A Statistical and Machine Learning Framework to Identify Determinants of Renal Treatment Choice in CKD Patients

[1] Abhijeet Jadhav, [2] Pratik Morade, [2] Het Keniya
[1] Dept of Community Medicine, MGM Medical College, Panvel, MGM Institute of Health Sciences, Kamothe, Navi Mumbai, India - 410209
[2]Dept of Dialysis Therapy Technology, MGM School of Biomedical Sciences, Kamothe, MGM Institute of Health Sciences, Kamothe, Navi Mumbai, India –410209

Corresponding Author: Abhijeet Jadhav

Abstract: Background: Selecting an appropriate renal replacement therapy (RRT) in chronic kidney disease (CKD) is influenced by multiple demographic, socioeconomic, and clinical factors. Understanding these determinants is essential to support individualized treatment decisions. To model relationships among health science variables, statistical and machine learning (ML) procedures provide a robust framework. **Objectives:** To identify key determinants influencing RRT modality choice among CKD-G5D patients using traditional statistical and machine learning approaches. **Methods:** A cross-sectional study was conducted on 241 patients with CKD-G5D. Baseline demographic, socioeconomic, and clinical variables were analyzed descriptively and comparatively using the Kruskal-Wallis and Chi-square tests. Predictive modelling was conducted using multinomial logistic regression and Random Forest algorithms. Model performance was assessed using confusion matrices, accuracy, and variable importance metrics. **Results:** Haemodialysis was the predominant modality (97.1%), followed by kidney transplantation (1.7%) and peritoneal dialysis (1.2%). Age and educational status showed significant associations with RRT choice ($p = 0.016$ and $p = 0.005$, resp).The Random Forest model achieved the highest predictive accuracy (98.6%), identifying duration of CKD and age as the most influential predictors. Younger patients with higher educational attainment were more likely to undergo transplantation, whereas older individuals predominantly received haemodialysis. **Conclusion:** Age and educational attainment are pivotal determinants guiding RRT modality among CKD-G5D patients. Machine learning approaches, particularly Random Forest, demonstrate strong potential to enhance predictive accuracy and support personalized, data-driven counselling in renal care.

Key words: CKD, RRT, Predictive Modelling, Machine Learning, Regression, Clinical Decision Support

1. Introduction

Chronic kidney disease (CKD) affects a substantial proportion of the global population and presents a growing burden in terms of morbidity, mortality and cost of care. For patients progressing towards end-stage kidney disease (ESKD), the selection of an appropriate renal treatment pathway including options such as haemodialysis, peritoneal dialysis, and transplantation represents a critical decision point. Despite this, the factors influencing the choice of renal replacement modality remain under-explored, especially when compared with the literature on when to initiate renal replacement therapy (RRT).Meanwhile, machine learning (ML) methods are increasingly applied in nephrology, for tasks such as CKD diagnosis, progression prediction and RRT initiation modelling. Bai et al. (2022) used ML to predict ESKD in CKD cohorts, achieving good sensitivity compared to traditional risk formulas. Delrue et al. (2024) summarised applications in the assessment of renal pathology and the prediction of functional decline.

In clinical practice, treatment choice for CKD patients is influenced by patient-level factors (age, comorbidities, laboratory values such as eGFR), sociodemographic factors (education, socioeconomic status, geography), provider and system factors (referral timing, centre policies, modality availability) and patient preferences. Traditional regression-based methods are well suited for identifying linear relationships but may struggle with complex, non-linear and interacting determinants. In contrast, ML frameworks allow managing high-dimensional data, capturing interactions and providing feature-importance insights (e.g., SHAP, LIME) that enhance model interpretability.

The present study aims to develop and validate a ML framework to identify and quantify the determinants influencing renal treatment selection among patients with CKD. By doing so, the study aims to enhance personalized decision support, identify modifiable factors influencing modality selection, and ultimately support improved alignment of therapy decisions with individual patient profiles.

