A feature selection based framework for histology image classification using global and local heterogeneity quantification

J. Coatelen [1,2,3,*] A. Albouy-Kissi [1,3] B. Albouy-Kissi [1,4] J.P. Coton [2] L. Sifre [2]
J. Joubert-Zakeyh [5] P. Dechelotte [5] A. Abergel [1,3,5]

***Abstract*— Biopsy remains the gold standard for the diagnosis of chronic liver diseases. However, the concordance between readers is subject to variability causing an increasing need of objective tissue description methods. A complete framework has been implemented to analyze histological images from any kind of tissue. Based on the *feature selection* approach, it computes the most relevant subset of descriptors in terms of classification from a wide initial list of local and global descriptors. In comparison with equivalent methods, this implementation is able to find lists of descriptors which are significantly shorter for an equivalent accuracy and furthermore it enables the classification of slides using combinations of global and local measurements. The results have pointed that it could reach an accuracy of 82.8% in a human liver fibrosis grading approach by selecting 6 descriptors from an initial set of 258 global and local descriptors.**

***Index Terms*— framework, fibrosis, feature selection, support vector machines, quantification**

I. INTRODUCTION

Biopsy is currently the gold standard to diagnose the amount of inflammation (also called grade) and fibrosis (also called stage of the disease) in chronic liver diseases. To evaluate the stage, two semi-quantitative scores are used by pathologists in every day practice: Metavir [1] and Ishak scores. There are many limitations to the liver biopsy: concordance between senior reader and junior reader is fair ($\kappa < 0.6$), concordance between two senior readers is fair to good according to the level of fibrosis (fibrosis is classified in 5 stages in the Metavir score: no fibrosis (F0), low fibrosis (F1), moderate fibrosis (F2), severe fibrosis (F3), cirrhosis (F4)).

These difficulties illustrate an increasing need to define objective tissue description methods. In image analysis objective approaches are implemented and compared to the gold standard. This leads to efficient techniques that allow feature extraction and quantification from histological images.

* Corresponding author
[1] Université d'Auvergne, 49 Boulevard Franois-Mitterrand, CS 60032, 63001 Clermont-Ferrand CEDEX 1, France
[2] HISTALIM, 126 rue Emile Baudot, 34000 Montpellier, France
[3] Image Science for Interventional Techniques, Btiment 3C, Faculté de médecine, 28 place Henri Dunant, 63001 Clermont-Ferrand, France
[4] Institut Pascal, CNRS UMR 6602, Université d'Auvergne, BP 10448, 63000 Clermont-Ferrand, France
[5] CHU Estaing, Clermont-Ferrand, France
Email addresses : jeremy.coatelen at etu.udamail.fr (J. Coatelen*) ; {adelaide.kissi, benjamin.albouy} at udamail.fr (A. Albouy-Kissi, B. Albouy-Kissi) ; {jpcoton, lsifre} at histalim.com (J.P. Coton, L. Sifre) ; {jjoubert, pdechelotte, aabergel} at chu-clermontferrand.fr (J. Joubert-Zakeyh, P. Dechelotte, A. Abergel)

A complete feature-selection framework has been implemented to make it as easy and flexible as possible to analyze histological images from any kind of tissue. The framework is used in an original manner to characterize liver fibrosis stage with morphological, structural and textural global and local features.

II. PREVIOUS RESEARCH

The *feature selection* approach is more and more popular in histology image analysis. Saeys et al. [2] put forward the two most common difficulties of its application to histology: large input dimensionality and small sample size. To deal with these problems, several solutions have been tested as using multiresolution approaches or improving the subset search algorithm. Kong et al. [3] present an automated grading algorithm of neuroblastoma based on multiresolution approach of the *bag-of-features* which can deal with high resolution images. As shown by Tabesh et al. [4], the feature selection method accuracy can reach 96.7% on classification of histological images of prostate according to the Gleason grading reference. The possibilities around this technique have been widely explored by Caicedo et al. [5] and the main advantage of their method is its adaptiveness to the particular contents of the image collection through the automated construction of a codebook [6]. Later, some work has been done on improving the algorithms as Ozçift [7] who uses Random Forest algorithm instead of the classical but efficient SVM algorithm [8], DiFranco et al. [9] who show how to increase the heat-map coherence using spatial filters and Raza et al. [10] who increase performance combining scale and rotation invariance for renal cell carcinoma subtype classification.

