

Identification of blood cell subtypes from images using an improved SSL algorithm

Ioannis E. Livieris^{1,*}, Emmanuel Pintelas², Andreas Kanavos¹, Panagiotis Pintelas³

¹*Department of Computer & Informatics Engineering, Technological Educational Institute of Western Greece, Greece, GR 263-34.*

²*Department of Electrical and Computer Engineering, University of Patras, GR 265-00, Greece.*

³*Department of Mathematics, University of Patras, GR 265-00, Greece.*

Abstract

Nowadays, the classification of blood cell subtypes constitutes a typical method for diagnosing many diseases, infections and inflammations. The application of an efficient cell classification method is considered essential in modern diagnostic medicine in order to increase the number of analyzed cells per patient and decrease the analysis time. The recent advances in digital technologies and the vigorous widespread of the Internet have ultimately led to the development of large repositories of images. Due to the effort and expense involved in labeling data, training datasets are of a limited size, while in contrast, electronic medical record systems contain a significant number of unlabeled images. Semi-supervised learning algorithms constitute the appropriate machine learning methodology to exploit the knowledge hidden in the unlabeled data with the explicit classification information of labeled data for building powerful and effective classifiers. In this work, we evaluate the performance of an ensemble semi-supervised learning algorithm for the classification of blood cell subtypes. The efficacy of the presented algorithm is illustrated by a series of experiments, demonstrating that reliable and robust prediction models could be developed by the adaptation of ensemble techniques in the semi-supervised learning framework.

Keywords: Blood cell subtypes, image classification, semi-supervised learning, self-labeled algorithms, ensemble learning.

1. Introduction

In recent years, a new era in the diagnostic medicine area has began by the adoption of machine learning and data mining techniques for the development of intelligent computational systems in order to extract useful and valuable information. Researchers have made significant efforts on the development of such systems which are able to efficiently analyze different types of medical images and extract useful knowledge [3, 10, 20, 26]. Therefore, the area of diagnostic medicine has massively changed from a rather qualitative science that was based on observations of whole organisms to a more quantitative science, which is also based on knowledge extraction from databases [15].

White blood cells, also called leukocytes, are the cells responsible for the protection of the human body against both infectious diseases and foreign invaders; thus they have been established as a significant part of the immune system [7, 9, 29]. There are five major types of subtypes of leukocytes: neutrophils, lymphocytes, monocytes, eosinophils and basophils which are distinguished by their physical and functional characteristics. A possible change in the number of different leukocytes subtypes in the blood is utilized as sign for various diseases. Therefore, the counting of blood cell subtypes in the bone marrow of a patient constitutes a very informative factor in clinical practice [19] since several blood-based diseases, infections and inflammations can often be early diagnosed by the characterization of patient blood samples. For example, patients with leukemia have often a higher level of lymphocytes due to malfunctioning of immune system and people suffering from allergies generally have an increase in their eosinophil counts.

*Corresponding author

Email address: livieris@teiwest.gr (Ioannis E. Livieris¹)

Therefore, blood cell classification and identification has acquired a lot of interest from laboratories and clinics since their proper counting can provide a powerful quantitative picture of people's health.

In general, there are two ways to classify patient's blood cells: the manual and the automated way. In the manual way, medical staff examines a sample of blood under a microscope and the identification between varying subtypes is accomplished based on characteristics of the cell morphology. Nevertheless, since the classification efficiency is highly dependent on the human experience, this process is a time consuming and repetitive task which can be influenced by operator's accuracy and tiredness [21]. The automated techniques were proposed in order to overcome the tedious and time-consuming task of human effort consumed with the manual way by utilizing machine learning and data mining [27].

With the vigorous development of the Internet and the widespread adoption of electronic medical records, research centers have accumulated large repositories of classified (labeled) images and mostly of unclassified (unlabeled) images from human experts. Hence, researchers have a significant potential to extract useful knowledge and transform biomedical research, by leveraging these images using machine learning methodologies.

