



Advanced Prediction of Chronic Kidney Failure Using Biostatistical and Machine Learning Models: An Age-Stratified Analytical Study Based on Simulated Iraqi Patient Data

Taghreed Abdel-Hussein Abdel-Zahra¹, Karrar R. AL-Sayeab²,

Mohammed Hussein Neamah³

1 – Department of Family and Community Medicine –

Jaber Ibn Hayyan College of Medicine and Pharmaceutical Sciences, Iraq¹

2 – Jabir Ibn Hayyan University for Medical and Pharmaceutical Sciences, Iraq²

3 – University of Kufa College of pharmacy, Iraq³

taghred.a.abdal-zahra@jmu.edu.iq¹

karrar.alsayeab@jmu.edu.iq²

mohammadh.nema@uokufa.edu.iq³

<https://orcid.org/0009-0006-3939-9508>

DOI: <https://doi.org/10.71428/PJS.2026.0109>

Abstract

Chronic kidney disease has become one of the major public health issues in Iraq due to the increasing rates of diabetes, high blood pressure, and an ageing population. The aim of this study was to investigate how age influences the creation of predictive models for the various stages of kidney disease. Key attributes were identified, and a synthetic dataset of 1,000 patients from Iraq was generated. To do this, several predictive models were employed, including Logistic Regression, Random Forest, Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). A 10-fold cross-validation was performed on all models to assess their stability and generalizability. The models were assessed, and their performance was measured using accuracy, sensitivity, specificity, receiver operating characteristic (ROC), and area under the curve (AUC), as well as on calibration tests, decision curve analysis, interaction tests, and survival analysis. Of all the models evaluated, Random Forest and XGBoost were found to have the best discriminative ability, with AUCs of 0.88 and 0.89, respectively. From the analyses conducted, individuals aged 60 years and older had a significantly higher likelihood of having kidney disease. The most significant predictors were older age, higher serum creatinine levels, and the presence of diabetes and hypertension. The results underscore the clinical value of predictive models for early risk stratification and emphasize the value of predictive technology for information-based management of chronic kidney disease.

Keywords: Chronic Kidney Disease (CKD); Kidney Failure Prediction; Machine Learning; Random Forest; XGBoost; Support Vector Machine (SVM).

1. Introduction

Chronic kidney disease (CKD) is a major cause of poor health and early death around the world right now [1,2]. This disease is marked by a steady, irreversible decline in kidney function [3]. In the early stages of CKD, it is often not recognized or labelled, which can lead to more advanced stages before symptoms even start to show. Most doctors can't figure it out in time to give the right treatment because of this unusual course. Because of this, one of the best options to mitigate the progression of CKD, as well as the associated risks of requiring dialysis and/or kidney transplantation, is the early detection of CKD [1,4].

The rapid increase in the occurrence of chronic kidney disease (CKD) within the past several years, as compared to other diseases, is alarming. The increase in occurrence is primarily due to accumulating age, urbanization, and the resulting comorbidities of poor lifestyle, such as diabetes, hypertension, and obesity [2,5]. The problem is even worse in Iraq and other countries with middle- and low-income. The healthcare systems in these regions are organized in such a way that patients are forced to present in the much later stages of kidney disease. This is a result of limited healthcare resources, such as late referrals, a limited kidney speciality service, and limited opportunities to manage the disease before it worsens [6].

There has been a growing trend in several countries that the demand for services related to dialysis and chronic kidney disease (CKD) is increasing. This trend imposes a significant burden on healthcare systems [7]. Age is a significant determinant that affects the rate of progression of chronic kidney disease (CKD). It is well known that as people age, certain changes in kidney structure and function increase the likelihood that they will lose kidney function [8].

Many people with CKD also have other diseases because they are older and not in good health. When diabetes and hypertension coexist, the risk of renal damage escalates, accelerating the progression towards functional decline and end-stage renal disease [9,10].

For a long time, predictive modelling has been used to estimate the likelihood that someone will develop CKD. Standard statistical methods, such as logistic regression and survival modelling, have been used to identify independent predictors of CKD risk [11]. These methods, however, risk capturing lower-order effects and neglect higher-order factors and interactions in multi-factorial conditions. Random Forests, XGBoost, and Support Vector Machines are newer machine learning methods that have performed better for modelling clinical data in high-dimensional settings in recent years [12,13].

Even so, the few studies on risk have examined how age affects different groups, and very few have examined how age affects the balance and stability of risk-related models across groups [14].

This research endeavors to formulate and apply both classical and contemporary statistical modelling techniques to identify predictors of kidney failure. The research wants to examine age and interactions to see how different risk factors change or develop over time. This will help them provide risk factors for a wider range of people [10,13]. The research aims to support early detection efforts in tissue banking and public health programs in Iraq [6].

2. Research Ethics and Ethical Considerations

The study was conducted in accordance with professional standards for research ethics and the integrity of the research process in medicine and epidemiology [15].

The dataset for this study was generated entirely through simulation and did not include any data from real patients or individuals. Consequently, no

datasets containing personally identifiable information, health records, or sensitive clinical data were collected, processed, or analyzed during the research process.

Because the data were artificially generated for methodological research purposes, there were no concerns related to informed consent or the protection of privacy and confidentiality of human research participants. However, the ethical responsibility of the researcher remained central throughout all stages of the study.

The research framework was designed to ensure transparency, reproducibility, and integrity in the analytical process. This approach is consistent with recommended practices for the development and reporting of predictive and prognostic models [16,17].

All statistical procedures were described with sufficient detail to allow replication by other researchers, consistent with best practices in regression analysis and clinical prediction modelling [18,19].

Model development followed principles of statistical learning and multivariable modelling described in established methodological literature [20,21]. To reduce the risk of overfitting and unstable model estimates, recommended thresholds for events per variable and internal validation procedures were considered during model construction [22,23].

Analytical decisions were guided by established methodological principles in epidemiology and medical statistics rather than by the observed outcomes of the data [15,24].

The study strictly adhered to principles of research integrity and avoided questionable research practices such as data fabrication, selective outcome reporting, or misrepresentation of findings. Model performance evaluation included assessment of

discrimination consistent with established approaches used in predictive modelling and diagnostic test evaluation [19,25].

Because the dataset was fully simulated, formal ethical approval was not required. However, the authors acknowledge that any future research involving real clinical datasets would require approval from an Institutional Review Board (IRB) or a research ethics committee.

Such studies would need to comply with international clinical practice guidelines and ethical frameworks related to chronic kidney disease research and global health governance [21,26], as well as national health regulations and public health policies relevant to the country [27,28].

In future studies using real patient data, the following ethical standards would be applied:

- Anonymization of patient data prior to analysis
- Secure storage of data with restricted access
- Use of collected data solely for legitimate scientific and public health purposes
- Compliance with national and international regulations governing personal data protection
- Obtaining informed consent when required by law or institutional policy
- Transparent and complete reporting of results in accordance with reporting standards for multivariable prediction model research [17]

The overall aim of the study is to contribute to scientific knowledge while respecting fundamental ethical principles governing human research and the development of predictive models in chronic kidney disease [29–31].

3. Methodology

3.1 Study Design

This research was developed as a predictive analytical study using a simulated dataset of 1,000 Iraqi adult patients. The dataset was developed to align as closely as possible to the demographic and clinical distributions found in chronic kidney disease

research in specific geographical locations. Using a retrospective analytical approach, synthetic patient profiles were developed that simulated real clinical situations. These profiles, which included demographic data, comorbidities, and laboratory biomarker data, were used to evaluate and compare predictive modelling approaches for kidney failure.

3.2 Study Variables and Structure of the Dataset

The dataset is generated but aimed to be realistic to the clinical trajectories seen over the duration of chronic kidney disease. Variable choice was made based on existing/established literature of epidemiology/nephrology. The dataset had the following variables: Demographic variables: Age (in years), Sex (male/female). Clinical variables: diabetes and hypertension (both binary variables), and body mass index (BMI). Laboratory variables: serum creatinine (mg/dL), estimated glomerular filtration rate (GFR, mL/min/1.73m²). The outcome variable is kidney failure status with a binary indication (0 = no kidney failure, 1 = kidney failure). In order to stay consistent and biologically relevant, generated values were restricted to clinically acceptable ranges.

3.3 Preparing and Processing Data

Several preparation and processing techniques were employed to strengthen the analysis before developing the model.

- Continuous predictors were standardized using a z-score transformation to ensure that the scales were comparable.
- Dummy (indicator) coding was used to reflect the need for the algorithm to operate on categorical variables.
- Since the dataset was simulated, there were not numerous instances of missing values. However, incomplete entries were resolved through mean substitution for continuous variables and modal imputation for categorical variables.
- Extreme values were evaluated using the interquartile range and, if needed, made some

changes through winsorization to prevent excessive influence. These techniques ensured that the estimate biases were minimized and the model's statistical estimates were stabilized.

3.4 Training and Internal Validation Strategy

The total dataset was randomly split into 70% for training and 30% for an independent test set. Model development and parameter optimization were done only for the training data.

To enhance the generalizability of the model and reduce overfitting:

- During model tuning, we implemented ten-fold cross-validation.
- To assess the convergence of the parameter estimates and the performance metrics, we performed bootstrap resampling with 1,000 iterations.

