

Model A: WADA Screening for Clean/Transfused

Table: Sort of P001 Sunderby Hematology Data Update_260326_CM Response: Clean/Transfused Validation: Validation 25_75 by ID
SORTED by Code and Time

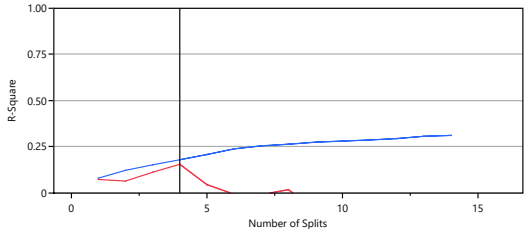
Details

Partition for Clean/Transfused

Validation: Validation 25_75 by ID

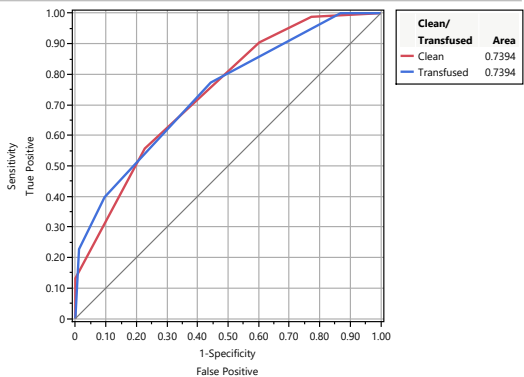
	RSquare	N	of Splits
Training	0.180	255	4
Validation	0.156	85	

Split History

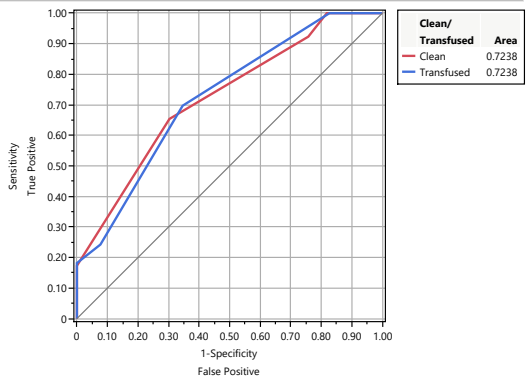


Validation Data in Red

Receiver Operating Characteristic on Training Data



Receiver Operating Characteristic on Validation Data



Bootstrap Forest for Clean/Transfused

Specifications

Target	Clean/Transfused	Training Rows:	255
Validation Column:	Validation 25_75 by ID	Validation Rows:	85
		Test Rows:	0
Number of Trees in the Forest:	100	Number of Terms:	4
Number of Terms Sampled per Split:	3	Bootstrap Samples:	255
		Minimum Splits per Tree:	10
		Minimum Size Split:	5

Overall Statistics

Measure	Training	Validation	Definition
Entropy RSquare	0.3526	0.1479	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.5041	0.2432	$(1 - (L(0)/L(\text{model}))^{2/n}) / (1 - L(0)^{2/n})$
Mean -Log p	0.4172	0.5692	$-\sum \log(p_{ij})/n$
RASE	0.3603	0.4476	$\sqrt{\sum (y_{ij} - p_{ij})^2/n}$
Mean Abs Dev	0.3065	0.3769	$\sum y_{ij} - p_{ij} /n$
Misclassification Rate	0.1804	0.3529	$\sum (p_{ij} \neq p_{\text{Max}})/n$
N	255	85	n

Number

of Trees

8

Confusion Matrix

Training			Validation		
Actual Clean/Transfused	Predicted Count		Actual Clean/Transfused	Predicted Count	
	Clean	Transfused		Clean	Transfused
Clean	157	10	Clean	46	6
Transfused	36	52	Transfused	24	9

