A Feature Selection Algorithm Based on Equal Interval Division and Conditional Mutual Information

Xiangyuan Gu[1] · Jichang Guo[1] · Tao Ming[1] · Lijun Xiao[1] · Chongyi Li[1]

Abstract
The performance of many feature selection algorithms is affected because of ignoring three-dimensional mutual information among features. Three-dimensional mutual information includes conditional mutual information, joint mutual information and three-way interaction information. Aiming at the limitation, this paper investigates feature selection based on three-dimensional mutual information. First, we propose an objective function based on conditional mutual information. Further, we propose a criterion to validate whether the objective function can guarantee the effectiveness of selected features. In the case that the objective function cannot guarantee the effectiveness of selected features, we combine a method of equal interval division and ranking with the objective function to select features. Finally, we propose a feature selection algorithm named EID-CMI. To validate the performance of EID-CMI, we compare it with several feature selection algorithms. Experimental results demonstrate that EID-CMI can achieve better feature selection performance.

1 Introduction

Feature selection is an important way of dimension reduction, which adopts metrics to measure features and selects informative features [1,2]. Feature selection algorithms can be classified into three categories: wrapper, embedded and filter. Since accuracy of classifier is used as the metric, wrapper and embedded algorithms are time-consuming. Filter algorithms are time-saving at the cost of performance. The most commonly used metrics in filter algorithms include Pearson correlation coefficient, Euclidean distance and mutual information. As mutual information can measure linear and nonlinear correlation, and has the invariance under space transformations [3], it has an advantage over other metrics. Therefore, many feature selection algorithms based on mutual information are proposed [4,5].

✉ Jichang Guo
  jcguo@tju.edu.cn

1   School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

MIM [6] is a basic algorithm based on mutual information and it calculates mutual information between each feature and the class label. On basis of MIM, some feature selection algorithms based on relevancy and redundancy are proposed, such as MIFS [7], MIFS-U [8], mRMR [9], CMI [10] and MIFS-CR [11]. These algorithms introduce mutual information between features and use it to describe redundancy. Furthermore, in our previous work [12], we investigate feature selection based on mutual information and propose an algorithm named EID-mRMR. EID-mRMR combines mutual information with a method of equal interval division and ranking to select features. Since three-dimensional mutual information, such as conditional mutual information, joint mutual information or three-way interaction information, is not exploited, the performance of these algorithms can be influenced.

There are some feature selection algorithms based on three-dimensional mutual information, such as DWFS [13], IWFS [14], JMI [15], CMIM [16], CIFE [17], JMIM [18] and RelaxMRMR [19]. All of these algorithms consider three-dimensional mutual information among features and the class label, while only RelaxMRMR considers three-dimensional mutual information among features.

Since our previous work [12] only considers mutual information and does not consider three-dimensional mutual information, the objective function of EID-mRMR misses some useful information and the performance is affected. To further improve the performance, feature selection based on three-dimensional mutual information is investigated. This work is its extension in three-dimensional mutual information and it combines three-dimensional mutual information among features with the method of equal interval division and ranking to select features. Specifically, an objective function based on conditional mutual information is presented. Then, the objective function is analyzed and a criterion that validates whether the objective function can guarantee the performance of selected features is achieved. Following that, in the case that the objective function cannot guarantee the performance of selected features, conditional mutual information among features and the class label as well as conditional mutual information among features is processed by the method of equal interval division and ranking. Finally, a feature selection algorithm named EID-CMI is proposed.

The main contributions of this paper are as follows: (1) Feature selection based on three-dimensional mutual information is investigated. (2) The method of equal interval division and ranking is adopted. (3) A feature selection algorithm based on conditional mutual information is proposed. (4) The proposed algorithm can achieve better feature selection performance.

The remainder of this paper is organized as follows: Sect. 2 presents some basic concepts of information theory. Related works of feature selection algorithms based on three-dimensional mutual information are introduced in Sect. 3. Section 4 presents the proposed algorithm. Experimental results are given and analyzed in Sect. 5. Section 6 gives conclusions and future work.

2 Some Basic Concepts of Information Theory

Mutual information is exploited to quantify the information shared by two variables, and mutual information $I(X; Y)$ between two variables X and Y can be defined as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{1}$$

A greater mutual information value indicates that two variables share more information.

Conditional mutual information is adopted to quantify the information by the other two variables when one variable is known and $I(X; Y|Z)$ can be defined as

$$I(X; Y|Z) = \sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} \tag{2}$$

Three-way interaction information is an extension of mutual information and $I(X; Y; Z)$ has the following relationship with conditional mutual information $I(X; Y|Z)$ [20].

$$I(X; Y; Z) = I(X; Y|Z) - I(X; Y) \tag{3}$$

where $I(X; Y; Z)$ can be positive, negative or zero.

3 Related Works

Feature selection algorithms based on three-dimensional mutual information include DWFS, IWFS, JMI, CMIM, CIFE, JMIM and RelaxMRMR. DWFS and IWFS belong to the same class, and they utilize symmetric uncertainty [21] between features and the class label to describe relevance, and adopt three-way interaction information among features and the class label to describe interaction. Owing to lacking some evidences in determining these two parts, their feature selection performance cannot be guaranteed.

JMI, CMIM, CIFE, JMIM and RelaxMRMR belong to the same class. Objective functions are the key of these algorithms and their objective functions are respectively presented as follows:

$$JMI = \arg \max_{f_i \in X} \left[I(c; f_i) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i) + \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i|c) \right] \tag{4}$$

$$CMIM = \arg \max_{f_i \in X} \left\{ \min_{f_s \in S} \left[I(c; f_i|f_s) \right] \right\} \tag{5}$$

$$CIFE = \arg \max_{f_i \in X} \left[I(c; f_i) - \sum_{f_s \in S} I(f_s; f_i) + \sum_{f_s \in S} I(f_s; f_i|c) \right] \tag{6}$$

$$JMIM = \arg \max_{f_i \in X} \left\{ \min_{f_s \in S} \left[I(f_i, f_s; c) \right] \right\} \tag{7}$$

$$RelaxMRMR = \arg \max_{f_i \in X} I(c; f_i) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i)$$

$$+ \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i|c) - \frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_i; f_k|f_s) \tag{8}$$

where $I(c; f_i)$ is mutual information between a candidate feature f_i and the class label c, $I(f_s; f_i)$ is mutual information between a selected feature f_s and f_i, $I(f_s; f_i|c)$ is conditional mutual information when c is given, $I(c; f_i|f_s)$ is conditional mutual information when f_s is given, $I(f_i, f_s; c)$ is joint mutual information, $I(f_i; f_k|f_s)$ is conditional mutual information when f_s is given, $|S|$ is the number of selected features, S is the selected feature set and X is the candidate feature set.

JMI, CMIM, CIFE and JMIM do not consider three-dimensional mutual information among features except for RelaxMRMR. Since three-dimensional mutual information among

features is considered, the objective function of RelaxMRMR contains more useful information.

4 The Proposed Feature Selection Algorithm

This section first presents our proposed objective function and a criterion that validates whether the objective function can guarantee the effectiveness of selected features. Then, a method of equal interval division and ranking is given. Finally, our proposed feature selection algorithm, EID-CMI, is presented.

4.1 The Proposed Objective Function and Criterion

First, we analyze the first three selected features. Since it can provide more information, the feature that has maximum mutual information with the class label c is the first to select and it is assumed to be f_s. In selecting the second feature, Eq. (9) is satisfied.

$$I(c; T) = I(c; f_i, f_s) = I(c; f_s) + I(c; f_i|f_s) \tag{9}$$

where f_i is a candidate feature. In order to guarantee $I(c; T)$ is maximum, since $I(c; f_s)$ is maximum, $I(c; f_i|f_s)$ should be maximum. The feature that has maximum conditional mutual information with c and f_s is selected.

In selecting the third feature, Eq. (10) is satisfied.

$$I(c; T) = I(c; f_i, f_s, f_k) \tag{10}$$

Since it is difficult to calculate Eq. (10), we introduce Eqs. (11) and (8) is its special case attaining maximum.

$$J(f_i) = I(c; f_i) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i) + \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i|c)$$
$$- \frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_i; f_k|f_s) \tag{11}$$

where f_k is a selected feature and S is the selected feature set.

Equation (12) is satisfied.

$$I(c; f_i) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i) + \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i|c)$$
$$= I(c; f_i) + \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i; c) = \frac{1}{|S|} \sum_{f_s \in S} I(c; f_i|f_s) \tag{12}$$

Equation (13) is introduced in selecting the third feature for simplifying computation and it can be understood as effective information that is brought in selecting features. The first part means the average value of relevant information, while the second part can be considered as the average value of redundant information.

$$J(f_i) = \frac{1}{|S|} \sum_{f_s \in S} I(c; f_i|f_s) - \frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s) \tag{13}$$

When the maximum of Eq. (13) is far greater than the secondary maximum, more information can be achieved by selecting the maximum and thus can guarantee the effectiveness of selected features. Otherwise, the advantage is not obvious. To guarantee the performance of selected features, a criterion that whether the difference between the maximum and secondary maximum is greater than a fixed value P is used to evaluate whether the objective function can guarantee the performance of selected features, as we do in our previous work [12]. If the difference is greater than P, Eq. (13) is exploited to select features; otherwise, the method of equal interval division and ranking is adopted to select features and it will be introduced in the next section.

4.2 Equal Interval Division and Ranking

In the case that the performance of selected features cannot be guaranteed, the method of equal interval division and ranking is exploited to process these two parts of Eq. (13), and CFF-CMI $(c; f_i|f_s)$ and FFF-CMI $(f_k; f_i|f_s)$ are obtained. Their computing processes are shown in Algorithms 1 and 2.

The number of dataset's features is taken as the number of subintervals. Take Algorithm 1 for example to analyze the method of equal interval division and ranking, the concrete practices are presented as follows: determining maximum and minimum of the interval of average value of $I(c; f_i|f_s)$ as interval values, and the interval values are divided equally. Then, the subintervals are indexed from 1 to the number of dataset's features, and the indexes of these subintervals are taken as the ordinal values of the values in the subintervals.

The method of equal interval division and ranking is adopted to process these two parts, and it can increase the difference among each part of values and guarantee the features in the same subinterval have the same priority. Therefore, Eq. (14) is adopted to process the case that the effectiveness of selected features cannot be guaranteed. When the number of features that satisfies the condition is 1, selecting the feature; otherwise, the feature that maximizes Eq. (13) is selected from these features.

$$J(f_i) = \arg\max_{f_i \in X} [CFF - CMI(c; f_i|f_s) - FFF - CMI(f_k; f_i|f_s)] \tag{14}$$

Algorithm 1 Compute CFF-CMI $(c; f_i|f_s)$;

Require: M: the number of dataset's features, $|S|$: the number of features in S, $\frac{1}{|S|}\sum_{f_s \in S} I(c; f_i|f_s)$.

