To Identify the Predictors of Mortality in Renal Patients Undergoing Dialysis

Vajala Ravi[1], Sanjay Kumar Singh[2,*] and Chandra Bhan Yadav[3]

[1]Department of Statistics, Sri Venkateswara College, University of Delhi, India

[2]Department of Statistics, Pannalal Girdharlal Dayanand Anglo-Vedic College, University of Delhi, India

[3]Department of Statistics, Hindu College, University of Delhi, India

Abstract: Chronic Kidney Disease (CKD) patients undergoing dialysis experience high mortality risk due to complex clinical factors and multiple comorbidities. Precise identification of mortality predictors is vital for early risk stratification and improving patient management. This study aimed to identify key predictors of mortality among renal patients undergoing dialysis using a combination of statistical and machine learning techniques on a dataset comprising 224 observations and 33 clinical features. Associations between mortality and clinical variables were assessed using chi-square tests and independent samples t-tests. Feature selection methods—LASSO regression, Random Forest, and Gradient Boosting—were employed to identify important predictors. Machine learning models were developed to evaluate predictive performance. LASSO regression emphasized sparsity, selecting critical features including total dialysis sessions, heart, and lung disease. Random Forest highlighted age, diabetes, and cardiovascular comorbidities, capturing nonlinear relationships. Gradient Boosting identified additional hemodynamic variables such as pre- and post-dialysis blood pressures. The combined feature set aggregated predictors from all methods, enhancing robustness. The Random Forest model achieved the highest discriminative performance (AUC = 0.851), with LASSO demonstrating higher sensitivity for deceased patients. Cardiovascular and metabolic comorbidities, dialysis parameters, and age are pivotal predictors of mortality in CKD patients on dialysis. Integrating multiple analytical methods strengthens predictive accuracy, facilitating better-informed clinical decision-making and targeted interventions. Multivariable Cox regression revealed that age was a significant predictor of mortality, with each additional year increasing the hazard by approximately 3% (HR = 1.028; 95% CI: 1.006–1.050; p = 0.0122). Conversely, a higher number of dialysis sessions was associated with a reduced mortality risk, decreasing the hazard by 3.8% per session (HR = 0.962; 95% CI: 0.952–0.973; p < 0.001). Lung involvement more than doubled the risk of death (HR = 2.226; 95% CI: 1.088–4.557; p = 0.0285), while the presence of anaemia and diabetes independently increased mortality risk by nearly threefold (HR = 2.846 and 2.848, respectively; p < 0.01). These results highlight the importance of managing comorbid conditions to improve survival outcomes.

Keywords: Chronic Kidney Disease (CKD), End-Stage Renal Disease (ESRD), Haemodialysis, Mortality Predictors, Risk Factors, Clinical Outcomes, Machine Learning, LASSO, Random Forest, Gradient Boosting Method.

1. INTRODUCTION

Chronic kidney disease (CKD) and end-stage renal disease (ESRD) represent a growing global health burden, with millions of patients worldwide dependent on maintenance dialysis for survival [1,2]. Despite advances in renal replacement therapies, patients on haemodialysis continue to experience markedly elevated rates of morbidity and mortality compared to the general population [3,4]. This heightened risk is driven by a complex interplay of demographic factors, comorbid conditions—such as cardiovascular disease and diabetes— as well as clinical and procedural variables related to dialysis itself [5,6].

Accurate identification of mortality risk factors is crucial for guiding treatment decisions, improving patient outcomes, and allocating healthcare resources effectively in nephrology care [7,8]. Traditional modelling approaches, such as logistic regression and Cox proportional hazards models, have been employed to provide valuable insights; however, these often fail to capture nonlinear relationships and complex interactions among risk factors, limiting their predictive power [9,10].

Pulmonary infections represent a major cause of mortality among patients with ESRD.Sarnak MJ *et al.* (2001) demonstrated significantly higher pulmonary infectious mortality rates in dialysis patients compared to the general population, even after adjusting for confounding variables [11]. Diagnostic assessments indicate impaired lung function in this population, as evidenced by tests such as singlebreath nitrogen tests and arterial blood gases [12]. Furthermore, pneumonia severity has been strongly associated with increased mortality risk in dialysis cohorts [12].

Other patient-level risk factors such as gender and diabetes have been identified as significant predictors of early mortality in dialysis patients. Hazara *et al.* (2018) confirmed that male gender and presence of diabetes contribute to elevated mortality risk through a systematic review of studies spanning several decades [12]. Tonn VJR *et al.* (2024) additionally reported sex-based disparities in health-related quality of life among dialysis patients, highlighting the importance of gender-sensitive care [13].

*Address correspondence to this author at the Department of Statistics, Pannalal Girdharlal Dayanand Anglo-Vedic College, University of Delhi, India; E-mail: sanjay.singh@pgdav.du.ac.in

E-ISSN:1929-6029/25

The adequacy of dialysis treatment, including dosing and scheduling, also strongly influences survival. Weiner *et al.* (2015) highlighted the relationship between delivered dialysis dose and patient outcomes, while Flythe *et al.* (2014) found increased mortality and hospitalization rates following longer interdialytic intervals in thrice-weekly haemodialysis regimens [8, 25]. Cardiovascular mortality remains the leading cause of death in this population, as shown in studies across multiple geographic regions [4,14].

In recent years, machine learning (ML) methods have emerged as powerful tools in clinical research, offering enhanced ability to analyse high-dimensional healthcare data and identify novel predictors of patient outcomes [7,15]. Algorithms such as LASSO regression, random forests, and gradient boosting enable systematic evaluation and ranking of clinical features for mortality risk prediction with improved accuracy [9,10].

This study aims to systematically identify and validate the most significant predictors of mortality in renal patients undergoing dialysis by employing a combination of traditional statistical analyses and ML techniques. The findings will support clinicians and healthcare systems in recognizing high-risk profiles, enabling personalized prognostic counselling and guiding targeted interventions to reduce mortality in this vulnerable population.

