Classification of Histological Images Based on the Stationary Wavelet Transform

M Z Nascimento[1], L Neves[2], S C Duarte[3], Y A S Duarte[3], V Ramos Batista[3]

[1] *Faculty of Computer Science (FACOM), Federal University of Uberlândia (UFU), Uberlândia, MG, Brazil*
[2] *Institute of Biosciences, Literature and Exact Sciences, Department of Computer Science and Statistics, São Paulo State University (UNESP), São José do Rio Preto, SP, Brazil*
[3] *Mathematics, Computer Science and Cognition Centre, Federal University of ABC (UFABC), Santo André, SP, Brazil*

E-mail: nascimento@facom.ufu.br

Abstract. Non-Hodgkin lymphomas are of many distinct types, and different classification systems make it difficult to diagnose them correctly. Many of these systems classify lymphomas only based on what they look like under a microscope. In 2008 the World Health Organisation (WHO) introduced the most recent system, which also considers the chromosome features of the lymphoma cells and the presence of certain proteins on their surface. The WHO system is the one that we apply in this work. Herewith we present an automatic method to classify histological images of three types of non-Hodgkin lymphoma. Our method is based on the Stationary Wavelet Transform (SWT), and it consists of three steps: 1) extracting sub-bands from the histological image through SWT, 2) applying Analysis of Variance (ANOVA) to clean noise and select the most relevant information, 3) classifying it by the Support Vector Machine (SVM) algorithm. The kernel types Linear, RBF and Polynomial were evaluated with our method applied to 210 images of lymphoma from the National Institute on Aging. We concluded that the following combination led to the most relevant results: detail sub-band, ANOVA and SVM with Linear and RBF kernels.

1. Introduction

Lymphoma is the term used for cancer when it happens in the lymph system, and it is classified as Hodgkin and non-Hodgkin (NHL). They are further subdivided into over 60 different subtypes [1]. This neoplasm can happen at any age. According to [1] in Canada lymphoma is the commonest cancer in people under 30, and it happens to one out of every 10 children diagnosed with cancer.

The most recent numbers of 2014 in the US are 731,277 people diagnosed with lymphoma, of which 558,340 are NHL [2]. In fact, the number of NHL-cases has been increasing. As for any kind of cancer, an early diagnosis can lead to a successful treatment and also a better quality of living for the patient.

Diagnoses and prognoses of lymphoma are obtained through the analysis of microscopic histological images carried out by a specialist. Despite the good knowledge of experienced pathologists, their diagnoses can be biased by several factors like subjectiveness, limitations of the human vision, and fatigue after several working hours. In order to reduce bias, a good alternative is the help of Computer Aided Diagnosis (CAD). These systems have been showing expressive success at classifying histological images correctly [3–5]. However, to the best of our knowledge in the specific case of lymphoma there is a lack of techniques devoted to extracting characteristics that help classify its subtypes.

In our work we present a method to extract characteristics of NHL and to classify three of its subtypes, namely Follicular Lymphoma (FL), Mantle Cell Lymphoma (MCL) and Chronic Lymphoid Leukemia (CLL). Our method is based on the Stationary Wavelet Transform (SWT) [6], on the Analysis of Variance (ANOVA) and on a Support Vector Machine (SVM) algorithm.

2. Materials and Methods

Three steps constitute the method: 1) extraction of sub-bands of the image with the SWT; 2) application of ANOVA to clean noise and select relevant information; 3) classification of lymphoma subtypes through SVM. Three different kernel types (Linear, RBF and Polynomial) were applied with our method. Its diagram is depicted in Figure 1.