III. APPROACH

A. Principle

The framework takes advantage of statistical learning and classification methods (Fig. 1). Thus, the results can be used for histology image segmentation and classification.

To this end, numerous morphological, textural and structural features are extracted from these images. Let N be the initial number of descriptors $(d_i)_{i\in[1,N]}$ (more than 200 kind of descriptors) which corresponds to feature quantitative measurements of the images. The larger the initial set of descriptor, the greater the number of possibilities to analyse the slides, with the advantage of a non-*a priori* approach, as discussed below. The descriptors can be global $(d_{g_i})_{i\in[1,N_g]}$

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

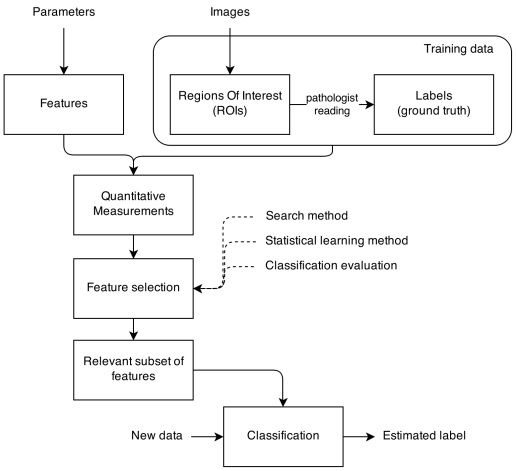

Fig. 1. Framework global structure.

or local $(d_{li})_{i\in[1,N_l]}$, with N_g and N_l respectively the number of global and local descriptors. A global descriptor gives a wide description of the image (e.g. global graph features, fractal dimension [11], proportion of stained tissue) while a local descriptor gives localized information (e.g. run length, co-occurrence matrix, histograms, texture spectrum, local binary pattern [12] or contour harmonics [13]). Each descriptor is an independent quantitative measure which is an other way to describe an image without *a priori* consideration. Local descriptors can be used either for the tissue identification (e.g. segmentation) or the global classification of the slide (e.g. grading approach).

Let $(S_j)_{j\in[1,Q]}$ be a set of Q histological slides as large and representative as possible. In a grading approach, for each slide one global image g_j and numerous local regions of interest (lROIs) $(l_{j,\bullet}) = (l_{j,k})_{k\in[1,M_j]}$ are extracted. A reading is done by pathologists (i.e. grade of fibrosis) and if C is the number of different grades of the pathology, let $L_j \in [1, C]$ be the label associated with S_j. For each slide S_j, global and local measurements are performed respectively on g_j and $(l_{j,\bullet})$. As the ground truth is defined for the whole slide, local measurements have to be considered all together and globally. The model extracts a series of distribution features (mean, standard deviation, kurtosis and skewness) from the computation of local descriptors on the slide lROIs $(d_l(l_{j,\bullet}) = d_l((l_{j,k}))_{k\in[1,M_j]}$ in $S_j)$. In summary, for each slide S_j the following measurements are performed:

$$d_{g_i}(g_j),\ i \in [1, N_g] \tag{1}$$

$$\left.\begin{array}{r} \text{mean}(d_{li}(l_{j,\bullet})) \\ \text{standard-deviation}(d_{li}(l_{j,\bullet})) \\ \text{kurtosis}(d_{li}(l_{j,\bullet})) \\ \text{skewness}(d_{li}(l_{j,\bullet})) \end{array}\right\} i \in [1, N_l] \tag{2}$$

