


Article

A Weighted Voting Ensemble Self-Labeled Algorithm for the Detection of Lung Abnormalities from X-rays

Ioannis E. Livieris^{1,*} , Andreas Kanavos¹, Vassilis Tampakas¹, Panagiotis Pintelas²

¹ Department of Computer & Informatics Engineering (DISK Lab), Technological Educational Institute of Western Greece, Antirion, GR 263-34, Greece; kanavos@ceid.upatras.gr, vtampakas@teimes.gr

² Department of Mathematics, University of Patras, GR 265-00, Greece; ppintelas@gmail.com

* Correspondence: livieris@teiwest.gr

Version March 13, 2019 submitted to Algorithms

Abstract: During last decades intensive efforts have been devoted to the extraction of useful knowledge from large volumes of medical data employing advanced machine learning and data mining techniques. Advances in digital chest radiography, have enabled research and medical centers to accumulate large repositories of classified (labeled) images and mostly of unclassified (unlabeled) images from human experts. Machine learning methods such as semi-supervised learning algorithms have been proposed as a new direction to address the problem of shortage of available labeled data, by exploiting the explicit classification information of labeled data with the information hidden in the unlabeled data. In the present work, we propose a new ensemble semi-supervised learning algorithm for the classification of lung abnormalities from chest X-rays based on a new weighted voting scheme. The proposed algorithm assigns a vector of weights on each component classifier of the ensemble based on its accuracy on each class. Our numerical experiments illustrate the efficiency of the proposed ensemble methodology against other state-of-the-art classification methods.

Keywords: Machine learning; semi-supervised learning; self-labeled algorithms; classifiers; ensemble learning; weighted voting; image classification; lung abnormalities.

1. Introduction

The automatic detection of abnormalities, diseases and pathologies constitutes a significant factor in computer-aided medical diagnosis and a vital component in radiologic image analysis. For over a century, radiology is a typical method for abnormality detection. A typical radiological examination is performed by utilizing a posterior-anterior chest radiograph, which is most commonly called *Chest X-Ray* (CXR). CXR imaging is widely used for health diagnosis and monitoring, due to its relatively low cost and easy accessibility, thus it has been established as the single most acquired medical image modality [1]. It constitutes a significant factor for the detection and diagnosis of several pulmonary diseases, such as tuberculosis, lung cancer, pulmonary embolism and interstitial lung disease [1]. However, due to increasing workload pressures, many radiologists today have to examine an enormous number of CXRs daily. Thus, a prediction system trained to predict the risk of specific abnormalities given a particular CXR image is considered essential for providing high quality medical assistance. More specifically, such a decision support system has the potential to support the reading workflow, improve efficiency and reduce prediction errors. Moreover, it could be used to enhance the confidence of the radiologist or prioritize the reading list where critical cases would be read first.

The significant advances in digital chest radiography and the continuously enlarged storage capabilities of electronic media, have enabled research centers to accumulate large repositories of

34 classified (labeled) images and mostly of unclassified (unlabeled) images from human experts. To
35 this end, researchers and medical staff were able to leverage and exploit these images by the adoption
36 of machine learning and data mining techniques for the development of intelligent computational
37 systems in order to extract useful and valuable information. As a result, the areas of biomedical
38 research and diagnostic medicine have been dramatically transformed, from rather qualitative
39 sciences which were based on observations of whole organisms to more quantitative sciences which
40 are now based on the extraction of useful knowledge from voluminous of data [2].

41 Nevertheless, distinguishing the various chest abnormalities from CXRs is a rather challenging
42 task, not only for a prediction model but even for an human expert. The progress in the field has been
43 hampered by the lack of available labeled images for efficiently training a powerful and accurate
44 supervised classification model. Moreover, the process of correctly labeling new unlabeled CXRs
45 usually incurs monetary costs and high time since it constitutes a long and complicated process and
46 requires the efforts of specialized personnel and expert physicians.

47 Semi-Supervised Learning (SSL) algorithms have been proposed as a new direction to address
48 the problem of shortage of available labeled data, comprising characteristics of both supervised
49 and unsupervised learning algorithms. These algorithms efficiently develop powerful classifiers
50 by meaningfully relating the explicit classification information of labeled data with the information
51 hidden in the unlabeled data [3,4]. Self-labeled algorithms probably constitute the most popular
52 class of SSL algorithms due to their simplicity of implementation, their wrapper-based philosophy
53 and good classification performance [2,5–8]. This class of algorithms exploits a large amount of
54 unlabeled data via a self-learning process based on supervised learners. In other words, they perform
55 an iterative procedure, enriching the initial labeled data, based on the assumption that their own
56 predictions tend to be correct.

57 Recently, Triguero et al. [9] proposed an in-depth taxonomy based on the main characteristics
58 presented in them and conducted a comprehensive research of their classification efficacy on several
59 datasets. Generally, self-labeled algorithm can be classified in two main groups: *Self-training* and
60 *Co-training*. In the original Self-training [10], a single classifier is iteratively trained on an enlarged
61 labeled dataset with its most confident predictions on unlabeled data while in Co-training [11], two
62 classifiers are separately trained utilizing two different views on a labeled dataset and then each
63 classifier augments the labeled data of the other with its most confident predictions on unlabeled
64 data. Along this line, several self-labeled algorithms have been proposed in the literature, while
65 some of them exploit ensemble methodologies and techniques.

66 Democratic-Co learning [12] is based on an ensemble philosophy since it uses three independent
67 classifiers following a majority voting and a confidence measurement strategy for predicting the
68 values of unlabeled examples. Tri-training algorithm [13] utilizes a bagging ensemble of three
69 classifiers which are trained on data subsets generated through bootstrap sampling from the original
70 labeled set and teach each other using on majority voting strategy. Co-Forest [14] utilizes bootstrap
71 sample data from the labeled set in order to train Random trees. At each iteration, each random
72 tree is reconstructed by newly selected unlabeled instances for its concomitant ensemble, utilizing a
73 majority voting technique. Co-Bagging [15] trains multiple base classifiers on bootstrap data created
74 by random resampling with replacement from the training set. Each bootstrap sample contains about
75 2/3 of the original training set, where each example can appear multiple times. Recently, a new
76 approach is given by Livieris et al. [2,8,16,17] and Livieris [18] in which some ensemble self-labeled
77 algorithms are proposed based on voting schemes. The proposed algorithms exploit the individual
78 predictions of the most efficient and frequently used self-labeled algorithms using simple voting
79 methodologies.

