

Article

Personalized Prediction of Kidney Function Decline and Network Analysis of the Risk Factors after Kidney Transplantation Using Nationwide Cohort Data

Moongi Simon Hong ^{1,2,†} , Yu-Ho Lee ^{3,†}, Jin-Min Kong ⁴, Oh-Jung Kwon ⁵, Cheol-Woong Jung ⁶, Jaeseok Yang ⁷, Myoung-Soo Kim ⁸, Hyun-Wook Han ^{1,2,*,†} , Sang-Min Nam ^{1,2,9,*,†} and Korean Organ Transplantation Registry Study Group [‡]

- ¹ Department of Biomedical Informatics, CHA University School of Medicine, CHA University, Seongnam 13488, Korea; moongi@chauniv.ac.kr
- ² Institute for Biomedical Informatics, CHA University School of Medicine, CHA University, Seongnam 13488, Korea
- ³ Division of Nephrology, Department of Internal Medicine, CHA Bundang Medical Center, CHA University, Seongnam 13496, Korea; borywork@chamc.co.kr
- ⁴ Department of Nephrology, BHS Hanseo Hospital, Busan 48253, Korea; drkongj@gmail.com
- ⁵ Department of Surgery, College of Medicine, Han Yang University, Seoul 04763, Korea; ojkwon@hanyang.ac.kr
- ⁶ Department of Surgery, Korea University Anam Hospital, Seoul 02841, Korea; cwjung@korea.ac.kr
- ⁷ Department of Internal Medicine, Yonsei University College of Medicine, Seoul 03722, Korea; jcyjs@yuhs.ac
- ⁸ Department of Surgery, Yonsei University College of Medicine, Seoul 03722, Korea; ysms91@yuhs.ac
- ⁹ Department of Ophthalmology, CHA Bundang Medical Center, CHA University, Seongnam 13496, Korea
- * Correspondence: stepano7@gmail.com (H.-W.H.); forustous@cha.ac.kr (S.-M.N.)
- † These authors contributed equally to this work.
- ‡ Membership of the Group is provided in the Acknowledgments section.



Citation: Hong, M.S.; Lee, Y.-H.; Kong, J.-M.; Kwon, O.-J.; Jung, C.-W.; Yang, J.; Kim, M.-S.; Han, H.-W.; Nam, S.-M.; Korean Organ Transplantation Registry Study Group. Personalized Prediction of Kidney Function Decline and Network Analysis of the Risk Factors after Kidney Transplantation Using Nationwide Cohort Data. *J. Clin. Med.* **2022**, *11*, 1259. <https://doi.org/10.3390/jcm11051259>

Academic Editor: Giordina Barbara Piccoli

Received: 2 February 2022

Accepted: 22 February 2022

Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: We developed a machine-learning-based model that could predict a decrease in one-year graft function after kidney transplantation, and investigated the risk factors of the decreased function. A total of 4317 cases were included from the Korean Organ Transplant Registry (2014–2019). An XGBoost model was trained to predict the recipient's one-year estimated glomerular filtration rate (eGFR) below 45 mL/min/1.73 m² using 112 pre- and peri-transplantation variables. The network of model factors was drawn using inter-factor partial correlations and the statistical significance of each factor. The model with seven features achieved an area under the curve of 0.82, sensitivity of 0.73, and specificity of 0.79. The model prediction was associated with five-year graft and rejection-free survival. Post-transplantation hospitalization >25 days and eGFR ≥ 88.0 were the prominent risk and preventive factors, respectively. Donor age and post-transplantation eGFR < 59.8 were connected to multiple risk factors on the network. Therefore, careful donor–recipient matching in older donors, and avoiding pre-transplantation risk factors, would reduce the risk of graft dysfunction. The model might improve long-term graft outcomes by supporting early detection of graft dysfunction, and proactive risk factor control.

Keywords: kidney transplantation; machine learning; risk factors; graft survival

1. Introduction

Kidney transplantation is the treatment of choice for patients with end-stage kidney disease, and offers improved survival compared with dialysis [1]. Although short-term renal allograft survival has increased substantially over the past three decades, long-term graft survival has remained unsatisfactory [2–4]. Therefore, identifying patients at risk of graft loss, and timely intervention, are necessary to prevent late graft failure, and improve long-term graft function [5].

Various factors are associated with late allograft failures, and the serum creatinine level at one year after transplantation can predict long-term renal allograft survival [6–8]. Therefore, measuring serum creatinine and estimated glomerular filtration rate (eGFR) is recommended to detect renal allograft dysfunction, according to the KDIGO (Kidney Disease: Improving Global Outcomes) guideline [8]. eGFR is a simple, inexpensive, and universally available method for the assessment of graft function, and can be calculated based on age, sex, and serum creatinine [9].

Notably, prospective observational studies have reported that one-year eGFR is the best predictor of long-term graft function after kidney transplantation [10,11]. Clinical factors associated with eGFR < 65 mL/min/1.73 m² at one-year post-transplantation included an advanced age of the donor and recipient, acute rejection within one year, delayed graft function, deceased donor, and the number of human leukocyte antigen (HLA) mismatches [10]. The risk of late renal allograft failure increased gradually with a lower one-year eGFR, particularly in recipients exhibiting a one-year eGFR of less than 45 mL/min/1.73 m² [11]. Therefore, prediction of the one-year post-transplant eGFR decrease may help to improve long-term graft function and survival.

Machine learning has been used to predict acute rejection, delayed graft function, graft survival, and chronic allograft nephropathy in kidney transplantation, and has advantages over conventional statistics in terms of better performance and the identification of complex associations among predictive factors [12]. A study attempted to predict one-year eGFR in 707 recipients who underwent deceased donor kidney transplantation [13]. The best model for predicting eGFR < 45 mL/min/1.73 m² was a Gaussian support vector machine with recursive feature elimination using five features (donor age, donor death code, recipient weight, recipient height, and recipient sex) from 56 input variables. However, the model performance was unsatisfactory (area under the curve (AUC) = 0.72) [13].

We designed this study to upgrade the model predicting one-year eGFR decline by incorporating up-to-date techniques, such as the extreme gradient boosting (XGBoost) machine learning algorithm [14], automated genetic programming (tree-based pipeline optimization tool, TPOT) for XGBoost hyperparameter tuning [15], and the Boruta algorithm for feature selection [16]. XGBoost enables high-performance boosting decision tree models, and dominates classification problems on structured datasets [17]. Additionally, a factor network was constructed using model-chosen factors, and risk-control targets were explored.

2. Materials and Methods

2.1. Development and External Validation Data

The Korean Organ Transplant Registry (KOTRY) is a prospective multicenter transplantation registry, which includes more than 900 clinical variables [18]. Data of 6129 recipients who underwent kidney transplantation from April 2014 to December 2019 were collected (Figure 1). After excluding 1812 cases with missing data or outliers (donor age under 14 years, donor eGFR over 200 mL/min/1.73 m²), a total of 4317 recipients were finally enrolled in this study. All available variables were selected from the recipient and donor in the pre- and peri-transplantation period before postoperative hospital discharge, and donor–recipient relationship variables and immunologic variables (ABO incompatibility, number of HLA mismatches) were extracted (Table 1). We calculated eGFR by the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) formula, whereas the Schwartz formula was applied to donors under 19 years old (Supplementary Materials). The model outcome was set as eGFR under 45 mL/min/1.73 m² at one year after kidney transplantation, and the development data were split 8:2 between the training and testing sets, with equal proportions of the model outcome.