Ongun et al. [18] developed an automated differential blood count system for feature extraction and classification of blood cells based on machine learning and data mining techniques. Motivated by the previous work, Osowski et al. [19] studied the application of a genetic algorithm for features selection and a support vector machine for the recognition of blood cells based on the images of the bone marrow aspirate. Their preliminary numerical experiments indicated that the use of the genetic algorithm for the selection of the diagnostic features constituted a significant role for improving the performance accuracy of the prediction model. Independently, Ramirez-Cortes et al. [24] proposed a methodology for the classification of leukocytes using the morphological pattern spectrum (pecstrum). Their experiments presented that the composed feature vector reported very good attributes to reflect the evolution in time of the white blood cells according to their maturity stage. In more recent works, Hegde et al. [8] proposed a robust image processing algorithm for nuclei detection and white blood cells classification based on features of the nuclei. More specifically, they utilized a novel image enhancement method to manage illumination variations and TissueQuant method to manage color variations for the detection of nuclei. The performance of their proposed method was presented against several state-of-the-art classification methods. Rawat et al. [25] presented a semi-automated technique for the identification and classification of white blood cells based on ensembles of binary artificial neural network classifiers. Their proposed method was evaluated utilizing a dataset containing 114 images, indicating the robustness and effectiveness of the proposed approach.

Nevertheless, the development of an accurate prediction model for cell classification is considered a rather difficult and challenging task. The main reason is that the progress in the medical field has been hampered by the lack of available labeled images for efficiently training an accurate classifier [15]. Furthermore, the process of correctly labeling new unlabeled images frequently requires the efforts of expert physicians and specialized personnel which constitutes a long and complicated process which will incur high time and monetary costs.

To address this problem, Semi-Supervised Learning (SSL) algorithms comprise the appropriate machine learning methodology for extracting useful knowledge exploiting both labeled and unlabeled data in order to build efficient classification models [34]. These algorithms constitute a combination of supervised and unsupervised learning, exploiting a small pool of labeled examples L , together with a large pool of unlabeled examples U , aiming to obtain better classification results. Their main objective is to efficiently combine the information hidden in the unlabeled data with the explicit classification information of labeled data. Self-labeled algorithms are generally considered the most popular class of SSL algorithms which follow an iterative procedure, aiming to obtain an enlarged labeled data set, in which they accept that their own predictions tend to be correct. From a theoretical point of view, Triguero et al. [28] proposed an in-depth taxonomy based on the main characteristics presented in them and conducted an exhaustive study of their classification efficacy on several datasets.

In this work, we evaluate and examine the performance of a new ensemble self-labeled algorithm, called EnSSL, for the classification of blood cell subtypes from images. The proposed algorithm combines the predictions of three of the most efficient and frequently used self-labeled algorithms, utilizing a maximum probability-based voting scheme. Our preliminary numerical experiments indicated the efficacy of the EnSSL, illustrating that reliable and robust classification models could be developed by the adaptation of ensemble methodologies in the semi-supervised learning framework.

The remainder of this paper is organized as follows: Section 2 presents a brief description of the self-labeled algorithms and the proposed ensemble semi-supervised classification algorithm. Section 3 presents a series of experiments

in order to evaluate the accuracy of the presented self-labeled algorithms on the classification of blood cell subtypes while Section 4 presents our conclusions.

2. On semi-supervised self-labeled classification

In this section, we present a brief description of semi-supervised classification and the most popular self-labeled algorithms proposed in the literature. Generally, self-labeled algorithms are considered a significant family of classification methods which progressively classify unlabeled data based on the most confident predictions, without making any specific assumptions about the input data.

Self-training algorithm [31] is considered as the simplest and one of the most efficient self-labeled algorithms. It is based on a wrapper philosophy which constitutes an iterative procedure of self-labeling unlabeled data. More specifically, in the self-training framework, a classifier is initially trained with a small number of labeled examples and at each iteration its training set is augmented gradually with the most confident predictions and then re-trained. However, an obvious disadvantage of self-training is that this methodology can lead to erroneous predictions if noisy examples are classified as the most confident examples and incorporated into the labeled training set.

