

Using Machine Learning to Examine Pre-Transplant Factors Influencing *De novo* HLA-Specific Antibody Development Post-Kidney Transplant

Authors:

Alex Rothwell¹, George Nita^{2,3}, Matthew Howse³, Dan Ridgway³, Abdul Hammad³, Sanjay Mehra³,
Andrew R Jones², Petra Goldsmith^{3*}

1. *Computational Biology Facility, Liverpool Shared Research Facilities, University of Liverpool, Liverpool, L69 7ZB, United Kingdom*

2. *Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, United Kingdom*

3. *Royal Liverpool University Hospital, Liverpool University Hospitals NHS Foundation Trust, Prescott Street, Liverpool L7 8XP, United Kingdom*

* Corresponding author: petra.goldsmith@liverpoolft.nhs.uk

Abstract

The development of *de novo* donor-specific antibodies (DSAs) against HLA is associated with premature graft failure in kidney transplantation. However, rates and factors influencing *de novo* DSA formation vary widely across the literature. We aimed to identify pre-transplant factors influencing the development of *de novo* HLA-specific antibodies following kidney transplantation using machine learning.

Data from 460 kidney transplant recipients at a single centre between 2009-2014 were analysed. Pre-transplant variables were collected, and post-transplant sera were screened for HLA antibodies. Positive samples were investigated using Single Antigen Bead (SAB) testing. Machine learning models (Classification and Regression Trees, Random Forest, XGBoost, CatBoost) were trained on a training set of pre-transplant data to predict *de novo* DSA formation, with and without SMOTE oversampling. Model performance was evaluated on an independent testing set using F1 scores, and feature importance was assessed using SHAP.

In the full cohort analysis, XGBoost models performed the best, with F1 scores of 0.54-0.59 without SMOTE and 0.72-0.79 with SMOTE. The strongest predictors were pre-transplant HLA antibodies, number of kidney transplants, cold ischemia time (CIT), recipient age and female gender. SHAP dependence plots showed that pre-existing HLA antibodies and past transplants increased the risk of *de novo* DSA development. In the unsensitised subgroup analysis, model performance was poor.

Machine learning models can be used to identify pre-transplant risk factors for *de novo* HLA-specific antibody development in kidney transplantation. Monitoring and risk-stratifying patients based on these factors may help guide preventive immunological strategies and recipient selection to improve long-term allograft outcomes.

Translational statement

This study identified pre-transplant risk factors for the development of *de novo* HLA-specific antibody in kidney transplantation. Monitoring and risk-stratifying patients based on these factors may help guide preventive immunological strategies and recipient selection to improve long-term allograft outcomes.

Introduction

De novo donor specific antibody (DSA) directed against HLA is associated with premature graft failure in renal transplantation^{1–11}. The clinical picture can vary: grafts may experience acute dysfunction, gradual decline, and even demonstrate no effect despite persistent DSA.

The rates of any *de novo* HLA-specific antibody (DSA and non-DSA) post-transplant vary enormously, ranging from 1.6–60%¹². There is evidence that developing *de novo* DSA, but not non-donor *de novo* HLA-specific antibody, has poorer outcomes¹³. However, Worthington et al identified that 50.9% of patients who lost their allografts produced HLA-specific antibody, compared to only 1.6% in those with a functioning allograft, and that 60% of those with graft loss and HLA-specific antibody developed antibody prior to graft loss³, not all of them donor-specific. Thus, all *de novo* HLA-specific antibody appears to be implicated in adverse outcomes in renal transplantation.

Several studies have shown that 7–11% of renal transplant recipients had circulating HLA-specific antibody but only 4–5.5% of these were post-transplant DSA^{4,5}. The overall frequency of HLA-specific antibody development in unsensitised pre-transplant patients was described as 14.7%⁹. Graft failure rates were higher in the *de novo* antibody group at 8.6% versus 3% in the group without antibodies⁹.

To understand which pre-transplant factors influence the development of *de novo* HLA-specific antibody following kidney transplant, we devised a study using a combination of univariate statistics and machine learning. Machine learning (ML) has been considered to be unhelpful in predicting

post-transplant outcomes, partly due to their “black-box” nature¹⁴. However, artificial intelligence (AI) and ML integration in enhancing perioperative care in transplantation, from donor and recipient selection to post-operative prevention of complications, and graft outcome prediction, remains a current research focus¹⁵. An increasing number of studies show promising applications of ML in developing prediction models in kidney transplantation. Ravindhran *et al.* published a recent systematic literature review and meta-analysis examining the current application of ML in kidney transplantation concluding that hybrid ML models, variations of Support Vector Machines (SVM), Random Forest (RF)-based models appear to perform the best in prediction of graft survival¹⁶. Principal component analysis (PCA)-derived algorithms can be used successfully to identify new patterns of reactivity that differ from patients’ historic HLA antibody pattern (Sn 100%, 95% CI, 73.54-100%, Sp 75%, 95% CI, 56.30%-92.54%)¹⁷. Growing evidence supports utilising ML in decision-making in kidney transplantation. Recurrent Neural Network (RNN) models can provide support in determining the most optimal donor-recipient pairing and informed-decision making¹⁸. Explainable AI (XAI) are systems that offer explanations for ML models’ outputs, aiming to improve interpretability, therefore increasing trust and confidence in predictions. This has been of particular interest in the medical field, where ML has the potential to benefit patients and health services, though with serious consequences if models perform poorly¹⁹. Our study aims to add to the body of literature supporting the integration of ML, by using XAI to identify pre-transplant variables leading to *de novo* HLA-specific antibody development.

Therefore, we attempt to describe the emergence of *de novo* HLA-specific antibody following renal transplant in an uncensored renal transplant cohort and an unsensitised subset, to examine the pre-transplant factors implicated in its development.

Materials and Methods

Data collection

Ethical approval was granted from the Health Research Authority National Research Ethics Service, study number 11/NW/0279 and the local Research and Development Department, who acted as sponsors, reference number 4049. The study participants had received a kidney transplant at the Royal Liverpool University Hospital between 2009 and 2014. Following informed consent, all samples available for each patient at 2 weeks, 1 month, 3 months, 6 months, 9 months, 12 months and subsequently yearly following transplant underwent a LabScreen® Mixed screening assay (LSM12, Lot 17) to broadly determine the presence or absence of Class I or Class II HLA-specific antibodies. The assay was carried out as per manufacturers specification and with the same methodology we have previously described^{20,21}. Samples with a positive result, defined as mean fluorescence intensity (MFI) ≥ 500 for class I and/or class II, were retained for single antigen bead (SAB) testing. For each patient, only the first positive sample for class I and class II were tested. The technique for testing with SAB was the same as that used for the LabScreen® Mixed assay, however, the threshold for a positive result was locally set at a normalised MFI of ≥ 2000 and the beads were specific for single Class I: HLA LabScreen® Single Antigen HLA Class I – Combi (lot numbers 9 and 10) and Class II: LABScreen® Single Antigen HLA Class II Group 1 (Lots 10 and 11).

Statistical Analyses

Clinical data, including recipient demographic data about all eligible participants, was collated from databases and software in the Royal Liverpool University Hospital and are listed in Table 1.

Statistical analysis was performed in Microsoft Excel. For univariate analysis, categorical variables were analysed with Chi Square tests and non-parametric continuous variables were analysed with Mann Whitney U tests. P-values were adjusted using the Benjamini-Hochberg procedure to control for family-wise error. None of the data was considered to be normally distributed. Ethnicity was

eliminated from multivariate and machine learning analysis due to the very small portion of Black, Asian and Minority Ethnic groups represented and intrinsic diversity in the subgroups precluding their fusion.

Two iterations of machine learning analysis were performed a) using the whole cohort b) using the cohort of unsensitised pre-transplant patients. The reason for performing two iterations is due to significant biases in both analyses i.e. in the total cohort, the results were heavily influenced by pre-formed HLA-specific antibody which appeared to account for almost all of the effect of the analysis, whereas in the unsensitised subset, analysis was limited in credibility by the small number of *de novo* HLA antibodies formed post-transplant. Thus, examining the same analysis from the two different cohorts aimed to provide evidence that addressed at least some of the foreseen bias encountered.

Table 1: Pre- Transplants Variables used for statistical and machine learning analyses.