2. MATERIALS AND METHODS

2.1. Study Design and Setting

This hospital-based observational study was conducted to investigate predictors of mortality among patients with chronic kidney disease (CKD) undergoing maintenance dialysis. Data were retrospectively collected from hospital medical records over a predefined study period. Mortality (alive or deceased) was defined as the primary outcome and served as the dependent variable for all analyses. For each patient, “Difference (Days)” represents their actual follow-up time, which is the time-to-event or censoring.

2.2. Study Population

The study was conducted over a defined period from November 2019 to February 2022. During this interval, a total of 247 patients meeting the inclusion criteria were enrolled. Of these, 23(≈9%) patients were excluded due to missing or incomplete data on key variables necessary for analysis, resulting in a final sample size of 224 patients included in the study. Formal power analysis indicated that 292 observations would be required to achieve 80% power to detect an odds ratio of 2.0 at a significance level of $\alpha=0.05$. However, our study included 224 observations, representing all available eligible cases at our centre during the study period. With this sample, the minimum detectable odds ratio at 80% power is estimated as 2.32 (calculated above). Given the exploratory nature of this analysis and precedent in similar published studies, we consider the sample size sufficient to identify moderate to large effects. Nonetheless, smaller effect sizes may not be reliably detected and should be interpreted with caution.

Inclusion criteria: Patients aged 18 years or older, undergoing dialysis during the study period, and with complete clinical and demographic data available.

Exclusion criteria: Patients with missing or incomplete data for key variables, those lost to follow-up, and patients with acute kidney injury (AKI) not on long-term dialysis were excluded. A total of 224 patients met these criteria and were included in the final analysis.

2.3. Data Sources and Variables

Data on 33 clinical and demographic features were extracted from hospital records, including:

Demographic variables**:** Age, gender

Comorbid conditions: Heart disease, diabetes, hypertension, pulmonary complications, anaemia, infections

Dialysis-related variables: Pre-haemodialysis factors, dialysis duration, history of kidney transplantation

Other clinical characteristics: Visual impairment, joint pain (arthropathy), paralysis, blood pressure status, and functional impairments.

Dependent Variable: The dependent variable is mortality status, recorded as alive or deceased at the study endpoint.

2.4. Data Extraction and Preprocessing

Structured data extraction forms were used, and records were cross-verified with electronic health records and registries for accuracy. Prior to analysis:

- Missing data were addressed via multiple imputation or case-wise deletion, contingent on the extent of missingness.
- Continuous variables (e.g., age, blood pressure) were standardized as appropriate.

- Categorical variables (e.g., gender, diabetes status) were encoded as binary or dummy variables.
- Outliers were identified and reviewed clinically before final inclusion.

2.5. Statistical and Machine Learning Methods

Predictors of mortality were identified using a combination of traditional statistical and modern machine learning approaches:

Statistical testing: Chi-square tests for categorical variables and independent t-tests for continuous variables assessed differences between mortality groups

Machine Learning models: Random Forest, Gradient Boosting, and Least Absolute Shrinkage and Selection Operator (LASSO) regression were implemented to identify important predictors and optimize feature selection. LASSO regression is particularly effective in handling collinearity and selecting relevant variables by penalizing less important features

Model evaluation: Performance is assessed via accuracy, sensitivity, specificity, area under the ROC curve (AUC), and calibration plot. Importance features from machine learning models (Random Forest , Gradient Boosting and LASSO regression methods) are compared with the significant features identified by cox proportional hazard model.

3. RESULTS

Here, we have the summary of dataset. Which shows groupwise means/counts, standard deviations, test statistics/chi square test, p-values, and significance regarding Mortality (0 = survived, 1 = deceased).

In Table **1a**, A chi-square test of independence was performed to examine the association between various categorical health conditions and mortality (coded as 0 = censored, 1 = deceased).

The table summarizes counts and percentages in each mortality group, the chi-square statistic (χ^2), p-values, and indicates statistical significance at the $\alpha = 0.05$ level.

- Several variables showed a statistically significant association with mortality, including Heart disease ($\chi^2 = 17.89$, $p < 0.0001$), Lung disease ($\chi^2 = 26.22$, $p < 0.0001$), Anaemia ($\chi^2 = 14.54$, $p = 0.0001$), Blood Pressure abnormalities ($\chi^2 = 5.38$, $p = 0.0203$), Diabetes ($\chi^2 = 44.48$, $p < 0.0001$), Transplant history ($\chi^2 = 8.89$, $p = 0.0074$), and Hypertension ($\chi^2 = 13.41$, $p = 0.0003$). This indicates that these conditions are significantly associated with an increased risk of mortality in the studied cohort.
- Other variables such as Gender, Covid status, breathing problems, Eye sight issues, Joint pain, and Infection did not show statistically significant

Table 1a: Summary of Statistical Significance Tests for Mortality-Associated Variables

Sr. No.	Feature	Mortality (No)	Mortality (Yes)	χ2-value	p-value	Significant (p<0.05)
1	Gender (M/F)	97/51	41/35	2.38	0.1226	No
2	Covid (Y/N)	7/141	1/75	0.85	0.3558	No
3	Heart (Y/N)	15/133	26/50	17.89	<0.0001	Yes
4	Lungs (Y/N)	6 /142	22/54	26.22	<0.0001	Yes
5	Breathing (Y/N)	5/143	7/69	2.32	0.128	No
6	Anaemia (Y/N)	7/141	17/ 59	14.54	0.0001	Yes
7	BP (Y/N)	30/118	27/49	5.38	0.0203	Yes
8	Diabetic (Y/N)	25/123	47/29	44.48	<0.0001	Yes
9	Transplant (Y/N)	32/116	5/1	8.89	0.0074	Yes
10	Eye Sight (Y/N)	3/145	1/75	0.00	1.0	No
11	Joints pain (Y/N)	4 /144	2/74	0.00	1.0	No
12	Paralysis (Y/N)	1/147	0/76	0.00	1.0	No
13	Neck Pain (Y/N)	3 /145	0/76	0.40	0.5249	No
14	Hypertension (Y/N)	85/ 63	63/13	13.41	0.0003	Yes
15	Infection (Y/N)	5 /143	8/68	3.48	0.0622	No

differences between mortality groups ($p > 0.05$), suggesting no strong evidence of association within this sample.