Co-training [2] constitutes a multi-view algorithm which can be considered as a different variant of self-training technique. This self-labeled algorithm is based on the assumption that the feature space can be divided in two conditionally independent views, each view being sufficient to train an efficient classifier. Under this assumption, two base learners are trained separately on each view, utilizing the initial labeled dataset and each base learner iteratively augments the training set of the other with its most confident predictions. Essentially, Co-training is a “*two-view weakly supervised algorithm*” since it uses the self-training approach on each view [17]. However, the assumption about the existence of sufficient and redundant views is a luxury hardly met in most real world scenarios [12, 13].

Tri-training algorithm [33] constitutes an improved single-view extension of the Co-training algorithm based on an ensemble methodology. It utilizes three classifiers which are trained on data subsets generated through bootstrap sampling from the original labeled training set. In each Tri-training round, if two classifiers agree on the labeling of an unlabeled instance while the third one disagrees, then these two classifiers will label this instance for the third classifier. It is worth noticing that this algorithm is based on the “*majority teach minority strategy*” which serves as an implicit confidence measurement avoiding thereby the use of complicated time-consuming approaches to explicitly measure the predictive confidence and as a result the training process is efficient [14].

Zhou and Goldman [32] proposed a multi-view learning algorithm, entitled *Democratic-Co learning*, which is based on the idea of incorporating majority voting in the SSL learning framework. This algorithm utilizes multiple algorithms for producing the necessary information and endorses a voted majority process for the final decision instead of demanding multiple views of the corresponding data. Motivated by the previous work, Li and Zhou [11] presented the *Co-Forest* algorithm. This algorithm trains Random trees on bootstrap data from the dataset and assigns a few unlabeled examples to each Random tree. Ultimately, the final decision is composed by a simple majority voting. A great asset on comparison with the rest self-labeled algorithms is the reduced fluctuations of its performance, under the condition that short number of labeled instances is provided. Furthermore, the default tactic of Random Tree classifier to construct trees from randomly chosen features of the basic feature vector means that no physical connection among the attributes of the collected data is required.

Co-Bagging [5] creates several base classifiers using the same learning algorithm on a bootstrap sample created by random resampling with replacement from the original training set. Each bootstrap sample contains about $2/3$ of the original training set, where each example can appear multiple times. This technique works well for unstable learning algorithms, where a small change in the input training set can lead to a major change in the output hypothesis.

2.1. Ensemble semi-supervised learning algorithm

In the sequel, we present a detailed description of the proposed SSL algorithm for the classification of blood cells images which is based on an ensemble philosophy.

Ensemble Semi-Supervised Learning algorithm (*EnSSL*) [12, 16] constitutes a SSL algorithm which efficiently exploits the individual prediction of three of the most popular self-labeled algorithms i.e. Self-training, Co-training and Tri-training utilizing a maximum probability-based voting scheme. The main difference between the selected self-labeled algorithms which constitute the ensemble are the utilized methodology to label unlabeled data since

Self-training and Tri-training are based on the single-view self-labeled technique, while Co-training is based on the multi-view self-labeled technique. A high-level description of the EnSSL algorithm is presented in Algorithm 1 which consists of two phases: *Training* phase and *Testing* phase.

In the Training phase, the self-labeled algorithms which constitute the ensemble are trained on using the same labeled L and unlabeled U datasets (Steps 1-3). In the Testing phase, EnSSL determines the final hypothesis on each unlabeled example x of the test set T , exploiting the individual predictions of the self-labeled algorithms. Initially the trained SSL algorithms are applied on each instance x in the test set (Step 6) and then the classifier which exhibits the most confident prediction over the unlabeled example x is selected (Step 7). 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. In this case, the output of the ensemble is defined as the combined predictions of three self-labeled learning algorithms via a simple majority voting. Finally, it is worth mentioning that the way in which the confidence predictions are measured depends on the type of utilized base learner (see [6] and the references there in).

Algorithm 1: **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-training( $L, U$ ).
2: Train Co-training( $L, U$ ).
3: Train Tri-training( $L, U$ ).