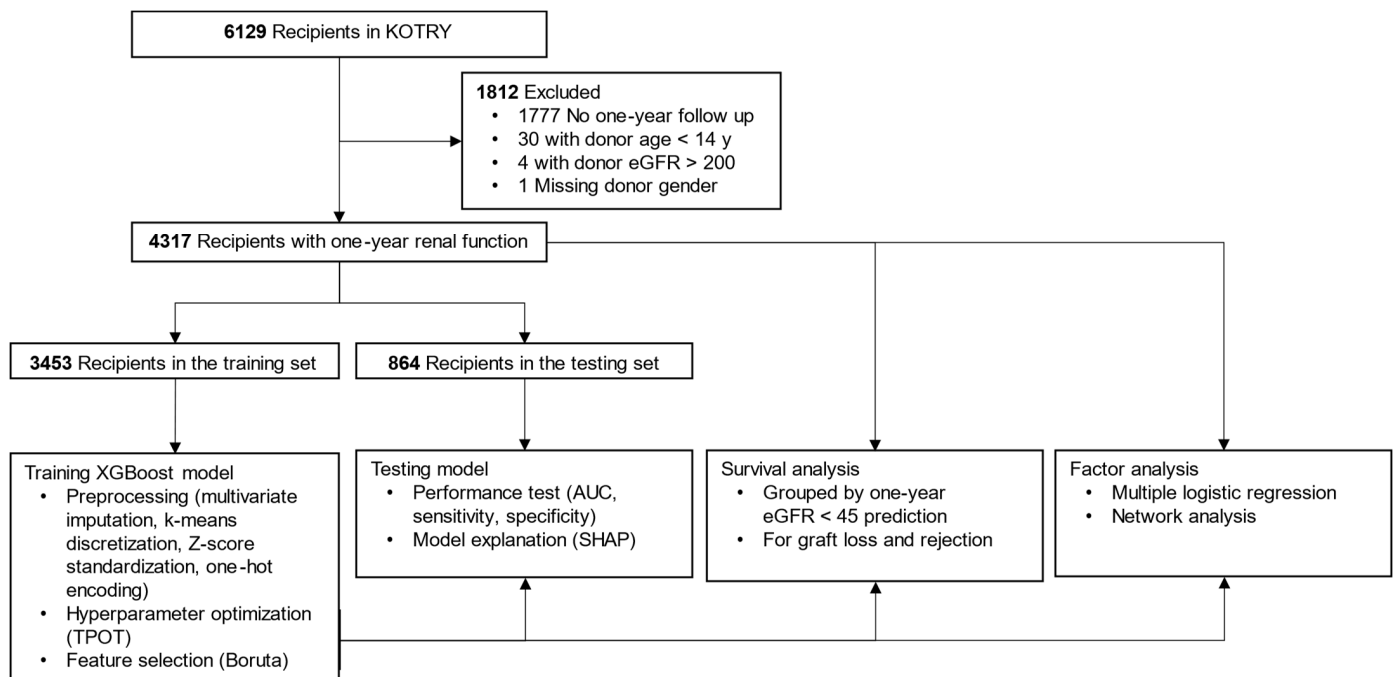


Figure 1. Flow diagram showing data processing and study methods. Abbreviations: KOTRY, Korean Organ Transplant Registry; eGFR, estimated glomerular filtration rate (mL/min/1.73 m²); XGBoost, extreme gradient boosting; TPOT, tree-based pipeline optimization tool; AUC, area under the curve; SHAP, shapley additive explanation.

Additionally, data of recipients who underwent kidney transplantation between 2010 and 2014, but were not included in the KOTRY cohort, were collected from the Bundang CHA Medical Center for external validation. The study was reviewed and approved by the Institutional Review Board of each transplantation center. The study was conducted following the Declaration of Helsinki, and adhered to the Declaration of Istanbul.

Table 1. Study variables.

Recipient Variables
General characteristics age, sex, BMI, smoking, primary renal disease, dialysis vintage, dialysis type, transplantation history, medication ^a
Comorbidities diabetes, hypertension, cardiovascular disease, tumor
Laboratory findings WBC count, neutrophil count (%), hemoglobin, hematocrit, platelet, BUN, creatinine, eGFR, uric acid, albumin, calcium ^b , phosphorus, fasting serum glucose, hsCRP ^c , PTH level ^c , total cholesterol, LDL, HDL, triglyceride
Immunology and immunosuppression ABO incompatibility; T and B cell crossmatch; HLA A, B, DR mismatch numbers; donor-specific antibody; desensitization; induction immunosuppression ^d ; maintenance immunosuppression ^e
Perioperative findings delayed graft function, creatinine at discharge ^c , eGFR at discharge ^c , post-transplantation stay ^c
Donor Variables
General characteristics age, sex, BMI, smoking, donor type ^f , cold ischemic time ^c , donation after cardiac death
Comorbidities diabetes, hypertension, cardiovascular disease, tumor

Table 1. *Cont.*

Laboratory findings WBC count ^c , neutrophil count (%) ^c , hemoglobin ^c , hematocrit ^c , platelet ^c , BUN ^c , creatinine ^c , eGFR ^c , uric acid ^c , albumin ^c , calcium ^{b,c} , phosphorus ^c , fasting serum glucose ^c , hsCRP ^c , total cholesterol ^c , proteinuria
Donor–recipient Relationship Variables age difference, sex match, height difference, weight difference, BSA ratio, viral serostatus ^g

^a Aspirin, statin, or vitamin D analog. ^b Corrected calcium = total calcium + $0.8 \times (4 - \text{albumin})$, where calcium in mg/dL, and albumin in g/dL. ^c Discretized to low, mid, high categories. ^d Basiliximab, anti-thymocyte globulin, or both. ^e Steroid, tacrolimus, cyclosporine, mycophenolate mofetil. ^f Deceased or living. ^g For hepatitis B virus, positive if donor hepatitis B surface antigen (HBsAg) is positive and recipient hepatitis B surface antibody (HbsAb) is negative. For hepatitis C virus, Epstein–Barr virus, and cytomegalovirus, positive if donor IgG serostatus is positive and recipient IgG serostatus is negative. Abbreviations: BMI, body mass index; WBC, white blood cell; HLA, human leukocyte antigen; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate; hsCRP, high-sensitivity c-reactive protein; PTH, parathyroid hormone; LDL, low-density lipoprotein; HDL, high-density lipoprotein; BSA, body surface area.

2.2. Data Preprocessing: Imputation, Discretization, One-Hot Encoding, and Standardization

Univariate and multivariate imputations were employed for missing categorical and continuous values, respectively. When the skewness or outliers needed to be addressed, the continuous variable was discretized into three categories, low, mid (reference), and high, using discretized k-means clustering (Table 1) [19]. Categorical variables with three or more unique values were one-hot encoded while dropping the reference category, and continuous variables were standardized using z-scores.

2.3. Training and Testing of the XGBoost Model

GPU-accelerated TPOT was used to tune nine hyperparameters of XGBoost (Supplementary Materials). Then, a subset of relevant features was identified using the Boruta-SHAP algorithm, and used to train a sparse model (Supplementary Materials). The performance of the sparse model was tested using metrics (AUC, sensitivity, specificity), and explained using Shapley additive explanation (SHAP) values in the test set [20]. A threshold was the Youden index for the training set, which was applied in testing and validation.

