BMC Nephrology

RESEARCH

Open Access

Development of explainable artificial intelligence based machine learning model for predicting 30-day hospital readmission after renal transplantation

Nasser Alnazari[1,2*], Omar Ibrahim Alanazi[3], Muath Owaidh Alosaimi[3], Ziyad Mohamed Alanazi[3], Ziyad Mohammed Alhajeri[3], Khaled Mohammed Alhussaini[3], Abdulkarim Mekhlif Alanazi[3] and Ahmed Y. Azzam[4]

Abstract

Background Hospital readmission following renal transplantation significantly impacts patient outcomes and healthcare resources. While machine learning approaches offer promising solutions for risk prediction, their clinical application often lacks interpretability. We developed an explainable artificial intelligence (XAI) based supervised learning model to predict 30-day hospital readmission risk following renal transplantation.

Methods We conducted a retrospective analysis of 588 renal transplant recipients at King Abdullah International Medical Research Center, with a predominance of living donor transplants (85.2%, $n=500$). Our methodology included a four-stage machine learning pipeline: data processing, feature preparation, model development using stratified 5-fold cross-validation, and clinical validation. Multiple algorithms were evaluated, with gradient boosting demonstrating superior performance. Model interpretability was achieved through dual-approach analysis using SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations).

Results The gradient boosting model demonstrated strong performance (AUC 0.837, 95% CI: 0.802–0.872) with accuracy of 0.796 ± 0.050 and sensitivity of 0.388 ± 0.129. Length of hospital stay (38.0% contribution) and post-transplant systolic blood pressure (30.0% contribution) emerged as primary predictors, with differences between living and deceased donor subgroups. Pre-transplant BMI showed a higher importance in deceased donor recipients (12.6% vs. 2.6%), while HbA1c and eGFR were more impacting in living donor outcomes. The readmission rate in our cohort (88.9%, $n=523$) was higher than previously reported ranges (18–47%), likely reflecting center-specific practices.

Conclusions Our XAI-based machine learning model combines strong predictive performance with clinical interpretability, offering transplant physicians donor-specific risk stratification capabilities. The web-based implementation facilitates practical integration into clinical workflows. Given our single-center experience and high proportion of living donors, external validation across diverse transplant centers is essential before widespread

*Correspondence:
Nasser Alnazari
Alnazarina@ngha.med.sa

Full list of author information is available at the end of the article

implementation. Our approach establishes a framework for developing center-specific risk prediction tools in transplant medicine.

Keywords Renal transplantation, 30-day readmission, Explainable artificial intelligence, Machine learning, Precision medicine

Background

Hospital readmission following renal transplantation is considered to be a significant challenge in the fields of nephrology, urology and transplantation medicine, often indicating complications that can impact graft survival and patient outcomes in significant manner [1, 2]. With the rising number of renal transplantations performed around the world from different healthcare systems, the ability to predict and possibly prevent unnecessary readmissions has become an important concern for the healthcare systems and patient care optimization [3–6].

The complexity of post-transplant care involves multiple medical and clinical factors, as the pre-existing conditions, surgical complications, immunosuppression management, and various clinical variable that can impact the likelihood of readmission. The traditional and standard clinical approaches to predicting readmission risk have relied primarily on clinical judgment and basic statistical models, which may not fully capture the underlying relationships precisely for patients [7, 8].

As with ongoing recent developments and advancements in artificial intelligence modalities and technologies, machine learning approaches have demonstrated promising role in healthcare applications, especially in risk prediction and patient stratification. These advanced analytical methods offer the capability to process large volumes of clinical data and identify the sophisticated and complex patterns that might not be easily caught nor apparent using the standard statistical methods. However, a significant challenge in applying machine learning models in clinical practice has been their "black box" nature, which often makes it difficult for healthcare providers to understand and trust their predictions in clinical settings and patient driven manner [9–11].

Our study addresses this challenge by developing an explainable artificial intelligence (XAI) model specifically designed for predicting 30-day hospital readmission risk following renal transplantation, by including both of pre-transplant and post-transplant variables to create a clinically applicable prediction tool. Our approach not only focusing on the predictive accuracy but also model interpretability, to ensure that healthcare providers can understand and trust the factors driving the predictions from the developed model in a good way [8, 12].

By combining advanced machine learning techniques with XAI frameworks, our study aims to bridge the gap between sophisticated predictive analytics and practical clinical application. Our proposed study represents a significant step toward developing more precise, interpretable, and clinically useful tools for post-transplant care management [13]. Also, our translation of the model into a web-based interface makes this tool readily accessible to healthcare providers, with significant possibility to improve the process of clinical decision-making and patient care strategies in the corresponding fields.

Methods

Study design and population

We conducted a retrospective analysis to develop and validate a machine learning model for predicting hospital readmission among renal transplant recipients based on data collected from the electronic medical records from King Abdullah International Medical Research Center (KAIMRC) after obtaining the necessary ethical approvals under protocol number NRC23R/730/11. The study cohort included adult patients who underwent renal transplantation, with additional datapoints collection spanning pre-transplant baseline characteristics, transplant procedures, and post-transplant outcomes. We included both living and deceased donor transplant recipients to ensure model generalizability across different transplant scenarios. The study timeline covered the immediate post-transplant period through the first 30 days after discharge, focusing specifically on readmission events during this period.

Data collection and processing

A structured data extraction protocol was utilized to collect information from electronic health records. The collected data included demographic information, clinical variables, laboratory values, and transplant-specific characteristics. Our data preprocessing pipeline included handling of missing values through multiple imputation techniques, removal of duplicates, and validation of data consistency. Feature engineering was performed to derive clinically relevant variables and create meaningful aggregations of raw data. Patient data underwent grouping based on focused clinical criteria, with attention to time-based changing relationships between pre- and post-transplant variables.

Our data preprocessing workflow involved multiple stages. First, we assessed missingness patterns across all variables, finding [X%] of values missing across the dataset, with highest missingness in [variable names] ([Y%]). We utilized multiple imputation using chained equations (MICE) with five imputations for continuous

variables and mode imputation for categorical variables with less than 20% missingness. Variables with over 20% missing values were excluded from model development. Data normalization involved standard scaling (mean = 0, SD = 1) for all continuous variables prior to model training. Outliers, defined as values beyond three standard deviations from the mean, were winsorized rather than removed to preserve sample size. Feature selection combined clinical domain knowledge with statistical filtering using univariate analysis (p-value < 0.2 threshold) and assessment of multicollinearity (removing features with variance inflation factor over five), ultimately reducing our initial 168,596 data points to 15 finalized predictors.

