ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Deep learning-based multimodal prediction of chronic kidney disease stage

Yan Zhang[1,2*], Xintong Zou[2,3], Zhuoyun Xie[2,3], Haole Huang[2,3], Ruihui Chen[2,3], Zihui Lin[2,3], Kaiting Wang[2,4], Jialin Zhao[2,4], Xiaoqing Liang[2,4], Yuhan Zhang[2,3], Ziyue Xiong[2,3], Xueyan Chen[2,4], Runxi Zhou[2,4], Huimin Li[5*], and Xiaoli Chu[6,7*]

[1]Guangdong Provincial Key Laboratory of Philosophy and Social Sciences, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

[2]Research Center of Intelligent Computing and Big Data Technology, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

[3]Department of Data Science and Big Data Technology, School of Statistics and Data Science, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

[4]Department of Big Data Management and Application, School of Big Data and Artificial Intelligence, Guangdong University of Finance and Economics, Guangzhou, Guangdong, China

[5]Department of Traditional Chinese Medicine Rehabilitation, Boai Hospital of Zhongshan, Zhongshan, Guangdong, China

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

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

(This article belongs to the *Special Issue: Artificial Intelligence in Traditional Chinese Medicine*)

***Corresponding authors:**
Yan Zhang
(zhangyan@gdufe.edu.cn);
Huimin Li
(lihuiminzs@yeah.net);
Xiaoli Chu
(xlchu@gzucm.edu.cn)

Citation: Zhang Y, Zou X, Xie Z, *et al*. Deep learning-based multimodal prediction of chronic kidney disease stage. *Artif Intell Health*. 2026;3(4):025490110. doi: 10.36922/AIH025490110

Received: December 2, 2025

Revised: January 25, 2026

Accepted: January 28, 2026

Published online: April 15, 2026

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

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

Abstract

Chronic kidney disease (CKD) constitutes a critical global public health challenge. Its early-stage symptoms are subtle and easily overlooked, frequently resulting in delayed diagnosis and escalated treatment expenditures. To enable timely early intervention, this study proposes a deep learning-based multimodal CKD stage prediction model, which integrates Western medical laboratory data and traditional Chinese medicine (TCM) symptom descriptions, thereby overcoming the inherent limitations of unimodal prediction approaches. For Western medical data, the synthetic minority oversampling technique (SMOTE) was employed to address class imbalance. Subsequently, a least absolute shrinkage and selection operator (LASSO) feature selection model was utilized to identify key biomarkers, including serum creatinine and serum chloride. A backpropagation neural network, enhanced with Adam optimization and regularization mechanisms, was constructed for predictive modeling. For TCM symptom texts (e.g., tongue manifestations and pulse conditions), the Google-pretrained Bidirectional Encoder Representations from Transformers (BERT) model was leveraged to learn latent semantic patterns in the textual data. Finally, a multimodal decision fusion strategy based on the attention mechanism was adopted to dynamically learn the relative importance of Western medical and TCM features in CKD staging prediction, culminating in the development of the proposed deep learning-driven multimodal CKD staging model. Validation results indicate that the proposed model outperforms classical machine learning models and unimodal models across multiple metrics, including accuracy, F1-score, area under

the curve, and convergence efficiency. These findings confirm its clinical feasibility and effectiveness, providing an innovative multimodal data-driven prediction framework that synergizes the strengths of Western medical quantitative testing and TCM qualitative diagnosis.

Keywords: Multimodal model; Chronic kidney disease; Improved backpropagation neural network; Self-attention mechanism; Bidirectional Encoder Representations from Transformers model

1. Introduction

1.1. Research background and significance

Currently, the number of kidney disease patients in China is close to 120 million, and the prevalence rate of adult chronic kidney disease (CKD) is as high as 10.8%.[1] The early symptoms of CKD are relatively hidden, and patients usually feel normal. Many patients do not go to the hospital until renal failure or uremia occurs. When the disease progresses to the advanced stage, the kidneys have already shown fibrosis, extensive glomerulosclerosis, and renal tubular atrophy. The risks of cardiovascular disease and death in such patients increase significantly, bringing severe disease and economic burdens to individuals and society. The causes of CKD are diverse, including primary glomerulonephritis, hypertension, renal arteriosclerosis, diabetic nephropathy, and secondary glomerulonephritis. CKD has a high prevalence rate, a low early diagnosis rate, and a low treatment rate. Therefore, timely diagnosis is critical.

Clinically, CKD is classified into stages G1 to G5 based on glomerular filtration rate (GFR), which is usually measured using methods such as creatinine clearance and the radioisotope tracer method. However, the diagnostic process is complicated. In primary hospitals, due to limited medical conditions, it is difficult to directly measure the GFR. Usually, auxiliary diagnostic and treatment methods, such as the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation to estimate GFR, urine albumin:creatinine ratio, and routine urine tests, are used to assess CKD stage. However, errors across measurement methods may affect measurement results, and factors such as individual differences and drug interference can make it challenging to diagnose early-stage CKD. Meanwhile, some CKD patients may rapidly deteriorate from G1 or G2 to G4.[2] Therefore, it is necessary to develop auxiliary diagnostic tools to help predict CKD stage.

Currently, various artificial intelligence algorithms are widely used to predict CKD risk. However, most models mainly rely on physical examination data, such as urine protein and serum creatinine. CKD is multifactorial, and it is not feasible to predict CKD solely from such data, as they do not account for multimodal data such as medical diagnoses. In this challenging field, a multimodal CKD stage prediction model based on deep learning can be helpful for the early diagnosis and monitoring of CKD, a more accurate assessment of renal function and pathological changes, and an early intervention in CKD.

1.2. Research status

In CKD research, traditional unimodal data analysis methods include computed tomography (CT) images, plasma protein profiles, genomic data, clinical characteristic data, and biomarker data. Hossain *et al.*[3] used CT images for CKD detection. Dubin *et al.*[4] used plasma protein data from the SomaScan platform in their research on CKD progression. Tin and Köttgen[5] and Robinson-Cohen *et al.*[6] mainly relied on genomic data for Genome-Wide Association Studies analysis of CKD and related traits. In addition, Hannan *et al.*[7] analyzed only the clinical data of patients in the Chronic Renal Insufficiency Cohort (CRIC) study. Ding *et al.*[8] reviewed previous research methods; analyzed the occurrence, risk factors, and epigenetic modifications of vascular calcification in CKD patients; and explored the mechanisms of vascular calcification in CKD and the associated epigenetic modifications. Venkatesan *et al.*[9] used the CKD clinical characteristics dataset from the University of California, Irvine (UCI) database for predictive analysis. Krukowski *et al.*[10] reviewed and analyzed previous research methods, sorted data on gut microbiota composition and uremic toxins, and explored the influence of gut microbiota factors on CKD.

Conventional unimodal analysis methods include deep learning, traditional machine learning algorithms, literature reviews, opinion presentation methods, prospective cohort study methods, systematic reviews, and meta-analyses. Kumar *et al.*[11] used a clinical dataset from Changhua Christian Hospital in Taiwan and a fuzzy deep neural network to predict CKD. Joshi *et al.*[12] conducted

a literature review and used an opinion presentation method to explore the relationship between CKD and dietary patterns. Macdougall[13] analyzed data on anemia treatment in CKD by reviewing clinical research. Świątek *et al.*[14] conducted a systematic review and meta-analysis to examine the impact of a vegetarian diet on CKD progression. Yamada and Nakano[15] used unimodal biomarker analysis to examine the relationship between CKD–mineral and bone disorder (MBD) and cardiovascular disease. Heo[16] used a prospective cohort study design in the United Kingdom Biobank to analyze self-reported questionnaire data. Mapuskar *et al.*[17]–conducted a review to integrate existing research data and explore the relationship between mitochondrial oxidative metabolism and CKD.

Although these methods have provided valuable insights into CKD research, they also share a limitation: they cannot comprehensively capture multidimensional features in complex CKD conditions. Unimodal methods have obvious limitations for investigating the multidimensional characteristics and pathological processes of CKD and are difficult to fully reflect the whole picture of CKD.

As CKD data accumulate continuously, unimodal learning methods can no longer fully exploit the rich information they contain. Scholars have gradually realized that multimodal data can more comprehensively tap the potential of CKD data. Multimodal data integrates information across multiple dimensions, such as imaging, biomarkers, genomics, and clinical data, thereby improving understanding and prediction accuracy of CKD. Saif *et al.*[18] used multimodal medical data from the Taiwan National Health Insurance Research Database and applied deep learning models, such as a convolutional neural network (CNN) and long short-term memory, to improve CKD early diagnosis. Similarly, Alsekait *et al.*[19] used multimodal clinical data from the UCI machine learning library and improved disease risk prediction accuracy through feature selection and deep learning integration. Akter *et al.*[20] combined clinical data from the Korean National Health Insurance Sharing Service website and applied supervised multitask learning, recurrent neural networks (RNNs), and multilayer perceptron (MLP) models to improve the applicability of precision medicine. Regarding imaging data, Ge *et al.*[21] combined ultrasound images with clinical characteristics and demonstrated robust predictive performance across different cohorts. Hou *et al.*[22] combined CT images and clinical features using the 3D U-Net architecture, thereby effectively improving the accuracy of renal health assessment. Chertow *et al.*[23] and Evenepoel *et al.*[24] enhanced the clinical applicability and reliability of model verification results through multicenter clinical trials and the integration of multiomics data. In a study of the relationship between gut microbiota and CKD, Ren *et al.*[25] combined multiple data sources and used two-sample Mendelian randomization and sensitivity analyses to deeply explore the causal relationship, strengthening the conclusions.

In addition to multimodal data fusion, method diversification offers greater opportunities in CKD research. In multimodal methods, deep learning (such as CNNs, RNNs, and MLPs), integrated learning, transfer learning, and multitask learning are widely used and have become important tools to improve model performance. For example, Venkatrao and Kareemulla[26] used the HDLNet model in their research and combined multiple data sources to ensure high model performance in auxiliary diagnosis. Joo *et al.*[27] combined deep learning algorithms with the Cox proportional hazards model to improve the accuracy of early CKD risk assessment and demonstrated the model's robustness across different datasets. Similarly, Liang *et al.*[28] combined multiple data sources with multiple machine learning and deep learning models, demonstrating excellent stability and consistency in disease risk prediction. Yang *et al.*[29] combined SERS spectroscopy data with deep learning models (such as improved AlexNet and ResNet) in research on immune diseases and CKD diagnosis to improve the accuracy of early disease detection. Bhattacharjee *et al.*[30] combined multi-source data from the lungs and kidneys and used the Xception deep learning model to improve early diagnosis of multiple diseases. Zhang[31] combined imaging data (e.g., ultrasound, CT, and magnetic resonance imaging) with deep learning architectures to support personalized medicine and enhance stability through data augmentation.

In clinical modeling tasks such as CKD prediction and staging, in addition to high predictive performance, model interpretability and feature learning paradigms have increasingly become important factors influencing their clinical application. Recently, Arif *et al.*[32] proposed an interpretable machine learning model for CKD risk assessment, achieving transparent interpretation of prediction results by analyzing the contributions of key clinical variables to the model output, thereby providing a solid research foundation for interpretable learning in data-driven CKD modeling. Furthermore, in another study, Arif *et al.*[33] constructed a robust machine learning framework for CKD early detection, enhancing model performance through systematic data preprocessing, feature selection, and hyperparameter optimization, underscoring the importance of the "high-quality feature engineering + robust modeling" approach in CKD risk prediction. These studies indicate that traditional supervised methods alone are insufficient to fully meet the demands of complex

clinical tasks, and that the choice of learning paradigm and feature representation capability are essential components in improving model generalizability and trustworthiness.

From the perspective of dependency on labeled data, machine learning methods can be divided into three categories: supervised, unsupervised, and semi-supervised. Supervised learning relies on large amounts of labeled samples. It optimizes an objective function to directly fit the prediction target, performing well in specific tasks but often limited by high labeling costs and cross-center data distribution discrepancies. In contrast, unsupervised deep learning has gained increasing attention in medical data analysis in recent years, aiming to learn latent structures or representations from high-dimensional unlabeled data to support subsequent analysis. Li *et al.*[34] employed unsupervised representation learning to extract patient features from sequential electronic health records (EHRs) and identified clinical subgroups through cluster analysis, revealing differences in clinical characteristics across patient populations and aiding understanding of underlying disease heterogeneity. More specifically, the "deep patient" study by Miotto *et al.*[35] adopted stacked denoising autoencoders to learn general patient representations, outperforming traditional feature-based methods in multiple disease prediction tasks, confirming the potential value of unsupervised representation learning for downstream predictive tasks. Moreover, in physiological signal modeling, such as electrocardiogram signals, Friedman *et al.*[36] demonstrated that unsupervised autoencoder-based approaches can generate richer latent representations that capture more disease-relevant patterns than conventional features, providing empirical support for the application of unsupervised learning in physiological data modeling. In summary, these studies indicate that unsupervised deep learning offers unique value in feature learning, latent structure discovery, and disease pattern mining.

However, as unsupervised methods do not inherently rely on manual labels, they typically focus more on exploring the data's latent structure than on directly optimizing explicit predictive objectives. In practical medical prediction tasks, purely unsupervised learning still faces challenges in terms of interpretability, prediction accuracy, and model stability. Given this, semi-supervised learning balances the contradiction between labeling requirements and predictive objectives by leveraging a small amount of high-quality labeled data and a large amount of unlabeled data through strategies such as consistency regularization and pseudo-labeling, making it more suitable for clinical prediction modeling. Given the current landscape, this study simultaneously addresses the accuracy and clinical interpretability of CKD staging prediction. It adopts a semi-supervised learning framework with manual verification and bias control and introduces a hybrid feature extraction strategy to enhance the model's ability to represent heterogeneous clinical features and improve prediction stability.

From the existing literature, integrating and analyzing multimodal data can not only fully leverage information from different data sources but also significantly improve the accuracy and robustness of the prediction model.[37] Despite significant progress, few scholars have conducted systematic, effective studies that combine Western medical test data with traditional Chinese medicine (TCM) symptom data. Therefore, this study aims to propose a multimodal CKD stage prediction model based on deep learning. The laboratory results from Western medicine and the disease symptom descriptions in TCM (including tongue manifestations, pulse conditions, and other symptoms unique to TCM) were selected to fully leverage the advantages of the combination of TCM and Western medicine and to significantly improve the accuracy and robustness of the prediction, thereby having important theoretical and application value for expanding the multimodal CKD stage prediction method. Disease symptom descriptions in TCM were selected because TCM employs multiparametric assessment, including inspection, auscultation, olfaction, inquisition, and palpation (MAP), to obtain informative cues for pattern differentiation and treatment.[38] An improved backpropagation (BP) neural network model was used to process Western medical blood examination results, and the Google-pretrained Bidirectional Encoder Representations from Transformers (BERT) model was used to process TCM symptom descriptions. Finally, the two models were fused to predict CKD stage.

First, the dataset was preprocessed to identify complications and symptoms highly correlated with CKD, and changes in patients' complications were analyzed. Since CKD has no apparent symptoms in the early stages and the number of patients seeking medical attention is small, the sample suffers from an imbalanced distribution, with a significantly higher number of late-stage patients than early-stage ones. Therefore, we applied the synthetic minority oversampling technique (SMOTE) to augment the clinical laboratory examination data. Next, we used the least absolute shrinkage and selection operator (LASSO) to select appropriate clinical indicators as input features for the model and established an improved BP neural network for CKD staging prediction, augmenting symptom description data to alleviate the issue of imbalanced sample categories. Then, we used a self-attention

mechanism to focus on patterns between symptoms and between symptoms and CKD stages in the diagnostic data, identifying potential symptom combinations that indicate mild, moderate, or severe CKD. BERT was built to deeply mine semantic information from textual data, enhancing the model's understanding of symptom descriptions for CKD staging prediction. Finally, we employed an attention-based multimodal fusion strategy[39] to adaptively learn the importance weights of Western medicine and TCM features in CKD stage prediction. On this basis, we constructed a deep learning-driven multimodal CKD stage prediction model, which significantly enhanced the model's decision-making performance, interpretability, and generalization capability.

