

ENHANCED VISUALIZATION OF MAGNETIC RESONANCE BRAIN TUMOR
SEGMENTATION VOLUMES FROM DEEP AUTOMATED MODELS

By

ANNA MARIE GANN

(Under the Direction of Tianming Liu)

ABSTRACT

In brain tumor diagnosis and treatment, manual segmentation is the “gold standard” approach, through which an expert annotates tumor regions manually. This process is becoming increasingly infeasible as patient data volumes exceed quantities which can be reliably segmented in reasonable periods of time. Additionally, brain tumors exhibit wide variation in type, extent and location, further complicating task. This renders manual segmentation a time-consuming and labor-intensive undertaking, shown to yield inconsistent results. Automated models, implemented with deep learning architectures have demonstrated a faster, more consistent segmentation approach. While the benefits of automated models have been established, they have historically failed to be integrated into clinical practice. Research indicates that bridging the clinical gap requires establishing “trust” and “transparency” between end users, clinicians, and these automated tools. This paper proposes color space visualization of automated segmentation results, designed to improve standard segmentation practices through collaborative effort between automation and expert knowledge.

INDEX WORDS: Artificial Intelligence, Deep Learning, Tumor Segmentation, Data Visualization, Image Processing, U-Net, Biomedical Image Analysis, Convolutional Neural Network

ENHANCED VISUALIZATION OF MAGNETIC RESONANCE BRAIN TUMOR
SEGMENTATION VOLUMES FROM DEEP AUTOMATED MODELS

by

ANNA MARIE GANN

BA, The University of Georgia, 2019

BA, The University of Georgia, 2019

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2020

© 2020

ANNA MARIE GANN

All Rights Reserved

ENHANCED VISUALIZATION OF MAGNETIC RESONANCE BRAIN TUMOR
SEGMENTATION VOLUMES FROM DEEP AUTOMATED MODELS

by

ANNA MARIE GANN

Major Professor: Tianming Liu
Committee: Frederick Maier
Delaram Yazdansepas

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

DEDICATION

This work is dedicated to two very important men in my family, a pair of the greatest motivators in my life for brain tumor research and advancement. In loving memory of my grandfather, Arvel Reese Gann, 1932-1999. In honor of my uncle, Trent Carriker, brain tumor survivor.



In Loving Memory of
Arvel Reese Gann
1932-1999



In Honor of
Trent Carriker
Brain Cancer Survivor

ACKNOWLEDGEMENTS

I would like to thank my family and teachers, who have supported my educational career over the years with their guidance, wisdom and patience in this process. It truly does take a village, and my village has shown unwavering support for which I am eternally grateful. I would like to recognize my parents, for opening every door for me along the way. I would also like to thank my thesis advisors, Dr. Liu, Dr. Maier and Dr. Yazdansepas for investing their time and effort in my studies, and for sharing their knowledge and wisdom with me along the way to help me reach my goals.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF FIGURES	ix
CHAPTERS	
1 INTRODUCTION AND PURPOSE	1
1.1 CONTRIBUTIONS	3
2 BRAIN AND CENTRAL NERVOUS SYSTEM TUMORS.....	4
2.1 CAUSE AND TREATMENT.....	4
2.2 PRIMARY AND METASTATIC TUMORS.....	5
2.3 TUMOR GRADING.....	6
2.4 TUMOR CLASSIFICATION.....	7
2.5 CHALLENGES TO TUMOR CLASSIFICATION	8
2.6 GLIOMA BRAIN TUMORS	9
3 DIAGNOSTIC MODALITIES AND BIOMEDICAL DATA.....	10
3.1 INVASIVE AND NON-INVASIVE DIAGNOSIS.....	10
3.2 COMPUTED TOMOGRAPHY	11
3.3 MAGNETIC RESONANCE IMAGING.....	11
3.4 IMAGE ANALYSIS AND SEGMENTATION.....	12
3.5 TUMOR SEGMENTATION WITH MRI.....	13

3.6	TUMOR SEGMENTATION METHODS	15
3.7	AUTOMATED DIAGNOSTIC TOOLS	17
3.8	CANCER DISPARITY	18
4	AUTOMATED AND SEMI-AUTOMATED SEGMENTATION	20
4.1	ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING	20
4.2	AUTOMATION AND LEARNING	21
4.3	EVALUATING LEARNING MODELS	22
4.4	SEGMENTATION TECHNIQUES	23
4.5	AUTOMATED AND SEMI-AUTOMATED	24
4.6	LIMITATIONS FOR CLINICAL LEARNING MODELS: DATA	25
4.7	LIMITATIONS: FALSE DISCOVERY	26
4.8	LIMITATIONS: THE BLACK BOX	26
5	RESEARCH AND CLINICAL REALITY	28
5.1	FUNDAMENTAL DIAGNOSTIC DIFFERENCES	28
5.2	CLINICAL CONCERNS	29
5.3	INTERPRETABLE DEEP LEARNING	30
5.4	HUMAN-ORIENTED DATA REPRESENTATION	31
5.5	COLOR DISCRIMINATION	33
5.6	MRI COLOR ENHANCEMENT	34
6	RESEARCH METHODS PART ONE: THE LEARNING MODEL	37
6.1	DATA	39
6.2	PREPROCESSING	41
6.3	MODEL ARCHITECTURE	41

6.4 TRAINING	46
6.5 SEGMENTATION RESULTS	47
6.6 DISCUSSION.....	48
7 RESEARCH METHODS PART TWO: VISUALIZATION	51
7.1 INPUT IMAGES	52
7.2 COLOR SPACE PROJECTION.....	54
7.3 GRAYSCALE SEGMENTATION RESULTS	54
7.4 COLOR PROJECTION SEGMENTATION.....	55
7.5 COLORIZATION AND VISUAL ENHANCEMENT METHODS...59	
7.6 ENHANCEMENT USING CONTOURS.....	65
7.7 ENHANCEMENT WITH OVERLAYS	67
8 RESULTS AND CONCLUSION.....	69
8.1 FUTURE WORK.....	71
REFERENCES	73

LIST OF FIGURES

	Page
Figure 1: Study Overview	38
Figure 2: Task Overview	39
Figure 3: 3D-ESPnet module	42
Figure 4: Efficient Spatial Pyramid	44
Figure 5: SSP and PSP Blocks.....	45
Figure 6: Input Data	53
Figure 7: Dimensional Representation FLAIR	53
Figure 8: Segmentation Results	55
Figure 9: HCL Sequential Perceptually Uniform and RGB Color Spaces	56
Figure 10: HCL Luminance Calculation.....	56
Figure 11: HCL Chroma Calculation.....	57
Figure 12: HCL Hue Calculation	57
Figure 13: Hue, Chroma and Luminance.....	57
Figure 14: Sequential Colormaps.....	58
Figure 15: HCL Map Variations	58
Figure 16: The Viridis Package	61
Figure 17: Colormap Reference	62
Figure 18: Color Space Results	63
Figure 19: Contour Results	66

Figure 20: Segmentation Overlay68

CHAPTER ONE

INTRODUCTION AND PURPOSE

Artificial Intelligence (AI) tools have revolutionized the field of data mining by enabling knowledge discovery in large, complex datasets with unprecedented speed and accuracy. In recent history, efforts have been made to apply AI deep learning (DL) models for biomedical image analysis. A specific instance is the use of deep learning models for image analysis in cancer diagnosis. In many cancer types, tumor segmentation is popular image analysis task, used to guide both diagnosis and treatment intervention. Tumor segmentation is an inherently complex task, further complicated in certain locations such as the brain. Current “gold standard” methods for brain tumor segmentation are becoming increasingly infeasible with the time and resources available. Volumes of patient medical imaging data are increasing exponentially, along with enhanced imaging protocols producing images with more detailed data than ever before. As the demand for biomedical image analysis increases, it is necessary to evaluate and improve current methods to ensure patient care is provided swiftly and knowledgeably for successful intervention.

A number of benchmark studies of deep learning models for brain tumor segmentation have received high accuracy and efficiency relative to the “gold standard” manual diagnosis. At the same time, these models have thus largely failed to be implemented across clinical institutions. Further research explains that while fast and

accurate diagnosis is ideal, further requirements for clinical diagnostic tools exist. Most frequently, clinical studies report interpretability and transparency as barriers inhibiting clinical trust of automated diagnostic tools. As benchmark DL models for brain tumor segmentation have failed to address further relevant clinical concerns, the models have also failed to gain popularity in use.

This paper addresses previous work on both state-of-the-art benchmark DL architectures for biomedical image analysis and the established need for clinical trust in diagnostic tools. The purpose of this study is to present an approach for using deep architectures for brain tumor segmentation in conjunction with enhanced explanatory visualization. The intention of this approach is to use expert evaluation, an element used in current clinical practice, to work alongside DL models in a combined effort to predict and evaluate tumor segmentation results. The goal, then, is to provide the best patient care possible by retaining the advantages of both automated and manual approaches for brain tumor segmentation.

1.1 CONTRIBUTIONS

The primary contribution of this work is the registration and enhancement module as part of a semi-automated brain tumor segmentation process. This tumor segmentation process is founded on principles of interpretable deep learning, in an effort to incorporate domain knowledge into real-world deep learning implementations. This semi-automated process involves deploying a benchmark deep automated model for brain tumor segmentation, followed by a registration and enhancement module to guide expert evaluation within the domain and ideally establish clinical trust in automated diagnostic

tools for biomedical image analysis. The enhancement module combines advancements in computer vision, human computer interaction and data visualization to improve the explanatory value of automated segmentation predictions for expert readers. Further work should implement these methods and tools in the real clinical setting so that the value and use may be evaluated.

CHAPTER TWO

BRAIN AND CENTRAL NERVOUS SYSTEM TUMORS

Brain tumors are a growing concern for society as they rate among the most common human diseases [7] and carry relatively poor prognosis for patients. A National Cancer Institute (NCI) initiative, the SEER program, indicates that in 2019, approximately 24,000 Americans were diagnosed with brain cancer while 18,000 Americans died of the disease [42]. This program also estimates that the median age at diagnosis of brain cancer patients is 59 years, and that only 32.9% of those diagnosed with brain cancer are expected to survive five years [42]. The frequency and gravity of brain tumors have identified the immediate need to improve brain cancer prognosis. Time is one of the most influential factors in brain cancer prognosis, relative to both diagnosis and treatment since brain cancer can progress rapidly and tumor stage at diagnosis strongly influences patient prognosis [42]. Relative to other cancer sites, brain tumors are associated with higher symptom burden due to the unique neurocognitive symptoms [4]. This is in part due to the restrictive location inside the skull which leaves little room for growth before brain function is affected [42, 55].

2.1 CAUSE AND TREATMENT

Most cancers are caused by a mutation of the Deoxyribonucleic Acid (DNA) sequence which alters the genes responsible for cellular reproduction [16]. DNA serves

as an instruction manual for genes, which carry out a number of cycles relevant to both cancer instance and growth. A tumor manifests as the result of a genetic mutation causing the uncontrolled growth of cells [16]. Other relevant DNA mutations consider the apoptosis, necrosis and angiogenesis cellular processes. Apoptosis and necrosis are both cell death cycles. Apoptosis is a “programmed cell death” which inhibits new gene transcription and complicates regulatory pathways [55]. Angiogenesis is a support cycle, in which genes establish vascular networks to support the needs of the body. In cancer incidence, angiogenesis allows cancer to spread and grow by establishing vascular networks to support newly formed tumors [16, 55]. A number of environmental, lifestyle, and diet factors can cause any of these mutations and result in cancer [16].

Treatment options for brain tumor patients consider both diagnostic information and relevant patient-specific information such as age, gender and religious restrictions. Most frequently, treatment options for brain tumors include surgical removal or resection, radiation therapy, chemotherapy, and other newer precision therapies such as intensity modulated proton therapy (IMPT).

2.2 PRIMARY AND METASTATIC TUMORS

Tumors are most generally categorized as *benign* or *malignant*. Malignancy indicates the degree of aggressiveness of the tumor [6], whereas benign tumors are not aggressive. Though not aggressive, benign brain tumors can pose detrimental threats to patient wellbeing due to their restrictive location in the brain and potential for functional deficit and elevated symptom burden [7], where room for tumor growth is limited by restricted neuroplasticity [52]. A further classification labels tumor as *primary*

or *metastatic*. Primary reflects a tumor which originates in the brain whereas metastatic tumors form as a result of cancer cells spreading from a separate primary tumor location in the body [3, 52]. Primary brain tumors carry high rates of mortality and morbidity, with an estimated a mortality rate of 60% [6]. Primary tumors are less common than metastasis, but present more frequently in older adults and children [84]. Primacy or metastasis, number of tumors, size and location is all taken into consideration for both treatment planning and determining patient prognosis [5].

2.3 TUMOR GRADING

Tumor malignancy is measured in grades I-IV, where lower grades indicate less aggressive tumors and grade increases with malignancy [6]. Grades I and II are referred to as ‘low grade’ while grades III and IV reflect ‘high grade’ tumors. In general, lower grades are associated with better prognosis and long-term survival. The grading criterion is largely based on four morphological criteria: cytological atypia, mitotic activity, microvascular proliferation, and necrosis [53]. Grade I tumors are benign, slow growing and do not exhibit any of the four morphological criteria [53]. Grade II tumors show only cytological atypia of the four criterion and can be either malignant or non-malignant. These tumors are generally slow growing but are known to recur as higher-grade tumors as the disease progresses [53]. Grade III indicates tumors which exhibit both anaplasia and mitotic activity, are malignant and also associated with recurrence as at a higher grade. Grade IV tumors present with anaplasia, mitotic activity, microvascular proliferation and/or necrosis. These tumors are aggressively malignant and exhibit rapid rates of reproduction [53].

2.4 TUMOR CLASSIFICATION

The World Health Organization published the most popular reference standards for tumor grading and categorization. A 2007 revision of these standards considers both brain tumors and central nervous system (CNS) tumors as a single categorization group. These standards delineate over 120 distinct classifications of brain and CNS tumors [3, 13]. The specifications for each category were updated recently in a 2016 revision. Prior this revision, brain neoplasms were classified largely on histopathologic analysis [12] followed by surgical biopsy or resection [6, 38]. This raised concerns that histological classification did not comprehensively account for the various differentiations which can co-exist within a single tumor [56]. Further concerns noted that pathologists attribute importance of the WHO grading criteria differently [32, 55, 56].

The 2016 update of the WHO standards intended to improve diagnostic accuracy, patient management and treatment response [53], with a number of notable changes. One change is the inclusion of molecular markers in conjunction with histological criteria for defining distinct tumor entities [54, 56]. Specifically, molecular markers are “layered” with histopathological criteria, such that a tumor classification might be made solely on histopathological considerations, and molecular markers provide additional clarification when required or suggested [56]. Current standards do not classify brain lesions solely on a molecular basis [53]. Molecular markers are intended to narrowly define distinct classifications with the intent of improving diagnostic accuracy and enabling better inter-laboratory comparison [53, 54]. Though the WHO provides detailed criterion for brain tumor diagnosis, tumor grading and categorization remains difficult as cellular structures become ill-defined and more difficult to distinguish at higher grades [16].

A final point on the 2016 revision is the addition of a ‘Not Otherwise Specified’ or NOS classification. The NOS designation is applied in cases where there is insufficient information to classify a tumor according to the established guidelines [37]. This can manifest in a number of instances, such as, when genetic testing is not available, genetic testing results are not compatible with other tumor classification guidelines, or in instances of uncertainty of tumor features due to insufficient sampling or artifacts which obstruct analysis [56]. This revision also includes distinct considerations for pediatric tumors separate from those of adult tumors [53].

2.5 CHALLENGES IN TUMOR CLASSIFICATION

Though molecular markers provide additional clarification for tumor categorization, there remain concerns surrounding the feasibility of obtaining molecular genetic information in the realistic clinical setting. Broadly, these concerns reason that many clinical institutions lack access to the necessary tools for genetic testing [55]. Further concerns mention that molecular genetic procedures are complicated and time-consuming, due to lack of a specified process for obtaining molecular information, which contribute to inter-observer and inter-institutional variability [53].

In addition to potentials for human error and variation, broad tumor variation further complicates classification. Brain tumors are notoriously heterogeneous, meaning that a single tumor can present differing histopathologic features in tissue samples taken from different locations as tumor biology varies throughout the extent of the lesion [38, 49]. At the same time, procedures involving extracting brain tissue carry inherent biological risk, affecting the practicality of obtaining more robust samples. Thus,

histopathological analysis further is dependent upon the extent to which the tumor samples are representative of the tumor as a whole.

2.6 GLIOMA BRAIN TUMORS

The dataset used for this study is composed of brain and CNS tumors classified in the Glioma group. Glioma account for over 70% of all brain tumors and are the most frequently diagnosed group of brain and CNS tumors [41, 46, 49], followed by the Meningioma group [41]. Glioma refers to a group of tumors which originate in the glial cells. There are three known variations of glial cells associated with cancer and tumors in the Glioma group are further classified based on the type of glial cell involved [63]. These classifications include Astrocytoma (g. I-IV), Oligodendroglioma (g. II-III), Ependymoma (g. I-III), and Glioblastoma (g. IV), where further subclassification for each of these four groups considers grade, age (pediatric or adult), and cellular features [62]. High grade Glioma present comparatively high mortality rates to other types of cancers, with a median survival rate of only two years [43]. Glioma are notoriously heterogeneous [38] and infiltrative lesions [46] which present challenges diagnosis and treatment decisions. The incidence and severity of Glioma tumors identify an immediate need for improving diagnosis and treatment of these lesions.

