

Figure S1. Prediction of the forest for different values of N% and AMW.

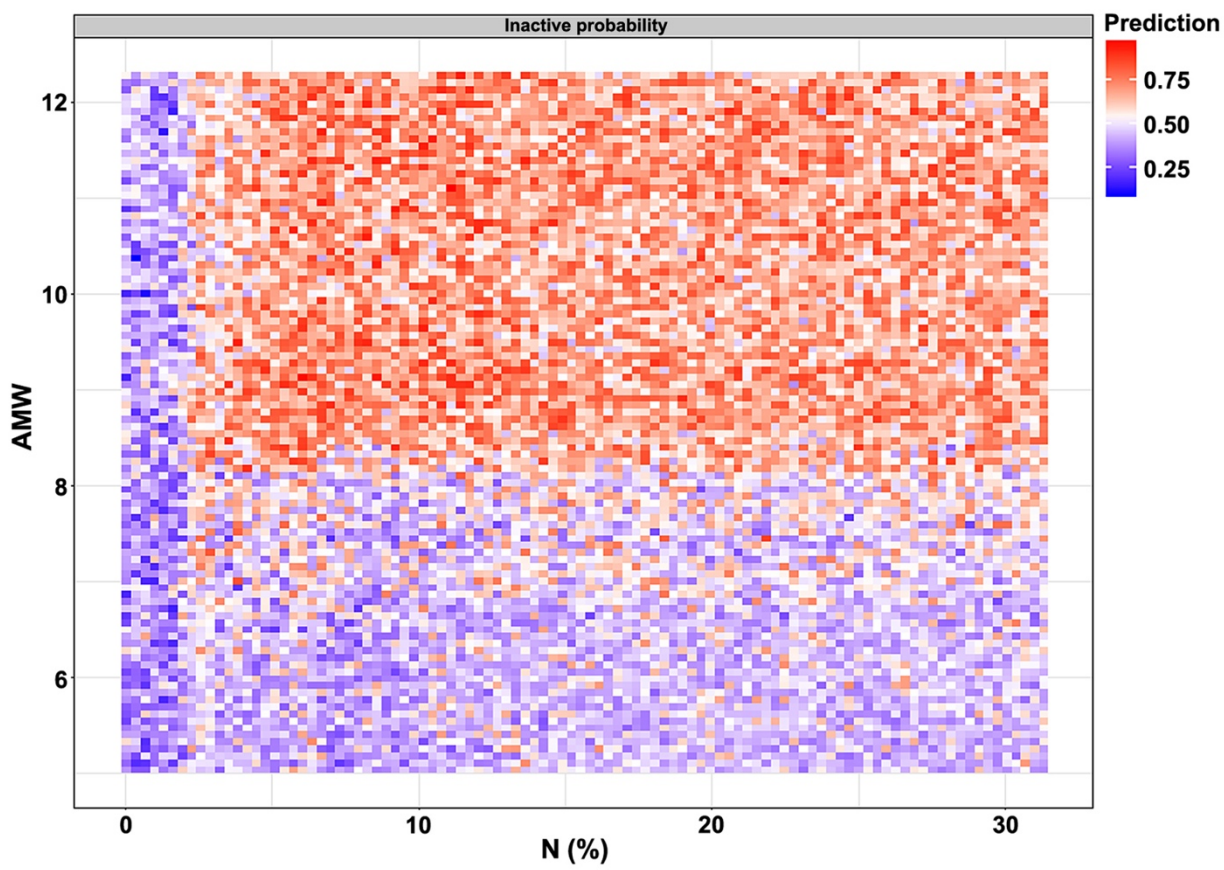


Figure S2. Prediction of the forest for different values of N% and nTA.