This paper has the following contributions: (i) A new strategy for CKD stage prediction based on simplified indicators is proposed by combining six key blood indicators, such as creatinine and chloride, effectively distinguishing CKD stages. (ii) A multimodal decision fusion strategy based on an attention mechanism is proposed. By adaptively learning the weight distributions of different modal features, the model dynamically integrates quantitative indicators from Western medicine and qualitative symptoms in TCM, significantly enhancing its interpretability and decision-making performance. (iii) A multimodal CKD stage prediction model combining TCM and Western medicine is proposed, which comprehensively utilizes laboratory examination results and disease symptom descriptions in TCM, combines the concepts of Western medicine diagnostics and TCM symptomatology, and improves the accuracy of medical diagnosis.

2. Methods

2.1. Overall framework of the model

This section introduces the overall framework of the proposed model, aiming to achieve more accurate and reliable CKD stage prediction through deep learning and multimodal fusion. The model adopts a modular design to build a multimodal prediction model in stages, represented in Figure 1. The model comprises:

- Input layer: Collect and preprocess multimodal data, including laboratory examination results (e.g., creatinine, chloride, and other blood examination results) and textual data (e.g., disease symptom descriptions), and pass them to the processing layer.
- Processing layer:
  (i) Data augmentation module. Augment the input data to address imbalanced sample sizes, provide rich data for model training, and analyze the main complications of CKD.
  (ii) Model training module. Based on an improved BP neural network, improve the traditional network structure and use blood examination data as model features. Based on the Google-pretrained BERT, leverage the self-attention mechanism to enhance understanding of symptoms and treat symptom descriptions as model features. The two models are trained independently using unimodal features.
- Fusion layer: Using the attention mechanism to adaptively weight and fuse the outputs of two independent models, the optimal combination of multimodal features is achieved through end-to-end learning, yielding a multimodal prediction model.
- Output layer: Use the multimodal prediction model to output the final CKD stage prediction result for the input multimodal features.

2.2. Data augmentation module

2.2.1. Data augmentation of blood examination data

We increased the number of training samples for blood examination data. SMOTE was used to address class imbalance in the current CKD samples.

To mitigate data leakage and ensure the objectivity of model performance evaluation, the raw data were first partitioned into a training set and a test set at a 7:3 ratio. All samples in the test set were sourced from real-world clinical data and were not subjected to any form of data augmentation. SMOTE was exclusively applied to the training set and independently executed within each training fold during model training (e.g., in cross-validation scenarios). This ensures that synthetic samples do not contaminate the test set or introduce information leakage, thereby guaranteeing the objectivity and reliability of model performance assessment.

The specific steps are as follows:

(i) Select a minority class sample x_i.
(ii) Find the k-nearest neighbors (KNNs) of x_i, and randomly select a sample x_j from the KNNs.
(iii) Generate a new synthetic sample x_{syn}, and its feature value is a linear interpolation between x_i and x_j. The generation process is represented as Equation 1:[40]

$$x_{syn} = x_i + \lambda(x_j - x_i) \quad \lambda \in [0, 1] \tag{1}$$

Before data augmentation, the proportions of samples for CKD stages 1 to 5 were 3.1%, 4.1%, 10.8%, 27.3%,

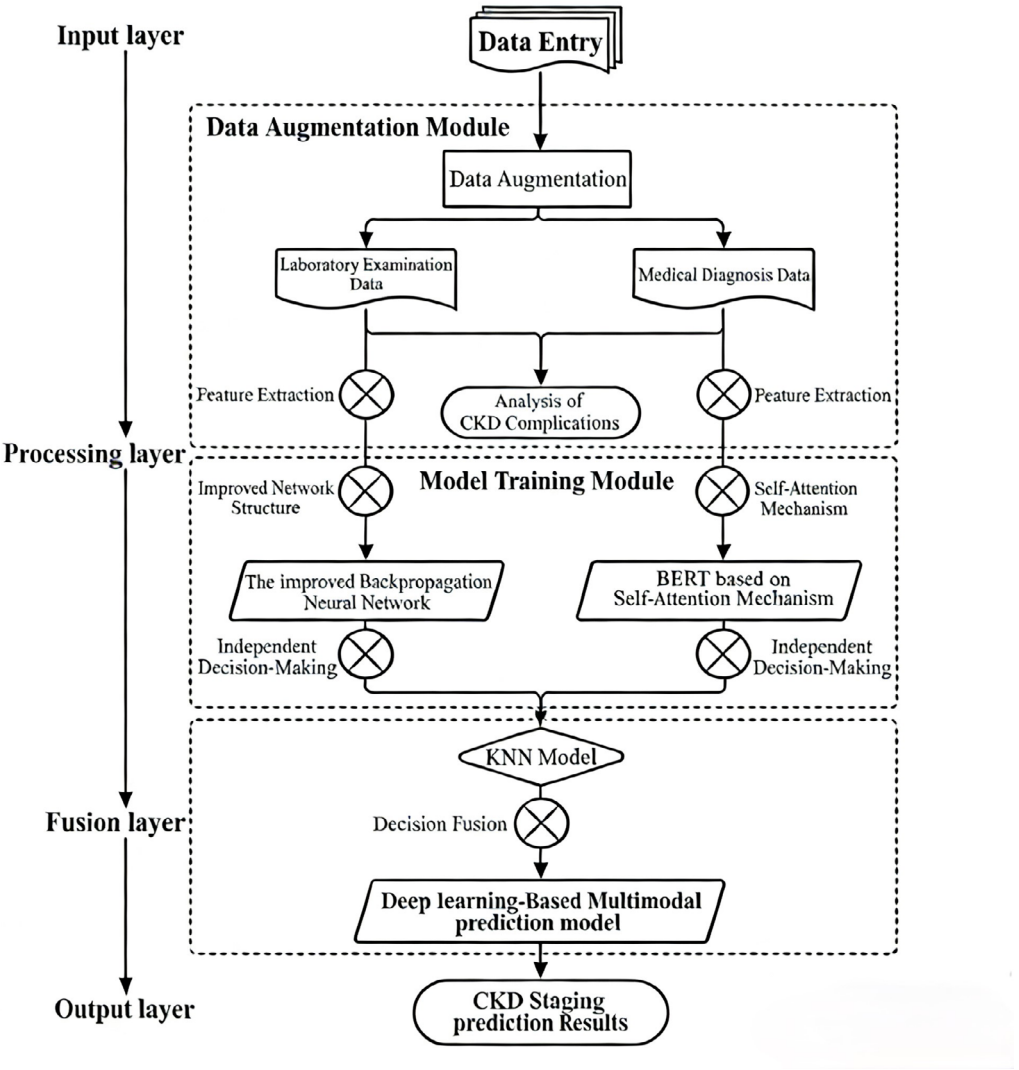


Figure 1. Overall framework diagram of the model
Abbreviations: BERT: Bidirectional Encoder Representations from Transformers; CKD: Chronic kidney disease; KNN: K-nearest neighbor.

and 54.7%, respectively. After applying SMOTE data augmentation to the blood examination indicators selected for feature selection, the sample sizes for stages G1 to G4 were increased to match that of stage G5, ensuring consistent sample sizes across stages.

2.2.2. Data augmentation of medical diagnosis data

The improved BP neural network was trained only on samples with clearly marked stages, and no unlabeled samples were used. To address the issue of some samples lacking CKD stage labels, a semi-supervised method was adopted to supplement the data. After the training was completed, the model was used to make a one-time prediction on samples without labeled stages, and a corresponding confidence score was output for each prediction. To reduce confirmation bias introduced during the self-labeling process, only samples with a prediction confidence of at least 75% were retained as pseudo-labeled data. The specific algorithm steps are shown in Algorithm A1 in the appendix.

The stage labels derived from the model predictions in this section were not used for either retraining the model or evaluating its performance. We entered the model-predicted stage results into the information database, without stage labels. On the independent validation set of labeled samples, the model achieved 98% accuracy in CKD staging prediction. This result indicates that, under the constraint of high confidence, the model's predictions can serve as a reasonable approximation of the ground-truth staging.

2.2.3. Analysis of chronic kidney disease complications

We used Word2Vec to represent the medical text describing complications and cosine similarity to quantify semantic similarity between embeddings. Subsequently, we applied principal component analysis (PCA) for dimensionality reduction and visualization.

First, we entered the text into Word2Vec. It is a technology for converting words into vectors. By training a shallow neural network to learn the distributed representation of words, it can capture the semantic relationships between words. Word2Vec has two main architectures for calculating the word embedding matrix: CBOW and Skip-Gram. Among them, CBOW uses context words to predict the center word, and Skip-Gram uses the center word to predict the context words. The training steps of Word2Vec are as follows:

(i) Convert the vocabulary into one-hot vector form.
(ii) Randomly initialize the weight matrices between the input layer and the hidden layer, as well as between the hidden layer and the output layer.
(iii) Adjust the hyperparameters of the model, such as the learning rate, window size, and word embedding dimension, to optimize the model performance.
(iv) For CBOW, calculate the average vector of the context words, the output according to the weight matrix, and the loss according to the cross-entropy loss function.
(v) Calculate the gradient of the weight matrix according to the loss function, and update the weight matrix using gradient descent.
(vi) Repeat the forward propagation and backward propagation processes until the loss function converges. After training, the vector representation of each word is extracted from the weight matrix. Then, the cosine similarity, as defined in Equation 2, calculates the similarity between two words. The closer the angle between two vectors is to 0, the closer the cosine value is to 1, indicating a higher similarity between the two vectors.

$$\cos\theta = \frac{\sum_1^n (A_i \times B_i)}{\sqrt{\sum_1^n A_i^2} \times \sqrt{\sum_1^n B_i^2}} \quad (2)$$

where A and B are n-dimensional vectors. In natural language processing, they refer to two word embeddings, and the cosine similarity represents the correlation between words.

Compared with the Euclidean distance metric, the cosine similarity pays more attention to the difference in direction between two vectors rather than the difference in distance or length, so it is more suitable for calculating the correlation of text.

Finally, since the word embeddings were high-dimensional vectors, it was necessary to perform dimensionality reduction analysis using PCA. PCA converts a set of variables into a set of linearly uncorrelated new variables, called principal components, through an orthogonal transformation, represented in Equation 3. PCA is often used for data dimensionality reduction. It can reduce the complexity of the dataset while retaining the variability of the original data as much as possible.

$$\Sigma v = \lambda v \quad (3)$$

where Σ is the covariance matrix, v is the eigenvector representing the direction of the principal component, and λ is the eigenvalue representing the variance of the principal component.

Based on this, we analyzed the correlations between various CKD complications. On one hand, for any two complications, if the cosθ value in Equation 2 is closer to 1, it indicates a stronger correlation between the two complications. On the other hand, all word vectors were reduced in dimensionality through PCA, retaining the first and second principal components, and projecting all complication word vectors into a 2D space to visually inspect the correlations between each complication word vector. Specifically, if the data for hypertension and diabetes mellitus are close to each other in the space, it suggests that these two conditions may have a strong correlation in CKD patients.

2.3. The improved backpropagation neural network

2.3.1. Extraction of Western medicine clinical indicators

Regarding the laboratory examination data, due to an excessive number of original features in the samples, the model is prone to overfitting on the training data. Therefore, it is necessary to select the features that are truly valuable for model prediction. After selecting six main numerical features, they are applied to all stages of the dataset. The training, validation, and test sets were divided at the sample level and shared the same feature space.

In the process of feature selection for numerical clinical indicators, due to the high dimensionality of the features and potential correlations among indicators, directly inputting all features into a neural network model is prone to overfitting. Therefore, we adopted LASSO regression with L1 regularization for feature selection.

Let the number of training samples be n, the original feature dimension be p, the j feature of the i sample be x_{ij}, and the corresponding predicted label be y_i. Then, the optimization objective function of the LASSO regression model is defined as follows:[41]

$$\hat{\beta} = arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \quad (4)$$

where β is the regression coefficient vector and λ ≥ 0 is the regularization parameter that controls the sparsity of the model.

In the feature selection stage, by solving the above optimization problem, when the regularization parameter λ attains its optimal value, the features with non-zero regression coefficients are considered to make a substantial contribution to the prediction task and are retained; those with zero regression coefficients are considered redundant and are eliminated. Therefore, the final selected feature set can be expressed as:

$$F = \{x_j \mid \beta_j \neq 0, j = 1, 2, \ldots, \quad (5)$$

Based on LASSO, Figure 2 presents the clinical indicators with the top 50% importance index.

As shown in Figure 2, the feature with the highest importance index is creatinine (0.103), followed by chloride (0.069). We used the top 10% importance indicators as the model's features, as shown in Table 1: creatinine, chloride, lymphocytes, red blood cells, hemoglobin, and blood urea nitrogen (BUN).

Table 1. Feature importance ranking

Importance ranking	Feature	Unit	Gini importance index
1	Creatinine	μmol/L	1.0368×10^{-1}
2	Chloride	mmol/L	6.9958×10^{-2}
3	Lymhocytes	10^9/L	4.1052×10^{-2}
4	Red blood cell	10^{12}/L	3.4246×10^{-2}
5	Hemoglobin	g/L	2.9959×10^{-2}
6	Blood urea nitrogen	mmol/L	2.6207×10^{-2}

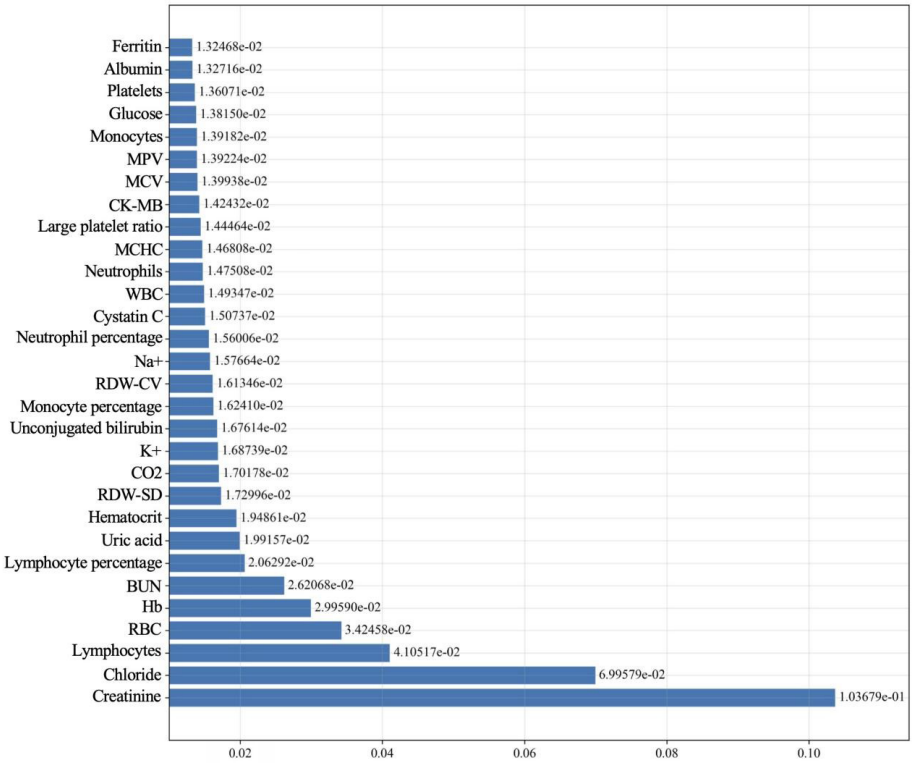


Figure 2. Bar chart of feature importance indices where the summation of all feature importance indices is 1
Abbreviations: BUN: Blood urea nitrogen; CK-MB: Creatine kinase-MB; Hb: Hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; MCV: Mean corpuscular volume; MPV: Mean platelet volume; RBC: Red blood cell; RDW-CV: Red cell distribution width-coefficient of variation; RDW-SD: Red cell distribution width-standard deviation; WBC: White blood cell.

