Received: 22 October 2022 | Revised: 2 March 2023 | Accepted: 13 March 2023

DOI: 10.1002/hed.27353

WILEY

ORIGINAL ARTICLE

Development of machine learning and multivariable models for predicting blood transfusion in head and neck microvascular reconstruction for risk-stratified patient blood management

Behrus Puladi MD, DMD[1,2,3] | **Mark Ooms MD**[1] | **Annette Rieg MD, PhD**[4] | **Max Taubert MD**[5] | **Ashkan Rashad MD, DMD**[2,3] | **Frank Hölzle MD, DMD, PhD**[2,3] | **Rainer Röhrig MD, PhD**[2,3] | **Ali Modabber MD, DMD, PhD**[1]

[1]Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Aachen, Germany

[2]Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany

[3]SMITH Consortium of the German Medical Informatics Initiative in Aachen, Aachen, Germany

[4]Department of Anaesthesiology, University Hospital RWTH Aachen, Aachen, Germany

[5]Center for Pharmacology, Department I of Pharmacology, Medical Faculty, University of Cologne, Cologne, Germany

Correspondence
Behrus Puladi, Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany.
Email: bpuladi@ukaachen.de

Funding information
Federal Ministry of Education and Research (Germany), Grant/Award Number: 01ZZ1803B

Abstract

Background: Although blood transfusions have adverse consequences for microvascular head and neck reconstruction, they are frequently administered. Pre-identifying patients would allow risk-stratified patient blood management.

Methods: Development of machine learning (ML) and logistic regression (LR) models based on retrospective inclusion of 657 patients from 2011 to 2021. Internal validation and comparison with models from the literature by external validation. Development of a web application and a score chart.

Results: Our models achieved an area under the receiver operating characteristic curve (ROC-AUC) of up to 0.825, significantly outperforming LR models from the literature. Preoperative hemoglobin, blood volume, duration of surgery and flap type/size were strong predictors.

Conclusions: The use of additional variables improves the prediction for blood transfusion, while models seems to have good generalizability due to surgical standardization and underlying physiological mechanism. The ML models developed showed comparable predictive performance to an LR model. However, ML models face legal hurdles, whereas score charts based on LR could be used after further validation.

Abbreviations: ALT, anterolateral thigh; ASA, American Society of Anesthesiologists physical status classification system; AUC, area under curve; CCI, Charlson comorbidity index; FFP, fresh frozen plasma; GBM, gradient boosting machine; HB, hemoglobin; HCT, Hematocrit; HR, hazard ratio; LR, logistic regression; ML, machine learning; NNET, neuronal network; OR, odds ratio; PC, platelet concentrate; PRBC, packed red blood cells; RDF, random decision forest; ROC, receiver operating characteristic; SVM, support vector machine.

KEYWORDS
blood transfusion, clinical prediction model, machine learning, microvascular reconstruction, patient blood management

1 | INTRODUCTION

Microvascular flaps are commonly used to cover defects in the head and neck region, especially those resulting from malignant neoplasms, and to enable aesthetic and functional reconstructions. This type of surgery is technically complex, results in at least two major wounds,[1] is accompanied by high blood loss, and often requires the administration of allogeneic blood.[2] However, allogeneic blood transfusions can lead to transfusion-induced immunomodulation[3,4] and thus have an unfavorable effect on overall survival (HR 2.011, 95% CI [1.274, 3.174]) and recurrence-free survival (HR 1.621, 95% CI [1.043, 2.520]).[5] Furthermore, a recent systematic review and meta-analysis showed an increased risk of medical and surgical complications (OR 1.64, 95% CI [1.23, 2.21]) and hospital readmission (OR 1.53, 95% CI [1.29, 1.81]) after blood transfusions in this patient population.[2] Even though blood transfusions should therefore be avoided as much as possible, on average, 29.2% of head and neck surgery patients receive a blood transfusion. Depending on the site and patient population, this rate can be as high as 90.3%.[2]

According to guidelines, transfusion of blood is triggered primarily by established thresholds for intraoperative and postoperative hemoglobin (HB) and hematocrit (HCT) levels, respectively.[6,7] A blood transfusion is generally mandatory if the HB value is below 7 g/dl and in elderly patients (>65 years) and patients with significant cardiovascular disease at an HB value of 8 g/dl. If there is evidence of anemic hypoxia, transfusion should be performed below an HB level of 10 g/dl.[7] Nevertheless, a systematic review and meta-analysis demonstrated that patient blood management (PBM) measures resulted in a 39% reduction in the blood transfusion rate (RR 0.61, 95% CI [0.55, 0.68]) compared to no PBM measures during surgical procedures.[8] PBM is based on three pillars: the optimization of red cell mass, the reduction of blood loss and bleeding, and the optimization of the patient's physiological tolerance to anemia.[9] Therefore, PBM includes various preoperative and intraoperative measures, such as iron substitution, erythropoietin treatment, autologous blood donation (not in tumor patients), hypotensive anesthesia, acute normovolemic hemodilution, or meticulous surgical techniques.[9]

In light of the current trend toward personalized medicine,[10] preoperative identification of patients who are at risk of the need for blood transfusion could help avoid PBM interventions in patients who do not benefit and intensify them in patients who do. This could help to efficiently reduce the adverse effects of blood transfusions such as overall and recurrence-free survival, medical and surgical complications, and readmissions.[2,5] In addition to the benefits to patient outcomes, prior identification can avoid unnecessary fixed pre-ordering of packed red blood cells (PRBCs) units, which leads to capacity commitment and to higher costs. If pre-ordered PRBCs are not administered, this can lead to a loss of units, which then must be discarded or, in the worst case, cannot be used to save the lives of other patients due to locking.[11]

However, due to a variety of parameters such as the type and extent of surgery, course of anesthesia, cardiopulmonary condition, and age, it is often difficult for a surgeon to confidently assess whether the HB or HCT cutoff value will be reached before the procedure.[11,12] Therefore, since the mid-1990s, attempts have been made to predict patients who are unable to compensate for blood loss by themselves using classical logistic regression (LR) models.[13] However, these LR models have several limitations. Either only a few patients with microvascular reconstructions ($n < 100$) were included, therefore making the model not fully applicable to patients with microvascular reconstructions in the head and neck region,[11,13] or predictions are limited only to cases with head and neck tumors, and therefore making the model not be generally applicable to microvascular reconstructions in the head and neck region.[12] Moreover, the choice of a regression formula for translation into clinical practice entails calculational complexity.[14]

In contrast to LR, prediction models based on machine learning (ML) have played an increasingly important role in recent years, as they are more flexible and scalable by comparison,[15] and literature has claimed that they offer the advantage of making more accurate predictions from unseen data.[16] In this regard, there exist many different ML algorithms for regression (e.g., blood volume in ml) and classification (e.g., blood transfusion yes/no). Neural networks (NNET) are inspired by biology and consist of layers of neurons connected with different weights. Depending on the input parameters, a prediction is made in the last (output) layer. Decision trees consist of a sequence of different nodes, each containing a parameter that is walked through and leads to a prediction. In this respect, the result of a random decision forest (RDF) is derived from the majority decision from a set of randomly generated and different decision trees. Gradient boosting

TABLE 1 Characteristics of the cohort

Variables		Blood transfusion at day of surgery		Blood transfusion within 3 days		
		Yes (no. of patients 266)	No (no. of patients 391)	Yes (no. of patients 419)	No (no. of patients 238)	Total (no. of patients 657)
Sex	Female	138 (46.8%)	157 (53.2%)	213 (72.2%)	82 (27.8%)	295 (44.9%)
	Male	128 (35.4%)	234 (64.6%)	206 (56.9%)	156 (43.1%)	362 (55.1%)
Age, year	Mean (SD)	65.3 (13.9)	61.7 (15.1)	64.5 (14.4)	60.8 (15.0)	63.2 (14.7)
Weight, kg	Mean (SD)	68.6 (14.9)	76.6 (16.3)	70.6 (15.5)	78.4 (16.3)	73.4 (16.2)
Blood volume, ml	Mean (SD)	4721 (819)	4347 (743)	4416 (762)	4840 (823)	4570 (810)
Preop. Hb, g/dl	Mean (SD)	12.6 (1.8)	13.9 (1.5)	12.9 (1.8)	14.2 (1.4)	13.4 (1.8)
Postop. Hb, g/dl	Mean (SD)	8.9 (1.0)	9.4 (1.2)	8.8 (1.0)	9.9 (1.1)	9.2 (1.2)
ASA	I	9 (21.4%)	33 (78.6%)	13 (31.0%)	29 (69.0%)	42 (6.4%)
	II	106 (34.6%)	200 (65.4%)	183 (59.8%)	123 (40.2%)	306 (46.6%)
	III	144 (48.6%)	152 (51.4%)	213 (72.0%)	83 (28.0%)	296 (45.1%)
	IV	7 (53.8%)	6 (46.2%)	10 (76.9%)	3 (23.1%)	13 (2.0%)
Weighted CCI	0	18 (6.8%)	34 (8.7%)	27 (6.4%)	25 (10.5%)	52 (7.9%)
	1–2	69 (25.9%)	182 (46.5%)	129 (30.8%)	122 (51.3%)	251 (38.2%)
	3–4	57 (21.4%)	93 (23.8%)	94 (22.4%)	56 (23.5%)	150 (22.8%)
	≥5	122 (45.9%)	82 (21.0%)	169 (40.3%)	35 (14.7%)	204 (31.1%)
Cancer	No	36 (37.1%)	61 (62.9%)	62 (63.9%)	35 (36.1%)	97 (14.8%)
	Head and neck	173 (38.6%)	275 (61.4%)	274 (61.2%)	174 (38.8%)	448 (68.2%)
	Other	57 (50.9%)	55 (49.1%)	83 (74.1%)	29 (25.9%)	112 (17.0%)
Flap type	Small soft	95 (27.5%)	251 (72.5%)	180 (52.0%)	166 (48.0%)	346 (52.7%)
	Osseous	33 (35.5%)	60 (64.5%)	60 (64.5%)	33 (35.5%)	93 (14.2%)
	Large soft	138 (63.3%)	80 (36.7%)	179 (82.1%)	39 (17.9%)	218 (33.2%)
Incision–suture time, minute	Mean (SD)	599 (140)	527 (114)	583 (133)	508 (109)	556 (130)
Neck dissection	None	25 (36.2%)	44 (63.8%)	41 (59.4%)	28 (40.6%)	69 (10.5%)
	Unilateral	117 (34.5%)	222 (65.5%)	189 (55.8%)	150 (44.2%)	339 (51.6%)
	Bilateral	124 (49.8%)	125 (50.2%)	189 (75.9%)	60 (24.1%)	249 (37.9%)
Tracheostomy	No	47 (37.6%)	78 (62.4%)	74 (59.2%)	51 (40.8%)	125 (19.0%)
	Yes	219 (41.2%)	313 (58.8%)	345 (64.8%)	187 (35.2%)	532 (81.0%)

