Noname manuscript No.
(will be inserted by the editor)

An improved self-labeled algorithm for cancer prediction

Ioannis Livieris · Emmanuel Pintelas · Andreas Kanavos · Panagiotis Pintelas

the date of receipt and acceptance should be inserted later

Abstract Nowadays, cancer constitutes the second leading cause of death globally. The application of an efficient classification model is considered essential in modern diagnostic medicine in order to assist experts and physicians to make more accurate and early predictions and reduce the rate of mortality. Machine learning techniques are being broadly utilized for the development of intelligent computational systems, exploiting the recent advances in digital technologies and the significant storage capabilities of electronic media. Ensemble learning algorithms and semi-supervised algorithms have been independently developed to build efficient and robust classification models from different perspectives. The former attempt to achieve strong generalization by using multiple learners while the latter attempt to achieve strong generalization by exploiting unlabeled data. In this work, we propose an improved semi-supervised self-labeled algorithm for the cancer prediction, based on ensemble methodologies. Our preliminary numerical experiments illustrate the efficacy and efficiency of the proposed algorithm, proving that reliable and robust prediction models could be developed by the adaptation of ensemble techniques in the semi-supervised learning framework.

Keywords Semi-supervised learning · ensemble learning · self-labeled algorithms · cancer prediction.

1 Introduction

Cancer is a horrifying disease causing a lot of human deaths every year. It has been documented that only cancer causes 13% of all deaths worldwide (DeSantis et al. 2014) and according to Global Health Observatory was responsible for 9.6 million deaths in 2018 (Sasieni & Parkin 2018). Some

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

E. Pintelas
Department of Electrical & Computer Engineering University of Patras, Greece, GR 265-00. E-mail: ece6835@upnet.gr

A. Kanavos
Department of Computer & Informatics Engineering (DISK Lab), Technological Educational Institute of Western Greece, Greece, GR 263-34. E-mail: kanavos@ceid.upatras.gr

P. Pintelas
Department of Mathematics, University of Patras, Greece, GR 265-00. E-mail: pintelas@upatras.gr

of the most common types of cancer are: breast cancer, cervical cancer, lung cancer, brain cancer, skin cancer and prostate cancer. According to National Cancer Institute, in all types of cancer the problem starts when the body's cells begin to divide without stopping and spread into surrounding tissues forming growths called tumors (Dollinger & Rosenbaum 2002). When these tumors grow, they can cause significant damage, squeezing and blocking small blood vessels in the area. In addition, when they have grown enough they can break off and travel to other places in the body through the blood, forming new tumors, which is known as metastasis. An early diagnosis of cancer can be crucially significant for patient's health and treatment.

Magnetic Resonance Imaging (MRI) scans, X-rays, computed tomography and ultrasound are common tools used to examine the body; thus they are widely used to assist experts diagnose possible cancer in the body. Unfortunately, a perfect diagnostic system has not yet been developed, because of the large number of interactive predictive variables involved while experts' decisions based on their personal knowledge memory and experience add uncertainty which may lead to errors (Morvan & Jenkins 2017). Therefore, the development of decision support tools for the accurate prediction of cancer is considered essential and has gained an intense interest (see Edwards & Elwyn (2009), Révész et al. (2017), Engelen et al. (2016) and the references therein).

Machine learning and data mining have become very popular in the diagnostic medicine area for the development of accurate prediction tools, assisting medical experts making decisions in order to treat patients efficiently (Fatima & Pasha 2017). They can extract useful and valuable information from large amounts of data, detecting key features and hidden patters in a variety of fields. This ability reveals their overall significance in general hence their adaptation in cancer diagnosis area is a growing trend and has gained a lot of popularity. In recent years, significant efforts have been devoted to the development of accurate prediction tools which distinguish cancer patients from healthy, revealing the future potential and interest of this trend. Abdel-Zaher & Eldeib (2016) used machine learning algorithms for predicting breast cancer based on a dataset of 683 patients. Naive Bayes and k-Nearest Neighbour classifiers were utilized for this task, resulting in achieving an accuracy of 97.51% of successful prediction. On another study, Esteva et al. (2017) trained a convolutional neural network for the prediction of skin cancer on a dataset of 129450 clinical images. The highest classification accuracy reported was 96%. Additionally, their model outperformed 21 board-certified dermatologists at skin cancer classification using photographic and dermoscopic images. Motivated by previous work, Vidić et al. (2018) utilized support vector machines for breast cancer classification using diffusion-weighted MRI histogram features on fifty-one patients achieving 96% prediction accuracy. On another study, Turgut et al. (2018) applied microarray data for the problem of breast cancer classification, utilizing several machine learning algorithms. From their numerical experiments, it was reported that Linear Regression classifier had the highest accuracy (99.23%). while with the application of two feature selection methods, support vector machines exhibited the highest classification performance. Recently, Hussain et al. (2018) used machine learning techniques by employing a combination of features extracting strategies on a dataset of 682 MRI from 20 patients, for prostate cancer classification achieving a prediction accuracy of 99.71%.