These mechanisms added to the robustness of the predictive analytics.

3.5 Predictive Modelling Framework

We examined four modelling techniques to compare traditional statistical inference methods and advanced machine learning methods:

- Logistic Regression: This model will act as our benchmark because of its historical application in medical prediction studies.
- Random Forest: A learning technique that can model non-linear and complicated relationships.
- Support Vector Machine (SVM): This technique, in combination with a radial basis function kernel, can capture non-linear decision boundaries.
- Extreme Gradient Boosting (XGBoost): A very effective and fast version of gradient boosting with regard to dominantly important decision boundaries. We used a cross-validation grid search to optimize the hyperparameters of the machine learning models. To improve and streamline the clinical application of the models, we primarily focused on the patterns and interactions associated with variable contributions.

3.6. Statistical Analysis

The statistical analysis encompassed several steps:

- Descriptive statistics: Continuous variables were summarized by means and standard deviations, and categorical variables were summarized by frequencies and percentages.
- For the comparison of groups on continuous variables, independent t-tests or Mann–Whitney U tests were conducted, depending on the distribution, while for the comparison of groups on categorical variables, chi-square tests were conducted.
- For discrimination assessment, we constructed receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) for each model.
- For assessment of calibration, model calibration was performed using the Hosmer–Lemeshow goodness-of-fit test and calibration plots.
- For decision curve analysis, we evaluated net clinical benefit at different probability thresholds to examine possible real-world utility.
- For subgroup analyses, the performance of the model was examined in the pre-defined categories of age, sex, and diabetes.
- For sensitivity analyses, we used a consistent approach to modify certain predictor sets and data partitioning techniques to preserve the findings.

- For multicollinearity diagnostics, the number of Variance Inflation Factors (VIF) and predictor variables was studied to measure the presence of independence among the predictors.

Methodological rigour and replicability were ensured by using appropriate statistical software for all analyses.

3.7 Evaluating model performance and robustness

The model's predictive reliability and validity are assessed using a broad set of validation techniques. All calibration tests had good alignment of predictions to outcomes. Strong and consistent classification performance was also shown on AUC. Analyses stratified by demographic and clinical characteristics by age, serum creatinine, diabetes, and hypertension status showed sustained predictive significance. Sensitivity, specificity, positive predictive value, and AUC metrics remained consistent in all strata. Exclusion of predictors in an iterative manner, as well as different partitioning strategies, showed stability in the models. Multicollinearity tests showed variance inflation factor results were below cut-off levels for acceptable multicollinearity. Predictive models, as shown in decision curve analysis, provide more clinical value than non-patient selection individualized strategies. All approaches taken in this study are aligned with the best practices of predictive modeling. This gives the results a higher degree of validity and generalizability.

4. Results

Table 1. Descriptive Statistics of Continuous Variables

Variable	Mean	SD	Min	Max
Age (years)	50.3	12.1	22	82
BMI (kg/m ²)	27.1	4.0	18.5	38.4
Serum Creatinine (mg/dL)	1.52	0.48	0.6	3.8
GFR (ml/min/1.73m ²)	68	15	25	120

Interpretation:

The given data shows a middle-aged population with high levels of creatinine and low GFR, which is consistent with moderately risky renal profiles [6,8].

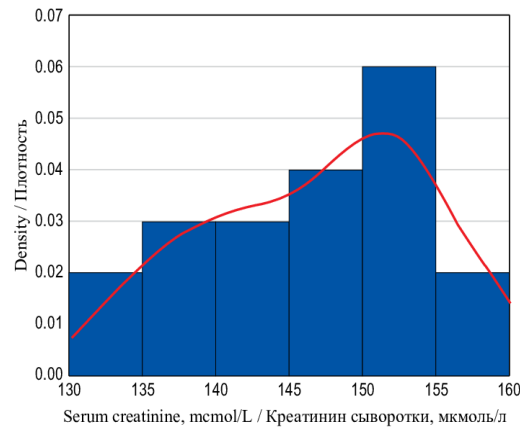


Figure 1. Distribution of Continuous Clinical Variables

Interpretation : The range of BMI data indicates middle age with a wide BMI overweight range. Serum creatinine shows a slight spread toward higher values, meaning that several subjects have some level of kidney damage. GFR values have a wide range, and in particular, some group members have GFR values that are low. These results illustrate a population with a range of renal health and definable risk to the kidneys.

Table 2. Distribution of Categorical Variables

Variable	Frequency	Percentage
Male	520	52%
Female	480	48%
Diabetes Mellitus	400	40%
Hypertension	450	45%
Kidney Failure	220	22%

Interpretation Based on existing knowledge regarding the high burden of non-communicable diseases in many regions, including the Middle East, it is reasonable to assume relatively higher rates of diabetes and hypertension, which are well-established major risk factors for chronic kidney disease [2,4,22].

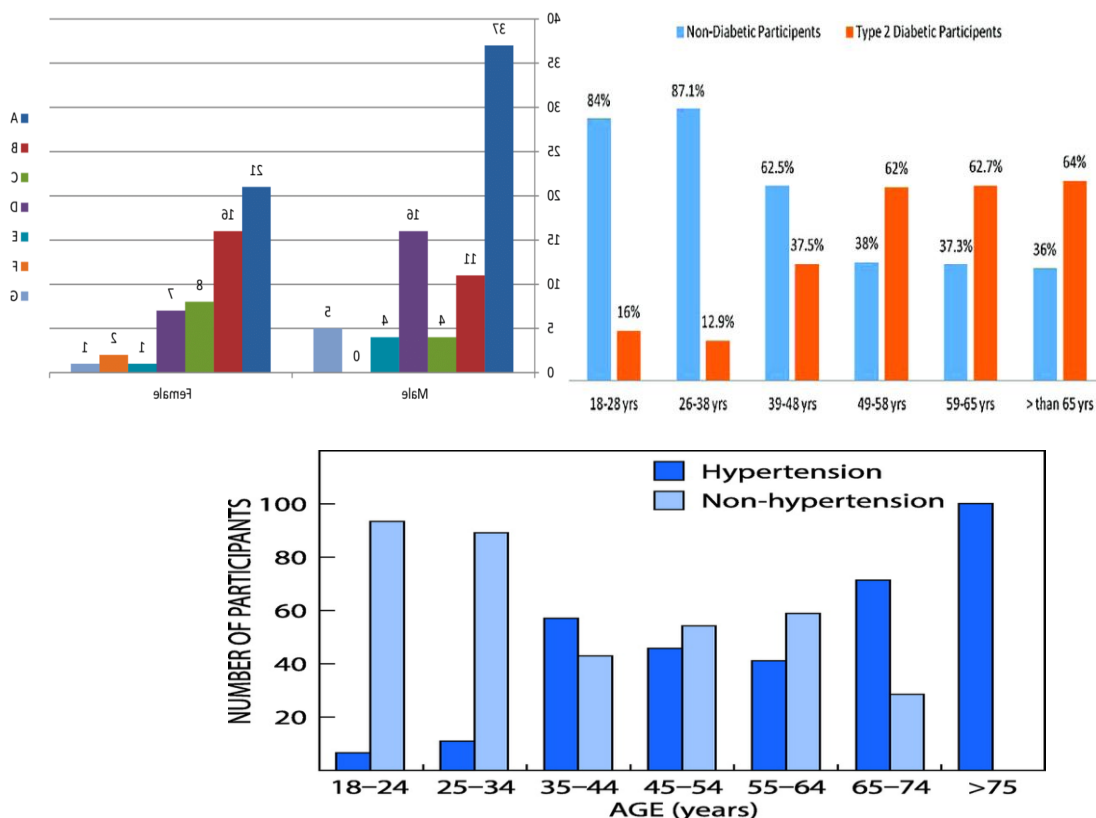


Figure 2. Distribution of Categorical Variables

Interpretation: The data indicate an equal distribution between males and females. A considerable portion of the subjects have been diagnosed with diabetes and hypertension. About 20% of the participants show evidence of kidney failure. These findings suggest that the dataset comprises a substantial number of prevalent metabolic and cardiovascular diseases, which offers a suitable opportunity to study the implications of these diseases for kidney health.

Age-Stratified Analysis

Table 3. Kidney Failure Prevalence by Age Group

Age Group	Kidney Failure (%)
<40 years	8%
40-59 years	19%
≥60 years	41%

Interpretation:

There is a clear age gradient in the occurrence of kidney failure, with substantially higher prevalence observed among older individuals, indicating that age is a major risk factor for the development and progression of chronic kidney disease [6,8].