machines (GBMs) work comparably by combining several weak models into one strong model (ensemble). In contrast, support vector machines (SVMs) generate hyperplanes/lines that divide the data into different classes (e.g., blood transfusion yes/no).[17]

However, ML has not yet been used to identify patients requiring blood transfusions that are undergoing microvascular reconstruction in the head and neck region. Consequently, can ML models, compared to LR models, better identify this patient group? How might such models be used clinically to enable targeted and intensified PBM?

We developed different ML models and a LR model to predict the need for blood transfusions, which was indicated by noncompensable blood loss by the patient. We chose clinically relevant time points including the day of surgery and within 3 days after surgery and compared our ML models by internal validation with traditional LR models from the literature on our dataset. To make the ML models more understandable, we determined the variable importance that describe the relevance of the parameters used in the prediction (like age, HB, etc.). Afterwards, we illustrated a possible application that uses ML models to recommend the use of PBM measures. Since ML models are considered medical devices, we also developed an easy-to-use score chart that can be used in clinical encounters without having to meet regulatory requirements.

TABLE 2 Predictors for blood transfusion (univariable analysis)

	At day of surgery (univariable analysis)			Within 3 days (univariable analysis)		
Variables	**Unadjusted odds ratio**	**95% CI**	***p*-value**	**Unadjusted odds ratio**	**95% CI**	***p*-value**
Demographic findings						
Sex						
Male	Reference category			Reference category		
Female	1.607	1.174–2.203	0.003	1.967	1.418–2.741	<0.001
Age, 10 years	1.188	1.065–1.330	0.002	1.183	1.062–1.320	<0.001
Weight, 10 kg	0.715	0.640–0.796	<0.001	0.738	0.664–0.818	<0.001
Height, 10 cm	0.690	0.576–0.823	<0.001	0.623	0.516–0.748	<0.001
BSA, 0.25 m^2	0.534	0.436–0.649	<0.001	0.532	0.434–0.647	<0.001
Blood volume, 1 L	0.546	0.441–0.671	<0.001	0.512	0.414–0.630	<0.001
Patient findings						
Preoperative Hb, g/dl	0.608	0.541–0.678	<0.001	0.601	0.532–0.675	<0.001
ASA						
I	Reference category			Reference category		
II	1.943	0.931–4.457	0.092	3.319	1.692–6.840	0.001
III	3.474	1.670–7.953	0.002	5.725	2.893–11.896	<0.001
IV	4.278	1.157–16.673	0.030	7.436	1.917–37.496	0.007
Weighted CCI						
0	Reference category			Reference category		
1–2	0.716	0.383–1.373	0.303	0.979	0.536–1.782	0.945
3–4	1.158	0.603–2.273	0.663	1.554	0.820–2.944	0.175
≥5	2.810	1.504–5.402	0.001	4.471	2.325–8.650	<0.001
Cancer						
No	Reference category			Reference category		
Yes, head and neck	1.066	0.680–1.690	0.783	0.889	0.559–1.396	0.613
Yes, other	1.756	1.012–3.072	0.046	1.616	0.895–2.936	0.112
Surgical procedure						
Flap type						
Small	Reference category			Reference category		
Osseous	1.453	0.887–2.352	0.132	1.677	1.049–2.717	0.033
Large	4.558	3.181–6.577	<0.001	4.233	2.845–6.415	<0.001
Surgery duration, 1 h	1.319	1.218–1.435	<0.001	1.356	1.244–1.484	<0.001
Neck dissection						
None	Reference category			Reference category		
Unilateral	0.928	0.544–1.609	0.785	0.860	0.504–1.451	0.576
Bilateral	1.746	1.014–3.060	0.047	2.151	1.220–3.766	0.007
Tracheostomy						
No	Reference category			Reference category		
Yes	1.161	0.780–1.743	0.465	1.271	0.850–1.891	0.238

Abbreviations: ASA, ASA physical status classification system; BSA, body surface area; CCI, Charlson comorbidity index (weighted).

TABLE 3 Multivariable prediction for blood transfusion (multivariable analysis)

	At day of surgery (multivariable analysis)			Within 3 days (multivariable analysis)		
Variables	**Adjusted odds ratio**	**95% CI**	***p*-value**	**Adjusted odds ratio**	**95% CI**	***p*-value**
Demographic findings						
Sex						
Male	Reference category			Reference category		
Female	1.159	0.749–1.791	0.506	1.546	0.979–2.451	0.062
Age, 10 years	1.099	0.942–1.284	0.231	1.055	0.897–1.24	0.517
Weight, 10 kg	0.708	0.61–0.818	<0.001	0.783	0.678–0.902	<0.001
Patient findings						
Preoperative Hb, g/dl	0.591	0.514–0.674	<0.001	0.582	0.5–0.672	0.001
ASA						
I	Reference category			Reference category		
II	1.101	0.428–3.063	0.846	2.6	1.086–6.514	0.036
III	1.358	0.516–3.838	0.547	3.126	1.258–8.076	0.016
IV	2.565	0.488–14.363	0.271	6.023	1.022–44.486	0.058
Cancer						
No	Reference category			Reference category		
Yes, head and neck	0.737	0.332–1.607	0.447	0.515	0.221–1.182	0.119
Yes, other	1.412	0.587–3.396	0.440	1.344	0.518–3.497	0.543
Surgical procedure						
Flap type						
Small	Reference category			Reference category		
Osseous	1.111	0.502–2.434	0.793	1.501	0.663–3.46	0.333
Large	3.626	2.325–5.707	<0.001	3.748	2.272–6.314	<0.001
Surgery duration						
≤8 h	Reference category			Reference category		
8–10 h	1.451	0.851–2.493	0.174	1.533	0.928–2.54	0.096
>10 h	5.829	3.265–10.638	<0.001	6.323	3.443–11.904	<0.001
Neck dissection						
None	Reference category			Reference category		
Unilateral	1.633	0.83–3.28	0.160	2.073	1.043–4.119	0.037
Bilateral	2.528	1.213–5.364	0.014	5.21	2.402–11.38	<0.001
Tracheostomy						
No	Reference category			Reference category		
Yes	1.182	0.635–2.226	0.600	1.194	0.634–2.249	0.581

Note: Probability was calculated using the following equation: Probability $= \frac{e^u}{(1+e^u)}$, where u for day of surgery i 6.37443 + 0.14736(female sex) + 0.09421(age) − 0.52541(preop. Hb in g/dl) − 0.34544(weight in kg) − 0.34544(ASA II) + 0.30622(ASA III) + 0.94210(ASA IV) − 0.30474(head&neck cancer) + 0.34521(other cancer) + 0.10525(osseous flap) + 1.28815(large flap) + 0.37214(op duration 8–10 h) + 1.76293(op duration >10 h) + 0.49036(unilateral neck dissection) + 0.92731(bilateral neck dissection) + 0.16753(performed tracheostomy) and u for within 3 days is 6.40294 + 0.43586(female sex) + 0.05314(age) − 0.54116 (preop. Hb in g/dl) − 0.24414(weight in kg) − 0.95549(ASA II) + 1.13962(ASA III) + 1.79561(ASA IV) − 0.66265(head&neck cancer) + 0.29553(other cancer) + 0.40614(osseous flap) + 1.32136(large flap) + 0.42695(op duration 8–10 h) + 1.84421(op duration >10 h) + 0.72879(unilateral neck dissection) + 1.65066 (bilateral neck dissection) + 0.17745(performed tracheostomy). e corresponds to Euler's number 2.71828.

Abbreviations: ASA, ASA physical status classification system.

