An Artificial Intelligence Approach to Tumor Volume Delineation

Finding the optimal treatment volume in radiotherapy – may multispectral MRI and deep learning segmentation be used as a tool in decision-making?

Marianne H. Hannisdal

Master Thesis
Master in Health Sciences
Radiographer / Medical dosimetrist

Department of Global Public Health and Primary Care
Faculty of Medicine

UNIVERSITY OF BERGEN

November 15th 2020

Acknowledgement

I would like to thank my main supervisor, Cecilie Brekke-Rygh, and my two co-supervisors Arvid Lundervold and Martha Chekenya Enger for giving me the opportunity to work with this project. Without your professionalism, supportive help, encouragement and continued guidance, this thesis would not have been possible. I am indebted to Martha Chekenya Enger for giving me the opportunity to join the team involved in the BORTEM-17 study. I would also like to thank my family and friends for their understanding and patience throughout the work with this thesis. I also owe great thanks to my leaders and co-workers at my workplace, Department of Oncology and Medical Physics, Haukeland University Hospital.

To all the aforementioned: Your support and our numerous professional discussions have been a great inspiration.

"*Education is not the filling of a bucket, but the lighting of a fire*" William Butler Yeats

Foreword

In my daily work at the Department of Oncology and Medical Physics, Haukeland University Hospital (HUH) as a medical dosimetrist, a part of my job is to prepare radiological image material for the oncologist to use for radiotherapy target delineation. The delineation determines the area that receives treatment, and it may have a large clinical impact on patient outcome and potential side-effects. The procedure might be challenging in many ways. In this work, I see first-hand how the radiotherapy software visualization presentation limits the image quality and hence the influences the image representation of the affected tissue, which again may lead to a poorer tool for treatment decision making. The applied software for delineation takes place is not high-end radiology software for diagnostic purposes, it is a radiotherapy planning software. Hence, it does not provide state-of the art image presentation tools like gamma grey-scale presentation function, nor does the oncologist have access to high-resolution screens for making the delineations. Even if the image-visualization and image-interpretation conditions and equipment might have limitations as to depict the true extent of the infiltrative malignant tissue in the brain, the oncologists are expected to delineate malignant tissue near critical organs with submillimeter precision. Overdiagnosis and overtreatment are well-known challenges in radiology. In the context of radiotherapy overtreatment means irradiating more tissue than needed for tumor control. When radiological material is subject to low specificity due to either imaging or visualization limitations, the chance for overtreatment increases. On this basis, the idea of my thesis came to life: if the conditions for qualitative interpretation are limited and deficient, could a quantitative machine learning approach be helpful?

Sammendrag

Bakgrunn: Stråleterapi målvolum blir tegnet manuelt av onkologer. Ved stråleterapi av glioblastom brukes 20 mm isotrop margin fra kontrastoppladende del av tumor fordi (i) det er vanskelig å spesifisere den ikke-kontrastoppladende delen av glioblastom i hjernevevet ved konvensjonell kvalitativ bildetolkning, og (ii) tilbakefall vanligvis kommer innenfor 20 mm fra primærtumor. Ved inntegning av målvolum innehar programvaren begrensninger i bildevisualisering, noe som gjør visuell tolkning vanskeligere for onkologene. Samtidig viser studier at det å begrense marginen til 10 mm ikke øker faren for hverken kantresidiver eller residiver utenfor målvolumet (1-3). Det er behov for en bedre metode for å spesifisere lokalisasjonen av tumorvolumet, slik at man kan begrense bivirkninger for pasientene.

Mål: 1: Å undersøke om tumorvolum segmentert med en robust kvantitativ maskinlæringsalgoritme er signifikant geometrisk korrelert med tumorvolum fra manuell inntegning, og dermed kan være nyttig som inntegningshjelp. 2: Å undersøke om maskinlæringsmodellen innehar prediktiv verdi i form av å inkludere fremtidig lokalisasjon for tilbakefall, i større grad enn dagens kliniske margin gjør.

Metode: Multispektral MRI fra 6 pasienter ble analysert med en tidligere publisert dyplæring algoritme, hvor output var to segmenter: a) kontrastoppladende, og b) ikke-kontrastoppladende tumordel. 1: Output ble sammenlignet med de respektive kliniske målvolumene ved bruk av Dice koeffisient. 2: Lokalisasjon av tumors tilbakefall ble sammenlignet med prediksjonen fra (i) klinisk målvolum, og (ii) maskinlærings output, ved Dice koeffisient.

