

# Feature Selection for MicroRNA Target Prediction

## Comparison of One-Class Feature Selection Methodologies

Malik Yousef<sup>1,2</sup>, Jens Allmer<sup>3,4</sup> and Waleed Khalifa<sup>1,2</sup>

<sup>1</sup>Computer Science, The College of Sakhnin, Sakhnin, 30810, Israel

<sup>2</sup>The Institute of Applied Research, The Galilee Society, P.O. Box 437 ShefaAmr, 20200, Israel

<sup>3</sup>Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir, 35430, Turkey

<sup>4</sup>Bionia Incorporated, IZTEKGEB A8, Urla, Izmir, 35430, Turkey

**Keywords:** MicroRNA Targets, One-Class, Two-Classes, Machine Learning, Feature Selection.

**Abstract:** Traditionally, machine learning algorithms build classification models from positive and negative examples. Recently, one-class classification (OCC) receives increasing attention in machine learning for problems where the negative class cannot be defined unambiguously. This is specifically problematic in bioinformatics since for some important biological problems the target class (positive class) is easy to obtain while the negative one cannot be measured. Artificially generating the negative class data can be based on unreliable assumptions. Several studies have applied two-class machine learning to predict microRNAs (miRNAs) and their target. Different approaches for the generation of an artificial negative class have been applied, but may lead to a biased performance estimate. Feature selection has been well studied for the two-class classification problem, while fewer methods are available for feature selection in respect to OCC. In this study, we present a feature selection approach for applying one-class classification to the prediction of miRNA targets. A comparison between one-class and two-class approaches is presented to highlight that their performance are similar while one-class classification is not based on questionable artificial data for training and performance evaluation. We further show that the feature selection method we tried works to a degree, but needs improvement in the future. Perhaps it could be combined with other approaches.

## 1 INTRODUCTION

MicroRNAs (miRNAs) are short (~21 nt) nucleotide sequences that are either co-transcribed during transcription or are organized in intergenic regions with their own promoters. One or more mature miRNAs are split from ~70-100 nucleotide long pre-miRNAs (hairpins) which consist of double-stranded region (stem) containing one or more loops and bulges. Interaction of a miRNA with its target messenger RNAs (mRNAs) leads to repressing or translation and causes mRNA degradation (Bartel et al., 2004). It has been shown that this process depends on binding of the miRNA to the 3'UTR of the target, which is a region searched by most target programs. However, recent findings (Lytle et al., 2007) suggest that microRNAs may affect gene expression by also binding to 5' UTRs of mRNA.

## 1.1 Target Identification

Numerous computational approaches have been proposed for the prediction of miRNA's targets (Yousef et al., 2009). All of these methods depend on the parameterization of the miRNA:mRNA duplex and related information. Currently, sequence complementarity, thermodynamic calculations, and evolutionary conservation between species are used to predict the miRNA-target structure (Bartel et al., 2004; Lai, 2004). MiRanda (John et al., 2004), for example, uses dynamic programming to find the optimum alignment between mature miRNAs and their target genes. Another tool, RNAhybrid (Bartel et al., 2004; Lai, 2004), employs the prediction of RNA secondary structure (like the Mfold algorithm (Zuker, 2003)) to evaluate target propensity. TargetScanS (Lewis et al., 2003), scores target sites based on their evolutionary conservation using multiple genomes (e.g.: human, mouse, rat, dog, and chicken). Similarly, PicTar (Krek et al., 2005) is

based on a statistical method using genome-wide alignments of related species. TargetBoost (Saetrom et al., 2005) uses machine learning based on sequence information to create weighted sequence motifs that extract a profile describing binding characteristics between miRNAs and their targets. Likewise, Sung-Kyu et al. (Kim et al., 2005) and Yan, X., et al (Yan et al., 2007), used machine learning algorithms (SVM and ensemble learning, respectively) to predict miRNA-mRNA duplexes. MicroTar (Thadani and Tammi, 2006) is a statistical computational tool, which in contrast to many competitors does not use sequence homology for the prediction of miRNA's targets. RNA22 (Miranda et al., 2006) is a program based on pattern discovery, scanning UTR sequences for targets. Yousef et al. also employed machine learning, to develop the NBmiRTar program (Yousef et al., 2007) which does not require sequence conservation, but generates a model from sequence and structure information. For more information please refer to Table 1 and Xiao et al. (Fan and Kurgan, 2014) who conducted a comprehensive review and assessment of existing computational tools of microRNA targets prediction in animals.

Recently, we compared one-class and two-class approaches (Yousef et al., 2010) and concluded that the advantage of one class methods is that they don't require the generation of an arbitrary negative class.

Table 1: Summary of the computational tools used for the prediction of microRNA targets.

