Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Haoyu Zhang Date

Invasive Front Analysis of Brain Tumors

By
Haoyu Zhang
Master of Science

Physics

_____

(Fereydoon Family)
Advisor

_____

(Yi Jiang)
Committee Member

_____

(Connie Roth)
Committee Member

_____

(H. George E. Hentschel)
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Invasive Front Analysis of Brain Tumors

By

Haoyu Zhang

B.S., University of Science and Technology of China, 2013

Advisor: Fereydoon Family

An abstract of

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Physics

2015

Abstract

Invasive Front Analysis of Brain Tumors

By

Haoyu Zhang

Glioma, a tumor that starts in the brain or spine, makes up about 30% of all tumors of the central nervous system and 80% of all malignant brain tumors. The mutations in isocitrate dehydrogenase 1 (IDH1) was found as a novel biomarker of gliomas. Sections of patient tumors of gliomas grades II-IV, stained with an antibody specific to the mutated IDH1 protein, were contributed by Andreas von Diemling and colleagues. We analyzed the glioma cell distributions in these slides and compared the cell distributions in the invasive front with different tumor types and grades. Our approach provides some basic understanding of the glioma tumor patterns and their invasive front, which will potentially help provide better therapy for patients, especially during surgery.

Invasive Front Analysis of Brain Tumors

By

Haoyu Zhang

B.S., University of Science and Technology of China, 2013

Advisor: Fereydoon Family

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Physics

2015

Contents

List of Figures

Chapter 1

Introduction

1.1 Glioma

Glioma is a tumor that starts at the brain or spine. It makes up about 30% of all brain and central nervous system tumors and 80% of all malignant brain tumors [1]. The speed of tumor growth increases with grades rising from I to IV, and the survival rate of patients decreases sharply as the grade increases. Malignant gliomas, the most common subtype of primary brain tumors, are aggressive, highly invasive, and neurologically destructive tumors considered to be among the deadliest of human cancers. In its most aggressive manifestation, glioblastoma (GBM), median survival time ranges from 9 to 12 months despite maximum treatment efforts—a statistical fact that has changed little over several decades of technological advances in neurosurgery, radiation therapy, and clinical trials of conventional and novel therapeutics. Gliomas have been defined pathologically as tumors that display histological, immunohistochemical, and ultrastructural evidence of glial differentiation. The most widely used scheme for classification and grading of gliomas is that of the World Health Organization (WHO). Gliomas are classified according to their hypothesized line of differentiation, that is, whether they display features of astrocytic,

oligodendroglial, or ependymal cells. They are called astrocytoma, and oligodendrocytoma, respectively. They are then graded on a scale of I to IV according to their degree of malignancy as judged by various histological features [2].

1.2 Approaches for distinguishing gliomas

The most common imaging approach for distinguishing gliomas are radiological methods, such as magnetic resonance imaging (MRI) and CAT scan, often with the help of biomarkers. The resolution for MRI, however, is limited, suffering from a poor tumor-edema differentiation [3]. Unlike other solid tumors, gliomas, though never metastasize outside the central nervous system, are intrinsically invasive. The tumor cells migrate and invade the surrounding tissue. Instead of a sharp boundary between tumor and surrounding tissue, in most primary solid tumors, gliomas have a diffuse interface between tumor and healthy tissue, featuring a varying tumor cell density [1]. This diffusive front is the main culprit of MRI resolution limit [4].

The mutations in isocitrate dehydrogenase (IDH) 1 and 2, which are enzymes that are encoded by the IDH1 gene, were first discovered in 2009 and became one of the most novel biomarkers for gliomas [7]. It occurs in up to 75% of low grade gliomas and secondary high grade gliomas, and more than 90% are mutations of the R132H type [7]. Clinical relevance and the highly homogeneous pattern of IDH1 mutations make it possible to develop a mutation-specific antibody. In previous work, researchers have already successfully produced a mouse monoclonal antibody

targeting the IDH1 R132H mutation, which could be used to detect the IDH1 mutations in gliomas [7]. Compared with other immunohistological approaches, the IDH1 clearly identifies glioma cells [7].

1.3 Reaction-Diffusion model for tumor growth

Building mathematical models is very important in biological research. The mathematical model we use is based on a well-known differential equation [8]:

$$\frac{\partial \mathrm{c}}{\partial \mathrm{t}} = \nabla \cdot (D \nabla c) + \rho c \ , \qquad (1)$$

where c(x,y,t) is the tumor cell density at location (x,y) at time t, ρ is the net proliferation rate, D is the diffusion coefficient which represents the degree of motility of glioma cells. Swanson et al. have used this model with realistic brain anatomy to guide brain tumor surgery [8]. The model is based on two assumptions: (1) Diffusion in tumors is Fickian diffusion and (2) the tissue environment surrounding the tumor is homogeneous so that the diffusion coefficient D is constant and uniform in the brain.

An interesting feature of the model is that the density of the tumor cells depends on the ratio ρ/D of growth rate ρ to the diffusion coefficient D. When ρ and D vary with a fixed ρ/D, the geometry of the tumor growth and invasion remain the same. And when ρ/D becomes large the cell density profile becomes steeper. Figure 5 shows a comparison of the cell density profile for tumors with high and low ρ/D. In our raw data, all pictures are taken at a fixed time, so we don't know how a given glioma behaves at different times. However, the density of each glioma could be

calculated, and using the relationship between ρ/D and density profiles we could compare features of different gliomas.

1.4 Questions and Motivation

The main question is how does the invasive front changes as tumors progress and how the spreading of the front depends on the brain tumor type? Our goal is to answer these questions by analyzing the cell density in high resolution patient slides. We investigate how ρ/D in density profiles in the invasive front depend on the types and grades of gliomas. The difficulty is in finding a quantitative method to specify and compare the steepness of different density profiles. Although the answer seems obvious in an ideal model like Figure 1, the cell density profiles in gliomas are usually irregular and cannot be easily fit to a unique straight line.

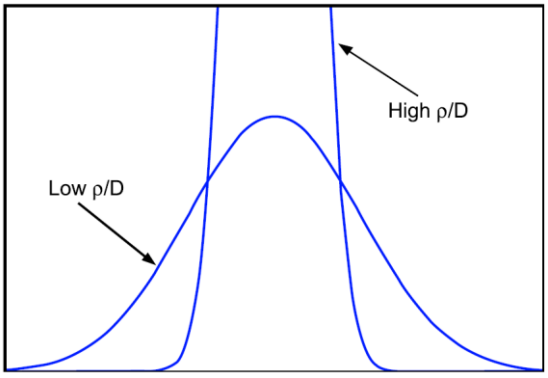


Figure 1. Cell density profiles for tumors with different ρ/D.

Applying the mathematical model to magnetic resonance image (MRI) of gliomas has been extensively studied. However, since IDH1 is a biomarker that has been

found within the last 5 years or so, few researches have applied the mathematical model directly to the gliomas stained with IDH1 antibody.