80 Motivated by these works, we propose a new semi-supervised self-labeled algorithm, which
81 is based on a sophisticated ensemble philosophy. The proposed algorithm exploits the individual
82 predictions of self-labeled algorithms, using a new weighted voting methodology. The proposed
83 weighted strategy assigns weights on each component classifier of the ensemble based on its accuracy

84 on each class. Our main aim is to measure the effectiveness of our weighted voting ensemble scheme
85 over the majority voting ensembles, using identical component classifiers in all cases. On top of
86 that we want to verify that powerful classification models could be developed by the adaptation of
87 advanced ensemble methodologies in the SSL framework. Our preliminary numerical experiments
88 prove the efficiency and the classification accuracy of the proposed algorithm, demonstrating that
89 reliable prediction models could be developed by incorporating ensemble methodologies in the
90 semi-supervised framework.

91 The remainder of this paper is organized as follows: Section 2 presents a brief survey of recent
92 studies concerning the application of machine learning for the detection of lung abnormalities from
93 X-rays. Section 3 presents a detailed description of the proposed weighted voting scheme and
94 ensemble algorithm. Section 4 presents a series of experiments carried out in order to examine and
95 evaluate the accuracy of the proposed algorithm against the most popular self-labeled classification
96 algorithms. Finally, Section 5 discusses the conclusions and some research topics for future work.

97 2. Related work

98 The significance of medical imaging for the diagnosis of diseases has been established, for the
99 treatment of chest pathologies and their early detection. During the last decades, the advances
100 of digital technology and chest radiography as well as the rapid development of digital image
101 retrieval have renewed the progress in new technologies for the diagnosis of lung abnormalities.
102 More specifically, research has been focused on the development of Computer-Aided Diagnostic
103 (CAD) models for abnormality detection in order to assist medical staff. Along this line, a variety
104 of methodologies have been proposed based on machine learning techniques, aiming on classifying
105 and/or detecting abnormalities in patients' medical images. A number of studies have been carried
106 out in recent years; some useful outcomes of them are briefly presented below.

107 Jaeger et al. [19] proposed a CAD system for tuberculosis in conventional posteroanterior chest
108 radiographs. Their proposed model initially utilizes a graph cut segmentation method to extract the
109 lung region from the CXRs and then a set of texture and shape features in the lung region is computed
110 in order to classify the patient as normal or abnormal. Their extensive numerical experiments on two
111 real-world datasets illustrated the efficiency of the proposed CAD system for tuberculosis screening,
112 achieving higher performance compared to that of human readings.

113 Melendez et al. [20] recommend a novel CAD system for detecting tuberculosis on chest
114 X-rays based on multiple-instance learning. Their proposed system is based on the idea of utilizing
115 probability estimations, instead of the sign of a decision function, to guide the multiple-instance
116 learning process. Furthermore, an advantage of their method is that it does not require labeling
117 of each feature sample during the training process but only a global class label characterizing a group
118 of samples.

119 Alam et al. [21] utilized a multi-class support vector machine classifier and developed an efficient
120 lung cancer detection and prediction model. The image enhancement and the image segmentation
121 have been done independently, in every stage of the classification process. Image scaling, color space
122 transformation and contrast enhancement have been utilized for image enhancement while threshold
123 and marker-controlled watershed have been utilized for segmentation. In the sequel, the support
124 vector machine classifier categorizes a set of textural features extracted from the separated regions of
125 interest. Based on their numerical experiments, the authors concluded that the proposed algorithm
126 can efficiently detect a cancer affected cell and its corresponding stage such as initial, middle, or final.
127 Furthermore, in case no cancer affected cell is found in the input image then it checks the probability
128 of lung cancer.

129 In more recent works, Madani [22] focused on the detection of abnormalities in chest X-ray
130 images, having available only a fairly small size dataset of annotated images. Their proposed
131 method deals with both problems of labeled data scarcity and data domain overfitting, by utilizing
132 Generative Adversarial Networks (GAN) in a SSL architecture. In general, GAN utilize two networks:

133 a generator which seeks to create as realistic images as possible and a discriminator which seeks to
134 distinguish between real data and generated data. Next, these networks are involved in a minimax
135 game to find the Nash equilibrium between them. Based on their experiments the author concluded
136 that the annotation effort is reduced considerably to achieve similar performance through supervised
137 training techniques.

138 In [2], Livieris et al. evaluated the classification efficacy of an ensemble SSL algorithm, called
139 CST-Voting, for CXR classification of tuberculosis. The proposed algorithm combines the individual
140 predictions of three efficient self-labeled algorithms i.e Co-training, Self-training and Tri-training
141 using a simple majority voting methodology. The authors presented some interesting results,
142 illustrating the efficiency of the proposed algorithm against several classical algorithms. Additionally,
143 their experiments lead them to the conclusion that reliable and robust prediction models could
144 be developed utilizing a few labeled and many unlabeled data. In [16] the authors extended the
145 previous work and proposed DTCo algorithm for the classification of X-rays. The proposed ensemble
146 algorithm exploits the predictions of Democratic-Co learning, Tri-training and Co-Bagging utilizing
147 a maximum-probability voting scheme. Along this line, Livieris et al. [17] proposed EnSL algorithm
148 which constitutes a generalized scheme of the previous works. More specifically, EnSL constitutes
149 a majority voting scheme of N self-labeled algorithms. Their preliminary numerical experiments
150 demonstrated that robust classification models could be developed by the adaptation of ensemble
151 methodologies in the SSL framework.

152 Guan and Huang [23] considered the problem of multi-label thorax disease classification on chest
153 X-ray images by proposing a Category-wise Residual Attention Learning (CRAL) framework. CRAL
154 predicts the presence of multiple pathologies in a class-specific attentive view, aiming to suppress the
155 obstacles of irrelevant classes by endowing small weights to the corresponding feature representation
156 while the same time, the relevant features would be strengthened by assigning larger weights. More
157 analytically, their proposed framework consists of two modules: feature embedding module and
158 attention learning module. The feature embedding module learns high-level features using a neural
159 network classifier while the attention learning module focuses on exploring the assignment scheme
160 of different categories. Based on their numerical experiments, the authors stated that their proposed
161 methodology constitutes a new state of the art.

162 3. A new weighted voting ensemble self-labeled algorithm

163 In this section, we present a detailed description of the proposed self-labeled algorithm, which
164 is based on an ensemble philosophy, entitled Weighed voting Ensemble Self-Labeled (WvEnSL)
165 algorithm.

166 Generally, the generation of an ensemble of classifiers considers mainly two steps: *Selection* and
167 *Combination*. The selection of the component classifiers is considered essential for the efficiency of
168 the ensemble and the key point for its efficacy is based on their diversity and their accuracy; while
169 the combination of the individual classifiers' predictions takes place through several techniques with
170 different philosophy [24,25].

