Combining Unsupervised Feature Learning and Riesz Wavelets for Histopathology Image Representation: Application to Identifying Anaplastic Medulloblastoma

Sebastian Otálora[1], Angel Cruz-Roa[1], John Arevalo[1], Manfredo Atzori[2], Anant Madabhushi[3], Alexander R. Judkins[4], Fabio González[1], Henning Müller[2], and Adrien Depeursinge[2,5]

[1] Universidad Nacional de Colombia, Bogotá, Colombia
[2] University of Applied Sciences Western Switzerland (HES-SO)
[3] Case Western Reserve University, Cleveland, OH, USA
[4] St. Jude Childrens Research Hospital from Memphis, TN, USA
[5] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Abstract. Medulloblastoma (MB) is a type of brain cancer that represent roughly 25% of all brain tumors in children. In the anaplastic medulloblastoma subtype, it is important to identify the degree of irregularity and lack of organizations of cells as this correlates to disease aggressiveness and is of clinical value when evaluating patient prognosis. This paper presents an image representation to distinguish these subtypes in histopathology slides. The approach combines learned features from (i) an unsupervised feature learning method using topographic independent component analysis that captures scale, color and translation invariances, and (ii) learned linear combinations of Riesz wavelets calculated at several orders and scales capturing the granularity of multiscale rotation-covariant information. The contribution of this work is to show that the combination of two complementary approaches for feature learning (unsupervised and supervised) improves the classification performance. Our approach outperforms the best methods in literature with statistical significance, achieving 99% accuracy over region-based data comprising 7,500 square regions from 10 patient studies diagnosed with medulloblastoma (5 anaplastic and 5 non-anaplastic).

1 Introduction

Medulloblastoma (MB) is a type of brain cancer that represent roughly 25% of all brain tumours in children, it grows in the cerebellum on the lower, rear portion of the brain. Classifying MB is useful to determine aggressive phenotypes that require intense and early treatments [5, 7]. There are subtypes of MB based on histological appearance. These include classical, anaplastic, and desmoplastic. In anaplastic MB, the degree of anaplasia correlates to disease aggressiveness and is of clinical value when evaluating patient prognosis [7]. The problem of

© Springer International Publishing Switzerland 2015
N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 581–588, 2015.
DOI: 10.1007/978-3-319-24553-9_71

distinguishing between anaplastic and non anaplastic MB is difficult mainly due to the complexity of the patterns found in the histopathology images and is due to cell organization, size, shape and orientation variability towards the different malignant grades of the tumor. Examples of several tissue patterns from cases of MB whole slide images (WSI) are shown in Fig. 1. Common approaches for

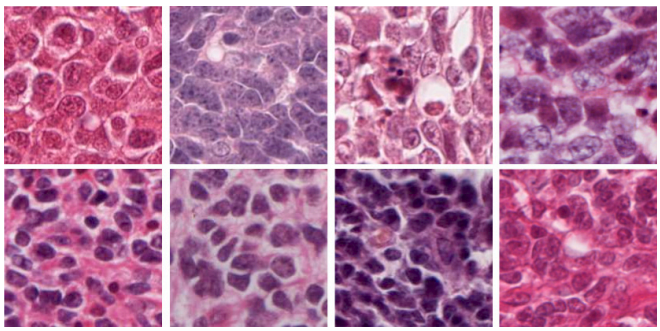

Fig. 1. Inter and intra class variability from MB tumor tiles of WSI at 40× magnification: anaplastic (Top), non-anaplastic (Bottom)

automatic tumor grading and phenotype differentiation rely on the identification of informative and discriminative features of the visual morphological patterns found in WSI. Unfortunately, they usually fail at capturing the variety of complex patterns present in the WSI, for example, considering only the subtle patterns captured by texture descriptors. This highlights the need for more powerful techniques or combinations of techniques to overcome the challenging tasks of automatical analysis of WSI [12].

Learning appropriate representations directly from data is a powerful machine learning strategy that has recently been applied with great success to pattern recognition problems, including image understanding and speech recognition [11, 10]. The representative techniques are mainly based on neural networks and can be grouped into two types: deep learning algorithms and unsupervised feature learning (UFL). An important assumption of these methods is that data patterns can be represented by the interaction of several factors at several hierarchically organized levels with semantically increasing content. Convolutional Neural Networks (CNN) are a representative technique where the features are hierarchically learned through several layers that combine convolution and pooling with non-linear functions [10].