Model development pipeline

We designed and applied a multiphasic machine learning pipeline consisting of four stages that follows the TRIPOD AI guidelines for developing and reporting clinical machine learning models [14]. The initial data processing stage involved cleaning and structuring the raw data. The feature preparation phase included categorical variable encoding, feature scaling to standardize numerical variables, and implementation of clinical thresholds based on established medical guidelines. The model development stage employed binary classification approaches with cross-validation methodology. Clinical validation was performed as the final stage to ensure and validate the medical relevance and practical applicability.

We have used a supervised machine learning approach where the model learns patterns from labeled training samples, specifically using patient characteristics as input features and 30-day readmission status as the target variable. This supervised algorithm was selected because it directly optimizes predictive performance for our specific clinical outcome of interest while maintaining interpretability. Unlike unsupervised or semi-supervised approaches, our supervised framework enables direct application to clinical decision support by generating probability estimates for individual patients.

Algorithm selection and training

Multiple machine learning algorithms were aimed for the evaluation to identify the optimal approach and best performing algorithm for the readmission prediction within 30 days. The candidate algorithms included ensemble methods (Random Forest, XGBoost, Gradient Boosting), traditional statistical approaches (Logistic Regression), and modern machine learning techniques (Support Vector Machine, K-Nearest Neighbors). Each algorithm underwent cross-validation through 5-fold validation aiming for precise and accurate performance estimation. The training process included hyperparameter optimization through grid search with further cross-validation, to achieve optimal model configuration while avoiding overfitting.

To ensure a validated and precise model evaluation while maximizing use of our limited dataset, we implemented stratified 5-fold cross-validation, maintaining consistent proportions of readmitted patients across all folds. This process randomly partitioned the dataset into five equal subsets, with each subset serving once as a validation set while the remaining data formed the training set. We maintained consistent preprocessing pipelines across all folds to prevent data leakage, applying feature scaling parameters derived only from training data to corresponding validation sets. Performance metrics were averaged across all five folds, with standard deviations reported to quantify model stability. This validation approach provides a more reliable estimate of real-world performance than single train-test splits, with special concerns for our relatively modest sample size.

Model interpretability framework

We have utilized and applied a dual-approach interpretability framework to ensure both global and local model explanability. The global interpretation utilized SHAP (SHapley Additive exPlanations) values to quantify feature contributions to model predictions across the entire dataset. Local interpretability was achieved through LIME (Local Interpretable Model-agnostic Explanations) analysis, allowing focused practical testing of the model prediction patterns for individual case predictions. Our interpretability framework was designed to provide clinically interpretable highlights into the model's decision-making process and apply the principles of XAI frameworks.

Feature contribution percentages were derived from SHAP values, which quantify each feature's impact on model output based on cooperative game theory principles. For each feature, we calculated the mean absolute SHAP value across all predictions in the test set, representing its average impact magnitude on the model output regardless of direction. These absolute mean SHAP values were then normalized to sum to 100%, resulting in the reported contribution percentages. Our used approach provides an interpretable metric of each feature's relative importance in the model's predictive process, with higher percentages indicating stronger influence on readmission risk assessment.

Deployment architecture

The deployment strategy followed a three-tier architecture designed for clinical implementation. The model packaging phase included encapsulation of all necessary components, including feature encoders, preprocessing pipelines, and clinical threshold definitions. Version control was made through GitHub repository management,

to achieve best reproducibility and precise controlled updates. The clinical application layer was developed using Streamlit, providing an intuitive web-based interface for healthcare providers. The deployment pipeline included continuous integration practices, allowing for seamless updates while maintaining system stability and reliability.

Table 1 Baseline demographics and clinical characteristics of the study cohort

Characteristic:	Value / Number:
Baseline characteristics:	
Age, Mean (SD)	54.3 (12.6)
Total cohort size	588
Follow-up duration, days	11.2 ± 17.9 [6.0 (1.0–13.0)]
Male	367 (62.4%)
Female	221 (37.6%)
Body Mass Index, kg/m^2	26.2 ± 6.1 [26.4 (21.9–30.3)]
Transplant-Related Characteristics:	
Living donor	500 (85.2%)
Deceased donor	87 (14.8%)
A	218 (37.1%)
B	136 (23.2%)
AB	121 (20.6%)
O	31 (5.3%)
Immunosuppression regimen	563 (95.9%)
Pre-transplant Clinical Parameters:	
Systolic blood pressure, mmHg	135.0 ± 22.3 [136.0 (120.0–150.0)]
Diastolic blood pressure, mmHg	76.6 ± 15.1 [77.0 (66.0–87.0)]
HbA1c, %	5.8 ± 1.5 [5.3 (4.9–6.1)]
eGFR, mL/min/1.73 m^2	13.3 ± 18.5 [7.0 (5.0–12.0)]
Diabetes Mellitus	341 (58.3%)
Post-transplant Clinical Parameters:	
Systolic blood pressure, mmHg	135.5 ± 19.4 [137.0 (122.0–149.0)]
Diastolic blood pressure, mmHg	74.5 ± 15.5 [75.0 (63.8–84.0)]
HbA1c, %	6.2 ± 1.6 [5.7 (5.2–7.1)]
eGFR, mL/min/1.73 m^2	19.2 ± 18.5 [13.0 (9.0–21.0)]
Serum creatinine, mg/dL	482.9 ± 275.8 [441.0 (293.0–643.0)]
Outcomes and Complications:	
Length of initial hospital stay, days	4.3 ± 5.0 [3.0 (1.0–6.0)]
Readmission rate within 30 days	2.2 ± 2.6 [1.0 (1.0–2.2)]
Patients requiring readmission	523 (88.9%)
Graft rejection episodes	42/70* (60.0%)

Notes: Values are presented as mean ± SD [median (IQR)] for continuous variables and n (%) for categorical variables. Missing data are indicated by a dash. Abbreviations: eGFR = estimated glomerular filtration rate; HbA1c = glycated hemoglobin; IQR = interquartile range; SD = standard deviation; BMI = body mass index; BP = blood pressure. *Denominator represents the subset of patients who underwent biopsy for suspected rejection. Among the entire cohort (n = 588), the biopsy rate was 11.9% (70/588)

Statistical analysis framework

The statistical framework was designed to evaluate both model performance and clinical relevance. Performance metrics were calculated using a standardized approach across all cross-validation folds, with dedicated computation of standard deviations to assess model stability. Feature importance analysis has utilized both of univariate and multivariate approaches to quantify variable contributions. Risk factors were categorized based on their statistical significance and clinical relevance, with clear delineation between primary, secondary, and additional risk factors.