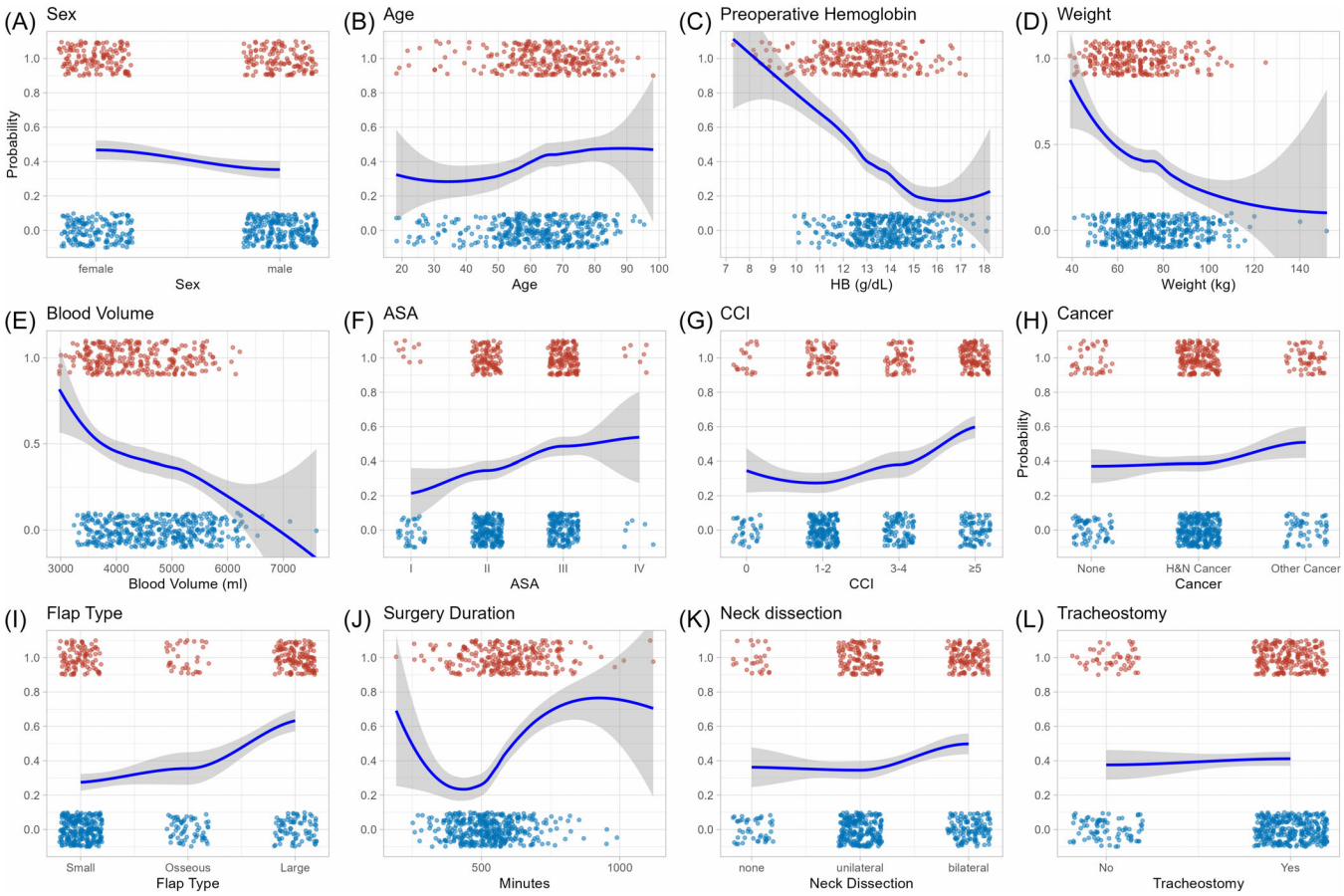


FIGURE 1 Probability of blood transfusions at day of surgery. (A–L) Locally weighted scatterplot smoothing (LOESS) curve in blue with 95% confidence interval (CI) in gray, plotting transfusion probability at surgery against different variables. The red points are cases in which a blood transfusion was given, and the blue points are cases in which no blood transfusion was necessary. Note that the different variables are not linearly associated with probability. This plot can be used to identify cut points where transfusion probability changes the most [Color figure can be viewed at wileyonlinelibrary.com]

2 | MATERIALS AND METHODS

2.1 | Patient cohort and procedure

A total of 657 cases aged 18 years and older treated with microvascular free flaps between 2011 and 2021 at the Department of Oral and Maxillofacial Surgery, RWTH Aachen University Hospital, were retrospectively reviewed. The surgeries were undertaken due malignant neoplasms (head and neck carcinoma, sarcoma, salivary gland carcinomas, basal cell carcinomas, and Merkel-cell carcinomas), benign neoplasms (ameloblastoma and myxomas), osteoradionecrosis/osteomyelitis, infections, trauma, and congenital malformations. All patients were treated according to the standardized treatment regimen of our department. The indications for blood transfusions were based on the relevant guidelines.[7] Outcome was defined as blood transfusions on the day of surgery and within 3 days.

The study was approved by the local ethics committee (approval number EK 057/21). This study follows the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.[18]

2.2 | Data collection and processing

The data were acquired by the local Data Integration Center (DIC). Patients who underwent microvascular free flap surgery were identified using ICPM codes (German modification of the International Classification of Procedures in Medicine). If a patient had undergone multiple microvascular reconstructions, only the first hospitalization was included.

Sex, age, and duration of surgery (incision–suture time) information was obtained from the local hospital information system. ASA category (I–IV), height, and weight were recorded in the last preoperative anesthesia record before surgery. Body surface area was calculated

based on weight and height. Blood volume was calculated based on weight, height, age, and sex.[19] Patients were classified by main medical diagnosis into head and neck cancer (C00–C14, and C30–C32), other cancer types (other C) and no cancer (no C codes). Using ICPM codes the number of neck dissection sides (5–403), tracheostomy (5–311 or 5–312) was determined. The weighted Charlson comorbidity index (CCI) was calculated with the ICD10 codes.[20]

The type of flap was determined from the operative report and divided into three categories: small/thin flaps (radial and ulnar forearm, upper arm and lower leg perforator and scapula soft tissue flaps), large/heavy flaps (anterolateral thigh, rectus abdominis, latissimus dorsi and multiple flaps), and osseous flaps (fibula, iliac crest, and scapula flaps).

The number of PRBC units administered on the day of surgery, within 3 days, and within 30 days, as well as the number of units of fresh frozen plasma (FFP) and platelet concentrate (PC) administered within 30 days were recorded. HB data were taken from the central laboratory. If absent from the laboratory, data were taken from the blood gas analyzer. The last deposited value up to 8 am on the day of surgery was taken as the preoperative HB value and thus before any surgical procedure. The earliest HB value on the first postoperative day was chosen as the postoperative HB value. For all collected data, missing data were imputed using the last available value before surgery.

2.3 | Development of prediction models

The final data were checked for zero and near-zero variance. Correlated predictors (height, weight, BMI, BSA, and blood volume) were identified and removed when appropriate in the final model development stage. The predictors used in the final models were: age (years),

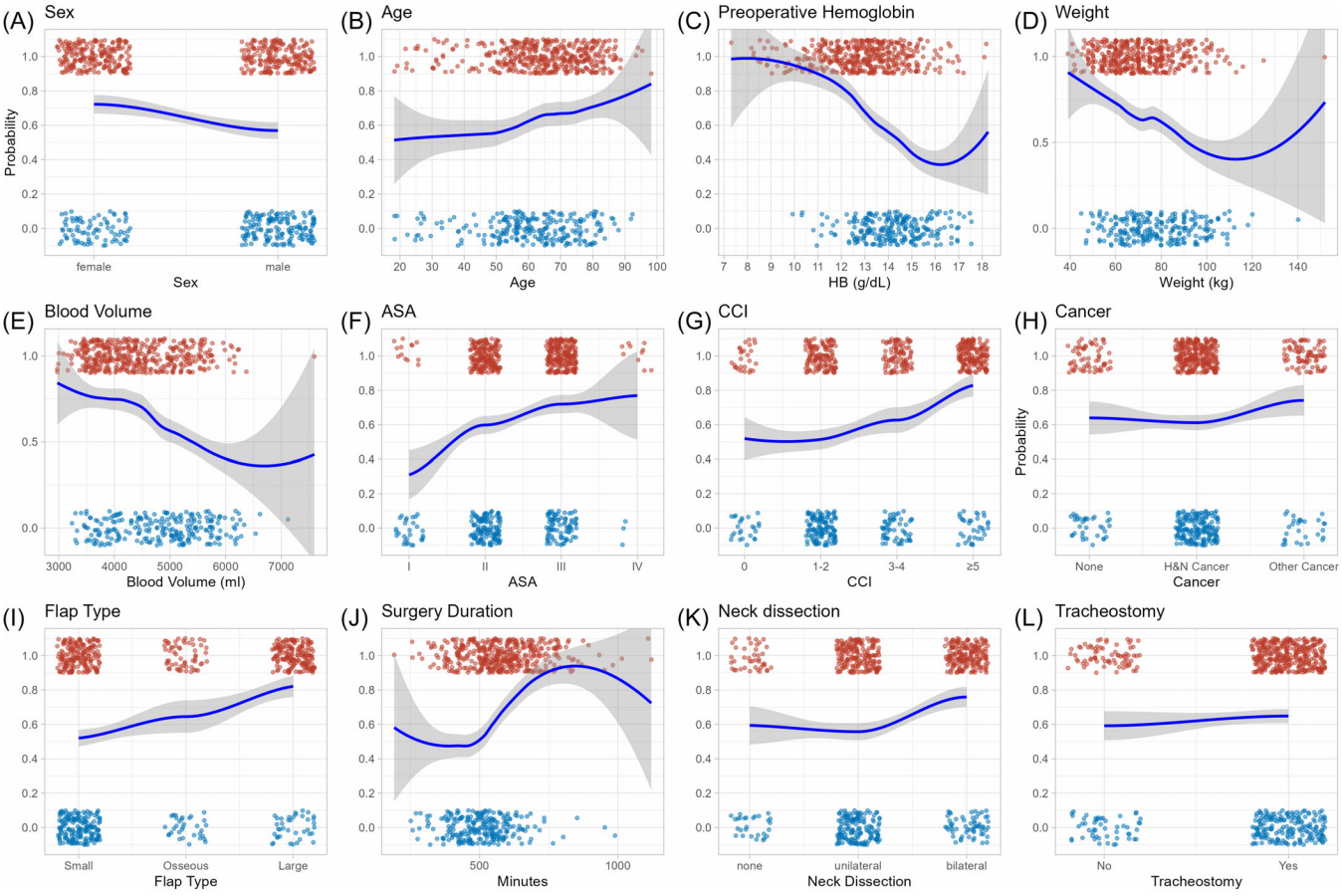


FIGURE 2 Probability for blood transfusions within 3 days. (A–L) Locally weighted scatterplot smoothing (LOESS) curve in blue with the 95% confidence interval in gray, plotting transfusion probability within 3 days against different variables. The red points are cases in which a blood transfusion was given, and the blue points are cases in which no blood transfusion was necessary. Note that the different variables are not linearly associated with probability. This plot can be used to identify cut points where transfusion probability changes the most [Color figure can be viewed at wileyonlinelibrary.com]

sex (female and male), blood volume (ml), ASA (0-IV), weighted CCI (0, 1–2, 3–4, ≥5), cancer type (none, head&neck, other), flap type (small, osseous, large), incision-suture time (short vs. long, threshold by median of all cases with 9 h), neck dissection sides (none, unilateral, bilateral), and tracheostomy (yes and no). Next, four popular models in ML—GBM, RDF, NNET, and SVM—were selected and trained and compared to a multivariable prediction model (LR).

