Building Diversified Multiple Trees for Classification in High Dimensional Noise Data

Jiuyong Li, Lin Liu, Jixue Liu and Ryan Green

School of Information Technology and Mathematical Sciences
University of South Australia, Australia

Abstract

It is common that a trained classification model is applied to the operating data that is deviated from the training data because of noise. This paper demonstrate an ensemble classifier, Diversified Multiple Trees (DMT) is more robust to classify noised data than other widely used ensemble methods. DMT is tested on three real world biological data sets from different laboratories in comparison with four benchmark ensemble classifiers. Experimental results show that DMT is significantly more accurate than other benchmark ensemble classifiers on noised test data. We also discussed a limitation of DMT and its possible variations.

1 Introduction

Classification aims at building models on known labeled data, and applying the models to unknown data to predict labels of data records. For example, a fraud detection system is built on known fraudulent cases, and is used to predict coming fraudulent cases. A fundamental assumption is that the system operating environment is very similar to the system building environment. Unfortunately, it is difficult to specify operating environment conditions precisely in the process of building systems. A system may be used in an environment that is different from the system building environment. No system works in an operating environment that is totally different from the developing environment. However, some systems are less sensitive than others regarding to changes in the environment.

The problem is generally categorised as robust classification. Term “Robust” has widely used in classification literature. The robustness indicates the capability of dealing with noises (or outliers) of a system. The impact of noises is twofold.

- noises make a data set un-learnable. No credible classifiers can be built from the noise data. In this case, the source of noises needs to be discovered and data needs to be cleaned. This is not a focus of our study.

- noises make a classifier unreliable. A classifier is good in the known training data, but does not make reliable predictions in an operating environment with noises. Our work is in this category. Correction of noised operating data is an approach to make a classifier reliable. In this paper we focus on building models for an imprecise operating environment without correction noised values.

The problem is preval. For example, biological samples may be obtained from different equipments using different processing procedures by different methods. However, we still wish model produced from one set of samples to be used in another set of samples. We wish to apply a model obtained from a small group to the general population. We wish to use a model built on the previous five years to the next two year data.
A practical setting of our discussions is based on biological data, where the dimension is normally very high and samples are relatively few, for example, thousands of attributes versus tens to a few hundreds of samples. Noises are unavoidable in biological experiments and can be introduced in multiple stages, biological sample preparation, experiment, data collection and data processing. Since the number of samples is normally small, minor noises can have big impact. For example, it has been found that one predictor found from one data set performs badly on the data of another data set [11].

Decision tree models are a typical example for this problem. A decision tree may perform badly in test data although it is validated very well in the validation data. This problem is evident when the dimension of a data set is high. One reason for this problem is that a decision tree uses a small subset of attributes for classification and any inconsistencies in these attributes between the training and test data will cause significant downgrade in its classification accuracy.

We demonstrate that a diversified multiple tree (DMT) approach [17] is more robust than other ensemble methods, such as AdaBoost [12], Bagging [6], and Random Forests (RF) [7], in an operating environment with noises. DMT is different from those ensemble methods, which make use of a large number of weak classifiers to improve classification accuracy. DMT utilises a small number of strong models to improve the robustness of a classification model. We use Diversified Multiple Tree instead of Maximally Diversified Multiple Decision Tree [17] for brevity.

2 Problem definition

Let $D$ be a training data set. We build model $M$ on $D$, and apply the model to a test data set, $D_T$. Some values in $D_T$ are noised. Noises make values in data deviate from their true values.

Note that noises can exist in training data set $D$ too. For example, let us assume that $D$ is noised and $D_T$ is not. From model $M$’s viewpoint, $D$ is its ground truth, and relatively $D_T$ is noised. Therefore, the assumption of $D_T$ being noised is a general one. Noises can be in both $D$ and $D_T$ but we offset all inconsistencies to $D_T$. Therefore, we only consider noises in $D_T$ in this paper. Another assumption is that noises in $D$ is not big enough to affect the learnability of data $D$.

Noises in $D_T$ will affect classification accuracy of model $M$ on $D_T$. The research questions is how to build a robust model $M$. The robustness of a model is its capacity for resisting noise values in the test data. In other words, a more robust model will make more accurate classifications on the noise test data than a less robust model.

Intuitively, an ensemble model containing more than one classification model will be more robust than a single classification model. This is because that noise values are more likely to affect a single model than two or more models simultaneously. In other words, a single model is easier to be affected by noises than an ensemble model.

