Biol Blood Marrow Transplant 000 (2020) 1–8

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Predicting the Availability of Hematopoietic Stem Cell Donors Using Machine Learning

Ying Li[1,*], Ausra Masiliune[1], David Winstone[1], Leszek Gasieniec[2], Prudence Wong[2], Hong Lin[1], Rachel Pawson[3], Guy Parkes[1], Andrew Hadley[4]

[1] *Department of Stem Cell Donation and Transplantation, NHS Blood and Transplant, Bristol, United Kingdom*
[2] *Department of Computer Science, University of Liverpool, Liverpool, United Kingdom*
[3] *Department of Clinical Haematology, Oxford University Hospitals NHS Foundation Trust, Bristol, United Kingdom*
[4] *Department of Specialist Patient Services, NHS Blood and Transplant, Bristol, United Kingdom*

Article history:
Received 18 December 2019
Accepted 29 March 2020

Keywords:
allogeneic hematopoietic stem cell transplantation
machine learning
donor availability
donor selection

A B S T R A C T

Hematopoietic stem cell transplantation (HSCT) is firmly established as an important curative therapy for patients with hematologic malignancies and other blood disorders. Apart from finding HLA-matched donors during the HSCT process, donor availability remains a key consideration as the time taken from diagnosis to transplant is recognized to adversely affect patient outcome. In this study, we aimed to develop and validate a machine learning approach to predict the availability of stem cell donors. We retrospectively collected a data set containing 10,258 verification typing requests made during the HSCT process in the British Bone Marrow Registry (BBMR) between January 1, 2013, and December 31, 2018. Three machine learning algorithms were implemented and compared, including boosted decision trees (BDTs), logistic regression, and support vector machines. Area under the receiver operating characteristic curve (AUC) was primarily used to assess the algorithms. The experimental results showed that BDTs performed better in predicting the availability of BBMR donors. The overall predictive power of the model, using AUC on the test cohort of 2052 records, was found to be 0.826. Our findings show that machine learning can predict the availability of donors with a high degree of accuracy. We propose the use of the BDT machine learning approach to predict the availability of BBMR donors and use the predictive scores during the HSCT process to ensure patients with blood cancers or disorders receive a transplant at the optimum time.

INTRODUCTION

Allogeneic hematopoietic stem cell transplantation (HSCT) is used to treat patients with a range of malignant and nonmalignant hematologic disorders as well as other specific disorders of the immune system. Patients require a detailed pretransplant assessment as well as investigations to assess their clinical status and their fitness to proceed to transplant. Allogeneic HSCT involves transferring the stem cells from a healthy donor into a patient's body after conditioning therapy (chemotherapy with or without total body irradiation) at a range of doses depending on the type and severity of the disease being treated. The improvement in outcomes after HSCT using unrelated donors (UDs) and the development of novel nontoxic preparative regimens make UD HSCT an option for patients who do not have an HLA-matched sibling [1,2].

Several variables have been demonstrated to have an association with adverse effects on patient outcome following HSCT. These include disease progression, donor and patient age, and donor-recipient sex mismatch [3-5]. The timing of the HSCT has also been reported to be a significant factor [6]. In a study of 8003 unrelated donor transplants by Pidala et al. [7], the overall survival rate at 5 years for patients with early-stage disease was found to be more than twice the rate of patients with advanced disease. Craddock et al. [8] found that a time from diagnosis to transplant of <4 months was significantly associated with improved overall survival and leukemia-free survival at 5 years. A study of 548 patients by Heemskerk et al. [9] found that 30% of patients became medically unfit while waiting for a UD HSCT. Taking into account factors such as disease risk, age, and sex, they concluded that reducing the time taken for donor provision was key to reducing rates of clinical deterioration.

A number of obstacles may be encountered in the provision of UD HSCT donors. One major point of delay is the verification typing (VT) stage [10]. VT includes the tests carried out on a fresh blood sample of a specific donor with the purpose of

Financial disclosure: See Acknowledgments on page 7.

*Correspondence and reprint requests: Ying Li, PhD, Department of Stem Cell Donation and Transplantation, NHS Blood and Transplant, 500 North Way, Bristol, BS34 7QH, UK.

E-mail address: ying.li@nhsbt.nhs.uk (Y. Li).

https://doi.org/10.1016/j.bbmt.2020.03.026
1083-8791/

verifying the identity and concordance of an existing HLA assignment. The purpose of this typing is to ensure that the volunteer is the same individual whose HLA typing was listed on the search report used to select the donor. Here, registries will need to be able to contact potentially matching donors—some of whom may have been on the register for several decades without regular contact—and to establish their willingness and fitness to donate before arranging for further blood samples for VT and testing for infectious disease markers. It may take several weeks to trace a donor with obsolete contact details, which may then only reveal that the donor may be medically ineligible to donate or they may have personal reasons as to why they no longer wish to donate, which will often be related to valid lifestyle issues such as family or travel.

Some particular characteristics are found to be associated with donor availability, including sex, age, time spent on register, and ethnicity [11-13]. Less committed blood donors are less likely to donate stem cells [14]. In a study looking at factors influencing donor willingness in the African American population [15], education and awareness of HSCT were found to have a positive correlation with a willingness to be a donor. In addition, certain psychosocial factors such as motivation, ambivalence, intrinsic commitment to donation, more realistic expectations, fewer medical concerns, and greater contact with the donor center were also associated with donor availability [16]. A recent study by Sivasankaran et al. [17] proposed a machine learning approach to predict the availability of every registered donor and to use these predictors during donor selection to reduce the time to transplant as much as possible.