Nevertheless, the development of an accurate prediction model for cancer prediction is considered a difficult and challenging task because in general the medical field is constrained by the lack of available labeled images (Livieris et al. 2018b). Furthermore, in order to train efficiently a classifier and be able to make accurate predictions, it often needs a large amount of labeled data. However, the process of correctly labeling new unlabeled images is often boring, expensive and time consuming as it also requires the efforts of human experts. To deal with this problem, Semi-Supervised Learning (SSL) algorithms constitute the appropriate machine learning methodology to extract useful information from labeled and unlabeled data. In contrast to traditional, supervised approaches, these algorithms exploit the explicit classification information of labeled data with the knowledge hidden in the un-

labeled data. Hence, these methods have received considerable attention due to their potential for reducing the effort of labeling data while still preserving competitive performance.

During the last decades, the development of an ensemble of classifiers has been also proposed as a new direction for obtaining higher classification accuracy. The basic idea of Ensemble Learning is the combination of a set of diverse prediction models, 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 Rokach (2010). It is worth noticing that although SSL and EL methodologies have been efficiently applied to a variety of real-world problems, they were almost developed separately. Recently, Zhou (2011) demonstrated that semi-supervised learning algorithms and ensemble learning algorithms are indeed beneficial to each other, and more efficient and robust classification algorithms can be developed. Along this line, Livieris et al. (2018a,b) and Livieris (2019) proposed to incorporate ensemble learning techniques in the self-labeled framework in order to develop efficient and robust classifiers.

In this work, we proposed a new semi-supervised algorithm, called iCST-Voting, for cancer prediction. The proposed algorithm combines the individual predictions of three of the most efficient self-labeled algorithms and exploits the efficiency of an ensemble as a base learner. Our preliminary experimental results reveal the efficacy of the proposed algorithm compared against state-of-the-art self-labeled methods.

The remainder of this paper is organized as follows: Section 2 presents a brief description of the self-labeled algorithms. Section 3 presents the proposed ensemble semi-supervised classification algorithm and Section 4 presents the datasets utilized in our framework. Section 5 presents a series of experiments in order to examine and evaluate the accuracy of the proposed algorithm against the most popular self-labeled classification algorithms. Finally, Section 6 sketches our concluding remarks and future work directions.

2 On semi-supervised self-labeled classification algorithms

Suppose that X denotes the domain of instances, in which a single instance $x \in X$ is represented as $x = \{x_1, x_2, \ldots, x_d, y\}$ with x belonging to a class y and a d-dimensional space in which x_i is the value of the i-th feature. Then, let us assume that the training set $L \cup U$ consists of a labeled set L of N_L instances where y is known and of an unlabeled set U of N_U instances where y is unknown with $N_L \ll N_U$. 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 basic aim of the semi-supervised classification is to obtain an efficient learning hypothesis utilizing the instances in the training set.

Self-labeled methods are prominent SSL methods which address the shortage of labeled data via a self-learning process based on supervised prediction models. This class of algorithms is characterized by their simplicity of implementation and their wrapper-based philosophy. In general, self-labeled methods are divided into *Self-training* (Yarowsky 1995) and *Co-training* (Blum & Mitchell 1998). Self-training is considered as one of the most simple and efficient self-labeled algorithm. This algorithm 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. Then, it trains one classifier in each specific view and the classifiers teach each other the most confidently predicted examples. Most self-labeled algorithms proposed in the literature are based on these algorithms while most of them are based on ensemble techniques.