2. Literature Review:

Early studies examining treatment modality choice in advanced CKD primarily focused on clinical and social determinants. Morton et al. (2012) demonstrated that socioeconomic status, nationality, language barriers, and functional dependence substantially influenced the choice between peritoneal dialysis (PD) and hemodialysis (HD), emphasizing the significant impact of social and contextual factors in renal replacement therapy (RRT) decision-making. Subsequent research by Morton et al. (2012) highlighted that patient preferences, including perceived survival benefit, frequency of hospital visits, and travel burden, also shape decisions regarding dialysis versus conservative care.

As CKD research evolved, attention expanded toward patient-centered and demographic determinants of RRT choice. Cho et al. (2022) and Arenas et al. (2024)

examined treatment modality selection among elderly CKD patients, identifying age, employment status, comorbidities, and planned dialysis initiation as major predictors of PD versus HD selection. In parallel, Cortvrindt et al. (2024) explored nephrologist perspectives, emphasizing that organizational infrastructure, cultural beliefs, and institutional policy also influence dialysis modality allocation. Collectively, these studies emphasize that treatment selection in CKD is multifactorial, shaped by both patientlevel and systemlevel factors.

In recent years, the application of machine learning (ML) has transformed CKD research. Debal et al. (2022) demonstrated that ML techniques specifically Random Forest, Support Vector Machine, and Decision Tree algorithms can accurately classify CKD stages, establishing the feasibility of ML approaches in nephrology. Tsai et al. (2023) applied ML-based variable selection on longitudinal laboratory data to identify key predictors of functional decline, such as estimated glomerular filtration rate (eGFR), and systolic blood pressure, in non-dialysis CKD patients. Building on this, Okita et al. (2024) developed and validated an ML model to predict time to initiation of RRT, which outperformed traditional eGFR-decline methods.

Although these studies highlight the predictive strength of ML in CKD progression, few have applied such techniques to model treatment modality choice. This gap justifies the present study's aimto integrate ML frameworks to identify the determinants of renal treatment choice in CKD patients, combining clinical, demographic, and socioeconomic predictors within a data driven analytical framework.

3. Materials and Methods

3.1 Study Design

This cross-sectional observational study was conducted among patients with chronic kidney disease stage G5D (CKD-G5D) attending nephrology outpatient clinics and dialysis units of three tertiary-care hospitals in Navi Mumbai, India, between March and December 2024. The study aimed to examine sociodemographic and clinical factors influencing patients choice of renal replacement therapy (RRT), including maintenance hemodialysis, peritoneal dialysis, and kidney transplantation.

3.2 Study Participants

All adult patients aged ≥18 years with a confirmed diagnosis of CKD-G5D were eligible. Patients with incomplete records or who had undergone multiple RRT modalities were excluded. A total of 241 patients were included in the final analysis after applying predefined inclusion and exclusion criteria.

3.3 Data Collection and Variables

Data were obtained from patient medical records and standardized structured questionnaires. The dataset included the following variable:

Dependent variable: Choice of RRT (Hemodialysis, Peritoneal Dialysis, or Kidney Transplantation)

Independent variables: Demographic and social factors: Age, Gender, Marital status, Educational status, Employment status, and Socio-economic status Clinical factor: Duration of CKD-G5D (years): All categorical variables were coded numerically, and continuous variables were summarized as means and standard deviations (SD).

3.4 Statistical Analysis

All statistical analyses were performed using R studio (R version 4.4.3).

3.4.1 Descriptive Statistics

Descriptive statistics were used to summarize baseline characteristics. Continuous variables were presented as mean ± SD, median (IQR) as appropriate while categorical variables were expressed as frequency and percentage.

3.4.2 Comparative Analysis

Comparative analyses were performed to evaluate the associations between RRT choice and explanatory variables. Continuous variables were compared using the Kruskal-Wallis rank-sum test, given non-normal distributions. Categorical variables were compared using the Chi-square test or Fisher's exact test where applicable. A p-value < 0.05 was considered statistically significant.

