COMPUTER AIDED IMAGE SEGMENTATION AND CLASSIFICATION FOR VIABLE AND NON-VIABLE TUMOR IDENTIFICATION IN OSTEOSARCOMA

HARISH BABU ARUNACHALAM[1]*, RASHIKA MISHRA[1], BOGDAN ARMASELU[1],
DR. OVIDIU DAESCU[1] and MARIA MARTINEZ[2]

[1]*Department of Computer Science,*
[2]*Department of Biomedical Engineering,*
University of Texas at Dallas, Richardson, TX
**Email: harishb@utdallas.edu*

DR. PATRICK LEAVEY, DR. DINESH RAKHEJA, DR. KEVIN CEDERBERG,
DR. ANITA SENGUPTA and MOLLY NI'SUILLEABHAIN
University of Texas Southwestern Medical Center, Dallas, TX

ABSTRACT: Osteosarcoma is one of the most common types of bone cancer in children. To gauge the extent of cancer treatment response in the patient after surgical resection, the H&E stained image slides are manually evaluated by pathologists to estimate the percentage of necrosis, a time consuming process prone to observer bias and inaccuracy. Digital image analysis is a potential method to automate this process, thus saving time and providing a more accurate evaluation. The slides are scanned in Aperio Scanscope, converted to digital Whole Slide Images (WSIs) and stored in SVS format. These are high resolution images, of the order of 10^9 pixels, allowing up to 40X magnification factor. This paper proposes an image segmentation and analysis technique for segmenting tumor and non-tumor regions in histopathological WSIs of osteosarcoma datasets. Our approach is a combination of pixel-based and object-based methods which utilize tumor properties such as nuclei cluster, density, and circularity to classify tumor regions as viable and non-viable. A K-Means clustering technique is used for tumor isolation using color normalization, followed by multi-threshold Otsu segmentation technique to further classify tumor region as viable and non-viable. Then a Flood-fill algorithm is applied to cluster similar pixels into cellular objects and compute cluster data for further analysis of regions under study. To the best of our knowledge this is the first comprehensive solution that is able to produce such a classification for Osteosarcoma cancer. The results are very conclusive in identifying viable and non-viable tumor regions. In our experiments, the accuracy of the discussed approach is 100% in viable tumor and coagulative necrosis identification while it is around 90% for fibrosis and acellular/hypocellular tumor osteoid, for all the sampled datasets used. We expect the developed software to lead to a significant increase in accuracy and decrease in inter-observer variability in assessment of necrosis by the pathologists and a reduction in the time spent by the pathologists in such assessments.

Keywords: Image segmentation, Otsu thresholding, Osteosarcoma, SVS image analysis

1. Introduction

Pathology Informatics, one of the fastest growing fields in medical informatics, deals with mining information from medical pathology data and images. It involves the use of computational methods and analytical processes to make informed decisions that serve as assistive tools in clinical diagnosis. Due to the complexity of medical data and given the expert knowledge required for such analyses, it is often difficult to replicate the work of pathologists and physicians.[1] Though there is substantial literature published in the area of tumor research[2,3]

the main challenge in the field is that all the methods are tumor specific which makes the development of one common method, that is applicable for all kinds of tumor, an arduous task. This necessitates the creation of ad hoc methods tied to each requirement, that consider signature of each tumor sample and incorporate tumor specific information such as the tumor spread, contextual information etc. Each tumor detection method utilizes specific information about the tumor and therefore one tumor identification approach may not be applicable for another. Hence it becomes a challenge to apply existing methods that work well for other types of tumors for Osteosarcoma detection, the tumor that is used in this study.

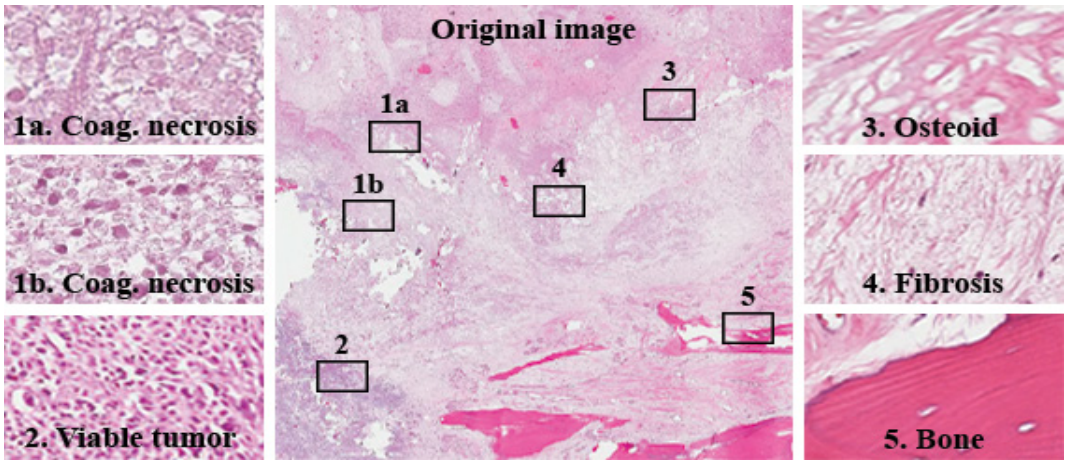

Fig. 1. WSI with different regions enlarged and their location in the image with figure 1a and 1b representing color and shape variations in coagulative necrosis regions for the same WSI. The numbered boxes represent the locations of histologically distinct regions in the image.

