

Supplemental 3

Feature Stability

Stability on Arita2018

LDA		0.54 AUC:0.83 (p=1.0) #Features: 64	0.85 AUC:0.8 (p=0.302) #Features: 16	0.74 AUC:0.8 (p=0.098) #Features: 64				
Logistic Regression		0.54 AUC:0.82 (p=0.642) #Features: 64	0.85 AUC:0.8 (p=0.334) #Features: 16			0.43 AUC:0.78 (p=0.052) #Features: 64		
Naive Bayes								
Neural Network	0.79 AUC:0.74 (p=0.051) #Features: 16	0.67 AUC:0.82 (p=0.687) #Features: 16			0.53 AUC:0.78 (p=0.114) #Features: 8			
RBF-SVM		0.62 AUC:0.79 (p=0.069) #Features: 32					0.45 AUC:0.77 (p=0.083) #Features: 8	
Random Forest	0.79 AUC:0.77 (p=0.116) #Features: 16	0.54 AUC:0.79 (p=0.086) #Features: 64						
SVM		0.54 AUC:0.82 (p=0.492) #Features: 64	0.85 AUC:0.8 (p=0.364) #Features: 16	0.74 AUC:0.79 (p=0.056) #Features: 64		0.43 AUC:0.78 (p=0.062) #Features: 64		
XGBoost		0.62 AUC:0.81 (p=0.433) #Features: 32					0.45 AUC:0.79 (p=0.181) #Features: 16	
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MRMR

Figure S16: Stability of the models on dataset Arita2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Carvalho2018

LDA	0.7 AUC:0.62 (p=0.238) #Features: 4	1.0 AUC:0.62 (p=0.223) #Features: 2	0.46 AUC:0.61 (p=0.152) #Features: 4		0.79 AUC:0.62 (p=0.207) #Features: 4		0.22 AUC:0.6 (p=0.144) #Features: 4	0.57 AUC:0.61 (p=0.159) #Features: 2
Logistic Regression	0.84 AUC:0.62 (p=0.214) #Features: 8	1.0 AUC:0.62 (p=0.228) #Features: 2	0.46 AUC:0.61 (p=0.144) #Features: 4		0.82 AUC:0.62 (p=0.221) #Features: 8	0.59 AUC:0.61 (p=0.128) #Features: 16	0.22 AUC:0.61 (p=0.166) #Features: 4	0.57 AUC:0.62 (p=0.21) #Features: 2
Naive Bayes	0.7 AUC:0.65 (p=0.211) #Features: 16	1.0 AUC:0.61 (p=0.145) #Features: 2	0.67 AUC:0.61 (p=0.131) #Features: 16		0.8 AUC:0.62 (p=0.229) #Features: 16			0.57 AUC:0.61 (p=0.189) #Features: 2
Neural Network	0.63 AUC:0.65 (p=0.547) #Features: 64	0.84 AUC:0.64 (p=0.474) #Features: 64	0.56 AUC:0.65 (p=0.746) #Features: 8	0.79 AUC:0.63 (p=0.339) #Features: 16	0.82 AUC:0.65 (p=0.748) #Features: 8	0.59 AUC:0.66 (p=0.848) #Features: 16	0.52 AUC:0.63 (p=0.303) #Features: 64	0.81 AUC:0.67 (p=1.0) #Features: 16
RBF-SVM	0.7 AUC:0.65 (p=0.612) #Features: 4	1.0 AUC:0.61 (p=0.215) #Features: 2	0.46 AUC:0.6 (p=0.116) #Features: 4		0.79 AUC:0.65 (p=0.642) #Features: 4	0.59 AUC:0.62 (p=0.195) #Features: 16	0.16 AUC:0.59 (p=0.093) #Features: 2	0.81 AUC:0.66 (p=0.705) #Features: 16
Random Forest	0.84 AUC:0.62 (p=0.202) #Features: 8	0.86 AUC:0.62 (p=0.218) #Features: 8	0.67 AUC:0.6 (p=0.085) #Features: 16		0.82 AUC:0.63 (p=0.327) #Features: 8		0.55 AUC:0.6 (p=0.055) #Features: 32	0.62 AUC:0.65 (p=0.59) #Features: 4
SVM	0.7 AUC:0.62 (p=0.208) #Features: 4	1.0 AUC:0.62 (p=0.242) #Features: 2	0.46 AUC:0.61 (p=0.138) #Features: 4		0.79 AUC:0.62 (p=0.204) #Features: 4	0.59 AUC:0.61 (p=0.129) #Features: 16	0.22 AUC:0.6 (p=0.14) #Features: 4	0.57 AUC:0.62 (p=0.218) #Features: 2
XGBoost	0.7 AUC:0.62 (p=0.254) #Features: 4	0.86 AUC:0.59 (p=0.054) #Features: 8	0.54 AUC:0.62 (p=0.296) #Features: 32	0.67 AUC:0.59 (p=0.062) #Features: 8	0.8 AUC:0.63 (p=0.318) #Features: 16	0.61 AUC:0.62 (p=0.317) #Features: 8		0.62 AUC:0.65 (p=0.673) #Features: 4
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMe

