A New Ensemble Semi-supervised Self-labeled Algorithm

Ioannis Livieris
Department of Computer & Informatics Engineering
Technological Educational Institute of Western Greece, Greece, GR 263-34
E-mail: livieris@teiwest.gr

Keywords: semi-supervised methods, self-labeled, ensemble methods, classification, voting

Received: March 13, 2018

As an alternative to traditional classification methods, semi-supervised learning algorithms have become a hot topic of significant research, exploiting the knowledge hidden in the unlabeled data for building powerful and effective classifiers. In this work, a new ensemble-based semi-supervised algorithm is proposed which is based on a maximum-probability voting scheme. The reported numerical results illustrate the efficacy of the proposed algorithm outperforming classical semi-supervised algorithms in term of classification accuracy, leading to more efficient and robust predictive models.

Povzetek: Razvit je nov delno nadzorovani učni algoritem s pomočjo ansamblov in glasovalno shemo na osnovi največje verjetnosti.

1 Introduction

The development of a powerful and accurate classifier is considered as one of the most significant and challenging tasks in machine learning and data mining [3]. Nevertheless, it is generally recognized that the key to recognition problems does not lie wholly in any particular solution since no single model exists for all pattern recognition problems [28, 15].

During the last decades, in the area of machine learning the development of an ensemble of classifiers has been proposed as a new direction for improving the classification accuracy. The basic idea of ensemble learning is the combination of a set of diverse prediction models, each of which solves the same original task, in order to obtain a better composite global model with more accurate and reliable estimates or decisions than can be obtained from using a single model [9, 28]. Therefore, several prediction models have been proposed based on ensembles techniques which have been successfully utilized to tackle difficult real-world problems [31, 14, 32, 30, 23, 27, 11]. Traditional ensemble methods usually combine the individual predictions of supervised algorithms which utilize only labeled data as training set. However, in most real-world classification problems, the acquisition of sufficient labeled samples is cumbersome and expensive and frequently requires the efforts of domain experts. On the other hand, unlabeled data are fairly easy to obtain and require less effort of experienced human annotators.

Semi-supervised learning algorithms constitute the appropriate and effective machine learning methodology for extracting useful knowledge from both labeled and unlabeled data. In contrast to traditional classification approaches, semi-supervised algorithms utilize a large amount of unlabeled samples to either modify or reprioritize the hypothesis obtained from labeled samples in order to build an efficient and accurate classifier. The general assumption of these algorithms is to leverage the large amount of unlabeled data in order to reduce data sparsity in the labeled training data and boost the classifier performance, particularly focusing on the setting where the amount of available labeled data is limited. Hence, these methods have received considerable attention due to their potential for reducing the effort of labeling data while still preserving competitive and sometimes better classification performance (see [18, 6, 7, 38, 17, 16, 21, 20, 22, 44, 45, 46, 43] and the references therein). The main issue in semi-supervised learning is how to exploit the information hidden in the unlabeled data. In the literature, several approaches have been proposed each with different philosophy related to the link between the distribution of labeled and unlabeled data [46, 4, 36].

Self-labeled methods constitute semi-supervised methods which address the shortage of labeled data via a self-learning process based on supervised prediction models. The main advantages of this class of methods are their simplicity and their wrapper-based philosophy. The former is related to the facility/comodity of application and implementation while the latter refers to the fact that any supervised classifier can be utilized, independent of its complexity [35]. In the literature, self-labeled methods are divided into self-training [41] and co-training [4]. Self-training constitutes an efficient semi-supervised method which iteratively enlarges the labeled training set by adding the most confident predictions of the utilized supervised classifier. The standard co-training method splits the feature space into two different conditionally independent views. Subsequently, it trains one classifier in each specific view and the classifiers teach each other the most confidently predicted examples. More sophisticated and advanced variants of this method do not require explicit feature splits or the iterative mutual-teaching procedure imposed by co-training, as they are commonly based on disagreement-based classifiers [44, 12, 36, 46, 45]

By taking these into consideration, ensemble methods and semi-supervised methods constitute two significant classes of methods. The former attempt to achieve strong classification performance by combining individual classifiers while the later attempt to enhance the performance of a classifier by exploiting the information in the unlabeled data. Although both methodologies have been efficiently applied to a variety of real-world problems during the last decade, they were almost developed separately. In this context, Zhou [43] advocated that ensemble learning and semi-supervised learning are indeed beneficial to each other and stronger learning machines can be generated by leveraging unlabeled data with the combination of diverse classifiers. More specifically, ensemble learning could be useful to semi-supervised learning since an ensemble of classifiers could be more accurate than an individual classifier. Additionally, semi-supervised learning could assist ensemble learning since unlabeled data can enhance the diversity of the base learner which constitute the ensemble and increase the ensemble's classification accuracy.

In this work, a new ensemble semi-supervised self-labeled learning algorithm is proposed. The proposed algorithm combines the individual predictions of three of the most representative SSL algorithms: Self-training, Co-training and Tri-training via a maximum-probability voting scheme. The efficiency of the proposed algorithm is evaluated on various standard benchmark datasets and the reported experimental results illustrate its efficacy in terms of classification accuracy, leading to more efficient and robust prediction models.

The remainder of this paper is organized as follows: Section 3 presents some elementary semi-supervised learning definitions and Section 4 presents a detailed description of the proposed algorithm. Section 5 presents the experimental results of the comparison of the proposed algorithm with the most popular semi-supervised classification methods on standard benchmark datasets. Finally, Section 6 discusses the conclusions and some research topics for future work.

2 Related work

Semi-Supervised Learning (SSL) and Ensemble Learning (EL) constitute machine learning techniques which were independently developed to improve the performance of existing learning methods, though from different perspectives and methodologies. SSL provides approaches to improve model generalization performance by exploiting unlabeled data; while EL explores the possibility of achieving the same objective by aggregating a group of learners. Zhou [43] presented an extensive analysis of how semi-supervised learning and ensemble learning can be efficiently fuse for the development of efficient prediction models. A number of rewarding studies which fuse and exploit their advantages have been carried out in recent years; some useful outcomes of them are briefly presented below.

Zhou and Goldman [42] have adopted the idea of ensemble learning and majority voting and proposed a new SSL algorithm which is based on the multi-learning approach. More specifically, this algorithm utilizes multiple algorithms for producing the necessary information and endorses a voted majority process for the final decision, instead of asking for more than one views of the corresponding data.

Along this line, Li and Zhou [17] proposed another algorithm, in which a number of Random trees are trained on bootstrap data from the dataset, named Co-Forest. The main idea of this algorithm is the assignment of a few unlabeled examples to each Random tree during the training process. Eventually, the final decision is composed by a simple majority voting. Notice that the utilization of Random Tree classifier for random samples of the collected labeled data is the main reason why the behavior Co-Forest is efficient and robust although the number of the available labeled examples is reduced. Xu et al. [40] applied this method for the predictions of protein subcellular localization providing some promising results.

Sun and Zhang [34] attempted to combine the advantages of multiple-view learning and ensemble learning for semi-supervised learning. They proposed a novel multiple-view multiple-learner framework for semi-supervised learning which adopted a co-training based learning paradigm in enlarging labeled data from a much larger set of unlabeled data. Their motivation is based on the fact that the use of multiple views is promising to promote performance compared with single-view learning because information is more effectively exploited; while at the same time, as an ensemble of classifiers is learned from each view, predictions with higher accuracies can be obtained than solely adopting one classifier from the same view. The experiments conduced on several datasets presented some encouraging results, illustrating the efficacy of the proposed method.

Roy et al. [29] presented a novel approach by utilizing a multiple classifier system in the SSL framework instead of using a single weak classifier for change detection in remotely sensed images. The proposed algorithm during the iterative learning process uses the agreement between all the classifiers which constitute the ensemble for collecting the most confident labeled patterns. The effectiveness of the proposed technique was presented by a variety of experiments carried out on multi-temporal and multi-spectral datasets.

In more recent works, Livieris et al. [21] proposed a new ensemble-based semi-supervised method for the prognosis of students' performance in the final examinations. They incorporated a ensemble of classifiers as base learner in the semi-supervised framework. Based on their numerical experiments, the authors concluded that ensemble methods and semi-supervised methodologies could efficiently combined to develop efficient prediction models. Motivated by the previous work, Livieris et al. [22] presented a new ensemble-based semi-supervised learning algorithm for the classification of chest X-rays of tuberculosis, presenting some encouraging results.

3 A review on semi-supervised self-labeled classification

In this section, we present a formal definition of the semi-supervised classification problem and briefly describe the most relevant self-labeled approaches proposed in the literature. Let $x_p = (x_{p1}, x_{p2}, \ldots, x_{pD}, y)$ be an example, where x_p belongs to a class y and a D-dimensional space in which x_{pi} is the i-th attribute of the p-th sample. Suppose L is a labeled set of N_L instances x_p with y known and U is an unlabeled set of N_U instance x_q with y unknown. Notice that the set $L \cup U$ consists the training set. Moreover, there exists a test set T of N_T unseen instances where y is unknown, which has not been utilized in the training stage. Notice that the aim of the semi-supervised classification is to obtain an accurate and robust learning hypothesis with the use of the training set.

