Prognostic Reporting of p53 Expression by Image Analysis in Glioblastoma Patients: Detection and Classification

[1]Mohammad F. Ahmad Fauzi, [2]Hamza N. Gokozan, [2,3]Christopher R. Pierson, [2]Jose J. Otero, [4]Metin N. Gurcan

[1]Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia
faizal1@mmu.edu.my
[2]Department of Pathology, The Ohio State University, Columbus, Ohio, USA
{Hamza.Gokozan,Jose.Otero}@osumc.edu
[3]Nationwide Children's Hospital Department of Pathology and Laboratory Medicine, Columbus, Ohio, USA
Christopher.Pierson@nationwidechildrens.org
[4]Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA
Metin.Gurcan@osumc.edu

Abstract. In this paper, we present a computer aided diagnosis system focusing on one important diagnostic branchpoint in clinical decision-making: prognostic reporting of p53 expression in glioblastoma patients. Studies in other tumor paradigms have shown that the staining intensity correlates with TP53 mutation status, and that gliomas show inter-tumoral heterogeneity in p53 mutation status. Increasing diagnostic accuracy by computer-aided image analysis algorithms would deliver an objective assessment of such prognostic biomarkers. We proposed a method for the detection and classification of positive and negative cells in digitized p53-stained images by means of a novel adaptive thresholding for the detection, and two-step rule based on weighted color and intensity for the classification. The proposed thresholding technique is able to correctly locate both positive and negative cells by effectively addressing the closely connected cells problem, and records a promising 85% average precision and 88% average recall rate. On the other hand, the proposed two-step rule achieves 81% classification accuracy, which is comparable with neuropathologists' markings.

Keywords: Cell detection, cell classification, digital pathology, p53 analysis

1 Introduction

In neurosurgery, if a decision is made to biopsy or excise an intracranial mass, the surgeon will often opt to submit a specimen to pathology for intraoperative consultation where a cytologic prep (smear) and/or frozen section is performed. Additional tissue is then submitted for formalin fixation and paraffin embedding (FFPE) where

pathologists report the tumor type, WHO grade, and additional prognostic markers such as Ki67 labeling index and p53 expression status.

Tumor protein p53 in humans is encoded by the TP53 gene. The p53 protein is crucial in multicellular organisms, where it regulates the cell cycle and functions as a tumor suppressor, preventing cancer. Studies in other tumor paradigms have shown that the staining intensity correlates with TP53 mutation status [1]. Furthermore, gliomas show inter-tumoral heterogeneity in TP53 mutation status [2].

Reporting p53 expression as a proxy for TP53 mutation status is highly subjective. Therefore, our goal is to develop image analysis algorithms to accurately and objectively quantify immunohistochemical p53, hence improving diagnostic reporting as well as clinical decision-making. These algorithms will aid and supplement pathological interpretation.

In this paper, we report the results of our preliminary work in classifying positively and negatively-stained cells in p53 immunohistochemical staining. Histopathological analysis of tissues has been gaining a lot of interest and is being applied to many different diseases such as follicular lymphoma [3-11], neuroblastoma [11-18] and glioblastoma [19]. Nevertheless, to the best of our knowledge, there has been no previous work in analyzing negative and positive staining intensities for reporting p53 expression.

The paper is organized as follows. Section 2 describes the methodology of the work where the ground truth generation and the proposed cell detection and classification processes are described. Section 3 discusses experimental evaluations of the proposed system against the ground truth. Finally conclusion and future works are discussed in Section 4.

2 Methodology

For the analysis of p53 immunohistochemical staining, slides were scanned at 40x magnification using a high-resolution whole slide scanner Aperio (Vista, CA) ScanScope™ at the resolution of 0.23 micrometer/pixel. The digitized slides were then transferred to a server where a web-based software is in place for the pathologists to independently mark the cells to determine the ground truth.

Retrospective cases were evaluated from the Pathology Tissue Archives of The Ohio State Unviersity Wexner Medical Center using an IRB approved protocol. Several cases of high grade glioma were procured. Inclusion criteria included the need for an unequivocal diagnosis and grade (confirmed by the neuropathologist) that had undergone p53 immunohistochemistry.