Figure S17: Stability of the models on dataset Carvalho2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Hosny2018A

LDA		0.59 AUC:0.6 (p=0.287) #Features: 8	0.14 AUC:0.58 (p=0.125) #Features: 4	0.22 AUC:0.58 (p=0.115) #Features: 1	0.51 AUC:0.57 (p=0.061) #Features: 8	0.49 AUC:0.59 (p=0.128) #Features: 16	0.29 AUC:0.58 (p=0.075) #Features: 32	0.77 AUC:0.6 (p=0.257) #Features: 64
Logistic Regression	0.64 AUC:0.59 (p=0.08) #Features: 16	0.59 AUC:0.61 (p=0.345) #Features: 8	0.14 AUC:0.59 (p=0.197) #Features: 4	0.22 AUC:0.58 (p=0.128) #Features: 1	0.56 AUC:0.58 (p=0.094) #Features: 16	0.49 AUC:0.59 (p=0.202) #Features: 16	0.29 AUC:0.59 (p=0.165) #Features: 32	0.77 AUC:0.6 (p=0.303) #Features: 64
Naive Bayes		0.74 AUC:0.58 (p=0.113) #Features: 2		0.53 AUC:0.57 (p=0.064) #Features: 8		0.49 AUC:0.57 (p=0.078) #Features: 16	0.25 AUC:0.59 (p=0.194) #Features: 16	
Neural Network	0.72 AUC:0.59 (p=0.354) #Features: 64	0.76 AUC:0.63 (p=0.673) #Features: 64	0.19 AUC:0.62 (p=0.596) #Features: 64	0.39 AUC:0.61 (p=0.409) #Features: 4	0.55 AUC:0.58 (p=0.139) #Features: 4	0.39 AUC:0.62 (p=0.503) #Features: 8	0.29 AUC:0.64 (p=1.0) #Features: 32	0.77 AUC:0.62 (p=0.541) #Features: 64
RBF-SVM	0.72 AUC:0.61 (p=0.148) #Features: 64	0.61 AUC:0.6 (p=0.232) #Features: 32	0.17 AUC:0.61 (p=0.403) #Features: 16	0.39 AUC:0.58 (p=0.127) #Features: 4	0.56 AUC:0.59 (p=0.163) #Features: 16	0.49 AUC:0.61 (p=0.343) #Features: 16	0.29 AUC:0.6 (p=0.14) #Features: 32	0.63 AUC:0.62 (p=0.509) #Features: 32
Random Forest	0.72 AUC:0.58 (p=0.077) #Features: 64	0.61 AUC:0.58 (p=0.079) #Features: 32		0.53 AUC:0.58 (p=0.111) #Features: 8		0.55 AUC:0.59 (p=0.106) #Features: 64		0.77 AUC:0.62 (p=0.451) #Features: 64
SVM	0.72 AUC:0.6 (p=0.207) #Features: 64	0.59 AUC:0.61 (p=0.417) #Features: 8	0.14 AUC:0.59 (p=0.16) #Features: 4	0.39 AUC:0.57 (p=0.104) #Features: 4	0.51 AUC:0.58 (p=0.082) #Features: 8	0.49 AUC:0.59 (p=0.194) #Features: 16	0.32 AUC:0.6 (p=0.306) #Features: 64	0.77 AUC:0.6 (p=0.246) #Features: 64
XGBoost	0.72 AUC:0.57 (p=0.063) #Features: 64		0.19 AUC:0.58 (p=0.092) #Features: 64	0.53 AUC:0.58 (p=0.122) #Features: 8	0.75 AUC:0.57 (p=0.081) #Features: 64	0.49 AUC:0.59 (p=0.18) #Features: 32		0.63 AUC:0.6 (p=0.25) #Features: 32
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MRFMR

Figure S18: Stability of the models on dataset Hosny2018A. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Hosny2018B