Tool Name	URL/ Reference
TargetScanS	<a href="http://genes.mit.edu/targetscan">http://genes.mit.edu/targetscan</a> (Lewis et al., 2005)
miRanda	<a href="http://www.microrna.org">http://www.microrna.org</a> (John et al., 2004)
PicTar	<a href="http://pictar.mdc-berlin.de/">http://pictar.mdc-berlin.de/</a> (Krek et al., 2005)
RNAhybrid	<a href="http://bibiserv.techfak.uni-bielefeld.de/rnahybrid">http://bibiserv.techfak.uni-bielefeld.de/rnahybrid</a> (Krüger and Rehmsmeier, 2006)
Diana-microT	<a href="http://diana.imis.athena-innovation.gr/DianaTools/index.php">http://diana.imis.athena-innovation.gr/DianaTools/index.php</a> (Kiriakidou et al., 2004)
Target Boost	<a href="http://www.interagon.com/demo">http://www.interagon.com/demo</a> (Saetrom et al., 2005)
Rna22	<a href="https://cm.jefferson.edu/rna22/">https://cm.jefferson.edu/rna22/</a> (Miranda et al., 2006)
MicroTar	<a href="http://tiger.dbs.nus.edu.sg/microtar/">http://tiger.dbs.nus.edu.sg/microtar/</a> (Thadani and Tammi, 2006)
NBmiRTar	<a href="http://wotan.wistar.upenn.edu/NBmiRTar">http://wotan.wistar.upenn.edu/NBmiRTar</a> (Yousef et al., 2007)
miRecords	<a href="http://mirecords.umn.edu/miRecords/">http://mirecords.umn.edu/miRecords/</a> (Xiao et al., 2009)

## 1.2 One Class Classification and Feature Selection

Supervised learning approaches for miRNA detection generally consider both positive and negative examples during training, testing, and application of the learned models. This binary (two-class) learning approach depends on the *a priori* knowledge of both classes in the examples. In contrast to binary learning strategies, one-class classification (OCC) uses only one class during training of the model. Anything not belonging to the class is rejected as an outlier by the trained model. For problems where the negative class cannot be unambiguously defined (e.g.: miRNA detection) one-class classification has received increasing attention (Crammer and Chechik, 2004; Gupta and Ghosh, 2005; Kowalczyk and Raskutti, 2002; Spinosa and Carvalho, 2005; Yousef et al., 2010) and has been successfully applied for example in text mining (Manevitz and Yousef, 2002), functional Magnetic Resonance Imaging (Thirion and Faugeras, 2004), signature verification (Koppel and Schler, 2004), and miRNA gene and target discovery (Yousef et al., 2008; Yousef et al., 2010). Recently, Khan and Madden, 2014 discussed OCC and presented a taxonomy based on the availability of training data, OCC algorithms, and the application domain.

Parameterization of biological information is of prime importance for proper learning, but often an abundance of features is derived. Therefore, feature selection, to determine the smallest subset of meaningful features, has to be performed. Feature selection for classification is well studied for two-class classification. Unfortunately, few methods are available for feature selection for OCC. Moreover, existing two-class feature selection methods may not be applicable to the OCC problem because they use two classes during feature ranking. Recently different studies suggest new or updated methods (Jeong et al., 2012; Lian, 2012; Lorena et al., 2014), for OCC feature selection such as the SVDD-radius-recursive feature elimination (Jeong et al., 2012).

Here, we present our feature selection approach for OCC for miRNA genes and target discovery and compare the results with two-class classification. Our feature selection approach leads to a good improvement of the OCC and for some cases it reached the performance of the two-class approach.

## 2 MATERIALS AND METHODS

### 2.1 MicroRNA Target Data

A collection of 326 confirmed MicroRNA targets (human, mouse, fruit fly, worm, and virus) were downloaded from the TarBase (Sethupathy et al., 2006) (TarBase\_V4, Tarbase flat file data as of 04/2007) web-site to serve as positive examples and 1,000 negative examples were chosen at random from the negative class pool generated for our previous study (Yousef et al., 2007).

