Explainable hemoglobin deferral predictions using machine learning models: interpretation and consequences for the blood supply

Marieke Vinkenoog[1,2], Matthijs van Leeuwen[2], and Mart P. Janssen[1]

[1]**Donor Medicine Research, Sanquin Research, Amsterdam, the Netherlands**
[2]**Leiden Institute of Advanced Computer Science, Leiden University, Leiden, the Netherlands**

ABSTRACT

Background - Accurate predictions of hemoglobin deferral for whole-blood donors could aid blood banks in reducing deferral rates and increasing efficiency and donor motivation. Complex models are needed to make accurate predictions, but predictions must also be explainable. Before the implementation of a prediction model, its impact on the blood supply should be estimated to avoid shortages.
Methods - Donation visits between October 2017 and December 2021 were selected from Sanquin's database system. The following variables were available for each visit: donor sex, age, donation start time, month, number of donations in the last 24 months, most recent ferritin level, days since last ferritin measurement, hemoglobin at *n*th previous visit (*n* between 1 and 5), days since the *n*th previous visit. Outcome hemoglobin deferral has two classes: deferred and not deferred. Support vector machines were used as prediction models, and SHapley Additive exPlanations values were used to quantify the contribution of each variable to the model predictions. Performance was assessed using precision and recall. The potential impact on blood supply was estimated by predicting deferral at earlier or later donation dates.
Results - We present a model that predicts hemoglobin deferral in an explainable way. If used in practice, 64% of non-deferred donors would be invited on or before their original donation date, while 80% of deferred donors would be invited later.
Conclusions - By using this model to invite donors, the number of blood bank visits would increase by 15%, while deferral rates would decrease by 60% (currently 3% for women and 1% for men).

INTRODUCTION

Sanquin, the Dutch national blood service, collects over 400 000 whole-blood donations from non-remunerated, voluntary blood donors every year. Women may donate a maximum of three times per year, and men five times. Hemoglobin levels are tested before every donation to prevent blood collection from donors with insufficient iron. The minimum hemoglobin level for blood donation is 7.8 mmol/L for women and 8.4 mmol/L for men; if the capillary hemoglobin test (HemoCue) shows a lower value, the donor is deferred for 3 months, that is, sent home without donating blood. If the hemoglobin value is more than 0.5 mmol/L below the donation threshold, the donor is referred to a donor physician. Additionally, since October 2017, ferritin levels have been measured in each new donor, as well as after every fifth donation in repeat donors. Donors are deferred for 6 months if their ferritin level is between 15 and 30 µg/L, or for 12 months if their ferritin level is below 15 µg/L. This ferritin deferral policy was implemented because hemoglobin is a poor indicator of iron stores, as iron deficient donors can still present with sufficient hemoglobin levels until the iron deficiency is very severe.

While it is important to defer donors that do not meet donation requirements, sending donors home without giving them the opportunity to donate is discouraging and costly. Previous studies have shown that donors are less likely to return to the blood bank after a deferral for low hemoglobin than after

a successful donation, especially if it concerns their first blood bank visit. [1] This is less likely after deferral for low ferritin levels, which occurs by letter after the donation, indicating that post-donation deferral is less demotivating for donors than on-site deferral. [2] The implementation of ferritin testing has had a considerable impact on the blood supply, as a large part of the existing donor population (53% of women and 42% of men) were found to have ferritin levels below 30 µg/L and had to be deferred. [3] However, this has had the intended positive impact on donor deferral rates due to low hemoglobin, which decreased from 8% for women and 3% for men in 2016 to 3% for women and 1% for men in 2021. [4]

Although percentage-wise, hemoglobin deferral rates are quite low in the Netherlands, they still amount to about 8000 deferrals each year, and there is a risk of permanently losing these donors. To reduce deferral rates and improve donor motivation, we should re-think hemoglobin deferral policies. One tool that can be used for this purpose is a hemoglobin deferral prediction model. Many of these prediction models have already been developed, including models that predict personalised donation intervals. [5, 6, 7] Prediction models can be used in the donor invitation process by predicting hemoglobin deferral for eligible donors and only inviting those donors that are predicted to not be deferred. Because deferred donors are only a small proportion of the total donor population, it has proven difficult to accurately identify them, and hence prediction models are not used in practice yet.