CHAPTER 3

DIAGNOSTIC MODALITIES AND BIOMEDICAL DATA

3.1 INVASIVE AND NON-INVASIVE DIAGNOSIS

Medical diagnostic procedures can be generally classified as invasive or non-invasive. Invasive procedures require biological samples to be extracted from the body or for some type of medical instrument to penetrate the body. Biopsy is popular invasive diagnostic procedure used in cancer diagnosis. Non-invasive diagnosis gathers information through questioning, physical examination, observation or biomedical imaging. Biomedical imaging is a popular resource used for cancer diagnosis, treatment planning and treatment response assessment since it provides quantitative imaging data with little or no biological risk to the patient.

Biomedical imaging is available in a number of modalities which are considered as structural or functional, dependent upon the type of information provided. Structural imaging depicts structural and anatomical information on the imaged organ or region. Functional imaging reflects functional information of the imaged organ through biomarkers and metabolic functions. Structural imaging modalities are routinely used in brain tumor diagnosis to gather information on the shape, extent, and position of the tumor in the brain [16, 84].

3.2 COMPUTED TOMOGRAPHY

The two most widely utilized imaging modalities for brain tumor diagnosis are Computed Tomography (CT) imaging and Magnetic Resonance Imaging (MRI), both of which provide structural information. CT imaging uses a series of subsequent and revolving x-ray scans to image the human body layer by layer [90]. CT is associated with enhanced biological risk for the use of series X-ray imaging, which requires exposing patients to ionized radiation. Doses of ionized radiation used in medical imaging have definitively been identified as a cause of cancer [28]. The concern for ionized radiation exposure is amplified in CT relative to X-ray, since a CT scan conducts a series of X-rays and therefore exposes patients to much higher doses of ionized radiation. It is estimated that a single chest CT scan exposes the patient to over 100 times more ionized radiation than a similar x-ray imaging procedure [28]. Further, the parameters for CT imaging protocols are not standardized the effective dose of ionized radiation exposure varies substantially within and across institutions [28]. A CT image reflects similar tissue information to an MR image but is generally regarded as inferior for cancer diagnosis due to the associated biological hazard [2].

3.3 MAGNETIC RESONANCE IMAGING

In contrast with CT, MRI is not associated with any known biological hazards. In a Magnetic Resonance Imaging scan, the patient lays inside the MR and the machine emits a strong magnetic field around the region being imaged. This magnetic field forces the protons in the body to align along a linear vector. Radio waves are then employed to deflect the vector of aligned protons. The intensity of this reflection is plotted in

grayscale to create a cross-sectional image of the organ [2]. Since MR images use signal intensity measured from a mechanical displacement, it is possible to identify necrotic from healthy tissue, as each have unique mechanical relaxation time [2].

An MR scan can be performed with a number of different parameters (sequences) each of which produce distinct variations of contrast in the resulting image [3]. A CT image reflects similar tissue information to an MR image but is generally regarded as inferior in the case of cancer diagnosis due to the associated biological hazard. For cancer diagnosis, common sequences include FLAIR, T1, T1-ce, and T2 [8, 16]. Fluid attenuated inversion recovery (FLAIR) is used to distinguish edema region from cerebrospinal fluid (CSF). FLAIR restricts the signal of water molecules flowing in the brain, allowing for CSF signals to be interpreted more clearly [8, 16]. T1 (also called T1-weighted) provides enhanced gray and white matter contrast and is used to annotate healthy brain tissue and view CSF [16]. T1-ce (contrast enhanced) uses the T1 sequence along with gadolinium contrast. Gadolinium is a contrast enhancement which is intravenously both before and during the scan and is used to enhance the contrast of tumor borders, necrotic core and active cell regions by making them appear brighter [3, 8, 83]. The T2 sequence is sensitive to water content and is used to visualize the edematous regions of the lesion with a bright contrast [3, 8, 16].

3.4 IMAGE ANALYSIS AND SEGMENTATION

Image Analysis is the process of extracting complex information from images [1]. Biomedical image analysis retrieves clinically relevant data from images and is to guide diagnostic decision making and treatment planning and treatment response assessment.

Segmentation is a sub-domain of image analysis which divides images into distinct but relatively homogeneous segments or regions [30, 67]. In general, image segmentation is useful for image understanding, feature extraction, and interpretation [10]. Image segmentation can be a very complex task, dependent upon the type of structure being segmented, size of the dataset and variability in the regions of interest [9].

Image segmentation involves pixel classification, where pixels are identified with one or more regions based on intensity, pattern recognition, or other quantifiable features such as texture [12, 68]. Segmentation methods are hard or soft. Hard image segmentation divides the image into regions which have zero-overlap, all pixels/voxels are identified with only one region separated by marked boundaries. Soft segmentation partitions the image into regions which overlap to some degree. This allows a single pixel to be associated with two or more regions as opposed to a single region in hard segmentation [60].

3.5 TUMOR SEGMENTATION WITH MRI

In biomedical image analysis, tumor segmentation is an applied image segmentation task which defines tumor borders and regions with the goal of identifying and analyzing tumor features as a guide for diagnosis and treatment [3, 14, 20]. Brain tumor segmentation methods are used for: tissue classification, tumor localization, volume estimation, delineation of blood vessels, surgical planning, atlas matching and medical image registration [10]. Some standard tumor segmentation methods are trained to identify healthy brain matter (white matter (WM), gray matter (GM) and CSF) and classify abnormal matter as pathological [26]. A number of models are available which

experts have designed to guide tumor segmentation based on morphological features and known anatomical structures [11].

Tumor segmentation is a highly technical and difficult task further complicated variation in brain tumor instances [10]. Brain tumor variability presents in a number of forms, including tumor heterogeneity, tumor infiltration, as well as size, shape and location. This makes it exceptionally difficult to establish generally applicable segmentation rules. Analysis of heterogeneous tumor images is challenging since the heterogeneity presents as slight various grayscale signal intensity across tumor regions. which may not be readily distinguishable to the human eye. This is an issue often related to the Glioma group, which have a lower blood-brain barrier than other tumors which inhibits the distribution of contrast during the imaging scan [49]. This is further complicated by the tendency of these tumors to be tumor infiltrative, where the necrotic core to diffuses into healthy tissue by extending tentacle-like structures [46, 49]. This makes delineating tumor borders difficult, as the borders become blurred in the diffusion process. Infiltrative tumors have been shown to present high uncertainty among expert neurosurgeons and neuroradiologists in defining tumor boundaries [49].

In addition to challenges in tumor variation, image analysis and segmentation is further complicated by factors associated with MRI including image noise, the partial volume effect, and hardware inconsistencies. Noise in MR imaging data complicates image segmentation by obscuring minute differences in the signal intensities which separate tumor regions. Noise and artifacts in MR scans have been shown to negatively influence segmentation results [12, 18]. The partial volume effect refers to blurred intensity between tissue classes which occurs at the border of distinct regions where a

single voxel represents more than one tissue type [8, 9]. The final and perhaps most severe challenge to MR analysis and effective tumor segmentation tools is the lack of standard hardware and institutional MR imaging protocols. Image intensities are not consistent across MR scanners [12], and it is often the case that different institutions use different imaging hardware [51]. In addition, there is no institutional standard for image acquisition parameters and contrast injection protocols [51]. Further the scale of voxel/pixel values is not standardized in MR imaging [46]. Each of these variations contributes to a lack of MRI generalization from which data generated from a single machine or institution can reflect different grayscale values for the very same tumor [51]. This makes it challenging to create effective tumor segmentation models which generalize well across institutions and establishes institutional bias which can directly influence the accuracy of the segmentation result [51].

3.6 TUMOR SEGMENTATION METHODS

Manual tumor segmentation (also called expert segmentation) is the current “gold standard” process across clinical institutions. [9, 12] This typically involves a radiologist sitting at a computer and using a mouse to define the region of interest (ROI) in a series of two-dimensional images taken from various angles. Images are evaluated one at a time, and once the ROI is identified, the radiologist annotates tumor features and borders [25]. An advantage of this process is the use of expert knowledge, which is readily available without sophisticated pre- and post-processing software [82]. A second advantage is that manual slice editing is performed on a “case-by-case” basis, where complex interpretations can be made without a robust training set to learn from [82].

While manual annotation is a popular and enduring approach, it suffers serious limitations and is rapidly becoming an unsustainable process. The concern for sustainability in this process is that the volume of patient imaging data being generated greatly exceeds the amount that experts can realistically analyze [18]. Manual slice editing is a tedious time-consuming task, even for experienced readers [3, 8, 33, 45, 84]. At the same time, a single medical image contains vast amounts of data [84] and the volume of patient data being generated is increasing exponentially [24]. It is estimated that a single colon cancer case generates over 20 Terabytes of data [82]. Thus, it is not feasible to manually annotate the vast, and rapidly growing volumes of imaging data in a reasonable amount of time [8, 24, 81].

In addition, manual segmentation produces accuracy concerns. Medical images contain data which the human visual system is unable to detect. A single MRI system can produce images equal to 65,535 distinct gray levels [12]. The human visual system restricts expert ability to detect minor differences in grayscale [82], and can negatively impact segmentation results [26]. This is, in addition to extenuating factors which inhibit image interpretation, such as structural noise in the images, incomplete visual search patterns, suboptimal image quality and fatigue [81]. Further accuracy concerns consider reader bias, where individuals attribute different levels of importance to diagnostic criteria and have varying levels of expertise [32].

These challenges, together with the inherent difficulty of the profession, make it unsurprising that studies show significant inter- and intra-reader variation in expert segmentation results [9, 20, 29, 32, 55, 81]. Further studies demonstrate that inter-reader variability has influences the quality of diagnosis [32] and is likely significant in the

impact of evaluating tumor response and progression [29]. Specific studies find manual segmentation to be biased toward larger models and in the process noted that the average time for a single manual tumor segmentation session took over two hours [33]. These concerns for data volumes, processing times, and variability identify a need for building efficient and accurate segmentation methods.

3.7 AUTOMATED DIAGNOSTIC TOOLS

This section focuses on AI tools such as DL as promising approach for improving current tumor segmentation methods. Perhaps the most severe limitation of manual segmentation is that volumes of patient data are expanding at rates which far exceed human resources needed to analyze them. At the same time, there is data present in MR images which is inaccessible to experts due to restrictions of the human visual system. Computational models, deep learning architectures specifically are powerful resources and practical application for efficient and effective biomedical image analysis [46, 81]. Computational models overcome several limitations of manual segmentation with efficient data processing powered by quantitative mathematical representation. DL models represent image data mathematically rather than visually. Thus enabling the identification of patterns in the pixel data relevant to classification and prediction without the inherent limitations of the human visual system [58]. Mathematical representation of pixel data enables quantitatively separation of pixel values which would have otherwise been inaccessible due to slight variation in value. In addition, these learning models use simultaneous layered functions which process images in a fraction of the time needed for manual segmentation while maintaining comparatively higher accuracy metrics. For

example, this study proposes an extended U-Net architecture, a popular implementation of deep learning architecture, which segments an entire brain with speeds ranging from as little as 25 seconds to 3 minutes [46]. These times are compelling in comparison to the average two hours needed for a single manual segmentation. Transitioning from manual segmentation to automated or semi-automated segmentation approaches will help radiologists and clinicians provide the best possible treatment options to patients faster, obtaining relevant tumor information more quickly to be used in treatment planning [18, 58].

3.8 CANCER DISPARITY

Cancer disparity identifies an additional purpose for transitioning from manual to automated and semi-automated cancer diagnostic approaches. In many regions across America and across the world, there exist rural areas with low medical infrastructure and fewer doctors than are needed to treat the population and residents lack accessible healthcare and hospital resources [15, 91]. Residents of rural and low-income areas are more likely to face a number of other challenges in meeting their healthcare needs. This is particularly relevant in the scope of cancer, where rural populations suffer from an increased risk of cancer death compared to urban populations [92] and low-income residents see an elevated cancer incidence [74]. Further, rural populations experience shorter survival times relative to metropolitan residents [74]. Real clinical circumstance often fails to meet the needs of rural populations. In turn, these populations are often diagnosed with tumors at a higher grade, presumably due to physical and economic

barriers which make early intervention more difficult for rural populations, which in turn see worse prognosis than metropolitan cancer patients [52].

It is hopeful that the implementation of automated tools can have significant benefit in areas of low medical infrastructure. Primarily, these tools are believed to be useful for helping clinicians work more efficiently, which is especially important in regions with low medical infrastructure. In addition, automated tools have potential to service low-income populations better, where studies indicate that automated diagnostic tools required for biomedical image analysis will probably be cheaper than expert image analysis [58]. This is a promising step for making healthcare resources more available to populations by offsetting the economic and physical limitations which these populations are known to suffer from. In turn, we hope to see increased survival and prognosis of rural cancer patients, by providing accessible treatment options.

CHAPTER 4

AUTOMATED AND SEMI-AUTOMATED SEGMENTATION

4.1 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial Intelligence is a field devoted to building and using “intelligent software to automate routine labor, understand speech or images, make diagnosis in medicine and support basic scientific research” [27]. AI is also known as “the field rapidly tackles and solved problems that are intellectually difficult for human beings but relatively straightforward for computers- problems that can be described by a list of formal, mathematical rules” [27]. Together, these ideas are the foundation of AI learning models for medical image analysis: to understand medical images and support medical diagnosis through heavy computational elements with a foundation of formal mathematical rules. While tumor segmentation is time-consuming and labor-intensive for radiologists, AI tools are able to tackle this task more efficiently.

The field of AI expands much further than the learning models discussed in this paper. Learning models are a large, but not comprehensive extension of AI. The core principle of learning models is to “learn” or make insight from input data without specific guidelines or instructions to complete a defined task [80]. Learning architectures use input data to build representative models to interpret the data and apply this knowledge for prediction and classification tasks. The key difference between the two categories of learning models is that deep learning refers to a specific set of models which loosely

simulate human decision making in a computational manner [80]. Unlike ML models, deep models obtain feature information hierarchically rather than linearly [27]. ML models are “linear” where structure defines a direct function mapping input data to output prediction or classification based on feature extraction. Deep learning utilizes “hidden” layers between the input and output vectors which allow more complicated concepts to be broken down into simpler ones built on top of one another [27]. Deep architectures learn complex concepts and relationships by breaking them down into simpler concepts and relationships and combining these simplified concepts for a greater feature understanding and representation of the total concept through multiple levels of composition [27].

4.2 AUTOMATION AND LEARNING

For the tumor segmentation task, MR images are used as input data to learn features, or variables and attributes in the image set. Once features have been extracted from the data, they are evaluated for relevance in the feature selection process. Feature relevance is measure of the extent to which a feature provides information relevant to solving the task at hand [73]. In short, some features provide information which carried weight in predictive output, while others provide little or no useful information to the task. The purpose feature selection is to reduce the dimensionality of the algorithm by reducing the search space. Considering only the most intuitive features and ignoring the less useful features minimizes the necessary computational resources of the model without reducing the accuracy.

Recently, popular approaches AI tools for segmentation implement variations of the deep learning architectures known as neural networks. Neural networks refer to a

family of models which mimic the structure of human neural system to learn from the data and make observations. By far the most popular in this family for medical image analysis are Convolutional Neural Networks (CNN) [11, 16]. This is because the hierarchical structure of CNNs utilize spatial and configural information from images [11]. High spatial resolution is necessary for characterizing features in heterogeneous tumors [65]. CNNs utilize this information using a convolution layers which build hierarchical feature map. This succession of layers obtains features which are invariant to translation and distortion and take into consideration a pixel's relevance to other neighboring pixels in determining a segmentation result [16].

A state-of-the-art learning network is the U-Net, an encoder-decoder network which models the convolution functions of a CNN along with pooling and up-sampling operations. These networks have quickly become very accomplished in image processing and classification tasks as the convolution function accounts for spatial information and the pooling and up-sampling operations enable high-level feature encoding while controlling dimensionality and computational resources. The result is that these models have achieved high model sensitivity and prediction accuracy [17, 61].

4.3 EVALUATING LEARNING MODELS

Machine learning models are evaluated on a number of metrics, but the most relevant to biomedical image analysis and tumor segmentation consider factors related to classification accuracy, algorithm performance and computational resources [12].

Sensitivity is a true positive fraction which represents the probability that a diagnostic test is positive, given that a person has the disease [47]. For tumor segmentation,

sensitivity is the probability that a pixel is classified as tumor given that it is in the necrotic zone. Specificity is a true negative fraction which reflects the probability that a diagnostic test is negative, given that a person does not have the disease [47]. For segmentation, this is the probability that a pixel is classified as healthy given that it is healthy tissue. Accuracy refers to the probability that a pixel is classified accurately, whether it be necrotic or healthy. Important factors for automated brain tumor segmentation models are accuracy of results, computation time, and robustness [8]. Computation time refers to the time it takes the model to generate the segmentation result, and robustness is a factor representing how well a model performs on all relevant data outside of the training data.