171 By taking these into consideration, the proposed algorithm is based on the idea of selecting
172 a set $C = (C_1, C_2, \dots, C_N)$ of N self-labeled classifiers by applying different algorithms (with
173 heterogeneous model representations) to a single dataset and the combination of their individual
174 predictions takes place through a new weighted voting methodology. It is worth noticing that
175 weighted voting is a commonly used strategy for combining predictions in pairwise classification
176 in which the classifiers are not treated equally. Each classifier is evaluated on a evaluation set D and
177 associated with a coefficient (weight), usually proportional to its classification accuracy.

178 Let us consider a dataset D with M classes, which is utilized for the evaluation of each
179 component classifier. More specifically, the performance of each classifier C_i , with $i = 1, 2, \dots, N$
180 is evaluated on D and a $N \times M$ matrix W is defined, as follows

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,M} \\ w_{2,1} & w_{2,2} & \dots & w_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \dots & w_{N,M} \end{bmatrix}$$

where each element $w_{i,j}$ is defined by

$$w_{i,j} = \frac{2p_j^{(C_i)}}{|D_j| + p_j^{(C_i)} + q_j^{(C_i)}}, \quad (1)$$

181 where D_j is the set of instances of the dataset belonging to the class j , $p_j^{(C_i)}$ are the number of correct
 182 predictions of classifier C_i on D_j and $q_j^{(C_i)}$ are the number of incorrect predictions of C_i that an instance
 183 belongs to class j . Clearly, each weight $w_{i,j}$ is the F_1 -score of classifier C_i for j class [26]. The rationale
 184 behind (1) is to measure the efficiency of each classifier, relative to each class j of the evaluation set D .

Subsequently, the class \hat{y} of each unknown instance x in the test set is computed by

$$\hat{y} = \arg \max_j \sum_{i=1}^N w_{i,j} \chi_A(C_i(x) = j),$$

185 where function $\arg \max$ returns the value of index corresponding to the largest value from array,
 186 $A = \{1, 2, \dots, M\}$ is the set of unique class labels and χ_A is the characteristic function which takes
 187 into account the prediction $j \in A$ of a classifier C_i on an instance x and creates a vector in which
 188 the j coordinate takes a value of one and the rest take the value of zero. At this point, it is worth
 189 mentioning that in our implementation we selected to evaluate the performance of each classifier of
 190 the ensemble on the initial training labeled set L .

191 A high-level description of the proposed framework is presented in Table 1 which consists
 192 of three phases: *Training*, *Evaluation* and *Weighted-Voting Prediction*. In the Training phase, the
 193 self-labeled algorithms, which constitute the ensemble are trained utilizing the same labeled L and
 194 unlabeled dataset U (Steps 1-3). Subsequently, in the Evaluation phase, the trained classifiers are
 195 evaluated using the training set L in order to calculate the weight matrix W (Steps 4-9). Finally, in
 196 the Weighted-Voting Prediction phase, the final hypothesis on each unlabeled example x of the test
 197 set combines the individual predictions of self-labeled algorithms utilizing the proposed weighted
 198 voting methodology (Steps 10-15). An overview of the proposed WvEnSL is depicted in Figure 1.

Table 1. WvEnSL framework

Input: L – Set of labeled instances (Training labeled set).
 U – Set of unlabeled instances (Training unlabeled set).
 T – Set of unlabeled test instances (Testing set).
 D – Set of instances for evaluation (Evaluation set).
 $C = (C_1, C_2, \dots, C_N)$ – Set of self-labeled classifiers which constitute the ensemble.

Output: The labels of instances in the testing set.

/ Phase I: Training */*

Step 1: for $i = 1$ to N do

Step 2: Train C_i using the labeled L and the unlabeled dataset U .

Step 3: end for

/ Phase II: Evaluation */*

Step 4: for $i = 1$ to N do

Step 5: Apply C_i on the evaluation set D .

199 **Step 6:** for $j = 1$ to M do
 200 **Step 7:** Calculate the weight

$$w_{i,j} = \frac{2p_j^{(C_i)}}{|D_j| + p_j^{(C_i)} + q_j^{(C_i)}}.$$

201 **Step 8:** end for
 202 **Step 9:** end for

203
 204 /* Phase III: Weighted-Voting Prediction */

205 **Step 10:** for each $x \in T$ do
 206 **Step 11:** for $i = 1$ to N do
 207 **Step 12:** Apply classifier C_i on x .
 208 **Step 13:** end for
 209 **Step 14:** Predict the label \hat{y} of x using