We present a novel machine learning hemoglobin deferral prediction model based on donor characteristics and donation history. New in our approach is that we use SHapley Additive exPlanations [8] to explain how the model uses the variables in its predictions and relate these explanations to known physiological processes. This gives valuable insight into the associations that are learned by the model; if prediction models are to be used to make decisions in practice, the user must understand how the model makes these decisions. Moreover, we show the potential impact that prediction models can have on the total blood supply, if these are to be used to guide donor invitations, by calculating deferral probabilities at multiple time points for each donor. By both explaining the predictions and assessing the impact of the model on the blood supply, we remove two important limitations that currently prevent blood services from implementing prediction models.

METHODS

Data

Data on blood bank visits by whole-blood donors were extracted from Sanquin's database system eProgesa, for donations. Only data from donors who explicitly provided informed consent for the use of their data for scientific research were used. This consent is given by more than 99% of all donors. For each visit, the following information was collected: donor sex, donor age, donation date, donation (registration) time, hemoglobin level and ferritin level. Ferritin is measured at every new donor intake and upon every fifth donation in repeat donors. Therefore, ferritin levels are unavailable for most donations. By using these data, predictor variables were calculated for each visit, as described in Table 1.

In total, 938 710 blood bank visits (excluding new donor intakes and donation types other than whole blood) by 241 131 unique donors were registered between October 2017 and December 2021. After excluding visits for which no previous ferritin measurement was available, 458 615 blood bank visits by 157 423 unique donors remained for the analysis.

The outcome variable *HbOK* is dichotomous; deferral (hemoglobin level below the eligibility threshold for donation) or non-deferral (hemoglobin equal to or above the threshold).

Analyses

Support vector machines (SVMs) [9] are used to predict hemoglobin deferral. SVMs are supervised machine learning models that find the optimal hyperplane separating the outcome classes based on the predictor variables of a so-called training set. After fitting the model on the training set, the model can predict the outcome class of unseen observations called the test set. It also gives the probability of an observation belonging to each outcome class. We chose SVMs as a classification algorithm because all predictor variables are numeric, and it is computationally less expensive than, for instance, K-nearest neighbours or (dynamic) linear mixed models.

For each sex, five SVMs were trained, named SVM-n for n between one and five, indicating the number of previous blood bank visits (*HbPrevn* and *DaysSinceHbn*) used as predictor variables. Donors are only included in SVM-n if they have at least n previous visits; therefore, sample sizes decrease from SVM-1 to SVM-5. Blood bank visits before 2021 were used as the training set, while visits in 2021 were

Variable	Unit or values	Description
Sex	male, female	Biological sex of the donor; separate models are trained for men and women
Age	years	Donor age at time of donation
Time	hours	Registration time when the donor arrived at the blood bank
Month	1–12	Month of the year that the visit took place
NumDon	count	Number of successful (collected volume >250 ml) whole-blood donations in the last 24 months
FerritinPrev	µg/L	Most recent ferritin level measured in this donor
DaysSinceFer	days	Time since this donor's last ferritin measurement
HbPrevn	mmol/L	Hemoglobin level at *n*th previous visit, for *n* between 1 and 5
DaysSinceHbn	days	Time since related hemoglobin measurement at *n*th previous visit, for *n* between 1 and 5

Table 1. All predictor variables used in the prediction models.

Metric	Outcome class	Definition
Precision	Deferral	The proportion of donations correctly classified as deferrals by the model, out of all donations classified as deferrals.
Recall	Deferral	The proportion of donations correctly classified as deferrals by the model, out of all donations classified as true deferrals.
Precision	Non-deferral	The proportion of true non-deferrals, out of all predicted non-deferrals.
Recall	Non-deferral	The proportion of predicted non-deferrals, out of all true non-deferrals.

Table 2. Interpretation of performance metrics.

used as the test set to validate performance on unseen data. This division was chosen over a random training/test division because if these models were used in practice, they would be trained on all historical data and applied to future data. We used a paired t-test to assess the difference in deferral rates between training and test sets of donors of the same sex with the same number of previous donations. To assess the generalisability of the model to new donors, we did a separate experiment in which the test set is comprised of the last blood bank visit of 20% of all unique donors, and the training set includes all donations from the remaining 80% of donors.

For each of the 10 models, that is, SVM-1 through SVM-5 for both sexes, hyperparameters were optimised separately, using stratified (on the outcome variable) five-fold cross-validation within the training set data (and thus not using the test data). Hyperparameters were optimised using grid search, using balanced accuracy as a scoring method, defined as the weighted average of recall in both classes (see Table 2 for the definition of recall). This method is especially suitable for imbalanced datasets because it uses class-balanced sample weights to determine the average recall.