2.4. Survival Analysis of the Model Prediction

Time-to-event analyses on death-censored graft loss and biopsy-proven rejection were performed according to the model prediction for the right-censored development data (Supplementary Materials). The total numbers of cases were 4305 and 4317 for graft and rejection-free survival, respectively. Twelve patients who experienced graft loss within one year were excluded from the graft survival analysis.

2.5. Statistical and Network Analyses of Model Factors

The statistical significance of the model or known factors was calculated using multiple logistic regression for the development data. In addition to model factors, the age and sex of the donor and recipient were included as covariates.

A partial correlation matrix of factors was then generated, and a thresholded extended Bayesian information criterion (EBIC) graphical lasso network was plotted (Supplementary Materials). The network was simplified by excluding spurious connections using graphical lasso regularization with EBIC model selection [21]. The node size of the factor was calculated using the effective size of the odds ratio, and each edge weight reflected the correlation strength between two nodes after controlling all other network correlations.

2.6. Statistics and Software

Two programming languages were used: Python (version 3.7.10, Python Software Foundation, Wilmington, DE, USA) for preprocessing, machine learning, and survival analysis; and R (3.6.3, R Foundation for Statistical Computing, Vienna, Austria) for multiple logistic regression and network analysis. DeLong's confidence intervals of AUC were

calculated using the pROC package of R [22]. Detailed software usage is described in Supplementary Materials.

3. Results

3.1. Selected Model Features

The mean one-year eGFR was 70 (SD 20) (95% CI, 69–70) mL/min/1.73 m², and the proportion of recipients exhibiting one-year eGFR < 45 mL/min/1.73 m² was 11% (378/3453) in the training set. Hyperparameter optimization of XGBoost was attempted with 112 variables, and none of the variables were dropped by the regularization of XGBoost. However, a sparse model with seven variables (6% of total variables), which included donor and recipient age, low eGFR at discharge, high eGFR at discharge, serum creatinine at discharge, post-transplantation stay, and height difference between donors and recipients, was built by feature selection using the Boruta algorithm.

3.2. Model Performance and Explanation

The prediction model with seven variables showed a moderate performance, with an AUC of 0.82 (95% CI, 0.77–0.86, DeLong), sensitivity of 0.73, and specificity of 0.79 on the test data (Figure 2). External validation with 51 recipients showed similar performance, with an AUC of 0.83, sensitivity of 0.76 (13/17), and specificity of 0.71 (24/34). According to the SHAP values, donor age had the most significant impact, and eGFR at discharge ranked second (Figure 2). Several factors had elongated tails in the SHAP summary plot, specifically, an elderly donor, eGFR at discharge < 59.8 mL/min/1.73 m², and a post-transplantation hospitalization longer than 25 days increased the probability of a decline in one-year eGFR up to 0.29, 0.15, and 0.15, respectively (Figure 2). A group of high SHAP probabilities for donor age was found on the plot with a median age of 72 (range, 70–81) years and their median SHAP value of 0.27 (range, 0.24–0.29) (Figure 2). By contrast, a young donor and eGFR at discharge ≥ 88 mL/min/1.73 m² reduced the probability up to 0.23 and 0.16, respectively (Figure 2). Notably, the SHAP value distribution of the recipient age was non-linear; the probability increased in both high (over 57 years) and low (less than 29 years) recipient ages (Figure 2).

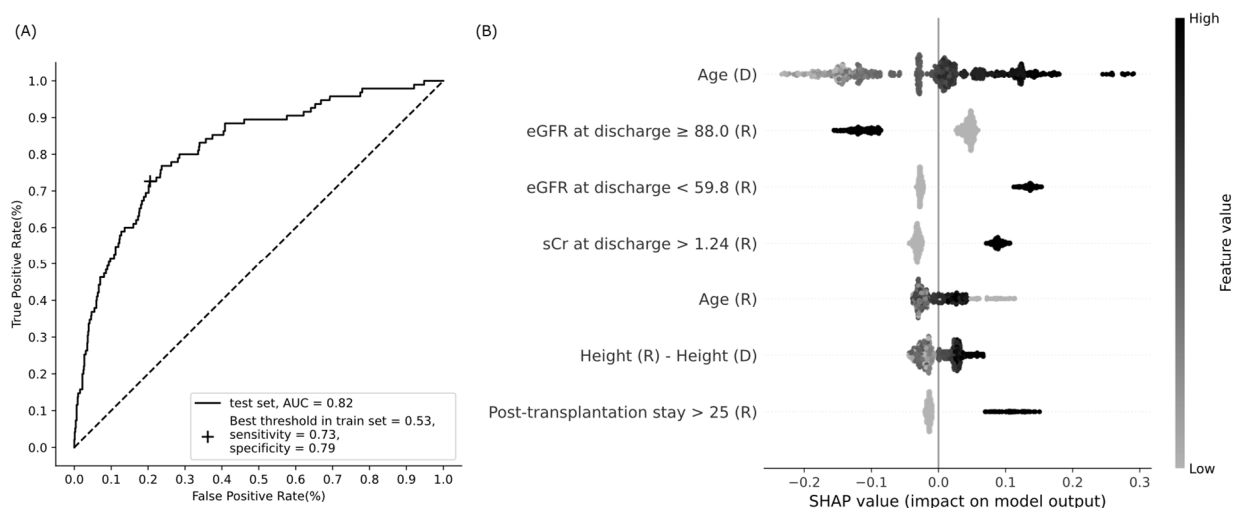


Figure 2. Performance of the XGBoost model to predict a decline in eGFR at one year after kidney transplantation. (A) Receiver operating characteristic curve; (B) summary plot of Shapley additive explanations (SHAP). Model features are sorted along the y axis of the summary plot by the sum of SHAP value (probability) magnitudes over all cases in the test data, and the distribution of each feature's impacts is plotted using SHAP values of individual cases. Binary features have 1 if present, or 0 if absent. Abbreviations: AUC, area under the curve; eGFR, estimated glomerular filtration rate (mL/min/1.73 m²); sCr, serum creatinine (mg/dL); D, donor; R, recipient.

3.3. Graft and Rejection-Free Survival for Model Prediction

The graft survival rate was significantly worse, and the cumulative incidence of rejection was significantly higher in recipients with predicted eGFR < 45 mL/min/1.73 m² than in those with predicted eGFR > 45 mL/min/1.73 m² (adjusted hazard ratios of 1.9 and 1.6, *p* = 0.02 and 0.007, respectively; Figure 3).