These features are all clinical indicators of blood. Specifically, creatinine and BUN reflect impaired renal clearance and azotemia, respectively. Chloride and red blood cell levels are directly related to electrolyte imbalance and anemia, which are common complications of CKD. These indicators not only provide an indirect assessment of GFR but also reflect the multi-system pathological changes in CKD patients, offering multidimensional medical evidence for staging prediction. The feature values were normalized to account for differences in scale and units across features, and the resulting data were used as the input vector for the prediction model.

Figures 3 and 4 illustrate the characteristics of the numerical model. As shown in Figure 3, the Shapley Additive Explanations (SHAP) distribution of creatinine indicates that the higher the creatinine values (red dots), the greater the corresponding SHAP values, meaning that high creatinine significantly increases the model's predicted CKD stage, demonstrating that serum creatinine is the most important feature. This is consistent with medical knowledge, as serum creatinine directly reflects glomerular filtration function: it rises during kidney dysfunction. Serum chloride also showed relatively high importance, with high values (red) typically corresponding

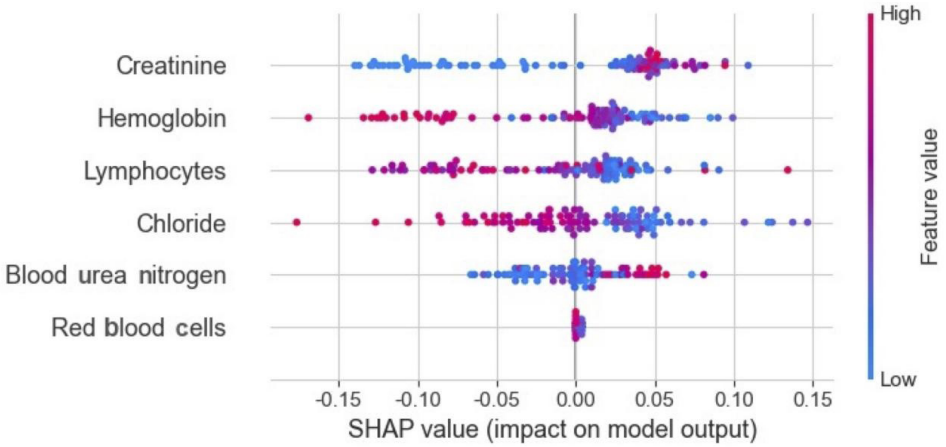


Figure 3. Shapley Additive Explanations (SHAP) summary plot of the numerical model

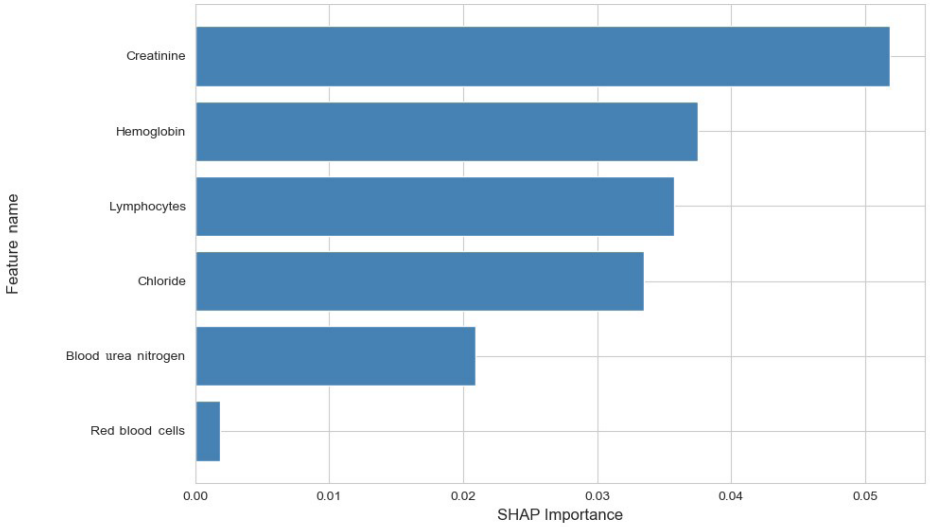


Figure 4. Feature importance of the numerical model

to positive SHAP values, suggesting that higher chloride levels are associated with the model's prediction of more severe CKD stages. Clinical studies have shown that high serum chloride levels are associated with an accelerated decline in estimated GFR and are a risk marker for CKD progression.[42] Lymphocyte count was of moderate importance, with high lymphocytes (red) in the SHAP plot often corresponding to low or negative SHAP values, and low lymphocytes (blue) corresponding to positive SHAP values, implying that lymphopenia is associated with more severe CKD, which aligns with the common low-grade inflammation and immune dysfunction in CKD patients. Other blood indicators (such as white blood cells, urea nitrogen, and hemoglobin) in Figure 4 showed relatively lower importance and were generally ranked lower in the SHAP ranking. Overall, the numerical model focuses on biochemical indicators directly related to renal function and homeostasis, highly consistent with the physiological mechanisms of CKD.

2.3.2. Construction of the network model

Considering the diversity of clinical data in blood examinations, the individual differences among patients, and other multidimensional factors, this paper selected the MLP-based BP neural network as the numerical feature modeling model and optimized the traditional BP neural network through training and regularization. By introducing the Adam optimizer, the Dropout mechanism, and L2 regularization, the convergence stability of the model during BP was effectively improved, thereby reducing prediction error. The model training process is as follows:

(i) Initialize the fully connected neural network and set a random dropout layer between the hidden layers to improve the robustness and generalization ability of the model.

(ii) During forward propagation, the input features are nonlinearly mapped through fully connected hidden layers. Each neuron in a hidden layer is fully connected to the neurons in the previous layer, thereby performing weighted combinations of the input features and learning latent feature representations of the data layer by layer.

(iii) As the number of network layers increases, the model provides a higher-level, more abstract representation of input features while preserving the fully connected architecture, thereby enhancing its ability to model features associated with CKD staging.

(iv) Calculate the loss function as in Equation 6, using the predicted value obtained from forward propagation and the sparse categorical cross-entropy, evaluate the model performance, and provide the direction for model optimization.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\hat{y}_{i,\ c_i}\right) \tag{6}$$

where N is the number of samples, and $\hat{y}_i$ is the ith element of the predicted vector.

(v) Calculate the gradient of the loss function concerning the weights and biases as in Equations 7 and 8 using BP to reduce prediction error.

$$\nabla w_{ji}^{(l)} = \frac{\partial \mathrm{L}}{\partial w_{ji}^{(l)}} = \frac{\partial \mathrm{L}}{\partial a_i^{(l)}} \cdot \frac{\partial a_i^{(l)}}{\partial z_i^{(l)}} \cdot \frac{\partial z_i^{(l)}}{\partial w_{ji}^{(l)}} \tag{7}$$

$$\nabla b_i^{(l)} = \frac{\partial \mathrm{L}}{\partial b_i^{(l)}} = \frac{\partial \mathrm{L}}{\partial a_i^{(l)}} \cdot \frac{\partial a_i^{(l)}}{\partial z_i^{(l)}} \cdot \frac{\partial z_i^{(l)}}{\partial b_i^{(l)}} \tag{8}$$

where L is the loss function, $\nabla w_{ji}^{(l)}$ and $\nabla b_i^{(l)}$ are the gradients of the weight and bias terms, respectively, and $z_i^{(l)}$ is the net input of the ith neuron in the lth layer.

(vi) Use the Adam optimizer to adjust the model's parameters adaptively and improve convergence performance[43].

$$m_t = \beta_1 m_{t\text{-}1} + \left(1 - \beta_1\right) g_t$$

$$v_t = \beta_2 v_{t\text{-}1} + \left(1 - \beta_2\right) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$W_t = W_{t\text{-}1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

where m_t is the first-order moment estimate of the gradient, v_t is the second-order moment estimate of the gradient, $\hat{m}_t$ is the corrected first-order moment estimate, $\hat{v}_t$ is the corrected second-order moment estimate, W_t is the updated weight, α is the learning rate, β_1 and β_2 are the exponential decay rates, and ϵ is the numerical stability term.

(vii) Use L2 regularization to prevent overfitting, as expressed in Equation 10.

$$\mathrm{L2} = \lambda \sum_{i=1}^{n} w_i^2 \tag{10}$$

where λ is the regularization parameter, and w_i is the *i*th element of the weight vector.

(viii) Update the weights and biases of the network using the calculated gradients and the predetermined learning rate, as in Equations

11 and 12. Then use them for subsequent forward propagation and iterate to optimize the performance of the network.

$$w_{ji}^{(l)} = w_{ji}^{(l)} - \alpha \nabla w_{ji}^{(l)} \quad (11)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \nabla b_i^{(l)} \quad (12)$$

where α is the learning rate.

The normalized blood test indicators (creatinine, chloride, lymphocytes, red blood cells, hemoglobin, and BUN) were used as input parameters for the model. Based on these input parameters, we constructed an improved BP neural network to learn the impact of different blood examination indicators on CKD staging. To prevent overfitting in CKD staging predictions, we applied L2 regularization (λ = 0.001) and a 20% random dropout strategy during training. The Adam optimizer was used to update the network's weights and biases to improve CKD staging prediction. After training, we obtained an improved BP neural network model that outputs CKD staging predictions (Figure 5).

2.4. Google-pretrained Bidirectional Encoder Representations from Transformers

2.4.1. Grading of traditional Chinese medicine disease symptoms

Traditional Chinese medicine disease symptoms are abnormal sensations and pathological changes experienced by patients and are one of the important bases of TCM diagnosis. TCM disease symptoms can provide information that is difficult to obtain through Western medicine diagnosis. Among them, the observation of tongue manifestations and pulse conditions in TCM is helpful for early disease detection and aligns with the principle of "preventive treatment of disease," emphasizing intervention prior to disease onset. For example, in early-stage CKD, the GFR and other indicators may not have decreased significantly, but slight changes in the patient's tongue coating may indicate early dysfunction of the spleen and stomach, and the slight changes in the pulse condition may indicate poor qi and blood circulation. Through these early signals, physicians can intervene promptly to prevent disease development. Therefore, TCM disease symptoms have unique advantages for predicting CKD stage and can provide important guidance for early identification, diagnosis, and treatment. At the same time, combining Western medical diagnosis with TCM symptomatology can complement diagnostic information and help comprehensively understand the patient's condition, thereby improving prediction accuracy.

We graded the symptoms of TCM diseases by severity and used them as input data for the model. In the information database, all symptom manifestations were divided into four severity levels, as shown in Table 2. After processing the information database, 27 disease symptoms were identified. After selecting 27 text-type features, they were applied to all stages of the dataset. The training, validation, and test sets were split at the sample level and shared the same feature space.

In Figure 6, the average importance of key terms, such as renal insufficiency and proteinuria, was relatively high. Renal insufficiency directly reflects decreased kidney function, the core pathology of CKD. According to the definition of CKD, a long-term reduction in GFR determines the disease stage. Hence, the model assigns a

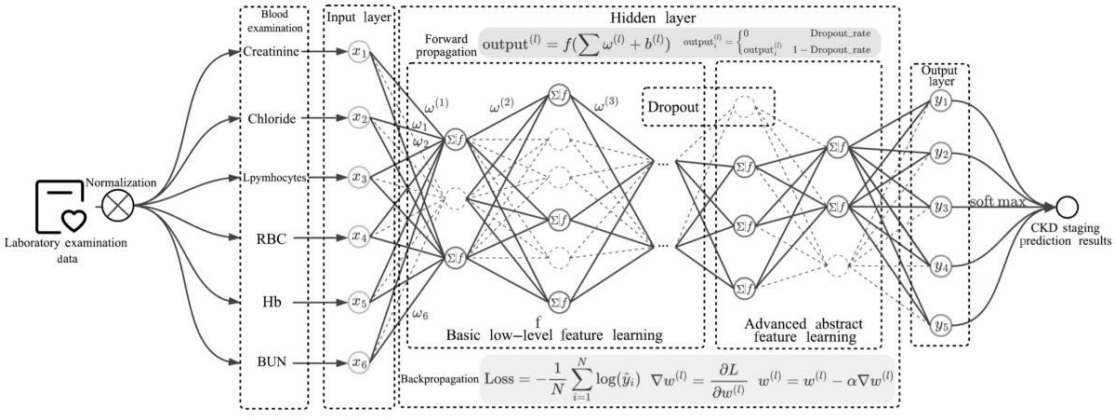


Figure 5. Processing flow of the improved backpropagation neural network
Abbreviations: BUN: Blood urea nitrogen; Hb: Hemoglobin; RBC: Red blood cell.

Table 2. Symptom grading table

Symptom	Symptom grade			
Red or slightly red tongue, little coating	Pale red, moist tongue, thin white coating, evenly distributed	Slightly red tongue, thin white coating, slightly dry	Obviously red tongue, little coating, dry tongue surface	Deep red tongue, little or no coating, dry tongue surface, even cracked
Thin deep pulse	Pulse position moderate, not floating, not sinking, pulse channel moderate, pulsation clear	Pulse position slightly deep, needs light pressing to feel, pulse channel slightly thin, pulsation clear	Pulse position obviously deep, needs heavy pressing to feel, pulse channel thin like a thread, pulsation weak	Pulse position deep, needs heavy pressing to the bone to feel, pulse channel extremely thin, pulsation weak, seeming to exist or not at all

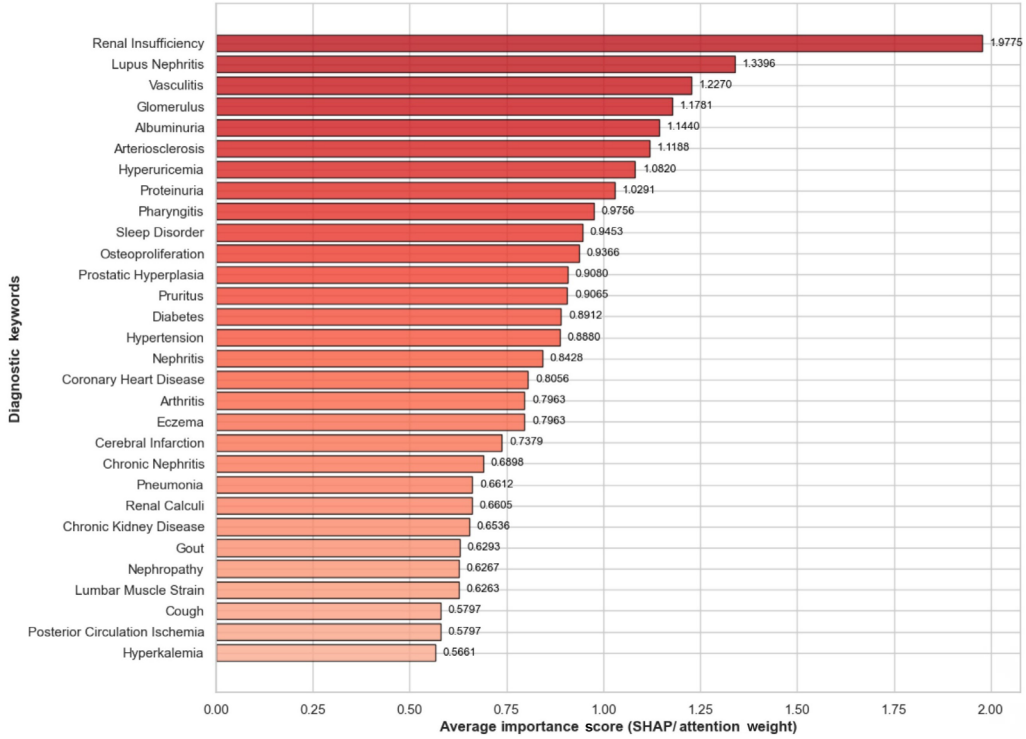


Figure 6. Keyword importance ranking of the BERT model for chronic kidney disease classification
Abbreviations: BERT: Bidirectional Encoder Representations from Transformers; SHAP: Shapley Additive Explanations.

high weight to renal insufficiency. Proteinuria is a hallmark symptom of kidney damage and disease progression. Numerous studies have shown that proteinuria is a predictor of CKD progression and prognosis.[44] The model's sensitivity to these professional terms indicates that it has learned to capture key information directly related to the severity of CKD from clinical texts. Other terms related to CKD progression (such as uremia and dialysis) might also be of concern, but statistical results showed that renal insufficiency and proteinuria were the most representative.