From the previous measure results and the ground truth, a *feature selection* algorithm is used. This algorithm does a statistical research on the initial list of features to obtain the most relevant association of features Π^*. The *feature selection* approach is defined by two functions: the search method and the evaluation method. From exhaustive search (very expensive) to efficient search techniques such as random search, there are numerous exploration methods. In this study, greedy algorithms are performed: at each stage, they make the locally optimal choice with the hope of finding a global optimum. Forward and backward greedy algorithms were used, the first starts from an empty set whereas the latter starts from the whole set of descriptors. At each step, a subset of descriptors Π is evaluated over the whole training data using machine or statistical learning techniques (e.g. Recursive Partitioning or Support Vector Machines). A *k-fold cross-validation* is used to ensure a robust evaluation of the resulting classification. The average of bad classification rates ε depends on Π. The exploration method gives the combination of descriptors Π^* such as $varepsilon$ is minimized ($\varepsilon = \varepsilon^*$ is minimal). Depending on the search algorithm, the minimum of error rate is not necessary reached at a global minima.

When the combination of descriptors is computed with an error rate sufficiently low, the framework is calibrated. It is then possible to use the classifier with these descriptors for the classification of new cases.

B. Heterogeneity descriptors

Numerous descriptors have been implemented to deal with histopathological slides. This list has been established by studying different ways to describe and quantify the heterogeneity on histological slides, as many ways to approach the problem from a different angle. Thus it is interesting to combine them in order to characterize a complex issue. These measurements are grouped into three categories (Table I):

The morphological features provide information about the shape of the objects in the image. The morphology is particularly useful to quantify shape or contour irregularities, the density or the type of the objects. In particular, this kind of heterogeneity criteria is relevant to evaluate the fibrosis expansion. This set of morphological features (Table I) includes the area, the concavity, the eccentricity or the roundness.

The image structure is also considered using a graph-based model of the image. From the histology image spectral data and an optional segmentation mask, a graph is extracted as follows. A set of primitives is generated from the local homogeneous regions. A Delaunay triangulation is performed to build a regular graph from the primitives. The non-relevant edges are filtered using the local context and the minimum spanning tree is optionally extracted. The structural features are then extracted from the resulting graph. The graph structure is appropriate to quantify complex structures or networks which particularly applies to fibrous tissue. The set of structural features is given in Table I.

Textural features are often used to describe the tissue. It is appropriate to extract the texture features for the tissue identification and classification, the definition of the region boundaries and the analysis of tissue components. This set of features includes the proportion of stained tissue, the histograms, the homograms [14], the Fourier distance [15], the autocorrelation [16], the texture spectrum [16], the local

TABLE I

TEXTURAL, MORPHOLOGICAL AND STRUCTURAL FEATURES.

Textural features		
Proportion of stained tissue	Texture spectrum [16]	Fourier distance [15]
Histograms & Homograms [14]	Fractal dimension [11]	Run-length matrix [17]
Local binary patterns [12]	co-occurrence matrix [9]	Autocorrelation [16]
Morphological features [8].		
Area	Perimeter	Aspect Ratio
Form factor (compactness)	Area equivalent diameter	Zernike moment
Perimeter equivalent diameter	Eccentricity	Rectangularity
Fourier edge descriptor [13]	Contour irregularity	Concavity
Convex area	Solidity	Roundness
Spot areas ratio (pigmentation ratio)	Area irregularity [18]	Orientation
Structural features: graph-based features [19]		
Number of nodes/edges	Number of cycles	Total length
Total weight	Degree distribution	Cost
Modularity	Shortest pathlength	Density
Graph spectral analysis	Distance to the nearest	Centrality
Spectral radius	Clustering coefficient	Detour Index
Network Density	Pi Index	Eta Index
Theta Index	Beta Index	Alpha Index
Gamma Index	Assortative coefficient	Hierarchy
Shimbel Index distribution	Eigen exponent	Hub Dependence
Average nearest neighbors degree distribution	Participation coefficient distribution	Number of connected components
Polygons and triangles features	Cohesion index distribution	Assortativity
Koenig number distribution	Sum of the eigenvalues in the spectrum	

binary patterns [12], the fractal dimension [11], the co-occurrence matrix [9] and the run length matrix [17].