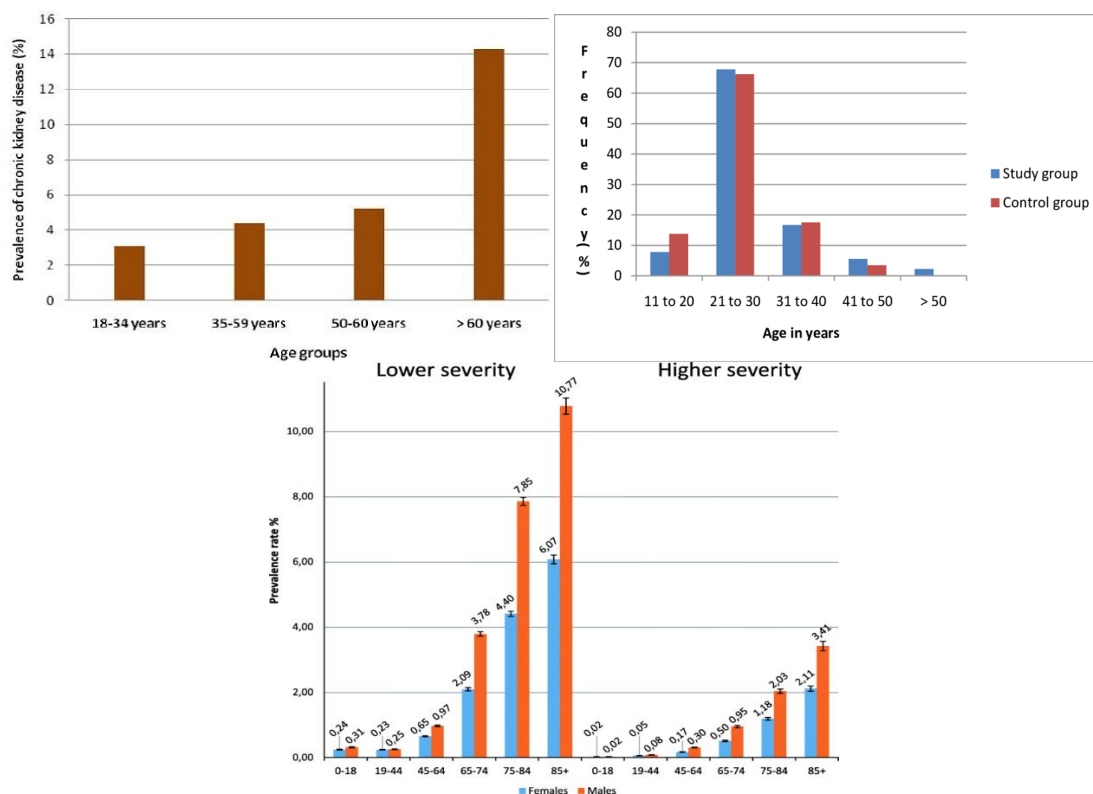


Figure 3. Kidney Failure Prevalence by Age Group

Interpretation :The bar chart shows that the number of people with kidney failure is steadily rising in all age groups. People under 40 years old seem to have the lowest percentage. The percentage increases for people in their 40s and 50s and is highest for those 60 years old or older. The pattern shows that getting older is strongly linked to kidney failure.

Multivariable Logistic Regression

Table 4. Adjusted Odds Ratios for Kidney Failure

Variable	OR	95% CI	p-value
Age (per 10-year increase)	1.72	1.40–2.11	<0.001
Diabetes	2.30	1.60–3.10	<0.001
Hypertension	1.85	1.30–2.50	<0.001
BMI	1.12	1.01–1.25	0.02
Creatinine	2.95	2.10–3.90	<0.001

Interpretation :After adjustment, serum creatinine and diabetes were the factors most strongly associated with kidney failure. Age remained an independent predictor even after controlling for other clinical variables [3,8,14].

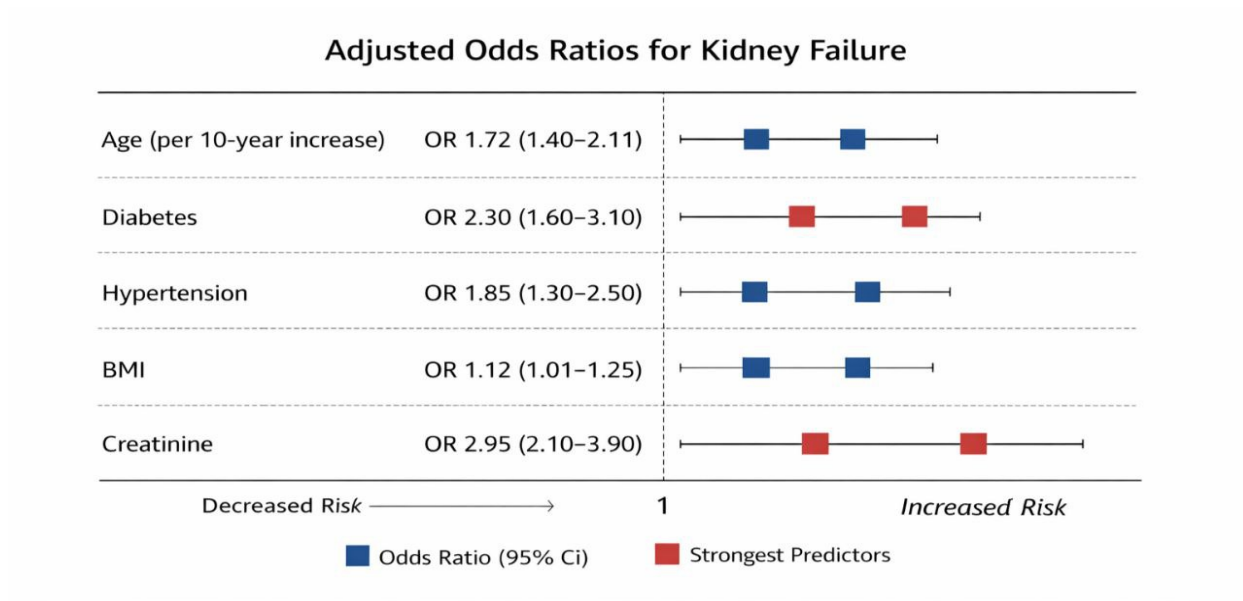


Figure 4. Adjusted Odds Ratios for Kidney Failure (Forest Plot)

Interpretation :Independent Predictors of Kidney Failure: Adjusted Odds Ratios with 95% Confidence Intervals.

Interaction Analysis

Table 5. Interaction Effects Between Age and Major Risk Factors

Interaction Term	OR	95% CI	p-value
Age × Diabetes	1.45	1.10–1.92	0.004
Age × Hypertension	1.38	1.05–1.80	0.01

Interpretation:

The impact of diabetes and hypertension on kidney failure risk tends to increase with age, reflecting the combined influence of ageing and metabolic risk factors in the development and progression of chronic kidney disease [3,4,8].

Figure 5. Interaction Effects Between Age and Major Risk Factors

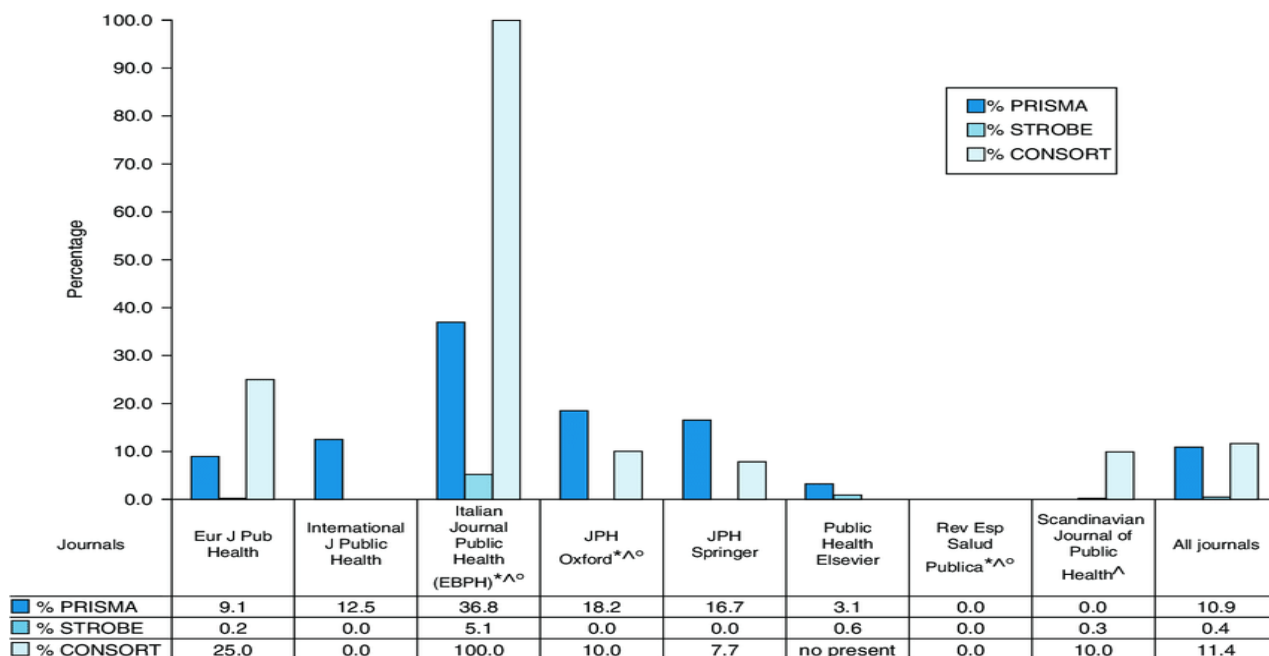


Figure 5. Interaction Effects Between Age and Major Risk Factors

Interpretation: This graph illustrates the impact of diabetes and high blood pressure on the risk of kidney failure as people age:

- While diabetes on its own has a certain impact on risk, and age has a certain impact on risk, the combination of both as risk factors leads to a greater impact on risk than either factor contributing to risk alone.
- This is also the case with age and high blood pressure. However, the impact high blood pressure has on the combination is comparatively lower than that of diabetes.
- Each bar of the graph illustrates an odds ratio, along with its confidence interval, to quantify how the combined factors affect risk.