Ensure: CFF-CMI $(c; f_i|f_s)$.
1: a = min $[\frac{1}{|S|}\sum_{f_s \in S} I(c; f_i|f_s)]$, b = max $[\frac{1}{|S|}\sum_{f_s \in S} I(c; f_i|f_s)]$;
2: c = (b-a) / (M-1);
3: **for** $f_i \in X$ **do**
4:     **for** j = 1: M **do**
5:         **if** $\frac{1}{|S|}\sum_{f_s \in S} I(c; f_i|f_s) \geq$ a - c + j*c **then**
6:             **if** $\frac{1}{|S|}\sum_{f_s \in S} I(c; f_i|f_s) <$ a + j*c **then**
7:                CFF-CMI $(c; f_i|f_s)$ = j;
8:             **end if**
9:         **end if**
10:    **end for**
11: **end for**

Algorithm 2 Compute FFF-CMI (f_k; $f_i|f_s$);

Require: M: the number of dataset's features, $|S|$: the number of features in S, $\frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s)$.

Ensure: FFF-CMI (f_k; $f_i|f_s$).

1: a = min $[\frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s)]$, b = max $[\frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s)]$;

2: c = (b-a) / (M-1);

3: **for** $f_i \in X$ **do**

4:     **for** j = 1: M **do**

5:         **if** $\frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s) \geq$ a - c + j*c **then**

6:             **if** $\frac{1}{|S||S-1|} \sum_{f_s \in S} \sum_{f_k \in S; k \neq s} I(f_k; f_i|f_s) <$ a + j*c **then**

7:             FFF-CMI (f_k; $f_i|f_s$) = j;

8:         **end if**

9:         **end if**

10:     **end for**

11: **end for**

4.3 Algorithmic Implementation

The proposed objective function and criterion are combined with the method of equal interval division and ranking, and a feature selection algorithm named EID-CMI is proposed and it is shown in Algorithm 3.

EID-CMI consists of two parts: in the first part (lines 1–10), initializing the candidate feature set X and the selected feature set S. Then, calculating $I(c; f_i)$ and selecting the feature with the maximum. Calculating $I(c; f_i|f_s)$ and taking the feature with the maximum from X; in the second part (lines 11–31), calculating $I(f_k; f_i|f_s)$, and the difference between the maximum and secondary maximum of Eq. (13), if the difference is greater than a fixed value P, selecting the feature with the maximum of Eq. (13); if not, calculating CFF-CMI (c; $f_i|f_s$), FFF-CMI (f_k; $f_i|f_s$) and the difference, and testing the number of the features with the maximum difference, if it is one, selecting the feature; otherwise, selecting the feature with the maximum of Eq. (13) from these features with the maximum difference. Following the above steps, it loops to select features until selecting a specified number of features N.

5 Experimental Results

This section describes the datasets and experimental settings, and compares EID-CMI with several algorithms.

5.1 The Datasets and Experimental Settings

These datasets in Table 1 are commonly used in feature selection and they come from UCI machine learning repository [22] and ASU feature selection datasets [23]. These datasets are from different fields, such as face image and digit recognition. In these datasets, their instances vary from 100 to 7000 and the number of their features varies from 57 to 10304. For these datasets with continuous features, minimum description length discretization method [24] is employed. Three classifiers, J48, IB1 and Naive Bayes, are adopted and their parameters are

---

Algorithm 3 EID-CMI: an algorithm based on equal interval division and conditional mutual information;

Require: M: the number of dataset's features, N: the number of features to be selected, P: a fixed value.
Ensure: S: the selected features.
1: initialize $S = \emptyset$ and $X = \{f_1, f_2, \ldots, f_M\}$;
2: **for** $f_i \in X$ **do**
3:    compute mutual information $I(c; f_i)$;
4: **end for**
5: take the feature f_l by maximizing mutual information from X and put it into S;
6: $f_s = f_l$;
7: **for** $f_i \in X$ **do**
8:    compute conditional mutual information $I(c; f_i|f_s)$;
9: **end for**
10: take the feature f_m by maximizing conditional mutual information from X and put it into S;
11: **while** $|S| \leq N$ **do**
12:    **for** $f_i \in X$ **do**
13:       **for** $f_s \in S$ **do**
14:          **for** $f_k \in S$ and $f_k \neq f_s$ **do**
15:             compute conditional mutual information $I(f_k; f_i|f_s)$;
16:          **end for**
17:       **end for**
18:    **end for**
19:    compute the difference between the maximum and secondary maximum of Eq.(13);
20:    **if** the difference is greater than P **then**;
21:       take the feature f_n that maximizes the condition from X and put it into S;
22:    **else**
23:       compute CFF-CMI $(c; f_i|f_s)$ and FFF-CMI $(f_k; f_i|f_s)$;
24:       find the number of the candidate features satisfying Eq.(14);
25:       **if** the number of features is more than 1 **then**
26:          take the feature f_p that maximizes Eq.(13) from these features from X and put it into S;
27:       **else**
28:          take the feature f_q that satisfies the condition from X and put it into S;
29:       **end if**
30:    **end if**
31: **end while**

---

set to WEKA's [25] default values. ASU feature selection software package [26] is utilized. N is set to 50.

For reducing the influence of randomness on final results, ten times of 10-fold cross-validation are conducted, and the mean values and standard deviations of ten results are taken as the final results. To determine whether the effectiveness of experimental results is significant, a one-sided paired t-test at 5% significance level is carried out, and the number of the datasets that EID-CMI performs better than/equal to/worse than other algorithms is presented by Win/Tie/Loss (W/T/L).

5.2 Experimental Results and Analysis

5.2.1 With Different *P* Values

To analyze the influence of *P* value on EID-CMI, we set the *P* value to 0.02, 0.03, 0.04, 0.05 and 0.06 respectively and conduct experiments with J48, IB1 and Naive Bayes. Calculating the accuracy of selected features from 1 to 50 for the three classifiers and the mean value is taken as the final result. Average performance of EID-CMI with different *P* values is

Table 1 Summary of datasets in the experiment

Datasets	Instances	Features	Classes	Types	Source
Spambase	4601	57	2	Continuous	UCI
Synthetic_control	600	60	6	Continuous	UCI
Mfeat_fou	2000	76	10	Continuous	UCI
Movement_libras	360	90	15	Continuous	UCI
Musk	476	166	2	Continuous	UCI
Mfeat_fac	2000	216	10	Continuous	UCI
Mfeat_pix	2000	240	10	Discrete	UCI
Semeion	1593	256	10	Discrete	UCI
USPS	256	9298	10	Continuous	ASU
lung_small	325	73	7	Discrete	ASU
Isolet	617	1560	26	Continuous	ASU
ORL	400	1024	40	Continuous	ASU
COIL20	1440	1024	20	Continuous	ASU
warpPIE10P	2420	210	10	Continuous	ASU
lung	3312	203	5	Continuous	ASU
gisette	7000	5000	2	Continuous	ASU
TOX_171	5748	171	4	Continuous	ASU
arcene	10000	200	2	Continuous	ASU
pixraw10P	10000	100	10	Continuous	ASU
orlraws10P	100	10304	10	Continuous	ASU

presented in Table 2. The value in the Avg. row presents the mean value of these twenty values.

In Table 2, there is a little difference among these Avg. values. When the P value is 0.04, EID-CMI achieves the best Avg. value. Therefore, in comparison with other feature selection algorithms, the P value of EID-CMI is set to 0.04.

5.2.2 Comparison with Feature Selection Algorithms

To validate feature selection performance of EID-CMI, several three-dimensional mutual information-based algorithms, JMI [15], CMIM [16], CIFE [17], JMIM [18], RelaxMRMR [19], DWFS [13] and IWFS [14], are compared. Average accuracies of selected features with J48, IB1 and Naive Bayes are presented in Tables 3, 4 and 5, respectively. Average performance of these twenty datasets with three classifiers is shown in Fig. 1. Calculating the accuracy of selected features from 1 to 50 and selecting the optimal value. The optimal accuracy of EID-CMI and other algorithms when using J48, IB1 and Naive Bayes is presented in Tables 6, 7 and 8. The running time of EID-CMI and other algorithms that select 50 features is shown in Table 9.

As shown in Table 3, CMIM, RelaxMRMR and EID-CMI can obtain better results than other algorithms in terms of the Avg. and W/T/L values. Compared with RelaxMRMR, the Avg. value of EID-CMI is 0.64% greater than RelaxMRMR. In the W/T/L values, the number of the datasets that EID-CMI is superior to RelaxMRMR is 7, while that of the datasets that

Table 2 Average performance with different P values

Datasets	0.02	0.03	0.04	0.05	0.06
Spambase	86.82	86.65	86.34	86.22	86.23
Synthetic_control	93.03	93.01	93.02	93.03	93.04
Mfeat_fou	77.26	77.24	77.19	77.16	77.06
Movement_libras	66.92	67.03	67.11	67.09	67.08
Musk	80.64	80.64	80.64	80.64	80.64
Mfeat_fac	89.66	89.68	89.69	89.71	89.70
Mfeat_pix	80.53	80.68	80.76	80.75	80.75
Semeion	70.91	71.36	71.77	71.93	71.99
USPS	85.27	85.27	85.27	85.27	85.27
lung_small	65.57	65.72	65.68	65.56	65.41
Isolet	75.28	75.31	75.25	75.20	75.17
ORL	68.69	68.91	69.14	69.05	69.12
COIL20	93.48	93.49	93.53	93.53	93.53
warpPIE10P	87.81	87.98	88.07	88.01	88.09
lung	92.11	91.94	91.87	91.84	91.76
gisette	90.97	91.12	91.17	91.17	91.17
TOX_171	71.03	71.05	70.71	70.79	70.71
arcene	77.60	77.61	77.61	77.61	77.61
pixraw10P	94.88	94.88	94.88	94.88	94.88
orlraws10P	86.81	86.83	86.88	86.89	86.94
Avg.	81.76	81.82	81.83	81.82	81.81

RelaxMRMR is superior to EID-CMI is 2, as presented, EID-CMI has more advantages than RelaxMRMR.

In Table 4, CMIM, RelaxMRMR and EID-CMI can obtain greater Avg. value. The W/T/L values show that RelaxMRMR and EID-CMI achieve better feature selection performance. Compared with Table 3, considering the Avg. values, EID-CMI gains more advantages than the other algorithms. In terms of the W/T/L values, EID-CMI outperforms other algorithms except IWFS with extra performance gain.

