
Supplementary Materials

In this section, the ICD-9 codes for the clinical attributes and outcomes are listed in Tables (A1). The results for Angina, Heart Failure, and Stroke outcomes are shown in Tables (A2-4). In Table (A5), the values of the “ROC Area” computed for all four outcomes (when applying SVM with both linear and RBF kernel functions) demonstrate the significance of the benefit obtained by the RBF kernel function (highlighted in bold) to the linear kernel function.

Table A1. ICD-9 Codes used to extract the records for the clinical attributes and outcomes

Attributes or Outcome	ICD-9 Code	Attributes or Outcome	ICD-9 Code
Angina	413	Heart Failure	428
Bariatric Surgery	43.89,44.38,44.39,44.68,44.95	Hypertension	401
Blindness	264.5,368.6,368.61,368.69,377.75,V19.0	Myocardial infarction	410
P Chronic obstructive pulmonary disease	490-492,494,496	Obesity	278.01,V85.35 - V85.4
Coronary Artery Disease	414	Sleep Apnea	780.5
Death	798	Stroke	43X.X0,43X.X1
Diabetes Manifestation	250.4X,250.5X,250.6X,250.7X	Tobacco Abuse	305.1
Diabetes T2DM	250.X0,250.X2	Transient ischemic Attack	435
Dyslipidemia	272		

Table A2. Results for Heart Failure outcome. U: Undersampling, O: Oversampling.

Classifier	Feature Selection	Acc		G-mean		Precision		Sensitivity		Specificity		F-measure		ROC area	
		U	O	U	O	U	O	U	O	U	O	U	O	U	O
Naive Bayes	Chi-Squared	0.72	0.73	0.72	0.73	0.72	0.74	0.72	0.71	0.72	0.75	0.72	0.73	0.79	0.80
	Information Gain	0.72	0.73	0.72	0.73	0.72	0.74	0.72	0.71	0.72	0.75	0.72	0.73	0.79	0.80
	CFS	0.71	0.73	0.70	0.73	0.69	0.72	0.76	0.77	0.65	0.70	0.72	0.74	0.78	0.81
RBFN	Chi-Squared	0.71	0.63	0.70	0.56	0.77	0.83	0.60	0.34	0.82	0.93	0.67	0.48	0.76	0.70
	Information Gain	0.71	0.69	0.70	0.68	0.77	0.78	0.60	0.55	0.82	0.84	0.67	0.64	0.76	0.75
	CFS	0.71	0.67	0.70	0.66	0.75	0.64	0.63	0.80	0.79	0.54	0.68	0.71	0.77	0.73
