

Features based on the percolation theory for quantification of non-Hodgkin lymphomas



Guilherme F. Roberto^{a,*}, Leandro A. Neves^a, Marcelo Z. Nascimento^b, Thaína A.A. Tosta^c, Leonardo C. Longo^a, Alessandro S. Martins^d, Paulo R. Faria^e

^a Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), R. Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, SP, Brazil

^b Faculty of Computation (FACOM) - Federal University of Uberlândia (UFU), Av. João Neves de Ávila 2121, BLB, 38400-902, Uberlândia, MG, Brazil

^c Center of Mathematics, Computing and Cognition, Federal University of ABC (UFABC), Av. dos Estados, 5001, 09210-580, Santo André, SP, Brazil

^d Federal Institute of Triângulo Mineiro (IFTM), R. Belarmino Vilela Junqueira S/N, 38305-200, Ituiutaba, MG, Brazil

^e Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia (UFU), Av. Amazonas, S/N, 38405-320, Uberlândia, MG, Brazil

ARTICLE INFO

Keywords:

Features
Percolation
Multiscale
Multidimensional
Lymphomas

ABSTRACT

Non-Hodgkin lymphomas are a health problem that affects over 70,000 people per year in the United States alone. The early diagnosis and the identification of this lymphoma are essential for an effective treatment. The classification of non-Hodgkin lymphomas is a task that continues to rank as one of the main challenges faced by hematologists, pathologists, as well as in the producing of computer vision methods due to its inherent complexity. In this paper, we present a new method to quantify and classify tissue samples of non-Hodgkin lymphomas based on the percolation theory. The method consists of associating multiscale and multidimensional approaches in order to divide the image into smaller regions and then verifying color similarity between pixels. A cluster labeling algorithm was applied to each region of interest to obtain the values for the number of clusters, occurrence of percolation and coverage ratio of the largest cluster. The method was tested on different classifiers aiming to differentiate three different groups of non-Hodgkin lymphomas. The obtained results (AUC rates between 0.940 and 0.993) were compared to those provided by methods consolidated in the Literature, which indicates that the percolation theory is a suitable approach for identifying these three classes of non-Hodgkin lymphomas, those being: mantle cell lymphoma, follicular lymphoma and chronic lymphocytic leukemia.

1. Introduction

Lymphomas are neoplasias present in cells that are part of the immunological system of the body and their early diagnosis is important for a more effective treatment [1]. Lymphomas are divided into two distinct groups, the Hodgkin lymphomas (HL) and the non-Hodgkin lymphomas (NHL) [2], which are classified according to its growth pattern and cytological features. In the United States, 8260 new cases of HL and 72,240 new cases of NHL are expected for the year 2017 [1]. The American Cancer Society also made an estimation of 20,140 deaths caused by NHL for the year 2017. In Brazil, the National Cancer Institute (INCA) estimated an occurrence of 2470 cases of HL and 10,240 cases of NHL in the year 2016 [3]. The main focus of research studies related to these forms of cancer is on non-Hodgkin Lymphomas, as these show a higher incidence rate.

The NHLs are divided into several subtypes including mantle cell lymphoma (MCL), follicular lymphoma (FL) and chronic lymphocytic leukemia (CLL). MCL represents 4% of all lymphomas in the USA and frequently affects people in their fifties. Morphologically, it is characterized by a diffuse or nodular infiltration of small monotonous lymphoid cells with scant cytoplasm and irregular nuclei (Fig. 1a) [4]. FL comprises of around 50% of adult lymphomas. It is composed of two germinal center-derived B cells: (1) centrocytes, which are characterized by irregular and cleaved nuclear contours and scant cytoplasm, and (2) centroblasts, which exhibit large nuclear, moderate amounts of cytoplasm, and usually more than two nucleoli per cell (Fig. 1b) [5]. CLL is the most common type of leukemia, and the affected patients are usually diagnosed at ages over 60. It is morphology characterized by the proliferation of mature, small lymphocytes depicting irregular nuclei, condensed chromatin, and a narrow border of cytoplasm, but

* Corresponding author. Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), R. Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, SP, Brazil.

E-mail address: guilherme@sjrp.unesp.br (G.F. Roberto).

<https://doi.org/10.1016/j.compbimed.2017.10.012>

Received 27 June 2017; Received in revised form 11 October 2017; Accepted 12 October 2017

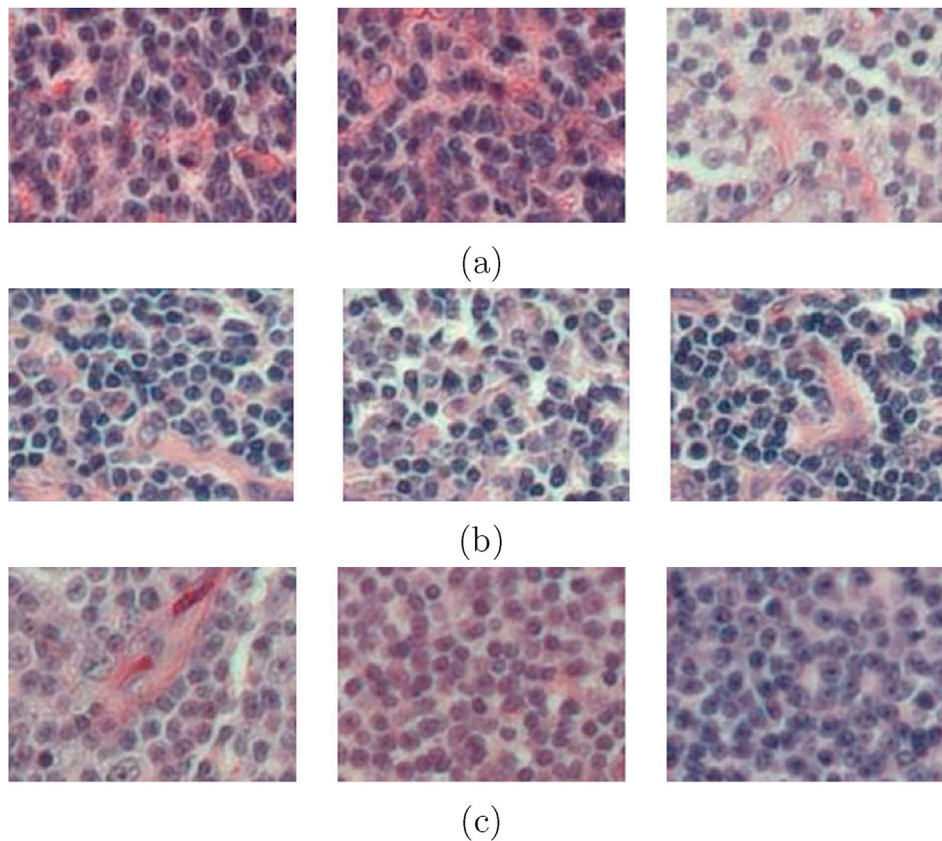


Fig. 1. Examples of different groups of histological NHL images, considering samples stained with hematoxylin-eosin (H&E): (a) MCL, (b) FL and (c) CLL.

inconspicuous nucleoli. Throughout the tumor, it is also possible to find large cells known as prolymphocytes, which show large nuclei, less condensed chromatin, and large nucleoli (Fig. 1c) [6]. Noteworthy here is that tissue samples stained with hematoxylin-eosin (H&E) have been considered by pathologists as one of the main means for the diagnose of NHL. However, in this type of image a single tumor can present different cytological features, hindering the classification task of the existent groups. These characteristics are of a challenging nature for hematologists, pathologists and computer vision methods, especially when attempting to identify and correctly classify each type of NHL [2,7]. Several studies have been published by researchers with the goal of improving the accuracy surrounding the differentiation amongst NHL classes. These studies focused on either the process of obtaining features from the images or the classification process, with the aim of increasing the accuracy of existent methods.

The authors Foran et al. [8] have developed a hybrid system that consists of the integration of a distributed telemicroscopy system and an intelligent image repository to extract spatial features from NHL images. The studies of Tuzel et al. [9] have proposed a method wherein basic texture elements (textons) were learned from the cell nuclei and cytoplasm. These textons were represented through histograms and serve as input for a support vector machine (SVM) for classifying lymphoma types. In Shamir et al. [10] a proposal is made that establishes a benchmark for biological image analysis, which included a NHL dataset. The authors applied the WND-CHARM classifier [11]. This classifier was based on nearest neighbor classification and was also applied by Ref. [2], where different color spaces were tested on NHL images that aimed at verifying which would provide the best differentiation ratio. Amongst the chosen color spaces for testing, the best accuracy was found on that which defined hematoxylin (H) and eosin (E) stains as two different color channels. Transforms such as Fourier and wavelet were applied to these images to generate a feature vector. The proposal put forward in Meng et al. [12] for an approach that consists of dividing the image into 25

sub-images and then extracting color and texture features from each. After applying the chi-squared attribute selection, the size of the feature vector was reduced from 12,625 to 50. The authors in Song et al. [13] proposed a method based on visual descriptors to extract features from grayscale images, including a NHL dataset. The features were extracted separately and each one went on to make up a different set that was given as input to a specific classification stage. In Codella et al. [14] 12 metrics were applied to segmented NHL images. The images were previously enhanced through different color and spatial configurations, generating five new images for each NHL image of the tested dataset. Different methods were used in these studies to quantify and extract features from the histological images, which considered the strategies commonly explored in the medical image context [15–18]. Moreover, these methods have provided interesting classification rates with an expressive number of features.

Although there is not a single and universal texture descriptor capable of performing the quantification of any kind of image, some researches have indicated that fractal techniques can provide better results for histological image quantification when compared to the previously mentioned [19–22], mainly due to the presence of stochastic properties and self-similarities. Amongst the main fractal features present in the literature, fractal dimension (FD), lacunarity (LAC) and percolation-based features stand out.