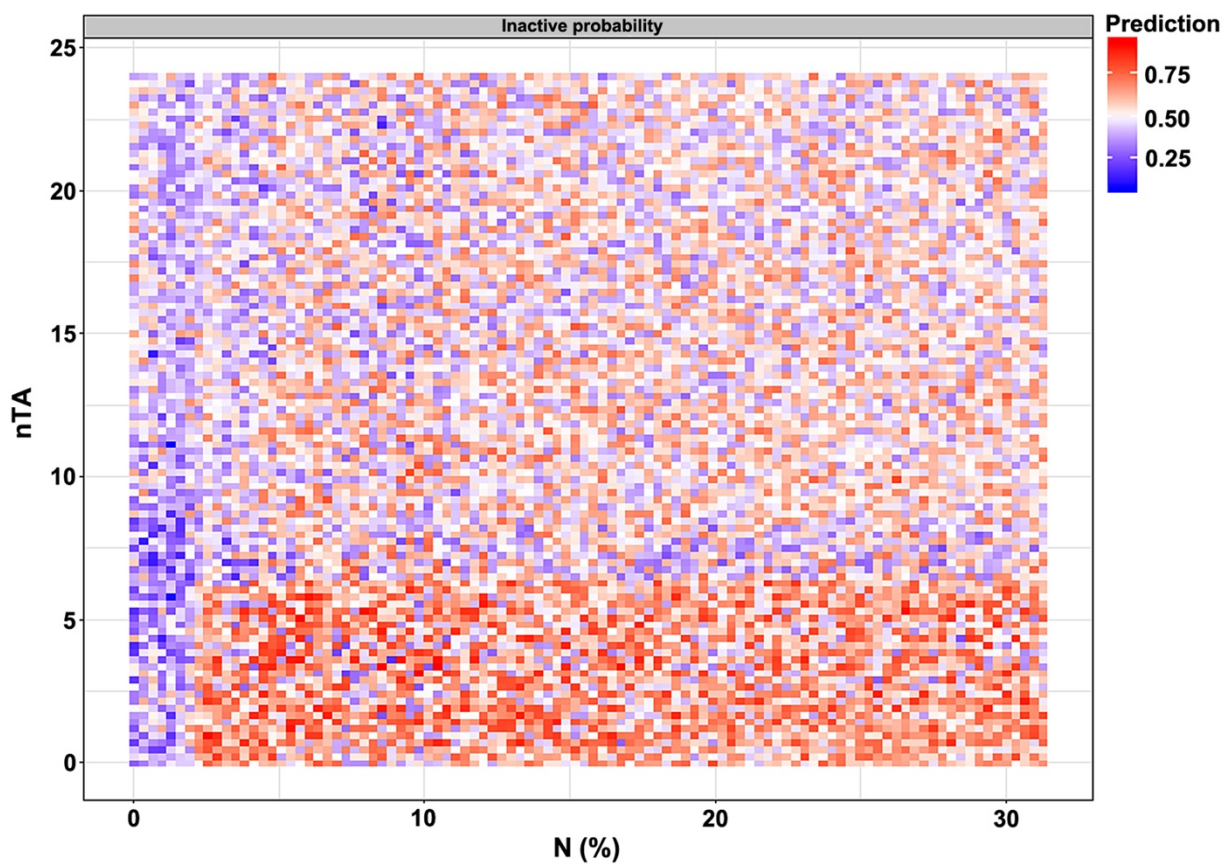


Figure S3. Prediction of the forest for different values of MCD and ARR.