Let us consider a simple vote mechanism of an ensemble model. Each alternative model makes a classification. The final classification of the ensemble model is the most frequent predicted class of all alternative models. For example, if we have 5 alternative classifiers in an ensemble model where each classifier is very accurate. For a test record, even if the noise affects the classification performance of two classifiers, the ensemble classifier still gives the correct classification.

The core for the robustness of an ensemble model is that the classification is based on the multiple alternative models. Assume that those alternative models are independent from each other for being affected by noises. In other words, noises are uncorrelated among attributes of a data set. Assume that all models are equally accurate and have an equal probability being affected by noises. Let the probability of one model affected by noises is $\alpha$. The probability of two models being affected by noises simultaneously is $\alpha^2$. Two
models are less likely being affected by noises than one model. Three models are less likely
being affected by noises than two models and so forth.

The independence of models being affected by noises is the key for the robustness of
an ensemble classifier. Let us assume that all attribute values have the equal chance for
being affected by noises. An intuitive implementation of independent models is to build
disjunct models which do not share attribute values in their decision logics. As a result,
disjunct alternative models are independent from each other for being affected by noises.
In the next section, we will discuss a decision tree based implementation.

3 Diversified Multiple Trees (DMT)

A decision tree is a popular data mining method, and is a form of human understandable
knowledge. A decision tree employs a divide and conquer scheme for tree building. Each
partition makes use of an attribute. Normally, a decision tree makes use of a subset of all
attributes in decision. When the number of attributes is large, a decision only makes use
of a small number of attributes. Other quality decision trees can be built on remaining
attributes. It is easy to build a set of trees that are disjunct.

A typical decision tree construction method is C4.5 [22]. C4.5 divides the training data
into some disjoint sub data sets based on distinct values of an attribute, which is selected by
the information gain ratio [22]. The sub data sets are then simultaneously divided by other
attributes recursively until each sub data set contains instances of one class, or nearly. In
classification, a coming unclassified instance is traced down a path from the root to a leaf
that contains the majority of instances of one class. The instance is classified by the class
at the leaf of the matched path.

A decision is unreliable in an operating environment with noises since noises in some
attributes may affect its performance significantly. For example, Figure 1 shows a decision
tree built on the Harvard data set (to be explained in the Experiments section) makes use
only one gene out of 11657 genes. If the data value of the gene is noised in a future coming
data set, the model make wrong predictions even though other genes can help make right
predictions. To make a tree model robust, multiple trees should be used.

Multiple trees will make a tree model robust if they do not use the same attributes. A
decision tree makes use of attributes for predictions explicitly, it is easy to build disjunct
decision trees to make a tree model robust. In some data sets, for example gene expression
data, the number of attributes is large, say 10,000 attributes. It easily builds many quality
trees on the data set without using overlapping attributes.

A method for building disjunct multiple trees is depicted in Algorithm 1. The idea
for Diversified Multiple Trees (DMT) algorithm is to build a set of disjunct trees. In this
process, used attributes by a tree are knocked off. As a result, all output trees are disjunct.
The base algorithm for decision tree construction is C4.5 [22]. The classification process
is based the simple vote mechanism.

The complexity of a decision tree construction algorithm is linear to the number of
attributes and the size of a data set, i.e. \( O(mn \log(n)) \), where \( n \) is the number of data
Algorithm 1 Diversified Multiple Trees (DMT)

**Training**

Input: data set $D$, integer $k$
Output: ensemble model $\mathcal{M}$

1: let $i = 0$
2: initiate ensemble model $\mathcal{M}$
3: while $i < k$ do
4:   Build a C4.5 tree $T_i$ on $D$
5:   remove all attributes used in $T_i$ from $D$
6:   let $i = i + 1$
7: add $M_i$ to the ensemble model $\mathcal{M}$
8: end while
9: output ensemble model $\mathcal{M}$.

**Classifying**

Input: a data record $r$ and an ensemble model $\mathcal{M}$
Output: class label of $r$

1: $C = \emptyset$
2: for each $M_i$ in ensemble model $\mathcal{M}$ do
3:   let $c_i = $ classification result of $M_i$ on $r$
4:   add $c_i$ to $C$
5: end for
6: output the most frequent class in $C$

objects and $m$ is the number of attributes. When we build $k$ trees, the complexity becomes $O(kmn \log(n))$. The algorithm is efficient.

