

# A Texture-based Classifier to Discriminate Anaplastic from Non-Anaplastic Medulloblastoma

Ying Lai<sup>a</sup>, Satish Viswanath<sup>a</sup>, Jennifer Baccon<sup>b</sup>, David Ellison<sup>c</sup>, Alexander R. Judkins<sup>d</sup>,  
and Anant Madabhushi<sup>a</sup>

<sup>a</sup> Department of Biomedical Engineering, Rutgers University, <sup>b</sup> Penn State Hershey Anatomic Pathology,

<sup>c</sup> St. Jude Children's Hospital from Memphis, Tennessee,

<sup>d</sup> Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Keck School of Medicine University of Southern California

**Abstract**—Medulloblastoma (MB) is the most common brain tumor in children. There are four distinct subtypes of MB, but patients with anaplastic/large cell have the worst prognosis. Since the morbidity is highly correlated with treatment for MB, the ability to distinguish aggressive (such as anaplastic/large cell) MB is crucial. We present a scheme that leverages quantitative image texture features (Haar, Haralick, and Laws) and classifier ensembles (random forests) to automatically classify histological images from MB resection as being anaplastic/large cell or non-anaplastic/large cell. Preliminary results for our scheme when applied to patch-based classification of MB specimens yield an AUC of 0.91.

## I. INTRODUCTION

Medulloblastoma (MB) is a malignant (World Health Organization grade IV), embryonal central nervous system (CNS) tumor comprised of primitive neuroectodermal cells involving the cerebellum. MB typically presents between 3-10 years of age and accounts for nearly 25% of all pediatric brain tumors. Currently, histopathological classification of MB includes four subtypes, one of which (anaplastic/large cell) is recognized as carrying a worse prognosis. However, risk based stratification of MB for treatment primarily relies upon the combination of age, extent of post-operative disease, and metastasis [1]. The average 5-year survival for MB ranges from >80% in average risk cases to approximately 30% in patients <3 years of age.

The morbidity associated with treatment for MB can be significant [2]. Therefore, improvement in risk based stratification of MB by including histopathological classifiers is important. Currently, several molecular abnormalities in MB are recognized to be associated with prognosis, for example i17q and MYCC or MYCN amplification [3]. In addition, genomic analysis of MB has established evidence for molecular subtypes of MB based on patterns of signaling pathway activation including Wnt, Shh and others [4]. The relationship between these molecular abnormalities and histopathological classification of MB is an area of active study. Currently, the presence of anaplastic/large cell features in MB is recognized to independently confer worse prognosis for the patient automating the recognition of anaplastic histopathological features is a first step to establishing tools that will further aid in the development of revised MB classification which incorporates molecular and signaling pathway information to improve current risk based stratification.

In previous work, texture feature extraction methods have successfully been applied to distinguish regions of prostate cancer on digitized biopsy specimens [5]. Based on the histopathological features of anaplasia in MB, we have developed an automated scheme to classify for anaplastic and non-anaplastic MB using Hematoxylin and Eosin (H&E) stained digitized biopsy samples. Our approach employs a suite of texture features (Haar, Haralick, and Laws) and ensemble classifiers in order to discriminate between anaplastic and non-anaplastic features in MB; the assumption being that these texture operators can quantitatively capture the subtle morphological differences between the two MB tissue classes.

## II. METHODOLOGY

### A. Data

Histopathological samples stained with H&E from 6 MB cases were obtained from the consultation files of St. Jude Children's Research Hospital. Representative slides were digitized by whole slide scanning using an Aperio ScanScope slide scanner at 200X magnification. Regions of tumor were then manually extracted which demonstrated MB free of necrosis, hemorrhage or technical artifacts which were then reviewed by neuropathologists and annotated for the presence or absence of anaplastic features. These regions were segmented into 100 patches, where each patch size was 500 x 500 pixels. Three of the 6 patients had anaplastic/large cell MB, and the remaining had non-anaplastic/large cell MB. In total, 300 patches were extracted from regions annotated as anaplastic/large cell MB, and 300 from non-anaplastic/large cell MB regions.

### B. Feature Extraction

The colored H&E patches were converted to HSV space, and texture features were extracted on a per pixel basis from the V channel (refer to Fig. 1).

**Haar wavelet features:** The Haar wavelet transform captures visually perceptible features of the shape and interior structure of objects at multiple scales. This feature extraction method entails averaging or differencing pairwise pixels recursively, producing a total of 7 features.

**Haralick features:** This set of features utilizes the assumption that the texture information in an image is contained in the spatial relationship which the gray levels in the image have to