The features of FD are used as a measure for evaluating the irregularity and the complexity of a fractal. On the other hand, the features of LAC are used as a measure for evaluating how the space is filled. The FD feature has been applied successfully in studies of prostatic cancer [23], relationship between aging and decreasing of the vascular complexity of the retina [24], analysis of periapical lesions [25], quantification of cellular rejection in myocardial biopsy of patients submitted to heart transplantation [26] and analysis of the brain white matter, age and sex [27].

The LAC feature has also provided relevant results, although it is not

as widespread as analyses considering FD. For instance, the LAC feature has been focused toward quantifying and classifying epileptic seizures [28], psoriatic lesions [29], prostate cancer [30] amongst other applications.

The performance of both FD and LAC can be enhanced by the application of multiscale and multidimensional approaches [31,32]. The application of the multiscale approach aims at dividing a complex problem into a simpler one. For an image, this can be adapted as the division of the image into smaller regions, allowing for each one to be analyzed individually. The multidimensional approach consists of mapping the color properties of the pixels in the image into n -dimensional vectors, aiming at obtaining numerical features from these properties. Both of these measures can be used separately or simultaneously to obtain information from an image. However, in some cases, different images can present very similar or even the same FD and LAC values, which may be a problem to classifiers when it comes to distinguishing between them.

Even though FD and LAC have not been applied to NHL histological images, features related to percolation theory may complement such evaluations, mainly when providing a greater distinguishing capacity of vascular structures on histological images [33]. Despite the quantification power of fractal techniques being enhanced with multiscale and multidimensional approaches [31,32], there are no methods in the Literature that focus on the percolation theory with the associations explored herein, as well as the differentiation amongst classes of NHL based on fractal measures. The percolation theory can support the image quantification and classification process in diagnostic support systems, with information concerning possible cluster features present on images. Moreover, even those studies, although few in number, have applied percolation as a quantifier, the results were interesting for tumor angiogenesis detection in vascular images [34]. In myocardial images of heart transplanted, percolation was also able to improve the understanding of the rejection process in each studied group [35]. However, in these works, multidimensional observations for the principles of percolation in colored images have not been considered. The development of an approach based on the percolation theory, multidimensional and multiscale with the goal of developing an extension for colored images is a relevant contribution to the area, and which holds the possibility of being tested on different groups of histological NHL images.

In this context, we propose a new method based on the percolation theory for quantifying and classifying colored images from three types of NHL. The method consists of associating the percolation theory with multiscale and multidimensional observations to provide different features: average number of clusters, which are determined by the similarity amongst pixels in the same region, quantifying a single structure in the image; occurrence of percolation, which consists of verifying whether a cluster spreads from one extremity of the region to the opposite one; and coverage ratio of the largest cluster, which has the goal of analyzing the difference between the main cluster of the region and the remaining clusters. The method was applied to histological NHL images, considering different classification methods. The distinctions amongst the groups of interest were evaluated using the area under the ROC curve (AUC) values.

Based on the approach presented in this paper, the main contributions are:

1. A new method that associates the percolation theory to multiscale and multidimensional approaches, a strategy that has not yet been explored in the literature;
2. Application of the proposed method on images with acknowledged scientific relevance, providing significant results that may support hematologists and pathologists for the classification of the three types of NHL;
3. Quantification and indication of the best association among techniques in order to distinguish and understand the behavior of MCL, FL and CLL classes.

2. Methodology

The proposed method is organized as follows: in [Section 2.1](#) the multiscale approach is defined in order to divide the image into smaller regions (with sizes in function of a scale L) and analyze each one; in [Section 2.2](#) the association of the multidimensional approach is described, which aims at evaluating the color similarity among pixels within the same region; in [Section 2.3](#) the procedure to associate the percolation theory is presented, with the results provided by multiscale and multidimensional approaches. The cluster labeling in a region and feature values are also presented in this step. In [Section 2.4](#) the employed image database and the quantitative evaluation are presented in order to apply and validate the proposed method. A summarized version of the proposed method is illustrated in [Fig. 2](#).

2.1. Multiscale approach

This step is defined to guarantee a multiscale analysis: the procedure is performed considering the gliding-box algorithm [36], in which a square box of side L , positioned on the upper left-hand corner of the image is given as input. The box goes through the whole image pixel by pixel and the value of L is increased after reaching the end. Initially, L is defined as 3 to represent the minimum scale: the increase always occurs in 2 units. Defining L as an odd number will guarantee that the box will have a central pixel to apply to the multidimensional analysis. This procedure is illustrated in [Fig. 3](#). In this work, the maximum chosen value of L was defined as 65 for all tests. This range provides sufficient measurements for the next steps. For each iteration the total number of boxes T on the image is given by L , width (W) and height (H) of the image. This relation is given in [Eq. \(1\)](#):

$$T = (H - L + 1) \times (W - L + 1). \quad (1)$$

2.2. Multidimensional approach

The second step of the proposed method consisted of defining the multidimensional analysis based on the approach described by Refs. [31,32]. For each square box of side L , the most relevant color channel according to the RGB color space was chosen from a comparison with the central pixel P_c of the box under analysis: the values of the color channels (r , g and b) and the coordinates (x and y) of the central pixel were assigned to a vector $F_c = f_c(x_c, y_c, r_c, g_c, b_c)$. Each pixel of the box was named as P and selected for comparison with the central pixel: the color channel values of P were assigned to the vector $F = f(x, y, r, g, b)$. Therefore, the vectors F and F_c were compared applying the Minkowski distance d , as shown in [Eq. \(2\)](#). An important adaptation has been considered in relation to the method defined by Refs. [31,32]. When the Minkowski distance d is smaller or equal to L , P receives the value -1 , which means the pixel represents a pore in that specific box. Otherwise, P receives the value 0, which can be understood as background. This procedure allowed for the application of different evaluations based on the percolation theory. The pixel selection process is illustrated in [Fig. 4](#). For instance, a box of size $L = 3$ with pixels valued after the application of this approach is shown in [Fig. 5](#).

$$|F - F_c| = \max |f(i) - f_c(i_c)| \leq L, \forall i = \overline{1, 5} \quad (2)$$

2.3. Features based on the percolation theory

The third step of our approach consisted of associating the percolation theory with the results provided by multiscale and multidimensional strategies, in order to enhance the potential of the features based on spatial percolation. The proposed model is based on the statement that, on a lattice, there is a path between two points α and β when there is at least one sequence of adjacent pixels (pores) between these points. In this case, α and β are connected. The percolation is defined when a set of

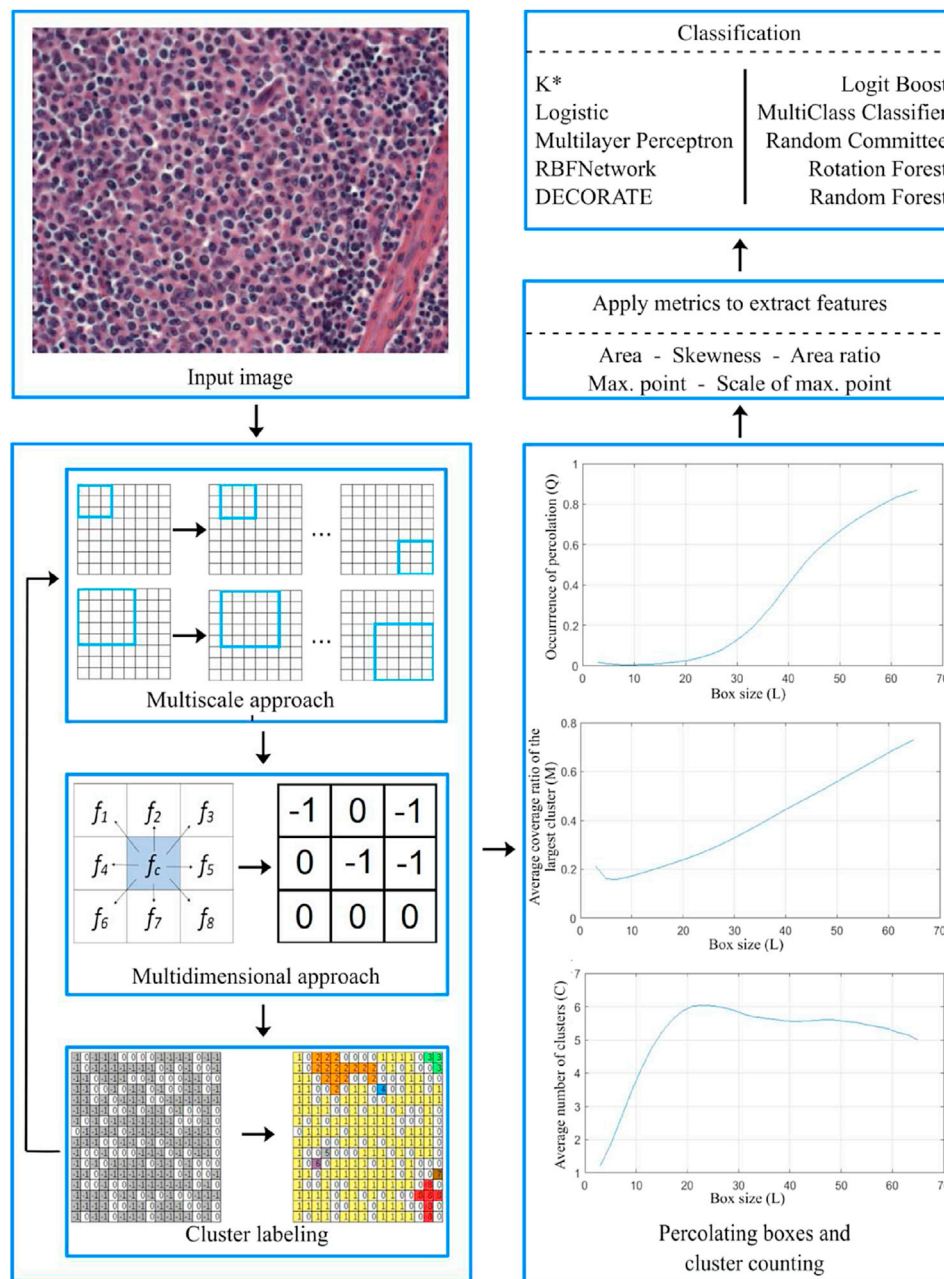


Fig. 2. A summary of the proposed method to quantify and classify the NHL images.

connected points (cluster) extends from one extremity of the system to the other [37], by considering it as a squared lattice. Thus, the height and width of the matrix, as well as the probability p were applied to define a pore whereby a supposed fluid may flow through it. The remaining spaces $(1 - p)$ correspond to solids where the fluid is not able to flow.