Alternative to the simple vote, various weighting schemes can be used in the classification stage. One is based on the precision of the leaf making the classification to integrate classifications of alternative trees. Since most data sets used in our experiments are small, we use Laplace accuracy:

$$\text{acc}_L = \frac{\#tp + 1}{\#tp + \#fp + c},$$

where $\#tp$ and $\#fp$ are the number of true positives and false positives, and $c$ is the number of classes in a data set. This scheme gives accurate classification high weights. However, such weights do not consider the number of instances in a decision leaf. To counter such a drawback, a support scheme weights the ratio of instances at a decision leaf, i.e. support $= \frac{\#fp}{n}$ where $n$ is the size of a data size. We will test both schemes in experiments.

Unlike Bagging [6] and Random Trees [10], DMT does not sample instances, but makes use of different set of attributes to build diversified trees. Unlike Random Forests and Random Trees [10] which make use of attributes randomly, DMT utilises the attributes in a systematical way. Unlike Bagging, Boosting [12], Random Forests and Random Trees which need a large tree community to make accurate classifications, DMT only needs a small number of high quality decision trees.

### 4 Experiments

Experiments are conducted in two parts. The first one is to test the robustness of DMT in comparison with other ensemble methods. The second part is to test the effectiveness of various classification weighting schemes of DMT. All experiments are conducted on real world high dimensional data sets.
Table 1: Data set description

| Name          | #Attr | Size | Classes | Comments |
|---------------|-------|------|---------|----------|
| Harvard (H)   | 11657 | 156  | 139/17  | Affymetric |
| Michigan (M)  | 6357  | 96   | 86/10   | Affymetric |
| Stanford (S)  | 11985 | 46   | 41/5    | cDNA      |

Table 2: A comparison of test accuracies on data sets of different laboratories. Accuracies are in percentage. The highest accuracy in each row is highlighted.

| tr/te | C4.5 | DMT-7 | Ada | Bag | RF | RT |
|-------|------|-------|-----|-----|----|----|
| H/M   | 99.0 | 99.0  | 99.0| 99.0| 93.8| 85.4|
| H/S   | 84.8 | 95.7  | 89.1| 84.8| 91.3| 82.6|
| M/H   | 96.8 | 98.7  | 96.8| 98.7| 89.1| 89.1|
| M/S   | 54.3 | 95.7  | 54.3| 93.5| 89.1| 78.2|
| S/H   | 80.8 | 92.9  | 80.8| 89.7| 89.1| 84.6|
| S/M   | 85.4 | 94.8  | 85.4| 89.6| 89.6| 86.5|
| Ave   | 83.5 | 96.1  | 84.2| 92.3| 90.3| 84.4|

Table 3: Wilcoxon signed ranks test for various methods. The alternative hypothesis is that the method to the left is more accurate than one at the top. Significant test results at 95% confidence level are highlighted.

| p-value | C4.5 | Ada | Bag | RF | RT |
|---------|------|-----|-----|----|----|
| DMT-7   | 0.030| 0.030| 0.050| 0.016| 0.016|
| Ada     | 0.5  | -   | -   | -  | -  |
| Bag     | 0.05 | 0.14| -   | 0.28| 0.016|
| RF      | 0.22 | -   | -   | -  | -  |
| RT      | 0.58 | -   | -   | -  | -  |

4.1 The robustness of DMT on noised test data

Data sets come from three laboratories studying the same type of lung cancer, called Harvard [4], Michigan [3], and Stanford [14]. They have been obtained from different patient samples and from two different Microarray platforms. There are some inconsistencies among data sets because of different experimental environments. For one data set, another data set is noised. We will test how DMT improves the classification accuracy when a classifier is trained by the data in one laboratory and tested on the data from another laboratory. A brief description of three data sets is listed in Table 1. We have preprocessed the Harvard, Michigan and Stanford data sets to make models built on them comparable. We have removed duplicated genes in the data sets since they correspond to different fragments of a gene and are unable to match across labs at the name level. We have removed genes that could not match genes from another laboratories. Finally, 1963 genes were left from each of three data sets. Three data sets were normalised by z-scores.