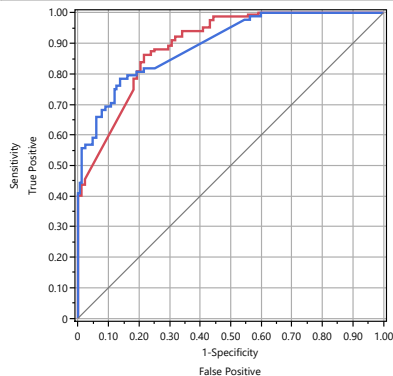
Actual Clean/Transfused	Predicted Rate		Actual Clean/Transfused	Predicted Rate	
	Clean	Transfused		Clean	Transfused
Clean	0.940	0.060	Clean	0.885	0.115
Transfused	0.409	0.591	Transfused	0.727	0.273

Model A: WADA Screening for Clean/Transfused

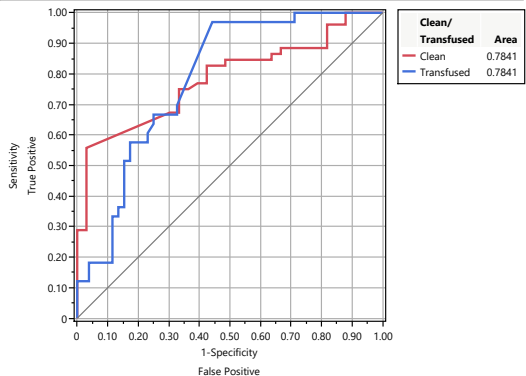
Details

Bootstrap Forest for Clean/Transfused

Receiver Operating Characteristic on Training Data



Receiver Operating Characteristic on Validation Data



Boosted Tree for Clean/Transfused

Specifications

Target	Clean/Transfused	Number of training rows:	255
Validation Column:	Validation 25_75 by ID	Number of validation rows:	85
Number of Layers:	43		
Splits per Tree:	2		
Learning Rate:	0.088		
Overfit Penalty:	0.0001		

Overall Statistics

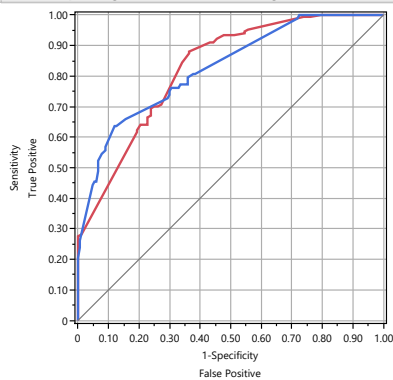
Measure	Training	Validation	Definition
Entropy RSquare	0.2370	0.0692	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.3633	0.1198	$(1 - L(0) / L(\text{model})) * (2/n) / (1 - L(0) * (2/n))$
Mean -Log p	0.4916	0.6217	$\sum -\text{Log}(p_{ij}) / n$
RASE	0.4031	0.4677	$\sqrt{\sum (y_{ij} - p_{ij})^2 / n}$
Mean Abs Dev	0.3539	0.3906	$\sum y_{ij} - p_{ij} / n$
Misclassification Rate	0.2157	0.3294	$\sum (p_{ij} \neq p_{\text{Max}}) / n$
N	255	85	n

Confusion Matrix

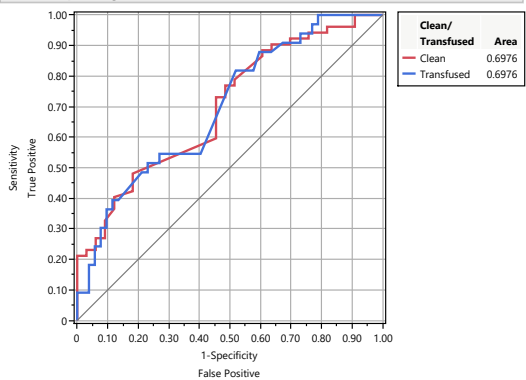
Training			Validation		
Actual	Predicted Count		Actual	Predicted Count	
Clean/Transfused	Clean	Transfused	Clean/Transfused	Clean	Transfused
Clean	156	11	Clean	49	3
Transfused	44	44	Transfused	25	8

Actual	Predicted Rate		Actual	Predicted Rate	
Clean/Transfused	Clean	Transfused	Clean/Transfused	Clean	Transfused
Clean	0.934	0.066	Clean	0.942	0.058
Transfused	0.500	0.500	Transfused	0.758	0.242

Receiver Operating Characteristic on Training Data



Receiver Operating Characteristic on Validation Data



Model A: WADA Screening for Clean/Transfused

Details

Neural

Validation Column: Validation 25_75 by ID

Model Launch

Hidden Layer Structure

Number of nodes of each activation type

Activation Sigmoid Identity Radial

Layer TanH Linear Gaussian

First 3 0 0

Second 0 0 0

Second layer is closer to X's in two layer models.