Osteosarcoma is the most common type of bone cancer that occurs in adolescents in the age of 10 to 14 years. The tumor usually arises in the long bones of the extremities in the metaphyses next to the growth plates.[4] What makes osteosarcoma analysis inherently challenging is that there is a high degree of histologic variability within the tumor (Figure 1) which is accentuated after therapy. In order to accurately identify tumor occurrence and estimate the treatment response, it is necessary to consider histologically distinct regions that include dense clusters of nuclei, fibrous tissues, blood cells, calcified bone segments, marrow cells, adipocytes, osteoblasts, osteoclasts, haemorrhagic tumor, cartilage, precursors, growth plates and osteoid (tumor osteoid and reactive osteoid) with and without cellular material. Each of these regions have different characteristic features that differ in color, shape, size, density, texture and area of occurrence. They also have significant differences in their biological features such as background stroma, presence or absence of certain cellular material, neighboring regions etc. There are also multiple color variations within the same dataset representing the same type of regions (Figure 1a and 1b), which makes segmentation based only on color ineffective. Due to the variable properties of the images, there is no one method that with certainty, can accurately segment regions and classify them. The literature available for osteosarcoma data image analysis/digital pathology is minimal which makes it critical to come up with methods that are applicable for this type of tumor analysis.

The goal of this paper is to present an approach to segment H&E[5] stained images into viable tumor and non-viable tumor regions using a combination of techniques (color segmentation, Otsu thresholding, and Flood-fill). It employs pixel-based and object-based segmentation by using color, shape and density parameters. The first step locates tumor and non-tumor regions and subsequent steps distinguish tumor into viable and non-viable regions. Color quantization in the approach accounts for variance in color distribution while shape properties such as area and circularity measures are utilized to accurately locate tumor. A further analysis performed on computed cluster data on resulting images characterize different regions in the image. The approach is shown to be robust and has a high accuracy overall in the datasets considered.

1.1. *Background and Setup*

We have assembled an investigative team of clinical scientists at University of Texas Southwestern Medical Center, Dallas and computer scientists at University of Texas at Dallas. Archival samples for 50 patients treated at Children's Medical Center, Dallas, between 1995 and 2015, have been identified. The treatment effect for each patient is estimated after surgical resection, by physically cutting the region of interest from the resected bone into pieces. These pieces are de-calcified, treated with H&E stain and converted to slides to be analyzed under microscope. Each patient case is represented by a single H&E stained slide at the time of biopsy when available and 8-50 H&E stained slides per case at time of resection when necrosis is determined. Each slide consists of 1-2cm X 1-2cm sections of the tumor in its widest coronal plane. Slides are scanned using an Aperio Scanscope at a magnification of upto 40X and stored in SVS format.[6] Each SVS WSI has a size between 150MB and 1.5GB and spans an order of 10^9 pixels. The experimental dataset consists of a subset of images from the above available patient datasets, manually annotated by pathologists.

1.2. *Related Work*

The Tumor identification/isolation problem has been well studied by researchers in the field of digital pathology[7–11] . The most common methods include image segmentation (region classification), image analysis (pattern analysis), regions of interest(ROI) identification, statistical analysis (such as number of clusters, mean size of groups) etc.[7,8] The methods used in all the studies include color based (pixel level), shape based (object level) and contextual information based methods.[8] Normally, pixel level methods, that make use of pixel level processing, form some of the basic approaches because they are the simplest but they are not the most efficient. Researchers have tried to analyze the pathological images based on quantitative metrics representing the spatial structure of histopathology imagery and include identifying structures such as nuclei, glands and lymphocytes etc. These spatial feature utilization[12] has become the backbone of histopathology image analysis techniques, as these are the prominent metrics that can yield maximum information. More advanced than the pixel based methods are the object based methods that make use of region growing and object identification utilizing shape properties. These methods provide better segmentation results than their pixel counterparts[2,13] however they are expensive in terms of the computational resources they use. Multi-level thresholding approach using Otsu segmentation promises good accuracy but when

there is a lot of noise, this method alone may not fare well. Pattern Recognition Image Analysis(PRIA) describes a pattern recognition method based on a genetic algorithm that evolves over multiple iterations and compares the results with GeNIE,[14] a bio-image analysis tool from Aperio and manual segmentation by pathologists. A recent work on Lung cancer[11] identifies 9879 image features and uses regularized classifiers to estimate patient prognosis.

