Ji, and K. Wei, "Cellular Senescence and Regulated Cell Death of Tubular Epithelial Cells in Diabetic Kidney Disease," Front Endocrinol (Lausanne), vol. 13, p. 924299, 2022. DOI: 10.3389/fendo.2022.924299.
[40] Y. Wang, H. Hamid. Reconstructing pharmaceutical service competency framework: development of AI-informed competency indicators and localized practices in China. Future Technology, 4(2), 61–75. DOI: 10.55670/fpll.futech.4.2.7.