4.4 SEGMENTATION TECHNIQUES

Tumor segmentation techniques vary in how features are considered to build segmentation models. This paper uses a thresholding technique. Thresholding techniques classify pixels in the image based on intensity and color information, where a threshold is set for each classification, and pixels are sorted accordingly [1]. Thresholding approaches are known to be most effective when the object and background, in this case the brain and other matter, are clearly separated [1]. Edge-based segmentation considers the edges and contours in the image and is known to fail if the image is too complex for borders to be clearly identified [1]. Edge-based segmentation is used in the second phase of this study for the visual enhancement module. Region based segmentation extracts pixels and compares these pixels with other neighboring pixels in the region to exploit spatial information within the regions [1]. Finally, volumetric segmentation is the rendering of a

three-dimensional volume from two-dimensional images by stacking consecutive images together [31]. Some consider volumetric segmentation a superior technique for biomedical image analysis as the segmentation result allows the interpreter to look through the volume, and because these models are capable of accommodating significant variability in biological structures over time and across individuals [31]. The model used in this study uses a thresholding technique to create a volumetric segmentation.

4.5 AUTOMATED AND SEMI-AUTOMATED

Automated methods for brain tumor segmentation are broadly classified as either semi-automatic or automatic based on the level of interaction between a human and the model. Semi-automatic methods involve some level of human interaction where fully automated segmentation models do not. This involvement varies between models, but common frameworks for semi-automated segmentation involve an expert input in one of ways: for outlining the region of interest (ROI) [12], setting the parameters of the model [8, 12], initializing the method [12], and analyzing the visual information for feedback and checking the accuracy of the segmentation result [8]. The general purpose of semi-automated segmentation is to exploit the advantages of both expert knowledge and computational performance. In these cases, the radiologist interacts with the model as a guide, making identifications to help the model achieve the best result, providing feedback to improve model robustness and evaluating the overall credibility of the model. As experts in the field, radiologists have extensive training and experience in tumor segmentation. Automated models streamline the computation process of segmentation but cannot replace expert experience and training. Semi-automation allows experts to

analyze, interpret and assess results, combining both expert knowledge and computational efficiency for obtaining the best possible result and establishing clinical trust. Fully automated segmentation models involve no human interaction. Advantages of these models include the full labor shift from manual to computational relieving experts from the most labor-intensive parts of diagnostic testing, and promoting accurate, consistent and reproducible segmentation results [25].

4.6 LIMITATIONS FOR CLINICAL LEARNING MODELS: DATA

While data surplus is a major challenge to manual tumor segmentation and image analysis, a lack of data availability for research is one of the greatest limitations to the advancement of learning models for biomedical image analysis. The amount of patient data in existence is in no way similar to the amount that any person, group, or clinical institution has access to. This is true for tumor imaging data in general as well as labeled, segmented data. Crowd-sourcing, or annotation of large-scale datasets through collaboration is a potential response to this problem, but is limited by the potential for accuracy loss from less-than-expert sources drawing ground-truth for training [35].

To construct a representative model, the number of images collected should be relatively high and instances of quite varied shapes should be selected [18, 30, 35, 82]. Patient data is private with legal, institutional and societal barriers limiting access [82]. For these reasons, it is difficult or even impossible for researchers to obtain medical images for training and testing samples [11] and expert segmentations to train with [33, 35]. Learning models require robust datasets in order to generalize well. A trained model cannot perform well on examples of classifications which it has not been exposed to in

training [13, 27]. The performance of a model is directly restricted to the variability in the training data [27]. Insufficiently deep architectures or insufficient training examples lead to models with poor generalization [11, 27]. Lack of available and robust data is a major limitation to model advancement and causes issues for model precision [82].

4.7 LIMITATIONS: FALSE DISCOVERY

Another concern with automated brain tumor models is the concern for “false discovery”, when the cost of misclassification is high [21, 22, 36]. It is proposed that the “black box” characteristic of ML models makes them prone to false discovery [18, 37], since it can be difficult to follow how the model classifies results from input to output. Brain tumor diagnosis is a “high stakes” application, where error resulting from a lack of model robustness or poor generalization can have a detrimental impact on overall patient care. False results can guide clinicians to make ill-informed patient care decisions, such as not treating a patient who the algorithm output led the clinician to believe the patient is healthy when the patient is actually diseased. This problem is clarified by studies which suggest physicians are highly influenced by computer aided diagnostic (CAD) systems [44].

4.8 LIMITATIONS: THE “BLACK BOX”

Perhaps the most common complaint of DL models is their “black box” nature, which refers to the opacity of these models [21] and the inherent difficulty in understanding and interpreting these models intuitively [11, 37]. The internal function of ML and DL frameworks make it difficult to determine the underlying reason or process

by which an output is determined [36]. This lack of transparency has historically deterred non-experts from understanding and applying these models [82]. The transparency limitation for non-expert interpretation is exacerbated for deep networks. There is a generally understood trade-off between shallow and deep networks (ML and DL), where deep architectures, the more complex of the two, yield higher accuracy and precision at the expense of interpretability [34]. Machine learning models, then, working with a more linear-functionality, have increased interpretability but decreased performance to their deep counterparts [34].

The inherent challenge of interpreting these models has generated criticism for their place in the medical domain, even in with the promise of better efficiency and accuracy. While automated segmentation provides quantitative and accurate prediction and classification results [35], critics maintain that the medical sector requires something more than accuracy and time-efficiency: transparency. Further, it is not clear what transparency and interpretability in these models would look like. Manual tumor segmentation and diagnosis refers to biological interpretation and predictive signatures [18]. Automated segmentation uses mathematical functions and matrix translations for segmentation results, without consideration of biomarkers or predictive signatures. The problem, then, becomes building credibility and transparency for the clinicians who use these models, but who have also not strayed from the clinical gold standard, which has a totally different functionality and decision-making process.

CHAPTER 5

RESEARCH AND CLINICAL REALITY

5.1 FUNDAMENTAL DIAGNOSTIC DIFFERENCES

Standard clinical diagnosis can be understood follow a model of diagnostic reasoning shaped by domain knowledge, specialized training and experience. Generally, this diagnostic reasoning model begins with a working hypothesis and proceeds to testing the hypothesis, acquiring and interpreting diagnostic data, and then confirming, rejecting, or forming a new hypothesis as information is gathered over time [89]. The initial hypothesis is typically established based on contributing factors from both patient circumstance and clinical experience. Generally, patient case factors such as observed or reported symptoms and patient/family medical history, are considered together with clinical experience in order to determine possible or probable diagnoses. This information is used to guide further diagnostic testing until a hypothesis can be confirmed or denied with some level of confidence.

Automated segmentation is a fundamentally different process which represents, learns and interprets and diagnostic data according to quantitative analysis. Learning models use mathematical functions to determine the correlation between variables [24]. This mathematical functionality enables automated segmentation models to describe feature relationships with high accuracy and thus establish segmentation results based quantitative data. Relative to current standard practices, this is done time and resource

efficiently. At the same time, the credibility of these models in the healthcare sector is arguably limited by their fundamental distinctions from clinical norms. A notable consequence of this distinction is that automated models do not consider causal factors as relevant to diagnostic predictions, where causal factors have historically carried vast importance in standard clinical diagnostic diagnosis.

5.2 CLINICAL CONCERNS

The invention of tools for clinical diagnosis is, relatively speaking, not a new idea. Automated learning models have been widely purposed for brain tumor classification and tumor grading tasks [38, 39, 40, 41, 45]. These automated tools have existed for some time now, with continued innovations promising unprecedented speed and accuracy in biomedical data mining tasks. In recent history, several of these models implemented for testing in clinical institutions in order to identify potential directions for model refining and improvement. While the research investment in these models is abundant, clinical diagnostic support systems have thus failed to be routinely implemented in healthcare institutions. One reason behind this failure is a lack of communication between researchers and clinicians, which has led to the engineering of automated models that fail to meet clinical standards for diagnostic tools [3, 19, 84, 89]. As the end users of automated diagnostic tools, clinicians play a key role in the tools' ability to thrive, being that clinicians ultimately decide whether or not to use these tools [9]. In order to bridge the gap between research and real-world clinical implementation, clinical needs must be addressed thoroughly in the development of all clinical diagnostic tools. Several studies have sought to clearly define clinical standards for diagnostic tools.

Largely, findings indicate benchmark automated models fail to meet clinical standards for interpretability and transparency [9, 23, 84] with additional desires for user-friendly interfaces [3, 9, 84].

5.3 INTERPRETABLE DEEP LEARNING

It has been established that clinicians want diagnostic tools which prioritize interpretability and ease of use [9] but remains unclear what this entails from a developmental standpoint. There is a lack of consensus on what constitutes ‘interpretable’ machine learning in any domain. It nonetheless remains that deep learning systems intended for the medical domain need be supplemented to improve human understanding and decision-making [19]. Certain works explain “interpretability” as loosely synonymous with explain-ability, and claim that the transparency standard for diagnostic tools can be met by providing clinicians with a validation standard for evaluating the model [19, 22]. In this sense, the “goal” is to provide clinicians with grounds for justifying the result. It is mentioned that this approach is useful for “high stakes” diagnostic cases where the cost of mistakes can be detrimental. This approach is further supported by the argument that “the human body is a black box” in which causal relationships can often not be identified for a number of reasons [19], and yet, diagnostic decisions are still made with some level of confidence.

Other works conclude that the diagnostic transparency standard can be met through expert interaction and enhanced data visualization to translate the complex logic behind the so called “black box models” [23]. Based on these findings, the understanding in this study is that an effective approach for enhancing the transparency of automated

diagnostic models establishes a standard which allows end users to evaluate model performance with reasons to confirm or deny the prediction, before any action is taken on behalf of the clinician [21]. From the clinical standpoint, transparency is valued as the foundation of a creditable and faithful diagnosis and serves as a reasonable defense against adversarial attacks [35]. Several works corroborate the idea that credible diagnostic tools do not require expert understanding of functionality, so long as they supplement clinical understanding with established standards for supporting or denying model results [34].

Healthcare institutions place a premium on the reasoning and comprehensibility of diagnostic systems [32]. The ability to explain diagnostic decisions is of significant interest to clinicians [32], possibly even more so in high stakes cases like brain tumor diagnosis, where the cost of being wrong is substantial. Transitioning automated diagnostic tools from research to clinical reality requires clinical motivation to do so. Establishing model credibility among non-expert users is a critical step toward implementing deep models into real-world situated use.

5.4 HUMAN-ORIENTED DATA REPRESENTATION

Visual interfaces with enhanced representation have the potential to add transparency to automated clinical tools for tumor segmentation by exploiting domain specific knowledge to gain credibility. Data visualization is concerned with the graphical representation of data set values and patterns. Effective data visualization models represent data intuitively well in terms of patterns and distinctions. Well-founded visual models have proven to be of significant value in enhancing knowledge discovery and

analysis of class-labeled data [36]. Recent studies have explored data visualization models as a tool for enhancing the explanatory context of diagnostic models and improving model understanding [22]. Several studies report that visual and interactive user interfaces, when implemented effectively, are effective in enabling users to integrate domain knowledge to interpret complex models [37].

The benefits of graphical model representation in enhancing the explanatory power of data are well explored. A highly relevant study finds that data visualization is a potentially powerful tool for enhancing visual reasoning and model credibility in automated models for diagnosis of breast cancer [36]. Further studies analyze the explanatory value of color differentiation in perception of model output for classification tasks. One particular study proposes a method for increasing visual reasoning with an interface which uses color differentiation to display quantitative and qualitative similarities among queries and class labels [87]. The authors of this study boast a “win-win” methodology for enhanced discriminatory power, on the grounds that the visual interface allows better formalization to support the visual reasoning process [87]. This methodology has potential to be further enriched in cases of expert use, where domain knowledge and enhanced visualization can be used as collaborative tools for assisting model understanding. These studies are among many which present significant evidence to support the use of visual models as a means of exploiting expert domain knowledge for overall increased visual reasoning and model understanding. Further, visual models are a practical approach for enhancing deep model understanding, being that visual enhancement modules are generally low-cost in terms of time-efficiency and computational resources.

5.5 COLOR DISCRIMINATION

It remains well established that data visualization has potential to be an effective tool for information gain and model understanding, but specific standards for “effectively” representing data remain unclear. An effective visual interface not only represents the data accurately and comprehensively but also in a manner which is intuitive for the end users. In context, an effective visual model for tumor segmentation visualizes the segmentation result in a manner which enables a radiologist to apply domain knowledge for evaluation as a standard for validating or invalidating model output. If successfully implemented, the intention is that the radiologist is able to determine model credibility in reference to expert knowledge of industry standards.

Biomedical images each contain vast amounts of data and standard visual representations of this data are not optimized for human visual interpretation. In this case, state-of-the-art visualization methods can play a key role in enhancing the discriminatory power of biomedical image representations. Previous work explains that the human visual system and color discrimination are highly relevant to enhancing information content in monochrome MR images for expert interpretation [60].

The limited explanatory content of monochrome medical images can generally be explained with color differentiation tasks. Color discrimination tasks rely on a person’s ability to detect small differences between two visual stimuli when examples of similar chromatic composition are presented [64]. A person with normal color vision can distinguish millions of colors [76]. Those with serious color deficiency can discriminate only a few hundred different colors [76]. However, this reflects general standards for distinct (not continuous or connected) color experiences, whereas color differentiation

becomes a far more challenging task when colors are presented in a continuous space, such as an image.

The human visual system projects all possible color impressions into three-dimensional space, and color perceptions differ on these three dimensions: intensity, the brightness factor of the perception, hue the color, and chroma (saturation), or the level of “colorfulness” [64, 77]. Several predefined color spaces have been created which represent colors according to numerical values. A popular “standard” color space is RGB (red, green, blue), another is HSV (hue, saturation, value) [77], a model intended to be designed to consider human sense of color [90]. Generally, color spaces differ upon scope, order and uniformity.

The human eye is unable to perceive continuous change in color and is only able to differentiate colors when there is significant change [60]. The threshold for change detection [64] is subjective to each person but generally within a standard margin. This margin was popularized by Ernst Weber, a 19th century psychologist who explained this concept as “Just Noticeable Difference” or “Weber’s Law”. Color spaces quantify color difference with distance formulas. Euclidean distance is a popular metric for cubic color representations such as RGB. In light of Weber’s Law, visual representations which maximize the instances of JND have enhanced discriminatory power relevant to color differentiation tasks.

5.6 MRI COLOR ENHANCEMENT

Since MR images are monochrome, pixel values are distinguished as variations of intensity using only a single hue. Applying the concept of uniform and distinguishable

color distance thresholds, the discriminatory power of MR images can be enhanced by maximizing the instances of JND over the image. The intention is to increase visual discrimination power by taking advantage of a larger range over the human visual system and expanding the differences between pixel values. At the same time, it is crucial to maintain uniformity and visual cues [78] in the translation so that the data is accurately represented according to the intensity information [77], but over a larger range of hues. By modeling the same data in a visual interface which utilizes all three aspects of color impression, the data takes a more human-centered approach to improving pixel discrimination for classification by mapping the pixels into a complete color space instead of along a single hue vector. This is done to enhance the difference (according to color distance) between similar values in the MR scan.

Several studies have considered the value of colorized tumor segmentation, both for the actual segmentation process and for interpretation of segmentation results. For studies which use colorization in the segmentation process, many report promising results with improved segmentation accuracy. This study reports an improved segmentation result obtained using color space translations as a preprocessing method [72]. Several similar studies report increased classification accuracy using color translated images for training segmentation models [72, 77, 90]. While this is a popular approach with established advantages, this comes at the expense of higher computational costs compared to monochrome images due to dimensionality expansion [67]. For this reason, the current study focuses on the value of color translations for segmentation evaluation.

Several earlier works address various applications for color translation of monochrome medical images for enhancing discrimination power. One approach for

transforming monochrome images uses 38 colors to represent the 38 basic types of soft tissues in the human brain [60]. This study reports that a single MRI sequence is insufficient for precise pathological evaluation, but that the colorized images enhance discriminatory power for information gain in pathological interpretation. The result, as reported, is an increased visualization of tissue density and opaqueness for easier analysis of brain regions individually and the brain in entirety [60]. One study uses a color space translation algorithm to evaluate head and neck tumors with a five-color-coded map and reports information gain in image interpretation and visualizing tumor heterogeneity [86]. A similar study maps two-dimensional (hue, intensity) pixel information into a 20-contrast scale and found similar improvements in enhancing data conspicuity and efficiency of interpretation [85]. Other studies encountered similar findings on the positive effects of color mapping MRI data for increasing interpretation performance of inexperienced readers [71]. These findings support motivations for using color space projection as a tool for information gain and model understating with clinical expert evaluation.