$$\hat{y} = \arg \max_j \sum_{i=1}^N w_{i,j} \chi_A(C_i(x) = j).$$

210 **Step 15:** end for

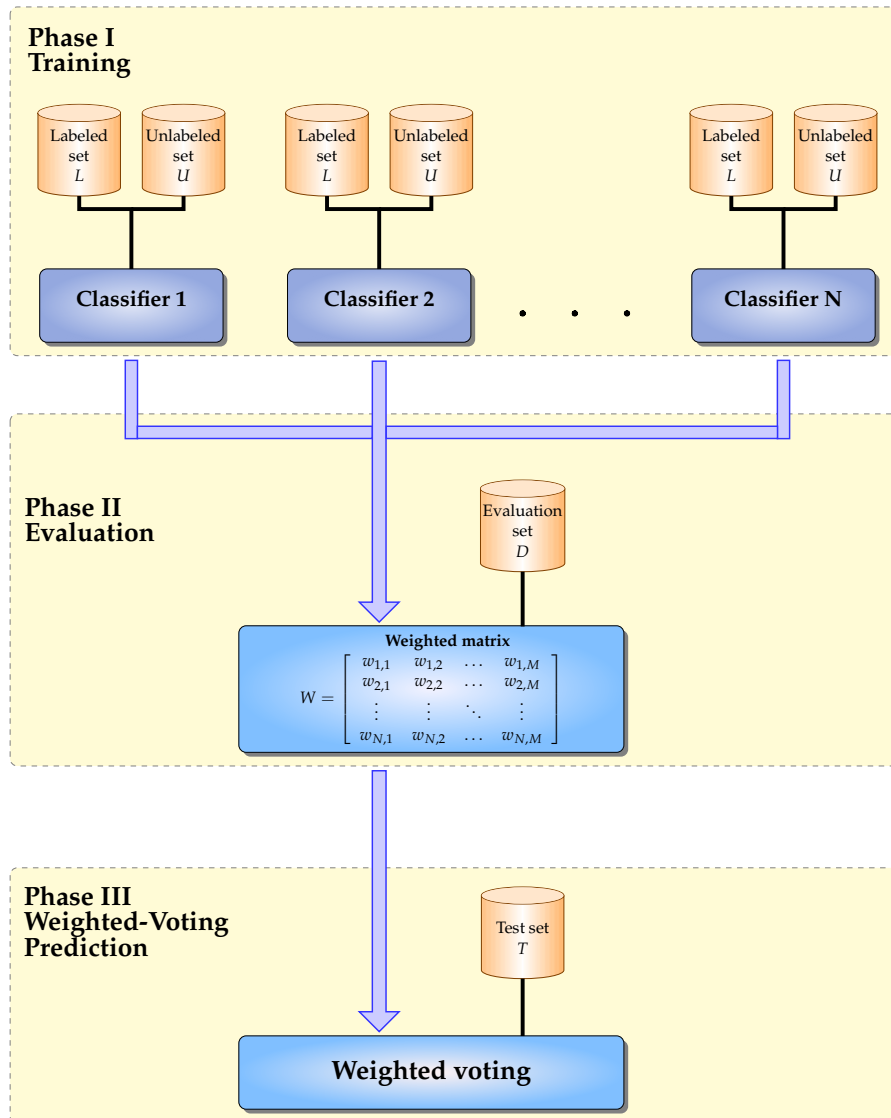


Figure 1. WvEnSL framework

213 4. Experimental methodology

214 In this section, we present a series of experiments in order to evaluate the performance of the
215 proposed WvEnSL algorithm for X-ray classification against the most efficient ensemble self-labeled
216 algorithms i.e. CST-Voting, DTCo and EnSL which utilize simple voting methodologies. The
217 implementation code was written in JAVA, making use of the WEKA 3.9 Machine Learning Toolkit
218 [27].

219 The performance of the classification algorithms is evaluated using the following performance
220 metrics: F -measure (F_1) and Accuracy (Acc). It is worth mentioning that F_1 consists of a harmonic
221 mean of precision and recall while Accuracy is the ratio of correct predictions of a classifier.

222 4.1. Datasets

223 The compared classification algorithms were evaluated utilizing the chest X-ray (Pneumonia)
224 dataset, the Shenzhen lung mask (Tuberculosis) dataset and the CT Medical images dataset.

- 225 • *Chest X-ray (Pneumonia) dataset*: The dataset contains 5830 chest X-ray images
226 (anterior-posterior) which were selected from retrospective cohorts of pediatric patients
227 of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou.
228 All chest X-ray imaging was performed as part of patients’ routine clinical care. For the analysis
229 of chest X-ray images, all chest radiographs were initially screened for quality control by
230 removing all low quality or unreadable scans. The diagnoses for the images were then graded
231 by two expert physicians before being cleared for training the artificial intelligence system. In
232 order to account for any grading errors, the evaluation set was also checked by a third expert.
233 The dataset was partitioned into two sets (training/testing). The training set consisting of 5216
234 examples (1341 normal, 3875 pneumonia) and the testing set with 624 examples (234 normal,
235 390 pneumonia) as in [28].
- 236 • *Shenzhen lung mask (Tuberculosis) dataset*: Shenzhen Hospital is one of the largest hospitals in
237 China for infectious diseases with a focus both on their prevention, as well as treatment. The
238 X-rays were collected within a one-month period, mostly in September 2012, as a part of the
239 daily routine, using a Philips DR Digital Diagnost system. The dataset was constructed by
240 manually-segmented lung masks for the Shenzhen Hospital X-ray set as presented in [29]. These
241 segmented lung masks were originally utilized for the description of the lung segmentation
242 technique in combination with lossless and lossy data augmentation. The segmentation masks
243 for the Shenzhen Hospital X-ray set were manually prepared by students and teachers of
244 the Computer Engineering Department, Faculty of Informatics and Computer Engineering,
245 National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” [29]. The
246 set contained 279 normal CXRs and 287 abnormal ones with tuberculosis. All classification
247 algorithms were evaluated using the stratified 10-fold cross-validation.
- 248 • *CT Medical images dataset*: This data collection contains 100 images [30] which constitute part of a
249 much larger effort, focused on connecting cancer phenotypes to genotypes by providing clinical
250 images matched to subjects from *the cancer genome Atlas* [31]. The images consist of the middle
251 slice of all Computed Tomography (CT) images taken from 69 different patients. The dataset
252 is designed to allow different methods to be evaluated for examining the trends in CT image
253 data associated with using contrast and patient age. The basic idea is to identify image textures,
254 statistical patterns and features correlating strongly with these traits and possibly build simple
255 tools for automatically classifying these images when they have been misclassified (or finding
256 outliers which could be suspicious cases, bad measurements, or poorly calibrated machines).
257 All classification algorithms were evaluated using the stratified 10-fold cross-validation.

258 The training partition was randomly divided into labeled and unlabeled subsets. In order to
 259 study the influence of the amount of labeled data, four different ratios (R) of the training data were
 260 used: 10%, 20%, 30% and 40%. Using the recommendation established in [9,32] in the division process
 261 we do not maintain the class proportion in the labeled and unlabeled sets since the main aim of
 262 semi-supervised classification is to exploit unlabeled data for better classification results. Hence, we
 263 use a random selection of examples that will be marked as labeled instances, and the class label of
 264 the rest of the instances will be removed. Furthermore, we ensure that every class has at least one
 265 representative instance.

266 4.2. Performance evaluation of WvEnSL against ensemble self-labeled algorithms

267 Next, we focus our interest on the experimental analysis for evaluating the classification
 268 performance of WvEnSL algorithm against the ensemble self-labeled algorithms CST-Voting and
 269 DTCo, which utilize simple voting methodologies. It is worth noticing that our main goal is to
 270 measure the effectiveness of the proposed weighted voting strategy over the simple majority voting;
 271 therefore we will compare ensembles using identical set of classifiers. This will eliminate the source
 272 of discrepancy originated from unequal classifiers. Thus, the difference in accuracy can solely be
 273 attributed to the difference of voting methodologies.

274 Furthermore, the base learners utilized in all self-labeled algorithms are the Sequential Minimum
 275 Optimization (SMO) [33], the C4.5 decision tree algorithm [34] and the k NN algorithm [35] as in
 276 [2,7–9]. which probably constitute the most effective and popular machine learning algorithms for
 277 classification problems [36].

