# New Results on Minimum Error Entropy Decision Trees

Joaquim P. Marques de Sá[1], Raquel Sebastião[2], João Gama[2], and Tânia Fontes[1]

[1] INEB-Instituto de Engenharia Biomédica, FEUP, Universidade do Porto,
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
{jmsa,trfontes}@fe.up.pt
[2] LIAAD - INESC Porto, L.A., Rua de Ceuta, 118, 6
4050-190 Porto, Portugal
{raquel,jgama}@liaad.up.pt

**Abstract.** We present new results on the performance of Minimum Error Entropy (MEE) decision trees, which use a novel node split criterion. The results were obtained in a comparive study with popular alternative algorithms, on 42 real world datasets. Carefull validation and statistical methods were used. The evidence gathered from this body of results show that the error performance of MEE trees compares well with alternative algorithms. An important aspect to emphasize is that MEE trees generalize better on average without sacrifing error performance.

**Keywords:** decision trees, entropy-of-error, node split criteria.

## 1 Introduction

Binary decision trees, based on univariate node splits, are popular tools in pattern recognition applications, given the semantic interpretation often assignable to nodal decision rules and fast computation. Available design algorithms for these trees are based on greedy construction of locally optimal nodes using some node split criterion. All node split criteria proposed until today are based, as far as we know, on estimates of class conditional input distributions at each node. A recent KDnuggets Poll (www.kdnuggets.com) disclosed that the most used analytic software were RapidMiner, R, and KNIME (freeware tools) and SPSS, SAS and Matlab (commercial tools). The decision tree algorithms in all these tools use "classic" split criteria (known since the seminal works on decision trees; for a survey see e.g. [1]): Gini, Information Gain and Twoing splitting rules.

In a recent paper [2] we proposed a new type of node split criterion that is not a "randomness measure" of class conditional distributions; instead, it is a "randomness measure" of nodal error distribution, concretely its entropy. We showed how to use this concept in the construction of "Minimum Error Entropy" (MEE) trees.

In the present paper we provide further comparative results on the application to real world datasets of MEE tree and competing tree design algorithms, using more datasets and sound validation and analysis methods, allowing, therefore, to reach a body of well-grounded conclusions concerning the advantages of using MEE trees.

C. San Martin and S.-W. Kim (Eds.): CIARP 2011, LNCS 7042, pp. 355–362, 2011.

## 2 MEE Trees

MEE trees are built by selecting, at each node a pair ($x$, $\omega$), where $x$ is a data feature and $\omega$ a class label, minimizing the error entropy. Consider a candidate split between a class $\omega_k \in \Omega$ and its complement $\overline{\omega}_k = \cup_{i \neq k} \omega_i$ using a rule $y$ based on the values of $x$ (e.g., $x < \Delta$). For any $x$, the rule $y$ produces a class assignment $\omega_y(x) \in \{\omega_k, \overline{\omega}_k\}$ which is compared to the true class label $\omega(x)$; an "error" variable $\omega(x) - \omega_y(x)$ is then defined. Denoting $T$ and $Y$ the random variables (r.v.) for a convenient coding of $\omega(x)$ and $\omega_y(x)$ (say, assigning 1 if $\omega = \omega_k$ and 0 otherwise), we then also have an r.v. of the "errors" (deviations), $E = T - Y$, taking value in {−1, 0, 1}, such that:

- $P(E = 1) = P(T = 1, Y = 0) \equiv P_{10}$ is the misclassification probability of $\omega_k$;
- $P(E = -1) = P(T = 0, Y = 1) \equiv P_{01}$ is the misclassification probability of $\overline{\omega}_k$;
- $P(E = 0) = 1 - P_{01} - P_{10}$ is the correct classification probability.

The MEE split rule consists of finding $y$ minimizing the error (Shannon) entropy:

$$EE \equiv EE(P_{01}, P_{10}) = -P_{01} \ln P_{01} - P_{10} \ln P_{10} - (1 - P_{01} - P_{10}) \ln(1 - P_{01} - P_{10}) .$$

The motivation for using MEE splits in decision trees stems from two main facts: by minimizing *EE* one is, in general, for not too overlapped distributions, favoring error distributions concentrated at the origin, with split points corresponding to the minimum probability of error; MEE will not work for largely overlapped $P(x|\omega_k)$ and $P(x \mid \overline{\omega}_k)$ distributions [3], providing a natural way when to stop tree growing, therefore inherently limiting the model complexity.