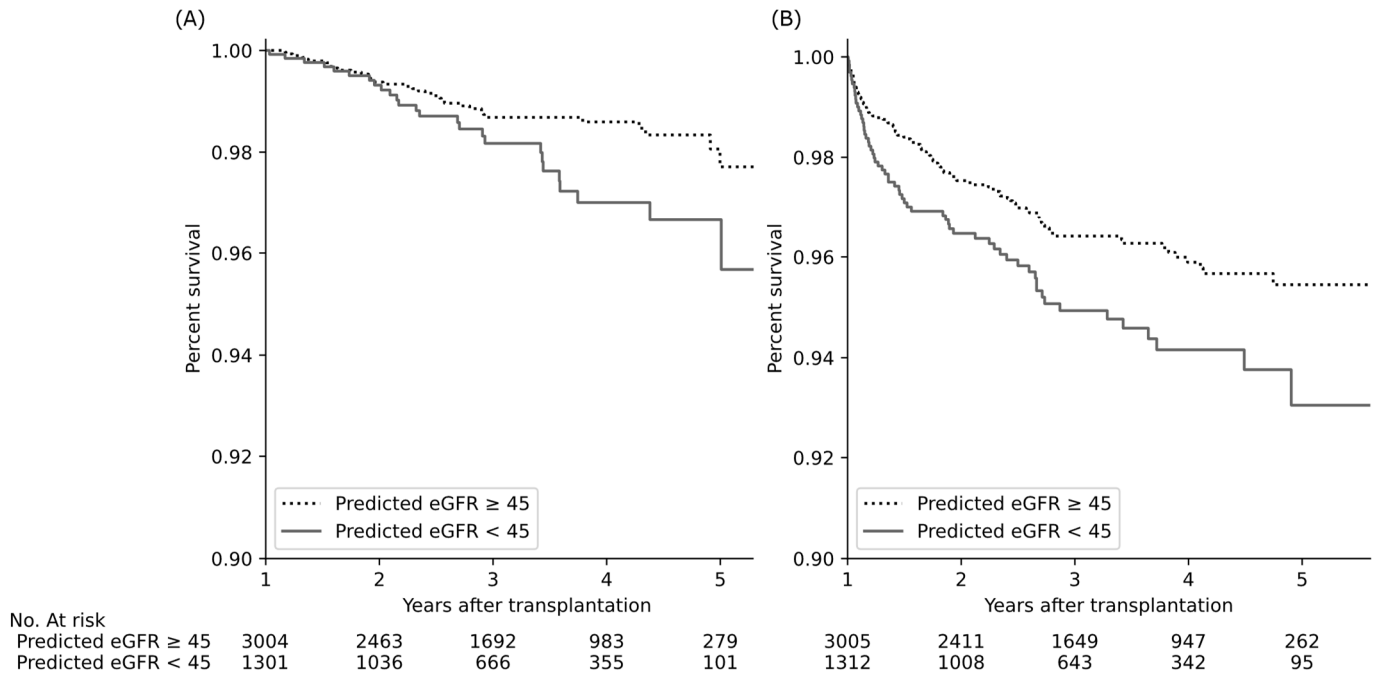


Figure 3. Kidney transplant outcomes according to predicted one-year eGFR levels of the recipient. Predicted eGFR decline was associated with a significant decrease in graft survival (A) and increase in cumulative incidence of biopsy-proven rejection (B) (adjusted hazard ratios of 1.9 and 1.6, *p* = 0.02 and *p* = 0.007, respectively, Cox proportional hazard regression with age and sex). Abbreviations: eGFR, estimated glomerular filtration rate (mL/min/1.73 m²).

3.4. Statistical Significance of Model Features

All seven model factors had a statistically significant association with a decline in one-year eGFR (Table 2). Both low and high recipient ages were significantly associated with reduced one-year eGFR when the recipient age was categorized into low (19–28 years), medium (29–57 years, reference), and high (58–76 years) groups according to SHAP value distribution (Figure 2 and Table 2).

Additionally, being a female recipient was a risk factor when the recipient’s sex was added as a covariate (Table 2). Among the reported non-model factors, that is, deceased donor, delayed graft function, and the number of HLA mismatches, the number of HLA mismatches remained a significant factor (Table 2) [10].

Table 2. Factor analysis for one-year renal allograft dysfunction (multiple logistic regression).

	eGFR ≥ 45 (N = 3453)	eGFR < 45 (N = 864)	OR (95% CI)	p Value
Categorical Factors	n (%)	n (%)		
eGFR (mL/min/1.73 m ²) ^a				
<59.8	871 (23%)	315 (67%)	2.1 (1.5–2.9)	<0.001
59.8–88.0	1725 (45%)	133 (28%)	NA	NA
≥88.0	1248 (32%)	25 (5.3%)	0.4 (0.2–0.6)	<0.001
Recipient age (year) ^b				
<29	214 (5.6%)	29 (6.1%)	2.3 (1.5–3.7)	<0.001
29–57	2755 (72%)	263 (56%)	NA	NA
>57	875 (23%)	181 (38%)	1.5 (1.2–1.9)	<0.001
Post-transplantation stay (day) ^a				
≤25	3501 (91%)	354 (75%)	NA	NA
>25	343 (8.9%)	119 (25%)	2.4 (1.8–3.2)	<0.001
Serum creatinine (mg/dL) ^a				
≤1.24	2787 (73%)	151 (32%)	NA	NA
>1.24	1057 (27%)	322 (68%)	2.0 (1.4–2.8)	<0.001
Female recipient ^c	1602 (42%)	162 (34%)	1.6 (1.2–2.2)	0.002
Deceased donor ^d	1331 (35%)	253 (53%)	1.2 (1.0–1.6)	0.08
Male donor ^c	2064 (54%)	238 (50%)	1.1 (0.8–1.5)	0.49
Delayed graft function ^d	116 (3.0%)	43 (9.1%)	1.1 (0.7–1.7)	0.68
Continuous Factors	mean (SE)	mean (SE)		
Donor Age (year)	46 (0.2)	56 (0.5)	1.9 (1.7–2.2) ^e	<0.001
Height (R)–Height (D) (cm)	−1 (0.2)	3 (0.6)	1.3 (1.1–1.5) ^e	0.007
HLA mismatch numbers ^d (range 0–6)	3.2 (0.03)	3.6 (0.07)	1.1 (1.0–1.3) ^e	0.04

^a Measured at discharge in recipients. ^b Not significant ($p = 0.39$) when a continuous factor. ^c Not model factors, but included to control sex effect. ^d Not model factors, but reported risk factors [10]. ^e Standardized odds ratio. Abbreviations: eGFR, estimated glomerular filtration rate; OR, odds ratio; CI, confidence intervals; SE, standard error; R, recipient; D, donor; HLA, human leukocyte antigen; NA, not available for a reference category.

3.5. Network Analysis of Model Predictors

The largest node was “eGFR at discharge ≥88.0”, and the most prominent risk node was “Post-transplantation stay >25” (Figure 4). In addition, two risk nodes could build up the risk effect together with other factors. The node with the highest potential was “eGFR at discharge <59.8”, which was positively connected to five risk nodes: “Post-transplantation stay >25”, “Serum creatinine at discharge >1.24”, “Donor age”, “Female recipient”, and “Recipient age > 57”, and negatively connected to a preventive factor, “eGFR at discharge ≥88.0”. Second, “Donor age” was positively associated with four risk nodes: “eGFR at discharge <59.8”, “Female recipient”, “Recipient age > 57”, and “Donor-recipient height difference”, and negatively associated with a preventive node, “eGFR at discharge ≥88.0”. Consequently, adjusted odds ratios of “eGFR at discharge <59.8” and “Donor age” were 2.1 and 1.9, respectively (multiple logistic regression, Table 2), but the crude odds ratios increased to 4.7 (95% CI, 3.8–5.8, 59.8–88.0 mL/min/1.73 m² as the reference) and 2.6 (95% CI 2.3–2.9), respectively, by the effect of associated risk factors (univariate logistic regression).