The British Bone Marrow Registry (BBMR) is a panel of blood donors who have volunteered to become hematopoietic stem cell donors. The BBMR provides UD HSCT donors to UK and overseas transplant centers. The BBMR has 370,757 active donors as of August 1, 2019, all recruited from blood donation sessions run by NHS Blood and Transplant (NHSBT). Multivariate analysis published by Switzer et al. [18] has shown that blood donors have a lower rate of attrition. However, our recent 5 years of data show that 36% of BBMR donors were not available at the VT stage. Although this is relatively good compared to the results published by Anthony Nolan and the National Marrow Donor Program (NMDP), which were 38% and 50%, respectively [12,17], it highlights the need for specific intervention programs to retain the BBMR donors who are at risk of dropping out. It is important to establish those factors that can predict BBMR donor availability to potentially simplify the transplant decision process and to minimize the risk of delays in transplantation. To our knowledge, no other stem cell registry that is integrated with blood donation and that only accepts blood donors on the register has published its donor availability statistics.

In this study, we use supervised machine learning techniques to train models by providing 5 years of donor information as the input and their corresponding responses to VT requests as target outputs. Three machine learning algorithms were implemented and compared, including boosted decision trees (BDTs) [19], logistic regression (LR) [20], and support vector machines (SVMs) [21]. Area under the receiver operating characteristic curve (AUC) was primarily used to assess the algorithms.

MATERIALS AND METHODS

We evaluated VT request data from January 1, 2013, to December 31, 2018. A total of 10,258 VT requests were made during this period. The models were trained and tested using a set of features extracted from the blood donor management system, the BBMR stem cell donor laboratory information management system (LIMS) system, and the 2011 UK Census [22] that captures donor information such as demographics, blood donation activities, medical deferrals, education background, and socioeconomic status. For the 2011 census, the smallest geographic unit for which outputs are published is the output area (OA) [23], which contains more than 100 persons and 40 households. Our donors are mapped to the OA level based on donors' home postcodes. It cannot be assumed that people have similar characteristics to those who live in the same area, but these area measures might be more valid than self-declared and unverified individual-level indicators.

In total, 12 features were captured for each donor, including the output variable (response to a VT request). Table 1 describes the features and the data types.

Machine learning deals with the usage of mathematical models on the data, meaning that it cannot be applied to data sets that have missing values. The general approach is to fill the missing value with a suitable value to substitute for the missing field. The NHS Give Blood App status is a categorical feature that indicates whether the status of using the app is active or inactive. It had 5529 missing values; this large number is attributed to the fact that the NHS Give Blood App was only introduced in 2014, and therefore, the missing values have been replaced with "unknown." The ethnicity feature had 297 donors missing values. We did not approach the donors to retrospectively collect the ethnicity information but replaced the missing ethnicity with "unknown." Moreover, 93 donors had missing postcodes, so we could not map the donors to the OA level. As a result, the missing values for social grade, property ownership, and education level were replaced with "unknown."

Once the data set was cleaned, we applied transformations to the data before they could be input into a machine learning algorithm. The categorical features were converted to ordinal numbers, and the noncategorical features were normalized to change the numeric values to a common scale between 0 and 1 using minimum-maximum normalization.

In our data set, the output is categorical, with positive or negative responses to a VT request. In the collected data set, there were 64% positive responses to VT and 36% negative responses. To overcome this imbalance problem, the synthetic minority oversampling technique [24] was used to create more copies of the underrepresented data set (negative responses) to balance our data. The trained models were BDTs, LR, and SVMs, and the modeling was conducted in Microsoft Azure (Microsoft, Redmond, WA, USA). For all 3 algorithms, we used binary classification, which is suitable to predict 2 possible outcomes (ie, either positive or negative response to a VT request in our case).

In the BDT model, the algorithm produces multiple decision trees where the newly created tree learns from the errors in the previously created tree. In each tree, the branch represents a choice between a number of alternatives of an attribute in the internal nodes leading to a final decision in the leaf node. The process of splitting based on decisions of different internal node features continues until a subset at a node has the same values of the target variable or when splitting no longer adds value to the predictions. The main goal of decision trees is to find the best split of each node of the tree. The final outcome prediction is assigned based on the weighted sum of the ensemble of created trees.

The LR model makes a prediction of a probability of an event by inputting independent variable values into a logistic regression equation. The coefficients of the equation are optimized during the training stage. Sigmoid function is used to map the linear combination of inputs into the range of [0,1], thus giving us the classification probabilities. In binary classification problems, the general rule is to use a probability threshold of 0.5 to make classification predictions. So, in our case, a record with predicted probability of >0.5 is classed as a positive response and probability of ≤ 0.5 is classed as a negative response.

In SVMs, an input record with n features is plotted as a point in an n-dimensional space with the value of each feature being the value of a particular coordinate. Then, classification is done by finding the hyperplane that separates the 2 classes (either positive or negative response) best. A hyperplane is a line that splits the input variable space. During the testing stage, the input records with known outcome are plotted in the same multidimensional space, and the predicted outcome is assign based on which side of the line the point belongs, thus giving the predictions of true positive, false positive, true negative, and false negative.

RESULTS

The entire data set of 10,258 records was randomly split into a training subset (n = 8206) and a testing subset (n = 2052)—an 80:20 split (training/testing). The training data set was used to train the selected models and the test data set was used to validate, evaluate, and compare the performances of the trained models.