For MB tumor differentiation, previous work [1] showed that the use of deep learning and UFL techniques outperforms classical texture descriptors. In [2], the authors made a comparative evaluation of several representation learning techniques including CNNs, Topographic Independent Component Analysis (TICA) and sparse Autoencoders (sAE), against MR8 and Haar texture descriptors. The results demonstrated the superior performance of the features learned by TICA, which builds a rotation and translation invariant representation of cell organizations in the anaplastic MB subtype.

Neural networks are the dominant approach for representation learning. However there are other representation learning strategies that are able to adapt conventional image descriptors to the needs of a particular image analysis task. In [3], the authors propose a multiscale texture signature learning approach using rotation-covariant Riesz wavelets, where most relevant combinations of orientations and scales are learned directly from the data. This approach outperformed state-of-the-art representations based on local binary patterns and grey level cooccurrence matrices for lung tissue classification [4]. Drawbacks of the data driven representations approaches are the amount of parameters involved that have to be manually tuned, which requires more time in model training. Some texture based representations fail to describe the feature patterns present in training samples that the data-driven approach is able to find [1].

In this work, we propose a joint framework for classification of MB WSI, where the invariant properties of TICA features and the multiscale rotation-covariant properties of Riesz wavelet features complement each other. We hypothesize that this fusion can lead to a better classification performance. This work join efforts of [3] and [1] in a simple manner to achieve the best accuracy reported for this histopathology WSI database.

2 Methodological Description

2.1 Topographic Independent Component Analysis

TICA is an unsupervised feature learning model, inspired by findings of the visual cortex behaviour. It groups activations of units in order to discover features that are rotation and translation invariant [1]. These are appropriate features for histopathology image characterization since shapes and cell organizations can be present regardless of the position or orientation of cells. Particularly, TICA organizes feature detectors in a square matrix for l groups such that adjacent feature detectors activate in a similar proportion to the same stimulus. To learn such groups, we need to optimize the cost function:

$$J_{\mathrm{TICA}}(\mathbf{W}) = \frac{\lambda}{2}\sum_{i=1}^{T}\left\|\mathbf{W}^{\mathrm{T}}\mathbf{W}x^{(i)} - x^{(i)}\right\|_2^2 + \sum_{i=1}^{m}\sum_{k=1}^{l}\sqrt{\mathbf{H}_k(\mathbf{W}x^{(i)})^2 + \epsilon} \qquad (1)$$

where $x^{(i)} \in \mathbb{R}^m$ is the *i-th* sample, T is the number of samples, $\mathbf{W} \in \mathbb{R}^{n\times m}$ is the matrix that encodes the features in each row, and $\mathbf{H} \in \{0,1\}^{l\times n}$ is the binary topographic organization where $\mathbf{H}_k^{(j)} = 1$, if the j-th feature detector, j-th row of $\mathbf{W}$, belongs to the k-th group, and 0 otherwise. This model sets $\mathbf{H}$ fixed while learning $\mathbf{W}$. In addition, TICA has two main computational advantages. First, the only parameters to be tuned are the regularization hyperparameter λ and the sparsity controller ϵ. Second, it is an unconstrained optimization problem, which can be solved efficiently by optimization techniques such as Limited memory-Broyden-Fletcher-Goldfarb-Shanno (L-BFGS).

2.2 Image Representation via Rotation-Covariant Riesz Wavelets

A fine to coarse wavelet image representation is obtained using Riesz wavelets. Class-wise texture signatures are learned for each scale as follows. First, image tiles are expressed in a feature space spanned by the energies of the responses of Riesz components. Then, an optimal linear combination of the Riesz components is learned using support vector machines (SVM) to maximize the margin between two classes. This linear combination yields class-wise texture signatures, that are locally aligned to maximize its response at the smallest scale, yielding rotation-covariant texture representations [3]. The wealth of the filterbank is controlled by the order N of the Riesz transform $\boldsymbol{\mathcal{R}}$, defined in the Fourier domain as:

$$\widehat{\mathcal{R}^{(n_1,n_2)}f}(\boldsymbol{\omega}) = \sqrt{\frac{n_1+n_2}{n_1!n_2!}}\frac{(-j\omega_1)^{n_1}(-j\omega_2)^{n_2}}{||\boldsymbol{\omega}||^{n_1+n_2}}\hat{f}(\boldsymbol{\omega}),$$