Boosting

Fit an additive sequence of models scaled by the learning rate.

Number of Models 0

Learning Rate 0.1

Fitting Options

☐ Transform Covariates

Penalty Method Squared

Number of Tours 1

Model NTanH(3)NBoost(2)

Training

Clean/Transfused

Measures	Value
Generalized RSquare	0.1221591
Entropy RSquare	0.0723219
RASE	0.4513221
Mean Abs Dev	0.4112363
Misclassification Rate	0.3202614
-LogLikelihood	90.343483
Sum Freq	153

Confusion Matrix

Actual	Predicted Count	
Clean/Transfused	Clean	Transfused
Clean	95	7
Transfused	42	9

Confusion Rates

Actual	Predicted Rate	
Clean/Transfused	Clean	Transfused
Clean	0.931	0.069
Transfused	0.824	0.176

Validation

Clean/Transfused

Measures	Value
Generalized RSquare	0.4391245
Entropy RSquare	0.3007129
RASE	0.3670258
Mean Abs Dev	0.3302412
Misclassification Rate	0.1587302
-LogLikelihood	27.531799
Sum Freq	63

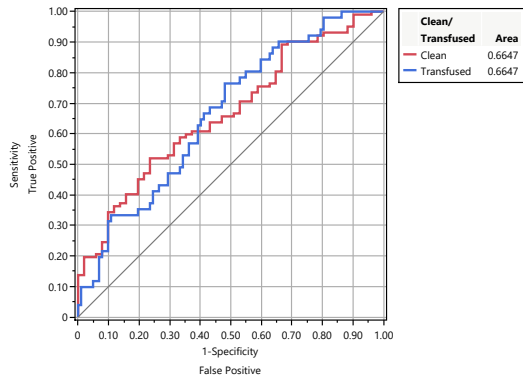
Confusion Matrix

Actual	Predicted Count	
Clean/Transfused	Clean	Transfused
Clean	43	0
Transfused	10	10

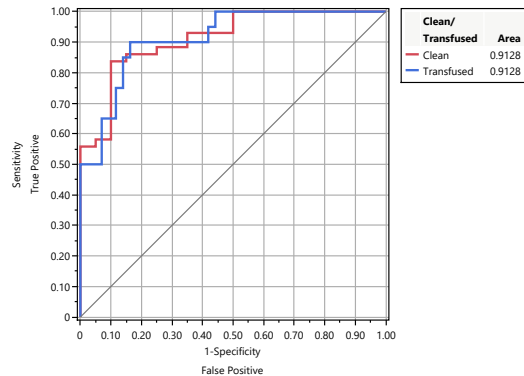
Confusion Rates

Actual	Predicted Rate	
Clean/Transfused	Clean	Transfused
Clean	1.000	0.000
Transfused	0.500	0.500

Receiver Operating Characteristic on Training Data



Receiver Operating Characteristic on Validation Data



Model A: WADA Screening for Clean/Transfused

Details

Support Vector Machine

Model Comparison

Show	Method	Kernel Function	Cost	Gamma	# SV	Training Misclassification Rate	Validation Misclassification Rate	Validation Generalized RSquare	Probability Threshold	Threshold Validation Misclassification Rate	Best
☒	Model 1	Radial Basis Function	1	0.25	110	0.31373	0.26984	0.35453	0.01	0.30159	Smallest Misclassification Rate Largest RSquare