- 278 • “CST-Voting (SMO)” stands for an ensemble of Co-training, Self-training and Tri-training with
 279 SMO as base learner using majority voting [2].
- 280 • “WvEnSL₁ (SMO)” stands for Algorithm WvEnSL using the same components classifiers as
 281 CST-Voting (SMO).
- 282 • “CST-Voting (C4.5)” stands for an ensemble of Co-training, Self-training and Tri-training with
 283 C4.5 as base learner using majority voting [2].
- 284 • “WvEnSL₁ (C4.5)” stands for Algorithm WvEnSL using the same components classifiers as
 285 CST-Voting (C4.5).
- 286 • “CST-Voting (k NN)” stands for an ensemble of Co-training, Self-training and Tri-training with
 287 k NN as base learner using majority voting [2].
- 288 • “WvEnSL₁ (k NN)” stands for Algorithm WvEnSL using the same components classifiers as
 289 CST-Voting (k NN).
- 290 • “DTCo” stands for an ensemble of Democratic-Co learning, Tri-training and Co-Bagging with
 291 C4.5 as base learner using majority voting [16].
- 292 • “WvEnSL₂” stands for Algorithm WvEnSL using the same components classifiers as DTCo.
- 293 • “EnSL” stands for an ensemble of Self-training, Democratic-Co learning, Tri-training and
 294 Co-Bagging with C4.5 as base learner using majority voting [17].
- 295 • “WvEnSL₃” stands for Algorithm WvEnSL using the same components classifiers as EnSL.

296 The configuration parameters for all supervised classifiers and self-labeled algorithms, utilized
 297 in our experiments, are presented in Table 2.

298 Tables 3, 4 and 5 presents the performance of all ensemble self-labeled methods on Pneumonia
 299 dataset, Tuberculosis dataset and CT Medical dataset, respectively. Notice that the highest
 300 classification performance for each ensemble of classifiers and performance metric is highlighted
 301 in bold. The aggregated results showed that the new weighted voting strategy exploits the
 302 individual predictions of each component classifier more efficiently than the simple voting schemes,

303 illustrating better classification performance. WvEnSL₃ exhibits the best performance, reporting
 304 the highest F_1 -score and accuracy, relative to all classification benchmarks and labeled ratio,
 305 followed by WvEnSL₂. In more detail, WvEnSL₃ demonstrates 82.53%-83.49%, 69.79%-71.73% and
 306 69%-77% classification accuracy for Pneumonia dataset, Tuberculosis dataset and CT Medical dataset,
 307 respectively; while WvEnSL₂ reports 81.89%-83.17%, 69.79%-71.55% and 67%-77%, in the same
 308 situations.

Table 2. Parameter specification for all the base learners and self-labeled methods used in the experimentation

Algorithm		Parameters
SMO	Supervised base learner	$C = 1.0$, Tolerance parameter = 0.001, Pearson VII function-based kernel, $\text{Epsilon} = 1.0 \times 10^{-12}$, Fit logistic models = true.
C4.5	Supervised base learner	Confidence level: $c = 0.25$, Minimum number of item-sets per leaf: $i = 2$, Prune after the tree building.
kNN	Supervised base learner	Number of neighbors = 3, Euclidean distance.
Self-training	Self-labeled (single classifier)	MaxIter = 40, $c = 95\%$.
Co-training	Self-labeled (multiple classifier)	MaxIter = 40, Initial unlabeled pool = 75
Tri-training	Self-labeled (multiple classifier)	No parameters specified.
Co-Bagging	Self-labeled (multiple classifier)	Committee members = 3, Ensemble learning = Bagging.
Democratic-Co	Self-labeled (multiple classifier)	Classifiers = kNN, C4.5, NB.
CST-Voting	Ensemble of self-labeled	No parameters specified.
DTCo	Ensemble of self-labeled	No parameters specified.
EnSL	Ensemble of self-labeled	No parameters specified.

Table 3. Performance evaluation of WvEnSL against ensemble self-labeled algorithms for Pneumonia dataset

Algorithm	Ratio = 10%		Ratio = 20%		Ratio = 30%		Ratio = 40%	
	F_1	Acc	F_1	Acc	F_1	Acc	F_1	Acc
CST-Voting (SMO)	83.08%	75.32%	83.26%	75.64%	83.39%	75.80%	83.39%	75.80%
WvEnSL ₁ (SMO)	83.39%	75.80%	83.48%	75.96%	83.76%	76.44%	83.85%	76.60%
CST-Voting (C4.5)	85.52%	79.97%	85.85%	80.45%	86.68%	81.73%	86.58%	81.57%
WvEnSL ₁ (C4.5)	85.65%	80.13%	86.08%	80.77%	86.78%	81.89%	86.92%	82.05%
CST-Voting (kNN)	82.91%	75.48%	83.09%	75.80%	83.15%	75.96%	83.73%	76.76%
WvEnSL ₁ (kNN)	83.63%	76.60%	83.73%	76.76%	83.95%	77.08%	84.23%	77.56%
DTCo	86.79%	81.41%	87.21%	82.05%	87.21%	82.05%	87.74%	82.85%
WvEnSL ₂	87.12%	81.89%	87.44%	82.37%	87.54%	82.53%	87.97%	83.17%
EnSL	87.19%	82.05%	86.92%	81.57%	87.34%	82.21%	87.61%	82.69%
WvEnSL ₃	87.51%	82.53%	87.70%	82.69%	88.23%	83.49%	88.17%	83.49%

Table 4. Performance evaluation of WvEnSL against ensemble self-labeled algorithms for Tuberculosis dataset

Algorithm	Ratio = 10%		Ratio = 20%		Ratio = 30%		Ratio = 40%	
	F_1	Acc	F_1	Acc	F_1	Acc	F_1	Acc
CST-Voting (SMO)	69.27%	69.43%	68.65%	68.37%	69.50%	69.61%	70.42%	70.32%
WvEnSL ₁ (SMO)	69.73%	69.79%	69.73%	69.79%	70.32%	70.32%	71.00%	70.85%
CST-Voting (C4.5)	66.67%	67.31%	68.19%	68.02%	67.51%	68.20%	69.52%	69.79%
WvEnSL ₁ (C4.5)	67.86%	68.20%	69.26%	69.26%	69.63%	69.79%	69.98%	70.14%
CST-Voting (kNN)	65.71%	66.08%	66.43%	66.96%	68.21%	68.55%	68.93%	69.26%
WvEnSL ₁ (kNN)	65.83%	66.25%	67.14%	67.49%	68.57%	68.90%	69.40%	69.61%
DTCo	69.73%	69.79%	69.96%	69.96%	71.45%	71.20%	71.80%	71.55%
WvEnSL ₂	69.73%	69.79%	70.19%	70.14%	71.58%	71.38%	71.80%	71.55%
EnSL	69.73%	69.79%	69.96%	69.96%	71.00%	70.85%	71.58%	71.38%
WvEnSL ₃	69.73%	69.79%	70.19%	70.14%	71.58%	71.38%	72.03%	71.73%