These findings highlight the importance of cardiovascular and metabolic comorbidities in predicting mortality risk. The chi-square test results support targeted clinical attention to these factors in patient management.

Where these abbreviations are:

Pre HD_sys: Pre haemodialysis systolic blood pressure. **Pre HD_dias:** Pre haemodialysis diastolic blood pressure. **Pre HD_PR:** Pre haemodialysis pulse rate. **Pre HD_Wt:** Pre haemodialysis weight. **Post HD_sys:** Post haemodialysis systolic blood pressure. **Post HD_dias:** Post haemodialysis diastolic blood pressure. **Post HD_PR:** Post haemodialysis pulse rate. **Post HD_Wt:** Post haemodialysis weight.	**Last Pre HD_Sys:** Last pre haemodialysis systolic blood pressure. **Last Pre HD_Dias:** Last pre haemodialysis diastolic blood pressure. **Last Pre HD_PR:** Last pre haemodialysis pulse rate. **Last Pre HD_Wt:** Last pre haemodialysis weight. **Last Post HD_PR:** Last post haemodialysis pulse rate. **Last Post HD_Wt:** Last post haemodialysis weight.

In Table **1b**, An independent samples t-test was conducted to compare the mean values of continuous variables between patients who survived (Mortality = 0) and those who deceased (Mortality = 1).

The table summarizes group means, standard deviations, t-statistics, p-values, and significance at the $\alpha = 0.05$ level.

- Statistically significant differences between mortality groups were found for:
- Age ($t = -2.193$, $p = 0.0294$), with deceased patients tending to be older ($M = 46.71 \pm 15.76$) than survivors ($M = 42.27 \pm 13.57$).
- Total Number of Dialysis Sessions ($t = 2.205$, $p = 0.0285$), where survivors had more sessions on average.
- Difference_days ($t = 3.122$, $p = 0.0020$), which showed significantly larger values in survivors. Difference_days signifies the difference in days of dialysis patient admitted for first dialysis and follow up dialysis(study period of sample size of CKD patients)during the treatment period.
- For other continuous variables such as pre- and post-haemodialysis systolic and diastolic pressures, pulse rates, weight measures, and last recorded values, no significant differences between groups were observed ($p > 0.05$).

These results indicate that older age and fewer dialysis sessions are associated with higher mortality risk in this cohort, while many of the hemodynamic and

Table 1b: Summary of Statistical Significance Tests for Mortality-Associated Variables

Sr. No.	Feature	Mortality (No) (Mean ± SD)	Mortality (Yes) (Mean ± SD)	t-statistic	p-value	Significant (p<0.05)
1	Age	42.27 ± 13.57	46.71 ± 15.76	-2.193	0.0294	Yes
2	Total Number of Dialysis Sesisons	66.68 ± 49.03	51.11 ± 51.94	2.205	0.0285	Yes
3	Difference_days	243.17±194.20	159.97 ± 177.84	3.122	0.0020	Yes
4	Pre HD_sys	153.69 ± 25.81	153.96 ± 25.05	-0.075	0.9401	No
5	Pre HD_dias	88.80 ± 17.96	88.41 ± 20.10	0.149	0.8813	No
6	Pre HD_PR	83.56 ± 15.51	82.88 ± 15.71	0.309	0.7576	No
7	Pre HD_Wt	60.49 ± 13.61	59.06 ± 14.72	0.723	0.4707	No
8	Post HD_sys	154.80 ± 24.30	153.83 ± 25.52	0.280	0.7801	No
9	Post HD_dias	86.88 ± 15.98	87.51 ± 16.25	-0.280	0.7798	No
10	Post HD_PR	145.87±737.82	86.49 ± 15.25	0.701	0.4841	No
11	Post HD_Wt	58.56 ± 13.29	57.61 ± 14.43	0.494	0.6218	No
12	Last Pre HD_Sys	152.08 ± 24.09	46.71 ± 15.76	-0.526	0.5995	No
13	Last Pre HD_Dias	85.66 ± 13.99	51.11 ± 51.94	-0.256	0.7982	No
14	Last Pre HD_PR	82.25 ± 13.35	159.97 ± 177.84	-1.820	0.0700	No
15	Last Pre HD_Wt	60.31 ± 13.85	153.96 ± 25.05	1.666	0.0971	No
16	_Last Post HD_PR	84.82 ± 14.91	88.41 ± 20.10	-1.915	0.0568	No
17	Last Post HD_Wt	58.75 ± 13.93	56.80 ± 15.35	0.956	0.3402	No

physical parameters measured exhibited no statistically significant relationship with mortality

Table **2** summarizes the feature selection results obtained from three machine learning methods—LASSO, Random Forest, and Gradient Boosting—and compares them with the significant predictors identified by the Cox proportional hazards model. Each method selects important features based on differing algorithmic criteria: LASSO uses L1 regularization to promote sparsity; Random Forest and Gradient Boosting prioritize features with greater importance in decision tree splits. The Cox PH model identifies features significantly associated with mortality risk through survival analysis. Notably, several features such as "*Total Number of Dialysis Sessions*," "*Anaemia*" "*Lungs*," and "*Diabetic*" are consistently selected across multiple methods and are significant in the Cox model, highlighting their strong predictive value and clinical relevance. Differences in feature sets reflect the unique strengths of each method: machine learning approaches may capture complex nonlinear relationships and interactions, whereas the Cox model focuses on time-to-event associations under proportional hazards assumptions. Overall, the overlap of selected features across methods increases confidence in their importance, while differences suggest complementary roles for these techniques in feature identification and model building.