5NN	Chi-Squared	0.68	0.88	0.68	0.88	0.68	0.87	0.72	0.90	0.65	0.86	0.70	0.88	0.76	0.95
	Information Gain	0.68	0.88	0.68	0.88	0.68	0.87	0.72	0.90	0.65	0.86	0.70	0.88	0.76	0.95
	CFS	0.71	0.87	0.71	0.87	0.73	0.84	0.67	0.92	0.75	0.82	0.70	0.87	0.76	0.95
C4.5 (J48)	Chi-Squared	0.70	0.93	0.70	0.93	0.71	0.94	0.70	0.92	0.71	0.94	0.70	0.93	0.75	0.97
	Information Gain	0.70	0.93	0.70	0.93	0.71	0.94	0.70	0.92	0.71	0.94	0.70	0.93	0.75	0.97
	CFS	0.71	0.91	0.71	0.91	0.75	0.92	0.64	0.91	0.78	0.92	0.69	0.91	0.75	0.96
SVM	Chi-Squared	0.71	0.88	0.71	0.88	0.74	0.91	0.65	0.88	0.77	0.91	0.69	0.87	0.71	0.88
	Information Gain	0.71	0.88	0.71	0.88	0.74	0.91	0.65	0.84	0.77	0.92	0.69	0.88	0.71	0.88
	CFS	0.72	0.87	0.72	0.87	0.75	0.89	0.67	0.86	0.77	0.89	0.70	0.87	0.72	0.88
LR	Chi-Squared	0.72	0.75	0.72	0.75	0.71	0.75	0.74	0.75	0.70	0.75	0.72	0.75	0.80	0.83
	Information Gain	0.72	0.75	0.72	0.75	0.71	0.75	0.74	0.75	0.70	0.75	0.72	0.75	0.80	0.83
	CFS	0.70	0.74	0.70	0.74	0.69	0.74	0.76	0.75	0.65	0.73	0.72	0.75	0.78	0.82
Random Forest	Chi-Squared	0.66	0.95	0.66	0.95	0.66	0.94	0.67	0.96	0.66	0.94	0.66	0.95	0.74	0.98
	Information Gain	0.66	0.95	0.66	0.95	0.66	0.94	0.67	0.96	0.66	0.94	0.66	0.95	0.74	0.98
	CFS	0.71	0.94	0.71	0.94	0.73	0.93	0.67	0.96	0.76	0.92	0.70	0.94	0.76	0.98
AdaBoostM1	Chi-Squared	0.70	0.79	0.70	0.79	0.70	0.81	0.71	0.76	0.70	0.82	0.71	0.78	0.79	0.88
	Information Gain	0.70	0.79	0.70	0.79	0.70	0.81	0.71	0.76	0.70	0.82	0.71	0.78	0.79	0.88
	CFS	0.69	0.79	0.69	0.79	0.67	0.81	0.77	0.76	0.62	0.82	0.72	0.78	0.78	0.88
Bagging	Chi-Squared	0.70	0.94	0.70	0.94	0.70	0.94	0.70	0.94	0.69	0.94	0.70	0.94	0.78	0.98
	Information Gain	0.70	0.94	0.70	0.94	0.70	0.94	0.70	0.94	0.69	0.94	0.70	0.94	0.78	0.98
	CFS	0.70	0.93	0.70	0.93	0.72	0.93	0.68	0.94	0.73	0.92	0.69	0.93	0.78	0.98