In summary, both numerical and text models exhibited a high degree of consistency with clinical medical understanding. Numerical models mainly rely on biochemical indicators reflecting kidney function (such as creatinine and chloride) and indicators of the inflammatory state, whereas text models focus on core concepts such as kidney function impairment and proteinuria. These features are the most critical factors in CKD staging and prognosis. The model output is consistent with clinical expectations: for example, when creatinine levels rise and

proteinuria occurs, the model predicts a higher CKD stage, which aligns with actual clinical logic. Therefore, these explainable analysis results verify the medical rationality and credibility of the model, providing support for its application in clinical settings.

2.4.2. Construction of the language model

Compared with other model structures, the self-attention mechanism can focus on key disease symptoms and assign higher weights, thereby playing an important role in the model's decision-making. At the same time, the self-attention mechanism can attend to combinations of symptoms to identify which combinations are more likely to indicate mild, moderate, or severe CKD. Therefore, the Transformer architecture equipped with the self-attention mechanism can learn the mapping relationship between TCM disease symptoms and CKD stages, enabling the prediction model to learn the relationship between symptoms from multiple angles, enrich the feature representation, and improve the model's ability to model the complex conditions of CKD, thereby improving CKD stage prediction accuracy.

We adopted BERT to capture context information on both sides at all levels simultaneously, thereby pretraining a deep, bidirectional text representation. The BERT-base, Chinese-pretrained model released by Google, which was pretrained on a large-scale general Chinese corpus, was used. Although this model was not specifically pretrained for TCM or medicine, it has demonstrated excellent generalization in Chinese-language understanding tasks. We stacked a classifier layer on BERT and fine-tuned it with task-related data to enable the model to learn feature representations relevant to TCM texts. To effectively integrate the description of TCM disease symptoms into this model, we used a BERT tokenizer to encode the text. The tokenizer divided the text into words or more fine-grained sub-word units and assigned a unique identifier to each unit. The encoding process is as follows and shown in Figure 7:

(i) Obtain the word embedding representation of the symptom description and map the text to a continuous vector space.
(ii) Use position encoding to retain the order information of the input text.
(iii) Use the multilayer Transformer encoder to encode the input text.
(iv) Take the average value of each dimension of the output of the encoder layer for average pooling processing to obtain the overall feature representation of the input text.
(v) Obtain the representation of the overall feature of the text after average pooling as the input tensor format of the BERT model.

In clinical diagnosis, a patient's chief complaints (such as a red tongue or fine, deep pulse) often provide important clues for the early identification of CKD. The

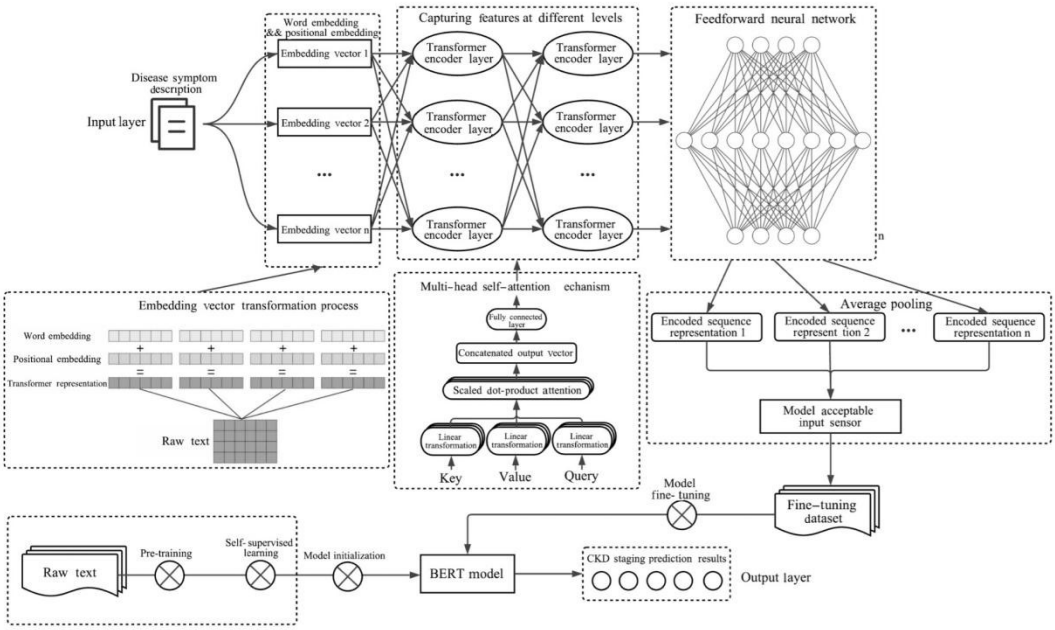


Figure 7. Processing flow of Google-pretrained Bidirectional Encoder Representations from Transformers (BERT)

BERT model can analyze patient symptom descriptions to uncover the relationship between symptom patterns and CKD staging. First, the BERT tokenizer processes the symptom text in EHRs, tokenizing it and converting it into embedded representations. For example, "the patient complains of increased nocturia" can be split into multiple tokens, which BERT then uses to learn contextual information. In the Transformer encoding layer, the self-attention mechanism captures the global dependencies between symptoms. For example, the co-occurrence of edema and gout may be related to CKD progression. Subsequently, the feedforward neural network processes the self-attention mechanism's output to map the symptom information to the corresponding CKD stage category. During model training, we used a fine-tuning strategy to optimize the parameters based on CKD-specific data, ultimately resulting in a BERT model capable of predicting CKD staging.

2.5. A multimodal stage prediction study of chronic kidney disease based on deep learning

2.5.1. Multimodal fusion strategy

To enhance the accuracy and interpretability of CKD staging prediction, we deeply integrated the feature representations of two modalities. Unlike simple decision-level fusion, this study adopted a feature-level fusion strategy that adaptively learns the relative importance of different modalities for CKD staging prediction through an attention mechanism. This approach can better exploit the complementary information between Western medical quantitative indicators and TCM qualitative symptoms, while providing an interpretable basis for model decisions (Figure 8).

2.5.2. Attention mechanism-based multimodal feature fusion[45]

(a) Feature representation extraction

Let the Western medicine feature extracted based on the improved BP neural network be represented as Hb ∈ Rdb, and the TCM feature extracted by the BERT model based on the self-attention mechanism be represented as Ht ∈ Rdt. Firstly, the two features were mapped to the same dimensional space through a fully connected layer:

$$Fb = WbHb + bb, \qquad Ft = WtHt + bt \tag{13}$$

where $F_b, F_t \in R^{d_f}$ are feature representations in the same dimensional space.

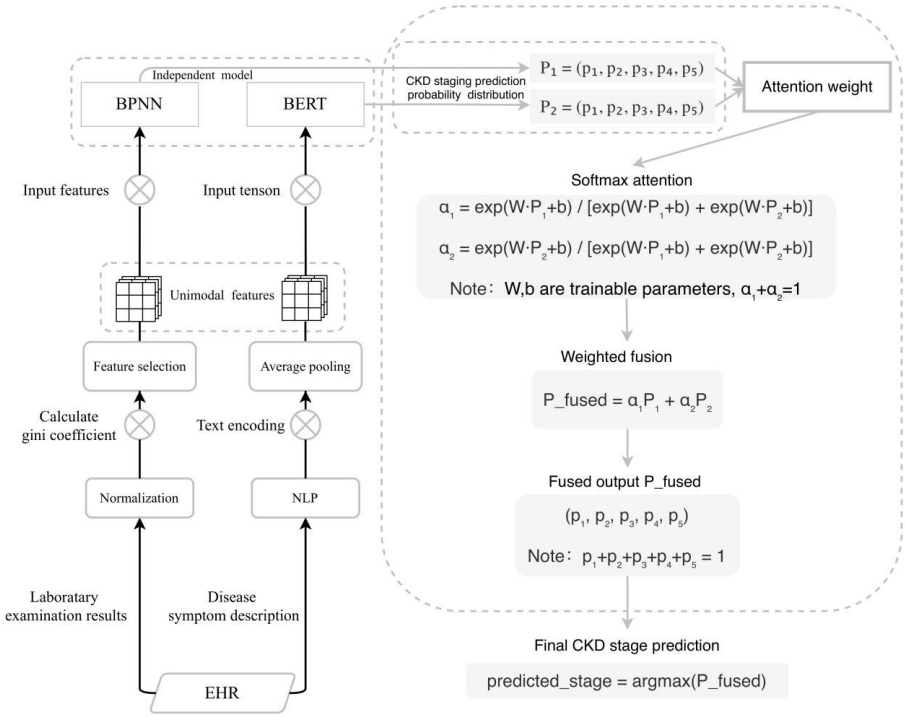


Figure 8. Fusion process of the multimodal chronic kidney disease (CKD) staging prediction model
Abbreviations: BERT: Bidirectional Encoder Representations from Transformers; BPNN: Backpropagation neural network; EHR: Electronic health records; NLP: Natural language processing.

(b) Calculation of attention weights[46]

The additive attention mechanism was adopted to calculate the weights of the features of the two modalities:

$$e_b = v^T \tanh\left(W_a F_b + b_a\right)$$

$$e_t = v^T \tanh\left(W_a F_t + b_a\right)$$

$$\alpha_b = \frac{\exp\left(e_b\right)}{\exp\left(e_b\right) + \exp\left(e_t\right)} \tag{14}$$

$$\alpha_t = \frac{\exp\left(e_t\right)}{\exp\left(e_b\right) + \exp\left(e_t\right)} \tag{15}$$

Among them, Wα ∈ Rdα × df, bα ∈ Rdα, v ∈ Rdα are learnable parameters, and α_b and α_t represent the attention weights of Western medicine and TCM features, respectively.

(c) Weighted feature fusion

We weighted and fused the features of the two modalities based on the attention weights:

$$F_{fused} = \alpha_b \cdot F_b + \alpha_t \cdot F_t \tag{16}$$

(d) Classification prediction

The fused features were input into the fully connected layer for CKD staging prediction:

$$P = Softmax\left(W_c F_{fused} + b_c\right) \tag{17}$$

Among them, $W_c \in R^{5 \times d_f}$ and $b_c \in R^5$ are the parameters of the classification layer, and $P \in R^5$ is the predicted probability distribution of each CKD stage.

(e) Analysis of algorithmic advantages

Compared with the original KNN fusion method, the fusion strategy based on the attention mechanism has the following advantages: (i) End-to-end learnability: The attention weights are automatically optimized through BP, eliminating the need for manual setting of weight parameters; (ii) Dynamic adaptability: It dynamically adjusts the importance of modalities based on the features of input samples, adapting to individual differences; (iii) Strong interpretability: The attention weights reflect the contribution of different modalities to the final decision, providing a basis for clinical doctors' decision-making; (iv) High computational efficiency: It avoids the high computational complexity of KNN in the prediction stage.

3. Empirical analysis

3.1. Data source

The data for this study came from the EHRs of kidney disease patients in a TCM hospital in Guangdong province between 2019 and 2022. According to the hospital's clinical practice guidelines, the internal medicine department primarily treats patients aged 15 and above; therefore, pediatric patients were excluded from this study's sample. As one of the largest and most visited TCM hospitals in the country, with an annual outpatient volume consistently exceeding 7 million for many years, it has long ranked first among TCM hospitals and among the top of all hospitals nationwide. It is also the regional TCM medical center in southern China. Its patients come from a wide range of sources, with 35–40% from outside Guangdong province, covering all provinces in the country. The large outpatient volume encompasses a full spectrum of diseases, from common and chronic to difficult and rare ones, providing a high-quality clinical data foundation with broad representativeness for this study. The original research sample included nearly 40,000 medical diagnosis text data recorded by TCM physicians in medical records based on observation, auscultation, olfaction, inquiry, and palpation (such as pulse diagnosis and tongue observation), and over 90,000 numerical laboratory results, involving nearly 70 clinical examination indicators. These were the original research sample, and an anonymized strategy based on the patient ID was used to protect patients' personal privacy and data security. Divide the dataset into a training set and a test set in a 7:3 ratio, and then take 1/5 of the training set as the validation set.

3.2. Data preparation

For the laboratory examination results, indicators that yielded only qualitative results were removed, and the remaining indicators were used as the original features. For the medical diagnosis data, the following preprocessing is performed:

(i) Text segmentation and filtering of stop words.
(ii) Removal of meaningless medical information, such as treatment methods and diagnosis time.
(iii) Separate the segmented text by spaces and organize it to form the original corpus. Then, a patient information database was constructed (Table 3).
(iv) Unify the text expressions with the same semantics. Perform K-means clustering at the character level so that texts within each cluster share the same semantics. For example, "hypertension" and "hypertension disease" were uniformly marked as "hypertension disease," and a new column, "hypertension disease," was added to the patient information database to record the hypertension level. Finally, the standardized

disease names were stored in the new column "complications," and the symptom descriptions were separated from diseases in the "diagnosis" column and stored separately in the "symptom" column to construct a new corpus containing all the texts corresponding to the "complications" and "symptom" columns in each row of the information database.

(v) Extract the diseases and symptoms highly associated with CKD. Use term frequency–inverse document frequency (TF–IDF) to evaluate the importance of a word to a corpus. Calculate the TF–IDF value of each word in the long text of each "diagnosis" column, take the top 50% of the words as keywords, reclassify them into complications and symptoms, and update the information database.

Table 3. Patient information database (template)

Patient ID	Gender	Age	Diagnosis
1	Male	51	Frequent urination, nephritis, impaired glucose tolerance
2	Female	49	Insomnia, anemia
…	…	…	…

3.3. Analysis of complications

3.3.1. Horizontal analysis of chronic kidney disease complications

We counted the etiological compositions of all complications in the information database. The top 5% of the etiologies are presented in Table 4. Among CKD patients, hypertension, gout, and diabetes mellitus were the most common complications, and the risk of a patient developing one of these complications reached 83%.

Then, we reduced the dimension of the complication word embeddings and projected them into a 2D image, as shown in Figure 9.

As shown in Figure 9, the data points are densely distributed within the region (±0.1, ±0.1) for both principal component (PC)1 and PC2, indicating an association between the complications, supporting further analysis of their correlation.

We counted co-occurring complications for a given CKD complication and calculated the weight of each complication relative to another complication. We retained only the binary complications with a correlation greater than 50%. The complication network diagram is shown in Figure 10.

As shown in Figure 10, hypertension, diabetes mellitus, and anemia were strongly associated with each other. Then, we extracted CKD stage information from the original patient samples and recorded it in the information database as pure numbers. Figure 11 shows the distribution diagram of patient numbers.

As shown in Figure 11, the number of patients in G5 was the largest, accounting for approximately 55% of the total patients, followed by G4 (27%) and G3 (7%). In contrast, the number of patients in G1 and G2 was relatively low, totaling only 7%. A plausible reason is that there are no obvious symptoms in the early stages; hence, the number of patients seeking medical treatment is low, or physicians do not mark the early stages.