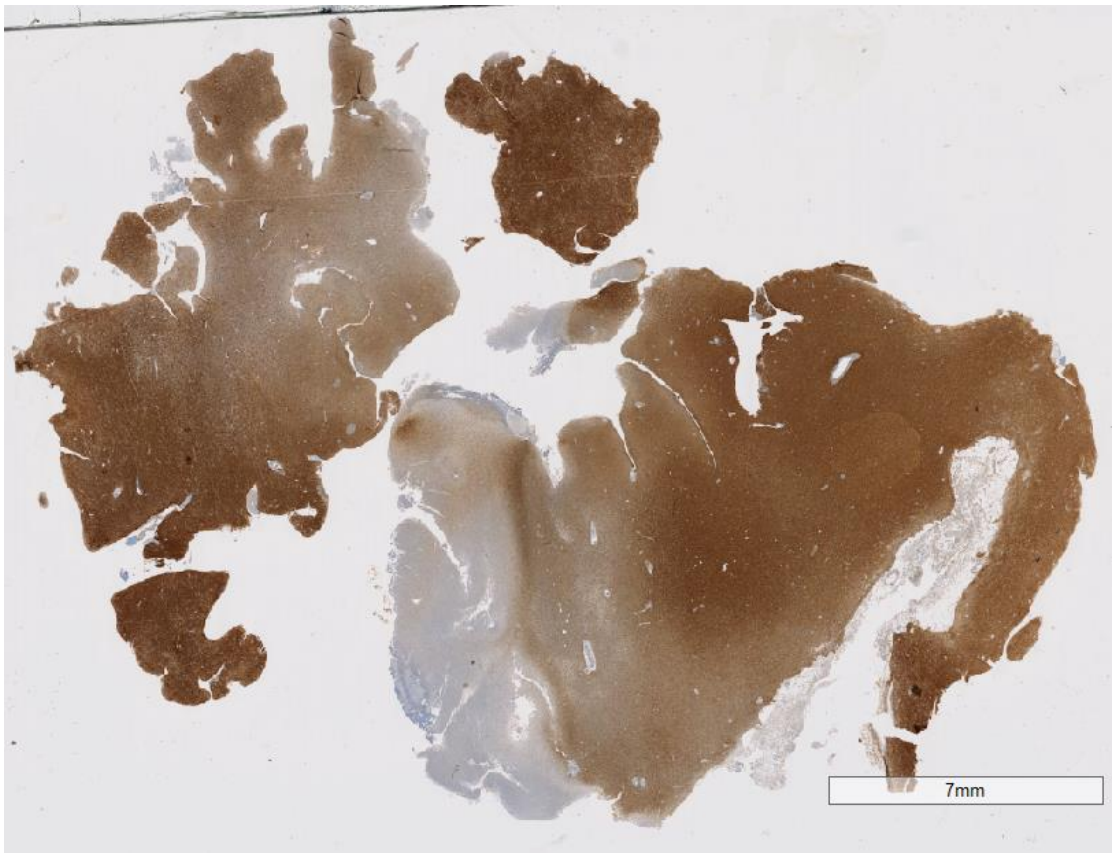
The dependence of a tumor on ρ/D is an important biomarker, especially during surgery. At low ρ/D the tumor is diffusive and cannot be completely removed during surgery. On the other hand, when ρ/D is large the tumor has a solid pattern with a sharp boundary and surgical procedures are more effective.

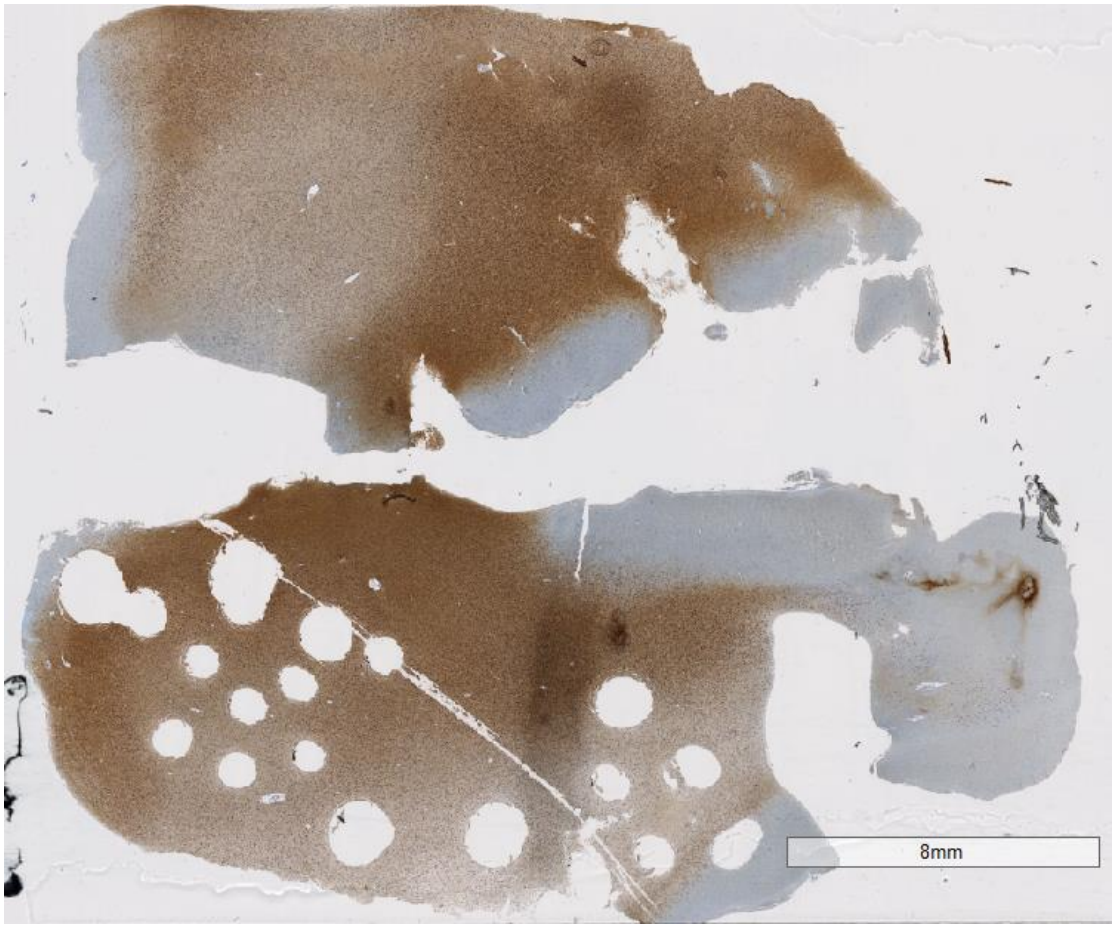
Chapter 2

Methods

2.1 Raw images

Sections of patient tumors of gliomas grades II-IV, stained with an antibody specific to the mutated IDH1 protein were contributed by Andreas von Diemling and colleagues [7]. The slides were scanned at the Winship Cancer Institute Pathology Core at 40X magnification in order to preserve the maximum amount of detail. There are 133 glioma images, each corresponding to one patient, comprising 8 different types. The types are astrocytoma grade II (AII), astrocytoma grade III (AIII), Oligoastrocytoma grade II (OAII), Oligoastrocytoma grade III (OAIII), Oligodendroglioma grade II (OII), Oligodendroglioma grade III (OIII), Glioblastoma multiforme grade (GBM) and Glioblastoma multiforme secondary (GBMII). About half of the images are low grade glioma (such as AII, OAII, OII), which are rare because low grade gliomas do not produce obvious clinical effect and are often not detected until they turn into higher grade tumors.