LDA						0.51 AUC:0.71 (p=0.495) #Features: 8	0.39 AUC:0.69 (p=0.212) #Features: 32	
Logistic Regression			0.76 AUC:0.69 (p=0.22) #Features: 64			0.51 AUC:0.72 (p=0.726) #Features: 8	0.39 AUC:0.69 (p=0.086) #Features: 32	
Naive Bayes	0.81 AUC:0.69 (p=0.074) #Features: 64	0.88 AUC:0.68 (p=0.051) #Features: 64				0.51 AUC:0.68 (p=0.139) #Features: 8		
Neural Network	0.76 AUC:0.73 (p=1.0) #Features: 32		0.33 AUC:0.68 (p=0.076) #Features: 8	0.76 AUC:0.67 (p=0.196) #Features: 64	0.57 AUC:0.66 (p=0.064) #Features: 32	0.45 AUC:0.71 (p=0.463) #Features: 4	0.39 AUC:0.69 (p=0.385) #Features: 32	0.64 AUC:0.68 (p=0.25) #Features: 32
RBF-SVM				0.76 AUC:0.65 (p=0.086) #Features: 64		0.51 AUC:0.65 (p=0.054) #Features: 8		0.46 AUC:0.66 (p=0.191) #Features: 8
Random Forest		0.43 AUC:0.65 (p=0.056) #Features: 2	0.36 AUC:0.67 (p=0.124) #Features: 4					
SVM				0.76 AUC:0.67 (p=0.1) #Features: 64		0.51 AUC:0.71 (p=0.403) #Features: 8		0.55 AUC:0.67 (p=0.07) #Features: 16
XGBoost	0.76 AUC:0.68 (p=0.096) #Features: 32	0.24 AUC:0.67 (p=0.054) #Features: 1	0.27 AUC:0.67 (p=0.124) #Features: 32	0.19 AUC:0.66 (p=0.079) #Features: 2				
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMR

Figure S19: Stability of the models on dataset Hosny2018B. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Hosny2018C

LDA								
Logistic Regression								
Naive Bayes								
Neural Network		0.34 AUC:0.57 (p=0.059) #Features: 2			0.35 AUC:0.56 (p=0.064) #Features: 32	0.27 AUC:0.57 (p=0.057) #Features: 2	0.19 AUC:0.58 (p=0.109) #Features: 4	
RBF-SVM	0.15 AUC:0.61 (p=0.254) #Features: 1	0.61 AUC:0.63 (p=0.261) #Features: 32		0.87 AUC:0.69 (p=1.0) #Features: 64	0.21 AUC:0.57 (p=0.052) #Features: 4	0.25 AUC:0.66 (p=0.557) #Features: 8	0.33 AUC:0.62 (p=0.291) #Features: 1	
Random Forest								
SVM		0.61 AUC:0.59 (p=0.077) #Features: 32		0.87 AUC:0.68 (p=0.824) #Features: 64		0.38 AUC:0.62 (p=0.254) #Features: 64		
XGBoost					0.15 AUC:0.57 (p=0.076) #Features: 1		0.19 AUC:0.59 (p=0.134) #Features: 4	
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMR

Figure S20: Stability of the models on dataset Hosny2018C. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Ramella2018

LDA			0.31 AUC:0.73 (p=0.137) #Features: 64	1.0 AUC:0.69 (p=0.062) #Features: 64		0.8 AUC:0.8 (p=0.804) #Features: 64		0.78 AUC:0.75 (p=0.188) #Features: 64
Logistic Regression	0.6 AUC:0.7 (p=0.091) #Features: 64		0.31 AUC:0.77 (p=0.5) #Features: 64	1.0 AUC:0.76 (p=0.436) #Features: 16		0.62 AUC:0.81 (p=1.0) #Features: 32	0.48 AUC:0.73 (p=0.158) #Features: 64	0.75 AUC:0.76 (p=0.235) #Features: 32
Naive Bayes								
Neural Network		0.38 AUC:0.7 (p=0.096) #Features: 8	0.31 AUC:0.71 (p=0.079) #Features: 64	1.0 AUC:0.73 (p=0.243) #Features: 64	0.44 AUC:0.71 (p=0.156) #Features: 16	0.8 AUC:0.79 (p=0.75) #Features: 64		0.78 AUC:0.74 (p=0.164) #Features: 64
RBF-SVM				1.0 AUC:0.71 (p=0.125) #Features: 16		0.62 AUC:0.74 (p=0.11) #Features: 32		
Random Forest								
SVM	0.6 AUC:0.71 (p=0.126) #Features: 64		0.31 AUC:0.75 (p=0.19) #Features: 64	1.0 AUC:0.77 (p=0.485) #Features: 16		0.62 AUC:0.81 (p=1.0) #Features: 32	0.48 AUC:0.72 (p=0.127) #Features: 64	0.75 AUC:0.77 (p=0.319) #Features: 32
XGBoost								
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMe

Figure S21: Stability of the models on dataset Ramella2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Lu2019