3.3.2. Vertical analysis of chronic kidney disease complications

Figure 12 shows the distribution of patients across CKD stages, re-evaluated using the augmented samples that

Table 4. Etiology composition table

Etiological composition	Average age (years)	Median age (years)	Gender ratio (male/female)	Proportion (%)
Hypertension	57	57	1.93	38.58
Gout	53	53	6.95	24.73
Diabetes mellitus	60	61	2.02	19.26
Anemia	57	57	1.1	17.43
Arthritis	59	60	0.95	11.69
Hyperuricemia	50	50	4.07	8.38

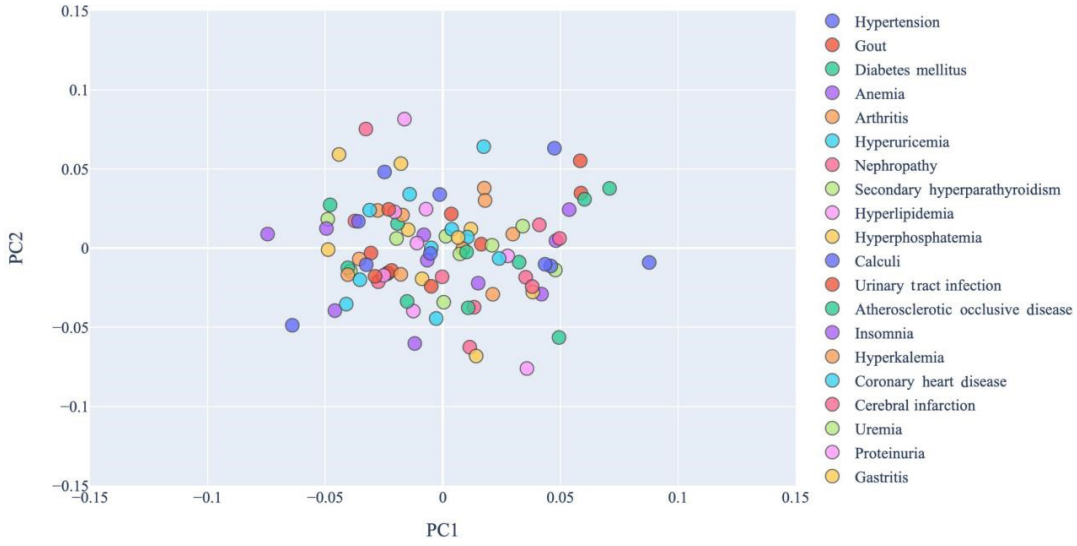


Figure 9. Visualization of word embedding dimensionality reduction
Abbreviation: PC: Principal component.

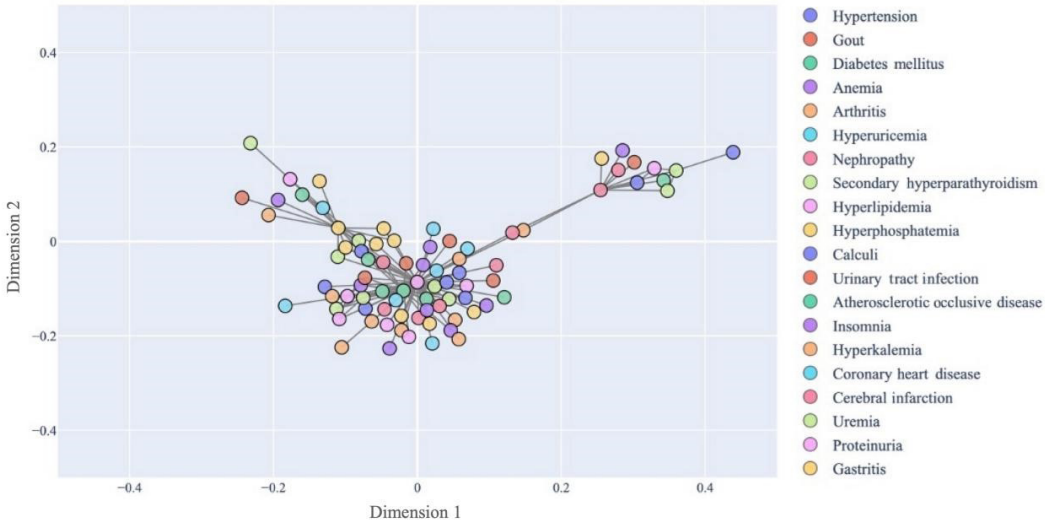


Figure 10. Network diagram of complications. The shorter the distance between nodes, the stronger the correlation

addressed class imbalance.

As shown in Figure 13, there were more male patients than female patients, and there were significantly more male patients in stage G3 than female patients. Meanwhile, the age distribution of patients was comparable, ranging from 20 to 90 years old.

As CKD progresses, the types of complications gradually increase. Figure 14 shows the top 10% complications in each CKD stage.

As shown in Figure 14, from G1 to G5, complications such as hypertension, diabetes mellitus, and hyperuricemia consistently accounted for a high proportion, indicating a universal association between these complications and CKD.

At the same time, with the progress of CKD, the correlation of some complications may change: (i) Changes in the frequency of complications. The frequency of some complications is relatively low in the early stage of CKD. However, their correlations in the late stage increase significantly. For example, anemia was not obvious in G1 and G2, but as renal function declined, its incidence

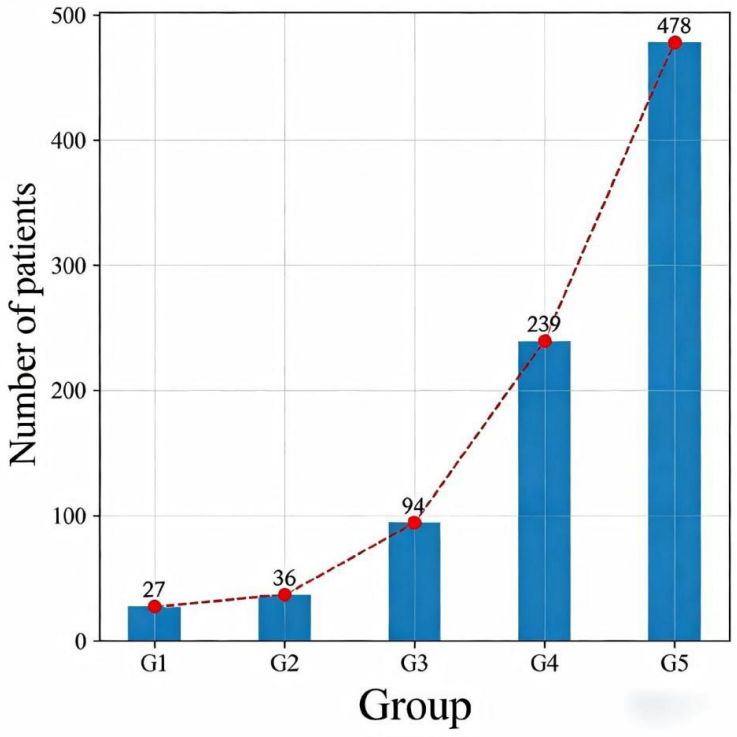


Figure 11. Patient distribution across different chronic kidney disease (CKD) stages. G1 to G5 denotes stages 1 to 5 of CKD based on glomerular filtration rate.

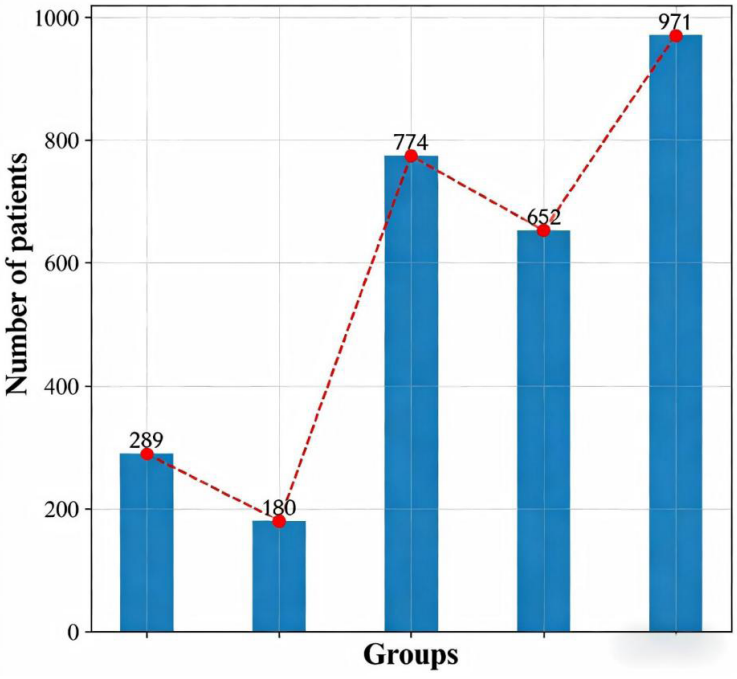


Figure 12. Distribution of chronic kidney disease patients after data augmentation

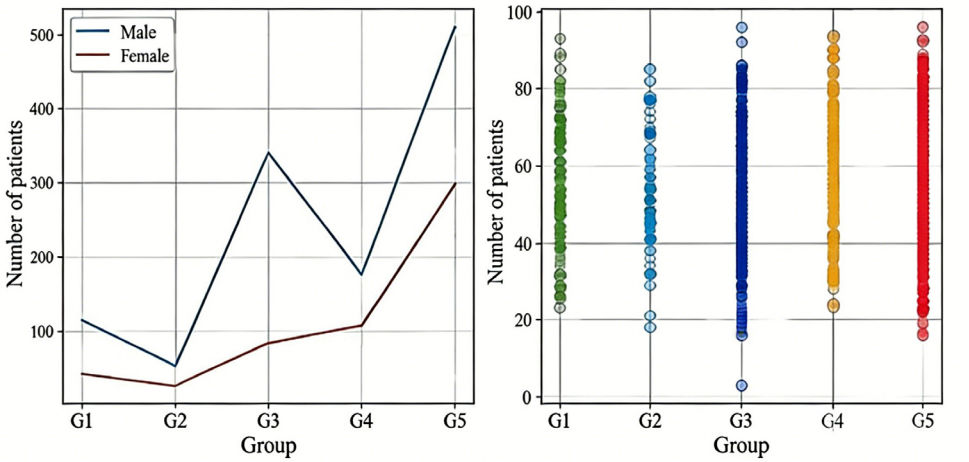


Figure 13. Results after data augmentation. (A) Patient number distribution. (B) Age distribution.

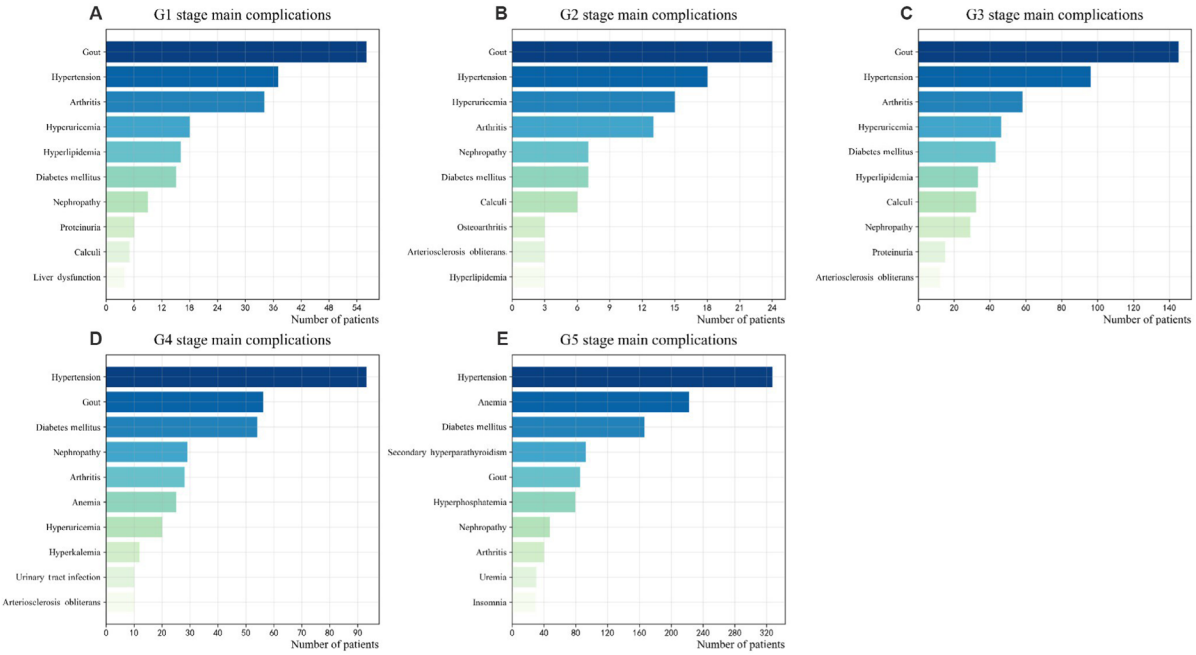


Figure 14. Major complications across different stages of CKD: (A) G1, (B) G2, (C) G3, (D) G4, and (E) G5.

increased significantly. (ii) Changes in the severity of complications. At different stages of CKD, the severity of the same complication may differ. Taking hypertension as an example, in the CKD early stage, hypertension was only slightly elevated. However, in the late stage, hypertension became more severe and difficult to control, requiring stronger drug treatment.

As shown in Figure 15, the incidence of different complications shows distinct trends across the five stages. In G1 and G2, since patients were at an early stage of CKD, the incidence of different etiologies was generally low. In these stages, renal function has not been significantly affected, and the body remains relatively stable. In G3 and G4, the situation begins to change. The incidence of various complications showed a slight to moderate increase, indicating that as renal function continues to deteriorate, the body gradually develops symptoms. In G5, CKD reached the end stage, and the incidence of all complications increased significantly, indicating that CKD and its complications have become more severe and complex. Patients in this stage need urgent intervention and may need to receive renal dialysis or kidney

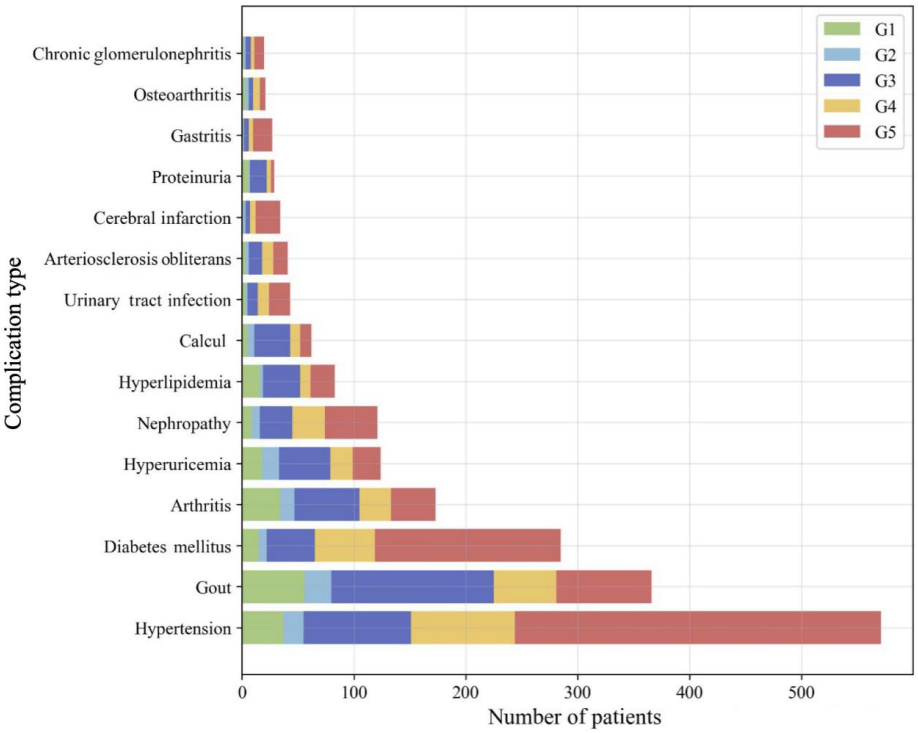


Figure 15. Changes in complications across different stages

transplantation.

In general, some complications, such as hypertension and diabetes mellitus, maintain a relatively high incidence in all CKD stages. In CKD treatment, special attention should be paid to controlling these common complications. At the same time, with CKD progression, the incidence of some complications increases, such as benign prostatic hyperplasia and kidney disease. This may be related to the gradual decline in renal function, which increases the incidence of these complications.

3.4. Model feature considerations

To consider the rationality of the selection of blood laboratory indicator features, it is necessary to observe which indicators physicians usually use as auxiliary examinations. According to clinical guidelines and expert consensus, creatinine, BUN, and uric acid are the most commonly used diagnostic indicators. Hence, we performed a binary statistical analysis on these indicators.