Precision and recall were determined and compared for training and test datasets for each model. Both metrics are calculated for both outcome classes. A practical interpretation of these metrics is given in Table 2.

To explain the model predictions, we used SHapley Additive exPlanations (SHAP) values, a model agnostic explainer. SHAP values show the contribution of each variable to the prediction for each individual observation, which is even more informative than coefficients returned by, for example, linear models. By summarizing observation-based contributions, we obtain variable importance measures for a model that does not have interpretable coefficients.

Potential impact on the blood supply

We assessed the potential impact of using SVMs to guide donor invitations by predicting deferral for all blood bank visits that took place in 2021 (the test set). For each observation, we used information of all previous blood bank visits (up to five) available as predictor variables. This means that SVM-1 is used when only one previous visit is available, SVM-2 if there are two previous visits, etc.

Model	Training		Test	
	Women	Men	Women	Men
SVM-1	128 173 (4084; 3.19%)	121 746 (1339; 1.10%)	110 372 (3696; 3.35%)	98 324 (1074; 1.09%)
SVM-2	83 532 (2884; 3.45%)	96 441 (1133; 1.17%))	85 131 (3065; 3.60%)	84 000 (984; 1.17%)
SVM-3	59 720 (2032; 3.40%)	79 690 (997; 1.25%)	67 167 (2451; 3.65%)	72 576 (902; 1.24%)
SVM-4	47 317 (1494; 3.16%)	67 934 (887; 1.31%)	54 090 (1874; 3.46%)	63 447 (806; 1.27%)
SVM-5	40 604 (1113; 2.74%)	59 611 (768; 1.29%)	45 208 (1378; 3.05%)	55 582 (699; 1.26%)

Table 3. Sizes of training and test datasets per model. The number and percentage of deferrals is given in brackets.

If prediction models are to be used in practice, they should estimate the deferral probability for different days in the future and invite a donor for the first occurrence where the non-deferral probability would exceed a preset value. To simulate this, we predicted hemoglobin deferral each week from 1 year before the original donation date to 1 year after, by adjusting all time-related variables. If the predicted donation interval were to be less than the minimum donation interval (57 days for men, 122 days for women), the latter would be applied.

We compare all original donation intervals with the donation intervals as proposed by the model. Dividing the sum of the original donation intervals by the sum of the model-guided donation intervals gives the relative change in blood bank visits per time unit and hence the relative yield of blood donations.

Software

All analyses were performed in Python 3.9, using modules numpy [10] and pandas [11] for data processing, sklearn [12] for model training and predictions, shap [8] for calculating SHAP values, and matplotlib [13] for creating graphs. The analysis code is available as a GitHub repository and indexed on Zenodo at https://doi-org.ezproxy.leidenuniv.nl/10.5281/zenodo.6938112.

RESULTS

Table 3 shows the sample sizes of training and test datasets for each model. Deferral rates in the training datasets are 3.19% (SD 0.28) for women and 1.22% (SD 0.09) for men; in the test sets, they are 3.42% (SD 0.24) for women and 1.21% (SD 0.08) for men. Using a paired t-test, the difference in deferral rate between the training and test datasets is significant for women ($p = 0.002$) but not for men ($p = 0.070$). No correction was made for the differing deferral rates, as the models are intended for future predictions, and in practice, the deferral rate of future blood bank visits is unknown. Also, a change in deferral rate should be correctly predicted by the model if the mechanism causing this change can be learned from the data. Deferral rates differ between models due to small differences in the data between subsets of the data (see Table 4). This is not a problem as long as the same associations between predictor variables and outcome are found in all subsets of the data, which is described in the feature importance part of the results.

Although the training datasets consist of 3 years of data, and the test datasets of only 1 year, their sizes are similar and sometimes the test dataset is even larger. This is because donations are only included from donors for whom at least one ferritin measurement was available. As ferritin screening was implemented using a stepped wedge approach (the first blood bank locations started in October 2017, but only in November 2019 all locations were included), the number of donors that could be included in the training dataset was limited. [4]

Marginal distributions of predictor variables are described in Table 4. As the number of previous donations increases, the median age increases from 30 to 36 years for women and from 34 to 38 for men. The median values of the last ferritin measurement decreased from 47 µg/L in SVM-1 to 39 µg/L in