LDA								0.6 AUC:0.71 (p=0.074) #Features: 16
Logistic Regression		0.81 AUC:0.72 (p=0.335) #Features: 64			0.72 AUC:0.7 (p=0.134) #Features: 16			0.6 AUC:0.73 (p=0.423) #Features: 16
Naive Bayes								
Neural Network	0.77 AUC:0.68 (p=0.057) #Features: 16	0.81 AUC:0.71 (p=0.188) #Features: 64	0.38 AUC:0.73 (p=0.433) #Features: 32	0.85 AUC:0.73 (p=0.392) #Features: 32	0.72 AUC:0.72 (p=0.269) #Features: 32	0.41 AUC:0.71 (p=0.294) #Features: 8	0.44 AUC:0.7 (p=0.107) #Features: 64	0.6 AUC:0.75 (p=0.902) #Features: 16
RBF-SVM		0.81 AUC:0.72 (p=0.239) #Features: 64	0.32 AUC:0.72 (p=0.283) #Features: 64	0.85 AUC:0.72 (p=0.235) #Features: 32	0.72 AUC:0.69 (p=0.101) #Features: 32		0.44 AUC:0.69 (p=0.086) #Features: 64	0.6 AUC:0.71 (p=0.121) #Features: 16
Random Forest				0.85 AUC:0.78 (p=1.0) #Features: 32		0.55 AUC:0.71 (p=0.062) #Features: 64		0.66 AUC:0.72 (p=0.113) #Features: 32
SVM		0.81 AUC:0.72 (p=0.281) #Features: 64	0.32 AUC:0.69 (p=0.08) #Features: 64		0.72 AUC:0.7 (p=0.123) #Features: 16			0.6 AUC:0.73 (p=0.302) #Features: 16
XGBoost				0.52 AUC:0.73 (p=0.493) #Features: 8		0.55 AUC:0.71 (p=0.218) #Features: 64		0.7 AUC:0.73 (p=0.416) #Features: 64
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMR

Figure S22: Stability of the models on dataset Lu2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Sasaki2019

LDA	0.79 AUC:0.62 (p=0.386) #Features: 4	0.74 AUC:0.61 (p=0.306) #Features: 2	0.35 AUC:0.61 (p=0.325) #Features: 1		0.82 AUC:0.64 (p=0.607) #Features: 2	0.15 AUC:0.54 (p=0.067) #Features: 1	0.38 AUC:0.58 (p=0.122) #Features: 4	0.49 AUC:0.64 (p=0.621) #Features: 1
Logistic Regression	0.79 AUC:0.62 (p=0.395) #Features: 4	0.74 AUC:0.61 (p=0.288) #Features: 4	0.35 AUC:0.61 (p=0.331) #Features: 1	1.0 AUC:0.57 (p=0.096) #Features: 64	0.82 AUC:0.64 (p=0.595) #Features: 2	0.24 AUC:0.54 (p=0.062) #Features: 2	0.38 AUC:0.59 (p=0.155) #Features: 4	0.49 AUC:0.64 (p=0.611) #Features: 1
Naive Bayes	0.73 AUC:0.62 (p=0.328) #Features: 32	0.74 AUC:0.61 (p=0.241) #Features: 16	0.25 AUC:0.61 (p=0.248) #Features: 64	1.0 AUC:0.55 (p=0.059) #Features: 64	0.69 AUC:0.64 (p=0.585) #Features: 16		0.38 AUC:0.6 (p=0.24) #Features: 4	0.49 AUC:0.64 (p=0.621) #Features: 1
Neural Network	0.79 AUC:0.62 (p=0.431) #Features: 4	0.74 AUC:0.65 (p=0.721) #Features: 4	0.31 AUC:0.63 (p=0.5) #Features: 8	1.0 AUC:0.65 (p=0.772) #Features: 64	0.65 AUC:0.64 (p=0.649) #Features: 4	0.31 AUC:0.66 (p=0.936) #Features: 4	0.44 AUC:0.66 (p=0.88) #Features: 64	0.49 AUC:0.66 (p=0.9) #Features: 1
RBF-SVM	0.59 AUC:0.6 (p=0.187) #Features: 8	0.74 AUC:0.58 (p=0.147) #Features: 16	0.28 AUC:0.61 (p=0.343) #Features: 2	1.0 AUC:0.63 (p=0.578) #Features: 8	0.69 AUC:0.61 (p=0.343) #Features: 16	0.31 AUC:0.62 (p=0.497) #Features: 4	0.44 AUC:0.59 (p=0.194) #Features: 64	0.49 AUC:0.59 (p=0.187) #Features: 1
Random Forest	0.59 AUC:0.63 (p=0.382) #Features: 8	0.63 AUC:0.59 (p=0.101) #Features: 8	0.31 AUC:0.6 (p=0.207) #Features: 8	1.0 AUC:0.62 (p=0.407) #Features: 64	0.69 AUC:0.64 (p=0.636) #Features: 16	0.31 AUC:0.61 (p=0.398) #Features: 4	0.41 AUC:0.57 (p=0.068) #Features: 32	0.49 AUC:0.62 (p=0.412) #Features: 1
SVM	0.79 AUC:0.63 (p=0.454) #Features: 4	0.74 AUC:0.64 (p=0.66) #Features: 16	0.35 AUC:0.59 (p=0.2) #Features: 1	0.99 AUC:0.59 (p=0.342) #Features: 32	0.82 AUC:0.64 (p=0.624) #Features: 2	0.15 AUC:0.56 (p=0.104) #Features: 1	0.3 AUC:0.56 (p=0.09) #Features: 2	0.49 AUC:0.63 (p=0.558) #Features: 1
XGBoost	0.59 AUC:0.67 (p=1.0) #Features: 8	0.74 AUC:0.63 (p=0.534) #Features: 2	0.25 AUC:0.63 (p=0.506) #Features: 4	1.0 AUC:0.65 (p=0.789) #Features: 64	0.69 AUC:0.65 (p=0.69) #Features: 16	0.31 AUC:0.63 (p=0.547) #Features: 4	0.38 AUC:0.63 (p=0.57) #Features: 4	0.49 AUC:0.63 (p=0.561) #Features: 1
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMR

Figure S23: Stability of the models on dataset Sasaki2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Toivonen2019

LDA	1.0 AUC:0.83 (p=0.525) #Features: 1		0.18 AUC:0.82 (p=0.455) #Features: 4		0.44 AUC:0.83 (p=0.411) #Features: 32	0.44 AUC:0.77 (p=0.187) #Features: 32	0.3 AUC:0.77 (p=0.175) #Features: 16	0.54 AUC:0.78 (p=0.173) #Features: 64
Logistic Regression	1.0 AUC:0.94 (p=0.619) #Features: 1	0.68 AUC:0.73 (p=0.054) #Features: 3	0.15 AUC:0.82 (p=0.455) #Features: 16		0.36 AUC:0.83 (p=0.408) #Features: 16	0.44 AUC:0.77 (p=0.188) #Features: 32	0.3 AUC:0.78 (p=0.22) #Features: 16	0.48 AUC:0.77 (p=0.114) #Features: 32
Naive Bayes	1.0 AUC:0.82 (p=0.463) #Features: 1	0.57 AUC:0.72 (p=0.052) #Features: 64	0.18 AUC:0.76 (p=0.112) #Features: 4			0.42 AUC:0.74 (p=0.059) #Features: 16	0.18 AUC:0.79 (p=0.091) #Features: 4	
Neural Network	1.0 AUC:0.83 (p=0.442) #Features: 1	0.59 AUC:0.74 (p=0.097) #Features: 32	0.15 AUC:0.83 (p=0.516) #Features: 16		0.44 AUC:0.85 (p=0.744) #Features: 32	0.44 AUC:0.81 (p=0.409) #Features: 32	0.16 AUC:0.8 (p=0.314) #Features: 2	0.54 AUC:0.82 (p=0.486) #Features: 64
RBF-SVM	0.44 AUC:0.77 (p=0.071) #Features: 8		0.18 AUC:0.79 (p=0.218) #Features: 4			0.44 AUC:0.79 (p=0.253) #Features: 32	0.3 AUC:0.72 (p=0.065) #Features: 16	0.48 AUC:0.74 (p=0.061) #Features: 32
Random Forest	1.0 AUC:0.85 (p=0.741) #Features: 1		0.15 AUC:0.78 (p=0.175) #Features: 16		0.37 AUC:0.83 (p=0.195) #Features: 8			0.34 AUC:0.74 (p=0.057) #Features: 4
SVM	1.0 AUC:0.84 (p=0.574) #Features: 1		0.15 AUC:0.83 (p=0.48) #Features: 16		0.44 AUC:0.86 (p=0.733) #Features: 32	0.44 AUC:0.78 (p=0.243) #Features: 32	0.27 AUC:0.79 (p=0.217) #Features: 8	0.48 AUC:0.8 (p=0.302) #Features: 32
XGBoost	1.0 AUC:0.83 (p=0.408) #Features: 1		0.15 AUC:0.81 (p=0.359) #Features: 16		0.37 AUC:0.87 (p=1.0) #Features: 8	0.44 AUC:0.76 (p=0.117) #Features: 32	0.3 AUC:0.76 (p=0.112) #Features: 16	0.48 AUC:0.75 (p=0.085) #Features: 32
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MRFMR

