scientific reports

Check for updates

OPEN

Machine learning enhanced immunologic risk assessments for solid organ transplantation

Eric T. Weimer[1]✉ & Katherine A. Newhall[2]

The purpose of this study was to enhance the prediction of solid-organ recipient and donor crossmatch compatibility by applying machine learning (ML). Prediction of crossmatch compatibility is complex and requires an understanding of the recipient and donor human leukocyte antigen (HLA) alleles and recipient HLA antibodies. An HLA allele imputation system that converts HLA antigens to alleles was developed to enhance the prediction's performance. The imputed and known HLA alleles were combined for recipient and donor with a recipient's HLA antibody profile. After processing, donor-specific antibodies were input into various ML models. Next, an ML model was developed and characterized based on determining donor-specific antibodies using the full HLA antibody profile of the recipient without laboratory interpretation. The models achieved an ROC-AUC of 0.975. These results demonstrate that the models can predict crossmatch reactivity and yield insight into the importance of specific HLA antibodies in the transplant-matching process. These data represent our understanding of personalized histocompatibility risk assessments.

Solid-organ transplantation is a life-saving treatment for many patients with chronic diseases. Currently, more than 114,000 patients are waiting to receive an organ. Kidney transplants are the most common type of transplant worldwide, with over 27,332 performed in the United States in 2023, representing a 7.18% increase from 2022. Advances in transplantation have reduced the risk of hyperacute allograft rejection by assessing a recipient's HLA antibody status and using physical crossmatching. Physical crossmatching is a test that examines the interaction between a recipient's serum and the donor's lymphocytes. A positive reaction indicates pre-existing immunity to the donor, usually due to pre-formed anti-HLA antibodies, and is generally considered a contraindication for kidney transplant[1].

A virtual crossmatch (VXM) requires at a minimum knowledge of a recipient's HLA antibody status and HLA genotyping for both the recipient and donor[2–4]. These factors and others are used to determine the relative immunologic risk of a potential transplant pair. Current practice is to further evaluate pair compatibility by a physical flow cytometric crossmatch (XM) involving the recipient's serum and the donor's lymphocytes (T and B cells) to identify potential hyperacute allograft rejection[5–8]. Both the virtual and physical crossmatches are time-consuming, as they are largely done by hand, leading to increased ischemic organ time. Additionally, the demand for VXM has dramatically increased recently due to changes in the deceased donor service area, as a potential deceased donor is available for a larger pool of recipients[9,10]. Within the transplant community, there is active debate about the use of VXM, the need for physical crossmatching, and the use of HLA eplets in matching or organ allocation. There have been attempts to devise various models to understand the complex relationship between HLA antibodies and their respective targets on donor cells[11–15]. Previous work has shown that manual VXM correctly predicts 89–97% of flow cytometric crossmatches[7,11,16–19]. However, that number is drastically lower for highly sensitized patients[9,19]. Our previous work established that mathematical modeling could yield powerful insights into crossmatch prediction and transplant biology[11]. However, to enable a deeper understanding of transplant immunologic compatibility more advanced methods are required. For example, several studies have shown that donor HLA expression impacts physical crossmatch outcomes[20,21].

Machine learning (ML) is well suited for complex biological interactions in which patterns are not readily observable, and mechanisms are not fully characterized. Currently, most studies have applied ML to six areas of transplantation: radiological or pathological evaluation, prediction of allograft survival, immunosuppression optimization, diagnosis of allograft rejection, or prediction of early allograft function[22–25]. The immunologic complexity of transplant pairs, therefore, lends itself to ML. There are numerous interactions between a recipient's immune system and the donor organ. Some of the known interactions include HLA antibodies, HLA genotype,

[1]Department of Pathology and Laboratory Medicine, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [2]Department of Mathematics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ✉email: eric.weimer@med.unc.edu

donor HLA expression, HLA antibody avidity, and HLA eplets. By combining ML with these known biological connections, observations previously poorly understood can be identified. This is particularly important for the application of HLA antibodies in the context of pre-transplant risk assessments.

To yield deeper insights into immunologic compatibility, we have developed and evaluated a digital alloimmune risk assessment (DARA) tool that utilizes machine learning with HLA antibodies to predict physical crossmatch compatibility. As the optimal determination of donor-specific antibodies (DSA) requires high-resolution HLA typing for both the recipient and donor[26], our DARA tool includes an HLA allele imputation system to determine HLA alleles (if unknown) from HLA serologic-level typing. It also consists of the estimation of physical crossmatch outcomes using machine learning models. This approach modernizes the current standard of care approach of using curated recipient HLA antibody profiles and the HLA genotype of the potential donor.

Results

Overview of machine learning models

To develop an unbiased, independent algorithm for crossmatch prediction, we analyzed flow cytometric crossmatches (XM) performed at UNC Hospitals from January 2015 to June 2024. XM cases with unexplained results or known false reactivity were excluded from the study (see "Materials and methods" for detailed exclusion criteria).

Our HLA antibody model automatically detected donor-specific HLA antibodies (DSAs) from the complete HLA antibody profile of each recipient. The mean fluorescence intensities (MFIs) of these allele-specific DSAs were used as features for machine learning (ML) training. Multiple ML models were evaluated, with the model yielding the highest custom metric score selected for subsequent comparative analyses.

For model training, we implemented the following:

1. Data Split: 75% of the XM cases were used for training, stratified based on recipient and XM result. This approach maintained a consistent ratio of positive XM between the training and test sets (Fig. 1). This approach also prevented any data leakage based on the recipient HLA antibodies and maximized data usage.
2. Training Dataset: The HLA antibody model was trained on 9,892 XM cases, comprising of:
   - 9292 negative cases (93.93%)
   - 600 positive cases (6.07%)
3. Cross-Validation: To prevent overfitting, five-fold cross-validation was used during model fitting and while testing various hyperparameter sets. Early stopping was used where appropriate.
4. Hyperparameter Selection: The hyperparameter set yielding the highest custom metric score was selected for the final model.
5. Threshold Optimization: Using the established hyperparameters, the optimal threshold for the f1-weighted score was determined. This optimized threshold was then applied in subsequent analyses.