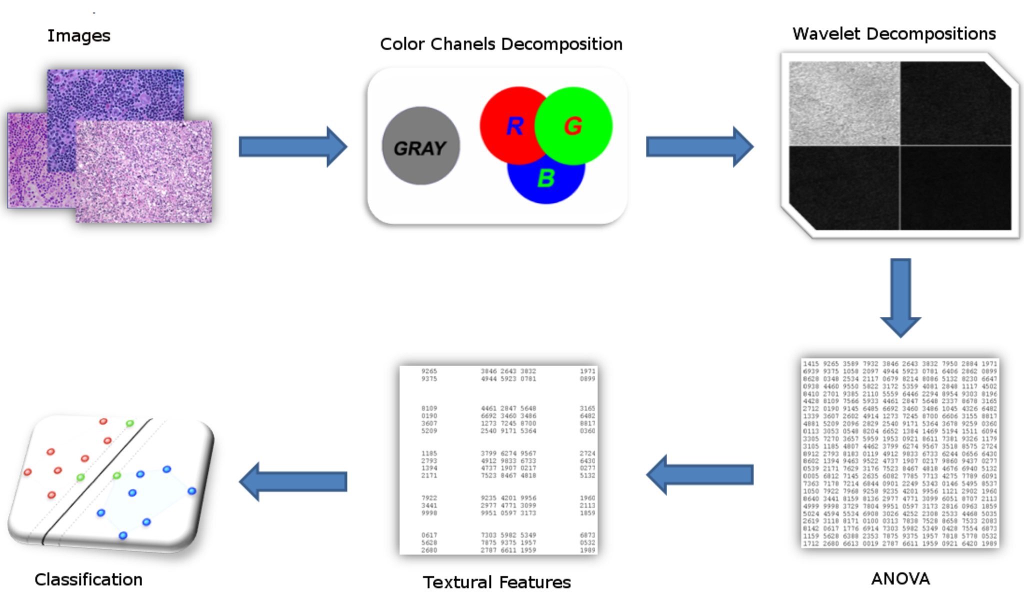

Figure 1: Diagram of the method.

3. Image Database

The database [7] was constructed by researchers of the *National Cancer Institute* and of the *National Institute on Aging*, both in the US. It contains images taken from 30 histological slides with lymph nodes stained with Haematoxylin and Eosin. They show 10 cases of each kind of lymphoma studied in our work (FL, MCL and CLL). The database contains 375 images, of which we have taken 210 at random: 70 of FL, 70 of MCL and 70 of CLL.

4. Extraction of Characteristics

We have chosen the SWT to extract sub-bands of the image because they are not downsampled by that transform, which is the case of the DWT. The SWT represents an image in the time-frequency domain (see [6] for details). In this way we kept information on low and high frequencies and obtained better results than with the DWT.

Journal of Physics: Conference Series **574** (2015) 012133 doi:10.1088/1742-6596/574/1/012133

Our SWT was implemented with Quadrature Mirror Filters, which consist of recursive high- and low-pass filtering. For the multiresolution analysis we applied the wavelet function Daubechies 8 to a single decomposition level. This resulted in four sub-bands of coefficients: a sub-band of approximation (the original image in low resolution), and three of details (horizontal, vertical and diagonal).

4.1. Reduction of Attributes and Classification

ANOVA is a statistical test that compares the averages of multiple groups. As a second step one applies ANOVA to the three sub-bands of detail separately. In this way we can clean noise and select relevant information [8, 9] in order to compare the results between the sub-bands.

The SVM classifier is based on machine learning techniques and statistical learning theory. Briefly saying, SVM tries to construct hyperplanes that maximize the separation between sets of data in order to group them into classes. It will fail if the sets are not linearly separable, but in this case SVM maps data through a *kernel function* within a space of characteristics of higher dimension that eventually makes data linearly separable [10].

As a third and last step the sub-bands that resulted from ANOVA were submitted to the SVM. Here we used the *n-fold cross validation method* with $n = 10$. It consists of taking 90% of the data as training set and the remaining 10% as test set. This is done 10 times by exchanging data between the 90% and the 10%, in such a way that in the end all data will have taken part in the test. We have used three different kernel functions in order to compare the final results: Linear, RBF and Polynomial.

5. Results

Table 1 summarizes the second step of our method. The third step ~~of our method~~ is summarized in Table 2, where we see that the Linear Kernel had the best performance. Figures 2(a-d) show the graphical confusion matrices for this kernel.