Figure S24: Stability of the models on dataset Toivonen2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Keek2020

LDA	0.33 AUC:0.65 (p=0.137) #Features: 4	0.36 AUC:0.64 (p=0.117) #Features: 1	0.33 AUC:0.65 (p=0.183) #Features: 2		0.27 AUC:0.65 (p=0.191) #Features: 1		0.3 AUC:0.68 (p=0.709) #Features: 8	0.48 AUC:0.64 (p=0.109) #Features: 8
Logistic Regression	0.61 AUC:0.66 (p=0.233) #Features: 16	0.83 AUC:0.65 (p=0.27) #Features: 64	0.28 AUC:0.65 (p=0.242) #Features: 9	0.9 AUC:0.65 (p=0.126) #Features: 64	0.27 AUC:0.65 (p=0.185) #Features: 1	0.17 AUC:0.68 (p=0.145) #Features: 2	0.3 AUC:0.68 (p=0.787) #Features: 8	0.48 AUC:0.64 (p=0.147) #Features: 8
Naive Bayes	0.7 AUC:0.65 (p=0.225) #Features: 32	0.83 AUC:0.66 (p=0.356) #Features: 64	0.31 AUC:0.66 (p=0.299) #Features: 32	0.9 AUC:0.64 (p=0.154) #Features: 64	0.64 AUC:0.65 (p=0.151) #Features: 64	0.23 AUC:0.65 (p=0.137) #Features: 4	0.3 AUC:0.67 (p=0.629) #Features: 8	
Neural Network	0.33 AUC:0.66 (p=0.338) #Features: 4	0.36 AUC:0.65 (p=0.213) #Features: 1	0.33 AUC:0.65 (p=0.134) #Features: 1		0.49 AUC:0.65 (p=0.195) #Features: 16	0.17 AUC:0.68 (p=0.56) #Features: 2	0.3 AUC:0.67 (p=0.655) #Features: 8	0.63 AUC:0.64 (p=0.21) #Features: 32
RBF-SVM	0.72 AUC:0.64 (p=0.063) #Features: 64	0.36 AUC:0.65 (p=0.213) #Features: 1	0.29 AUC:0.63 (p=0.054) #Features: 64	0.9 AUC:0.63 (p=0.053) #Features: 64	0.31 AUC:0.65 (p=0.153) #Features: 4	0.17 AUC:0.66 (p=0.107) #Features: 2	0.3 AUC:0.66 (p=0.403) #Features: 8	0.73 AUC:0.67 (p=0.459) #Features: 64
Random Forest	0.72 AUC:0.64 (p=0.18) #Features: 64	0.83 AUC:0.67 (p=0.466) #Features: 64	0.28 AUC:0.64 (p=0.168) #Features: 8		0.64 AUC:0.66 (p=0.441) #Features: 64	0.37 AUC:0.64 (p=0.074) #Features: 32	0.4 AUC:0.64 (p=0.118) #Features: 32	0.73 AUC:0.68 (p=0.725) #Features: 64
SVM	0.61 AUC:0.66 (p=0.374) #Features: 16	0.36 AUC:0.65 (p=0.143) #Features: 1	0.33 AUC:0.65 (p=0.247) #Features: 2		0.31 AUC:0.65 (p=0.221) #Features: 4	0.17 AUC:0.69 (p=1.0) #Features: 2	0.3 AUC:0.66 (p=0.715) #Features: 8	0.48 AUC:0.64 (p=0.162) #Features: 8
XGBoost	0.61 AUC:0.65 (p=0.28) #Features: 16	0.5 AUC:0.63 (p=0.132) #Features: 16	0.28 AUC:0.65 (p=0.305) #Features: 8		0.35 AUC:0.65 (p=0.236) #Features: 8	0.37 AUC:0.63 (p=0.069) #Features: 32	0.4 AUC:0.66 (p=0.445) #Features: 32	0.73 AUC:0.64 (p=0.156) #Features: 64
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMe

Figure S25: Stability of the models on dataset Keek2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Li2020