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

```

3. Experimental methodology

In this section, we conducted a series of experiments in order to evaluate the performance of EnSSL against the most popular and frequently utilized self-labeled algorithms in terms of classification accuracy. Accuracy is one of the most frequently used measures for assessing the overall effectiveness of a classification algorithm and is defined as the percentage of correctly classified instances. The experiments in our study took place in two distinct parts:

- In the first part, we evaluated the classification performance of EnSSL against its component SSL algorithms and in particular Self-training, Co-training and Tri-training.
- In the second part, we compared its performance against some state-of-the-art self-labeled algorithms, namely Co-Forest, Co-Bagging and Democratic-Co learning.

The implementation code was written in JAVA, making use of the WEKA Machine Learning Toolkit [6]. In order to study the influence of the amount of labeled data, four different ratios (R) of the training data were used, i.e. 10%, 20%, 30% and 40%. The configuration parameters for all SSL algorithms, utilized in our experiments, are presented in Table 1 while all base learners were used with their default parameter settings included in the WEKA 3.9 software [6]. Moreover, similar to Blum and Mitchell [2], a limit to the number of iterations of all self-labeled algorithms is established. This strategy has also been adopted by many researchers [12–16, 28].

Self-labeled algorithm	Parameters
Self-training	Threshold = 0.95.
Co-training	Initial unlabeled pool = 75.
Tri-training	No parameters specified.
Co-Forest	Number of Random Forest classifiers = 6, Threshold = 0.75.
Co-Bagging	Committee members = 3, Ensemble learning = Bagging.
Democratic-Co	Classifiers = NB, k NN, C4.5.

Table 1: Parameter specification for all self-labeled algorithms

3.1. Dataset

All algorithms evaluated their classification performance on the blood cells images dataset¹. This dataset contains 12515 augmented images of blood cells of four different cell types which was partitioned into two sets (training/testing). The training set consisting of 5216 examples (2510 eosinophils, 2489 lymphocytes, 2482 monocytes, 2547 neutrophils) and the testing set with 2487 examples (623 eosinophils, 620 lymphocytes, 620 monocytes, 624 neutrophils)

3.2. Performance evaluation of SSL algorithms

In the sequel, we focus our interest on the experimental analysis for evaluating the classification performance of EnSSL against its component self-labeled methods, i.e. Self-training, Co-training and Tri-training. All SSL algorithms were evaluated by deploying as base learners the Naive Bayes, the Sequential Minimum Optimization, the C4.5 decision tree and the k NN algorithm [1]. These algorithms probably constitute the most effective and popular machine learning algorithms for classification problems [30]. A brief description of the utilized supervised classifiers is given below:

- Naive Bayes (NB) [4] classifier constitutes one of the most popular classification technique for data mining and machine learning. The basic aim of this classifier is to construct a rule which will allow us to assign future objects to a class, assuming independence of attributes when probabilities are established. For continuous data, we follow a typical assumption in which continuous values associated with each class are distributed according to a Gaussian distribution. Notice that the probabilities' extraction is straightforward, due to the fact that this method explicitly computes the probability belonging to each class for the given test instance.