In addition, a score chart was developed as described by Zhang et al.[21] for predicting PRBCs transfusion on the day of surgery and within 3 days. For this purpose, clinically easy-to-use predictors were first categorized based on their transfusion probability using feasible cutoffs (e.g. surgery duration ≤8 h, 8–10 h and >10 h).[14,21] A model was then derived using LR. The coefficients were multiplied by 4 to avoid rounding to 0 for small values. The intercept was not taken into account, and a suitable threshold was chosen based on a sensitivity rate of about 90%.[21]

Although a fixed split between training and test datasets (e.g., 80%–20%) is not recommended in smaller datasets for validation, it is often used.[22] The performance of the model, however, can be either better or worse, depending on which cases are included in the training and test datasets purely by chance. Therefore, to validate the models, we performed a bootstrapping with 500 repetitions (out-of-sample bootstrap), which is recommended as a state-of-the-art method for internal validation and should in no way be misinterpreted as data augmentation.[22] With the out-of-sample bootstrap

TABLE 4 Number of transfused PRBCs

Number of transfused PRBCs		No (no. of patients 238)	1–2 PRBCs (no. of patients 236)	3–4 PRBCs (no. of patients 111)	≥5 PRBCS (no. of patients 72)	Total (no. of patients 657)
PRBCs at day of surgery	Mean (SD)	0.0 (0.0)	0.6 (0.7)	1.7 (1.2)	4.7(2.4)	1.0 (1.7)
PRBCs at Day 1 and 2 after surgery	Mean (SD)	0.9 (0.7)	1.8 (1.1)	2.5 (2.1)	0.9 (1.2)	0.9 (0.7)
PRBCs at Day 3–30 after surgery	Mean (SD)	0.2 (0.8)	0.5 (2.1)	0.8 (2.1)	1.2 (2.7)	0.5 (1.9)
Preop. Hb (g/dl)	Mean (SD)	14.2 (1.4)	13.2 (1.5)	12.8 (1.8)	12.1 (2.1)	13.4 (1.8)
Postop. Hb (g/dl)	Mean (SD)	9.9 (1.1)	8.8 (0.9)	8.7 (0.9)	8.8 (1.1)	9.2 (1.2)
Flap type	Small	166 (69.7%)	132 (55.9%)	40 (36.0%)	8 (11.1%)	346 (52.7%)
	Osseous	33 (13.9%)	35 (14.8%)	14 (12.6%)	11 (15.3%)	93 (14.2%)
	Large	39 (16.4%)	69 (29.2%)	57 (51.4%)	53 (73.6%)	218 (33.2%)
Incision–suture time (min)	Mean (SD)	508 (109)	554 (122)	606 (122)	641 (160)	556 (130)

Abbreviation: PRBC, packed red blood cells.

TABLE 5 Flap type and blood transfusion

Flap type		Small flap (no. of patients 348)	Osseous flap (no. of patients 93)	Large flap (no. of patients 216)	Total (no. of patients 657)
Cancer	No	13 (3.8%)	65 (69.9%)	19 (8.7%)	97 (14.8%)
	Head and neck	291 (84.1%)	19 (20.4%)	138 (63.3%)	448 (68.2%)
	Other	42 (12.1%)	9 (9.7%)	61 (28.0%)	112 (17.0%)
Incision–suture time (min)	Mean (SD)	531 (122)	590 (108)	580 (143)	556 (130)
PRBCs within 30 days	Mean (SD)	1.6 (2.9)	2.3 (2.5)	3.9 (3.8)	2.4 (3.3)
	Range	0–28	0–12	0–24	0–28
	Sum	548	215	842	1605
FFPs within 30 days	Mean (SD)	0.2 (1.1)	0.5 (1.6)	1.0 (2.1)	0.5 (1.6)
	Range	0–10	0–8	0–12	0–12
	Sum	83	49	214	346
PCs within 30 days	Sum	11	0	12	23

Abbreviations: FFP, fresh frozen plasma; PC, platelet concentrate; PRBC, packed red blood cells.

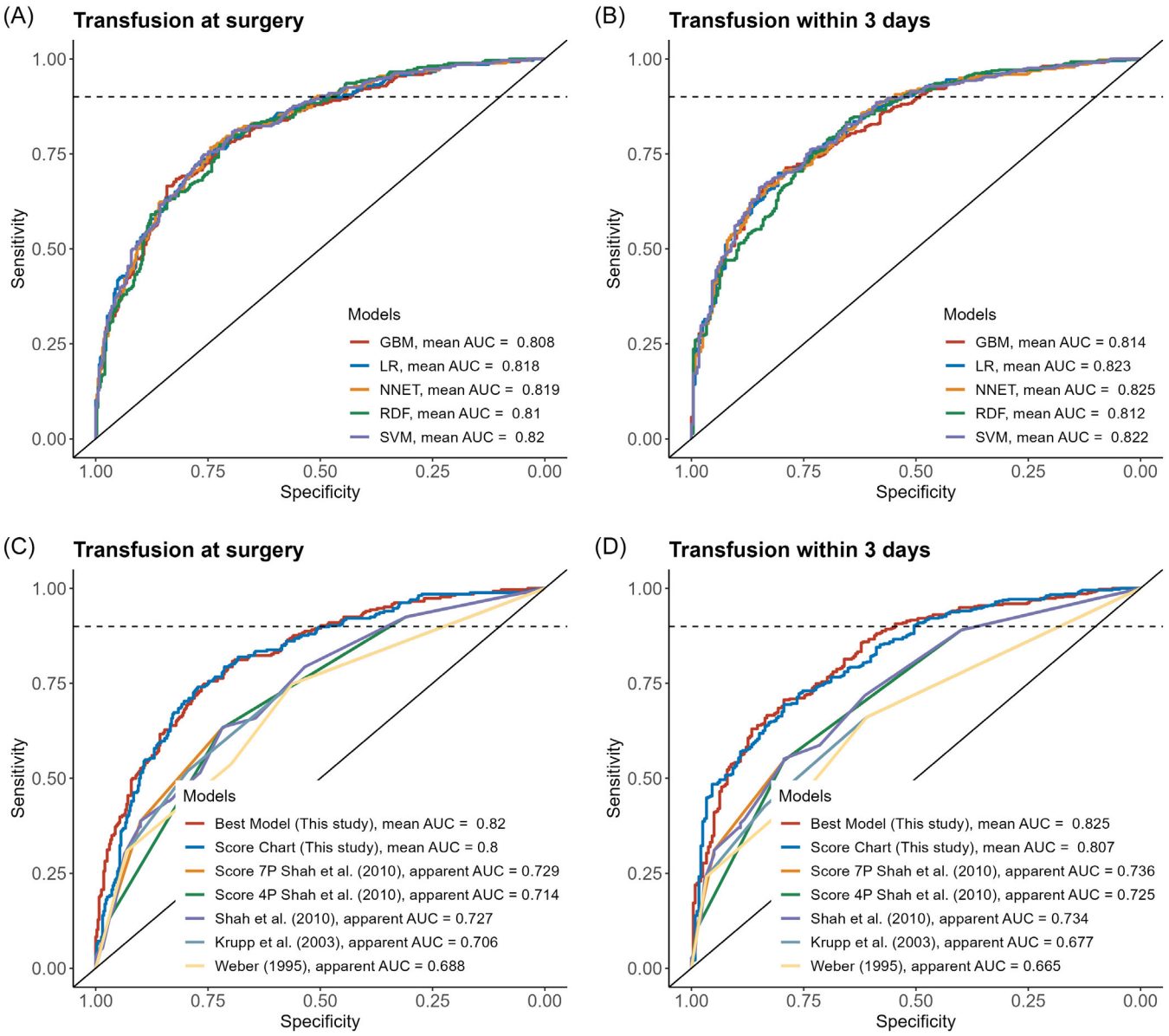


FIGURE 3 Prediction of packed red blood cell transfusion. (A, B) The receiver operating characteristics (ROC) curve of average predictions (mean AUC) using bootstrapping (500 repetitions) for five different machine learning (ML) models for predicting the need for transfusion of PRBCs on the day of surgery (A) and transfusion within 3 days (B). The models used were gradient boosting machine (GBM), logistic regression (LR), random decision forest (RDF), neural network (NNET) and support vector machine (SVM). (C, D) Our best ML model compared to a derived score and three models that have been described in the literature[11–13] both for predicting transfusion of PRBCs on the day of surgery (C) and within 3 days (D). The external models/scores were validated on our data (apparent AUC), which corresponds to external validation. (A–D) The horizontal dashed line illustrates a sensitivity level of 90%. The mean area under the curve (AUC) is indicated next to the legend. The *x*-axis corresponds to specificity, and the *y*-axis corresponds to sensitivity [Color figure can be viewed at wileyonlinelibrary.com]

repeated 500 times, approximately 2/3 of the data (when subjects are selected multiple times) is used for training and 1/3 of the data not used for training is used for testing in each repetition and is therefore not seen during training, but is used for internal validation. Finally, an average prediction is calculated based on all tests per case.[23] This approach allows the estimation of how well the models would perform on unseen data.[14,22]

