

Table S1 – Neural Network Outputs

Predicted scores	
T 0 Scale 1 High risk 0	0.0073
T 0 Scale 1 High risk 1	0.7745
T 1 Scale 1 High risk 0	0.1163
T 1 Scale 1 High risk 1	0.2592
T 2 Scale 1 High risk 0	0.1717
T 2 Scale 1 High risk 1	0.4648
T 3 Scale 1 High risk 0	0.2301
T 3 Scale 1 High risk 1	0.9653
T 4 Scale 1 High risk 0	0.1112
T 4 Scale 1 High risk 1	0.8739
T 0 Scale 2 High risk 0	0.3323
T 0 Scale 1 High risk 1	0.6453
T 1 Scale 2 High risk 0	0.2497
T 1 Scale 2 High risk 1	0.3489
T 2 Scale 2 High risk 0	0.3848
T 2 Scale 2 High risk 1	0.4623
T 3 Scale 2 High risk 0	0.6101
T 3 Scale 2 High risk 1	0.8974
T 4 Scale 2 High risk 0	0.4644
T 4 Scale 2 High risk 1	0.1951
T 0 Scale 3 High risk 0	0.5945
T 0 Scale 3 High risk 1	0.4578
T 1 Scale 3 High risk 0	0.5242
T 1 Scale 3 High risk 1	0.4286
T 2 Scale 3 High risk 0	0.531
T 2 Scale 3 High risk 1	0.609
T 3 Scale 3 High risk 0	0.4783
T 3 Scale 3 High risk 1	0.5757
T 4 Scale 3 High risk 0	0.4613
T4 Scale 3 High risk 1	0.6844

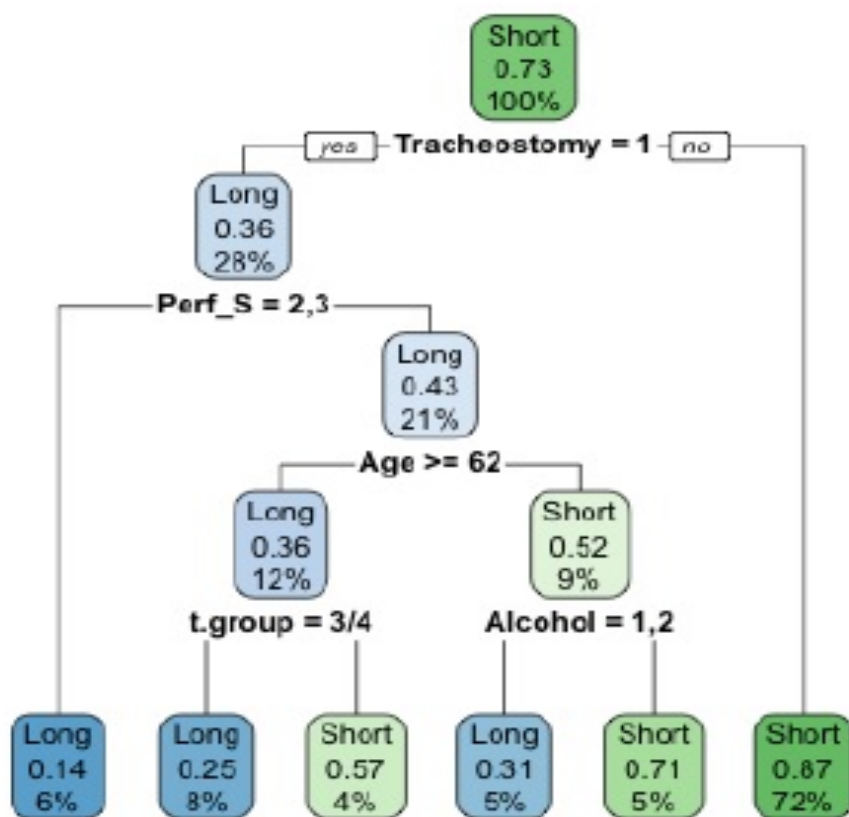


Figure S1 Decision tree output.

Table S2 Positivity of surgical margins Bayes Probability Table

Lip	Oral Cavity	Pharynx (inc tonsil)	Nasopharynx	Hypopharynx	Supraglottis	Larynx	Subglottis	Para-nasal sinuses	Neck only	Salivary gland	Other
0.00	0.05	0.74	0.07	0.01	0.00	0.00	0.07	0.01	0.03	0.02	0.01
0.00	0.03	0.58	0.07	0.01	0.01	0.00	0.10	0.01	0.15	0.01	0.02
T = 0	T=1	T=2	T=3	T=4							
0.05	0.36	0.27	0.09	0.23							
0.11	0.17	0.22	0.10	0.41							
No Extracapsular spread	Extracapsular spread										
0.83	0.17										
0.63	0.37										

Table S3 Length of Hospital Stay Linear Regression model (when expected length of stay <15 days)