Self-labeled techniques constitute a significant family of classification methods which progressively classify unlabeled data based on the most confident predictions and utilize them to modify the hypothesis learned from labeled samples. Therefore, the methods of this class accept that their own predictions tend to be correct, without making any specific assumptions about the input data. In the literature, a variety of self-labeled methods has been proposed each with different philosophy and methodology on exploiting the information hidden in the unlabeled data. In this work, we focus our attention to Self-training, Co-training and Tri-training which constitute the most efficient and commonly used self-labeled methods [21, 20, 22, 35, 37, 36].

3.1 Self-Training

Self-training [41] is generally considered as the simplest and one of the most efficient SSL algorithms. This algorithm is a wrapper based SSL approach which constitutes an iterative procedure of self-labeling unlabeled data. According to Ng and Cardie [25] "*self-training is a single-view weakly supervised algorithm*" which is based on its own predictions on unlabeled data to teach itself. Firstly, an arbitrary classifier is initially trained with a small amount of labeled data, constituting its training set which is iteratively augmented using its own most confident predictions of the unlabeled data. More analytically, each unlabeled instance which has achieved a probability over a specific threshold *ConLev* is considered sufficiently reliable to be added to the labeled training set and subsequently the classifier is retrained.

Clearly, the success of Self-training is heavily depended on the newly-labeled data based on its own predictions, hence its weakness is that erroneous initial predictions will probably lead the classifier to generate incorrectly labeled data [46]. A high-level description of Self-training algorithm is presented in Algorithm 1.

Algorithm 1: **Self-training**

Input: L − Set of labeled instances.
U − Set of unlabeled instances.
ConLev − Confidence level.
C − Base learner.

Output: Trained classifier.

1 : **repeat**
2 : Train C on L.
3 : Apply C on U.
4 : Select instances with a predicted probability more than *ConLev* per iteration (x_{MCP}).
5 : Remove x_{MCP} from U and add to L.
6 : **until** some stopping criterion is met or U is empty.

3.2 Co-training

Co-training [4] is a SSL algorithm which utilizes two classifiers, each trained on a different view of the labeled training set. The underlying assumptions of the Co-training approach is that feature space can be split into two different conditionally independent views and that each view is able to predict the classes perfectly [33]. Under these assumptions, two classifiers are trained separately for each view using the initial labeled set and then iteratively the classifiers augment the training set of the other with the most confident predictions on unlabeled examples.

Essentially, Co-training is a "*two-view weakly supervised algorithm*" since it uses the self-training approach on each view [25]. Blum and Mitchell [4] have extensively studied the efficacy of Co-training and they concluded that if the two views are conditionally independent, then the use of unlabeled data can significantly improve the predictive accuracy of a weak classifier. Nevertheless, the assumption about the existence of sufficient and redundant views is a luxury hardly met in most real world scenarios. Algorithm 2 presents a high-level description of Co-training algorithm.

Algorithm 2: **Co-training**

Input: L − Set of labeled instances.
U − Set of unlabeled instances.
C_i − Base learner ($i = 1, 2$).

Output: Trained classifier.

Create a pool U' of u examples by randomly choosing from U.
repeat
    Train C_1 on $L(V_1)$.
    Train C_2 on $L(V_2)$.
    for each classifier C_i **do** $(i = 1, 2)$
        C_i chooses p samples (P) that it most confidently labels as positive and n instances (N) that it most confidently labels as negative from U.
        Remove P and N from U'.
        Add P and N to L.
    end for
    Refill U' with examples from U to keep U' at constant size of u examples.
until some stopping criterion is met or U is empty.

Remark: V_1 and V_2 are two feature conditionally independent views of instances.

3.3 Tri-Training

Tri-Training [44] consists of an improved version of Co-Training which overcomes the requirements for multiple sufficient an redundant feature sets. This algorithm constitutes a bagging ensemble of three classifiers, trained on the data subsets generated through bootstrap sampling from the original labeled training set. In case two of the three classifiers agree on the categorization of an unlabeled instance, then this is considered to be labeled and augment the third classifier with the newly labeled example. The efficiency of the training process is based on the strategy the "*majority teach minority*" which avoids the use of a complicated time consuming approach to explicit measure the predictive confidence, serving as an implicit confidence measurement,

In contrast to several SSL algorithms, Tri-training does not require different supervised algorithms as base learners which leads to greater applicability in many real world classification problems [12, 46, 19]. A high-level description of Tri-training is presented in Algorithm 3.

Algorithm 3: **Tri-training algorithm**

Input: L − Set of labeled instances.
U − Set of unlabeled instances.
C_i − Base learner ($i = 1, 2, 3$).

Output: Trained classifier.

for $i = 1, 2, 3$ **do**
    $S_i = \text{BootstrapSample}(L)$.
    Train C_i on S_i.
end for

repeat
    for $i = 1, 2, 3$ **do**
        $L_i = \emptyset$.
        for $u \in U$ **do**
            if $C_j(u) = C_k(u)$ **then** $(j, k \neq i)$
                $L_i = L_i \cup (u, C_j(u))$.
            end if
        end for
    end for
    for $i = 1, 2, 3$ **do**
        Train C_i on S_i.
    end for
until some stopping criterion is met or U is empty.

4 An ensemble semi-supervised self-labeled algorithm

In this section, the proposed ensemble SSL algorithm is presented which is based on the hybridization of ensemble learning with semi-supervised learning. Generally, the development of an ensemble of classifiers consists of two main steps: *selection* and *combination*.

The selection of the appropriate component classifiers which constitute the ensemble is considered essential for its efficiency and the key points for its efficacy is based on the diversity and the accuracy the component classifiers. A commonly and widely utilized approach is to apply diverse classification algorithms (with heterogeneous model representations) to a single dataset [24]. Moreover, the combination of the individual predictions of the classification algorithms takes place through several methodologies and techniques with different philosophy and performance [28, 9].

By taking these into consideration, the development of an ensemble of classifiers is considered to be constituted by the SSL algorithms: Self-training, Co-training and Tri-training. These algorithms are self-labeled algorithms which exploit the hidden information in unlabeled data with complete different methodologies since Self-training and Tri-training are single-view methods while Co-training is a multi-view method.

A high-level description of the proposed Ensemble Semi-supervised Self-labeled Learning (EnSSL) algorithm is presented in Algorithm 4 which consists of two phases: *Training* phase and *Testing* phase.

In the Training phase, the SSL algorithms which constitute the ensemble are trained independently, using the same labeled L and unlabeled U datasets (steps 1-3). Clearly, the total computation time of this phase is the sum of computation times associated with each component SSL algorithm. In the Testing phase, initially the trained SSL algorithms are applied on each instance in the testing set (step 6). Subsequently, the individual predictions of the three SSL algorithms are combined via a maximum probability-based voting scheme. More specifically, the SSL algorithm which exhibits the most confident prediction over an unlabeled example of the test set is selected (step 8). In case the confidence of the prediction of the selected classifier meets a predefined threshold (*ThresLev*) then the classifier labels the example otherwise the prediction is not considered reliable enough (step 9). In this case, the output of the ensemble is defined as the combined predictions of three SSL learning algorithms via a simple majority voting, namely the ensemble output is the one made by more than half of them (step 11). This strategy has the advantage of exploiting the diversity of the errors of the learned models by using different classifiers and it does not require training on large quantities of representative recognition results from the individual learning algorithms.

Algorithm 4: **EnSSL**

Input: L − Set of labeled training instances.
U − Set of unlabeled training instances.
T − Set of test instances.
ThresLev − Threshold level.

Output: The labels of instances in the testing set.

/* Phase I: Training phase */
1: Train Self-train(L, U).
2: Train Co-train(L, U).
3: Train Tri-train(L, U).

/* Phase II: Testing phase */
5: **for** each x from T **do**
6: Apply Self-train, Co-train, Tri-train classifiers on x.
7: Find the classifier C^* with the highest confidence prediction on x.
8: **if** (Confidence of $C^* \geq$ *ThresLev*) **then**
9: C^* predicts the label y of x.
10: **else**
11: Use majority vote to predict the label y of x.
12: **end if**
13: **end for**

5 Experimental results

In this section, the classification performance of the proposed algorithm is compared with that of Self-training, Co-training and Tri-training on 40 benchmark datasets from KEEL repository [2] in terms of classification accuracy.

Each self-labeled algorithm was evaluated deploying as base learners:

- C4.5 decision tree algorithm [26].
- RIPPER (JRip) [5] as the representative of the classification rules.
- kNN algorithm [1] as instance-based learner.