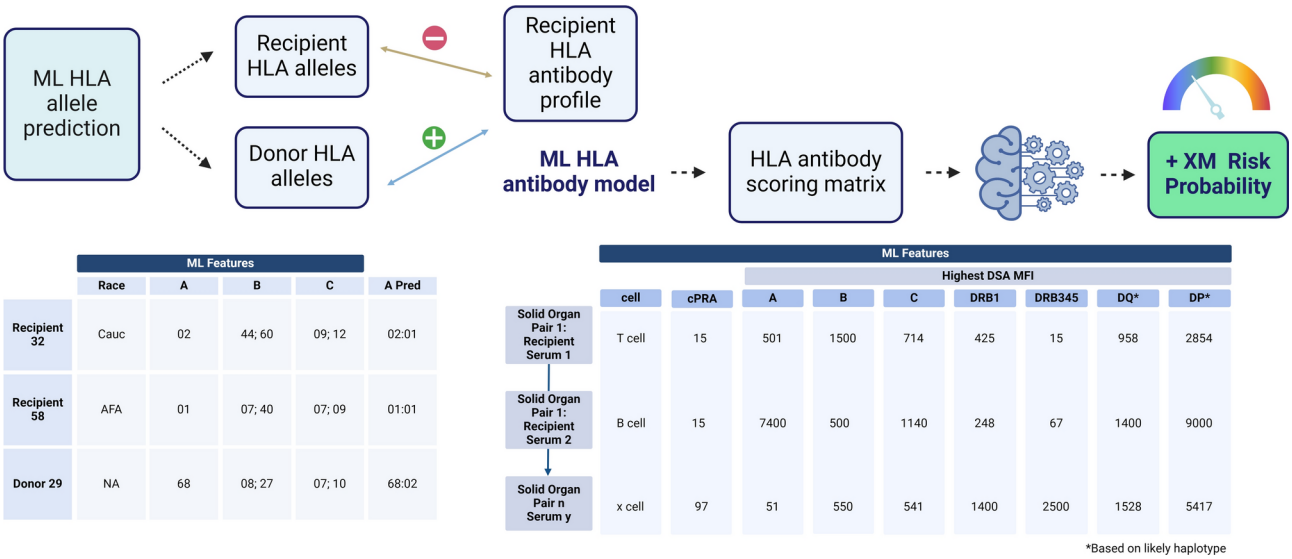

Fig. 1. Overview of machine learning (ML) models. The ML models begin with HLA allele imputation for either recipients or donors. The imputed HLA alleles are then used to determine donor-specific antibodies (DSAs) in the HLA antibody model. The HLA antibody mean fluorescent intensity (MFI) is identified using the donor-specific HLA alleles from single-antigen bead testing and used to generate an HLA antibody scoring matrix. The scoring matrix used for ML training (75% of total data) using the optimal hyperparameters that generated the highest custom scorer using five-fold cross-validation. Each trained ML model was tested with an independent dataset (25% of data) using their optimal parameters to determine the positive crossmatch risk.

Effect of HLA allele imputation on ML model performance

To enable use of historical data, HLA allele prediction ML models were developed. The models are specific to each HLA locus and yielded an internally validated average two-allele accuracy of 91.4%, greater than 99% with at least one allele accuracy (manuscript in process). To assess the impact of HLA allele-level prediction on the performance of our HLA antibody ML model, we compared model performance with and without HLA imputation. The results of this comparison are summarized in Table 1. Imputation led to a decrease in the HLA antibody model's sensitivity by 6.3 percentage points (from 78.9 to 72.6%). There was no effect on the specificity. The Brier score loss decreased from 0.059 to 0.026 with imputation, indicating an increase in overall prediction performance. These observations are consistent with previous manual assessments of the impact of recipient and donor HLA alleles on donor-specific antibody evaluation and assignment. Together, these results support that the imputation accuracy does not detrimentally affect the accuracy of the downstream ML model.

HLA antibody machine learning model characteristics

To address the imbalance in our dataset, where negative crossmatch (XM) results significantly outnumbered positive ones, we employed various sampling techniques. These techniques were applied only to the training data, while all models were evaluated using the same unmodified test data to ensure fair comparison. Synthetic Minority Oversampling Technique (SMOTE) was used to generate synthetic samples within the minority class (positive XM results). SMOTE creates new, synthetic examples in the feature space of the minority class, effectively increasing its representation in the dataset[27]. Random Undersampling randomly downsamples the majority class (negative XM cases). By reducing the number of negative XM cases, the technique aims to balance the dataset and *potentially* improve the model's ability to identify positive cases[28].

Figure 2 illustrates the performance of the models under different sampling conditions. Overall, model performance was consistent across all the sampling techniques as evident by the similar shapes of the detection error tradeoff (DET) and precision-recall curves. These data indicates that the original model was relatively robust to class imbalance.

HLA antibody-based machine learning model

Having established the overall performance of our HLA antibody-based machine learning model, we next sought to understand how individual HLA antibodies contribute to the model's predictions. To achieve this, permutation feature importance, a model-agnostic technique that evaluates the importance of machine learning features (Fig. 3A) was used. This method assesses how the model's performance decreases when a single feature (DSA to a specific HLA loci) is randomly altered, providing insight into each antibody's relative impact on the model's predictions.

It is well established that the determination of HLA antibodies using solid-phase assays can lead to over-interpretation of a recipient's reactivity. There are several overreactive beads and HLA professionals often make manual interpretations of these complex reactivity patterns. To evaluate the impact of individual HLA antibodies on ML performance, a range of specific HLA antibody MFIs were analyzed. Specific HLA antibodies were incrementally increased from their minimum to maximum MFI value in the dataset with an absolute maximum of 10,000 MFI (Fig. 3B). These simulated the presence of a single donor-specific antibody. Each HLA locus-specific antibody's impact on ML prediction was assessed. Not surprisingly, increasing the antibody MFI value increased the likelihood of a positive crossmatch prediction for most of the loci.