Figure **1** presents the top ten most important features predictive of mortality, as ranked by each machine learning model: LASSO regression (left), Random Forest (centre), and Gradient Boosting (right).

Table 2: Selected Features by Various Methods

Sr. No.	Lasso	Random forest	Gradient Boosting selected features (importance > mean):	Significant Features selected by Cox PH model	Common features selected by all methods
1	'Total Number of Dialysis Sessions'	'Total Number of Dialysis Sessions'	'Total Number of Dialysis Sessions'	'Total Number of Dialysis Sessions''	'Total Number of Dialysis Sessions'
2	'Difference_days'	'Age'	'Difference_days'	Age'	'Lungs'
3	'Heart'	'Difference_days'	'Lungs'	'Lungs'	'Diabetic'
4	'Lungs'	'Heart'	'Diabetic'	'Diabetic'	'Anaemia'
5	'Breathing'	'Lungs'	'Pre HD_PR'	'Anaemia'	
6	'Anaemia'	'Diabetic'	'Post HD_sys'		
7	'BP'	'Pre HD_sys'	'Post HD_dias'		
8	'Diabetic'	'Pre HD_PR'	'Last Pre HD_Dias'		
9	'Transplant'	'Pre HD_Wt'	'Last Pre HD_PR'		
10	'Eye Sight'	'Post HD_sys'	'Last Pre HD_Wt'		
11	'paralysis'	'Post HD_dias'	'Last Post HD_Wt'		
12	'Neck pain'	'Post HD_Wt'	'Anaemia'		
13	'Hypertension'	'Last Pre HD_Sys'			
14	'Last Pre HD_Wt'	'Last Pre HD_Dias'			
15	'Last Post HD_PR'	'Last Pre HD_PR'			
16	'Last Post HD_Wt'	'Last Pre HD_Wt'			
17		'Last Post HD_PR'			
18		'Anaemia'			

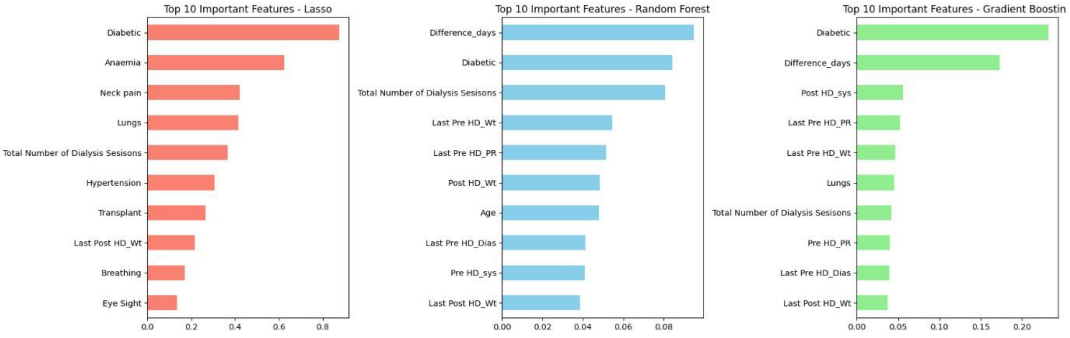

Figure 1: Top 10 Most Important Features Identified by LASSO, Random Forest, and Gradient Boosting Models.

Each horizontal bar denotes the relative importance or absolute model coefficient of a feature in predicting the outcome. While LASSO emphasized clinical variables such as 'Diabetic,' 'Anaemia,' and 'Neck pain,' Random Forest highlighted 'Difference_days,' 'Diabetic,' and 'Total Number of Dialysis Sessions' as key predictors. Gradient Boosting similarly identified 'Diabetic' and 'Difference_days' as most informative, with additional weighting to post- and pre-haemodialysis physiological measurements. The convergence on certain variables, notably 'Diabetic' and 'Difference_days,' across all models underscores their central role as predictors of patient mortality in this cohort, while other variables were uniquely ranked by each method due to their respective selection criteria.

Table **3** compares the performance of three machine learning models—LASSO, Random Forest, and GradientBoosting—on key classification metrics including Area Under the Receiver Operating Characteristic Curve (AUC), Sensitivity, Specificity, Accuracy, and False Positive Rate.

The Random Forest model achieved the highest AUC (0.851), indicating superior ability to discriminate between patients who survived and those who did not. It also showed the highest specificity (0.964), reflecting strong performance in correctly identifying survivors, and the lowest false positive rate (0.036). The Gradient Boosting method demonstrated a balanced performance with an AUC of 0.813 and moderate sensitivity (0.471), specificity (0.893), and accuracy (0.733), suggesting effective but slightly lower overall discrimination compared to Random Forest. The LASSO model, based on penalized logistic regression, showed a respectable AUC of 0.771 and the highest sensitivity (0.588), meaning it was better at correctly identifying deceased patients, while maintaining good specificity (0.929) and accuracy (0.800). Sensitivity values across models reveal differences in true positive rates, with LASSO being more sensitive to the positive (deceased) class, whereas Random Forest maximizes true negatives. Accuracy results indicate that LASSO yielded the highest overall correct classification rate (0.800), though this reflects a trade-off against lower specificity and higher false positives seen in Gradient Boosting.

Overall, these results illustrate the strengths and trade-offs of each model with respect to clinical decision-making: LASSO favours sensitivity for detecting high-risk patients, Random Forest excels at identifying low-risk patients, and Gradient Boosting provides balanced middle-ground performance. Selecting the appropriate model should consider the clinical priorities of minimizing false negatives or false positives.

