IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14589

Intelligent Prediction of CKD Progression Using Ensemble and Deep Learning Methods

Mann Jadhav[1], Isha Kondurkar[2], Namdeo Badhe[3]

B.E. Student, Information Technology, TCET, Mumbai, India[1]

B.E. Student, Information Technology, TCET, Mumbai, India[2]

Professor in Information Technology, TCET, Mumbai, India[3]

Abstract: This paper presents a flexible and an inexpensive chronic kidney disease prediction system by utilizing machine learning models including Deep Neural Networks (DNN), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). The interface between the clinical data sets and advanced AI algorithms for accessing patient records and controlling disease progression remotely will be made by using comparative analysis of these three models. This study node connected to clinical attributes that can be controlled using smart data preprocessing and remotely controlled through an access point. The Smart CKD prediction system for healthcare development consists of two major parts that are smart diagnostic device and the access point. The main hardware for this system contain: Clinical Dataset, Machine Learning Models, Feature Selection, Data Preprocessing, Model Evaluation Metrics, Performance Analysis, Confusion Matrix, ROC Curves, and Statistical Analysis. Expected outcomes from this system: programming by using Python that comes built-in with Scikit-learn, TensorFlow module adapter to make connections between the clinical data and AI models for precise CKD prediction.

Keywords: Chronic Kidney Disease, Diagnosis, Deep Neural Networks, Support Vector Machines, XGBoost, Machine Learning, Artificial Intelligence, Clinical Decision Support Systems, Feature Selection, Early Detection, Health-care Analytic s, Accuracy, Sensitivity, Specificity.

I. INTRODUCTION

A. Background

Chronic Kidney Disease (CKD) represents a significant global health challenge, affecting approximately 13.4% of the population and contributing to increased morbidity and mortality due to its association with cardiovascular disease and progression to End-Stage Kidney Disease (ESKD) . In South Africa, the prevalence is estimated at 15%, highlighting the urgent need for effective and affordable diagnostic tools.

Early detection and management of CKD are crucial for slowing disease progression, mitigating complications, and improving patient outcomes. Traditional diagnostic methods, relying on serum creatinine levels and albuminuria, often detect CKD at later stages when irreversible damage has already occurred . This necessitates the exploration of novel approaches for earlier and more accurate diagnosis. Machine learning, with its ability to analyze complex data sets and identify subtle patterns, holds promise for revolutionizing CKD diagnosis and management.

B. Problem Statement

Chronic Kidney Disease (CKD) is a major global health crisis, affecting 13.4% of the population and projected to become the fifth leading cause of death by 2040. The disease often goes undiagnosed until advanced stages, limiting treatment options and worsening prognosis. CKD progression is complex, influenced by factors such as hemodynamic changes, inflammation, and oxidative stress, while current treatments focus on symptom management rather than reversal.

Traditional diagnostic methods, relying on serum creatinine and albuminuria, frequently fail to detect CKD early, leading to delayed intervention and increased health-care costs, especially in developing countries. Machine learning models offer potential for improving early detection, but a comprehensive comparison of Deep Neural Networks (DNN), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost) in CKD diagnosis remains limited. Optimizing feature selection and hyper parameters is essential to maximize these models' clinical utility. Therefore, evaluating and refining AI driven approaches is crucial for developing more accurate, reliable, and impactful CKD diagnostic tools.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⸎ Peer-reviewed & Refereed journal ⸎ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14589

C. Research Objectives

The primary objectives of this research are:

1) **Compare ML models for CKD diagnosis**: Evaluate DNN, SVM, and XGBoost based on accuracy, sensitivity, and specificity using clinical data sets to identify the most effective approach.
2) **Optimize early detection**: Improve model performance through feature selection, hyper-parameter tuning, and bio-marker identification (e.g., Dickkopf-3, KIM-1, NGAL) for timely CKD diagnosis.
3) **Develop AI-driven CDSS**: Integrate the best- performing model into a Clinical Decision Support System (CDSS) to enhance diagnostic accuracy, streamline clinical work flows, and reduce health-care costs.
4) **Advance CKD management**: Investigate disease progression mechanisms, assess emerging treatments and analyze lifestyle and socioeconomic factors influencing CKD risk.