Details on the practical application of these principles to the construction of MEE trees are given in [2], showing that MEE trees can be applied to data described by either numerical or nominal features, and to any number of classes. The MEE tree algorithm written in Matlab, together with its description, is available at http:// gnomo.fe.up.pt/~nnig/. The main steps of the MEE tree algorithm are as follows (further details in [2]):

1. At each tree node we are given an $n \times f$ ($n$ cases, $f$ features) matrix $X$ and an $n \times c$ ($n$ cases, $c$ classes) matrix $T$, filled with zeros and ones. A univariate split $y$ minimizing *EE* is searched for in the $f \times c$-dimensional space.
2. For that purpose, the error rates $P_{10} = n_{10}/n$, $P_{01} = n_{01}/n$ ($n_{tt'}$: number of class $t$ cases classified as $t'$) are computed for each candidate class label $t$.
3. The rule minimizing (the empirical) *EE* is assigned to the node and if a stopping criterion is satisfied the node becomes a leaf. Otherwise, the left and right node sets are generated and steps 1 and 2 iterated.

A leaf is reached whenever a lower bound on the number of instances is reached or min*EE* occurs at interval ends, corresponding to the large distribution overlap case.

An important aspect shown in [2] is that MEE trees are quite insensitive to pruning; i.e., in general, the tree built without pruning is the same or almost the same as the one to which pruning was applied. This is a consequence of what was said above:

for largely overlapped $P(x \mid \omega_k)$ and $P(x \mid \overline{\omega}_k)$ distributions one is able to detect an invalid MEE point and stop node splitting. Therefore, MEE enforces simple models with good generalization ability.

## 3 Materials and Methods

### 3.1 Real World Datasets and Experimental Setup

The MEE algorithm was applied to the 42 datasets presented in Table 1, and the results confronted with those obtained using the CART-Gini, CART-Information-Gain and CART-Twoing algorithms (available in Matlab) and the popular C4.5 algorithm (available in Weka). All algorithms were run with unit misclassification costs, estimated priors and the same minimum number of instances at a node: 5.

All datasets are from the well-known UCI repository [4], except the colon, central nervous system and leukemia datasets which are from the Kent Ridge Biomedical Dataset (http://datam.i2r.a-star.edu.sg/datasets/krbd).

The CART and MEE algorithms were run with Cost-Complexity Pruning (CCP) with the 'min' criterion and 10-fold cross-validation. The C4.5 algorithm was run with Pessimistic Error Pruning (PEP) at 25% confidence level [1].