- Sequential Minimal Optimization (SMO) [22] is an efficient algorithm for training Support Vector Machines (SVM). It was originally proposed by Platt [22] and has been established as one of the simplest and fastest method for training a SVM. The main idea of this algorithm is derived from solving dual quadratic optimization

¹<https://www.kaggle.com/paultimothymooney/blood-cells>

problem by optimizing the minimal subset including two elements at each iteration. The advantages of SMO are its simplicity of implementation and its low memory requirements, which allows to handle very large training sets.

- C4.5 [23] constitutes one of the most effective and efficient classification algorithm for building decision trees. This algorithm induces classification rules in the form of decision trees for a given training set. More analytically, it categorizes instances to a predefined set of classes according to their attribute values from the root of a tree down to a leaf. The accuracy of a leaf corresponds to the percentage of correctly classified instances of the training set.
- k NN [1] constitutes a representative instance-structured learning algorithm based on dissimilarities among a set of instances. It belongs to the lazy learning family of methods [1], which do not build a model during the learning process. According to k NN algorithm, characteristics extracted from classification process, viewing the entire distance among new individual which should be classified and earlier individuals and then the nearest k category is used. As a result of this process, test data belongs to the nearest k neighbor category which has more members in certain class. The main advantages of the k NN classification algorithm is its easiness and simplicity of implementation and the fact that it provides good generalization results during classification assigned to multiple categories.

Table 2 presents the classification performance of Self-training, Co-training, Tri-training and EnSSL, relative to all labeled ratios. Notice that the highest classification accuracy is highlighted in bold for each base learner. Firstly, it is worth noticing that all self-labeled algorithms exhibit the best performance using k NN as base learner. EnSSL exhibits the best performance, relative to all base learners and all utilized labeled ratio. Moreover, using k NN as base learner presents the highest classification performance, correctly classifying 93.29% of the test instances using 40% as labeled ratio. Finally, a more representative visualization of the accuracy of the compared self-labeled algorithm is presented in Figure 1. Each box-plot presents the accuracy for each tested algorithm according to the supervised base learner and labeled ratio.

Base learner	Ratio	Self-training	Co-training	Tri-training	EnSSL
NB	10%	54.60%	59.35%	60.11%	61.64%
	20%	56.82%	60.80%	60.72%	61.88%
	30%	57.54%	60.47%	61.36%	69.52%
	40%	60.03%	60.23%	61.16%	75.23%
SMO	10%	48.97%	48.01%	49.06%	50.02%
	20%	48.61%	48.17%	49.34%	50.66%
	30%	50.10%	49.74%	49.78%	50.99%
	40%	51.23%	50.10%	50.18%	51.75%
C4.5	10%	86.85%	86.37%	85.85%	90.47%
	20%	86.97%	86.25%	87.45%	90.55%
	30%	87.58%	86.25%	89.10%	90.59%
	40%	87.41%	86.25%	89.71%	90.79%
k NN	10%	88.46%	88.86%	75.39%	91.27%
	20%	88.46%	89.06%	76.52%	91.68%
	30%	88.42%	89.43%	81.34%	92.48%
	40%	88.46%	89.91%	84.64%	93.29%

Table 2: Performance evaluation of Self-training, Co-training, Tri-training and EnSSL

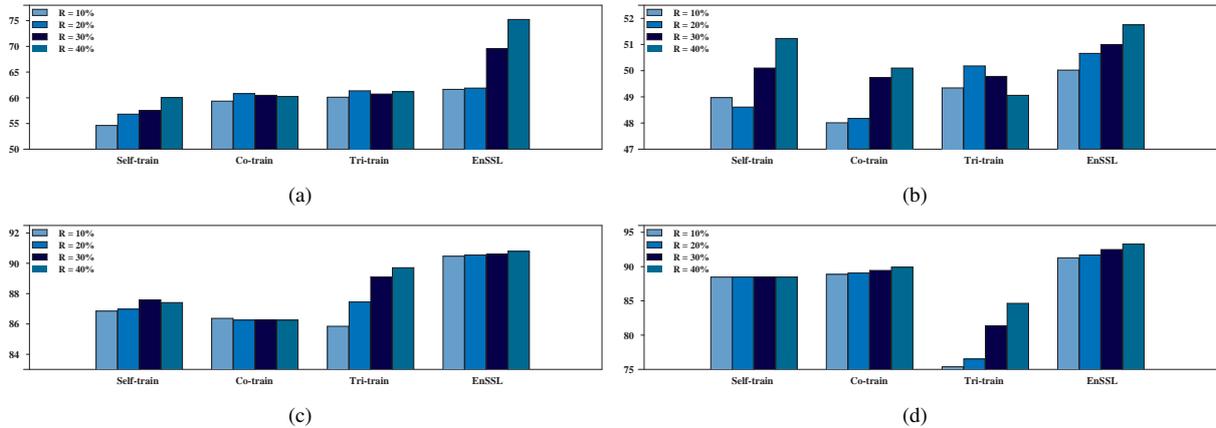


Figure 1: Box plot for each labeled ratio using as base learner (a) NB (b) SMO (c) C4.5 (d) k NN

Subsequently, we evaluated the classification performance of the presented ensemble algorithm EnSSL, against some other state-of-the-art self-labeled algorithms such as Co-Forest, Co-Bagging and Democratic-Co learning. Notice that EnSSL utilizes k NN as base learner which exhibited the best performance.