These algorithms probably constitute three of the most effective and most popular data mining algorithms for classification problems [39]. In order to study the influence of the amount of labeled data, four different ratios of the training data were used: 10%, 20%, 30% and 40%. Moreover, we compared the classification performance of the proposed algorithm for each utilized base learner against the corresponding supervised learner.

The implementation code was written in JAVA, using WEKA Machine Learning Toolkit [13]. The configuration parameters of all the SSL methods and base learners used in the experiments are presented in Tables 1 and 2, respectively. It is worth noticing that the base learners were utilized with their the default parameter settings included in the WEKA software in order to minimize the effect of any expert bias by not attempting to tune any of the algorithms to the specific datasets.

Table 3 presents a brief description of the datasets structure i.e. number of instances (#Instances), number of attributes (#Features) and number of output classes (#Classes). The datasets considered contain between 101 and 7400 instances, the number of attributes ranges from 3 to 90 and the number of classes varies between 2 and 15.

SSL Algorithm	Parameters
Self-training	Maximum number of iterations $= 40$. $c = 95\%$.
Co-training	Maximum number of iterations $= 40$. Initial unlabeled pool $= 75$.
Tri-training	No parameters specified.
EnSSL	*ThresLev* $= 95\%$.

Table 1: Parameter specification for all SSL algorithms employed in the experimentation.

Base learner	Parameters
C4.5	Confidence factor used for pruning $= 0.25$. Minimum number of instances per leaf $= 2$. Number of folds used for reduced-error pruning $=$ 3. Pruning is performed after tree building.
JRip	Number of optimization runs $= 2$. Number of folds used for reduced-error pruning $=$ 3. Minimum total weight of the instances in a rule $=$ 2.0. Pruning is performed after tree building.
kNN	Number of neighbors $= 3$. Euclidean distance.

Table 2: Parameter specification for all base learners employed in the experimentation.

Dataset	#Instances	#Features	#Classes
automobile	159	15	2
appendicitis	106	7	2
australian	690	14	2
automobile	205	26	7
breast	286	9	2
bupa	345	6	2
chess	3196	36	2
contraceptive	1473	9	3
dermatology	358	34	6
ecoli	336	7	8
flare	1066	9	2
glass	214	9	7
haberman	306	3	2
heart	270	13	2
housevotes	435	16	2
iris	150	4	3
led7digit	500	7	10
lymph	148	18	4
mammographic	961	5	2
movement	360	90	15
page-blocks	5472	10	5
phoneme	5404	5	2
pima	768	8	2
ring	7400	20	2
satimage	6435	36	7
segment	2310	19	7

(continued).

Dataset	#Instances	#Features	#Classes
sonar	208	60	2
spambase	4597	57	2
spectheart	267	44	2
texture	5500	40	11
thyroid	7200	21	3
tic-tac-toe	958	9	2
titanic	2201	3	2
twonorm	7400	20	2
vehicle	846	18	4
vowel	990	13	11
wisconsin	683	9	2
wine	178	13	3
yeast	1484	8	10
zoo	101	17	7

Table 3: Brief description of datasets.

Tables 4-7 present the experimental results using 10%, 20%, 30% and 40% labeled ratio, respectively regarding all base learners.

Table 8 presents the number of wins of each one of the tested algorithms according to the supervised classifier used as base learner and utilized the ratio of labeled data in the training, while the best scores are highlighted in bold. It should be mentioned that draw cases between algorithms have not been encountered. Clearly, the presented results illustrated that EnSSL is the most effective method in all cases except the one using kNN as base learner with a labeled ratio of 30%. In this case, Tri-training performs better in 13 datasets, followed by EnSSL (9 wins). It is worth noticing that

– Depending upon the the ratio of labeled instances in the training set, EnSSL illustrates the highest classification accuracy in 46.2% of the datasets for 10% labeled ratio, 40% of the datasets for labeled ratio 20%, 44.4% of the datasets for labeled ratio 30% and 44.4% of the datasets for 40% labeled ratio. Obviously, EnSSL exhibits better classification accuracy for 10% and 40% labeled ratio.

– Regarding the base classifier, EnSSL (C4.5) presents the best classification accuracy in 14, 20, 21 and 19 of the datasets using a labeled ratio of 10%, 20%, 30% and 40%, respectively. EnSSL (JRip) prevails in 18, 14, 16 and 16 of the datasets using a labeled ratio of 10%, 20%, 30% and 40%, respectively. EnSSL (kNN) exhibit the best performance in 11, 9, and 17 of the datasets using a labeled ratio of 10%, 20%, 30% and 40%, respectively. Hence, EnSSL performs better using C4.5 and JRip as base learners.

Dataset	C4.5	Self (C4.5)	Co (C4.5)	Tri (C4.5)	EnSSL (C4.5)	JRip	Self (JRip)	Co (JRip)	Tri (JRip)	EnSSL (JRip)	kNN	Self (kNN)	Co (kNN)	Tri (kNN)	EnSSL (kNN)
automobile	64,21%	71,63%	71,58%	66,46%	69,79%	64,88%	69,08%	70,33%	64,63%	65,33%	61,75%	72,29%	64,13%	69,00%	74,13%
appendicitis	76,27%	81,09%	83,00%	82,00%	82,00%	83,91%	82,09%	81,00%	83,09%	83,09%	82,00%	85,82%	85,82%	85,82%	85,82%
australian	84,20%	85,80%	85,65%	87,10%	86,67%	85,22%	85,65%	85,36%	86,23%	86,38%	83,19%	83,91%	85,36%	83,77%	84,93%
banana	74,40%	74,58%	74,85%	75,00%	74,85%	73,19%	72,89%	73,15%	73,25%	73,30%	72,38%	72,89%	73,15%	73,25%	73,30%
breast	70,22%	75,87%	75,54%	73,82%	75,54%	68,45%	69,91%	67,81%	73,12%	69,56%	73,03%	72,41%	73,09%	73,45%	73,45%
bupa	56,24%	57,98%	57,96%	57,96%	58,57%	56,24%	58,57%	57,96%	57,96%	57,96%	56,24%	58,57%	57,96%	57,96%	57,96%
chess	98,97%	99,41%	97,62%	99,44%	99,41%	97,97%	99,09%	97,68%	99,09%	99,19%	93,90%	96,34%	90,02%	96,56%	96,40%
contraceptive	48,75%	49,69%	50,98%	50,37%	50,30%	43,04%	43,65%	46,64%	46,57%	46,77%	48,95%	50,84%	51,12%	51,59%	51,12%
dermatology	92,60%	94,54%	90,17%	94,54%	95,36%	85,76%	87,15%	86,06%	89,61%	91,00%	94,79%	97,25%	94,53%	97,24%	96,97%
ecoli	79,77%	80,37%	74,99%	80,97%	79,78%	78,83%	77,99%	75,88%	79,48%	78,88%	80,93%	80,97%	77,37%	82,15%	82,15%
flare	72,23%	74,66%	71,76%	73,73%	74,10%	68,38%	71,20%	67,18%	70,44%	70,36%	72,04%	74,95%	63,32%	73,92%	74,20%
glass	63,51%	67,81%	62,73%	64,48%	67,32%	61,21%	68,25%	62,64%	55,30%	64,09%	64,03%	72,51%	71,56%	72,97%	73,44%
haberman	71,90%	72,24%	70,24%	70,24%	70,24%	70,91%	71,57%	70,26%	70,56%	70,90%	71,55%	70,89%	73,88%	74,20%	74,20%
heart	78,54%	78,57%	76,89%	80,53%	81,52%	78,92%	80,89%	80,23%	80,90%	81,23%	80,87%	79,88%	80,86%	81,19%	80,20%
housevotes	96,52%	96,56%	94,84%	93,51%	95,69%	96,96%	96,56%	96,58%	93,51%	95,69%	91,34%	91,85%	91,85%	91,85%	91,85%
iris	92,67%	94,00%	95,33%	94,67%	94,00%	92,00%	93,33%	91,33%	90,00%	94,00%	92,67%	93,33%	93,33%	95,33%	94,67%
led7digit	69,80%	71,80%	58,60%	53,20%	69,40%	68,00%	70,60%	69,00%	34,20%	69,80%	72,60%	73,00%	56,00%	53,00%	69,40%
lymph	70,95%	74,38%	73,76%	73,71%	73,71%	72,90%	74,29%	75,05%	72,29%	74,38%	76,95%	78,48%	80,57%	81,19%	80,48%
mammographic	82,41%	83,49%	83,01%	84,22%	84,34%	82,41%	83,25%	82,29%	83,86%	83,73%	82,05%	82,65%	82,29%	83,73%	83,25%
movement	40,28%	56,94%	50,00%	35,83%	52,78%	29,44%	56,94%	49,17%	31,94%	48,89%	40,28%	65,00%	56,94%	59,72%	65,56%
page-blocks	95,39%	96,58%	95,71%	96,49%	96,71%	95,96%	96,09%	95,65%	96,36%	96,47%	96,05%	96,27%	95,34%	96,27%	96,16%
phoneme	80,33%	81,79%	80,13%	81,24%	81,98%	79,40%	81,35%	80,16%	80,46%	81,46%	80,26%	82,27%	81,25%	81,87%	82,14%
pima	74,47%	73,81%	73,81%	74,46%	74,20%	74,47%	73,29%	72,90%	73,81%	73,16%	72,69%	72,38%	73,03%	73,15%	73,54%
ring	80,41%	80,82%	80,91%	81,20%	83,54%	91,84%	92,47%	92,62%	92,61%	93,08%	62,15%	61,66%	60,51%	62,19%	61,05%
satimage	83,20%	84,38%	83,98%	84,65%	85,39%	83,31%	83,62%	84,15%	83,43%	84,80%	88,48%	89,25%	88,47%	89,03%	89,46%
segment	92,55%	94,42%	90,30%	93,90%	94,89%	91,82%	90,87%	86,15%	90,09%	92,77%	93,33%	93,12%	90,52%	93,29%	93,77%
sonar	67,43%	73,57%	68,67%	71,19%	71,19%	68,86%	77,05%	72,69%	74,71%	76,12%	70,69%	78,95%	74,10%	73,67%	76,05%
spambase	91,55%	92,72%	91,13%	92,79%	92,89%	90,68%	92,37%	91,55%	91,89%	92,83%	92,39%	93,02%	92,33%	93,22%	93,31%
spectheart	67,50%	68,75%	70,00%	70,00%	70,00%	63,75%	72,50%	70,00%	71,25%	71,25%	63,75%	66,25%	68,75%	68,75%	68,75%
texture	84,55%	87,87%	86,02%	86,65%	88,95%	84,73%	86,91%	86,33%	86,20%	89,64%	94,75%	96,07%	95,13%	95,78%	96,22%
thyroid	99,17%	99,32%	98,72%	99,24%	99,28%	98,89%	99,17%	98,42%	99,17%	99,24%	98,43%	98,76%	98,53%	98,69%	98,87%
tic-tac-toe	81,73%	83,60%	85,70%	85,27%	85,38%	97,08%	97,49%	97,91%	97,60%	97,49%	97,29%	99,06%	98,75%	98,64%	98,96%
titanic	77,15%	76,83%	77,60%	77,65%	77,82%	77,06%	77,19%	76,92%	77,65%	77,69%	77,06%	76,83%	77,69%	77,60%	77,65%
twonorm	78,99%	79,54%	79,50%	79,51%	82,19%	83,99%	84,82%	84,39%	84,19%	86,61%	93,39%	93,59%	93,69%	93,70%	94,61%
vehicle	66,55%	70,33%	66,78%	68,66%	70,44%	62,17%	60,87%	60,04%	61,34%	60,99%	64,90%	70,69%	67,97%	69,38%	70,33%
vowel	97,27%	98,28%	97,57%	98,28%	98,28%	96,96%	98,18%	97,17%	98,28%	98,28%	95,85%	97,57%	95,85%	97,47%	97,57%
wisconsin	94,57%	94,56%	93,57%	94,13%	94,56%	93,99%	95,85%	93,84%	94,98%	95,12%	96,42%	96,70%	96,28%	96,70%	96,70%
wine	84,28%	89,90%	78,01%	88,79%	89,90%	86,44%	89,28%	86,41%	89,87%	90,98%	93,20%	95,52%	94,97%	95,52%	95,52%
yeast	75,13%	74,93%	74,86%	74,86%	74,86%	75,07%	74,19%	75,74%	75,13%	75,20%	75,21%	74,19%	75,07%	75,27%	75,14%
zoo	93,09%	92,09%	89,18%	92,09%	92,09%	84,09%	86,09%	87,09%	86,09%	86,09%	90,09%	95,09%	84,27%	95,09%	95,09%