Validation Column: Validation 25_75 by ID

Model Launch

Kernel Function

Radial Basis Function

☐ Tuning Design

Cost

1

Gamma

0.25

Support Vector Machine Model 1

Model Summary

Estimation Details

Response	Clean/Transfused	Cost	1
Validation Method	Validation Column	Gamma	0.25
Kernel Function	Radial Basis Function		
Measure	Training	Validation	
Number of rows	153	63	
Sum of Frequencies	153	63	
Misclassification Rate	0.3137255	0.2698413	
Number of Support Vectors	110	110	

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.1243	0.2333	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
Generalized RSquare	0.2033	0.3545	$(1 - (L(0)/L(\text{model}))^{(2/n)}) / (1 - L(0)^{(2/n)})$
Mean -Log p	0.5574	0.4791	$-\sum (y(i) - p(i)) / n$
RASE	0.4316	0.3942	$\sqrt{\sum (y(i) - p(i))^2 / n}$
Mean Abs Dev	0.3762	0.3393	$\sum y(i) - p(i) / n$
Misclassification Rate	0.2680	0.2540	$\sum (p(i) \neq pMax) / n$
N	153	63	n

Confusion Matrix

Set Probability Threshold

Probability Threshold 0.01

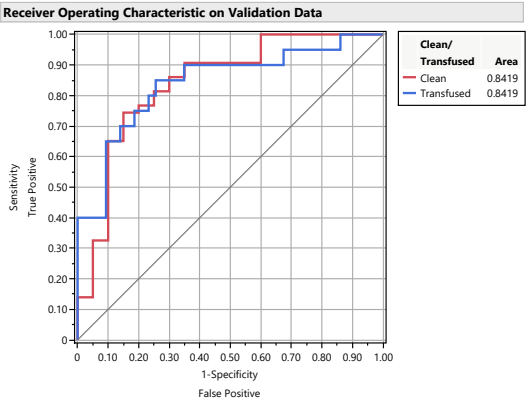
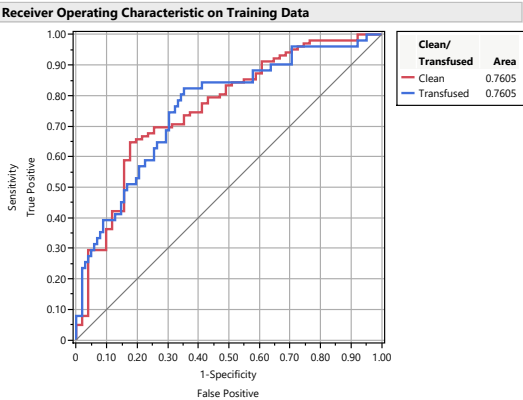
Training

Actual	Predicted Rate	Misclassification Rate
Clean/Transfused	Clean Transfused	
Clean	1.000 0.000	0.3203
Transfused	0.961 0.039	
Actual	Predicted Count	
Clean/Transfused	Clean Transfused	
Clean	102 0	
Transfused	49 2	

Validation

Actual	Predicted Rate	Misclassification Rate
Clean/Transfused	Clean Transfused	
Clean	1.000 0.000	0.3016
Transfused	0.950 0.050	
Actual	Predicted Count	
Clean/Transfused	Clean Transfused	
Clean	43 0	
Transfused	19 1	

ROC Curve for Clean/Transfused = Clean



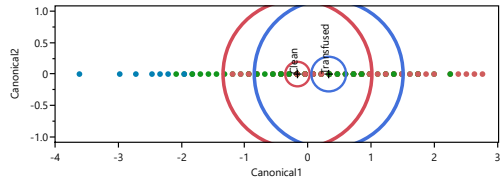
Model A: WADA Screening for Clean/Transfused

Details

Discriminant Analysis

Discriminant Method: Linear
Classification: Clean/Transfused
Validation: Validation 25_75 by ID

Canonical Plot



Score Summaries

Columns

Hemoglobin (g·L⁻¹)

		Number	Percent	Entropy	
Source	Count	Misclassified	Misclassified	RSquare	-2LogLikelihood
Training	153	66	43.1373	-0.0416	202.875
Validation	69	15	21.7391	0.15990	