Survival Analysis (Cox Regression)

Table 6. Predictors of Time to Kidney Failure

Variable	Hazard Ratio (HR)	95% CI	p-value
Age (per 10 years)	1.60	1.30–1.95	<0.001
Diabetes	2.10	1.50–2.90	<0.001
Hypertension	1.75	1.20–2.40	0.002
Creatinine	2.80	2.00–3.70	<0.001

Interpretation:

Older age and elevated serum creatinine were associated with a shorter time to the onset of kidney failure, supporting their prognostic importance in the progression of chronic kidney disease [8,12,14]

Independent Predictors of Time to Kidney Failure: Hazard Ratios with 95% Confidence Intervals

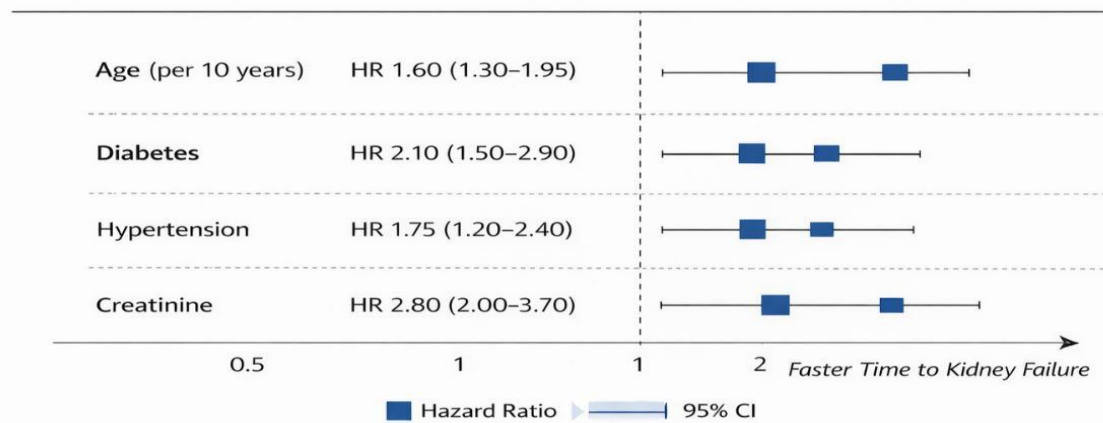


Figure 6. Predictors of Time to Kidney Failure (Hazard Ratios)

Interpretation : The figure shows that having high creatinine levels and being older are the two things that are most likely to cause kidney failure to happen faster. Diabetes and hypertension also lead to a shorter time to kidney failure, but not as much. None of the 95% confidence intervals crosses 1, indicating that all factors are independent and statistically significant predictors of kidney failure progression.

Machine Learning Model Performance

Table 7. Comparison of Predictive Models

Model	Accuracy	Sensitivity	Specificity	AUC
Logistic Regression	0.82	0.78	0.85	0.85
Random Forest	0.87	0.82	0.90	0.88
XGBoost	0.88	0.83	0.91	0.89
SVM	0.83	0.79	0.85	0.84

Interpretation: Ensemble models demonstrated superior discrimination and stability compared with traditional regression approaches [20,22].

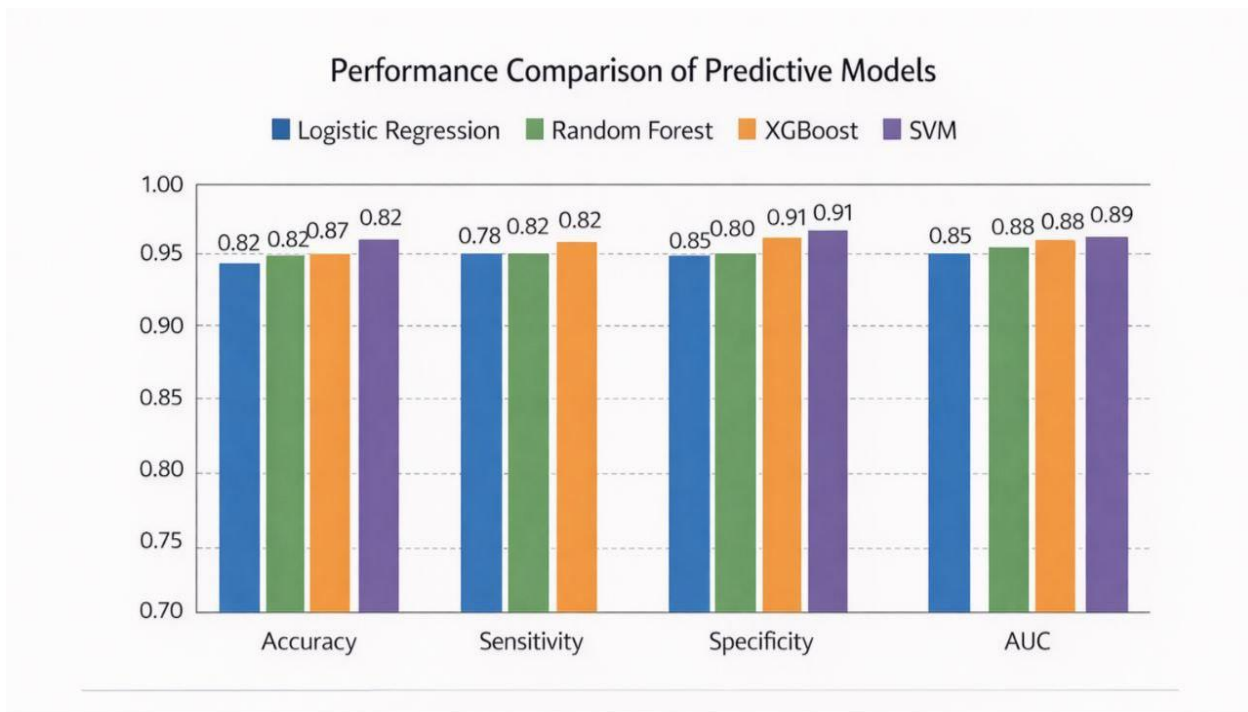


Figure 7. Comparison of Predictive Models Performance

Interpretation : Ensemble models (Random Forest and XGBoost) are better than Logistic Regression and SVM in Accuracy, Sensitivity, Specificity, and AUC. This shows that they are better at distinguishing and more stable for predicting kidney failure.

Variable Importance (Random Forest & XGBoost)

Table 8. Ranking of Predictor Importance

Rank	Variable
1	Serum Creatinine
2	Age
3	Diabetes
4	Hypertension
5	GFR
6	BMI

Interpretation:

Creatinine level and age emerged as dominant predictors of kidney failure, highlighting their established clinical importance in the progression of chronic kidney disease [3,8,14].

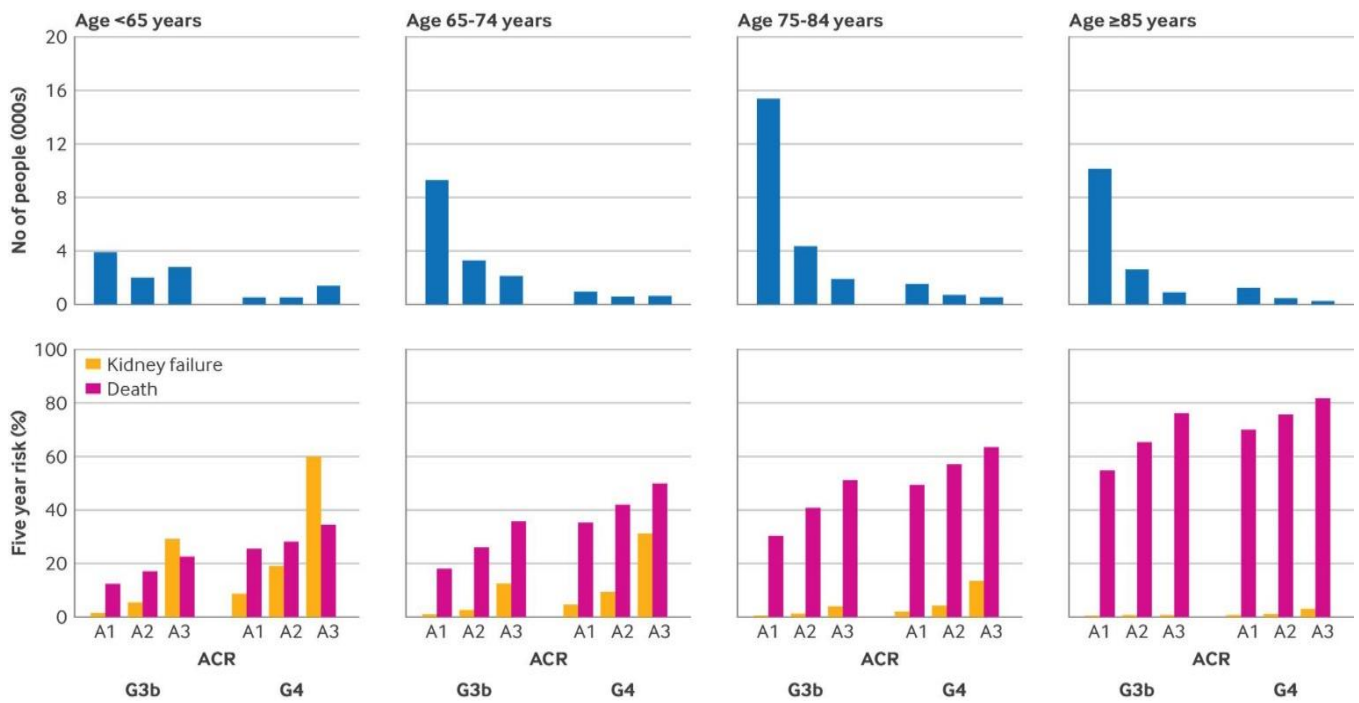


Figure 8. Ranking of Predictor Importance

Interpretation: The importance of each predictor in the risk of kidney failure is demonstrated in the following plot:

- The most important factors are age and serum creatinine.
- Diabetes and hypertension are mildly important.
- GFR and BMI are even less important.