Table 5. Performance evaluation of WvEnSL against ensemble self-labeled algorithms for CT Medical dataset

Algorithm	Ratio = 10%		Ratio = 20%		Ratio = 30%		Ratio = 40%	
	F_1	Acc	F_1	Acc	F_1	Acc	F_1	Acc
CST-Voting (SMO)	66.67%	66.00%	70.00%	70.00%	73.08%	72.00%	75.00%	74.00%
WvEnSL ₁ (SMO)	68.00%	68.00%	71.29%	71.00%	73.79%	73.00%	75.73%	75.00%
CST-Voting (C4.5)	67.96%	67.00%	71.84%	71.00%	73.79%	73.00%	73.79%	73.00%
WvEnSL ₁ (C4.5)	69.90%	69.00%	73.79%	73.00%	75.00%	74.00%	75.73%	75.00%
CST-Voting (kNN)	66.00%	66.00%	69.90%	69.00%	73.79%	73.00%	73.27%	73.00%
WvEnSL ₁ (kNN)	66.67%	67.00%	70.59%	70.00%	72.00%	72.00%	74.75%	75.00%
DTCo	66.02%	65.00%	69.90%	69.00%	72.55%	72.00%	74.29%	73.00%
WvEnSL ₂	67.33%	67.00%	71.29%	71.00%	72.55%	72.00%	76.92%	76.00%
EnSL	64.08%	63.00%	71.84%	71.00%	74.29%	73.00%	74.29%	73.00%
WvEnSL ₃	69.90%	69.00%	75.73%	75.00%	76.47%	76.00%	77.67%	77.00%

309 The statistical comparison of several classification algorithms over multiple datasets is
 310 fundamental in the area of machine learning and it is usually performed by means of a statistical test
 311 [2,7–9]. Since our motivation stems from the fact that we are interested in evaluating the rejection
 312 of the hypothesis that all the algorithms perform equally well for a given level based on their
 313 classification accuracy and highlighting the existence of significant differences between our proposed
 314 algorithm and the classical self-labeled algorithms, we utilized the non-parametric Friedman Aligned
 315 Ranking (FAR) [37] test.

Let r_i^j be the rank of the j -th of k learning algorithms on the i -th of M problems. Under the null-hypothesis H_0 , which states that all the algorithms are equivalent, the Friedman aligned ranks test statistic is defined by:

$$F_{AR} = \frac{(k-1) \left[\sum_{j=1}^k \hat{R}_j^2 - (kM^2/4)(kM+1)^2 \right]}{\frac{kM(kM+1)(2kM+1)}{6} - \frac{1}{k} \sum_{i=1}^M \hat{R}_i^2}$$

316 where \hat{R}_i is equal to the rank total of the i -th dataset and \hat{R}_j is the rank total of the j -th algorithm.
 317 The test statistic F_{AR} is compared with the χ^2 distribution with $(k-1)$ degrees of freedom. It is

318 worth noticing that, FAR test does not require the commensurability of the measures across different
 319 datasets, since it is non-parametric, neither assumes the normality of the sample means, and thus, it
 320 is robust to outliers.

Additionally, in order to identify which algorithms report significant differences, the Finner test [38] with a significance level $\alpha = 0.05$, is applied as a post-hoc procedure. More analytically, the Finner procedure adjusts the value of α in a step-down manner. Let p_1, p_2, \dots, p_{k-1} be the ordered p -values with $p_1 \leq p_2 \leq \dots \leq p_{k-1}$ and H_1, H_2, \dots, H_{k-1} be the corresponding hypothesis. The Finner procedure rejects H_1-H_{i-1} if i is the smallest integer such that $p_i > 1 - (1 - \alpha)^{(k-1)/i}$, while the adjusted Finner p -value is defined by:

$$p_F = \min \left\{ 1, \max \left\{ 1 - (1 - p_j)^{(k-1)/j} \right\} \right\},$$

321 where p_j is the p -value obtained for the j -th hypothesis and $1 \leq j \leq i$. It is worth mentioning that the
 322 test rejects the hypothesis of equality when the p_F is less than α .

323 The control algorithm for the post-hoc test is determined by the best (lowest) ranking obtained
 324 in each FAR test. Moreover, the adjusted p -value with Finner's test (p_F) was presented based on
 325 the corresponding control algorithm at the α level of significance while the post-hoc test rejects the
 326 hypothesis of equality when the value of p_F is less than the value of α . It is worth mentioning that
 327 the FAR test and the Finner post-hoc test were performed based on the classification accuracy of each
 328 algorithm over all datasets and labeled ratio.

329 Table 6 presents the information of the statistical analysis performed by nonparametric multiple
 330 comparison procedures for all ensemble self-labeled algorithms. The interpretation of Table 6
 331 demonstrates that WvEnSL₃ reports the highest probability-based ranking by statistically presenting
 332 better results, followed by WvEnSL₂ and WvEnSL₁ (C4.5). Moreover, it is worth mentioning that all
 333 weighted voting ensemble outperformed the corresponding ensemble which utilize classical voting
 334 schemes. Finally, based on the statistical analysis, we can easily conclude that the new weighted
 335 voting scheme had a significant impact on the performance of all ensemble of self-labeled algorithms.

Table 6. Friedman Aligned Ranking (FAR) test and Finner post-hoc test

Algorithm	FAR	Finner Post-Hoc Test	
		p_F -value	Null Hypothesis
WvEnSL ₃	15.667	-	-
WvEnSL ₂	34.958	0.174312	accepted
WvEnSL ₁ (C4.5)	44.208	0.049863	rejected
EnSL	47.958	0.029437	rejected
DTCo	51.125	0.018734	rejected
CST-Voting (C4.5)	64.042	0.001184	rejected
WvEnSL ₁ (SMO)	71.417	0.000194	rejected
CST-Voting (SMO)	88.292	0.000001	rejected
WvEnSL ₁ (kNN)	89.083	0.000001	rejected
CST-Voting (kNN)	98.250	0.000001	rejected

336 4.3. Performance evaluation of WvEnSL against classical supervised algorithms

337 Next, we compare the classification performance of the proposed algorithm against the classical
 338 supervised classification algorithms: SMO, C4.5 and kNN. Moreover, we compare the performance
 339 of iCST-Voting against the ensemble of classifiers (Voting) which combines the individual predictions
 340 of the supervised classifiers utilizing a simple majority voting strategy. It is worth noticing that

- 341 • we selected WvEnSL₃ from all versions of the proposed algorithm since it presented the best
 342 overall performance.
- 343 • all supervised algorithms were trained using with 100% of the training set while WvEnSL₃ was
 344 trained using $R = 40\%$ of the training set.