Table 3 reports the performance of each tested self-labeled algorithm, regarding each labeled ratio. As above mentioned, the accuracy measure of the best performing algorithm is highlighted in bold. Clearly, the presented ensemble self-labeled algorithm illustrates the best performance, independent of the utilized labeled ratio.

Ratio	Co-Forest	Co-Bagging (NB)	Co-Bagging (SMO)	Co-Bagging (C4.5)	Co-Bagging (k NN)	Democratic-Co	EnSSL (k NN)
10%	72.58%	58.71%	46.16%	70.37%	66.34%	55.69%	91.27%
20%	80.94%	51.67%	47.65%	74.59%	70.37%	60.72%	91.68%
30%	83.72%	56.69%	50.30%	82.43%	72.38%	61.60%	92.48%
40%	86.45%	59.75%	51.63%	82.03%	75.39%	61.12%	93.29%

Table 3: Performance evaluation of Co-Forest, Co-Bagging, Democratic-Co learning and EnSSL

4. Conclusions

In this work, we evaluated the performance of an ensemble SSL algorithm, entitled EnSSL, for the classification of blood cell subtypes from images. The proposed ensemble algorithm combines the individual predictions of three of the most efficient and popular self-labeled algorithms, i.e. Self-training, Co-training and Tri-training, utilizing a maximum probability-based voting scheme. Our preliminary numerical experiments indicated the efficacy of the EnSSL, illustrating that reliable and robust classification models could be developed by the adaptation of ensemble methodologies in the semi-supervised learning framework.

Our future work is focused on enhancing the classification efficiency of EnSSL utilizing more efficient and sophisticated self-labeled algorithms using the presented maximum probability-based voting scheme. Furthermore, another interesting aspect is focusing on expanding our experiments and applying further the proposed algorithm to several biomedical datasets for image classification.

References

- [1] D. Aha. *Lazy learning*. Dordrecht: Kluwer academic publishers, 1997.
- [2] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *11th annual conference on computational learning theory*, pages 92–100, 1998.

- [3] D. Di Nardo, G. Gambarini, R. Costantini, L. Testarelli, L. Piasecki, and D. Al-Sudani. 3D clinical evaluation of unusual anatomy of a maxillary second molar: a case report. *Biomedical Journal of Scientific & Technical Research*, 2:1–4, 2018.
- [4] P. Domingos and M. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, 29:103–130, 1997.
- [5] M.F.A. Hady and F. Schwenker. Combining committee-based semi-supervised learning and active learning. *Journal of Computer Science and Technology*, 25(4):681–698, 2010.
- [6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA data mining software: An update. *SIGKDD explorations newsletters*, 11:10–18, 2009.
- [7] N.A.M. Hamed. Persistent eosinophilia is a challenging problem. *Biomedical Journal of Scientific & Technical Research*, 1(3):557–560, 2017.
- [8] R.B. Hegde, K. Prasad, H. Hebbar, and B.M.K. Singh. Development of a robust algorithm for detection of nuclei and classification of white blood cells in peripheral blood smear images. *Journal of medical systems*, 42(6):110, 2018.
- [9] T. Honda and H. Inagawa. Lipopolysaccharide-treated human monocytes regulate gene expressions after interactions with human adipocytes. *Biomedical Journal of Scientific & Technical Research*, 7(4):1–5, 2018.
- [10] S. Kumar, Z. Khan, and A. Jain. A review of content based image classification using machine learning approach. *International Journal of Advanced Computer Research*, 2(3):55, 2012.
- [11] M. Li and Z.H. 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.
- [12] I.E. Livieris. A new ensemble self-labeled semi-supervised algorithm. *Informatica*, (accepted for publication), 2018.
- [13] I.E. Livieris, K. Drakopoulou, V. Tampakas, T. Mikropoulos, and P. Pintelas. An ensemble-based semi-supervised approach for predicting students' performance. In *Research on e-Learning and ICT in education*. Elsevier, 2018.