We firstly test the robustness of DMT in comparison with a single decision tree and other randomisation based ensemble methods, namely AdaBoost [12], Bagging [6], Random Forests (RF) [7], and Random Trees (RT) [10]. The number of interactions of AdaBoost is 100. The number of trees of Bagging, Random Forests and Random Trees is 100 respectively. We have used Weka implementation [15] for experiments. Our DMT has also been implemented as a Weka API and plugged to Weka. The number of diversified trees is set as 3, 7, 13 and 21. We have built a classifier model on one data set and test the model on two other data sets among Harvard, Michigan and Stanford data sets. Because of different experimental environments, test data sets are considered as being noised. The experimental results are listed in Table 2.
Figure 2: Test accuracy of multiple diversity trees in comparison to various methods at different noise level

In Table 2, we see that accuracies of single trees of Harvard to Michigan and Michigan to Harvard are higher than other since both laboratories have utilised the same Microarray platform, and the inconsistencies between them are small. In contrast, inconsistencies between other pairs are larger, and the accuracies of single trees are lower too. DMT and all other ensemble methods improve accuracies over the single tree. The improvement of DMT is the most significant. To further confirm this conclusion, we have conducted Wilcoxon signed ranks test [24] to compare the performance of these methods. Demšar [9] has recommended that Wilcoxon signed ranks test is a robust non-parametric test for statistical comparisons of different classifiers. Wilcoxon signed ranks test results are listed in Table 3. We see that the DMT tree is more robust than all other methods at significance level of 95%. The second most robust method is Bagging, but it is not significantly more robust than AdaBoost and Random Forests. Interestingly, AdaBoost is sensitive to noises. Its performance will deteriorate greatly with the increase of noises as shown in the following experiment.

To further test the robustness of DMT and various ensemble models, we have added noises on 5%, 10%, 20%, 30%, and 50% percent of randomly selected attributes. The added noises follow a \((0, \sigma)\) distribution, where \(\sigma\) is the standard deviation of a selected
attribute, implemented according to Box Muller transformation [5]. A reported result is the average of 100 tests on the noised test data sets. Test results of robustness of 7-DMT tree with other ensemble methods are listed in Figure 2 and the test results of robustness of various DMT trees are listed in Figure 3.

Figure 2 shows that DMT is consistently more robust than other ensemble classifiers with the increase of added noises. A statistical test has confirmed that DMT classifiers are significantly more accurate than other ensemble methods in noise added data. Among four ensemble methods, Random Forests and Bagging are the best. The performance of Random Forests is very stable with the increase of noises. Bagging is not as stable as Random Forests but its average accuracy is as high as that of Random Forests. The performance of Random Trees is also stable but it has a lower accuracy than Bagging and Random Forests. Adaboost performs inconsistently in noised data and this is consistent with previous results [19]. Adaboost sometimes performs badly, for example, in Michigan to Stanford pair. Consider that the number of DMT trees is small and each alternative tree is the heuristically best possible one on the complete data set. The interpretability of DMT trees are good. In contrast, a Bagging model contains 100 trees on the randomly sampled data sets and a Random Forests contains 100 random trees. The interpretability of each alternative tree in a Bagging model and a Random Forests is not as good as an alternative tree in DMT. Therefore, DMT is more robust and has better interpretability than other ensemble methods.

Figure 3 shows that all DMTs with different sizes are more robust than a single decision tree. Added noises do not affect their classification accuracies much. Look at two test results of Harvard to Michigan and Michigan to Harvard. Since both laboratories make use of the same Microarray platform, their data are quite consistent when there is no added noises. The performance of a single decision tree is quite good. However, with the increase of noises, the accuracy of a single tree reduces greatly. In contrast, DMT trees maintain a similar accuracy as for that of no noise added data. Added noises make slight difference in classification accuracies for DMTs, and hence DMT trees are robust. The robustness of DMT trees does not necessarily need a large number of alternative trees. In overall, 7-DMT and 13-DMT are the bets among all DMTs. This has been typified in results of Stanford to Harvard and Standard to Michigan. Theoretically, a large number of diversified trees will increase the robustness. However, there may not be the same high quality trees as the first few trees on a data set when the number of trees increases. Low quality trees reduce classification accuracies. Therefore, the number of DMT trees should not be large.

4.2 A counter example

A base of the robustness of DMT trees is that a data set supports a number of quality alternative trees. Thus is true for many high dimensional biological data sets. When this is not true, we should expect that DMT does not work. We have identified an example and show it in Figure 4.