1.3. *Challenges*

A majority of pathological images are in proprietary format and there is no common standard for these images, which makes it difficult for researchers collaborating towards common goals to share information. Openslide[15] and VIPS[16] are image processing libraries that help to narrow down the gap by a small margin, however, the main problem of handling different imaging formats remains the same.

Pathology is a relatively subjective field dependent on the opinions of trained pathologists resulting in discrepancy in the accuracy of different image analysis approaches available for pathological images. A study on renal cell carcinoma[10] performed analysis on pathologists and found a high degree of subjectivity in their evaluation. Therefore, a standard objective procedure is recommended, which is important not only from a clinical standpoint but also for developing quality research application that will be reliable and independent of varying views of pathologists.[7] The size of images generated in these studies is large and as a result the algorithms for smaller images may not scale up due to memory issues.[17] Hence, there is a need for a standard approach that will process images one tile at a time and at the same time can scale up without loss of accuracy. Most of the tools that are developed by researchers in academia cater to a subset of problems. ImageJ[18] has many inbuilt image processing algorithms, however, is limited in its use to process proprietary formats and large files. PRIA by Webster et.al[1] is an advanced method, but it fails to perform well in identifying necrosis, which is one of the main tasks in this study. CellProfiler,[19] a tool for high throughput image analysis is good at identifying cellular objects and calculating their properties. However, the results of our trials with CellProfiler are inconclusive in identifying cellular objects in Osteosarcoma. Object based methods[2,13] work well on images with well-defined shapes but need pre-configured training sets. Machine learning approaches such as Bayesian classifier, Support Vector Machines[9] etc. are effective but would need a large annotated training data and the training phase is very time consuming.[7] Some of the related works[9–11] in lung and breast cancer focus on identifying properties and features of nuclei. Necrotic regions in this study do not necessarily have nuclei and hence the above works address only a part of the problem. The presence of a high degree of variability in the shapes, in Osteosarcoma datasets, makes the above methods unlikely to perform well. Another issue with WSI images is the color variance between different features,[9] which causes active tumor cells to have different color signatures, and thus segmentation based only on color is less accurate.

Given these challenges, in the next few sections we describe our approach explaining the algorithm and our results.

2. The Approach

We illustrate in Figure 2, the complete procedure, which includes K-means color segmentation, multi-level Otsu segmentation, Flood-fill clustering and statistical analysis. Based on inputs from pathologists, we define the different tumor regions with the following properties which are then quantified in the segmentation and classification approach.

(1) **Viable Tumor:** Nuclei densely aggregated together
(2) **Non-Viable Tumor:** Cells and tissues in the stage of recovery or dead
   (a) *Coagulative Necrosis:* disintegrated nuclei but with less color density than viable tumor.
   (b) *Fibrosis:* Fibrous collagen (protein) produced by fibroblasts (benign cells).
   (c) *Acellular/Hypocellular Tumor Osteoid (subsequently designated as Osteoid in this paper):* Eosinophilic/pink extracellular protein matrix produced by tumor cells.

The viable tumor and coagulative necrosis regions resemble each other in terms of high color density, closer to blue while fibrosis and osteoid have brighter color shade, closer to pink. The above regions are grouped together into two intermediary classes Ψ_1 and Ψ_2 based on high intra-class similarities, as follows:

(1) $\Psi_1 = \{Viable\ tumor, Coagulative\ necrosis\}$
(2) $\Psi_2 = \{Fibrosis,\ Osteoid\}$

The images in Ψ_1 are analyzed in terms of shape and density properties to classify them as viable tumor and non-viable tumor, while those in Ψ_2 are by default classified as non-viable tumor.