The Avg. and W/T/L values in Table 5 show that CMIM, RelaxMRMR and EID-CMI achieve better results. Different from Tables 3 and 4, in terms of the Avg. values, EID-CMI achieves more advantages than other algorithms. The W/T/L values show that EID-CMI is superior to other algorithms except for CMIM with more performance gain.

Tables 6, 7 and 8 show that EID-CMI performs better than other algorithms in general. Take Table 7 for example, in the twenty datasets, the number of the datasets that EID-CMI can obtain better results than other algorithms is 9. Furthermore, EID-CMI can achieve greater Avg. values than other algorithms with the three classifiers.

In Table 9, owing to calculating conditional mutual information among features and the class label as well as conditional mutual information among features, from the Avg. value, we can see that EID-CMI and RelaxMRMR are more time-consuming than the other algorithms. Considering that it only calculates conditional mutual information among features and the class label, CMIM takes the least time in feature selection.

As shown in Fig. 1, EID-CMI achieves better feature selection performance, while the other algorithms cannot handle well in some datasets. For example, JMI cannot obtain the desired

Table 3 Margins (mean ± std. dev.) between the performance of algorithms when using J48

Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	92.02 ± 0.08	92.00 ± 0.07	92.11 ± 0.07	90.26 ± 0.13	91.96 ± 0.08	91.90 ± 0.12	91.95 ± 0.06	92.07 ± 0.10
Synthetic_control	91.10 ± 0.48	88.47 ± 0.46	91.31 ± 0.40	88.88 ± 0.45	90.32 ± 0.62	89.89 ± 0.65	86.00 ± 0.31	91.08 ± 0.50
Mfeat_fou	75.13 ± 0.47	75.33 ± 0.50	75.24 ± 0.48	72.29 ± 0.41	75.37 ± 0.50	72.65 ± 0.35	75.20 ± 0.47	75.40 ± 0.55
Movement_libras	64.08 ± 0.77	58.43 ± 1.11	62.63 ± 1.14	61.56 ± 1.15	60.99 ± 1.33	62.07 ± 0.96	59.54 ± 1.32	60.16 ± 0.85
Musk	80.42 ± 0.80	78.04 ± 0.50	80.07 ± 0.81	78.55 ± 0.94	80.21 ± 0.88	80.20 ± 0.77	79.24 ± 0.61	80.28 ± 0.64
Mfeat_fac	85.91 ± 0.19	84.45 ± 0.17	85.15 ± 0.22	82.49 ± 0.51	85.52 ± 0.44	84.24 ± 0.33	85.18 ± 0.31	85.89 ± 0.36
Mfeat_pix	74.79 ± 0.18	71.62 ± 0.26	73.92 ± 0.75	70.06 ± 0.54	73.21 ± 0.46	71.63 ± 0.40	73.57 ± 0.27	74.20 ± 0.42
Semeion	70.00 ± 0.50	61.49 ± 0.21	68.56 ± 0.38	62.03 ± 0.72	64.03 ± 0.19	61.33 ± 0.51	65.37 ± 0.46	65.40 ± 0.21
USPS	84.39 ± 0.14	80.21 ± 0.07	83.08 ± 0.07	81.66 ± 0.15	81.88 ± 0.15	81.81 ± 0.18	81.21 ± 0.09	83.50 ± 0.17
lung_small	45.25 ± 3.38	44.73 ± 4.30	43.63 ± 3.25	42.13 ± 2.56	43.04 ± 2.50	40.81 ± 3.38	44.00 ± 3.03	45.59 ± 4.20
Isolet	72.08 ± 0.33	62.65 ± 0.62	69.32 ± 0.47	63.84 ± 0.62	70.78 ± 0.38	63.38 ± 0.57	64.58 ± 0.63	68.63 ± 0.59
ORL	51.79 ± 1.73	52.22 ± 1.35	52.09 ± 1.01	51.47 ± 1.05	52.38 ± 1.64	50.74 ± 1.67	52.41 ± 1.04	52.47 ± 1.37
COIL20	90.19 ± 0.17	87.70 ± 0.29	88.37 ± 0.49	84.87 ± 0.39	87.87 ± 0.66	86.43 ± 0.40	88.12 ± 0.40	89.21 ± 0.33
warpPIE10P	78.39 ± 0.89	79.01 ± 1.19	79.74 ± 1.16	80.11 ± 2.34	80.81 ± 1.10	81.13 ± 1.99	79.19 ± 1.08	78.73 ± 1.62
lung	89.96 ± 1.02	88.92 ± 1.28	90.10 ± 0.87	86.09 ± 1.50	87.34 ± 0.90	86.40 ± 1.62	88.08 ± 1.45	89.35 ± 0.87
gisette	92.76 ± 0.06	92.28 ± 0.06	91.83 ± 0.11	92.41 ± 0.10	92.66 ± 0.09	92.17 ± 0.16	91.20 ± 0.06	92.64 ± 0.09
TOX_171	60.47 ± 2.31	59.32 ± 1.60	59.33 ± 1.55	58.95 ± 2.43	61.74 ± 1.83	61.20 ± 2.77	57.40 ± 1.22	59.49 ± 1.81
arcene	77.67 ± 1.46	73.83 ± 1.58	73.71 ± 1.47	75.34 ± 2.25	76.51 ± 2.14	76.54 ± 1.73	72.77 ± 1.71	77.98 ± 1.49
pixraw10P	91.37 ± 0.96	91.30 ± 1.45	91.03 ± 0.86	90.23 ± 2.27	90.44 ± 1.91	91.36 ± 3.24	92.40 ± 1.85	91.74 ± 1.29
orlraws10P	75.77 ± 2.39	73.58 ± 2.29	76.03 ± 3.41	71.67 ± 2.87	74.49 ± 1.91	71.67 ± 3.76	73.86 ± 2.65	76.90 ± 2.13
Avg.	77.18	74.78	76.36	74.24	76.08	74.88	75.06	76.54
W/T/L	–	12/7/1	9/9/2	16/3/1	13/5/2	15/4/1	14/5/1	7/11/2

Table 4 Margins (mean ± std. dev.) between the performance of algorithms when using 1B1

Datasets	IED-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	84.56 ± 0.11	85.35 ± 0.09	83.89 ± 0.11	83.12 ± 0.15	83.98 ± 0.08	85.18 ± 0.08	84.77 ± 0.12	84.95 ± 0.09
Synthetic_control	93.47 ± 0.28	90.40 ± 0.33	93.60 ± 0.26	86.44 ± 0.39	90.30 ± 0.36	87.69 ± 0.47	89.92 ± 0.29	93.36 ± 0.27
Mfeat_fou	79.75 ± 0.16	80.19 ± 0.21	80.17 ± 0.21	78.42 ± 0.25	80.25 ± 0.18	78.62 ± 0.23	80.05 ± 0.22	80.22 ± 0.21
Movement_libras	81.12 ± 0.96	74.16 ± 1.07	79.77 ± 1.18	78.81 ± 1.09	78.26 ± 1.04	79.23 ± 1.18	76.82 ± 1.08	78.12 ± 0.86
Musk	82.10 ± 0.81	79.88 ± 0.49	80.61 ± 0.82	82.01 ± 0.59	81.11 ± 0.54	81.77 ± 0.60	80.11 ± 0.55	81.92 ± 0.73
Mfeat_fac	92.18 ± 0.12	90.75 ± 0.17	91.50 ± 0.18	91.08 ± 0.21	91.82 ± 0.18	91.41 ± 0.15	91.52 ± 0.15	92.13 ± 0.08
Mfeat_pix	81.60 ± 0.19	70.40 ± 0.20	79.86 ± 0.16	68.65 ± 0.35	76.95 ± 0.19	69.90 ± 0.33	75.68 ± 0.28	77.72 ± 0.21
Semeion	71.59 ± 0.33	56.89 ± 0.20	69.39 ± 0.50	59.28 ± 0.65	61.25 ± 0.17	57.70 ± 0.49	62.38 ± 0.38	62.28 ± 0.16
USPS	87.32 ± 0.10	80.14 ± 0.08	85.62 ± 0.08	84.02 ± 0.09	83.09 ± 0.08	82.30 ± 0.21	82.19 ± 0.13	84.79 ± 0.07
lung_small	73.87 ± 1.82	69.44 ± 1.64	71.85 ± 1.41	63.46 ± 2.13	69.82 ± 2.14	64.07 ± 2.31	70.26 ± 1.41	71.73 ± 2.00
Isolet	75.97 ± 0.22	66.33 ± 0.33	73.84 ± 0.33	68.10 ± 0.33	75.11 ± 0.44	66.83 ± 0.31	67.65 ± 0.36	71.91 ± 0.30
ORL	82.14 ± 0.73	76.83 ± 0.94	81.51 ± 0.91	80.47 ± 0.91	80.06 ± 0.70	80.95 ± 0.84	78.97 ± 0.93	78.46 ± 0.66
COIL20	96.09 ± 0.14	95.26 ± 0.13	95.97 ± 0.11	95.36 ± 0.13	95.66 ± 0.07	95.48 ± 0.10	95.56 ± 0.12	95.83 ± 0.12
warpPIE10P	95.30 ± 0.50	91.77 ± 0.60	93.49 ± 0.36	93.85 ± 0.64	93.12 ± 0.35	93.60 ± 0.71	92.26 ± 0.39	92.89 ± 0.46
lung	92.16 ± 0.61	92.29 ± 0.69	92.20 ± 0.70	89.64 ± 1.22	91.77 ± 0.92	88.73 ± 0.74	91.83 ± 0.75	92.93 ± 0.47
gisette	91.20 ± 0.14	91.68 ± 0.06	90.73 ± 0.12	91.11 ± 0.08	92.28 ± 0.10	91.24 ± 0.20	90.22 ± 0.11	91.80 ± 0.10
TOX_171	82.93 ± 2.03	79.70 ± 2.24	81.99 ± 1.42	83.22 ± 2.23	84.59 ± 1.38	82.12 ± 2.37	76.63 ± 1.81	80.45 ± 2.16
arcene	81.57 ± 0.60	79.70 ± 1.28	81.53 ± 0.56	80.24 ± 1.16	81.36 ± 0.68	80.88 ± 1.68	80.07 ± 1.03	81.52 ± 0.59
pixraw10P	95.75 ± 1.08	95.40 ± 0.81	95.02 ± 1.03	81.58 ± 2.62	88.80 ± 0.99	86.07 ± 1.38	95.05 ± 0.73	96.41 ± 0.74
orlraws10P	91.71 ± 0.91	90.94 ± 1.29	91.51 ± 1.07	82.03 ± 1.65	86.64 ± 1.65	81.32 ± 0.93	91.48 ± 1.00	91.83 ± 0.83
Avg.	85.62	81.88	84.70	81.04	83.31	81.25	82.67	84.06
W/T/L	–	15/2/3	13/6/1	17/3/0	15/2/3	15/4/1	16/2/2	12/4/4