The effect of HLA class I antibodies on ML prediction was assessed using only T cell responses, while HLA class II antibodies were assessed using B cell responses. (Fig. 3). Consistent with previous reports regarding HLA locus-specific expression, HLA-B, HLA-A, and HLA-DRB1 had the most effect on the model's prediction further suggesting a connection between HLA protein expression and HLA antibody binding capacity (Fig. 3A)[20,21]. Of note, HLA antibodies targeting HLA-C, HLA-DP, or HLA-DRB345 did not individually impact the prediction, suggesting a reliance on other HLA antibodies (Fig. 3A,B). Consistent with the overreaction in the antibody assay, antibodies targeting HLA-DQ and -DRB1 required higher MFI values to affect the ML prediction (Fig. 3B). These observations are consistent with the over-reactivity of several HLA-DRB1 and HLA-DQ alleles in the HLA antibody assay and emphasize the need to list unacceptable HLA antibodies at the HLA antigen level. These data provide a data-driven approach to the individualized listing of unacceptable HLA antibodies. Additionally, the data demonstrate the feasibility of the models to learn representative transplant biology and aid in the understanding of global model predictions.

HLA antibody model performance

To ensure that the ML model scores were consistent with basic transplant immunobiology, we evaluated the ML model performance on T cell prediction using HLA class II antibody data. Since T cells lack HLA class II

Imputation	PPV	Sensitivity	Specificity	Brier score loss	Brier score loss per XM class	Total number of validation pairs
No	0.856	0.789	0.993	0.059	Negative XM: 0.001	2604 (5.6% + XM)
					Positive XM: 0.073	
Yes	0.858	0.726	0.993	0.026	Negative XM: 0.023	3285 (6.8% + XM)
					Positive XM: 0.027	

Table 1. Effect of HLA allele imputation on ML model performance.

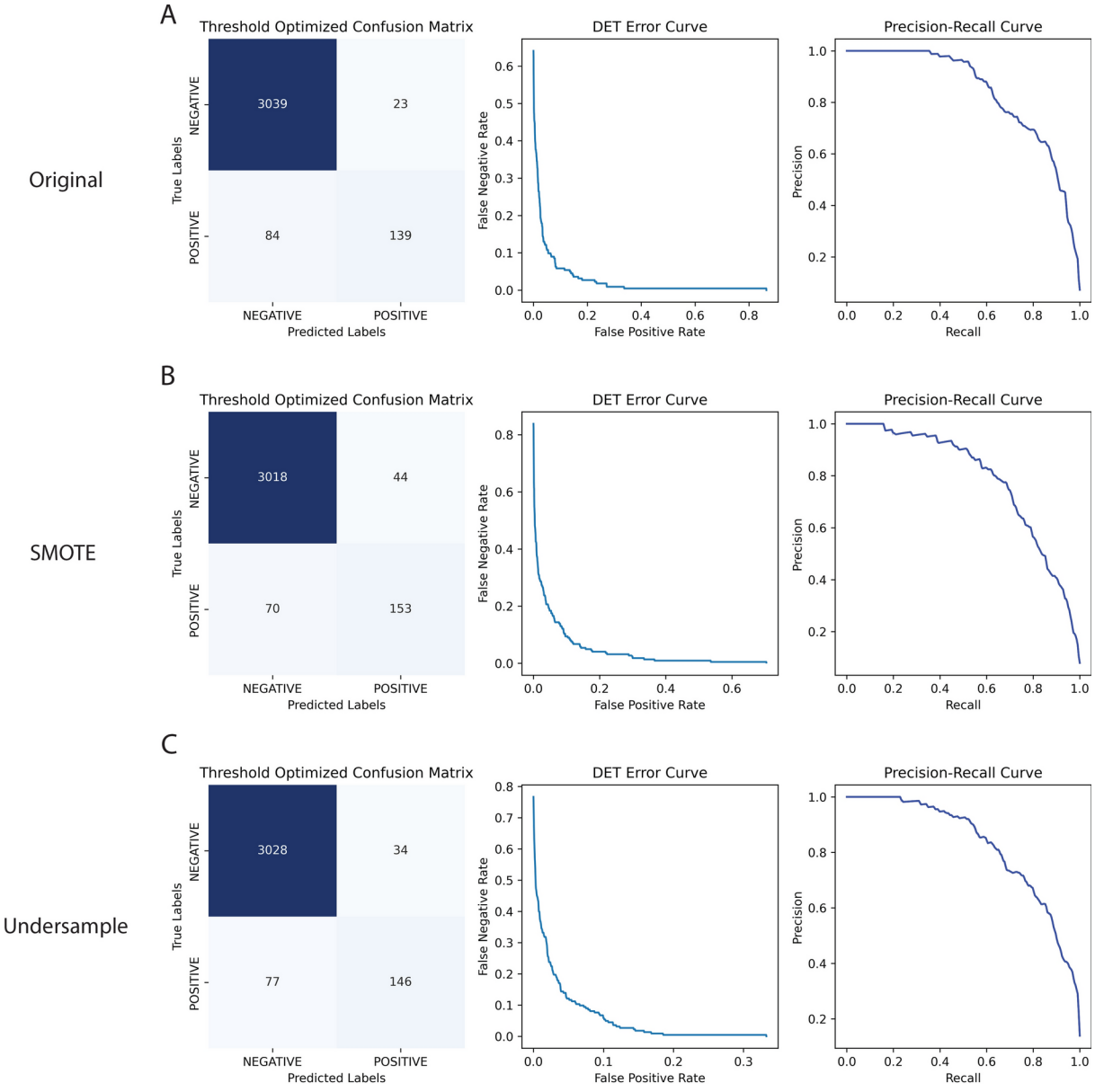

Fig. 2. Performance of the HLA antibody model. (**A**) Original unmanipulated dataset. (**B**) Using Synthetic Majority Oversampling Techniques (SMOTE) to equalize the positive XM cases. (**C**) Random Undersampling Technique. Left are confusion matrices. Center are detection error tradeoff (DET) graphs. Right are precision-recall curves. All graphs were made using the optimal hyperparameter and optimal f1-weighted threshold ML model with the independent test dataset. All sampling techniques were only applied to training data.

expression, HLA class II data should not affect model predictions. Using only HLA class II data yielded a low average positive crossmatch probability of 0.001 (data not shown).