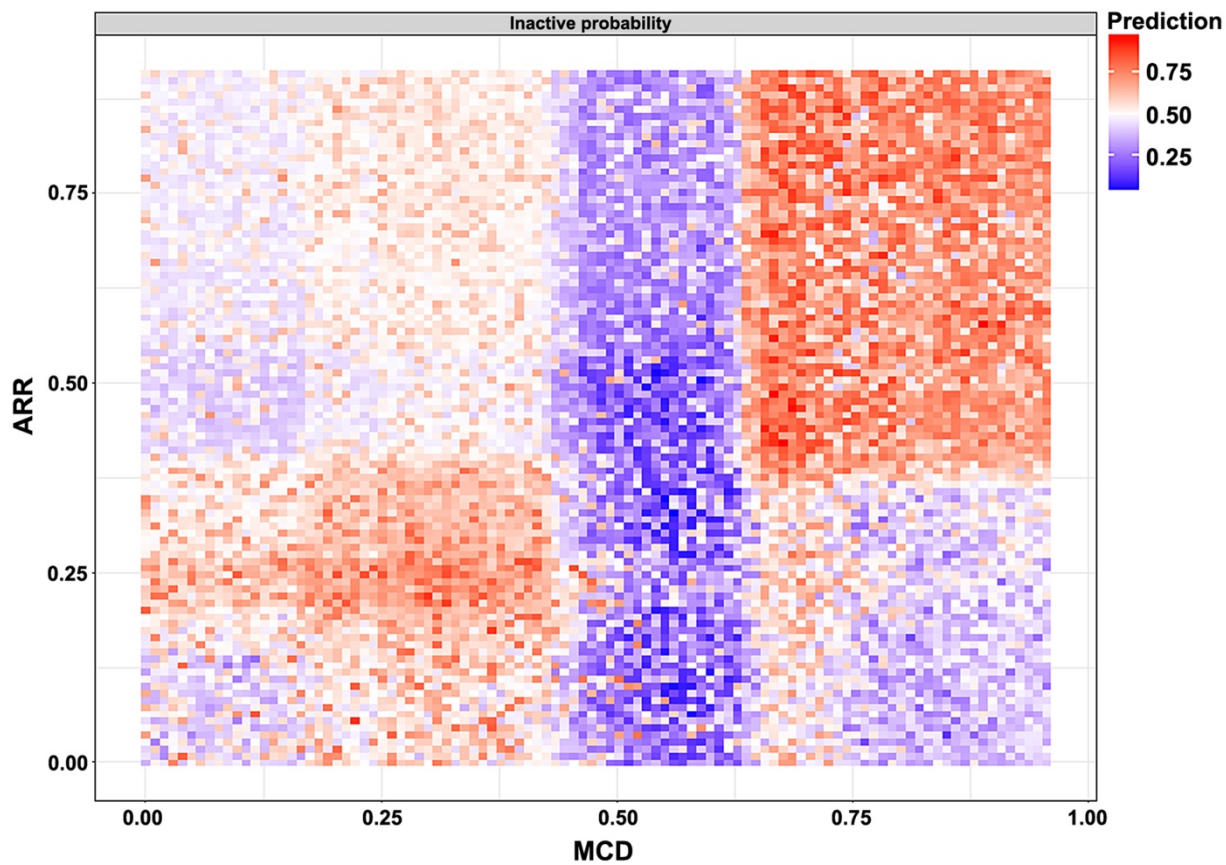


Table S1. Performance of selected classification models with positive predictive value (PPV) higher than 75% for the 10-fold nested cross-validation.

	Model	Specificity	Sensitivity	PPV ^d	Balanced accuracy	MMCE
1.	Topological descriptors – RF* (1) ^a	0.9374	0.3583	0.8424	0.6479	0.3022
2.	Topological descriptors – RF (2)	0.9298	0.3628	0.7964	0.6463	0.3105
3.	Topological descriptors – RF (1) over ^b	0.9148	0.5752	0.8749	0.745	0.2548
4.	Topological descriptors – RF (1) smote ^c	0.8946	0.499	0.8158	0.6968	0.3086
5.	Walk and path – RF (1)	0.9465	0.285	0.7587	0.6158	0.3231
6.	Information indices – RF (1)	0.9486	0.3434	0.8368	0.646	0.3003
7.	Information indices – RF (2)	0.9685	0.3448	0.8848	0.6566	0.2878
8.	Information indices – RF (1) over	0.9022	0.634	0.8715	0.7681	0.2319
9.	Information indices – RF (1) smote	0.9023	0.5438	0.851	0.723	0.2776
10.	Information indices – BST** (1) smote	0.78	0.7536	0.7803	0.7668	0.2344
11.	2D-autocorrelation - RF (1)	0.927	0.3414	0.776	0.6342	0.3063
12.	2D-autocorrelation - RF (2)	0.9687	0.3005	0.8707	0.6346	0.3063
13.	2D-autocorrelation - RF (2) over	0.9453	0.611	0.9201	0.7782	0.2289
14.	2D-autocorrelation - RF (2) smote	0.9174	0.4858	0.8583	0.7016	0.2993
15.	Burden eigenvalues – RF (2)	0.941	0.3373	0.7943	0.6391	0.3063
16.	Burden eigenvalues – RF (2) over	0.8803	0.6373	0.8417	0.7588	0.2427
17.	Burden eigenvalues – RF (2) smote	0.8445	0.6265	0.8057	0.7355	0.2641
18.	P-VSA-like – RF (1)	0.9327	0.3528	0.7825	0.6428	0.3058
19.	P-VSA-like – RF (2)	0.9332	0.3716	0.7996	0.6524	0.2967
20.	P-VSA-like – RF (2) over	0.9149	0.6159	0.8891	0.7654	0.2369
21.	P-VSA-like – RF (2) smote	0.8919	0.5541	0.8273	0.723	0.283
22.	Eta indices – RF (2)	0.9384	0.3807	0.8394	0.6596	0.2872
23.	Edge adjacency – RF (1)	0.9412	0.3453	0.8242	0.6432	0.307
24.	Edge adjacency – RF (2)	0.9301	0.3652	0.8006	0.6477	0.3038
25.	Edge adjacency – RF (1) over	0.9031	0.6477	0.8635	0.7754	0.2239
26.	Edge adjacency – SVM*** (2) over	0.7663	0.7113	0.7519	0.7388	0.2696
27.	Global – BST (1) over	0.793	0.8137	0.7899	0.8034	0.1994
28.	Global – BST (1) smote	0.7974	0.7957	0.7927	0.7966	0.202

*RF – random forest classifier; **BST – gradient boosting classifier; *** SVM – support vector machines; ^a numbers in brackets indicate the subset of features selected by the different feature selection algorithms (1 – random forest importance and information gain; 2 – symmetrical uncertainty); ^bover - training set balanced through oversampling; ^csmote – training set balanced through the smote technique (*synthetic minority oversampling technique*). ^dPPV – positive predictive value. The first in the name of each model indicates the block of descriptors used for its building.