CHAPTER 6

RESEARCH METHODS PART ONE: THE SEGMENTATION TASK

The purpose this study is to propose a semi-automated approach for brain tumor segmentation which combines state-of-the-art deep learning and computer vision innovation with methods of current clinical practice in a collaborative effort to identify more-efficient tumor segmentation methods. Manual segmentation, the current “gold standard” practice, is unsustainable due to concerns for relevant time-efficiency in cancer diagnosis and rapidly growing quantities of patient imaging data to be processed [18]. At the same time, manual segmentation is a historically relevant process with established clinical trust, while more time and resource-efficient benchmark automated tumor segmentation approaches have failed to gain popularity in clinical practice. Prior studies indicate that deep learning tumor segmentation models “fail to thrive” as a result of unmet clinical standards for diagnostic tools [3, 19, 84, 89]. While current benchmark models exceed clinical standards of speed and accuracy [46], additional concerns for interpretability, transparency and a user-friendly interface remain largely unaddressed [3, 9, 23, 84]. This study presents enhanced visualization and export collaboration as an approach to interpretable deep learning and adapting benchmark deep learning segmentation approaches to meet clinical industry standards for diagnostic tools.

The Methods used in this study are divided into two subsequent tasks: the segmentation task and the registration and enhancement task. In the segmentation task, a

benchmark deep learning model for brain tumor segmentation is trained with and evaluated on a brain tumor dataset. In the registration and enhancement task, the resulting segmentation predictions are registered as images and the data is translated into a visual representation intended to enhance discriminatory power and perceptual uniformity according to standards of human visual perception. *Figure 1* gives an overview of the entire study broken down into the segmentation and registration and enhancement tasks. *Figure 2* explains each of the subtasks.

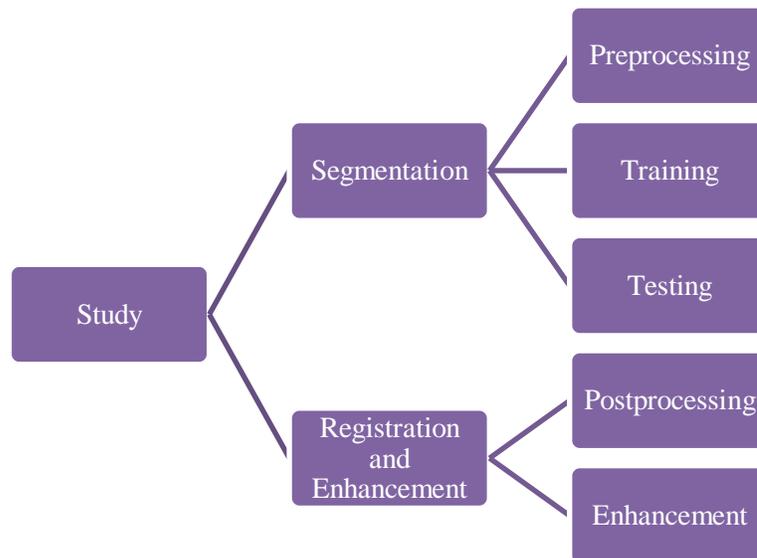


Figure 1: Study Overview

This chapter focuses on the segmentation task and explains the first three tasks in the total five-task overview presented in *Figure 2*. The intention of the segmentation task is to utilize recent innovations in deep learning models such as the U-Net for brain tumor segmentation. This, and similar automated architectures for brain tumor segmentation complete the task with better time efficiency and accuracy relative to the currently popular manual segmentation process.

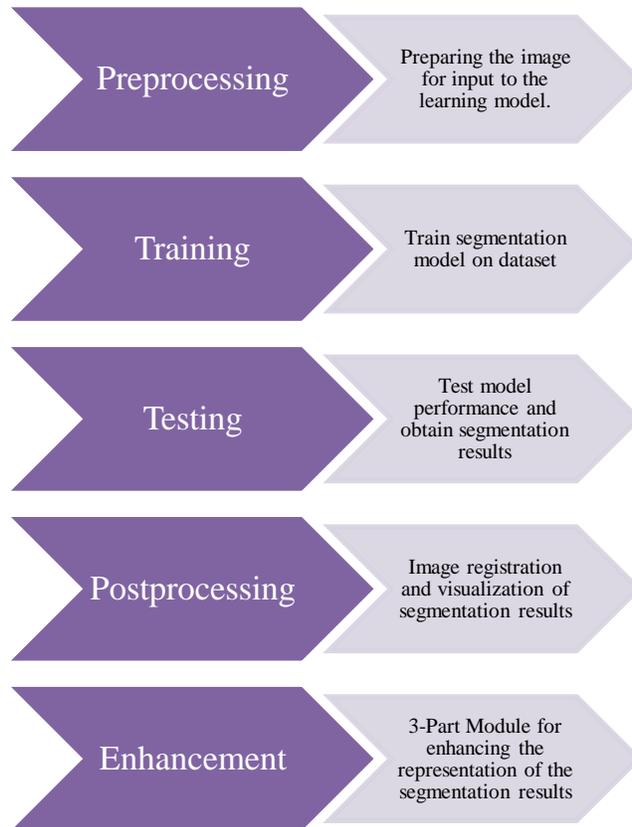


Figure 2: Task Overview

6.1 DATA

This method is trained and validated with imaging volumes provided by the Medical Image Computing and Computer Assisted Intervention (MICCAI) society for the annual Brain Tumor Segmentation (BraTS) Challenge [43, 48, 49, 50, 57]. The BraTS challenge is one of the largest research initiatives for automated and semi-automated segmentation of brain tumors. The BraTS challenge dataset is uniquely beneficial for ML and DL models as it overcomes several limitations of medical image computing datasets. The performance of automated models is directly dependent upon the quality of data that is used to train the model. Ideally, an automated model will be able to perform well if trained with a data set which is heterogeneous, representative and comprehensive.

In reality, such high caliber and robust datasets are rare in the medical field, where patient privacy regulations heavily restrict the sharing and distribution of medical data. Inter-institutional variation aids the trained models in obtaining better generalization by decreasing the learning effect of randomness and increasing objectivity to the imaging protocols of specific institutions. Further the BraTS datasets are released with segmentation and ground truth labels which have been annotated through expert collaboration, reducing the potential impact of human error. Each dataset is annotated by 1-4 experts following the same established guidelines before ultimately being approved by a board-certified neuro-radiologist. It is believed that this effort of clear segmentation criterion and collaboration reduces the potential bias of human error in the established ground truth. Finally, this dataset is comparatively large, with many volumes of patient data available for model training, validation and testing.

This implementation uses the 2018 BraTS Challenge set for training and the 2019 set for validation. Both of these sets consider Glioma tumors, the most widespread category of brain lesion. Each set contains both glioblastoma or high-grade glioma (GBM/HGG) and low-grade glioma (LGG) volumes which have been pathologically confirmed. The 2018 set is sourced from 19 separate clinical institutions and includes n=210 HGG patient volumes and n=75 LGG patient volumes totaling to n=285. The 2019 set contains n=259 HGG patient volumes and n=76 LGG patient volumes for a total of 335. Each patient volume contains 4 multimodal pre-operative scans of file type NIfTI (.nii). Each patient volume contains a single example of four distinct MR imaging protocol: native (T1), post-contrast T1-weighted with gadolinium (T1Gd), T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR).

6.2 PREPROCESSING

The BraTS dataset is preprocessed before distribution. This preliminary preprocessing performs skull-stripping by utilizing the brain extraction tool from the Oxford center for Functional MRI of the Brain (FMRIB) software library, commonly known as FSL. This preprocessing also includes re-orienting each scan according to the left-posterior-superior coordinate system and co-registering to a uniform T1 anatomical template with the Linear Registration Tool (FLIRT) provided by FMRIB. Finally, images are uniformly interpolated to a voxel resolution of 1mm^3 [49]. The resulting images are annotated with three defined segmentation labels, each differentiating distinct tumor regions and attributes. These labels identify the enhancing tumor core (ET, label 4), the non-enhancing tumor regions (NET/NCR, label 1) and the peritumoral edema (ED label 2). To enhance data visualization, this approach renders a volumetric segmentation. The imaging volumes are also normalized using mix-max normalization, which rescales the imaging data to a refined range based on the minimum and maximum voxel values observed.

6.3 MODEL ARCHITECTURE

To obtain a segmentation result to enhance, I implemented the approach: “3D-ESPNet with Pyramidal Refinement for Volumetric Brain Tumor Image Segmentation” proposed by Neuchterlein and Sachin in [93]. This approach awarded second in the 2018 BraTS challenge, and the source code is available here [93, 94, 95]. This model is an adapted U-Net architecture, which the authors call 3D-ESPNet. The main differentiation between 3D-ESPNet and U-Net, according to the authors is the use

of efficient convolutional blocks instead of stacked convolutional layers to learn and represent features. A visual model of the 3D-ESPNet network is reproduced from the authors in *Figure 3*.

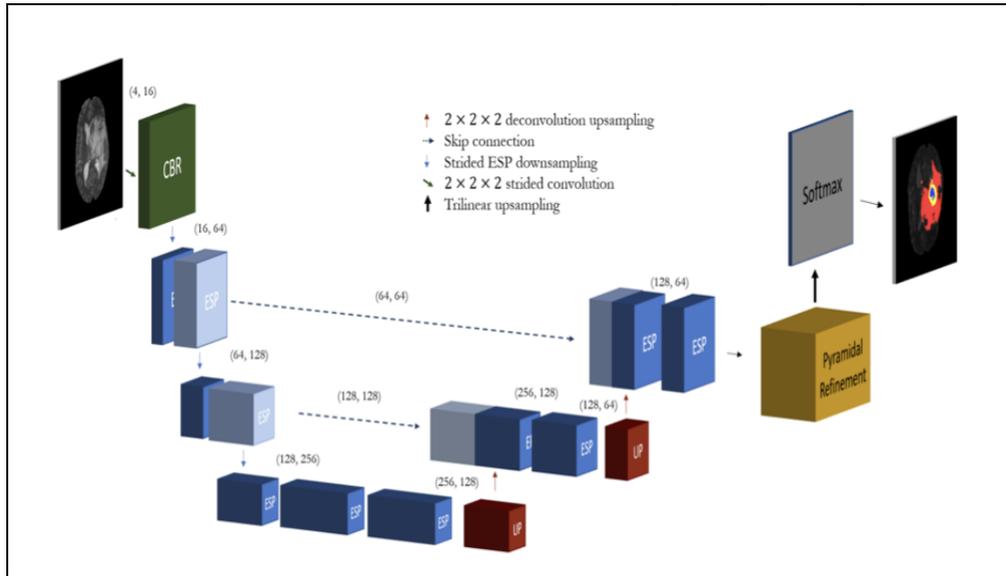


Figure 3: 3D-ESPnet Module

The 3D-ESPNet approach implements an end-to-end system design using a 3-dimensional adaptation of the Efficient Spatial Pyramid module (ESP) proposed in [94], followed by pyramidal refinement. End-to-End is a system design principle first popularized by Saltzer, Reed and Clark’s paper, “End-To-End Arguments in System Design”. This principle suggests organizing functions within system modules based on rational principles of cost evaluation, where the cost of implementing functions at low levels of a system may exceed the value of those functions at the same level [96].

The ESP module [94] is a deep model designed with convolutional factorization in order to reduce the computational cost of deploying the model while maintaining accuracy, relative to other deep models for segmentation tasks. Per the authors, ESPNet is

“fast, small, low power and low latency, yet still preserves segmentation accuracy” [94]. Convolutional factorization is a technique for reducing the computational complexity of operations by dividing the convolution operation into multiple steps. The convolution factorization used here splits a standard convolution operation into two steps, a pointwise convolution followed by a spatial pyramid of dialed convolutions. Pointwise convolutions refer to a 1×1 convolutional layer consisting also of a convolutional filter of the same size which considers only a single point per channel at a time [97]. Pointwise convolutions are widely used for parameter reduction in deep learning architectures by adjusting the number of channels or dimensions in feature maps in order to optimize computational efficiency. Dilated convolutions are used for image registration, to enhance image resolution by inserting zeroes between voxels in convolutional kernels [98]. Dilated convolution operations are determined by a dilation rate which specifies the number of zeroes inserted between the image voxels. This network maps the standard 2-dimensional convolutional operations ($n \times n$) to 3-dimensional convolutions ($n \times n \times n$) to map the features into volumetric space.

The convolutions here are employed as an encoder-decoder network, where the network learns feature representations in the encoder phase and decodes these representations in the following phase. In the encoder phase, the network performs a single stride convolution (a convolution operation in which $\text{stride}=1$, the default value and stride define the step size of the kernel) followed by three ESP convolution operations which use a 3-dimensional kernel and a stride of 2. The ESP module is reproduced from the authors for reference in *Figure 4* below. The variation in kernel and stride values allows the encoder to learn feature representations in multiple scales. It is

important to note that the ESP convolution maps features into convolutional blocks as opposed to the standard method of stacking convolutional layers. The decoder phase upsamples the feature representations output from the encoder with $3 \times 3 \times 3$ deconvolution kernels followed by trilinear upsampling layer. The decoder outputs a feature map to be passed to the following pyramidal refinement module.

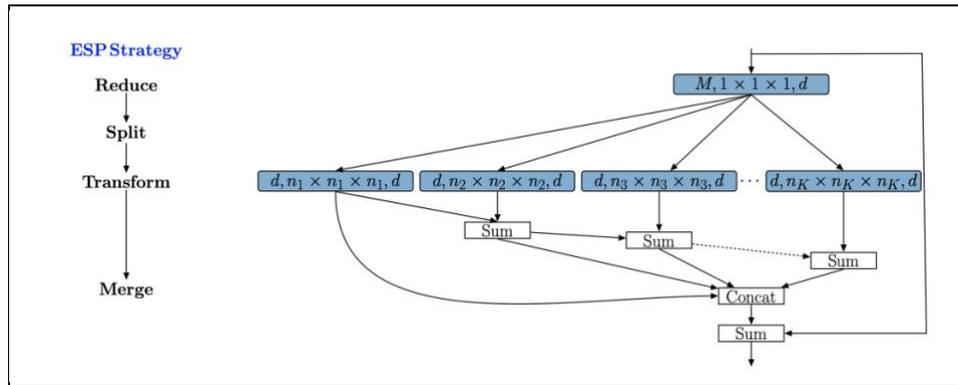


Figure 4: Efficient Spatial Pyramid

The pyramidal refinement module is comprised of three layers: the projection layer, the spatial pyramid pooling (SSP) block, and the pyramid pooling module (PSP) block. A visualization of this module is reproduced from the authors in *Figure 5*. Per the authors this module combines techniques for both feature map-based and convolutional kernel-based pooling methods, effectively establishing feature representations in convolutional blocks opposed to the standard stacking of convolutional layers in order to efficiently establish spatial information within the images. In the projection layer, the ESP block is mapped into C-dimensional space (where C= number of classes). This is done using a 3-dimensional pointwise convolutional layer followed by batch normalization and Rectified Linear Unit (ReLU). Batch normalization is used to adjust

the means and variances or activations of the input layers. ReLU is an activation function which outputs zero if input is negative and outputs identity (the input value) otherwise. Following this is the SPP block. In this pooling layer, low dimensional feature vectors (resulting from the ReLU operation) are sub-sampled with convolutional kernels of varying dimensions. The output of this pooling block is the sum of the varied convolution operations. The final PSP block takes as input the feature representation maps and divides these maps into C-branches (where C= number of classes). The feature maps are then down-sampled, where each separate branch down-sampled with a different pooling rate. The feature maps are then transformed with pointwise convolutions and upsampled to the resolution of the input feature maps with bilinear interpolation. The resulting feature maps are then merged with the input maps (of the same resolution) and an output feature representation is returned. A visual model of the SPP and PSP blocks is reproduced from the authors in *Figure 5*. In this network, the pyramidal refinement module (comprised of the aforementioned projection, SPP and PSP blocks) is completed with a final layer, called the “classification layer”. The classification layer implements an additional SSP block and upsamples the output by a factor of two with trilinear interpolation (as opposed to bilinear interpolation in the PSP block).

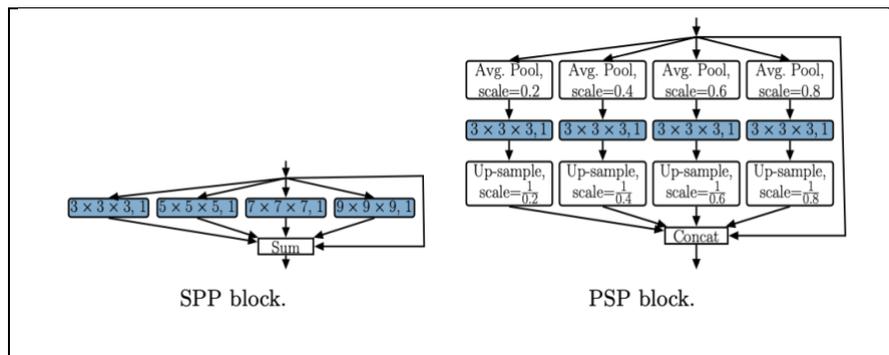


Figure 5: SSP and PSP Blocks

6.4 TRAINING

As a training set, I used the complete BraTS 2018 set including both HGG and LGG volumes. For validation, the BraTS 2019 set was used. I opted to use a separate validation set rather than partitioning the training set (as the authors did) in order to maximize the quantity of training data for the model to learn from and to maintain class balance.