Table 1
Description of the Features and the Data Types

Feature Name	Description	No. of Missing Records	Data Type	Mean	Range
Sex	Donor's sex	0	Binary nominal	—	Male or female
Ethnicity	Self-declared donor ethnicity	297	Multinominal	—	Bangladeshi, Indian, Pakistani, Asian other, black African, black Caribbean, black other, Chinese, mixed other, mixed white/Asian, mixed white/black African, mixed white/black Caribbean, unknown, British, Irish, white other
Age when selected	Age of donor at the time of VT request	0	Numerical	35.15	18-60 years old
Length on registry	Time period in years, calculated from the date when donor joined the BBMR to the date of VT request	0	Numerical	7.96	0-30 years
NHS Give Blood App status	If a user of the NHS Give Blood App	5529	Multinominal	—	Active, inactive, or unknown
No. of days since last donor contact	Number of days since the most recent donor contact to the date of VT request	0	Numerical	612.50	1-8299 days
Blood donation team	Blood donation team hosting the most recent blood donation appointment that donor had attended prior being selected for VT	0	Multinominal	—	120 blood donation teams, such as London, Birmingham, Manchester
Blood donation reliability score [14]	The blood donor reliability score relating to blood donation ranging from 1 (best) to 5 (worst)	0	Multinominal	—	1, 2, 3, 4 or 5
Social grade	Approximated socioeconomic classification produced by the ONS (UK Office for National Statistics). See details in Appendix B	93	Multinominal	—	AB, C1, C2, DE, or unknown
Property ownership	Percentage of people living within an area who solely or partially own their property	93	Numerical	66.94%	1.2%-100%
Education level	Percentage of people living within an area whose highest qualification is level 2 and above	93	Numerical	60.18%	17.9%-100%
Outcome	Outcome of VT request. A categorical variable is used to indicate whether the donor provided a VT sample or did not	0	Binary nominal	—	Yes or no

Once each model was trained, we used it to make predictions and to generate a confusion matrix on the testing data. The confusion matrix was used to calculate the classification accuracy, sensitivity, precision, and F1 scores as well as to plot the receiver operating characteristic curve for the model. Ten-fold cross-validation was performed on training data set to assess the variability of the training data set and the reliability of the machine learning (ML) models trained using that data. The training data were divided into 10 folds, and then the model-fitting procedure was repeated for a total of 10 times, with each fit being performed on a training set consisting of 90% of the total training set selected at random, with the remaining 10% used as a holdout set for validation. When the building and evaluation process is complete for all folds, a set of performance metrics (accuracy, sensitivity, precision, F1 score, AUC) is generated for each fold. The mean of the fold AUCs is the cross-validated AUC estimate. We reviewed these metrics and did not observe that any single fold has particularly high or low accuracy. The trained model was then applied on testing data set, and the performance metrics were similar to what we achieved on the training data. This confirmed that the model learned well from the training data and confirmed that our data set is representative and the proposed model works well for different variations of the data.

All of the abovementioned performance metrics were used to compare the models and to find the one that is best suited for our donor availability data. BDTs had the highest scores compared with all the other models, so we used this model for further analysis. Table 2 shows the computed metrics of the models measured on the training and testing data sets.

The confusion matrix of BDTs generated from the predictions on the testing data is shown in Figure 1a. In our case, we were more focused on identifying the donors who will not proceed with the VT process, which is the true-negative case of our model's predictions. The receiver operating characteristic curve for the BDT model on testing data was plotted and is shown in Figure 1b.

The Azure machine learning built-in module, permutation feature importance [25,26], was used to identify the relative influence of features in the prediction of donor availability. The features were plotted in the order of significance, as shown in Figure 2.

Apart from producing an overall predictive score for donor availability, we also used the BDT model to predict the subcategories of negative responses, including medical deferral, ability to contact, and personal commitment. The prediction results on testing data were inconsistent across the categories, but it shows promise of using the proposed model to predict

Table 2
Computed Metrics of Models Measured on Training and Testing Data Sets

	Accuracy		Sensitivity		Precision		F1 Score		AUC	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
BDT	0.770	0.742	0.741	0.730	0.797	0.765	0.757	0.747	0.860	0.826
LR	0.695	0.683	0.683	0.671	0.693	0.696	0.688	0.683	0.764	0.748
SVM	0.673	0.661	0.717	0.697	0.656	0.659	0.685	0.678	0.734	0.721

Accuracy is the percentage of predictions that are correct ((TP + TN)/(TP + TN + FP + FN)). Sensitivity is the percentage of positive cases that were predicted as positive (TP/(TP + FN)). Precision is the percentage of positive predictions that are correct (TP/(TP + FP)). F1 score is the harmonic mean of sensitivity and precision (2 × sensitivity × precision/(sensitivity + precision)). AUC is the area under the receiver operating characteristic (ROC) curve, calculated from the ROC plot.

the ability to contact unavailability category. The prediction accuracy for medical, ability to contact, personal commitment is 0.685, 0.875, and 0.636, respectively. See Appendix A for additional details.

Machine learning is frequently referred to as a "black box" (ie, data go in, decisions come out, but the processes between input and output are opaque). To have a better understanding of why decisions are made by the BDT model, additional subsidiary analyses were done in subgroups of the features. The blood donation team was excluded due to the high dimension of the feature. Fisher exact test was used for comparison of the subgroups, and we consider a *P*value of less than .05 significant. The analyses were carried out with R v3.6.1 (R Foundation for Statistical Computing, Vienna, Austria), and the results are summarized in Table 3.