Table S2. Performance of all classification models for the external validation set (different descriptor blocks, sampling techniques, and classification algorithms).

Model	Specificity	Sensitivity	PPV	Balanced accuracy	MMCE
Constitutional indices - RF	0.9516	0.4546	0.8696	0.7031	0.2547
Constitutional indices – RF over	0.7903	0.500	0.6286	0.6452	0.3302
Constitutional indices – RF smote	0.8387	0.500	0.6875	0.6694	0.3019
Constitutional indices - BST	0.7419	0.5227	0.5897	0.6323	0.3491
Constitutional indices – BST over	0.7742	0.5227	0.6216	0.6485	0.3302
Constitutional indices – BST smote	0.6774	0.6364	0.5833	0.6569	0.3396
Constitutional indices - SVM	0.8065	0.6136	0.6923	0.7100	0.2736
Constitutional indices – SVM over	0.6613	0.7045	0.5962	0.6829	0.3208
Constitutional indices – SVM over	0.7742	0.6591	0.6744	0.7166	0.2736
Constitutional indices – knn	0.8065	0.5682	0.6757	0.6873	0.2925
Constitutional indices – knn over	0.6935	0.7045	0.6200	0.6990	0.3019
Constitutional indices – knn smote	0.6613	0.6364	0.5714	0.6488	0.3491
Ring descriptors – RF(1)	0.8871	0.4318	0.7308	0.6595	0.3019
Ring descriptors – RF(2)	0.8548	0.3636	0.6400	0.6092	0.3491
Ring descriptors – RF(1) over	0.8871	0.4545	0.7407	0.6708	0.2925
Ring descriptors – RF(1) smote	0.8226	0.4773	0.6562	0.6499	0.3208
Ring descriptors – BST (1)	0.7581	0.5455	0.6154	0.6518	0.3302
Ring descriptors – BST (2)	0.7903	0.6136	0.6750	0.7020	0.2830
Ring descriptors – BST (1) over	0.7903	0.5682	0.6579	0.6793	0.3019
Ring descriptors – BST (1) smote	0.7258	0.6591	0.6304	0.6924	0.3019
Ring descriptors – SVM (1)	0.8065	0.3864	0.5862	0.5964	0.3679
Ring descriptors – SVM (2)	0.7903	0.6136	0.6750	0.7020	0.2830
Ring descriptors – SVM (1) over	0.6290	0.7045	0.5741	0.6668	0.3396
Ring descriptors – SVM (1) smote	0.6935	0.6136	0.5870	0.6536	0.3396
Ring descriptors – SVM (2) smote	0.6290	0.7500	0.5893	0.6895	0.3208
Ring descriptors – knn (1)	0.7581	0.5000	0.5946	0.6290	0.3491
Ring descriptors – knn (2)	0.7742	0.5227	0.6216	0.6485	0.3302
Ring descriptors – knn (1) over	0.6290	0.7500	0.5893	0.6895	0.3208
Ring descriptors – knn (1) smote	0.7581	0.5000	0.5946	0.6290	0.3491
Topological descriptors – RF (1)	0.9194	0.5000	0.8148	0.7097	0.2547