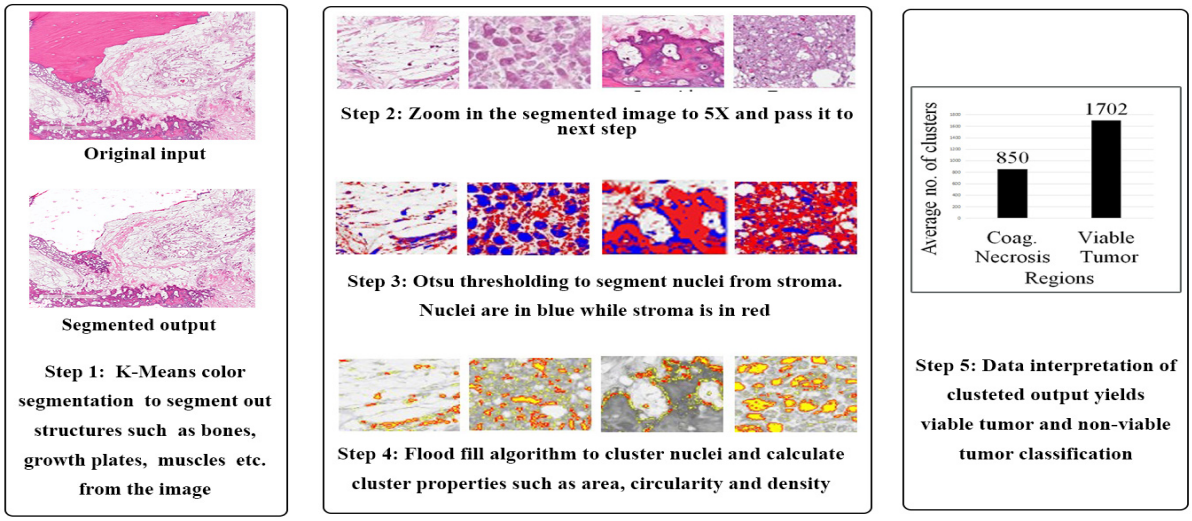

Fig. 2. **Algorithm pipeline**

2.1. *The Algorithm Pipeline*

The following is the general algorithm pipeline.
Input: Unprocessed SVS image

Output: Color segmented image and mapped regions identified as tumor (viable and non-viable) and non-tumor

Steps

For each SVS image given as input, at the eye fit (scaling factor 1X) zoom level, do the following:

(1) Run K-means color segmentation with K=3.
(2) For each of the tumor regions identified in step 1, increase the scaling factor to 5X.
(3) On a window size of 512 * 512, on the original 5X scaled images, do the following:
    (a) Generate red-blue segmentation using 2-level Otsu thresholding.
    (b) Compute the percentage of red pixels and blue pixels in each image.
    (c) Images with higher percentage of blue pixels fall into the Ψ_1 class
    (d) Images with higher percentage of red pixels fall into the Ψ_2 class
    (e) Create a tumor map of the pixel information by including pixel location, color values and class label
(4) For each entry in the tumor map,
    (a) Run Flood-fill algorithm to identify boundaries and group pixels in the cluster.
    (b) Remove clusters that are smaller than minimum cluster size and larger than maximum cluster size to remove false positives.
    (c) Output the remaining pixels in the original image along with cluster-mapped pixels.
(5) Run data analysis and classify images as viable and non-viable tumor based on the cluster data output.

2.2. *Color Segmentation*

At eye-fit level(scaling factor 1X), a 3-means color segmentation process is used to distinguish the given image into tumor and non-tumor image. Since all the WSIs are H&E stained, the pixels are made up of variants of Red and Blue channels. Hence it is imperative that more focus is given to these two channels. Given an Image I, of width I_w and height I_h, made up of $I_w * I_h$ pixels, each pixel P^i is represented by $\{P^i_r, P^i_g, P^i_b\}$, where P^i_r, P^i_g and P^i_b denote the red, blue and green channel values of i^{th} pixel. Let the set $C \epsilon (C^w, C^p, C^b)$ represent the cluster centers of white, pink and blue regions. These color values are taken from the empirical analysis of the stained images. Each P^i in the image is assigned to one of the cluster centers C^k by calculating distances between the pixel and the centroids. The distance is given by subtracting the color channel differences between the pixel and centroid. If $\phi(P^i)$ represents the cluster value for pixel i, then

$$\phi(P^i) \ = \arg\min_{C^j} \ \delta(P^i, C^j) \tag{1}$$

where,

$$\delta(P^i, C^j) \ = \ \sqrt{(P^i_r - P^k_r)^2 \ + \ (P^i_g - P^k_g)^2 \ + \ (P^i_b - P^k_b)^2} \tag{2}$$

where P^k_r, P^k_g, P^k_b represent RGB color channel values of the pixel, of k^{th} cluster centroid. The centroids are initialized with random values and each pixel P^i in I is classified. The pixel

values of centroids are then updated as follows:

$$P_r^k = \frac{1}{N_k}\sum_j P_r^j;\ P_g^k = \frac{1}{N_k}\sum_j P_g^j;\ P_b^k = \frac{1}{N_k}\sum_j P_b^j; \tag{3}$$

Where N_k is the number of pixels classified under k^{th} centroid. The algorithm is run for Γ iterations until there are no more changes to the clusters. The clusters represented by centroids C^b and C^p are regions of potential tumor whereas C^w represents non-tumor. The blue and pink clusters are further investigated at a higher level of magnification for detailed classification. After K-means, the data is passed on to the next step, with the following values populated for each pixel P^i. Map M contains $\{ m_{p_1}, m_{p_2}...\}$ and each m_{p_i} =*(pixel-location, color value, label)*