A comparison of sensitivity (Sens), specificity (Spec), negative predictive value (NPV), positive predictive value (PPV), and likelihood ratios (LR) was performed between the current manual expert-assigned VXM, summed MFI thresholding approach[11], and the HLA antibody ML model. A summary is shown in Table 2. The current expert VXM had excellent specificity and NPV from the ease of predicting negative FCXM when no HLA antibodies are present. However, the expert VXM had low sensitivity and PPV. The ML model slightly outperformed the summed MFI thresholding approach for NPV and sensitivity. While the thresholding approach had a higher PPV, consistent with applying a conservative threshold (Table 2). Specificity was equivalent across all the approaches evaluated.

To evaluate the diagnostic ability of each model to predict crossmatch outcomes, receiver-operating-characteristic (ROC) curves and the area under the curve (AUC) were determined (Fig. 4). The AUC for the HLA antibody ML model was 0.975 compared to the manual VXM AUC of 0.602 (Fig. 4). Importantly, all the techniques (SMOTE and Undersampling) to address the data imbalance did not significantly alter the model

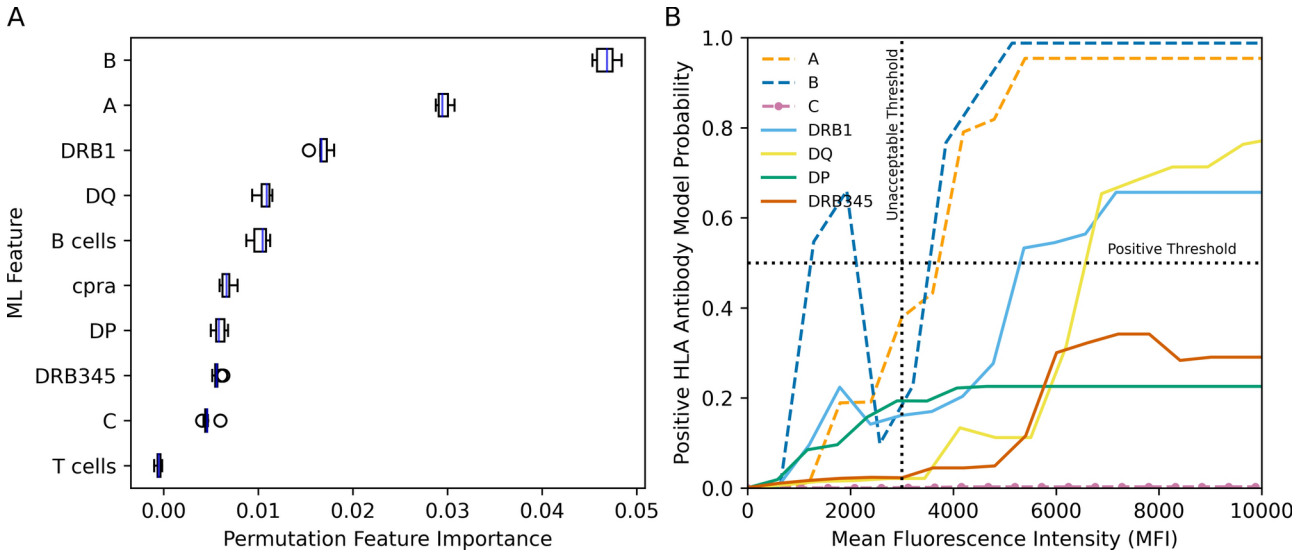

Fig. 3. Effect of HLA antibodies in ML model prediction. (**A**) ML feature importance using the permutation technique. (**B**) Stimulation of a single DSA and evaluation of the impact of various HLA antibodies on model prediction. Only the indicated variable was changed while all others were kept constant. The variable was changed from 0 to 10,000 mean fluorescent intensity (MFI). The influence on positive prediction using probability was evaluated. HLA class I (dashed lines) or class II (solid lines) antibodies impact on ML crossmatch prediction. The vertical line represents the institutional unacceptable threshold (3000 MFI). The horizontal line represents the traditional ML model predictive positivity cutoff value of 0.5. HLA class I antibodies were assessed using T cells and HLA class II antibodies were assessed using B cells.

	Expert VXM	Summed MFI thresholding model	Machine learning: HLA antibody (original)	Machine learning: HLA antibody (SMOTE)
PPV	0.490	0.931	0.858	0.778
NPV	0.970	0.961	0.973	0.977
Spec	0.980	0.995	0.993	0.986
Sen	0.385	0.606	0.726	0.735
LR+	19.25	121.20	103.71	52.50
LR−	0.628	0.396	0.276	0.269

Table 2. ML model diagnostic performance metrics. *PPV* positive predictive value, *NPV* negative predictive value, *Spec* specificity, *Sens* sensitivity, *LR* + positive likelihood ratio, *LR*− negative likelihood ratio.