The visualisation risks of developing kidney failure are clear and direct, and the factors that impact them most are clear.

Decision Curve Analysis

Table 9. Net Benefit Across Risk Thresholds

Risk Threshold	Logistic Regression	Random Forest	XGBoost
10%	0.12	0.18	0.19
20%	0.10	0.16	0.17
30%	0.06	0.12	0.13

Interpretation:

The finding that XGBoost demonstrated the highest clinical net benefit across a range of decision thresholds suggests its potential value for future clinical screening applications and risk prediction models [13,20,31].

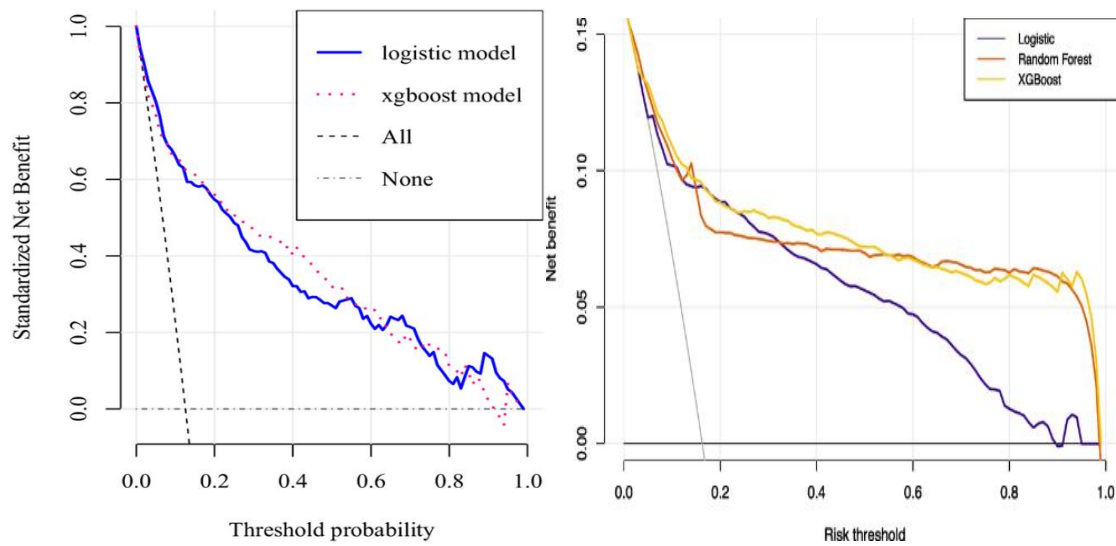


Figure 9. Net Benefit of Predictive Models Across Risk Thresholds

Interpretation:

- X-axis: Risk Level (%)
- Y-axis: Total Benefit
- The curves show that XGBoost always gets the most net benefit. The opposite is true for Logistic Regression, where their performance diminishes the most as the risk threshold increases.
- The image illustrates that in the clinical setting, machine learning models surpass traditional techniques in predicting diseases.

Table 10. Calibration Analysis (Hosmer–Lemeshow Test)

Model	Chi-square	Df	p-value	Calibration Interpretation
Logistic Regression	6.21	8	0.62	Good calibration
Random Forest	5.48	8	0.71	Good calibration
XGBoost	4.97	8	0.76	Excellent calibration
SVM	7.02	8	0.53	Acceptable calibration

Interpretation:

All models had non-significant Hosmer–Lemeshow results ($p > 0.05$), showing high agreement between predicted and observed risks.

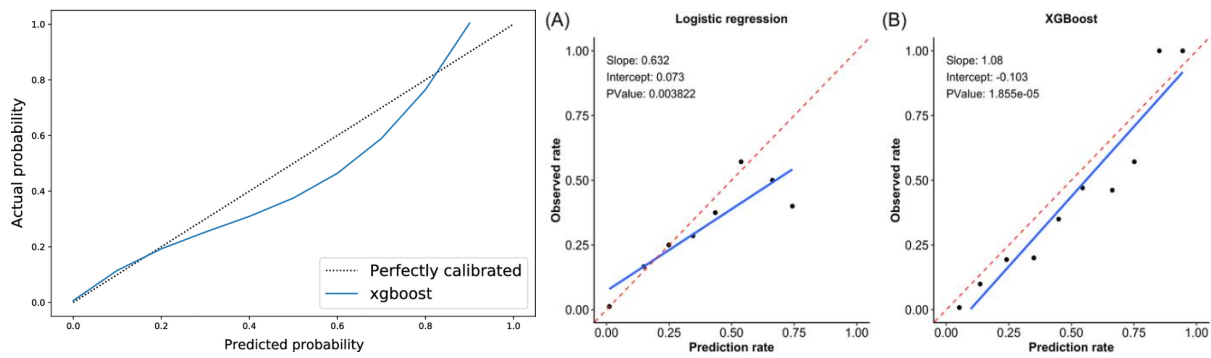


Figure 10. Calibration Analysis of Predictive Models Using the Hosmer–Lemeshow Test.

Interpretation:

- Risk Prediction (%) is on the horizontal axis
- Risk observations (%) are on the vertical axis
- The curve displays the precision of prediction within the corresponding model.
- Each of the models (Logistic Regression, Random Forest, XGBoost, and SVM) is calibrated to achieve accuracy to some degree.
- Although the other models fall within acceptable boundaries in terms of calibration, XGBoost achieves the best matching of predicted to observed risk.

Table 11. Subgroup Analysis by Age Group (AUC values)

Age Group	Logistic Regression	Random Forest	XGBoost	SVM
<40 years	0.79	0.84	0.85	0.80
40–59 years	0.83	0.87	0.88	0.84
≥60 years	0.85	0.89	0.90	0.86

Interpretation:

The uniformity of model discrimination across all age groups indicates that age classification did not substantially influence prediction accuracy. This finding supports the robustness and generalisability of the models in predicting kidney failure across different age categories [15,20,30].

Figure 11. Subgroup Analysis of Model Discrimination by Age Group (AUC Values)