345 Table 7 presents the performance of the proposed algorithm WvEnSL₃ against the supervised
 346 algorithms SMO, C4.5, kNN and Voting on Pneumonia dataset, Tuberculosis dataset and CT Medical
 347 dataset. As above mentioned, the highest classification performance for each labeled ratio and
 348 performance metric is highlighted in bold. The aggregated results show that WvEnSL₃ is the most
 349 efficient algorithm since it illustrates the best overall classification performance. More specifically,
 350 WvEnSL₃ exhibits the highest F_1 -score and classification accuracy on Pneumonia and Tuberculosis
 351 datasets while for CT Medical dataset, WvEnSL₃ reports the second best performance, considerably
 352 outperformed by C4.5.

Table 7. Performance evaluation WvEnSL₃ against state-of-the-art supervised algorithms on Pneumonia dataset, Tuberculosis dataset and CT Medical dataset

Algorithm	Pneumonia		Tuberculosis		CT Medical	
	F_1	Acc	F_1	Acc	F_1	Acc
SMO	74.03%	76.76%	71.41%	71.37%	74.91%	75.00%
C4.5	72.41%	74.83%	62.32%	62.36%	79.82%	80.00%
3NN	72.32%	74.51%	67.51%	67.49%	67.08%	67.00%
Voting	73.34%	76.12%	71.00%	71.02%	74.07%	74.00%
WvEnSL ₃	88.17%	83.49%	72.03%	71.73%	77.67%	77.00%

353 5. Conclusions

354 In this work, we proposed a new weighted voting ensemble self-labeled algorithm for the
 355 detection of lung abnormalities from X-rays, entitled WvEnSL. The proposed algorithm combines
 356 the individual predictions of self-labeled algorithms utilizing a new weighted voting methodology.
 357 The significant advantage of WvEnSL is that weights assigned on each component classifier of the
 358 ensemble are based on its accuracy on each class of the dataset.

359 For testing purposes, the algorithm was extensively evaluated using the chest X-rays
 360 (Pneumonia) dataset, the Shenzhen lung mask (Tuberculosis) dataset and the CT Medical images
 361 dataset. Our numerical experiments indicated better classification accuracy of the WvEnSL and
 362 demonstrated the efficiency of the new weighted voting scheme, as statistically confirmed by the
 363 Friedman Aligned Ranks nonparametric test as well as the Finner post hoc test. Therefore, we
 364 can conclude that the new weighted voting strategy had a significant impact on the performance
 365 of all ensembles of self-labeled algorithms, exploiting the individual predictions of each component
 366 classifier more efficiently than the simple voting schemes. Finally, it is worth mentioning that efficient
 367 and powerful classification models could be developed by the adaptation of ensemble methodologies
 368 in the SSL framework.

369 In our future work, we intend to pursue extensive empirical experiments to compare the
 370 proposed WvEnSL with other algorithms, belonging to different SSL classes and evaluate its
 371 performance using various component self-labeled algorithms and base learners. Furthermore,
 372 since our preliminary numerical experiments are quite encouraging, our next step is to explore the
 373 performance of the proposed algorithm on imbalanced datasets [39,40] and incorporate our proposed
 374 methodology for multi-target problems [41–43]. Additionally, another interesting aspect is the use
 375 of other component classifiers in the ensemble and enhance our proposed framework with more
 376 sophisticated and theoretically sound criteria for the development of an advanced weighted voting
 377 strategy. Finally, we intent to investigate and evaluate different strategies for the selection of the
 378 evaluation set.

379 **Author Contributions:** I.E. Livieris, A. Kanavos, V. Tampakas and P. Pintelas conceived of the idea, designed and
 380 performed the experiments, analyzed the results, drafted the initial manuscript and revised the final manuscript.

381 **Funding:** This research received no external funding.

382 **Conflicts of Interest:** The authors declare no conflict of interest.

383

- 384 1. Van Ginneken, B.; Stegmann, M.B.; Loog, M. Segmentation of anatomical structures in chest radiographs
385 using supervised methods: a comparative study on a public database. *Medical image analysis* **2006**,
386 *10*, 19–40.
- 387 2. Livieris, I.; Kanavos, A.; Tampakas, V.; Pintelas, P. An ensemble SSL algorithm for efficient chest X-ray
388 image classification. *Journal of Imaging* **2018**, *4*.
- 389 3. Zhu, X.; Goldberg, A. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence*
390 *and machine learning* **2009**, *3*, 1–130.
- 391 4. Chapelle, O.; Scholkopf, B.; Zien, A. Semi-supervised learning. *IEEE Transactions on Neural Networks* **2009**,
392 *20*, 542–542.
- 393 5. Levatic, J.; Dzeroski, S.; Supek, F.; Smuc, T. Semi-supervised learning for quantitative structure-activity
394 modeling. *Informatica* **2013**, *37*, 173.
- 395 6. Levatic, J.; Ceci, M.; Kocev, D.; Dzeroski, S. Semi-supervised classification trees. *Journal of Intelligent*
396 *Information Systems* **2017**, *49*, 461–486.
- 397 7. Livieris, I.; Kanavos, A.; Tampakas, V.; Pintelas, P. An auto-adjustable semi-supervised self-training
398 algorithm. *Algorithm* **2018**, *11*.
- 399 8. Livieris, I.; Kiriakidou, N.; Kanavos, A.; Tampakas, V.; Pintelas, P. On ensemble SSL algorithms for credit
400 scoring problem. *Informatics* **2018**, *5*.
- 401 9. Triguero, I.; García, S.; Herrera, F. Self-labeled techniques for semi-supervised learning: taxonomy,
402 software and empirical study. *Knowledge and information systems* **2015**, *42*, 245–284.
- 403 10. Yarowsky, D. Unsupervised word sense disambiguation rivaling supervised methods. Proceedings of
404 the 33rd annual meeting of the association for computational linguistics, 1995, pp. 189–196.
- 405 11. Blum, A.; Mitchell, T. Combining labeled and unlabeled data with co-training. 11th annual conference
406 on computational learning theory, 1998, pp. 92–100.
- 407 12. Zhou, Y.; Goldman, S. Democratic co-learning. 16th IEEE International Conference on Tools with Artificial
408 Intelligence (ICTAI). IEEE, 2004, pp. 594–602.
- 409 13. Zhou, Z.; Li, M. Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on*
410 *Knowledge and Data Engineering* **2005**, *17*, 1529–1541.
- 411 14. Li, M.; Zhou, Z. Improve computer-aided diagnosis with machine learning techniques using undiagnosed
412 samples. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **2007**,
413 *37*, 1088–1098.
- 414 15. Hady, M.; Schwenker, F. Combining committee-based semi-supervised learning and active learning.
415 *Journal of Computer Science and Technology* **2010**, *25*, 681–698.
- 416 16. Livieris, I.; Kotsilieris, T.; Anagnostopoulos, I.; Tampakas, V. DTCo: An ensemble SSL algorithm for
417 X-rays classification. In *Advances in Experimental Medicine and Biology*; Springer, 2018.
- 418 17. Livieris, I.; Kanavos, A.; Pintelas, P. Detecting lung abnormalities from X-rays using and improved SSL
419 algorithm. *Electronic Notes of Theoretical Computer Science*, (accepted for publication) **2019**.
- 420 18. Livieris, I. A new ensemble self-labeled semi-supervised algorithm. *Informatica*, (accepted for publication)
421 **2018**.
- 422 19. Jaeger, S.; Karargyris, A.; Candemir, S.; Folio, L.; Siegelman, J.; Callaghan, F.; Xue, Z.; Palaniappan, K.;
423 Singh, R.; Antani, S.; Thoma, G.; Wang, Y.; Lu, P.; McDonald, C. Automatic tuberculosis screening using
424 chest radiographs. *IEEE Transactions on Medical Imaging* **2014**, *33*, 233–245.
- 425 20. Melendez, J.; van Ginneken, B.; Maduskar, P.; Philipsen, R.; Reither, K.; Breuninger, M.; Adetifa, I.; Maane,
426 R.; Ayles, H.; Sánchez, C. A novel multiple-instance learning-based approach to computer-aided detection
427 of tuberculosis on chest X-rays. *IEEE Transactions on Medical Imaging* **2015**, *34*, 179–192.
- 428 21. Alam, J.; Alam, S.; Hossan, A. Multi-Stage Lung Cancer Detection and Prediction Using Multi-class
429 SVM Classifier. 2018 International Conference on Computer, Communication, Chemical, Material and
430 Electronic Engineering. IEEE, 2018, pp. 1–4.