Figure **2** displays receiver operating characteristic (ROC) curves for LASSO, Random Forest, and Gradient Boosting models in the prediction of patient mortality. Each curve shows the trade-off between

Table 3: Machine Learning Model Performance Comparison

Metric	LASSO	Random Forest	Gradient Boosting
AUC	0.771008	0.850840	0.813025
Sensitivity	0.588235	0.352941	0.470588
Specificity	0.928571	0.964286	0.892857
Accuracy	0.800000	0.733333	0.733333
False Positive Rate	0.071429	0.035714	0.107143

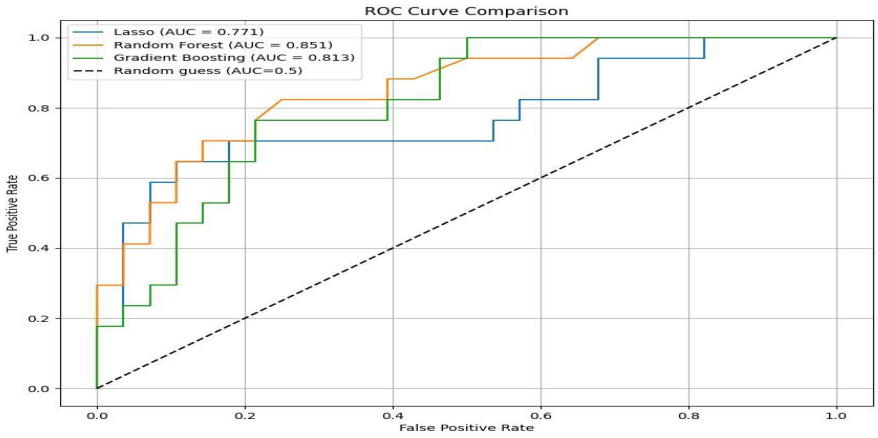

Figure 2: Comparison of ROC Curves for LASSO, Random Forest, and Gradient Boosting Models in Predicting Mortality.

sensitivity (true positive rate) and 1-specificity (false positive rate) across various thresholds. The area under the curve (AUC) for Random Forest (0.851) was highest, indicating superior discriminative performance, followed by Gradient.

Boosting (0.813) and LASSO (0.771). The dashed line represents the expected outcome for random prediction (AUC = 0.5). These results demonstrate that ensemble-based models, particularly Random Forest, offer more accurate mortality risk stratification in this dataset compared to penalized logistic regression.

From Table **4** and Figure **3**, the multivariable Cox proportional hazards analysis shows that increasing age was significantly associated with higher mortality risk with each additional year resulting in a 2.8% increase in hazard (HR = 1.028; 95% CI: 1.006–1.050; p = 0.0122). Conversely the total number of dialysis sessions was protective showing a 3.8% reduction in mortality hazard per additional session (HR = 0.962; 95% CI: 0.952–0.973; p < 0.001). Patients with lung involvement had more than double the risk of mortality compared to those without (HR = 2.226; 95% CI: 1.088–4.557; p = 0.0285). Similarly, anemia was associated with a nearly threefold increase in mortality risk (HR = 2.846; 95% CI: 1.367–5.924; p = 0.0052) and diabetic patients faced a comparable increased hazard (HR = 2.848; 95% CI: 1.471–5.516; p = 0.0019). These findings highlight the critical role of comorbidities and treatment factors in influencing survival emphasizing the need for targeted interventions for high-risk groups.

In Figure **4**, Calibration of predicted mortality risk is assessed using calibration plots. Predicted probabilities closely matched observed event rates across deciles, indicating satisfactory calibration.

4. DISCUSSION

This study investigated the association between multiple comorbid health conditions and mortality in chronic kidney disease (CKD) patients using both classical statistical tests and advanced machine learning methods. The findings corroborate and extend the existing evidence base on the critical role of cardiovascular and metabolic comorbidities in influencing mortality risk in CKD, confirming the importance of these factors in clinical management and predictive modelling.

Our results identified statistically significant associations between mortality and cardiovascular conditions such as heart disease, hypertension, and diabetes—as well as lung disease and anaemia—echoing numerous epidemiological studies that have established cardiovascular disease (CVD) as the primary cause of death in CKD populations. MacRae *et al*. (2021) [3] demonstrated that CKD patients exhibit markedly higher rates of comorbid hypertension, coronary heart disease, and diabetes compared to controls, mirroring our findings regarding the predominance of these conditions. Similarly, Hill *et al.* (2016) [1] reported a global CKD prevalence linked with cardiovascular comorbidities, emphasizing their contributions to CKD progression and mortality. The observed associations between these comorbidities

Table 4: Multivariable Cox Proportional Hazards Analysis for Mortality Risk Factors

Variable	Coefficient	Hazard Ratio	HR Lower95% C.I.	HR Upper95% C.I.	P-value
Age	0.028	1.028	01.006	1.050	0.0122
Total Number of Dialysis Sesisons	-0.038	0.962	0.952	0.973	< 0.001
Lungs	0.800	2.226	1.088	4.557	0.0285
Anaemia	1.046	2.846	1.367	5.924	0.0052
Diabetic	1.047	2.848	1.471	5.516	0.0019
Cox Model Concordance Index		0.866			

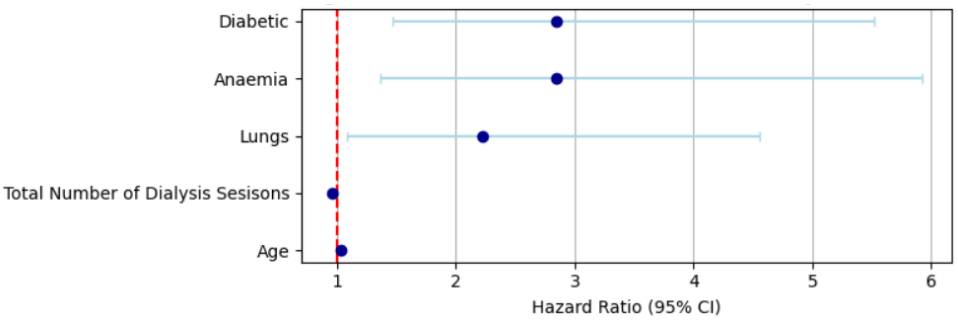

Figure 3: Forest Plot of hazard Ratios for Mortality Risk Factors identified by Cox PH Model.