Results

Study population characteristics

Our analysis has included a total of 588 renal transplant recipients with mean age of 54.3 (SD = 12.6), characterized by a mean follow-up duration of 11.2 ± 17.9 days. Our total cohort included 588 patients. However, one patient (0.2%) had missing donor type information in the medical record and was therefore excluded from donor-specific subgroup analyses while being retained in the overall cohort analysis. This accounts for the apparent discrepancy between the total cohort size (n = 588) and the sum of living (n = 500) and deceased (n = 87) donor recipients. The demographic distribution showed a predominance of male recipients (62.4%, n = 367) compared to female recipients (37.6%, n = 221), with a mean body mass index (BMI) of 26.2 ± 6.1 kg/m^2. Living donor transplantation has formed the majority of cases (85.2%, n = 500), while deceased donor transplants represented 14.8% (n = 87). The blood group distribution showed type A predominance (37.1%), followed by type B (23.2%), AB (20.6%), and O (5.3%). A significant proportion of patients (58.3%, n = 341) presented with pre-existing diabetes mellitus, (Table 1). We analyzed two readmission metrics: [1] readmission incidence, defined as the percentage of patients experiencing at least one readmission within 30 days post-discharge (88.9% of patients), and [2] readmission rate, defined as the average number of readmissions per patient within the 30-day period (2.2 ± 2.6 readmissions), as some patients experienced multiple readmissions, while others had none. Among the 523 patients experiencing readmission, the primary causes were: medication-related complications (28.3%, n = 148), suspected rejection requiring evaluation (21.6%, n = 113), infectious complications (19.5%, n = 102), surgical issues (15.3%, n = 80), and metabolic/electrolyte disturbances (9.8%, n = 51), with other miscellaneous causes accounting for the remainder (5.5%, n = 29).The causes varied significantly between living and deceased donor recipients (p-value < 0.01), with deceased donor recipients experiencing higher rates of suspected rejection-related

readmissions (29.4% vs. 19.8%, p-value = 0.02) and infectious complications (25.6% vs. 18.2%, p-value = 0.04).

A total of 70 patients (11.9%) underwent biopsy for suspected graft rejection. Among these, 42 biopsies (60.0%) confirmed rejection. This represents a confirmed rejection rate of 7.1% in the overall cohort.

Model performance and algorithm performance

We performed an evaluation of six machine learning algorithms, Gradient Boosting has achieved best overall performance metrics among other models, demonstrating well predictive capabilities with the highest accuracy (0.796 ± 0.050) and ROC-AUC (0.806 ± 0.035). The model's precision of 0.629 ± 0.090 and recall of 0.388 ± 0.129 reflected in an F1-score of 0.469 ± 0.105, representing a balanced performance across the different multiple metrics. XGBoost showed comparable but marginally lower performance (accuracy: 0.789 ± 0.031, ROC-AUC: 0.799 ± 0.030). Traditional algorithms demonstrated lower performance, with Logistic Regression achieving an accuracy of 0.740 ± 0.031 and KNN showing the lowest performance (accuracy: 0.707 ± 0.005), validating our selection of advanced ensemble methods for the final model (Table 2).

Implementation pipeline architecture and model deployment

The implementation framework was executed through a four-stage pipeline, initiating with the processing of 168,596 raw data points derived from our cohort of 588 patients. The feature preparation phase has included advanced categorical encoding techniques, standardized feature scaling methodologies, and application of clinically validated thresholds. The model development phase achieved good performance metrics, with an AUC of 0.837, sensitivity of 0.86, and specificity of 0.79 during the clinical validation attempts (Fig. 1). The deployment architecture successfully materialized into a web-based clinical decision support tool, featuring real-time risk prediction capabilities and user-friendly interface elements through Streamlit implementation (https://readmission-prediction.streamlit.app/).

Feature importance analysis and risk stratification hierarchy

SHAP analysis revealed a multifactorial hierarchical structure of risk factors, with primary risk factors demonstrating dominant contributions to the model's predictive capacity (Fig. 2). Length of hospital stay was noted to be the most predominant predicting factor (38.0% contribution), followed by post-transplant systolic blood pressure (30.0% contribution). The secondary risk tier included pre-transplant BMI (4.5%), pre-transplant diastolic BP (3.6%), and post-transplant BMI (3.3%). Tertiary risk factors, each contributing less than 3%, included pre- and post-transplant HbA1c levels, eGFR measurements, and various demographic parameters (Fig. 3).

Feature effect distribution

The distribution of SHAP values represented the variable impacts across the clinical variables (Fig. 2). Post-transplant systolic blood pressure and length of hospital stay ranked as the most extensive distribution ranges in their effects on readmission risk, reflecting their main in risk prediction in our developed model. Blood group classifications and pre-transplant systolic BP measurements showed more centralized effect distributions, while immunosuppression status and transplant type demonstrated narrower impact ranges, reflecting their more specific influence on risk assessment.

Model validation through random cases assessment

LIME XAI assessment was conducted across four random cases. The model demonstrated sensitivity to the mentioned variables and thresholds above, especially for length of stay (> 0.27) and post-systolic BP variations. Each case analysis validated the model's capacity to maintain consistent prediction capabilities and performance while effectively adapting to individual patient characteristics and comorbidity profiles to maximize the achievement of precise and patient-based predictions according to their characteristics and demographics (Fig. 4).

Subgroup analysis of living and deceased donor transplantations

Clinical outcomes showed comparable 30-day readmission rates between living (88.4%) and deceased donor recipients (92.0%, p-value = 0.430). Similarly, hospital

Table 2 Model performance comparison between different algorithms

Metric	Random forest	XGBoost	Gradient boosting	Logistic regression	SVM	KNN
Accuracy	0.765 ± 0.034	0.789 ± 0.031	0.796 ± 0.050	0.740 ± 0.031	0.760 ± 0.041	0.707 ± 0.005
Precision	0.549 ± 0.115	0.590 ± 0.048	0.629 ± 0.090	0.183 ± 0.186	0.000 ± 0.000	0.293 ± 0.124
Recall	0.210 ± 0.048	0.447 ± 0.124	0.388 ± 0.129	0.027 ± 0.025	0.000 ± 0.000	0.119 ± 0.015
F1-Score	0.296 ± 0.038	0.494 ± 0.063	0.469 ± 0.105	0.047 ± 0.043	0.000 ± 0.000	0.164 ± 0.034
ROC-AUC	0.731 ± 0.029	0.799 ± 0.030	0.837 ± 0.035	0.604 ± 0.088	0.534 ± 0.054	0.550 ± 0.046