### 2.2 Structure and Sequence Features Target

Feature extraction was done according to (Yousef et al., 2007). The miRNA:mRNA duplex was partitioned into two, the seed (5' 8 nt of the miRNA) and out-seed (3' remainder). 57 structural features were extracted for these parts. Here sequence features (words) are defined as short sequences having lengths equal to or less than 3 which leads to 84 features in total. The complete length of the feature vector was thus 141 (141 = 57 + 84). Supplementary Table 1 categorizes the features and Supplementary Table 5 presents the complete list of features and their ranking using 4 different methods (<http://bioinformatics.iyte.edu.tr/supplements/binfo2016>)

### 2.3 Feature Selection for One-Class Classification

Feature selection has been well studied for the two-class classification problem, while few methods are available for feature selection in respect to OCC. Unfortunately, existing two-class feature selection methods cannot be applied to feature selection for OCC because they also use two classes for the ranking of the features. Recently different studies suggest novel or updated methods for feature selection under the premise of OCC (Bailey and Elkan, 1994; Goymer, 2006; Hall et al., 2009; Lorena et al., 2014; Novak, 2006; Xuan et al., 2011). We considered these methods in this study but only compare to the Pearson method (Lorena et al., 2014) since no significant performance difference was seen among the suggested feature selection methods (Supplementary File 2 in <http://bioinformatics.iyte.edu.tr/supplements/binfo2016>). The Pearson correlation measure allows detection of linear relation among features. The pair-wise distances among all feature were calculated using

Pearson correlation. Features with lower correlation were preferred during feature selection.

### 2.4 Zero-Norm Feature Selection

We define for each feature's vector a zero-norm to be the non-zero values over all positive examples. Our approach to performing feature selection based on zero-norm is to remove a feature whose vector values are all zero. Moreover, we have defined a  $\#(v)$  as the number of values with non-zero value. For example if  $v = (0.4, 0, 0.6, 0, 0, 0.8, 0, 1, 1.4)$  the value of  $\#(v)$  is 5. Furthermore, we define different levels for thresholds of  $\#(v)$  to determine the relevance of a feature in order to remove it from the set of features (valid for both, the positive and the negative dataset). For example given a threshold of 3 for  $\#(v)$  the feature is not selected if  $\#(v)$  is less than the threshold. We considered the following thresholds: 3, 5, 10, 15, 20, 25, 30, and 35. Obviously, only positive data was used for feature selection.

### 2.5 One-Class Classifiers

Two-class classification depends on properly assigned examples from both the positive (miRNA) and negative (non-miRNA) classes in order to build a classifier that can effectively discriminate between them. One-class classification employs only the information of one class (target) during training of the model which then is able to recognize the examples belonging to that class and rejecting others as outliers.

Many one-class classification algorithms are available and we chose three one-class algorithms for comparison. In the following the algorithms will be briefly described, but more information is available in (Schölkopf et al., 2001; Tax, 2001). The LIBSVM library (Chang and Lin, 2011) was used for the implementation of the SVM-based (two-class) classifiers and DDtools (Tax, 2015) were used to implement all other selected OCCs. WEKA software (Witten et al., 2011) was used as implementation of the two-class classifiers enabling comparison with existing tools like the popular SVM method (Schölkopf et al., 1999; Vapnik, 1995). In the following OC-Gaussian (2.5.1), OC- $k$ Means (2.5.2), OC- $k$ NN (2.5.3), and for comparison the two-class classifiers NB (2.6.1), SVM (2.6.2), random forest (2.6.3), and C4.5 (2.6.4) are briefly described.

#### 2.5.1 One-Class Gaussian

This OCC algorithm (OC-Gaussian) uses a density estimation model which is under the assumption of a

multivariate normal distribution and that the probability density function can be calculated for a given test sample in  $n$ -dimensional space and compared to the training sample distribution (Yousef et al., 2007). Here we use  $g$  to depict the density.

### 2.5.2 One-Class $k$ Means

Kmeans is a well-known clustering algorithm which can partition data into  $k$  clusters. Using OC- $k$ Means we divide the data into  $k$  clusters. For an unknown sample  $z$  the distance  $d(z)$  is calculated to all  $k$  clusters. Generally, the class is assigned by returning the label of the closest cluster. In this case learned clusters are from the target class and thus if the unknown example is closer to the clusters than a threshold, they are assigned the target class or otherwise receive the label 'unknown'.

### 2.5.3 One-Class $k$ -Nearest Neighbour

As a modification of the two-class nearest neighbour classifier the one-class nearest neighbour classifier (OC- $k$ NN) learns from positive examples only. OCC- $k$ NN stores all positive training examples as its model. When classifying an unknown example  $z$ , the distance to its nearest neighbour  $y$  ( $y = \text{NN}(z)$ ) is calculated as  $d(z,y)$ . In case the distance to  $y$  is smaller than to any of  $y$ 's nearest neighbours, the example is classified as  $y$ . Here we consider the average distance of the  $k$  nearest neighbours in the OC- $k$ NN implementation.

## 2.6 Two Class Methods

Two-class classification methods were selected based on their popularity in bioinformatics. Currently, we see a rise of the use of random forest-based machine learning can be seen while a decrease for the use of simple decision trees (e.g.: C4.5) is apparent.