Democratic-Co learning (Zhou & Goldman 2004) consists a single view algorithm utilizing multiple learners for producing the necessary information. More specifically, starts by training all learners on the labeled dataset and then they are used for predicting a label for every example on the unlabeled dataset. If the majority of the learners agree on the label of an example and the sum of

the mean confidence of the majority is greater of the minority, then this example can be labeled for the minority of learners which disagree. This process is repeated, till there is no example in the unlabeled dataset. A rather representative approach which is based on the ensemble philosophy is the *Tri-training* algorithm (Zhou & Li 2005) which constitutes an improved single-view extension of the Co-training algorithm. This algorithm utilizes a labeled dataset to initially train three classifiers which are used to make predictions on an unlabeled dataset. If two classifiers agree on labeling an example, then this is labeled for the third classifier too. This strategy gives the advantage, comparing with the self and co-training algorithms, to avoid the explicitly measuring of the confidence of the labeling of the classifiers since such measuring is sometimes a quite complicated and time-consuming process. Along this line, Li & Zhou (2007) proposed *Co-Forest* algorithm, in which a number of Random trees are trained on bootstrap data from the dataset. 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 of a simple majority voting. *Co-Bagging* (Hady & Schwenker 2010) algorithm utilizes a number of base classifiers on a bootstrap sample created by random resampling with replacement from the original training set. Each sample has around 2/3 of the original training set. It is worth mentioning that this method works well for unstable learning algorithms, where a small change in the input data can lead to a big change in the output.

3 iCST algorithm

Recently, research focus on incorporating ensemble learning techniques in the self-labeled framework in order to build efficient and powerful classifiers.

Livieris et al. (2018b) proposed the *CST-Voting* algorithm which exploit the individual predictions of the self-labeled algorithms: Co-training, Self-training and Tri-training utilizing simple majority voting. Along this line, Livieris (2019) combined the predictions of the same self-labeled algorithms utilizing a maximum-probability voting scheme. Livieris et al. (2018a) utilized an ensemble of classifiers as base learners to increase the efficiency of semi-supervised self-labeled methods.

Motivated by the previous works, we consider to extend the previous works and improve the efficiency of CST-Voting by utilizing an ensemble of classifiers as a base learner in all its component self-labeled algorithms. A high-level description of the proposed algorithm, entitled iCST-Voting (improved CST-Voting), is presented in Algorithm 1. It consists of two phases: the *Training* phase and *Voting* phase.

Initially, in the Training phase, the self-labeled algorithms i.e. Co-training, Self-training and Tri-training, which constitute the ensemble are trained utilizing the same labeled L and unlabeled U datasets (Steps $1-3$). Notice that all self-labeled algorithms wrap around an ensemble of supervised classifiers. Subsequently, in the Voting phase, the final hypothesis on each unlabeled example x of the test set combines the individual predictions of self-labeled algorithms utilizing a majority voting methodology (Steps $4-7$).

Algorithm 1. iCST-Voting

Input: L −Set of labeled instances.
U − Set of unlabeled instances.
C − Ensemble of supervised base learners.

/* Phase I: Training */
1: Train Self-training(L,U) using C as base learner.
2: Train Co-training(L,U) using C as base learner.
3: Train Tri-training(L,U) using C as base learner.

/* Phase II: Voting */
4: for each $x \in T$ **do**
5: Apply Self-training, Co-training, Tri-training on x.
6: Use majority vote to predict the label y^* of x.
7: end for

Output: The labels of instances in the testing set.

4 Datasets

The efficient of the proposed algorithm iCST-Voting was evaluated using the Wisconsin Diagnosis Breast Cancer dataset, the Coimbra dataset and the CT Medical images dataset.