Madelon data set from UCI ML repository [1] is an artificial dataset containing data points grouped in 32 clusters placed on the vertices of a five dimensional hypercube and randomly labeled as two classes. The data points are described by 500 features (attributes). The training data set contains 2000 data points (1000/1000) and the test data set contains 1800 (900/900) data points. The performance of MDT is bad in all aspects. The accuracy of un-noised and noised test data of DMT is lower than a single decision tree.

Let us compare the size of trees in the MDT classifiers of Madelon and of Harvard in Table 4. We see that tree size of Madelon is large in the first four trees, and very small after the fourth tree. Firstly, a large tree indicates that it is difficult to build a good classifier on the training data. The size increases in the subsequent trees and this shows that the classification becomes even more difficult in the remaining attributes. Secondly, a tree of size one classifies a test instance to the largest distributed class. In this data set, the classification is a random choice since both classes are equally distributed. With the above knowledge,
we easily understand the performance in Figure 4. The first tree is the most accurate one and all subsequent trees are less accurate than their previous tree in the first four trees. Therefore, all DMT classifiers are less accurate than the first decision tree. Further, a large tree is more likely affected by noises than a small tree. DMT tree performance deteriorates with the increase of noises. For DMTs whose sizes are greater than 8, their classification accuracy is around 50%.

| Tree size | 1st | 2nd | 3rd | 4th | 5th | 7th | 13th | 21th |
|-----------|-----|-----|-----|-----|-----|-----|------|------|
| Madelon   | 259 | 379 | 409 | 437 | 1   | 1   | 1    | 1    |
| Harvard   | 3   | 3   | 3   | 5   | 7   | 7   | 5    | 9    |

Table 4: Tree size of DMT on Madelon data set in comparison with Harvard data set.

In contrast, trees of Harvard data set are consistently small. This means that many alternative trees are very good for the classification of Harvard data set. Many quality small trees make DMT work. However, when good classification attributes are inadequate, alternative trees are significantly worse than the first tree. As a result, DMT does not work as in Madelon data set.
4.3 Variations of DMT

We have discussed a number of possible variations for the construction of DMT classifiers: simple vote, weighted vote with Laplace accuracy at a predictive leaf, and weighted vote with the support at a predictive leaf.

We test these variations on the 10 data sets with 3, 5, 11 and 21 trees. The 10 data sets have obtained from Kent Ridge [18] (the first six) and research literature [21, 23, 20, 8] (the last four). A brief description of the data sets is listed in Table 5. The 10 cross validation accuracies of various variations are listed in Figure 5.

From Figure 5, we see that the difference between simple vote and the Laplace weighted vote is marginal, and that the support weighted vote is worse than both. Therefore, simple vote is a simple yet accurate choice. In addition, all DMT classifiers in this experiment are significantly more accurate than a single decision tree without noises.

5 Related work

In this section, we review ensemble classification methods used in this paper.

Bagging [6], Boosting [13], Random Forests [7] and Random Trees [10] are four major randomisation based ensemble classification methods in the machine learning field. We present a brief discussions for each in the following.

Bagging was proposed by Leo Breiman [6]. Bagging uses a bootstrap technique to re-sample a training data set. Bootstrap sampling is random sampling with replacement. A set of alternative trees are built on a set of re-sampled data sets. Each tree will gives a predicted class to a coming test instance. The final predicted class for the test instance is determined by the majority predicted class by all alternative trees.
Boosting method was first developed by Freund and Schapire [13]. Boosting trains a sequence of classifiers on a set of data sets with different distribution ratios. The first classifier is constructed from the original data set where every record has an equal distribution ratio of 1. In the following training data sets, the distribution ratios are assigned differently among records. The distribution ratio of a record is reduced if the record has been correctly classified, and is increased otherwise. A weighted voting method is used in the committee decision.

Random Forests was proposed by Leo Breiman [7]. This method combines Bagging and Random Subspace methods [16]. A random forests classifier consists of a number of randomly generated trees. Each tree is constructed from a bootstrap sample of the original data set. At each node of a tree, the data partition attribute is selected from a subset of randomly chosen attributes.

Random Trees was proposed by Dietterich [10]. In random tree construction, at each node in the decision tree one of randomly selected attribute from the twenty best tests is selected to partition data at the node. With continuous attributes, it also produces twenty best splits of attributes, and the one used is randomly selected in the top 20.