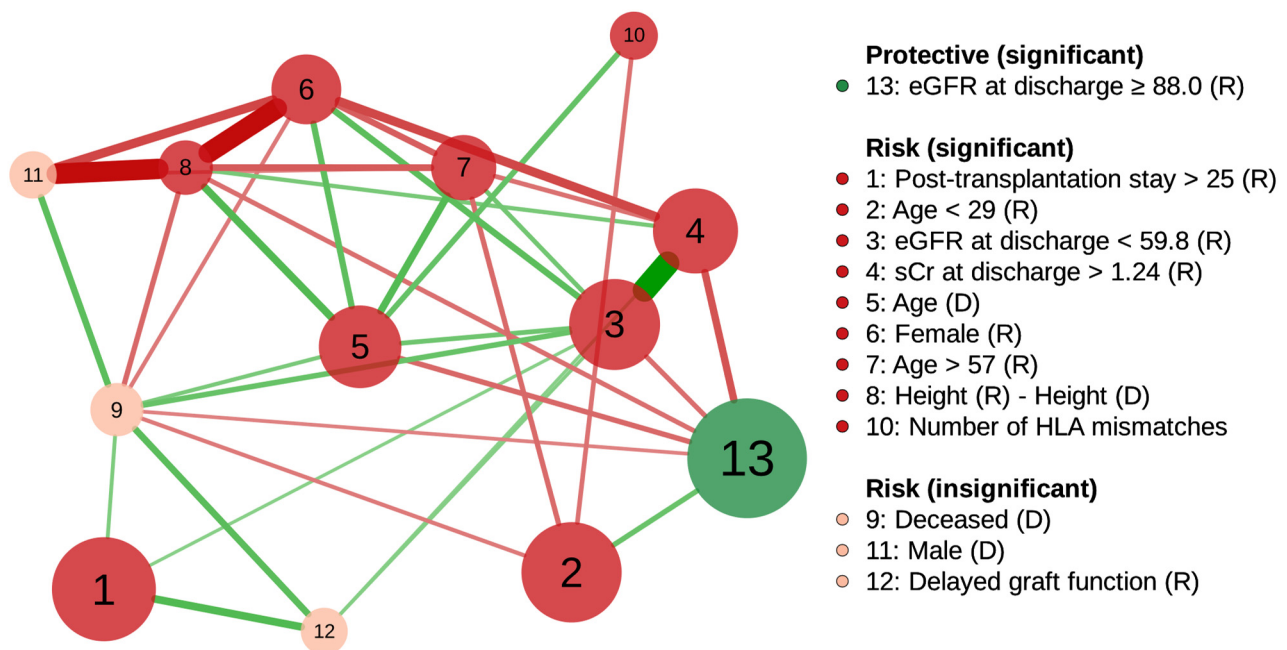


Figure 4. Partial correlation network of factors associated with one-year renal allograft dysfunction using graphical lasso regularization. Node size is proportional to the effective size of the odds ratio. The nodes are positively (risk) or negatively (protective) related to graft function decline. Green and red edges represent positive and negative correlations between the nodes, respectively. The edge with the highest absolute weight has full-color saturation and the widest width. Abbreviations: R, recipient; D, donor; eGFR, estimated glomerular filtration rate ($\text{mL}/\text{min}/1.73 \text{ m}^2$); sCr, serum creatinine (mg/dL); HLA, human leukocyte antigen.

Two insignificant risk nodes, “Deceased donor” and “Delayed graft function”, were simultaneously connected to two significant risk nodes: “Post-transplantation stay >25” and “eGFR at discharge <59.8” (Figure 4). “Deceased donor” without “eGFR at discharge <59.8”, and “Delayed graft function” without “eGFR at discharge <59.8” and “Post-transplantation stay >25” were statistically significant ($p = 0.03$ and $p = 0.049$, respectively; multiple logistic regression).

4. Discussion

4.1. Principal Findings

One-year eGFR < 45 $\text{mL}/\text{min}/1.73 \text{ m}^2$ was successfully predicted using seven factors (donor and recipient ages, low and high eGFR levels after transplantation, high serum creatinine after transplantation, height difference between the donor and recipient, and post-transplantation stay), and the predicted decline in one-year eGFR was related to long-term allograft outcomes. Two risk factors, donor age and low eGFR after transplantation, were noticeable on the network because of their multiple positive connections to other risk factors.

4.2. Improving the Performance of the XGBoost Model

A higher performance (AUC = 0.8) of the model was obtained when predicting the one-year eGFR < 45 $\text{mL}/\text{min}/1.73 \text{ m}^2$ than in the previous report (AUC = 0.7) [13]; however, the current and previous studies could not be directly compared because of the difference in development data size (4317 vs. 707), transplantation period (2014–2019 vs. 1998–2008), ethnicity (Asian vs. European), donor type (living and deceased vs. deceased only), and machine learning type (classifier vs. regressor). However, the following techniques were used to improve the model performance.

4.2.1. Automated Machine Learning

A developer configures the XGBoost hyperparameters to obtain the best performance, and the results can vary according to the developer's experience [17]. This study used the GPU-accelerated TPOT, one of the most effective automated machine learning tools. TPOT uses genetic algorithms to find the best combination of hyperparameters, and GPUs increase the chance that TPOT finds more advanced combinations by accelerating its computation time [23].

4.2.2. Addressing the Imbalanced Classification Problem

The minority class is more difficult to predict because a model may focus on learning the characteristics of the abundant cases from the majority class [24]. This study had a moderate imbalance problem of 11% of the one-year eGFR < 45 mL/min/1.73 m² (positive class), which was addressed using class-weighted XGBoost [24]. In detail, the model was eight times more likely to correct errors in the positive class, and successfully achieved balanced sensitivity and specificity.

4.3. Clinical Relevance of Discretized Factors

Some numerical variables were discretized to categorical variables, and the data-driven discretization criteria had clinical relevance. For example, eGFR < 59.8 and ≥ 88.0 mL/min/1.73 m² correspond to GFR categories in chronic kidney disease, that is, "mildly to moderately decreased" (GFR 45–59, G3 or higher category) and "normal or high" (GFR ≥ 90 , G1 category), respectively [25]. For post-transplantation stay >25 days, longer than 14 days is a risk factor of five-year mortality, and longer than 30 days was reported in 1% of recipients [26].

4.4. Factors Associated with a Decline in One-Year Renal Allograft Function

4.4.1. Donor Age, the Most Influential Factor in the Model Prediction

Donor age has been reported as the most critical factor governing graft survival and patient mortality after kidney transplantation [1]. "Donor age" had long tails in low and high SHAP values on the summary plot, which meant that the probability varied considerably with age (Figure 2). Age > 70 years required special attention because of a dramatic increase in the probability of one-year eGFR decrease in this group (Figure 2), and this finding has been consistently demonstrated [27–29].

4.4.2. eGFR at Discharge, the Most Statistically Significant Factor

The factor eGFR at discharge ≥ 88.0 mL/min/1.73 m² showed the strongest association with a one-year eGFR decrease, and eGFR at discharge <59.8 mL/min/1.73 m² was a potent risk factor (Table 2 and Figure 2). The KDIGO guideline recommends daily measurement of serum creatinine, and calculation of eGFR until hospital discharge after kidney transplantation to detect acute changes in renal allograft function, such as acute rejection, obstruction, urine leak, vascular compromise, and recurrent infection diseases, which are more common early after transplantation [30].

4.4.3. Other Significant Model Factors

"Post-transplantation stay >25" was a significant risk node, and positively connected to "eGFR at discharge <59.8" on the factor network, which means that hospitalization length might be associated with early renal allograft function (Figures 2 and 4). Prolonged hospitalization can be caused by early renal allograft dysfunction, and patients with poor fitness or health for transplantation tend to have an extended length of stay during transplant admission [1,30].

A tall recipient for a short donor was another significant risk factor (Table 2 and Figure 2). Since body height is correlated with kidney length, the results imply that a relatively small donor kidney may not be sufficient for a tall recipient [31,32]. An explanation for the kidney size mismatch problem is nephron dosing, suggesting that the

graft size should match the recipient's physiologic demand to prevent renal damage by glomerular hyperfiltration.