- *Wisconsin Diagnosis Breast Cancer (WDBC) dataset*: The collected data results of the efforts made at the University of Wisconsin Hospital for the diagnosis and prognosis of breast tumors solely based on FNA test. This test involves fluid extraction from a breast mass using a small gauge needle and then visual inspection of the fluid under a microscope. The dataset contains of 569 instances (357 benign - 212 malignant), where each instance has 32 attributes and represents FNA test measurements for one diagnosis case.
- *Coimbra dataset*: This dataset is constituted by 10 predictors, all quantitative, and a binary dependent variable, indicating the presence or absence of breast cancer (Patrício et al. 2018). The predictors are anthropometric data and parameters which can be gathered in routine blood analysis. Prediction models based on these predictors, if accurate, can potentially be used as a biomarker of breast cancer.
- *CT Medical images dataset*: This data collection[1] contains 100 images (Albertina et al. 2016) which constitute part of a much larger effort, focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from *the cancer genome Atlas* (Clark et al. 2013). The images consist of the middle slice of all Computed Tomography (CT) images taken where valid age, modality and contrast tags could be found which results in 475 series from 69 different patients. The dataset is designed to allow different methods to be evaluated for examining the trends in CT image data associated with, using contrast and patient age.

5 Experimental results

In this section, we present a series of experiments in order to evaluate the performance of the proposed algorithm iCST-Voting for cancer prediction against the most efficient and frequently utilized self-labeled algorithms i.e. Self-training, Co-training, Tri-training, Co-Bagging, CST-Voting, Co-Forest and Democratic-Co learning. The first five self-labeled methods were evaluated by deploying as base learners the Sequential Minimum Optimization (SMO) (Platt 1998), the Naive Bayes (NB) (Domingos & Pazzani 1997) and the kNN algorithm (Aha 1997). These supervised classifiers probably constitute the most effective and popular machine learning algorithms for classification problems (Wu et al. 2008). In order to study the influence of the amount of labeled data, three different ratios (R) of the training data were used, i.e. 20%, 30% and 40% and all self-labeled algorithms were evaluated using the stratified 10-fold cross-validation.

[1] `https://www.kaggle.com/kmader/siim-medical-images/home`

The implementation code was written in JAVA, making use of the WEKA 3.9 Machine Learning Toolkit (Hall et al. 2009). The configuration parameters for all supervised classifiers and self-labeled algorithms, utilized in our experiments, are presented in Table 1.

Algorithm		Parameters
SMO	Supervised base learner	$C = 1.0$, tolerance parameter = 0.001. Epsilon = 1.0×10^{-12}, Kernel type = polynomial, Polynomial degree = 1, Fit logistic models = true.
NB	Supervised base learner	No parameters specified.
kNN	Supervised base learner	Number of neighbors = 3, Euclidean distance.
Self-training	Self-labeled (single classifier)	MaxIter = 40. $c = 0.95$.
Co-training	Self-labeled (multiple classifier)	MaxIter = 40. Initial unlabeled pool = 75.
Tri-training	Self-labeled (multiple classifier)	No parameters specified.
CST-Voting	Ensemble of self-labeled	No parameters specified.
Democratic-Co	Self-labeled (multiple classifier)	Classifiers = kNN, C4.5, NB.
Co-Bagging	Self-labeled (multiple classifier)	Committee members = 3, Ensemble learning = Bagging.
Co-Forest	Self-labeled (multiple classifier)	Number of Random Forest classifiers = 6. Threshold = 0.75.

Table 1 Parameter specification for all the base learners and self-labeled methods used in the experimentation

The performance of the classification algorithms was evaluated utilizing the performance metrics: F-measure (F_1) and Accuracy (Acc) which are respectively defined by

$$F_1 = \frac{2T_P}{2T_P + F_N + F_P}, \qquad Acc = \frac{T_P + T_N}{T_P + T_N + F_P + F_N},$$

where T_P stands for the number of instances which have been correctly classified as positive, T_N stands for the number of instances which have been correctly classified as negative, F_P (type I error) stands for the number of instances which have been wrongly classified as positive, F_N (type II error) stands for the number of instances which have been wrongly classified as negative. Notice that F_1 consists of a harmonic mean of precision and recall while Accuracy is the ratio of correct predictions of a classifier.

Table 2 presents the classification performance of the self-labeled algorithms on F_1 metric, relative to all labeled ratios. Notice that the highest classification accuracy is highlighted in bold for each base learner. Clearly, the proposed algorithm iCST-Voting presented the best performance regarding F_1 metric, outperforming all self-labeled algorithms, relative to all dataset sets and utilized labeled ratio.

