

Table SI. K-Nearest neighbors Classification [EGFR vs. negative group (TN)].

K-Nearest neighbors classification							
Nearest neighbors	Weights	Distance	n (Train)	n (Validation)	n (Test)	Validation accuracy	Test accuracy
8	rectangular	Euclidean	48	11	14	0.6780	0.7857

The model is optimized with respect to the validation set accuracy.

Table SII. Confusion matrix.

Category		Predicted	
		EGFR	TN
Observed	EGFR	6	3
	TN	0	5

Table SIII. Model performance metrics.

Performance metrics	EGFR	TN	Average / Total
Support	9	5	14
Accuracy	0.7857	0.7857	0.7857
Precision (Positive Predictive Value)	1.0000	0.6250	0.8661
Recall (True Positive Rate)	0.6667	1.0000	0.7857
False Positive Rate	0.0000	0.3333	0.1667
False Discovery Rate	0.0000	0.3750	0.1875
F1 Score	0.8000	0.7692	0.7890
Matthews Correlation Coefficient	0.6455	0.6455	0.6455
Area Under Curve (AUC)	0.9667	0.9667	0.9667
Negative Predictive Value	0.6250	1.0000	0.8125
True Negative Rate	1.0000	0.6667	0.8333
False Negative Rate	0.3333	0.0000	0.1667
False Omission Rate	0.3750	0.0000	0.1875
Threat Score	2.0000	0.8333	1.4167
Statistical Parity	0.4286	0.5714	1.0000

All metrics are calculated for every class against all other classes.

Table SIV. K-Nearest neighbors classification (ALK vs. negative group).

K-Nearest neighbors classification							
Nearest neighbors	Weights	Distance	n (Train)	n (Validation)	n (Test)	Validation accuracy	Test accuracy
9	rectangular	Euclidean	39	9	11	0.4583	0.7273

The model is optimized with respect to the validation set accuracy.

Table SV. Confusion matrix.

Category		Predicted	
		ALK	TN
Observed	ALK	6	3
	TN	0	5

Table SVI. Model performance metrics.

Performance metrics	ALK	TN	Average / Total
Support	5	6	11
Accuracy	0.7273	0.7273	0.7273
Precision (Positive Predictive Value)	0.7500	0.7143	0.7305
Recall (True Positive Rate)	0.6000	0.8333	0.7273
False Positive Rate	0.1667	0.4000	0.2833
False Discovery Rate	0.2500	0.2857	0.2679
F1 Score	0.6667	0.7692	0.7226
Matthews Correlation Coefficient	0.4485	0.4485	0.4485
Area Under Curve (AUC)	0.8167	0.8167	0.8167
Negative Predictive Value	0.7143	0.7500	0.7321
True Negative Rate	0.8333	0.6000	0.7167
False Negative Rate	0.4000	0.1667	0.2833
False Omission Rate	0.2857	0.2500	0.2679
Threat Score	0.7500	1.0000	0.8750
Statistical Parity	0.3636	0.6364	1.0000

All metrics are calculated for every class against all other classes.

Table SVII. K-Nearest neighbors classification (EGFR vs ALK group).

K-Nearest neighbors classification							
Nearest neighbors	Weights	Distance	n (Train)	n (Validation)	n (Test)	Validation accuracy	Test accuracy
9	rectangular	Euclidean	51	5	14	0.6071	0.8571

The model is optimized with respect to the validation set accuracy.

Table SVIII. Confusion matrix.

Category		Predicted	
		ALK	EGFR
Observed	ALK	4	2
	EGFR	0	8

Table SIX. Model performance metrics.

Performance metrics	ALK	EGFR	Average / Total
Support	6	8	14
Accuracy	0.8571	0.8571	0.8571
Precision (Positive Predictive Value)	1.0000	0.8000	0.8857
Recall (True Positive Rate)	0.6667	1.0000	0.8571
False Positive Rate	0.0000	0.3333	0.1667
False Discovery Rate	0.0000	0.2000	0.1000
F1 Score	0.8000	0.8889	0.8508
Matthews Correlation Coefficient	0.7303	0.7303	0.7303
Area Under Curve (AUC)	0.8333	0.8333	0.8333
Negative Predictive Value	0.8000	1.0000	0.9000
True Negative Rate	1.0000	0.6667	0.8333
False Negative Rate	0.3333	0.0000	0.1667
False Omission Rate	0.2000	0.0000	0.1000
Threat Score	2.0000	2.0000	2.0000
Statistical Parity	0.2857	0.7143	1.0000

All metrics are calculated for every class against all other classes.