3.4.1. Relationships between creatinine, blood urea nitrogen, and uric acid

As shown in Figure 16A, the creatinine values of most patients fell within the range of 0 to 250, while the uric acid values fluctuated between 400 and 800. In Figure 16B, most of the data were distributed within the ranges of 0 to 500 for creatinine and 0 to 20 for BUN. As creatinine increased, BUN increased linearly, indicating a linear relationship. In Figure 16C, uric acid remained within the range of 400 to 800, whereas the fluctuation of BUN was large, and its distribution was not consistent with that of uric acid. Based on the above observations, we focused on identifying the relationship between creatinine and BUN.

3.4.2. Changes in creatinine and blood urea nitrogen across chronic kidney disease stages

As shown in Figure 17, as CKD progressed, creatinine and BUN levels increased. In the early stage, they were mostly in the lower range; in the late stage, they increased significantly. Meanwhile, the slope of the creatinine–BUN relationship appeared to decrease. To further analyze this phenomenon, a regression equation was constructed.

As shown in Figure 18, the p-values of all regression lines were less than the significance level of 0.01. Therefore, regardless of CKD stage, even when the slope decreased, the regression equation remained linear, indicating that different levels of creatinine and BUN can better distinguish CKD and play an important role in evaluating renal function and the severity of CKD.

Given individual differences, creatinine and BUN levels in patients at the same stage may vary greatly, attributable to variables such as age and gender. Therefore, further

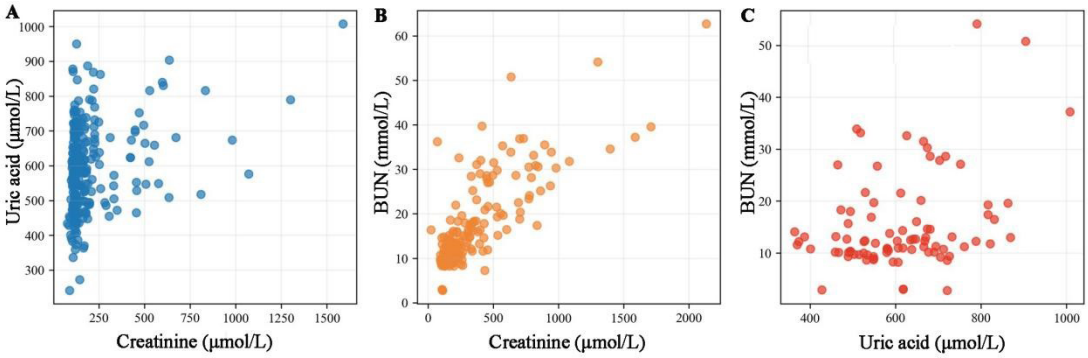


Figure 16. Scatter plots of (A) uric acid vs. creatinine, (B) blood urea nitrogen (BUN) vs. creatinine, and (C) BUN vs. uric acid.

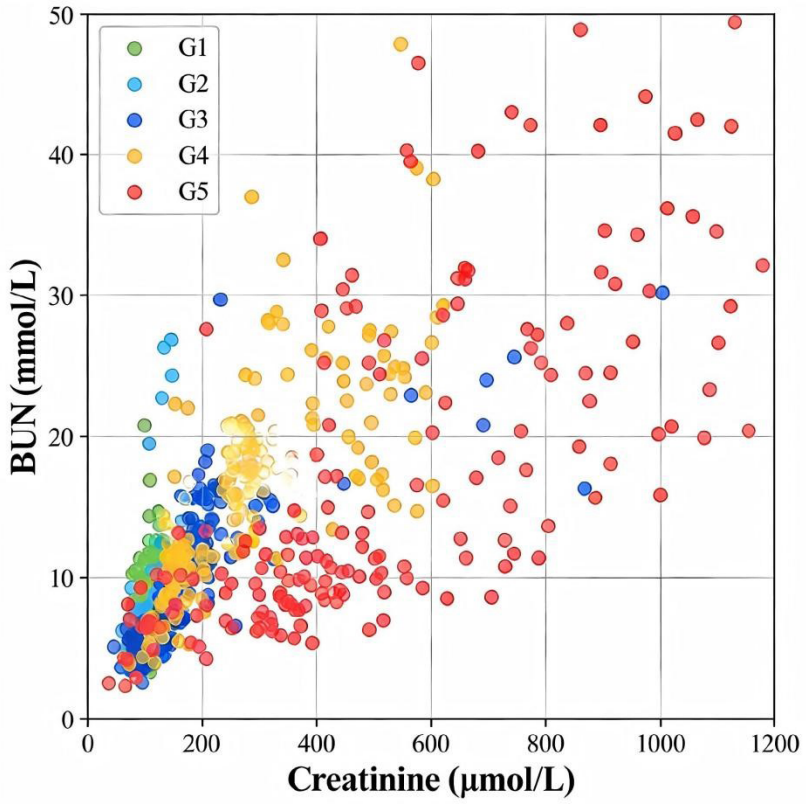


Figure 17. Scatter plot of creatinine vs. blood urea nitrogen (BUN) across chronic kidney disease stages

exploration was needed.

As shown in Figure 19, with increasing age, creatinine and BUN did not change significantly. Therefore, age is not considered the main factor affecting creatinine and BUN levels.

As shown in Figure 20, the fluctuation of creatinine–BUN levels in males was slightly larger than that in females, but the slopes were the same. Therefore, gender is not considered to have a significant impact on creatinine and BUN.

In conclusion, creatinine and BUN are important indicators for distinguishing CKD stages, showing that our indicator selection is feasible.

3.5. Experimental environment and parameter settings

The operating system used in this experiment was Windows 11 (64-bit), the CPU was AMD Ryzen 7 5800H with Radeon Graphics, the graphics card was NVIDIA GeForce GTX 3050, and the programming versions were Python 3.11.5, PyTorch 2.3.0, CUDA 11.3, and Transformers 4.39.1.

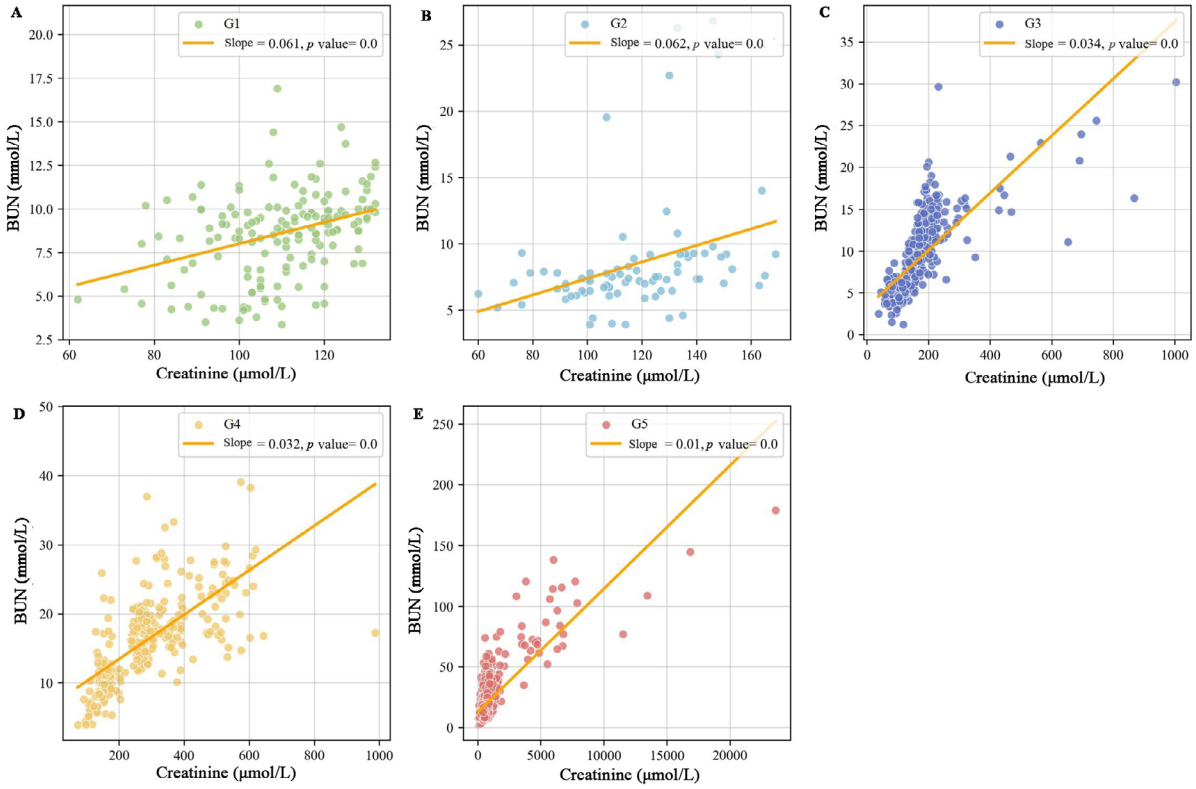


Figure 18. Scatter plots of creatinine vs. blood urea nitrogen (BUN) levels across chronic kidney disease stages: (A) G1, (B) G2, (C) G3, (D) G4, and (E) G5.

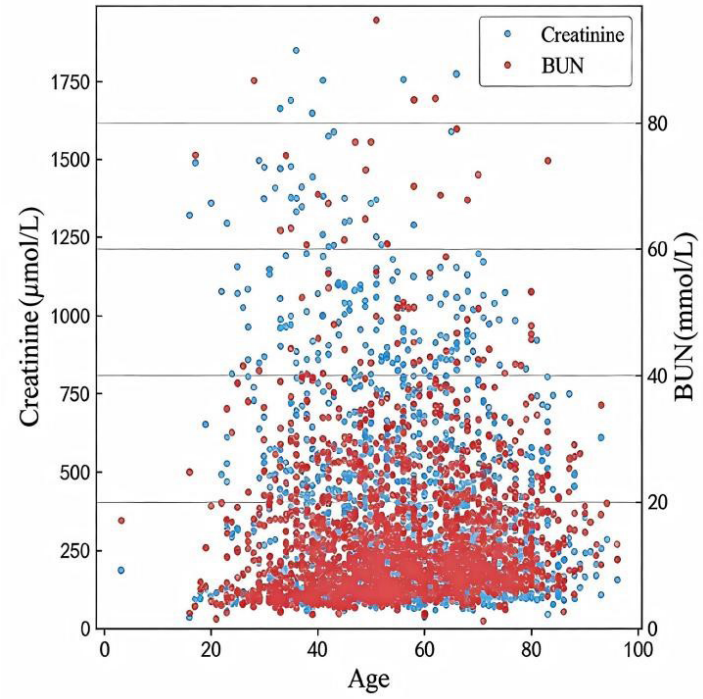


Figure 19. Scatter distribution of blood creatinine and blood urea nitrogen (BUN) across age

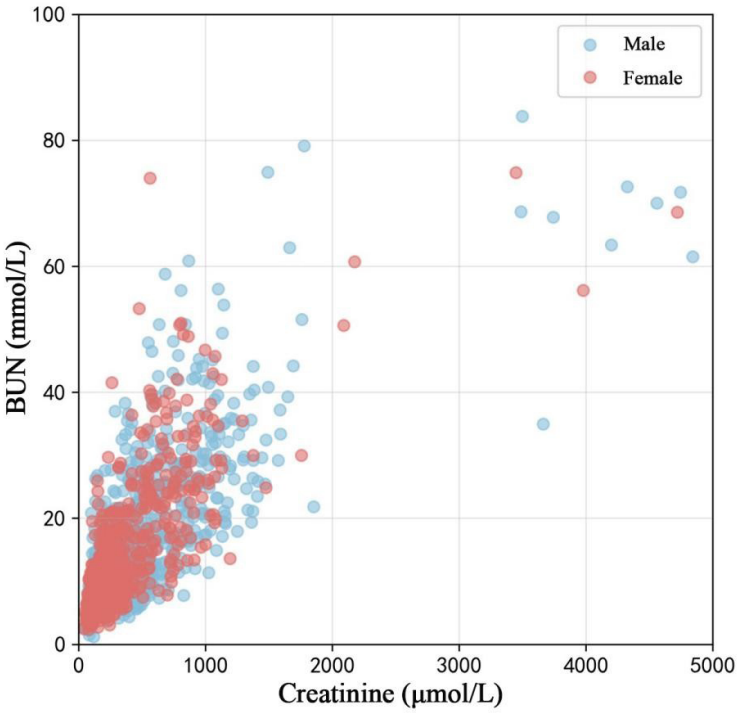


Figure 20. Relationship between creatinine and blood urea nitrogen (BUN) across gender

This experiment predicted CKD staging results and generated the corresponding probability distribution. For the BERT model, the iteration was set to 100, the batch size to 8, the learning rate to 1×10^{-5}, and the optimizer to Adam. For the BP neural network model, the loss function was sparse categorical cross-entropy, the kernel regularizer was the L2 norm, the dropout rate was 0.2, the number of epochs was 100, the batch size was 32, the learning rate was 1×10^{-3}, and the optimizer was Adam.

3.6. Analysis of empirical results

We dynamically updated the weights for both the improved BP neural network model and the Google-pretrained BERT. Furthermore, we used the KNN model to integrate the top K probable values from the probability distributions of different weight combinations, given a K value. The corresponding stages were determined by these probability values, and the final stage prediction result was the most frequently occurring stage. This study leveraged the complementarity between independent models, removed their respective shortcomings through combination, and improved the performance and generalization ability of the model.

To compare the accuracy of the fused model, we designed a 10-fold cross-validation comparison experiment. The comparison models included classic machine learning models, existing unimodal models, and the proposed deep learning-based multimodal model in this study. Among them, the existing unimodal models comprised a widely used traditional detection method and a recently proposed hybrid fuzzy deep neural network model.[47]

As shown in Table 5, the accuracy of classic machine learning models was generally low, indicating that their structures were simple or unsuitable for clinical data with nonlinear relationships and interactions. They did not have sufficient ability to capture data patterns, and their accuracy was low. Compared with the 90% accuracy of existing unimodal models, the accuracy of the multimodal prediction model after fusing the two independent models was close to 98%, surpassing that of any single-modality model. This improvement is mainly attributed to the complementarity of multimodal data at the information level: numerical clinical indicators can objectively reflect the physiological state of patients, while symptom and complication information supplement the characteristics of disease progression at the subjective and pathological levels. Through feature-level or decision-level fusion, redundant information between different modalities was suppressed, while discriminative features were enhanced.

Then, we defined the categories Class 1 to Class 5 corresponding to the G1 to G5 of CKD stages, and compared the F1 values across models in each class (Table 6).

Table 5. Comparison of accuracy across different prediction models

CKD stage prediction model	Accuracy
Decision tree model	79.5%
Support vector machine model	61.3%
Feedforward neural network model	81.2%
Hybrid on fuzzy deep neural network model	93.5%
Traditional radioimmunoassay method	91.2%
Improved BP neural network model	93.2%
BERT model	95.1%
Deep learning-based multimodal model	97.9%

Abbreviations: BERT: Bidirectional Encoder Representations from Transformers; BP: Backpropagation; CKD: Chronic kidney disease.

As shown in Table 6, the classic machine learning models performed poorly at CKD staging prediction. The highest F1 score occurred when the feedforward neural network predicted stage G2. The F1 values of these models in other categories were relatively low. For the support vector machine, the F1 scores were 0 in all stages except G3 and G5. This could be due to the model's inability to identify the optimal hyperplane in the class-imbalanced clinical data, leading it to favor the more prevalent classes. The existing unimodal models exhibited a relatively low F1 score in stage G2 and a relatively high F1 score in stage G1, indicating they cannot fully distinguish between G1 and G2 and are more likely to classify early-stage patients as G1. This is mainly because the G1 and G2 stages show high similarity across multiple clinical indicators and symptom manifestations, and their differences often lie in slight numerical fluctuations or latent pathological changes. Traditional and single-modal models had difficulty capturing these subtle and nonlinear characteristic differences and were therefore more likely to misclassify some G2-stage patients as G1-stage. The proposed model overcame the shortcomings of existing unimodal models, thereby achieving accurate classification of early-stage patients. At the same time, the model maintained relatively high F1 scores in G3 to G5.