In this implementation, a GeForce RTX 2080 Ti with 64GB DDR4 Graphics Processing Unit (GPU) was used. On this machine, Anaconda version 4.8.2 was built with Python version 3.7.4 (released July 2019). Using Anaconda, I created an environment for this module to install and access the various Python libraries and dependencies which were required. The authors originally deployed this model using PyTorch, an open-source machine learning library, version 0.3.1, with Compute Unified Device Architecture (CUDA), a parallel computing platform and programming model for GPU computing. This was originally done in 2018 for the BraTS challenge. The authors recommend using PyTorch version 0.3.1 (released February 2018) which is compatible with CUDA 9.1 binaries (released December 2017). In this work, I modified the source code for compatibility with more recent Anaconda and Python distributions, as necessary for resolving incompatibilities with older libraries. Specifically, I encountered incompatibilities with CUDA version 9.1 upon initial install and upgraded to PyTorch 0.4.1 and CUDA 9.2. Since the imaging volumes are of file type NIFTI1 (.nii) Python package NiBabel (version 2.3.0) was used to gain read and write permissions. For scientific computing purposes, Numpy version 1.7.1 Anaconda distribution was installed

in the environment. To work in conjunction with Numpy, scikit-image version 0.15.0 was used for image processing with Numpy arrays.

I trained with a class size of 4 (representing the data labels), using a batch size of 128 images at full resolution with a learning rate of .0005 to 500 epochs. The number of parameters used was 3,626,584. With each epoch, a loss function was calculated using the mean intersection over union was recorded. Mean intersection over union, also known as the Jaccard index, is an evaluation metric for segmentation which reflects the percent overlap between the predicted and the target through a quantitative measure of common pixels between the two. The mean IoU represents the average over the four class labels at each epoch. Training was initiated through a virtual private network (VPN) service and lasted slightly under 7 days. Evaluation took under an hour. The authors report a significantly shorter training time of approximately 5 hours under different conditions. This vast difference can presumably be attributed to a number of contributing factors. They train with less data, using the BraTS 2018 training set with an 80:20 split for training and validation. They also train at full resolution, with a batch size of 4 where I used 128. Additionally, they train to 300 epochs using a learning rate of $10e-4$ for the first 200 epochs and a rate of $10e-5$ for the remaining 100 where I trained to 500 with a learning rate of $5e-4$ constant.

6.5 SEGMENTATION RESULTS

The segmentation result is quantitatively evaluated using dice score calculations for each class label. The authors report dice scores of .74 for enhancing tumor (ET), 0.88 for whole tumor (WT) and .81 for core mask. These scores reflect performance on the

BraTS 2018 official. test/evaluation set and achieved second place in the official challenge. Dice score calculations on my implementation achieve .92 on the WT label, .88 for the CM label and .73 for ET and also reflect model performance on the BraTS 2018 official test set.

6.6 DISCUSSION

Relative to manual segmentation as the current standard clinical practice, these are promising results. While my initial model training was time-consuming compared to a typical manual segmentation session, where model training took a week and a typical manual segmentation session lasts 2-3 hours (on average), observed testing times were consistently brief at around 20 minutes. Additionally, manual segmentation also observes long training windows, where neuroradiologists undergo multiple years of instruction on manual segmentation before results are produced. At the same time, testing windows are much shorter as observed for automated segmentation methods when compared to manual.

Further, automated approaches have the potential to present more consistent results by approaching the segmentation task from a strictly quantitative standpoint. In context, this means that automated networks for tumor segmentation use image processing libraries, as Numpy and Sci-kit image are used here, to represent image properties as values in arrays. Representing the tumor images mathematically enables the network to establish tumor features through functional differentiation between numerical values. These values can then be attributed to tumor features and matched to labels, through repeated exposure in the form of batches of imaging data being “seen” by the

model as it learns. Simply put, automated models learn to attribute distinct tumor features (and their corresponding labels) mathematically, and in this process, thresholds are established for assigning labels and segmentation boundaries, where the threshold is met (activated) when the features are observed. It is the fundamental working aspects of these models which attribute to them consistency in the segmentation process and consistency and reproducibility in the output segmentation results.

One of the greatest benefits of artificial learning models is the ability to process large quantities of data quickly and to make objective observations from the data. However, these models are not perfect. These models are engineered by humans, as is the data that these models learn from, and each of these factors imbeds varying level of bias into the models themselves and in turn the results. In addition, automated models are constructed to learn from what is observed, meaning that these models fail when tested on examples not represented in the training data. Clinicians, on the other hand, are adaptable. Clinicians, like all humans, continue to learn their entire lives. Clinicians are able to adapt and adjust as necessary and this, combined with expert experience and training, is indispensable in the medical field.

Still, a concern for overall generalization remains, where automated models can be limited by objectivity to the point that they are not (easily) adaptable. In context, the model implemented in this study was trained with glioma data sourced from 19 clinical institutions with varying head MRI protocols and varying hardware signatures. It could be the case that this model, when tested on glioma MRI data sourced from an unseen institution, may not perform as well as observed with data sourced from the training institutions. This overgeneralization issue happens when the model attributes feature

importance to randomness or abstraction in the training data which is not representative of the testing data due to variations in imaging protocols, hardware, etc. This issue can be resolved or negated by re-training the model with additional data, but that process can again be time consuming. However, current trends in hardware and software development are working to mitigate the issue of generalization, through faster, and more efficient models and computing machinery. It is becoming an increasingly feasible options to 'retrain' and adapt models quickly and efficiently.

CHAPTER 7

RESEARCH METHODS PART TWO: VISUALIZATION

Studies in interpretable deep learning applications for clinical diagnosis show that data visualization has significant value in enhancing the explanatory content of diagnostic models and improving model understanding [22, 36]. In addition, effective visual interfaces have been identified as a powerful tool for enhancing visual reasoning and model credibility [36] as well as displaying quantitative similarities among class labels through color differentiation [87]. The contribution of the registration and enhancement module is to present visualizations of the benchmark 3D-ESPNet segmentation result for the purpose of exploiting expert domain knowledge.

This chapter focuses on the main contribution of this study, the registration and enhancement task. The following segments outline the process of framing and enhancing the segmentation result for the purpose of expert interpretation. The visual representation is ‘enhanced’ according to standards of human visual perception and increased discrimination power. The purpose of this approach is to utilize advantages of expert interpretation in the manual segmentation by exploiting principles of visual perception to maximize the discriminatory power of visual data representations.

In the registration task, the segmentation prediction from the 3D-ESPNet segmentation is visualized as an overlay on each of the four corresponding input images. This is done for the purpose of baseline evaluation and standard MR representation. The

subsequent enhancement module utilizes three different methods for post-processing the monochrome representations in accordance with visual perception and computer vision standards. The three enhancement methods implemented use color space translations, overlays and contours to modify and ideally the representation of the MR scans and segmentation results for the purpose of expert evaluation.

7.1 IMAGE REGISTRATION

Provided here is a single patient volume from the BraTS 2018 challenge training set. Each patient volume includes single scan from each of the following MRI sequences: T1-ce, T2, T1 and FLAIR in *Figure 6*. Also provided in each patient volume is a segmentation volume with labels. The purpose of varying MRI imaging sequences is that each sequence is used to enhance visualization of distinct brain and tumor structures. The biomedical scans are volumetric, or 3-dimensional, and reproduced here is a 2-dimensional slice of each 3-dimensional volume.

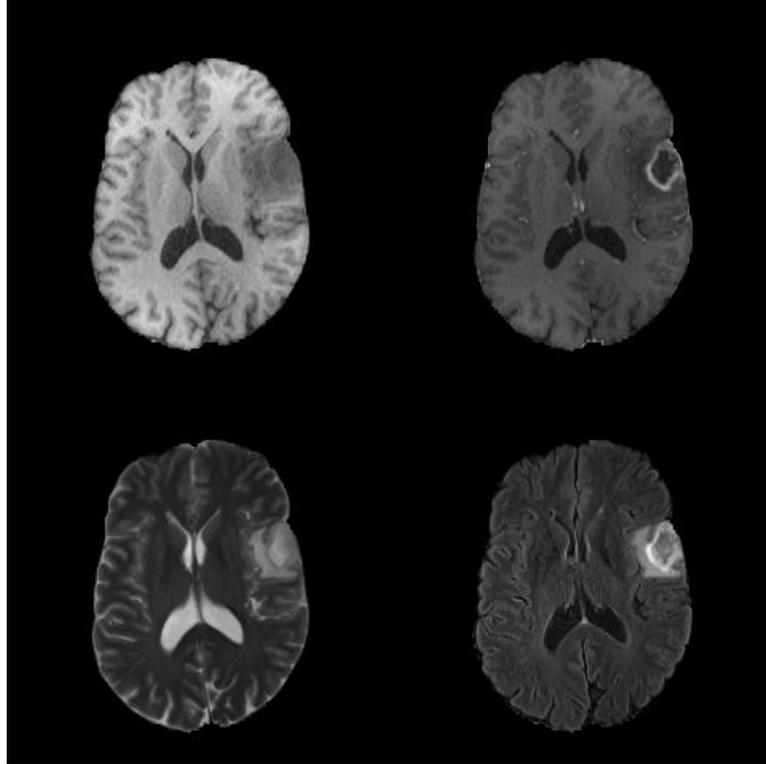


Figure 6: Input Data
 Top row from left: t1, t1-ce
 Bottom row from left: t2, flair

The ground truth segmentation for this patient volume is reproduced in *Figure 7* below layered on the FLAIR sequence for context visualization. These figures depict the original BraTS files, preprocessed uniformly with the BraTS preprocessing modules with the aforementioned skull stripping, co-registration to the same anatomical template and interpolation to the same resolution.

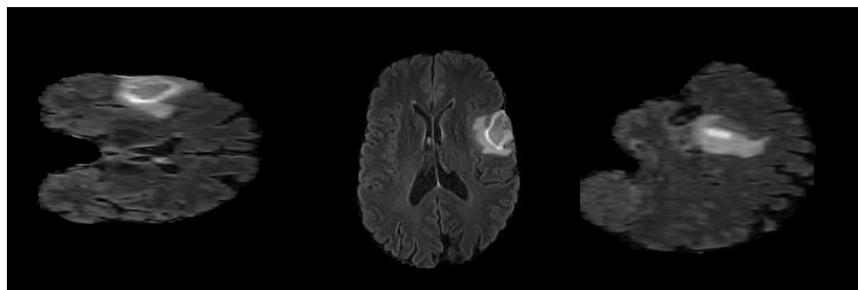


Figure 7: Dimensional Representation FLAIR
 From left: Coronal slice 104, horizontal slice 47, sagittal slice 79

7.2 COLOR SPACE PROJECTION

The images used for training and testing the deep network are represented in grayscale color space. Grayscale images represent information on the dimensions of luminance and intensity, where each pixel in the images can be explained with the numeric values representing these two factors. In this light, grayscale images are 2-dimensional, not in reference to volumetric space, but in reference to the two dimensions of values along the vectors of luminance and intensity which represent the image voxels. Image computing refers to the field of representing, processing and interpreting images. For computing purposes, and in the learning, network used here specifically, images are represented as matrices (or NumPy arrays). In these arrays, image pixels are indexed individually as numeric values. In grayscale computing, these numeric values represent information on the luminance and intensity of each voxel.

7.3 GRAYSCALE SEGMENTATION RESULTS

The original segmentation results obtained and reproduced in grayscale are reproduced below. *Figure 8* (left) depicts the segmentation result alone while *Figure 8* (right) depicts the segmentation result layered with a FLAIR volume of the same patient volume. This represents the input data for the color space projection model and serves as a base-line representation of the non-enhanced data visualization of the segmentation task. For the purpose of interpretation, it is useful to note here that the segmentation result is layered with the FLAIR representation in the identical position as the result is represented alone.

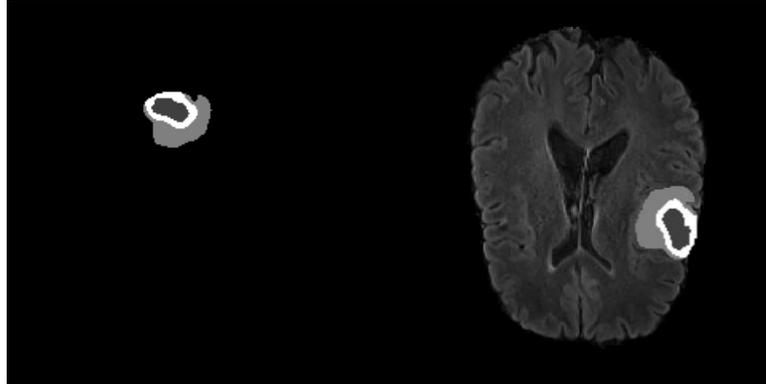


Figure 8: Segmentation Results

7.4 COLOR PROJECTION SEGMENTATION

The following approach converts the gray-scale results into color space using palettes descending from the HCL color model proposed in 2005 by Sarifuddin and Missaou [59]. The Hue-Chroma-Luminance model was designed to exploit the advantages of preexisting Red-Green-Blue and CIELAB (L, a*, b*) color spaces. HCL is a perceptually uniform color space, where perceptual uniformity refers to a measure of consistency over the perceived similarity/difference of sets of equidistant points across the color space [59]. This color space is constructed as a “natural representation of color models” centered around the physiological perception of the human eye [59]. Each coordinate (pixel value) in HCL space has attributes “Hue”, “Chroma” and “Luminance”. Hue is understood to refer to the full color spectrum as the “dominant color” perceived in a color experience. Chroma is the “colorfulness” attribute of a color experience, and Luminance is the “brightness” attribute [70]. The structure of HCL space and coordinate calculations centers on the reasoning that the average human visual experience reacts to color intensity logarithmically, rather than linearly as visualized in *Figure 9*, Reproduced from HCI wizard GUI [59, 99].

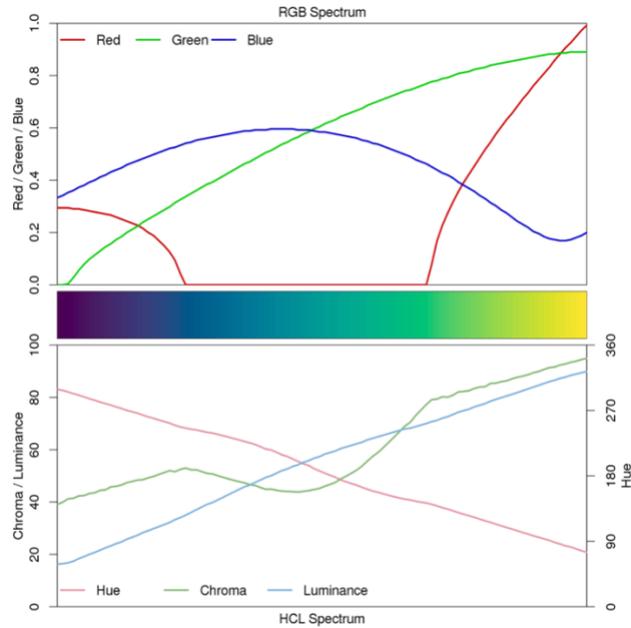


Figure 9: HCL Sequential Perceptually Uniform and RGB Color Spaces

To calculate the Luminance value of an HCL coordinate, the L value is calculated as a linear combination of Minimum and maximum RGB values (“black” is the min, RGB value and “white” the max) *Figure 10*.

$$L = \frac{Q \cdot \text{Max}(R, G, B) + (1 - Q) \cdot \text{Min}(R, G, B)}{2}$$

Figure 10: HCL Luminance Calculation

In this calculation, $Q = e^{\alpha\gamma}$ functions as a tuning parameter to account for variations in saturated hues and hues with a large white component. Chroma is calculated using *Figure 11* where elements (R, G, B) are combinations of RGB space components, including red-green, green-blue and blue-red. Finally, the Hue attribute ranges from -90° to 90° and is calculated as *Figure 12*. *Figure 13* provides a visual reference for Hue, Chroma and Luminance attributes and is reproduced here from [70].

$$C = \frac{Q \cdot (R - G) + |G - B| + |B - R|}{3}$$

Figure 11: HCL Chroma Calculation

$$H = \arctan\left(\frac{G - B}{R - G}\right)$$

Figure 12: HCL Hue Calculation

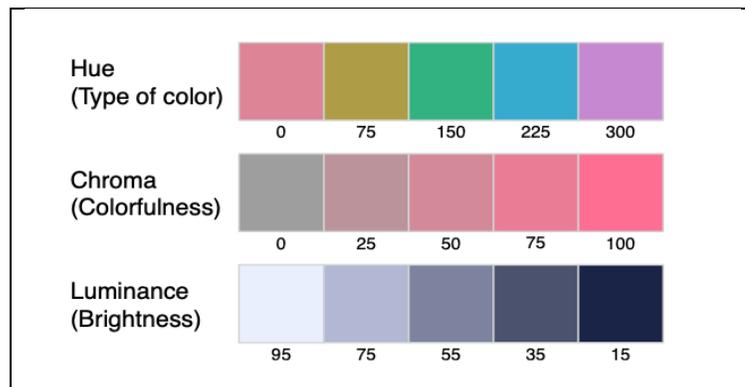


Figure 13: Hue, Chroma and Luminance

A colormap uses a palette of colors and mapping function to map data values to color [69]. Broadly, three classes of color palettes have descended from HCL space; qualitative, sequential and diverging. The three classes are differentiated by the trajectories of each the HCL components, with the Luminance value carrying the most importance in the class distinction. In a qualitative colormap, each color in the palette is given the same perceptual weight. This type of colormap distinguishes classes using distinct hues with equal chroma and luminance values [70]. This kind of map is useful for representing unordered categorical data. A sequential palette uses a monotonic function

trajectory for luminance values (increasing or decreasing) to represent ordered numerical data as a sequence. Without the chroma attribute, a sequential palette would correspond to grayscale [70] as visualized in *Figure 14*, reproduced from [101]. A diverging palette is modeled as a dual-direction monotonic sequential mapping function, where values diverge to two extremes from a central point [70]. A diverging palette is used to code numerical information around a central value [75]. *Figure 15* provides visual reference for comparison between the classes of colormaps and is reproduced from [100].