Topological descriptors – RF (2)	0.9194	0.5227	0.8214	0.7210	0.2453
Topological descriptors – RF (1) over	0.9355	0.5682	0.8621	0.7518	0.2170
Topological descriptors – RF (1) smote	0.9516	0.5909	0.8966	0.7713	0.1981
Topological descriptors – BST (1)	0.8871	0.7045	0.8158	0.7958	0.1887
Topological descriptors – BST (2)	0.8065	0.6364	0.7000	0.7214	0.2642
Topological descriptors – BST (1) (over)	0.7903	0.7955	0.7292	0.7929	0.2075
Topological descriptors – BST (1) (smote)	0.9194	0.6818	0.8571	0.8006	0.1793
Topological descriptors – SVM (1)	0.8871	0.6818	0.8108	0.7845	0.1981
Topological descriptors – SVM (2)	0.8710	0.6364	0.7778	0.7537	0.2264
Topological descriptors – SVM (1) over	0.8065	0.7500	0.7333	0.7782	0.2170
Topological descriptors – SVM (1) smote	0.7742	0.7500	0.7021	0.7621	0.2358
Topological descriptors – knn (1)	0.8548	0.6591	0.7632	0.7570	0.2264
Topological descriptors – knn (2)	0.8548	0.5455	0.7273	0.7001	0.2736
Topological descriptors – knn (1) over	0.8548	0.6591	0.7632	0.7570	0.2264

Model	Specificity	Sensitivity	PPV	Balanced accuracy	MMCE
Topological descriptors – knn (1) smote	0.8387	0.6591	0.7436	0.7489	0.2358
Walk and path – RF (1)	0.9516	0.2727	0.8000	0.6122	0.3302
Walk and path – RF (1) over	0.6774	0.6818	0.6000	0.6796	0.3208
Walk and path – RF (1) smote	0.6613	0.6818	0.5882	0.6716	0.3302
Walk and path – BST (1)	0.7742	0.5455	0.6316	0.6598	0.3208
Walk and path – BST (1) over	0.7097	0.6591	0.6170	0.6844	0.3113
Walk and path – BST (1) smote	0.8548	0.5455	0.7273	0.7001	0.2736
Walk and path – SVM (1)	0.9194	0.3409	0.7500	0.6301	0.3208
Walk and path – SVM (1) over	0.6774	0.7500	0.6226	0.7137	0.2925
Walk and path – SVM (1) smote	0.6774	0.7500	0.6226	0.7137	0.2925
Walk and path – knn (1)	0.7742	0.5227	0.5610	0.6162	0.3679
Walk and path – knn (1) over	0.7097	0.5227	0.5610	0.6162	0.3679
Walk and path – knn (1) smote	0.7581	0.6591	0.6591	0.7086	0.2830
Information indices – RF (1)	1.0000	0.5000	1.0000	0.7500	0.2075
Information indices – RF (2)	0.9839	0.5227	0.9583	0.7533	0.2076
Information indices – RF (1) over	1.0000	0.5227	1.0000	0.7614	0.1981
Information indices – RF (1) smote	1.0000	0.5682	1.0000	0.7841	0.1792
Information indices – BST (1)	0.9355	0.6818	0.8824	0.8086	0.1698
Information indices – BST (2)	0.9677	0.6136	0.9310	0.7907	0.1792
Information indices – BST (1) over	0.9355	0.7954	0.8974	0.8655	0.1226
Information indices – BST (1) smote	0.9355	0.7500	0.8919	0.8427	0.1415
Information indices – SVM (1)	0.9516	0.3864	0.8500	0.6690	0.2830