### 2.6.1 Naïve Bayes

Naïve Bayes is a classification algorithm based on posterior probabilities (Mitchell, 1997) and can, therefore, provide a probability for the membership of an unknown example. Important is the assumption that the features are conditionally independent given the class which may not hold in this case.

We used the Rainbow program (McCallum, 1996) to train the naïve Bayes classifier. To combine the numeric features identified in the miRNA-target duplex with the sequence features (words) in the target candidate sequence, a dictionary of all the

unique words was generated and the frequency of each word in the sequence was used.

### 2.6.2 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) have been employed in bioinformatics (Donaldson et al., 2003; Haussler, 1999; Pavlidis et al., 2001). Linear SVMs are usually defined as SVMs with linear kernel. The training data for linear SVMs could be linear non-separable and then soft-margin SVM could be applied. Linear SVM separates the two classes in the training data by producing the optimal separating hyper-plane with a maximal margin between the class 1 and class 2 samples; given a proper training set.

### 2.6.3 Random Forest

Random forests are an ensemble of tree predictors. Each tree depends on the values of a random vector sampled independently for all trees in the forest assuring same distribution for all trees (Breiman, 2001). The improvement in the classification accuracy is due to the growing of an ensemble of trees that vote for the most popular class. Random forests are becoming increasingly popular because of their ability to deal with small sample size and high-dimensional space.

### 2.6.4 C4.5

C4.5 is a decision tree algorithm, developed by (Quinlan, 1993). A decision tree is a simple structure where non-terminal nodes represent tests on one or more attributes and terminal nodes (leaves) reflect decision outcomes.

## 2.7 Classification Performance Evaluation

To evaluate classification performance, we used the data generated from the positive class and 1,000 negative examples. The negative class is not used for training of the one-class classifiers, but merely for estimating the classification specificity.

Each one-class algorithm was trained using 90% of the positive class and the remaining 10% were used for sensitivity evaluation. The randomly selected 1,000 negative examples were used for the evaluation of specificity. The whole process was repeated 100 times in order to evaluate the stability of the methods. The procedure is depicted as a flowchart in Figure 1.

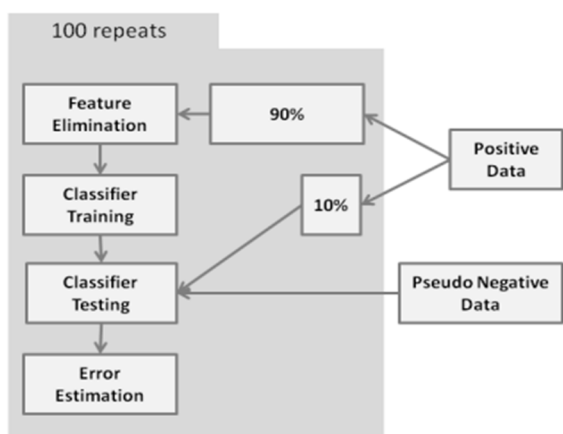


Figure 1: Training and testing procedure for the one-class classifiers. Competing two-class classifiers also received negative data during training.

### 3 RESULTS AND DISCUSSION

#### 3.1 Zero-Norm Feature Selection

Since the current standard approach in miRNA prediction is using two-class classification, the OCCs are compared to two-class classifiers.

Feature selection effectiveness is only presented for OCC since feature selection is well established for two-class classification. Table 2 shows the effectiveness of OC-*k*Means, OC-*k*NN, and OC-Gaussian for different number of selected features using the zero-norm feature selection method. Nine different feature set sizes were tested with their associated zero-norm thresholds. OC-*k*Means and OC-*k*NN achieved similar maximum accuracy (96.65 and 96.8, respectively) while OC-Gaussian was somewhat less accurate (94.4). In general, accuracy rises to the maximum and then decreases with the number of features although some outliers can be seen (Table 2). OC-*k*NN shows best performance for the unfiltered feature set at a *k* of four (Table 2). The number of clusters is generally at four, but for some feature sets it increases to six and even ten. The related method OC-*k*Means interestingly performs best in a range from 18-35 for *k* inversely related with the number of features. The lowest accuracy for OC-*k*Means (95.68) and OC-*k*NN (94.8) was still better than the best accuracy for OC-Gaussian. Perhaps, the density estimation for OC-Gaussian is not as effective as the clustering methods in OC-*k*Means and OC-*k*NN. This is also seen by the range of accuracies achieved by these different methods.

The accuracy spread for OC-*k*Means (~1) and

OC-*k*NN (2) is lower than OC-Gaussian (8). Given this data and feature selection method, OC-Gaussian doesn't seem to be performing well. On the other hand, OC-*k*Means and OC-*k*NN perform well and feature selection was most effective for OC-*k*Means.