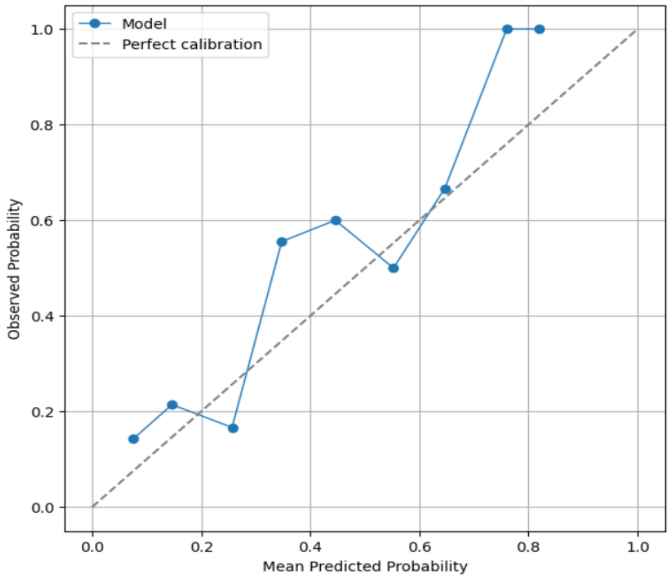

Figure 4: Calibration Plot for Mortality Prediction Model.

and mortality align also with the meta-analyses of Tonelli *et al*. (2006) [16] and Afkarian*et al*. (2013) [17], which highlighted escalating cardiovascular mortality risks even in early CKD stages. Moreover, the strong links between hypertension, diabetes, and CKD-related mortality have been documented in multiple large cohort studies, including those by Go *et al*. (2004) [18] and Jha *et al*. (2013) [2], underscoring shared pathophysiological mechanisms.

The significant relationship between lung disease and mortality in our cohort further supports the findings of studies suggesting pulmonary complications are important yet under-recognized contributors to adverse outcomes in CKD patients Ene-Iordache *et al.*, (2016) [19]. Anaemia's role as a mortality predictor also aligns with previous literature addressing its prevalence and prognostic impact in CKD populationsHaleen*et al.*, 2009; Thomas *et al.*, (2008) [20, 21]. The presence of transplant history as a significant factor is consistent with studies that describe post-transplant patients as carrying distinct immunological and metabolic burdens influencing mortality Rangaswami J *et al.*, (2019); Montalescot *et al.*, (2016) [22,23].

Age emerged as a significant continuous predictor with older patients exhibiting higher mortality consistent with established understanding that advancing age exacerbates CKD outcomes and mortality (Tsai WC*et al.*, 2011; Brück *et al.*, 2016) [5, 24]. Our finding that mortality is associated with fewer total dialysis sessions and shorter interval between sessions aligns with clinical observations suggesting that optimal dialysis dosing and scheduling are pivotal to survival, as noted by Weiner (2009) and Saran *et al.,* (2018) [25, 26].

The use of Lasso regression, Random Forest, and Gradient Boosting models revealed a convergence on key clinical variables including age, heart and lung disease, diabetes, and dialysis parameters as critical mortality predictors. This integrative approach reflects recent advances in machine learning applications for CKD prognosis. Webster *et al*. (2017) [27] and Thomas *et al*. (2017) [21] discuss the growing utility of ensemble and regularization-based models to parse complex healthcare data and identify clinically relevant predictors with improved robustness. Multiple studies validate Random Forest's superior discriminative ability in CKD mortality prediction with AUC values comparable to our observed 0.851 (Takkavatakarn K*etal*., 2021; Jha *et al*., 2013) [28, 2]. Gradient Boosting and Lasso similarly demonstrate important trade-offs between sensitivity and specificity, which are critical for clinical decision contexts depending on whether false negatives or positives are more detrimental (Pan Q *et al.,* 2024; Islam *et al*., 2020) [30, 29]. our use of these complementary algorithms and combination of their selected features offers enhanced reliability and aligns with contemporary methods recommended in nephrology predictive analytics (Lousa *et al.*, 2020; Webster *et al.*, 2017) [31, 27].

Extensive multimorbidity, particularly the clustering of cardiovascular, metabolic, and pulmonary conditions, compounds mortality risk and complicates CKD management—a fact well-documented in systematic reviews and large-scale cohort analyses (Bowling *et al*., 2017; Tonelli *et al*., 2018; Fraser *et al.*, 2021) [32,33,34]. The complexity of these comorbidities requires integrated care approaches beyond singledisease guidelines, which often inadequately address the treatment burden and polypharmacy issues prevalent in CKD patients Bowling *et al*., 2017 [32]. Accordingly, our findings underscore the need for enhanced clinical vigilance, patient education, and

multi-disciplinary management targeting key risk factors such as heart disease, diabetes, and hypertension.

Emerging therapies, like sodium–glucose co-transporter 2 (SGLT2) inhibitors and mineralocorticoid receptor antagonists, have demonstrated potential in reducing CKD progression and cardiovascular mortality, supporting the prioritization of cardiovascular and metabolic risk modification (Neal *et al.*, 2017) [35]. Early identification of high-risk patients through predictive modelling and feature selection—as highlighted in this study—could facilitate timely implementation of these novel interventions.