Training			Validation		
Actual	Predicted Count		Actual	Predicted Count	
Clean/Transfused	Clean	Transfused	Clean/Transfused	Clean	Transfused
Clean	56	46	Clean	35	10
Transfused	20	31	Transfused	5	19

Actual	Predicted Rate		Actual	Predicted Rate	
Clean/Transfused	Clean	Transfused	Clean/Transfused	Clean	Transfused
Clean	0.549	0.451	Clean	0.778	0.222
Transfused	0.392	0.608	Transfused	0.208	0.792

Nominal Logistic Fit for Clean/Transfused

Effect Summary

Source	Logworth	PValue
Epo (UL ⁻¹)	0.520	0.30218
Hemoglobin (g·L ⁻¹)	0.397	0.40069
OFF-hr	0.361	0.43594
Reticulocytes/Erythrocytes (%)	0.342	0.45455

Converged in Gradient, 5 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	5.447144	4	10.89429	0.0278*
Full	91.939524			
Reduced	97.386668			

RSquare (U)	0.0559
AICc	194.287
BIC	209.031
Observations (or Sum Wgts)	153

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.0559	0.2450	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0955	0.3697	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6009	0.4718	$\sum -\log(p[j])/n$
RASE	0.4573	0.3909	$\sqrt{\sum (y[j]-p[j])^2/n}$
Mean Abs Dev	0.4167	0.3349	$\sum y[j]-p[j] /n$
Misclassification Rate	0.3268	0.2063	$\sum (p[j] \neq pMax)/n$
N	153	63	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	148	91.939524	183.879
Saturated	152	0.000000	
Fitted	4	91.939524	0.0241*

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-19.388242	12.496665	2.41	0.1208
Hemoglobin (g·L ⁻¹)	0.92532783	1.1010701	0.71	0.4007
Reticulocytes/Erythrocytes (%)	-0.7007444	0.9370027	0.56	0.4545
Epo (UL ⁻¹)	-0.040876	0.039617	1.06	0.3022
OFF-hr	-0.8549706	1.0974368	0.61	0.4359

For log odds of Transfused/Clean

Effect Wald Tests

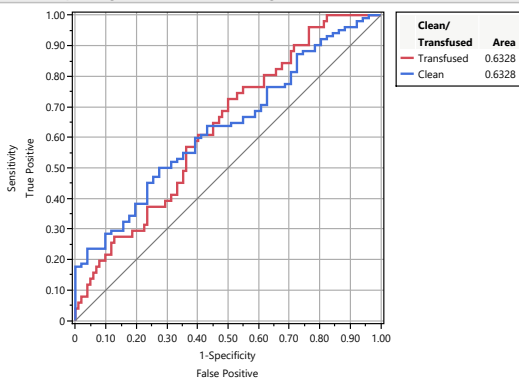
Source	Nparm	Wald		
		DF	ChiSquare	Prob>ChiSq
Hemoglobin (g·L ⁻¹)	1	1	0.70625469	0.4007
Reticulocytes/Erythrocytes (%)	1	1	0.55929061	0.4545
Epo (UL ⁻¹)	1	1	1.06457074	0.3022
OFF-hr	1	1	0.60693663	0.4359

Model A: WADA Screening for Clean/Transfused

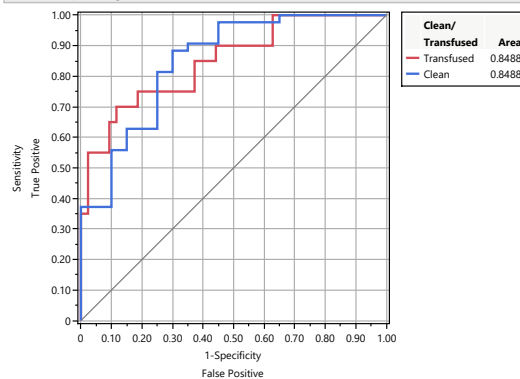
Details

Nominal Logistic Fit for Clean/Transfused

Receiver Operating Characteristic on Training Data



Receiver Operating Characteristic on Validation Data



Generalized Regression for Clean/Transfused = Transfused

Model Comparison

Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Validation Generalized RSquare
☒	Binomial	Lasso	Validation Column	4	192.73246	204.58395	0.0905149	0.4009006