LDA	0.77 AUC:0.78 (p=0.075) #Features: 64	0.72 AUC:0.84 (p=0.197) #Features: 32	0.48 AUC:0.77 (p=0.07) #Features: 16		0.77 AUC:0.78 (p=0.095) #Features: 32	0.64 AUC:0.76 (p=0.051) #Features: 16		
Logistic Regression	0.77 AUC:0.81 (p=0.139) #Features: 64	0.72 AUC:0.89 (p=1.0) #Features: 32	0.45 AUC:0.79 (p=0.097) #Features: 32		0.77 AUC:0.8 (p=0.157) #Features: 32	0.87 AUC:0.81 (p=0.183) #Features: 64		0.6 AUC:0.78 (p=0.051) #Features: 16
Naive Bayes								
Neural Network	0.86 AUC:0.83 (p=0.361) #Features: 32	0.81 AUC:0.87 (p=0.702) #Features: 64	0.48 AUC:0.8 (p=0.211) #Features: 16		0.79 AUC:0.8 (p=0.173) #Features: 64	0.64 AUC:0.87 (p=0.69) #Features: 16		0.6 AUC:0.87 (p=0.718) #Features: 16
RBF-SVM		0.79 AUC:0.8 (p=0.152) #Features: 8	0.48 AUC:0.77 (p=0.096) #Features: 16			0.64 AUC:0.89 (p=0.887) #Features: 16		0.6 AUC:0.83 (p=0.291) #Features: 16
Random Forest	0.86 AUC:0.77 (p=0.068) #Features: 32	0.81 AUC:0.84 (p=0.34) #Features: 64	0.45 AUC:0.78 (p=0.067) #Features: 32		0.79 AUC:0.77 (p=0.053) #Features: 64	0.64 AUC:0.8 (p=0.151) #Features: 16		0.7 AUC:0.83 (p=0.298) #Features: 32
SVM	0.77 AUC:0.8 (p=0.121) #Features: 64	0.72 AUC:0.89 (p=0.729) #Features: 32	0.48 AUC:0.8 (p=0.165) #Features: 16		0.77 AUC:0.82 (p=0.251) #Features: 32	0.87 AUC:0.81 (p=0.201) #Features: 64		
XGBoost		0.81 AUC:0.82 (p=0.236) #Features: 64	0.39 AUC:0.82 (p=0.194) #Features: 64			0.52 AUC:0.79 (p=0.065) #Features: 4		
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MRFMR

Figure S26: Stability of the models on dataset Li2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Park2020

LDA	0.73 AUC:0.64 (p=0.443) #Features: 64	0.71 AUC:0.63 (p=0.22) #Features: 32	0.16 AUC:0.62 (p=0.144) #Features: 32	0.87 AUC:0.62 (p=0.113) #Features: 64	0.73 AUC:0.63 (p=0.199) #Features: 64	0.49 AUC:0.63 (p=0.364) #Features: 32	0.36 AUC:0.62 (p=0.154) #Features: 64	0.51 AUC:0.61 (p=0.051) #Features: 64
Logistic Regression	0.73 AUC:0.64 (p=0.509) #Features: 64	0.77 AUC:0.64 (p=0.617) #Features: 64	0.17 AUC:0.62 (p=0.147) #Features: 64	0.87 AUC:0.62 (p=0.125) #Features: 64	0.73 AUC:0.63 (p=0.292) #Features: 64	0.55 AUC:0.64 (p=0.631) #Features: 64	0.36 AUC:0.63 (p=0.209) #Features: 64	0.51 AUC:0.61 (p=0.062) #Features: 64
Naive Bayes	0.73 AUC:0.62 (p=0.092) #Features: 64	0.69 AUC:0.62 (p=0.083) #Features: 16		0.87 AUC:0.61 (p=0.096) #Features: 64	0.73 AUC:0.62 (p=0.137) #Features: 64		0.3 AUC:0.61 (p=0.068) #Features: 32	
Neural Network	0.73 AUC:0.65 (p=1.0) #Features: 64	0.67 AUC:0.62 (p=0.143) #Features: 4	0.17 AUC:0.63 (p=0.414) #Features: 64	0.76 AUC:0.61 (p=0.091) #Features: 32	0.73 AUC:0.63 (p=0.385) #Features: 64	0.55 AUC:0.64 (p=0.741) #Features: 64	0.36 AUC:0.63 (p=0.483) #Features: 64	
RBF-SVM	0.73 AUC:0.62 (p=0.229) #Features: 64			0.87 AUC:0.61 (p=0.073) #Features: 64	0.73 AUC:0.61 (p=0.079) #Features: 64	0.49 AUC:0.61 (p=0.142) #Features: 32		
Random Forest	0.73 AUC:0.61 (p=0.079) #Features: 64	0.77 AUC:0.61 (p=0.079) #Features: 64				0.49 AUC:0.61 (p=0.063) #Features: 32	0.36 AUC:0.61 (p=0.061) #Features: 64	
SVM	0.73 AUC:0.64 (p=0.374) #Features: 64	0.77 AUC:0.63 (p=0.24) #Features: 64	0.16 AUC:0.61 (p=0.068) #Features: 32	0.87 AUC:0.61 (p=0.088) #Features: 64	0.73 AUC:0.64 (p=0.325) #Features: 64	0.49 AUC:0.63 (p=0.351) #Features: 32	0.36 AUC:0.62 (p=0.263) #Features: 64	
XGBoost	0.73 AUC:0.62 (p=0.232) #Features: 64	0.77 AUC:0.62 (p=0.217) #Features: 64		0.87 AUC:0.62 (p=0.227) #Features: 64	0.73 AUC:0.62 (p=0.191) #Features: 64	0.55 AUC:0.63 (p=0.296) #Features: 64	0.36 AUC:0.6 (p=0.051) #Features: 64	
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MRFMR