3.4.3 Predictive Modelling Framework

To identify predictors of RRT modality, machine learning models were developed using 70:30 random training-test data split. Multinomial logistic regression served as the baseline parametric model. The Random Forest algorithm was employed as a non-parametric ensemble learning method. Model performance was evaluated using confusion matrices, overall accuracy, sensitivity, specificity, and Kappa statistic. For the Random Forest model, variable importance was assessed based on the Mean Decrease in Gini and the Mean Decrease in Accuracy indices. The modelling framework followed standard supervised learning workflows, including feature encoding, model training, validation, and interpretation.

3.5 Ethical Considerations

Ethical clearance for the study was obtained from the Institutional Ethics Committee of the participating academic institution (Approval No: MGM/DCH/IEC/IN/SBS/56/01/2024). The present study is based on a secondary analysis of de-identified data that were collected between March and December 2024 under the approved protocol, without any modification to the original study objectives, participant population and variables. Informed consent had been obtained from all participants during the primary phase of data collection. Participant confidentiality and data anonymity were maintained throughout the research process. The study was conducted in accordance with the ICMR National Ethical Guidelines for Biomedical and Health Research Involving Human Participants and the ethical principles outlined in the Declaration of Helsinki (2013 revision).

Results

4.1 Descriptive Analysis

A total of 241 patients with CKD stage G5D were included in the study. The mean age of participants was $\mathbf{54.9 \pm 15.1}$ years (range 21-92 years), and $\mathbf{65.1}\%$ were male. The median duration of CKDwas 2 years (IQR 0.7- 4.0 years). The majority of participants were married($\mathbf{80.8}\%$), had attained secondary education or above$\mathbf{74}\%$, and $\mathbf{86}\%$ were employed.

When examining patterns in treatment modality selection, haemodialysis was the predominant choice($\mathbf{97.1}\%$), followed by kidney transplantation ($\mathbf{1.7}\%$) and peritoneal dialysis($\mathbf{1.2}\%$). Most patients belonged to the lower ($\mathbf{41}\%$) and middle ($\mathbf{23.9}\%$) socioeconomic classes, with only $\mathbf{2.6}\%$ in the upper category.

Table 1 presents overall demographic and clinical characteristics of the study participants, organized according to their selected renal replacement therapy (RRT) modality.

Table 1: Baseline Demographic and Clinical Characteristics of CKD Patients Stratified by Treatment Choice

Variable	Hemodialysis (n=234)	Peritoneal Dialysis (n=3)	Kidney Transplant (n=4)	p-value
Age (years), mean±SD	55.39 ± 14.63	45.33 ± 20.43	33.25 ± 7.50	0.006
Gender	Male 153 (65.4 %) Female 81 (34.6 %)	Male 1 (33.3 %) Female 2 (66.7 %)	Male 3 (75.0 %) Female 1 (25.0 %)	0.469
Marital status, n (%)	Married 189 (80.8%)	Married 2 (66.7%)	Married 3 (75.0%)	0.815
Educational status	Primary 9.0% Secondary 13.2% Higher secondary 41.5% Graduate 10.3% Postgraduate 20.5% Other 5.6%	Higher secondary 33.3 % Graduate 66.7%	25% Higher secondary 25% Graduate 50% Postgraduate	0.005
Employment status	Employed 86.3%	100 %	50 %	0.092

Socio-economic status	Low 41.0% Lower-middle 20.1% Middle 23.9% Upper-middle 12.4% Upper 2.6%	Lower-middle33.3% Middle66.7%	Low50% Middle25% Upper-middle25%	0.686
Duration of CKD (years), mean ± SD	3.16 ± 3.71	2.70 ± 3.75	1.20 ± 1.26	0.563