7mm


8mm

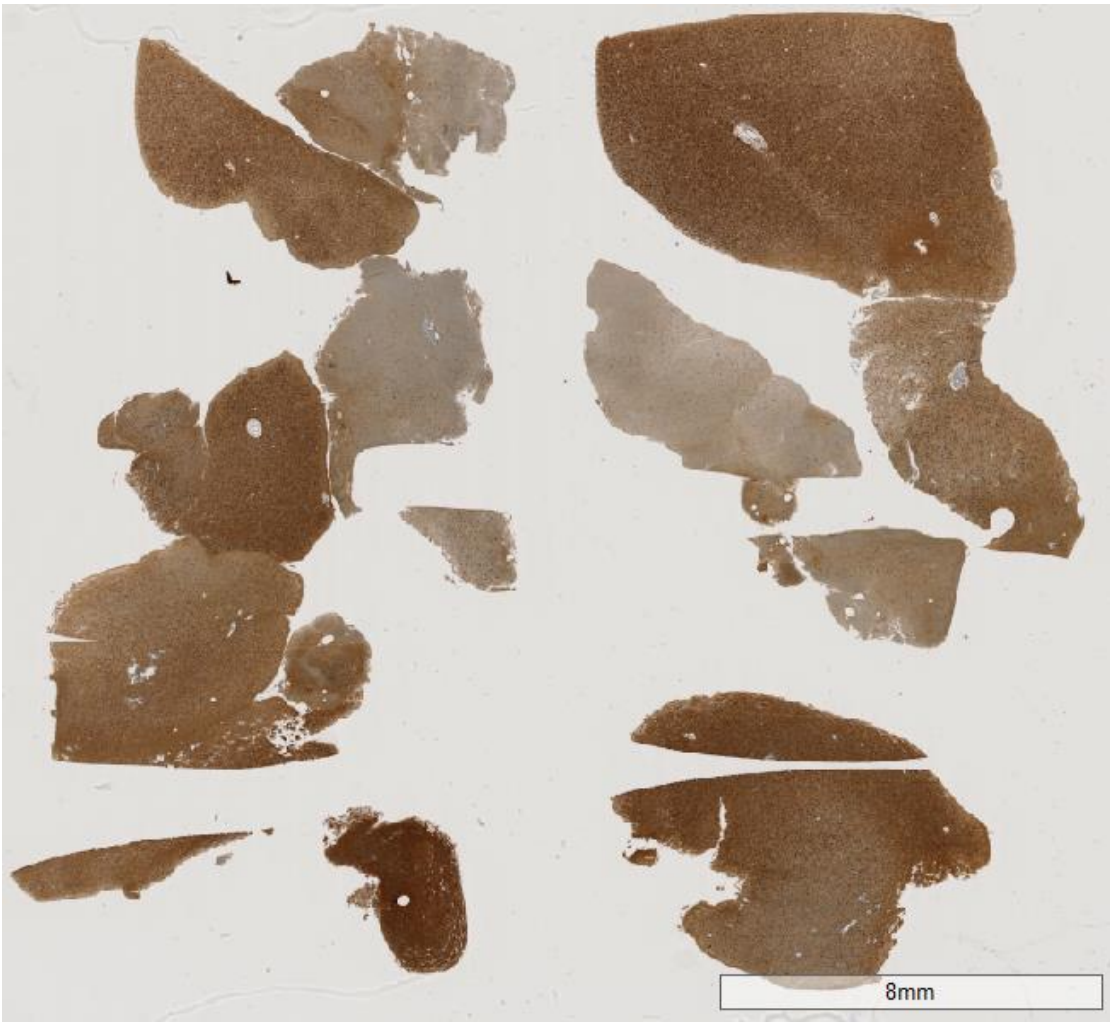


Figure 2: Examples of raw images

2.2 Cutting images

We mainly focus on the cell density distribution near the boundary of glioma (invasive front) which is the part of the glioma that connects to the healthy brain tissue. The cell density profile near the tumor boundary shows how the tumor grows and invades the healthy brain. The cell density inside the glioma is more complex and currently there are no universally accepted understanding of the variations of the density inside the glioma. Figure 3 shows the type of area we are interested in. Inside

a glioma, the cell density may go up and down multiple times. But we only consider the area inside the tumor near the boundary with the healthy brain. Near this boundary the glioma cell density is almost zero and does not across any peaks.

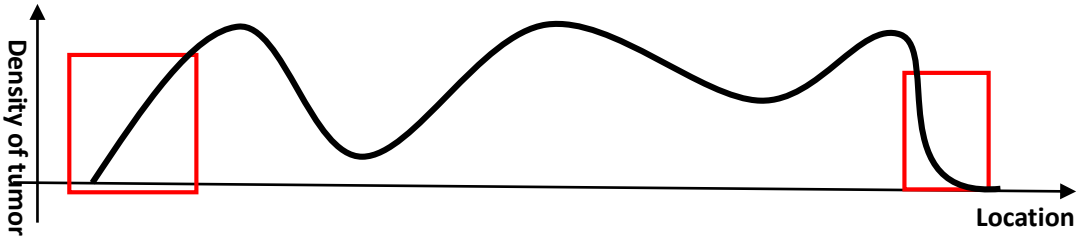


Figure 3: Density may change multiple times inside a glioma. The red rectangles show the kind of area we are interested in.

The raw images are cut into several small images due limitations in the computer memory size and processing speed. However, finding the place to cut the images could be difficult. The reason is that sometimes it is not clear where is the boundary of the glioma in IDH1 images. Figure 4 shows a comparison of an MRI and an IDH1 image of glioma. In the MRI image we can clearly identify the glioma (the white area) and the healthy brain. In IDH1 images, on the other hand, the glioma tissue may have been reshaped and separated into several cuts during surgery. Sometimes it is hard to distinguish the areas of the healthy brain versus the glioma. We have been able to overcome this problem by following a suggestion by clinical neruopathoogist Daniel J. Brat from the Winship Cancer Center. He has suggested that the best instrument for distinguishing the place to cut an image is the eye.

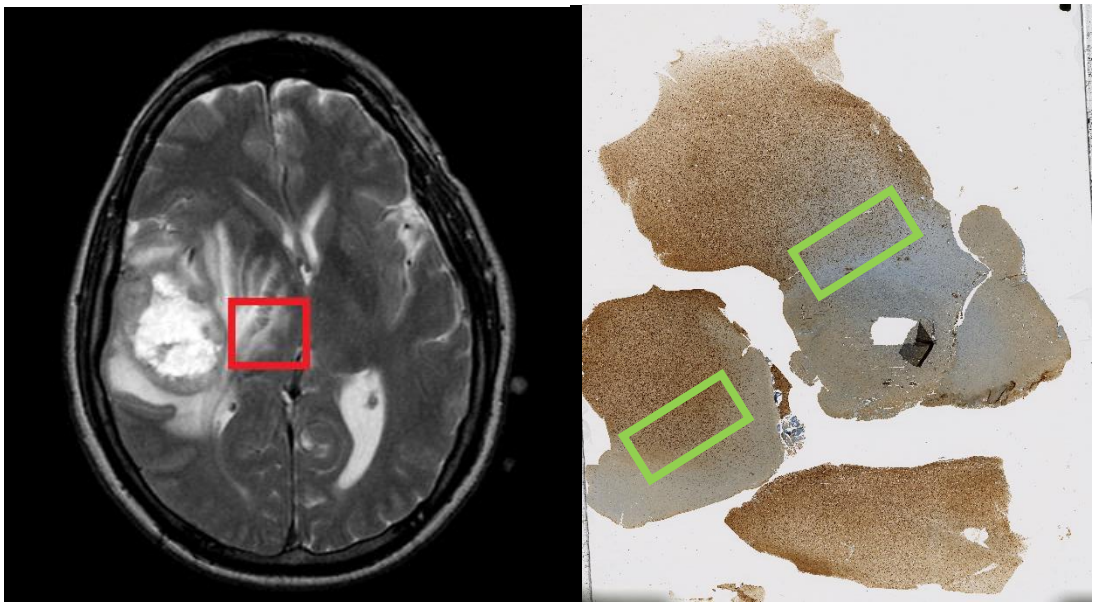


Figure 4: (a) MRI of glioma, the red rectangle shows the area we want to analyze as the invasive front of glioma. (b) Image of IDH1 stained glioma: brown is IDH1 mutant stain, indicating glioma cells, and blue indicates normal tissue. The green rectangle indicates a region that corresponds to an invasive front that we analyze.

2.3 Image Processing

We take a series of steps to transform the original images to binary images in which we use white to represent the cells and black to represent the background. At first, we convert the images to grayscale and segment them by thresholding. Most of automatic threshold determining algorithms such as (maximum likelihood algorithm) would fail because the shape and size of tumor cells, the color of background and cells vary greatly. The reason is that the glioma morphology changes with different patients (who have different types and grades), and the staining process is done manually which brings a lot of uncertainty to color. As a result, we select a threshold manually. Then, we delete small objects by removing connected objects below the threshold of 30 pixel and connectivity of 8. Figure 5 shows the result of these steps, and Figure 6 provides examples of segmentation results with the original images overlaid with the

outlines of the binary image shown in white.

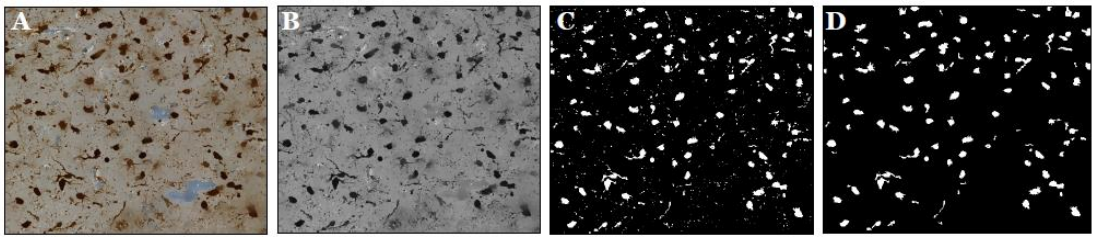


Figure 5: A. original image B. Grayscale image C. Binary image before morphological operations D. Image after morphological operations

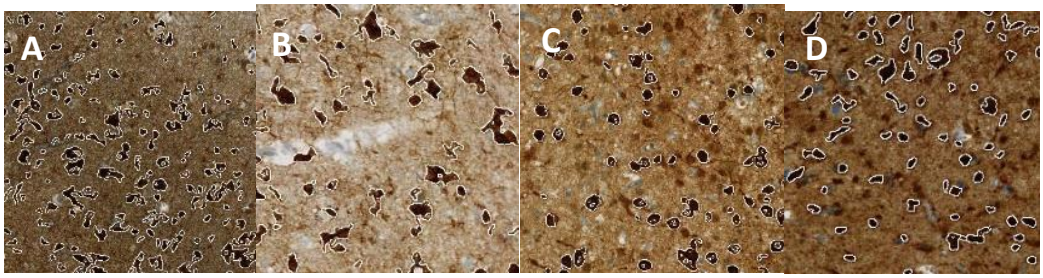


Figure 6: Examples of segmentation results with the original images overlaid with binary image outlines (white)

Converting color images to grayscale and then setting up a threshold is a common method to do the segmentation. However, it may cause a problem in our case, because it would be difficult to distinguish between the brown (tumor cells) and dark blue (background). As a result, some blue areas may be treated as cells which are actually just the background. Our solution is to search the blue background and lighten the color using the RGB values before converting the images to grayscale. After doing that, the light blue areas will be treated as background in segmentation.