Age
Gender
Ethnicity
Mode of Dialysis
Number of kidney transplants
HLA Mismatches (A, B, Cw, DR, DQ)
Type of transplant
Amino Acid Mismatch

Cold Ischaemic Time (CIT)
Induction Immunosuppression
Pre-Transplant HLA-Specific Antibody

Machine Learning

Four different classification algorithms were trained on the pre-transplant factors listed to predict whether transplant patients would or would not develop *de novo* HLA-specific antibody.

Classification and Regression Trees (CART), Random Forest (RF), CatBoost (CB) and XGBoost (XGB) are tree-based algorithms that can be used for classification. CART are binary trees that mathematically determine the best split on single features. Using the classifiers for prediction involves following the series of rules generated by the tree, e.g. if number of kidney transplants is > 0 and gender is female, then “does not develop *de novo* HLA-specific antibody post-transplant” may be predicted. RFs create an ensemble of trees, each tree is varied due to having been limited to searching over a random subset of features on a random sample of training data when generating decision rules, with the output being the class voted by most trees²². XGB is a gradient boosting algorithm that works by combining a series of weak tree classifiers which sequentially aim to further minimise the training error of the previous tree, ultimately resulting in a strong classifier²³. CB is also a gradient boosting algorithm, similar to XGB. Though it has been shown to outperform XGB at times, especially in the presence of categorical features, which CB is capable of handling natively without requiring them to be encoded numerically. For example, through the use of one hot encoding which involves creating a new column for each categorical value then assigning a binary value to represent presence or absence²⁴. RF and gradient boosting algorithms are less likely to overfit training data than a single decision tree²⁵.

Ten- fold cross-validation was used to train and measure model performance. For CART and RF, there was a 90/10 train-test split in each fold. For XGB and CB, there was a 72/18/10 train-test-validation split in each fold, as 10% was used for validation, and of the resulting 90%, 80% was used for training and the rest for testing. Hyperparameters of XGB and CB were tuned using randomised grid-search prior to model training.

Classification algorithms were trained twice, once with the training data being upsampled and balanced through the use of Synthetic Minority Oversample Technique (SMOTE), and once without. Imbalanced classes, i.e. one class being substantially less represented in a training dataset than others, can lead to poorer classifier accuracy for minority classes ²⁶. SMOTE creates synthetic data for the minority class by randomly placing samples along an imaginary line between similar minority samples, with the intention that increasing the number of training instances of a particular class will increase the classifiers sensitivity to them ²⁷. In the whole cohort, there were 115 patients that did develop *de novo* HLA-specific antibody post-transplant and 345 that did not. Therefore, SMOTE was applied to the class that did develop *de novo* HLA-specific antibody post-transplant, using synthetic data to increase the size of the class to 345, the same size as the class that did not.

Precision is the proportion of instances of a class which are identified correctly, and recall is the proportion of actual class samples which are identified correctly. F1 score is the harmonic mean of precision and recall. F1 score is a preferred evaluation metric when classes are imbalanced and was therefore calculated to evaluate model performance ²⁸. A two-way ANOVA was performed to analyse the effect of classification algorithm and whether SMOTE was used on model F1 score, with Tukey HSD. Of the models trained using the whole cohort, the top three best performing models exhibiting the highest F1 score, were chosen for further evaluation by XAI.

SHapley Additive exPlanation (SHAP) values use a game theoretic approach to calculate the importance of each features contribution to a final model prediction output. For each feature of each training data point, a SHAP value was calculated, where the SHAP value corresponds to how

the data point influenced the model output²⁹. A SHAP value of 0 has no impact on model outcome, a negative value impacted the model to predict the outcome as the patient not developing *de novo* HLA-specific antibody, whereas a positive value impacted the model to predict the outcome as the patient developing *de novo* HLA-specific antibody, while the magnitude of the value reflects the strength of this effect. SHAP summary plots and dependence plots were visually inspected to understand how the model attributed each feature's importance as the value of the feature changed.

Machine learning analysis was performed in Python (version 3.6). The sklearn package was used for CART and RF, and xgboost package for XGB. (catboost 1.0.4; sklearn 0.24.2; xgboost 1.5.1).

Data in figures and text is presented as mean \pm standard deviation (SD) unless otherwise specified.

Results

Four hundred and sixty patients were identified for inclusion in the analysis. Demographic, pre-transplant and post-transplant variables are shown in Table 2.

Table 2: Univariate Analysis for Pre-transplant Variables

Variable (N=460)	<i>De novo</i> HLA-Specific Antibody Yes (N=115)	<i>De novo</i> HLA-Specific Antibody No (N=345)	Test (p-value)	Adjusted p-value
Gender				
Male	57	230	Chi Square	<i>P</i> = 0.01
Female	58	115	<i>P</i> = 0.01	
Ethnicity				
White	106	304	Chi Square	

NR	4	21	(White versus	
Asian (non-Chinese)	1	3	Other)	
Asian (Chinese)	0	5	P = 0.23	P = 0.25
Black	1	2		
Other	3	10		
Age				
Median	48	52	Mann White U	
IQR	37.8, 56.4	40.4, 61.0	Test	
Range	18.3, 76.1	19.1, 79.8	P = 0.03	P = 0.06
Mode of Dialysis				
Pre-dialysis	20	90	Chi Square	
HD	67	171	P = 0.04	P = 0.06
CAPD	24	81		
NR	4	3		
No. of Transplants				
1	70	321	Chi Square	
2	36	21	P < 0.001	P < 0.001
3	8	2		
4	1	1		
HLA Mismatches (A, B, Cw, DR, DQ) (n)				
Median	4	5	Mann Whitney U	
IQR	2,5	3, 6	Test	
Range	0,10	0, 10	P < 0.001	P = 0.01
Type of Transplant				

Live Donor	36	130	Chi Square	
Deceased Donor	79	215	P = 0.22	P = 0.27
Class I Amino Acid Mismatch (n)				
Median	25	29	Mann Whitney U	
IQR	9, 35	19, 37	Test	
Range	0, 61	0, 63	P = 0.01	P = 0.01
Class II Amino Acid Mismatch (n)				
Median	12	15	Mann Whitney U	
IQR	0, 24.5	1, 25	Test	
Range	0, 46	0, 47	P = 0.27	P = 0.27
Cold Ischaemic Time (mins)				
Median	789	694	Mann Whitney U	
IQR	393.5, 1000.5	230, 932	Test	
Range	130, 1940	37, 1710	P < 0.001	P = 0.02
Induction Immunosuppression				
Depleting Antibody	66	136	Chi Square	
Non-depleting Antibody	49	209	P < 0.001	P = 0.01
Pre-Transplant Antibody				
Yes	80	32	Chi Square	

No	35	313	<i>P</i> =< 0.001	<i>P</i> < 0.001
----	----	-----	--------------------------	-------------------------

Subgroup Analysis

As referred to in the methods, a subset analysis was then performed of the same analysis for the cohort of patients who did not have any pre-transplant sensitisation which provided 348 patients for analysis. Univariate analysis is shown in Table 3.

Table 3: Subgroup univariate analysis for unsensitised pre-transplant patients for Pre-transplant Variables

Variable (N=348)	<i>De novo</i> HLA-Specific Antibody Yes (N=35)	<i>De novo</i> HLA-Specific Antibody No (N=313)	Test (p-value)	Adjusted p-value
Gender				
Male	22	215	Chi Square	
Female	13	98	P = 0.48	P = 0.60
Ethnicity				
White	33	276	Chi Square	
NR	1	18	(White versus	
Asian (non-Chinese)	0	2	Other)	
Asian (Chinese)	1	5	P = 0.28	P = 0.51
Black	0	2		
Other	0	10		
Age				
Median	48	52	Mann White U	
IQR	35.3, 54.7	41.0, 61.2	Test	

Range	18.3, 74	19.1, 79.8	<i>P</i> < 0.05	P = 0.30
Mode of Dialysis				
Pre-dialysis	9	82	Chi Square	
HD	12	151	P = 0.24	P = 0.51
CAPD	13	77		
NR	1	3		
No. of Transplants				
1	30	295	Chi Square	
>1	5	18	P = 0.05	P = 0.30
HLA Mismatches (A, B, Cw, DR, DQ) (n)				
Median	5	5	Mann Whitney U	
IQR	3, 5.5	3, 6	Test	
Range	0, 10	0, 10	P = 0.75	P = 0.82
Type of Transplant				
Live Donor	11	117	Chi Square	
Deceased Donor	24	196	P = 0.49	P = 0.60
Class I Amino Acid Mismatch (n)				
Median	30	29	Mann Whitney U	
IQR	19.5, 37	20, 38	Test	
Range	0, 53	0, 63	P = 0.91	P = 0.91
Class II Amino Acid Mismatch (n)				
Median	23	17		

IQR	4, 27.5	2, 25	Mann Whitney U	
Range	0, 46	0, 47	Test	P = 0.51
			P = 0.15	
Cold Ischaemic Time (mins)				
Median	659	676	Mann Whitney U	
IQR	305, 809.5	230, 922	Test	
Range	130, 1145	37, 1710	P = 0.39	P = 0.60
Induction Immunosuppression				
Depleting Antibody	17	122	Chi Square	
Non-depleting Antibody	18	191	P = 0.27	P = 0.51

Machine Learning

Total Patient Cohort Analysis

On models trained and evaluated using data from the entire patient cohort, a two-way ANOVA showed that the classification algorithm had a statistically significant effect on model F1 score ($p < 0.001$), as did using SMOTE to upsample the minority class ($p = 0.05$) (Figure 1). However, the interaction between these terms was not significant ($F(3, 72) = 0.18, p = 0.9$).