It is important to note that the organization of the system and consequently of the clusters can be significantly changed in relation to different values of p . For instance, the presence of a percolating cluster is guaranteed when the value of p is greater than the percolation threshold [38]: research studies have defined the percolation threshold as $p = 0.59275$ [39,40]. This value was considered in order to avoid an additional step to guarantee the verification of the existence of percolation when the value of p is greater or equal to this threshold. For instance, in Fig. 6 three lattices are presented with different values of p . The gray pixels represent the pores, where the supposed fluid may flow through. Percolation occurs in lattice (c), for these have a neighboring point set (in gray) that extends through two extremities of the system. In the proposed

model each pixel represents one space of the lattice.

Percolating clusters were obtained considering the Hoshen-Kopelman cluster labeling algorithm [41], which was chosen due to its simple adaptability for images. In the Hoshen-Kopelman approach, each element P of the matrix was defined as 0 or -1 : in terms of percolation, if $P = -1$, the element represents a pore (gray pixels) in Fig. 6, as defined previously. After labeling each pore, a cluster was defined if the same label was associated to a set of pixels. When the labeling for the first cluster is finished, the algorithm advances to the next element valued as -1 (corresponding to a cluster that has not yet been labeled). The Hoshen-Kopelman algorithm is presented in Algorithm 1. A matrix before and after the labeling process is illustrated in Fig. 7.

From the relationships presented previously, the cluster average C , the percolating boxes ratio Q and the average coverage ratio of the largest cluster M are functions proposed herein to represent the association within the percolation theory with multiscale and multidimensional approaches. The number of clusters in a single box is given by c_i . As the

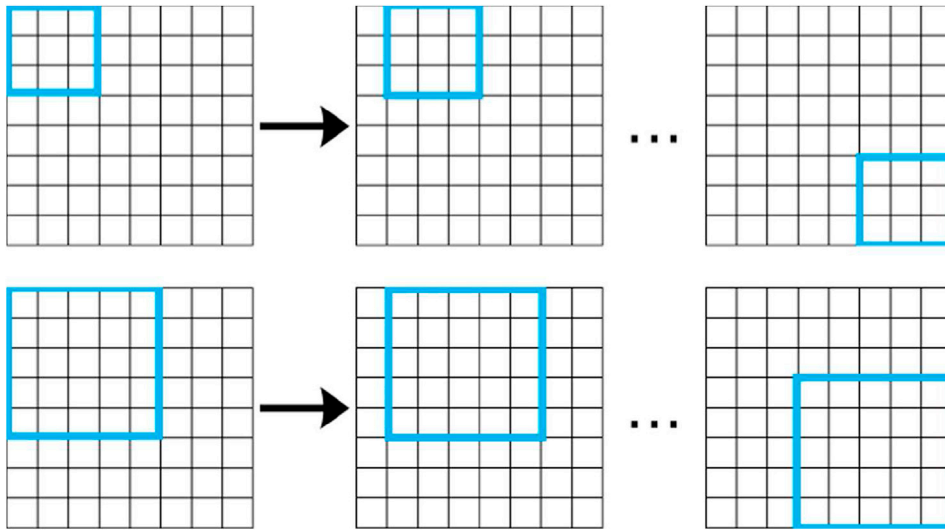


Fig. 3. Illustration of the procedure in which the box goes through the whole image pixel by pixel, considering square boxes with different sizes of L (3 and 5).

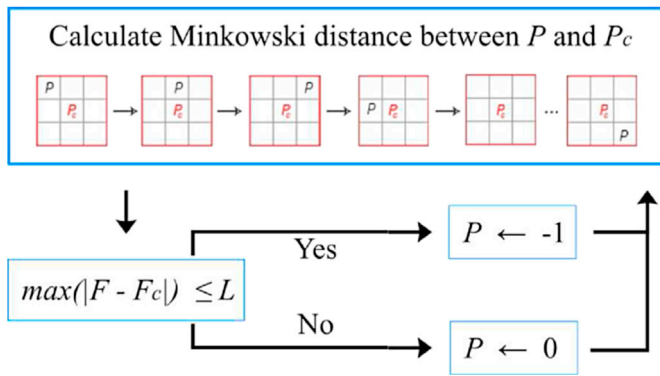


Fig. 4. Illustration of the pixel selection process, considering comparisons between the vectors $F_c = f(x_c, y_c, r_c, g_c, b_c)$ and $F = f(x, y, r, g, b)$ by the Minkowski distance.

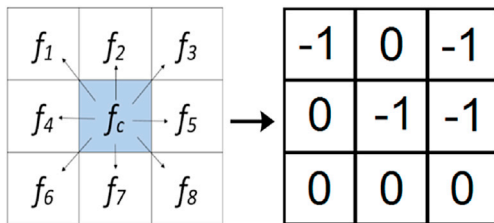


Fig. 5. Examples of a box of size $L = 3$ with pixels valued after the application of the multidimensional approach, considering the conditions: pore is defined with the value -1 ; background is defined with the value 0 .

images have different dimensions, a normalization was necessary: the sum of the number of clusters in a scale L was divided by the number of boxes T (see Eq. (1)), thus obtaining the average number of clusters per box $C(L)$ as given by Eq. (3).

$$C(L) = \frac{\sum_{i=1}^T c_i}{T} \quad (3)$$

Algorithm 1. Hoshen-Kopelman algorithm applied to define the labeling of pores and clusters [41]

```

1: Label ← 0
2: for i = 1 to i = height do
3:   for j = 1 to j = width do
4:     Label ← Label + 1
5:     LABEL_CLUSTER(i, j)
6:   end for
7: end for
8: LABEL_CLUSTER(i, j){
9:   site(i, j) ← Label
10:  if i < height and site(i + 1, j) = -1 then
11:    LABEL_CLUSTER(i + 1, j)
12:  else if j < width and site(i, j + 1) = -1 then
13:    LABEL_CLUSTER(i, j + 1)
14:  else if i > 1 and site(i - 1, j) = -1 then
15:    LABEL_CLUSTER(i - 1, j)
16:  else if j > 1 and site(i, j - 1) = -1 then
17:    LABEL_CLUSTER(i, j - 1)
18:  end if
19: }

```

In a similar manner, the percolating boxes ratio Q was obtained by counting the number of boxes that have percolated in a scale L . A box q_i was considered as percolating if the ratio between the number of pixels labeled as pores (Ω_i) and the total amount of pixels in the box (L^2) exceeded the theoretical percolation threshold p [39], as shown in Eq. (4). Thus, $Q(L)$ was obtained with the division of the total number of percolating boxes q_i by the total amount of boxes T (Eq. (1)) in a scale L , as given in Eq. (5): q_i is defined as 1, if the box has percolated, or 0, otherwise.

$$q_i = \begin{cases} 1, & \frac{\Omega_i}{L^2} \geq 0.59275 \\ 0, & \frac{\Omega_i}{L^2} < 0.59275 \end{cases} \quad (4)$$

$$Q(L) = \frac{\sum_{i=1}^T q_i}{T} \quad (5)$$

The average coverage ratio of the largest cluster M was obtained by identifying the coverage ratio of the largest cluster in each box evaluated on the L scale. Thus, the average coverage ratio was calculated by dividing the sum of the coverage in each box by the total amount of boxes T (Eq. (1)). The coverage of a box i was given by the division of its largest

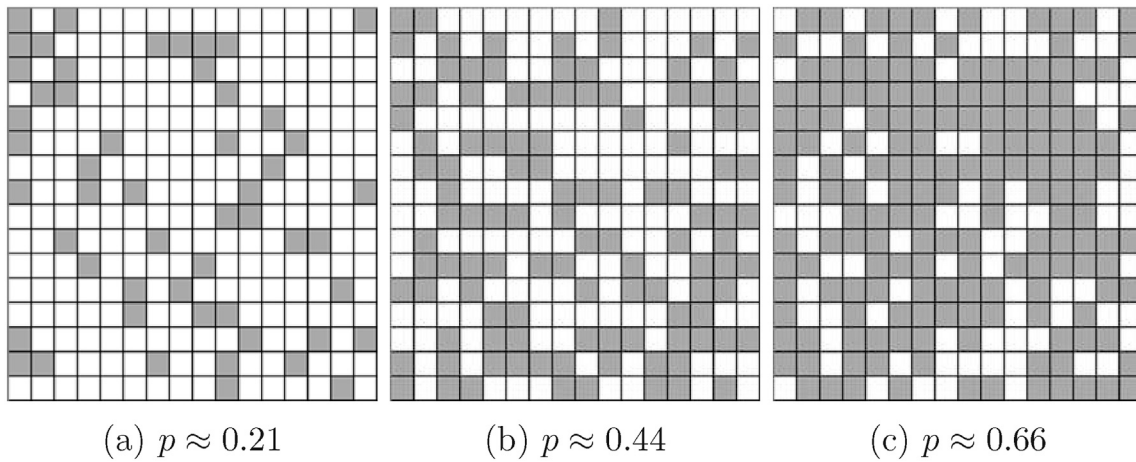


Fig. 6. Three lattices with different p values: percolation occurs in lattice (c).