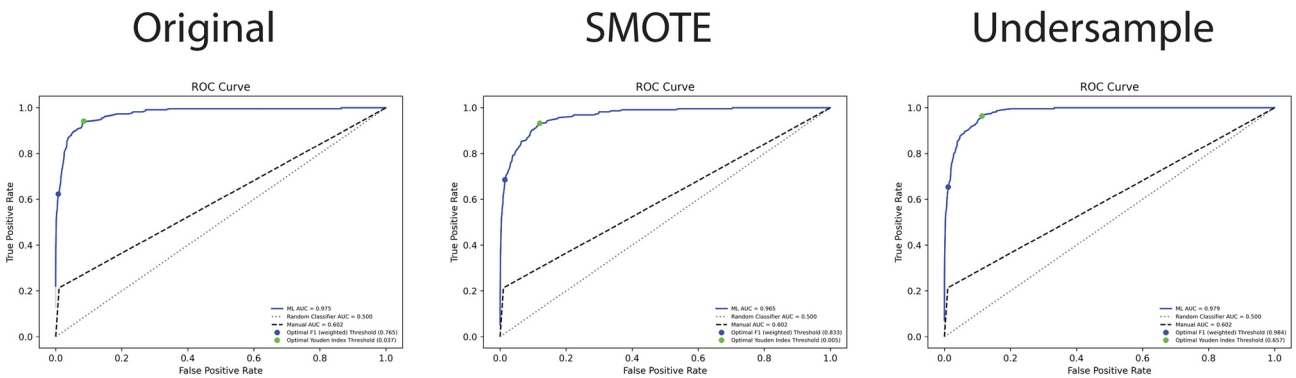

Fig. 4. ML model performance. ROC analysis using either unmanipulated original data (left), SMOTE (center), or undersampling (right). The blue line represents the ROC curve for that model. The dashed line represents the manual VXM process, and the dotted line represents random ROC performance. The blue dot represents the optimal f1-weighted threshold while the green dot represents the optimal Youden index threshold.

performance (AUC 0.965–0.979). These data strongly indicate that using optimized prediction thresholds the HLA antibody ML model can accurately predict crossmatch outcomes with minimal human interaction.

Discussion

As the demand for pretransplant immunologic risk assessments has increased, so has the need and requirement for more accurate and rapid tools. Here we describe the development and evaluation of a digital alloimmune risk assessment tool (DARA) to address these challenges that used machine learning models to predict allele-level HLA typing (as needed), gather and analyze DSA, and predict physical crossmatch results (Fig. 1). As stated by previous researchers, accurate prediction of HLA alleles remains a substantial challenge in clinical practice[29]. With the growth of high-resolution HLA typing for solid organ transplant recipients and donors, the accuracy of the ML model will likely move closer to that observed with the non-imputed dataset (Table 1). Our DARA achieved a ROC-AUC of 0.975, outperforming the current standard of care VXM methods (Fig. 4).

Importantly, the HLA antibody model's ability to maintain high specificity (99.3%) while improving sensitivity suggests that it could reduce the number of unnecessary physical crossmatches without increasing the risk of unexpected positive crossmatches. This may streamline the allocation process, reduce cold ischemia times, and improve graft outcomes[30].

A benefit of applying machine learning to pre-transplant immunologic compatibility is the ability to learn complex patterns and reduce the human bias in interpreting them. The current approach to assessing HLA antibodies has numerous established issues[31–33]. However, current physical crossmatch assessments rely heavily on human interpretation of complex HLA antibodies to estimate immunologic compatibility and risk[9,19,32]. The HLA antibody ML approach outperformed the manual human expert prediction, particularly when predicting positive results (Table 2). Expectedly, experts outperformed the ML model in predicting negative results (Table 2)[19,32] given that experts can readily understand the relationship between low or near-zero HLA antibody levels and negative crossmatch results. Given the issues with HLA antibody detection and the observed phenomenon of shared epitope spreading[31,34,35], it is not surprising that the HLA antibody model had lower sensitivity performance. Potential ways of addressing this deficiency are cross-reactive epitope group (CREG) analysis, HLA eplet-based analysis, and HLA serotype-based analysis[12,36].

The HLA antibody model enables understanding of the impact specific HLA antibodies have on crossmatch results. The structure of the HLA antibody scoring matrix likely allows for more generalizability of the ML model across different sites as only a single DSA value is required for each HLA locus. While still locally accurate, HLA antigens or alleles can be used as features for a similar ML model; however, those models have limited practical application due to the lack of depth to each HLA antigen or allele. Importantly, the HLA antibody model has learned which HLA antibodies are commonly considered over-reactive in the antibody detection assay (Fig. 3). In particular, the rank order of the HLA antibodies corresponds to the established HLA antigen expression profile on lymphocytes[20,37,38]. For instance, HLA-DQ antibodies required higher MFI values (approximately 6,000) to affect predictions, aligning with known issues of over-reactivity in single antigen bead assays. Lastly, the HLA match status of the potential pair is not directly evaluated or considered by the ML model. For pairs where a recipient and donor are matched at several HLA loci, regardless of the DSA MFI detected, those DSA would be reduced 100-fold when the ML model predicts the likelihood of a positive XM (see Material and Methods). This process ensures that highly matched pairs will appropriately be handled by the ML model when considering DSA MFI at the predictor.

The current practice of transplant centers is to list unacceptable HLA mismatches based on a recipient's HLA antibody profile using an MFI cutoff value; however, in many cases, this cutoff value is applied across all HLA loci rather than individualized to specific HLA loci. The HLA antibody model enables a deeper understanding of which HLA antibodies should be avoided given a recipient's medical status and how best to manage the associated immunologic risk. Consistent with published reports[39,40], higher levels of HLA-DQ antibodies were necessary to induce a positive crossmatch prediction (Fig. 3B). Interestingly, antibodies targeting HLA-C did not induce a positive crossmatch prediction even at high levels. This observation may be from the linkage disequilibrium between HLA-B and HLA-C, and Fig. 3B shows the model places more importance on antibodies targeting HLA-B. The reduced impact of HLA-C antibodies on crossmatch prediction is consistent with the over-representation of HLA-C and the over-reactivity of some HLA-DQ proteins on the solid phase antibody detection assay compared to their biological expression.

We addressed crossmatch outcome imbalance in the data using techniques such as SMOTE and random undersampling but did not find significant improvement in the model's performance. The AUC remained consistently high (0.975–0.979) across all approaches (Fig. 4). This resilience to class imbalance is likely due to several factors including threshold optimization and inherent data characteristics. It is likely the original data may have captured the important patterns distinguishing positive and negative crossmatches, even with the outcome imbalance. This suggests that the features used for model prediction are highly informative for crossmatch prediction.