We then record the location of glioma cells in the binary images. The most direct method is to calculate the center of mass of each connected area in binary images, and use that location to represent the location of the cells. However, due to the segmentation restriction, if multiple cells are close to each other, we can not distinguish them as multiple connected areas. As a result, they will be treated as one cell. Figure 7 shows this situation. If we only use a threshold to segment an image,

this problem could always happen. Some researchers try to figure out a solution for this kind of problem, but none of these methods have been applied to glioma images because glioma cells are much more complex and more varied than other cells.

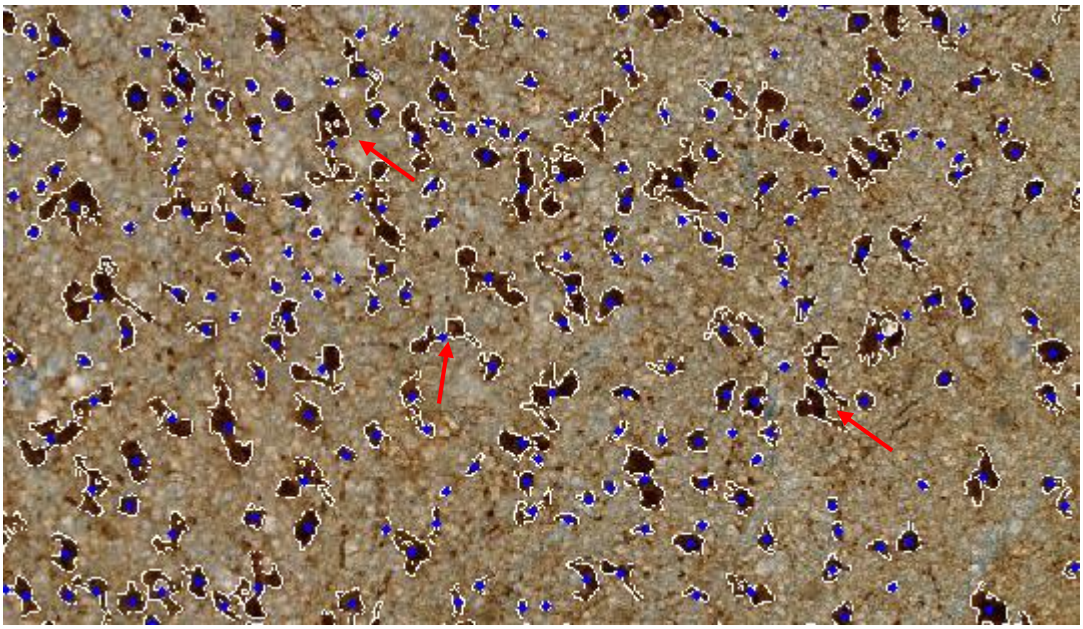


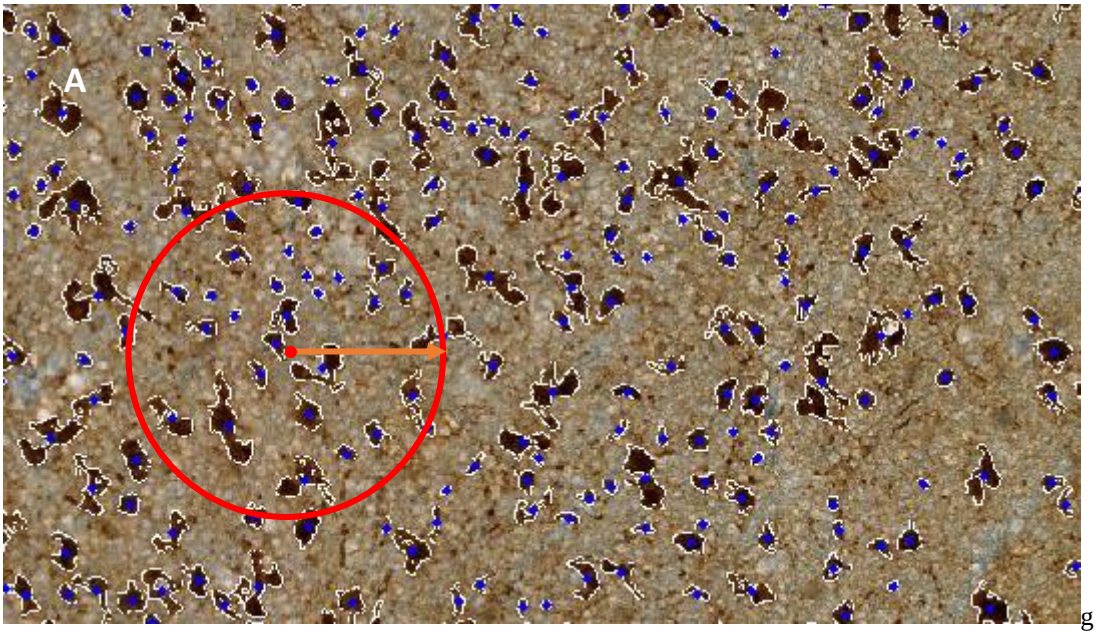
Figure 7 Blue points are the center of mass of each connected area. Each connected area (brown) will be treated as a cluster of cells (red arrow)

Our method shows not only the location of each connected area, but also how many pixels there are in each area. The basic idea is to get the density of cells in a given area. It is not necessary to know whether a pixel belongs to a particular cell or another cell, because we only need to know if it is part of the tumor and should be counted in the calculation of the density. Our method makes cell segmentation much easier, and it should provide the same result as the previous method. Furthermore, even if one could successfully separate the cells, our method would still be a valid approach, because it is the cell areas that contribute to the tumor density.

2.4 Definition and Calculation of Density

We use a box counting method to calculate the cell density. We draw a circle with a radius that we call the “boxsize”. We then determine the location of each center of mass within the circle and the total areas that belong to these centers. We add up the pixel values of all the cell areas and divide it by the area of the circle to determine the cell density. In our definition, the density is simply the fraction or the percentage of the area covered by the cells and consequently it is a dimensionless quantity.

In order to improve the efficiency of the calculations, we use the K Dimensional Tree method [15] to do the range query for finding the number of centers of mass inside a circle. This is a very efficient method, becasue f the number of center of mass is n, then the computatioal time cost for this method would be proportional to $\sqrt{n}$ instead of n [15].



g

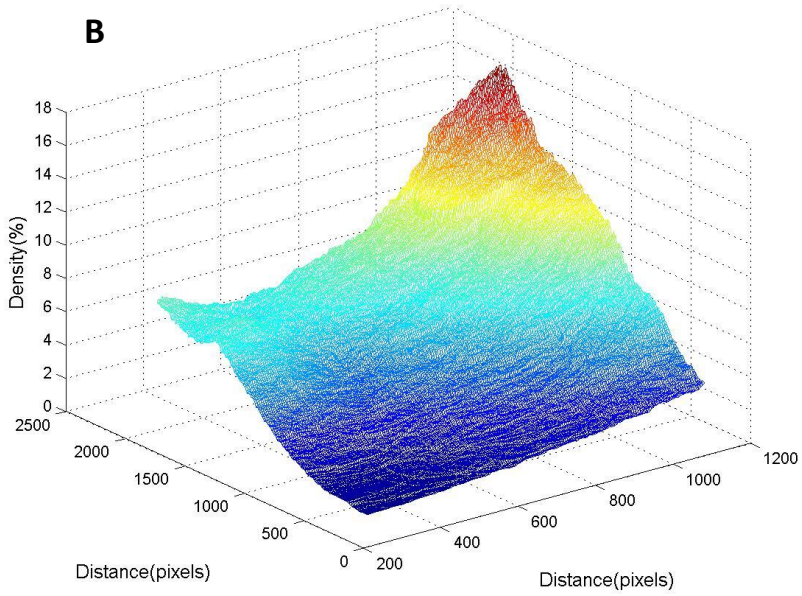


Figure 8 (A) The cell density is calculated by dividing the areas belonging to the center of mass within the circle divided by the area of the circle. The orange arrow shows the boxsize. (B) Results of density calculation as a function of x and y distance in a typical image.