The feature "number of days since last donor contact" plays a more significant role than other features. It is an indication of a donor's current status, and it also highlights the importance to establish a recent contact with the donors. It is also found that "NHS Give Blood App status" has a relatively high influence in the prediction of a donor's availability. The NHS Give Blood App was launched in 2014 and has significantly changed the way many donors make appointments and keep track of their donation history, rewards received, and communications from the NHSBT.

In medical practice and biomedical research, self-identified ethnicity is frequently collected and often serves as a proxy for genetic ancestry [27]. However, it remains a challenging area due to errors in self-identified information and complex ancestry information [28]. A person's ethnic identity is part of a wider social process and is influenced by their own perceptions of ethnicity and what they perceive others' perceptions are within their particular community. Also, a person's responses can change over time [29]. The ethnicity categories used in our blood donor management system are the same as the ethnic groups used in the 2011 census. We observed that the VT outcome for the mixed ethnic groups does not show any statistical significance when compared to white British. This implies that how they were raised and where they grew up may have more influence on their donation behaviors than the self-declared ethnicity.

There were significant association between 2 Caucasian origins and donor availability when compared to white British (white other and white Irish, $P < .001$). In the NHSBT, white other is normally used to indicate that donors originated from the European Union who are not of the English, Welsh, Scottish, or Irish ethnic groupings. The lower donor availability rate may be due to donors already being registered in their own countries or donors moving back to their own countries. Further study is required to better understand the impact.

We found that the predominant reason for donor attrition among ethnic minorities (Pakistani, Indian, Bangladeshi, black Caribbean, Asian other, black African, Chinese, black other) was the inability to contact donors. This could imply that there

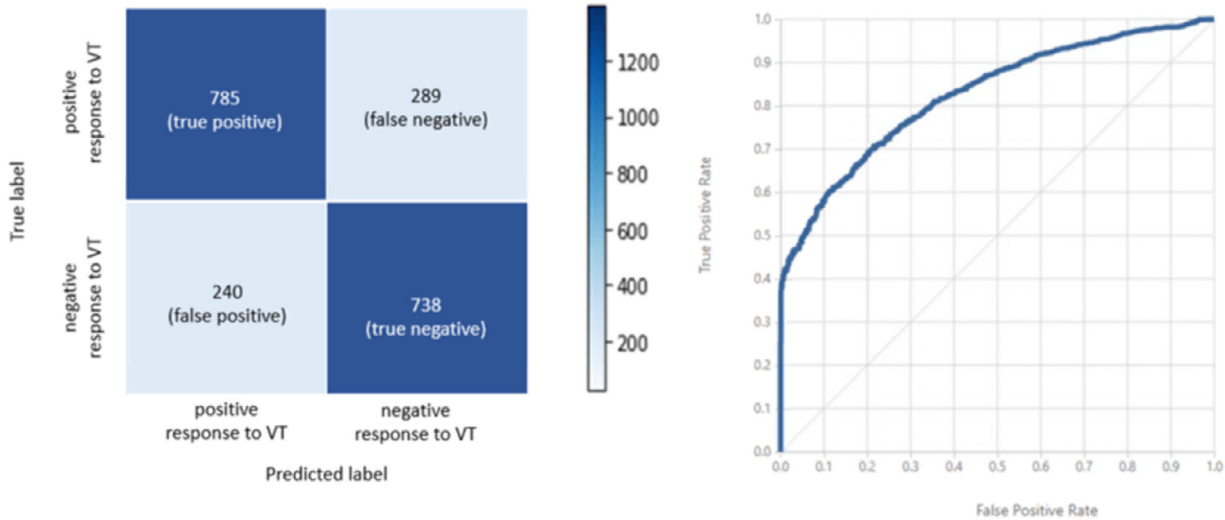

Figure 1. (a) Confusion matrix of the BDT model on testing data. True positive (TP) means predicted positive and actual positive; true negative (TN) means predicted negative and actual negative; false positive (FP) means predicted positive and actual negative; false negative (TN) means predicted negative but actual positive. (b) Receiver operating characteristic (ROC) curve for the BDT model on testing data. ROC is a 2-dimensional graph in which the true-positive rate (TP/(TP + FN)) is plotted on the y-axis and false-positive rate (FP/(FP + TN)) is plotted on the x-axis. To generate the entire ROC curve, the true-positive rate and the false-positive rate for all possible classification thresholds that range from 0 and 1 are plotted. We used the default value, which is 100 for the number of thresholds in Microsoft Azure.

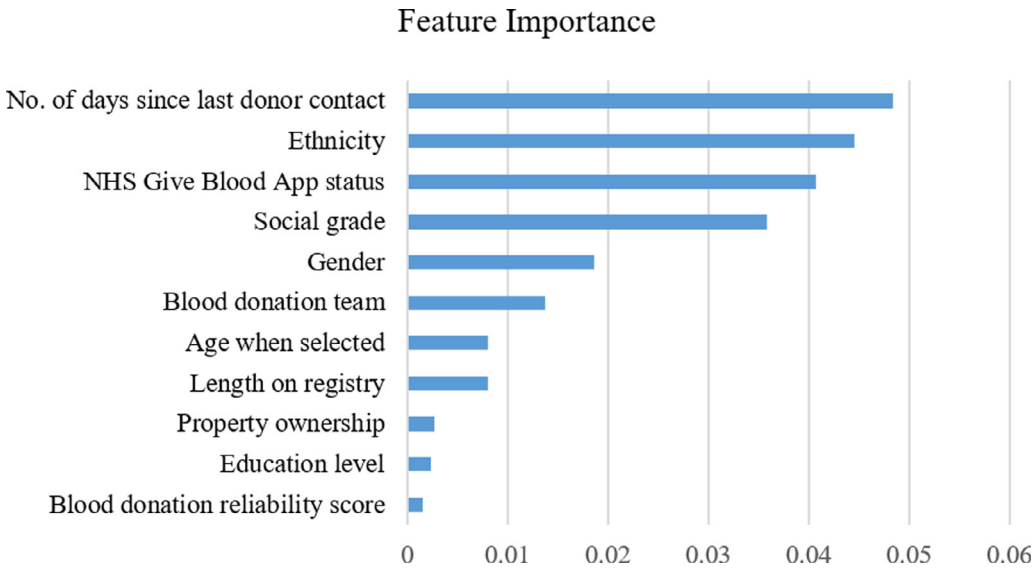

Figure 2. Feature importance for the BDT model.