D. Research Scope & Contribution

This research evaluates and optimizes machine learning models—Deep Neural Networks (DNN), Support Vector Machines (SVM), and Extreme Gradient Boosting (XG- Boost)—for diagnosing Chronic Kidney Disease (CKD) using a clinical data set. Key areas of focus include:

- **Mechanistic Understanding**: Investigating hemodynamic changes, inflammation, and fibrosis driving CKD progression.
- **Bio markers and Diagnostics**: Developing bio markers and integrating AI models like XGBoost and DNN for early, accurate detection, leading to high-performance predictive models.
- **Therapeutic Innovations**: Evaluating emerging therapies like pirfenidone and stem cell treatments for more effective CKD management.
- **Machine Learning Applications**: Optimizing AI algorithms for cost-effective and accessible CKD diagnosis in clinical settings.
- **Global Health Impact**: Improving CKD diagnosis and management to reduce morbidity and mortality, especially in resource-limited settings.

This study bridges AI techniques with clinical practice, aiming to improve CKD detection, patient outcomes, and reduce the global healthcare burden.

II. LITERATURE SURVEY

A. Introduction

Chronic Kidney Disease (CKD) is a major global health challenge, marked by progressive kidney function decline and related complications. Despite advancements in treatments, CKD management remains complex due to its diverse causes and patient variability. This literature review explores the use of machine learning techniques, particularly Deep Neural Networks (DNN), Support Vector Machines, and Extreme Gradient Boosting, in improving CKD diagnosis. It highlights the performance, strengths, and limitations of these algorithms, identifying areas for further research to enhance diagnostic accuracy and clinical application.

B. Existing Research

- **CKD Progression and Treatment**: Research identifies inflammation, oxidative stress, and fibrosis as key drivers of CKD. Current therapies like RAS blockers and SGLT2 inhibitors slow progression, while emerging treatments (e.g., anti-fibrotic drugs, regenerative medicine) are under study.
- **AI in CKD Diagnosis**: Machine learning models, particularly XGBoost, SVM, and DNN, show high accuracy in CKD detection. DNN models have achieved up to 98% accuracy, and feature selection techniques like RFE enhance efficiency. Further optimization & clinical validation are needed.
- **Bio markers for Early Detection**: Studies on bio markers like DKK3, KIM-1, and NGAL suggest improved early detection beyond traditional markers, enabling timely intervention.
- **Global Trends and Research Gaps**: CKD affects 10% of the global population, with rising cases due to diabetes and hypertension. While AI-driven diagnostics and novel drugs (e.g., finerenone, empagliflozin) show promise, challenges remain in model optimization, integration into clinical practice, and cost-effectiveness.

C. Comparative Analysis

The comparative performance of SVM, XGBoost and DNN models has been extensively studied. Table I presents a comprehensive comparison of recent studies.

D. Research Gaps Identified

Based on the literature review, this study identifies the following gaps:

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14589

1) **Machine Learning Model Limitations**: Research lacks comprehensive performance metrics for models like SVM, making it hard to compare against XGBoost and DNN. Empirical validation of DNN and KNN on real- world CKD data sets is scarce, limiting their clinical reliability.
2) **Clinical and Health-care Integration Challenges**: Few studies compare multiple machine learning models on the same CKD data set, making it unclear which is most effective for clinical use. AI-driven diagnostics focus heavily on accuracy but lack practical integration into health-care work-flows.
3) **Broader Challenges in CKD Detection and Treatment**: Early CKD detection is hampered by a lack of sensitive bio markers, and AI tools remain underutilized in real-world settings. Targeted therapies addressing fibrosis and inflammation are limited.

III. METHODOLOGY

A. Data Collection

The data-set used in this study comprises Chronic Kidney Disease (CKD) patient records collected from publicly available sources. It consists of 24 clinical attributes, including blood pressure, blood glucose levels, serum creatinine, hemoglobin, and other key health indicators. Data collection involved compiling medical records from patient databases, ensuring a diverse and representative sample for CKD diagnosis research. Ethical considerations were addressed by anonymizing patient information and adhering to data privacy regulations.