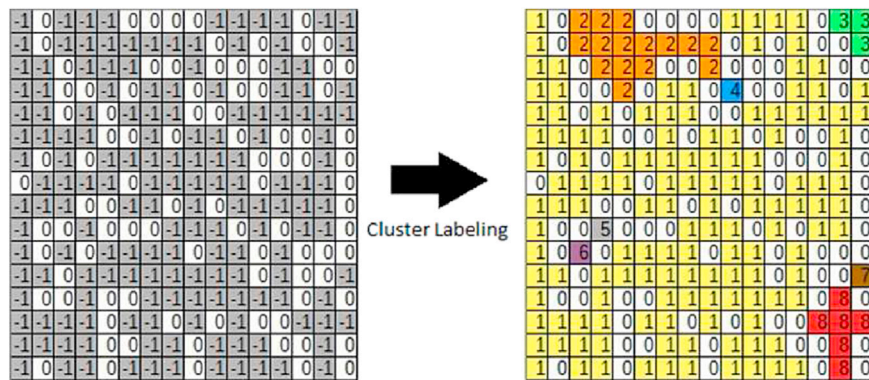


Fig. 7. Example of a matrix before (left) and after (right) the application of the Hoshen-Kopelman algorithm [41].

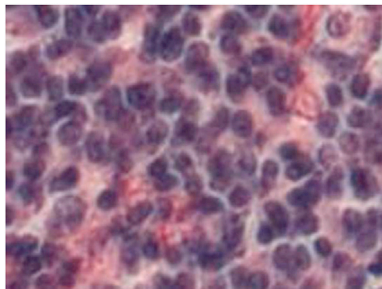


Fig. 8. Example of a histological lymphoma image.

cluster (m_i) by the number of pixels in the box (L^2), as shown in Eq. (6).

$$M(L) = \frac{\sum_{i=1}^T m_i}{T} \quad (6)$$

In this context a histological lymphoma image is given as an example in Fig. 8 and the functions $C(L)$, $Q(L)$ and $M(L)$ were calculated from the variation of the scale L , which are shown in Figs. 9–11. The behavior of each function can be represented by a scalar value and enable the composition of a feature vector.

We applied metrics to obtain scalar values from each function according to that presented by Ref. [42]. The metrics used were area under curve, skewness, area ratio, maximum point and the scale of the maximum point. Thus, considering a function f for mapping $C(L)$, $Q(L)$ or $M(L)$, the area A that is under f and between two points (a and b) over the x axis was calculated considering Eq. (7). As the analyzed functions are

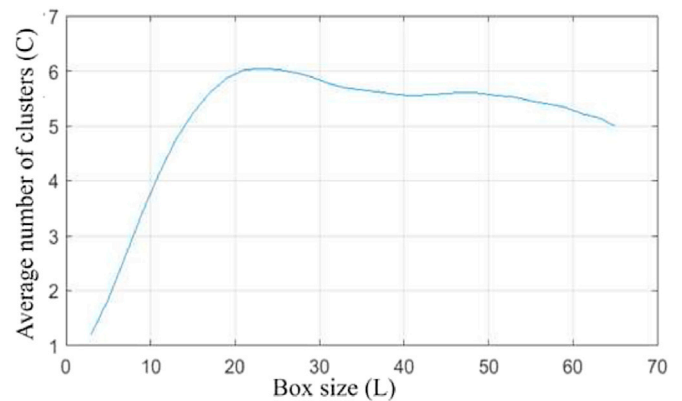


Fig. 9. Average number of cluster for the image shown in Fig. 8.

discrete, the A was determined by the numerical integration and by applying the trapezoidal method. The integration with points equally spaced was performed according to Eq. (8), wherein the space between each point corresponds to its scalar value and N is the number of samples. In the functions $C(L)$, $Q(L)$ and $M(L)$ the variables a and b represent the minimum and maximum values of L , respectively. In this case, the values 3 and 65 were assigned to a and b , respectively.

$$A = \int_a^b f(x) dx \quad (7)$$

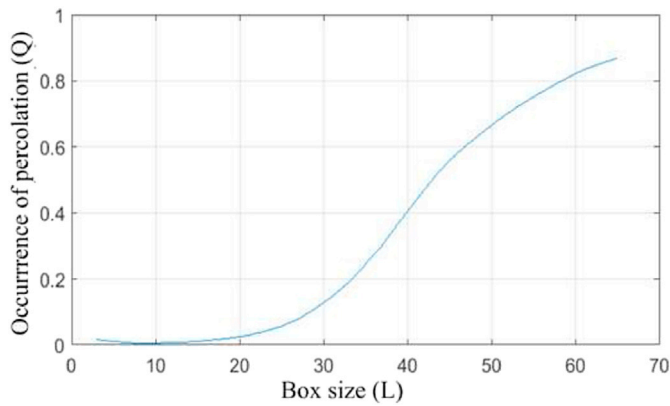


Fig. 10. Occurrence of percolation in function of the multiscale and multidimensional approaches for the image shown in Fig. 8.

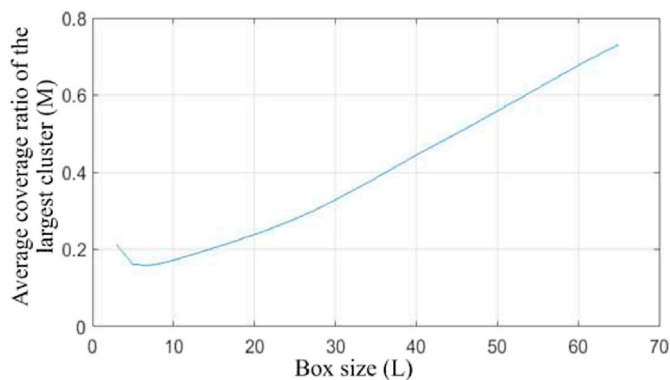


Fig. 11. Average coverage of the largest cluster in function of the multiscale and multidimensional approaches for the image shown in Fig. 8.

$$\int_a^b f(x) dx \approx \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(x_n) + f(x_{n+1})) \quad (8)$$

Skewness is an asymmetry measure of a sample in relation to its average: for negative skewness, the sample is concentrated to the left of the average value; for positive skewness, the sample is concentrated to the right of the average value. On a perfectly symmetrical sample, the skewness is valued at 0. For a sample with N values, the skewness O is given by Eq. (9), in which \bar{x} is the sample average and x_i indicates the value of the sample referent to the i^{th} value of x .

$$O = \frac{\frac{1}{N} \sum_{i=a}^b (x_i - \bar{x})^3}{\sqrt[3]{\left[\frac{1}{N-1} \sum_{i=a}^b (x_i - \bar{x})^2 \right]^3}} \quad (9)$$

The area ratio R was defined considering the right-side and left-side areas under the function curve. Its value increases as the texture gets heterogeneity in larger scales [42]. As $A_{a,b}$ corresponds to the area under the curve formed by a function between the points a and b over the x axis, as previously defined on Eq. (7), the area ratio is given by Eq. (10).

$$R = \frac{A_{(\frac{b}{2}+1,b)}}{A_{(a,\frac{b}{2})}} \quad (10)$$

The maximum point of each function was also considered to define two other measures, the observation scale and the value of the function on the point given as reference. Thus, the application of these metrics on the functions $C(L)$, $Q(L)$ and $M(L)$ allowed for the definition of a vector composed of 15 features, as illustrated on Table 1.

2.4. Image dataset and evaluation measures

The method was tested on NHL images according to the studies performed by researchers of the National Cancer Institute and from the National Institute on Aging, both in the United States [43,44]. The image database considered for the tests has 173 histological NHL images (99 representing the MCL group, 62 representing the FL group and 12 representing the CLL group). The images were digitally acquired, through a light microscope (Zeiss AxioScope) with a 20× objective and a colored digital camera (AXio Cam MR5) attached. The regions of interest of each plate were selected by specialists, digitally photographed and recorded without compression, with the RGB color model, a spatial resolution of 1388×1040 pixels and a 24 bit quantization ratio.

The performance of the method was evaluated by applying different classifiers and considering the cross validation strategy to define the generalization capacity of the proposed method. The feature vectors were divided into 10 disjoint groups with 90% of the data for training and the 10% for validation. This strategy was repeated 10 times [45]. In the application of the 10-fold cross-validation, the groups of data used in the training and testing procedure were randomly defined. The overall performance of the algorithms was evaluated by means of examining the ROC area index over the testing output values. The tests were performed with the following classifiers: K^* [46], Logistic [47], Multilayer Perceptron [48], RBFNetwork [49], DECORATE [50], Logit Boost [51], Multi-Class Classifier [52], Random Committee [53], Rotation Forest [54] and Random Forest [55].

The classifiers were selected to represent different classifying models in order to test the proposed method under a variety of circumstances. For instance, the K^* classifier belongs to the Lazy Classifying Model. This model stores all training samples and builds a general model when a new sample needs to be classified [56]. These classifiers are usually faster at training, but slower at predicating time. Another type of classifier available are the function-based classifiers [47–49]. These classifiers are based on artificial neural networks, these perform calculations and estimates on the input vectors through several layers of weighted functions, with the aim of approximating non-linear functions in order to classify instances. In this paper, the tested function-based classifiers are Logistic, Multilayer Perceptron and RBFNetwork. Another tested classifying model was that based on trees, which consists of mapping the features values to classes in order to classify both trained and untrained instances. In this mapping the leaf nodes are labeled with a class and the other nodes are linked to subtrees that are based on feature values of an instance [57]. In this paper, this model was represented by the Random Forest algorithm. Classifying models defined as meta-classifiers were also applied to complete our tests. Meta-classifiers are designated for finding the most suitable classifier for each set and then select the most fitting

Table 1

Features defined by the application of our method, considering the association of the percolation theory with the multiscale and multidimensional approaches.