Tukey post-hoc test revealed significant pairwise differences between the F1 score of models generated by CART and RF algorithms compared to the F1 score of models generated by CB and XGB (CART vs. CB: -0.17 F1, CART vs. XGB: -0.22 F1, RF vs. CB: -0.09 F1, RF vs. XGB: -0.15). XGB models were the most accurate (with SMOTE: 0.59 ± 0.09 , without SMOTE: 0.54 ± 0.12), followed by CB

(with SMOTE: 0.55 ± 0.06 , without SMOTE: 0.48 ± 0.05), before RF (with SMOTE: 0.44 ± 0.08 , without SMOTE: 0.40 ± 0.09), and CART (with SMOTE: 0.36 ± 0.11 , without SMOTE: 0.33 ± 0.15).

Of the models trained and evaluated using entire patient cohort data, the models with the three highest F1 scores were further interrogated. Each of these models was an XGB model, one was trained using SMOTE on the minority class (F1: 0.79), the other two were trained without (F1: 0.74 and 0.72). Feature importance was calculated for each of these models using gain, the average improvement in model accuracy when a feature is included (Figure 2). The two most important features ranked by average increase in gain across the three models were number of pre-transplant HLA-specific antibodies, both Class I and Class II. SHAP dependence plots were visually inspected for the models with the three highest F1 scores to understand how the model attributed each feature's importance as the value of the feature data changed.

Pre-transplant HLA Specific Antibody

SHAP dependence plots suggest a relationship between the number of pre-transplant HLA specific antibodies, both Class I and Class II, and the SHAP value of this data (Figure 3). When any pre-transplant HLA-specific antibodies are identified i.e. the count is greater than 0, the associated SHAP values are greater than 0, indicating the model was influenced to predict that the patient did develop post-transplant HLA-specific antibodies. Inversely, when no pre-transplant HLA-specific antibodies were present, i.e. the count was 0, then the associated SHAP values were less than 0, indicating the model was influenced to predict that the patient did not develop post-transplant HLA-specific antibodies.

Number of Kidney Transplants

The number of kidney transplants the patient had undergone was the third most important feature in the top three most accurate models when ranked by gain. When the number of kidney transplants the patient had received was 1 i.e. the kidney transplant they received was their only transplant, the

SHAP value was less than 0, indicating the model was influenced to predict that these patients did not develop post-transplant *de novo* HLA-specific antibody (Figure 4). Inversely, if they had received a previous transplant, the SHAP values were greater than 0 indicating the model was influenced to predict those patients did develop *de novo* HLA-specific antibody post-transplant.

Cold Ischaemic Times

SHAP dependence plots of CIT showed a similar effect across two of the three best performing models. Figure 5 shows a dependence plot of one of the models where shorter CIT, typically seen in living donors, resulted in negative SHAP values, influencing the model to predict that the patient did not develop HLA-specific antibodies.

As CIT increased in the living donors, so did SHAP value. However, longer CIT, typically seen in cadaveric donors, also resulted in negative SHAP values as CIT increased which is somewhat unexpected. Therefore, overall, the shortest and longest CIT times resulted in influencing the models to predict that patients did not develop *de novo* HLA-specific antibody post-transplant.

Age

Dependence plots of age showed a similar effect to CIT, where both the extreme ranges of young and old were associated with the models predicting that the individual would not develop HLA-specific antibody post-transplant. The predominant effect, however, was for older adults over the age of 65 (Figure 6) who appear to develop less *de novo* HLA-specific antibody post-transplant.

Gender

Despite females being significantly more likely to develop *de novo* HLA-specific antibody post-transplant ($P=0.01$), SHAP dependence plots showed no clear patterns suggesting gender was influencing model output. Figure 7 shows the proportion of males and females who did and did not develop *de novo* HLA-specific antibody post-transplant.

Subgroup Analysis

Models trained using only the unsensitised subgroup had low F1 scores (0.15 ± 0.15 - across all models, with and without the use of SMOTE), therefore these models were not further interrogated using XAI as they could not be trusted to return a reliable prediction (Supplementary Figure S1).

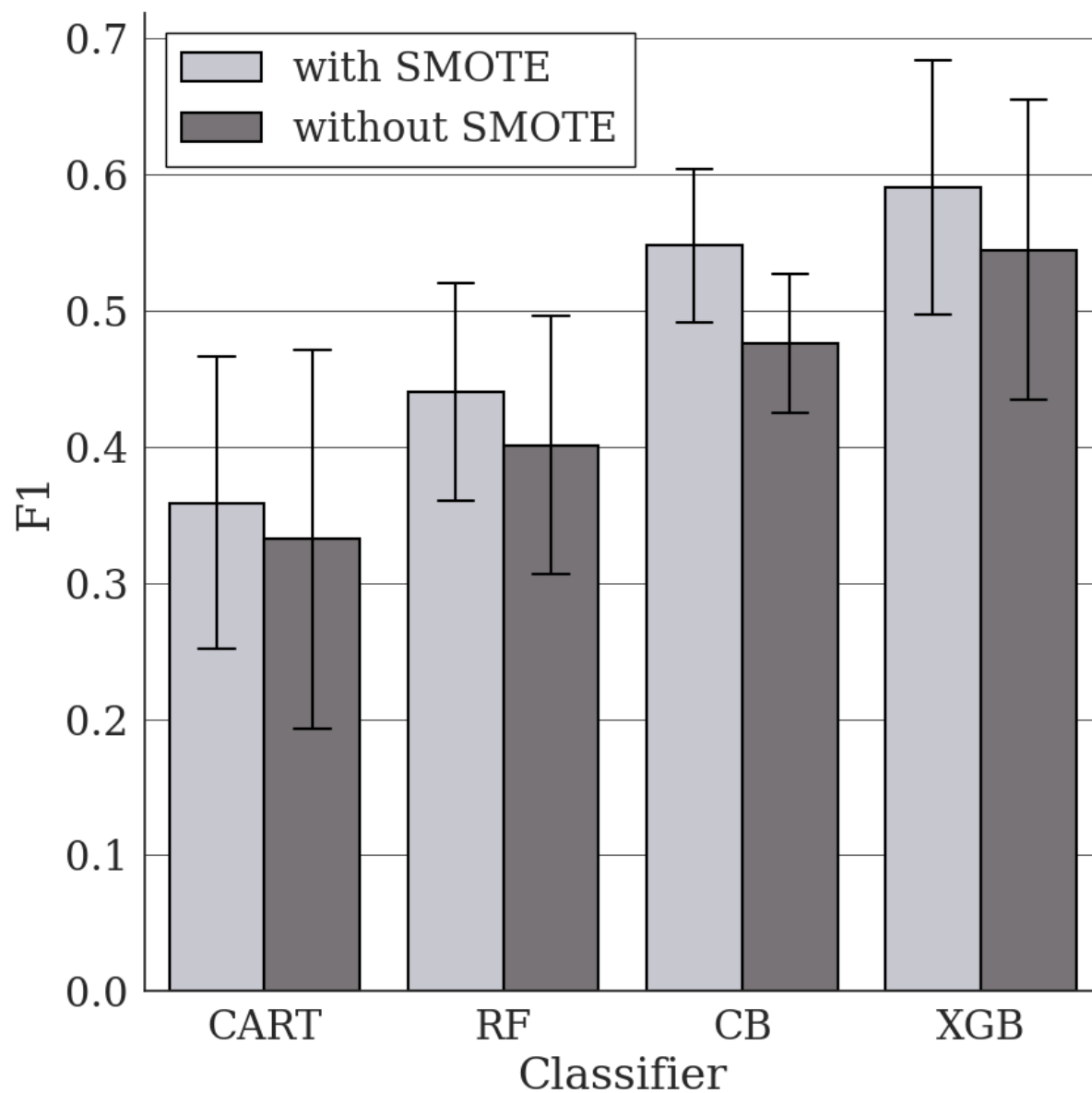


Figure 1: Mean (\pm SD) F1 score for classifier models on the whole cohort. Models were cross-validated over ten folds containing data from 460 patients. This was repeated twice, once using SMOTE during training to upsample the minority class (those who did develop *de novo* HLA-specific antibody) to the same size as the majority class, and once without.