**Table 1.** Datasets. The number of categorical features is given inside parentheses.

| | Arrythmya | Balance | Car | Clev. HD2 | Clev.HD5 | CNS | Colon |
|---|---|---|---|---|---|---|---|
| No. cases | 452 | 625 | 1278 | 297 | 297 | 60 | 62 |
| No. features | 274 (54) | 4 (4) | 6 (6) | 13 (8) | 13 (8) | 7129 (0) | 2000 (0) |
| No. classes | 9 | 3 | 4 | 2 | 5 | 2 | 2 |
| | Cork stop. | Credit | CTG | Dermatol. | E-coli | Flags | H. surv |
| No. cases | 150 | 653 | 2126 | 358 | 327 | 194 | 306 |
| No. features | 10 (0) | 15 (9) | 21 (0) | 34 (33) | 5 (0) | 26 (20) | 3 (0) |
| No. classes | 3 | 2 | 10 | 6 | 5 | 7 | 2 |
| | Heart | Image Seg. | Landsat | Led | Leukemia | LRS | Lymphog. |
| No. cases | 270 | 2310 | 6435 | 200 | 72 | 531 | 148 |
| No. features | 13 (8) | 18 (0) | 36 (0) | 7 (7) | 7129 (0) | 101 (1) | 18 (17) |
| No. classes | 2 | 7 | 6 | 10 | 2 | 6 | 3 |
| | Mammog. | Monk | Mushrrom | Ozone | Page blks | Parkinsons | Pen Digits |
| No. cases | 830 | 556 | 8214 | 1847 | 5473 | 195 | 10992 |
| No. features | 4 (3) | 6 (6) | 21 (21) | 72 (0) | 10 (0) | 22 (0) | 16 (0) |
| No. classes | 2 | 2 | 2 | 2 | 5 | 2 | 10 |
| | P. Diabetes | P. Gene | Robot-1 | Spect-Heart | Spectf-Heart | Swiss HD | Synth. Chart |
| No. cases | 768 | 106 | 88 | 267 | 267 | 120 | 600 |
| No. features | 8 (0) | 57 (57) | 90 (0) | 22 (22) | 44 (0) | 7 (4) | 60 (0) |
| No. classes | 2 | 2 | 4 | 2 | 2 | 5 | 6 |
| | Thyroid | VA HD | Wdbc | Wpbc | Wine | Yeast | Zoo |
| No. cases | 215 | 186 | 569 | 194 | 178 | 1479 | 101 |
| No. features | 5 (0) | 6 (4) | 30 (0) | 32 (0) | 13 (0) | 6 (0) | 16 (16) |
| No. classes | 3 | 5 | 2 | 2 | 3 | 9 | 7 |

### 3.2 Statistical Methods

Ten-fold crossvalidation (CV10) was applied to all datasets and tree design algorithms. According to the theoretical analysis of crossvalidation [5], ten is a sensible

choice for the number of folds. A more recent work, [6], also confirmed the good performance of CV10 when compared with alternative validation methods.

Besides average test error estimates we also computed average design (resubstitution) error estimates, allowing us to evaluate generalization. Statistics regarding tree sizes in the cross-validation experiments were also computed.

All results obtained for the five methods were evaluated following recommendations in [7-10], by namely performing: counts of wins and losses with chi-square test; multiple sign test comparing each method against MEE; Friedman test; post-hoc Dunn-Sidak test for multiple comparison; post-hoc Finner test for comparison of each method against MEE. The post-hoc tests are only performed when a significant Friedman $p < 0.05$ is found. The Finner test for post-hoc comparisons of a proposed method against another was analyzed in [8] and found to be more powerful than competing tests.

## 4 Results

### 4.1 Error Rates

Table 2 presents the cross-validation estimates of the error rate, with the best MEE solution found for class unions up to [$c$/2]. The total number of wins (smallest error) and losses are also shown in Table 2 with the chi-square test $p$: no significant difference is found relative to the equal distribution hypothesis.

The Friedman test did not detect significant differences ($p = 0.453$) for these 42 datasets. The mean ranks for the five methods (following from now on Table 2 order) are: 2.98, 3.16, 3.27, 2.68 and 2.92.

The comparison between MEE vs any of the other algorithms, with the signs used in the multiple sign test [8], was also performed. Denoting by $e$ the error rate, the null hypothesis of the test is $H_0$: $e_j \leq e_{MEE}$, where $j$ is any algorithm MEE is compared with. $H_0$ is rejected whenever the sum of minuses is below a certain critical value, which is in this case 16 at $p = 0.05$. Since all sums of minuses (resp. 19, 17, 17, 23) are above the critical value, we conclude that MEE performs similarly to any of the other algorithms.

### 4.2 Generalization

Denoting by $e_R$ and $e_{CV}$ respectively the mean training set error rate and the mean test set (CV10) error rate, we computed $D = |e_R - e_{CV}| / \bar{s}$, using the pooled standard deviation $\bar{s}$. $D$ reflects the generalization ability of the classifiers. The Friedman test found a significant difference ($p \approx 0$) of the methods for the $D$ values (mean ranks: 2.69, 2.91, 3.01, 4.22 and 2.16); the post-hoc Dunn-Sidak test revealing a significant difference between MEE vs C4.5 and Twoing (see Fig. 1). The post-hoc Finner test found, in fact, significantly better generalization of MEE vs *any* of the other methods.