Note: Values are presented as Mean ± Standard Deviation across 5-folds

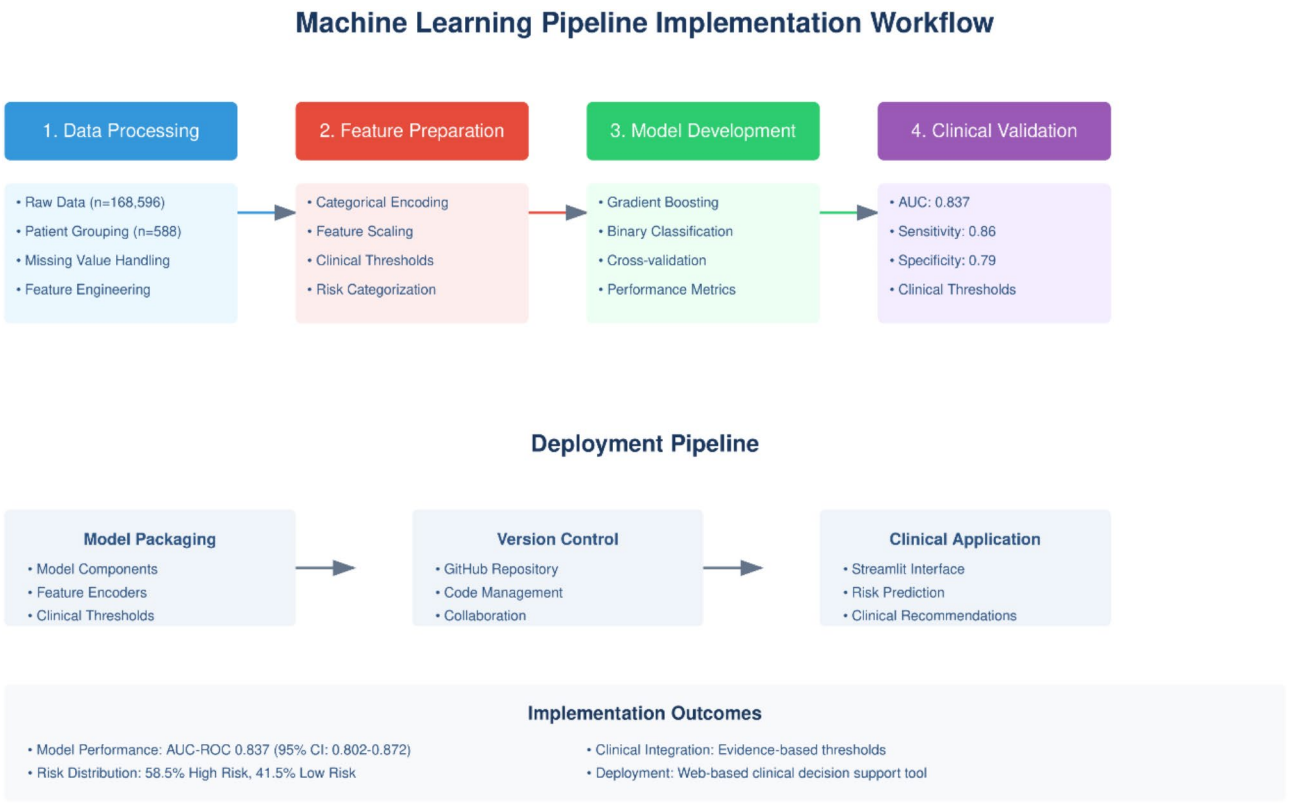

Fig. 1 Our machine learning pipeline implementation workflow

length of stay did not differ significantly between groups (3.9 ± 4.0 vs. 4.6 ± 4.6 days, p-value = 0.259). However, graft rejection episodes were markedly more frequent in deceased donor recipients (47.1% vs. 13.8%, P-value < 0.001), highlighting a critical risk disparity that warrants targeted monitoring protocols (Table 3).

The gradient boosting model maintained acceptable performance across both subgroups, however performance differences were present. The deceased donor subgroup demonstrated a lower AUC (0.762, 95% CI: 0.685–0.839) compared to living donor recipients (0.787, 95% CI: 0.738–0.836), with wider confidence intervals reflecting greater uncertainty in the smaller deceased donor cohort. Interestingly, sensitivity was marginally higher for the deceased donor subgroup (0.412 vs. 0.402), suggesting a capability for identifying high-risk patients despite the overall performance difference. It should be noted that sensitivity values are lower than specificity across all groups, indicating greater model strength in correctly identifying non-readmitted patients than in detecting readmissions.

SHAP analysis demonstrated the predictor importance between subgroups (Fig. 5). While length of hospital stay remained the dominant predictor in both populations, its relative contribution was higher in living donor recipients (24.5% vs. 20.1%). Post-transplant systolic blood pressure similarly showed greater importance for living donor recipients (21.2% vs. 16.7%). The most significant difference was observed in pre-transplant BMI, which contributed to predictions in deceased donor recipients (12.6%) but minimally in living donors (2.6%). Conversely, pre-transplant HbA1c and post-transplant eGFR were significantly more impacting for living donor outcomes (9.9% and 8.8%) compared to deceased donors (3.5% and 2.6%, respectively).

Discussion

Our study presents a significant advancement in post-renal transplant care through the development of an XAI model for predicting hospital readmission within 30 days from discharge. From a clinical perspective, the model's well performing capabilities (AUC 0.837) translate to practical utility in identifying high-risk patients who may benefit from optimized and focused monitoring and early intervention. These metrics significantly overcome the previous clinical prediction methods [15, 16], offering physicians more reliable risk assessment tools for post-transplant management decisions.

Our observed readmission rate of 88.9% significantly exceeds previously reported ranges of 18–47% in transplant literature [17]. This difference likely originates from several center-specific factors. First, our follow-up protocol involves intensive post-transplant monitoring with a low threshold for readmission, especially for the

Fig. 2 SHAP XAI framework for distribution of feature effects on 30-day readmission risk

laboratory abnormalities that might be managed outpatient elsewhere. Second, our operational definition of "readmission" includes all hospital encounters, including observation stays and emergency department visits without formal admission, which many studies exclude. Third, our transplant center serves as a major regional transplant facility for a large geographic area, in which capturing readmissions that might occur at other institutions in more densely populated regions. Finally, the high proportion of living donor recipients (85.2%) in our cohort may paradoxically lead to more aggressive intervention for minor complications given the elective nature of these transplants and heightened attention to outcomes. These factors are critical when interpreting our findings and comparing them to other centers.