The one-vs.-rest strategy was used to adjust the multi-class prediction models in Table 5. The receiver operating characteristic (ROC) and precision–recall (PR) curves for each model across the different categories are shown in Figures 21 and 22, respectively.

This study adopted the internationally recognized Kidney Disease: Improving Global Outcomes (KDIGO) clinical guidelines as the gold standard for CKD staging. The predicted staging results were compared with this gold standard. The diagnostic accuracy of the model was evaluated using the ROC curve and the area under the curve (AUC). The ROC curve reflects the consistency between the model's predicted results and the gold standard at different discrimination thresholds. A larger AUC value indicates greater consistency between the model's predicted staging results and the gold-standard staging and stronger diagnostic accuracy.

As shown in Figure 21, there were significant differences in the diagnostic consistency between classical machine learning models and the KDIGO gold standard across different CKD staging categories. For example, although the AUC values of the feedforward neural network in Class 1 and Class 2 both reached 1, the AUC in Class 3 was only 0.85. This indicates that the model cannot distinguish different categories well, resulting in poor classification performance in some stages. By comparing the existing unimodal models with the proposed model, it was found that their ROC curves across different categories were similar and remained within 0.97–0.99, indicating strong generalization and high prediction stability. However, the AUC values of the existing unimodal models were not high enough, indicating that the prediction accuracy is average. Specifically, they showed high AUC values in

Table 6. Comparison of F1 scores across different prediction models

CKD stage prediction model	Class 1	Class 2	Class 3	Class 4	Class 5
Decision tree	0.90	0.67	0.82	0.73	0.90
Support vector machine	0	0	0.63	0	0.93
Feedforward neural network	1	0.96	0.48	0.66	0.34
Hybrid on fuzzy deep neural network model	0.93	0.54	0.90	0.78	0.94
Traditional radioimmunoassay method	0.93	0.50	0.90	0.79	0.95
Deep learning-based multimodal model	1	1	0.97	0.93	0.93

Abbreviation: CKD: Chronic kidney disease.

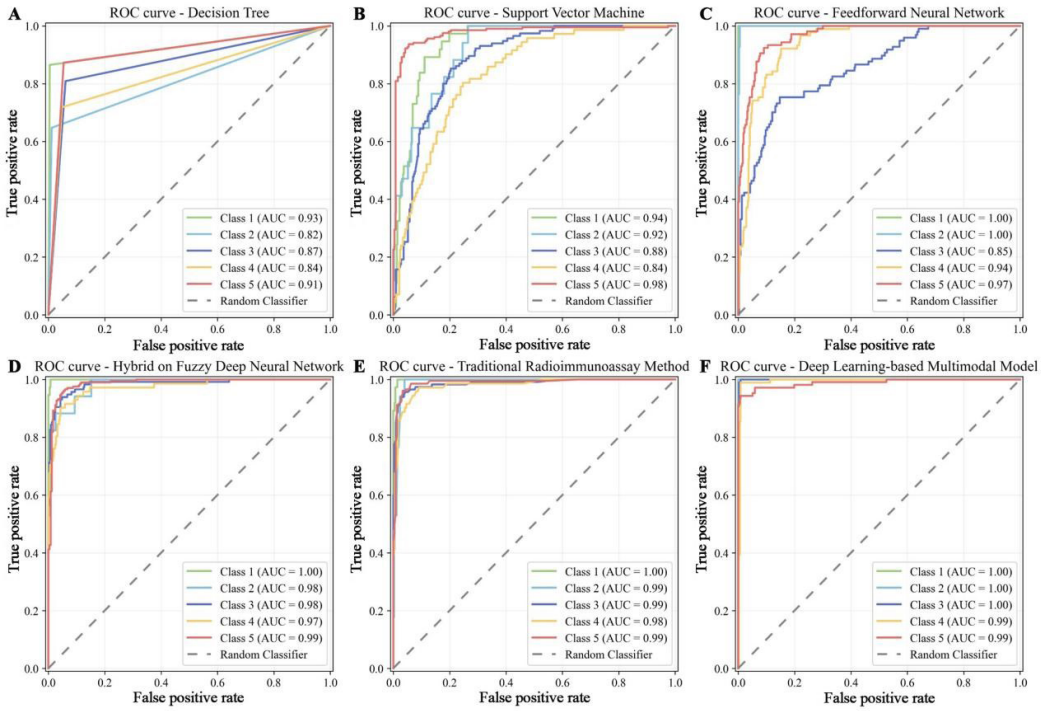


Figure 21. Receiver operating characteristic (ROC) curves of different prediction models: (A) decision tree, (B) support vector machine, (C) feedforward neural network, (D) hybrid on fuzzy deep neural network, (E) traditional radioimmunoassay method, and (F) deep learning-based multimodal model. Abbreviation: AUC: Area under the curve.

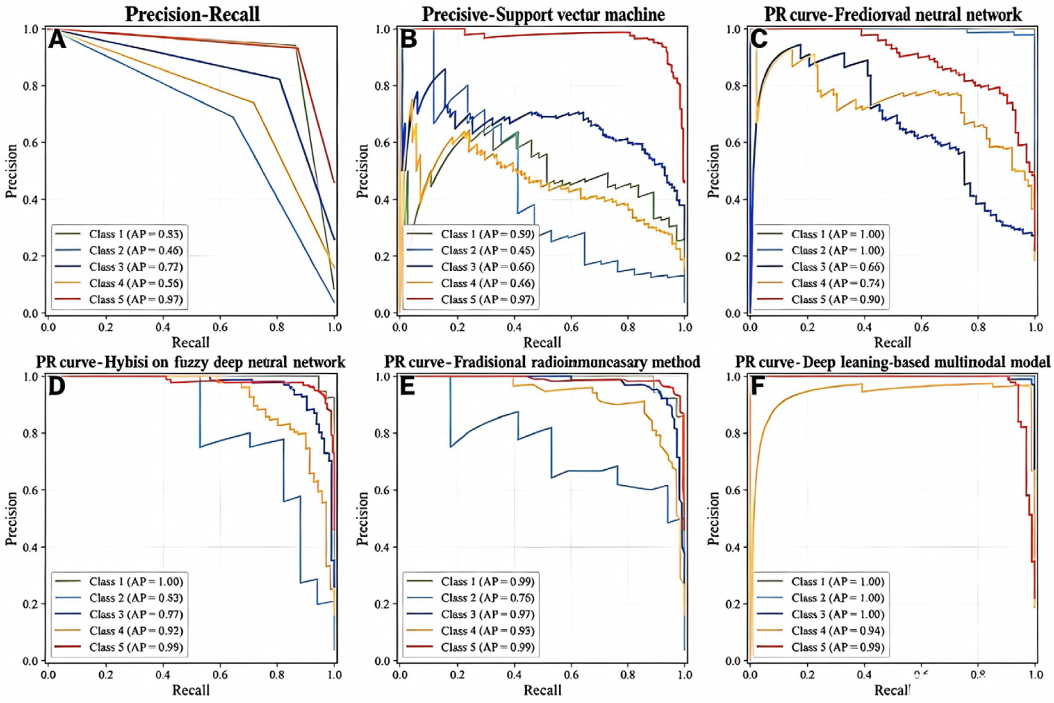


Figure 22. Precision–recall (PR) curves of different prediction models, where AP represents the average precision rate, which is the area under the PR curve: (A) decision tree, (B) support vector machine, (C) feedforward neural network, (D) hybrid on fuzzy deep neural network, (E) traditional radioimmunoassay method, and (F) deep learning-based multimodal model.

Class 1; that is, the prediction accuracy is high only in stage G1 and cannot be maintained in other stages. In other words, these models can only identify whether a patient has CKD but cannot accurately predict the specific stage of the patient. This is because the existing unimodal models are all based on a single modality and cannot capture higher-dimensional patient factors, resulting in low prediction accuracy. On the contrary, the multimodal model proposed in this paper maintained high AUC values and high consistency with the gold standard across all CKD stages. Its values in Classes 1–3 were all 1, and the AUC values in Classes 4 and 5 were also close to 1. This indicates that the model has a high diagnostic accuracy in all stages. This result fully demonstrates the effectiveness of the multimodal strategy in improving the diagnostic accuracy of CKD staging, enabling the model's predictions to achieve higher AUC across categories and achieve the optimal diagnostic performance.

As shown in Figure 22, as recall increased, the precision of classic machine learning models dropped rapidly and fluctuated, indicating that when the prediction threshold is reduced, the models are easily affected by data noise, resulting in high error rates. Although existing unimodal models solved this problem to a certain extent, their macro-average precision was only about 0.93, which was low. They could not distinguish between stages G1 and G2 well, resulting in very low AP values in Class 2. Notably, the key to treating CKD liesin early detection. Therefore, reducing the misdiagnosis rate at an early stage is of great significance for the diagnosis and treatment of CKD. The proposed model combined deep learning with a multimodal strategy, and the macro-average precision exceeded 0.98, achieving strong discrimination across categories.

In addition, we compared the loss curves of the models in Table 5 after 100 iterations to assess their performance and convergence on the training and validation sets (Figure 20). In terms of training efficiency, different models varied in computational complexity and convergence characteristics. Traditional machine learning models, such as decision trees and support vector machines, achieved relatively low training costs but showed limited performance in the CKD staging prediction task. Unimodal deep learning models, such as feedforward neural networks and improved BP networks, typically require 50–60 training rounds to converge. In contrast, the proposed deep learning-based multimodal model, despite its more complex structure, showed faster convergence trends during training. Its validation performance tended to stabilize after 30 to 40 rounds and converged within 50 rounds, with a smoother training process. These results indicate that multimodal feature fusion not only enhances prediction performance but also improves the optimization efficiency of the model to a certain extent.

As shown in Figure 23, although the decision tree converged on the training set, its performance on the validation set was poor, indicating that it is too simple to learn useful information from the training set. The loss curves for the support vector machine on the training and validation sets were highly similar, indicating strong generalization performance. However, the curve only converged to 1 at the end, and the goodness-of-fit was not ideal. Although the loss value of the feedforward neural network converged to 0, it fluctuated violently with increasing iterations, indicating that the model had learned noise and random fluctuations, resulting in overfitting. Although the existing unimodal models had reduced overfitting to some extent, the loss curves on the training and validation sets still differed, indicating insufficient generalization performance. The loss curve of the proposed model closely matched the validation and training sets and steadily converged to 0, indicating strong generalization performance, high goodness-of-fit, and suitability for CKD stage prediction.

In summary, the proposed model demonstrated better performance across evaluation indicators, including precision, F1 score, convergence, and goodness-of-fit, than other single-modality models. This is precisely the advantage of applying the multimodal strategy, which can comprehensively consider the multidimensional factors of patients to provide more accurate predictions, indicating that the model has high practical value. At the same time, the model combined the dual-system concepts of Western medicine diagnostics and TCM symptomatology, thereby providing strong support for achieving more accurate and comprehensive medical diagnosis.

4. Discussion

This study demonstrated that the proposed deep learning-based multimodal model significantly outperformed classical machine learning methods and existing unimodal models for CKD staging prediction, especially in overall accuracy and category-level F1 scores. This performance improvement not only validates the effectiveness of the multimodal fusion strategy but also further indicates that the CKD staging prediction task has a distinct characteristic of relying on multi-source heterogeneous information.

From a clinical perspective, the early stages of CKD (G1–G2) are highly similar across multiple biochemical indicators and symptom manifestations, with differences often reflected only in minor numerical fluctuations

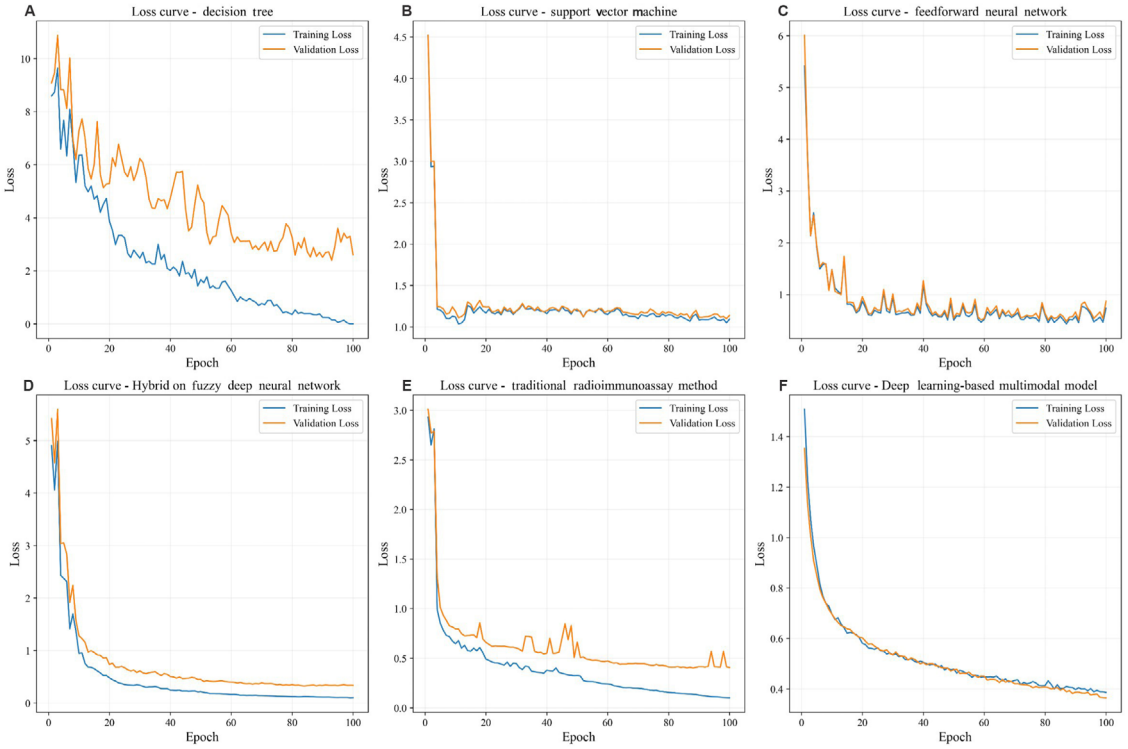


Figure 23. Loss curves of different prediction models: (A) decision tree, (B) support vector machine, (C) feedforward neural network, (D) hybrid on fuzzy deep neural network, (E) traditional radioimmunoassay method, and (F) deep learning-based multimodal model.

or subtle pathological changes. Therefore, traditional or unimodal models struggle to capture fine-grained, nonlinear feature differences, leading to insufficient accuracy in predicting early-stage patients. The proposed model, by integrating structured clinical indicators with semantic information, such as symptoms and complications, achieves complementary enhancement of features from different modalities, thereby effectively improving discrimination ability at early stages.

Furthermore, ROC and PR curve analyses indicate that the proposed model maintained high diagnostic consistency and stability across all CKD staging categories, suggesting that its predictive ability is not limited to identifying whether a patient is ill but can also accurately determine the specific stage of the disease. This is of great significance for the early screening, risk assessment, and individualized intervention of CKD.

It is worth noting that the proposed multimodal model shows faster convergence and a more stable optimization process during training, indicating that multimodal fusion not only enhances prediction performance but also improves feature representation quality and model learning efficiency.

5. Conclusion

This paper proposed a deep learning-based multimodal CKD stage prediction model to achieve early diagnosis, accurate staging, and personalized treatment of CKD patients. Based on the EHRs of kidney disease patients from a hospital in Guangdong province, this paper conducted analyses and modeling by applying CKD complications, symptoms, clinical examination indicators, and CKD staging. It not only selected the laboratory examination results from Western medicine and the disease symptom descriptions from TCM but also constructed two independent models—an improved BP neural network and the Google-pretrained BERT—to process unimodal features. After decision fusion, the proposed deep learning-based multimodal CKD stage prediction model was constructed. It is superior to most methods for predicting CKD staging from unimodal data in terms of accuracy and F1 score. The multimodal modeling framework did not rely on specific age groups or single disease scenarios. Its core idea is to enhance the characterization of disease progression features by integrating multi-source clinical information. Therefore, when pediatric CKD data or other chronic disease data are introduced, the model can

be extended and applied through retraining or transfer learning, demonstrating excellent universality and potential for promotion.