C. Implementation specifications

The framework is neutral, for instance, it does not define the segmentation thresholds, the filters, the graphs, the size of the regions of interest or the parameters of the morphological operators by itself. Anything that can remove its neutrality is given as input parameter or input data.

The framework calibration will always depend on training data. Therefore, if it is calibrated to recognize glioblastoma, it will not be appropriate to deal with fibrosis. A calibration is required for each different case. Similarly, particular attention should be paid to the training data specificity.

The framework is thought as an association of modules in such a way that every implemented technique can be substituted by another equivalent technique. It provides some flexibility in the analysis process. Thus, it is possible to redefine the training data, the parameters, the initial list of descriptors, the exploration strategy, the training method and the classification evaluation function.

IV. RESULTS

Several benchmark datasets from the R package mlbench were used to validate the *feature selection* implementation and the results are compared to those obtained by Blachnik [20] (Table II). Two statistical classifiers are used: rpart (recursive partitioning and regression tree) and svm (support vector machine), respectively from rpart and e1071 R packages. For both methods, the parameters are set to default values which are detailed in [21] and [22]. Two backward elimination search functions (greedy algorithm) are tested: an in-house recursive algorithm (RB) and the backward search function from the FSelector R package (FS). Acc and F are respectively Accuracy ($Acc = 1 - \varepsilon$) and Fraction of selected features. The framework results

TABLE II

CLASSIFICATION METHODS IN COMPARISON WITH BLACHNIK [20](B).

Benchmark	B		rpart+FS		rpart+RB		svm+FS		svm+RB	
dataset	Acc	F	Acc	F	Acc	F	Acc	F	Acc	F
Iris	0.953	0.75	0.964	0.75	0.955	0.25	0.966	1	0.963	0.5
Ionosphere	0.878	1	0.899	0.906	0.882	0.281	0.945	0.938	0.916	0.218
Sonar	0.668	0.933	0.735	0.95	0.662	0.717	0.844	0.967	0.813	0.183
Wine	0.973	0.846	0.921	0.846	0.922	0.384	0.990	1	0.860	0.308
Breast Cancer	0.978	0.778	0.955	0.889	0.949	0.444	0.968	1	0.967	0.667

FS: FSelector backward elimination algorithm. RB: in-house recursive algorithm of the backward elimination approach. Acc: accuracy. F: fraction of descriptors.

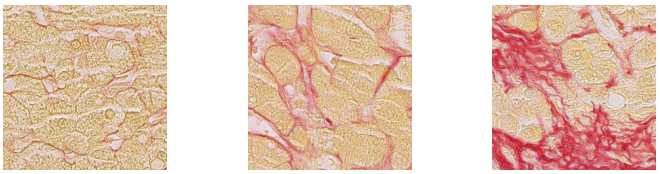

Fig. 2. Examples of human liver tissue. From left to right: low, average and high proportion of collagen - of fibrosis - stained with Sirius red.

are comparable with the reference results. The accuracy can keep close to the reference value with a number of selected descriptors significantly lower than the reference.

The framework was firstly tested with 8 biopsies of human liver (Sirius red stain). For each slide, 20 local regions of interest were extracted to gather 160 hepatic tissue images (of any kind). Each image is labeled by an expert according to the quantity of stained collagen in the image: "low proportion" (LP, 58 images), "average proportion" (AP, 71 images) and "high proportion" (HP, 31 images) (Fig. 2). As a first step, the classes LP and AP were grouped which gives 129 images labeled LP+AP and 31 images labeled HP. The classification results are shown in table III. The method rpart+RB gives an accuracy greater than 95% with 20 features. In a second step, the analysis was performed with the non-grouped images, with 58 images labeled LP, 71 images labeled AP and 31 images labeled HP (Table III). The error rate had a very significant increase for every methods in comparison with previous results. A minimal error rate approximatively equal to 16% is reached with a subset of 4 features with the method svm+RB. This increase is mainly due to difficulty to accurately distinguish LP and AP labeled images (even qualitatively) by the reference pathologists.