Hovedresultat: 1: Gjennomsnittlig Dice koeffisient ble påvirket av gjennomgående volumetriske forskjeller, men viste en gjennomsnittlig korrelasjon på 0,54 med en Sann Positiv rate på 0,88 ($p<0,001$). 2: Ingen av metodene viste signifikant prediktiv verdi (p=ns).

Konklusjon: 1: For pasientene i denne studien var det en signifikant korrelasjon mellom det manuelt inntegnede volumet og volumet segmentert ved maskinlæring, målt ved Dice-koeffisient (DC=54%) ($p<0,001$). 2: Sammenligning av de respektive metodenes prediktive evne til å inkludere fremtidig plassering av tilbakefall viste ingen signifikant forskjell (p=ns).

Nøkkelord: Glioblastom, Kunstig Intelligens, Maskinlæring, Stråleterapi, Magnetisk Resonans

Abstract

Background: Radiotherapy target volume is determined manually by oncologists. For glioblastoma, isotropic margin of 20 mm from the visible tumor is used because (i) conventional MRI and interpretation methods have limitations as to specify the location of non-enhancing infiltrative glioblastoma, and (ii) recurrence commonly occur within 20 mm from the primary tumor. Software-related limitations for image visualization in the radiotherapy planning program, adds to the challenge of optimal visual interpretation and accurate delineation for the oncologists. Also, recurrence pattern studies show that limiting margins to 10 mm induce no increase in edge-recurrences or out-target recurrence (1-3). A better method is needed to specify the radiotherapy target volume and avoid unnecessary neurocognitive defects.

Aim: Q1: Investigate if a quantitative machine learning model for segmenting malignant tissue from multispectral MRI is correlated to the manual delineations, and thereby potentially feasible as oncologist support tool. Q2: Investigate if the machine learning model has longitudinal predictive value better than the standard manual clinical margin.

Methods: Multispectral MRI from six patients were analyzed using deep learning, from which the enhancing core and non-enhancing glioblastoma compartments were derived and compared (using the Dice-coefficient) to the manually delineated target volumes. The Dice-coefficient correlation between the recurrent tumor site, and (i) the clinical target volume and (ii) the machine learning-derived was compared, respectively.

Main results: Q1: The mean Dice-coefficient showed an agreement of 0.54 with a True Positive rate of 0.88 ($p < 0.001$). Q2: Neither methods showed significant predictive value (p=ns).

Conclusion: Q1: For the six patients in this study, there was a moderate concordance in the detected extent of malignant tissue comparing the two methods with Dice-coefficient (DC=54%) ($p < 0.001$). Q2: For comparing the two method's ability to include the site of tumor recurrence, no discrepancy between methods were detected (p=ns).

Key words: Glioblastoma, Artificial Intelligence, Machine Learning, Radiotherapy, Magnetic Resonance Imaging

Abbreviations

ABTC - Adult Brain Tumor Consortium

AI – Artificial Intelligence

ADC – apparent diffusion coefficient

ALARA – As Low As Reasonable Achievable

BraTS – Brain Tumor Segmentation

CE – contrast enhanced

CNR – contrast-to-noise ratio

CSF – cerebrospinal fluid

CTV – clinical target volume

CTVprimary – clinical target volume of the primary tumor site

CTVrecurrent – clinical tumor volume of the recurrent tumor

DC – Dice-coefficient

DICOM - Digital Imaging and COmmunication in Medicine

DIR – deformable image registration

DIPA – data protection impact assessment

DKFZ - Division of Medical Image Computing at the German Cancer Research Center

DPO – Data Protection Officer

DWI – diffusion weighted imaging

EC – enhancing core

EFRS - European Federation of Radiographer Societies

EPI – echo planar imaging

FLAIR - Fluid Attenuated Inversion Recovery

GBM - glioblastoma

GTV – gross tumor volume

HD-GLIOprimary – tumor volume on "MRIprimary" derived with machine learning using the HD-GLIO algorithm

HD-GLIOrecurrent – tumor volume on "MRIrecurrent" derived with machine learning using the HD-GLIO algorithm

HUH – Haukeland University Hospital

ISRRT - The International Society of Radiographers and Radiological Technologists

Interfractional – between fractions

Intrafractional – during the daily fraction of dose (treatment) is given, usually between one to several minutes