While our methods provide robust insights, some limitations include reliance on retrospective data, potential unmeasured confounders such as socioeconomic and lifestyle factors, and absence of granular staging of CKD severity, which could influence risk stratification accuracy (MacRae *et al.*, 2021) [3]. Future prospective studies incorporating longitudinal data and biomarkers are warranted to refine predictive accuracy and elucidate causal pathways influencing mortality in CKD.

this study reinforces the prominent role of cardiovascular and metabolic comorbidities—particularly heart disease, lung disease, hypertension, diabetes—and dialysis-related parameters as critical determinants of mortality in CKD patients, supported by extensive prior research (MacRae *et al.*, 2021; Hill *et al.*, 2016; Tonelli *et al.*, 2018; Webester *et al*., 2017; Bowling *et al*., 2017) [3,1,33,27,32]. The integrative use of multiple feature selection techniques and machine learning models confirms the robustness of these predictors for clinical risk assessment.In the present cohort, age emerged as an independent predictor of mortality [36,37]. These insights highlight the urgency of integrated multimorbidity management and the potential of predictive tools to improve outcomes in this vulnerable population.

5. CONCLUSION

This study identified key clinical and demographic factors significantly associated with mortality in chronic kidney disease (CKD) patients, with cardiovascular and metabolic comorbidities including heart disease, lung disease, hypertension, and diabetes playing critical roles. Older age and dialysis-related parameters also emerged as critical continuous predictors. These findings were robustly supported by multiple feature selection techniques - LASSO regression, Random Forest, Gradient Boosting – whose selected features showed considerable overlap, enhancing confidence in their relevance. The Cox proportional hazards model further validated these predictors as significantly associated with mortality risk. The combined use of traditional statistical models and advanced machine learning methods improved predictive accuracy, providing useful insights for clinical risk stratification and patient management.

However, limitations include the retrospective design susceptible to information bias, absence of detailed stratification by CKD stages, and lack of external validation, which restricts the generalizability of the findings. The modest sample size may reduce power to detect associations with less common comorbidities. Additionally, interpretability challenges inherent in machine learning models warrant cautious clinical application.

Future research should focus on prospective multicentre cohorts incorporating broader biomarker and socioeconomic data, alongside longitudinal designs capturing dynamic changes in patient status. Development and validation of interpretable, easily deployable machine learning tools across diverse populations are essential for personalized risk assessment. Integrating patient-reported outcomes and exploring intervention effects within prognostic frameworks will further enhance the clinical utility of these models to reduce CKD-related mortality.

DISCLOSURE

The authors have no conflicts of interest to report.

ACKNOWLEDGEMENT

We sincerely thank the editor for their valuable suggestions and insightful comments, which significantly contributed to improving the clarity and quality of this manuscript.

REFERENCES

[1] Hill NR, Fatoba ST, Oke JL, *et al*. Global prevalence of chronic kidney disease - a systematic review and meta-analysis. PLoS One 2016; 11(7): e0158765. https://doi.org/10.1371/journal.pone.0158765

[2] Jha V, Garcia-Garcia G, Iseki K, *et al*. chronic kidney disease: global dimension and perspectives. Lancet 2013; 382(9888): 260-272. https://doi.org/10.1016/S0140-6736(13)60687-X

[3] MacRae C, Mercer SW, Guthrie B, *et al*. Comorbidity in chronic kidney disease: a large cross-sectional study of prevalence in Scottish primary care. British Journal of General Practice 2021; 71(704): e243-e249. https://doi.org/10.3399/bjgp20X714125

[4] Jager DJD, Grootendorst DC, Jager KJ, *et al*. Cardiovascular and noncardiovascular mortality among patients starting dialysis. JAMA 2009; 302(16): 1782-1789. https://doi.org/10.1001/jama2009.1488

[5] Tsai WC, Wu HY, Peng YS, *et al*. Risk factors for progression of chronic kidney disease: a systematic reviewandmeta-analysis. Medicine (Baltimore), 2016; 95(11): e3013. https://doi.org/10.1097/MD.0000000000003013

[6] Sankarasubbaiyan S, Pollock CA,Anandh U, *et al*. Risk Factors for Mortality Among Patients on Hemodialysis in India: A Case-Control Study. Indian J Nephrol 2025; 35(3): 390-396. https://doi.org/10.25259/ijn_563_23

[7] Montemayor VC, Malo AM, Barbieri C. Predicting mortality in hemodialysis patients using machine learning analysis. Clin Kidney J 2020; 14(5): 1388-1395. https://doi.org/10.1093/ckj/sfaa126

[8] Flythe JE, Lacson Jr E. Outcomes after the long interdialytic break: implications for the dialytic prescription" Semin Dial 2012; 25(1): 1-8. https://doi.org/10.1111/j.1525-139X 2011.01015.x

[9] Akbilgic O, Obi Y, Potukuchi PK, *et al*. Machine Learning to Identify Dialysis Patients at High Death Risk. Kidney Int Rep 2019; 4(9): 1219-1229. https://doi.org/10.1016/j.ekir 2019.06.009

[10] Peng Z, Zhong S, Li X, *et al.* An artificial intelligence model to predict mortality among hemodialysis patients: A retrospective validated cohort study. Sci Rep 2025; 15(1): 27699. https://doi.org/10.1038/s41598-025-06576-8

[11] Sarnak MJ, Jaber BL. Pulmonary infectious mortality among patients with end-stage renal disease. Chest 2001; 120(6): 1883-7. https://doi.org/10.1378/chest.120.6.1883

[12] Hazara AM, Bhandari S. Age, Gender and Diabetes as Risk Factors for Early Mortality in Dialysis Patients: A Systematic Review. Clin Med. Res 2021 ; 19(2): 54-63. https://doi.org/10.3121/cmr2020.1541

[13] Tonn VJR, MacRae JM.Sex and gender differences in health-related quality of life in individuals treated with incremental and conventional haemodialysis. Clin Kidney J 2024; 17(10): sfae273. https://doi.org/10.1093/ckj/sfae273