Call:					
lm(formula	=	dave.formula,	data	=	ls.train)
Residuals:					
	Min	1Q	Median	3Q	Max
	-13.576	-4.348	-1.301	1.891	39.385
Coefficients:					
	Estimate	Std.	Error	t	Pr(> t)
(Intercept)	-6.9888	2.82555	-2.473	0.01395	0.05
Age	0.10962	0.03654	3	0.00293	0.01
t.group3/4	0.09353	1.10087	0.085	0.93235	
Perf_S1	1.08614	1.06775	1.017	0.30989	
Perf_S2	2.24747	1.38265	1.625	0.10514	
Perf_S3	1.6625	2.0878	0.796	0.42651	
Tracheostomy1	6.0708	1.42193	4.269	2.65E-005	0.001
High_risk1	3.21311	1.23278	2.606	0.00962	0.01
ScaleofSurgery2	3.79137	1.32831	2.854	0.00462	0.01
ScaleofSurgery3	8.85271	1.56941	5.641	3.99E-008	0.001
Alcohol2	-0.89413	1.12228	-0.797	0.42627	
Alcohol3	2.3253	1.47868	1.573	0.1169	
Alcohol4	2.20704	1.37171	1.609	0.1087	
Alcohol5	3.23868	1.9429	1.667	0.0966	

Table S4 ROC curve analysis

ROC Curve analysis					
Dependent Y	Flap failure				
Method	Enter				
Sample size	1593				
Positive cases ^a	75 (4.71%)				
Negative cases ^b	1518 (95.29%)				
^a fLAP_FAILURE = 1					
^b fLAP_FAILURE = 0					
Overall Model Fit					
Null model -2 Log Likelihood	604.795				
Full model -2 Log Likelihood	559.712				
Chi-squared	45.084				
DF	1				
Significance level	P < 0.0001				
Cox & Snell R ²	0.0279				
Nagelkerke R ²	0.08833				
Coefficients and Standard Errors					
Variable	Coefficient	Std. Error	Wald	P	
p	6.01799	0.83259	52.2441	<0.0001	
Constant	-3.54525	0.15781	504.7088	<0.0001	
Odds Ratios and 95% Confidence Intervals					
Variable	Odds ratio	95% CI			
p	410.7502	80.3268 to 2100.3663			
Hosmer & Lemeshow test					
Chi-squared	6.996				
DF	8				
Significance level	P = 0.5371				
Group	Y=0		Y=1		Total
	Observed	Expected	Observed	Expected	
1	171	168.047	2	4.953	173
2	109	106.784	1	3.216	110
3	155	156.228	6	4.772	161
4	170	167.695	3	5.305	173
5	157	157.797	6	5.203	163
6	151	150.713	5	5.287	156
7	148	149.357	7	5.643	155
8	155	153.43	5	6.57	160
9	148	150.647	11	8.353	159
10	154	157.301	29	25.699	183
Classification table (cut-off value p=0.1)					
Actual group	Predicted group		Percent correct		
	0	1			
Y = 0	1456	62	95.92%		
Y = 1	56	19	25.33%		
Percent of cases correctly classified			92.59%		
ROC curve analysis					
Area under the ROC curve (AUC)	0.719				
Standard Error	0.0319				
95% Confidence interval	0.696 to 0.741				
Brier's Score	0.44				

Appendix 1

A confusion matrix or contingency table. The different types of errors can be summarized in a matrix as (where n is the number of observations).

	positive label	negative label
predicted positive	TP/n	FP/n
predicted negative	FN/n	TN/n

TP = # true positives, FP = # false positives, TN = # true negatives, FN = # false negatives

Sensitivity (also known as recall) = $TP / (TP + FN)$ = (number of true positive assessment) / (Number of all positive assessment)

Specificity = $TN / (TN + FP)$ = (number of true negative assessment) / (number of all negative assessment)

Accuracy = $(TN + TP) / (TN + TP + FN + FP)$ = (number of correct assessments) / number of all assessments

Positive predictive value (also known as precision) = $TP / (TP + FP)$

Negative predictive value = $TN / (TN + FN)$

F1 score = $2 \cdot TP / (2 \cdot TP + FP + FN)$

A plot of the true positive rate (TPR) versus the false positive rate (FPR) is called a receiver operating characteristic (ROC) curve:

True positive rate = TP / # positives; false positive rate = FP / # negatives

Error types in a two-class problem

- False positives (type I error): True label is -1, predicted label is +1.
- False negative (type II error): True label is +1, predicted label is -1.

Error rate ER = $\frac{\text{\# wrong predictions}}{\text{\# observations}} = \frac{FP + FN}{FP + FN + TP + TN}$

Does not distinguish errors between classes.

Relevance

Distinction between error types is crucial, e.g., if:

- Classes differ significantly in size;
- One type of error has worse consequences than other.

Hosmer-Lemeshow Goodness of Fit Test

This is a statistical test for 'goodness of fit' for logistic regression models. It is used frequently in risk prediction models. It measures the concordance of the observed event rates and the expected event rates in subgroups of the model population. When the expected rates and observed event rates in subgroups are similar ($P > 0.05$) the model is described as well calibrated.

$$H = \sum_{q=1}^{10} \left(\frac{(\text{Observed. } A - \text{Expected. } A)^2}{\text{Expected. } A} + \frac{(\text{Observed. not. } A - \text{Expected. not. } A)^2}{\text{Expected. not. } A} \right)$$

Brier's Score

$$BS = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

in which f_t is the predicted probability, o_t is the actual outcome of the event at instance t (0 if it does not occur, 1 if it does occur) and N is the number of patient care episodes. It is, in effect, the mean squared error of the forecast.