From a clinical perspective, CKD is characterized by slow progression and significant individual differences, making its staging difficult to determine based on a single laboratory indicator. Traditional assessment mainly relies on objective test results, such as GFR, serum creatinine, and urine protein. However, in the early stages of the disease, these indicators show limited changes, resulting in insufficient sensitivity for identifying early risks. In contrast, information at the symptom level can reflect dynamic changes in the overall functional status of patients and often provides supplementary clues about disease progression. This paper combined data from both Chinese and Western medicine to achieve a multidimensional description of CKD clinical states. The results showed that this multimodal model has better discrimination ability in early staging, especially in distinguishing adjacent stages such as G1 and G2, and maintains stable performance in the middle and late stages, providing a new technical approach for early risk identification and clinical decision support in CKD. It helps to make timely interventions and formulate more personalized treatment plans to slow or prevent the progression of the disease, thereby improving treatment outcomes and enhancing the efficiency and quality of medical services.

Acknowledgments

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

Funding

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

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Yan Zhang, Xintong Zou
Formal analysis: Zhuoyun Xie, Kaiting Wang
Investigation: Ruihui Chen, Zihui Lin
Methodology: Ziyue Xiong, Jialin Zhao, Huimin Li and Xiaoli Chu
Project administration: Haole Huang
Resources: Huimin Li, Xiaoli Chu
Writing–original draft: Xiaoqing Liang, Yuhan Zhang
Writing–review & editing: Xueyan Chen, Runxi Zhou, Huimin Li, Xiaoli Chu
All medical interpretations and clinical conclusions were thoroughly reviewed and confirmed by clinical physicians (Huimin Li and Xiaoli Chu). The manuscript was finalized only after clinical validation and formal approval by these medical collaborators.

Ethics approval and consent to participate

Written informed consent was obtained from all participants included in the study prior to participation.

Consent for publication

All patients in this study gave their oral consent for the anonymous use of their data for scientific publication during their clinical visits.

Availability of data

The dataset used and analyzed in this study is non-public data, and its ownership belongs to Guangdong Provincial Hospital of Traditional Chinese Medicine. Due to patient privacy concerns and the hospital's data-sharing policy, the original data are not publicly available. If there is a legitimate research need, please contact the corresponding author of this paper or the Data Management Committee of Guangdong Provincial Hospital of Traditional Chinese Medicine. The relevant data may be used under the condition of passing the ethical review of the institution

and signing a data use agreement.

Further disclosure

Part of findings was presented and published in: 2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS) at Nanjing, China from May 23-25, 2025, A Multimodal Stage Prediction Study of Chronic Kidney Disease based on Deep Learning (doi: 10.1109/EEISS65394.2025.11085892); 2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS) at Nanjing, China from May 23-25, 2025, Research on Chronic Kidney Disease Staging Prediction and Prescription Recommendation Model Based on Multimodal Data Fusion (doi: 10.1109/EEISS65394.2025.11085645).

References

1. Zhang L, Wang F, Wang L, *et al*. Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet*. 2012;379(9818):815-822.

   doi: 10.1016/S0140-6736(12)60033-6

2. Shlipak MG, Tummalapalli SL, Boulware LE, *et al*. The case for early identification and intervention of chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int*. 2021;99(1):34-47.

   doi: 10.1016/j.kint.2020.10.012

3. Hossain MS, Hassa SMN, Al-Amin M, Rahaman MN, Hossain R, Hossain MI. Kidney Disease Detection from CT Images using a customized CNN model and Deep Learning. In: *2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS)*. IEEE; 2023:1-6.

   doi: 10.1109/aicaps57044.2023.10074314

4. Dubin RF, Deo R, Ren Y, *et al*. Proteomics of CKD progression in the chronic renal insufficiency cohort. *Nat Commun*. 2023;14(1):6340.

   doi: 10.1038/s41467-023-41642-7

5. Tin A, Köttgen A. Genome-wide association studies of CKD and related traits. *Clin J Am Soc Nephrol*. 2020;15(11):1643-1656.

   doi: 10.2215/cjn.00020120

6. Robinson-Cohen C, Triozzi JL, Rowan B, *et al*. Genome-wide association study of CKD progression. *J Am Soc Nephrol*. 2023;34(9):1547-1559.

   doi: 10.1681/asn.0000000000000170

7. Hannan M, Chen J, Hsu J, *et al*. Frailty and cardiovascular outcomes in adults with CKD: Findings from the Chronic Renal Insufficiency Cohort (CRIC) study. *Am J Kidney Dis*. 2024;83(2):208-215.

   doi: 10.1053/j.ajkd.2023.06.009

8. Ding N, Lv Y, Su H, *et al*. Vascular calcification in CKD: New insights into its mechanisms. *J Cell Physiol*. 2023;238(6):1160-1182.

   doi: 10.1002/jcp.31021

9. Venkatesan VK, Ramakrishna MT, Izonin I, Tkachenko R, Havryliuk M. Efficient Data Preprocessing with Ensemble Machine Learning Technique for the Early Detection of Chronic Kidney Disease. *Appl Sci*. 2023;13(5):2885.

   doi: 10.3390/app13052885

10. Krukowski H, Valkenburg S, Madella AM, *et al*. Gut microbiome studies in CKD: opportunities, pitfalls and therapeutic potential. *Nat Rev Nephrol*. 2023;19(2):87-101.

    doi: 10.1038/s41581-022-00647-z

11. Kumar K, Pradeepa M, Mahdal M, RajaRao MVLN, Ramesh JVN. A Deep Learning Approach for Kidney Disease Recognition and Prediction through Image Processing. *Appl Sci*. 2023;13(6):3621.

    doi: 10.3390/app13063621

12. Joshi S, Kalantar-Zadeh K, Chauveau P, Carrero JJ. Risks and Benefits of Different Dietary Patterns in CKD. *Am J Kidney Dis*. 2023;81(3):352-360.

    doi: 10.1053/j.ajkd.2022.08.013

13. Macdougall IC. Anaemia in CKD—treatment standard. *Nephrol Dial Transplant*. 2024;39(5):770-777.

    doi: 10.1093/ndt/gfad250

14. Świątek Ł, Jeske J, Miedziaszczyk M, Idasiak-Piechocka I. The impact of a vegetarian diet on chronic kidney disease (CKD) progression–a systematic review. *BMC Nephrol*. 2023;24(1):168.

    doi: 10.1186/s12882-023-03233-y

15. Yamada S, Nakano T. Role of Chronic Kidney Disease (CKD)–Mineral and Bone Disorder (MBD) in the Pathogenesis of Cardiovascular Disease in CKD. *J Atheroscler Thromb*. 2023;30(8):835-850.

    doi: 10.5551/jat.RV22006

16. Heo GY, Koh HB, Kim HJ, *et al*. Association of Plant Protein Intake with Risk of incident CKD: a UK Biobank study. *Am J Kidney Dis*. 2023;82(6):687-697.e1.

    doi: 10.1053/j.ajkd.2023.05.007

17. Mapuskar KA, Vasquez-Martinez G, Mayoral-Andrade G, Tomanek-Chalkley A, Zepeda-Orozco D, Allen BG. Mitochondrial oxidative metabolism: an emerging therapeutic target to improve CKD outcomes. *Biomedicines*. 2023;11(6):1573.

    doi: 10.3390/biomedicines11061573

18. Saif D, Sarhan AM, Elshennawy NM. Deep-kidney: an

effective deep learning framework for chronic kidney disease prediction. *Health Inf Sci Syst*. 2023;12(1):3.

doi: 10.1007/s13755-023-00261-8

19. Alsekait DM, Saleh H, Gabralla LA, *et al*. Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models. *Appl Sci*. 2023;13(6):3937.

doi: 10.3390/app13063937

20. Akter S, Ahmed M, Al Imran A, *et al*. CKD.Net: A novel deep learning hybrid model for effective, real-time, automated screening tool towards prediction of multi stages of CKD along with eGFR and creatinine. *Expert Syst Appl*. 2023;223:119851.

doi: 10.1016/j.eswa.2023.119851

21. Ge XY, Lan ZK, Lan QQ, Lin HS, Wang GD, Chen J. Diagnostic accuracy of ultrasound-based multimodal radiomics modeling for fibrosis detection in chronic kidney disease. *Eur Radiol*. 2023;33(4):2386-2398.

doi: 10.1007/s00330-022-09268-3

22. Hou Z, Zhang G, Ma Y, *et al*. Development of a multimodal kidney age prediction based on automatic segmentation CT image in patients with normal renal function. *Clin Kidney J*. 2023;16(11):2091-2099.

doi: 10.1093/ckj/sfad167

23. Chertow GM, Correa-Rotter R, Vart P, *et al*. Effects of Dapagliflozin in Chronic Kidney Disease, With and Without Other Cardiovascular Medications: DAPA-CKD Trial. *J Am Heart Assoc*. 2023;12(9):e028739.

doi: 10.1161/JAHA.122.028739

24. Evenepoel P, Stenvinkel P, Shanahan C. Inflammation and gut dysbiosis as drivers of CKD–MBD. *Nat Rev Nephrol*. 2023;19(10):646-657.

doi: 10.1038/s41581-023-00736-7

25. Ren F, Jin Q, Jin Q, *et al*. Genetic evidence supporting the causal role of gut microbiota in chronic kidney disease and chronic systemic inflammation in CKD: a bilateral two-sample Mendelian randomization study. *Front Immunol*. 2023;14:1287698.

doi: 10.3389/fimmu.2023.1287698

26. Venkatrao K, Kareemulla S. HDLNET: a hybrid deep learning network model with intelligent IoT for detection and classification of chronic kidney disease. *IEEE Access*. 2023;11:99638-99652.

doi: 10.1109/ACCESS.2023.3312183

27. Joo YS, Rim TH, Koh HB, *et al*. Non-invasive chronic kidney disease risk stratification tool derived from retina-based deep learning and clinical factors. *NPJ Digit Med*. 2023;6(1):114.

doi: 10.1038/s41746-023-00860-5

28. Liang P, Yang J, Wang W, *et al*. Deep learning identifies intelligible predictors of poor prognosis in chronic kidney disease. *IEEE J Biomed Health Inform*. 2023;27(7):3677-3685.

doi: 10.1109/JBHI.2023.3266587

29. Yang J, Chen X, Luo C, *et al*. Application of serum SERS technology combined with deep learning algorithm in the rapid diagnosis of immune diseases and chronic kidney disease. *Sci Rep*. 2023;13(1):15719.

doi: 10.1038/s41598-023-42719-5

30. Bhattacharjee A, Rabea S, Bhattacharjee A, *et al*. A multi-class deep learning model for early lung cancer and chronic kidney disease detection using computed tomography images. *Front Oncol*. 2023;13:1193746.

doi: 10.3389/fonc.2023.1193746

31. Zhang M, Ye Z, Yuan E, *et al*. Imaging-based deep learning in kidney diseases: recent progress and future prospects. *Insights Imaging*. 2024;15(1):50.

doi: 10.1186/s13244-024-01636-5

32. Arif MS, Rehman AU, Asif D. Explainable machine learning models for chronic kidney disease prediction. *Algorithms*. 2024;17(10):443.

doi: 10.3390/a17100443

33. Arif MS, Mukheimer A, Asif D. Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model. *Big Data Cogn Comput*. 2023;7(3):144.

doi: 10.3390/bdcc7030144

34. Li J, Zakka KR, Booth J, *et al*. Discovering patient groups in sequential electronic healthcare data using unsupervised representation learning. *BMC Med Inform Decis Mak*. 2025;25(1):45.

doi: 10.1186/s12911-024-02812-9

35. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6(1):26094.

doi: 10.1038/srep26094

36. Friedman SF, Khurshid S, Venn RA, *et al*. Unsupervised deep learning of electrocardiograms enables scalable human disease profiling. *NPJ Digit Med*. 2025;8:23.

doi: 10.1038/s41746-024-01418-9

37. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(2):423-443.

doi: 10.1109/TPAMI.2018.2798607

38. Li W, Ge X, Liu S, Xu L, Zhai X, Yu L. Opportunities and challenges of traditional Chinese medicine doctors in the era of artificial intelligence. *Front Med*. 2024;10:1336175.

doi: 10.3389/fmed.2023.1336175

39. Zhai Y, Gu J, Zong F, Jiang W. Fine grained cross-modal retrieval algorithm for IETM with attention mechanism fused. *Syst Eng Electron*. 2023;45(12):3915-3923.

doi: 10.12305/j.issn.1001-506X.2023.12.21

40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357.

doi: 10.1613/jair.953

41. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer; 2009.

doi: 10.1007/978-0-387-84858-7

42. Khatri M, Zitovsky J, Lee D, Nayyar K, Fazzari M, Grant C. The association between serum chloride levels and chronic kidney disease progression: a cohort study. *BMC Nephrol*. 2020;21(1):165.

doi: 10.1186/s12882-020-01828-3

43. Kingma D, Ba J. Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR). *arXiv*. Preprint posted online January 30, 2017.

doi: 10.48550/arXiv.1412.6980

44. Bezinque A, Parker J, Babitz SK, Noyes SL, Hu S, Lane BR. Proteinuria impacts patient survival differentially based on clinical setting: A retrospective cross-sectional analysis of cohorts from a single health system: Retrospective cohort study. *Ann Med Surg*. 2019;45:120-126.

doi: 10.1016/j.amsu.2019.07.029

45. Feng Y, Shen J, Xu H, Wang R, Liu T, Zhang Y. Cross fusion multimodal sentiment analysis model based on translate mechanism. *Data Anal Knowl Discov*. 2025;9(3):16-27.

doi: 10.11925/infotech.2096-3467.2024.0003

46. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. International Conference on Learning Representations (ICLR). *arXiv*. Preprint posted online May 19, 2014.

doi: 10.48550/arXiv.1409.0473

47. Li F, Zheng T, He N, *et al*. Data-driven hybrid neural fuzzy network and ARX modeling approach to practical industrial process identification. *IEEE/CAA J Autom Sin*. 2022;9(9):1702-1705.

doi: 10.1109/JAS.2022.105821

Appendix

Algorithm A1. Augmentation of medical diagnostic data

Input: Unlabeled sample data and prediction model

Output: Predicted categories of all samples

pred_proba = model.predict_proba(data);

//Step 1. Compute model-predicted probability distributions for all sample categories.

category = ∅;

for (i = 1; xi ∈ pred_proba; i++) do

  flag = 0;

  for (j = 1; yj ∈ xi; j++) do

    if (yj > 0.75) then

      category = category ∪ {argmax(xi)};

      flag = 1;

    end if

//Step 2. Retain the model-predicted category when there exists a probability value greater than 0.75 in the probability distribution.

  end for

  if (flag == 0) then

    category = category ∪ {0};

  end if

//Step 3. Otherwise, withhold the prediction results.

return category.

Algorithm A2. Attention mechanism-based multimodal fusion algorithm

Input: The feature matrix of Western medicine H_b, the feature matrix of traditional Chinese medicine H_t, and the CKD label Y

Output: The fused prediction result P and the attention weights α

1: Initialize model parameters W_b, W_t, W_a, v, W_c

2: Set hyperparameters d_f=256, d_a=128, lr=0.001, epochs=100

3: The data is divided into a training set, a validation set, and a test set.

4: for epoch = 1 to epochs do

5: for batch (h_b, h_t, y) in train_loader do

6: # Feature mapping

7: F_b = W_b(h_b); F_t = W_t(h_t)

8: # Attention weight calculation

9: e_b = v·tanh(W_a·F_b); e_t = v·tanh(W_a·F_t)

10: α_b, α_t = softmax(e_b, e_t)

11: # Weighted fusion

12: F_fused = α_b·F_b + α_t·F_t

13: # Classification prediction

14: P = softmax(W_c·F_fused)

15: # Calculate the loss and update the parameters.

16: loss = CrossEntropy(P, y)

17: loss.backward()

18: optimizer.step()

19: # Testing phase

20: Use the trained model to predict on the test set and output the final result.