2.1 Ground Truth Generation

To determine the ground truth for p53 staining, two board-certified neuropathologists and one pathology research fellow reviewed several scanned/digitized slides immunostained with p53. First, to determine inter-observer variability, the three pathologists scored separately the status of each cell in a high-power field (HPF) --at

40x magnification-- as “negative stain,” “weak intensity,” “moderate intensity,” or “strong intensity.” Then, using a similar grading scheme, one of the pathologists proceeded to mark another five HPFs from five different cases. In this preliminary work, the objective is to classify the positively and negatively-stained cells, hence the “weak intensity”, “moderate intensity” and “strong intensity” marked cells are grouped together as positively-stained cells.

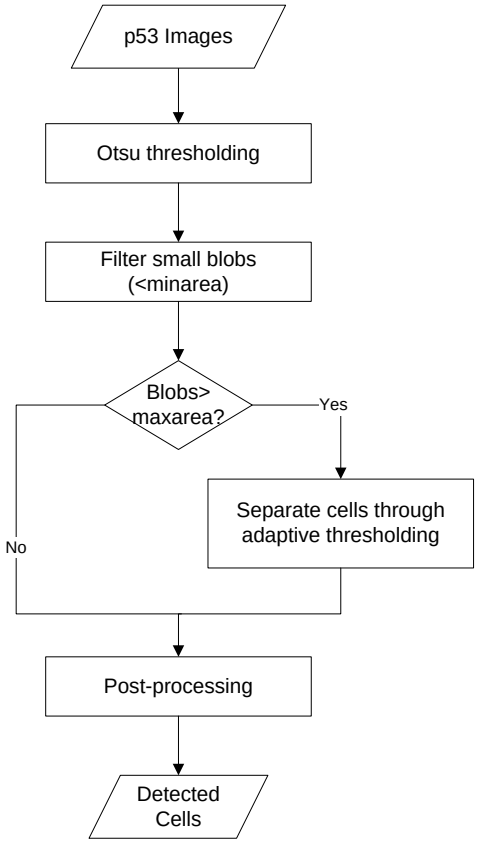

Fig. 1. Cells detection process

2.2 Cells Detection

Fig. 1 shows the flowchart of the proposed cell detection process. Given a digitized p53 stained image, we first convert the image into gray scale by only considering the luminance channel. Otsu thresholding is then applied to convert the luminance image into binary, separating the cells (blobs) from the background. The binary image is then cleaned by filtering very small and isolated blobs (blobs smaller than a certain minimum threshold, minarea). While Otsu thresholding is a very simple and straightforward technique in detecting dark objects from a light background, due to the lack

of spatial information, it is prone to grouping closely connected cells together as a single cell, which will affect the cell counting process significantly.

To address this, we propose a novel adaptive thresholding scheme for any detected blobs with total area greater than a particular maximum threshold maxarea. The adaptive thresholding checks if a large detected blob consists of multiple cells by searching for potential valley within the blob, and proceeds to segment the blobs further if they do. For a 40x magnification image used in our experiment, a suitable minarea and maxarea threshold was found to be 300 and 1000 pixels (~70 and 230 μm2), respectively.

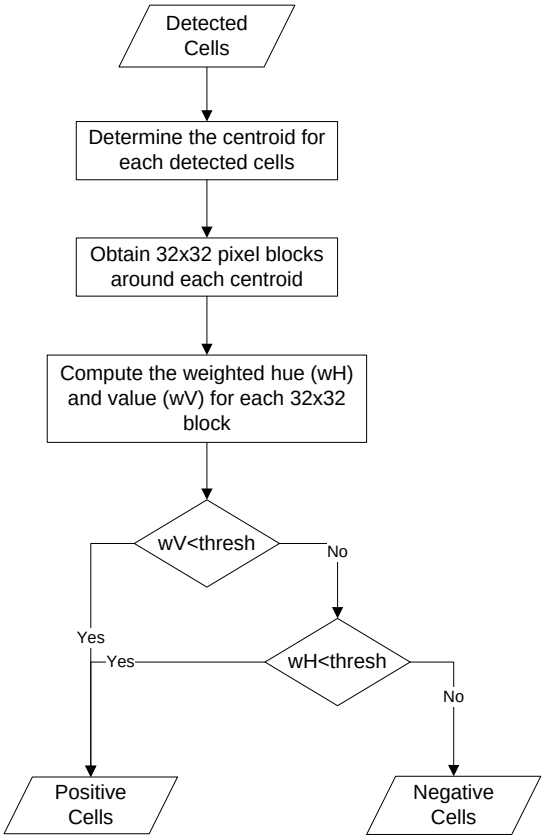

Fig. 2. Positive-negative cells classification process

2.3 Cells Classification

Fig. 2 shows the flowchart of the proposed positive-negative cell classification process. The classification of the cells into positive or negative will be based on the intensity and color of the cells, hence the image is first converted into HSV (Hue-Saturation-Value) color model. For each of the detected cells found in the previous

section, their centroid is determined and 32x32 pixel blocks are extracted around each centroid. The weighted Hue and Value are calculated for each block and these values are used for classifying the cells. The weights used are inversely proportional to the pixels' distance to the centroid, with those closer to the center of the block receiving higher weight, and those further from the center receiving less weight.

