Table 4: SVM results: Comparing RBF kernel, GMML, SimpleMKL, Laplacian kernel and Isolation Kernel. GMML and SimpleMKL have "out of memory" error on 2 datasets; SimpleMKL cannot return any result within 24 hours on 5 datasets. The first and second subsets consists of 2-class datasets having less than and more than 5000 points, respectively; the third subset consists of multi-class datasets.

	#inst	#fea	#class	RBF	GMML	SimpleMKL	Laplacian	Isolation
GPS	163	6	2	0.855±0.018	0.855±0.028	0.867±0.018	0.836±0.026	0.842±0.022
Heart	270	13	2	0.833±0.027	0.844±0.034	0.826±0.035	0.852±0.024	0.852±0.025
Breast	277	9	2	0.750±0.018	0.768±0.016	0.746±0.009	0.761±0.012	0.771±0.015
Ionosphere	351	34	2	0.949±0.006	0.941±0.020	0.958±0.009	0.946±0.014	0.955±0.011
Vote	435	16	2	0.943±0.013	0.949±0.010	0.940±0.007	0.956±0.008	0.961±0.013
ILPD	583	11	2	0.716±0.004	0.726±0.008	0.720±0.003	0.713±0.002	0.720±0.008
WBC	683	9	2	0.972±0.003	0.974±0.004	0.969±0.004	0.977±0.003	0.975±0.007
Austra	690	14	2	0.857±0.016	0.864±0.014	0.854±0.015	0.859±0.017	0.871±0.011
German	1000	24	2	0.753±0.008	0.754±0.006	0.705±0.003	0.759±0.006	0.767±0.012
Parkinson	1040	28	2	0.999±0.001	1.000±0.000	0.998±0.002	1.000±0.000	1.000±0.000
QSAR	1055	42	2	0.880±0.007	0.870±0.005	0.864±0.007	0.873±0.009	0.869±0.012
Messidor	1151	19	2	0.690±0.021	0.734±0.014	0.673±0.023	0.692±0.019	0.694±0.018
Spam	4141	58	2	0.932±0.004	0.912±0.003	> 24 hours	0.944±0.002	0.940±0.003
Wilt	4839	5	2	0.946±0.039	0.984±0.001	> 24 hours	0.946±0.021	0.985±0.014
Mushrooms	8124	112	2	1.000±0.000	1.000±0.000	> 24 hours	1.000±0.000	1.000±0.000
Phishing	11055	30	2	0.965±0.001	0.959±0.003	> 24 hours	0.968±0.002	0.967±0.001
a8a	32561	123	2	0.846±0.003	memory error	memory error	0.846±0.003	0.847±0.003
IJCNN	49990	22	2	0.980±0.001	memory error	memory error	0.978±0.001	0.978±0.002
Urban	168	147	9	0.835± 0.022	0.841±0.029	0.882±0.027	0.847±0.033	0.841±0.020
Air	359	64	3	0.956±0.010	0.933±0.012	0.942±0.013	0.967±0.008	0.967±0.013
Forest	523	27	4	0.895 ±0.014	0.910±0.011	0.964±0.011	0.891±0.015	0.912±0.011
Vowel	528	10	11	0.983±0.010	0.979±0.010	0.982±0.007	0.979± 0.010	0.989±0.007
Corel	10000	67	100	0.364±0.005	0.387±0.005	> 24 hours	0.457±0.003	0.466±0.004
Isolation has #wins/#draws/#losses				19/1/3	14/3/4	11/1/4	13/5/5	-

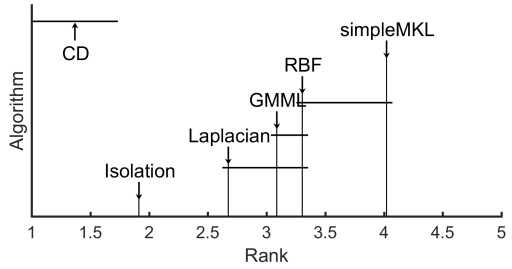

Figure 6: Nemenyi Test at 0.05 significance level. If two algorithms are connected by a CD (critical difference) line, then there is no significant difference between them.

Table 5: Runtime (second). Total time for training & testing.

	RBF	GMML	SimpleMKL	Laplacian	Isolation
IJCNN	54+14	mem err	mem err	56+15	4297+716
Corel	26+7	23+5	>24 hrs	27+9	3975+810

Note that GMML and SimpleMKL could not complete on IJCNN because of their large memory requirements; and SimpleMKL could not complete within 24 hours on Corel.

6 RELATION WITH RF KERNEL

6.1 Breiman's explanation of RF classifier behaviour in terms of kernel

By defining RF kernel of two points as the average number of shared leaf nodes of Random Forest, Breiman [2] explains the behavior of the RF classifier in terms of this kernel: as a means to locate boundaries, which trade-offs between the kernel's 'symmetry' and 'skewness'. They correspond to correlation and strength of the ensemble, respectively. 'Symmetry' kernel is produced from completely random splits; and 'skewed' kernel is generated from splits which favour pure nodes, which has high strength or classification accuracy. This skewness is also conjectured to enable nonlinear classification [2].

This interpretation has inspired others to use RF as a similarity measure in distance-based neighbourhood methods. Breiman and Cutler [3] first describe two methods to generate RF similarity: one generates RF similarity from a labelled dataset; and the other from an unlabelled dataset. Shi and Horvath [13] applied the second method for tumor discovery using the RF similarity in a distance-based clustering algorithm. Davis and Ghahramani [6] attempted to generalise RF kernel to use different partitioning methods; however, evaluations of their validity are not provided.

The applications of RF similarity has been limited for two reasons. First, the theory requires that the trees are trained using bootstrap