for all the combinations of (n_1, n_2) with $n_{1,2} \in \mathbb{N}$ such that $n_1 + n_2 = N$. The vector $\boldsymbol{\omega}$ is composed of $\omega_{1,2}$ corresponding to the frequencies along the two image axes and $\hat{f}(\boldsymbol{\omega})$ is the Fourier transform of $f(\boldsymbol{x})$. The Riesz transform yields $N+1$ distinct components behaving as N-th order directional differential operators when coupled with a multi-resolution framework based on isotropic band-limited wavelets ψ_s, with $s = 1, \cdots, S$ the number of iterations of the wavelet transform. An interesting property of Riesz wavelets is that the response of each component $\mathcal{R}^{(n_1,n_2)}$, rotated by an arbitrary angle θ, can be derived from a linear combination of the responses from all components of the filterbank. This steerability property is leveraged to obtain rotation covariant-texture features [3]. For a class c, the multiscale texture signature Γ_c^N is defined as:

$$\Gamma_c^N = w_1\left(\mathcal{R}^{(N,0)}\right)_{s=1} + w_2\left(G * \mathcal{R}^{(N-1,1)}\right)_{s=1} + \cdots + w_{SN+S}\left(G * \mathcal{R}^{(0,N+1)}\right)_{s=S}.$$

SVMs are used to find the optimal weights $\boldsymbol{w}^{\mathrm{T}} = (w_1 \ldots w_{SN+S})$[3]. Finally, the local orientations of each signature are optimized to maximize their local magnitude at the first scale $\Gamma_{c,1}^N$. The final representation is a vector of dimensionality $(N+1) \times J \times k$, where k is the number of classes.

2.3 Fusing UFL and Riesz Features

Once we have the features for each tissue tile, the features are concatenated. The latter is used as input to feed a standard softmax classifier with weight decay regularization. Two outputs of the classifier represent the probability for a tissue tile being considered as anaplastic or not, respectively. To train the models weights Θ that map the fused features into the anaplastic probability, the following cost function is minimized with an L-BFGS procedure:

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k}\mathbb{I}\{y^{(i)}=j\}\log\left(\frac{\exp\Theta_j x^{(i)}}{\sum_{l=1}^{k}\exp\Theta_l x^{(i)}}\right)\right] + \frac{\rho}{2}\sum_{i=1}^{k}\sum_{j=1}^{n}\Theta_{ij}^2,$$

where m stands for the number of samples, and k is the number of classes, and ρ is the weight decay parameter that penalizes large values for parameters. The fused representation of an unseen test tissue tile $x^{(p)} \in \mathbb{R}^{(N+1)Jk+l}$ is classified as anaplastic (or non-anaplastic) by calculating a probability:

$$p(y^p = 1|x^p; \Theta) = \frac{\exp\left(\Theta_1 x^{(p)}\right)}{\sum_{l=1}^{2} \exp\left(\Theta_l x^{(p)}\right)}.$$

A tile belongs to the anaplastic class if $p(y^p|x^p; \Theta) > 0.5$ and non-anaplastic otherwise.

3 Experimental Results and Discussion

The workflow of the proposed approach is summarized in Fig. 2. As first step, we compute the UFL features learned by TICA and the supervised features learned with Riesz wavelets for each image as described in Sections 2.1 and 2.2. Once TICA and Riesz wavelets are computed a final step of supervised classification is made using the combination of the computed features in a concatenated vector as input for a standard softmax classifier as described in Section 2.3. Parameter tuning is presented in Section 3.2.

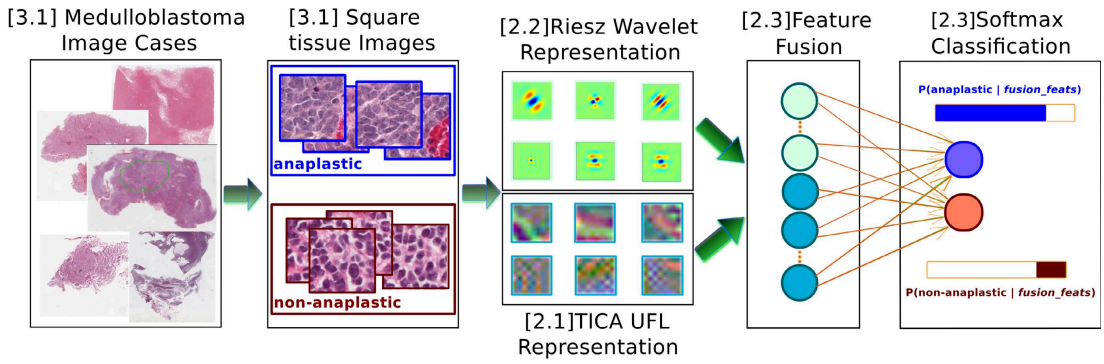

Fig. 2. Flowchart for MB feature extraction and classification for both learned representations: Riesz and TICA, the details of each stage are described in subsections.