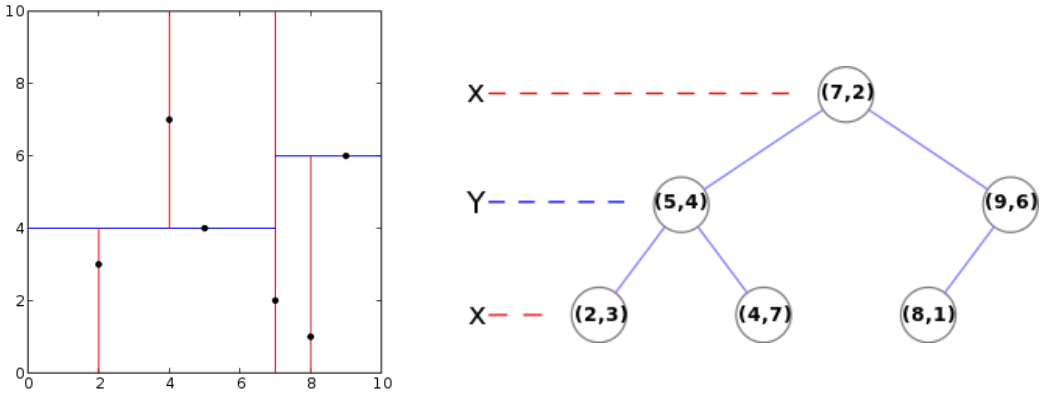


Figure 9 K Dimensional Tree

One of the drawbacks of our method is the boundary effect. If part of a cell is inside the circle but the center of mass is outside the circle, then the cell will not contribute to the density. This will both underestimate the density and also make the density discontinuous. Why don't we instead of recording the locations of the center of mass,

directly count how many pixels belong to the cells inside a circle? This seems accurate in theory, however, this idea does not work in actual calculations due to the computational restrictions. Each connected area has about 100 pixels on average, so even with the most advanced range query algorithm, it will be 10 times slower than searching for the centers of mass. Another disadvantage of these query algorithms is the cost in memory, if one tries to make the algorithm efficient. Thus, querying pixels will easily run out of memory. Although counting pixels is not a practical approach, there are still limitations to our approach due to boundary effects.

2.5 Finding the direction of diffusion

After calculating the density, we use the result to calculate the cell density profile. The cell density profile, however, are different in different directions. Clearly, our interest is in determining the direction in which the tumor grows and finding this direction is nontrivial. Figure 10(A) shows an ideal case of glioma diffusion, where the boundary of the glioma is a straight line and the regions that are the same distance from the boundary have the same cell density. In Figure 10, darker colors show regions with higher cell density. In this case, the direction of diffusion is perpendicular to the front boundary, and the cell density profile in diffuse direction has the steepest slope.

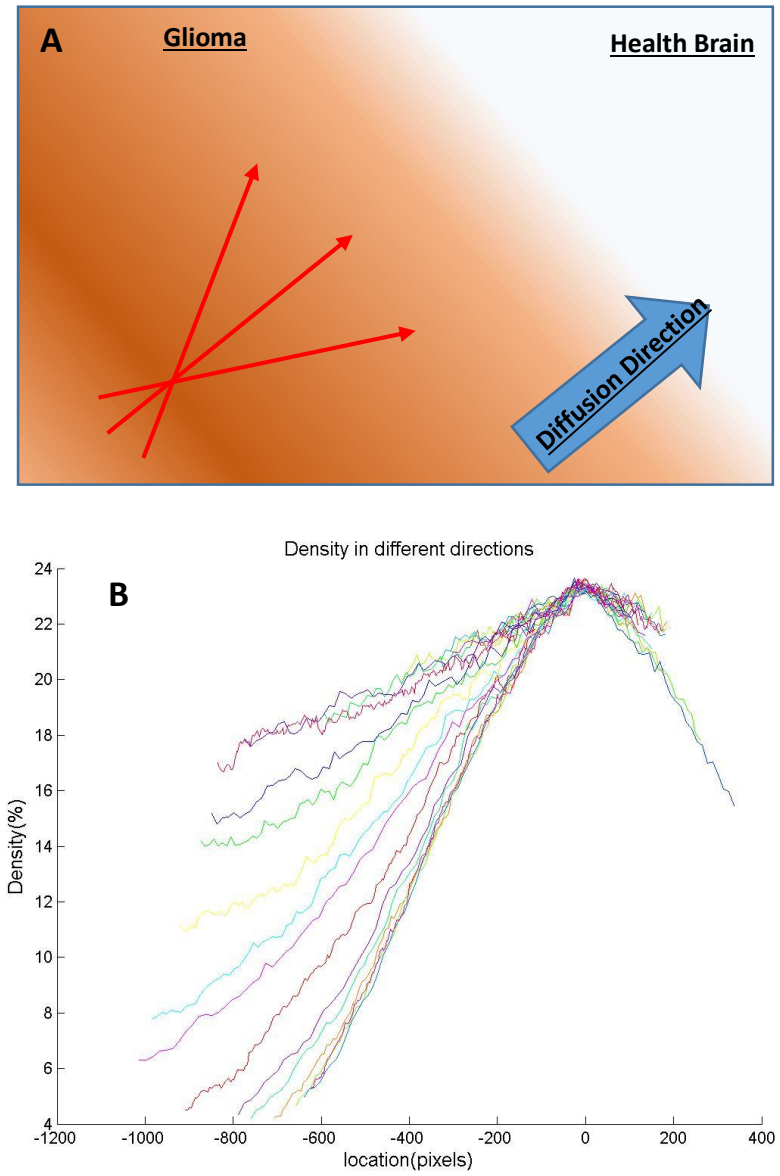


Figure 10 Finding the direction of glioma diffusion: (A) Ideal case (B) Real calculation of the cell density in different directions.

In reality, the tumor boundary is not a line, and the cell density distribution is also more complex so that the direction of the diffusion may not be very clear to the eye. To determine the diffusion direction we calculate the slope in different direction and define the diffusion direction as the profile with the steepest slope. We first determine

the “peak” of the distribution in each sample by identifying the maximum of the cell density. Then we calculate the cell density profiles of each sample along several different directions with the angle of the direction increasing by 10 degrees each time. Finally, we choose the steepest cell density profile as the density profile in the diffuse direction for the sample. The figure 10(b) shows our calculation of the direction of diffusion in an actual image.

2.6 Calculating the steepness of cell density profiles

To calculate the steepness of the cell density we assume that the density profiles to be almost linear except near the boundary. We use linear regression to determine the slope of the cell density profiles and use the slope of the fitting function to represent the steepness of the cell density profiles, as shown in Figure 11. The steepness is determined by the ratio ρ/D of the glioma density to its diffusion constant, which is related to the kinetic properties of the glioma.

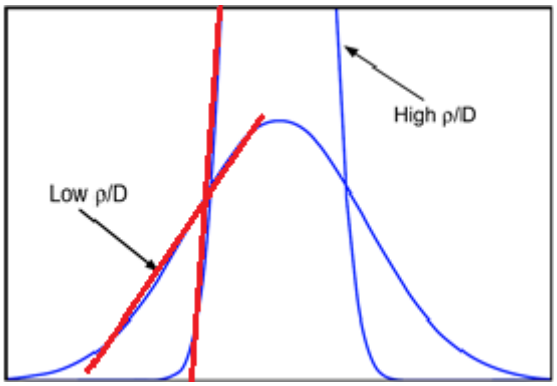


Figure 11 Calculating the steepness

2.7 Boundary effect

We purpose two methods to overcome boundary effect in our density calculations. The first approach is to optimize the boxsize selection. When a box size is small, the boundary effect is large. As the boxsize gets larger, the cell density profile becomes smother, thus reducing the boundary effect. On the other hand, if the boxsize is too large, the cell density is just equal to the average cell density and not the local density of the cells, which we are trying to calculate. As a result, selecting a proper boxsize is very important. In Figure 11, we show our approach for selecting an optimum boxsize. Figure 12(b) shows a plot of the slope of cell density profiles for a given sample vs different boxsizes. We find that when the boxsize is under 400 pixels the slopes fluctuate, but have very similar values. On the other hand, when the boxsize is larger than 400 pixels, the slope keeps decreasing. As an optimum value, we select 400 pixels as our boxsize. As can be seen from figure 12(a), when the boxsize is equal to 400 pixels, the box includes a proper amount of cells and it seems to properly reflect the cell density of the center of the box.