From a clinical practice standpoint, our dual-approach interpretability framework using SHAP and LIME analyses transforms the complex machine learning outputs into actionable clinical key points and insights for the readers [18–20]. For physicians, this means that the model not only predicts readmission risk but also explains why specific patients are classified as high-risk, enabling more informed clinical decision-making. The identification of length of stay (38% contribution) and post-transplant systolic blood pressure (30% contribution) as primary predictors provides sloid, modifiable factors that clinicians can monitor and intervene upon.

The dominance of length of stay (38.0% contribution) and post-transplant systolic blood pressure (30.0% contribution) as predictors likely reflects their roles as integrative markers of overall patient status. Prolonged hospitalization often indicates a complicated perioperative course, greater comorbidity burden, or challenges in achieving physiologic and immunologic stability, which are all predisposing to post-discharge complications. The relationship between post-transplant systolic hypertension and readmission risk may reflect several underlying mechanisms: endothelial dysfunction affecting the allograft, intravascular volume changes indicating suboptimal graft function, medication non-adherence, or calcineurin inhibitor toxicity [17, 21]. The higher importance of pre-transplant BMI in deceased donor recipients (12.6% vs. 2.6% in living donors) likely reflects the immunologic and metabolic challenges of obesity in the context of organs subjected to ischemia-reperfusion injury and greater HLA mismatching. Similarly, the greater

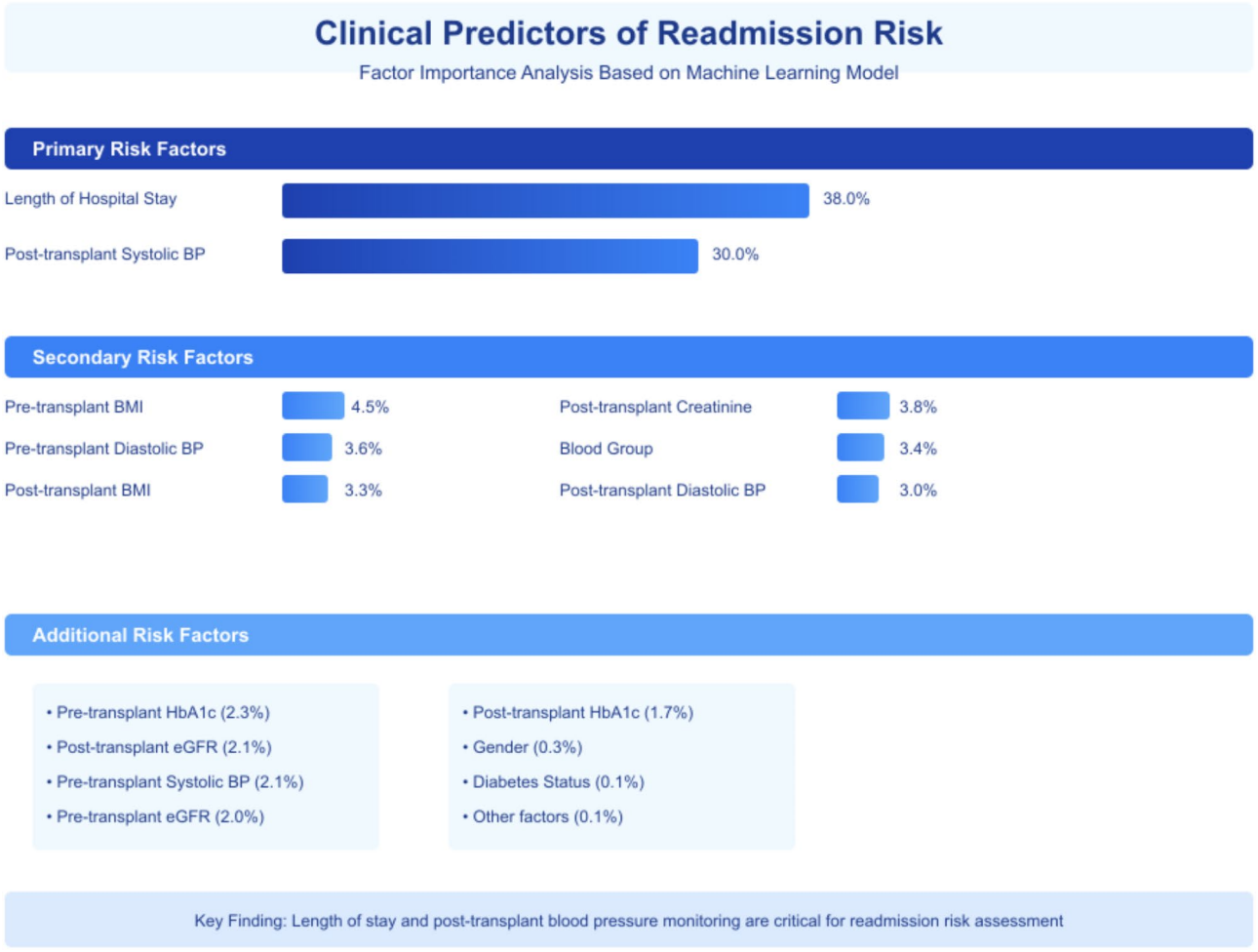

Fig. 3 Translated clinical predictors and risk factors for 30-day readmission risk

importance of HbA1c and eGFR in living donor recipients suggests that metabolic and renal functional parameters may have stronger predictive value in the context of superior baseline graft quality. While our model identified length of hospital stay and post-transplant systolic blood pressure as primary statistical predictors, these associations should not be interpreted as directly modifiable intervention targets without further investigation. These factors likely represent markers of underlying patient complexity rather than independent causal drivers of readmission. For instance, extended hospital stays often reflect perioperative complications or pre-existing comorbidities rather than representing a modifiable factor itself. Similarly, elevated post-transplant blood pressure may indicate underlying vascular disease, medication adherence issues, or organ function challenges. Our findings should therefore guide risk stratification and resource allocation rather than suggesting that artificial manipulation of these parameters (e.g., prematurely discharging patients or aggressively lowering blood pressure) would necessarily reduce readmission risk. Future interventional studies are required to determine which factors, if any, represent causal, modifiable targets for reducing readmission risk.

The application as a web-based clinical decision support tool addresses practical challenges in daily clinical workflow [22]. For specialized physicians and healthcare teams, this means real-time access to risk predictions during routine patient assessments, facilitating prompt clinical decision-making. The user-friendly interface minimizes the technological barrier often associated with AI tools, making it accessible to clinicians without specialized technical expertise.