Model Launch

Response Distribution

Binomial

Estimation Method

Lasso

☐ Adaptive

Validation Method

Validation Column

☐ Early Stopping

Validation Column: Validation 25_75 by ID

Binomial Lasso with Validation Column

Model Summary

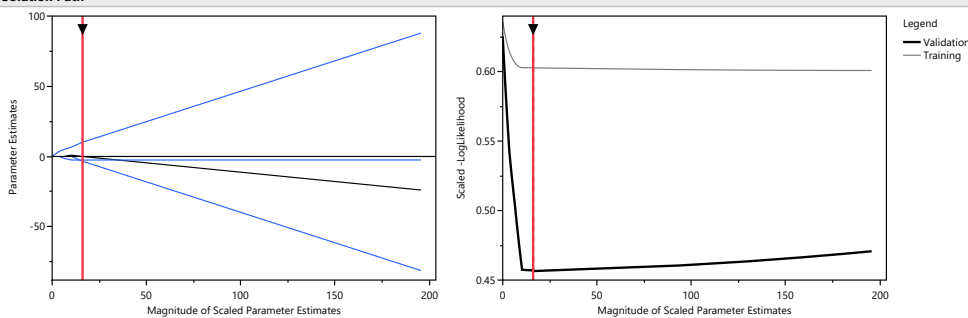
Response	Clean/Transfused	Number of Grid Points	150
Distribution	Binomial	Minimum Penalty Fraction	1e-4
Estimation Method	Lasso	Grid Scale	Square Root
Validation Method	Validation Column		
Probability Model Link	Logit		

Validation Column: Validation 25_75 by ID

Measure	Training	Validation
Number of rows	153	63
Sum of Frequencies	153	63
-LogLikelihood	92.231097	28.758571
Number of Parameters	4	4
BIC	204.58395	74.089682
AICc	192.73246	66.206798
Generalized RSquare	0.0905149	0.4009006
Lambda Penalty	0.0024856	.

Estimation Details

Solution Path



Parameter Estimates for Original Predictors

Term	Wald			Prob >	Lower 95%	Upper 95%
	Estimate	Std Error	ChiSquare	ChiSquare		
Intercept	-10.57066	3.30914	10.204091	0.0014*	-17.05646	-4.08487
Hemoglobin (g L ⁻¹)	0.1028528	0.0704307	2.1325918	0.1442	-0.035189	0.2408945
Reticulocytes/Erythrocytes (%)	0	0	0	1.0000	0	0
Epo (UL ⁻¹)	-0.041607	0.0406397	1.0481874	0.3059	-0.12126	0.038045
OFF-hr	-0.034483	0.0687443	0.2516173	0.6159	-0.16922	0.1002533

Validation

Method		N	Sum Wgt	Entropy		Misclassification			Generalized
				RSquare	Rate	AUC	RASE	RSquare	
Neural Boosted	†	63	63.0000	0.3007	0.1587	0.9128	0.36703	0.4391	
Generalized Regression Lasso	†	63	63.0000	0.2696	0.1746	0.9012	0.38064	0.4009	
Nominal Logistic		63	63.0000	0.2450	0.2063	0.8488	0.39085	0.3697	
Support Vector Machines		63	63.0000	0.2333	0.2540	0.8419	0.39421	0.3545	
Discriminant		69		0.1599	0.2174				
Decision Tree	†	85	85.0000	0.1559	0.3412	0.7238	0.44498	0.2550	
Bootstrap Forest	†	85	85.0000	0.1479	0.3529	0.7841	0.44758	0.2432	
Boosted Tree	†	85	85.0000	0.0692	0.3294	0.6976	0.46766	0.1198	

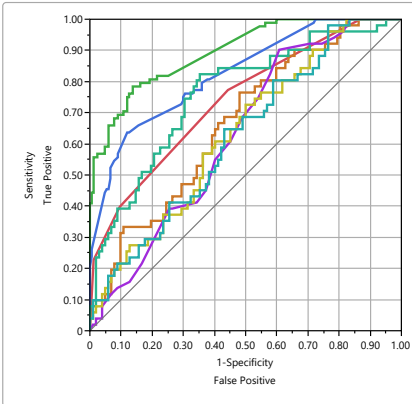
† indicates that the method uses validation data to tune model.