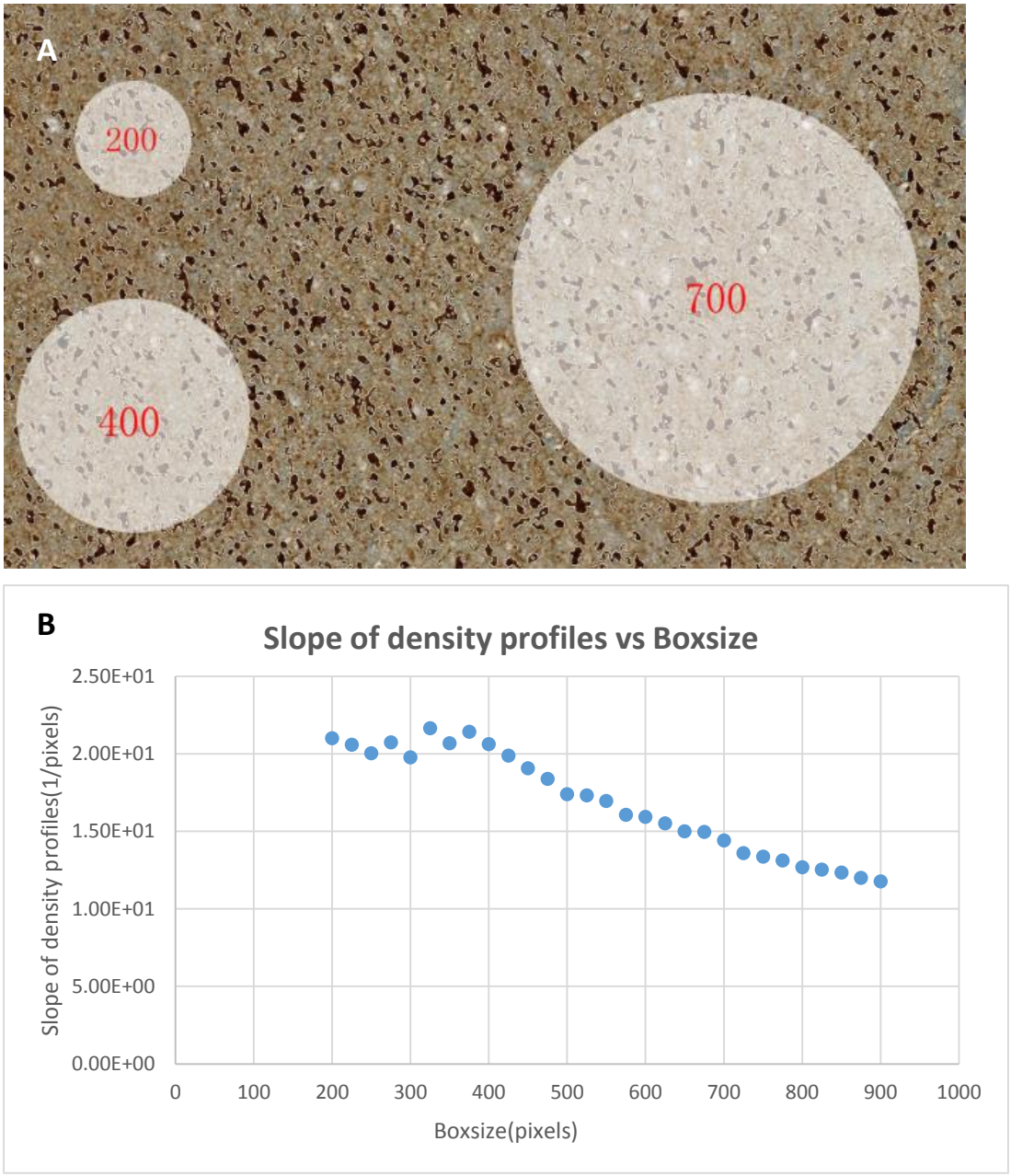


Figure 12 Boxsize selection: (A) Examples of boxsizes with 100, 200 and 700 pixels radii. (B) Cell density vs slope of cell density profiles

Another method we apply to overcome the boundary effect is splitting bigger areas into smaller areas. Due to our segmentation method, there are some areas which are too big. They will enhance the boundary effect, when the centers of mass of these areas are near the box boundary. Including or not including the big areas will lead to big changes in the cell density. The method we use is to separate the areas that are

larger than twice the average area size. For example, if an area is more than twice and less than three times the average, we separate it into two areas. On the other hand, if an area is more than three times and less than four times the average area, we separate it into three areas. We calculate the center of mass for each smaller area, and use the smaller areas to calculate the cell density. Our results indicate that these methods are effective in reducing the boundary effect.

Chapter 3

Results and Analysis

3.1 Pipeline of our analysis

We have analyzed 70 samples and 147 cuts in six different grades and types. We excluded some samples that did not have a clear boundary between glioma and healthy brain tissue. We also did not use the GBM and GBMII data because we only have one GBM sample and that may not be typical.

We made 2~3 cuts for each sample based as discussed before, we then carried out the image processing and the density calculation for each cut. We calculated the cell density profiles in different directions in each cut and selected the steepest one to represent the cut. Finally, we calculated the slope for the density profiles. Figure 13 shows the pipeline of our analysis.

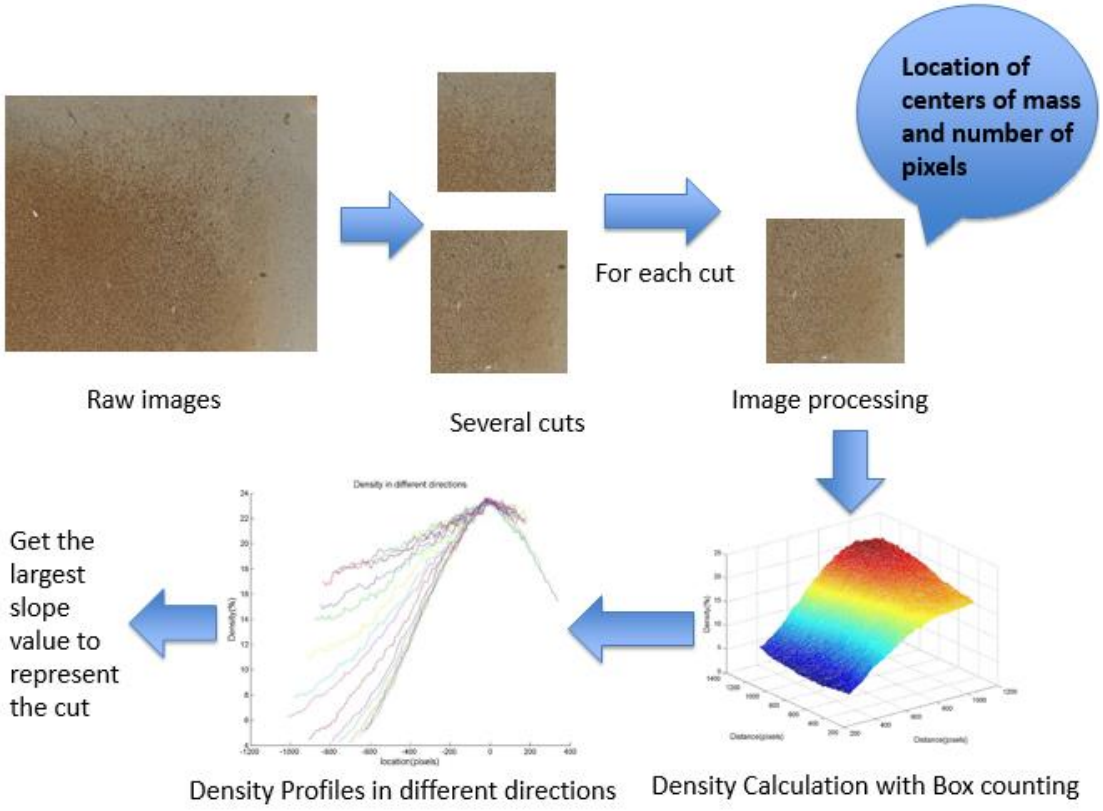


Figure 13 Pipeline of our analysis

3.2 The distribution of slope for samples in the same glioma type and different grades

Figure 14 shows the distribution of slopes for gliomas of the same type and different grades. We find that, for type A and OA, there are clear differences in the slope distribution. The lower grade gliomas tend to have a higher frequency in the high slope region. However, for type O, the OIII's slope distribution is more concentrated and OII's distribution is more diffused. Consequently, it is not possible to unequivocally say which type has a higher slope.