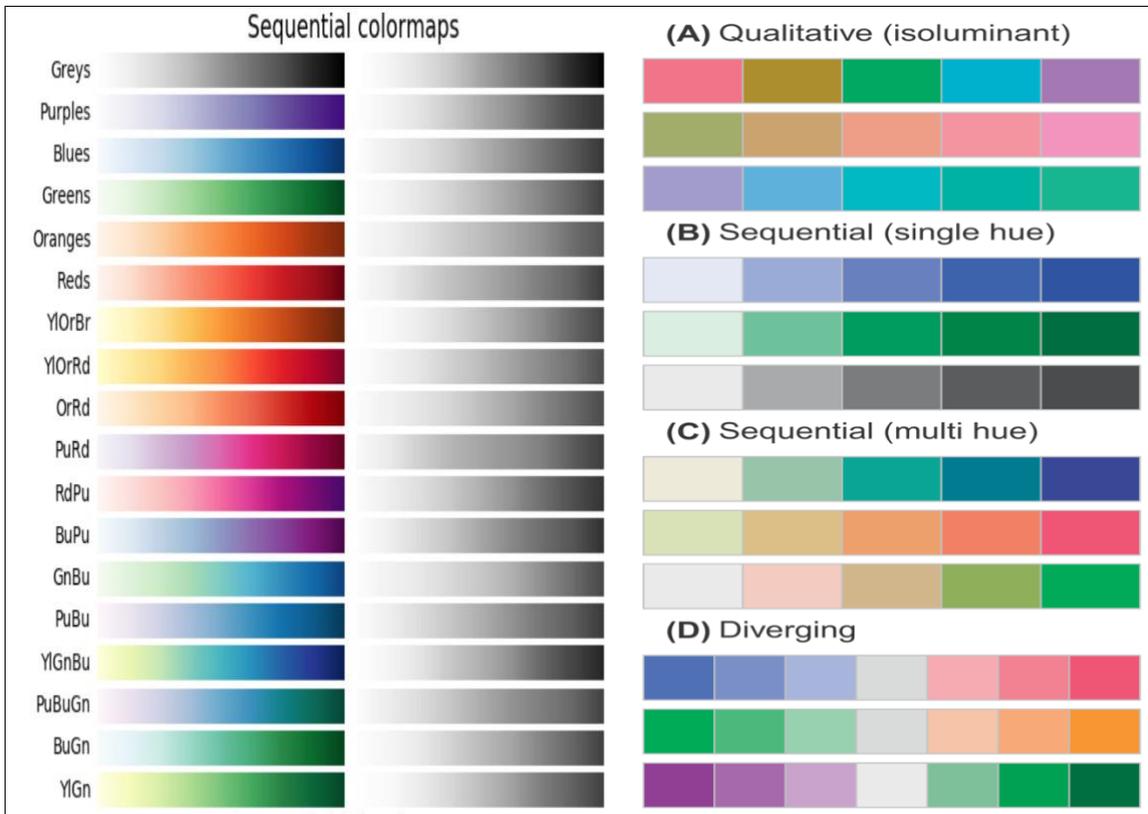


Figure 14: Sequential Colormaps

Figure 15: HCL Map Variations

An effective colormap for any task space is dependent upon the structure and composition of the dataset, with consideration for factors of numerical ordering, scaling, and spatial distribution [79]. A good colormap is a tool for effectively representing and

communicating data, and a bad map can lead to misrepresenting or misunderstanding data [69]. Principles of design in colormaps consider order, smoothness, uniformity and discriminative power [88]. A colormap with order progresses through color values with direction. Sequential and divergent are colormaps have order, whereas qualitative maps do not. Smoothness is the extent to which a colormap has identifiable boundaries. Perceptual uniformity is a measure of constancy between equidistant values, where two pairs of distinct values with the same difference measure also have the same impression of “sameness”. Finally, discriminative power refers to the range of the map, and is measured by a count of ‘just noticeable difference’ (JND) instances [88]. JND refers to the average minimum threshold by which two values can be consistently differentiated from one another.

7.5 COLORIZATION AND VISUAL ENHANCEMENT METHODS

In this study, the goal is to enhance the visualization of brain imaging data and the original data representation is grayscale. A brain/head MRI maps grayscale as a representation of signal intensity produced by the mechanical reaction of the tissue within the imaged region to the strong magnetic forces emitted by the MR scanner. Since distinct tissue compositions (more or less dense tissue/healthy or tumorous tissue) also have distinct mechanical reactions to the magnetic fields, signal intensity in the resulting image reflects anatomical features. In the same scope, grayscale images are represented computationally as numerical arrays. In this representation, anatomical features then also use numerical representation. An effective colormap for this data enhances the visualization while also retaining predetermined structure. The intent is to attain

information gain through enhanced visualization while preventing information loss in the translation.

Grayscale is a single-channel scalar representation for imaging data and contains only a single luminance value for each voxel. This means that grayscale is a single-hue sequential colormap where the hue is gray. Since the grayscale image has ordered numerical values, a sequential colormap can be used to translate the values into a multi-value color space with a uniform function for mapping one-dimensional luminance values into three-dimensional HCL coordinate space. The purpose of doing this is to maintain the inherent distribution of the data and represent the same distribution with more color values across over the same range to increase the discriminatory power of the color space by using values with the same distribution but with greater distance between them. Using a wider color-space distance between pixel values applies the concept of just noticeable difference, where the distance between pixels is increased to enhance the difference between visually similar color impressions. Mapping one-dimensional grayscale values into 3-dimensional HCL space exponentially affects the quantity of JND instances within the image.

For this study, five distinct sequential and perceptually uniform color-mapping functions are applied to volumes from the 3D-ESPNet segmentation results in an effort to enhance the visualization of the segmentation result to increase interpretability and establish model credibility. Each of the mapping functions is a variant using the HCL model, but the parameters of the methods differ slightly and can be used to enhance different image features, similar to the use of separate MRI sequences for distinct contrast enhancement. The colormaps used in this study are reproduced in *Figure 16* as

reproduced from [102] and come from the ‘viridis’ package, including ‘viridis’, ‘magma’, ‘plasma’, ‘inferno’ and ‘cividis’ [102]. These maps are designed to be both sequential and perceptually uniform. In addition, these maps account from the most common forms of color blindness (all 5) and color vision deficiency (the cividis map). The cividis map is a 2018 variation of the 2015 viridis package, adapted for color deficiency [66, 104].

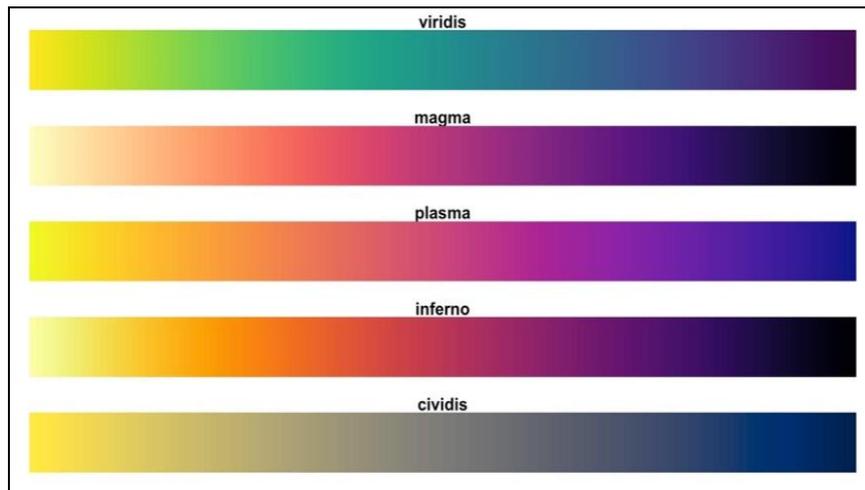


Figure 16: The Viridis Package

This color-mapping method uses Python 3.7 and Python libraries OpenCV, NumPy, Nibabel and Matplotlib for image processing. This method processes a single patient volume in each iteration. For each patient directory, 5 Nift1 files are read using Nibabel into separate 3-dimensional [240] x [240] x [155] Numpy arrays. A linear normalization is then used to scale the grayscale voxel values from (min)(max) into 32-bit floating point format range 0.0-1.0. The normalization is necessary for properly assigning input-output values to the colormap since the HCL conversion relies on RGB value calculations, and the 32-bit float RGB assignment requires 0-1 input. At this point, Matplotlib is used to store n=1000 equidistant values from each color space into a look-

up-table (LUT) for each space. For reference, the color spaces used are represented in with $n=3$ and $n=20$ colors respectively in *Figure 17* reproduced from Pypi Palettable [103] (the cividis map was not available through this source).

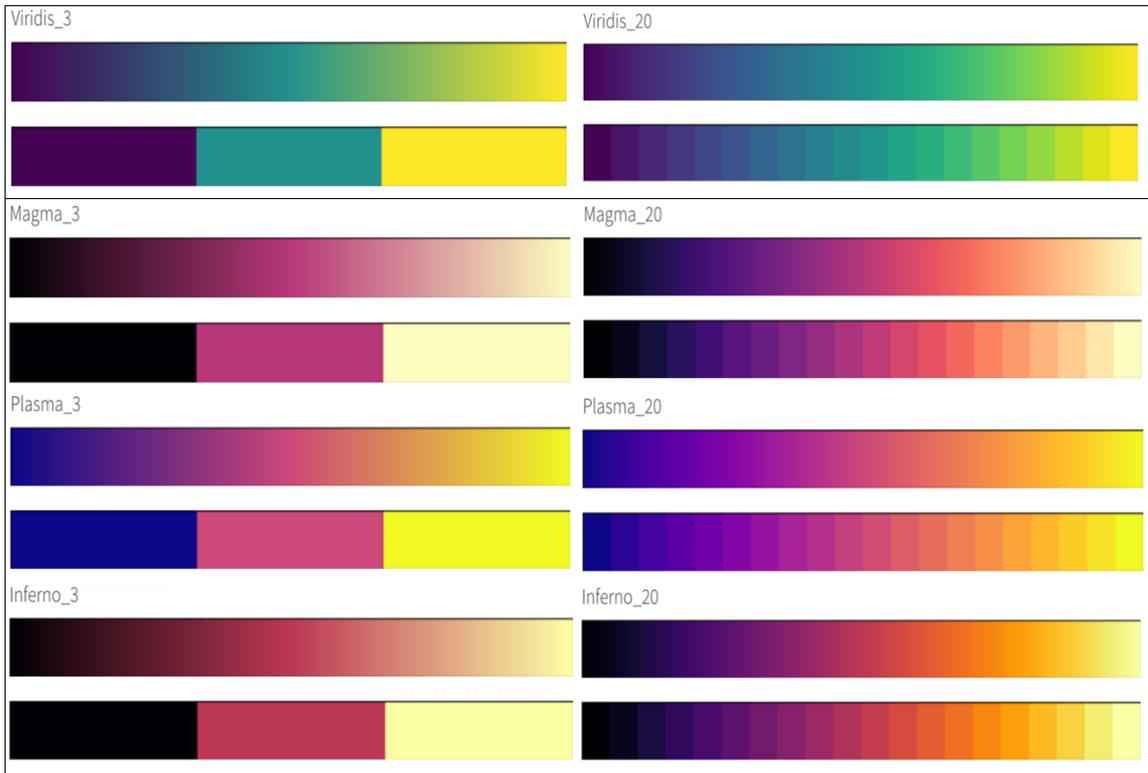


Figure 17: Colormap Reference

All 5 volumes (t1, t1-ce, t2, flair and segmentation result) are then input into the conversion function which uses the stored color values as reference to supply the values. With each color space ($n=5$) processing each image in a single patient volume ($n=5$), the output is 25 distinct volumetric scans, or 3,875 (25 images with a depth of 155 in dimension 3) two-dimensional image slices. The automated segmentation result is processed along with the MRI scans to this point. The result is 4 MR sequence images and 1 automated segmentation image which have undergone visual enhancement

individually. Slices from the resulting images from this stage of enhancement are reproduced in *Figure 18*.

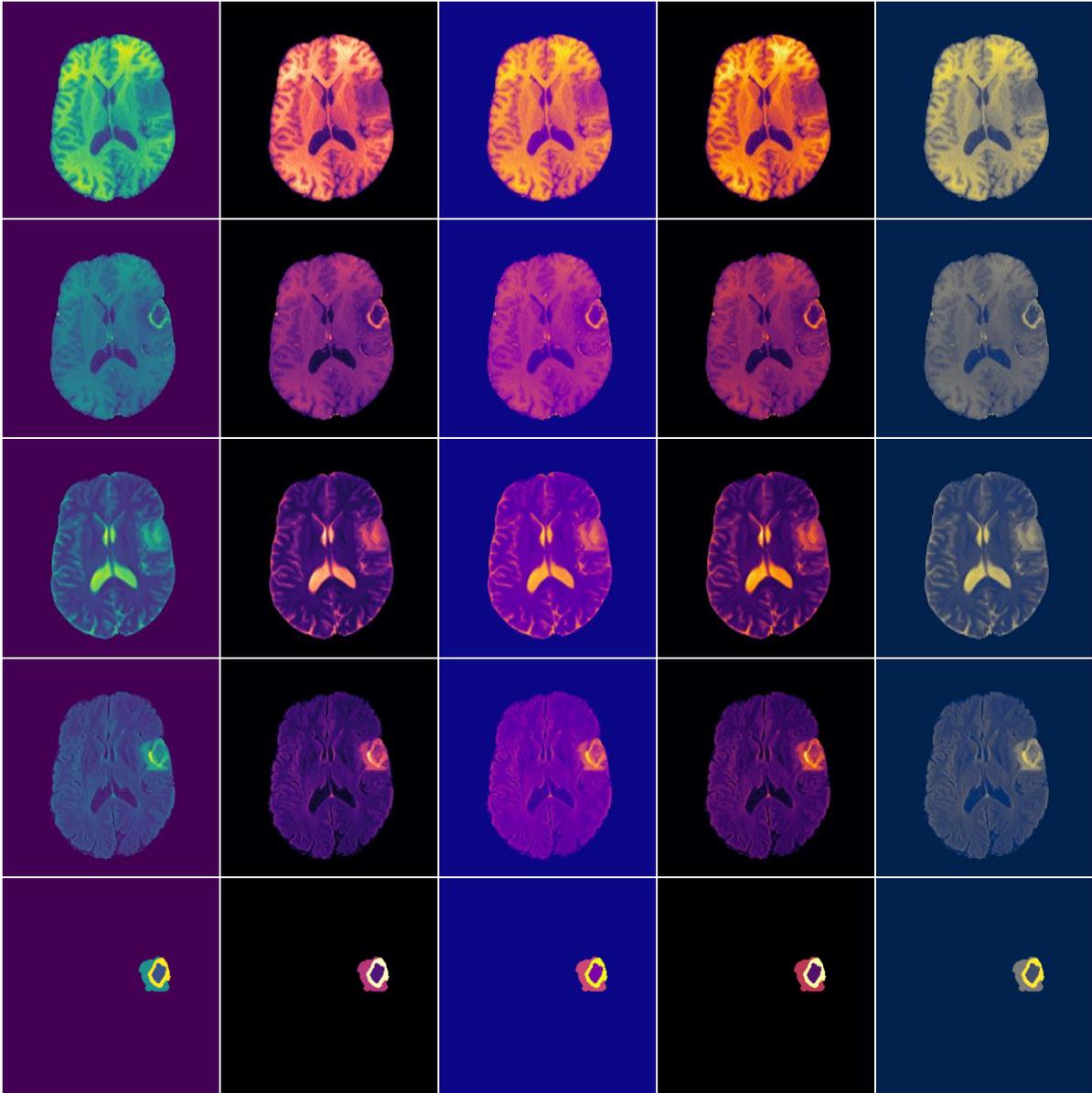


Figure 18: Color Space Results
from left → right viridis: magma, plasma, inferno, cividis
from top to bottom: t1, t1-ce, t2, flair, segmentation
image slice 71/155

This represents the first stage of enhanced visualization. At this stage, the tumor segmentation result and brain scans can be compared side-by-side with an enhanced view. The purpose of providing the segmentation result beside the full brain scan, without any overlay or label is to provide a baseline enhancement for expert analysis. The intention of this study is to address current limitations of automated tools for implementation in the real-world clinical sector. Current limitations of automated tools for brain tumor segmentation are associated with a lack of established clinical trust and approval of the validity and reliability of these methods. This stage of image enhancement seeks to establish clinical trust through quantitative analysis and ease-of-transition from manual to automated segmentation. The manual segmentation process has been a clinical “gold standard” for many years, and this process relies totally on expert image analysis and interpretation. Expert neuroradiologists endure many years of training and education to interpret these images and annotate both brain and tumor structures. For many clinical professionals, the manual segmentation process is not only a standard, but also a trustworthy process. It is a process though which an expert uses training and logical reasoning to assign tumor labels. The result of a manual segmentation session is not only a segmentation result, but also hours of reasoning and explanation to support or defend this result.