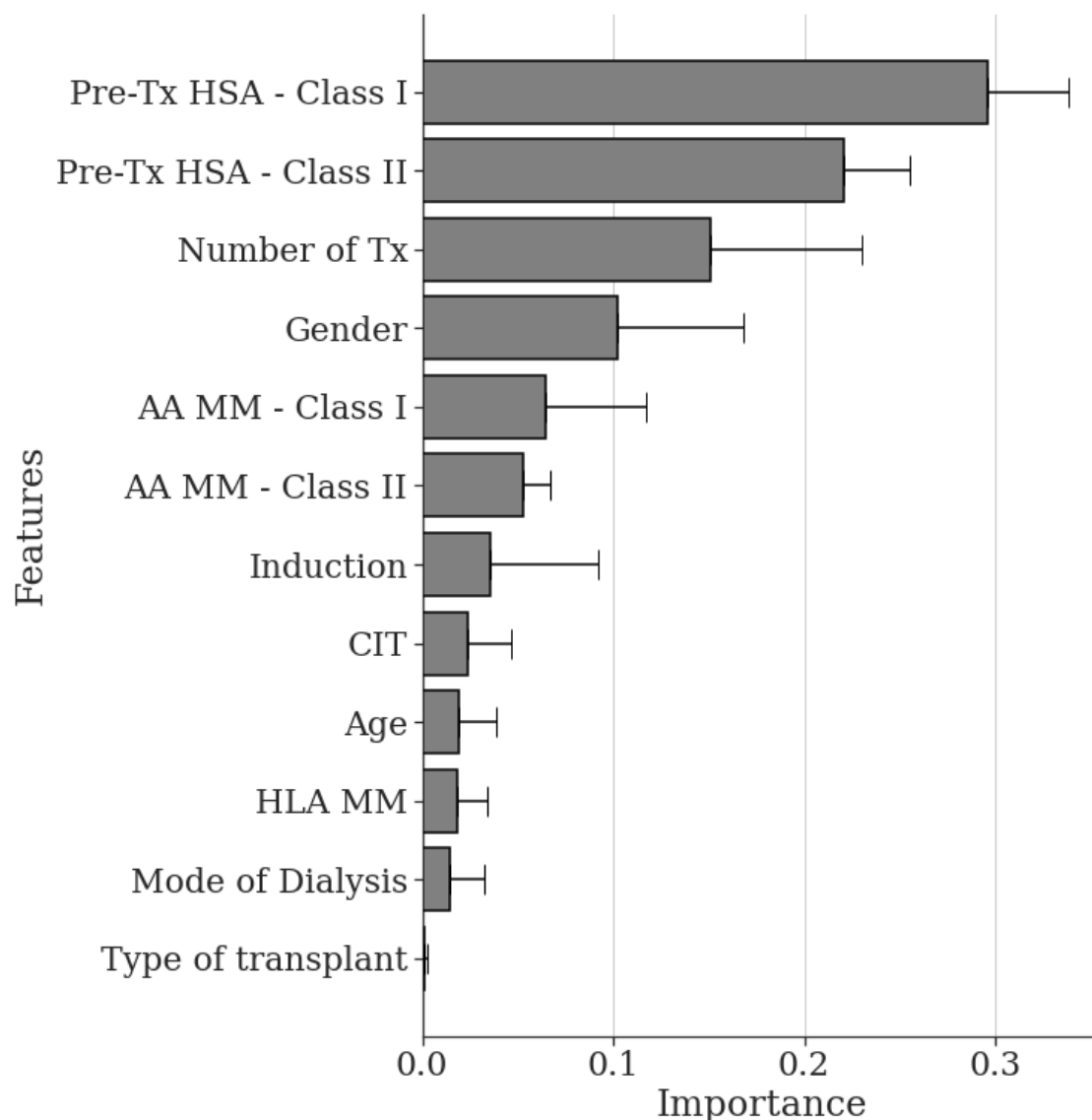


Figure 2: Mean (\pm SD) gain of each feature across the three most accurate models trained on the whole cohort. Gain is the average improvement in model accuracy when a feature is included in the model and is a measure of the importance of that feature to the model.

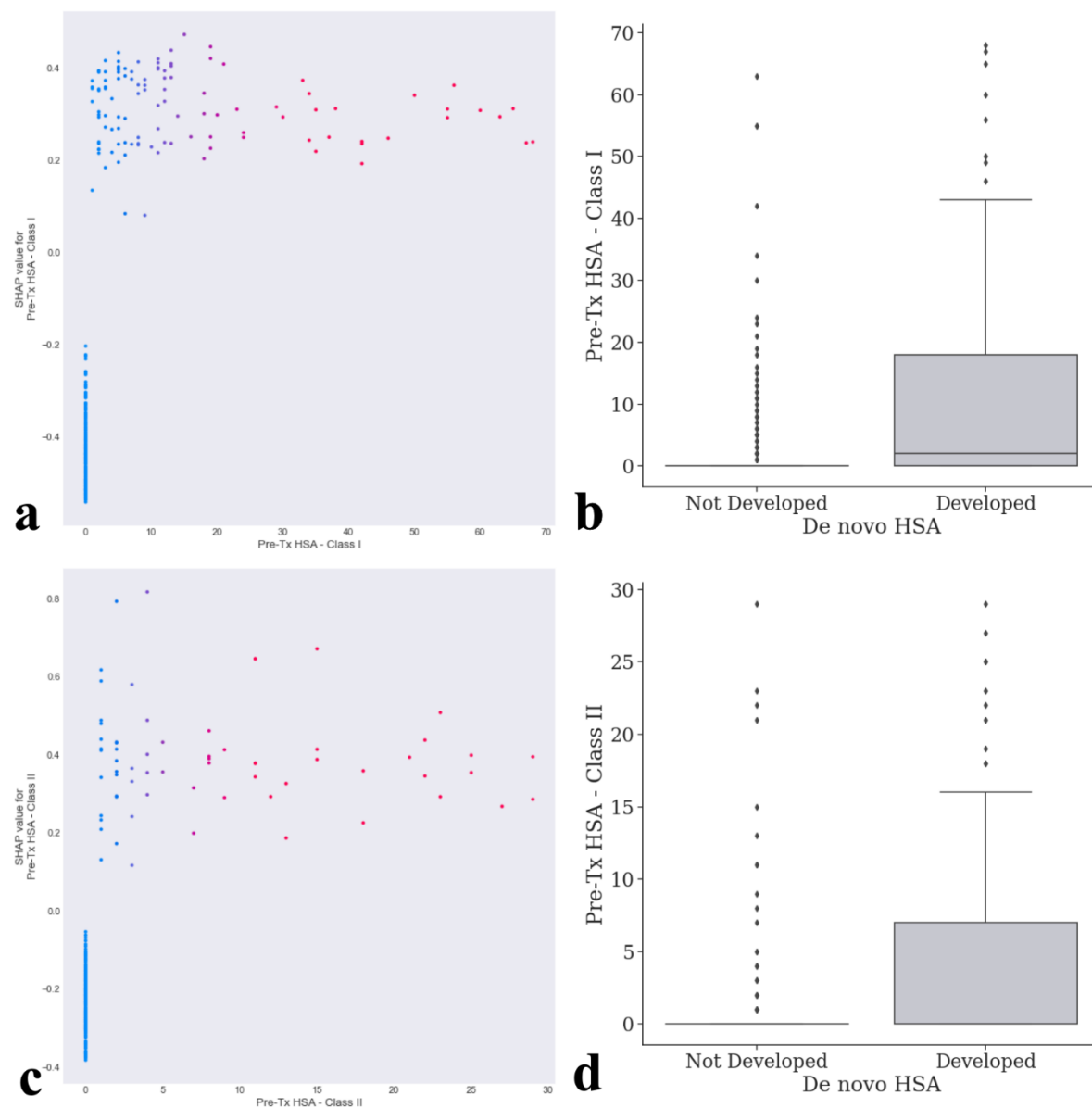


Figure 3: a and c) Example SHAP dependence plots from one of the three best performing models for Pre-transplant HLA Specific Antibody counts (Class I and Class II, respectively) vs SHAP value for Pre-transplant HLA Specific Antibody (Class I and Class II), showing how Pre-transplant HLA Specific Antibody, influences model output. SHAP values of 0, have no impact on model output, SHAP values of <0 indicate the model was influenced to predict that the patient did develop HLA-specific antibody post-transplant, however SHAP values of >0 indicate the model was influenced to predict the inverse. Only training data that was used for that model is plotted, synthetic data generated by SMOTE was not plotted. b and d) boxplots plots showing Pre-transplant HLA Specific Antibody (Class