Table 4: Classification accuracy (labeled ratio 10%).

Dataset	C4.5	Self (C4.5)	Co (C4.5)	Tri (C4.5)	EnSSL (C4.5)	JRip	Self (JRip)	Co (JRip)	Tri (JRip)	EnSSL (JRip)	kNN	Self (kNN)	Co (kNN)	Tri (kNN)	EnSSL (kNN)
automobile	66,08%	77,29%	62,75%	73,50%	76,00%	65,42%	69,67%	64,67%	71,50%	74,04%	64,17%	68,46%	65,92%	72,25%	74,08%
appendicitis	80,09%	81,09%	83,00%	82,91%	82,91%	83,91%	82,09%	82,00%	82,91%	82,00%	83,09%	86,82%	86,73%	85,82%	85,82%
australian	86,09%	86,67%	86,23%	87,10%	87,68%	85,51%	86,09%	85,80%	86,23%	86,09%	84,93%	85,94%	83,04%	84,06%	85,07%
banana	74,62%	74,57%	75,23%	75,08%	78,26%	73,36%	72,75%	74,21%	73,79%	75,13%	74,55%	72,75%	74,21%	73,79%	75,13%
breast	70,23%	74,16%	71,31%	75,54%	75,64%	69,24%	72,07%	68,51%	71,70%	71,01%	73,12%	70,68%	71,69%	72,75%	72,75%
bupa	57,41%	58,27%	57,96%	57,96%	58,57%	57,10%	58,27%	57,96%	57,96%	57,96%	57,10%	57,41%	57,96%	57,96%	57,96%
chess	99,00%	99,41%	98,18%	99,37%	99,41%	98,87%	99,09%	98,15%	99,03%	99,06%	94,90%	95,99%	91,02%	96,71%	96,40%
contraceptive	50,44%	50,17%	50,84%	50,44%	50,71%	43,04%	42,57%	46,64%	46,36%	45,75%	50,51%	50,37%	51,93%	49,83%	50,71%
dermatology	93,41%	92,63%	89,32%	93,99%	94,81%	85,77%	88,52%	85,49%	89,05%	91,52%	94,79%	96,97%	95,32%	96,97%	97,24%
ecoli	80,02%	79,48%	76,79%	79,19%	80,06%	80,62%	78,89%	77,66%	78,01%	78,58%	80,94%	79,20%	80,07%	81,29%	81,58%
flare	73,17%	75,42%	72,70%	73,35%	74,29%	68,95%	73,17%	72,70%	71,85%	73,73%	72,51%	74,29%	68,48%	73,36%	73,45%
glass	65,52%	67,34%	63,70%	64,96%	70,24%	63,12%	64,94%	65,02%	62,21%	66,47%	67,81%	66,84%	71,58%	69,13%	72,97%
haberman	72,24%	70,24%	70,24%	70,24%	70,24%	71,27%	70,24%	70,27%	69,91%	70,24%	71,87%	70,59%	73,56%	73,56%	73,24%
heart	79,25%	77,89%	77,60%	79,22%	80,20%	80,88%	78,58%	76,89%	79,56%	79,57%	80,92%	81,53%	82,86%	80,86%	81,52%
housevotes	96,52%	96,56%	95,69%	93,51%	95,69%	96,96%	96,99%	96,99%	93,08%	94,38%	91,79%	91,85%	91,85%	91,85%	91,85%
iris	94,00%	94,00%	93,33%	93,33%	93,33%	93,33%	93,33%	91,33%	93,33%	93,33%	93,33%	93,33%	94,00%	93,33%	94,67%
led7digit	70,40%	71,00%	65,60%	68,00%	70,20%	69,60%	70,00%	70,80%	58,80%	70,40%	73,00%	73,80%	67,00%	69,40%	71,20%
lymph	71,57%	75,71%	72,43%	74,43%	76,43%	74,48%	72,43%	76,38%	73,76%	75,10%	79,19%	79,81%	83,24%	81,19%	81,14%
mammographic	83,61%	82,65%	82,65%	84,10%	83,37%	83,25%	83,37%	82,89%	83,73%	83,61%	83,01%	83,49%	82,29%	83,98%	83,25%
movement	50,00%	59,17%	47,50%	47,22%	57,50%	43,33%	54,17%	51,94%	21,39%	45,83%	57,22%	63,06%	55,83%	61,11%	65,00%
page-blocks	96,36%	96,75%	96,02%	96,58%	96,78%	96,22%	96,49%	95,74%	96,55%	96,71%	96,13%	96,40%	95,69%	96,18%	96,16%
phoneme	80,51%	81,33%	80,00%	81,20%	81,79%	79,94%	81,12%	80,11%	81,05%	81,55%	81,25%	82,12%	81,49%	81,81%	82,35%
pima	74,48%	74,33%	73,15%	73,29%	73,81%	74,62%	74,73%	73,41%	73,28%	73,67%	73,47%	74,07%	73,54%	73,68%	73,67%
ring	81,00%	80,69%	81,12%	80,91%	83,76%	92,28%	92,62%	92,16%	93,01%	93,14%	62,20%	61,36%	60,58%	62,38%	61,04%
satimage	83,29%	84,57%	84,27%	84,15%	84,90%	83,40%	83,23%	83,00%	83,73%	84,55%	88,90%	89,28%	88,50%	89,42%	89,65%
segment	93,46%	94,37%	91,17%	94,03%	94,59%	92,16%	91,21%	88,96%	90,48%	92,47%	92,34%	92,90%	91,21%	93,64%	93,55%
sonar	70,76%	71,24%	73,12%	73,62%	76,07%	70,71%	69,81%	75,07%	70,26%	69,83%	74,50%	75,98%	74,64%	78,86%	79,88%
spambase	92,28%	92,89%	91,87%	92,81%	92,85%	90,94%	92,55%	91,78%	92,52%	92,89%	92,85%	93,18%	92,81%	93,39%	93,70%
spectheart	71,25%	68,75%	71,25%	70,00%	68,75%	65,00%	71,25%	70,00%	71,25%	71,25%	66,25%	66,25%	66,25%	67,50%	68,75%
texture	86,36%	87,29%	86,29%	87,42%	88,76%	85,33%	86,53%	86,13%	86,51%	89,31%	94,49%	96,27%	95,58%	96,05%	96,56%
thyroid	99,21%	99,32%	98,96%	99,25%	99,31%	99,01%	99,17%	98,54%	99,13%	99,19%	98,58%	98,65%	98,96%	98,58%	98,79%
tic-tac-toe	82,36%	86,11%	85,28%	84,96%	87,47%	97,39%	97,70%	98,02%	98,01%	97,91%	98,12%	98,12%	97,07%	98,64%	98,33%
titanic	77,19%	77,06%	77,19%	77,65%	77,24%	77,15%	77,46%	75,69%	77,65%	77,65%	77,15%	76,92%	77,06%	77,33%	76,96%
twonorm	79,74%	79,58%	79,39%	79,64%	82,70%	84,11%	83,72%	84,16%	84,07%	86,62%	93,50%	93,73%	93,61%	93,73%	94,69%
vehicle	68,56%	71,26%	66,78%	70,09%	71,62%	62,54%	60,17%	59,92%	61,11%	60,63%	65,37%	67,50%	67,73%	70,21%	69,97%
vowel	97,87%	98,08%	98,48%	98,38%	98,58%	97,77%	98,18%	98,08%	98,18%	98,18%	96,76%	96,86%	96,66%	97,17%	97,47%
wisconsin	94,70%	94,28%	94,57%	94,13%	94,42%	94,42%	95,71%	95,56%	95,99%	95,70%	96,42%	96,85%	96,56%	96,85%	96,70%
wine	88,82%	89,90%	87,61%	85,42%	87,68%	89,90%	88,76%	84,15%	89,93%	89,90%	93,24%	95,52%	94,41%	95,52%	95,52%
yeast	75,34%	76,07%	74,39%	75,00%	74,73%	75,20%	75,80%	75,14%	74,80%	75,20%	75,47%	74,86%	75,34%	75,41%	75,20%
zoo	94,00%	92,09%	82,18%	89,09%	91,09%	86,09%	84,18%	89,00%	86,09%	86,09%	92,09%	95,09%	81,27%	94,18%	94,18%