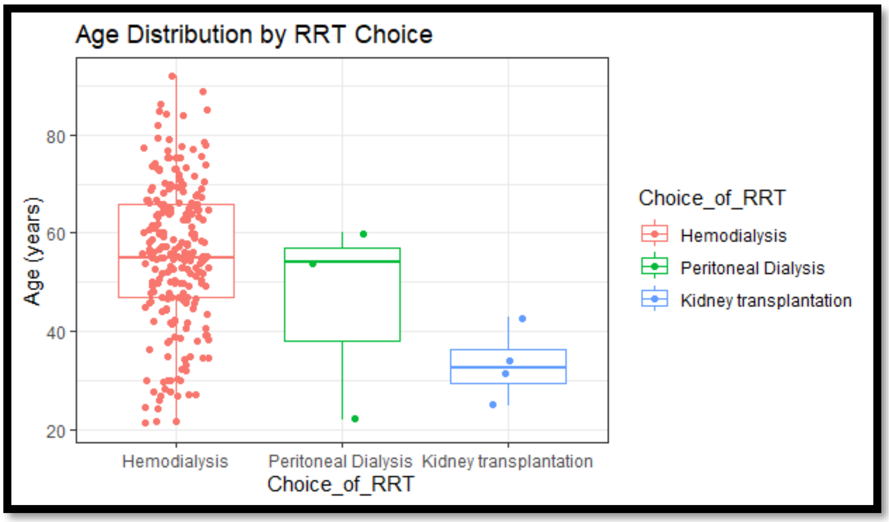
4.2 Comparative Analysis

Comparative analyses were performed to assess differences across renal replacement therapy groups. Continuous variables were analyzed using the Kruskal–Wallis rank-sum test because data were not normally distributed. Categorical variables were compared using the Chi-square test (with Fisher's exact test applied where cell counts were< 5). There was a statistically significant difference in age among treatment groups($\chi^2 = 8.22, df = 2, p = 0.016$), indicating that younger patients were more likely to undergo kidney transplantation, whereas older individuals predominantly received hemodialysis. Similarly, educational status was significantly associated with treatment choice($\chi^2 = 25.07, df = 10, p = 0.005$), with higher education levels observed among transplant and peritoneal dialysis patients.

No significant associations were found between gender($\chi^2 = 1.51, p = 0.469$), marital status($p = 0.815$), employment status($p = 0.092$), socio-economic status($p = 0.686$), or duration of CKD($\chi^2 = 1.91, p = 0.385$)and the choice of renal replacement therapy.

Figure 1. Boxplot showing age distribution across renal replacement therapy (RRT) modalities. Patients opting for kidney transplantation were notably younger, whereas those undergoing hemodialysis or peritoneal dialysis were generally older. The median age for the transplant group was markedly lower than for other modalities, underscoring that age is a key determinant of treatment suitability and choice in CKD. This pattern reflects global trends where younger, medically fit patients are more likely to pursue transplantation. These results indicate that age and level of educational attainment are the primary demographic determinants influencing the selection of renal treatment modality in CKD patients.

Figure 1. Age distribution by RRT Choice



4.3 Predictive Modelling and Machine Learning Framework

4.3.1 Model Development

To identify determinants influencing renal replacement therapy (RRT) choice among chronic kidney disease (CKD) patients, a supervised classification framework was implemented. The dependant variable examined was the selection of Renal Replacement Therapy (including hemodialysis, peritoneal dialysis, or kidney transplantation). Independent variables included Age, Gender, Level of education, Employment status, Socio-economic status, and Duration of CKD (CKDG5D), chosen based on their significance identified through descriptive and comparative analysis.

Categorical variables were encoded to factors, while continuous variables were standardized before creating the model. The dataset was divided randomly into training (70%) and validation (30%) subsets. Two predictive algorithms were tested:

1. Multinomial Logistic Regression
2. Random Forest Classifier (ensemble learning model)

Models were evaluated based on accuracy, sensitivity, specificity, Kappa coefficient, and confusion matrix performance.

4.3.2 Model Results

a. Multinomial Logistic Regression

The selection of RRT was influenced by the most significant independent variables, which were investigated using a multinomial logistic regression. This was dependent on the choice of RRT (Hemodialysis, Peritoneal Dialysis and Kidney Transplantation). Age, Gender, Educational Status, Employment Status and Socio-economic Status were the predictor/independent variables added based on clinical and statistical relevance of the disease (CKD G5D).