- [14] I.E. Livieris, K. Drakopoulou, V. Tampakas, T. Mikropoulos, and P. Pintelas. Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research*, 2018.
- [15] I.E. Livieris, A. Kanavos, V. Tampakas, and P. Pintelas. An ensemble SSL algorithm for efficient chest X-ray image classification. *Journal of Imaging*, 4(7), 2018.
- [16] I.E. Livieris, V. Tampakas, N. Kiriakidou, T. Mikropoulos, and P. Pintelas. Forecasting students' performance using an ensemble SSL algorithm. In *8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion*, pages 1–5. ACM, 2018.
- [17] V. Ng and C. 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, 2003.
- [18] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksaç, and S. Beksaç. Feature extraction and classification of blood cells for an automated differential blood count system. In *Proceedings of International Joint Conference on Neural Networks*, volume 4, pages 2461–2466. IEEE, 2001.
- [19] S. Osowski, R. Siroic, T. Markiewicz, and K. Siwek. Application of support vector machine and genetic algorithm for improved blood cell recognition. *IEEE Transactions on Instrumentation and Measurement*, 58(7):2159–2168, 2009.
- [20] A.N. Papadopoulos, M.E. Plissiti, and D.I. Fotiadis. Medical-image processing and analysis for CAD systems. *Medical Image Analysis Methods*, pages 51–86, 2005.
- [21] V. Piuri and F. Scotti. Morphological classification of blood leucocytes by microscope images. *Computational Intelligence for Measurement Systems and Applications*, pages 103–108, 2004.
- [22] J. Platt. Using sparseness and analytic QP to speed training of support vector machines. In M.S. Kearns, S.A. Solla, and D.A. Cohn, editors, *Advances in neural information processing systems*, pages 557–563, 1999.
- [23] J.R. Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann, San Francisco, 1993.
- [24] Juan Manuel Ramirez-Cortes, Pilar Gomez-Gil, Gabriel Sanchez-Perez, and Cesar Prieto-Castron. Shape-based hand recognition approach using the morphological pattern spectrum. *Journal of Electronic Imaging*, 18(1):013012, 2009.
- [25] J. Rawat, A. Singh, H.S. Bhaduria, J. Virmani, and J.S. Devgun. Application of ensemble artificial neural network for the classification of white blood cells using microscopic blood images. *International Journal of Computational Systems Engineering*, 4(2-3):202–216, 2018.
- [26] P. Smitha, L. Shaji, and M.G. Mini. A review of medical image classification techniques. In *International conference on VLSI, Communication & Intrumnatation*, pages 34–38, 2011.
- [27] N. Theera-Umpon and P.D. Gader. System-level training of neural networks for counting white blood cells. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(1):48–53, 2002.
- [28] I. Triguero, S. García, and F. Herrera. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. *Knowledge and information systems*, 42(2):245–284, 2015.
- [29] M. Udvardy. Newly recognised or new blood transfusion transmitted infections. challenges, to be faced by clinicians, transfusion experts, which certainly needs refreshed blood banking skills and clinical approaches. *Biomedical Journal of Scientific & Technical Research*, 5(5):1–4, 2018.
- [30] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A.F.M. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand, and D. Steinberg. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.
- [31] D. 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.
- [32] Y. Zhou and S. Goldman. Democratic co-learning. In *16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 594–602. IEEE, 2004.
- [33] Z.H. Zhou and M. Li. Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on Knowledge and Data Engineering*, 17(11):1529–1541, 2005.
- [34] X. Zhu and A.B. Goldberg. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130, 2009.