Table 5: Classification accuracy (labeled ratio 20%).

Dataset	C4.5	Self (C4.5)	Co (C4.5)	Tri (C4.5)	EnSSL (C4.5)	JRip	Self (JRip)	Co (JRip)	Tri (JRip)	EnSSL (JRip)	kNN	Self (kNN)	Co (kNN)	Tri (kNN)	EnSSL (kNN)
automobile	74,21%	73,46%	72,92%	77,29%	79,21%	67,92%	63,42%	70,38%	71,54%	72,83%	65,50%	61,63%	69,17%	70,96%	70,33%
appendicitis	82,00%	83,09%	83,00%	84,82%	84,00%	83,91%	83,91%	84,82%	83,82%	83,82%	85,73%	86,73%	86,73%	84,91%	86,73%
australian	85,94%	86,52%	85,80%	86,81%	86,67%	85,65%	85,94%	85,65%	85,80%	85,51%	84,20%	83,91%	85,07%	84,06%	85,64%
banana	74,70%	74,58%	75,36%	74,70%	78,81%	73,45%	72,89%	73,70%	73,11%	76,11%	74,66%	72,89%	73,70%	73,11%	76,11%
breast	70,32%	75,20%	74,16%	75,54%	75,74%	69,54%	75,17%	69,95%	71,32%	72,03%	73,23%	73,09%	71,69%	73,09%	72,75%
bupa	57,10%	57,98%	57,96%	57,96%	58,57%	57,41%	57,98%	55,67%	57,96%	57,96%	57,41%	55,92%	57,96%	57,96%	57,96%
chess	99,12%	99,41%	98,28%	99,41%	99,44%	98,90%	99,00%	98,12%	99,22%	99,31%	94,96%	94,15%	92,49%	96,71%	95,93%
contraceptive	50,85%	49,82%	50,91%	50,17%	51,72%	46,50%	44,60%	47,39%	46,98%	46,43%	51,39%	49,21%	51,66%	52,20%	51,11%
dermatology	94,80%	93,15%	90,97%	94,53%	95,08%	87,67%	88,81%	86,35%	87,40%	89,08%	95,88%	96,43%	96,15%	97,24%	96,97%
ecoli	80,06%	79,15%	77,07%	78,87%	78,57%	80,66%	79,51%	79,79%	76,53%	77,12%	81,24%	79,80%	80,37%	80,70%	80,70%
flare	73,63%	74,48%	74,20%	73,45%	73,73%	69,13%	71,00%	70,64%	70,55%	71,95%	72,61%	73,35%	71,57%	74,11%	73,73%
glass	66,47%	61,19%	65,95%	69,74%	70,15%	63,16%	63,66%	65,06%	67,40%	68,83%	69,70%	63,68%	60,80%	71,99%	70,65%
haberman	72,24%	71,86%	70,24%	70,24%	70,24%	71,91%	71,86%	70,90%	70,24%	70,24%	72,89%	70,91%	72,57%	73,54%	72,56%
heart	79,90%	76,27%	79,87%	78,88%	80,22%	81,23%	79,59%	79,22%	82,22%	81,87%	82,22%	80,19%	83,84%	81,52%	81,84%
housevotes	96,52%	96,56%	96,99%	96,56%	96,56%	96,96%	96,99%	96,56%	96,99%	96,99%	92,21%	91,85%	91,85%	92,26%	91,85%
iris	94,00%	94,00%	94,00%	93,33%	94,00%	93,33%	93,33%	92,00%	94,00%	93,33%	93,33%	94,00%	94,00%	92,00%	93,33%
led7digit	71,20%	70,40%	69,20%	71,00%	71,00%	70,40%	69,20%	71,60%	69,00%	71,00%	73,20%	73,60%	70,80%	70,80%	71,80%
lymph	76,33%	73,62%	76,43%	72,38%	71,71%	74,90%	75,76%	79,76%	75,86%	77,14%	79,81%	79,14%	77,86%	81,19%	80,52%
mammographic	83,73%	83,98%	82,05%	84,22%	84,10%	83,61%	84,10%	82,29%	84,10%	84,22%	83,37%	83,86%	82,53%	83,73%	83,96%
movement	55,28%	58,89%	51,67%	50,56%	61,39%	51,39%	54,44%	50,00%	38,33%	53,06%	59,11%	63,06%	54,44%	58,06%	63,61%
page-blocks	96,38%	96,47%	96,38%	96,69%	96,87%	96,29%	96,36%	96,11%	96,38%	96,60%	96,20%	96,20%	95,92%	96,33%	96,34%
phoneme	81,05%	81,01%	80,11%	81,31%	81,42%	80,61%	80,55%	80,64%	80,88%	81,44%	81,68%	81,98%	81,35%	82,20%	82,14%
pima	75,53%	74,84%	73,68%	74,72%	75,24%	75,25%	73,80%	72,65%	72,37%	73,02%	74,48%	74,51%	74,20%	72,76%	74,71%
ring	81,23%	80,30%	81,43%	81,03%	83,15%	92,59%	92,88%	91,80%	92,59%	92,88%	62,36%	61,15%	60,65%	62,26%	60,80%
satimage	84,29%	84,48%	84,41%	84,69%	85,18%	83,43%	83,39%	83,36%	83,56%	84,91%	88,90%	89,08%	88,98%	89,45%	89,76%
segment	93,68%	94,03%	91,73%	94,37%	94,76%	92,64%	91,13%	87,88%	90,30%	92,77%	92,55%	92,51%	90,82%	93,55%	93,55%
sonar	72,62%	71,69%	74,57%	76,10%	74,17%	74,55%	74,14%	71,69%	74,10%	76,50%	74,52%	77,50%	76,43%	72,21%	74,10%
spambase	92,70%	92,70%	92,13%	92,92%	92,87%	92,15%	91,78%	91,83%	92,31%	92,44%	92,98%	92,55%	92,94%	93,37%	93,26%
spectheart	71,25%	71,25%	68,75%	67,50%	68,75%	68,75%	70,00%	71,25%	71,25%	71,25%	70,00%	71,25%	68,75%	67,50%	68,75%
texture	86,44%	87,80%	86,73%	86,76%	88,85%	86,25%	86,44%	87,45%	86,56%	88,95%	95,64%	95,89%	95,85%	96,16%	96,40%
thyroid	99,25%	99,17%	99,22%	99,32%	99,28%	99,07%	99,04%	99,17%	99,00%	99,13%	98,61%	98,33%	98,68%	98,63%	98,64%
tic-tac-toe	83,30%	84,96%	85,80%	85,38%	88,41%	97,81%	97,70%	97,60%	98,02%	97,70%	98,54%	96,45%	97,07%	98,85%	98,85%
titanic	77,15%	77,28%	77,46%	77,10%	77,24%	77,24%	77,24%	77,46%	77,51%	77,24%	77,17%	77,19%	77,46%	77,19%	77,06%
twonorm	79,85%	79,53%	79,68%	81,18%	83,59%	84,82%	83,93%	84,73%	84,91%	87,38%	93,72%	93,88%	93,73%	93,93%	94,89%
vehicle	68,68%	70,45%	69,15%	69,74%	71,75%	62,77%	58,52%	60,64%	60,76%	60,76%	67,73%	66,20%	67,86%	70,21%	69,04%
vowel	97,87%	97,47%	97,67%	97,98%	97,87%	97,77%	97,57%	97,98%	98,38%	98,28%	97,07%	96,86%	96,05%	97,97%	97,77%
wisconsin	94,99%	94,99%	94,13%	94,42%	94,85%	95,28%	96,42%	94,41%	94,99%	94,98%	96,57%	96,70%	96,56%	96,70%	96,70%
wine	89,35%	88,79%	87,61%	91,57%	91,57%	91,57%	87,58%	88,73%	88,79%	88,17%	94,35%	96,08%	96,63%	95,52%	96,08%
yeast	75,41%	74,73%	75,20%	75,20%	75,54%	76,08%	75,13%	75,00%	75,54%	76,21%	75,68%	74,53%	74,59%	75,20%	75,07%
zoo	94,00%	93,09%	88,09%	94,00%	95,00%	87,09%	87,09%	81,18%	86,09%	86,09%	93,01%	94,09%	88,27%	93,09%	93,09%