For visualization, the average prediction was used for internal validation and sensitivity, specificity, and area under the receiver operating characteristic (ROC-AUC) curves were calculated. Finally, our models were compared to existing statistical models found in the literature: Weber, Krupp et al., and Shah et al.[11–13] For this purpose, we applied the published LR models on our patient collective. Since T classification (T1/T2

TABLE 6 Internal validation of blood transfusion prediction performance by bootstrapping

Prediction	Model	ROC-AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	Specificity at a sensitivity of 90%[a]
Blood transfusion at day of surgery	GBM	0.808 [0.75–0.86]	0.837 [0.76–0.91]	0.611 [0.5–0.72]	0.482
	LR	0.818 [0.76–0.86]	0.821 [0.75–0.89]	0.647 [0.54–0.75]	0.508
	NNET	0.819 [0.77–0.87]	0.836 [0.75–0.91]	0.623 [0.51–0.73]	0.5
	RDF	0.81 [0.76–0.86]	0.883 [0.8–0.96]	0.51 [0.34–0.66]	0.471
	SVM	0.82 [0.77–0.87]	0.824 [0.75–0.89]	0.644 [0.54–0.75]	0.508
Blood transfusion within 3 days	GBM	0.814 [0.76–0.86]	0.849 [0.77–0.92]	0.563 [0.43–0.68]	0.517
	LR	0.823 [0.78–0.87]	0.834 [0.75–0.91]	0.621 [0.51–0.72]	0.503
	NNET	0.825 [0.78–0.87]	0.873 [0.79–0.94]	0.57 [0.45–0.69]	0.539
	RDF	0.812 [0.76–0.86]	0.926 [0.86–0.98]	0.438 [0.29–0.58]	0.476
	SVM	0.822 [0.78–0.87]	0.848 [0.77–0.92]	0.594 [0.47–0.71]	0.531

Note: Best models in bold.

Abbreviations: GBM, gradient boosting machine; LR, logistic regression; NNET, neuronal network; RDF, random decision forest; ROC-AUC, receiver operating characteristic-area under curve; SVM, support vector machine.

[a]Threshold with a target sensitivity of 90% evaluated by bootstrapping.

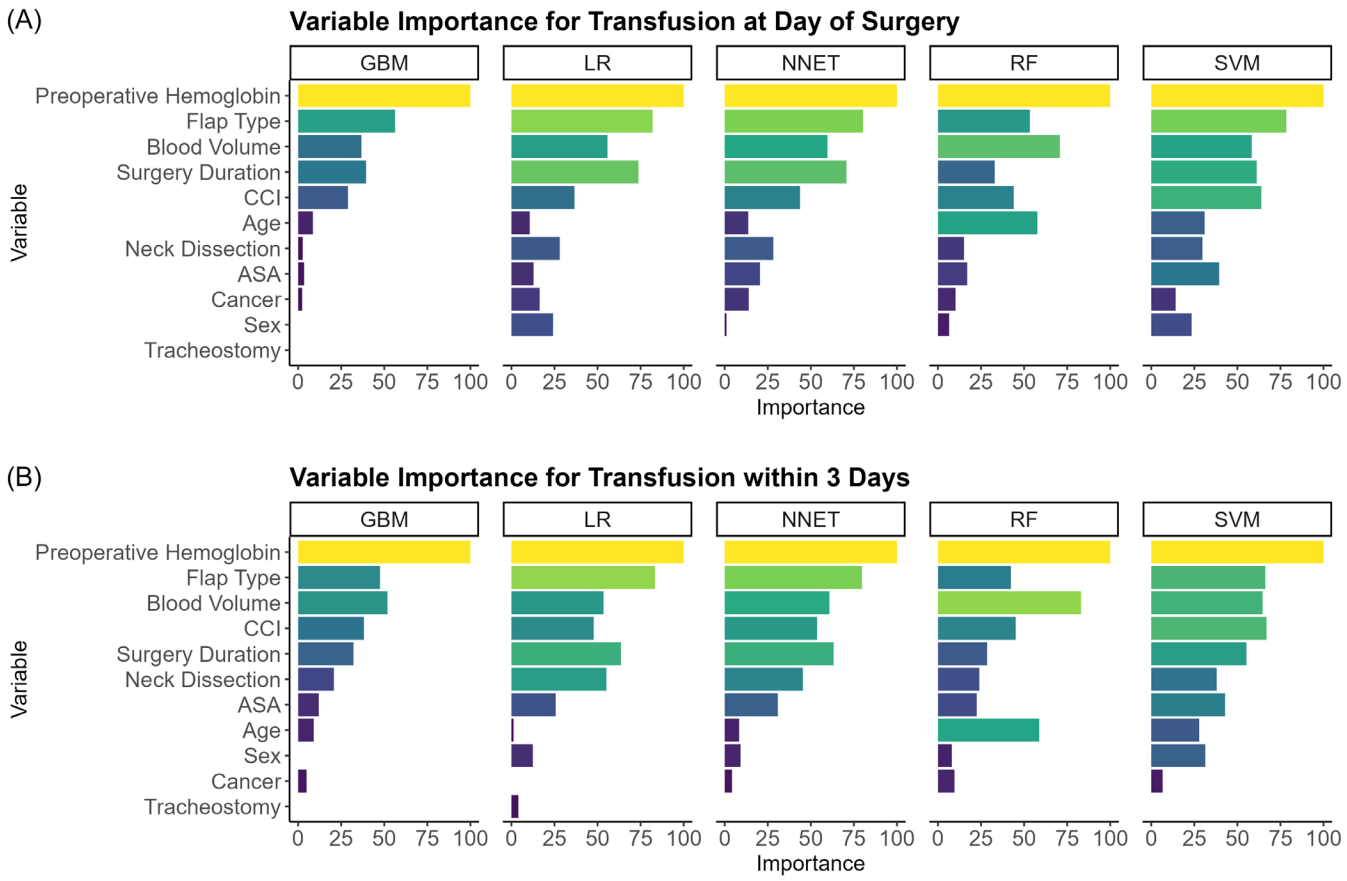


FIGURE 4 Variable importance. (A, B) The *y*-axis shows the different variables (predictors). On the *x*-axis, their importance is shown. The higher the importance, the greater the contribution of the corresponding variable to the model's decision. Preoperative hemoglobin, in particular, was, along with flap type, surgery duration, sneck dissection, and weight, one of the most important variables in the majority of models [Color figure can be viewed at wileyonlinelibrary.com]

vs. T3/T4) was not available for every patient case (e.g., reconstruction-only cases) and there are differences in T classification depending on the entity and version, we declared flap type as a surrogate parameter (small/osseous flaps as T1/T2 and large flaps as T3/T4).