The model was estimated using the maximum likelihood method and expressed as:

$$
\log \frac{P(Y = K)}{P(Y = Hemodialysis)} = \beta_0^{(k)} + \beta_1^{(k)}(Age) + \beta_2^{(k)}(Gender) + \beta_3^{(k)}(Educational\ Status) + \beta_4^{(k)}(Employment\ Status) + \beta_5^{(k)}(\text{Socio} - \text{economic Status}) + \beta_1^{(k)}(\text{Duration of CKD G5D})
$$

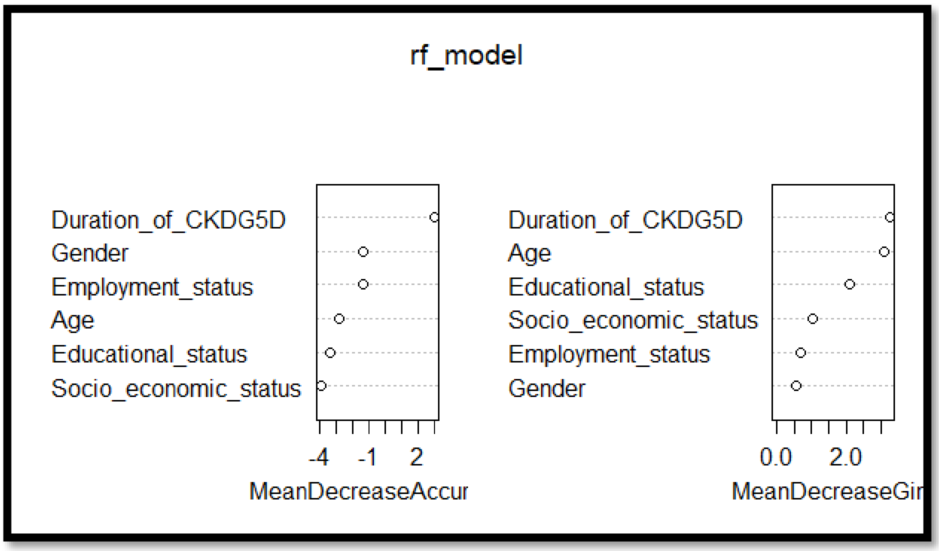
Where,
Y = RRT Choice
k =Peritoneal Dialysis or Kidney Transplantation
$X_1, X_2, \dots, X_p$ =Predictors
β_{ik}= coefficient for predictor X_i for category k
Where, Y represents the RRT modality and Hemodialysis was used as the reference category.
The multinomial model converged successfully (AIC = 71.07; residual deviance = 15.07) and achieved an overall accuracy of 94.4%(95% CI = 86.2 − 98.4%). However, the Kappa statistic (**−0.02**) indicated poor agreement beyond chance, mainly due to the models limited discrimination for peritoneal dialysis and transplant categories. The model exhibited excellent identification for hemodialysis patients (sensitivity = 0.96) but whereas transplantation and peritoneal dialysis groups showed limited models ability to accurately classify these categoriesdue to small sample sizes and class imbalance.
Among predictors, age and educational status emerged as significant contributors, suggesting that younger and more educated patients were more likely to undergo transplantation, while older individuals predominantly received hemodialysis.

b. Random Forest Classifier Model

A Random Forest classification model was also constructed to enhance prediction accuracy and capture non-linear relationships between predictors and RRT choice. The model was trained using the same independent variables with 500 decision trees (ntree = 500), employing bootstrapped sampling and Gini impurity for node splitting. Model performance was evaluated on the test dataset using a confusion matrix, overall accuracy, and Kappa statistic.

The Random Forest achieved an accuracy of 98.6% (95% CI: 92.4–99.9%), confirming strong predictive performance for the hemodialysis group, though limited classification was observed for minority classes due to dataset imbalance (hemodialysis: 97%; peritoneal dialysis: 1.2%; transplantation: 1.6%).