Information indices – SVM (2)	0.9194	0.2954	0.7222	0.6074	0.3396
Information indices – SVM (1) over	0.7742	0.5455	0.6316	0.6598	0.3208
Information indices – SVM (1) smote	0.6452	0.6364	0.5600	0.6408	0.3585
Information indices – knn (1)	0.9023	0.3864	0.7391	0.6448	0.3113
Information indices – knn (2)	0.9516	0.7273	0.9143	0.8394	0.1415
Information indices – knn (1) over	0.8226	0.7955	0.7609	0.8090	0.1887
Information indices – knn (1) smote	0.9194	0.7273	0.8649	0.8233	0.1604
2D-matrix – RF (1)	0.9032	0.4091	0.7500	0.6562	0.3019
2D-matrix – RF (2)	0.8871	0.4545	0.7407	0.6708	0.2925
2D-matrix – RF (1) over	0.8871	0.4773	0.7500	0.6822	0.2830
2D-matrix RF (1) smote	0.8548	0.5000	0.7097	0.6774	0.2925
2D-matrix BST (1)	0.6774	0.5682	0.5556	0.6228	0.3679
2D-matrix BST (2)	0.7097	0.6591	0.6170	0.6844	0.3113
2D-matrix BST (2) over	0.7258	0.7045	0.6458	0.7152	0.2830
2D-matrix BST (1) smote	0.7097	0.7273	0.6400	0.7185	0.2830
2D-matrix SVM (1)	0.8226	0.6591	0.7250	0.7408	0.2453
2D-matrix SVM (2)	0.8226	0.7045	0.7381	0.7636	0.2264
2D-matrix SVM (2) over	0.7903	0.7273	0.7111	0.7588	0.2358
2D-matrix SVM (2) smote	0.7419	0.7727	0.6800	0.7573	0.2453
2D-matrix knn (1)	0.8548	0.5455	0.7273	0.7001	0.2736
2D-matrix knn (2)	0.8065	0.5455	0.6667	0.6760	0.3019
2D-matrix knn (1) over	0.6452	0.7955	0.6140	0.7203	0.2925
2D-matrix knn (2) smote	0.7903	0.7045	0.7045	0.7474	0.2453
2D-autocorrelation - RF (1)	0.9355	0.3864	0.8095	0.6609	0.2924
2D-autocorrelation - RF (2)	0.9677	0.4091	0.9000	0.6884	0.2642
2D-autocorrelation - RF (2) over	0.9032	0.5000	0.7857	0.7016	0.2642
2D-autocorrelation - RF (2) smote	0.9194	0.4773	0.8077	0.6983	0.2642
2D-autocorrelation - BST (1)	0.7097	0.6364	0.6087	0.6730	0.3208

Model	Specificity	Sensitivity	PPV	Balanced accuracy	MMCE
2D-autocorrelation - BST (2)	0.6613	0.6136	0.5625	0.6375	0.3585
2D-autocorrelation - BST (2) over	0.6613	0.6364	0.5714	0.6488	0.3491
2D-autocorrelation - BST (2) smote	0.5968	0.6591	0.5370	0.6279	0.3774
2D-autocorrelation - SVM (1)	0.6935	0.6591	0.6042	0.6763	0.3208
2D-autocorrelation - SVM (2)	0.8226	0.6364	0.7179	0.7295	0.2547
2D-autocorrelation - SVM (2) over	0.6935	0.7045	0.6200	0.6990	0.3019
2D-autocorrelation - SVM (2) smote	0.7258	0.6818	0.6383	0.7038	0.2925
2D-autocorrelation - knn (1)	0.7903	0.6591	0.6905	0.7247	0.2642
2D-autocorrelation - knn (2)	0.8226	0.7727	0.7556	0.7977	0.1981
2D-autocorrelation - knn (2) over	0.7419	0.8182	0.6923	0.7801	0.2264