All these randomisation-based ensemble methods are based on a large set of weak learners to improve classification accuracy of tree based classifiers. Their performance on un-noised data has been comprehensively evaluated at [2]. In contrast, DMT [17] makes use of a small set of disjunct strong trees and it has been shown to outperform AdaBoost, Bagging, and Random Forests for microarray data classification. In this paper, it has been shown to outperform AdaBoost, Bagging, Random Forests and Random Trees on noised test data sets.

6 Conclusions

This paper has demonstrated a Diversified Multiple Trees (DMT) approach is able to classify noised test data that is deviated from the training data better than some widely used ensemble methods. We have tested DMT on three real world biological data sets from different laboratories in comparison with four benchmark ensemble classifiers, and experimentally shown that DMT is significantly more robust than other benchmark ensemble classifiers on noised test data. We have also discussed a limitation of DMT when a training data set does not support many simple and quality decision trees. We have further demonstrated that the DMT is a simple yet effective design among a number of possible variations. DMT makes use of a small number of quality trees to tolerate noises in test data. It is more robust than other large ensemble classifiers and has better interpretability. It is promising in many real world applications where data dimensional is high and noises are present.
References

[1] Asuncion, A. and Newman, D.J. UCI machine learning repository, 2007. URL http://archive.ics.uci.edu/ml/.

[2] Banfield, Robert E., H., Lawrence O., B., Kevin W., and Kegelmeyer, W.P. A comparison of decision tree ensemble creation techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29:173–180, 2007. ISSN 0162-8828.

[3] Beer, D. and et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nature Medicine, 8(8):816–824, 2002.

[4] Bhattacharjee, A. and et al. Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses. Proceedings of the National Academy of Sciences, 98(24):13790–13795, 2001.

[5] Box, G. E. P. and Muller, M. E. A note on the generation of random normal deviates. The Annals of Mathematical Statistics, 29(2):610–611, 1958.

[6] Breiman, L. Bagging predictors. Machine Learning, 24(2):123–140, 1996.

[7] Breiman, L. Random forests. Machine Learning, 45(1):5–32, 2001.

[8] Chen, X. and et al. Gene expression patterns in human liver cancers. Molecular Biology of the Cell, 13:1929–1939, 2002.

[9] Demšar, J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7:1–30, 2006.

[10] Dietterich, T. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning, 40 (2):139–158, 2000.

[11] Ein-Dor, L. and Zuk, O. and Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proceedings of National Academy of Sciences, 103:59235928, 2006.

[12] Freund, Y. and Schapire, R. E. Experiments with a new boosting algorithm. In Proceedings of IEEE International Conference on Machine Learning, pp. 148–156, Bari, Italy, 1996.

[13] Freund, Y. and Schapire, R. E. Experiments with a new boosting algorithm. In International Conference on Machine Learning, pp. 148–156, 1996.

[14] Garber, M. and et al. Proceedings of the National Academy of Sciences, 98(24):13784–13789, 2001.

[15] Hall, Q. and Frank, E. and Holmes, G. and Pfahringer, B. and Reutemann, P. and Witten, I. The weka data mining software: An update. SIGKDD Explorations, 11(1), 2009.

[16] Ho, T. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8):832–844, 1998.

[17] Hu, H. and Li, J. and Wang, H. and Daggard, G. and Shi, M. A maximally diversified multiple decision tree algorithm for microarray data classification. In Proceedings of 2006 Workshop on Intelligent Systems for Bioinformatics (WISB 2006), pp. 35-38, Sydney, 2006. CRPIT 73, ACS.
[18] Li, J. and Liu, H.  
http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html.

[19] Long, P. M. and Servedio, R. A.  Random classification noise defeats all convex potential boosters.  In Proceedings of international conference on Machine learning (ICML), pp. 608–615; New York, NY, USA, 2008.  ACM. ISBN 978-1-60558-205-4.

[20] Ma, X. and et al.  A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.  Cancer Cell, 5:607–616, 2004.

[21] Pomeroy, S. L. and et al.  Prediction of central nervous system embryonal tumour outcome based on gene expression.  Nature, 415:436–442, 2002.

[22] Quinlan, J. R.  C4.5: Programs for Machine Learning.  Morgan Kaufmann, San Mateo, California, 1993.

[23] Singh1, D. and et al.  Gene expression correlates of clinical prostate cancer behavior.  Cancer Cell, 1:203–209, 2001.

[24] Triola, M. M. and Triola, M. F.  Biostatistics for the biological and health sciences.  Addison-Wesley, Boston, USA, 2nd edition, 2005.