Table 6: Classification accuracy (labeled ratio 30%).

Dataset	C4.5	Self (C4.5)	Co (C4.5)	Tri (C4.5)	EnSSL (C4.5)	JRip	Self (JRip)	Co (JRip)	Tri (JRip)	EnSSL (JRip)	kNN	Self (kNN)	Co (kNN)	Tri (kNN)	EnSSL (kNN)
automobile	74,25%	72,33%	77,33%	75,46%	81,13%	70,88%	59,71%	68,46%	70,96%	71,58%	67,92%	65,33%	64,75%	67,21%	69,75%
appendicitis	83,82%	81,09%	85,73%	82,00%	82,00%	83,91%	81,09%	83,82%	83,00%	83,00%	85,81%	82,09%	85,82%	84,91%	85,82%
australian	86,23%	85,80%	86,09%	87,54%	87,10%	85,65%	85,36%	85,94%	86,38%	85,36%	85,38%	84,93%	84,06%	84,20%	86,78%
banana	74,79%	74,66%	75,77%	74,72%	80,53%	73,47%	72,74%	73,55%	72,81%	75,70%	74,94%	72,74%	73,55%	72,81%	75,70%
breast	70,95%	71,34%	75,20%	75,16%	75,16%	70,41%	70,68%	70,33%	71,70%	70,67%	73,04%	72,73%	72,38%	72,75%	73,08%
bupa	58,04%	54,75%	57,67%	57,96%	58,57%	57,44%	54,75%	57,67%	55,67%	57,96%	57,54%	55,34%	57,67%	57,96%	57,96%
chess	99,22%	99,25%	99,03%	99,41%	99,41%	99,00%	99,19%	98,62%	99,12%	99,16%	95,71%	93,55%	93,30%	96,65%	95,96%
contraceptive	51,41%	48,00%	51,73%	50,03%	51,52%	46,87%	42,84%	46,98%	47,05%	46,88%	51,96%	47,93%	51,11%	52,07%	51,93%
dermatology	95,08%	93,46%	92,05%	94,26%	95,38%	87,71%	87,98%	88,25%	89,08%	90,17%	96,14%	96,43%	95,59%	97,24%	97,24%
ecoli	81,84%	77,67%	80,63%	79,48%	80,34%	81,22%	79,49%	77,69%	80,37%	79,80%	82,04%	80,96%	79,46%	80,69%	82,47%
flare	73,82%	73,63%	73,07%	74,29%	74,10%	69,23%	68,86%	71,76%	69,79%	70,64%	73,27%	73,17%	72,32%	73,64%	73,36%
glass	70,65%	61,58%	67,38%	68,72%	72,01%	66,76%	55,13%	67,79%	61,77%	67,79%	73,42%	62,19%	70,17%	73,40%	74,78%
haberman	73,53%	73,53%	71,90%	70,24%	70,24%	72,20%	72,86%	70,94%	69,27%	69,27%	72,91%	72,22%	73,87%	74,20%	74,20%
heart	80,23%	74,94%	77,95%	77,90%	80,88%	81,55%	80,26%	82,47%	82,22%	83,52%	82,87%	81,53%	82,52%	80,86%	82,49%
housevotes	96,56%	94,82%	96,12%	96,56%	96,56%	96,96%	96,99%	96,56%	96,56%	96,56%	92,23%	91,85%	92,26%	91,85%	91,85%
iris	94,00%	94,00%	93,33%	93,33%	93,33%	94,00%	94,00%	86,67%	93,33%	93,33%	94,00%	94,00%	94,00%	92,67%	93,33%
led7digit	71,40%	68,60%	68,40%	70,40%	70,80%	70,80%	69,60%	68,80%	70,80%	71,00%	73,40%	74,00%	72,00%	71,80%	72,20%
lymph	76,33%	75,10%	74,29%	75,05%	75,05%	76,24%	76,43%	77,86%	75,76%	77,24%	80,52%	76,43%	79,81%	81,86%	81,86%
mammographic	83,73%	83,61%	82,29%	84,10%	84,10%	83,86%	83,61%	82,89%	84,22%	83,49%	83,37%	82,29%	82,29%	83,61%	83,13%
movement	55,83%	58,89%	51,11%	55,00%	59,17%	52,44%	50,28%	50,00%	49,17%	52,78%	61,39%	53,89%	58,89%	65,28%	62,78%
page-blocks	96,42%	96,56%	96,36%	96,77%	96,91%	96,34%	96,34%	96,29%	96,24%	96,34%	96,31%	96,27%	96,05%	96,31%	96,40%
phoneme	81,11%	80,51%	80,66%	81,20%	81,25%	80,90%	80,05%	80,48%	81,03%	81,18%	82,14%	81,61%	81,53%	82,11%	82,20%
pima	74,87%	73,54%	74,33%	73,16%	74,20%	76,05%	73,80%	73,81%	73,16%	74,33%	74,57%	74,19%	74,34%	73,02%	74,84%
ring	82,45%	80,91%	80,97%	81,16%	83,32%	92,69%	92,96%	91,64%	92,74%	93,19%	62,72%	60,47%	60,47%	62,32%	60,49%
satimage	84,38%	84,34%	84,55%	84,24%	85,10%	83,74%	84,48%	83,71%	83,73%	85,00%	88,92%	88,81%	89,20%	89,45%	89,73%
segment	94,20%	93,46%	92,03%	93,72%	94,20%	93,03%	90,35%	90,87%	90,26%	91,82%	92,99%	92,08%	92,12%	93,42%	93,07%
sonar	73,17%	71,74%	72,71%	72,69%	73,67%	76,00%	70,81%	72,71%	71,29%	76,26%	75,02%	77,00%	74,14%	75,57%	77,50%
spambase	92,81%	92,41%	92,11%	92,72%	92,76%	92,26%	91,87%	91,87%	92,05%	92,37%	93,02%	92,65%	93,22%	93,18%	93,41%
spectheart	72,50%	66,25%	71,25%	68,75%	68,75%	68,75%	72,50%	70,00%	70,00%	71,25%	70,00%	67,50%	70,00%	68,75%	68,75%
texture	87,05%	87,85%	87,05%	87,56%	88,89%	86,89%	86,42%	86,45%	87,24%	89,16%	95,91%	95,69%	95,84%	96,09%	96,31%
thyroid	99,25%	99,08%	99,25%	99,22%	99,25%	99,17%	99,07%	99,07%	99,17%	99,18%	98,69%	98,50%	98,54%	98,63%	98,78%
tic-tac-toe	83,51%	84,34%	85,90%	85,70%	88,93%	98,02%	97,49%	97,60%	97,70%	97,81%	98,64%	93,73%	97,29%	98,85%	98,43%
titanic	77,60%	77,46%	77,87%	77,51%	77,92%	77,60%	77,46%	77,96%	77,92%	77,92%	77,60%	77,65%	77,96%	77,19%	78,01%
twonorm	80,11%	80,04%	80,19%	80,22%	82,82%	84,89%	83,65%	84,18%	83,95%	86,07%	94,11%	94,03%	93,91%	93,84%	95,03%
vehicle	70,34%	69,25%	69,40%	68,45%	70,68%	64,88%	57,68%	60,88%	60,05%	60,88%	68,20%	67,60%	68,08%	70,09%	69,38%
vowel	98,08%	97,77%	97,98%	98,28%	98,18%	97,98%	98,28%	97,87%	98,18%	98,18%	97,57%	96,36%	97,67%	97,47%	97,37%
wisconsin	94,99%	94,28%	94,85%	94,99%	94,99%	95,99%	95,56%	94,70%	95,27%	95,27%	97,42%	96,42%	96,99%	96,70%	96,70%
wine	90,39%	88,79%	88,24%	88,79%	90,49%	91,57%	88,20%	85,39%	90,36%	88,73%	94,87%	94,97%	95,52%	95,52%	95,52%
yeast	75,35%	74,66%	75,20%	75,27%	75,60%	76,08%	73,91%	75,34%	74,93%	76,27%	76,08%	73,85%	75,40%	75,34%	75,40%
zoo	95,00%	90,09%	91,09%	93,00%	92,00%	87,09%	87,09%	85,09%	87,09%	87,09%	93,01%	90,18%	92,09%	92,09%	93,09%