Table 5 Margins (mean ± std. dev.) between the performance of algorithms when using Naïve Bayes

Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	82.45 ± 0.17	86.38 ± 0.08	86.04 ± 0.09	65.80 ± 0.13	83.95 ± 0.14	72.40 ± 0.21	86.18 ± 0.10	84.23 ± 0.09
Synthetic_control	94.51 ± 0.12	88.94 ± 0.36	94.71 ± 0.29	93.52 ± 0.45	93.93 ± 0.36	93.67 ± 0.46	85.85 ± 0.28	93.39 ± 0.27
Mfeat_fou	76.71 ± 0.20	76.35 ± 0.16	76.26 ± 0.17	73.42 ± 0.31	76.41 ± 0.18	73.69 ± 0.27	76.23 ± 0.16	76.45 ± 0.18
Movement_libras	56.15 ± 0.80	47.62 ± 1.07	54.77 ± 0.91	53.08 ± 0.95	54.07 ± 1.06	52.72 ± 1.01	48.36 ± 1.04	52.62 ± 0.89
Musk	79.39 ± 0.67	77.78 ± 0.39	78.97 ± 0.57	75.69 ± 0.71	78.90 ± 0.50	77.61 ± 0.77	78.11 ± 0.51	79.38 ± 0.59
Mfeat_fac	90.99 ± 0.15	86.33 ± 0.21	88.67 ± 0.11	85.83 ± 0.23	88.18 ± 0.16	86.78 ± 0.23	87.88 ± 0.17	89.76 ± 0.07
Mfeat_pix	85.88 ± 0.19	76.21 ± 0.17	82.31 ± 0.13	72.42 ± 0.20	78.42 ± 0.17	72.48 ± 0.17	79.10 ± 0.32	81.42 ± 0.10
Semeion	73.71 ± 0.24	61.64 ± 0.13	71.89 ± 0.36	56.39 ± 0.87	65.32 ± 0.15	55.08 ± 0.44	64.79 ± 0.21	65.89 ± 0.13
USPS	84.09 ± 0.06	75.08 ± 0.08	78.60 ± 0.10	73.88 ± 0.16	75.66 ± 0.07	74.41 ± 0.16	75.21 ± 0.11	81.75 ± 0.05
lung_small	77.92 ± 1.99	72.85 ± 1.40	75.66 ± 1.96	72.00 ± 3.24	75.11 ± 2.01	71.70 ± 3.52	74.36 ± 1.49	75.97 ± 2.05
Isolet	77.72 ± 0.29	63.56 ± 0.40	71.98 ± 0.37	69.27 ± 0.33	73.66 ± 0.54	68.79 ± 0.48	64.37 ± 0.67	70.57 ± 0.50
ORL	73.47 ± 1.02	59.66 ± 0.81	70.09 ± 0.48	70.63 ± 0.85	70.85 ± 0.94	70.24 ± 0.90	61.42 ± 0.78	65.39 ± 0.89
COIL20	94.30 ± 0.22	90.58 ± 0.39	93.59 ± 0.24	89.00 ± 0.40	92.39 ± 0.31	90.69 ± 0.31	91.72 ± 0.31	93.84 ± 0.16
warpPIE10P	90.51 ± 0.81	82.03 ± 0.78	86.51 ± 1.02	84.06 ± 1.09	85.33 ± 1.28	83.86 ± 1.29	82.00 ± 0.80	86.20 ± 1.09
lung	93.49 ± 0.60	92.31 ± 0.54	93.05 ± 0.61	91.44 ± 1.16	93.26 ± 1.16	90.98 ± 1.16	92.31 ± 0.58	93.47 ± 0.48
gisette	89.55 ± 0.18	86.85 ± 0.03	88.48 ± 0.04	84.17 ± 0.10	87.67 ± 0.11	86.22 ± 0.18	86.04 ± 0.05	88.67 ± 0.04
TOX_171	68.74 ± 1.35	63.68 ± 1.40	68.11 ± 1.90	66.89 ± 2.46	68.64 ± 1.95	66.51 ± 2.46	61.02 ± 1.46	66.14 ± 1.82
arcene	73.59 ± 1.24	71.06 ± 0.57	70.76 ± 0.41	71.43 ± 1.16	74.00 ± 1.27	72.62 ± 1.54	70.28 ± 0.51	73.83 ± 1.23
pixraw10P	97.51 ± 0.86	96.75 ± 0.74	96.84 ± 0.73	89.41 ± 1.93	91.44 ± 1.82	87.43 ± 2.26	96.41 ± 0.95	97.32 ± 1.02
orlraws10P	93.16 ± 1.14	90.67 ± 1.46	92.77 ± 1.27	79.71 ± 1.85	87.40 ± 2.11	79.36 ± 1.34	90.82 ± 1.45	92.84 ± 0.96
Avg.	82.69	77.32	81.00	75.90	79.73	76.36	77.62	80.46
W/T/L	–	19/0/1	14/4/2	20/0/0	16/3/1	19/1/0	19/0/1	14/5/1

Table 6 Classification accuracy (%) of the selected optimal features when using J48

Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	93.07 (49)	92.98 (41)	**93.16** (26)	91.89 (45)	93.13 (19)	93.10 (47)	92.96 (43)	93.12 (47)
Synthetic_control	93.70 (31)	93.43 (50)	93.77 (28)	92.13 (40)	93.20 (48)	92.93 (49)	93.77 (50)	**93.80** (43)
Mfeat_fou	77.44 (23)	77.76 (13)	77.77 (10)	74.25 (28)	77.77 (10)	74.71 (19)	77.72 (18)	**77.83** (18)
Movement_libras	**69.14** (31)	65.50 (49)	66.89 (45)	65.58 (43)	65.92 (50)	65.69 (38)	66.50 (50)	65.50 (50)
Musk	82.98 (48)	82.38 (50)	**83.39** (50)	80.91 (40)	82.96 (42)	82.72 (32)	83.09 (48)	83.11 (43)
Mfeat_fac	**89.30** (49)	88.65 (49)	88.41 (50)	85.20 (49)	88.46 (27)	86.99 (28)	88.71 (48)	88.94 (26)
Mfeat_pix	79.02 (44)	78.20 (48)	77.98 (50)	75.32 (48)	78.46 (26)	77.35 (48)	**79.03** (44)	78.27 (49)
Semeion	**76.43** (48)	71.49 (50)	76.01 (50)	67.28 (48)	73.08 (50)	66.76 (32)	74.62 (49)	72.42 (50)
USPS	**88.77** (50)	85.10 (50)	88.39 (50)	86.03 (46)	87.36 (50)	86.65 (50)	85.57 (50)	87.97 (49)
lung_small	47.39 (15)	**48.05** (30)	45.70 (14)	44.14 (45)	44.61 (50)	42.30 (43)	45.71 (20)	47.55 (18)
Isolet	79.45 (40)	68.09 (50)	79.13 (50)	71.02 (48)	**80.35** (50)	69.57 (50)	72.22 (49)	74.90 (50)
ORL	55.68 (46)	**57.65** (47)	55.85 (50)	55.25 (49)	56.22 (32)	53.97 (42)	56.43 (38)	56.88 (43)
COIL20	**93.38** (31)	90.99 (49)	91.68 (48)	88.90 (50)	90.83 (48)	89.56 (50)	91.61 (40)	92.23 (25)
warpPIE10P	81.43 (25)	82.38 (37)	82.48 (41)	82.48 (36)	83.86 (50)	**84.57** (47)	82.52 (45)	81.67 (18)
lung	91.04 (27)	90.53 (45)	**91.56** (48)	87.19 (7)	87.90 (49)	87.44 (3)	89.34 (43)	91.32 (6)
gisette	**94.22** (50)	93.33 (50)	93.87 (50)	93.50 (43)	93.92 (45)	93.23 (50)	92.82 (48)	93.87 (41)
TOX_171	63.12 (25)	61.83 (30)	62.83 (32)	60.75 (16)	**65.52** (35)	63.44 (48)	60.05 (8)	61.74 (49)
arcene	79.95 (25)	75.85 (22)	75.35 (20)	76.65 (19)	78.30 (14)	77.70 (25)	74.60 (13)	**80.40** (24)
pixraw10P	92.40 (9)	91.90 (14)	92.20 (28)	91.10 (2)	92.00 (2)	92.00 (8)	**94.00** (23)	93.30 (4)
orlraws10P	77.40 (20)	75.30 (38)	78.00 (28)	73.30 (48)	76.30 (50)	74.30 (50)	75.50 (39)	**79.40** (33)
Avg.	**80.27**	78.57	79.72	77.14	79.51	77.75	78.84	79.71

Bold indicates the better values

Table 7 Classification accuracy (%) of the selected optimal features when using IB1

Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	90.70 (44)	90.65 (50)	90.55 (50)	88.87 (47)	**90.92** (24)	89.96 (31)	90.62 (50)	90.74 (44)
Synthetic_control	97.88 (50)	98.45 (49)	98.42 (42)	95.83 (50)	97.17 (50)	96.07 (50)	**98.47** (48)	97.98 (46)
Mfeat_fou	83.29 (21)	84.07 (12)	84.15 (18)	82.09 (18)	84.13 (16)	82.45 (19)	84.01 (16)	**84.33** (17)
Movement_libras	**85.19** (36)	82.72 (48)	83.86 (22)	84.17 (50)	83.92 (49)	84.17 (50)	83.64 (50)	82.94 (50)
Musk	85.76 (47)	83.57 (50)	83.66 (49)	**85.92** (50)	83.86 (42)	84.60 (41)	83.09 (50)	85.48 (50)
Mfeat_fac	**96.40** (31)	95.60 (50)	96.11 (42)	95.75 (48)	96.02 (49)	95.93 (48)	96.12 (38)	96.33 (42)
Mfeat_pix	**93.92** (49)	89.34 (50)	91.70 (50)	84.42 (50)	90.85 (50)	80.87 (50)	90.17 (50)	91.14 (47)
Semeion	**86.55** (50)	76.69 (50)	85.08 (49)	70.62 (50)	81.43 (50)	69.15 (47)	81.73 (50)	78.47 (50)
USPS	**95.89** (50)	87.97 (50)	95.63 (50)	92.51 (50)	93.33 (50)	92.24 (50)	90.85 (49)	93.26 (50)
lung_small	**83.54** (50)	77.02 (41)	82.18 (48)	74.25 (49)	77.54 (50)	74.70 (50)	78.98 (48)	80.61 (50)
Isolet	85.88 (48)	75.62 (50)	**87.10** (50)	77.56 (50)	86.18 (45)	74.89 (50)	77.31 (50)	81.72 (50)
ORL	**93.28** (49)	88.23 (50)	92.37 (48)	90.90 (50)	90.40 (50)	91.85 (50)	90.23 (50)	89.12 (50)
COIL20	99.86 (47)	99.51 (48)	**99.93** (48)	99.27 (49)	99.44 (49)	99.31 (50)	99.88 (47)	99.72 (50)
warpPIE10P	**99.33** (37)	95.81 (48)	97.67 (46)	97.95 (50)	97.81 (50)	97.76 (49)	96.67 (47)	96.76 (50)
lung	93.89 (22)	94.81 (43)	94.04 (47)	92.42 (44)	93.77 (41)	90.56 (50)	94.32 (50)	**94.97** (40)
gisette	94.03 (49)	93.69 (50)	93.88 (50)	93.40 (50)	**95.00** (50)	93.45 (50)	93.03 (48)	93.79 (34)
TOX_171	92.99 (50)	87.56 (46)	90.17 (50)	92.58 (50)	**94.57** (44)	92.40 (50)	84.92 (50)	87.67 (50)
arcene	83.55 (50)	82.00 (47)	**84.90** (41)	83.00 (50)	84.15 (50)	83.00 (36)	82.75 (50)	84.10 (49)
pixraw10P	97.80 (46)	97.70 (45)	96.80 (47)	90.00 (2)	90.90 (35)	89.20 (49)	96.90 (43)	**98.00** (38)
orlraws10P	**96.30** (43)	94.40 (50)	96.10 (50)	86.70 (38)	89.30 (18)	85.10 (47)	95.10 (43)	95.30 (50)
Avg.	**91.80**	88.77	91.21	87.91	90.03	87.38	89.44	90.12

Bold indicates the better values

Table 8 Classification accuracy (%) of the selected optimal features when using Naive Bayes

Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	90.97 (12)	90.80 (28)	91.01 (26)	84.36 (5)	91.25 (18)	86.42 (4)	90.81 (36)	**91.64** (19)
Synthetic_control	97.37 (26)	95.83 (48)	**97.57** (23)	96.75 (29)	96.93 (42)	96.83 (49)	95.27 (50)	96.48 (45)
Mfeat_fou	**79.42** (21)	79.17 (19)	78.89 (24)	75.92 (18)	79.18 (20)	76.64 (18)	79.24 (19)	79.32 (23)
Movement_libras	**61.19** (43)	56.08 (50)	59.53 (50)	59.81 (50)	59.33 (50)	59.36 (50)	56.78 (50)	56.86 (41)
Musk	**82.96** (50)	79.87 (50)	80.70 (42)	77.48 (50)	81.68 (49)	79.96 (50)	80.51 (50)	82.82 (49)
Mfeat_fac	**95.61** (50)	90.85 (48)	93.39 (48)	89.91 (48)	91.73 (23)	90.91 (44)	92.36 (35)	93.78 (50)
Mfeat_pix	**92.54** (50)	86.34 (50)	87.62 (28)	82.87 (50)	85.02 (50)	78.87 (50)	86.87 (50)	88.52 (50)
Semeion	**83.84** (50)	73.57 (50)	82.12 (50)	60.23 (50)	76.96 (50)	59.43 (32)	77.01 (50)	75.78 (50)
USPS	**88.05** (47)	79.37 (50)	83.21 (50)	79.48 (50)	81.19 (50)	79.82 (50)	78.65 (50)	85.30 (31)
lung_small	**87.46** (47)	81.86 (50)	85.71 (49)	82.77 (50)	83.88 (50)	81.98 (50)	83.23 (50)	84.43 (48)
Isolet	**88.37** (50)	70.11 (50)	82.27 (50)	78.91 (50)	83.20 (50)	77.47 (50)	71.10 (38)	78.12 (50)
ORL	**88.10** (50)	73.45 (50)	83.52 (48)	84.45 (50)	83.53 (49)	84.53 (50)	74.03 (50)	78.33 (50)
COIL20	**98.76** (50)	95.81 (50)	98.62 (49)	95.09 (50)	96.76 (49)	96.06 (49)	96.95 (50)	98.39 (50)
warpPIE10P	**97.71** (50)	90.43 (49)	92.43 (40)	90.86 (50)	91.33 (49)	88.95 (50)	89.95 (50)	92.05 (50)
lung	**95.32** (47)	93.88 (34)	94.86 (42)	94.03 (48)	94.86 (50)	93.10 (50)	94.13 (30)	94.76 (42)
gisette	**91.34** (50)	87.53 (50)	90.67 (50)	86.81 (2)	88.31 (9)	87.26 (17)	86.81 (2)	89.54 (39)
TOX_171	74.68 (49)	67.77 (50)	73.61 (50)	73.46 (48)	**75.66** (46)	73.51 (50)	67.38 (50)	72.27 (48)
arcene	75.15 (46)	71.95 (26)	71.95 (16)	72.40 (23)	**75.75** (50)	73.75 (37)	71.50 (9)	75.05 (49)
pixraw10P	**99.10** (48)	97.70 (26)	99.00 (50)	90.40 (5)	93.20 (29)	89.00 (36)	97.70 (8)	98.50 (22)
orlraws10P	**98.30** (46)	95.60 (46)	97.80 (43)	83.90 (33)	91.70 (25)	82.40 (28)	95.30 (45)	97.70 (45)
Avg.	**88.31**	82.90	86.22	81.99	85.07	81.81	83.28	85.48

Bold indicates the better values

Table 9 Running time (s) of EID-CMI and other algorithms

| Datasets	EID-CMI	JMI	CMIM	CIFE	DWFS	IWFS	JMIM	RelaxMRMR
Spambase	7.35	0.56	0.49	0.53	0.73	0.75	0.59	7.38
Synthetic_control	1.32	0.11	0.10	0.11	0.14	0.15	0.12	1.36
Mfeat_fou	4.87	0.44	0.39	0.43	0.58	0.58	0.48	4.95
Movement_libras	2.03	0.13	0.10	0.12	0.17	0.17	0.13	2.08
Musk	5.27	0.23	0.16	0.21	0.33	0.33	0.25	5.34
Mfeat_fac	19.11	0.86	0.70	0.84	1.30	1.30	1.00	19.55
Mfeat_pix	23.08	0.74	0.52	0.70	1.28	1.28	0.90	24.12
Semeion	19.89	0.64	0.43	0.60	1.10	1.10	0.77	20.54
USPS	130.70	7.41	6.45	7.07	9.19	9.23	7.82	127.57
lung_small	7.07	0.25	0.15	0.22	0.38	0.37	0.27	7.23
Isolet	51.68	3.17	2.71	3.06	4.35	4.35	3.52	52.56
ORL	36.17	1.67	1.30	1.56	2.49	2.49	1.92	37.88
COIL20	94.50	5.04	4.40	4.80	6.78	6.77	5.56	101.10
warpPIE10P	69.90	2.77	1.90	2.44	3.97	3.94	3.00	71.62
lung	96.01	3.77	2.60	3.34	5.47	5.43	4.10	97.41
gisette	2263.15	66.04	51.60	61.18	101.36	100.36	73.24	2335.22
TOX_171	157.68	6.01	3.97	5.29	8.61	8.57	6.40	156.69
arcene	292.25	10.25	6.73	9.04	14.80	14.79	10.88	283.98
pixraw10P	278.15	10.40	6.91	8.79	14.37	14.21	10.99	273.19
orlraws10P	272.35	10.58	7.15	9.21	14.66	14.47	11.26	276.07
Avg.	191.63	6.55	4.94	5.98	9.60	9.53	7.16	195.29
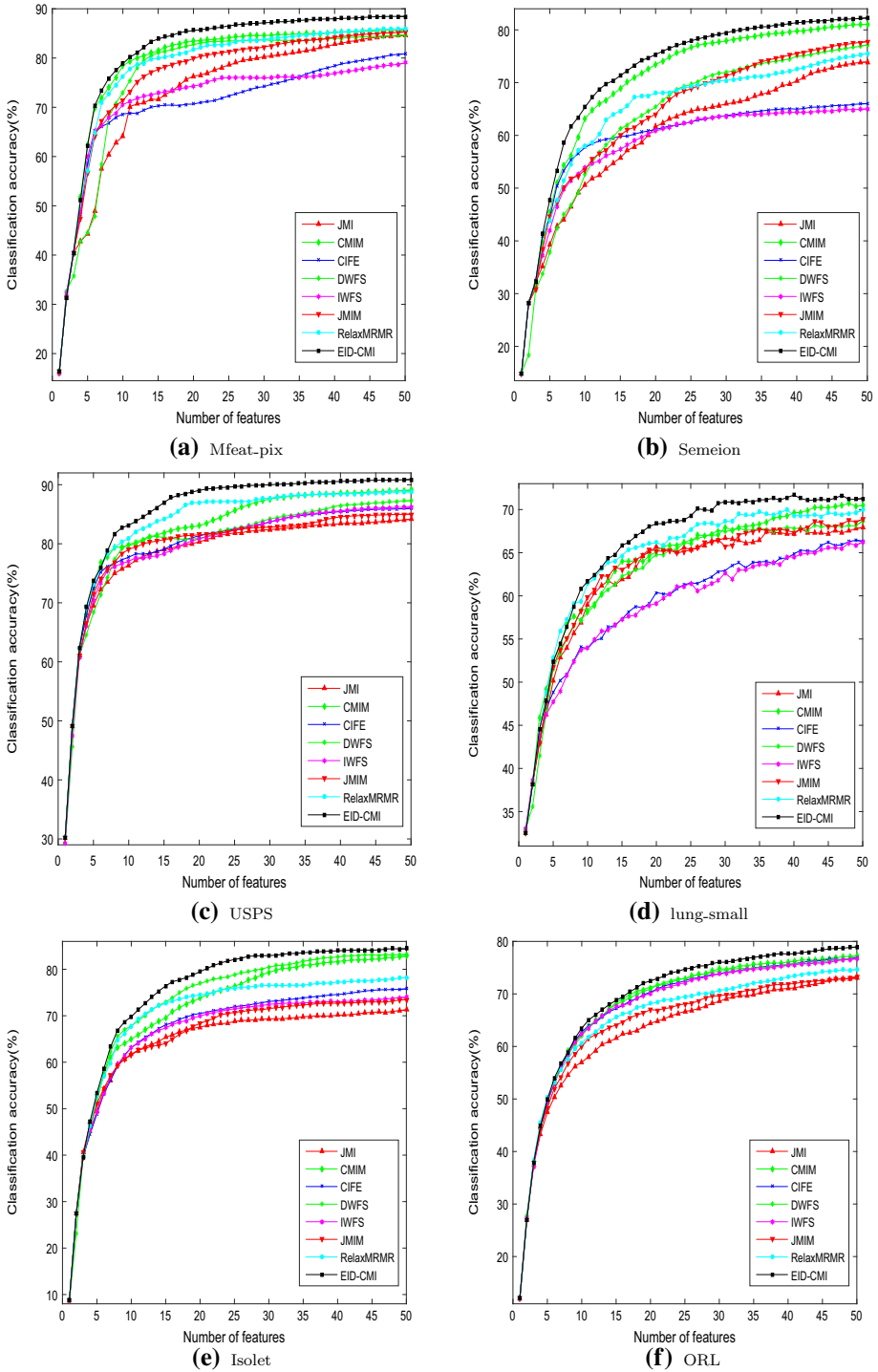
Fig. 1 Average performance comparisons of algorithms with the three classifiers

performance in Movement_libras, Musk, Mfeat_fac, Semeion and ORL. RelaxMRMR does not achieve the desired results in Movement_libras, Semeion and ORL.