While the processes for manual and automated segmentation are fundamentally different, processing and evaluation of the produced segmentation results can be done with relative similarity. This allows the expert reading the segmentation result to judge the result similarly to how they would might judge their own segmentation work. In a traditional manual session, an expert will sit at a screen, view the patient MRI scans, and

carefully read the scans, identify points of interest and eventually determine tumor landmarks. With automated segmentation models powered by deep learning architectures, it is possible to streamline the often hours-long process of drawing the tumor segmentation. At the same time, it is not necessary to streamline the evaluation steps. With enhanced visualization of both the MR scans and predicted segmentation boundaries, resident expert clinicians can carry out segmentation evaluation and critiquing as would be done with manual segmentation, but with enhanced visual representation to support the segmentation.

7.6 ENHANCEMENT USING CONTOURS

In the second phase of image registration and enhancement, an edge detection algorithm is run on the predicted segmentation result for each of the patient volumes. Using OpenCV (open source computer vision for python) a Canny edge detection algorithm is run on the segmentation files. This edge detection uses thresholding of pixel values to establish edges. This is done for the grayscale segmentation files, ignoring the color segmentations temporarily. For each segmentation, a binary image mask is output containing the edges. From these binary masks, a function is then run to learn contours with an established hierarchy from the images. The hierarchy considers the segmentation regions separately so that contours can be drawn to reflect the three distinct segmentation labels. Once the contours are learned, the contours are then drawn onto the colored image using a line size of $n=0.25$ so as to highlight and define the segmentation regions without obscuring the enhanced data in and around the regions. This allows the clinician to visualize the segmentation result with exact reference to extent and boundaries

throughout the entire image volume (all 3 dimensions). The contour representations of the images from *Figure 18* are visualized in *Figure 19*.

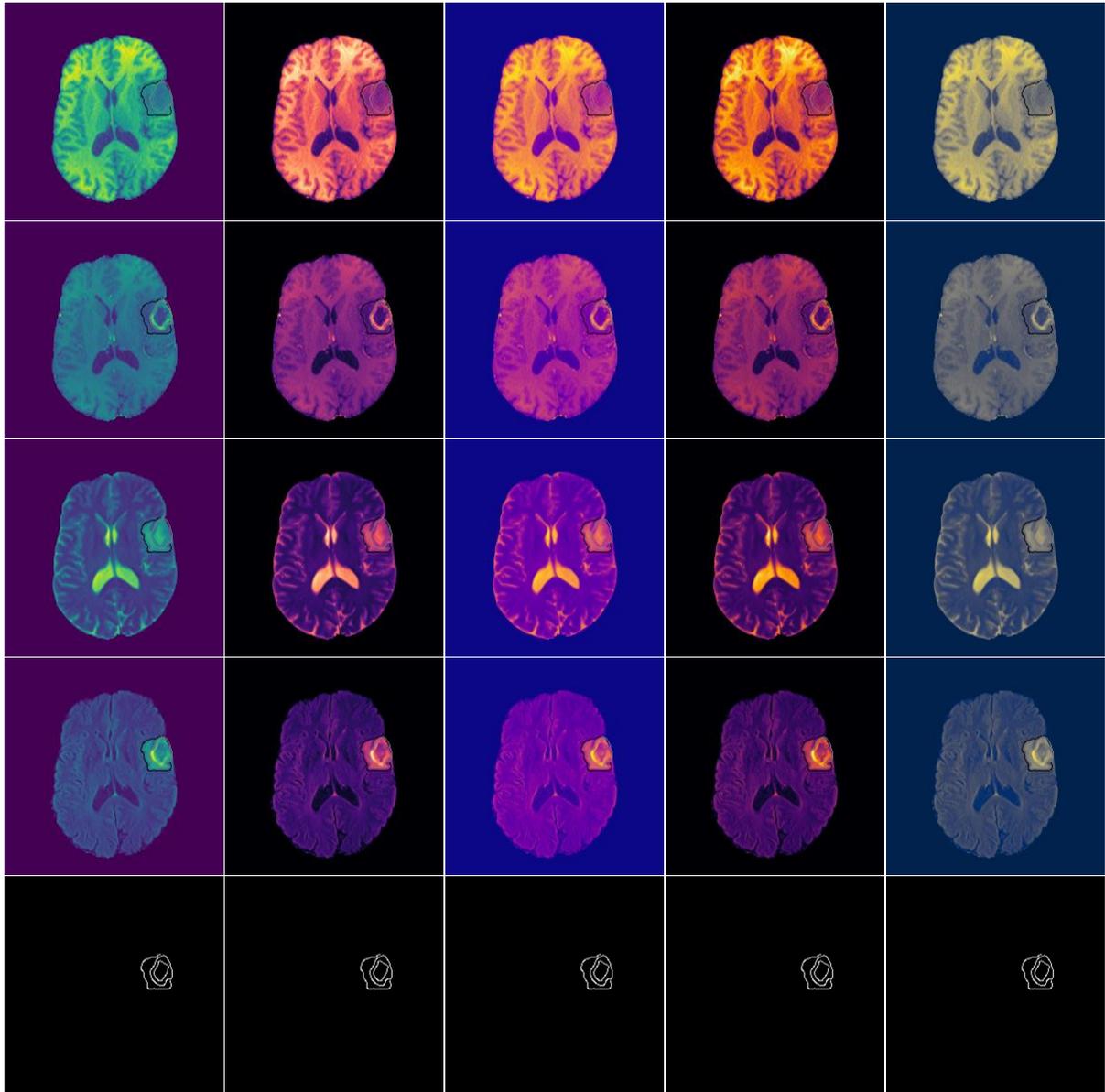


Figure 19: Contour Results
From left: viridis, magma, plasma, inferno, cividis
From top: t1, t1-ce, t2, flair, segmentation mask binary output edge detection

7.7 ENHANCEMENT WITH OVERLAYS

In the final visual perspective enhancement, the colorized segmentation result is presented as an overlay in a merged image. The base image in each of these cases is the color-mapped projection of the input image with an opacity of 100%. For the segmentation overlay, the background of the image is unweighted in this case. It was necessary to “unweight” the segmentation background so that the pixel values did not obscure/alter/negatively affect the values and presentation of the brain and relevant structures under the segmentation. This is because the segmentation dimensions (since they consider only the tumor regions) are much smaller than the dimensions of the entire brain, so merging the two images without accounting for the difference in size and extent would overlay background pixel values (black) onto the full extent of the brain not included in the tumor segmentation. This would negatively affect the context of the result and the integrity of the intensity distribution of the image pixels. To determine the most visually intuitive segmentation overlay representation, I experimented with segmentation overlays at opacity=25%, opacity=50% and opacity=75%. The resulting images were subjectively similar, but it was eventually decided that an opacity level of 75% in the overlay presented the clearest visualization of the automated segmentation result inside the brain scan, with a trade-off at the expense of any image data under/behind/obscured by the segmentation overlay. For this reason, the contour method is ideal for presenting the segmentation without information lost or obscured by the segmentation overlay. This is from the perspective that evaluating the segmentation result might require visualizing the scan both with and without the segmentation in order to determine if the segmentation labels are appropriate. This enhancement is visualized in *Figure 20* below.

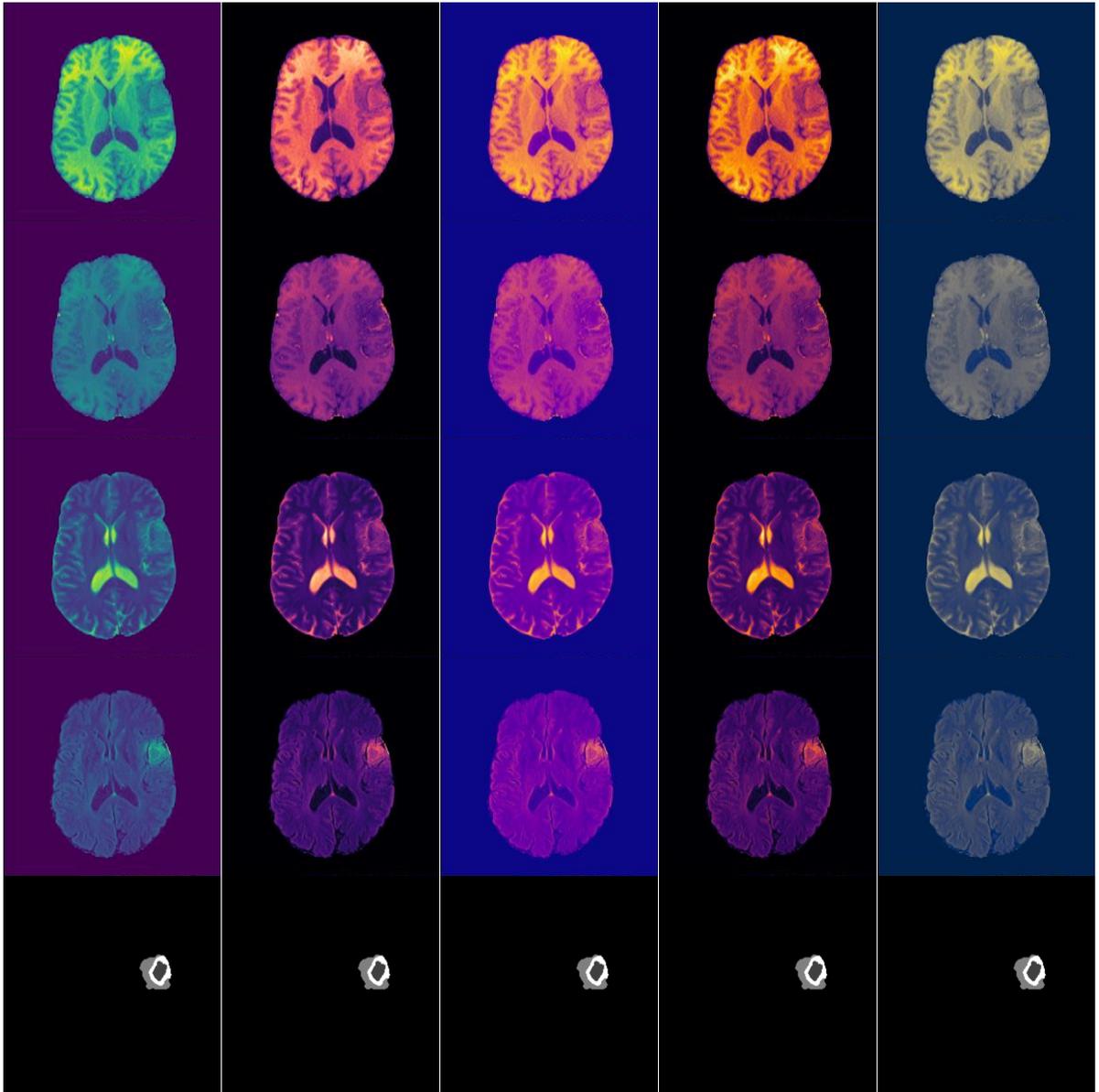


Figure 20: Segmentation Overlay
 From left: viridis, magma, plasma, inferno, cividis
 From top: t1, t1-ce, t2, flair, segmentation grayscale

CHAPTER 8

RESULTS AND CONCLUSION

The motivation of this study is to recognize clinical and general industry concerns for transparency and understanding in real-world deep learning applications. Significant research indicates transparency and understanding [34, 36, 82] as major limitations for non-expert use of deep learning applications. In the clinical domain, and specifically in high stakes diagnostic situations such as brain tumors, a premium is placed on transparency and interpretability. This dual-phase tumor segmentation process presented in this study is constructed to overcome the limitations of manual segmentation while also maintaining the advantages. The segmentation task trains a benchmark deep learning model, the 3D-ESPNet to segment brain tumor regions from labeled MR scans and tests model predictions. Relative to the clinical “gold standard” manual segmentation, state-of-the-art deep learning networks complete the segmentation task with improved time-efficiency and consistently, but lack the clinical transparency observed in manual segmentation. The registration and enhancement module proposes an approach for attributing transparency to deep learning brain tumor segmentation approaches as part of a semi-automated process which provides expert clinician with a standard by which to evaluate the automated model through enhanced and human-centered data visualization.

A standard for accepting or denying a diagnostic result need not rely on understanding the specific methodologies used to obtain the result. Pregnancy tests, blood

tests and strep throats swabs are prime examples of this idea. Clinicians do not need to understand the underlying chemical processes of these exams in order to interpret the results since the tests have established guidelines for interpretation of exam results. This study works to establish the same idea in the scope of semi-automated and automated models for biomedical analysis and brain tumor segmentation. The technique used in this study proposes a dual-part method for establishing clinical trust and reliability in consideration for automated diagnostic models. This approach first uses an automated model to obtain brain tumor segmentation results with increased speed and accuracy related to current methods and follows with a post-processing of the results for the purpose of expert evaluation as used in current practice.

Translating the gray-scale MRI segmentation results in color space takes advantage of expert radiologist segmentation knowledge by using the human visual spectrum as a guideline for enhancing the contrast of different tumor regions for manual visual analysis and interpretation. Allowing clinical experts to consider the segmentation results through a presentation which is tailored to take advantage of the human visual spectrum as opposed to grayscale allows the experts to differentiate between differing MRI signal levels which they would not have been able to do in a traditional manual segmentation setting.

Several studies have indicated the value of enhanced data visualization for interpretation of images. Studies have produced similar findings on the value of automated models in general and deep learning architectures specifically in achieving very high accuracy in image processing tasks with fast computation times. The contribution of this work is applying human-visual and perceptual scientific findings

along with benchmark model results for a combined semi-automated approach to establishing clinical trust in automated brain tumor segmentation models. The overarching goal is to establish a collaborative relationship between research and the realities of clinical situation to engineer methods to improve patient care and prolong prognosis of brain cancer patients.

8.1 FUTURE WORK

The major contribution of this work is a semi-automated brain tumor segmentation process which uses the image registration and enhancement module to post-process automated tumor segmentation predictions for expert evaluation. The registration and enhancement module creates a baseline monochrome representation of the MR scans and the segmentation, and then three additional enhanced visualizations for the purpose of integrating domain knowledge into the segmentation process. The first representation translates the monochrome standard representations into multi-chrome color space which is both sequential and perceptually uniform. The second uses thin contours to identify segmentation boundaries on the MR scans for perspective. The final representation overlays the segmentation prediction in monochrome on the multi-chrome MR representation. The enhanced visualizations are intended for use as a model evaluation metric, where clinical professionals with relevant domain knowledge evaluate model performance on the segmentation task. The purpose of enhancing the standard visual representations is to increase the explanatory value of the predictions according to the human visual spectrum.

As the end users of tumor segmentation tools, the clinical perspective is crucial to determining their value and overall use. Future work should be done to implement these methods into clinical practice in order to determine the value of these enhancements. This evaluation would provide crucial insight into the effective information gain of each enhancement both in combination and individually, such that more precise visual enhancements might be made in the future. The resulting data would then show which enhancement methods are more or less effective relative to the distinct MR parameters used. It may be found that colorization, contours and overlays vary in explanatory value for MR parameters, or that colorization methods vary in the same way. Further work could consider these methods over multiple brain tumor classifications in addition to the Glioma group, or other tumor locations throughout the body.