Table 7: Classification accuracy (labeled ratio 40%).

	10%			20%			30%			40%		
SSL Algorithm	C4.5	JRip	kNN	C4.5	JRip	kNN	C4.5	JRip	kNN	C4.5	JRip	kNN
Self-Train	11	9	8	9	6	7	1	5	4	0	5	1
Co-Train	4	5	2	2	6	4	3	5	4	3	3	2
Tri-Train	4	3	8	2	4	7	7	5	**13**	3	4	8
Supervised	4	4	0	4	5	2	4	5	4	7	8	4
EnSSL	**14**	**18**	**11**	**20**	**14**	**15**	**21**	**16**	9	**19**	**16**	**17**

Table 8: Total wins of each SSL algorithm.

The statistical comparison of multiple algorithms over multiple data sets is fundamental in machine learning and usually it is typically carried out by means of a nonparametric statistical test. Therefore, the Friedman Aligned-Ranks (FAR) test [8] is utilized in order to conduct a complete performance comparison between all algorithms for all the different labeled ratios. Its application will allow us to highlight the existence of significant differences between the proposed algorithm and the classical SSL algorithms and evaluate the rejection of the hypothesis that all the classifiers perform equally well for a given level. Notice that FAR test is considered to be one of the most well-known tools for multiple statistical comparison tests when comparing more than two methods [10]. Furthermore, the Finner test is applied as a post hoc procedure to find out which algorithms present significant differences.

Ratio	Classifier (C4.5)	Friedman Ranking	Finner post-hoc test p-value	Null Hypothesis
10%	EnSSL	58.4375		
	Self-training	76.625	0.049750	rejected
	Tri-training	94.7875	0.037739	rejected
	Co-training	128.225	0.025321	rejected
	Supervised	144.425	0.012741	rejected
20%	EnSSL	56.6		
	Self-training	83.8	0.045583	rejected
	Tri-training	103.85	0.037739	rejected
	Supervised	115.4875	0.025321	rejected
	Co-training	142.7625	0.012741	rejected
30%	EnSSL	57.575		
	Tri-training	93.5375	0.044582	rejected
	Supervised	108.85	0.037739	rejected
	Self-training	109.2625	0.025321	rejected
	Co-training	133.275	0.012741	rejected

(continued).

Ratio	Classifier (C4.5)	Friedman Ranking	Finner post-hoc test p-value	Null Hypothesis
40%	EnSSL	58.475		
	Supervised	77.45	0.142611	accepted
	Tri-training	106.9625	0.000239	rejected
	Co-training	116.2	0.000016	rejected
	Self-training	143.4125	0.000000	rejected

Table 9: FAR test and Finner post hoc test (C4.5).

Tables 9, 10 and 11 present the information of the statistical analysis performed by nonparametric multiple comparison procedures for each base learner. The best(lowest) ranking obtained in each FAR test determines the control algorithm for the post hoc test. Moreover, the adjusted p-value with Finner's test (Finner APV) is presented based on the control algorithm, at $\alpha = 0.05$ level of significance. Clearly, the proposed algorithm exhibits the best overall performance, outperforming the rest SSL algorithms, since it reports the highest probability-based ranking, presenting statistically better results, relative to all labeled ratio.

6 Conclusions & future research

In this work, a new ensemble semi-supervised algorithm is proposed based on a voting methodology. The proposed algorithm combines the individual predictions of three SSL algorithms: Co-training, Self-training and Tri-training via a maximum-probability voting scheme. The numerical experiments and the presented statistical analysis indicate that the proposed algorithm EnSSL outperforms its component SSL algorithms, confirming its efficacy.

An interesting direction for future work is the development of a parallel implementation of the the proposed algorithm. Notice that the implementation of each one of its component based learners in parallel machines constitutes a significant aspect to be studied, since a huge amount of

Ratio	Classifier (JRip)	Friedman Ranking	Finner post-hoc test p-value	Null Hypothesis
10%	EnSSL	62.2625		
	Self-training	81.5375	0.136404	accepted
	Tri-training	100.2625	0.004429	rejected
	Co-training	121.0125	0.136404	rejected
	Supervised	137.425	0.000000	rejected
20%	EnSSL	69.25		
	Self-training	95.225	0.044749	rejected
	Tri-training	102.35	0.014031	rejected
	Supervised	116.7	0.000492	rejected
	Co-training	118.975	0.000488	rejected
30%	EnSSL	66.225		
	Supervised	99.9625	0.009140	rejected
	Tri-training	104.175	0.004484	rejected
	Self-training	109.25	0.001771	rejected
	Co-training	122.8875	0.000048	rejected
40%	EnSSL	64.925		
	Supervised	76.1	0.387887	accepted
	Tri-training	107.875	0.001206	rejected
	Co-training	121.175	0.000028	rejected
	Self-training	132.425	0.000001	rejected

Table 10: FAR test and Finner post hoc test (JRip).

Ratio	Classifier (kNN)	Friedman Ranking	Finner post-hoc test p-value	Null Hypothesis
10%	EnSSL	59.65		
	Tri-training	73.825	0.273404	accepted
	Self-training	89.3375	0.028959	rejected
	Co-training	129.8375	0.00000	rejected
	Supervised	149.85	0.00000	accepted
20%	EnSSL	59.5125		
	Tri-training	79.1625	0.128941	accepted
	Self-training	103.55	0.00089	rejected
	Co-training	130.075	0.00000	rejected
	Supervised	130.2	0.00000	accepted
30%	EnSSL	70.9625		
	Tri-training	86.9875	0.045642	rejected
	Supervised	101.175	0.026013	rejected
	Self-training	117.8625	0.000581	rejected
	Co-training	125.5125	0.0001	rejected
40%	EnSSL	61.9875		
	Supervised	74.3375	0.33996	accepted
	Tri-training	92.2625	0.02568	rejected
	Co-training	124.225	0.000003	rejected
	Self-training	149.6875	0.000000	rejected

Table 11: FAR test and Finner post hoc test (kNN).

data can be processed in significantly less computational time. Since the experimental results are quite encouraging, a next step could be the evaluation of the proposed algorithm in specific scientific fields applying real world datasets, such as the educational, health care, etc.