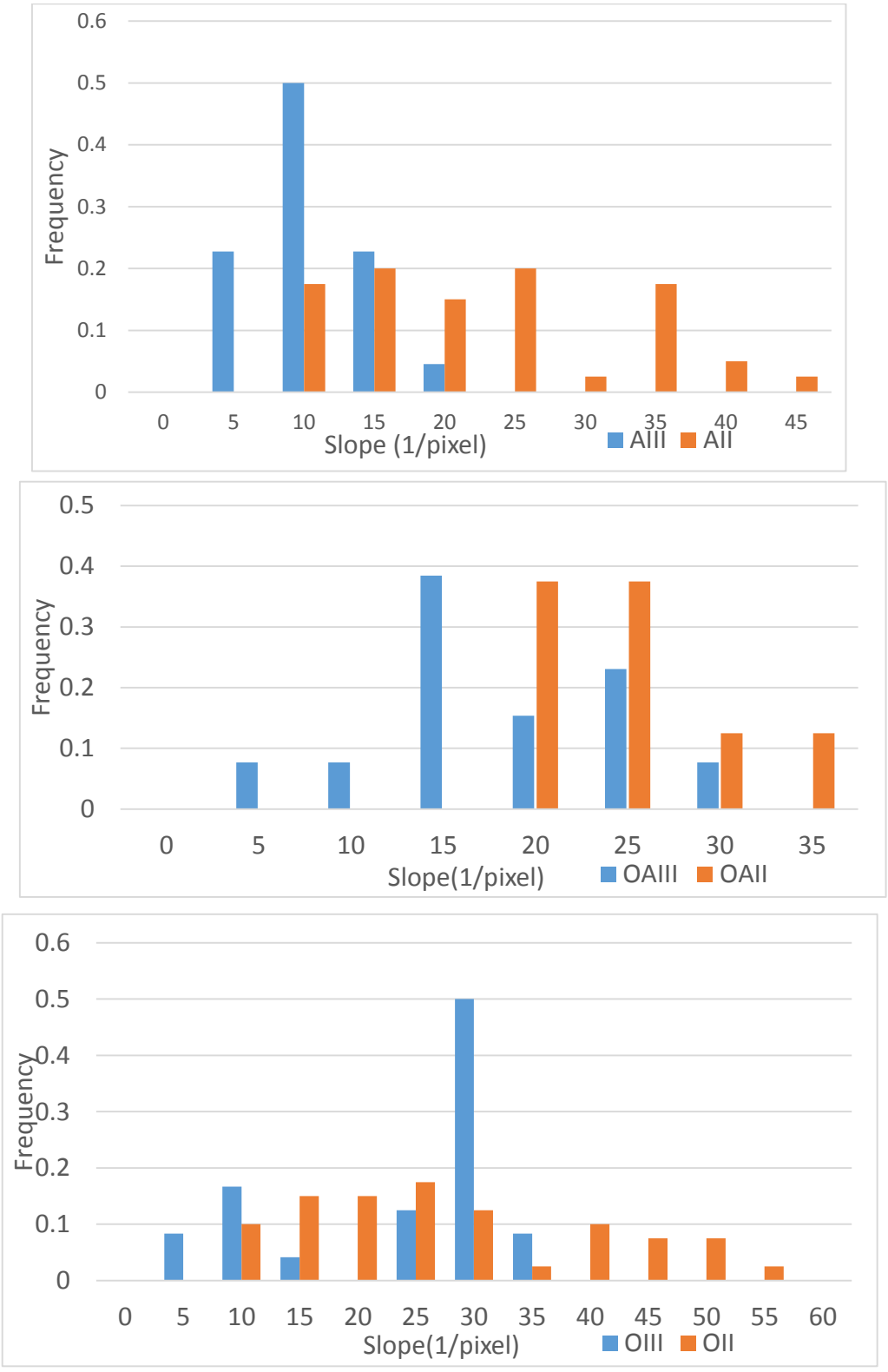
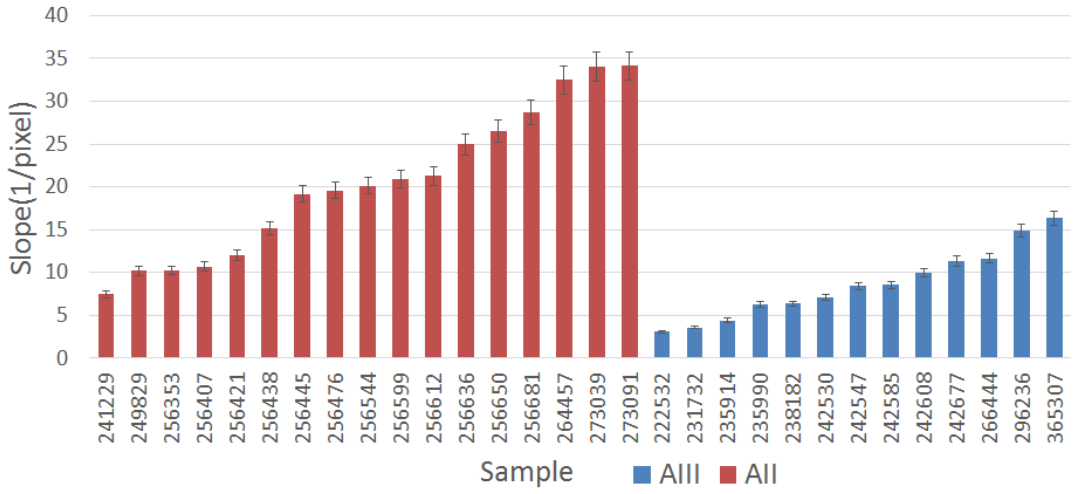


Figure 14 The distribution of slope for samples in same type and different grades

3.3 Average slopes for glioma in same sample and different grades

We make 2 to 3 cuts in each sample and then average the slope of the cuts from the same sample. Figure 15 shows our results for different samples. We see that, for type A and OA, most of the lower grade samples have a higher slope than higher grade samples. But, for type O, the coincidence between higher grade and lower grade is larger.