are engagement barriers with donors from ethnic minorities, and this possibility should be addressed in future research. Bangladeshi and black other ethnic groups failed to show a significant ethnicity-availability association, which may be a result of a small number of records included in these groups. We did not group them to broader ethnic groups as this would result in reduced granularity of information about donors' ethnic background and would introduce investigator bias into ethnicity grouping. Also, for machine learning, it is important to include all information as precise as possible so that it can learn from past experience.

DISCUSSION

Machine learning is a rapidly growing tool, which is being used to predict the effectiveness and outcomes in various treatment areas [30-32]. A recent large study by Sivasankaran et al. [17] evaluated 178,249 VT requests. The overall predictive power of using AUC on a test cohort of 44,544 requests was found to be 0.77. This demonstrated the potential in using ML to predict donor availability. They included both domestic (NMDP) data and data from international collaborating donor centers, but several features such as recommitted response, self-online registration, and postrecruitment survey are exclusive to the NMDP only. In this study, we presented a machine learning approach to analyze not only donor characteristics and behaviors but also socioeconomic data. Unavailability of the donor showed association with lower social grade (odds ratio, 1.40; $P < .001$; social grade DE versus AB). The proposed BDT machine learning model performed well in predicting BBMR donor availability. The overall predictive power of the model, using AUC on the test cohort of 2052 records, was found to be 0.826. It also shows promise of using the proposed model to predict the ability-to-contact unavailability category with a classification accuracy of 0.875.

The proposed BDT model calculated the probability of getting the positive class of the output variable on which it makes the final class predictions. If the probability is greater than or equal to 0.5, then it predicts the outcome to be positive, and if the probability is less than 0.5, then it predicts the outcome to be negative. These probabilities can be interpreted as the donors' availability score. We have initiated a pilot project in the BBMR using the predictive tool to select donors for HLA typing improvement. Donor availability score and other characteristics were used during the selection process. It would be beneficial to focus on the donors who are more likely to donate.

We propose the use of the BDT machine learning approach to predict the availability of donors and use the predictive scores during the HSCT process. Apart from finding HLA-matched donors during the HSCT process, donor availability remains a key consideration as the time taken from diagnosis to transplant is recognized to adversely affect patient outcome. Individual consideration of each applicable characteristic is laborious. A single score for each potential donor can simplify the donor selection process and assist the clinicians to make decisions. Ultimately, such interventions should reduce delays in unrelated hematopoietic stem cell donor provision.

Our study also had several limitations. First, the BBMR is a population of exclusive blood donors, which is atypical of most current worldwide registries. Some of the features that act as the predictors in the proposed BDT model are related to blood donation activities and behaviors (eg, NHS Give Blood App status, blood donation team, and blood donation reliability score). Moreover, there are 120 blood donation teams in our data set. We did not perform statistical analyses on this feature due to the high dimension. Blood donation team is an indication of a donor's location; further studies (eg, geographical study of places and the relationships between donors and their environments) are needed to better understand why they have an impact on the prediction. However, mobile app and location of donors are not exclusive features to blood donors. In addition, blood donation reliability score was at the low end of feature importance, and we think this is probably due to the strong positive correlation with the number of days since last donor contact and length on registry, with Pearson correlation coefficients of 0.77 and 0.47, respectively. A side study to exclude the blood donation-related features was performed; the overall AUC predictive power of the model was reduced but still achieved an AUC of 0.804. Therefore, this proof-of-principle exercise suggests that the proposed BDT machine learning model may have wider applications in other registries.

Second, the socioeconomic data we collected in this study are based on donors' home postcodes mapped to the OA level from the 2011 census. It cannot be assumed that people have similar characteristics to those who live in the same area. However, these area measures might be more valid than self-