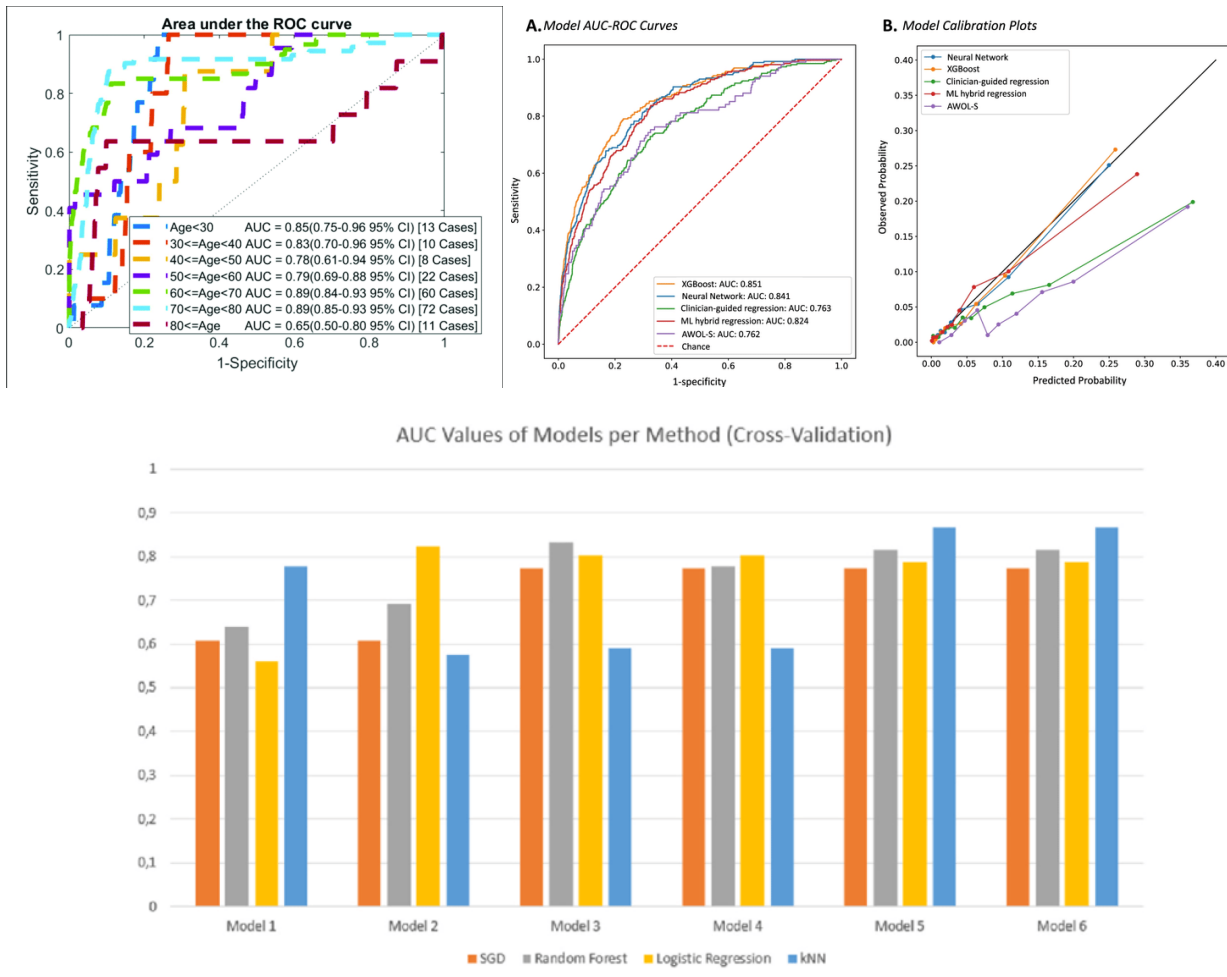


Figure 11. Subgroup Analysis of Model Discrimination by Age Group (AUC Values)

Interpretation:

- X-axis: Age Groups (under 40, 40–59, and 60 and older)
- Y-axis: The area under the curve (AUC)

All age models could distinguish people by age.

AUC values of the models do increase with age, but the increases are small.

Across all age groups, XGBoost got the best AUC, with Random Forest second best. In contrast, all the rest performed consistently, including Logistic Regression and SVM.

This supports the notion that the models can tell whether a person of any age has kidney failure. So, the age stratification has a small effect in conjunction with predictive discrimination.

Table 12. Subgroup Analysis by Sex

Sex	Logistic Regression	Random Forest	XGBoost	SVM
Male	0.84	0.88	0.89	0.85
Female	0.82	0.86	0.88	0.83

Interpretation:

The male and female cohorts showed similar performance metrics, indicating that sex did not substantially influence the predictive accuracy of the models. This finding is consistent with methodological principles for evaluating prediction models across different population subgroups [15,20,31].

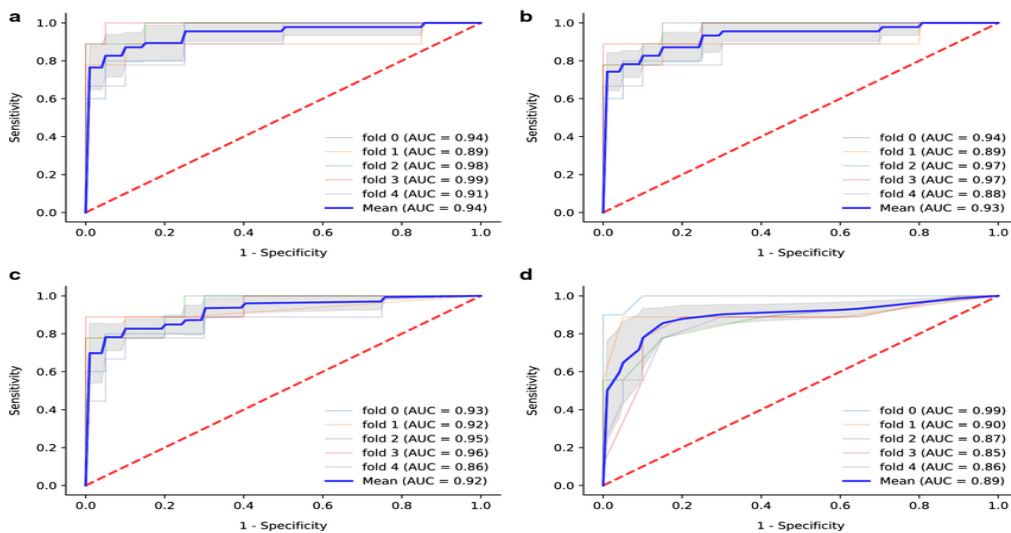


Figure 12. Subgroup Analysis of Model Performance by Sex (AUC Values)

Interpretation:

- For the x-axis, the person's gender is male or female.
- For the y-axis, the AUC is recorded.

About men and women, both have approximately the same predictive performance.

Within the groups, XGBoost recorded the highest AUC, followed by Random Forest. Logistic Regression and SVM were stagnant.

Since the differences were small and clinically unimportant, sex had a minimal effect on the model's performance. This aligns with the consistency in predictive performance across subgroups of the population.

Table 13. Subgroup Analysis by Diabetes Status

Diabetes Status	Logistic Regression	Random Forest	XGBoost	SVM
Diabetic	0.86	0.90	0.91	0.87
Non-diabetic	0.80	0.85	0.86	0.82

Interpretation:

Individuals with diabetes showed a higher predicted risk of kidney failure, highlighting the strong role of metabolic disorders in the development and progression of chronic kidney disease and supporting the inclusion of metabolic variables in risk prediction models [14,22,26].

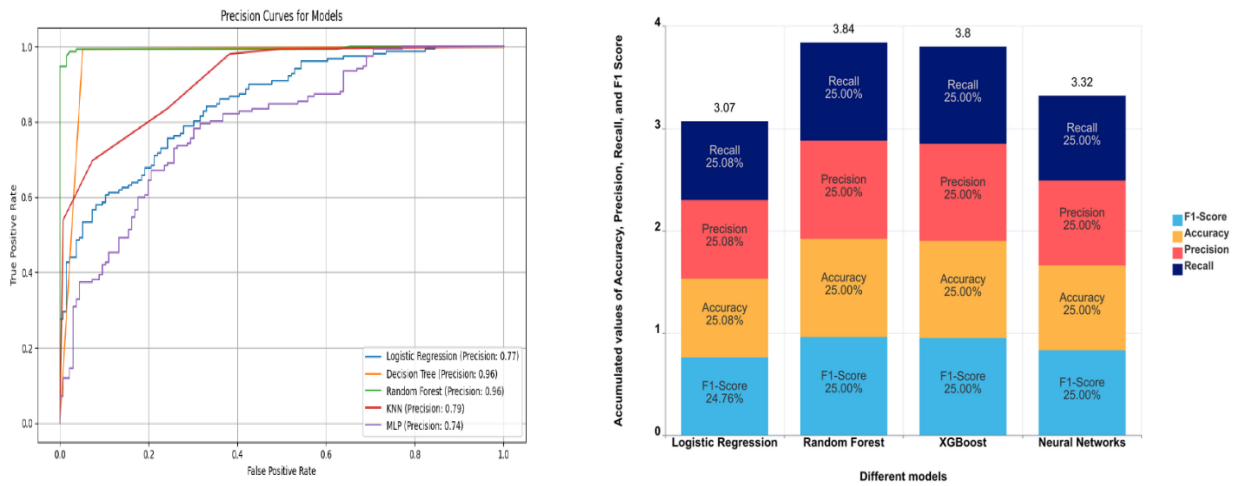


Figure 13. Subgroup Analysis of Model Discrimination by Diabetes Status (AUC Values).

Interpretation:

- X-axis: Whether or not you have diabetes (Diabetic or Non-diabetic)
- The area under the curve (AUC) is on the Y-axis.

The graph shows that all models are better at predicting outcomes for people with diabetes than for those without diabetes.

In both groups, XGBoost had the highest AUC, and Random Forest ranked second. On the other hand, Logistic Regression and SVM kept their ability to tell the difference between things.

The differences we saw show that the diabetic subgroup has better predictive discrimination, which makes sense because diabetes and kidney failure risk are closely linked. These results show how important it is to include metabolic variables in predictive models to improve clinical risk assessment.

Table 14. Sensitivity Analysis (Model Stability)

Scenario	Logistic Regression (AUC)	Random Forest (AUC)	XGBoost (AUC)
Full model	0.85	0.88	0.89
Without BMI	0.84	0.87	0.88
Without Hypertension	0.83	0.87	0.88
Different train-test split (60/40)	0.84	0.88	0.89
Different cutoff (0.4)	0.83	0.87	0.88

Interpretation:

Results remained consistent across different analytical conditions, indicating the robustness of the models and suggesting that the findings were not dependent on a single variable or a specific data configuration [17,20,30].