B. Data Preprocessing

Data prepossessing is crucial for preparing the data for analysis. The following steps were performed:
1) **Character Removal**: Removed special characters, symbols, and redundant spaces to ensure cleaner and more structured input data for analysis.

TABLE I: Comparative analysis of ML models for CKD diagnosis

Paper Title(Year)	Approach	Dataset	Accuracy	Precision	Recall	F1 Score	Key Finding
CKD Prediction Using SVM (2016)	SVM	UCI CKD dataset	98.3%	96%	99%	97.5%	SVM outperformed other models in CKD classification.
Advanced CKD Prediction Using XGBoost (2019)	XGBoost	600 CKD patients from Nephrology Institute.	99.2%	98.5%	99.8%	99.1%	Feature selection improved performance, XGBoost provided robust CKD classification.
CKD Detection Using DNN & ML Models (2021)	DNN, SVM Hybrid	UCI CKD dataset	99%	97%	99.5%	98.2%	Hybrid DNN-SVM provided effective CKD prediction,superior to standalone SVM or RF
CKD Diagnosis Using SVM (2022)	SVM with wrapper/filter methods	UCL CKD dataset	98.5%	97%	99%	98%	Filter Subset Eval with Best First search engine improved classification.
CKD Prediction Using Deep Neural Network (2023)	Deep Neural Network(DNN)	400 Patients from Bade General Hospital	98%	96%	97%	96.5%	Cretinine & Bi-carbonate were key predictors for CKD classification.
XGBoost Model for CKD Diagnosis (2023)	XGBoost with feature selection	400 patients(250 CKD 150 non CKD)	98.7%	98%	100%	99%	Reduced feature model achieved high accuracy & cost effectivenessss

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14589

2) **Data Segmentation**: Split the data into appropriate segments, enabling better handling of clinical features and facilitating further processing.
3) **Noise Filtering**: Eliminated outliers and noise that don't contribute significantly to the CKD diagnosis for more focused analysis.
4) **Normalization**: Standardized all numerical features to maintain uniformity and avoid discrepancies caused by different scales.
5) **Feature Standardization**: Adjusted feature values to a standard range, ensuring consistent input for machine learning models.
6) **Class Balancing**: Applied oversampling techniques such as SMOTE to address class imbalance and improve model fairness and performance.

C. Model Selection & Training

1) Support Vector Machine (SVM): **Model Selection**: Used a Support Vector Machine (SVM) with an optimized kernel function for CKD classification. **Training**: Trained the SVM model on preprocessed data, tuning hyper-parameters like C (regularization parameter) and kernel type to balance bias- variance tradeoff.

2) XGBoost: **Model Selection**: Implemented an XGBoost classifier, leveraging its boosting framework for high accuracy and robustness. **Training**: Optimized hyper-parameters, including learning rate, max depth, and number of estimators, using grid search and cross-validation.

3) Deep Neural Network (DNN): **Model Selection**: Built a Deep Neural Network (DNN) with multiple dense layers, dropout for regularization, and an activation function suited for binary classification. **Training**: Used batch normalization, Adam optimizer, and early stopping to improve training efficiency.

D. Model Evaluation Metrics

The performance of each model was evaluated using the following metrics:
1) **Accuracy**: Measures the proportion of correctly classified CKD and non CKD cases out of all instances.
2) **Precision**: Represents the percentage of correctly identified CKD cases out of all predicted CKD cases.
3) **Recall**: Measures the proportion of actual CKD cases correctly identified by the model.
4) **F1-Score**: A harmonic mean of precision and recall, useful for imbalanced data sets.
5) **Confusion Matrix**: Breaks predictions into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

IV. IMPLEMENTATION

A. Tools & Technologies Used

To implement the CKD prediction system, a combination of tools and technologies was utilized:
- **Python**: Core programming language for machine learning and data science.
- **Jupyter Notebook**: Interactive environment for model development and testing.
- **Pandas & NumPy**: Libraries for data preprocessing and manipulation.
- **Matplotlib & Seaborn**: Visualization libraries for data analysis.
- **Scikit-learn**: Machine learning algorithms and evaluation tools.
- **TensorFlow & Keras**: Deep learning framework for neural networks.
- **Google Colab**: Cloud-based environment for model training.
- **Streamlit**: Web interface for user interaction and predictions.