Model A: WADA Screening for Clean/Transfused

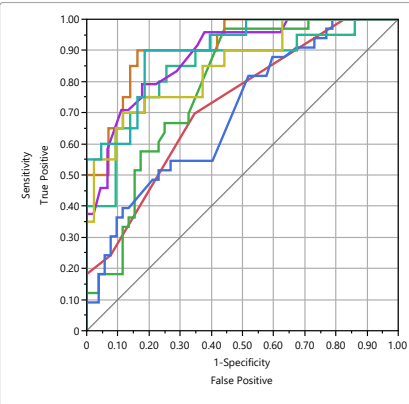
Sum Freq is suppressed when it is the same as N.

ROC Curve

Training



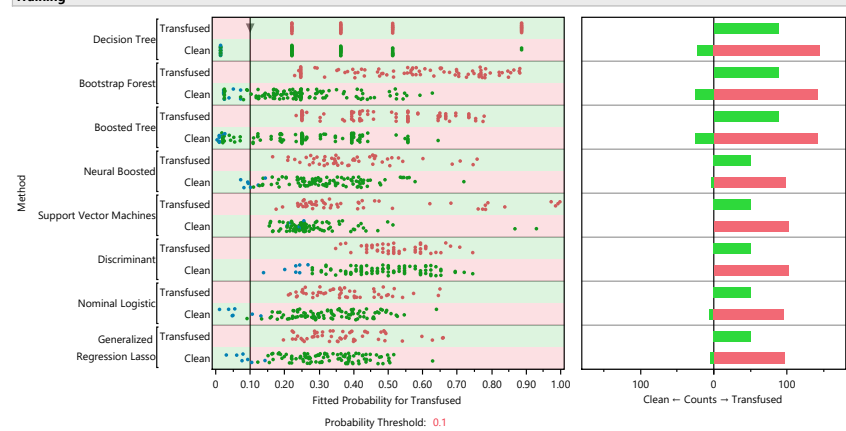
Validation



Method	AUC	
	Training	Validation
Decision Tree	0.7394	0.7238
Bootstrap Forest	0.8960	0.7841
Boosted Tree	0.8259	0.6976
Neural Boosted	0.6647	0.9128
Support Vector Machines	0.7605	0.8419
Discriminant	0.6300	0.8847
Nominal Logistic	0.6328	0.8488
Generalized Regression Lasso	0.6228	0.9012