Metric	Function
Area	$C(L)$
Skewness	
Area Ratio	
Maximum Point	
Scale of the Maximum Point	
Area	$Q(L)$
Skewness	
Area Ratio	
Maximum Point	
Scale of the Maximum Point	
Area	$M(L)$
Skewness	
Area Ratio	
Maximum Point	
Scale of the Maximum Point	

results before proceeding to the next iteration [50], as these are based on a combination series of different classifiers applied on ensembles of instances, usually in an iterative manner. In our proposal this model was represented by DECORATE, LogitBoost, Random Committee, MultiClass Classifier and Rotation Forest.

3. Results

The discriminative capacity of the proposed method and the best combination of techniques were obtained by analyzing segmented and unsegmented images. Different segmentation strategies were applied to histological NHL images, in order to verify the capacity of the method under different circumstances. The images were segmented by specialists - golden standard (Fig. 12) as reported by Ref. [58] and by methods developed from the combination of different techniques. The chosen methods those as defined by Oliveira et al. [59], Vahadane and Sethi [60], Wienert et al. [61] and Sertel et al. [62]. Examples of segmented images with each one of these methods are illustrated in Figs. 13–16, respectively.

The segmented and unsegmented images were given as input to the proposed method and the results obtained from the possible combinations amongst NHL groups and classifiers are shown on Tables 2–5. The combinations that provided the best AUC values are highlighted in bold.

From these results, one notes that the best AUC rates for the comparison, MCL versus FL versus CLL, did not present great difference between the unsegmented images and the segmented images by the specialist) (Table 2). The AUC rates were of approximately 0.940. This value was surpassed by the two-class comparison, MCL versus FL (Table 3) and MCL versus CLL (Table 4). Considering only the unsegmented images, the best result was obtained by the comparison of MCL versus FL (Table 3) and with the Rotation Forest classifier, obtaining the AUC value of 0.965. When these two groups are compared considering the segmented images, the best result was obtained by using the Random Forest classifier and segmentation performed by the specialist, which obtained the AUC of 0.950. The comparisons between FL versus CLL (Table 5) were those that provided the least relevant AUC rates, even so the rate was higher than 0.920, considering the segmented images by the specialist and the Logit Boost classifier. On the other hand, the proposed method presented the best performance for the comparisons considering MCL versus CLL (Table 4). The highest AUC rates (0.993) were achieved with images segmented by the specialist and applying the Logistic and MultiClass classifiers.

4. Discussion

The performance obtained by the method was compared to that obtained through the FD and LAC features [31,32], both multiscale and multidimensional. The results were provided by applying the classifiers described in Subsection 2.4. An overview of these results is presented in Fig. 17. Each column of the graph represents the average AUC value among the four tested class comparisons for classifications of segmented and unsegmented images. Besides providing better results than FD and

LAC in all tested cases, the best results occurred on images segmented by the specialist: the average AUC rates were 0.950. In this case, the proposed method provided the best performance in relation to FD: the AUC rate was approximately 17% superior.

Through an analysis of the statistic test procedure of [63], we applied the Friedman test to the results provided by the proposed method. The Friedman test is a non-parametric static method that ranks k algorithms for n given datasets in a way that the best performing algorithm acquires rank 1 and the k^{th} performing algorithm acquires rank k . If different algorithms have the same performance on a single dataset, average ranks are assigned [64]. To compare the relevance of the three tested quantification methods, as well as the classifiers performance, we applied the test to rank combinations among the methods and classifiers. Therefore, the methods (FD, LAC and the proposed approach) along with the ten classifiers were ranked by the Friedman test, resulting in a ranking with 30 combinations. We used 24 datasets as input, which consisted of a combination of the six different types of segmentation with the four different comparisons tested for each case. The average ranking is shown on Table 6. One can conclude from these results that, except for the RBFNetwork classifier, our method will provide better classification results, as the first nine positions in the ranking consisted of a combination made up of any of the other classifiers combined with our proposed image quantification method based on percolation.

These relevant results can be associated to the following statements: (1) in the context of histological images, researches have shown that vascular and histological structures may present a spatial behavior similar to percolation models [33,65–67]. Thus, percolation-based features may have more capacity for the processing of such characteristics; (2) the number of features provided by the proposed method (15 features) allows for a more complete evaluation of the features present on the images, wherein the low number of features on the LAC vector (4 features) and the FD vector (1 feature) may be a restricting factor for the classification. Thus, the obtained features may support the understanding of the presence, organization and extension of clusters. Moreover, the quantification presented here are much more complementary than ratable, representing an important advance for studies considering NHL. We believe that our method provided the desired robustness for similar applications with valuable information contained in each test and comparison.

Another verified case consisted of applying a method for identifying the most relevant features for the study of NHL classes. The chosen method was the support vector machine attribute evaluation (SVM AE) [68]. The test was applied to the feature vectors obtained from the images segmented by the specialist, as these presented the best results (Fig. 17). The AUC rates of the classifications performed with the most relevant features are shown on Table 7.

A noteworthy point on Table 7 is that most of the AUC rates have not been changed in function of the reduction of the number of features. As an example, the distinction between the classes MCL versus CLL was performed with only three features, with effective contributions from metrics applied on the functions $C(L)$, $Q(L)$ and $M(L)$. In this case, the AUC rate value of 0.993 has not changed. Similar behavior is observed on

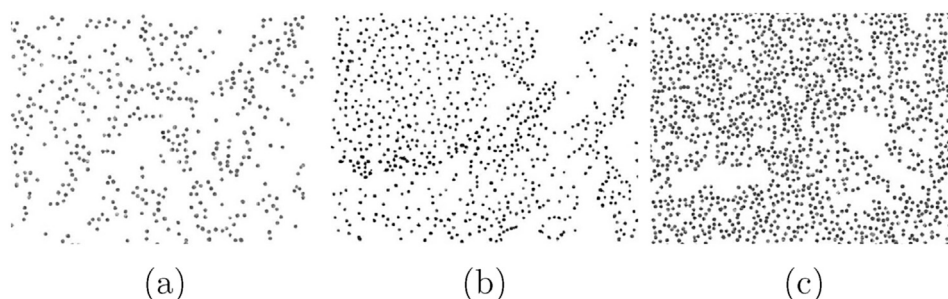


Fig. 12. NHL images segmented by the specialists [58]: (a) MCL class, (b) FL class and (c) CLL class.

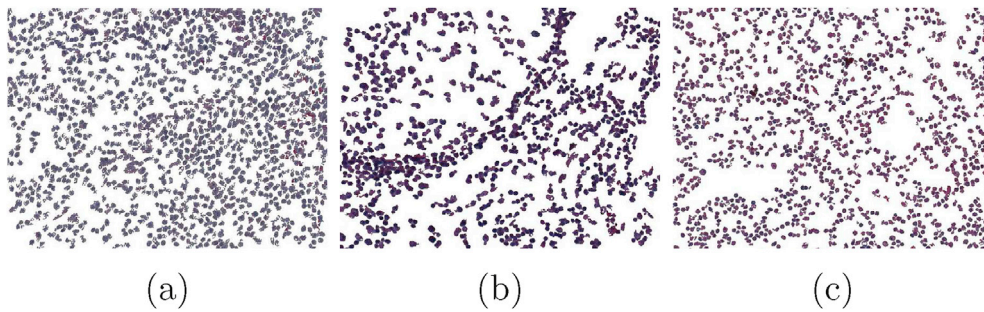


Fig. 13. NHL images segmented by using the Oliveira et al. algorithm [59]: (a) MCL class, (b) FL class and (c) CLL class.

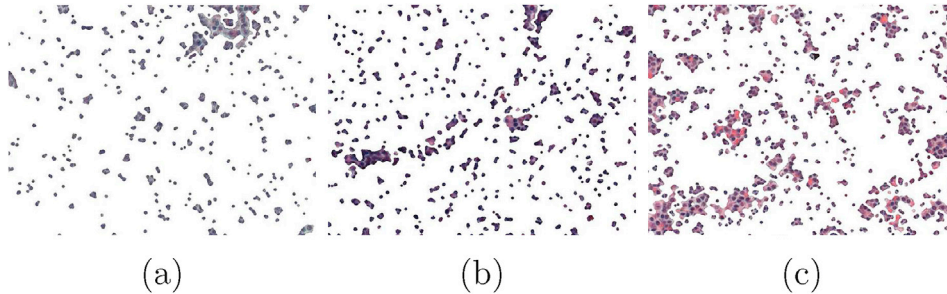


Fig. 14. NHL images segmented by using the Vahadane and Sethi algorithm [60]: (a) MCL class, (b) FL class and (c) CLL class.

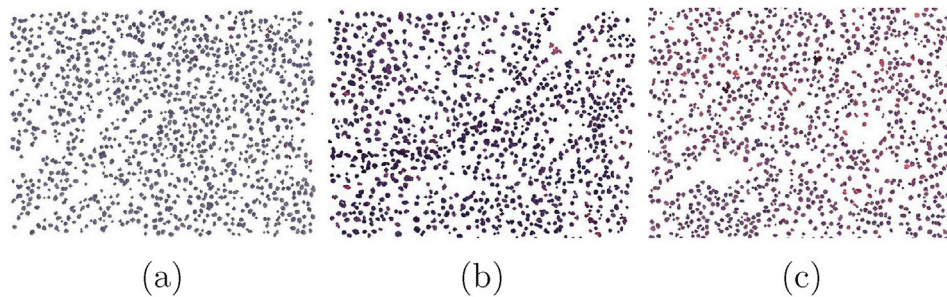


Fig. 15. NHL images segmented by using the Wienert et al. algorithm [61]: (a) MCL class, (b) FL class and (c) CLL class.