I and Class II) data, each data point and the overall data for the whole cohort, separated by those who did and those who did not develop *de novo* HLA-specific antibody. Horizontal line in the boxplot represents the median, the box limits represent the interquartile range (IQR), whiskers represent 1.5 x IQR, with points outside this range independently plotted.

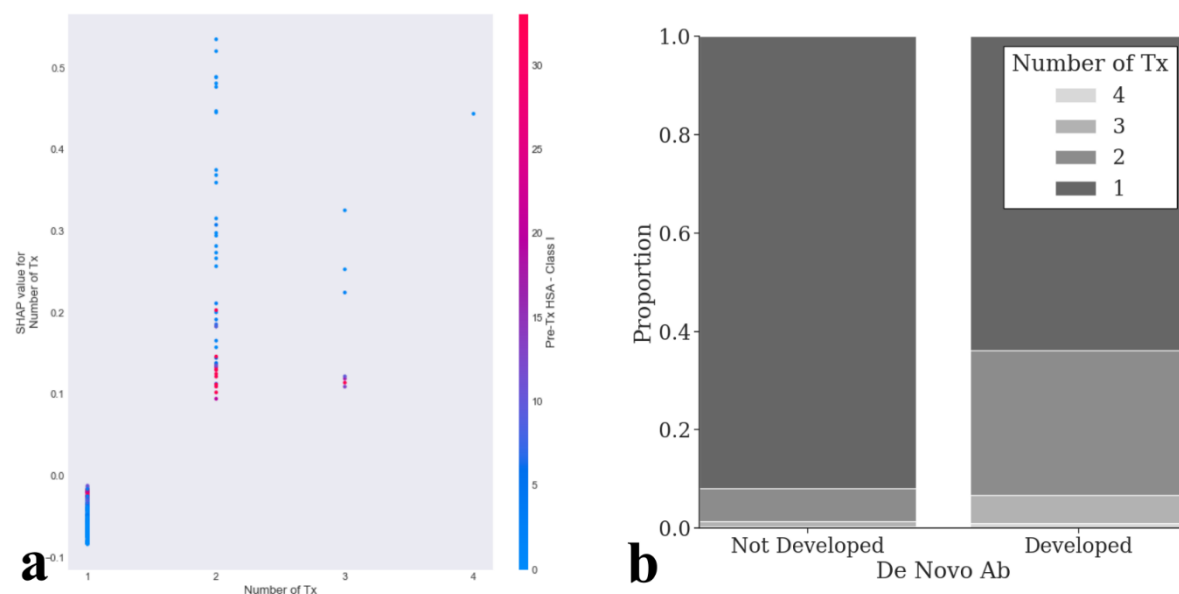


Figure 4: a) An example SHAP dependence plot from one of the three best performing models for Number of Kidney Transplants vs SHAP value for Number of Kidney Transplants, showing how Number of Transplants influences model output. SHAP values of 0, have no impact on model output, SHAP values of <0 indicate the model was influenced to predict that the patient did develop HLA-specific antibody post-transplant, however SHAP values of >0 indicate the model was influenced to predict the inverse. The point of the colour represents the number of Pre-transplant HLA specific antibody – Class I, the feature that approximately had the strongest interaction with Number of Kidney Transplants. Only training data that was used for that model is plotted, synthetic data generated by SMOTE was not plotted. b) boxplots plots showing Number of Kidney Transplants data, each data point and the overall data for the whole cohort, separated by those who did and those who did not develop *de novo* HLA-specific antibody. Horizontal line in the boxplot represents the median (in both cases 0), the box limits represent the interquartile range (IQR), whiskers represent $1.5 \times$ IQR, with points outside this range independently plotted.

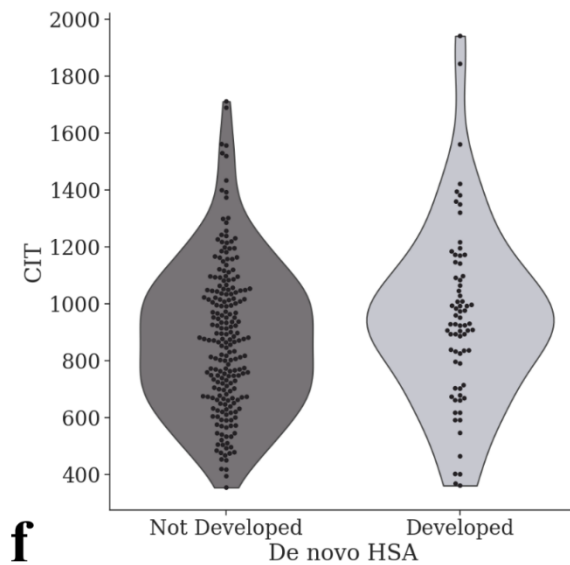
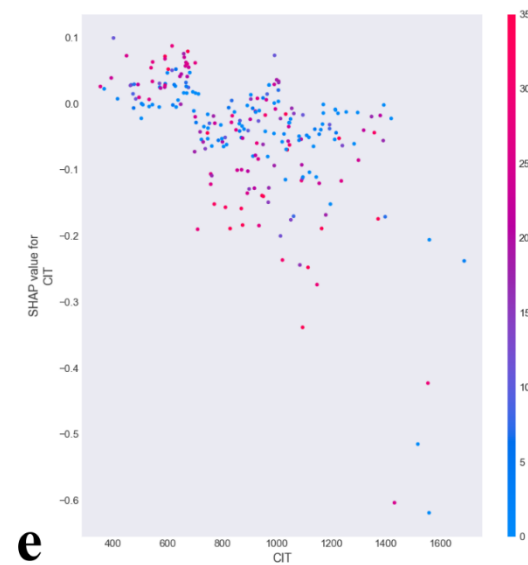
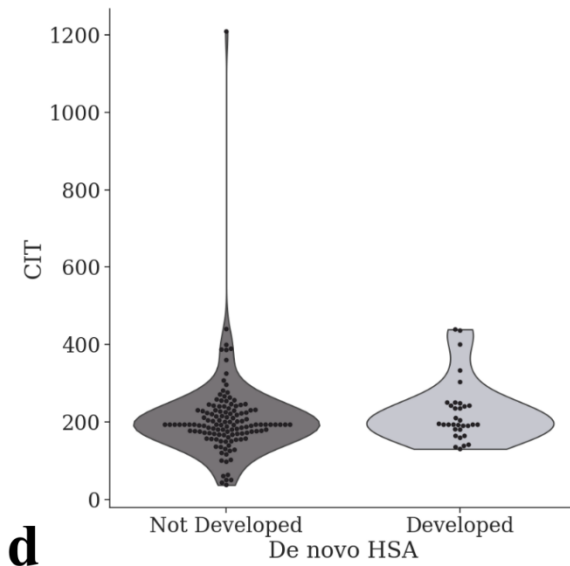
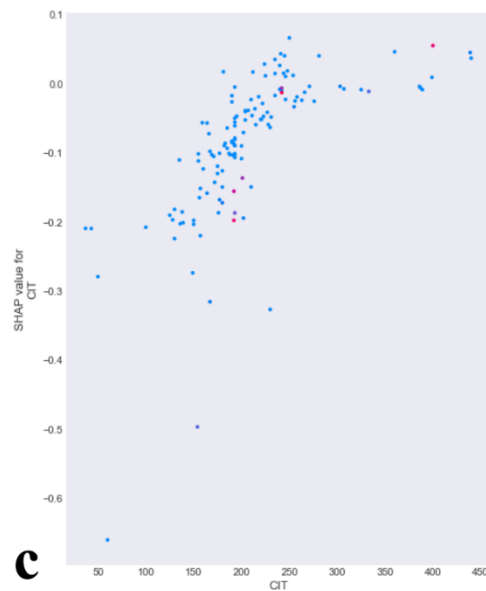
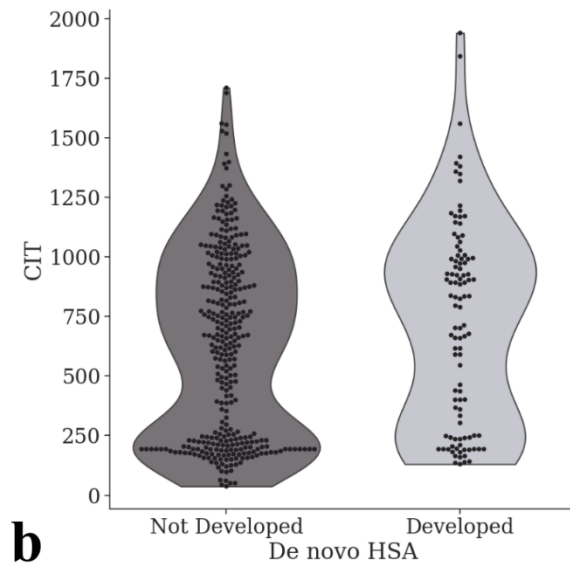
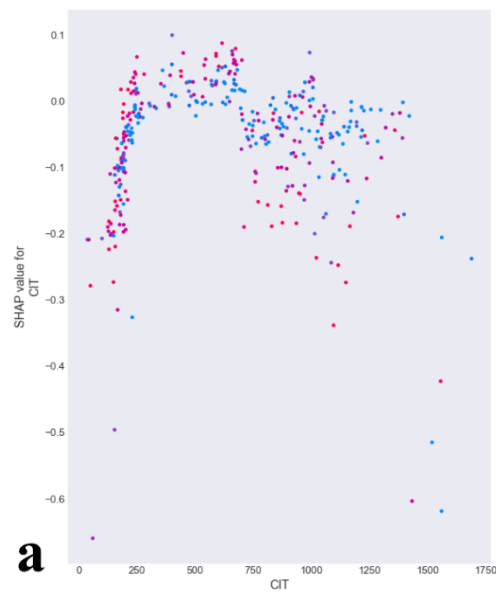