- 431 22. Madani, A.; Moradi, M.; Karargyris, A.; Syeda-Mahmood, T. Semi-supervised learning with generative
432 adversarial networks for chest X-ray classification with ability of data domain adaptation. 15th IEEE
433 International Symposium on Biomedical Imaging. IEEE, 2018, pp. 1038–1042.
- 434 23. Guan, Q.; Huang, Y. Multi-label chest X-ray image classification via category-wise residual attention
435 learning. *Pattern Recognition Letters* **2018**.
- 436 24. Dietterich, T. Ensemble methods in machine learning. *Multiple Classifier Systems*; Kittler, J.; Roli, F., Eds.
437 Springer Berlin Heidelberg, 2001, Vol. 1857, pp. 1–15.
- 438 25. Rokach, L. *Pattern classification using ensemble methods*; World Scientific Publishing Company, 2010.
- 439 26. Powers, D.M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and
440 correlation. *Journal of Machine Learning Technologies* **2011**, *2*, 37–63.
- 441 27. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. The WEKA data mining software:
442 An update. *SIGKDD explorations newsletters* **2009**, *11*, 10–18.
- 443 28. Kermany, D.; Goldbaum, M.; Cai, W.; Valentim, C.; Liang, H.; Baxter, S.; McKeown, A.; Yang, G.; Wu, X.;
444 Yan, F. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**,
445 *172*, 1122–1131.
- 446 29. Stirenko, S.; Kochura, Y.; Alienin, O.; Rokovyi, O.; Gang, P.; Zeng, W.; Gordienko, Y. Chest X-ray analysis
447 of tuberculosis by deep learning with segmentation and augmentation. *arXiv preprint arXiv:1803.01199*
448 **2018**.
- 449 30. Albertina, B.; Watson, M.; Holback, C.; Jarosz, R.; Kirk, S.; Lee, Y.; Lemmerman, J. Radiology data from
450 the cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. *The Cancer Imaging Archive*
451 **2016**.
- 452 31. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle,
453 M. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.
454 *Journal of Digital Imaging* **2013**, *26*, 1045–1057.
- 455 32. Wang, Y.; Xu, X.; Zhao, H.; Hua, Z. Semi-supervised learning based on nearest neighbor rule and cut
456 edges. *Knowledge-Based Systems* **2010**, *23*, 547–554.
- 457 33. Platt, J. *Advances in Kernel Methods - Support Vector Learning*; MIT Press: Cambridge, Massachusetts, 1998.
- 458 34. Quinlan, J. *C4.5: Programs for machine learning*; Morgan Kaufmann: San Francisco, 1993.
- 459 35. Aha, D. *Lazy learning*; Dordrecht: Kluwer academic publishers, 1997.
- 460 36. Wu, X.; Kumar, V.; Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.; Ng, A.; Liu, B.; Yu, P.; Zhou,
461 Z.; Steinbach, M.; Hand, D.; Steinberg, D. Top 10 algorithms in data mining. *Knowledge and information*
462 *systems* **2008**, *14*, 1–37.
- 463 37. Hodges, J.; Lehmann, E. Rank methods for combination of independent experiments in analysis of
464 variance. *The annals of mathematical statistics* **1962**, *33*, 482–497.
- 465 38. Finner, H. On a monotonicity problem in step-down multiple test procedures. *Journal of the American*
466 *statistical association* **1993**, *88*, 920–923.
- 467 39. Li, S.; Wang, Z.; Zhou, G.; Lee, S. Semi-supervised learning for imbalanced sentiment classification. IJCAI
468 proceedings-international joint conference on artificial intelligence, 2011, Vol. 22, p. 1826.
- 469 40. Jeni, L.A.; Cohn, J.F.; De La Torre, F. Facing imbalanced data – Recommendations for the use
470 of performance metrics. Humaine Association Conference on Affective Computing and Intelligent
471 Interaction. IEEE, 2013, pp. 245–251.
- 472 41. Levatić, J.; Ceci, M.; Kocev, D.; Džeroski, S. Self-training for multi-target regression with tree ensembles.
473 *Knowledge-Based Systems* **2017**, *123*, 41–60.
- 474 42. Levatić, J.; Kocev, D.; Džeroski, S. The importance of the label hierarchy in hierarchical multi-label
475 classification. *Journal of Intelligent Information Systems* **2015**, *45*, 247–271.
- 476 43. Levatić, J.; Kocev, D.; Ceci, M.; Džeroski, S. Semi-supervised trees for multi-target regression. *Information*
477 *Sciences* **2018**, *450*, 109–127.