A recipient age under 29 years or over 57 years was significantly related to a decline in one-year eGFR (Table 2). A recipient age ≥ 60 years is a risk factor for one-year eGFR decline, and affects the five-year eGFR by lowering one-year eGFR [10]. In addition, previous studies have demonstrated that renal allograft survival is the worst in young recipients between 10 and 20 years of age [33,34]. It should also be noted that the proposed network showed that the recipient age < 29 years was associated with high eGFR at discharge, whereas the recipient age > 57 years was related to low eGFR at discharge (Figure 4). Accordingly, low and high recipient ages would have different reasons for a one-year eGFR drop. The increased risk of graft failure in young adults is partly explained by non-adherence to immunosuppressive medications, and vulnerability to graft rejection [35,36].

4.4.4. Significant Non-Model Factors

Two significant non-model factors, the female recipient and the number of HLA mismatches, were included as a covariate and known factor, respectively. Female recipients show higher graft failure risks than male recipients, especially when their donors are male, presumably because of sex-determined minor histocompatibility antigens, or the influence of sex hormones on immune activation [37,38]. HLA mismatch is a known risk factor for low one-year eGFR [10] and graft loss one-year post-transplantation [39].

4.5. Application of Factor Network for Finding Control Targets and Confounders

Two significant risk nodes, "eGFR at discharge < 59.8 " and "Donor age", had multiple positive connections to other risk factors in the network (Figure 4). These connections explained the increase in the factors' unadjusted odds ratios to 2.2 and 1.4 times their adjusted odds ratios, respectively. Therefore, controlling these two risk factors might considerably reduce the risk of declining one-year eGFR.

For "eGFR at discharge < 59.8 ", there might be a causal relationship with pre-transplantation risk factors, such as "Donor age", "Female recipient", and "Recipient age > 57 ", which occurred temporally before "eGFR at discharge < 59.8 " (Figure 4). For example, a graft from an older donor reduced the number of viable nephrons, and led to early graft dysfunction [40].

"Donor age" was positively connected to "Female recipient", "Recipient age > 57 ", "Height(R)–Height(D)", and "Number of HLA mismatches", which means that care should be taken when matching a recipient with an older donor (Figure 4). For example, better HLA matching offsets the disadvantages of the older donor in deceased donor kidney transplantation [41]. Lastly, the positive connection between "Donor age" and "eGFR at discharge < 59.8 " should be emphasized because donor age could affect the one-year eGFR and long-term outcome through low eGFR at discharge (Figures 3 and 4).

The factor network revealed that two known risk nodes, "Deceased donor" and "Delayed graft function", could be confounded by "eGFR at discharge < 59.8 " and "Post-transplantation stay > 25 " through their positive connections (Figure 4). For a deceased donor, a prolonged ischemia–reperfusion injury is usually an inevitable event affecting early kidney allograft function [42]. The injury also causes delayed graft function (see the positive connection between nodes 9 and 12 in Figure 4), which is reported to be associated with a decreased eGFR at discharge, and increased length of stay after transplantation [26,43].

4.6. Limitations

First, the relationships among the factors on the network should be considered associative, not causal. However, causal relationships between factors can be suggested if they had temporality and plausibility [44]. Second, the results may change because of the different ethnicities of the recipients and donors, and the discharge policies after transplantation. Third, there might be unfound pre-, peri-, and post-transplantation factors, such as medication adherence, that would significantly alter the one-year eGFR [13].

5. Conclusions

An XGBoost-based model successfully predicted the one-year eGFR reduction in kidney transplant recipients with good performance. In addition to the reported factors, novel pre- and peri-transplantation factors were discovered by selecting all relevant features from nationwide cohort data. Donor age and eGFR at discharge had the highest impact on model prediction, and could be primary risk factor control targets. Therefore, this study could help control one-year eGFR-associated risk factors before and after kidney transplantation, and allow physicians to plan preemptive measures when the graft function would be compromised. A practice using the proposed model and network could provide clinical benefits, as model-predicted graft dysfunction is related to long-term graft survival and rejection.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11051259/s1>; eGFR calculation, TPOT, XGBoos, feature selection using the Boruta algorithm, survival analysis, multiple logistic regression, network graph, and Scikit-learn tools [14–17,45–50].

Author Contributions: H.-W.H. and S.-M.N. conceived the study. J.-M.K., O.-J.K., C.-W.J., J.Y., M.-S.K. and Korean Organ Transplantation Registry Study Group collected and processed the data. Y.-H.L. and S.-M.N. contributed to the study design, and ensured medical accuracy. M.S.H. and S.-M.N. developed the model, and performed the analysis. All authors contributed to the study, and critically revised and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT, MSIT); S.M.N. was funded by NRF-2019R1C1C1007663; H.W.H. was funded by NRF-2020R1F1A1068423; and Y.H.L. was funded by NRF-2021R1G1A1014115. KOTRY cohort was funded by the Research of Korea Centers for Disease Control and Prevention Agency grant funded by the Korean government (Ministry of Health and Welfare, MOHW): 2014-ER6301-00, 2014-ER6301-01, 2014-ER6301-02, 2017-ER6301-00, 2017-ER6301-01, and 2017-ER6301-02.

Institutional Review Board Statement: The study was reviewed and approved by the Institutional Review Board of each transplantation center. The study was conducted following the Declaration of Helsinki, and adhered to the Declaration of Istanbul.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon reasonable research proposal to the Korean Organ Transplantation Registry (<https://www.kotry.org/eng/>, (accessed on 24 February 2022)).

Acknowledgments: The KOTRY was established in 2014, and has managed nationwide cohorts of kidney, liver, pancreas, heart, and lung transplants. The kidney subcommittee consists of 41 researchers from 39 transplant centers. The names of the KOTRY participating hospitals and principal investigators of the kidney subcommittee are as follows: BHS Hanseo Hospital (Jin Min Kong), Han Yang University Hospital (Oh Jung Kwon), Korea University Anam Hospital (Myung-Gyu Kim, Cheol-Woong Jung), Wonju Severance Christian Hospital (Sung Hoon Kim), Inje University Busan Paik Hospital (Yeong Hoon Kim), Bongseng Memorial Hospital (Joong Kyung Kim), Kyungpook National University Hospital (Chan-Duck'Kim), Bucheon St. Mary's Hospital (Ji Won Min), Chonbuk National University Hospital (Sung Kwang Park), Gil Medical Center (Yeon Ho Park), Ajou University Hospital (Inwhee Park), Samsung Medical Center (Park Jae Berm), Konkuk University Hospital (Jung Hwan Park), Yeungnam University Hospital (Jong-Won Pa'k), Eunpyeong St. Mary's Hospital (Tae Hyun Ban), Pusan National University Hospital (Sang Heon Song), Ewha Womans University Medical Center (Seung Hwan Song), Kosin University Gaspel Hospital (Ho Si' Shin), Seoul St. Mary's Hospital (Chul Woo Yang), Incheon St. Mary's Hospital (Hye Eun Yoon), Chungnam National University Hospital (Kang Wook Lee), Maryknoll Medical Center (Dong Ryeol Lee), Pusan National University Yangsan Hospital (Dong Won Lee), Kangdong Sacred Heart Hospital (Samuel Lee), Kyung Hee University Hospital at Gangdong (Sang-Ho Lee), CHA Bundang Medical Center (Yu Ho Lee), SNU Boramae Medical Center (Jung Pyo Lee), Myongji Hospital (Jeong-Hoon Lee), Soonchunhyang University Seoul Hospital (Jin Seok Jeon), Inje University Ilsan Paik Hospital (Heungman Jun), Kyung