Our study has several limitations. The allele prediction model impacts all downstream analyses. While the ML models are more robust than traditional methods, they have not been hardcoded to understand if the recipient is non-sensitized (no HLA antibodies) leading to a negative crossmatch. Additionally, the model has been trained and tested using internal data; validation through external data would enhance the confidence in the model prediction.

The work shown here is the initial step to a larger integration of data into the pre-transplant immunologic risk assessment that may include repeat HLA and eplet mismatches, HLA antibody quality, and other relevant parameters.

Materials and methods

Study design

Flow cytometric crossmatches (FCXM) performed at UNC Health from January 2016 to May 2024 were analyzed. Initially, 16,061 FCXM results were extracted from the laboratory information system (HistoTrac, version 2.52.9). After removing known false-negative, false-positive, inconclusive, incomplete, autologous, and uninterpretable FCXM, 13,158 cases (12,354 negative and 804 positive) remained for analysis.

Recipient and donor HLA typing data were obtained from HistoTrac. HLA antibody data from January 2013 to May 2024 were extracted from Fusion (version 4.2, OneLambda). HLA antibody mean fluorescence intensity (MFI) was determined using the OneLambda single antigen bead assay, performed as previously described[11].

FCXM procedures used pronase-treated lymphocytes, with positive cutoffs determined using normal human serum according to established laboratory practices[11]. All data were de-identified and processed using Python (version 3.8.15) in local and Microsoft Azure Cloud environments.

Machine learning (ML) datasets and training

The analysis included:

- 13,158 unique flow cytometric crossmatches
- 1903 unique recipients
- 2513 unique donors
- 5703 unique sera
- 29,070 HLA typing data points at the highest available resolution for all HLA loci

Data were split 75/25 for training and testing. Multiple supervised learning models were tested including Random Forest, XGBoost, Balanced Random Forest, Random Undersampling Classifier, and Logistical Regression Classifier with the best chosen using a custom-designed training metric. This metric provided differential penalties for false positive and false negative predictions, emphasizing the reduction of false negatives. The model was further optimized by determining the threshold that maximized the f1-weighted score, which balances precision and recall for imbalanced datasets.

HLA antibody profiles were obtained directly from Fusion (version 4.2). Donor-specific antibodies (DSA) were automatically determined by matching donor HLA alleles to recipient HLA antibodies in the single-antigen bead (SAB) assay. When a donor HLA allele was not represented in the SAB testing, a dictionary of HLA alleles and their SAB equivalents (generated by amino acid alignment using BioPython version 1.76) was used to select the closest match. For DQ and DP, the most likely donor heterodimer was determined using global DQA1-DQB1 haplotype frequency and DPA1-DPB1 haplotype frequency. The highest DSA MFI was used for each HLA locus. Any DSA that was also a recipient-specific (i.e. donor and recipient are an allele level match) antibody was reduced 100-fold before model training. The recipient's current calculated panel reactive antibody (cPRA) was also included as an ML feature, with missing values imputed using the mean cPRA of the entire dataset. If a donor lacked all DRB3/4/5, an MFI value of minus 1 (− 1) was used. HLA allele prediction.

Eight separate XGBoost models (one for each major HLA locus) were trained to predict HLA alleles when not previously known. These models used self-identified Race as a feature for allele prediction. For class I models, HLA antigen level data were used as inputs to predict a single HLA allele. HLA-DPB1 was not predicted, the allele for HLA-DPB1 was presumed to be the most common allele from real-time PCR-detected alleles. HLA-DPB1 was typed using LinkSeq HLA-ABCDRDQDP SABR 384 typing kit (One Lambda). Only missing alleles were predicted; known alleles were used directly in ML training and evaluation for crossmatch prediction.

Each imputation model was trained on high-resolution HLA genotyping data at two-field resolution that was converted to HLA serologic equivalents for training and evaluation. The training dataset varied by HLA locus and ranged from 16,164 to 60,072 high-resolution patients. The evaluation dataset of each model was HLA locus dependent and ranged from 4041 to 15,019. The data were split 80% for training and 20% for testing. Each model was generated using five-fold cross-validation.

Crossmatch ML training and assessment

Data were prepared by predicting missing donor and recipient HLA information using the ML models. Features were created using the highest donor specific HLA antibody MFI and accounted for DQ and DP heterodimer formation given the donor and recipient HLA genotypes, with cell type (T-cell or B-cell) treated as one of the features, allowing the model to predict both simultaneously. Race was not included as a feature for crossmatch prediction.

Model performance was assessed using true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) against flow cytometry crossmatch results. Precision (TP/(TP + FP)) and recall (TP/(TP + FN)) were calculated.

To address the imbalanced nature of the dataset, two techniques were applied:

1. Synthetic Minority Over-sampling Technique (SMOTE): The minority class (positive XM cases) was over-sampled to equal the negative XM cases in the training data.
2. Random undersampling: The majority class (negative XM cases) was down sampled in the training data.

ML model evaluation was always performed using the unmanipulated original dataset.

Individual HLA antibody impact on ML prediction probability

To evaluate the relative importance and impact of individual HLA antibodies on positive crossmatch outcomes, we employed Partial Dependence Plots (PDPs). PDPs are a model-agnostic method that helps visualize the marginal effect of a single feature on the predicted outcome of a machine-learning model while accounting for the average effects of all other features.

For each HLA antibody feature in our trained model: (a) a range of MFI values was created, starting from the minimum observed in the dataset up to either the maximum observed value or 10,000 MFI, whichever was lower. (b) This range was divided into equally spaced intervals (e.g., 100 points) to create a smooth curve.

For each point in this MFI range: (a) The entire dataset was copied, replacing the original MFI value for the antibody of interest with the current point value. (b) Predictions were made using our trained model on these modified datasets. (c) The average predicted probability of a positive crossmatch was calculated across all these predictions.