Negative p53-stained cells tend to be blue (higher Hue) and less intense (higher Value), while positive p53-stained cells tend to be brown (lower Hue) with varying intensities. Based on these properties, we propose a two-step classification rule:

1. If the weighted Value (wV) for a block is less than a particular threshold (darker), the block will be classified as containing positive cell, regardless of its weighted Hue (wH).
2. Otherwise, the classification depends on weighted Hue, with wH less than a particular threshold means the block contains positive cells, and wH more than the threshold means it contains negative cells.

From experiment, suitable threshold value for both wH and wV is found to be 70.

3 Experimental Results

We first analyze the inter-observer variability during ground truth generation before discussing the performance of the proposed cell detection and classification methods against the ground truths.

3.1 Inter-Observer Variability

Table 1 shows the summary of cell marking between the three pathologists. Overall the three pathologists are all able to detect more than 1000 cells within the image, although the classification into positive and negative cells differs quite significantly. Table 2 further examines the pathologist agreement by looking at their classification of each cell. Each column gives the number of cells where the pathologists are in agreement with each other. For example, in the first row, pathologists A and B gave the same classification for 966 cells (207 for positive cells, and 759 for negative cells), which is equivalent to 83.8% of the 1153 total cell marked. From the table, we can see that the agreements between any two pathologists are between 78.8% and 83.8%, which drops to 70.6% when all three pathologists are considered. This shows that even expert pathologists differs quite significantly in classifying positive and negative cells in digitized p53-stained images.

Table 1. Positive and Negative Cell Counts by Different Pathologists

	Positive	Negative	Total
Pathologist A	781	325	1106
Pathologist B	868	285	1153
Pathologist C	720	378	1098

Table 2. Marking Agreement Between Pathologists

	Positive	Negative	Total	Pctg
A&B	207	759	966	83.8
A&C	235	673	908	78.8
B&C	220	691	911	79.0
A&B&C	173	641	814	70.6

3.2 Cells Detection

Table 3 summarizes the performance of the proposed cells detection technique. 'Cases' refers to the different digitized tissue, 'GT' is the ground truth by pathologist A, B or C, 'TP' is the true positive (correctly detected cells), 'FP' is the false positive (incorrectly detected cells) and 'FN' is the false negative (missing or undetected cells). Precision and recall are defined by:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

The proposed detection method recorded 85% precision and 88% recall rate. The recall rate in particular is very promising as all cases recorded more than 80% rate. This suggests that the proposed system is able to detect as much cells as identified by the pathologists. The precision rate on the other hand is slightly lower, suffering from slightly high false positive cases. Some of this is also caused by the pathologists missing some cells when marking the ground truth. Note that there could be more than 1000 cells in one viewed field causing the pathologists to unintentionally miss some of the cells.

Table 3. Performance of the Proposed Cell Detection Technique

Cases	GT	TP	FP	FN	Precision	Recall
1	A	844	188	32	0.96	0.82
1	B	878	200	49	0.95	0.81
1	C	852	187	49	0.95	0.82
2	A	420	77	43	0.91	0.85
3	A	221	12	73	0.75	0.95
4	A	257	15	30	0.90	0.94
5	A	185	9	82	0.69	0.95
6	A	172	23	69	0.71	0.88
Average					**0.85**	**0.88**

3.3 Cells Classification

Table 4 shows the classification accuracy for the six images. The classification accuracy is defined by:

$$Accuracy = \frac{Correctly\ classified\ cells}{Total\ number\ of\ cells}$$

As seen from the table, the accuracy varies between 0.76 and 0.95, with an average of 0.81. Considering the agreement between any two pathologist is between 78.8% and 83.8%, and between all three pathologists is 70.6%, the results obtained is very promising, suggesting that the proposed system measures up well with the given task. For case 1, we can also compare how the system measures against the different ground truth of the three pathologists, and in this particular case, it tends to agree more with pathologist C. Fig. 3 shows one example (Case 4) of the detection and classification against the available ground truth. The positive and negative cells identified by the system are marked with green and blue squares respectively. For reference purposes, the positive and negative cells identified by the pathologist are marked with green and blue asterisks respectively.