2.3. *Otsu multi-level threshold segmentation*

A 2-level threshold segmentation is used in the next step. A window of 512 x 512 is considered and the color image is converted to 24 bit grayscale image, with more weight to blue channel. This is due to the fact that tumor regions have more blue channel values than non-tumor regions. The gray scale values for two level threshold are represented as [1,2,...t] and [t+1,....L] respectively and the weighted class variance is calculated as

$$\sigma_w^2(t) \ = \ q_1(t)\sigma_1^2(t) \ + \ q_2(t)\sigma_2^2(t) \tag{4}$$

where

$$q_1(t) \ = \sum_{i=1}^{t} P(i);\ q_2(t) \ = \sum_{i=t+1}^{L} P(i) \tag{5}$$

are the class probabilities and the intra-class variance is given by

$$\mu_1(t) = \sum_{i=1}^{t} \frac{i * P(i)}{q_1(t)} \tag{6}$$

Now the image contains two classes of pixels following a bi-modal histogram. We calculate the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal. Otsu thresholding creates a red-blue color map where red signifies the non-viable tumor pixel and blue is the viable tumor or coagulative necrosis pixel. The data from this stage is exported by updating the values in map M as m_{p_i}=(pixel, red/blue value, original color value, label)

2.4. *Calculating clusters*

This stage calculates blue clusters and their properties. We run Flood-fill algorithm to identify boundaries and compute clusters. Viable tumor and coagulative necrosis always are in the blue region and contain cellular structures within them of non-uniform circularity. Along with each cluster, the size of the cluster in terms of area a, circularity c, and average color of cluster are calculated. Clusters that are less than minimum cluster size (50 pixels) and greater than maximum cluster size (300 pixels) are discarded as they represent false positives. The output of this stage is a map M' (Cluster label, Start Point, Centroid, Circularity, Area, List of Points, Color) and an image of clusters with red borders.

3. Results

The application was created using Openslide-Java for processing SVS images, ImageJ and Java Advanced Imaging for basic image processing tasks and C# .NET to perform Otsu segmentation and Flood-fill. The dataset included 120 images of 1160 x 640 resolution and all data samples were manually analyzed and classified by pathologists at UT Southwestern. The output of the application was compared with the classified images for verification and was validated by the team at UT Southwestern.

The choice of parameters such as window size in Otsu step (512*512), minimum and maximum cluster sizes (50 pixels and 300 pixels) etc. are selected based on empirical values for which the accuracy is maximized. Each figure (Figure 3 - Figure 6) consists of an original image I from the dataset, shown in (a), the output of Otsu method applied on I, (b) and clustering of cells by flood fill method,(c). Each cluster inside the red boundary in the images is defined in memory as a map data structure, Cluster(*Centroid,OriginalColor,Area,Perimeter,Circularity*).

3.1. *Viable Tumor*

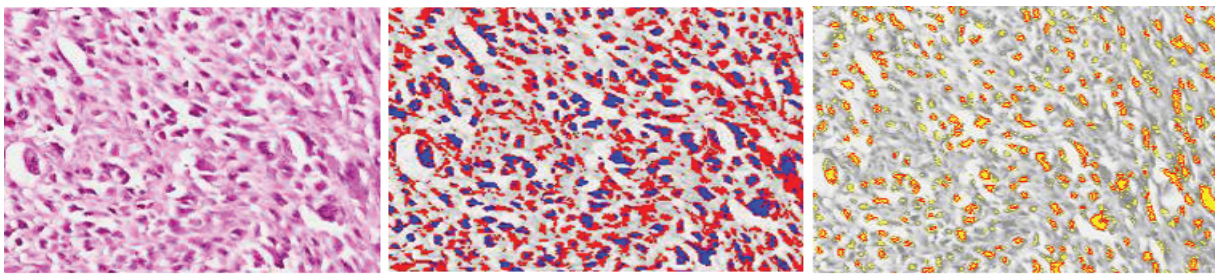

Fig. 3. Region: Viable tumor. **(a)** original image for viable tumor, **(b)** Otsu output showing more blue color, (**c**) Cell clustering using flood fill showing computed clusters.

Figure 3(a) shows that viable tumor has dense nuclei with more blue color. Otsu segmentation captures the nuclei with high accuracy represented by blue regions, as seen in 3(b). The percentage of blue pixels higher than the percentage of red pixels is a significant indicator for classifying this region into class Ψ_1. Clustered cellular information from Flood-fill 3(c) shows that there are more cells in the viable tumor region than the others. (see Figure 8(a)).

3.2. *Coagulative Necrosis*

Figure 4(a) shows coagulative necrosis containing cells with disintegrated nuclear matter, which makes the image appear brighter than viable tumor. Otsu segmentation step 4(b) yields higher percentage of blue pixels than red pixels, which is a key parameter in deciding regions belonging to class Ψ_1. Cluster properties in 4(c) show that the cell clusters are less dense and are distant from each other.