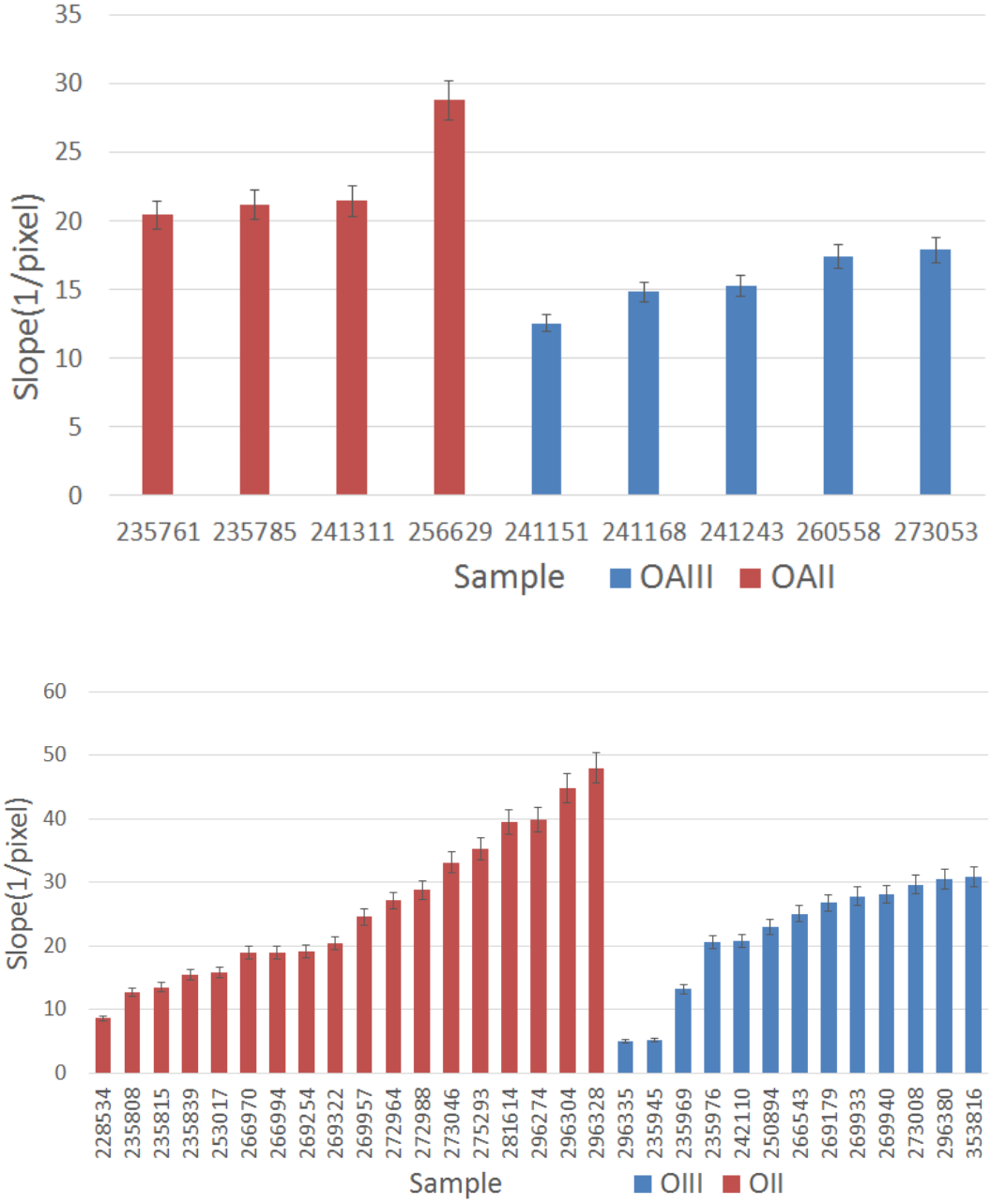


Figure 15 Average slopes for glioma in same sample and different grades

3.4 Average slope of gliomas in different types and grades

Figure 16 shows the average slope for gliomas in different types and grades. Although the difference between OII and OIII is smaller than other types of gliomas,

we find a consistent trend that a lower grade glioma has a higher slope or higher ratio of ρ/D. Since the high grade glioma is developed from low grade glioma, our results indicate that a lower grade glioma will become more diffusive and grows slower as with time. In statistics, the p-value is the probability that data at least as surprising as the observed sample results would be generated under a model of random chance (determined by the null hypothesis). We have calculated the p-value to see if the differences between the mean values of the slopes of AII and AIII; OAII and OAIII, OII and OIII are significant. The results are 9.26E-09, 0.0208, 0,0152, respectively, which is much smaller than 0.05, indicating that the differences are significant.

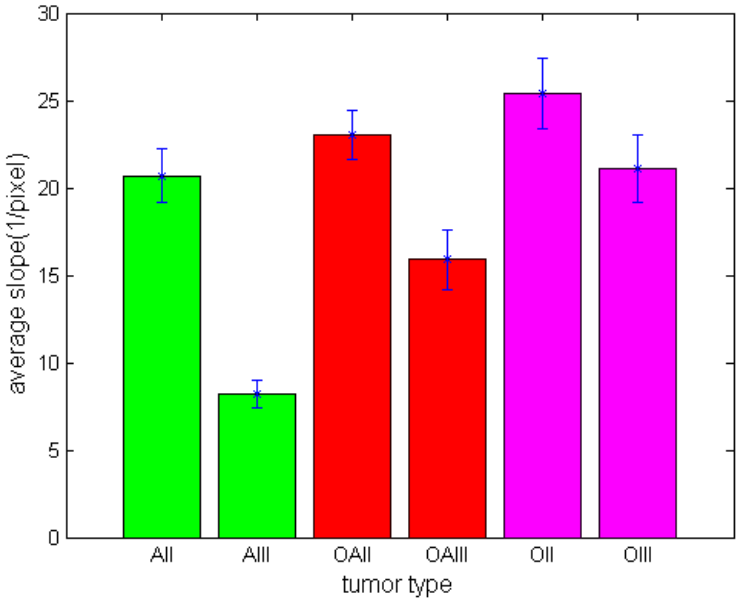


Figure 16 Average slope of gliomas in different types and grades

Chapter 4

Conclusions and Discussion

Our analysis of patient slides has enabled us for the first time to investigate and characterize the invasive front of gliomas of different types and grades. Higher grade gliomas have shallower invasive front, corresponding to more spreading and less proliferation. This trend remains the same across all three glioma types from grade II to grade III. Our approach provides some basic understanding of the growth of glioma tumors and their invasive front, which will potentially help provide better therapy for patients, especially during surgery.

This work may be continued in several directions. First it would be informative to segment cells instead of segmenting pixels. As we have discussed, segmenting pixels and cells may be both acceptable, but this should be checked.

The second work could be to find a better way to grade the segments. In pathology, a tumor will be graded as a high grade glioma if parts of it have a high grade tumor, which means another part could be a lower grade glioma. Thus, the grade may not be consistent. That is also why the slopes have a relatively large range. If we could find grades for each segment instead of samples, our result will be more representative.

Bibliography

1. Omuro A, DeAngelis L M. Glioblastoma and other malignant gliomas: a clinical review[J]. Jama, 2013, 310(17): 1842-1850.
2. Elizabeth A. Maher et al., Malignant glioma: genetics and biology of a grave matter, Genes & Dev. 2001. 15: 1311-1333
3. Young R J, Knopp E A. Brain MRI: tumor evaluation[J]. Journal of Magnetic Resonance Imaging, 2006, 24(4): 709-724.
4. Upadhyay N, Waldman A D. Conventional MRI evaluation of gliomas[J]. 2014.
5. Wang C H, Rockhill J K, Mrugala M, et al. Prognostic significance of growth kinetics in newly diagnosed glioblastomas revealed by combining serial imaging with a novel biomathematical model[J]. Cancer research, 2009, 69(23): 9133-9140.
6. Cohen A L, Holmen S L, Colman H. IDH1 and IDH2 mutations in gliomas[J]. Current neurology and neuroscience reports, 2013, 13(5): 1-7.
7. Capper D, Zentgraf H, Balss J, et al. Monoclonal antibody specific for IDH1 R132H mutation[J]. Acta neuropathologica, 2009, 118(5): 599-601.
8. Swanson K R, Bridge C, Murray J D, et al. Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion[J]. Journal of the neurological sciences, 2003, 216(1): 1-10.
9. Gu S, Chakraborty G, Champley K, et al. Applying a patient-specific bio-mathematical model of glioma growth to develop virtual [18F]-FMISO-PET

images[J]. Mathematical Medicine and Biology, 2012, 29(1): 31-48.

10. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1990, 12(7): 629-639.
11. Weickert, J. Anisotropic Diffusion in Image Processing. Teubner-Verlag, Stuttgart, Germany, 1998.
12. Weickert, J., ter Haar Romeny, B.M., Viergever, M.A. “Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing.” Volume 7, No. 3. 1998.
13. D'Almeida, F. Nonlinear Diffusion Toolbox. Available at: http://www.mathworks.com/matlabcentral/fileexchange/3710-nonlinear-diffusion-toolbox.
14. Niemisto, A. (2004). HistThresh Toolbox. Avalable at: http://www.cs.tut.fi/%7Eant/histthresh/.
15. Botev Z I, Grotowski J F, Kroese D P. Kernel density estimation via diffusion[J]. The Annals of Statistics, 2010, 38(5): 2916-2957.
16. Arteta C, Lempitsky V, Noble J A, et al. Learning to detect cells using non-overlapping extremal regions[M]//Medical image computing and computer-assisted intervention–MICCAI 2012. Springer Berlin Heidelberg, 2012: 348-356.
17. Bentley J L. Multidimensional binary search trees used for associative searching[J]. Communications of the ACM, 1975, 18(9): 509-517.