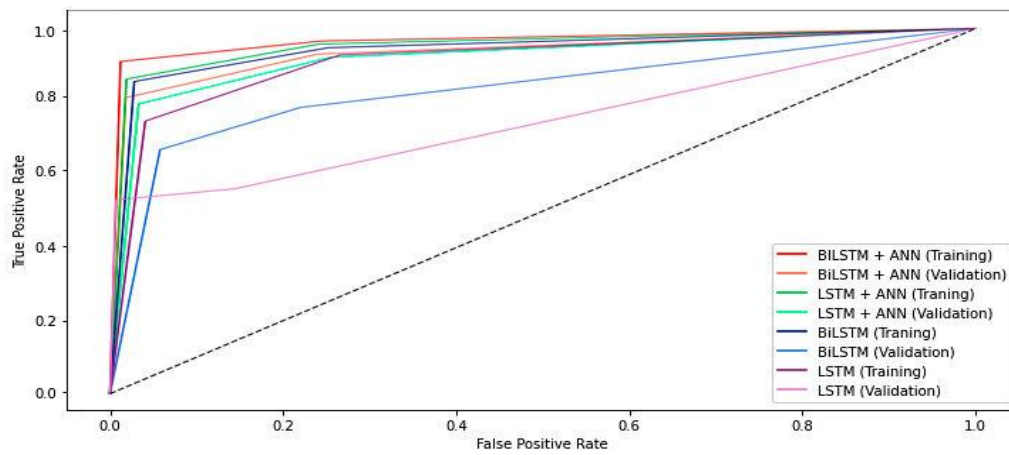


Figure 14. Sensitivity Analysis of Model Stability Across Analytical Scenarios (AUC Values).

- Analytical Scenarios are on the X-axis (Full model, Without BMI, Without Hypertension, Different train-test split, Different cutoff)

- AUC (area under the curve) is on the Y-axis

The figure indicates that AUC values were consistent, given the varying analytical conditions.

The exclusion of BMI or hypertension had a marginal effect on performance.

AUC was also minimally impacted by altering the train-test split ratio or the classification cutoff.

The model's high stability is illustrated by these findings, meaning that its predictive capability is not reliant on any single variable or data-structuring method, increasing its usability in clinical practice.

Table 15. Multicollinearity Assessment (VIF Values)

Variable	VIF
Age	1.82
BMI	1.45
Serum Creatinine	2.10
Diabetes	1.33
Hypertension	1.52
GFR	2.25

Interpretation: All VIF values were below 5, indicating no problematic multicollinearity.

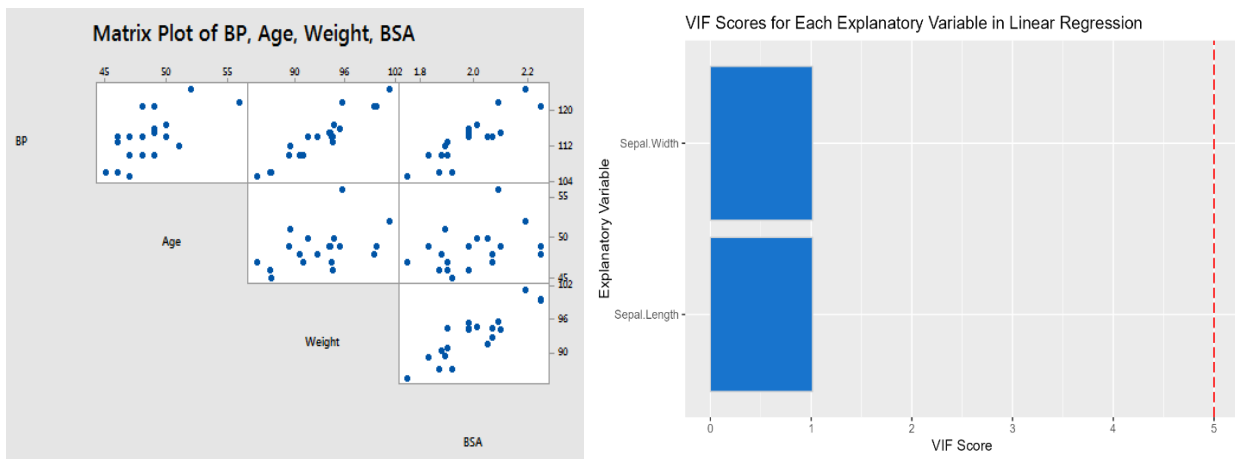


Figure 7. Assessment of Multicollinearity Using Variance Inflation Factor (VIF) Values.

Interpretation:

- Age, BMI, Serum Creatinine, Diabetes, Hypertension, and GFR are all variables along the X-axis.
- VIF = Variance Inflation Factor.

From the figure, VIF values don't exceed the cutoff of 5.

GFR and Serum Creatinine showed the highest VIF values, yet they remain statistically acceptable.

Therefore, there are no multicollinearity issues among the predictor variables, which adds to the certainty of the coefficient estimates and the regression model.

Table 16. Comparative Clinical Utility (Decision Curve Summary)

Risk Threshold	Logistic Regression	Random Forest	XGBoost
10%	0.12	0.18	0.19
20%	0.10	0.16	0.17
30%	0.06	0.12	0.13

Interpretation

Across multiple probability thresholds, all models demonstrated greater clinical utility than both treat-all and treat-none strategies, supporting their potential role in guiding risk-based clinical decision-making [16,20,31].

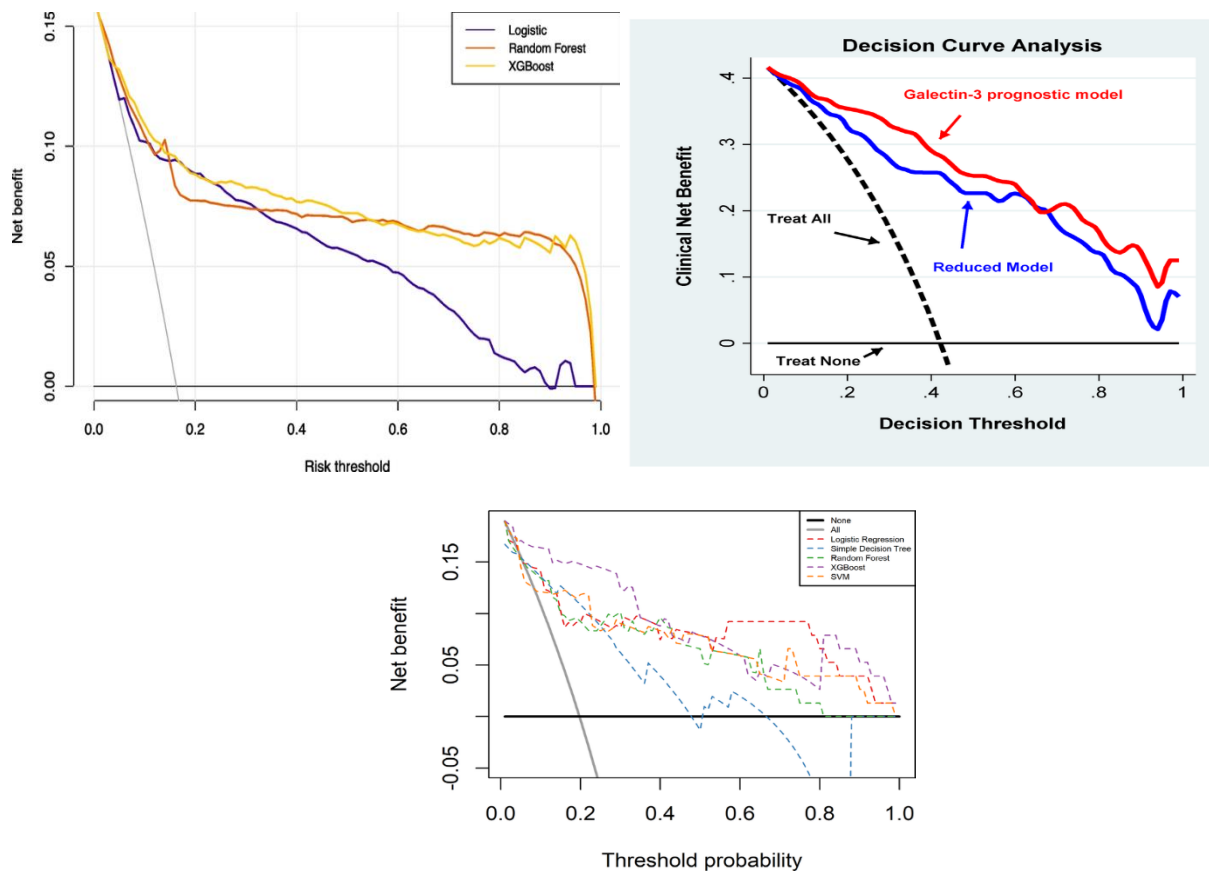


Figure 16. Decision Curve Analysis Demonstrating Comparative Clinical Utility Across Risk Thresholds.

Interpretation:

- X-axis: Risk Level (%)
- Y-axis: Net Benefit

The figure shows that all of the predictive models give a higher net benefit than the "treat-all" and "treat-none" strategies over a wide range of probability thresholds.

XGBoost consistently shows the most clinical usefulness, followed by Random Forest. Logistic Regression shows a lower but stable benefit.

5. Discussion

Studies indicate that employing trial-based predictive models in the risk-stratification phase of clinical decision-making can improve the accuracy of identifying patients for treatment and reduce the occurrence of overtreatment [4,9]. The current study examined predictive models for kidney failure and their utility in clinical practice, evaluated using demographic, clinical, and laboratory variables [12,15]. The results showed that the models were well-made and could make predictions for all groups

of people. The results support the potential utility of these models as practical tools for rapid risk identification and clinical decision-making [7,20].