Table A3. Results for Angina outcome.U: Undersampling, O: Oversampling.

Classifier	Feature Selection	Acc		G-mean		Precision		Sensitivity		Specificity		F-measure		ROC area	
		U	O	U	O	U	O	U	O	U	O	U	O	U	O
Naive Bayes	Chi-Squared	0.74	0.74	0.72	0.74	0.68	0.72	0.90	0.80	0.57	0.68	0.77	0.76	0.66	0.82
	Information Gain	0.74	0.74	0.72	0.74	0.68	0.72	0.90	0.80	0.57	0.68	0.77	0.76	0.66	0.82
	CFS	0.74	0.73	0.72	0.73	0.68	0.76	0.90	0.68	0.57	0.78	0.77	0.72	0.66	0.82
RBFN	Chi-Squared	0.73	0.64	0.70	0.62	0.67	0.68	0.90	0.51	0.55	0.76	0.77	0.58	0.64	0.70
	Information Gain	0.73	0.71	0.70	0.69	0.67	0.85	0.90	0.52	0.91	0.77	0.65	0.64	0.80	
	CFS	0.73	0.74	0.70	0.74	0.67	0.80	0.90	0.65	0.55	0.84	0.77	0.72	0.64	0.82
5NN	Chi-Squared	0.73	0.97	0.70	0.97	0.67	0.96	0.90	0.97	0.55	0.96	0.77	0.97	0.65	0.99
	Information Gain	0.73	0.97	0.70	0.97	0.67	0.96	0.90	0.97	0.55	0.96	0.77	0.97	0.65	0.99
	CFS	0.73	0.96	0.70	0.96	0.67	0.95	0.90	0.98	0.55	0.95	0.77	0.97	0.65	0.99
C4.5 (J48)	Chi-Squared	0.73	0.99	0.70	0.99	0.67	0.99	0.90	1.00	0.55	0.99	0.77	0.99	0.64	1.00
	Information Gain	0.73	0.99	0.70	0.99	0.67	0.99	0.90	1.00	0.55	0.99	0.77	0.99	0.64	1.00
	CFS	0.73	0.99	0.70	0.99	0.67	0.99	0.90	0.99	0.55	0.99	0.77	0.99	0.64	1.00
SVM	Chi-Squared	0.74	0.98	0.72	0.98	0.68	0.98	0.90	0.97	0.57	0.98	0.77	0.98	0.74	0.98
	Information Gain	0.73	0.98	0.71	0.98	0.67	0.98	0.90	0.97	0.56	0.98	0.77	0.98	0.74	0.98
	CFS	0.73	0.97	0.71	0.97	0.67	0.98	0.90	0.97	0.56	0.98	0.77	0.97	0.74	0.97
LR	Chi-Squared	0.74	0.79	0.72	0.78	0.68	0.77	0.90	0.82	0.57	0.75	0.77	0.79	0.66	0.85
	Information Gain	0.74	0.78	0.72	0.78	0.68	0.76	0.90	0.82	0.57	0.74	0.77	0.79	0.66	0.85
	CFS	0.74	0.78	0.72	0.78	0.68	0.76	0.90	0.83	0.57	0.74	0.77	0.79	0.66	0.85
Random Forest	Chi-Squared	0.73	1.00	0.70	1.00	0.67	0.99	0.90	1.00	0.90	0.99	0.55	1.00	0.66	1.00
	Information Gain	0.73	1.00	0.70	1.00	0.67	0.99	0.90	1.00	0.55	0.99	0.77	1.00	0.66	1.00
	CFS	0.73	1.00	0.70	1.00	0.67	0.99	0.90	1.00	0.55	0.99	0.77	1.00	0.66	1.00
AdaBoostM1	Chi-Squared	0.74	0.84	0.72	0.84	0.68	0.83	0.90	0.86	0.57	0.83	0.77	0.85	0.67	0.93
	Information Gain	0.73	0.84	0.71	0.84	0.67	0.83	0.90	0.86	0.56	0.83	0.77	0.85	0.67	0.93
	CFS	0.73	0.84	0.71	0.84	0.67	0.83	0.90	0.86	0.56	0.83	0.77	0.85	0.67	0.93
Bagging	Chi-Squared	0.73	1.00	0.70	1.00	0.67	0.99	0.90	1.00	0.55	0.99	0.77	1.00	0.69	1.00
	Information Gain	0.73	1.00	0.70	1.00	0.67	0.99	0.90	1.00	0.55	0.99	0.77	1.00	0.69	1.00
	CFS	0.73	0.99	0.70	0.99	0.67	0.99	0.90	1.00	0.55	0.99	0.77	0.99	0.69	1.00

Table A4. Results for Stroke outcome. U: Undersampling, O: Oversampling.