The resulting curve shows how the predicted probability of a positive crossmatch changes as the MFI of a specific HLA antibody varies, while averaging out the effects of all other features. This process was repeated for each HLA antibody feature in our model.

ML feature permutation determination

To assess the relative importance of features in our model, we employed permutation importance, a model-agnostic method that measures the impact of each feature on model performance. This technique operates by randomly shuffling the values (HLA antibody MFIs) of a single feature while keeping all others constant, then evaluating the change in the model's performance metric. The process is repeated for each feature independently. The permutation process was repeated 10 times for each feature to obtain a distribution of importance scores, enhancing the reliability of our estimates. The analysis was conducted after fitting the model to the entire dataset, providing a comprehensive view of feature importance across all available data.

Detection error tradeoff (DET) and precision-recall curves

The DET curve was generated to visualize the tradeoff between false negative rate (FNR) and false positive rate (FPR) across various decision thresholds. The model's predicted probabilities for the test set were obtained. These probabilities were compared against a range of thresholds to compute FNR and FPR pairs. The curve was plotted with FPR on the x-axis and FNR on the y-axis. Lower curves indicate better model performance, with the optimal point balancing FNR and FPR according to our specific clinical requirements.

The Precision-Recall curve was used to assess the model's performance, particularly in the context of the imbalanced dataset. For both curves, the optimal threshold was determined before these analyses using a custom metric that balanced sensitivity and specificity according to clinical priorities. This threshold was applied consistently across all evaluations to ensure coherence between model training, validation, and performance visualizations.

Thresholding model metrics

For the same set of donors and patients used in crossmatch ML training, total class I and class II donor-specific HLA antibody MFI were calculated from patient SAB data. Thresholds applied to the sum of the MFI values were tested in increments of 10, seeking to maximize the number of TP + TN. Values above the threshold were predicted positive, and values below were predicted negative. The optimal threshold for T-cell crossmatches (using class I antibody totals) was 4990, and for B-cell crossmatches (using class I plus class II antibody totals) was 9350.

Expert VXM

1,751 VXM were performed at UNC Health from January 2016 to February 2024 and were exported from HistoTrac. Equivocal VXM interpretations were removed. Data were connected to FCXM by unique identifiers. FCXM cases with inconclusive, incomplete, and uninterpretable results were removed.

Statistics

Scikit-learn (version 1.2.0) was used for ML training, testing, and metrics. XGBoost (version 1.6.1), imblearn (version 0.10.1), and optuna (version 4.0.0) were used for ML training, testing, and evaluation purposes.

Data availability

All data, code, and materials used in the analysis will be available to any researcher for reproducing or extending the analysis. Access to the materials mentioned will require a materials transfer agreement (MTA). Please contact the corresponding author for data and material requests.

Received: 6 November 2024; Accepted: 25 February 2025
Published online: 07 March 2025

References

1. Tambur, A. R. et al. Sensitization in transplantation: assessment of risk (STAR) 2017 working group meeting report. *Am. J. Transplant.* **18**, 1604–1614 (2018).
2. Roll, G. R. et al. A virtual crossmatch-based strategy facilitates sharing of deceased donor kidneys for highly sensitized recipients. *Transplantation* **104**, 1239–1245 (2020).
3. Eby, B. C. et al. Virtual HLA crossmatching as a means to safely expedite transplantation of imported pancreata. *Transplantation* **100**, 1103–1110 (2016).