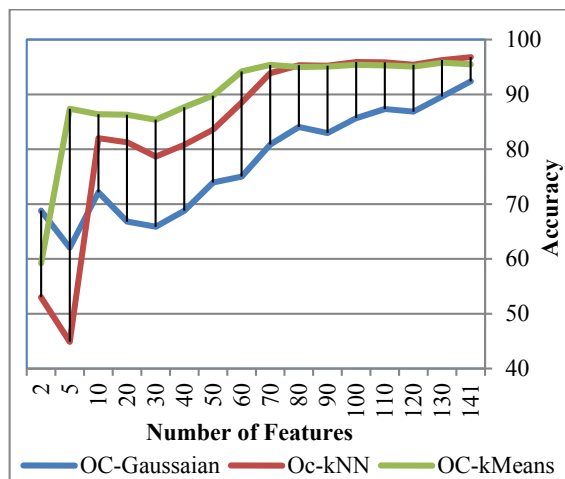


Figure 2: Classifier performance in respect to number of selected features using Pearson feature selection.

It is always possible that several features that individually have not much discriminative power together are very potent at discriminating among classes. This problem cannot be captured with the feature selection method employed here. Unfortunately, feature selection is NP-hard (Amaldi and Kann, 1998) and it is not possible to test all combination of features at all *k* or *g*. This may explain the outliers that can be seen in Table 2.

#### 3.2 Pearson-based Feature Selection

For comparison with our zero-norm method, we performed feature selection results using the Pearson approach (Figure 2). OC-*k*NN achieves highest accuracy (96.82) followed by OC-*k*Means (95.8) and OC-Gaussian (92.39). The accuracy spread for the selected features between 60 and 141 is 6 for OC-*k*Means, ~13 for OC-*k*NN, and ~19 for OC-Gaussian (details in Supplementary Table 2 in <http://bioinformatics.iyte.edu.tr/supplements/binfo2016>).

Most interestingly, the Pearson method shows best performance for OC-*k*NN and OC-Gaussian with all features (141) selected. This indicates that for our data the Pearson-base feature selection method was not successful. For OC-*k*Means some feature selection was achieved, however, which shows that in principle the Pearson-based method was correctly applied. Whether this is a problem of the features, the data, or a weakness of the feature selection

Table 2: The accuracy performance of one-class classifiers in respect to selected features at different zero-norm thresholds. Number of clusters ( $k$ ) was tested between 1 and 150 and density ( $g$ ) between 0.01 and 1. Highest accuracy is highlighted in grey. ACC: accuracy, SE: sensitivity, SP: specificity.

Features	Zero-norm Thres -hold	OC- $k$ Means				OC- $k$ NN				OC-Gaussian			
		$k$	ACC	SE	SP	$k$	ACC	SE	SP	$g$	ACC	SE	SP
141	0	18	96.2	82.8	96.6	4	96.8	90.3	97.0	0.6	92.5	82.5	92.8
119	3	19	96.1	80.7	96.6	4	96.7	90.6	96.9	0.4	94.4	80.0	94.8
113	5	20	96.12	80.5	96.62	6	96.0	90.1	96.2	0.5	92.7	83.1	93.0
101	10	25	96.65	80.93	97.15	4	96.3	89.6	96.5	0.5	90.5	83.6	90.7
90	15	20	95.1	78.15	96.9	4	95.8	90.8	96.0	0.3	92.6	81.4	93.0
81	20	25	96.08	79.18	96.62	4	95.7	89.7	95.9	0.3	91.5	81.8	91.8
75	25	25	95.95	79.72	96.46	4	95.4	90.8	95.6	0.2	91.9	80.3	92.2
66	30	25	95.75	79.93	96.25	6	94.8	90.7	95.0	0.2	89.4	83	89.6
58	35	25	95.68	79.38	96.2	10	95.2	91.1	95.3	0.2	86.4	84.1	86.4

methodology cannot be deduced from our study, but in the future we will investigate this issue further. For OC- $k$ Means the accuracy doesn't seem to have a correlation with the number of features while for OC- $k$ NN and OC-Gaussian there is a minimum between the two maxima with steady decrease and increase correlated with the number of features.

While the results that this method achieves are somewhat comparable to our zero-norm feature selection method, OC- $k$ Means achieves about 1% less accuracy for Pearson feature selection when compared to zero-norm feature selection. The performance of OC- $k$ NN is the same for both feature selection algorithms and OC-Gaussian performs about 2% worse when using Pearson-based feature selection.

In summary, feature selection methods can be effective but work differently. A combination of methods may, therefore, be more successful than the methods that were compared here. Compared to the Pearson method, the zero-norm approach appeared to be more stable.