REFERENCES

- [1] R. Preetha and G. Suresh, "Performance analysis of fuzzy c means algorithm in automated detection of brain tumor," in *2014 World Congress on Computing and Communication Technologies*, 2014: IEEE, pp. 30-33.
- [2] A. Berger, "How does it work? Magnetic resonance imaging," *BMJ: British Medical Journal*, vol. 324, no. 7328, p. 35, 2002.
- [3] M. K. Abd-Ellah, A. I. Awad, A. A. Khalaf, and H. F. Hamed, "A review on brain tumor diagnosis from MRI images: Practical implications, key
- [4] J. Afseth, L. Neubeck, T. Karatzias, and R. Grant, "Holistic needs assessment in brain cancer patients: A systematic review of available tools," *European*
- [5] E. ERDIŞ, B. YÜCEL, S. Urvay, and A. MAYADAĞLI, "Overview of Survival and Related Parameters in Malignancies with Brain Metastasis," *TURKISH JOURNAL OF ONCOLOGY*, vol. 31, no. 3, 2016.
- [6] R. Ardhini and D. Tugasworo, "Epidemiology of primary brain tumors in dr. Kariadi Hospital Semarang in 2015-2018," in *E3S Web of Conferences*, 2019, vol. 125: EDP Sciences, p. 16004.
- [7] I. Alnaami, L. Sarhan, A. Alqahtani, A. Alghamdi, S. Alkhashrami, and O. Mostafa, "Does brain tumor epidemiology differ from place to another? Saudi single tertiary care center experience," *Biomedical Research*, vol. 29, no. 14, pp. 2982-2987, 2018.
- [8] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of MRI-based brain tumor segmentation methods," *Tsinghua Science and Technology*, vol. 19, no. 6, pp. 578-595, 2014.
- [9] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426-1438, 2013.
- [10] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "A review on brain tumor segmentation of MRI images," *Magnetic resonance imaging*, 2019.
- [11] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221-248, 2017.
- [12] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139-161, 2018.
- [13] S. Iqbal *et al.*, "Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation," *Microscopy research and technique*, 2019.
- [14] R. Thillaikkarasi and S. Saravanan, "An Enhancement of Deep Learning Algorithm for Brain Tumor Segmentation Using Kernel Based CNN with M-SVM," *Journal of medical systems*, vol. 43, no. 4, p. 84, 2019.
- [15] G. Soto-Romero, C. Escriba, P. Acco, E. Campo, and J.-Y. Fourniols, "Artificial Intelligence for Health: Towards more intelligence or more artificial medicine?" 2018.

- [16] G. S. Tandel *et al.*, "A Review on a deep learning perspective in brain cancer classification," *Cancers*, vol. 11, no. 1, p. 111, 2019.
- [17] H. Gao and S. Ji, "Graph u-nets," *arXiv preprint arXiv:1905.05178*, 2019.
- [18] N. Boyko, T. Sviridova, and N. Shakhovska, "Use of machine learning in the forecast of clinical consequences of cancer diseases," in *2018 7th Mediterranean Conference on Embedded Computing (MECO)*, 2018: IEEE, pp. 1-6.
- [19] M. Sendak *et al.*, "" The Human Body is a Black Box": Supporting Clinical Decision-Making with Deep Learning," *arXiv preprint arXiv:1911.08089*, 2019.
- [20] S. Schamoni, H. A. Lindner, V. Schneider-Lindner, M. Thiel, and S. Riezler, "Leveraging implicit expert knowledge for non-circular machine learning in sepsis prediction," *Artificial intelligence in medicine*, vol. 100, p. 101725, 2019.
- [21] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2018: ACM, pp. 559-560.
- [22] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use," *arXiv preprint arXiv:1905.05134*, 2019.
- [23] J. C. Ginestra *et al.*, "Clinician perception of a machine learning–based early warning system designed to predict severe sepsis and septic shock," *Critical care medicine*, vol. 47, no. 11, pp. 1477-1484, 2019.
- [24] M. Rowe, "An introduction to machine learning for clinicians," *Academic Medicine*, vol. 94, no. 10, pp. 1433-1436, 2019.
- [25] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical image analysis*, vol. 1, no. 2, pp. 91-108, 1996.
- [26] A. Ortiz, J. Górriz, J. Ramírez, D. Salas-Gonzalez, and J. M. Llamas-Elvira, "Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies," *Applied Soft Computing*, vol. 13, no. 5, pp. 2668-2682, 2013.
- [27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [28] R. Smith-Bindman *et al.*, "Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer," *Archives of internal medicine*, vol. 169, no. 22, pp. 2078-2086, 2009.
- [29] R. M. Hayward, N. Patronas, E. H. Baker, G. Vézina, P. S. Albert, and K. E. Warren, "Inter-observer variability in the measurement of diffuse intrinsic pontine gliomas," *Journal of neuro-oncology*, vol. 90, no. 1, pp. 57-61, 2008.
- [30] A. Ben Rabeh, F. Benzarti, and H. Amiri, "Segmentation of brain MRI using active contour model," *International Journal of Imaging Systems and Technology*, vol. 27, no. 1, pp. 3-11, 2017.
- [31] S. Lakare and A. Kaufman, "3D segmentation techniques for medical volumes," *Center for Visual Computing, Department of Computer Science, State University of New York*, pp. 59-68, 2000.
- [32] E. Papageorgiou *et al.*, "Brain tumor characterization using the soft computing technique of fuzzy cognitive maps," *Applied Soft Computing*, vol. 8, no. 1, pp. 820-828, 2008.

- [33] J. E. Cates, R. T. Whitaker, and G. M. Jones, "Case study: an evaluation of user-assisted hierarchical watershed segmentation," *Medical Image Analysis*, vol. 9, no. 6, pp. 566-578, 2005.
- [34] R. Elshawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in HealthCare A Comparative Study of Local Machine Learning Interpretability Techniques," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, 2019: IEEE, pp. 275-280.
- [35] B. Liu *et al.*, "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning aided decision support: three decades development course and future prospect," *arXiv preprint arXiv:1901.07858*, 2019.
- [36] T. Brito-Sarracino, M. R. dos Santos, E. F. Antunes, I. B. de Andrade Santos, J. C. Kasmanas, and A. C. P. de Leon Ferreira, "Explainable Machine Learning for Breast Cancer Diagnosis," in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, 2019: IEEE, pp. 681-686.
- [37] J. Zhang, Y. Wang, P. Molino, L. Li, and D. S. Ebert, "Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 364-373, 2018.
- [38] E. I. Zacharaki *et al.*, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609-1618, 2009.
- [39] E. I. Zacharaki *et al.*, "MRI-based classification of brain tumor type and grade using SVM-RFE," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009: IEEE, pp. 1035-1038.
- [40] R. Verma *et al.*, "Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images," *Academic radiology*, vol. 15, no. 8, pp. 966-977, 2008.
- [41] M. Sasikala and N. Kumaravel, "A wavelet-based optimal texture feature set for classification of brain tumours," *Journal of medical engineering & technology*, vol. 32, no. 3, pp. 198-205, 2008.
- [42] L. A. Ries *et al.*, "SEER cancer statistics review, 1975-2003," 2006.
- [43] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993-2024, 2014.
- [44] T. Lindow, J. Kron, H. Thulesius, E. Ljungström, and O. Pahlm, "Erroneous computer-based interpretations of atrial fibrillation and atrial flutter in a Swedish primary health care setting," *Scandinavian journal of primary health care*, vol. 37, no. 4, pp. 426-433, 2019.
- [45] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016: Springer, pp. 424-432.
- [46] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18-31, 2017.

- [47] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digital Signal Processing*, vol. 20, no. 2, pp. 433-441, 2010.
- [48] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.
- [49] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, p. 170117, 2017.
- [50] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, vol. 286, 2017.
- [51] E. A. AlBadawy, A. Saha, and M. A. Mazurowski, "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," *Medical physics*, vol. 45, no. 3, pp. 1150-1158, 2018.
- [52] M. Hayat, *Methods of cancer diagnosis, therapy, and prognosis: liver cancer*. Springer Science & Business Media, 2009.
- [53] A. Gupta and T. Dwivedi, "A simplified overview of World Health Organization classification update of central nervous system tumors 2016," *Journal of neurosciences in rural practice*, vol. 8, no. 04, pp. 629-641, 2017.
- [54] T. Komori, "The 2016 WHO classification of tumours of the central nervous system: the major points of revision," *Neurologia medico-chirurgica*, vol. 57, no. 7, pp. 301-311, 2017.
- [55] D. Schiffer, *Brain tumor pathology: current diagnostic hotspots and pitfalls*. Springer Science & Business Media, 2006.
- [56] D. N. Louis *et al.*, "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta neuropathologica*, vol. 131, no. 6, pp. 803-820, 2016.
- [57] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, vol. 286, 2018.
- [58] A. S. Adamson and H. G. Welch, "Machine Learning and the Cancer-Diagnosis Problem-No Gold Standard," *The New England journal of medicine*, vol. 381, no. 24, p. 2285, 2019.
- [59] M. Sarifuddin and R. Missaoui, "A new perceptually uniform color space with associated color similarity measure for content-based image and video retrieval," in *Proc. of ACM SIGIR 2005 workshop on multimedia information retrieval (MMIR 2005)*, 2005, pp. 1-8.
- [60] M. Attique *et al.*, "Colorization and automated segmentation of human T2 MR brain images for characterization of soft tissues," *PloS one*, vol. 7, no. 3, p. e33616, 2012.
- [61] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [62] "Childhood Astrocytomas Treatment (PDQ®)–Health Professional Version," *National Cancer Institute*. [Online]. Available:

- <https://www.cancer.gov/types/brain/hp/child-astrocytoma-treatment-pdq>. [Accessed: 09-Apr-2020].
- [63] "Glioma," *Mayo Clinic*, 04-Apr-2020. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/glioma/symptoms-causes/syc-20350251>. [Accessed: 09-Apr-2020].
- [64] M. da Fonseca and I. Samengo, "Derivation of human chromatic discrimination ability from an information-theoretical notion of distance in color space," *Neural Computation*, vol. 28, no. 12, pp. 2628-2655, 2016.
- [65] E. Furman-Haran, D. Grobgeld, R. Margalit, and H. Degani, "Response of MCF7 human breast cancer to tamoxifen: evaluation by the three-time-point, contrast-enhanced magnetic resonance imaging method," *Clinical cancer research*, vol. 4, no. 10, pp. 2299-2304, 1998.
- [66] J. R. Nuñez, C. R. Anderton, and R. S. Renslow, "Optimizing colormaps with consideration for color vision deficiency to enable accurate interpretation of scientific data," *PloS one*, vol. 13, no. 7, 2018.
- [67] T. Horiuchi, "Grayscale image segmentation using color space," *IEICE transactions on information and systems*, vol. 89, no. 3, pp. 1231-1237, 2006.
- [68] M. Mignotte, "Segmentation by fusion of histogram-based k -means clusters in different color spaces," *IEEE Transactions on image processing*, vol. 17, no. 5, pp. 780-787, 2008.
- [69] R. M. Karim, O.-H. Kwon, C. Park, and K. Lee, "A Study of Colormaps in Network Visualization," *Applied Sciences*, vol. 9, no. 20, p. 4228, 2019
- [70] A. Zeileis *et al.*, "colorspace: A toolbox for manipulating and assessing colors and palettes," *arXiv preprint arXiv:1903.06490*, 2019
- [71] J. Veltman *et al.*, "The additional value of three time point color coding in dynamic contrast-enhanced MRI of the breast for inexperienced and experienced readers," *European journal of radiology*, vol. 74, no. 3, pp. 514-518, 2010.
- [72] C.-M. Wang, C.-T. Kuo, and D.-P. Yang, "Application of color spaces fusion approach in MRI classification," in *2011 International Conference on Machine Learning and Cybernetics*, 2011, vol. 4: IEEE, pp. 1672-1677.
- [73] M. C. Popescu and L. M. Sasu, "Feature extraction, feature selection and machine learning for image classification: A case study," in *2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, 2014: IEEE, pp. 968-973.
- [74] C. J. Bradley, C. W. Given, and C. Roberts, "Disparities in cancer diagnosis and survival," *Cancer*, vol. 91, no. 1, pp. 178-188, 2001.
- [75] V. Kwok *et al.*, "Learning new color names produces rapid increase in gray matter in the intact adult human cortex," *Proceedings of the National Academy of Sciences*, vol. 108, no. 16, pp. 6686-6688, 2011.
- [76] I. Langer and A. R. Várkonyi-Kóczy, "Artificial lighting to enhance color perception," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, 2012: IEEE, pp. 1-5.
- [77] T. Li, Y. Wang, C. Chang, N. Hu, and Y. Zheng, "Color-appearance-model based fusion of gray and pseudo-color images for medical applications," *Information Fusion*, vol. 19, pp. 103-114, 2014.

- [78] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1537-1552, 2010.
- [79] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang, "Mapping color to meaning in colormap data visualizations," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 810-819, 2018.
- [80] H. Reese, "Understanding the differences between AI, machine learning, and deep learning," *Techrepublic*. Retrieved from <https://www.techrepublic.com/article/understanding-the-differences-between-ai-machine-learning-and-deep-learning>, 2017.
- [81] D. Sheth and M. L. Giger, "Artificial intelligence in the interpretation of breast cancer on MRI," *Journal of Magnetic Resonance Imaging*, 2019.
- [82] A. Holzinger *et al.*, "Towards the augmented pathologist: Challenges of explainable-ai in digital pathology," *arXiv preprint arXiv:1712.06657*, 2017.
- [83] L. Alamo and U. Fischer, "Contrast-enhanced color Doppler ultrasound characteristics in hypervascular breast tumors: comparison with MRI," *European radiology*, vol. 11, no. 6, pp. 970-977, 2001.
- [84] M. Angulakshmi and G. Lakshmi Priya, "Automated brain tumour segmentation techniques—A review," *International Journal of Imaging Systems and Technology*, vol. 27, no. 1, pp. 66-77, 2017.
- [85] K. Weiss, S. Stiving, E. Herderick, J. Cornhill, and D. Chakeres, "Hybrid color MR imaging display," *American Journal of Roentgenology*, vol. 149, no. 4, pp. 825-829, 1987.
- [86] J. Yuan, S. Chow, D. Yeung, and A. King, "A five-colour colour-coded mapping method for DCE-MRI analysis of head and neck tumours," *Clinical radiology*, vol. 67, no. 3, pp. 216-223, 2012.
- [87] J.-B. Lamy, B. Sekar, G. Guezennec, J. Bouaud, and B. Séroussi, "Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach," *Artificial intelligence in medicine*, vol. 94, pp. 42-53, 2019.
- [88] C. Ware, T. L. Turton, R. Bujack, F. Samsel, P. Shrivastava, and D. H. Rogers, "Measuring and modeling the feature detection threshold functions of colormaps," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 9, pp. 2777-2790, 2018.
- [89] R. A. Miller and A. Geissbuhler, "Clinical diagnostic decision support systems—an overview," in *Clinical decision support systems*: Springer, 1999, pp. 3-34.
- [90] C.-M. Wang, C.-T. Kuo, and D.-P. Yang, "Application of color spaces fusion approach in MRI classification," in *2011 International Conference on Machine Learning and Cybernetics*, 2011, vol. 4: IEEE, pp. 1672-1677.
- [91] A. Delavar, O. M. Al Jammal, K. R. Maguire, A. R. Wali, and M. H. Pham, "The impact of rural residence on adult brain cancer survival in the United States," *Journal of neuro-oncology*, vol. 144, no. 3, pp. 535-543, 2019.
- [92] C. J. Bradley, C. W. Given, and C. Roberts, "Disparities in cancer diagnosis and survival," *Cancer*, vol. 91, no. 1, pp. 178-188, 2001.
- [93] N. Nuechterlein and S. Mehta, "3D-ESPNet with pyramidal refinement for volumetric brain tumor image segmentation," in *International MICCAI Brainlesion Workshop*, 2018: Springer, pp. 245-253.

- [94] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 552-568.
- [95] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore, and L. Shapiro, "Y-Net: joint segmentation and classification for diagnosis of breast biopsy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018: Springer, pp. 893-901.
- [96] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end arguments in system design," *ACM Transactions on Computer Systems (TOCS)*, vol. 2, no. 4, pp. 277-288, 1984.
- [97] P. Mazumder, P. Singh, and V. Namboodiri, "Cpwc: Contextual point wise convolution for object recognition," *arXiv preprint arXiv:1910.09643*, 2019.
- [98] H. Zhou *et al.*, "Multi-scale dilated convolution neural network (MSD-CNN) for image artifact correction of limited-angle tomography," *IEEE Access*, 2019.
- [99] J. C. Fisher, R. Stauffer, and A. Zeileis, *HCL Wizard - Somewhere over the Rainbow*. [Online]. Available: <http://hclwizard.org/hclwizard/>. [Accessed: 08-Apr-2020].
- [100] R. Stauffer, G. J. Mayr, M. Dabernig, and A. Zeileis, "Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations," *Bulletin of the American Meteorological Society*, vol. 96, no. 2, pp. 203-216, 2015.
- [101] J. Hunter and D. Dale, "The Matplotlib User's Guide," *Matplotlib 0.90. 0 user's guide*, 2007.
- [102] S. Garnier, "viridis: default color maps from 'matplotlib'R package version 0.3. 4 <https://CRAN.R-project.org/package=viridis>, 2016.
- [103] S. Eichstädt and I. Smith, "PyDynamic Documentation," 2019.
- [104] N. Smith and S. van der Walt, "A better default colormap for Matplotlib," *SciPy2015*, 2015.