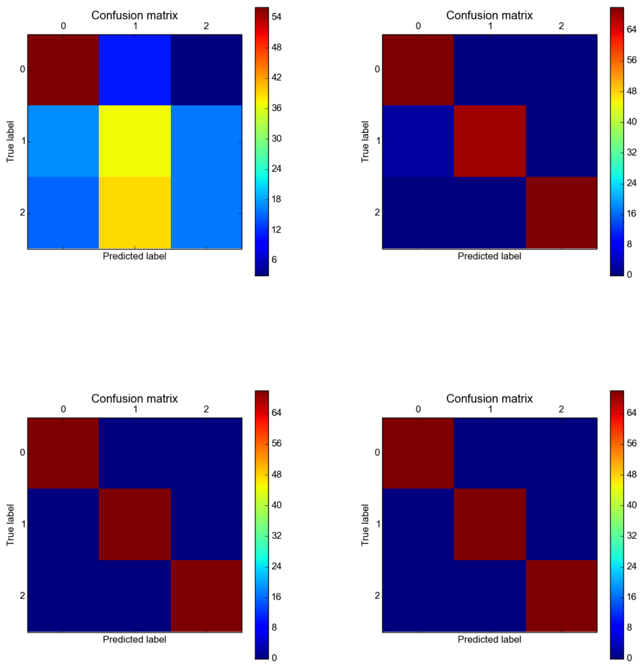

Figure 2: Confusion Matrices for (a) Approximation, (b) Horiz. (c) Vert. and (d) Diag. Details.

Table 1: Amount of attributes after ANOVA and their percentage reduction.

Sub-band	Attributes	ANOVA	Reduction (%)
Approximation	1,443,520	1,429,769	0.95
Horizontal Details	1,443,520	74,745	94.80
Vertical Details	1,443,520	75,321	94.80
Diagonal Details	1,443,520	34,236	97.60

Table 2: Results after applying SWT.

		Without Reduction		Reduction by ANOVA	
Sub-band	SVM-*Kernel*	Accuracy	Deviation	Accuracy	Deviation
Approximation	Linear	0.519	0.079	0.519	0.079
	RBF	0.514	0.044	0.514	0.044
	Polynomial	0.533	0.077	0.533	0.077
Horiz. Details	Linear	0.329	0.015	0.990	0.020
	RBF	0.433	0.065	0.948	0.035
	Polynomial	0.333	0.000	0.333	0.000
Vert. Details	Linear	0.333	0.000	1.000	0.000
	RBF	0.490	0.075	0.933	0.051
	Polynomial	0.333	0.000	0.333	0.000
Diag. Details	Linear	0.490	0.032	1.000	0.000
	RBF	0.319	0.032	0.886	0.060
	Polynomial	0.333	0.000	0.338	0.015

6. Conclusion

We presented a method to extract attributes through the analysis of textures based on SWT and ANOVA. The best results were obtained when we applied ANOVA to the vertical and diagonal coefficients, namely 100% of accuracy.

References

[1] Cancer Research UK 2014 Different types of non Hodgkin lymphoma we beat Cancer
[2] Leukemia & Lymphoma Society Team 2014 lls.org the Leukemia & Lymphoma Society (LLS) is the world's largest voluntary health agency dedicated to blood cancer URL `http://www.lls.org`
[3] Sertel O, Syst J K *et al.* 2009 *Journal of Signal Processing Systems* **55**
[4] Al-Kadi O S 2010 *Pattern Recognition* **43** 2043–2053 ISSN 0031-3203
[5] Krishnan M M R, Pal M, Bomminayuni S K, Chakraborty C, Paul R R, Chatterjee J and Ray A K 2009 *Computers in Biology and Medicine* **39** 1096–1104 ISSN 0010-4825 URL `http://www.sciencedirect.com/science/article/pii/S0010482509001668`
[6] Mallat S 1999 *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, CA)
[7] on Aging N I 2008 Lymphoma - NIA Intramural Research Program Laboratory of Genetics URL `http://ome.grc.nia.nih.gov/iicbu2008/lymphoma/index.html`
[8] Vieira S 2006 *Análise de variância: ANOVA* (Atlas)
[9] Surendiran B, Sundaraiah Y and Vadivel A 2009 *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09. International Conference on* pp 175–177
[10] Osta H, Qahwaji R and Ipson S 2008 *Systems, Signals and Devices, 2008. IEEE SSD 2008. 5th International Multi-Conference on* pp 1–6