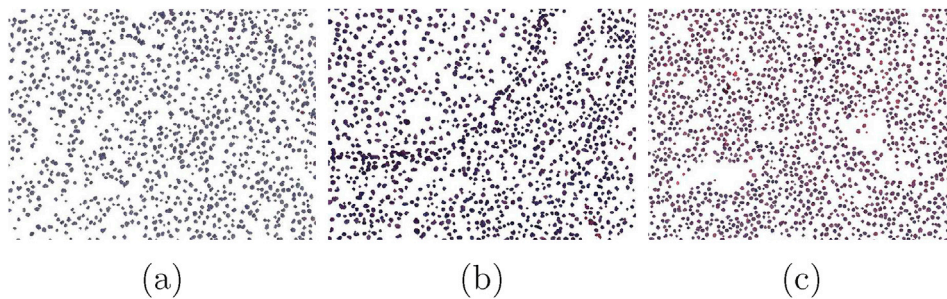


Fig. 16. NHL images segmented by using the Sertel et al. algorithm [62]: (a) MCL class, (b) FL class and (c) CLL class.

the comparison MCL versus FL versus CLL. One notes the need of more features to keep an AUC value of 0.943. In this case, the selected features are composed of two metrics from each of the functions $C(L)$, $Q(L)$ and $M(L)$. The comparisons MCL versus FL and FL versus CLL are shown to be where the need of a larger set of features is necessary, which may indicate a greater difficulty in terms of the distinction amongst the FL class against the others. Another remark is that the functions $Q(L)$ and $M(L)$ were those that provided more contributions with features for the classification: 19 features out of 25. These features are related to the dimension of

the clusters ($M(L)$) and the occurrence of percolation ($Q(L)$), which may indicate that these histological characteristics are important for the quantification of the classes MCL, FL and CLL.

We also analyzed different performance measures other than the AUC for the results generated by the classification of NHL images segmented by the specialist using the features selected by SVMFAE. These measures are sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). To illustrate the performance of these other measures, results obtained from 2-class comparisons are shown

Table 2

AUC rates obtained by comparisons among MCL versus FL versus CLL groups, considering unsegmented and segmented images of NHL lymphoma.

	Unsegmented	Oliveira et al. [59]	Vahadane and Sethi [60]	Wienert et al. [61]	Sertel et al. [62]	Specialist
K*	0.940	0.844	0.715	0.745	0.754	0.916
Logistic	0.903	0.751	0.735	0.712	0.745	0.899
Multilayer Perceptron	0.930	0.812	0.786	0.746	0.796	0.922
RBFNetwork	0.853	0.770	0.723	0.607	0.648	0.774
DECORATE	0.934	0.864	0.846	0.805	0.908	0.944
Logit Boost	0.885	0.796	0.770	0.760	0.830	0.906
MultiClass Classifier	0.913	0.756	0.755	0.713	0.763	0.924
Random Committee	0.913	0.810	0.831	0.788	0.882	0.928
Rotation Forest	0.925	0.855	0.777	0.803	0.855	0.923
Random Forest	0.935	0.853	0.842	0.826	0.875	0.939

The combinations that provided the best AUC values are highlighted in bold.

Table 3

AUC rates calculated by comparisons performed with MCL versus FL, considering unsegmented and segmented images of NHL lymphoma.

	Unsegmented	Oliveira et al. [59]	Vahadane and Sethi [60]	Wienert et al. [61]	Sertel et al. [62]	Specialist
K*	0.951	0.849	0.717	0.748	0.769	0.922
Logistic	0.946	0.787	0.750	0.672	0.742	0.905
Multilayer Perceptron	0.946	0.809	0.766	0.702	0.783	0.896
RBFNetwork	0.846	0.797	0.729	0.667	0.739	0.810
DECORATE	0.955	0.889	0.808	0.811	0.892	0.948
Logit Boost	0.923	0.768	0.741	0.739	0.847	0.895
MultiClass Classifier	0.946	0.787	0.750	0.672	0.742	0.905
Random Committee	0.948	0.824	0.797	0.789	0.834	0.932
Rotation Forest	0.965	0.886	0.775	0.787	0.861	0.935
Random Forest	0.948	0.855	0.832	0.801	0.894	0.950

The combinations that provided the best AUC values are highlighted in bold.

Table 4

AUC rates obtained by comparisons of MCL versus CLL with on unsegmented and segmented images of NHL lymphoma.

	Unsegmented	Oliveira et al. [59]	Vahadane and Sethi [60]	Wienert et al. [61]	Sertel et al. [62]	Specialist
K*	0.882	0.881	0.833	0.886	0.818	0.978
Logistic	0.727	0.718	0.795	0.915	0.928	0.993
Multilayer Perceptron	0.705	0.849	0.795	0.901	0.902	0.987
RBFNetwork	0.792	0.809	0.642	0.669	0.795	0.899
DECORATE	0.853	0.796	0.937	0.876	0.978	0.965
Logit Boost	0.910	0.856	0.848	0.863	0.953	0.987
MultiClass Classifier	0.727	0.718	0.795	0.915	0.928	0.993
Random Committee	0.857	0.801	0.893	0.902	0.918	0.949
Rotation Forest	0.934	0.902	0.803	0.902	0.856	0.991
Random Forest	0.876	0.890	0.907	0.896	0.907	0.985

The combinations that provided the best AUC values are highlighted in bold.

Table 5

Comparisons of FL versus CLL on unsegmented and segmented images of NHL lymphoma.

	Unsegmented	Oliveira et al. [59]	Vahadane and Sethi [60]	Wienert et al. [61]	Sertel et al. [62]	Specialist
K*	0.876	0.792	0.596	0.845	0.636	0.832
Logistic	0.791	0.612	0.818	0.882	0.813	0.890
Multilayer Perceptron	0.813	0.774	0.837	0.789	0.844	0.726
RBFNetwork	0.610	0.761	0.623	0.681	0.871	0.858
DECORATE	0.747	0.663	0.887	0.855	0.819	0.894
Logit Boost	0.838	0.783	0.842	0.782	0.872	0.928
MultiClass Classifier	0.791	0.612	0.818	0.882	0.813	0.890
Random Committee	0.684	0.686	0.818	0.776	0.819	0.905
Rotation Forest	0.888	0.722	0.854	0.853	0.817	0.909
Random Forest	0.757	0.780	0.872	0.848	0.868	0.888

The combinations that provided the best AUC values are highlighted in bold.

on [Table 8](#).

The evaluation of these metrics on a 2-class problem requires one class to be considered as the positive case and the other as the negative case. In the results presented on [Table 8](#), the class on the left was considered as positive and the class on the right was considered as negative.

The sensitivity corresponds to the test capacity of identifying the true positive values. The higher the sensitivity value of a test, the higher will be the chance of subjects that present the positive case being correctly identified. In the same way, the specificity corresponds to the test capacity of identifying true negative values. A specific test will rarely identify negative cases as positive. One notes that the highest values

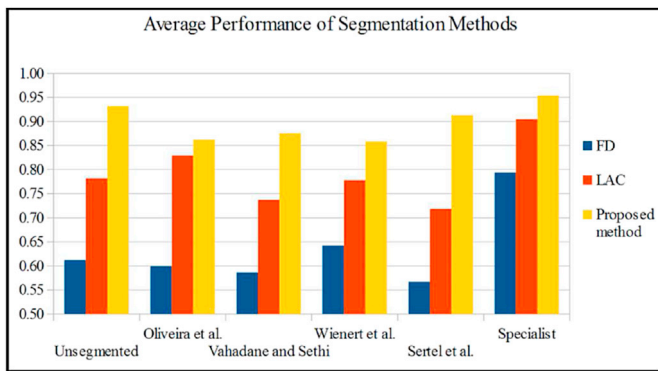


Fig. 17. Average AUC rates considering FD, LAC and the proposed method for each segmentation method.

Table 6
Average rankings of the classifications considering the Friedman test.

Method	Classifier	Ranking
Proposed method	Random Forest	3.8
	Rotation Forest	4.2
	DECORATE	4.5
	Random Committee	6.0
	Logit Boost	6.3
	K*	7.7
	Multilayer Perceptron	8.0
	MultiClass Classifier	8.5
	Logistic	9.0
LAC	K*	10.9
	Rotation Forest	13.5
Proposed method	RBFNetwork	13.7
LAC	Multilayer Perceptron	14.2
	Random Forest	14.2
	MultiClass Classifier	14.6
	Logistic	14.7
	RBFNetwork	14.8
	Logit Boost	15.1
	DECORATE	17.2
	Random Committee	17.3
FD	K*	22.8
	Multilayer Perceptron	23.5
	RBFNetwork	23.5
	Logit Boost	23.7
	Logistic	23.9
	MultiClass Classifier	24.0
	DECORATE	24.5
	Random Forest	25.0
Rotation Forest	27.9	
Random Committee	27.9	

obtained by sensitivity and specificity were of 0.980 and 0.923. These values correspond to the specificity of the comparisons MCL versus CLL and FL versus CLL. As in both comparisons the CLL class was considered as the negative case, we can conclude that the CLL class presents the best classification ratio when the proposed method is applied. The second best performance was presented by the MCL class, wherein both comparisons (MCL versus FL and MCL versus CLL), were considered as the positive case. The sensitivity values in these comparisons were of 0.884 and 0.833 respectively. The less relevant performance occurred in the FL class, when compared to the CLL class. The sensitivity value obtained in this comparison was 0.778, considering the FL class as the positive case. This may indicate a greater difficulty in identifying images from the FL class using percolation-based features. This fact is reinforced by the results presented on Table 7, as the comparisons with the FL class were those that required more attributes to obtain relevant classification ratios.

PPV and NPV correspond to the proportion of cases classified by a test

Table 7
AUC rates of the classifications performed with the most relevant features and segmented images by the specialists.