2D-autocorrelation - knn (2) smote	0.7903	0.7500	0.7174	0.7702	0.2264
Burden eigenvalues – RF (1)	0.9355	0.5227	0.8518	0.7291	0.2358
Burden eigenvalues – RF (2)	0.9516	0.4773	0.8750	0.7144	0.2453
Burden eigenvalues – RF (2) over	0.9516	0.5909	0.8966	0.7713	0.1981
Burden eigenvalues – RF (2) smote	0.9355	0.5682	0.8621	0.7518	0.2170
Burden eigenvalues – bst (1)	0.7581	0.6818	0.6667	0.7199	0.2736
Burden eigenvalues – bst (2)	0.8226	0.7045	0.7381	0.7636	0.2264
Burden eigenvalues – bst (2) over	0.7742	0.7045	0.6889	0.7394	0.2547
Burden eigenvalues – bst (2) smote	0.8226	0.7273	0.7442	0.7749	0.2170
Burden eigenvalues – SVM (1)	0.7097	0.6591	0.6170	0.6844	0.2830
Burden eigenvalues – SVM (2)	0.7581	0.6591	0.6591	0.7086	0.2830
Burden eigenvalues – SVM (2) over	0.6290	0.7273	0.5818	0.6782	0.3302
Burden eigenvalues – SVM (2) smote	0.6290	0.7273	0.5818	0.6782	0.3302
Burden eigenvalues – knn (1)	0.8226	0.5909	0.7027	0.7067	0.2736
Burden eigenvalues – knn (2)	0.8548	0.6364	0.7568	0.7456	0.2358
Burden eigenvalues – knn (1) over	0.7419	0.7045	0.6596	0.7232	0.2736
Burden eigenvalues – knn (1) smote	0.7097	0.6818	0.6250	0.6957	0.3019
P-VSA-like – RF (1)	0.9783	0.6562	0.9545	0.8173	0.1538
P-VSA-like – RF (2)	0.9783	0.6875	0.9565	0.8329	0.1410
P-VSA-like – RF (2) over	0.9783	0.7812	0.9615	0.8798	0.1026
P-VSA-like – RF (2) smote	0.9783	0.9062	0.9667	0.9423	0.0513
P-VSA-like – BST (1)	0.7581	0.6591	0.6591	0.7086	0.2830
P-VSA-like – BST (2)	0.7419	0.5682	0.6098	0.6551	0.3302
P-VSA-like – BST (2) over	0.7742	0.6136	0.6585	0.6939	0.2925
P-VSA-like – BST (2) smote	0.6935	0.6591	0.6042	0.6763	0.3208
P-VSA-like – SVM (1)	0.8871	0.5909	0.7879	0.7390	0.2358
P-VSA-like – SVM (2)	0.7903	0.6818	0.6977	0.7361	0.2547
P-VSA-like – SVM (1) over	0.7097	0.7273	0.6400	0.7185	0.2830
P-VSA-like – SVM (1) smote	0.7581	0.7273	0.6809	0.7427	0.2547
P-VSA-like – knn (1)	0.8387	0.6136	0.7297	0.7262	0.2547
P-VSA-like – knn (2)	0.8710	0.6364	0.7778	0.7537	0.2264
P-VSA-like – knn (1) over	0.7903	0.7045	0.7045	0.7474	0.2453
P-VSA-like – knn (1) smote	0.8226	0.6591	0.7250	0.7408	0.2453
Eta indices – RF (1)	0.8710	0.4091	0.6923	0.6400	0.3208
Eta indices – RF (2)	0.9032	0.4318	0.7600	0.6675	0.2924
Eta indices – RF (2) over	0.8710	0.4318	0.7037	0.6514	0.3113
Eta indices – RF (2) smote	0.8710	0.5455	0.7500	0.7082	0.2642
Eta indices – BST (1)	0.7419	0.6818	0.6522	0.7119	0.2830
Eta indices – BST (2)	0.6774	0.5455	0.5455	0.6114	0.3774
Eta indices – BST (2) over	0.6290	0.7045	0.5741	0.6668	0.3396