Classifier	Feature Selection	Acc		G-mean		Precision		Sensitivity		Specificity		F-measure		ROC area	
		U	O	U	O	U	O	U	O	U	O	U	O	U	O
Naive Bayes	Chi-Squared	0.67	0.74	0.67	0.74	0.63	0.73	0.84	0.77	0.55	0.71	0.72	0.75	0.65	0.83
	Information Gain	0.67	0.74	0.67	0.74	0.63	0.73	0.84	0.77	0.55	0.71	0.72	0.75	0.65	0.83
	CFS	0.67	0.74	0.67	0.74	0.63	0.73	0.84	0.75	0.55	0.72	0.72	0.74	0.65	0.84
RBFN	Chi-Squared	0.69	0.67	0.66	0.67	0.64	0.65	0.86	0.76	0.51	0.59	0.73	0.70	0.67	0.75
	Information Gain	0.69	0.71	0.66	0.69	0.64	0.82	0.86	0.53	0.51	0.88	0.73	0.65	0.67	0.79
	CFS	0.69	0.71	0.66	0.70	0.64	0.67	0.86	0.81	0.51	0.60	0.73	0.74	0.67	0.82
5NN	Chi-Squared	0.65	0.96	0.62	0.96	0.61	0.96	0.86	0.97	0.44	0.96	0.71	0.96	0.60	0.99
	Information Gain	0.65	0.96	0.62	0.96	0.61	0.96	0.86	0.97	0.44	0.96	0.71	0.96	0.60	0.99
	CFS	0.65	0.96	0.62	0.96	0.61	0.95	0.86	0.98	0.44	0.94	0.71	0.96	0.60	0.99
C4.5 (J48)	Chi-Squared	0.64	0.99	0.61	0.99	0.60	0.99	0.84	0.99	0.44	0.99	0.70	0.99	0.60	1.00
	Information Gain	0.64	0.99	0.61	0.99	0.60	0.99	0.84	0.99	0.44	0.99	0.70	0.99	0.60	1.00
	CFS	0.64	0.99	0.61	0.99	0.60	0.99	0.84	0.99	0.44	0.99	0.70	0.99	0.60	1.00
SVM	Chi-Squared	0.67	0.97	0.65	0.97	0.63	0.98	0.84	0.97	0.51	0.98	0.72	0.97	0.67	0.97
	Information Gain	0.67	0.97	0.65	0.97	0.63	0.98	0.84	0.97	0.51	0.98	0.72	0.97	0.67	0.97
	CFS	0.67	0.97	0.65	0.97	0.63	0.97	0.84	0.97	0.51	0.97	0.72	0.97	0.67	0.97
LR	Chi-Squared	0.67	0.74	0.65	0.74	0.63	0.75	0.84	0.74	0.51	0.75	0.72	0.74	0.63	0.84
	Information Gain	0.67	0.74	0.65	0.74	0.63	0.74	0.84	0.74	0.51	0.75	0.72	0.74	0.63	0.84
	CFS	0.67	0.73	0.65	0.73	0.63	0.73	0.84	0.74	0.51	0.73	0.72	0.73	0.63	0.83
Random Forest	Chi-Squared	0.67	1.00	0.65	1.00	0.63	0.99	0.84	1.00	0.51	0.99	0.72	1.00	0.64	1.00
	Information Gain	0.67	1.00	0.65	1.00	0.63	0.99	0.84	1.00	0.51	0.99	0.72	1.00	0.64	1.00
	CFS	0.67	1.00	0.65	1.00	0.63	0.99	0.84	1.00	0.51	0.99	0.72	1.00	0.64	1.00
AdaBoostM1	Chi-Squared	0.67	0.83	0.65	0.83	0.63	0.82	0.84	0.85	0.51	0.82	0.72	0.84	0.64	0.92
	Information Gain	0.67	0.83	0.65	0.83	0.63	0.82	0.84	0.85	0.51	0.82	0.72	0.84	0.64	0.92
	CFS	0.67	0.82	0.65	0.82	0.63	0.81	0.84	0.84	0.51	0.80	0.72	0.82	0.64	0.91
Bagging	Chi-Squared	0.67	0.99	0.65	0.99	0.63	0.99	0.84	1.00	0.51	0.99	0.72	0.99	0.67	1.00
	Information Gain	0.67	0.99	0.65	0.99	0.63	0.99	0.84	1.00	0.51	0.99	0.72	0.99	0.67	1.00
	CFS	0.67	0.99	0.65	0.99	0.63	0.99	0.84	1.00	0.51	0.99	0.72	0.99	0.67	1.00

Table A5. Comparative ROC area results for SVM using Linear kernel against RBF kernel

Outcome	Linear kernel	RBF kernel
Diabetes	0.62	0.70
Heart Failure	0.75	0.88
Stroke	0.75	0.97
Angina	0.79	0.98