Table 3
Comparison of the Subgroups of Features Used for Modeling

Features	Total No. of VT Requests (% of All Requests)	No. of Positive Responses to VT (% of Category)	No. of Negative Responses to VT (% of Category)	Odds Ratio for Donor Attrition (95% Confidence Interval)	*P* Value
No. of days since last donor contact					
Less than 6 months*	3176 (31.0)	2523 (79.4)	653 (20.6)	1.00	—
6 months to 2 years	2134 (20.8)	1442 (67.6)	692 (32.4)	2.61 (2.33-2.91)	<.001
2+ years	4948 (48.2)	2605 (52.6)	2343 (47.4)	3.47 (3.13-3.86)	<.001
Ethnicity					
White British*	8911 (86.9)	6006 (67.4)	2905 (32.6)	1.00	—
Asian Bangladeshi	14 (0.1)	6 (42.9)	7 (57.1)	2.41 (0.69-8.69)	.136
Asian Indian	129 (1.3)	65 (50.4)	64 (57.1)	2.04 (1.41-2.92)	<.001
Asian Pakistani	48 (0.5)	24 (50.0)	25 (50.0)	2.15 (1.18-3.95)	.009
Asian other	61 (0.6)	20 (32.8)	41 (67.2)	4.23 (2.42-7.64)	<.001
Black African	51 (0.5)	21 (41.2)	30 (58.8)	2.95 (1.63-5.43)	<.001
Black Caribbean	93 (0.9)	44 (47.3)	49 (52.7)	2.30 (1.50-3.55)	<.001
Black other	9 (0.1)	3 (33.3)	6 (66.7)	4.13 (0.88-25.6)	.067
Chinese	31 (0.3)	12 (38.7)	19 (61.3)	3.27 (1.51-7.40)	.002
Mixed other	60 (0.6)	36 (60.0)	24 (40.0)	1.38 (0.79-2.38)	.269
Mixed white/Asian	56 (0.5)	44 (78.6)	12 (21.4)	0.56 (0.27-1.09)	.086
Mixed white/black African	18 (0.2)	12 (66.7)	6 (33.3)	1.03 (0.32-2.98)	1
Mixed white/black Caribbean	66 (0.6)	39 (59.1)	27 (40.9)	1.43 (0.84-2.40)	.187
Unknown	297 (2.9)	137 (46.1)	160 (53.9)	2.41 (1.90-3.07)	<.001
White Irish	107 (1.0)	57 (53.3)	50 (46.7)	1.81 (1.21-2.71)	.002
White other	307 (3.0)	164 (53.4)	143 (46.6)	1.80 (1.42-2.28)	<.001
NHS Give Blood App status					
Active*	4564 (44.5)	3491 (76.5)	1073 (23.5)	1.00	—
Inactive	165 (1.6)	110 (66.7)	55 (33.3)	1.62 (1.15-2.29)	.005
Unknown	5529 (53.9)	2969 (53.7)	2560 (46.3)	2.81 (2.57-3.06)	<.001
Social grade					
AB*	2521 (24.6)	1685 (66.8)	836 (33.2)	1.00	—
C1	4459 (43.5)	2916 (65.4)	1543 (34.6)	1.07 (0.96-1.18)	.227
C2	872 (8.5)	551 (63.2)	321 (36.8)	1.17 (1.00-1.38)	.051
DE	2313 (22.5)	1366 (59.1)	947 (40.9)	1.40 (1.24-1.57)	<.001
Unknown	93 (0.9)	52(55.9)	41 (44.1)	1.59 (1.02-2.46)	.033
Sex					
Male*	6269 (61.1)	4310 (68.8)	1959 (31.2)	1.00	—
Female	3989 (38.9)	2260 (56.7)	1729 (43.3)	1.68 (1.55-1.83)	<.001
Age when selected					
18-30 years	3088 (30.1)	2124 (68.8)	964 (31.2)	1.00	—
31-40 years	3335 (32.5)	2050 (61.5)	1285 (38.5)	1.38 (1.24-1.53)	<.001
41-50 years	2726 (26.6)	1732 (63.5)	994 (36.5)	1.26 (1.13-1.41)	<.001
50+ years	1109 (10.8)	664 (59.9)	445 (40.1)	1.48 (1.28-1.71)	<.001
Length on registry					
0-5 years	3263 (31.8)	2391 (73.3)	872 (26.7)	1.00	—
6-10 years	3590 (35.0)	2057 (57.3)	1533 (42.7)	2.04 (1.84-2.27)	<.001
10+ years	3405 (33.2)	2122 (62.3)	1283 (37.7)	1.66 (1.49-1.84)	<.001
Education level					
% level 2 and above ≥60*	5429 (52.9)	3612 (66.5)	1817 (33.5)	1.00	—
% level 2 and above <60	4736 (46.2)	2906 (61.4)	1830 (38.6)	1.25 (1.15-1.36)	<.001
Unknown	93 (0.9)	52 (55.9)	41 (44.1)	1.57 (1.01-2.41)	.035
Property ownership					
% own property ≥67*	5767 (56.2)	3851 (66.8)	1916 (33.2)	1.00	—
% own property <67	4398 (42.9)	2667 (60.6)	1731 (39.4)	1.30 (1.20-1.42)	<.001
Unknown	93 (0.9)	52 (55.9)	41 (44.1)	1.58 (1.02-2.44)	.035
Blood donation reliability score					
1*	2416 (23.6)	1888 (78.1)	528 (21.9)	1.00	—

(*continued*)

Table 3 (*Continued*)

Features	Total No. of VT Requests (% of All Requests)	No. of Positive Responses to VT (% of Category)	No. of Negative Responses to VT (% of Category)	Odds Ratio for Donor Attrition (95% Confidence Interval)	*P* Value
2	1839 (17.9)	1344 (73.1)	495 (26.9)	1.31 (1.14-1.52)	<.001
3	646 (6.3)	450 (69.7)	196 (30.3)	1.55 (1.27-1.89)	<.001
4	645 (6.3)	428 (66.4)	217 (33.6)	1.81 (1.49-2.20)	<.001
5	4712 (45.9)	2460 (52.2)	2252 (47.8)	3.27 (2.92-3.67)	<.001

Fisher exact test was used for comparison of the subgroups, and *P* value of less than .05 is considered significant.

* Reference category.

declared and unverified individual-level indicators. We noticed that social grade, education level, and property ownership were contributing toward output prediction. This merits further research to better understand why they have impact on the prediction. As an exploratory analysis, this suggests that using census data as a proxy of socioeconomic data could be an alternative to collecting such information for the specific individuals.