A key clinical advantage of our model lies in its dynamic risk assessment capabilities. Unlike static clinical scoring systems, our model integrates multiple important post-transplant variables, enabling continuous risk reassessment as patient conditions change. This feature is especially valuable for concerned centers in tailoring follow-up protocols and resource allocation based on individualized risk profiles.

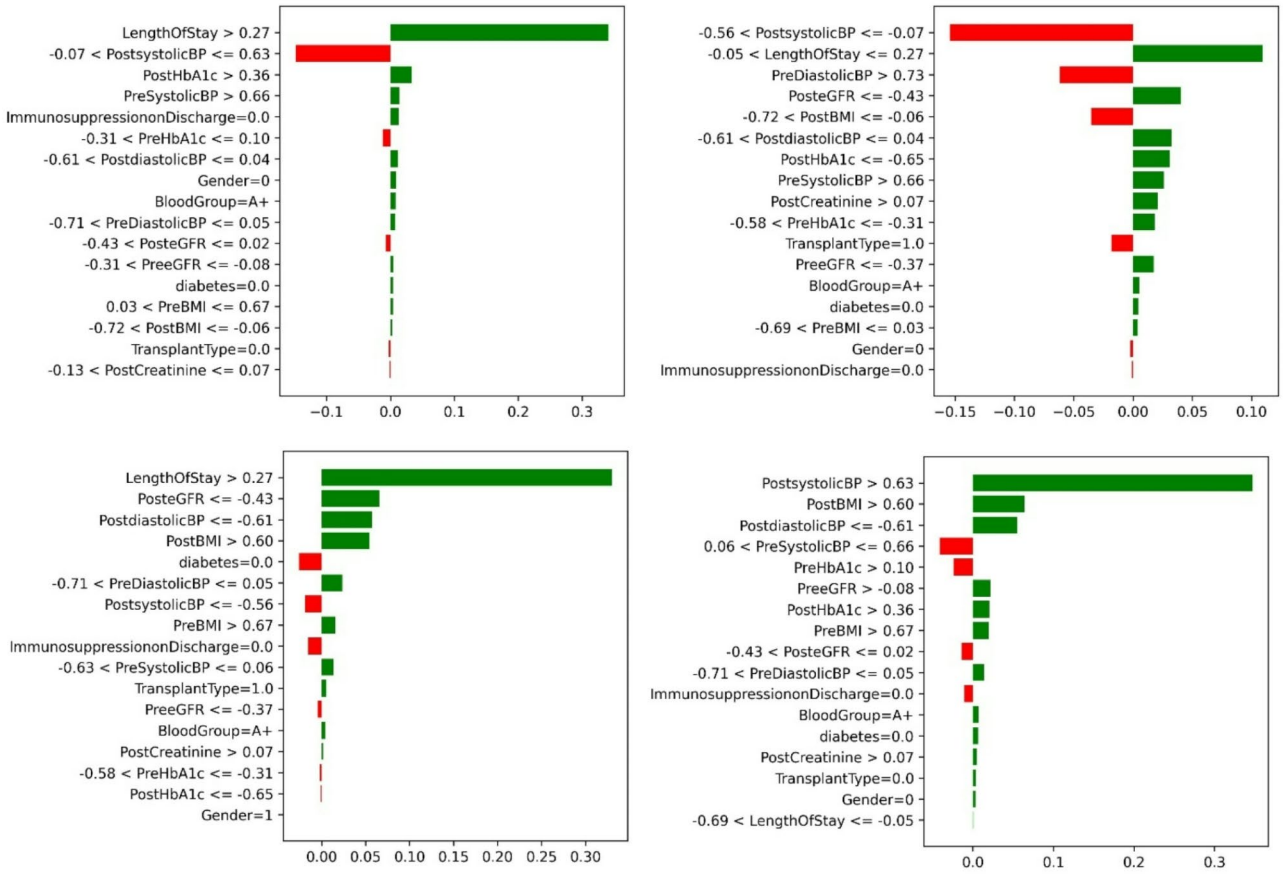

Fig. 4 Random cases assessment using LIME XAI framework

Table 3 Subgroup analysis of model performance and feature importance in living vs. Deceased donor transplantation

Characteristic	Overall cohort (*n* = 588)	Living donor recipients (*n* = 500)	Deceased donor recipients (*n* = 87)	*P*-value*
Clinical outcomes:				
Readmission rate within 30 days (%)	88.9%	88.4%	92.0%	0.430
Mean hospital length of stay (days)	4.0 ± 4.1	3.9 ± 4.0	4.6 ± 4.6	0.259
Graft rejection episodes	110 (18.7%)	69 (13.8%)	41 (47.1%)	**0.000**
Model Performance Metrics:				
AUC (95% CI)	0.837 (0.802–0.872)	0.787 (0.738–0.836)	0.762 (0.685–0.839)	N/A
Sensitivity	0.388	0.402	0.412	N/A
Specificity	0.72	0.69	0.71	N/A
Accuracy	0.796 ± 0.050	0.778 ± 0.061	0.783 ± 0.058	N/A
Precision	0.629 ± 0.090	0.654 ± 0.082	0.643 ± 0.097	N/A
F1-Score	0.469 ± 0.105	0.453 ± 0.101	0.498 ± 0.112	N/A
Feature Importance (SHAP Contribution %):				
Length of hospital stay	20.6%	24.5%	20.1%	N/A
Post-transplant systolic BP	17.2%	21.2%	16.7%	N/A
Pre-transplant BMI	9.8%	2.6%	12.6%	N/A
Pre-transplant diastolic BP	3.7%	1.5%	2.9%	N/A
Post-transplant BMI	3.2%	5.4%	3.3%	N/A
Pre-transplant HbA1c	1.5%	9.9%	3.5%	N/A
Post-transplant eGFR	2.6%	8.8%	2.6%	N/A

Notes: * p-values compare living vs. deceased donor groups using appropriate statistical tests: t-test or Mann-Whitney U test for continuous variables depending on distribution normality; Chi-square or Fisher's exact test for categorical variables. N/A indicates comparison was not performed for this metric