Previous visits	≥ 1	≥ 2	Women ≥ 3	≥ 4	≥ 5
Age	30 (23–47)	32 (24–48)	34 (25–50)	35 (26–51)	36 (37–52)
NumDon	1 (0–3)	2 (1–3)	3 (2–4)	3 (2–4)	3 (3–4)
FerritinPrev	47 (33–74)	46 (33–70)	44 (32–65)	41 (31–59)	39 (29–55)
DaysSinceFer	237 (125–420)	329 (197–497)	383 (260–547)	400 (230–572)	372 (204–567)
HbPrev1	8.5 (8.1–8.9)	8.5 (8.1–8.9)	8.5 (8.1–8.9)	8.5 (8.1–8.9)	8.5 (8.1–8.9)
DaysSincePrev1	135 (105–196)	154 (132–211)	158 (132–217)	167 (133–224)	173 (133–236)
HbPrev2		8.5 (8.1–8.9)	8.5 (8.1–8.9)	8.5 (8.1–8.8)	8.5 (8.1–8.8)
DaysSincePrev2		302 (255–412)	328 (271–445)	336 (273–468)	349 (280–493)
HbPrev3			8.5 (8.1–8.8)	8.4 (8.1–8.8)	8.4 (8.1–8.8)
DaysSincePrev3			482 (398–644)	511 (420–674)	528 (430–696)
HbPrev4				8.4 (8.1–8.8)	8.4 (8.1–8.8)
DaysSincePrev4				674 (553–871)	709 (581–904)
HbPrev5					8.4 (8.1–8.8)
DaysSincePrev5					877 (721–1107)
Previous visits	≥ 1	≥ 2	Men ≥ 3	≥ 4	≥ 5
Age	34 (26–48)	35 (27–49)	36 (27–50)	37 (28–51)	38 (28–51)
NumDon	3 (1–5)	4 (2–5)	4 (3–6)	5 (3–6)	5 (4–6)
FerritinPrev	77 (44–141)	66 (40–126)	57 (38–108)	52 (36–89)	47 (35–73)
DaysSinceFer	200 (100–335)	232 (151–365)	257 (177–378)	271 (186–385)	267 (173–387)
HbPrev1	9.4 (9.0–9.9)	9.4 (9.0–9.9)	9.4 (9.0–9.8)	9.4 (8.9–9.8)	9.4 (8.9–9.8)
DaysSincePrev1	81 (63–133)	90 (67–147)	92 (69–160)	98 (70–168)	105 (70–176)
HbPrev2		9.4 (9.0–9.8)	9.4 (9.0–9.8)	9.4 (8.9–9.8)	9.4 (8.9–9.8)
DaysSincePrev2		185 (128–287)	196 (147–302)	210 (153–315)	219 (158–330)
HbPrev3			9.4 (9.0–9.8)	9.4 (9.0–9.8)	9.4 (8.9–9.8)
DaysSincePrev3			302 (225–441)	322 (238–463)	335 (245–485)
HbPrev4				9.4 (8.9–9.8)	9.4 (8.9–9.8)
DaysSincePrev4				424 (315–600)	444 (330–620)
HbPrev5					9.4 (8.9–9.8)
DaysSincePrev5					552 (416–752)

Table 4. Marginal distributions of predictor variables, represented by median and interquartile ranges.

SVM-5 for women and from 77 to 47 µg/L for men. The median time between consecutive donations increases from SVM-1 to SVM-5, while previous hemoglobin levels are consistent across models, as well as different numbers of previous visits.

Accuracy and model fit

Figure 1 compares precision and recall for class non-deferral across all models. Performance on the training and test sets are similar, indicating that the models are well-fitted. Both precision and recall increase as more previous blood bank visits are used to make predictions. Re-running all models only on donors with at least five previous blood bank visits did not change this observed increase in performance. The models handle the difference between the proportions of deferral in the training and test set very well: comparing the observed difference in deferral proportion in the training and test set to the predicted difference, the mean difference of these differences is only 0.05 percentage points (maximum: 0.12 percentage points). This indicates that the models are robust against (modest) changes in deferral rates.

Performance on a test set of unseen donors

Precision and recall for both outcome classes are similar for the different types of splits in training and test set. Table 5 shows the comparison in performance between the time split and the random split, as described in the methods section. Metrics are shown for SVM-5; the differences are smaller for all other models. For women, the random split has a higher precision and recall than the time split. For men, this is the other way around. For both sexes, the differences are minimal.