References

[1] David W. Aha. *Lazy Learning*. Dordrecht: Kluwer Academic Publishers, 1997. https://doi.org/10.1007/978-94-017-2053-3

[2] Jesús Alcalá-Fdez, Alberto Fernández, Julián Luengo, Joaquín Derrac, Salvador García, Luciano Sánchez, and Francisco Herrera. Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic & Soft Computing*, 17, 2011. https://doi.org/10.1109/nwesp.2011.6088224

[3] Ethem Alpaydin. *Introduction to Machine Learning*. MIT Press, Cambridge, 2nd edition, 2010. https://doi.org/10.1017/s0269888906220745

[4] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *11th annual conference on Computational learning theory*, pages 92–100. ACM, 1998. https://doi.org/10.1109/icdm.2001.989574

[5] William W. Cohen. Fast effective rule induction. In *International Conference on Machine Learning*, pages 115–123, 1995. https://doi.org/10.1016/b978-1-55860-377-6.50023-2

[6] Bozidara Cvetkovic, Boštjan Kaluza, Mitja Luštrek, and Matjaz Gams. Semi-supervised learning for adaptation of human activity recognition classifier to the user. In *Proceddings of International Joint Conference on Artificial Intelligence*, pages 24–29, 2011.

[7] Asif Ekbal and Sivaji Bandyopadhyay. Named entity recognition using appropriate unlabeled data, post-processing and voting. *Informatica*, 34(1), 2010.

[8] Helmut Finner. On a monotonicity problem in step-down multiple test procedures. *Journal of the American Statistical Association*, 88(423):920–923, 1993. https://doi.org/10.2307/2290782

[9] Matjaž Gams. *Weak intelligence: through the principle and paradox of multiple knowledge*. Nova Science, 2001.

[10] Salvador García, Alberto Fernández, Julián Luengo, and Francisco Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180(10):2044–2064, 2010. https://doi.org/10.1016/j.ins.2009.12.010

[11] Hristijan Gjoreski, Boštjan Kaluža, Matjaž Gams, Radoje Milić, and Mitja Luštrek. Context-based ensemble method for human energy expenditure estimation. *Applied Soft Computing*, 37:960–970, 2015. https://doi.org/10.1016/j.asoc.2015.05.001

[12] Tao Guo and Guiyang Li. Improved tri-training with unlabeled data. *Software Engineering and Knowledge Engineering: Theory and Practice*, pages 139–147, 2012. https://doi.org/10.1007/978-3-642-25349-2_19

[13] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: An update. *SIGKDD Explorations Newsletters*, 11:10–18, 2009. https://doi.org/10.1145/1656274.1656278

[14] Kyaw Kyaw Htike. Hidden-layer ensemble fusion of MLP neural networks for pedestrian detection. *Informatica*, 41(1), 2017.

[15] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. McGraw Hill, John Wiley & Sons, Inc., second edition, 2014. https://doi.org/10.1002/9781118914564

[16] Jurica Levatić, Sašo Džeroski, Fran Supek, and Tomislav Šmuc. Semi-supervised learning for quantitative structure-activity modeling. *Informatica*, 37(2), 2013.

[17] Ming Li and Zhi-Hua Zhou. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(6):1088–1098, 2007. https://doi.org/10.1109/tsmca.2007.904745

[18] Chang Liu and Pong C. Yuen. A boosted co-training algorithm for human action recognition. *IEEE transactions on circuits and systems for video technology*, 21(9):1203–1213, 2011. https://doi.org/10.1109/tcsvt.2011.2130270

[19] Ioannis E. Livieris, Ioannis Dimopoulos, Thedore Kotsilieris, and Panagiotis Pintelas. Predicting length of stay in hospitalized patients using ssl algorithms. In *ACM 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion*, pages 1–8, 2018. https://doi.org/10.1145/3218585.3218588

[20] Ioannis E. Livieris, Konstantina Drakopoulou, Vassilis Tampakas, Tassos Mikropoulos, and Panagiotis Pintelas. Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research*, 2018. https://doi.org/10.1177/0735633117752614

[21] Ioannis E. Livieris, Konstantina Drakopoulou, Vassilis Tampakas, Tassos Mikropoulos, and Panagiotis Pintelas. *Research on e-Learning and ICT in Education*, chapter An ensemble-based semi-supervised approach for predicting students' performance, page 25-42. Springer, 2018. https://doi.org/10.1007/978-3-319-95059-4_2

[22] Ioannis E. Livieris, Andreas Kanavos, Vassilis Tampakas, and Panagiotis Pintelas. An ensemble SSL algorithm for efficient chest x-ray image classification. *Journal of Imaging*, 4(7), 2018. https://doi.org/10.3390/jimaging4070095

[23] Ioannis E. Livieris, Tassos Mikropoulos, and Panagiotis Pintelas. A decision support system for predicting students' performance. *Themes in Science and Technology Education*, 9:43–57, 2016.

[24] Christopher J. Merz. Using correspondence analysis to combine classifiers. *Machine Learning*, 36:33–58, 1999. https://doi.org/10.1023/A:1007559205422

[25] Vincent Ng and Claire Cardie. Weakly supervised natural language learning without redundant views. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 94–101. Association for Computational Linguistics, 2003. https://doi.org/10.3115/1073445.1073468

[26] J. Ross Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann, San Francisco, 1993. https://doi.org/10.1007/BF00993309

[27] Matteo Re and Giorgio Valentini. *Advances in Machine Learning and Data Mining for Astronomy*, chapter Ensemble methods: A review, pages 563–594. Chapman & Hall, 2012. https://doi.org/10.1201/b11822-34

[28] Lior Rokach. *Pattern Classification Using Ensemble Methods*. World Scientific Publishing Company, 2010. https://doi.org/10.1142/7238

[29] Moumita Roy, Susmita Ghosh, Ashish Ghosh. A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system. *Information Sciences*, 269:35–47, 2014. https://doi.org/10.1016/j.ins.2014.01.037

[30] S.K. Satapathy, A.K. Jagadev, and S. Dehuri. An empirical analysis of different machine learning techniques for classification of EEG signal to detect epileptic seizure. *Informatica*, 41(1), 2017.

[31] Sandeep Kumar Satapathy, Alok Kumar Jagadev, and Satchidananda Dehuri. Weighted majority voting based ensemble of classifiers using different machine learning techniques for classification of EEG signal to detect epileptic seizure. *Informatica*, 41(1):99, 2017.

[32] Gasper Slapničar, Mitja Luštrek, and Matej Marinko. Continuous blood pressure estimation from PPG signal. *Informatica*, 42(1), 2018.

[33] Shiliang Sun and Feng Jin. Robust co-training. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(07):1113–1126, 2011. https://doi.org/10.1142/s0218001411008981

[34] Shiliang Sun and Qingjiu Zhang. Multiple-view multiple-learner semi-supervised learning. *Neural processing letters*, 34(3):229, 2011. https://doi.org/10.1007/s11063-011-9195-8

[35] Isaac Triguero, Salvador García, and Francisco Herrera. SEG-SSC: A framework based on synthetic examples generation for self-labeled semi-supervised classification. *IEEE Transactions on Cybernetics*, 45:622–634, 2014. https://doi.org/10.1109/tcyb.2014.2332003

[36] Isaac Triguero, Salvador García, and Francisco Herrera. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. *Knowledge and Information Systems*, 42(2):245–284, 2015. https://doi.org/10.1007/s10115-013-0706-y

[37] Isaac Triguero, José A. Sáez, Julián Luengo, Salvador García, and Francisco Herrera. On the characterization of noise filters for self-training semi-supervised in nearest neighbor classification. *Neurocomputing*, 132:30–41, 2014. https://doi.org/10.1016/j.neucom.2013.05.055

[38] Julius Venskus, Povilas Treigys, Jolita Bernatavičienė, Viktor Medvedev, Miroslav Voznak, Mindaugas Kurmis, and Violeta Bulbenkienė. Integration of a self-organizing map and a virtual pheromone for real-time abnormal movement detection in marine traffic. *Informatica*, 28(2):359–374, 2017.

[39] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, and Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008. https://doi.org/10.1201/9781420089653

[40] Qian Xu, Derek Hao Hu, Hong Xue, Weichuan Yu, and Qiang Yang. Semi-supervised protein subcellular localization. *BMC bioinformatics*, 10(1):S47, 2009. https://doi.org/10.1186/1471-2105-10-s1-s47

[41] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting of the association for computational linguistics*, pages 189–196, 1995. https://doi.org/10.3115/981658.981684

[42] Yan Zhou and Sally Goldman. Democratic co-learning. In *16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 594–602. IEEE, 2004. https://doi.org/10.1109/ictai.2004.48

[43] Zhi-Hua Zhou. When semi-supervised learning meets ensemble learning. In *Frontiers of Electrical and Electronic Engineering in China*, volume 6, pages 6–16. Springer, 2011. https://doi.org/10.1007/s11460-011-0126-2

[44] Zhi-Hua Zhou and Ming Li. Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on Knowledge and Data Engineering*, 17(11):1529–1541, 2005. https://doi.org/10.1109/tkde.2005.186

[45] Xiaojin Zhu. Semi-supervised learning. In *Encyclopedia of Machine Learning*, pages 892–897. Springer, 2011.

[46] Xiaojin Zhu and Andrew B. Goldberg Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130, 2009. https://doi.org/10.2200/s00196ed1v01y200906aim006