Figure S27: Stability of the models on dataset Park2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Song2020

LDA								
Logistic Regression	0.92 AUC:0.96 (p=0.109) #Features: 64		0.75 AUC:0.96 (p=0.112) #Features: 64					
Naive Bayes								
Neural Network	0.92 AUC:0.96 (p=0.058) #Features: 64		0.75 AUC:0.96 (p=0.115) #Features: 64			0.88 AUC:0.96 (p=0.224) #Features: 64	0.8 AUC:0.98 (p=1.0) #Features: 32	
RBF-SVM								
Random Forest								
SVM	0.92 AUC:0.97 (p=0.178) #Features: 64		0.75 AUC:0.96 (p=0.104) #Features: 64				0.78 AUC:0.96 (p=0.054) #Features: 64	
XGBoost				0.92 AUC:0.96 (p=0.114) #Features: 32				
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMR

Figure S28: Stability of the models on dataset Song2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

Stability on Veeraraghavan2020

LDA			0.48 AUC:0.57 (p=0.302) #Features: 8			0.15 AUC:0.57 (p=0.293) #Features: 1		
Logistic Regression			0.48 AUC:0.58 (p=0.409) #Features: 8	0.57 AUC:0.5 (p=0.057) #Features: 64		0.15 AUC:0.57 (p=0.268) #Features: 1		
Naive Bayes	0.79 AUC:0.55 (p=0.247) #Features: 64		0.48 AUC:0.56 (p=0.273) #Features: 8	0.57 AUC:0.53 (p=0.14) #Features: 64	0.47 AUC:0.53 (p=0.159) #Features: 8	0.19 AUC:0.58 (p=0.358) #Features: 2	0.37 AUC:0.51 (p=0.071) #Features: 16	
Neural Network	0.41 AUC:0.55 (p=0.188) #Features: 16	0.31 AUC:0.57 (p=0.288) #Features: 4	0.34 AUC:0.53 (p=0.083) #Features: 64	0.83 AUC:0.63 (p=0.914) #Features: 32	0.57 AUC:0.58 (p=0.329) #Features: 64	0.62 AUC:0.61 (p=0.581) #Features: 2	0.24 AUC:0.64 (p=0.957) #Features: 4	0.31 AUC:0.58 (p=0.376) #Features: 8
RBF-SVM	0.02 AUC:0.54 (p=0.097) #Features: 1	0.31 AUC:0.57 (p=0.324) #Features: 4	0.29 AUC:0.52 (p=0.08) #Features: 32	0.91 AUC:0.63 (p=0.911) #Features: 64	0.08 AUC:0.57 (p=0.309) #Features: 1	0.62 AUC:0.63 (p=0.896) #Features: 2	0.4 AUC:0.55 (p=0.134) #Features: 8	0.31 AUC:0.55 (p=0.182) #Features: 8
Random Forest		0.3 AUC:0.58 (p=0.382) #Features: 2		0.49 AUC:0.59 (p=0.44) #Features: 1		0.62 AUC:0.62 (p=0.495) #Features: 2	0.19 AUC:0.54 (p=0.131) #Features: 2	
SVM	0.5 AUC:0.5 (p=0.059) #Features: 32	0.79 AUC:0.57 (p=0.286) #Features: 64		0.91 AUC:0.57 (p=0.322) #Features: 64	0.57 AUC:0.56 (p=0.215) #Features: 64		0.15 AUC:0.58 (p=0.384) #Features: 1	0.73 AUC:0.54 (p=0.163) #Features: 64
XGBoost	0.49 AUC:0.51 (p=0.058) #Features: 64	0.59 AUC:0.56 (p=0.317) #Features: 16		0.67 AUC:0.62 (p=0.706) #Features: 16		0.62 AUC:0.64 (p=1.0) #Features: 2	0.4 AUC:0.6 (p=0.588) #Features: 8	0.11 AUC:0.54 (p=0.205) #Features: 1
	Anova	Bhattacharyya	ET	FCBF	Kendall	LASSO	MM	MFRMe

Figure S29: Stability of the models on dataset Veeraraghavan2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