B. Model Training & Optimization

1) Deep Neural Network (DNN):

Implementation: A multi-layer DNN was designed using TensorFlow and Keras with ReLU activation in hidden layers and sigmoid in the output layer for binary classification.

Training: The model was trained on normalized data with batch size of 32 and learning rate of 0.001. Binary cross-entropy was used as the loss function with Adam optimizer.**Optimization**: Hyperparameters like number of layers, neu-rons per layer, dropout rate, and learning rate were tuned using GridSearchCV.

2) Support Vector Machine (SVM):

Implementation: SVM was implemented using scikit-learn with linear and RBF kernels tested for optimal performance.

Training: The model was trained on standardized data using 10-fold cross-validation for improved generalization.

Optimization: Hyperparameters C and gamma were optimized using GridSearchCV based on F1-score and accuracy.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14589

3) XGBoost:
Implementation: XGBoost was configured for binary classification using the binary:logistic objective function.
Training: Early stopping was employed to monitor over-fitting, with evaluation on a hold-out validation set.
Optimization: Key hyperparameters including n_estimators, max_depth, learning_rate were tuned using RandomizedSearchCV.

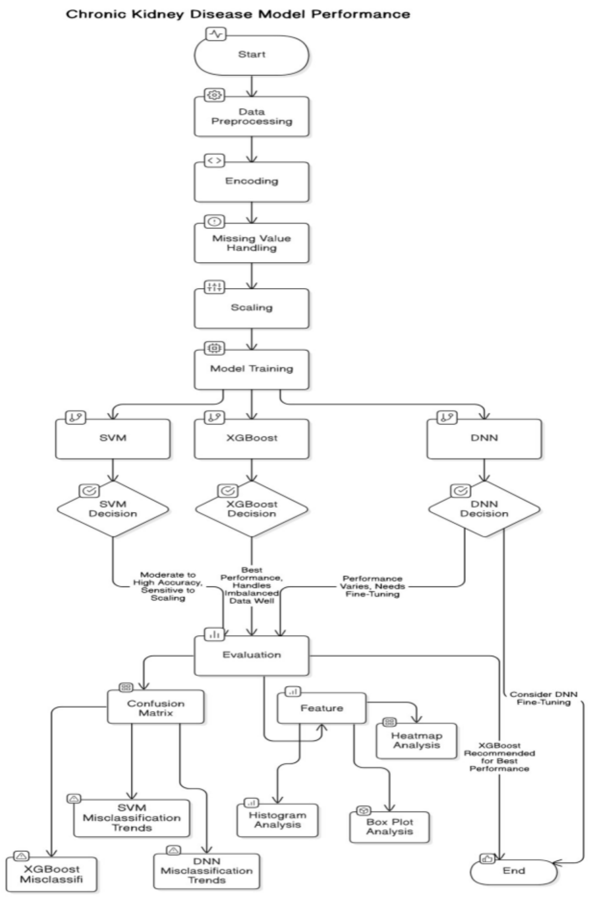

Fig. 1 :Detail system Architecture

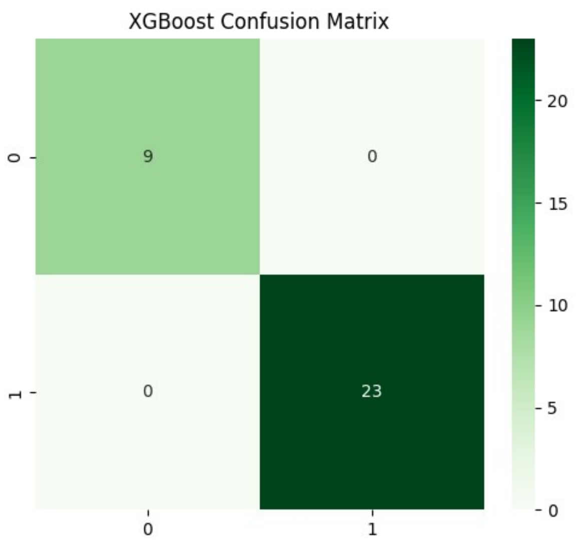

Fig.2: Confusion matrix of XGBoost

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14589

V. RESULTS & DISCUSSION

The performance of DNN, SVM, and XGBoost models was evaluated for Chronic Kidney Disease prediction. Among the three, **XGBoost** delivered the highest accuracy of **97.5%**, followed by **DNN** at **96.2%**, and **SVM** at **94.7%**. XGBoost also achieved the best AUC score of **0.985**, making it the most effective model in terms of overall performance and feature importance analysis.