Eta indices – BST (2) smote	0.6935	0.6818	0.6122	0.6877	0.3113

Model	Specificity	Sensitivity	PPV	Balanced accuracy	MMCE
Eta indices – SVM (1)	0.8065	0.6364	0.7000	0.7214	0.2642
Eta indices – SVM (2)	0.8387	0.6136	0.7297	0.7262	0.2547
Eta indices – SVM (1) over	0.7258	0.7273	0.6531	0.7265	0.2736
Eta indices – SVM (1) smote	0.7742	0.7273	0.6957	0.7507	0.2453
Eta indices – knn (1)	0.7742	0.5000	0.6111	0.6371	0.3396
Eta indices – knn (2)	0.7903	0.5909	0.6667	0.6906	0.2925
Eta indices – knn (1) over	0.7097	0.5682	0.5814	0.6389	0.3491
Eta indices – knn (1) smote	0.7097	0.6136	0.6000	0.6617	0.3302
Edge adjacency – RF (1)	0.9839	0.4545	0.9524	0.7192	0.2358
Edge adjacency – RF (2)	0.9839	0.3864	0.9444	0.6851	0.2642
Edge adjacency – RF (1) over	0.9516	0.4545	0.8696	0.7031	0.2547
Edge adjacency – RF (1) smote	0.8710	0.4091	0.6923	0.6400	0.3208
Edge adjacency – BST (1)	0.7097	0.7955	0.6604	0.7526	0.2547
Edge adjacency – BST (2)	0.9023	0.6364	0.8235	0.7698	0.2076
Edge adjacency – BST (1) over	0.7419	0.7727	0.6800	0.7573	0.2453
Edge adjacency – BST (1) smote	0.6774	0.7500	0.6226	0.7137	0.2925
Edge adjacency – SVM (1)	0.9194	0.5682	0.8333	0.7438	0.2264
Edge adjacency – SVM (2)	0.9355	0.5454	0.8571	0.7405	0.2264
Edge adjacency – SVM (1) over	0.9023	0.6364	0.8235	0.7698	0.2076
Edge adjacency – SVM (1) smote	0.8548	0.6364	0.7568	0.7456	0.2358
Edge adjacency – knn (1)	0.8387	0.6136	0.7297	0.7262	0.2547
Edge adjacency – knn (2)	0.8710	0.5909	0.7647	0.7309	0.2453
Edge adjacency – knn (1) over	0.8226	0.6136	0.7105	0.7181	0.2642
Edge adjacency – knn (1) smote	0.8065	0.6591	0.7073	0.7328	0.2547
Global – RF (1)	0.9032	0.3864	0.7391	0.6448	0.3113
Global – RF (2)	0.8548	0.5000	0.7097	0.6774	0.2925
Global – RF (1) over	0.8387	0.4318	0.6552	0.6353	0.3302
Global – RF (1) smote	0.8226	0.4773	0.6562	0.6499	0.3208
Global – BST (1)	0.9194	0.8636	0.8837	0.8915	0.1038
Global – BST (2)	0.8387	0.4773	0.6774	0.6580	0.3113
Global – BST (1) over	0.8871	0.9318	0.8542	0.9095	0.0943
Global – BST (1) smote	0.9032	0.9318	0.8723	0.9175	0.0849
Global – SVM (1)	0.9355	0.7273	0.8889	0.8314	0.1509
Global – SVM (2)	0.8387	0.6591	0.7436	0.7489	0.2358
Global – SVM (1) over	0.8871	0.7727	0.8293	0.8299	0.1604
Global – SVM (1) smote	0.8710	0.7500	0.8049	0.8105	0.1792
Global – knn (1)	0.9516	0.7727	0.9189	0.8622	0.1226
Global – knn (2)	0.8548	0.5682	0.7353	0.7115	0.2642
Global – knn (1) over	0.9032	0.9773	0.8775	0.9402	0.0660
Global – knn (1) smote	0.8871	0.7727	0.8293	0.8299	0.1604