**Table 2.** CV10 estimates of test set Pe-(std) with wins (bold) and losses (italic)

| | Arrythmya | Balance | Car | Clev. HD2 | Clev.HD5 | CNS |
|---|---|---|---|---|---|---|
| Gini | 0.3518 (0.022) | **0.1952 (0.016)** | 0.0434 (0.005) | 0.2357 (0.025) | *0.4680 (0.029)* | 0.3500 (0.062) |
| Info Gain | 0.3606 (0.023) | 0.2528 (0.017) | 0.0457 (0.005) | 0.2593 (0.025) | **0.4512 (0.029)** | **0.3000 (0.059)** |
| Twoing | 0.3628 (0.023) | 0.2176 (0.017) | **0.0405 (0.005)** | *0.2761 (0.026)* | 0.4613 (0.029) | 0.3167 (0.060) |
| C4.5 | *0.3934 (0.023)* | 0.2192 (0.017) | 0.0434 (0.005) | **0.1987 (0.023)** | 0.4577 (0.029) | *0.4833 (0.065)* |
| MEE | **0.3208 (0.022)** | *0.3120 (0.019)* | *0.0718* (0.006) | 0.2222 (0.024) | 0.4646 (0.029) | 0.3667 (0.062) |
| | Colon | Cork stop. | Credit | CTG | Dermatol. | E-coli |
| Gini | *0.2581 (0.056)* | **0.1133 (0.028)** | 0.1363 (0.013) | 0.1689 (0.008) | **0.0531 (0.012)** | *0.1927 (0.022)* |
| Info Gain | *0.2581 (0.056)* | *0.1400 (0.028)* | 0.1363 (0.013) | 0.1877 (0.008) | 0.0587 (0.012) | 0.1896 (0.022) |
| Twoing | 0.2419 (0.054) | 0.1267 (0.027) | 0.1363 (0.013) | 0.1811 (0.008) | 0.0670 (0.013) | 0.1682 (0.021) |
| C4.5 | 0.2419 (0.054) | **0.1133 (0.026)** | **0.1332 (0.013)** | **0.1731 (0.008)** | *0.0991 (0.016)* | 0.1713 (0.021) |
| MEE | **0.1935 (0.050)** | 0.1200 (0.027) | *0.1424 (0.014)* | *0.1891 (0.008)* | 0.0559 (0.012) | **0.1315 (0.019)** |
| | Flags | H. surv | Heart | Image Seg. | Landsat | Led |
| Gini | 0.4794 (0.036) | 0.2680 (0.025) | 0.2037 (0.025) | 0.0403 (0.004) | **0.1294 (0.004)** | 0.3100 (0.033) |
| Info Gain | 0.4639 (0.036) | **0.2647 (0.025)** | 0.2444 (0.026) | 0.0368 (0.004) | 0.1361 (0.004) | 0.3050 (0.033) |
| Twoing | *0.4845 (0.036)* | 0.2745 (0.026) | *0.2481 (0.026)* | 0.0485 (0.004) | 0.1406 (0.004) | 0.3200 (0.033) |
| C4.5 | **0.4022 (0.035)** | *0.3070 (0.026)* | **0.2000 (0.024)** | **0.0385 (0.004)** | 0.1324 (0.004) | *0.5869 (0.035)* |
| MEE | 0.4691 (0.036) | **0.2647 (0.025)** | 0.2444 (0.026) | *0.0589 (0.005)* | *0.1566 (0.005)* | **0.3000 (0.032)** |
| | Leukemia | LRS | Lymphog. | Mammog. | Monk | Mushrrom |
| Gini | 0.1806 (0.045) | 0.1450 (0.015) | *0.2754 (0.037)* | 0.2084 (0.014) | 0.1007 (0.013) | 0.0004 (0.000) |
| Info Gain | *0.1944 (0.047)* | 0.1450 (0.015) | *0.2754 (0.037)* | 0.2157 (0.014) | 0.1187 (0.014) | **0.0000 (0.000)** |