Table 3 presents the classification accuracy of all self-labeled algorithms based on the performance metric Acc, regarding each labeled ratio. As mentioned above, the accuracy measure of the best-performing algorithm is highlighted in bold. The aggregated results showed that iCST-Voting was by far the most efficient and robust method relative to all benchmarks. In more detail, iCST-Voting exhibited the highest classification accuracy regarding all utilized labeled ratio. Furthermore, based on the discussion, we conclude that the incorporation of an ensemble of supervised classifiers as a base learner improved the performance of the classical algorithm CST-Voting.

Algorithm	WDBC			Coimbra			CT Medical		
	$R = 20\%$	$R = 30\%$	$R = 40\%$	$R = 20\%$	$R = 30\%$	$R = 40\%$	$R = 30\%$	$R = 30\%$	$R = 40\%$
Self (SMO)	95.24%	94.63%	94.34%	69.92%	66.18%	65.19%	70.59%	74.29%	75.00%
Co (SMO)	94.81%	94.81%	94.81%	66.67%	66.67%	66.67%	55.56%	55.56%	57.14%
Tri (SMO)	95.01%	95.01%	94.96%	68.75%	69.06%	66.67%	70.59%	73.08%	75.73%
CoBag (SMO)	95.24%	94.63%	94.34%	69.92%	66.18%	69.06%	70.59%	71.84%	75.00%
CST (SMO)	95.49%	95.47%	96.67%	69.17%	69.01%	69.06%	72.55%	74.29%	75.00%
Self (NB)	91.04%	91.69%	91.04%	69.23%	63.25%	71.54%	73.08%	74.51%	75.00%
Co (NB)	91.25%	91.25%	92.02%	60.99%	63.31%	62.50%	56.86%	60.00%	62.00%
Tri (NB)	91.98%	92.42%	92.45%	66.17%	67.72%	69.70%	73.79%	74.51%	75.73%
CoBag (NB)	91.08%	91.25%	91.80%	71.32%	64.46%	72.00%	73.79%	74.51%	73.79%
CST (NB)	91.94%	92.16%	92.49%	68.15%	70.87%	70.23%	73.79%	74.51%	75.73%
Self (kNN)	92.45%	94.00%	91.87%	71.54%	68.15%	69.86%	68.69%	72.00%	74.00%
Co (kNN)	92.42%	92.42%	92.42%	49.57%	61.76%	65.73%	43.01%	58.59%	65.22%
Tri (kNN)	93.05%	93.56%	92.53%	69.23%	69.57%	67.77%	69.90%	69.90%	71.15%
CoBag (kNN)	91.73%	93.05%	90.95%	69.77%	69.06%	73.33%	58.06%	67.35%	71.29%
CST (kNN)	93.33%	94.00%	92.60%	70.87%	68.57%	72.59%	69.90%	73.79%	75.73%
Co-Forest	93.08%	93.59%	93.81%	50.85%	67.21%	71.43%	66.67%	67.96%	68.63%
Democratic Co	94.69%	95.90%	95.90%	71.07%	72.73%	74.80%	70.59%	71.84%	73.08%
iCST	**96.15%**	**96.63%**	**97.36%**	**75.20%**	**77.78%**	**79.69%**	**73.79%**	**75.73%**	**78.10%**

Table 2 Performance of the self-labeled algorithms based on the F_1 performance metric for each labeled ratio