In a second time, the method was evaluated in the context of a grading approach using sixty-eight human liver biopsies (Sirius Red stain) labeled by pathologists using gold standard Metavir scores. Considering possible variabilities in biopsy labeling, the ground truth was defined as follows ($C = 3$): L_1 (scores F0 and F1, 17 images), L_2 (scores F2 and F3, 32 images) and L_3 (scores F4, 19 images). A large initial set of descriptors ($N_g = 2$ and $N_l = 256$), a forward search (FSelector R package) and a SVM based evaluator were used for this test. The size of the local regions of interest (lROIs) was 250 μm and the magnification was x5. The feature selection framework gives a subset of six descriptors ($F = 1.875\%$) with a total accuracy of 82.759%. The final subset is composed of the following descriptors: skewness of the co-occurrence matrix contrast, the co-occurrence matrix

TABLE III

EVALUATION OF THE TWO-CLASSES (LEFT) CLASSIFICATION VERSUS THREE-CLASSES (RIGHT) CLASSIFICATION OF LIVER TISSUE IMAGES.

	two-classes			three-classes		
Method	Acc	BER	F	Acc	BER	F
rpart+FS	0.9419	0.1092	0.9778	0.7463	0.2060	0.9556
rpart+RB	0.9573	0.0887	0.4444	0.7567	0.1987	0.5333
svm+FS	0.9316	0.1379	1	0.7749	0.2179	0.9556
svm+RB	0.9386	0.1195	1	0.8375	0.1585	0.0889

Acc: Accuracy. F: fraction of descriptors. BER: balanced error rate.

dissimilarity, the histogram kurtosis, the histogram standard deviation and the homogram standard deviation distributions, and the standard deviation of the homogram median value distribution.

V. DISCUSSION

The components of a system to classify histological images by quantifying textural, morphological and structural global and local features have been presented. From a wide initial list of descriptors and without *a priori*, it computes the most relevant subset of descriptors in terms of classification. In an original approach, distributions are defined from the local features which are extracted from the regions of interest of the slides and global features are extracted from these distributions (mean, standard deviation, skewness and kurtosis). Thus, even if the ground truth is a global information (grading approach) any kind of local feature takes part in the *feature selection* process with the global features. Benchmark tests have shown that, in comparison with equivalent methods, this *feature selection* implementation is able to find lists of descriptors which are significantly shorter for an equivalent accuracy. Moreover, the results have pointed that for the classification of human liver histological slides in terms of grading (simplified Metavir scale), the framework could reach an accuracy of 82.8% after selecting only 1.87% of the initial set of descriptors. However, it has been seen that the error rate could increase significantly (from 5% to 16%) when the ground truth is defined on an homogeneous dataset (when labeling at least two classes which are difficult to separate). Further work has to be done to improve the framework accuracy and robustness, including an assessment of the method in larger dataset.

REFERENCES

[1] M-C Rousselet, S Michalak, F Dupré, A Croué, P Bedossa, J-P Saint-André, and P Calès, "Sources of variability in histological scoring of chronic viral hepatitis.," *Hepatology (Baltimore, Md.)*, vol. 41, no. 2, pp. 257–64, Feb. 2005.

[2] Y Saeys, I Inza, and P Larrañaga, "A review of feature selection techniques in bioinformatics.," *Bioinformatics (Oxford, England)*, vol. 23, no. 19, pp. 2507–17, Oct. 2007.

[3] O. Sertel, J. Kong, H. Shimada, U V Catalyurek, J.H. H Saltz, M.N. N Gurcan, and K.L. Boyer, "Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation," *Pattern Recognition*, vol. 42, no. 6, pp. 1080–1092, June 2009.