| Twoing | 0.1667 (0.044) | **0.1431 (0.015)** | **0.2061 (0.033)** | 0.2072 (0.014) | *0.1331 (0.014)* | **0.0000 (0.000)** |
| C4.5 | **0.1389 (0.041)** | *0.2917 (0.020)* | 0.2528 (0.036) | **0.2000 (0.014)** | 0.1115 (0.013) | **0.0000 (0.000)** |
| MEE | 0.1667 (0.044) | 0.1638 (0.016) | 0.2500 (0.036) | *0.2386 (0.015)* | **0.0989 (0.013)** | *0.0009 (0.000)* |
| | Ozone | Page blks | Parkinsons | Pen Digits | P. Diabetes | P. Gene |
| Gini | **0.0693 (0.006)** | 0.0342 (0.002) | *0.1590 (0.026)* | 0.0418 (0.002) | **0.2487 (0.016)** | 0.2547 (0.042) |
| Info Gain | **0.0693 (0.006)** | 0.0347 (0.002) | 0.1487 (0.025) | **0.0357 (0.002)** | 0.2695 (0.016) | *0.3208 (0.045)* |
| Twoing | 0.0731 (0.006) | *0.0365 (0.003)* | 0.1487 (0.025) | 0.0378 (0.002) | 0.2695 (0.016) | 0.2642 (0.043) |
| C4.5 | *0.0785 (0.006)* | **0.0281 (0.002)** | **0.1242 (0.024)** | 0.0418 (0.002) | 0.2578 (0.016) | 0.2547 (0.042) |
| MEE | 0.0704 (0.006) | 0.0347 (0.002) | 0.1436 (0.025) | *0.0666 (0.002)* | *0.3216 (0.017)* | **0.1698 (0.036)** |
| | Robot-1 | Spect-Heart | Spectf-Heart | Swiss HD | Synth. Chart | Thyroid |
| Gini | 0.2727 (0.047) | 0.2022 (0.025) | 0.2097 (0.025) | 0.6117 (0.044) | 0.1150 (0.013) | 0.0844 (0.019) |
| Info Gain | 0.2841 (0.048) | 0.2060 (0.023) | **0.2060 (0.025)** | 0.6083 (0.045) | 0.0817 (0.011) | *0.1023 (0.021)* |
| Twoing | **0.1932 (0.042)** | *0.2210 (0.025)* | *0.2172 (0.025)* | *0.6250 (0.044)* | *0.1200 (0.013)* | 0.0977 (0.020) |
| C4.5 | *0.3500 (0.051)* | **0.1873 (0.024)** | 0.2135 (0.025) | **0.5847 (0.045)** | 0.0833 (0.011) | **0.0558 (0.016)** |
| MEE | 0.2614 (0.047) | 0.1985 (0.024) | **0.2060 (0.025)** | 0.6083 (0.045) | **0.0617 (0.010)** | 0.0977 (0.020) |
| | VA HD | Wdbc | Wpbc | Wine | Yeast | Zoo |
| Gini | *0.7527 (0.032)* | 0.0650 (0.010) | **0.2371 (0.031)** | 0.1067 (0.023) | 0.4219 (0.013) | 0.1683 (0.037) |
| Info Gain | *0.7527 (0.032)* | *0.0721 (0.011)* | 0.2474 (0.031) | **0.0562 (0.017)** | 0.4206 (0.013) | 0.1683 (0.037) |
| Twoing | 0.7419 (0.032) | 0.0685 (0.011) | **0.2371 (0.031)** | 0.0787 (0.020) | 0.4381 (0.013) | 0.1386 (0.034) |
| C4.5 | 0.7366 (0.032) | 0.0650 (0.010) | *0.3144 (0.031)* | 0.0899 (0.021) | **0.4077 (0.013)** | *0.3069 (0.046)* |
| MEE | **0.7097 (0.033)** | **0.0615 (0.010)** | **0.2371 (0.031)** | *0.1180 (0.024)* | *0.5335 (0.013)* | **0.1089 (0.031)** |