Decision Thresholds

Training



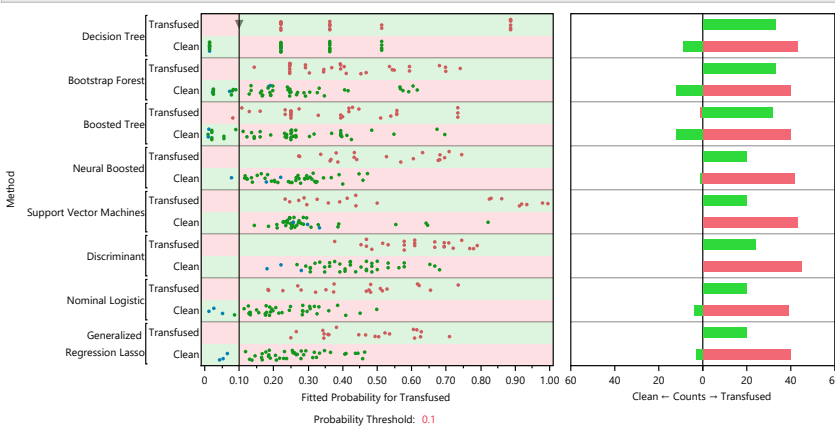
Method	Actual	Predicted		Rates	
		Clean	Transfused	Clean	Transfused
Decision Tree	Transfused	0	88	0.000	1.000
	Clean	22	145	0.132	0.868
Bootstrap Forest	Transfused	0	88	0.000	1.000
	Clean	25	142	0.150	0.850
Boosted Tree	Transfused	0	88	0.000	1.000
	Clean	25	142	0.150	0.850
Neural Boosted	Transfused	0	51	0.000	1.000
	Clean	4	98	0.039	0.961
Support Vector Machines	Transfused	0	51	0.000	1.000
	Clean	0	102	0.000	1.000
Discriminant	Transfused	0	51	0.000	1.000
	Clean	0	102	0.000	1.000
Nominal Logistic	Transfused	0	51	0.000	1.000
	Clean	6	96	0.059	0.941
Generalized Regression Lasso	Transfused	0	51	0.000	1.000
	Clean	5	97	0.049	0.951

Metrics

Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Decision Tree	88	0	145	22	1.0000	0.1317	0.3777	0.4314	0.5483	0.2231
Bootstrap Forest	88	0	142	25	1.0000	0.1497	0.3826	0.4431	0.5535	0.2393
Boosted Tree	88	0	142	25	1.0000	0.1497	0.3826	0.4431	0.5535	0.2393
Neural Boosted	51	0	98	4	1.0000	0.0392	0.3423	0.3595	0.51	0.1159
Support Vector Machines	51	0	102	0	1.0000	0.0000	0.3333	0.3333	0.5	.
Discriminant	51	0	102	0	1.0000	0.0000	0.3333	0.3333	0.5	.
Nominal Logistic	51	0	96	6	1.0000	0.0588	0.3469	0.3725	0.5152	0.1429
Generalized Regression Lasso	51	0	97	5	1.0000	0.0490	0.3446	0.366	0.5126	0.13

Where Transfused is considered Positive. Probability Threshold: 0.100

Validation



Method	Actual	Predicted		Rates	
		Clean	Transfused	Clean	Transfused
Decision Tree	Transfused	0	33	0.000	1.000
	Clean	9	43	0.173	0.827
Bootstrap Forest	Transfused	0	33	0.000	1.000
	Clean	12	40	0.231	0.769
Boosted Tree	Transfused	1	32	0.030	0.970
	Clean	12	40	0.231	0.769
Neural Boosted	Transfused	0	20	0.000	1.000
	Clean	1	42	0.023	0.977
Support Vector Machines	Transfused	0	20	0.000	1.000
	Clean	0	43	0.000	1.000
Discriminant	Transfused	0	24	0.000	1.000
	Clean	0	45	0.000	1.000
Nominal Logistic	Transfused	0	20	0.000	1.000
	Clean	4	39	0.093	0.907
Generalized Regression Lasso	Transfused	0	20	0.000	1.000
	Clean	3	40	0.070	0.930

Metrics

Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Decision Tree	33	0	43	9	1.0000	0.1731	0.4342	0.4941	0.6055	0.2741
Bootstrap Forest	33	0	40	12	1.0000	0.2308	0.4521	0.5294	0.6226	0.323
Boosted Tree	32	1	40	12	0.9697	0.2308	0.4444	0.5176	0.6095	0.2714
Neural Boosted	20	0	42	1	1.0000	0.0233	0.3226	0.3333	0.4878	0.0866
Support Vector Machines	20	0	43	0	1.0000	0.0000	0.3175	0.3175	0.4819	.
Discriminant	24	0	45	0	1.0000	0.0000	0.3478	0.3478	0.5161	.
Nominal Logistic	20	0	39	4	1.0000	0.0930	0.339	0.381	0.5063	0.1776
Generalized Regression Lasso	20	0	40	3	1.0000	0.0698	0.3333	0.3651	0.5	0.1525