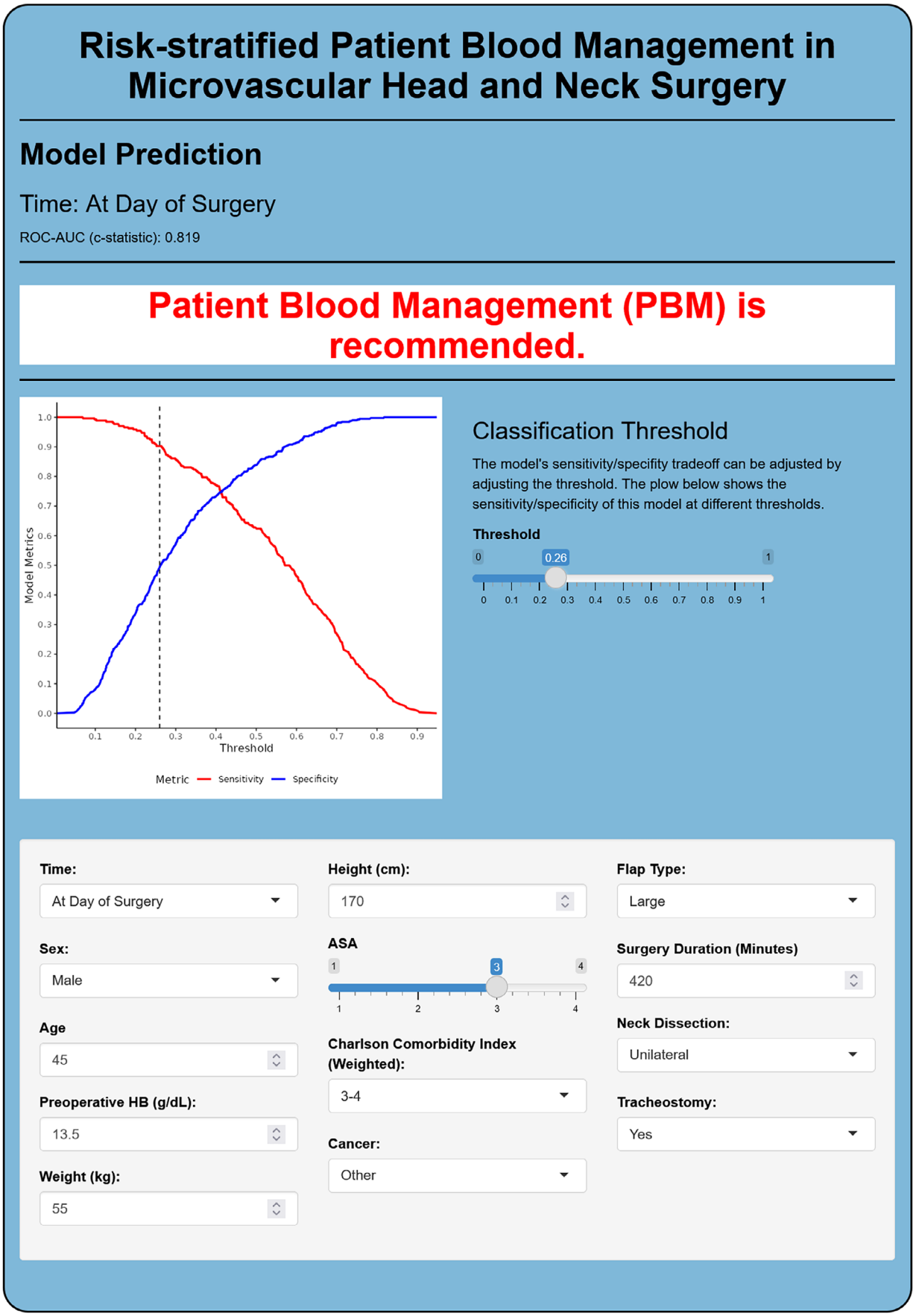
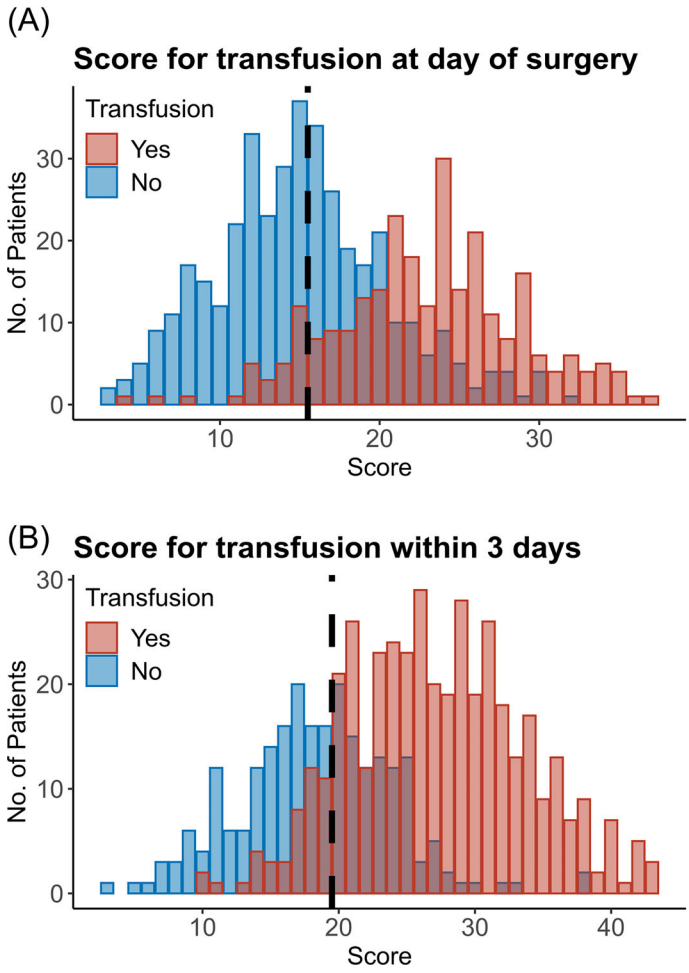


FIGURE 5 Legend on next page.



(C)

Transfusion Score for Microvascular Reconstruction of the Head & Neck

*Add the values for the patient either "At surgery" or "Within 3 days". If the calculated value is above the limit of 15 or 19, the blood loss is probably not compensable without measures. *Value ranges that are not specified are always counted as 0.*

Parameter	Range*	At surgery	Within 3 days
Sex	female	1	2
Age (Years)	56-75	2	0
	≥76	2	1
Preoperative Hemoglobin (g/dL)	≤11	12	13
	11.1-12	9	11
	12.1-13	6	7
	13.1-14	5	5
	14.1-15	2	3
Weight (kg)	≤60	5	4
	60.1-70	3	3
	70.1-80	2	1
ASA	II	0	4
	III	1	5
	IV	3	7
Cancer	None	2	2
	H&N	0	0
	Other	3	4
Flap type	Osseous	0	2
	Large	5	5
Surgery Duration (Hours)	8.1-10	1	2
	>10	7	7
Neck dissection	unilateral	2	3
	bilateral	4	7
Tracheostomy	Yes	1	1
Patient Blood Management Necessary	**No**	**≤15**	**≤19**
	Yes	**≥16**	**≥20**

FIGURE 6 Score chart for patient blood management. (A, B) Plot of the distribution of the number of patient cases on the day of surgery and within 3 days that needed (red) or did not need (blue) transfusion of PRBCs. The dashed vertical line represents the score achieving a sensitivity level of about 90%. On the *x*-axis are the score points calculated from the clinical decision aid (see panel C). (C) The clinical decision aid can be used to determine whether PRBC is needed on the day of surgery or within 3 days with a sensitivity level of about 90%. Based on the parameters, points are added from the corresponding ranges. If the patient does not fall into any range of a parameter, no points are given for the parameter. A calculated score of 16 or more (at surgery) or 20 or more (within 3 days) indicates that blood loss will not be compensable, so patient blood management (PBM) would be necessary. Example calculation at surgery: male sex (=0 points), age 45 years (=0 points), preoperative hemoglobin 13.5 g/dl (=5 points), weight 55 kg (=5 points), ASA III (=1 point), H&N cancer (=0 points), flap type large (=5 points), expected surgery duration 7 h (=0 points), neck dissection unilateral (=2 points), tracheostomy: yes (=1 point), which would yield $0 + 0 + 5 + 5 + 1 + 0 + 5 + 0 + 2 + 1 = 19$ points. Since this is greater than the cutoff of 15 points, blood loss on the day of surgery is unlikely to be compensable by the patient, so PBM would be beneficial [Color figure can be viewed at wileyonlinelibrary.com]

2.4 | Statistical analysis

For data provenance, data processing was performed after delivery from the DIC using the programming language R (R Core Team, 2021, www.r-project.org) in a hard-coded processing script. Odds ratios (ORs) were calculated from the LR coefficients. p values <0.05 were considered significant. The ML and LR models were developed using the Caret package.[24] Hyperparameter tuning was performed on the high-performance computer of RWTH Aachen University. The ROC-AUC was compared to the pROC package with bootstrapping.[25]

FIGURE 5 Web application for patient blood management. This is the representation of an application for the prediction of the need for blood transfusions. In the bottom area, the predictors can be specified. At the top, the recommendation of the ML model (NNET) for blood transfusions is displayed depending on a defined threshold. For demonstration purposes, the web application can be viewed under the following web link (please note the disclaimer!): https://omfs.shinyapps.io/blood_loss/ [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 7 Comparison of blood transfusion prediction with models from literature

Model	Study	At day of surgery ROC-AUC [95% CI]	Within 3 days ROC-AUC [95% CI]	Original reported ROC-AUC in literature [95% CI]	No. of patients with flaps
Best models	This study	0.82 [0.77–0.87]	0.825 [0.78–0.87]		657
Score chart		0.8 [0.75–0.84]	0.807 [0.75–0.86]		
7-point score	Shah et al.[12]	0.729	0.736	0.747 [0.710–0.782]	585
4-point score		0.714	0.725	0.744 [0.706–0.779]	
Model		0.727	0.734	0.754 [NA]	
Model	Krupp et al.[11]	0.706	0.677	0.72 [0.62–0.81]	88
Model	Weber[13]	0.688	0.665	NA	100

Abbreviations: NA, not available; ROC-AUC, area under the receiver operating characteristic.

3 | RESULTS

3.1 | Transfusion of allogeneic blood

Of the 657 patients included in the study, 44.9% were women, and 55.1% were men. The mean age was 63.2 years (range: 18–98 years). Most patients were classified as ASA II (46.6%) or ASA III (45.1%). The majority of cases were associated with treatment for head and neck cancers (68.2%), followed by other tumor entities (17%) and nontumors (14.8%). The majority of patients underwent reconstruction using small soft tissue flaps (52.7%), followed by large soft tissue flaps (33.2%) and osseous flaps (14.2%). The mean incision–suture time was 556 min (just over 9 h). Most patients underwent unilateral (51.6%) or bilateral (37.9%) neck dissection. Tracheostomy was performed in 81% of the patients. In this context, 40.5% of all patients received blood transfusions on the day of surgery and another 23.3% required blood transfusions on the second or third day after surgery; thus, 63.8% of all patients received blood transfusions within the first 3 days. There was no missing data (Table 1).

Variables associated with blood transfusion on the day of surgery ($OR_{Surgery}$) and within the first 3 days ($OR_{3\text{-}Days}$) were the female sex ($OR_{Surgery}$ 1.607, $OR_{3\text{-}Days}$ 1.967), increased age ($OR_{Surgery}$ 1.188, $OR_{3\text{-}Days}$ 1.183),low body weight ($OR_{Surgery}$ 0.715, $OR_{3\text{-}Days}$ 0.738), low body height ($OR_{Surgery}$ 0.690, $OR_{3\text{-}Days}$ 0.623), low body surface area ($OR_{Surgery}$ 0.534, $OR_{3\text{-}Days}$ 0.532), low blood volume ($OR_{Surgery}$ 0.546, $OR_{3\text{-}Days}$ 0.512), low preoperative HB ($OR_{Surgery}$ 0.608, $OR_{3\text{-}Days}$ 0.601), high ASA category ($OR_{Surgery}$ 1.943–4.278, $OR_{3\text{-}Days}$ 3.319–7.436), high CCI category ($OR_{Surgery}$ 0.716–2.810, $OR_{3\text{-}Days}$ 0.979–4.471), other cancers aside from those in the head and neck ($OR_{Surgery}$ 1.756, $OR_{3\text{-}Days}$ 1.616), large soft tissue flap usage ($OR_{Surgery}$ 4.558, $OR_{3\text{-}Days}$ 4.233), and bilateral neck dissection ($OR_{Surgery}$ 1.746, $OR_{3\text{-}Days}$ 2.151) (Table 2). A multivariable analysis showed similar results (Table 3). In this regard, only variables shown in the score chart were included due to the collinearity of different predictors (e.g., weight, height, blood volume, and body surface area).