| | Gini | Info Gain | Twoing | C4.5 | MEE | p |
|---|---|---|---|---|---|---|
| Wins | 7 | 9 | 6 | 14 | 13 | 0.27 |
| Losses | 6 | 8 | 9 | 10 | 12 | 0.70 |

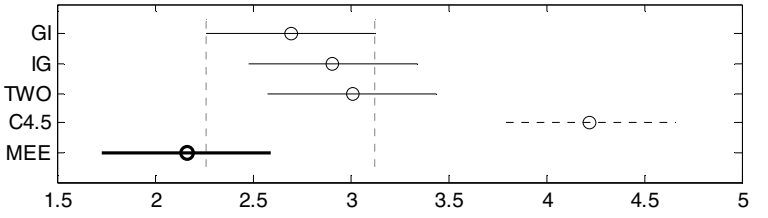


**Fig. 1.** Dunn-Sidak comparison intervals for the *D* scores

### 4.3 Tree Sizes

Table 3 shows the averages and ranges of tree sizes achieved in the cross-validation experiments by all algorithms. The total number of wins (smallest average tree size) and losses are also shown in Table 3 with the chi-square *p*. A significant difference is found relative to the equal distribution hypothesis. Performing the multiple sign test as in the preceding section, we found a significant difference of MEE vs C4.5: smaller trees on average for MEE. The Friedman test also found a significant difference ($p \approx 0$) with mean ranks 2.51, 2.43, 2.80, 4.35 and 2.92. The post-hoc comparisons tests confirmed the conclusions of the multiple sign test (see Fig. 2).

We re-analyzed the test set error rates and the *D* scores for the 20 datasets where MEE found the smallest average tree size. We arrived essentially to the same conclusions as in 4.1 and 4.2.

We also analyzed the tree size ranges (see Table 3), since a significantly smaller range of tree sizes is a symptom of a more stable algorithm [11-12]. We didn't find any statistically significant diference of the tree size ranges, either for the Friedmann ($p = 0.3$) or the multiple sign test.