In addition to these algorithms above, EID-CMI is compared with the algorithm in [27], MCMI [28], MRI [29], MIM [6] and EID-mRMR [12]. In the algorithm in [27], the parameter c is set to 0.4 and η is set to 0.2. In EID-mRMR, the parameter P is set to 0.5. The accuracy of the features selected by EID-CMI and other algorithms with J48, IB1 and Naive Bayes is presented in Tables 10, 11 and 12. The accuracy of the optimal features selected by EID-CMI and other algorithms when using J48, IB1 and Naive Bayes is presented in Tables 13, 14 and 15. The running time of EID-CMI and other algorithms that select 50 features is given in Table 16.

In Table 10, EID-mRMR and EID-CMI can achieve better Avg. and W/T/L values. The number of the datasets that EID-CMI is superior to EID-mRMR is 8, while that of the datasets that EID-CMI is inferior to EID-mRMR is 2, as shown, EID-CMI can achieve better feature selection performance than EID-mRMR.

As shown in Table 11, EID-mRMR and EID-CMI can obtain better Avg. and W/T/L values. Compared with Table 10, in terms of the Avg. values, EID-CMI outperforms other algorithms except EID-mRMR with more performance gain. In the W/T/L values, EID-CMI achieves more advantages than the algorithm in [27], MRI and EID-mRMR.

The Avg. and W/T/L values in Table 12 show that EID-mRMR and EID-CMI achieve better feature selection performances. Different from Tables 10 and 11, in the Avg. values, EID-CMI gains more advantages than other algorithms except for EID-mRMR. In terms of the W/T/L values, EID-CMI performs better than MCMI, MRI and MIM.

Tables 13, 14 and 15 show that EID-CMI and EID-mRMR perform better than other algorithms in general. Take Table 13 for example, in the twenty datasets, the number of the datasets that EID-CMI can obtain better results than other algorithms is 7. Furthermore, EID-CMI and EID-mRMR obtain greater Avg. values with the three classifiers.

In Table 16, considering that both conditional mutual information among features and the class label and conditional mutual information among features are calculated, EID-CMI takes more time than the other algorithms in feature selection. Since it only calculates mutual information between features and the class label, MIM spends the least time.

Compared with other algorithms, EID-CMI and RelaxMRMR consider three-dimensional mutual information among features. Due to considering three-dimensional mutual information among features, the objective functions contain more useful information and they can achieve better results than the majority of algorithms at the cost of more time-consuming. As EID-CMI adopts the objective function of RelaxMRMR to select features when the objective function can guarantee the performance of selected features, and combines the objective function with equal interval division and ranking to select features when the objective function cannot guarantee the performance of selected features, EID-CMI can obtain better feature selection performance than RelaxMRMR.

6 Conclusions and Future Work

This paper introduces a method of equal interval division and ranking to investigate feature selection based on three-dimensional mutual information and proposes a feature selection algorithm named EID-CMI. To verify the performance, EID-CMI is applied to eight UCI datasets, twelve ASU datasets and three classifiers, and compared with several feature selec-

Table 10 Margins (mean ± std. dev.) between the performance of algorithms when using J48

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	92.02 ± 0.08	92.06 ± 0.10	91.75 ± 0.08	92.08 ± 0.06	91.79 ± 0.07	91.94 ± 0.07
Synthetic_control	91.10 ± 0.48	90.42 ± 0.31	84.20 ± 0.58	90.86 ± 0.49	82.52 ± 0.48	89.89 ± 0.51
Mfeat_fou	75.13 ± 0.47	75.26 ± 0.52	75.28 ± 0.53	75.31 ± 0.52	75.25 ± 0.49	74.93 ± 0.45
Movement_libras	64.08 ± 0.77	60.24 ± 0.93	54.84 ± 1.12	59.06 ± 1.01	52.47 ± 1.08	63.74 ± 0.82
Musk	80.42 ± 0.80	80.47 ± 0.89	78.17 ± 0.56	78.80 ± 0.54	77.26 ± 0.47	80.69 ± 0.69
Mfeat_fac	85.91 ± 0.19	85.63 ± 0.31	82.49 ± 0.28	85.53 ± 0.33	82.22 ± 0.25	85.96 ± 0.30
Mfeat_pix	74.79 ± 0.18	73.22 ± 0.42	64.52 ± 0.32	73.37 ± 0.35	62.35 ± 0.21	74.51 ± 0.45
Semeion	70.00 ± 0.50	63.86 ± 0.23	58.61 ± 0.12	63.72 ± 0.21	57.22 ± 0.17	69.89 ± 0.50
USPS	84.39 ± 0.14	83.50 ± 0.12	72.88 ± 0.08	80.97 ± 0.07	64.67 ± 0.07	84.21 ± 0.16
lung_small	45.25 ± 3.38	45.43 ± 4.99	43.57 ± 3.89	41.55 ± 4.41	43.64 ± 3.48	43.34 ± 5.06
Isolet	72.08 ± 0.33	68.75 ± 0.58	53.53 ± 0.56	67.46 ± 0.29	49.53 ± 0.58	71.56 ± 0.59
ORL	51.79 ± 1.73	52.33 ± 1.49	49.94 ± 1.05	52.43 ± 1.35	47.83 ± 1.09	51.65 ± 1.95
COIL20	90.19 ± 0.17	89.26 ± 0.38	83.67 ± 0.31	87.91 ± 0.34	81.60 ± 0.28	89.56 ± 0.36
warpPIE10P	78.39 ± 0.89	78.10 ± 1.41	77.80 ± 0.70	79.96 ± 1.50	76.77 ± 1.12	79.05 ± 0.78
lung	89.96 ± 1.02	89.11 ± 1.19	86.33 ± 1.64	89.42 ± 1.05	85.90 ± 1.36	89.74 ± 1.14
gisette	92.76 ± 0.06	92.50 ± 0.07	92.33 ± 0.04	92.86 ± 0.11	90.19 ± 0.07	92.42 ± 0.10
TOX_171	60.47 ± 2.31	58.93 ± 1.48	58.79 ± 1.78	60.22 ± 1.83	55.91 ± 2.08	60.71 ± 1.73
arcene	77.67 ± 1.46	77.62 ± 1.58	74.02 ± 1.67	73.96 ± 1.92	72.20 ± 1.29	77.49 ± 1.41
pixraw10P	91.37 ± 0.96	88.98 ± 1.52	91.64 ± 1.53	88.92 ± 2.50	87.76 ± 1.93	89.31 ± 2.14
orlraws10P	75.77 ± 2.39	75.82 ± 2.53	71.19 ± 2.89	75.40 ± 3.46	65.45 ± 1.74	78.08 ± 1.91
Avg.	77.18	76.07	72.28	75.49	70.13	76.93
W/T/L	–	11/7/2	16/3/1	12/4/4	18/1/1	8/10/2

Table 11 Margins (mean ± std. dev.) between the performance of algorithms when using IB1

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	84.56 ± 0.11	84.26 ± 0.18	86.04 ± 0.15	85.81 ± 0.12	85.49 ± 0.11	84.43 ± 0.15
Synthetic_control	93.47 ± 0.28	92.32 ± 0.22	87.83 ± 0.25	93.45 ± 0.15	87.52 ± 0.27	92.66 ± 0.28
Mfeat_fou	79.75 ± 0.16	80.13 ± 0.19	80.09 ± 0.19	80.21 ± 0.21	80.10 ± 0.19	79.62 ± 0.20
Movement_libras	81.12 ± 0.96	77.44 ± 0.95	67.65 ± 0.93	76.21 ± 1.17	67.16 ± 0.95	80.75 ± 1.03
Musk	82.10 ± 0.81	82.00 ± 0.70	79.44 ± 0.68	80.10 ± 0.55	79.20 ± 0.51	82.28 ± 0.74
Mfeat_fac	92.18 ± 0.12	92.03 ± 0.06	88.48 ± 0.15	91.71 ± 0.12	88.58 ± 0.09	92.26 ± 0.14
Mfeat_pix	81.60 ± 0.19	76.04 ± 0.27	59.95 ± 0.32	75.81 ± 0.17	57.62 ± 0.15	81.36 ± 0.29
Semeion	71.59 ± 0.33	60.03 ± 0.22	52.62 ± 0.32	60.11 ± 0.25	52.38 ± 0.21	71.20 ± 0.37
USPS	87.32 ± 0.10	85.38 ± 0.06	67.91 ± 0.08	82.18 ± 0.08	62.25 ± 0.12	87.13 ± 0.07
lung_small	73.87 ± 1.82	71.44 ± 1.85	65.18 ± 2.09	70.88 ± 1.81	62.29 ± 1.40	72.90 ± 2.34
Isolet	75.97 ± 0.22	71.88 ± 0.26	55.15 ± 0.41	70.79 ± 0.42	51.25 ± 0.56	75.41 ± 0.33
ORL	82.14 ± 0.73	78.33 ± 0.54	74.12 ± 0.51	78.17 ± 0.60	71.60 ± 0.47	81.54 ± 0.80
COIL20	96.09 ± 0.14	95.91 ± 0.11	91.44 ± 0.38	95.94 ± 0.09	87.73 ± 0.19	96.09 ± 0.09
warpPIE10P	95.30 ± 0.50	93.05 ± 0.37	90.51 ± 0.78	92.30 ± 0.43	88.84 ± 0.56	94.81 ± 0.25
lung	92.16 ± 0.61	93.04 ± 0.50	89.49 ± 0.55	92.70 ± 0.86	89.87 ± 0.70	92.05 ± 0.42
gisette	91.20 ± 0.14	90.78 ± 0.19	91.48 ± 0.07	92.70 ± 0.05	89.06 ± 0.09	91.29 ± 0.16
TOX_171	82.93 ± 2.03	80.14 ± 1.73	73.51 ± 1.47	82.00 ± 1.71	68.18 ± 1.26	83.19 ± 1.74
arcene	81.57 ± 0.60	81.12 ± 0.66	78.56 ± 1.21	80.27 ± 1.46	76.84 ± 1.14	80.77 ± 0.78
pixraw10P	95.75 ± 1.08	95.89 ± 0.87	92.98 ± 1.75	94.76 ± 1.00	90.61 ± 0.99	96.29 ± 0.67
orlraws10P	91.71 ± 0.91	92.06 ± 0.99	87.85 ± 2.02	89.71 ± 1.44	81.26 ± 2.08	92.60 ± 1.21
Avg.	85.62	83.66	78.01	83.29	75.89	85.43
W/T/L	–	14/4/2	17/0/3	14/3/3	18/0/2	11/7/2