These associations were also illustrated visually (Figures 1 and 2). Therefore, depending on the predictor, the probability for a blood transfusion showed an almost linear increase, as, for example, with decreasing preoperative Hb level (Figure 1C) or a sudden increase with respect to the surgery duration at just over about 500 min (≈8 h) (Figure 1J).

In this context, none or 1–2 PRBCs were administered mainly in patients who required small and osseous flaps, whereas 1–4 PRBCs were more likely to be administered in patients with large flaps. Additionally, ≥5 PRBCs were administered, especially in patients with large flaps, but were also less frequent in patients with small or osseous flaps (Table 4). The associations between flap type and the need for transfusion are most remarkable (Figures 1I and 2I). On average, 1.6 PRBCs were required for small flaps, 2.3 PRBCs for osseous flaps, and 3.9 PRBCs for large flaps. Two-thirds of patients who needed large flaps received a blood transfusion on the day of surgery, whereas this was the case in only a third of patients who needed small or osseous flaps (Table 5).

3.2 | Prediction models

In predicting the need for transfusion of PRBCs on the day of surgery and within 3 days, the ML models were comparable with ROC-AUCs between 0.808 and 0.825 (Figure 3A,B and Table 6). Among these, the non-tree-based models (NNET and SVM) had slightly higher ROC-AUCs compared to the tree-based models (GBM and RDF). The LR model achieved a ROC-AUC of 0.818 and 0.823 (Figure 3A,B and Table 6). There were no

significant differences in ROC-AUC between the developed ML/LR models ($p > 0.05$). In all models, preoperative HB levels were the most important predictor. Additionally, flap type, blood volume, surgery duration, age, and neck dissection were important predictors (Figure 4).

The best ML models, which were integrated into a web application, were used to provide predictions for clinicians (Figure 5). To develop a score chart, two LRs were trained with categorized predictors (e.g., age groups instead of age), and subsequently, score charts were derived from the final models (Figure 6). The score charts yielded a slightly lower ROC-AUC of 0.8 (on the day of surgery) and 0.807 (within 3 days), respectively. Again, preoperative HB was the most important predictor. In addition, flap type, surgery duration, neck dissection, and weight were important predictors (ie, score items; Figure 6C).

Finally, to make our models and score charts comparable, three external models and two external scores were validated against our data (≙external validation), achieving an ROC-AUC between 0.665 and 0.736, and compared with our best ML models and the derived score chart (Table 7 and Figure 3C,D). In a cross-wise comparison by validation on our dataset, all our models (ML, LR, and Score Chart) had significantly higher ROC-AUC than the external LR models from the literature ($p < 0.05$).

4 | DISCUSSION

Although various studies have shown that blood transfusion has an adverse effect on surgical outcomes,[2–5] compensating for blood loss in patients undergoing microvascular reconstruction of the head and neck via allogeneic blood transfusion has remains the standard of care.[2] However, prior identification of compensatory capacity could lead to risk-stratified PBM measures that could reduce the use of allogeneic blood transfusion[8] and thus avoid related adverse effects.[2,5]

However, good predictions are only possible if corresponding strong predictors exist.[14] Our study showed that compensatory or intrinsic factors (preoperative HB level, blood volume, sex, age, and health status) and aggravating or extrinsic factors (duration of surgery, invasiveness of the procedure—including extent of neck dissection—and type of flap) were strong predictors of an increased need for blood transfusion (Figures 1 and 2). These results are consistent with other studies that found low preoperative HB, osseous flap, underweight, and female sex as factors for increased blood transfusion demand.[11–13,26]

The aforementioned predictors can be rationally explained. A low preoperative HB level correlates to a smaller blood reserve, and a low weight, shorter stature, or the combination (body surface area) is accompanied by lower blood volume. In this context, women often have a lower weight and, by nature, often a lower HB level.[19] Older age is associated with lower blood volume.[27] As a result, these patients are usually less able to compensate for blood loss. Therefore, a critical HB level is reached earlier, which ultimately triggers the need for blood transfusions. Furthermore, the decreased cardiopulmonary fitness that occurs in people of older ages or in association with a higher ASA or CCI category results in the need for blood transfusions even at higher HB thresholds.[7] In addition, the invasiveness of the surgical procedure itself has an aggravating effect due to increased blood loss. Prolonged duration of surgery is associated with longer exposure to potentially bleeding wounds and more surgical steps that bleed. In this regard, a bilateral neck dissection corresponds to a larger wound area. Consistent with this, large tumors (T3/T4) are significantly associated with the need for blood transfusions.[11–13,26]

However, the use of tumor size via the T classification to estimate resection volume and thus bleeding potential has certain limitations. Depending on the tumor entity, the T classification may correspond to different sizes. Moreover, a high T classification can also be a measure of the depth of invasion.[28] Small tumors with a large depth of invasion have a high T classification but are not necessarily associated with a significantly larger resection volume. Furthermore, T classification primarily describes the resection, but not the size and type of donor site resulting from the raising of a flap.

Therefore, we took a different approach. The results of our study show that the size and type of the flap itself are among the strongest predictors (Figure 1I and 2I), along with preoperative HB level, blood volume, and duration of surgery (Figures 1 and 2). This is because these predictors define both the resulting donor site and the resection volume, and thus, they can provide a good estimate of the bleeding potential. Furthermore, the underlying disease is also an important predictor of the type of intervention. In this context, it is interesting to note that head and neck tumors have the lowest risk of bleeding compared to other tumors and nontumors. One reason could be that head and neck tumors are treated intraorally, while other tumor types, such as melanomas, usually affect the extraoral region. In the latter, microvascular flaps are mostly used only when local flaps are no longer suitable. In line with this, large flaps are used more frequently in other tumors (54.5%) than in head and neck tumors (30.8%), while nontumors are treated predominantly with osseous

reconstructions (67% of cases), which increase the bleeding potential (Figures 1H and 2H).

Based on these predictors, we developed predictive models that can forecast the need for blood transfusions preoperatively. Our developed ML models showed excellent performance with ROC-AUCs ranging from 0.808 to 0.825 and were comparable to our LR model (Table 6). We could not establish an advantage of ML models over classical LR models. This is also consistent with a systematic review and meta-analysis that found no clear performance advantage of ML models over LR models.[29]

However, all of our models (ML, LR, and score chart) outperformed those from the literature with a maximum reported ROC-AUC of 0.754 (Table 7) and during external validation on our dataset. Furthermore, unlike the models in the literature,[11–13] our models are generally applicable to any type of microvascular reconstruction in the head and neck region. Reasons for better performance include the use of additional strong predictors, and we did not exclude nonsignificant predictors. Unlike in inferential statistics, insignificant predictors can also contribute to prediction models.[14]

Nevertheless, before applying prediction models, some considerations must be made. Due to site-specific factors, such as the underlying patient population, the experience of the surgical team, and the local treatment philosophy, predictive models should be externally validated.[14] Older models from the literature were also applicable to our cohort and resulted in an almost identical ROC-AUC to the original reported values (Table 7). Possible reasons for this are that head and neck microvascular reconstructions are subject to global surgical standardizations,[30] and the administration of blood transfusions is triggered by broadly applied HB or HCT thresholds based on underlying physiological mechanisms.[6,7,31,32] These results suggest that the above models are more generalizable, as previously speculated,[11] making the use of predictive models particularly interesting in head and neck surgery. Furthermore, especially with ML models, due to their "black box" character, the underlying factors determining a decision are not so easily recognizable.[17] However, the use of variable importance helps make ML models' decisions explainable (Figure 4). The greater the variable's importance, the more it contributes to the decision of the ML model.[17] Furthermore, one must consider how an ML model can be applied in practice. For this purpose, a web application, as shown in Figure 5, could be used (https://omfs.shinyapps.io/blood_loss/; please note the disclaimer). Since these ML models (including our web application) are classified as medical devices according to the FDA in the US or according to Council Directive 93/42/EEC in EU countries, high regulatory hurdles must first be overcome before they can be applied in clinical use.[33]

However, the use of a score chart would already be an alternative that could be applied in clinical settings.[14] Therefore, we derived a score chart from our LR models (Figure 6C). For practical application, we converted the continuous variables into categorized variables. A detailed description of the method can be found in Zhang et al.[21] Although there was a slight loss of information due to the categorization of variables, our score chart still showed a much better ROC-AUC of 0.807 than even the best score chart from the literature with an ROC-AUC of 0.747 (Table 7).