	Classifier	AUC	Features	Function
MCL versus FL versus CLL	DECORATE	0.943	Area Ratio	M(L)
			Skewness	M(L)
			Max. Point	C(L)
			Skewness	Q(L)
			Scale of Max. Point	Q(L)
MCL versus FL	Random Forest	0.952	Area	C(L)
			Skewness	M(L)
			Scale of Max. Point	Q(L)
			Scale of Max. Point	M(L)
			Skewness	Q(L)
			Max. Point	M(L)
			Area Ratio	Q(L)
			Area	M(L)
			Area Ratio	M(L)
			Area Ratio	M(L)
MCL versus CLL	Logistic	0.993	Area Ratio	M(L)
			Max. Point	C(L)
			Skewness	Q(L)
FL versus CLL	Logit Boost	0.940	Area	C(L)
			Max. Point	C(L)
			Scale of Max. Point	Q(L)
			Area Ratio	Q(L)
			Area Ratio	C(L)
			Skewness	M(L)
			Scale of Max. Point	M(L)
			Max. Point	Q(L)

Table 8
Results obtained from 2-class comparisons: sensitivity, specificity, positive predictive value and negative predictive value.

	Sensitivity	Specificity	PPV	NPV
MCL versus FL	0.884	0.862	0.920	0.807
MCL versus CLL	0.833	0.980	0.833	0.980
FL versus CLL	0.778	0.923	0.583	0.968

in relation to the real total of cases on a given dataset. Unlike the sensitivity and specificity, PPV and NPV may change according to the number of cases on each class. It is shown on Table 8 that the best performance occurred once again on the CLL class (0.980 in the comparison MCL versus CLL and 0.968 in the comparison FL versus CLL), as this presented the best NPV when considered as the negative case. In the comparisons wherein the MCL class was considered as the positive case, the obtained PPV was also relevant: 0.920 when compared to the FL class and 0.833 when compared to the CLL class. Once again, the FL class provided the least relevant performance in both tested comparisons (0.807 when compared to the MCL class and 0.583 when compared to the CLL class). This also reinforced the remark that this class is the most difficult to identify when the proposed method is applied to NHL images segmented by the specialist.

Finally, we present an illustrative overview to show the good quality of our method when analyzed from the point of view of other approaches focused on the theme. This illustrative overview is propitious to forming an understanding of the potential of the features developed from percolation theory. The results are shown on Table 9, considering as a reference the accuracy rates reported by the studies. One notes here that the proposed method provided an average accuracy of approximately 92.00% for two-class comparisons and features obtained from images segmented by the specialist. These results are relevant and compatible with that observed in the literature, for example those described by Refs. [8–10]. Despite our method presenting an accuracy rate smaller than that obtained by Refs. [2,12–14], the differences were not significant when analyzed by the Friedman test ($P_f = 0.05$), with all pairwise comparisons

Table 9
Illustrative overview of accuracy rates from different approaches for NHL images.

Method	Features	Colored	Segmented	Accuracy	AUC
Foran et al. [8]	4 (3)	Yes	Yes	89.00%	–
Tuzel et al. [9]	75	No	Yes	89.00%	–
Shamir et al. [10]	1025	Yes	No	85.00%	–
Meng et al. [12]	12,625 (50)	Yes	No	92.70%	–
Orlov et al. [2]	1025 (12–200)	Yes	No	99.00%	–
Song et al. [13]	9872	No	No	96.80%	–
Codella et al. [14]	216	Yes	Yes	95.50%	–
Proposed Method	15 (3–8)	Yes	Yes	87.58%–96.40% (average \approx 92.00%)	average \approx 0.967

(Conover), and the Kruskal-Wallis test ($P_k = 0.9746$), considering all pairwise comparisons (Dwass-Steel-Christchlow-Fligner). In these tests the significance level was of 0.05. Moreover, these accuracy values should not be used for comparison as to defining the best one, as this would in itself be a fruitless task. For instance, in our study, the dataset was composed of an unequal number of images for each NHL class, which may influence the accuracy values. This fact is also verified through the referenced studies. A more precise measure is found in the AUC value, according to that described by Refs. [69–71]. However, the AUC values were not provided by the studies considered for this illustrative overview, hindering a more suitable analysis. Even so, an important contribution of our method is to provide robust results with a reduced set of features, besides allowing for the identification of the most relevant features for the study of NHL histological images.

5. Conclusion

In this paper we proposed a new method based on the percolation theory to quantify colored NHL images. The percolation-based techniques available in the Literature were limited to binary or grayscale images. We showed that percolation is well suited for application as a complementary measure for other fractal-based characterization methods to improve the differentiation among different classes of NHL. The results were important with significant differentiation rates: the AUC values were over 0.94. The indication of the best association of segmentation method, classifier and features to distinguish amongst NHL classes was achieved testing different combinations. On average, the best classification results were obtained by classifying images segmented by the specialist. We then applied an attribute selection method (SVMFE) to define the most relevant features. For the comparison between the three NHL classes, an AUC value of 0.943 was obtained with the classifier DECORATE and considering only six features, which comprise of two metrics from each function $C(L)$, $Q(L)$ and $M(L)$. The best AUC value was provided by the classifier Logistic with a set considering only three features: the AUC rate was of 0.993 for the comparison considering MCL versus CLL. We also verified that the occurrence of percolation and the cluster dimension are the histological features with the highest relevance to differentiation amongst NHL. When our results were compared to those provided by the fractal features with highest exploration in the literature, such as FD and LAC [31,32], respectively, the method indicated a better performance in nine out of ten tested classifiers.

These contributions may encourage new research studies to apply the

proposed method as a viable alternative in order to complement the quantification performed with FD and LAC. Moreover, the initial proposal of creating a method able to distinguish among the three groups of NHL was achieved, for the method obtained important results when applied to the correspondent image database, obtaining precision values in the classification superior than those of other methods [2,8,9] with less features in the generated feature vector.

In future works we intend to apply the proposed method on many different kinds of dataset other than clinical images, evaluating their results with other techniques, besides FD and LAC. Some studies can also be performed considering the influence of noise in the quantification process in order to avoid features with overestimated values. The analysis of other color spaces apart from RGB and the calculation of different distances for pixel comparison other than the Minkowski distance can be performed in order to verify the influence of these parameters when combined with the percolation theory. Moreover, we intend to verify our method on different types of medical images in order to assess its generality when applied to different problems and define its performance quality within each context.

Conflict of interest

None.

Acknowledgment

The authors thank to CAPES (33004153073P9) and FAPEMIG (TEC - APQ-02885-15 project) for financial support.

References

- [1] A.C. Society, Cancer Statistics Center, <https://cancerstatisticscenter.cancer.org>, Accessed: 2017-02-09.
- [2] N.V. Orlov, W.W. Chen, D.M. Eckley, T.J. Macura, L. Shamir, E.S. Jaffe, I.G. Goldberg, Automatic classification of lymphoma images with transform-based global features, *IEEE Trans. Inf. Technol. Biomed.* 14 (4) (2010) 1003–1013.
- [3] I. I. N. do Câncer, Estimativa — 2016-Incidência de câncer no Brasil, Ministério da Saúde, 2016.
- [4] J.M. Vose, Mantle cell lymphoma: 2012 update on diagnosis, risk-stratification, and clinical management, *Am. J. Hematol.* 87 (6) (2012) 604–609.
- [5] L. Xerri, S. Dirnhofer, L. Quintanilla-Martinez, B. Sander, J.K. Chan, E. Campo, S.H. Swerdlow, G. Ott, The heterogeneity of follicular lymphomas: from early development to transformation, *Virchows Arch.* 468 (2) (2016) 127–139.
- [6] M. Hallek, Chronic lymphocytic leukemia: 2013 update on diagnosis, risk stratification and treatment, *Am. J. Hematol.* 88 (9) (2013) 803–816.
- [7] F.P. de Souza Santos, G. dos Santos Fernandes, Linfomas não-Hodgkin, http://www.medicinanet.com.br/conteudos/reviseos/99/linfomas_nao_hodgkin.htm, Accessed: 2016-05-04.
- [8] D.J. Foran, D. Comaniciu, P. Meer, L.A. Goodell, Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemedicine, *IEEE Trans. Inf. Technol. Biomed.* 4 (2000) 265–273.
- [9] O. Tuzel, L. Yang, P. Meer, D.J. Foran, Classification of hematologic malignancies using texton signatures, *Pattern Analysis Appl.* 10 (4) (2007) 277–290.
- [10] L. Shamir, N. Orlov, D.M. Eckley, T.J. Macura, I.G. Goldberg, IICBU 2008: a proposed benchmark suite for biological image analysis, *Med. Biol. Eng. Comput.* 46 (9) (2008) 943–947.
- [11] N. Orlov, L. Shamir, T. Macura, J. Johnston, D.M. Eckley, I.G. Goldberg, WND-CHARM: multi-purpose image classification using compound image transforms, *Pattern Recognit. Lett.* 29 (11) (2008) 1684–1693.
- [12] T. Meng, L. Lin, M.-L. Shyu, S.-C. Chen, Histology image classification using supervised classification and multimodal fusion, in: *Multimedia (ISM), 2010 IEEE International Symposium on, IEEE, 2010*, pp. 145–152.
- [13] Y. Song, W. Cai, H. Huang, D. Feng, Y. Wang, M. Chen, Bioimage classification by subcategory discriminant transform of high dimensional visual descriptors, *BMC Bioinforma.* 17 (1) (2016) 465.
- [14] N. Codella, M. Moradi, M. Matasar, T. Syeda-Mahmood, J.R. Smith, Lymphoma diagnosis in histopathology using a multi-stage visual learning approach, in: *SPIE Medical Imaging, International Society for Optics and Photonics, 2016*, 97910H–97910H.
- [15] Y. Zhang, Z. Dong, L. Wu, S. Wang, A hybrid method for MRI brain image classification, *Expert Syst. Appl.* 38 (8) (2011) 10049–10053.
- [16] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Fast wavelet-based image characterization for highly adaptive image retrieval, *IEEE Trans. Image Process.* 21 (4) (2012) 1613–1623.