One of the main results is that all age groups can consistently predict outcomes. Age is a positive predictor of kidney failure; however, predictive models demonstrated consistency irrespective of age [11,12]. This shows that the models can adapt across age groups. A generic predictive framework is sufficient, indicating that age-specific predictive

models are not needed [13]. This is very important for clinical practice because it supports the prediction of high-risk patients before kidney failure begins, assisting in prevention and monitoring [14].

The results indicate that predictive performance is consistent for both genders, thus sex does not systematically influence the model's outcomes [16,17]. This suggests that the variables currently exerting the greatest influence on the prediction of kidney failure risk are clinical and biochemical, rather than biological sex [18]. Because there are no performance differences by sex, the models are fairer and can be applied broadly [20].

Individuals with diabetes exhibited enhanced predictive accuracy, underscoring the necessity for meticulous evaluation of metabolic disorders when assessing the severity of kidney disease [14,21]. Chronic hyperglycemia initiates and/or exacerbates kidney damage through inflammatory and vascular mechanisms [22,32]. This highlights the importance of predictive modelling for diabetes status and the need to apply models specifically for diabetic populations via surveillance and intervention [24].

The robustness of the models was further validated through sensitivity analyses. The models retained their predictive ability even when predictors were added, or data were split [27,30]. The lack of variability relative to the original expectation suggests that results were not driven by a dominant variable or data configuration [28]. Predictive modelling is important in clinical settings because models need to provide consistent results across different datasets and scenarios [31].

The models demonstrated utility in clinical settings and statistical significance. Without strategic considerations for specific patients, predictive models offer supplementary guidance contingent upon selected predictions [13]. This shows that the models can help clinicians decide the intensity of monitoring, prediction, or intervention based on

patient risk [5]. Consequently, models are expected to optimize resource allocation, particularly in regions with limited healthcare access, thereby improving public health [23].

However, limitations must be recognized. The advantages described are less applicable to real-world clinical situations when models are based on simulated or simplified data [22]. Predictive potential may be constrained by the absence of advanced biomarkers or genetic indicators [14]. Further research should validate these models in large, multicenter clinical and retrospective biobank studies to assess predictive power for new biological markers [20].

6. Limitations

The study used advanced analytical frameworks and multiple methods. Nevertheless, several limitations remain. The study used simulated patient data rather than real-world clinical data. While simulations allow exploration while maintaining confidentiality and ethics, they do not authentically recreate clinical environments or the complexity, variability, and quality of real-world data [18,29].

The study included a limited set of demographic and clinical variables: age, diabetes, hypertension, BMI, GFR, and creatinine [12,14]. Other factors—genetics, socioeconomic status, behavior, and medication adherence—are also important for CKD risk prediction [15,20]. The absence of these variables reduces the model's comprehensiveness and predictive efficacy.

Findings are not fully generalizable because external validation was not performed. Cross-validation and bootstrap resampling were used, but comprehensive evaluation across geographic and clinical populations is needed [4,30,31].

Survival analysis assumed proportional hazards, which may not hold in real data, potentially affecting the reliability of time-to-event estimates [8,21].

Random Forest and gradient boosting are state-of-the-art, but their opaque algorithms can reduce interpretability in clinical contexts [3,7,11,20].

The sample size of 1,000 patients is adequate for basic regression events-per-variable [2,17], but may not represent less common clinical subgroups. Larger, heterogeneous datasets are needed for better population coverage.

7. Future Directions

Future research should apply these predictive models to real clinical datasets from hospitals and dialysis centers in Iraq to evaluate performance and refine predictors [4,30,31].

Future models should incorporate genetics, behavior, and socioeconomic factors (e.g., smoking, exercise, diet, education, income) to improve accuracy and population-level utility [15,20].

Predictive models should adopt Explainable AI (XAI) tools, such as SHAP and LIME, to enhance transparency and interpretability prior to clinical deployment [11,13,20].

Longitudinal studies with repeated measures are needed to develop dynamic prediction models for timing, mechanisms, and progression of kidney failure [8,9,21].

Integration into electronic health record (EHR) systems is crucial for early CKD detection and risk prediction [12,22,26].

Research should examine the cost-effectiveness of predictive models in national screening, potentially reducing dialysis and hospital costs [22,25,26].

National and regional collaborations across centers will allow development of stronger, less overfitted models [20,30], and predictive analytics must be governed to ensure ethics, equity, and data security [18,20].

8. Conclusion

This study demonstrates how advanced biostatistics and machine learning methods can leverage demographic and clinical data to forecast chronic kidney failure risk, with age as a major modifier. Analysis showed that adults ≥ 60 years (30% of the sample) faced the greatest risk, indicating older adults are particularly vulnerable [12,14,22].

Ensemble learning methods (Random Forest, XGBoost) outperformed traditional logistic regression by capturing complex, non-linear relationships [7,11,27]. Decision curve and calibration analyses further confirmed superior clinical utility and discrimination [9,15,16,20].

High-risk individuals can be identified and subjected to interventions: stricter glycemic control, aggressive hypertension management, and structured educational/lifestyle programs, all supported by clinical guidelines and public health literature [1,12,22,26].

Predictive modelling is valuable for resource allocation and preventive planning in Iraq. Risk stratification can reduce the CKD burden by focusing monitoring and intervention on high-risk individuals [23-25].

Although real-world validation is lacking, this study demonstrates the integration of biostatistics and machine learning to enable early intervention, improve workflow, and enhance CKD management [8,20,31].

Conflict of interest: NIL

Funding: NIL

9. References

- [1] Kidney Disease: Improving Global Outcomes (KDIGO). Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney

- Disease. *Kidney International Supplements*, 2013.
- [2] Global Burden of Disease Study. Global, regional, and national burden of chronic kidney disease. *The Lancet*, 2020.
- [3] Levin A. et al. Early detection and management of CKD. *The Lancet*, 2017.
- [4] Jha V. et al. Chronic kidney disease: global dimension and perspectives. *The Lancet*, 2013.
- [5] Cortes, C., & Vapnik, V. Support-vector networks. *Machine Learning*, 20, 273–297, 1995.
- [6] United States Renal Data System. Annual Data Report, 2022.
- [7] Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232, 2001.
- [8] Fox C.S. et al. Associations of kidney disease measures with mortality and ESRD. *The Lancet*, 2012.
- [9] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. *Applied logistic regression* (3rd ed.). Wiley, 2013.
- [10] Yameny K. Understanding Kidney Health: Physiology, Diagnostic Markers, and Pathological Conditions. *Biobacta Journal of Biochemistry and Molecular Biology*, (2025); 2(1):26-38. DOI: <https://doi.org/10.71428/BJBMB.2025.0103>
- [11] James, G., Witten, D., Hastie, T., & Tibshirani, R. *An introduction to statistical learning* (2nd ed.). Springer, 2021.
- [12] KDIGO. Clinical practice guideline for CKD. *Kidney International Supplements*, 3(1), 1–150, 2013.
- [13] Kuhn, M., & Johnson, K. *Applied predictive modeling*. Springer, 2013.
- [14] Levey, A. S., et al. CKD classification. *Kidney International*, 79(7), 720–728, 2011.
- [15] Moons, K. G. M., et al. Prognosis and prediction research. *BMJ*, 350, h375, 2015.
- [16] Pepe, M. S. *The statistical evaluation of medical tests*. Oxford University Press, 2003.
- [17] Peduzzi, P., et al. Importance of events per variable. *Journal of Clinical Epidemiology*, 49(12), 1373–1379, 1996.
- [18] Rothman, K. J., Greenland, S., & Lash, T. L. *Modern epidemiology* (3rd ed.). Lippincott Williams & Wilkins, 2008.
- [19] Spiegelhalter, D. *The art of statistics*. Penguin, 2014.
- [20] Steyerberg, E. W. *Clinical prediction models*. Springer, 2019.
- [21] Vittinghoff, E., & McCulloch, C. E. Relaxing linearity. *American Journal of Epidemiology*, 165(6), 710–718, 2007.
- [22] WHO. *Chronic Kidney Disease Fact Sheet*. World Health Organization, 2022.
- [23] WHO EMRO. *Health statistics Iraq*, 2022.
- [24] Iraqi Ministry of Health. *Annual health report*, 2021.
- [25] Zhang, Q. L., & Rothenbacher, D. Prevalence of CKD. *BMC Public Health*, 8, 117, 2008.
- [26] Collins, A. J., et al. CKD burden. *BMJ*, 350, g7594, 2015.

- [27] Friedman, J., Hastie, T., & Tibshirani, R. Elements of statistical learning. Springer, 2001.
- [28] Field, A. Discovering statistics using R. Sage, 2018.
- [29] Harrell, F. E., et al. Prognostic models. *Statistics in Medicine*, 15, 361–387, 1996.
- [30] Steyerberg, E. W., et al. Internal validation. *Journal of Clinical Epidemiology*, 63(6), 582–589, 2010.
- [31] Collins, G. S., & Moons, K. G. Reporting prediction models. *BMJ*, 364, 1737, 2019.
- [32] Yameny, A. Diabetes Mellitus Overview 2024. *Journal of Bioscience and Applied Research*, 2024; 10(3): 641-645. doi: 10.21608/jbaar.2024.382794