Figure 2. Variable importance plot of predictors influencing renal replacement therapy (RRT) choice



According to the Variable importance plot in the Random Forest model, Figure 2 demonstrates that Duration of CKD (Mean Decrease Gini = 3.27) and Age (3.11) are the most influential predictors, followed by Educational Status (2.10) and Socio-economic Status (1.02). According to the findings, patient age, disease chronicity, and socioeconomic and educational factors are important factors that impact treatment modality selection.

4.3.3 Model Interpretation and Insights

The random forest model exhibited superior predictive accuracy and interpretability compared to the multinomial regression. The findings suggest that younger age, longer disease duration, and higher education levels were associated with a greater likelihood of choosing advanced RRT modalities, such as transplantation, while lower socioeconomic and employment status favoured hemodialysis. The model underscores the multidimensional nature of renal treatment decision-making, where clinical, educational, and social variables collectively guide therapy choice. These insights can inform personalized treatment counselling and health resource planning for CKD management.

5. Discussion

This study explored demographic, socioeconomic, and clinical determinants influencing the choice of RRT among patients with chronic kidney disease stage G5D (CKD-G5D) using both classical statistical methods and machine-learning approaches. The findings indicate that age and educational status were the most significant determinants of treatment choice, whereas other sociodemographic variables such as gender, employment status, and socioeconomic status exhibited weaker significant associations.

In the present cohort, hemodialysis remained the overwhelmingly preferred modality (≈ 97%), whereas peritoneal dialysis and renal transplantation were chosen by only a small minority. Similar trends have been reported across developing regions, including studies from India, Southeast Asia, and parts of Africa, where infrastructural limitations, late referral, and financial constraints contribute to under-utilization of home-based dialysis and transplantation (Kumar et al., 2021; Abraham et al., 2020; Yeates et al., 2022). Conversely, in high-income countries, patient autonomy and health-literacy interventions have led to greater uptake of peritoneal dialysis (Davison &Jassal, 2020; Karopadi et al., 2023).

The Kruskal-Wallis analysis indicated that younger patients were significantly more likely to undergo kidney transplantation ($p = 0.006$), consistent with literature suggesting age as a decisive biological and psychosocial determinant (Tonelli et al., 2019; Wong et al., 2022). Educational attainment was also significantly associated with RRT modality ($p = 0.005$), supporting the premise that awareness, comprehension of therapy options, and ability to navigate healthcare systems strongly influence modality selection (Okpechi et al., 2021). Gender, employment, and socioeconomic status, although intuitively relevant, did not reach statistical significance, possibly reflecting homogeneity in the studied population or systemic constraints limiting free choice irrespective of background.

The machine-learning framework provided insights where the Random Forest model achieved high classification accuracy (98.6%) and highlighted Age, Educational status, and Duration of CKD-G5D as the most influential predictors, aligning closely with findings from the inferential analyses. Although the Multinomial Logistic Regression produced comparatively lower accuracy (94.3%), its parameter estimates supported the of predictor importance. These findings emphasize the potential of ensemble learning algorithms to enhance traditional regression models for clinical decision-support in nephrology (Nguyen et al., 2023; Khazaei et al., 2024). Integration of such models into patient-counselling workflows or referral pathways may facilitate early education and more individualized planning of treatment modality.

Limitations and Future Directions

Although this study applied statistical and machine learning methods, few limitations should be acknowledged. The cross-sectional design restricts temporal and causal interpretation, and the marked class imbalance with hemodialysis may have affected model calibration and reduced accuracy for transplantation and peritoneal dialysis. Also, unmeasured confounders such as physician preference, family support, health literacy, and comorbidity burden were not captured, potentially limiting the explanatory completeness. Despite these constraints, the study clearly demonstrates that younger age and higher education remain dominant determinants of RRT choice in CKD-G5D patients. The integration of machine learning frameworks offers a pragmatic, data-driven approach to personalize pre-dialysis counselling and strengthen evidence-based decision making in renal care.

Conflict of Interest: Authors asserts they have no conflicting interests. No financial or non-financial associations existed that might be considered to have affected the research presented in this manuscript.