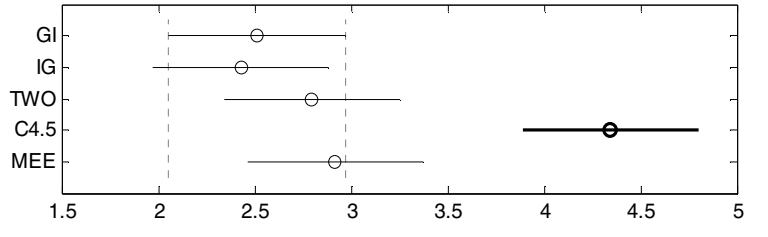


**Fig. 2.** Dunn-Sidak comparison intervals for the average tree size

**Table 3.** Average-(ranges) of tree sizes with wins (bold) and losses (italic)

| | Arrythmya | Balance | Car | Clev. HD2 | Clev.HD5 | CNS |
|---|---|---|---|---|---|---|
| Gini | 12.2 (10) | **26.6 (10)** | 66.2 (38) | **6.8 ( 8)** | **5.6 (16)** | **1.6 (2)** |
| Info Gain | 12.8 ( 8) | **26.6 (20)** | 58.0 (34) | 7.8 ( 8) | 6.8 ( 8) | 2.0 (2) |
| Twoing | **11.6 (14)** | 27.6 ( 8) | 66.0 (26) | 8.8 (10) | 5.8 (12) | 2.4 (2) |
| C4.5 | *37.8 (24)* | 43.6 (12) | 71.4 (12) | *19.2 (10)* | *34.8 (16)* | *6.2 (4)* |
| MEE | 36.2 (10) | *90.6 (52)* | *115.0 (48)* | 18.8 (12) | 22.0 (42) | 3.2 (4) |
| | Colon | Cork stop. | Credit | CTG | Dermatol. | E-coli |
| Gini | **3.0 (0)** | **5.0 (0)** | **3.0 ( 0)** | 74.6 (46) | 13.4 (6) | 10.8 (12) |
| Info Gain | **3.0 (6)** | **5.0 (0)** | **3.0 ( 0)** | 70.2 (52) | 15.8 (6) | 10.2 (14) |
| Twoing | **3.0 (4)** | **5.0 (0)** | **3.0 ( 0)** | *60.0 (30)* | *16.2 (2)* | 11.0 (14) |
| C4.5 | *5.8 (2)* | *5.8 (6)* | *24.8 (20)* | 136 (26) | **13.0 (0)** | *17.2 ( 8)* |
| MEE | 3.2 (2) | **5.0 (0)** | 12.4 (24) | **56.8 (16)** | 14.1 (2) | **9.4 ( 2)** |
| | Flags | H. surv | Heart | Image Seg. | Landsat | Led |
| Gini | 12.8 (12) | *4.2 (28)* | 11.2 (12) | *79.8 (58)* | **108.8 (80)** | 22.0 (24) |
| Info Gain | **10.0 (14)** | **1.0 ( 0)** | 10.8 (32) | 55.2 (36) | 131.2 (98) | **20.6 ( 8)** |
| Twoing | 10.6 (20) | 2.6 ( 8) | **8.4 (10)** | 70.4 (74) | 110.4 (98) | *25.6 (20)* |
| C4.5 | *27.8 ( 6)* | *4.2 ( 6)* | 17.2 ( 8) | 59.8 (18) | *331.6 (58)* | 22.0 ( 6) |
| MEE | 20.0 (36) | **1.0 ( 0)** | *19.4 (14)* | **32.4 ( 8)** | 125.6 (56) | 23.0 ( 4) |
| | Leukemia | LRS | Lymphog. | Mammog. | Monk | Mushrrom |
| Gini | **3.0 (0)** | 11.6 ( 4) | **6.0 ( 4)** | **5.8 (14)** | 28.8 (28) | 18.2 (8) |
| Info Gain | **3.0 (0)** | 11.8 ( 6) | **6.0 ( 4)** | 6.2 ( 6) | 30.0 (24) | **16.2 (2)** |
| Twoing | **3.0 (0)** | **11.2 ( 2)** | 6.4 ( 4) | 6.2 ( 6) | **27.0 (24)** | 18.8 (2) |
| C4.5 | *3.8 (2)* | *28.8 (12)* | 13.8 ( 8) | 16.8 ( 4) | 31.0 (14) | 23.0 (0) |
| MEE | **3.0 (0)** | 24.2 ( 6) | *17.6 (22)* | *24.6 (12)* | 33.6 (28) | *42.6 (16)* |
| | Ozone | Page blks | Parkinsons | Pen Digits | P. Diabetes | P. Gene |
| Gini | **1.0 ( 0)** | 23.4 (28) | 5.4 ( 8) | 336.8 (184) | 6.0 ( 4) | 7.8 (10) |
| Info Gain | **1.0 ( 0)** | 23.4 (22) | 7.6 (10) | 327.6 (216) | 6.6 (14) | **5.6 ( 8)** |
| Twoing | 3.4 (24) | 22.6 (24) | 8.8 (18) | *371.2 (152)* | 8.4 (36) | 7.0 ( 8) |
| C4.5 | *54.0 (34)* | *54.8 (16)* | *15.0 ( 4)* | 271.4 ( 32) | *30.6 (28)* | *11.4 ( 6)* |
| MEE | **1.0 ( 2)** | **20.8 ( 6)** | **3.0 ( 0)** | **233.0 ( 60)** | **2.2 ( 4)** | 8.6 ( 8) |
| | Robot-1 | Spect-Heart | Spectf-Heart | Swiss HD | Synth. Chart | Thyroid |
| Gini | **8.2 (4)** | 6.0 (14) | 1.6 ( 6) | 1.8 ( 8) | 24.8 (20) | 8.0 ( 6) |
| Info Gain | 8.8 (4) | 13.4 (24) | **1.0 ( 0)** | **1.0 ( 0)** | 37.4 (18) | 7.4 ( 8) |
| Twoing | 9.4 (2) | **5.8 (20)** | 2.2 (12) | 1.6 ( 6) | *39.2 (24)* | *9.2 (12)* |
| C4.5 | *10.4 (4)* | 13.6 ( 4) | *28.4 (12)* | *17.8 (16)* | 28.8 ( 6) | 9.0 ( 4) |
| MEE | **8.2 (2)** | *23.2 (28)* | **1.0 ( 0)** | **1.0 ( 0)** | **22.0 ( 2)** | **5.0 ( 0)** |
| | VA HD | Wdbc | Wpbc | Wine | Yeast | Zoo |
| Gini | **5.8 (18)** | 10.2 (14) | **1.0 ( 0)** | *11.8 (16)* | 22.4 (22) | 10.6 (6) |
| Info Gain | 10.0 (18) | 10.2 (14) | 1.6 ( 6) | 8.4 ( 2) | **22.0 (16)** | 10.8 (4) |
| Twoing | 8.6 (28) | 11.0 (20) | **1.0 ( 0)** | 7.8 ( 4) | 17.8 (12) | **10.4 (2)** |
| C4.5 | *26.0 (20)* | *15.4 (10)* | *13.8 (26)* | 8.8 ( 2) | *135.2 (40)* | *11.0 (0)* |
| MEE | 23.4 (12) | **5.2 ( 2)** | **1.0 ( 0)** | **6.6 ( 4)** | 25.4 (10) | *11.0 (0)* |