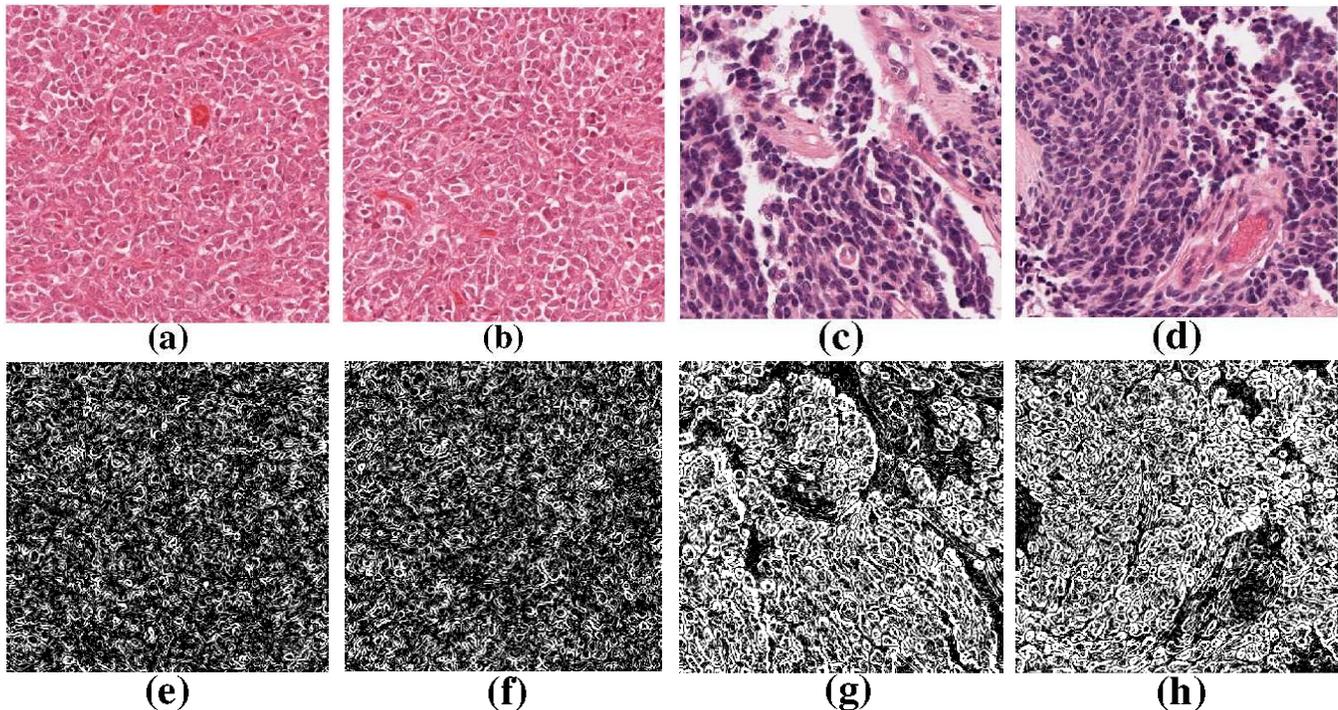


Fig. 1: (a), (b) are original anaplastic/large cell patches. (c), (d) are original non-anaplastic/large cell patches. (e)-(h) show corresponding Haralick texture responses for (a)-(d).

one another. 13 such features were extracted.

**Laws features:** Laws features are comprised by edges, spots, ripples, waves, low pass, or average gray value via a set of one-dimensional filters. 25 different two-dimensional filter responses were calculated by multiplying these filters against each other.

Features	Average AUC
Haar	$0.7173 \pm 0.205$
Haralick	<b><math>0.9146 \pm 0.0676</math></b>
Laws	$0.7991 \pm 0.1097$

TABLE I: Averages and standard deviations of AUC values via 25 runs of 3 fold cross validation of random forests in conjunction with different textural features to differentiate anaplastic and non-anaplastic MB.

### III. EXPERIMENTAL RESULTS

We evaluated each of the 3 different texture feature vectors (Haar<sup>1</sup>, Haralick<sup>2</sup>, and Laws<sup>3</sup>) in conjunction with a Random Forest classifier, which utilizes an ensemble approach to combine the results from multiple decision tree classifiers in order to obtain an improved classification result.

Evaluation was done via 25 runs of 3-fold cross validation, on a per-patch basis. The average Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve was calculated over all cross-validation runs, for each comparison considered (refer to Table I).

### IV. CONCLUDING REMARKS

In this work, we have presented preliminary results for a novel classification scheme to identify anaplastic/large cell

features of MB. We found that the Haralick feature set in combination with a random forest ensemble classifiers most successfully differentiates between anaplastic/large cell and non-anaplastic/large cell MB (AUC=0.91). In future work, we plan to extend and develop the scheme to use multi-class classification to quantify disease risk on digitized MB tissue specimens, as well as validate our results on a larger cohort of data.

### REFERENCES

- [1] D.W. Ellison, et al, *Definition of Disease-Risk Stratification Groups in Childhood Medulloblastoma Using Combined Clinical, Pathologic, and Molecular Variables*. J Clin Oncol, 2010.
- [2] E. C. Halperin, C. A. Perez, & L. W. Brady, *Perez and brady's principles and practice of radiation oncology*. Philadelphia: Lippincott Williams & Wilkins, 2008.
- [3] P. Gibson, et al, *Subtypes of medulloblastoma have distinct developmental origins*. Nature 468(7327), 1095-9, 2010.
- [4] D.W. Parsons, et al, *The genetic landscape of the childhood cancer medulloblastoma*. Science, 331(6016), 435-9, 2011.
- [5] S. Doyle, C. Rodriguez, and A. Madabhushi, *Detecting Prostatic Adenocarcinoma From Digitized Histology Using a Multi-Scale Hierarchical Classification Approach*. Proc IEEE EMBS, 4759-4762, 2006.