Table 4. Summary of Classification Accuracy

Cases	GT	Correctly Classified	Incorrectly Classified	Accuracy
1	A	652	192	0.77
1	B	681	197	0.78
1	C	691	161	0.81
2	A	224	96	0.77
3	A	193	28	0.87
4	A	209	48	0.81
5	A	175	10	0.95
6	A	130	42	0.76
Average				**0.81**

4 Conclusion

We have proposed a method for the detection and classification of positive and negative cells in digitized p53-stained images by means of Otsu and adaptive thresholding for the detection, and two-step rule based on weighted color and intensity for the classification. The proposed method is able to achieve 85% average precision and 88% average recall rate for cells detection, and 81% accuracy in classifying them into positively or negatively stained cells.

Considering the confusing nature of differentiating weakly stained positive cells from certain negative cells and even non-cells, and the fact that inter-observer variability is rather high (only 70% agreement among three pathologist), the proposed system provides promising results for detection and classification tasks. We are cur-

rently working on further classifying the positively-stained cells into different staining strength (weak, moderate, strong), which is even more challenging as the inter-observer variability is much higher.

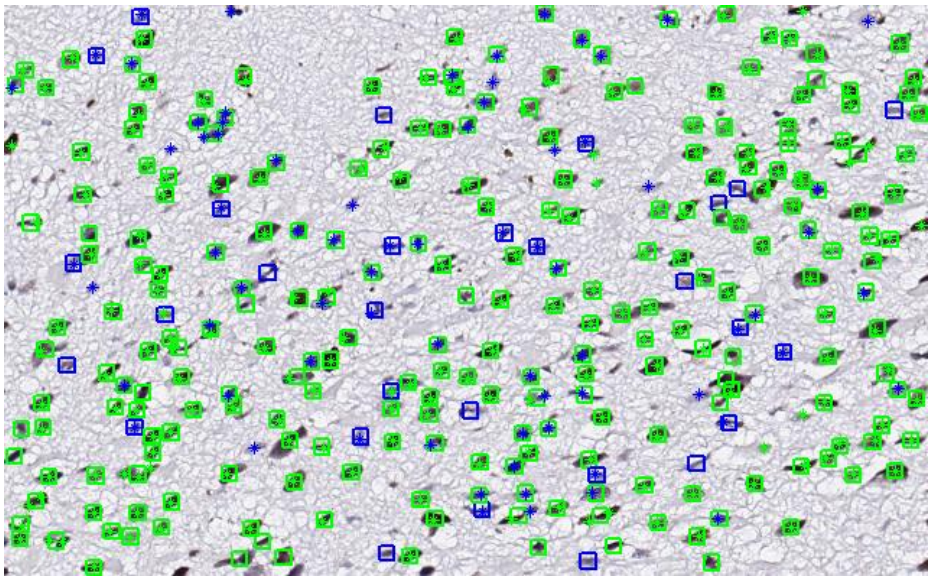

Fig. 3. Example of detection and classification by the system against the available ground truth (Case 4), where the positive and negative cells identified by the system are marked with green and blue squares respectively. The positive and negative cells identified by the pathologist are marked with green and blue asterisks respectively.

References

1. A. Yemelyanova, R. Vang, M. Kshirsagar, D. Lu, M.A. Marks, I.M. Shih, R.J. Kurman, "Immunohistochemical staining patterns of p53 can serve as a surrogate marker for TP53 mutations in ovarian carcinoma: an immunohistochemical and nucleotide sequencing analysis," Mod Pathol 24:1248-1253 (2011).
2. Z.P. Ren, T. Olofsson, M. Qu, G. Hesselager, T. Soussi, H. Kalimo, A. Smits, M. Nister, "Molecular genetic analysis of p53 intratumoral heterogeneity in human astrocytic brain tumors," J Neuropathol Exp Neurol 66:944-954 (2007).
3. O. Sertel, J. Kong, U. Catalyurek, G. Lozanski, J. Saltz, and M. Gurcan, "Histopathological Image Analysis using Model-based Intermediate Representations and Color Texture: Follicular Lymphoma Grading," Journal of Signal Processing Systems 55, 169-183 (2009).
4. O. Sertel, G. Lozanski, A. Shana'ah, and M.N Gurcan, "Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation," IEEE Trans Biomed Eng 57(10), 2613-6 (2010).
5. O. Sertel, J. Kong, G. Lozanski, A. Shanaah, A. Gewirtz, F. Racke, J. Zhao, U. Catalyurek, J.H. Saltz, and M. Gurcan, "Computer-assisted grading of follicular lymphoma: High grade differentiation," Modern Pathology 21, 371A-371A (2008).