Table 12 Margins (mean ± std. dev.) between the performance of algorithms when using Naive Bayes

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	82.45 ± 0.17	82.80 ± 0.10	85.87 ± 0.08	85.49 ± 0.13	85.13 ± 0.10	83.09 ± 0.25
Synthetic_control	94.51 ± 0.12	93.06 ± 0.26	83.71 ± 0.37	93.86 ± 0.17	81.67 ± 0.24	94.24 ± 0.20
Mfeat_fou	76.71 ± 0.20	76.41 ± 0.17	76.33 ± 0.16	76.33 ± 0.17	76.28 ± 0.17	76.64 ± 0.24
Movement_libras	56.15 ± 0.80	52.62 ± 0.96	41.48 ± 0.84	51.76 ± 1.05	40.20 ± 0.54	56.18 ± 1.49
Musk	79.39 ± 0.67	79.65 ± 0.33	75.98 ± 0.91	78.48 ± 0.44	76.29 ± 0.43	79.43 ± 0.53
Mfeat_fac	90.99 ± 0.15	89.90 ± 0.15	82.63 ± 0.33	88.62 ± 0.14	82.55 ± 0.20	90.90 ± 0.10
Mfeat_pix	85.88 ± 0.19	80.01 ± 0.16	65.99 ± 0.45	78.97 ± 0.10	63.89 ± 0.17	85.89 ± 0.12
Semeion	73.71 ± 0.24	64.32 ± 0.10	57.51 ± 0.12	64.26 ± 0.22	56.63 ± 0.13	73.76 ± 0.39
USPS	84.09 ± 0.06	82.15 ± 0.07	60.39 ± 0.06	76.03 ± 0.10	55.59 ± 0.07	83.96 ± 0.08
lung_small	77.92 ± 1.99	74.92 ± 1.83	68.80 ± 2.14	75.07 ± 2.11	64.59 ± 1.49	76.83 ± 2.12
Isolet	77.72 ± 0.29	70.95 ± 0.50	50.54 ± 0.48	70.02 ± 0.45	46.65 ± 0.60	76.70 ± 0.29
ORL	73.47 ± 1.02	66.12 ± 0.88	52.08 ± 1.13	65.85 ± 1.01	48.71 ± 0.97	72.86 ± 1.20
COIL20	94.30 ± 0.22	93.92 ± 0.13	81.21 ± 1.15	93.21 ± 0.23	78.70 ± 0.31	94.45 ± 0.13
warpPIE10P	90.51 ± 0.81	86.82 ± 0.77	77.15 ± 0.88	84.95 ± 0.97	74.26 ± 0.70	90.13 ± 0.82
lung	93.49 ± 0.60	93.62 ± 0.65	88.31 ± 0.50	92.96 ± 0.58	88.19 ± 0.75	93.65 ± 0.80
gisette	89.55 ± 0.18	89.42 ± 0.06	84.99 ± 0.08	87.59 ± 0.02	85.79 ± 0.02	90.20 ± 0.04
TOX_171	68.74 ± 1.35	66.48 ± 1.59	57.94 ± 0.95	67.38 ± 1.86	55.11 ± 0.67	69.73 ± 1.49
arcene	73.59 ± 1.24	73.66 ± 1.26	69.74 ± 0.67	71.17 ± 0.97	69.54 ± 0.45	73.13 ± 1.55
pixraw10P	97.51 ± 0.86	98.20 ± 0.34	90.84 ± 0.75	97.27 ± 0.44	90.93 ± 0.94	98.16 ± 0.15
orlraws10P	93.16 ± 1.14	93.26 ± 1.22	79.64 ± 2.94	93.01 ± 1.35	70.25 ± 2.56	93.40 ± 1.42
Avg.	82.69	80.41	71.56	79.61	69.55	82.67
W/T/L	–	14/4/2	19/0/1	17/2/1	19/0/1	8/8/4

Table 13 Classification accuracy (%) of the selected optimal features when using J48

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	93.07 (49)	93.09 (19)	92.95 (36)	**93.13** (21)	92.94 (38)	93.03 (49)
Synthetic_control	**93.70** (31)	93.68 (34)	93.57 (50)	93.63 (35)	93.62 (50)	92.87 (50)
Mfeat_fou	77.44 (23)	77.82 (11)	77.72 (10)	77.68 (10)	**77.84** (10)	77.09 (32)
Movement_libras	**69.14** (31)	66.06 (46)	65.17 (47)	64.17 (50)	64.94 (48)	68.50 (29)
Musk	82.98 (48)	83.39 (27)	82.47 (50)	82.37 (46)	82.15 (50)	**83.41** (48)
Mfeat_fac	**89.30** (49)	88.86 (29)	87.53 (50)	88.77 (49)	86.86 (50)	89.03 (43)
Mfeat_pix	**79.02** (44)	78.55 (47)	78.25 (50)	**79.02** (40)	75.41 (48)	78.81 (49)
Semeion	76.43 (48)	72.58 (50)	68.80 (50)	71.61 (50)	68.29 (50)	**76.75** (50)
USPS	**88.77** (50)	88.32 (50)	81.58 (50)	85.40 (50)	74.45 (49)	88.57 (50)
lung_small	47.39 (15)	**48.27** (25)	46.20 (19)	42.57 (25)	45.64 (14)	45.95 (21)
Isolet	**79.45** (40)	75.48 (48)	61.21 (48)	75.27 (48)	60.91 (45)	79.44 (50)
ORL	55.68 (46)	**56.65** (50)	55.85 (50)	56.30 (40)	54.98 (47)	55.95 (50)
COIL20	**93.38** (31)	92.44 (50)	88.93 (49)	91.27 (42)	88.08 (48)	92.81 (46)
warpPIE10P	81.43 (25)	80.71 (10)	82.19 (43)	**84.00** (39)	82.05 (46)	83.48 (46)
lung	91.04 (27)	90.78 (7)	88.46 (49)	90.73 (27)	87.41 (30)	**91.47** (7)
gisette	94.22 (50)	93.70 (50)	93.72 (47)	93.92 (26)	91.81 (50)	**94.24** (43)
TOX_171	63.12 (25)	61.67 (23)	61.70 (25)	62.91 (20)	59.64 (30)	**63.37** (20)
arcene	79.95 (25)	80.15 (41)	75.60 (37)	75.20 (12)	74.10 (25)	**80.50** (26)
pixraw10P	92.40 (9)	90.80 (5)	**92.80** (42)	90.80 (12)	88.90 (9)	91.40 (8)
orlraws10P	77.40 (20)	77.70 (11)	73.60 (37)	77.30 (42)	69.00 (43)	**81.40** (31)
Avg.	80.27	79.54	77.42	78.80	75.95	**80.40**

Bold indicates the better values

Table 14 Classification accuracy (%) of the selected optimal features when using IB1

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	90.70 (44)	**91.01** (48)	90.64 (39)	90.80 (47)	90.67 (50)	90.70 (50)
Synthetic_control	97.88 (50)	97.87 (49)	98.40 (50)	98.43 (47)	**98.55** (49)	97.58 (49)
Mfeat_fou	83.29 (21)	**84.38** (17)	84.27 (17)	84.09 (16)	84.06 (12)	83.15 (23)
Movement_libras	**85.19** (36)	83.14 (50)	82.39 (49)	82.92 (49)	82.64 (47)	85.03 (37)
Musk	85.76 (47)	85.78 (50)	83.07 (35)	82.98 (50)	83.57 (48)	**86.18** (50)
Mfeat_fac	96.40 (31)	96.38 (48)	93.94 (50)	95.97 (28)	93.77 (50)	**96.48** (30)
Mfeat_pix	93.92 (49)	90.54 (50)	85.86 (50)	90.43 (50)	82.05 (50)	**93.94** (50)
Semeion	**86.55** (50)	78.49 (50)	72.85 (50)	77.08 (50)	75.17 (50)	86.40 (47)
USPS	**95.89** (50)	94.25 (50)	79.81 (50)	90.40 (50)	73.27 (50)	95.74 (50)
lung_small	83.54 (50)	79.84 (44)	77.39 (50)	83.68 (50)	72.27 (50)	**83.70** (48)
Isolet	85.88 (48)	81.75 (50)	65.51 (50)	81.75 (50)	65.60 (50)	**86.10** (50)
ORL	**93.28** (49)	88.83 (50)	86.30 (49)	88.65 (45)	85.00 (50)	93.03 (49)
COIL20	99.86 (47)	99.87 (48)	98.31 (50)	**99.91** (50)	96.65 (50)	99.87 (46)
warpPIE10P	**99.33** (37)	97.05 (46)	95.57 (50)	95.76 (49)	95.14 (49)	99.29 (41)
lung	93.89 (22)	94.82 (47)	92.50 (33)	**94.87** (47)	93.82 (48)	93.80 (8)
gisette	94.03 (49)	93.42 (50)	94.38 (47)	**95.30** (49)	91.98 (49)	93.93 (50)
TOX_171	**92.99** (50)	89.01 (50)	82.72 (49)	90.70 (44)	79.38 (50)	92.45 (50)
arcene	**83.55** (50)	**83.55** (50)	81.70 (50)	82.20 (42)	79.45 (50)	83.35 (24)
pixraw10P	97.80 (46)	**98.80** (45)	94.50 (36)	98.10 (42)	92.20 (44)	98.50 (47)
orlraws10P	96.30 (43)	96.40 (50)	91.80 (50)	93.70 (50)	88.90 (50)	**97.20** (41)
Avg.	91.80	90.26	86.60	89.89	85.21	**91.82**