Hee University Hospital (Kyunghwan Jeong), Ewha Womans University Mokdong Hospital (Ku Yong Chung), Ulsan University Hospital (Hong Rae Cho), Gangnam Severance Hospital (Man Ki Ju), Seoul National University Bundang Hospital (Dong-Wan Chae), Chonnam National University Hospital (Soo Jin Na Choi), Asan Medical Center (Duck Jong Han), Keimyung University School of Medicine (Seungyeup Han), Severance Hospital (Jaeseok Yang, Kyu Ha Huh), Seoul National University Hospital (Curie Ahn).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patzer, R.E.; Hamoda, R.E.; Knechtle, S.J. Results of Renal Transplantation. In *Kidney Transplantation—Principles and Practice*, 8th ed.; Elsevier: Philadelphia, PA, USA, 2019; pp. 684–708.
2. Hariharan, S.; McBride, M.A.; Cherikh, W.S.; Tolleris, C.B.; Bresnahan, B.A.; Johnson, C.P. Post-transplant renal function in the first year predicts long-term kidney transplant survival. *Kidney Int.* **2002**, *62*, 311–318. [[CrossRef](#)] [[PubMed](#)]
3. Lamb, K.E.; Lodhi, S.; Meier-Kriesche, H.U. Long-term renal allograft survival in the United States: A critical reappraisal. *Am. J. Transplant.* **2011**, *11*, 450–462. [[CrossRef](#)] [[PubMed](#)]
4. Coemans, M.; Susal, C.; Dohler, B.; Anglicheau, D.; Giral, M.; Bestard, O.; Legendre, C.; Emonds, M.P.; Kuypers, D.; Molenberghs, G.; et al. Analyses of the short- and long-term graft survival after kidney transplantation in Europe between 1986 and 2015. *Kidney Int.* **2018**, *94*, 964–973. [[CrossRef](#)] [[PubMed](#)]
5. First, M.R. Renal function as a predictor of long-term graft survival in renal transplant patients. *Nephrol. Dial. Transplant.* **2003**, *18* (Suppl. 1), i3–i6. [[CrossRef](#)] [[PubMed](#)]
6. Pascual, M.; Theruvath, T.; Kawai, T.; Tolckoff-Rubin, N.; Cosimi, A.B. Strategies to improve long-term outcomes after renal transplantation. *N. Engl. J. Med.* **2002**, *346*, 580–590. [[CrossRef](#)] [[PubMed](#)]
7. Galichon, P.; Xu-Dubois, Y.C.; Finianos, S.; Hertig, A.; Rondeau, E. Clinical and histological predictors of long-term kidney graft survival. *Nephrol. Dial. Transplant.* **2013**, *28*, 1362–1370. [[CrossRef](#)]
8. Kasiske, B.L.; Zeier, M.G.; Chapman, J.R.; Craig, J.C.; Ekberg, H.; Garvey, C.A.; Green, M.D.; Jha, V.; Josephson, M.A.; Kiberd, B.A.; et al. KDIGO clinical practice guideline for the care of kidney transplant recipients: A summary. *Kidney Int.* **2010**, *77*, 299–311. [[CrossRef](#)]
9. Webster, A.C.; Nagler, E.V.; Morton, R.L.; Masson, P. Chronic Kidney Disease. *Lancet* **2017**, *389*, 1238–1252. [[CrossRef](#)]
10. Salvadori, M.; Rosati, A.; Bock, A.; Chapman, J.; Dussol, B.; Fritsche, L.; Kliem, V.; Lebranchu, Y.; Oppenheimer, F.; Pohanka, E.; et al. Estimated one-year glomerular filtration rate is the best predictor of long-term graft function following renal transplant. *Transplantation* **2006**, *81*, 202–206. [[CrossRef](#)]
11. Schnitzler, M.A.; Lentine, K.L.; Axelrod, D.; Gheorghian, A.; You, M.; Kalsekar, A.; L'Italien, G. Use of 12-month renal function and baseline clinical factors to predict long-term graft survival: Application to BENEFIT and BENEFIT-EXT trials. *Transplantation* **2012**, *93*, 172–181. [[CrossRef](#)]
12. Senanayake, S.; White, N.; Graves, N.; Healy, H.; Baboolal, K.; Kularatna, S. Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models. *Int. J. Med. Inform.* **2019**, *130*, 103957. [[CrossRef](#)] [[PubMed](#)]
13. Lasserre, J.; Arnold, S.; Vingron, M.; Reinke, P.; Hinrichs, C. Predicting the outcome of renal transplantation. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 255–262. [[CrossRef](#)] [[PubMed](#)]
14. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
15. Olson, R.S.; Moore, J.H. TPOT: A tree-based pipeline optimization tool for automating machine learning. In *Automated Machine Learning*; Hutter, F., Kotthoff, L., Vanschoren, J., Eds.; The Springer Series on Challenges in Machine Learning; Springer: Cham, Switzerland, 2019; Volume 17, pp. 151–160.
16. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
17. Brownlee, J. XGBoost with Python: Gradient Boosted Trees with XGBoost and Scikit-Learn. *Machine Learning Mastery*, 2018; Available online: https://books.google.com.hk/books?hl=zh-CN&lr=&id=HgmqDwAAQBAJ&oi=fnd&pg=PP1&dq=XGBoost+with+Python:+Gradient+Boosted+Trees+with+XGBoost+and+Scikit-Learn&ots=nMhHjaQeOC&sig=rIp3pYcC1oRHhQ5y0gwV9KM04Ts&redir_esc=y#v=onepage&q=XGBoost%20with%20Python%3A%20Gradient%20Boosted%20Trees%20with%20XGBoost%20and%20Scikit-Learn&f=false (accessed on 1 February 2022).
18. Yang, J.; Jeong, J.C.; Lee, J.; Kim, Y.H.; Paik, H.C.; Kim, J.-J.; Park, H.-Y.; Kim, M.S.; Ahn, C. Design and methods of the Korean organ transplantation registry. *Transplant. Direct* **2017**, *3*, e191. [[CrossRef](#)] [[PubMed](#)]
19. Guo, Q.; Lu, X.; Gao, Y.; Zhang, J.; Yan, B.; Su, D.; Song, A.; Zhao, X.; Wang, G. Cluster analysis: A new approach for identification of underlying risk factors for coronary artery disease in essential hypertensive patients. *Sci. Rep.* **2017**, *7*, 43965. [[CrossRef](#)] [[PubMed](#)]
20. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)]