6. O. Sertel, J. Kong, G. Lozanski, U. Catalyurek, J.H. Saltz, and M.N. Gurcan, "Computerized microscopic image analysis of follicular lymphoma," Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis 6915, 1-11 (2008).
7. S.S. Samsi, A.K. Krishnamurthy, M. Groseclose, R.M. Caprioli, G. Lozanski, and M.N. Gurcan, "Imaging mass spectrometry analysis for follicular lymphoma grading," Proceedings of Annual Int Conf of the IEEE Engineering in Medicine and Biology Society, 6969-6972 (2009).
8. S. Samsi, G. Lozanski, A. Shana'ah, A.K. Krishanmurthy, and M.N Gurcan, "Detection of follicles from IHC-stained slides of follicular lymphoma using iterative watershed," IEEE Trans Biomed Eng 57(10), 2609-2612 (2010).
9. M. Oger, P. Belhomme, and M.N. Gurcan, "A general framework for the segmentation of follicular lymphoma virtual slides," Comput Med Imaging Graph 36(6), 442-51 (2012).
10. K. Belkacem-Boussaid, O. Sertel, G. Lozanski, A. Shana'aah, and M. Gurcan, "Extraction of color features in the spectral domain to recognize centroblasts in histopathology," Proceedings of Annual Int Conf of the IEEE Engineering in Medicine and Biology Society, 3685-3688 (2009).
11. H.C. Akakin and M.N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," IEEE Trans Inf Technol Biomed 16(4), 758-769 (2012).
12. G. Teodoro, R. Sachetto, O. Sertel, M.N. Gurcan, W. Meira, U. Catalyurek, and R. Ferreira, "Coordinating the Use of GPU and CPU for Improving Performance of Compute Intensive Applications," Proceedings of IEEE Int Conf on Cluster Computing and Workshops, 437-446 (2009).
13. O. Sertel, J. Kong, H. Shimada, U.V. Catalyurek, J.H. Saltz, and M.N. Gurcan, "Computer-aided Prognosis of Neuroblastoma on Whole-slide Images: Classification of Stromal Development," Pattern Recognition 42(6), 1093-1103 (2009).
14. A. Ruiz, O. Sertel, M. Ujaldon, U.V. Catalyurek, J. Saltz, and M.N. Gurcan, "Stroma classification for neuroblastoma on graphics processors," International Journal of Data Mining and Bioinformatics 3(3), 280-298 (2009).
15. A. Ruiz, J. Kong, M. Ujaldon, K. Boyer, J. Saltz, and M.N Gurcan, "Pathological image segmentation for neuroblastoma using the GPU," Proceedings of 5th IEEE Int Symp on Biomedical Imaging: From Nano to Macro, 296-299 (2008).
16. M. Gurcan, T. Pan, H. Shimada, and J.H Saltz, "Image analysis for neuroblastoma classification: Hysteresis thresholding for cell segmentation," Proceedings of APIII, Vancouver, BC (2006).
17. B. Cambazoglu, O. Sertel, J. Kong, J.H. Saltz, M.N. Gurcan, and U.V. Catalyurek, "Efficient processing of pathological images using the grid: Computer-aided prognosis of neuroblastoma," Proceedings of Challenges of Large Scale Applications in Distributed Environments (CLADE), Monterey Bay, CA., 35-41 (2007).
18. O. Sertel, U.V. Catalyurek, H. Shimada, and M.N. Gurcan, "A Combined Computerized Classification System for Whole-slide Neuroblastoma Histology: Model-based Structural Features," International Conference on Medical Image Computing and Computer Assisted Intervention, 7-18 (2009).
19. M.F.A Fauzi, H.N. Gokozan, B. Elder, V.K. Puduvalli, J.J. Otero, M.N. Gurcan, "Classification of Glioblastoma and Metastasis for Neuropathology Intraoperative Diagnosis: A Multi-resolution Textural Approach to Model the Background", Proceedings of SPIE Medical Imaging 2014: Digital Pathology (2014).