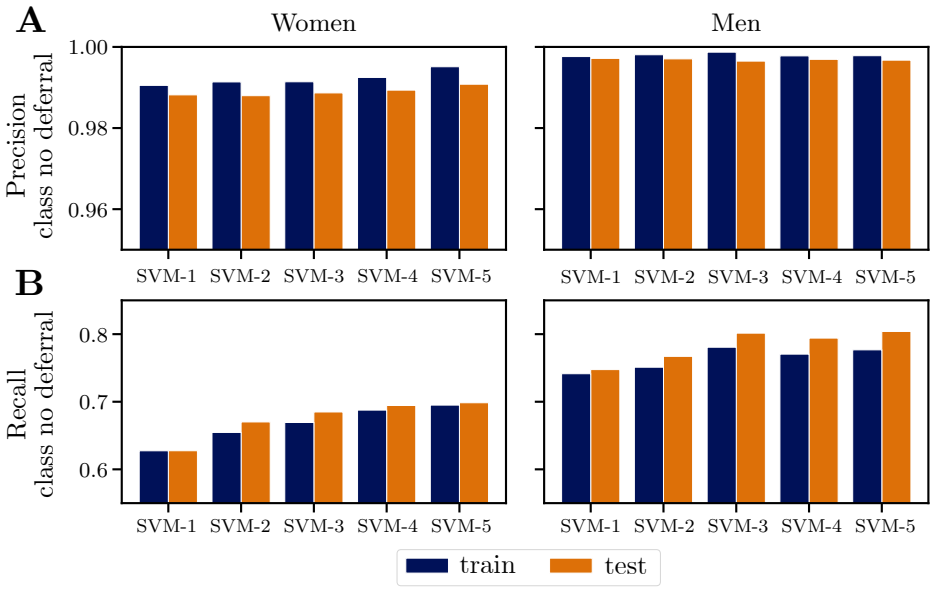

Figure 1. Performance metrics for all models. (A): Precision of class non-deferral; the proportion of successful donations among all predicted non-deferrals. The complement of the precision is the deferral rate, should the model be used to guide invitations. (B): Recall of class non-deferral; the proportion of successful donations that are predicted correctly. The complement of the recall is the proportion of missed donations, should the model be used to guide invitations. Note that the y-axes in are zoomed in to highlight the differences between various models.

Sex	Metric	Time split	Random split	Difference
Women	Precision	0.991	0.994	-0.003
	Recall	0.698	0.701	-0.003
Men	Precision	0.997	0.996	+0.001
	Recall	0.804	0.791	+0.013

Table 5. Precision and recall for outcome class non-deferral, compared between two different training/test splits.

Feature importance and explanation of predictions

SHAP values were computed based on a random subset of 100 donations in the test set. Figure 2 shows the SHAP summary plot for the SVM-5 models, the summary plots for the other eight models are included in the online supplement of the published paper.

For all models, the most important predictor variable is the previous hemoglobin measurement (*HbPrev1*), and in general, more recent measurements are more important than earlier ones. The time since the previous hemoglobin measurements also ranks high on feature importance, but their chronological order is less well-preserved than the *HbPrev* variables.

The association between the feature value and impact on the prediction is as expected for most variables. For hemoglobin measurements, higher values are associated with predicted non-deferral. For *DaysSinceHb*, longer times since the previous hemoglobin measurement are indicative of predicted non-deferral. However, *DaysSinceHb4* shows the opposite association, meaning that when the fourth previous measurement was long ago, the chance of predicted non-deferral becomes lower, while higher would be expected.

Variable *NumDon* has the expected impact on prediction in all models but SVM-5 for female donors; in all other models, a higher number of recent donations shifts the prediction towards deferral. In most models, the number of donations is a more important predictor for men than for women, and it is always less important than all *HbPrev* variables.

The variable *FerritinPrev* shows the same association with the prediction as *HbPrev* variables: higher ferritin levels are associated with predicted non-deferral. Ferritin is a more important predictor for men than for women. For both sexes, the time since the previous ferritin measurement is more important than the actual ferritin level, and a higher value for *DaysSinceFer* makes predicted deferral more likely.