The separation of positive versus 'unknown' class is better when fewer features are used in training (Figure 3). The linear projection in Figure 3 visually confirms that the classes are better separated in fewer dimensions, however, the separation could be better and selection of proper features may improve the situation.

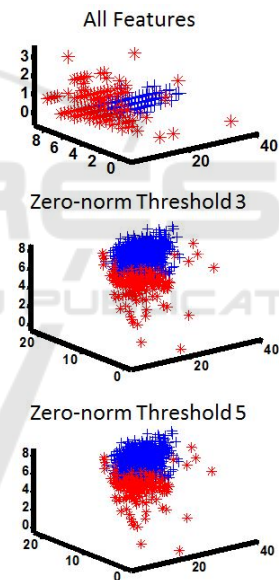


Figure 3: Linear projection of high dimensional data into two dimensions with red representing the positive and blue representing the 'unknown' class. Top pane shows all features (141), middle pane shows projection when features are selected using a zero-norm threshold of 3 (119 features), and the last pane shows the distribution after filtering using a threshold of 5 (113 features).

### 3.3 Comparison to Two-Class Classification

Since two-class classification is the *de facto* standard for miRNA target prediction, there is a need to compare the accuracy achieved using OCC to the one

that can be reached using two-class classification. We chose a few representative two-class classification algorithms which are also popular in bioinformatics.

Unfortunately, any comparison between one-class and two class classification’s effectiveness must remain biased unless perfectly known examples are available for both classes. For miRNA target prediction, this is not possible since only when a miRNA is co-expressed with its target can the effect be measured on the protein level. This in turn means that any genomic sequence can be a target until it was shown that it is not via co-expression with all known miRNAs. This is extremely difficult for any organism and futile for any higher organisms.

Table 3: Accuracy performance of two-class classification algorithms in respect to selected features using the zero-norm feature selection approach. ACC: accuracy, #: number (Further details in Supplementary Table 4, <http://bioinformatics.iyte.edu.tr/supplements/binfo2016/>).

Feat - ures	Zero-Norm Thres-hold	Random Forest	LIB-SVM	kNN	C4.5
#		ACC	ACC	ACC	ACC
119	3	98.53	99.30	93.33	96.38
113	5	98.51	99.27	93.19	96.4
101	10	98.63	99.36	93.38	96.31
90	15	98.61	99.46	93.96	96.49
81	20	98.78	99.45	93.98	96.44
75	25	98.71	99.48	93.99	96.13
66	30	98.71	99.48	94.83	96.83
58	35	98.79	99.52	94.77	96.4

Since truly negative data cannot be determined, it is necessary to create an artificial negative dataset for training and testing of the two-class classifiers. This would pose no problem if targeting would be understood in detail which would make machine learning unnecessary. Thus, for all artificial negative datasets the content of false negative data is unknown. This causes a bias when comparing to one class classification where the positive class is comparably well defined.

Nonetheless, a comparison may be informative and, therefore, we performed two-class classification on the same data, using the same features, as we employed for OCC. Additionally, we employ the same zero-norm feature selection method (Table 3). Table 3 shows the performance of the two-class classification algorithms when feature selection is only based on the information from the positive class

(exactly as done for OCC, above). There doesn’t seem to be a clear influence of the feature selection on the performance of the two-class classification results and down to 58 features they keep their accuracy more or less constant within a range of less than 1% accuracy.

The best accuracy for the two-class classification methods are achieved with fewer features, something we would have expected to see in the OCC analysis as well. Overall, the performance of the two-class classification seems to be better than the OCC results. C4.5 and kNN perform worse than OC-kNN and equal to OC-kMeans. Random forest and SVM are up to ~2.5% more accurate than the OCC models. This view is likely biased since the OCC accuracy measures were established using artificial negative data. This will overestimate the accuracy of the two-class classification and underestimate the accuracy of the one-class models.

## 4 CONCLUSIONS

We intended to show that one-class classification can be used for miRNA target prediction. For this we first attempted feature selection and were partially successful. From our data it seems clear that proper selection of features is important and has a positive influence on classification accuracy (Tables 2 and 3). A number of problems complicate feature selection, however:

- 1) not all features are known
- 2) negative data is artificial and of unknown quality
- 3) feature selection is NP-hard and many features have already been proposed.

Since feature selection methods perform differently (Table 2 and Figure 2), we are optimistic that a combination of feature selection methods may in the future propose a minimal feature set with maximum accuracy. We believe that problems 1) and 2) can be solved, but have no hope for the third issue.

It was our aim to point out that relying on artificial negative data may be dangerous and that OCC can perform at a similar accuracy as two-class classification despite biased accuracy estimates. The performance difference among methods is only up to ~2.5% (Tables 2 and 3).