Algorithm	WDBC			Coimbra			CT Medical		
	$R = 20\%$	$R = 30\%$	$R = 40\%$	$R = 20\%$	$R = 30\%$	$R = 40\%$	$R = 30\%$	$R = 30\%$	$R = 40\%$
Self (SMO)	96.49%	96.13%	95.78%	68.10%	60.34%	59.48%	70.00%	73.00%	74.00%
Co (SMO)	96.13%	96.13%	96.13%	54.31%	54.31%	54.31%	60.00%	60.00%	61.00%
Tri (SMO)	96.31%	96.31%	96.31%	65.52%	62.93%	62.93%	70.00%	72.00%	75.00%
CoBag (SMO)	96.49%	96.13%	95.78%	68.10%	60.34%	62.93%	70.00%	71.00%	74.00%
CST (SMO)	96.66%	96.66%	97.54%	64.66%	62.07%	62.93%	72.00%	73.00%	74.00%
Self (NB)	93.32%	93.85%	93.32%	65.52%	62.93%	69.83%	72.00%	74.00%	74.00%
Co (NB)	93.50%	93.50%	94.02%	52.59%	56.03%	58.62%	56.00%	60.00%	62.00%
Tri (NB)	94.02%	94.38%	94.38%	61.21%	64.66%	65.52%	73.00%	74.00%	75.00%
CoBag (NB)	93.32%	93.50%	93.85%	68.10%	62.93%	69.83%	73.00%	74.00%	73.00%
CST (NB)	94.02%	94.20%	94.38%	62.93%	68.10%	66.38%	73.00%	74.00%	75.00%
Self (kNN)	94.38%	95.61%	94.02%	69.83%	62.93%	62.07%	69.00%	72.00%	74.00%
Co (kNN)	94.38%	94.38%	94.38%	49.14%	55.17%	57.76%	47.00%	59.00%	68.00%
Tri (kNN)	94.90%	95.25%	94.55%	65.52%	63.79%	66.38%	69.00%	69.00%	70.00%
CoBag (kNN)	93.85%	94.90%	93.32%	66.38%	62.93%	65.52%	61.00%	68.00%	71.00%
CST (kNN)	95.08%	95.61%	94.55%	68.10%	62.07%	68.10%	69.00%	73.00%	75.00%
Co-Forest	94.90%	95.25%	95.43%	50.00%	65.52%	68.97%	65.00%	67.00%	68.00%
Democratic Co	96.13%	97.01%	97.01%	69.83%	71.55%	73.28%	70.00%	71.00%	72.00%
iCST	**97.19%**	**97.54%**	**98.07%**	**73.28%**	**75.86%**	**77.59%**	**73.00%**	**75.00%**	**77.00%**

Table 3 Performance of the self-labeled algorithms based on the *Acc* performance metric for each labeled ratio

6 Conclusions and future work

In this work, we present a new ensemble semi-supervised self-labeled algorithm, entitled iCST-Voting, for the prediction of cancer. The proposed ensemble algorithm exploits the individual predictions of three of the most efficient and popular self-labeled algorithms: Self-training, Co-training and Tri-training, utilizing an ensemble as base learner to increase their classification accuracy. Our experiments indicated the efficiency and robustness of the proposed algorithm, resulting in an increased classification accuracy comparing against the most popular self-labeled algorithms.

Our future work is focused on improving the prediction accuracy of our framework by combining the prediction of the self-labeled algorithms using a more sophisticated weighted voting methodology. Additionally, another interesting aspect is concentrated on extending our experiments by applying the proposed algorithm to several biomedical datasets for image classification.

References

Abdel-Zaher, A. M. & Eldeib, A. M. (2016), 'Breast cancer classification using deep belief networks', *Expert Systems with Applications* **46**, 139–144.

Aha, D. (1997), *Lazy learning*, Dordrecht: Kluwer academic publishers.

Albertina, B., Watson, M., Holback, C., Jarosz, R., Kirk, S., Lee, Y. & Lemmerman, J. (2016), 'Radiology data from the cancer genome atlas lung adenocarcinoma [TCGA-LUAD] collection', *The Cancer Imaging Archive* .

Blum, A. & Mitchell, T. (1998), Combining labeled and unlabeled data with co-training, *in* '11th annual conference on computational learning theory', pp. 92–100.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D. & Pringle, M. (2013), 'The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository', *Journal of Digital Imaging* **26**(6), 1045–1057.

DeSantis, C., Ma, J., Bryan, L. & Jemal, A. (2014), 'Breast cancer statistics, 2013', *CA: a cancer journal for clinicians* **64**(1), 52–62.

Dollinger, M. & Rosenbaum, E. (2002), *Everyone's Guide to Cancer Therapy;: How Cancer Is Diagnosed, Treated, and Managed Day to Day*, Andrews McMeel Publishing.

Domingos, P. & Pazzani, M. (1997), 'On the optimality of the simple Bayesian classifier under zero-one loss', *Machine learning* **29**, 103–130.

Edwards, A. & Elwyn, G. (2009), *Shared decision-making in health care: Achieving evidence-based patient choice*, Oxford University Press.