3.1 Medulloblastoma Dataset

Our MB database is from St. Jude Childrens Research Hospital in Memphis where a neuropathologist manually annotated the cancerous regions of 10 pathology slides, 5 diagnosed as anaplastic and 5 as non-anaplastic MB. Slides were stained with hematoxylin and eosin (H&E) and digitization was done on an Aperio Scanner obtaining WSI with a resolution of 80,000×80,000 pixels. Each image can have several cancerous regions, which were manually annotated. For training, we randomly extracted a total of 7,500 square regions (750 per case) of 200×200 pixels of the tumor regions (3,750 anaplastic and 3,750 non-anaplastic).

3.2 Experimental Setup

In order to evaluate the advantages of the presented approach we compare the performance of the state-of-the art work on this database [1, 7]. In [7], the authors used a bag-of-visual-words approach using texton-based features obtained from Haar wavelets and MR8 filterbanks. In [1], the authors perform an extensive comparison between UFL techniques (sAE and TICA) and CNN, showing that the TICA outperforms by a considerable margin the previous approaches in [7]. Our evaluations follow the same leave-two-patients-out cross-validation scheme that consist of 20 trials where for each 4 of the non-anaplastic and 4 of the anaplastic cases are randomly selected for training whereas the remaining 2 slides (i.e., 1 non-anaplastic and 1 anaplastic) are used for validation. The architectures and setup for the approaches are summarized as:

- 2-Layer CNN: The best CNN model reported in baseline [1] was reproduced, consisting of a 2-layer architecture with 225 features in the first layer, and 225 units on the fully-connected layer. The feature kernel was of 8×8 pixels and a pool size of 2×2.
- TICA: The best of the three different TICA models reported in baseline [1] were reproduced: $\text{TICA}^{225}_{F:8,P:1}$ using a pool size of 1 and using 225 features with a feature kernel size of 8×8.
- Riesz wavelets: For the presented Riesz wavelet representation we explore the order N of the wavelet as well as the number of scales J as $N \in [1, 2, 3, 4, 5]$ and $J \in [1, 2, 3, 4]$. We also tried concatenations of orders and scales.
- TICA + Riesz fusion: For the combination of the best approaches we propose to build a joint vector 240 features for each of the tiles composed of the 225 best features found by TICA and the 15 features of the concatenation of the Riesz wavelets of orders 1,2,3 and scales 2,2,1 respectively.

For softmax the weight decay parameter was explored logarithmically in the range [1e-10, 100]. The performance measures are accuracy, sensitivity and specificity.

3.3 Results

Table 1 presents the results of the approaches. The best results were obtained by combining TICA and Riesz wavelets. We show some qualitative digital annotation results on 2 sample test cases in Fig. 3. We compare the statistical significance of the results of the combined representation and the TICA results with the Kruskal-Wallis test that calculates the average rank of the accuracy results of two approaches on the 20 trials, and compute the p-value for the null hypothesis that the two set of results comes from the same distribution. The test gives us a p-value of 2.5394^{-7} at a 1% significance level, hence we reject the null hypothesis.

Table 1. MB classification performance (baseline, Riesz, fusion). The measures are averaged over the 20 test runs with standard deviation where available.

Method	Accuracy	Sensitivity	Specificity
TICA + Riesz$[N_3^1, N_2^2, N_1^2]$	**0.997 ± 0.002**	**0.995 ± 0.004**	**1 ± 0**
TICA [1]	0.972 ± 0.018	0.977 ± 0.021	0.967 ± 0.031
Riesz $[N_3^1, N_2^2, N_1^2]$ [3]	0.964 ± 0.038	0.999 ± 0.001	0.932 ± 0.07
Riesz $[N_3^1]$ [3]	0.958 ± 0.062	0.963 ± 0.05	0.916 ± 0.125
Riesz $[N_2^2]$ [3]	0.94 ± 0.02	0.94 ± 0.02	0.3 ± 0.04
2-Layer *CNN* [1]	0.90 ± 0.1	0.89 ± 0.18	0.9 ± 0.0.3
sAE [1]	0.90	0.87	0.93
BOF + *A2NMF* (Haar) [2]	0.87	0.86	0.87
Riesz $[N_1^2]$ [3]	0.85 ± 0.23	0.9 ± 0.15	0.7 ± 0.47
BOF + K - NN (Haar) [7]	0.80	-	-
BOF + K - NN (MR8) [7]	0.62	-	-