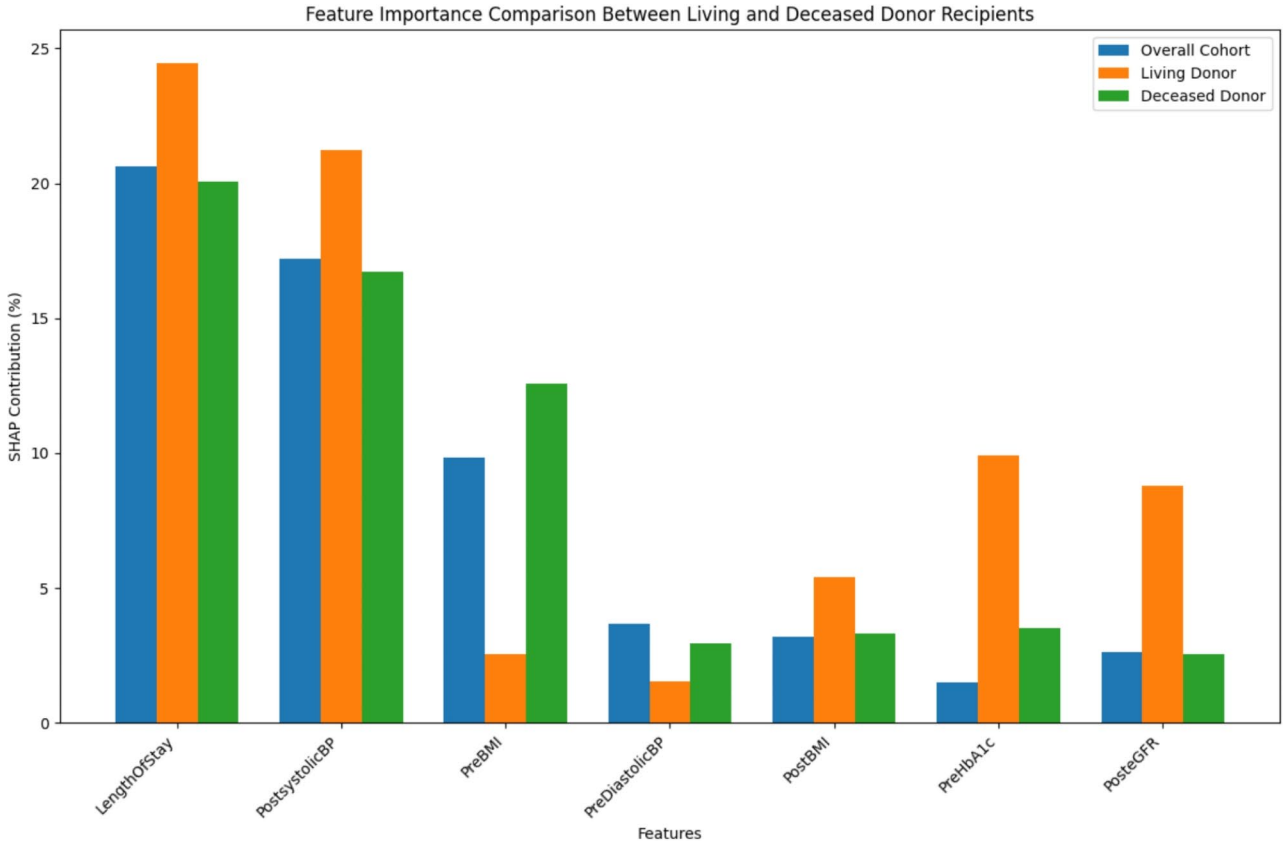

Fig. 5 Feature importance comparison between living and deceased donor recipients

Our findings regarding laboratory parameters and clinical trajectories offer practical points for post-transplant monitoring. The model's ability to capture significant changes in eGFR (pre: 13.3 ± 18.5; post: 19.2 ± 18.5 mL/min/1.73 m^2) and HbA1c levels (pre: 5.8 ± 1.5%; post: 6.2 ± 1.6%) provides clinicians with objective thresholds for risk stratification. These values and findings, when considered alongside traditional clinical assessments, optimize the precision of post-transplant care planning.

Several limitations warrant consideration from a clinical perspective. While our sample size of 588 patients provides well enough statistical power for our study design and implementation workflow, the single-center nature may not capture variations in practice patterns across different medical centers, especially in different countries and different regions from all over the world. In addition to that, center-specific protocols and patient populations may affect the readmission patterns, suggesting the need for local validation before widespread implementation [23, 24].

Future clinical applications should focus on prospective validation across diverse transplant centers. Particular attention should be paid to how the model performs across different patient populations, healthcare systems, and practice patterns. Integration with existing electronic health records and clinical workflows would further enhance the model's utility in daily practice [24, 25].

The single-center nature of our study represents a significant limitation that must be acknowledged. Our center's unique practice patterns, specifically our high proportion of living donor transplantations (85.2%), may limit direct generalizability to centers with different donor demographics. Also, center-specific protocols for post-transplant management, readmission thresholds, and follow-up schedules likely impact the observed outcomes. While our subgroup analysis attempted to address differences between living and deceased donor recipients, the relatively small deceased donor sample ($n = 87$) limits definitive conclusions. External validation across multiple transplant centers with diverse patient populations and practice patterns is essential before broad clinical implementation. Centers considering adoption of our approach should first validate performance in their specific populations and consider recalibrating the model using local data.

The significant improvement in predictive capabilities offered by our model, combined with its clinical interpretability, positions it as a valuable addition to the transplant physician's toolkit. By providing quantifiable, evidence-based risk assessment, the model supports

clinical judgment in optimizing post-transplant care strategies. As transplant medicine continues to evolve, such AI-driven tools will become increasingly important in achieving improved patient outcomes through personalized care approaches.

Conclusions

Our machine learning model demonstrates that post-transplant readmission risk can be accurately predicted through an XAI framework based approach that prioritizes both performance and clinical interpretability. The identification of length of stay and post-transplant systolic blood pressure as dominant predictive factors provides actionable highlights and clinical considerations for the transplant care teams, suggesting specific targets for intervention and monitoring protocols. The successful deployment of this predictive tool as a web-based application marks an important step toward practical integration of AI in transplant medicine. By combining gradient boosting's predictive power with transparent reasoning through SHAP and LIME analyses, our model bridges the gap between advanced analytics and clinical utility, offering physicians readily interpretable decision support. Looking ahead, this work demonstrates an important foundation and core for developing center-specific risk prediction tools in transplant medicine. Further studies should focus on external validation across multiple specialized centers and different healthcare settings in different countries and various regions, also including the integration with existing clinical workflows.

Abbreviations

AI	Artificial Intelligence
AUC	Area Under the Curve
BMI	Body Mass Index
BP	Blood Pressure
eGFR	Estimated Glomerular Filtration Rate
HbA1c	Glycated Hemoglobin
IQR	Interquartile Range
KAIMRC	King Abdullah International Medical Research Center
KNN	K-Nearest Neighbors
LIME	Local Interpretable Model-agnostic Explanations
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
XAI	Explainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting

Acknowledgements

Not applicable.