We know that for women, higher age makes deferral less likely (due to menopause), and the SHAP values confirm this relation. For men, age is one of the least important predictors, and there is no clear direction of the relation. The month of donation is of medium importance for both sexes, with predicted deferral being more likely earlier in the year. This captures the seasonal effect of temperature on hemoglobin as measured by the HemoCue. Donating earlier in the day (i.e., a lower value for variable *Time*) increases the likelihood of predicted non-deferral, which is supported by previous research showing that hemoglobin levels are highest in the morning and decrease throughout the day. [14]

Impact on blood supply

Figure 3 shows the cumulative count of donors as invited by the models relative to their original donation date. Once the model predicts non-deferral, it never predicts deferral at a later date. Of non-deferred donors, 50% would be invited more than 2 weeks earlier by the model, and 26% within 2 weeks from around the original donation date. Only 5% would not be invited within a year, causing a successful donation to be missed. Of deferred donors, only 13% would be invited earlier, while 40% would be invited over 3 months later. 28% would not be invited within 1 year. The majority of donors would be invited around their original donation date. For many donors, the original donation date was shortly after the minimum donation interval had passed, and as such, there was no room to invite them earlier.

Because the true hemoglobin level of donors on days other than their original donation date is unknown, we must make assumptions about the accuracy of the predictions in order to calculate a hypothetical number of donations and deferrals. In the most optimistic scenario, all donors who were not deferred on their original donation date would also not be deferred if they were invited earlier; and all donors who were deferred on their original donation date but are invited later by the model would not be deferred by then. In that scenario, only 5% of successful donations would be lost because the model would (incorrectly) not invite those donors, while the deferral rate would decrease by 60% (from 3% to 1% for women and from 1% to 0.4% for men).

We estimate the impact on the blood supply by comparing the length of the original donation interval to the donation interval as suggested by the model. For women, the median time between two donations decreases from 157 to 127 days using the prediction model. For men, the median time decreases from 92 to 63 days. Therefore, the total number of blood bank visits per time unit would increase by a maximum of 15%. This assumes that all donors who responded to the original invitation would also respond to the invitation if it would be sent at an earlier or later date. We also assume that all donors visit the blood bank within 1 week of the invitation. With the original invitations, 15% of donors that responded to the invitation visited the blood bank within 8 days, so the 15% increase in visits is likely to be a small