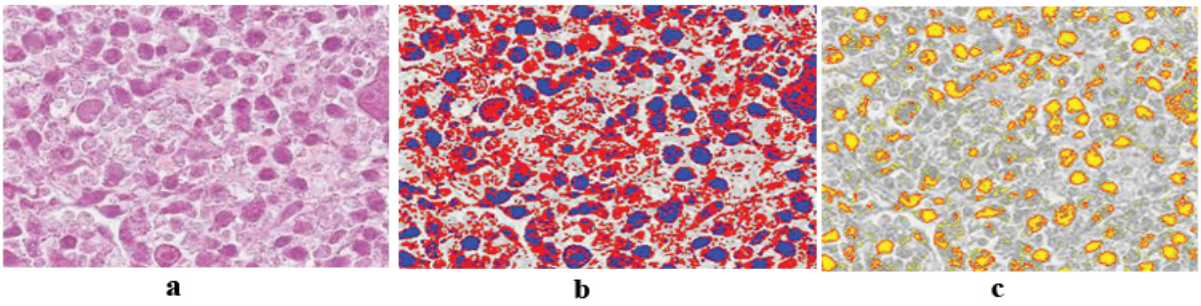

Fig. 4. Region: Coagulative necrosis. (**a**) original image for coagulative necrosis. (**b**) output of Otsu showing more blue than red, (**c**) Cell clustering using flood fill similar to viable tumor

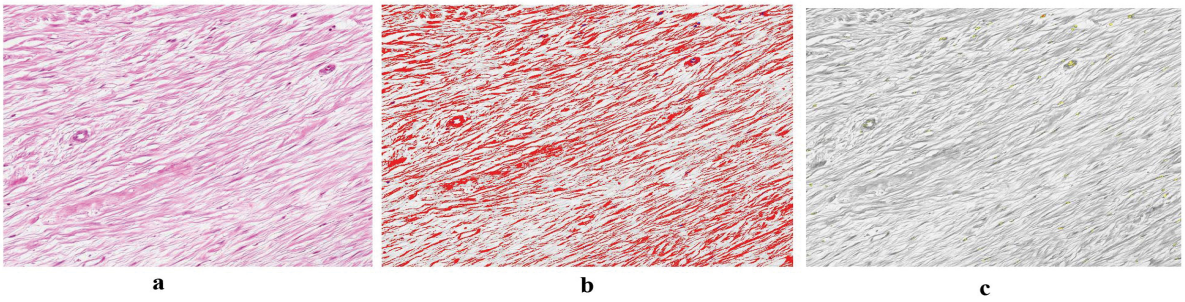

Fig. 5. Region: Fibrosis. (**a**) original image for fibrosis, (**b**) shows Otsu output for Fibrosis where there is a higher percentage of red pixels than blue, (**c**) cell clustering using flood fill, showing presence of fewer cells

3.3. *Fibrosis*

Figure 5(a) shows fibrosis region represented by strand like structures and absence of cells and nuclei. Due to this characterstic, Otsu in 5(b) produces higher percentage of red than blue pixels. This result distinguishes fibrosis from images in class Ψ_1. Flood-fill on this output, seen in 5(c), produces lesser number of clusters compared to viable tumor and coagulative necrosis.

3.4. *Osteoid*

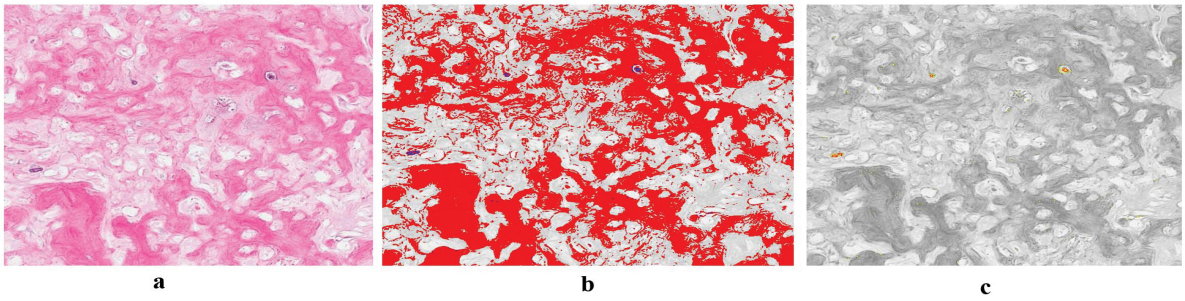

Fig. 6. Region: Osteoid. (**a**) original image for osteoid, (**b**) Otsu output for osteoid characterized by higher percentage of red than blue pixels. (**c**) Cell clustering flood fill output marked by absence of cells.

Figure 6(a) shows osteoid from the dataset, characterized by pink regions, background

stroma and absence of cells and nuclei similar to fibrosis. Running Otsu on this image produces 6(b), with high percentage of red pixels due to absence of cells. This result makes osteoid to be grouped in Ψ_2 distinguishing them from images in class Ψ_1. Since osteoid is an extracellular protein matrix, it remains after tumor cells have undergone necrosis. However, there maybe interspersed cells found in the matrix. Flood-fill on this output, shown in 6(c), captures scattered cells that are less dense, unlike viable tumor and coagulative necrosis.