[4] A Tabesh, M Teverovskiy, H-Y Pang, V P Kumar, D Verbel, A Kotsianti, and O Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images.," *IEEE transactions on medical imaging*, vol. 26, no. 10, pp. 1366–78, Oct. 2007.

[5] JC Caicedo, A Cruz, and FA Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Artificial Intelligence in Medicine*, Carlo Combi, Yuval Shahar, and Ameen Abu-Hanna, Eds., vol. 5651, pp. 126–135. Springer Berlin Heidelberg, 2009.

[6] A Cruz-Roa, JC Caicedo, and F a González, "Visual pattern mining in histology image collections using bag of features.," *Artificial intelligence in medicine*, vol. 52, no. 2, pp. 91–106, June 2011.

[7] Akin Ozçift, "Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis.," *Computers in biology and medicine*, vol. 41, no. 5, pp. 265–71, May 2011.

[8] M. Muthu Rama Krishnan, Chandan Chakraborty, Ranjan Rashmi Paul, and Ajoy K. Ray, "Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1062–1077, Jan. 2012.

[9] M D DiFranco, G O'Hurley, E W Kay, R W G Watson, and P Cunningham, "Ensemble based system for whole-slide prostate cancer probability mapping using color texture features.," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 35, no. 7-8, pp. 629–45, 2011.

[10] S H Raza, R M Parry, R a Moffitt, A N Young, and M D Wang, "An analysis of scale and rotation invariance in the bag-of-features method for histopathological image classification.," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 14, no. Pt 3, pp. 66–74, Jan. 2011.

[11] J. Coatelen, A. Albouy-Kissi, B. Albouy-Kissi, J P. Coton, L. Sifre, P. Dechelotte, and A. Abergel, "Automatic scale-independent morphology-based quantification of liver fibrosis," in *SPIE Medical Imaging*, Metin N. Gurcan and Anant Madabhushi, Eds. Mar. 2014, p. 904111, International Society for Optics and Photonics.

[12] H Qureshi, O Sertel, N Rajpoot, R Wilson, and M Gurcan, "Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification.," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11, no. Pt 2, pp. 196–204, Jan. 2008.

[13] G H Granlund, "Fourier Preprocessing for Hand Print Character Recognition," *IEEE Transactions on Computers*, vol. C-21, no. 2, pp. 195–201, 1972.

[14] H D Cheng, X H Jiang, and J Wang, "Color image segmentation based on homogram thresholding and region merging," *Pattern Recognition*, vol. 35, no. 2, pp. 373–393, 2002.

[15] D. Gibson and P.A. a Gaydecki, "Definition and application of a fourier domain texture measure: applications to histological image segmentation.," *Computers in biology and medicine*, vol. 25, no. 6, pp. 551–7, Nov. 1995.

[16] N Sebe and M S Lew, "Texture Features for Content-Based Retrieval," in *Principles of Visual Information Retrieval*, Michael S. Lew PhD, Ed., pp. 51–85. Springer London, 2001.

[17] X Tang, "Texture information in run-length matrices.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 7, no. 11, pp. 1602–9, Jan. 1998.

[18] P-W Huang and Y-H Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognition*, vol. 43, no. 4, pp. 1550–1563, Apr. 2010.

[19] A Madabhushi, S Agner, A Basavanhally, S Doyle, and G Lee, "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, vol. 35, no. 7-8, pp. 506–14, 2011.

[20] M Blachnik, "Comparison of Various Feature Selection Methods in Application to Prototype Best," in *Computer Recognition Systems 3*, Marek Kurzynski and Michal Wozniak, Eds., pp. 257–264. Springer Berlin Heidelberg, 2009.

[21] D Meyer, E Dimitriadou, K Hornik, A Weingessel, and F Leisch, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2014, R package version 1.6-2.

[22] T M Therneau, B Atkinson, and B Ripley, *rpart: Recursive Partitioning*, 2011, R package version 3.1-50.