| | Gini | Info Gain | Twoing | C4.5 | MEE | p |
|---|---|---|---|---|---|---|
| Wins | 15 | 15 | 11 | 1 | 20 | 0.00 |
| Losses | 3 | 0 | 5 | 27 | 9 | 0.00 |

## 5 Conclusions

The present work provided a substantial body of results concerning classification experiments carried out in 42 real-world datasets, with varied number of cases, classes and features, by trees designed using the MEE approach and the popular CART-Gini, CART-Information-Gain, CART-Twoing and C4.5 algorithms. The statistical analysis of the test set (CV10) error rates, obtained in these experiments showed that the MEE algorithm competes well with the other algorithms.

Moreover, we have obtained in the present work statistically significant evidence that the MEE algorithm produces, on average, smaller trees than the popular C4.5 algorithm.

As to the generalization issue, MEE trees were found to generalize better than those produced by the other algorithms without sacrifice on performance.

These features of the MEE tree design, particularly the better generalization of MEE tree solutions, together with their relative insensibility to pruning shown elsewhere [2], are of importance in many pattern recognition applications.

**Aknowledgement.** The work of Raquel Sebastião is supported by the Portuguese Foundation for Science and Technology (FCT) under the PhD Grant SFRH/BD/41569/2007.

## References

1. Rokach, L., Maimon, O.: Decision Trees. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook. Springer, Heidelberg (2005)
2. Marques de Sá, J.P., Sebastião, R., Gama, J.: Tree Classifiers Based on Minimum Error Entropy Decisions. Can. J. Artif. Intell., Patt. Rec. and Mach. Learning (in press, 2011)
3. Silva, L., Felgueiras, C.S., Alexandre, L., Marques de Sá, J.: Error Entropy in Classification Problems: A Univariate Data Analysis. Neural Computation 18, 2036–2061 (2006)
4. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2010), http://www.ics.uci.edu/~mlearn/MLRepository.html
5. Kearns, M.: A Bound on the Error of Cross Validation Using the Approximation and Estimation Rates, with Consequences for the Training-Test Split. Neural Computation 9, 1143–1161 (1997)
6. Molinaro, A.M., Simon, R., Pfeiffer, R.M.: Prediction Error Estimation: A Comparison of Resampling Methods. Bioinformatics 21, 3301–3307 (2005)
7. Demšar, J.: Statistical Comparisons of Classifiers over Multiple Data Sets. J. of Machine Learning Research 7, 1–30 (2006)
8. García, S., Fernández, A., Luengo, J., Herrera, F.: Advanced Nonparametric Tests for Multiple Comparisons in the Design of Experiments in Computational Intelligence and Data Mining: Experimental Analysis of Power. Information Sciences 180, 2044–2064 (2010)
9. Hochberg, Y., Tamhane, A.C.: Multiple Comparison Procedures. John Wiley & Sons, Inc. (1987)
10. Salzberg, S.L.: On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach. Data Mining and Knowledge Discovery 1, 317–328 (1997)
11. Jensen, D., Oates, T., Cohen, P.R.: Building Simple Models: A Case Study with Decision Trees. In: Liu, X., Cohen, P., Berthold, M. (eds.) IDA 1997. LNCS, vol. 1280, pp. 211–222. Springer, Heidelberg (1997)
12. Li, R.-H., Belford, G.G.: Instability of Decision Tree Classification Algorithms. In: Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 570–575 (2002)