3.5. *Data Interpretation*

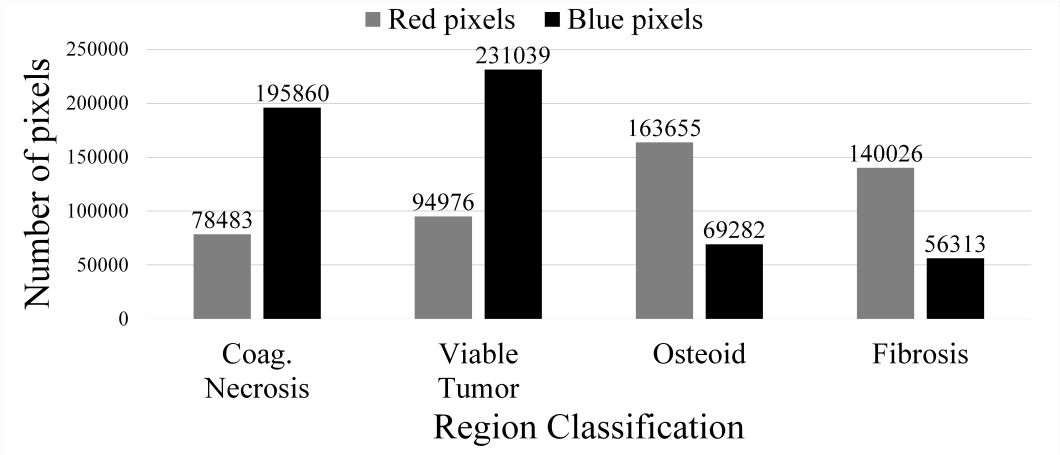

Fig. 7. Region wise average red and blue pixel count

A plot of average pixel count for classification regions shows that viable tumor and coagulative necrosis regions have more blue pixels, while fibrosis and osteoid regions have more red pixels. This result from Otsu step divides the image into prominent blue and red regions (see Figure 7). The regions that get classified under Ψ_1 have more cells than the regions under Ψ_2, and therefore have more blue pixels than red. Thus, images with viable tumor and coagulative necrosis regions can be classified into Ψ_1, while fibrosis and osteoid can be classified into Ψ_2.

A further analysis on average cell counts shows that viable tumor has 1702 cell clusters, while coagulative necrosis has 850 cells (see Figure 8(a)). This further distinguishes images in Ψ_1 into viable tumor and coagulative necrosis more accurately. We calculated the average density of cells in a 32x32 window as shown in Figure 8(b). It is observed that viable tumor has a cellular density of 2.4 while coagulative necrosis has 1.17. This important characteristic differentiates viable tumor from coagulative necrosis. The findings conclude that viable tumor is more dense and has closely aggregated cells than coagulative necrosis, the result of which has been used in classification.

It can be seen that fibrosis and osteoid regions have low cell clusters and high background stroma, hence concurring with the previous findings that these segmented images have less blue and more red pixels. Thus, the images in Ψ_2, that were identified as fibrosis and osteoid, have been categorized as non-viable tumor. Furthermore, in class Ψ_1, the cellular density distinguishes viable tumor from coagulative necrosis, which can be used to classify coagulative

necrosis under non-viable tumor.

The accuracy of the method has been measured and is found in Table 1.

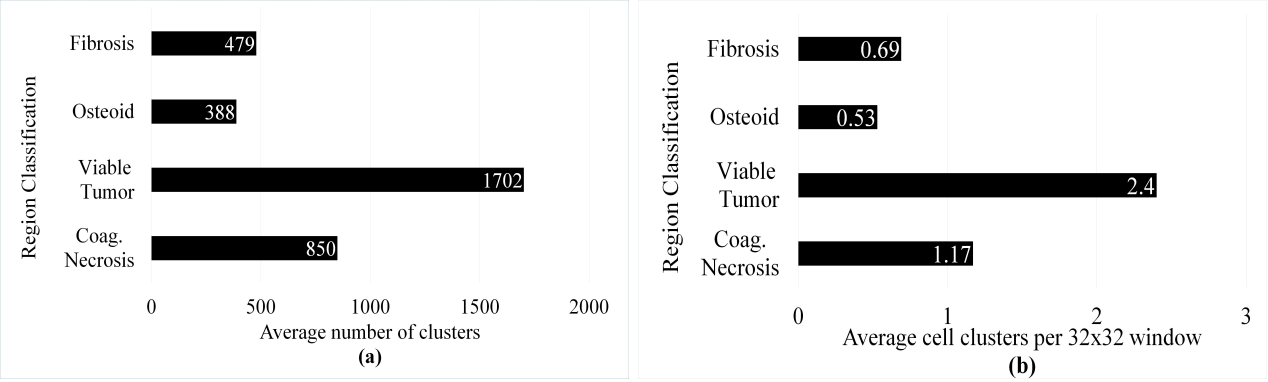

Fig. 8. (a) Region wise cellular count.(b) Region wise average cellular density count per 32x32 window.