Engelen, A., Vanderhaegen, J., Van Poppel, H. & Van Audenhove, C. (2016), 'Patients views on using decision support tools: A systematic review', *European Journal for Person Centered Healthcare* **4**(1), 61–186.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S. (2017), 'Dermatologist-level classification of skin cancer with deep neural networks', *Nature* **542**(7639), 115.

Fatima, M. & Pasha, M. (2017), 'Survey of machine learning algorithms for disease diagnostic', *Journal of Intelligent Learning Systems and Applications* **9**(01), 1.

Hady, M. & Schwenker, F. (2010), 'Combining committee-based semi-supervised learning and active learning', *Journal of Computer Science and Technology* **25**(4), 681–698.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. (2009), 'The WEKA data mining software: An update', *SIGKDD explorations newsletters* **11**, 10–18.

Hussain, L., Ahmed, A., Saeed, S., Rathore, S., Awan, I. A., Shah, S. A., Majid, A., Idris, A. & Awan, A. A. (2018), 'Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies', *Cancer Biomarkers* (Preprint), 1–21.

Li, M. & Zhou, Z. (2007), '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.

Livieris, I. (2019), 'A new ensemble self-labeled semi-supervised algorithm', *Informatica* pp. 1–14, (to be appear).

Livieris, I., Drakopoulou, K., Tampakas, V., Mikropoulos, T. & Pintelas, P. (2018a), An ensemble-based semi-supervised approach for predicting students' performance, *in* 'Research on e-Learning and ICT in education', Elsevier.

Livieris, I., Kanavos, A., Tampakas, V. & Pintelas, P. (2018b), 'An ensemble SSL algorithm for efficient chest X-ray image classification', *Journal of Imaging* **4**(7).

Morvan, C. & Jenkins, W. J. (2017), *Judgment under Uncertainty: Heuristics and Biases*, Macat Library.

Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seiça, R. & Caramelo, F. (2018), 'Using resistin, glucose, age and BMI to predict the presence of breast cancer', *BMC cancer* **18**(1), 29.

Platt, J. (1998), *Advances in Kernel Methods - Support Vector Learning*, MIT Press, Cambridge, Massachusetts.

Révész, D., Engelhardt, E., Tamminga, J., Schramel, F. M., Onwuteaka-Philipsen, B. D., van de Garde, E., Steyerberg, E. W., Jansma, E., De Vet, H. C. & Coupé, V. M. (2017), 'Decision support systems for incurable non-small cell lung cancer: A systematic review', *BMC medical informatics and decision making* **17**(1), 144.

Rokach, L. (2010), *Pattern classification using ensemble methods*, World Scientific Publishing Company.

Sasieni, P. D. & Parkin, D. M. (2018), 'Global perspectives surrounding cancer prevention and screening', *Cancer Prevention and Screening: Concepts, Principles and Controversies* pp. 1–18.

Turgut, S., Dağtekin, M. & Ensari, T. (2018), Microarray breast cancer data classification using machine learning methods, *in* '2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)', IEEE, pp. 1–3.

Vidić, I., Egnell, L., Jerome, N. P., Teruel, J. R., Sjøbakk, T. E., Østlie, A., Fjøsne, H. E., Bathen, T. F. & Goa, P. E. (2018), 'Support vector machine for breast cancer classification using diffusion-weighted mri histogram features: Preliminary study', *Journal of Magnetic Resonance Imaging* **47**(5), 1205–1216.

Wu, X., Kumar, V., Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G., Ng, A., Liu, B., Yu, P., Zhou, Z., Steinbach, M., Hand, D. & Steinberg, D. (2008), 'Top 10 algorithms in data mining', *Knowledge and information systems* **14**(1), 1–37.

Yarowsky, D. (1995), Unsupervised word sense disambiguation rivaling supervised methods, *in* 'Proceedings of the 33rd annual meeting of the association for computational linguistics', pp. 189–196.

Zhou, Y. & Goldman, S. (2004), Democratic co-learning, *in* '16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)', IEEE, pp. 594–602.

Zhou, Z. (2011), When semi-supervised learning meets ensemble learning, Vol. 6, Springer, pp. 6–16.

Zhou, Z. & Li, M. (2005), 'Tri-training: Exploiting unlabeled data using three classifiers', *IEEE transactions on knowledge and data engineering* **17**(11), 1529–1541.