- [17] D. Marín, A. Aquino, M.E. Gegúndez-Arias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features, *IEEE Trans. Med. Imag.* 30 (1) (2011) 146–158.
- [18] M.M. Matsumoto, C.M. Sehgal, J.K. Udupa, Local binary pattern texture-based classification of solid masses in ultrasound breast images, in: *SPIE Medical Imaging, International Society for Optics and Photonics*, 2012, 83201H–83201H.
- [19] P.P. Ohanian, R.C. Dubes, Performance evaluation for four classes of textural features, *Pattern Recognit.* 25 (8) (1992) 819–833.
- [20] R. Lopes, N. Betrouni, Fractal and multifractal analysis: a review, *Med. Image Anal.* 13 (4) (2009) 634–649.
- [21] R. Lopes, P. Dubois, I. Bhoui, M. Bedoui, S. Maouche, N. Betrouni, Local fractal and multifractal features for volumic texture characterization, *Pattern Recognit.* 44 (8) (2011) 1690–1697.
- [22] L. Li, L. Chang, S. Ke, D. Huang, Multifractal analysis and lacunarity analysis: a promising method for the automated assessment of muskmelon (*Cucumis melo* L.) epidermis netting, *Comput. Electron. Agric.* 88 (2012) 72–84.
- [23] E. Yu, J.P. Monaco, J. Tomaszewski, N. Shih, M. Feldman, A. Madabhushi, Detection of prostate cancer on histopathology using color fractals and Probabilistic Pairwise Markov models, in: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE, 2011*, pp. 3427–3430.
- [24] M.Z.C. Azemin, D.K. Kumar, T.Y. Wong, J.J. Wang, P. Mitchell, R. Kawasaki, H. Wu, Age-related rarefaction in the fractal dimension of retinal vessel, *Neurobiol. Aging* 33 (1) (2012) 194–e1.
- [25] E. Soğur, B.G. Baksö, H.-G. Gröndahl, B.H. Şen, Pixel intensity and fractal dimension of periapical lesions visually indiscernible in radiographs, *J. Endod.* 39 (1) (2013) 16–19.
- [26] L. Neves, F. Oliveira, F. Peres, R. Moreira, A. Moriel, M. De Godoy, L.M. Junior, Maximum entropy, fractal dimension and lacunarity in quantification of cellular rejection in myocardial biopsy of patients submitted to heart transplantation, in: *Journal of Physics: Conference Series*, vol. 285, IOP Publishing, 2011, 012032.
- [27] S. Farahibozorg, S. Hashemi-Golpayegani, J. Ashburner, Age- and sex-related variations in the brain white matter fractal dimension throughout adulthood: an MRI study, *Clin. Neuroradiol.* 25 (1) (2015) 19–32.
- [28] W. Zhou, Y. Liu, Q. Yuan, X. Li, Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG, *IEEE Trans. Biomed. Eng.* 60 (12) (2013) 3375–3381.
- [29] M. Ivanovici, N. Richard, H. Decean, Fractal dimension and lacunarity of psoriatic lesions—A colour approach, *Medicine* 6 (4) (2009) 7.
- [30] L.A. Neves, M. Nascimento, D. Oliveira, A. Martins, M. Godoy, P. Arruda, D. de Santi Neto, J.M. Machado, Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer, *Expert Syst. Appl.* 41 (11) (2014) 5017–5029.
- [31] M. Ivanovici, N. Richard, Fractal dimension of color fractal images, *IEEE Trans. Image Process.* 20 (1) (2011) 227–235.
- [32] M. Ivanovici, N. Richard, The lacunarity of colour fractal images, in: *Image Processing (ICIP), 2009 16th IEEE International Conference on, IEEE, 2009*, pp. 453–456.
- [33] J.W. Baish, R.K. Jain, Fractals and cancer, *Cancer Res.* 60 (14) (2000) 3683–3688.
- [34] L. Ichim, R. Dobrescu, Characterization of tumor angiogenesis using fractal measures, in: *Control Systems and Computer Science (CSCS), 2013 19th International Conference on, IEEE, 2013*, pp. 345–349.
- [35] V. Carvalho, L. Neves, M. De Godoy, R. Moreira, A. Moriel, L.M. Junior, Aprendizado de Máquina Simbólico e Técnicas Fractais Para Caracterizar Rejeição em Biópsia Miocárdica, in: *V Latin American Congress on Biomedical Engineering CLAIB 2011 May 16–21, 2011, Springer, Habana, Cuba, 2013*, p. 272.
- [36] R.F. Voss, Random fractals: characterization and measurement, in: *Scaling Phenomena in Disordered Systems*, Springer, 1991, pp. 1–11.
- [37] J.W. Essam, Percolation theory, *Rep. Prog. Phys.* 43 (7) (1980) 833.
- [38] P. Dean, A new Monte Carlo method for percolation problems on a lattice, in: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 59, Cambridge Univ Press, 1963, pp. 397–410.
- [39] N. Bird, E. Perrier, Multiscale percolation properties of a fractal pore network, *Geoderma* 160 (1) (2010) 105–110.
- [40] R.M. Ziff, Test of scaling exponents for percolation-cluster perimeters, *Phys. Rev. Lett.* 56 (6) (1986) 545.
- [41] J. Hoshen, R. Kopelman, Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm, *Phys. Rev. B* 14 (8) (1976) 3438.
- [42] A. Cálman, M. Ivanovici, Psoriasis image analysis using color lacunarity, in: *Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on, IEEE, 2012*, pp. 1401–1406.
- [43] N.C. Institute, <http://www.cancer.gov/>, Accessed: 2016-05-13.
- [44] N.I. on Aging, <https://www.nia.nih.gov/>, Accessed: 2016-05-13.
- [45] F. Moayedi, Z. Azimifar, R. Boostani, S. Katebi, Contourlet-based mammography mass classification using the SVM family, *Comput. Biol. Med.* 40 (4) (2010) 373–383.
- [46] J.G. Cleary, L.E. Trigg, et al., K*: an instance-based learner using an entropic distance measure, in: *Proceedings of the 12th International Conference on Machine Learning*, vol. 5, 1995, pp. 108–114.
- [47] S. le Cessie, J. van Houwelingen, Ridge estimators in logistic regression, *Appl. Stat.* 41 (1) (1992) 191–201.
- [48] M.W. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences, *Atmos. Environ.* 32 (14) (1998) 2627–2636.
- [49] M.J. Orr, et al., *Introduction to Radial Basis Function Networks*, 1996.
- [50] P. Melville, R.J. Mooney, Constructing diverse classifier ensembles using artificial training examples, in: *IJCAI*, vol. 3, 2003, pp. 505–510. CiteSeer.
- [51] J. Friedman, T. Hastie, R. Tibshirani, Additive Logistic Regression: a Statistical View of Boosting, *Tech. Rep.*, Stanford University, 1998.
- [52] E. Frank, L. Trigg, R. Kirkby, MultiClass Classifier, <http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/MultiClassClassifier.html>, Accessed: 2016-01-19.
- [53] E. Frank, Random Committee, <http://weka.sourceforge.net/doc.dev/weka/classifiers/meta/RandomCommittee.html>, Accessed: 2016-01-19.
- [54] J.J. Rodriguez, L.I. Kuncheva, C.J. Alonso, Rotation forest: a new classifier ensemble method, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (10) (2006) 1619–1630.
- [55] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [56] W. Perrizo, Q. Ding, A. Denton, Lazy classifiers using P-trees, in: *CAINE, 2002*, pp. 176–179.
- [57] J.R. Quinlan, Learning decision tree classifiers, *ACM Comput. Surv. (CSUR)* 28 (1) (1996) 71–72.
- [58] T.A.A. Tosta, P.R. Faria, L.A. Neves, M.Z. do Nascimento, Computational method for unsupervised segmentation of lymphoma histological images based on fuzzy 3-partition entropy and genetic algorithm, *Expert Syst. Appl.* 81 (2017) 223–243.
- [59] D.L.L. De Oliveira, M.Z. Do Nascimento, L.A. Neves, M.F. De Godoy, P.F.F. De Arruda, D. de Santi Neto, Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy, *Expert Syst. Appl.* 40 (18) (2013) 7331–7340.
- [60] A. Vahadane, A. Sethi, Towards generalized nuclear segmentation in histological images, in: *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on, IEEE, 2013*, pp. 1–4.
- [61] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert, F. Klauschen, Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach, *Sci. Rep.* 2 (2012) 503.
- [62] O. Sertel, U.V. Catalyurek, G. Lozanski, A. Shanaah, M.N. Gurcan, An image analysis approach for detecting malignant cells in digitized h&e-stained histology images of follicular lymphoma, in: *Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010*, pp. 273–276.
- [63] S. Garcia, F. Herrera, An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons, *J. Mach. Learn. Res.* 9 (Dec) (2008) 2677–2694.
- [64] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (Jan) (2006) 1–30.
- [65] Y. Gazit, D.A. Berk, M. Leunig, L.T. Baxter, R.K. Jain, Scale-invariant behavior and vascular network formation in normal and tumor tissue, *Phys. Rev. Lett.* 75 (12) (1995) 2428.
- [66] J.W. Baish, Y. Gazit, D.A. Berk, M. Nozue, L.T. Baxter, R.K. Jain, Role of tumor vascular architecture in nutrient and drug delivery: an invasion percolation-based network model, *Microvasc. Res.* 51 (3) (1996) 327–346.
- [67] O. Craciunescu, S. Das, S. Clegg, Dynamic contrast-enhanced MRI and fractal characteristics of percolation clusters in two-dimensional tumor blood perfusion, *J. Biomech. Eng.* 121 (5) (1999) 480–486.
- [68] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, *Mach. Learn.* 46 (2002) 389–422.
- [69] L. Nanni, C. Fantozzi, N. Lazzarini, Coupling different methods for overcoming the class imbalance problem, *Neurocomputing* 158 (2015) 48–61.
- [70] C.L. Chen, A. Mahjoubfar, L.-C. Tai, I.K. Blaby, A. Huang, K.R. Niazi, B. Jalali, Deep learning in label-free cell classification, *Sci. Rep.* 6.
- [71] A. Lumini, L. Nanni, Overview of the combination of biometric matchers, *Inf. Fusion* 33 (2017) 71–85.