To detect as many patients who need a blood transfusion as possible, we chosen a threshold for our score chart that achieved a sensitivity of at least 90% (Figure 6A,B). With a prevalence of 40.5% of blood transfusions on the day of surgery and 63.8% within 3 days, in our cohort, this would result in 4.1% and 6.4% of all cases, respectively, that had noncompensable blood loss but were not identified as such. If the score were used to decide, for example, to perform only a type and screen instead of a type and crossmatch, and thus not to reserve PRBCs in advance, this would be acceptable in a maximum care hospital with a corresponding large blood bank. This would be further mitigated by the fact that bleeding on the day of surgery and on the 1st and 2nd postoperative days are of a different nature, with the latter being much more harmless. Along with sensitivity, specificity plays a major role in reducing the number of patients who receive PBM overtreatment. With a target sensitivity of 90%, our models outperformed the specificity described in the literature (Figure 3C,D and Table 6). The application of our score chart with an example is described in the legend of Figure 6.

Overall, our prediction models could help to better differentiate between patients who need blood transfusion and those, who do not, thus helping to establish a risk-stratified PBM by target and intensified measurements and address alternative tasks, such as preoperative iron substitution or erythropoietin treatment, allogenic blood transfusions, hypotensive anesthesia, acute normovolemic hemodilution, or meticulous surgical techniques.[9] While the routine type and crossmatch of two PRBC units in all patients is associated with several disadvantages,[11] our predictive models could also better narrow down the patient group requiring type and screen.

Our study is not without limitations. First, we did not include blood clotting disorders as predictors, because there are numerous different types. This may lead to an even further improvements in the models. However, the use of preoperative INR, partial thromboplastin time, and platelet count did not lead to further improvements in

the model in our case (not shown). Furthermore, with limited exceptions, the radial forearm flap in our department is a small flap, while ALT is a large flap. This must always be considered during the application of the models. However, we believe that the ALT donor site, regardless of flap size, has a higher bleeding potential due to the exposure of the vascular pedicle in greater depth. Finally, it must be noted that random events, such as accidental opening of a vessel (e.g., the carotid artery), cannot be predicted with the given predictors.

5 | CONCLUSION

The use of additional clinical variables leads to improved predictions of blood transfusion demand. The raised flap type and size are strong predictors of both donor site area and head and neck resection volume and thus an expression of the bleeding potential. Therefore, all of our developed prediction models significantly outperformed those derived from the literature when validated on our dataset and according to their reported performance. Furthermore, our models are applicable to any kind of microvascular head and neck reconstruction. Interestingly, prediction models for microvascular head and neck surgery seem likely to be more generalizable because of the surgical standardization of such procedures and underlying physiological mechanism. The developed ML models had comparable predictive performance to an LR model. While the ML models face major regulatory hurdles, a developed score charts based on LR could be implemented after external validation to identify patients who would benefit most from PBM measures, ultimately reducing the adverse effects of blood transfusions.

AUTHOR CONTRIBUTIONS

Conceptualization: Behrus Puladi, Ali Modabber, and Rainer Röhrig. *Methodology*: Behrus Puladi. *Software*: Behrus Puladi. *Validation*: Max Taubert and Behrus Puladi. *Formal analysis*: Behrus Puladi, Mark Ooms, Annette Rieg, Max Taubert, Ashkan Rashad, Frank Hölzle, Rainer Röhrig, and Ali Modabber. *Investigation*: Behrus Puladi. *Resources*: Rainer Röhrig, Frank Hölzle, and Behrus Puladi. *Data curation*: Behrus Puladi and Mark Ooms. *Writing – original draft preparation*: Behrus Puladi. *Writing – review and editing*: Behrus Puladi, Mark Ooms, Annette Rieg, Max Taubert, Ashkan Rashad, Frank Hölzle, Rainer Röhrig, and Ali Modabber. *Visualization*: Behrus Puladi. *Supervision*: Rainer Röhrig, Ali Modabber, and Frank Hölzle. *Project administration*: Behrus Puladi. *Funding acquisition*: Behrus Puladi. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We thank Ewout Steyerberg for his helpful discussion with us while writing the manuscript. Thanks to Irina Lutz and Albert Esser for data acquisition (Data Integration Center of the University Hospital RWTH Aachen). Simulations were performed with computing resources granted by RWTH Aachen University under project rwth1262. This publication of the SMITH consortium was supported by the German Federal Ministry of Education and Research under grant number 01ZZ1803B. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The study approved by the Institutional Review Board (or Ethics Committee) of University Hospital RWTH Aachen (approval code EK 057/21).

ORCID

Behrus Puladi https://orcid.org/0000-0001-5909-6105
Mark Ooms https://orcid.org/0000-0002-9785-9502
Annette Rieg https://orcid.org/0000-0002-7043-5494
Max Taubert https://orcid.org/0000-0001-8925-7782
Ashkan Rashad https://orcid.org/0000-0003-3743-8658
Frank Hölzle https://orcid.org/0000-0002-2298-3517
Rainer Röhrig https://orcid.org/0000-0002-0032-5118
Ali Modabber https://orcid.org/0000-0003-2160-4527

REFERENCES

1. Sweeny L, Rosenthal EL, Light T, et al. Outcomes and cost implications of microvascular reconstructions of the head and neck. *Head Neck*. 2019;41:930-939.
2. Giovacchini F, Bensi C, Paradiso D, Docimo R, Tullio A. Association between blood transfusions and complications in head and neck reconstruction: a systematic review and meta-analysis. *Eur Arch Otorhinolaryngol*. 2021;278:2171-2185.
3. Taniguchi Y, Okura M. Prognostic significance of perioperative blood transfusion in oral cavity squamous cell carcinoma. *Head Neck*. 2003;25:931-936.
4. Vamvakas EC, Blajchman MA. Transfusion-related immunomodulation (TRIM): an update. *Blood Rev*. 2007;21:327-348.
5. Spanier G, Böttcher J, Gerken M, et al. Prognostic value of perioperative red blood cell transfusion and anemia on survival and recurrence in oral squamous cell carcinoma. *Oral Oncol*. 2020;107:104773.

6. Kim MJ, Woo K-J, Park BY, Kang SR. Effects of transfusion on free flap survival: searching for an optimal hemoglobin threshold for transfusion. *J Reconstr Microsurg*. 2018;34:610-615.
7. Bundesärztekammer. *Querschnitts-Leitlinien zur Therapie mit Blutkomponenten und Plasmaderivaten: Gesamtnovelle 2020 mit 22 Tabellen*. 5th ed. Deutscher Ärzteverlag; 2020:337.
8. Althoff FC, Neb H, Herrmann E, et al. Multimodal patient blood management program based on a three-pillar strategy: a systematic review and meta-analysis. *Ann Surg*. 2019;269:794-804.
9. Desai N, Schofield N, Richards T. Perioperative patient blood management to improve outcomes. *Anesth Analg*. 2018;127:1211-1220.
10. Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: data science enabling personalized medicine. *BMC Med*. 2018;16:150.
11. Krupp NL, Weinstein G, Chalian A, Berlin JA, Wolf P, Weber RS. Validation of a transfusion prediction model in head and neck cancer surgery. *Arch Otolaryngol Head Neck Surg*. 2003;129:1297-1302.
12. Shah MD, Goldstein DP, McCluskey SA, et al. Blood transfusion prediction in patients undergoing major head and neck surgery with free-flap reconstruction. *Arch Otolaryngol Head Neck Surg*. 2010;136:1199-1204.
13. Weber RS. A model for predicting transfusion requirements in head and neck surgery. *Laryngoscope*. 1995;105:1-17.
14. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer; 2019:558.
15. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. 2019;20:e262-e273.
16. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18:e323.
17. Kuhn M, Johnson K. *Applied Predictive Modeling*. Springer; 2013:615.
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102:148-158.
19. Pearson TC, Guthrie DL, Simpson J, et al. Interpretation of measured red cell mass and plasma volume in adults: Expert Panel on Radionuclides of the International Council for Standardization in Haematology. *Br J Haematol*. 1995;89:748-756.
20. Gasparini A. Comorbidity: an R package for computing comorbidity scores. *J Open Source Software*. 2018;3:648.
21. Zhang Z, Zhang H, Khanal MK. Development of scoring system for risk stratification in clinical medicine: a step-by-step tutorial. *Ann Transl Med*. 2017;5:436.
22. Steyerberg EW. Validation in prediction research: the waste by data splitting. *J Clin Epidemiol*. 2018;103:131-133.
23. John CR. MLeval: Machine learning model evaluation; 2021. https://cran.r-project.org/web/packages/MLeval/index.html.
24. Kuhn M. Building predictive models in R using the caret package. *J Stat Soft*. 2008;28:1-26.
25. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77.
26. Perisanidis C, Mittlböck M, Dettke M, et al. Identifying risk factors for allogenic blood transfusion in oral and oropharyngeal cancer surgery with free flap reconstruction. *J Oral Maxillofac Surg*. 2013;71:798-804.
27. Davy KP, Seals DR. Total blood volume in healthy young and older men. *J Appl Physiol*. 1985;1994(76):2059-2062.
28. Amin MB, Greene FL, Edge SB, eds. *AJCC Cancer Staging Manual*. AJCC; 2017:1024.
29. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.
30. Kansy K, Mueller AA, Mücke T, et al. Microsurgical reconstruction of the head and neck region: current concepts of maxillofacial surgery units worldwide. *J Craniomaxillofac Surg*. 2015;43:1364-1368.
31. Rossmiller SR, Cannady SB, Ghanem TA, Wax MK. Transfusion criteria in free flap surgery. *Otolaryngol Head Neck Surg*. 2010;142:359-364.
32. Frank SM, Rothschild JA, Masear CG, et al. Optimizing preoperative blood ordering with data acquired from an anesthesia information management system. *Anesthesiology*. 2013;118:1286-1297.
33. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digital Health*. 2021;3:e195-e203.

How to cite this article: Puladi B, Ooms M, Rieg A, et al. Development of machine learning and multivariable models for predicting blood transfusion in head and neck microvascular reconstruction for risk-stratified patient blood management. *Head & Neck*. 2023;45(6):1389-1405. doi:10.1002/hed.27353