Funding Statement: This research received no external funding.
Acknowledgements: The authors gratefully acknowledge the support of the participating academic institution and affiliated tertiary care hospitals for facilitating data access and providing the necessary infrastructure for this study. We also express our sincere appreciation to the hospital staff for their cooperation and assistance in enabling efficient data collection.

References:

1. Abraham G, et al. Choice of renal replacement therapy among patients with chronic kidney disease: perspectives from India. *Clin Kidney J.* 2020; 13(4):553–60. pubmed.ncbi.nlm.nih.gov
2. Arenas MD, et al. Social determinants of health influencing the choice of RRT in advanced CKD. *Nefrología*. 2024; 44(4):560–567.
3. Bai Q, Su C, Tang W, et al. Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports*. 2022; 12:8377.
4. Cho JH, et al. Factors Affecting Selection of a Dialysis Modality in Elderly … Front Med. *2022; 9:919028.*
5. Cortvrindt C, et al. Factors influencing renal replacement therapy modality choice from the nephrologist's perspective. *J Nephrol.* 2024; 37(3):635–645.
6. Davison SN, Jassal SV. Supportive care: integration of palliative care principles into the management of advanced chronic kidney disease. *Semin Dial.* 2020; 33(5):369–80. pubmed.ncbi.nlm.nih.gov
7. Debal DA et al. Chronic kidney disease prediction using machine learning … *J Big Data*. 2022; 9:110.
8. Delrue C, et al. Application of Machine Learning in Chronic Kidney Disease. PMC. 2024. (PMC10967786) PMC
9. Indian Council of Medical Research. *National Ethical Guidelines for Biomedical and Health Research Involving Human Participants*. New Delhi: ICMR; 2017 (updated 2023).
10. Karopadi AN, et al. Determinants of dialysis modality selection and outcomes in developing countries: a multicenter perspective. *Perit Dial Int.* 2023; 43(1):45–54. pubmed.ncbi.nlm.nih.gov

11. Khazaei M, et al. Machine learning approaches for clinical decision-making in nephrology: a systematic review. *BMC Med Inform DecisMak.* 2024; 24:29. pubmed.ncbi.nlm.nih.gov
12. Kumar SVR B, Karthikeyan B, Nair SV, Ramasamy A, Khan S, Periasamy S. A study of factors affecting dialysis recovery time in haemodialysis patients in India. *Indian J Nephrol.* 2021; 31(5):460–6. pubmed.ncbi.nlm.nih.gov
13. Morton RL, Snelling P, Webster AC, et al. Factors influencing patient choice of dialysis versus conservative care to treat end-stage kidney disease. CMAJ. 2012;184(5):E277–E283.
14. Nguyen H, et al. Machine learning for chronic kidney disease prediction and progression modeling. *Front ArtifIntell.* 2023; 6:1150092. pubmed.ncbi.nlm.nih.gov
15. Okita J, Nakata T, Uchida H, et al. Development and validation of a machine learning model to predict time to renal replacement therapy in patients with chronic kidney disease. *BMC* Nephrol. 2024; 25:101.
16. Okpechi IG, et al. Global variation in access to renal replacement therapy and associated outcomes. *Kidney Int Rep.* 2021; 6(11):2815–26. pubmed.ncbi.nlm.nih.gov
17. Tonelli M, et al. Comparison of dialysis modalities and mortality outcomes in adults with end-stage kidney disease. JAMA Netw Open. 2019; 2(5):e193013. pubmed.ncbi.nlm.nih.gov
18. Tsai MH, et al. An integrated machine learning predictive scheme … *Frontiers in Medicine*. 2023.
19. Wong G, et al. Decision-making and care transitions for patients with advanced kidney disease: a global review. Am J Kidney Dis. 2022; 79(3):384–95. pubmed.ncbi.nlm.nih.gov
20. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA. 2013; 310(20):2191–2194.
21. Yeates K, et al. Global kidney health atlas: structures, organization, and services for kidney care across countries and regions. Kidney Int Rep. 2022; 7(5):1204–15. pubmed.ncbi.nlm.nih.gov.