4. Amico, P., Hönger, G., Steiger, J. & Schaub, S. Utility of the virtual crossmatch in solid organ transplantation. *Curr. Opin. Organ Transplant.* **14**, 656–661 (2009).
5. Takeda, A. et al. Acute humoral rejection of kidney allografts in patients with a positive flow cytometry crossmatch (FCXM). *Clin. Transplant.* **14**(Suppl 3), 15–20 (2000).
6. Downing, J. The lymphocyte crossmatch by flow cytometry for kidney transplantation. *Methods Mol. Biol.* **882**, 379–390 (2012).
7. Sullivan, H. C. et al. (F)Utility of the physical crossmatch for living donor evaluations in the age of the virtual crossmatch. *Hum. Immunol.* **79**, 711–715 (2018).
8. Liwski, R. S. et al. Rapid optimized flow cytometric crossmatch (FCXM) assays: The Halifax and Halifaster protocols. *Hum. Immunol.* **79**, 28–38 (2018).
9. Morris, G. P., Phelan, D. L., Jendrisak, M. D. & Mohanakumar, T. Virtual crossmatch by identification of donor-specific anti-human leukocyte antigen antibodies by solid-phase immunoassay: a 30-month analysis in living donor kidney transplantation. *Hum. Immunol.* **71**, 268–273 (2010).
10. Morris, A. B., Sullivan, H. C., Krummey, S. M., Gebel, H. M. & Bray, R. A. Out with the old, in with the new: Virtual versus physical crossmatching in the modern era. *HLA* **94**, 471–481 (2019).
11. Weimer, E. T. & Newhall, K. A. Development of data-driven models for the flow cytometric crossmatch. *Hum. Immunol.* **80**, 983–989 (2019).
12. Norin, A. J. et al. Determination of unacceptable antigens by summation of anti-HLA eplet antibody strength (MFI) based on single antigen bead assays: Excellent correlation with negative cell based cross matches. *Hum. Immunol.* **83**, 482–493 (2022).
13. Valentin, M. O. et al. Improving the access of highly sensitized patients to kidney transplantation from deceased donors: the spanish PATHI program with allocation based on the virtual crossmatch. *Transplantation* **108**, 787–801 (2024).
14. Wrenn, S. M. et al. Improving the performance of virtual crossmatch results by correlating with nationally-performed physical crossmatches: Obtaining additional value from proficiency testing activities. *Hum. Immunol.* **79**, 602–609 (2018).
15. Johnson, C. P. et al. Renal transplantation with final allocation based on the virtual crossmatch. *Am. J. Transplant.* **16**, 1503–1515 (2016).
16. Peräsaari, J. P., Jaatinen, T. & Merenmies, J. Donor-specific HLA antibodies in predicting crossmatch outcome: Comparison of three different laboratory techniques. *Transpl. Immunol.* **46**, 23–28 (2018).
17. Piazza, A. et al. Virtual crossmatch in kidney transplantation. *Transplant. Proc.* **46**, 2195–2198 (2014).
18. Chowdhry, M., Agrawal, S., Thakur, Y., Guleria, S. & Sharma, V. Implication of a positive virtual crossmatch with negative flow crossmatch: A mind-boggler. *Asian J. Transfus. Sci.* **14**, 79–82 (2020).
19. Olszowska-Zaremba, N., Zagożdżon, R. & Gozdowska, J. Accuracy of virtual crossmatch (VXM) prediction of physical crossmatch (PXM) results of donor specific antibody (DSA) in routine pretransplant settings-a single-center experience. *Transpl. Immunol.* **72**, 101583 (2022).
20. Badders, J. L., Jones, J. A., Jeresano, M. E., Schillinger, K. P. & Jackson, A. M. Variable HLA expression on deceased donor lymphocytes: Not all crossmatches are created equal. *Hum. Immunol.* **76**, 795–800 (2015).
21. Montgomery, M. C., Liu, C., Petraroia, R. & Weimer, E. T. Using nanopore whole-transcriptome sequencing for human leukocyte antigen genotyping and correlating donor human leukocyte antigen expression with flow cytometric crossmatch results. *J. Mol. Diagn.* **22**, 101–110 (2020).
22. Miller, R. J. H. et al. Temporal shift and predictive performance of machine learning for heart transplant outcomes. *J. Heart Lung Transplant.* **41**, 928–936 (2022).
23. Tian, D. et al. Machine learning-based prognostic model for patients after lung transplantation. *JAMA Netw. Open* **6**, e2312022 (2023).
24. Mark, E., Goldsman, D., Gurbaxani, B., Keskinocak, P. & Sokol, J. Using machine learning and an ensemble of methods to predict kidney transplant survival. *PLoS One* **14**, e0209068 (2019).
25. Becker, J. U. et al. Artificial intelligence and machine learning in nephropathology. *Kidney Int.* **98**, 65–75 (2020).
26. Senev, A., Emonds, M.-P. & Naesens, M. Second field high-resolution HLA typing for immunologic risk stratification in kidney transplantation. *Am. J. Transplant.* **21**, 3502–3503 (2021).
27. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *Jair* **16**, 321–357 (2002).
28. Shao, H., Liu, X., Zong, D. & Song, Q. Optimization of diabetes prediction methods based on combinatorial balancing algorithm. *Nutr. Diabetes* **14**, 63 (2024).
29. Engen, R. M., Jedraszko, A. M., Conciatori, M. A. & Tambur, A. R. Substituting imputation of HLA antigens for high-resolution HLA typing: Evaluation of a multiethnic population and implications for clinical decision making in transplantation. *Am. J. Transplant.* **21**, 344–352 (2021).
30. Baxter-Lowe, L. A., Cecka, M., Kamoun, M., Sinacore, J. & Melcher, M. L. Center-defined unacceptable HLA antigens facilitate transplants for sensitized patients in a multi-center kidney exchange program. *Am. J. Transplant.* **14**, 1592–1598 (2014).
31. Sullivan, H. C., Liwski, R. S., Bray, R. A. & Gebel, H. M. The road to HLA antibody evaluation: do not rely on MFI. *Am. J. Transplant.* **17**, 1455–1461 (2017).
32. Ellis, T. M. et al. Diagnostic accuracy of solid phase HLA antibody assays for prediction of crossmatch strength. *Hum. Immunol.* **73**, 706–710 (2012).
33. Greenshields, A. L. & Liwski, R. S. The ABCs (DRDQDPs) of the prozone effect in single antigen bead HLA antibody testing: Lessons from our highly sensitized patients. *Hum. Immunol.* **80**, 478–486 (2019).
34. Garcia-Sanchez, C., Usenko, C. Y., Herrera, N. D. & Tambur, A. R. The shared epitope phenomenon-A potential impediment to virtual crossmatch accuracy. *Clin. Transplant.* **34**, e13906 (2020).
35. Sullivan, H. C., Krummey, S. M., Gebel, H. M. & Bray, R. A. The utility of second single antigen bead assay: Clearing the water or stirring up mud?. *Hum. Immunol.* **81**, 663–670 (2020).
36. Osoegawa, K. et al. A new strategy for systematically classifying HLA alleles into serological specificities. *HLA* **100**, 193–231 (2022).
37. Hughes, A. E. O., Montgomery, M. C., Liu, C. & Weimer, E. T. Allele-specific quantification of human leukocyte antigen transcript isoforms by nanopore sequencing. *Front. Immunol.* **14**, 1199618 (2020).
38. Cornaby, C., Montgomery, M. C., Liu, C. & Weimer, E. T. Unique molecular identifier-based high-resolution HLA typing and transcript quantitation using long-read sequencing. *Front. Genet.* **13**, 901377 (2022).
39. Bettinotti, M. P., Zachary, A. A. & Leffell, M. S. Clinically relevant interpretation of solid phase assays for HLA antibody. *Curr. Opin. Organ Transplant.* **21**, 453–458 (2016).
40. Raghavan, M., Yarzabek, B., Zaitouna, A. J., Krishnakumar, S. & Ramon, D. S. Strategies for the measurements of expression levels and half-lives of HLA class I allotypes. *Hum. Immunol.* **80**, 221–227 (2019).

Author contributions

Conceptualization: E.T.W., K.A.N. Methodology: E.T.W., K.A.N. Investigation: E.T.W., K.A.N. Visualization: E.T.W., K.A.N. Writing—original draft: E.T.W., K.A.N. Writing—review and editing: E.T.W., K.A.N.

Funding
No external funding was used in this study.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to E.T.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.