The current results show that it is possible to train a classifier based only on positive examples yielding a competitive performance. Moreover, using zero-norm feature selection with the one-class approaches is able to improve the performance and approach to two-class performance levels. Clearly OCC is more

sensitive to non-relevant features than two-class classification. However, since the process of obtaining reliable biological data that defines the negative class is a time consuming, if not an impossible, endeavour a successful application of OCC can reduce this cost and provide important tools for classification of biological data and prediction of unknown data.

## ACKNOWLEDGEMENTS

The work was supported by the Scientific and Technological Research Council of Turkey [grant number 113E326] to JA.

## REFERENCES

- Amaldi, E., and Kann, V. (1998). On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1-2), 237–260. doi:10.1016/S0304-3975(97)00115-1.
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 2, 28–36. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7584402>.
- Bartel, D. P., Lee, R., and Feinbaum, R. (2004). MicroRNAs : Genomics , Biogenesis , Mechanism , and Function Genomics : The miRNA Genes, 116, 281–297.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324.
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1–27. doi:10.1145/1961189.1961199.
- Cramer, K., and Chechik, G. (2004). A needle in a haystack: local one-class optimization. In R. Greiner and D. Schuurmans (Eds.), *Proceedings of the 21st International Conference on Machine Learning (ICML-04)*. Retrieved from <http://www.machinelearning.org/proceedings/icml2004/papers/239.ps>.
- Donaldson, I., Martin, J., de Bruijn, B., Wolting, C., Lay, V., Tuekam, B., ... Hogue, C. W. V. (2003). PreBIND and Textomy--mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics*, 4, 11.
- Fan, X., and Kurgan, L. (2014). Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Briefings in Bioinformatics*. doi:10.1093/bib/bbu044.
- Goymer, P. (2006). Different treatment. *Nature Reviews Cancer*, 6(2), 94–95. doi:10.1038/nrc1808.
- Gupta, G., and Ghosh, J. (2005). Robust one-class clustering using hybrid global and local search. In *Proceedings of the 22nd international conference on Machine learning - ICML '05* (pp. 273–280). New York, New York, USA: ACM Press. doi:10.1145/1102351.1102386.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, 11(1), 10. doi:10.1145/1656274.1656278.
- Hausler, D. (1999). Convolution Kernels on Discrete Structures. In *Technical Report UCSCRL9910 UC*, 23(1), 1–38. Retrieved from <http://eprints.kfupm.edu.sa/32597/>
- Jeong, Y.-S., Kang, I.-H., Jeong, M.-K., and Kong, D. (2012). A New Feature Selection Method for One-Class Classification Problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1500–1509. doi:10.1109/TSMCC.2012.2196794.
- John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biology*, 2(11), e363. doi:10.1371/journal.pbio.0020363.
- Khan, S. S., and Madden, M. G. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review*, 29(03), 345–374. doi:10.1017/S026988891300043X.
- Kim, S.-K., Nam, J.-W., Lee, W.-J., and Zhang, B.-T. (2005). A Kernel Method for MicroRNA Target Prediction Using Sensible Data and Position-Based Features. In *2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (pp. 1–7). IEEE. doi:10.1109/CIBCB.2005.1594897.
- Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes and Development*, 18(10), 1165–1178. doi:10.1101/gad.1184704.
- Koppel, M., and Schler, J. (2004). Authorship verification as a one-class classification problem. In *Twenty-first international conference on Machine learning - ICML '04* (p. 62). New York, New York, USA, Alberta, Canada: ACM Press. doi:10.1145/1015330.1015448.
- Kowalczyk, A., and Raskutti, B. (2002). One Class SVM for Yeast Regulation Prediction. *SIGKDD Explorations*, 4(2), 99–100.
- Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., ... Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nature Genetics*, 37(5), 495–500. doi:10.1038/ng1536.
- Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*, 34(Web Server issue), W451–454. doi:10.1093/nar/gkl243.



