Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\sigma \in [0.01, 100]$	84.46	48.48	9.39	38.6	93.94	64.81	52.6	80.29	68.53	21.44	65.62	11.04	4.72	80.06	82.03
Feature $\in [2, 10]$	6	6	2	2	4	2	2	3	5	4	4	5	9	2	4
$\beta \in [0.01, 100]$	9.84	4.77	6.54	2.62	0.3	4.44	0.56	8.84	5.23	4.22	6.53	1.35	3.23	6.65	8.01

Table 1: Optimal values of MMCC parameters for 15 two-class datasets

Dataset	16	17	18	19	20	21	22	23	24	25
$\sigma \in [0.01, 100]$	6.63	96.61	59.31	29.25	31.52	65.92	72.78	20.63	0.33	88.65
Feature $\in [2, 10]$	7	2	8	3	6	2	7	5	8	29
$\beta \in [0.01, 100]$	7.54	9.08	7.15	35.98	6.89	0.89	19.03	14.99	8.44	29.041

Table 2: Optimal values of MMCC parameters for 10 multi-class datasets

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
MMCC	0.44	2.21	0.5	0.07	0.08	0.03	0.01	0.01	0.03	0.02	0.02	0.22	0.18	0.23	0.22
MIM	0.54	0.74	0.69	0.28	0.34	0.23	0.06	0.06	0.04	0.04	0.03	0.3	0.2	1.75	0.53
MRMR	0.44	1.23	0.77	105.04	114.3	60.96	0.18	0.140	0.14	0.04	0.16	0.35	1669.5	0	204.63
CMIM	0.11	0.26	0.24	120.87	116.05	70.5	0.03	0.03	0.01	0.01	0.16	0.06	165.57	1418.78	140.62
JMI	0.9	2.23	1.11	1.64	3.31	0.81	0.04	0.04	0.07	0.03	0.05	0.47	2.58	6.74	3.9

Table 3: The time (in seconds) comparison of MMCC, MIM [36], MRMR [37], CMIM [38] and JMI [39] on 15 two-class datasets.

Dataset	16	17	18	19	20	21	22	23	24	25
MMCC	0.07	0.01	0.006	2.07	0.58	0.05	0.43	0.70	0.89	1.13
MIM	0.15	0.08	0.01	4.57	6.34	0.57	0.43	0.63	0.56	3.85
MRMR	0.21	0.08	0.03	NA*	NA	NA	75.78	24.66	0.75	NA
CMIM	0.04	0.05	0.03	NA	NA	1018.98	187.97	89.07	0.15	NA
JMI	0.39	0.04	0.01	NA	NA	5.15	6.29	12.05	1.83	NA

* indicates the *out of memory* error while running the program

Table 4: The time (in seconds) comparison of MMCC, MIM [36], MRMR [37], CMIM [38] and JMI [39] on 10 multi-class datasets.

Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ANHD	0.007	0.007	0.008	0.014	0.025	0.025	0.041	0.041	0.083	0.083	0.09	0.008	0.050	0.014	0.014

(a) ANHD on two-class datasets

Dataset	16	17	18	19	20	21	22	23	24	25
ANHD	0.025	0.044	0.7	0.003	0.014	0.020	0.014	0.008	0.008	0.014

(b) ANHD on multi-class datasets

Table 5: Exploring the stability of 15 two-class (above) and 10 multi-class (bottom) datasets via Average Normalized Hamming Distance (ANHD) [40].

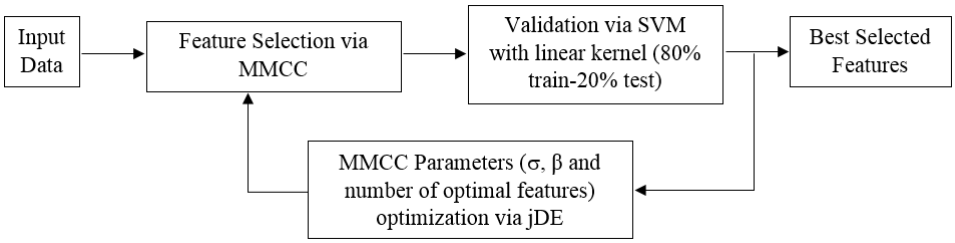

Figure 1: The procedure of MMCC for selection optimal parameters and features.

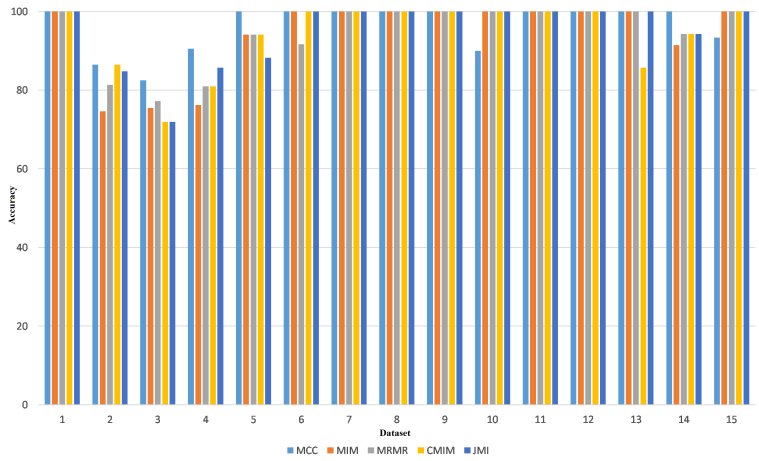

Figure 2: Maximum accuracy in 10,000 runs of SVM classification on selected features for 15 two-class datasets. The methods are MMCC, MIM [36], MRMR [37], CMIM [38] and JMI [39].

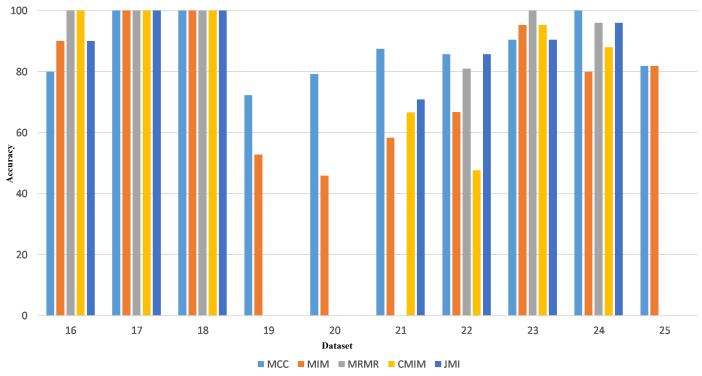

Figure 3: Maximum accuracy in 10,000 runs of SVM classification on selected features for 10 multi-class datasets. The methods are MMCC, MIM [36], MRMR [37], CMIM [38] and JMI [39].