[14] Johnson DW, DentH, Hawley CM, *et al*. Associations of Dialysis Modality and Infectious Mortality in Incident Dialysis Patients in Australia and New Zealand.American Journal of Kidney Diseases 2009; 53(2): 290-297. https://doi.org/10.1053/j.ajkd2008.06.032

[15] Noh J, Park SY, Bae W, *et al*. Predicting early mortality in haemodialysis patients: a deep learning approach using a nationwide prospective cohort in South Korea. Sci Rep 2024; 14: 29658. https://doi.org/10.1038/s41598-024-80900-6

[16] Comorbidity as a driver of adverse outcomes in people with chronic kidney disease. Tonelli M, Wiebe N, Guthrie B, *et al.* Kidney International 2015 ; 88(4): 859-66. https://doi.org/10.1038/ki2015.228

[17] AfkarianM, Zelnick LR, HallY N, *et al*. (2013). Clinical manifestations of kidney disease among US adults with diabetes, 1988-2014. JAMA 2016; 316(6): 602-610. https://doi.org/10.1001/jama2016.10924

[18] Go AS, Chertow GM, Fan D, *et al*. (2004). Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. The New England Journal of Medicine 351(13): 1296-1305. https://doi.org/10.1056/NEJMoa041031

[19] Iordache BE, Perico N, Bikbov B, etal.Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study.Lancet Global Health 2016; 4: e307-e319. https://doi.org/10.1016/S2214-109X(16)00071-1

[20] Haleen HV, Jackson J, Spinowitz B, *et al*. Impact of chronic kidney disease and anemia on health-related quality of life and work productivity: analysis of multinational real-world data. BMC Nephrol 2020; 21(1): 88. https://doi.org/10.1186/s12882-020-01746-4

[21] Thomas B, Matsushita K, Abate KH, *et al*. Global Cardiovascular and Renal Outcomes of Reduced GFR. J Am Soc Nephrol 2017; 28(7): 2167-2179.

[22] Rangaswami J, Mathew OR, Parasuraman R, *et al*. Cardiovascular disease in the kidney transplant recipient: epidemiology, diagnosis and management strategies. Nephrology Dialysis Transplantation 2019 ; 34(5): 760-773. https://doi.org/10.1093/ndt/gfz053

[23] Montalescot L, Dorard G, Speyer E. Patient perspectives on chronic kidney disease and decision-making about treatment. Discourse of participants in the French CKD-REIN cohort study. J Nephrol 2022 ; 35(5): 1387-1397. https://doi.org/10.1007/s40620-022-01345-6

[24] Brück K, Stel, VS, Gambaro G, *et al*. (CKD prevalence varies across the European general population. Journal of the American Society of Nephrology 2016 ; 27(7): 2135-47. https://doi.org/10.1681/ASN 2015050542

[25] Weiner DE. The dialytic prescription: intensifying dialysis to improve outcomes. Seminars in Dialysis 2009; 22(5): 543-547.

[26] Saran R, Robinson B, Abbott K. *et al*. US Renal Data System 2017 Annual Data Report: Epidemiology of kidney disease in the United States. American Journal of Kidney Diseases 2018; 71(3 Suppl 1): A7. https://doi.org/10.1053/j.ajkd2018.03.001

[27] Webster AC, Nagler EV, Morton RL, *et al*. Chronic kidney disease. The Lancet 2017; 389(10075), 1238-1252. https://doi.org/10.1016/S0140-6736(16)32064-5

[28] Takkavatakarn K, Oh W, Cheng E, *et al.* Machine learning models to predict end-stage kidney disease in chronic kidney disease stage 4. BMC Nephrology 2023; 24(1): 376. https://doi.org/10.1186/s12882-023-03424-7

[29] Islam Ariful, Majumder ZH, Hussein A.Chronic kidney disease prediction based on machine learning algorithms. J Pathol Inform 2023; 14: 100189. https://doi.org/10.1016/j.jpi2023.100189

[30] Pan Q, Tong M. Artificial intelligence in predicting chronic kidney disease prognosis. A systematic review and meta-analysis. Ren Fail 2024; 46(2): 2435483. https://doi.org/10.1080/0886022X2024.2435483

[31] Lousa I, Reis F. New Potential Biomarkers for Chronic Kidney Disease Management—A Review of the Literature.International Journal of Molecular Sciences 2020; 22(1): 43. https://doi.org/10.3390/ijms22010043

[32] Bowling CB, Plantinga L, Phillips LS, *et al.* Association of Multimorbidity with Mortality and Healthcare Utilization in Chronic Kidney Disease. Journal of the American Geriatrics Society 2017; 65(4): 704-711. https://doi.org/10.1111/jgs.14662

[33] Tonelli M, Weibe N, James M, *et al*. A population-based cohort study defines prognoses in severe chronic kidney disease. Kidney Int 2018; 93(5): 1217-1226. https://doi.org/10.1016/j.kint2017.12.013

[34] Fraser SD, Roderick PJ, May CR. The burden of comorbidity in people with chronic kidney disease stage 3: a cohort study. BMC Nephrology 2015; 16: 193. https://doi.org/10.1186/s12882-015-0189-z

[35] Neal B, Perkovic V, Mahaffey KW, *et al*. Canagliflozin and cardiovascular and renal events in type 2 diabetes. The New England Journal of Medicine 2017; 377(7): 644-657. https://doi.org/10.1056/NEJMoa1611925

[36] Kucirka LM, Grams ME, Lessler J, *et al*. Association of race and age with survival among patients undergoing dialysis. JAMA 2011; 10(306): 620-6. https://doi.org/10.1001/jama2011.1127

[37] Sikole A, Nikolov V, Dzekova P, *et al*. Survival of patients on maintenance haemodialysis over a twenty-year period. Prilozi 2007; 28: 99-110.

Received on 14-10-2025 Accepted on 16-11-2025 Published on 08-12-2025

https://doi.org/10.6000/1929-6029.2025.14.68