21. Pereira-Morales, A.J.; Adan, A.; Forero, D.A. Network analysis of multiple risk factors for mental health in young Colombian adults. *J. Ment. Health* **2019**, *28*, 153–160. [CrossRef]
22. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [CrossRef]
23. Becker, N. Faster AutoML with TPOT and RAPIDS. Available online: <https://medium.com/rapids-ai/faster-automl-with-tpot-and-rapids-758455cd89e5> (accessed on 1 February 2022).
24. Brownlee, J. Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning. Machine Learning Mastery, 2020; Available online: https://books.google.com.hk/books?hl=zh-CN&lr=&id=jaXJDwAAQBAJ&oi=fnd&pg=PP1&dq=Imbalanced+Classification+with+Python:+Better+Metrics,+Balance+Skewed+Classes,+Cost-Sensitive+Learning%3B&ots=CfKBfMVWYO&sig=cwCeJ2HQep_lqicBFpAxaFvONwo&redir_esc=y#v=onepage&q=Imbalanced%20Classification%20with%20Python%3A%20Better%20Metrics%2C%20Balance%20Skewed%20Classes%2C%20Cost-Sensitive%20Learning%3B&f=false (accessed on 1 February 2022).
25. Deceased, G. Chapter 1: Definition and classification of CKD. *Kidney Int. Suppl. (2011)* **2013**, *3*, 19–62. [CrossRef]
26. McAdams-DeMarco, M.A.; King, E.A.; Luo, X.; Haugen, C.; DiBrito, S.; Shaffer, A.; Kucirka, L.M.; Desai, N.M.; Dagher, N.N.; Lonze, B.E.; et al. Frailty, length of stay, and mortality in kidney transplant recipients: A national registry and prospective cohort study. *Ann. Surg.* **2017**, *266*, 1084–1090. [CrossRef]
27. Marconi, L.; Figueiredo, A.; Campos, L.; Nunes, P.; Roseiro, A.; Parada, B.; Mota, A. Renal transplantation with donors older than 70 years: Does age matter? *Transplant. Proc.* **2013**, *45*, 1251–1254. [CrossRef] [PubMed]
28. Englum, B.R.; Schechter, M.A.; Irish, W.D.; Ravindra, K.V.; Vikraman, D.S.; Sanoff, S.L.; Ellis, M.J.; Sudan, D.L.; Patel, U.D. Outcomes in kidney transplant recipients from older living donors. *Transplantation* **2015**, *99*, 309–315. [CrossRef] [PubMed]
29. Noppakun, K.; Cosio, F.G.; Dean, P.G.; Taler, S.J.; Wauters, R.; Grande, J.P. Living donor age and kidney transplant outcomes. *Am. J. Transplant.* **2011**, *11*, 1279–1286. [CrossRef] [PubMed]
30. Kasiske, B.L.; Zeier, M.G.; Chapman, J.R.; Craig, J.C.; Ekberg, H.; Garvey, C.A.; Green, M.D.; Jha, V.; Josephson, M.A.; Kiberd, B.A.; et al. Kidney Disease: Improving Global Outcomes Transplant Work, G. KDIGO clinical practice guideline for the care of kidney transplant recipients. *Am. J. Transplant.* **2009**, *9* (Suppl. S3), S1–S155. [CrossRef] [PubMed]
31. Emamian, S.A.; Nielsen, M.B.; Pedersen, J.F.; Ytte, L. Kidney dimensions at sonography: Correlation with age, sex, and habitus in 665 adult volunteers. *Am. J. Roentgenol.* **1993**, *160*, 83–86. [CrossRef]
32. Brenner, B.M.; Cohen, R.; Milford, E. In renal transplantation, one size may not fit all. *J. Am. Soc. Nephrol.* **1992**, *3*, 162–169. [CrossRef] [PubMed]
33. Andreoni, K.A.; Forbes, R.; Andreoni, R.M.; Phillips, G.; Stewart, H.; Ferris, M. Age-related kidney transplant outcomes: Health disparities amplified in adolescence. *JAMA Intern. Med.* **2013**, *173*, 1524–1532. [CrossRef] [PubMed]
34. Pankhurst, T.; Evison, F.; Mytton, J.; Williamson, S.; Kerecuk, L.; Lipkin, G. Young adults have worse kidney transplant outcomes than other age groups. *Nephrol. Dial. Transplant.* **2020**, *35*, 1043–1051. [CrossRef]
35. Foster, B.J. Heightened graft failure risk during emerging adulthood and transition to adult care. *Pediatric. Nephrol.* **2015**, *30*, 567–576. [CrossRef]
36. Bestard, O.; Sarwal, M.M. Antibody-mediated rejection in young kidney transplant recipients: The dilemma of noncompliance and insufficient immunosuppression. *Pediatr. Nephrol.* **2015**, *30*, 397–403. [CrossRef]
37. Lepeyre, F.; Dahhou, M.; Zhang, X.; Boucquemont, J.; Sapir-Pichhadze, R.; Cardinal, H.; Foster, B.J. Association of Sex with Risk of Kidney Graft Failure Differs by Age. *J. Am. Soc. Nephrol.* **2017**, *28*, 3014–3023. [CrossRef] [PubMed]
38. Tan, J.C.; Wadia, P.P.; Coram, M.; Grumet, F.C.; Kambham, N.; Miller, K.; Pereira, S.; Vayntrub, T.; Miklos, D.B. H-Y antibody development associates with acute rejection in female patients with male kidney transplants. *Transplantation* **2008**, *86*, 75–81. [CrossRef] [PubMed]
39. Foroutan, F.; Friesen, E.L.; Clark, K.E.; Motaghi, S.; Zyla, R.; Lee, Y.; Kamran, R.; Ali, E.; De Snoo, M.; Orchanian-Cheff, A. Risk factors for 1-year graft loss after kidney transplantation: Systematic review and meta-analysis. *Clin. J. Am. Soc. Nephrol.* **2019**, *14*, 1642–1650. [CrossRef] [PubMed]
40. Cravedi, P.; Ruggenti, P.; Remuzzi, G. Old donors for kidney transplantation: How old? *Gerontology* **2011**, *57*, 513–520. [CrossRef] [PubMed]
41. Foster, B.J.; Dahhou, M.; Zhang, X.; Platt, R.W.; Hanley, J.A. Relative importance of HLA mismatch and donor age to graft survival in young kidney transplant recipients. *Transplantation* **2013**, *96*, 469–475. [CrossRef] [PubMed]
42. Ponticelli, C. Ischaemia-reperfusion injury: A major protagonist in kidney transplantation. *Nephrol. Dial. Transplant.* **2014**, *29*, 1134–1140. [CrossRef] [PubMed]
43. Kasiske, B.L.; Israni, A.K.; Snyder, J.J.; Skeans, M.A.; Peng, Y.; Weinhandl, E.D. A simple tool to predict outcomes after kidney transplant. *Am. J. Kidney Dis.* **2010**, *56*, 947–960. [CrossRef]
44. Rothman, K.J.; Greenland, S. Causation and causal inference in epidemiology. *Am. J. Public Health* **2005**, *95* (Suppl. S1), S144–S150. [CrossRef]
45. Levey, A.S.; Stevens, L.A.; Schmid, C.H.; Zhang, Y.; Castro, A.F., III; Feldman, H.I.; Kusek, J.W.; Eggers, P.; Van Lente, F.; Greene, T. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **2009**, *150*, 604–612. [CrossRef]
46. Schwartz, G.J.; Munoz, A.; Schneider, M.F.; Mak, R.H.; Kaskel, F.; Warady, B.A.; Furth, S.L. New equations to estimate GFR in children with CKD. *J. Am. Soc. Nephrol.* **2009**, *20*, 629–637. [CrossRef]

47. Keany, E. BorutaShap 1.0.15. Available online: <https://pypi.org/project/BorutaShap> (accessed on 1 February 2022).
48. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
49. Lachenbruch, P.A.; Mickey, M.R. Estimation of error rates in discriminant analysis. *Technometrics* **1968**, *10*, 1–11. [[CrossRef](#)]
50. Epskamp, S.; Cramer, A.O.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* **2012**, *48*, 1–18. [[CrossRef](#)]