Figure 5: a) An example SHAP dependence plot from one of the three best performing models for Cold Ischaemic Time (mins) vs SHAP value for Cold Ischaemic Time (mins), showing how Cold Ischaemic Time influences model output. SHAP values of 0, have no impact on model output, SHAP values of >0 indicate the model was influenced to predict that the patient did develop HAL-specific antibody post-transplant, however SHAP values of <0 indicate the model was influenced to predict the inverse. The point of the colour represents the number of Class II Amino Acid mismatches, the feature that approximately had the strongest interaction with Cold Ischaemic Time. Only training data that was used for that model is plotted, synthetic data generated by SMOTE was not plotted. b) violin plots showing Cold Ischaemic Time data, each data point and the overall distribution of data for the whole cohort, separated by those who did and those who did not develop *de novo* HLA-specific antibody. c) and d) show the same plots as a) and b), respectively, though showing only data from living donors, whereas e) and f), again show the same plots, however with only data from cadaveric donors.

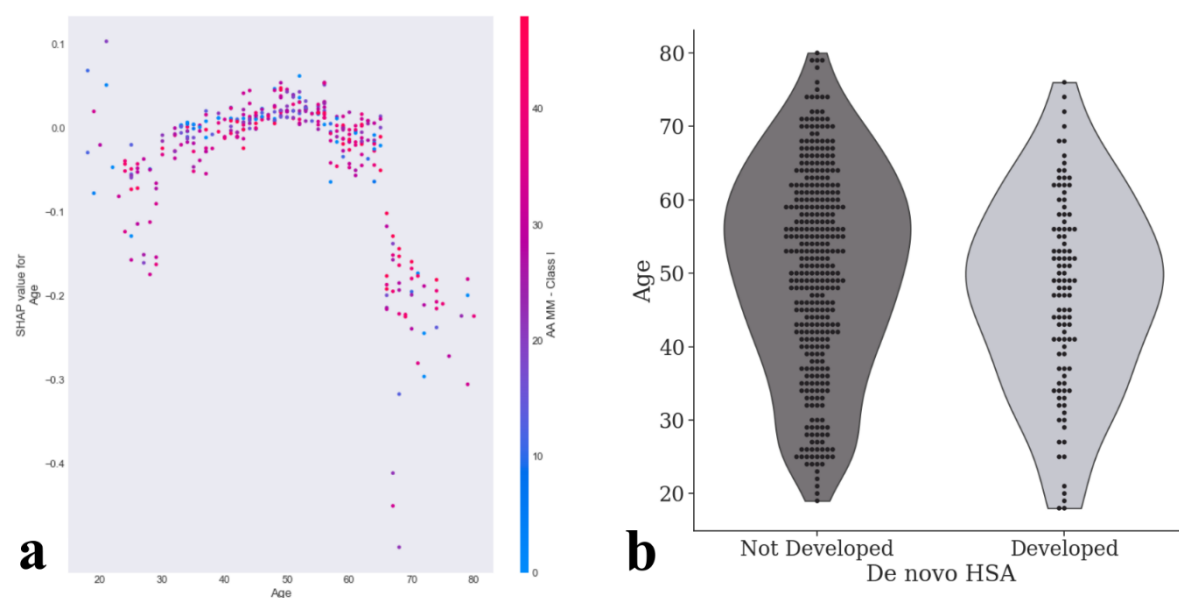


Figure 6: a) An example SHAP dependence plot from one of the three best performing models for Age vs SHAP value for Age, showing how Age influences model output. SHAP values of 0, have no impact on model output, SHAP values of >0 indicate the model was influenced to predict that the patient did develop HLA-specific antibody post-transplant, however SHAP values of <0 indicate the model was influenced to predict the inverse. The point of the colour represents the number of Class I Amino Acid mismatches, the feature that approximately had the strongest interaction with Age. Only training data that was used for that model is plotted, synthetic data generated by SMOTE was not plotted. b) violin plots showing Age data, each data point and the overall distribution of data for the whole cohort, separated by those who did and those who did not develop *de novo* HLA-specific antibody.