Third, the findings in our study show the feasibility and promise of using ML to predict donor availably. However, several challenges need be addressed before the clinical application of the method. The first challenge to apply this model into practice is information technology development. Our data were collected from 3 different systems (blood donor management system, the BBMR LIMS, and the 2011 UK census database) to train the model. However, substantial effort is needed to synchronize the 3 systems and embed the predictor in an easy-access format so that it can be used effectively. The second challenge is that thorough validation of the proposed ML model is needed before clinical adoption. The proposed BDT model achieved a high degree of accuracy, but false positives (ie, donors predicted to be available but actually unavailable) could result in false hopes to patients. Also, false negatives (ie, donors predicted to be unavailable but actually available) might be neglected during the donor selection process. The final challenge to consider is to engage with the clinicians and specialists to gain their acceptance of integrating a predictive score to assist their clinical decision-making process. The predictive scores need to be integrated appropriately with their workflow, without having an extra load of work to maintain with the new solution.

In conclusion, maximizing donor availability is key to ensuring patients with blood cancers or disorders receive a transplant at the optimum time as delays adversely affect patient outcomes. BBMR used machine learning to analyze donor characteristics, socioeconomic data, blood donation activities, and behaviors and has developed a tool that predicts donor availability with a high degree of accuracy. Further studies are needed to estimate the cost-effectiveness of incorporating a machine learning-based model in practice, and our BDT machine learning model needs to be improved before clinical applications and general applications.

ACKNOWLEDGMENTS

The authors thank the University of Liverpool for their helpful collaboration on machine learning knowledge transfer.

Financial disclosure: This work was funded by the British Bone Marrow Donor Appeal (BBMDA).

Conflict of interest statement: There are no conflicts of interest to report.

Authorship statement: Y.L. conceived and designed the study. Y.L. and A.M. developed and validated the machine learning models, under the supervision of L.G. and P.W., and with clinical input from R.P., A.M., D.W., and H.L. collected the data and ensured quality of the data given to the analysis. Y.L., G.P., and A.H. applied and secured the funding of this project. Y.L. wrote the first draft of the article, which was critically revised and approved by all authors.

APPENDIX A

There are 2052 (20% of the total records) in the testing subset, including 1117 positive responses and 935 negative responses. The prediction accuracy for medical, ability to contact, and personal commitment is 0.685, 0.875, and 0.636, respectively. We also have an “other” unavailability category, which is frequently used in the BBMR but does not provide a meaningful difference from the rest of the categories. The results are summarized in Table 4.

We think the inconsistent accuracy is mainly due to (1) data quality, (2) features selection, and (3) the nature of unavailability reason.

1) Data quality. We are aware that pregnancy-related unavailable reasons have been recorded inconsistently. Sometimes they are recorded in the medical category, and sometimes they are recorded in the “other” category. There was lack of information in the other category when we collected the data, so we were unable to differentiate the other category from the rest. We eliminated the other category and introduced a pregnancy category last year. We will retrain our model once sufficient granularity of data is available. In contrast, the ability-to-contact category is clear and the data quality is relatively good, which might explain the high prediction accuracy.

Table 4
Summary of the Prediction Accuracy for Each Unavailability Category on Testing Data (n = 2052)

Unavailable Reason	No. of Negative Responses to VT in the Testing Set (% All Requests)	True Negative (% of Category)	False Positive (% of Category)
Medical	384 (41.1)	263 (68.5)	121 (31.5)
Ability to contact	279 (29.8)	244 (87.5)	35 (12.5)
Personal commitment	107 (11.4)	68 (63.6)	39 (36.4)
Other	165 (17.6)	120 (72.7)	45 (27.2)
Total	935	695	240

True negative means predicted negative response to a VT and actual negative response to a VT. False positive means predicted positive response to a VT and actual negative response to a VT.

2) Features selection. To improve the prediction accuracy for each category, it might be more appropriate to use a different set of features to train the model for a specific objective, instead of using a set of generic features to predict on all 3 categories. Employment status, household income and density are available from the 2011 UK census data, which might be useful to predict the medical unavailability. This requires further study.

3) The nature of unavailability reason. For the personal commitment category, we have observed that many donors are really willing to donate and eager to help a patient, but their family circumstances and difficulties (eg, loss of family member, young children, carer responsibilities) prevent them from doing so. Such errors in the prediction are inevitable.

APPENDIX B

Approximated social grade [33] is a classification system designed by the Office for National Statistics, which groups people aged 16 years and older into 6 possible categories (A, B, C1, C2, D, and E) based on their socioeconomic status, derived from the British National Readership Survey. For the 2011 census, categories A and E make up a very small proportion of the UK population, so the first 2 categories and the last 2 categories were combined, which is most widely known as the 4-way classification (AB, C1, C2, DE). The description of the social grade can be found in Table 5.