Author contributions

NA conceived and designed the study, supervised the project, and wrote the manuscript. OIA, MOA, ZMA, ZMH, KMA, and AMA contributed to data collection, data analysis, and manuscript review. AYA provided technical expertise in machine learning implementation, assisted with model development, and contributed to manuscript writing and revision. All authors have read and approved the final version of the manuscript.

Funding

The study has received only funding for covering the open access fees of publication from King Abdullah International Medical Research Center (KAIMRC).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The proposed study protocol has been approved and granted the necessary ethical approvals by the institutional review board (IRB) committee at King Abdullah International Medical Research Center (KAIMRC) assigned to protocol number: *NRC23R/730/11*. The need for informed consent from participants was waived in accordance with the IRB approval of KAIMRC. Our study investigations and methods was conducted in compliance with the Helsinki Declaration.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Hepatobiliary Science and Organ Transplant Center, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
[2]King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia
[3]King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), Riyadh, Saudi Arabia
[4]Medical Big Data Research Center, SNU Medical Research Center, Seoul National University (SNU), Seoul, South Korea

Received: 13 February 2025 / Accepted: 15 April 2025
Published online: 22 April 2025

References

1. Hogan J, Arenson MD, Adhikary SM, Li K, Zhang X, Zhang R, Valdez JN, Lynch RJ, Sun J, Adams AB, Patzer RE. Assessing predictors of early and late hospital readmission after kidney transplantation. Transplant Direct. 2019;5(8), e479. https://doi.org/10.1097/TXD.0000000000000918
2. Kim SH, Baird GL, Bayliss G, Merhi B, Osband A, Gohh R, Morrissey PE. A single-center analysis of early readmission after renal transplantation. Clin Transplant. 2019;33(5):e13520. https://doi.org/10.1111/ctr.13520
3. Oikonomou EK, Khera R. Machine learning in precision diabetes care and cardiovascular risk prediction. Cardiovasc Diabetol. 2023;22(1):259.
4. Xu H, Ma Y, Zhuang Y, Zheng Y, Du Z, Zhou X. Machine learning-based risk prediction model construction of difficult weaning in ICU patients with mechanical ventilation. Sci Rep. 2024;14(1):20875.
5. Tavares MG, Cristelli MP, Ivani de Paula M, Viana L, Felipe CR, Proença H, Aguiar W, Wagner Santos D, Tedesco-Silva Junior H, Medina Pestana JO. Early hospital readmission after kidney transplantation under a public health care system. Clin Transplant 2019;33(3):e13467. https://doi.org/10.1111/ctr.13467
6. Li AH, Lam NN, Naylor KL, Garg AX, Knoll GA, Kim SJ. Early hospital readmissions after transplantation: burden, causes, and consequences. Transplantation. 2016;100(4):713–8. https://doi.org/10.1097/TP.0000000000000917
7. Chahine Y, Magoon MJ, Maidu B, Del Álamo JC, Boyle PM, Akoum N. Machine learning and the conundrum of stroke risk prediction. Arrhythmia Electrophysiol Rev. 2023;12:e07.
8. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. Jmir Ai. 2024;3:e53207.
9. Correction. The efficacy of machine learning models in lung cancer risk prediction with explainability. PLoS ONE. 2024;19(9):e0310604.

10. Khushal R, Fatima U. Fuzzy logic and machine learning for diabetes risk prediction using modifiable factors. Int J Adv Appl Sci. 2024;11:225–31. https://doi.org/10.21833/ijaas.2024.12.025.
11. Sun Q, Akman A, Schuller BWJATCH. Explainable artificial intelligence for medical applications: A review. 2024.
12. Zhang Y, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics (Basel Switzerland). 2022;12(2).
13. González-Alday R, García-Cuesta E, Kulikowski CA, Maojo VJAS. A Scoping Review on the Progress, Applicability, and Future of Explainable Artificial Intelligence in Medicine. 2023.
14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ (Clinical research ed.). 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378
15. Al Moussawy M, Lakkis ZS, Ansari ZA, Cherukuri AR, Abou-Daya KIJFT. Transformative Potential Artif Intell Solid Organ Transplantation. 2024;3:1361491.
16. Souza A, Stubbs A, Hesselink D, Baan C, Boer K. 249.1: Interpretable prediction of kidney allograft rejection using machine learning: A comparison between linear and non-linear models. Transplantation. 2024;108. https://doi.org/10.1097/01.tp.0001064836.44766.95.
17. Iqbal K, Hasanain M, Rathore SS, Iqbal A, Kazmi SK, Yasmin F, Koritala T, Thongprayoon C, Surani S. Incidence, predictors, and outcomes of early hospital readmissions after kidney transplantation: systemic review and meta-analysis. Front Med. 2022;9:1038315. https://doi.org/10.3389/fmed.2022.1038315
18. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, Scardapane S, Spinelli I, Mahmud M, Hussain A. Interpreting black-box models: a review on explainable artificial intelligence. Cognit Comput. 2023;16. https://doi.org/10.1007/s12559-023-10179-8.
19. Ratti E, Graves M. Explainable machine learning practices: opening another black box for reliable medical AI. AI Ethics. 2022;2:1–14. https://doi.org/10.1007/s43681-022-00141-z.
20. Srinivasu PN, Sandhya N, Jhaveri R, Raut R. From blackbox to explainable AI in healthcare: existing tools and case studies. Mob Inf Syst. 2022;2022:1–20. https://doi.org/10.1155/2022/8167821.
21. McAdams-DeMarco MA, Law A, Salter ML, Chow E, Grams M, Walston J, et al. Frailty and early hospital readmission after kidney transplantation. Am J Transplantation: Official J Am Soc Transplantation Am Soc Transpl Surg. 2013;13(8):2091–5.
22. Gotlieb N, Azhie A, Sharma D, Spann A, Suo N-J, Tran J, et al. The promise of machine learning applications in solid organ transplantation. Npj Digit Med. 2022;5:89.
23. Fabreti-Oliveira RA, Nascimento E, de Melo Santos LH, de Oliveira Santos MR, Veloso AA. Predicting kidney allograft survival with explainable machine learning. Transpl Immunol. 2024;85:102057.
24. Ali H, Shroff A, Fülöp T, Molnar MZ, Sharif A, Burke B, et al. Artificial intelligence assisted risk prediction in organ transplantation: a UK Live-Donor kidney transplant outcome prediction tool. Ren Fail. 2025;47(1):2431147.
25. Peloso A, Naesens M, Thaunat O. The dawn of a new era in kidney transplantation: promises and limitations of artificial intelligence for precision diagnostics. Transpl International: Official J Eur Soc Organ Transplantation. 2023;36:12010.

Publisher's note