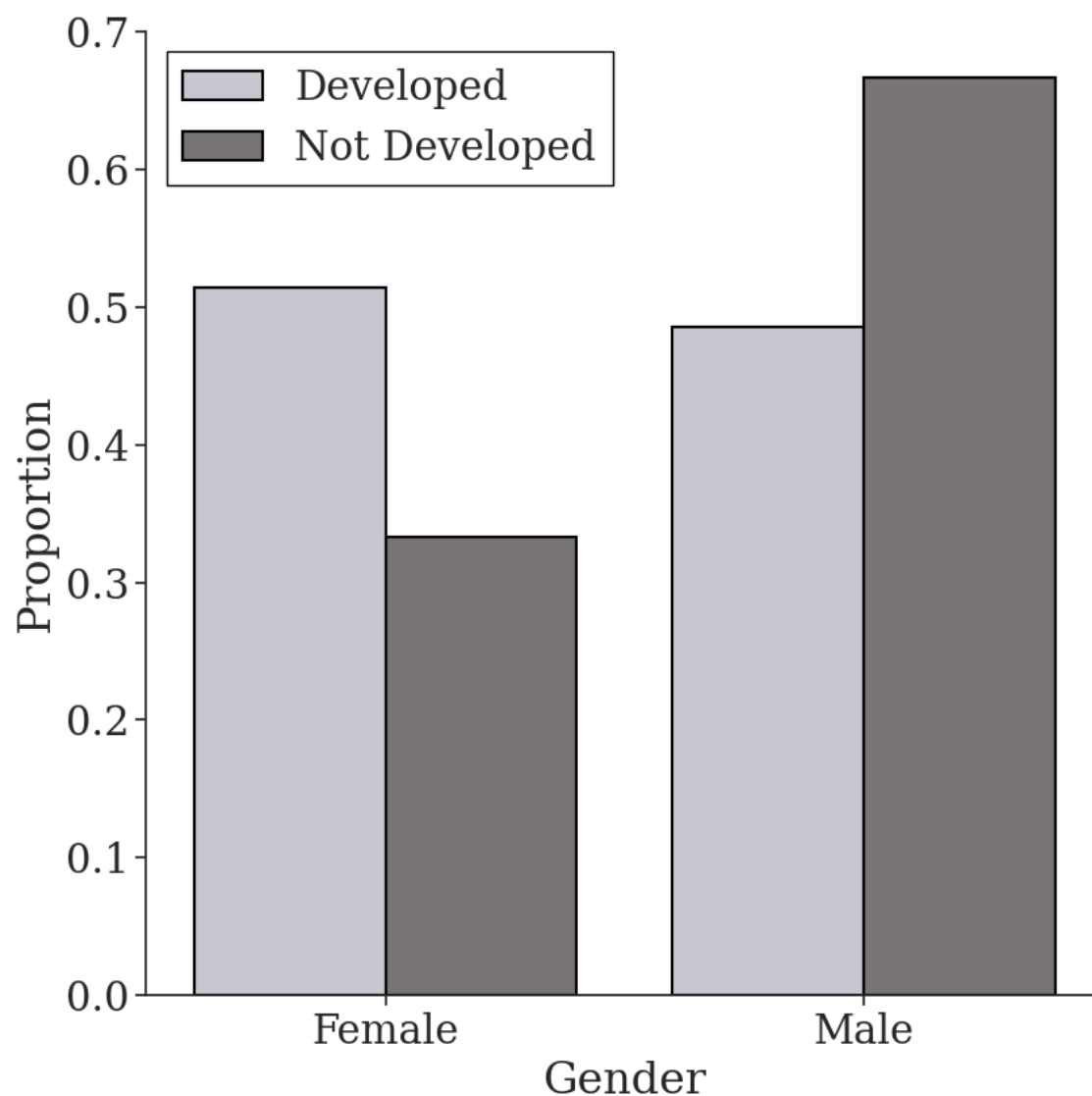


Figure 7: Proportion of patients in the whole cohort (460 patients) who did and did not develop *de novo* HLA-specific antibody, separated by gender.

Discussion

The aim of this study was to examine pre-transplant factors that affect *de novo* HLA-specific antibody development following kidney transplant. This was performed through the use of univariate statistics and interrogating trained machine learning classifier models using XAI.

The most important factor positively influencing formation of post-transplant HLA-specific antibody was the presence of pre-transplant HLA-specific antibody and was consistent across both univariate statistics and XAI outcomes. The reasons for this are multifactorial but possible explanations include the semi-quantitative nature of SAB testing and lack of reproducibility of MFIs³⁰; natural and dynamic variation in circulating HLA-specific antibody³¹ and patients who have had prior sensitising events being primed to develop more HLA-specific antibody. Machine learning models trained in the unsensitised cohort were inaccurate and unreliable, suggesting that the presence of pre-transplant HLA-specific antibody is the most reliable clinical marker.

Also appearing to generate a significant effect on the development of post-transplant HLA-specific antibody was the number of kidney transplants that a patient had received. Patients for whom the kidney transplant was their first transplant were significantly less likely to develop post-transplant HLA-specific antibody. SHAP values reflected this, with any previous transplants pushing the models to predict that the patient would develop *de novo* HLA-specific antibody. This observation is possibly unsurprising since prior transplantation is a significant sensitising event and will correlate closely with the presence of pre-transplant HLA-specific antibody. Similarly, females were significantly more likely to develop *de novo* HLA-specific antibody than males presumably due to their higher exposure to sensitising events in the context of pregnancy, childbirth and miscarriage.

A smaller, but notable, effect was observed with recipient age. Those who were older were less likely to develop post-transplant *de novo* HLA-specific antibody, though this was not significant. This is probably due to immunosenescence which is observed as a dysfunctional and less responsive immune system with advancing age³² and raises the question about the value of immunosuppression minimisation in this cohort to mitigate against side effects and infection³³.

CIT was significantly lower in the group that did not develop post-transplant HLA-specific antibody. It is a complex, bimodal variable which is illustrated in the SHAP plot (Figure 5) and from which we can make some observations. Firstly, we observe longer CIT for living donors was associated with the

model predicting post-transplant *de novo* HLA-specific antibody development. This is probably due to longer live donor CITs being observed with transplants in the kidney sharing scheme which are normally retrieved at a site distant to the implanting hospital, and indicate a degree of immunological complexity by nature of being in the scheme. Another factor increasing CIT in the context of living kidney donation is surgical complexity which could indicate a second or subsequent transplant and its association with pre-transplant HLA-specific antibody. Next, we observed that longer CIT in cadaveric donation was associated with the model predicting that post-transplant *de novo* HLA-specific antibody would not be developed. This observation is less obvious to rationalise but is most prominent at the longest CIT which may indicate a late offer that would only be accepted for an excellent graft and a straightforward recipient from the surgical and immunological perspective aiming to achieve the best possible graft outcomes.

Patients who received a non-depleting induction immunosuppressant were significantly less likely to develop post-transplant *de novo* HLA-specific antibody. In our centre, non-depleting induction therapy is only used for standard immunological risk transplants, so any high-risk transplant including re-transplants and those with pre-transplant HLA-specific antibody would receive depleting induction therapy. Thus, a depleting induction agent can be considered to be a surrogate for pre-transplant sensitisation.

Several groups have suggested that where there is serological evidence of non-DSA *de novo* HLA-specific antibody, that DSA may be bound to the transplanted kidney thereby reducing the serum load, making it undetectable on serological testing^{34–37}. In our view, an important next step would be to perform a paired serum and intra-graft biopsy study (DSA elution and histological features of rejection) to examine whether non-DSA *de novo* HLA-specific antibody is, indeed, a marker for intra-graft DSA, with or without biopsy and clinical evidence of rejection.

This paper demonstrates how the combined use of ML and SHAP presents a powerful framework for firstly, modelling complex relationships, and secondly, identifying the key factors influencing these

relationships. In the present study, the identification of the complex relationships CIT and age have with post-transplant *de novo* HLA-specific antibody formation was only possible due to the use of these methods. While caution should be exercised in interpreting these results, due to the limitations of SHAP, further discussed below, this paper demonstrates how this method enhances data interpretation beyond that possible with basic statistical tests.

Limitations

In our study, the number of patients developing HLA-specific antibody were a reasonable proportion of the total cohort, but interpretation of the unsensitised subgroup was limited due to the small numbers of *de novo* HLA-specific antibody formed in these patients. Pre-transplant HLA-specific antibody appears to be the predominant predictor of post-transplant HLA-specific antibody, but with the caveats described of assay cut-off, MFI variability and lack of reproducibility introducing bias and raising the question of what is actually *de novo*.

While univariate statistics generally supported the findings and patterns identified by SHAP dependence plots, it is worth noting that SHAP interpretations are only as useful as the accuracy of the models. To mitigate this, only the best performing models were interrogated using XAI. Secondly, analysis took place using only data derived from a single-centre and using relatively few patients. Future research should explore these relationships using a larger cohort.

Conclusion

In conclusion, we have demonstrated that pre-transplant sensitisation is the biggest predictor of post-transplant *de novo* HLA-specific antibody formation. We have reasonable evidence that female gender, induction immunosuppression and a history of previous transplants are other factors which influence post-transplant antibody formation, though our study cannot draw robust conclusions regarding DSA formation due to the small sample size with notable effects seen in age and CIT. In an unsensitised cohort, machine learning models are imprecise, but multivariate analysis indicates age,

transplant type and cold ischaemic time being predominant pre-transplant factors influencing post-transplant *de novo* HLA-specific antibody formation.

Disclosure

There are no known conflicts of interest associated with this publication.

Data Availability Statement

Code for running experiments is available on GitHub at

(https://github.com/CBFLivUni/ML_Kidney_Tx). Data may be shared upon reasonable request to the corresponding author.