Table 5
Description of the Approximated Social Grade and the Percentage of UK Population in Each Grade

Social Grade	Description	% UK Population
AB	Higher and intermediate managerial, administrative, professional occupations	22.17
C1	Supervisory, clerical and junior managerial, administrative, professional occupations	30.84
C2	Skilled manual occupations	20.94
DE	Semiskilled and unskilled manual occupations, unemployed, and lowest grade occupations	26.05

REFERENCES

1. Ballen KK, King RJ, Chitphakdithai P, et al. The national marrow donor program 20 years of unrelated donor hematopoietic cell transplantation. *Biol Blood Marrow Transplant*. 2008;14:2–7.
2. Gratwohl A, Pasquini M, Aljurf M, et al. One million haemopoietic stem-cell transplants: a retrospective observational study. *Lancet Haematol*. 2015;2:e91–e100.
3. Schetelig J, de Wreede LC, van Gelder M, et al. Risk factors for treatment failure after allogeneic transplantation of patients with CLL: a report from the European Society for Blood and Marrow Transplantation. *Bone Marrow Transplant*. 2017;52:552–560.
4. Mehta J, Gordon LI, Tallman MS, et al. Does younger donor age affect the outcome of reduced-intensity allogeneic hematopoietic stem cell transplantation for hematologic malignancies beneficially? *Bone Marrow Transplant*. 2006;38:95–100.
5. Kollman C, Howe CW, Anasetti C, et al. Donor characteristics as risk factors in recipient after transplantation of bone marrow from unrelated donors: the effect of donor age. *Blood*. 2001;98(7):2043–2051.
6. Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*. 2007;110(13):4576–4583.
7. Pidala J, Lee SJ, Ahn KW, et al. Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation. *Blood*. 2014;124(16):2596–2606.
8. Craddock C, Labopin M, Pillai S, et al. Factors predicting outcome after unrelated donor stem cell transplantation in primary refractory acute myeloid leukaemia. *Leukemia*. 2011;25:808–813.
9. Heemskerk MB, van Walraven SM, Cornelissen JJ, et al. How to improve the search for an unrelated haematopoietic stem cell donor: faster is better than more!. *Bone Marrow Transplant*. 2005;35:645–652.
10. Lown RN, Shaw BE. Beating the odds: factors implicated in the speed and availability of unrelated haematopoietic cell donor provision. *Bone Marrow Transplant*. 2013;48:210–219.
11. Switzer GE, Bruce JG, Myaskovsky L, et al. Race and ethnicity in decisions about unrelated hematopoietic stem cell donation. *Blood*. 2013;121:1469–1476.
12. Lown RN, Marsh SG, Switzer GE, et al. Ethnicity, length of time on the register and sex predict donor availability at the confirmatory typing stage. *Bone Marrow Transplant*. 2014;49:525–531.
13. Shaw BE, Logan BR, Spellman SR, et al. Development of an unrelated donor selection score predictive of survival after HCT: donor age matters most. *Biol Blood Marrow Transplant*. 2018;24(5):1049–1056.
14. Balassa K, Griffiths A, Winstone D, Li Y, Rocha V, Pawson R. Attrition at the final donor stage among unrelated haematopoietic stem cell donors: the British Bone Marrow Registry experience. *Transfusion Medicine*. 2019;29(5):332–337.
15. Onitilo AA, Lin YH, Okonofua EC, Afrin LB, Ariail J, Tilley BC. Race, education, and knowledge of bone marrow registry: indicators of willingness to donate bone marrow among African Americans and Caucasians. *Transplant Proc*. 2004;36(10):3212–3219.
16. Switzer GE, Dew MA, Goycoolea JM, Myaskovsky L, Abress L, Confer DL. Attrition of potential bone marrow donors at two key decision points leading to donation. *Transplantation*. 2004;77:1529–1534.
17. Sivasankaran A, Williams E, Albrecht M, Switzer GE, Cherkassky V, Maiers M. Machine learning approach to predicting stem cell donor availability. *Biol Blood Marrow Transplant*. 2018;24(12):2425–2432.
18. Switzer GE, Dew MA, Stukas AA, Goycoolea JM, Hegland J, Simmons RG. Factors associated with attrition from a national bone marrow registry. *Bone Marrow Transplant*. 1999;24:313–319.
19. Freund Y, Schapire RE. A decision-theoretic generalization of on-Line learning and an application to boosting. *J Comput System Sci*. 1997;55(1):119–139..
20. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol*. 1996;49(11):1225–1231.
21. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20:273–297.
22. Office for National Statistics. 2011Census: population and household estimates for wards and output areas in England and Wales. Available at: http://www.ons.gov.uk/ons/rel/census/2011-census/population-and-household-estimates-for-england-and-wales/index.html. Accessed 16 July 2012.
23. Office for National Statistics. 2011Census: output area (OA). Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#output-area-oa. Accessed 18 January 2016.
24. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority over-sampling technique. *J Artificial Intelligence Res.*. 2002;16:321–357.
25. Breiman L. Random forests. *Machine Learning*.. 2001;45(1):5–32.
26. Fisher A, Rudin C, Dominici F. Model class reliance: variable importance measures for any machine learning model class, from the "Rashomon" perspective. 2018. http://arxiv.org/abs/1801.01489.
27. Dorsey R, Graham G. New HHS data standards for race, ethnicity, sex, primary language, and disability status. *JAMA*. 2011;306(21):2378–2379.
28. Hollenbach JA, Saperstein A, Albrecht M, et al. Race, ethnicity and ancestry in unrelated transplant matching for the National Marrow Donor Program: a comparison of multiple forms of self-identification with genetics. *PLoS ONE*.. 2015;10:(8) e0135960.
29. Gullickson A, Morning A. Choosing race: multiracial ancestry and identification. *Soc Sci Res*. 2011;40:498–512.
30. Yoo KD, Noh J, Lee H, et al. A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: a multicentre cohort study. *Sci Rep*.. 2017;7:8904.
31. Liu X, Fae L, Kale A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health*. 2019;1(6):e271–e297.
32. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388–2396.
33. Lambert H, Moy C. Social grade allocation to the 2011 census. 2013. Available at: https://www.mrs.org.uk/pdf/Social%20Grade%20Allocation%20for%202011%20Census.pdf. Accessed March 2013.