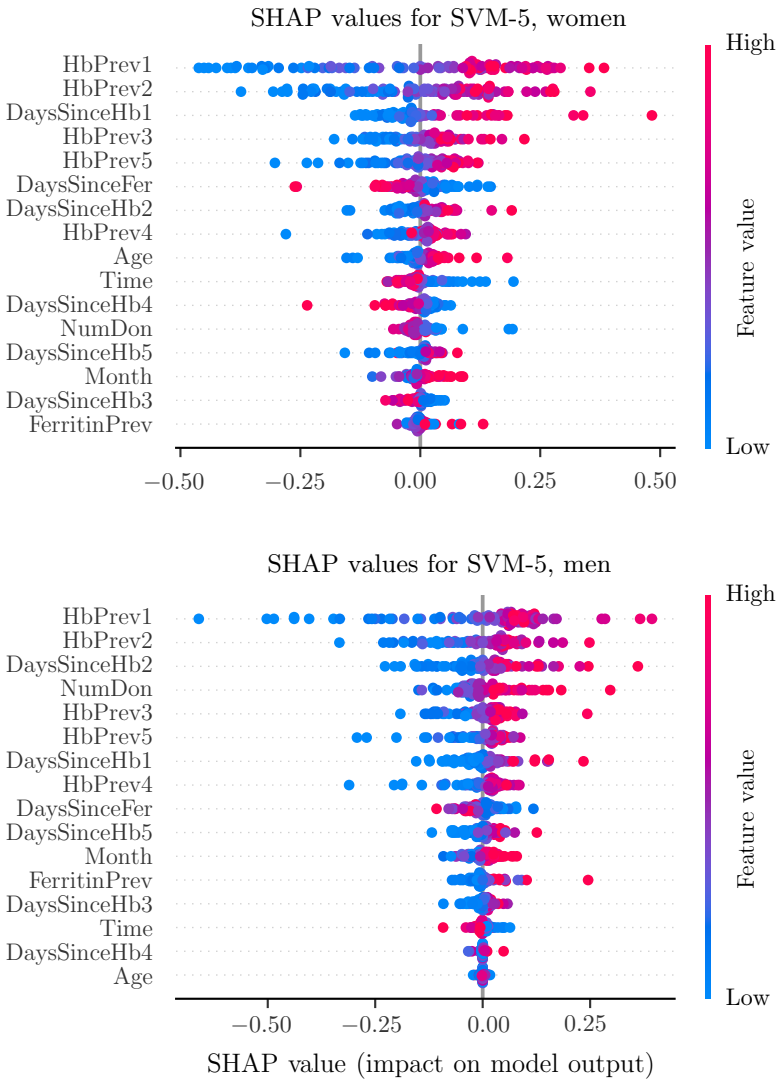

Figure 2. SHAP summary plots for predictions made by SVM-5, on 100 random donations from the test set. Each point represents one single observed donation. The location on the x-axis indicates the contribution of the predictor variable on the prediction (positive value: indicative class non-deferral, negative: indicative of class deferral) while the colour of the point indicates the relative value of the feature in that observation. The features on the y-axis are ordered by their relative importance, measured as the mean absolute SHAP value.

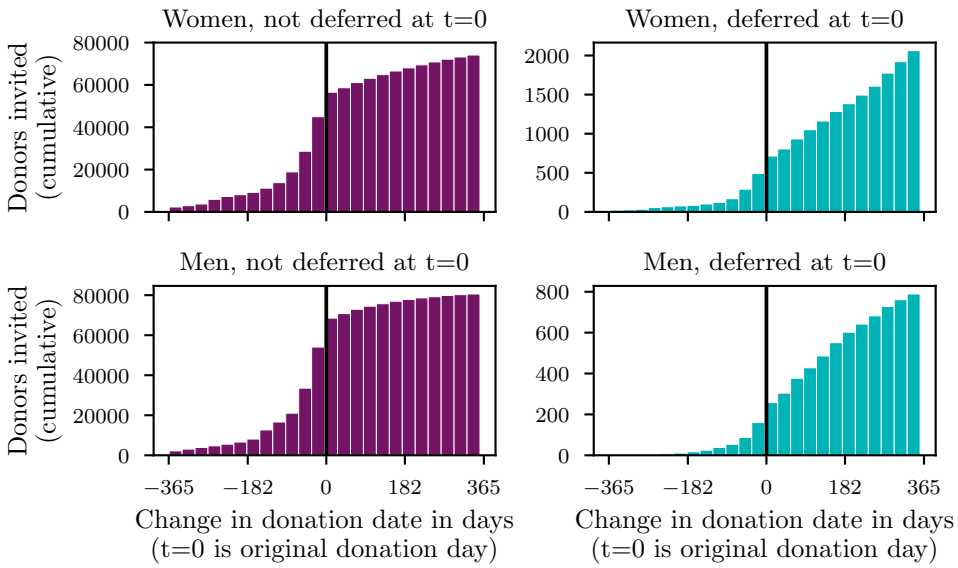

Figure 3. Cumulative distribution of the timing of donor invitations on basis of first predicted hemoglobin level above the donation threshold relative to the original donation date.

overestimation. These assumptions may not hold for mobile donation sites but are reasonable for all regular donation sites, where 95.3% of all visits in our data occurred.

DISCUSSION

This study presents an explainable machine learning approach to predict hemoglobin deferral in whole-blood donors using the information on previous donations and various donor characteristics. We show that we can prevent up to 60% of on-site low hemoglobin deferrals using the model to guide donor invitations.

To our knowledge, this is the first model using machine learning for explainable hemoglobin deferral prediction. An explainable model outcome is crucial for prediction models that are to be used in the context of a decision-support system concerning humans. SHAP values show that our models are able to learn biologically sensible associations. They support findings from other prediction models that found the previous hemoglobin value to be the best predictor for future deferral. We add to this by showing that including more previous donations will improve these predictions.

Although most associations found by SHAP values can be explained biologically, some seem to be caused by organisational policies. Higher values for *DaysSinceFer* are associated with predicted deferral; the opposite association is found for *DaysSinceHb* variables. For donors with fewer than five donations since the start of ferritin testing, the only ferritin measurement is the one taken at their new donor intake, and therefore the time since that previous ferritin measurement is equal to the time since their new donor intake. It is known that deferral becomes more likely once a donor has been donating for a longer period of time.

The precision of class deferral is low, meaning that the predicted deferral is wrong for a substantial proportion of donors. However, by predicting deferral for different timepoints, we see a clear difference between deferred and non-deferred donors: non-deferred donors are in many cases invited earlier than their original donation date by the model, while deferred donors are mostly invited later or not at all, thereby reducing the deferral rate. In non-deferred donors, the median donation interval becomes shorter if invitations were guided by the model, and thus the number of blood bank visits per time unit would increase.