References

1. Wiebe C, Gibson IW, Blydt-Hansen TD, et al. Evolution and clinical pathologic correlations of *de novo* donor-specific HLA antibody post kidney transplant. *American Journal of Transplantation*. 2012;12(5):1157-1167. doi:10.1111/j.1600-6143.2012.04013.x
2. Martin S, Dyer PA, Mallick NP, Gokal R, Harris R, Johnson RWG. Posttransplant antidonor lymphocytotoxic antibody production in relation to graft outcome. *Transplantation*. 1987;44(1):50-53. doi:10.1097/00007890-198707000-00012
3. Worthington JE, Martin S, Al-Husseini DM, Dyer PA, Johnson RWG. Posttransplantation production of donor HLA-specific antibodies as a predictor of renal transplant outcome. *Transplantation*. 2003;75(7):1034-1040. doi:10.1097/01.TP.0000055833.65192.3B
4. Hourmant M, Cesbron-Gautier A, Terasaki PI, et al. Frequency and clinical implications of development of donor-specific and non-donor-specific HLA antibodies after kidney transplantation. *Journal of the American Society of Nephrology*. 2005;16(9):2804-2812. doi:10.1681/ASN.2004121130
5. Cardarelli F, Pascual M, Tolckoff-Rubin N, et al. Prevalence and significance of anti-HLA and donor-specific antibodies long-term after renal transplantation. *Transplant International*. 2005;18(5):532-540. doi:10.1111/j.1432-2277.2005.00085.x
6. Campos ÉF, Tedesco-Silva H, Machado PG, Franco M, Medina-Pestana JO, Gerbase-DeLima M. Post-transplant anti-HLA class II antibodies as risk factor for late kidney allograft failure. *American Journal of Transplantation*. 2006;6(10):2316-2320. doi:10.1111/j.1600-6143.2006.01503.x

7. Li X, Ishida H, Yamaguchi Y, Tanabe K. Poor graft outcome in recipients with de novo donor-specific anti-HLA antibodies after living related kidney transplantation. *Transplant International*. 2008;21(12):1145-1152. doi:10.1111/j.1432-2277.2008.00755.x
8. Hidalgo LG, Campbell PM, Sis B, et al. De novo donor-specific antibody at the time of kidney transplant biopsy associates with microvascular pathology and late graft failure. *American Journal of Transplantation*. 2009;9(11):2532-2541. doi:10.1111/j.1600-6143.2009.02800.x
9. Terasaki PI, Ozawa M. Predicting Kidney Graft Failure by HLA Antibodies: A Prospective Trial. *American Journal of Transplantation*. 2004;4(3):438-443. doi:10.1111/j.1600-6143.2004.00360.x
10. Lachmann N, Terasaki PI, Budde K, et al. Anti-human leukocyte antigen and donor-specific antibodies detected by luminex posttransplant serve as biomarkers for chronic rejection of renal allografts. *Transplantation*. 2009;87(10):1505-1513. doi:10.1097/TP.0b013e3181a44206
11. Everly MJ, Everly JJ, Arend LJ, et al. Reducing de novo donor-specific antibody levels during acute rejection diminishes renal allograft loss. *American Journal of Transplantation*. 2009;9(5):1063-1071. doi:10.1111/j.1600-6143.2009.02577.x
12. Akalin E, Pascual M. Sensitization after kidney transplantation. *Clin J Am Soc Nephrol*. 2006;1(3):433-440. doi:10.2215/CJN.01751105
13. Aubert O, Loupy A, Hidalgo L, et al. Antibody-mediated rejection due to preexisting versus de Novo donor-specific antibodies in kidney allograft recipients. *Journal of the American Society of Nephrology*. 2017;28(6):1912-1923. doi:10.1681/ASN.2016070797
14. Bae S, Massie AB, Caffo BS, Jackson KR, Segev DL. Machine learning to predict transplant outcomes: helpful or hype? A national cohort study. *Transplant International*. 2020;33(11):1472-1480. doi:10.1111/tri.13695
15. Ramalheite L, Almeida P, Ferreira R, Abade O, Teixeira C, Araújo R. Revolutionizing Kidney Transplantation: Connecting Machine Learning and Artificial Intelligence with Next-Generation Healthcare—From Algorithms to Allografts. *BioMedInformatics*. 2024;4(1):673-689. doi:10.3390/biomedinformatics4010037
16. Ravindhran B, Chandak P, Schafer N, et al. Machine learning models in predicting graft survival in kidney transplantation: meta-analysis. *BJS Open*. 2023;7(2). doi:10.1093/bjsopen/zrad011
17. Cornaby C, Weimer ET. Utilizing principal component analysis in the identification of clinically relevant changes in patient HLA single antigen bead solid phase testing patterns. Chegou NN, ed. *PLoS One*. 2023;18(10):e0288743. doi:10.1371/journal.pone.0288743
18. Paquette FX, Ghassemi A, Bukhtiyarova O, et al. Machine Learning Support for Decision-Making in Kidney Transplantation: Step-by-step Development of a Technological Solution. *JMIR Med Inform*. 2022;10(6). doi:10.2196/34554

19. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. *Sensors*. 2023;23(2). doi:10.3390/s23020634
20. Goldsmith P, Lowe D, Wong C, et al. Investigating complement mediated interference in class I HLA-specific antibodies following renal transplantation. *Transpl Immunol*. 2020;62:101310. doi:10.1016/j.trim.2020.101310
21. Goldsmith P, Lowe D, Wong C, et al. Investigating the relationship between class I HLA-specific immunoglobulin-G subclasses, Pan-IgG single antigen bead assays and complement mediated interference in sera from renal transplant recipients. *Transpl Immunol*. 2020;63:101332. doi:10.1016/j.trim.2020.101332
22. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
23. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Vol 13-17-August-2016. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785
24. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. Published online October 24, 2018. Accessed August 6, 2024. <http://arxiv.org/abs/1810.11363>
25. Oguz BU, Shinohara RT, Yushkevich PA, Oguz I. Gradient boosted trees for corrective learning. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 10541 LNCS. Springer Verlag; 2017:10-18. doi:10.1007/978-3-319-67389-9_24
26. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*. 2013;14:106. doi:10.1186/1471-2105-14-106
27. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2011;16:321-357. doi:10.1613/jair.953
28. Seo S, Kim Y, Han HJ, et al. Predicting Successes and Failures of Clinical Trials With Outer Product-Based Convolutional Neural Network. *Front Pharmacol*. 2021;12. doi:10.3389/fphar.2021.670670
29. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. *Adv Neural Inf Process Syst*. 2017;2017-December:4766-4775. Accessed August 6, 2024. <http://arxiv.org/abs/1705.07874>
30. Sullivan HC, Gebel HM, Bray RA. Understanding solid-phase HLA antibody assays and the value of MFI. *Hum Immunol*. 2017;78(7-8):471-480. doi:10.1016/j.humimm.2017.05.007
31. May FNJ, Rees MT, Griffin S, Fildes JE. Understanding immunological response to desensitisation strategies in highly sensitised potential kidney transplant patients. *Transplant Rev*. 2021;35(2):100596. doi:10.1016/j.trre.2021.100596

32. Lee KA, Flores RR, Jang IH, Saathoff A, Robbins PD. Immune Senescence, Immunosenescence and Aging. *Frontiers in Aging*. 2022;3. doi:10.3389/fragi.2022.900028
33. Dreyer GJ, de Fijter JW. Transplanting the Elderly: Mandatory Age- and Minimal Histocompatibility Matching. *Front Immunol*. 2020;11:520737. doi:10.3389/fimmu.2020.00359
34. Adeyi OA, Girnita AL, Howe J, et al. Serum analysis after transplant nephrectomy reveals restricted antibody specificity patterns against structurally defined HLA class I mismatches. *Transpl Immunol*. 2005;14(1):53-62. doi:10.1016/j.trim.2005.01.001
35. Terasaki PI. Humoral theory of transplantation. *American Journal of Transplantation*. 2003;3(6):665-673. doi:10.1034/j.1600-6143.2003.00135.x
36. Martin L, Guignier F, Mousson C, Rageot D, Justrabo E, Rifle G. Detection of donor-specific anti-HLA antibodies with flow cytometry in eluates and sera from renal transplant recipients with chronic allograft nephropathy. *Transplantation*. 2003;76(2):395-400. doi:10.1097/01.TP.0000078895.24606.45
37. Lee PC, Terasaki PI, Takemoto SK, et al. All chronic rejection failures of kidney transplants were preceded by the development of HLA antibodies. *Transplantation*. 2002;74(8):1192-1194. doi:10.1097/00007890-200210270-00025

Acknowledgements

This work was supported by Kidney Research Northwest (formerly Mersey Kidney First) Grant Number 42/13