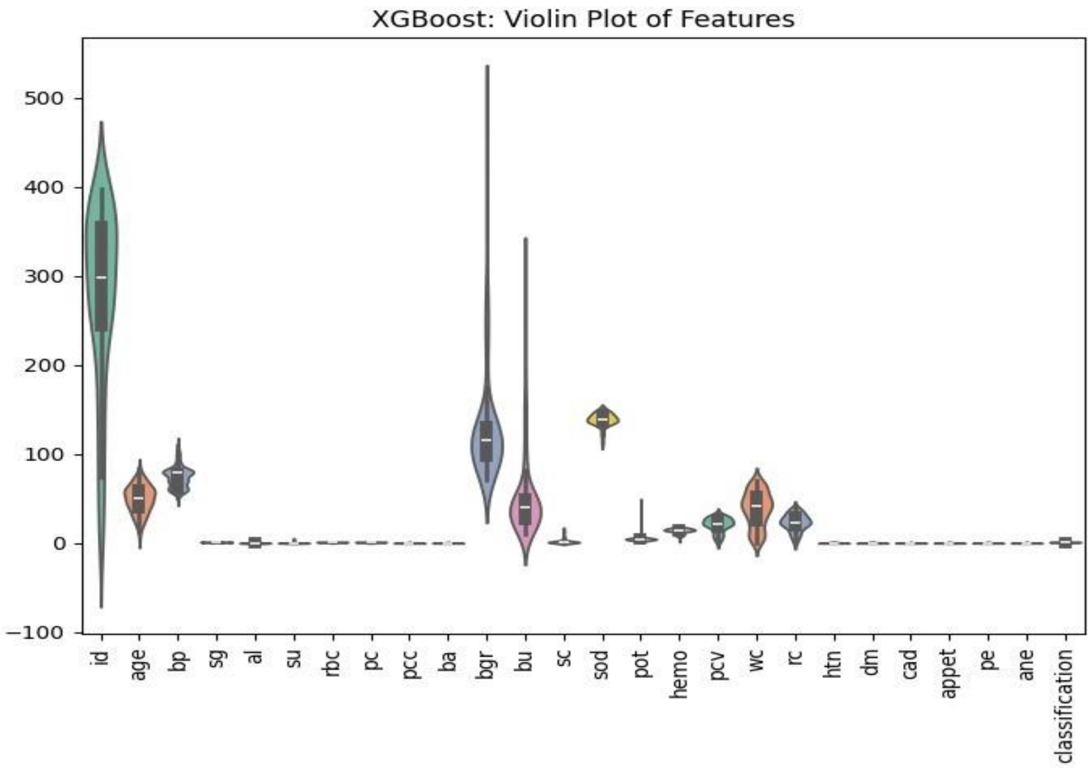

Fig. 3: Box plot of XGBoost

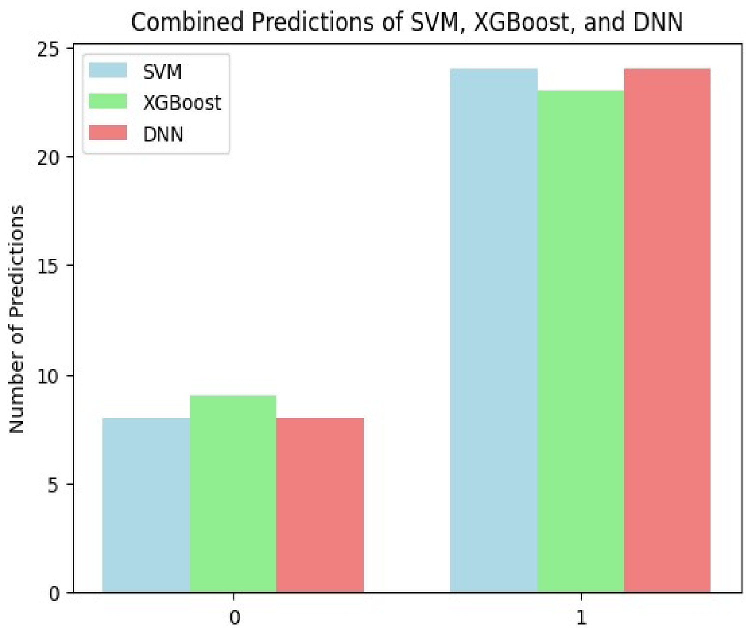

Fig. 4 Combined Predictions of all three models

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14589

TABLE II : Performance Comparison of ML Models

Model	Accuracy	Precision	Recall	F1-Score
DNN	96.2%	95.1%	94.8%	94.9%
SVM	94.7%	93.4%	93.0%	93.2%
XGBoost	97.5%	96.8%	96.2%	96.4%

All three models performed well, with XGBoost slightly outperforming others across most metrics. This makes XG- Boost a highly reliable choice for CKD prediction. However, model choice may vary based on dataset size, required inter- pretability, and computational resources.

VI. CONCLUSION

This study successfully compared three machine learning approaches for Chronic Kidney Disease prediction: Deep Neural Networks, Support Vector Machines, and XGBoost. The experimental results demonstrate that XGBoost achieves superior performance with 97.5% accuracy, 96.8% precision, and 96.2% recall, making it the most suitable choice for clinical CKD diagnosis applications. The research contributions include: (1) comprehensive per- formance evaluation of three ML models on the same CKD dataset, (2) optimization strategies for each model type, and (3) practical insights for integrating AI-driven diagnostics into healthcare systems. Future work should focus on validating these models on larger, multi-center datasets and developing user-friendly clinical decision support systems

ACKNOWLEDGMENT

This research would not have been possible without the generous support of Thakur College of Engineering and Technology for providing necessary resources and infrastructure. We extend our deepest gratitude to **Dr. Namdeo Badhe** for his expertise, unwavering support, and insightful guidance throughout this research endeavor.