We can only calculate the accuracy of deferral predictions on the original donation date, as hemoglobin levels on other days are unknown. As hemoglobin levels slowly increase after a donation, non-deferred donors would also not be deferred if they were invited later. If they are invited earlier, we cannot know if their hemoglobin level is already above the deferral threshold. The same applies to deferred donors

that are invited later by the model - it is plausible that their hemoglobin levels are above the threshold then, but not certain. Based on accuracy measures of predictions on the original donation dates, we can be fairly confident that the predictions are reliable.

Incorporating prediction models in hemoglobin deferral policies could bring many benefits to blood banks, but it is important to think about how they should be used. If the model is used in practice, the change in policy will lead to changes in the data. Models would therefore need updating by re-training on a regular basis. Additionally, it would be wise not to outsource invitations to the model completely, as that would hinder the model's ability to learn from its mistakes. Although deferrals incorrectly predicted to be non-deferrals would be discovered, we would never know how many donors were incorrectly not invited. This can be prevented by sending part of the invitations without using the model's predictions. In addition to using the model to predict deferral outcomes, the model can also be used to return a deferral probability, allowing blood banks to incorporate this probability in their risk assessment when inviting donors.

Our model is limited to predictor variables that are presently collected by Sanquin. Additional variables could be considered to improve prediction accuracy. Donor height and weight (optionally BMI or total blood volume), as well as smoking status, are examples known to be related to iron levels and are relatively easy to be included. Information on iron-related genetic markers or donor diet may also improve accuracy but are expensive to collect.

Based on the results of this study, we conclude that using prediction models to guide donor invitations would bring multiple advantages to blood banks: lower deferral rates combined with shorter donation intervals would result in motivated and healthy donors, as well as a steady blood supply.

REFERENCES

[1] Brian Custer, Artina Chinn, Nora V Hirschler, Michael P Busch, and Edward L Murphy. "The consequences of temporary deferral on future whole blood donation". In: *Transfusion* 47.8 (2007), pp. 1514–1523.

[2] Marloes LC Spekman, Steven Ramondt, and Maike G Sweegers. "Whole blood donor behavior and availability after deferral: consequences of a new ferritin monitoring policy". In: *Transfusion* 61.4 (2021), pp. 1112–1121.

[3] Marieke Vinkenoog, Katja van den Hurk, Marian van Kraaij, Matthijs van Leeuwen, and Mart P Janssen. "First results of a ferritin-based blood donor deferral policy in the Netherlands". In: *Transfusion* 60.8 (2020), pp. 1785–1792.

[4] Maike G. Sweegers, Saurabh Zalpuri, Franke A. Quee, Elisabeth M. J. Huis in 't Veld, Femmeke J. Prinsze, Emiel O. Hoogendijk, Jos W. R. Twisk, Anton W. M. van Weert, Wim L. A. M. de Kort, and Katja van den Hurk. "Ferritin measurement IN Donors—Effectiveness of iron Monitoring to diminish iron deficiency and low haemoglobin in whole blood donors (FIND'EM): study protocol for a stepped wedge cluster randomised trial". In: *Trials* 21.1 (Oct. 2020), p. 823.

[5] W Alton Russell, David Scheinker, and Brian Custer. "Individualized risk trajectories for iron-related adverse outcomes in repeat blood donors". In: *Transfusion* 62.1 (2022), pp. 116–124.

[6] AM Baart, WLAM De Kort, KGM Moons, and Y Vergouwe. "Prediction of low haemoglobin levels in whole blood donors". In: *Vox Sanguinis* 100.2 (2011), pp. 204–211.

[7] Kazem Nasserinejad, Joost van Rosmalen, Wim de Kort, Dimitris Rizopoulos, and Emmanuel Lesaffre. "Prediction of hemoglobin in blood donors using a latent class mixed-effects transition model". In: *Statistics in medicine* 35.4 (2016), pp. 581–594.

[8] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[9] William S Noble. "What is a support vector machine?" In: *Nature biotechnology* 24.12 (2006), pp. 1565–1567.

[10] Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362.

[11] Wes McKinney et al. "Data structures for statistical computing in python". In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX, 2010, pp. 51–56.

[12] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[13] John D Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in science & engineering* 9.03 (2007), pp. 90–95.

[14] Lauren Berkow. “Factors affecting hemoglobin measurement”. In: *Journal of clinical monitoring and computing* 27 (2013), pp. 499–508.