- Lai, E. C. (2004). Predicting and validating microRNA targets. *Genome Biology*, 5(9), 115. doi:10.1186/gb-2004-5-9-115.
- Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1), 15–20. doi:10.1016/j.cell.2004.12.035.
- Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7), 787–798. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14697198>.
- Lian, H. (2012). On feature selection with principal component analysis for one-class SVM. *Pattern Recognition Letters*, 33(9), 1027–1031. doi:10.1016/j.patrec.2012.01.019.
- Lorena, L. H. N., Carvalho, A. C. P. L. F., and Lorena, A. C. (2014). Filter Feature Selection for One-Class Classification. *Journal of Intelligent and Robotic Systems*, 1–17. doi:10.1007/s10846-014-0101-2.
- Lytle, J. R., Yario, T. A., and Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23), 9667–9672. doi:10.1073/pnas.0703820104.
- Manevitz, L. M., and Yousef, M. (2002). One-Class SVMs for Document Classification. *The Journal of Machine Learning Research*, 2, 139–154. Retrieved from <http://dl.acm.org/citation.cfm?id=944808>.
- McCallum, A. K. (1996). Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. Retrieved from <http://www.cs.cmu.edu/~mccallum/bow>.
- Miranda, K. C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A. M., ... Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, 126(6), 1203–17. doi:10.1016/j.cell.2006.07.031.
- Mitchell, T. (1997). *Machine Learning*.
- Novak, K. (2006). Taking out the trash. *Nature Reviews Cancer*, 6(2), 92–92. doi:10.1038/nrc1807.
- Pavlidis, P., Weston, J., Jinsong, C., and Grundy, W. N. (2001). Gene functional classification from heterogeneous data. In *Proceedings of the Fifth International Conference on Computational Molecular Biology* (pp. 242–248). Retrieved from <https://noble.gs.washington.edu/papers/exp-phylo.pdf>.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Saetrom, O., Snøve, O., and Saetrom, P. (2005). Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms. *RNA*, 11(7), 995–1003. doi:10.1261/rna.7290705.
- Schölkopf, B., Burges, C. J. C., and Smola, A. J. (1999). *Advances in Kernel Methods*. Cambridge, MA: MIT Press.
- Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution. *Neural Comp.*, 13(7), 1443–1471.
- Sethupathy, P., Corda, B., and Hatzigeorgiou, A. G. (2006). TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, 12(2), 192–7. doi:10.1261/rna.2239606.
- Spinosa, E. J., and Carvalho, A. C. P. L. F. de. (2005). Support vector machines for novel class detection in Bioinformatics. *Genetics and Molecular Research [electronic Resource] : GMR.*, 4(3), 608–615.
- Tax, D. M. J. (2001). *One-class classification*. Technical University Delft. Retrieved from ISBN: 90-75691-05-x.
- Tax, D. M. J. (2015). DDtools, the Data Description Toolbox for Matlab.
- Thadani, R., and Tammi, M. T. (2006). MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics*, 7 Suppl 5, S20. doi:10.1186/1471-2105-7-S5-S20.
- Thirion, B., and Faugeras, O. (2004). Feature characterization in fMRI data: The Information Bottleneck approach. *Medical Image Analysis*, 8(4), 403–419. doi:10.1016/j.media.2004.09.001.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, New York, USA: Springer-Verlag. Retrieved from <http://dl.acm.org/citation.cfm?id=211359>.
- Witten, I. H., Frank, E., and Hall, M. A. (2011). Introduction to Weka. In *Data Mining: Practical Machine Learning Tools and Techniques* (pp. 403–406). Elsevier. doi:10.1016/B978-0-12-374856-0.00010-9.
- Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*, 37(Database issue), D105–10. doi:10.1093/nar/gkn851.
- Xuan, P., Guo, M., Liu, X., Huang, Y., Li, W., and Huang, Y. (2011). PlantMiRNAPred: efficient classification of real and pseudo plant pre-miRNAs. *Bioinformatics (Oxford, England)*, 27(10), 1368–76. doi:10.1093/bioinformatics/btr153.
- Yan, X., Chao, T., Tu, K., Zhang, Y., Xie, L., Gong, Y., ... Peng, X. (2007). Improving the prediction of human microRNA target genes by using ensemble algorithm. *FEBS Letters*, 581(8), 1587–93. doi:10.1016/j.febslet.2007.03.022.
- Yousef, M., Jung, S., Kossenkov, A. V., Showe, L. C., and Showe, M. K. (2007). Naïve Bayes for microRNA target predictions--machine learning for microRNA targets. *Bioinformatics (Oxford, England)*, 23(22), 2987–92. doi:10.1093/bioinformatics/btm484.
- Yousef, M., Jung, S., Showe, L. C., and Showe, M. K. (2008). Learning from positive examples when the negative class is undetermined--microRNA gene identification. *Algorithms for Molecular Biology*, 3, 2. doi:10.1186/1748-7188-3-2.
- Yousef, M., Najami, N., and Khalifa, W. (2010). A Comparison Study Between One-Class and Two-Class Machine Learning for MicroRNA Target Detection. *Journal of Biomedical Science and Engineering*.

- Yousef, M., Showe, L., and Showe, M. (2009). A study of microRNAs in silico and in vivo: Bioinformatics approaches to microRNA discovery and target identification. *FEBS Journal*. doi:10.1111/j.1742-4658.2009.06933.x.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415. doi:10.1093/nar/gkg595.