Bold indicates the better values

Table 15 Classification accuracy (%) of the selected optimal features when using Naïve Bayes

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	90.97 (12)	89.97 (18)	90.61 (34)	**91.00** (17)	90.63 (37)	90.26 (12)
Synthetic_control	97.37 (26)	96.82 (41)	95.53 (50)	97.07 (37)	95.77 (50)	**97.98** (31)
Mfeat_fou	**79.42** (21)	79.15 (18)	79.32 (19)	79.08 (24)	78.96 (26)	79.34 (21)
Movement_libras	**61.19** (43)	57.11 (43)	55.61 (50)	56.86 (32)	56.11 (50)	61.03 (37)
Musk	**82.96** (50)	82.77 (50)	79.49 (49)	80.32 (36)	79.98 (50)	82.77 (50)
Mfeat_fac	**95.61** (50)	94.28 (36)	87.89 (50)	92.74 (36)	87.12 (38)	95.48 (49)
Mfeat_pix	92.54 (50)	88.49 (50)	83.54 (50)	86.72 (50)	81.51 (50)	**92.78** (50)
Semeion	83.84 (50)	75.52 (50)	71.37 (50)	74.29 (50)	72.27 (50)	**83.97** (50)
USPS	88.05 (47)	86.27 (45)	68.99 (50)	80.96 (50)	64.25 (50)	**88.13** (50)
lung_small	87.46 (47)	83.52 (45)	80.07 (50)	85.45 (48)	76.43 (49)	**87.52** (49)
Isolet	**88.37** (50)	79.29 (50)	59.42 (50)	78.35 (49)	59.43 (50)	87.90 (50)
ORL	**88.10** (50)	78.78 (49)	65.63 (50)	78.00 (50)	61.73 (50)	88.05 (50)
COIL20	98.76 (50)	98.52 (50)	89.65 (50)	98.49 (50)	87.18 (50)	**99.12** (50)
warpPIE10P	**97.71** (50)	92.57 (49)	88.52 (50)	90.57 (47)	87.62 (50)	97.19 (39)
lung	95.32 (47)	95.05 (44)	89.98 (50)	94.81 (41)	90.83 (41)	**95.56** (48)
gisette	91.34 (50)	90.24 (47)	87.45 (46)	88.30 (49)	86.75 (43)	**91.82** (50)
TOX_171	74.68 (49)	72.10 (50)	61.03 (45)	72.81 (46)	59.67 (50)	**76.30** (49)
arcene	**75.15** (46)	74.75 (8)	70.50 (18)	72.85 (48)	70.40 (16)	74.30 (19)
pixraw10P	99.10 (48)	**100.00** (32)	93.40 (5)	98.90 (23)	92.40 (7)	99.60 (47)
orlraws10P	98.30 (46)	**98.50** (39)	83.10 (41)	98.30 (47)	74.40 (49)	98.20 (41)
Avg.	88.31	85.69	79.05	84.79	77.67	**88.36**

Bold indicates the better values

Table 16 Running time (s) of EID-CMI and other five algorithms

Datasets	EID-CMI	The algorithm in [27]	MCMI	MRI	MIM	EID-mRMR
Spambase	7.35	0.64	0.48	0.67	0.29	0.36
Synthetic_control	1.32	0.12	0.10	0.13	0.06	0.08
Mfeat_fou	4.87	0.45	0.39	0.51	0.28	0.33
Movement_libras	2.03	0.13	0.10	0.15	0.05	0.08
Musk	5.27	0.23	0.17	0.30	0.05	0.12
Mfeat_fac	19.11	0.86	0.72	1.13	0.28	0.45
Mfeat_pix	23.08	0.72	0.53	1.07	0.01	0.21
Semeion	19.89	0.62	0.44	0.91	0.01	0.20
USPS	130.70	7.56	6.63	9.10	4.16	4.96
lung_small	7.07	0.25	0.15	0.31	0.01	0.11
Isolet	51.68	3.17	2.72	3.87	1.58	2.05
ORL	36.17	1.68	1.31	2.12	0.39	0.93
COIL20	94.50	5.06	4.42	6.32	2.26	3.06
warpPIE10P	69.90	2.72	1.90	3.41	0.42	1.93
lung	96.01	3.73	2.60	4.61	0.64	3.17
gisette	2263.15	65.68	53.08	93.20	10.71	27.74
TOX_171	157.68	5.91	4.01	7.21	0.86	6.69
arcene	292.25	10.11	6.77	12.59	1.23	16.63
pixraw10P	278.15	10.10	6.79	12.42	1.29	18.06
orlraws10P	272.35	10.40	7.02	12.77	1.52	18.43
Avg.	191.63	6.51	5.02	8.64	1.31	5.28

tion algorithms. Experimental results demonstrate that EID-CMI can achieve better feature selection effectiveness.

Since it achieves better feature selection performance, EID-CMI can be applied to some fields, such as text processing [30], network anomaly detecting [31], underwater objects recognition and classification [32,33], steganalysis [34,35] and information retrieval [36]. Considering that IED-CMI is more time-consuming, we will consider both performance and time consumption to investigate feature selection in the next stage.

References

1. Bolon CV, Sanchez MN, Alonso BA (2015) Recent advances and emerging challenges of feature selection in the context of big data. Knowl-Based Syst 86:33–45
2. Borja SP, Veronica BC, Amparo AB (2017) Testing different ensemble configurations for feature selection. Neural Process Lett 46:1–24
3. Estevez PA, Tesmer M, Perez CA, Zurada JM (2009) Normalized mutual information feature selection. IEEE Trans Neural Netw 20(2):189–201
4. Vergara JR, Estevez PA (2014) A review of feature selection methods based on mutual information. Neural Comput Appl 24(1):175–186

5. Brown G, Pocock A, Zhao MJ, Lujan M (2012) Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. J Mach Learn Res 13:27–66

6. Lewis DD (1992) Feature selection and feature extraction for text categorization. In: Proceedings of the Workshop on speech and natural language, pp 212–217

7. Battiti R (1994) Using mutual information for selecting features in supervised neural net learning. IEEE Trans Neural Netw 5(4):537–550

8. Kwak N, Choi CH (2002) Input feature selection for classification problems. IEEE Trans Neural Netw 13(1):143–159

9. Peng HC, Long FH, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238

10. Foithong S, Pinngern O, Attachoo B (2012) Feature subset selection wrapper based on mutual information and rough sets. Expert Syst Appl 39(1):574–584

11. Wang ZC, Li MQ, Li JZ (2015) A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure. Inf Sci 307:73–88

12. Gu XY, Guo JC, Xiao LJ, Ming T, Li CY (2020) A feature selection algorithm based on equal interval division and minimal-redundancy maximal-relevance. Neural Process Lett 51(2):1237–1263

13. Sun X, Liu YH, Xu MT, Chen HL, Han JW, Wang KH (2013) Feature selection using dynamic weights for classification. Knowl-Based Syst 37:541–549

14. Zeng ZL, Zhang HJ, Zhang R, Yin CX (2015) A novel feature selection method considering feature interaction. Pattern Recogn 48(8):2656–2666

15. Yang HH, Moody J (1999) Data visualization and feature selection: new algorithms for non-gaussian data. In: Proceedings of conference on neural information processing systems

16. Fleuret F (2004) Fast binary feature selection with conditional mutual information. J Mach Learn Res 5:1531–1555

17. Lin DH, Tang X (2006) Conditional infomax learning: an integrated framework for feature extraction and fusion. In: Proceedings of European conference on computer vision, pp 68–82

18. Bennasar M, Hicks Y, Setchi R (2015) Feature selection using joint mutual information maximisation. Expert Syst Appl 42(22):8520–8532

19. Vinh NX, Zhou S, Chan J, Bailey J (2016) Can high-order dependencies improve mutual information based feature selection. Pattern Recogn 53:46–58

20. Jakulin A, Bratko T (2004) Testing the significance of attribute interactions. In: Proceedings of international conference on machine learning, pp 409–416

21. Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5:1205–1224

22. Lichman M (2013) UCI Machine learning repository. University of California, Irvine, School of Information and Computer Sciences

23. Li JD, Cheng KW, Wang SH, Morstatter F, Trevino RP, Tang JL, Liu H (2018) Feature selection: a data perspective. ACM Comput Surv 50(6)

24. Fayyad U, Irani KB (1993) Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of international joint conference on artificial intelligence, pp 1022–1027

25. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. ACM SIGKDD Explor Newsl 11(1):10–18

26. Zhao Z, Morstatter F, Sharma S, Alelyani S, Anand A, Liu H (2010) Advancing feature selection research. ASU feature selection repository 1–28

27. Herman G, Zhang B, Wang Y, Ye GT, Chen F (2013) Mutual information-based method for selecting informative feature sets. Pattern Recogn 46(12):3315–3327

28. Ren JF, Jiang XD, Yuan JS (2015) Learning LBP structure by maximizing the conditional mutual information. Pattern Recogn 48(10):3180–3190

29. Wang J, Wei JM, Yang ZL, Whang SQ (2017) Feature selection by maximizing independent classification information. IEEE Trans Knowl Data Eng 29(4):828–841

30. Tang B, Kay S, He HB (2016) Toward optimal feature selection in Naive Bayes for text categorization. IEEE Trans Knowl Data Eng 28(9):2508–2521

31. Zhang F, Chan PPK, Biggio B, Yeung DS, Roli F (2016) Adversarial feature selection against evasion attacks. IEEE Trans Cybernet 46(3):766–777

32. Fei T, Kraus D, Zoubir AM (2012) A hybrid relevance measure for feature selection and its application to underwater objects recognition. In: Proceedings of international conference on image processing, pp 97–100

33. Fei T, Kraus D, Zoubir AM (2015) Contributions to automatic target recognition systems for underwater mine classification. IEEE Trans Geosci Remote Sens 53(1):505–518

34. Gu XY, Guo JC (2019) A study on Subtractive Pixel Adjacency Matrix features. Multimedia Tools Appl 78(14):19681–19695
35. Gu XY, Guo JC, Wei HW, He YH (2020) Spatial-domain steganalytic feature selection based on three-way interaction information and KS test. Soft Comput 24(1):333–340
36. Veronica BC, Noelia SM, Amparo AB (2013) A review of feature selection methods on synthetic data. Knowl Inf Syst 34(3):483–519

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.