Table 1. Quantitative metric comparison of classification regions

Region Type	Quantitative Metrics (Average)				
	red pixels count	blue pixels count	cell clusters	cell density (32x32 window)	Classification accuracy(%)
Viable Tumor	94976	231039	1702	2.4	100
Coagulative Necrosis	78483	195860	850	1.17	100
Osteoid	163655	69282	388	0.53	93
Fibrosis	140026	56313	479	0.69	89

4. Limitations and Future Improvements

The approach presented in this paper is limited to Osteosarcoma tumor identification. The current method works well in the given sampled datasets. We propose to extend it to all images in the dataset by incorporating contextual information that yield additional data. RGB color channels used in the experiments are affected by color variance in images and hence we would replace them with LAB colorspace. We plan to identify more relevant features from the images and subsequently use machine learning algorithms on them to improve classification accuracy.

Acknowledgments

This research was partially supported by NSF award IIP1439718 and CPRIT award RP150164. We would like to thank Dr.Lan Ma, University of Maryland and Dr.Riccardo Ziraldo, University of Texas at Dallas for their helpful discussions. We also would like to thank John-Paul Bach and Sammy Glick from UT Southwestern Medical Center for their help with datasets.

References

1. J. D. Webster, A. M. Michalowski, J. E. Dwyer, K. N. Corps, B.-R. Wei, T. Juopperi, S. B. Hoover, R. M. Simpson *et al.*, *Journal of pathology informatics* **3**, p. 18 (2012).
2. J. I. Zhongwu Wang, John R. Jensen, *Environmental Modelling and Software* **25**, 1149 (October 2010).
3. R. Harrabi and E. B. Braiek, *EURASIP Journal on Image and Video Processing* **2012**, 1 (2012).
4. G. Ottaviani and N. Jaffe, *The Epidemiology of Osteosarcoma*, in *Pediatric and Adolescent Osteosarcoma*, eds. N. Jaffe, S. O. Bruland and S. Bielack (Springer US, Boston, MA, 2010), Boston, MA, pp. 3–13.
5. A. H. Fischer, K. A. Jacobson, J. Rose and R. Zeller, *Cold Spring Harbor Protocols* **2008**, pdb (2008).
6. Aperio svs tiff format `http://openslide.org/formats/aperio/`.
7. M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot and B. Yener, *IEEE Reviews in Biomedical Engineering* **2**, 147 (2009).
8. T. H. S. Sonal Kothari, John H Phan and M. D. Wang, *J Am Med Inform Assoc.* **20**, 1099 (November 2013).
9. A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot and A. Madabhushi, *IEEE Transactions on Biomedical Engineering* **57**, 642 (March 2010).
10. T. J. Fuchs and J. M. Buhmann, *Computerized Medical Imaging and Graphics* **35**, 515 (2011).
11. K.-H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin and M. Snyder, *Nature Communications* **7** (2016).
12. K. L. Weind, C. F. Maier, B. K. Rutt and M. Moussa, Invasive carcinomas and fibroadenomas of the breast: comparison of microvessel distributions–implications for imaging modalities. *Radiology* **208**1998. PMID: 9680579.
13. T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. van der Meer, H. van der Werff, F. van Coillie *et al.*, *ISPRS Journal of Photogrammetry and Remote Sensing* **87**, 180 (2014).
14. S. P. Brumby, N. R. Harvey, S. J. Perkins, R. B. Porter, J. J. Szymanski, J. P. Theiler and J. J. Bloch, Genetic algorithm for combining new and existing image processing tools for multispectral imagery, in *AeroSense 2000*,
15. A. Goode, B. Gilbert, J. Harkes, D. Jukic, M. Satyanarayanan *et al.*, Openslide: A vendor-neutral software foundation for digital pathology *Journal of pathology informatics* **4** (Medknow Publications, 2013).
16. K. Martinez and J. Cupitt, Vips-a highly tuned image processing software architecture, in *IEEE International Conference on Image Processing 2005*, 2005.
17. F. Wang, T. W. Oh, C. Vergara-Niedermayr, T. Kurc and J. Saltz, Managing and querying whole slide images, in *SPIE Medical Imaging*, 2012.
18. M. D. Abramoff, P. J. Magalhaes and S. J. Ram, *Biophotonics international* **11**, 36 (2004).
19. T. R. Jones, I. H. Kang, D. B. Wheeler, R. A. Lindquist, A. Papallo, D. M. Sabatini, P. Golland and A. E. Carpenter, *BMC bioinformatics* **9**, p. 1 (2008).