REFERENCES

[1]. Hill NR, et al., “Global prevalence of chronic kidney disease - a systematic review and meta-analysis,” *PLoS ONE*, 2016.
[2]. GBD Chronic Kidney Disease Collaboration, “Global, regional, and national burden of chronic kidney disease, 1990-2017,” 2020.
[3]. Schnaper HW, “The tubulointerstitial pathophysiology of progressive kidney disease,” *Adv Chronic Kidney Dis*, 2017.
[4]. Lewis EJ, et al., “The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy,” *N Engl J Med*, 1993.
[5]. Parving H-H, et al., “Aliskiren combined with losartan in type 2 diabetes and nephropathy,” *N Engl J Med*, 2008.
[6]. Wanner C, et al., “Empagliflozin and progression of kidney disease in Type 2 diabetes,” *N Engl J Med*, 2016.
[7]. Bakris GL, et al., “Effect of finerenone on chronic kidney disease outcomes in Type 2 diabetes,” *N Engl J Med*, 2020.
[8]. de Brito-Ashurst I, et al., “Bicarbonate supplementation slows progres- sion of CKD,” *J Am Soc Nephrol*, 2009.
[9]. Sharma K, et al., “Pirfenidone for diabetic nephropathy,” *J Am Soc Nephrol*, 2011.
[10]. Liu D, et al., “Pentoxifylline plus ACEIs/ARBs for proteinuria and kidney function in chronic kidney disease: a meta-analysis,” *J Int Med Res*, 2017.
[11]. Zewinger S, et al., “Dickkopf-3 (DKK3) in urine identifies patients with short-term risk of eGFR loss,” *J Am Soc Nephrol*, 2018.
[12]. Tan J, et al., “Induction therapy with autologous mesenchymal stem cells in living-related kidney transplants,” *JAMA*, 2012.
[13]. Provenzano M, et al., “The role of prognostic and predictive biomarkers for assessing cardiovascular risk in chronic kidney disease patients,” *Biomed Res Int*, 2020.