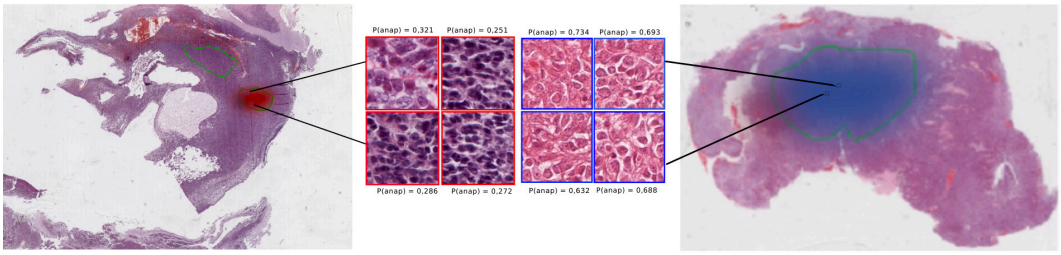

Fig. 3. Predictions over two WSIs, non-anaplastic MB (left) and anaplastic (right).

4 Concluding Remarks

We present a feature fusion between unsupervised feature learning and supervised Riesz wavelet representation that captures subtle pattern of textures as well as high level features, allowing to create a more separable feature space where the differentiation of medulloblastoma into anaplastic and non-anaplastic can be made with high classification accuracy outperforming any other result previously described in the literature. To our knowledge this is the first time that a feature fusion method is presented between UFL and the Riesz wavelets in the context of histopathology image analysis showing the complementarity between these learned features for the challenging task of tumour differentiation, we are currently working on extending the method to other patch-based histopathology image analysis problems with larger cohorts of patients.

Acknowledgments. This work was supported by the SNSF (PZ00P2_154891), by the Administrative Department of Science, Technology and Innovation of Colombia (Colciencias) (1225-569-34920) and by Microsoft Research LACCIR (R1212LAC006). Otlora and Cruz-Roa thank for the Young Researcher and Doctoral Fellowship grants (645/2014, 528/2011). The authors thank for K40 Tesla GPU donated by NVIDIA used for the training process.

References

[1] Cruz-Roa, A., Arevalo, J., Basavanhally, A., et al.: A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation. In: Proc. SPIE 9287, pp. 92870G–92870G–6 (2015)

[2] Cruz-Roa, A., González, F., Galaro, J., Judkins, A.R., Ellison, D., Baccon, J., Madabhushi, A., Romero, E.: A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. MICCAI 2012, pp. 157–164. Springer, Heidelberg (2012)

[3] Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D., Müller, H.: Rotation–covariant texture learning using steerable Riesz wavelets. IEEE Transactions on Image Processing 23(2), 898–908 (2014)

[4] Depeursinge, A., Foncubierta–Rodriguez, A., Van de Ville, D., Müller, H.: Multiscale lung texture signature learning using the riesz transform. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 517–524. Springer, Heidelberg (2012)

[5] Ellison, D.W.: Childhood medulloblastoma: novel approaches to the classification of a heterogeneous disease. Acta Neuropathologica 120(3), 305–316 (2010)

[6] Fuchs, T.J., Buhmann, J.M.: Computational pathology: Challenges and promises for tissue analysis. Computerized Medical Imaging and Graphics 35(7–8), 515–530 (2011)

[7] Galaro, J., Judkins, A., Ellison, D., Baccon, J., Madabhushi, A.: An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas. In: EMBC, 2011 Annual International Conference of the IEEE, pp. 3443–3446 (2011)

[8] Gurcan, M.N., Boucheron, L.E., Can, A., et al.: Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering 2, 147–171 (2009)

[9] Kothari, S., Phan, J.H., Stokes, T.H., Wang, M.D.: Pathology imaging informatics for quantitative analysis of whole-slide images. Journal of the American Medical Informatics Association 20(6), 1099 (2013)

[10] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105 (2012)

[11] Ranzato, M., Huang, F.J., Boureau, Y.L., LeCun, Y.: Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)

[12] Wang, H., Cruz-Roa, A., Basavanhally, A., et al.: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging 1(3), 34003 (2014)