MGMT - O6 methylguanine DNA methyltransferase

ML – machine learning

MRI – magnetic resonance imaging

MRIrecurrent – MRI-dataset acquired at time of tumor recurrence

MRIprimary – MRI dataset acquired at time of primary tumor

NE – Non-Enhancing (tumor)

NIFTI - Neuroimaging Informatics Technology Initiative (fileformat)

OS – overall survival

PACS - Picture Archiving and Communication System

Peritumoral – the peripheric area around the tumor, or tumor cavity

PTV – Planned target volume

RANO - Response Assessment in Neuro Oncology

RECIST - Response Evaluation Criteria in Solid Tumors

RF – radio frequency

RT – radiation therapy

SAR – specific absorption rate

SNR – specific absorption rate

SUS – Stavanger University Hospital

Target volume – volume defining where dose should be set in radiotherapy (CTV/PTV)

Tumor cavity - cavitation in the brain from surgical removed tumor

VEGF - vascular endothelial growth factor

VPF - vascular permeability factor

VOI – Volume Of Interest

QIB - Quantitative Imaging Biomarker

Table of contents

List of figures

List of tables

1 Introduction

Glioblastoma (GBM) is the most common and aggressive primary brain tumor in adults, and every year there are almost 300 new cases diagnosed in Norway (4). Typical focal symptoms are related to rapidly increased intracranial pressure, including persistent headache, early morning nausea, vomiting, and neurological symptoms dependent on the size and location of the tumor (5). Magnetic resonance imaging (MRI) is an important part of the diagnostic process and tentative diagnosis is indicated by contrast enhanced MRI. The appearance of GBM on MR images is dependent on image weighting. Typical image characteristics reflects the underlying pathology are hypodense necrotic core that is surrounded by a peripheral ring of contrast-enhancing tumor on post-contrast T1 weighted images. This is often surrounded by an area of poorly circumscribed, vasogenic edema and infiltrative malignant tissue presented as hyperintense signal on T2/FLAIR-images (6). The final diagnosis is confirmed following histopathology and molecular genetic analysis of stereotactic biopsy or resected tissue sample (7). The prognosis for patients diagnosed with GBM is dismal, where median survival is 14.6 months and 5-year overall survival (OS) is approximately 4.5 % (8).

The primary treatment for GBM is maximum safe surgical resection, and limited-volume external beam radiotherapy (RT) with concurrent Temozolomide chemotherapy (8). Along with age, Karnofsky Performance Score (KPS), extent of tumor resection, the promoter methylation status of the DNA repair gene O^6-methylguanine-DNA methyltransferase (*MGMT*) is the strongest molecular prognostic factor in GBM, as well as predictive marker of response to Temozolomide. Patients harboring tumors with an unmethylated *MGMT* promoter have the lowest survival prognosis (9). Treatment efficacy is measured by OS and radiographic response in terms of progression-free survival (PFS) (10). In most cases, the tumor recur within 20 mm from the primary site (1, 11, 12). GBM is highly resistant to radiation, therefore tumor recurrence within the brain regions receiving full RT doses remains a challenge to effective management of the disease (13). Nevertheless, RT increases survival benefit over surgery alone (14), even though the radioresistant tumor biology is limits the prospects of tumor control with RT alone (13, 14), as a result, treatment is largely palliative. There is a need of improved treatment strategies for this patient group, but equally important is taking into account that the aggressive radiation treatment approach induce severe side-effects due to cerebral radionecrosis. The patients' quality of life during the remaining life expectancy should also be weighted highly. The challenge in radiotherapy

remains then to provide treatment that is both limited in terms of sparing normal tissue, and effective in terms of tumor control.

There is a strong consensus in both the literature and the community that target delineation is the weakest link in RT (15). A review article from 2016 shows great variations in target delineation practice between the European treatment institutions (11). In radiotherapy of glioblastoma there has not yet been established a definite contouring guideline consensus, but there are three major guidelines available; the Radiotherapy and Oncology Group (RTOG), the European Organization for Research and Treatment of Cancer (EORTC), and the European Society for Radiotherapy and Oncology Advisory Committee on Radiation Oncology Practice (ESTRO-ACROP). They vary in how they define the target volumes but they have in common that they advise the use of 20 mm margin around the morphologically visible pathology on anatomical MRI images. The margin of 20 mm is intended to cover the area of infiltrating tumor cells, and is based on histological post-mortem findings, as well as tumor recurrence studies which state that 78 % of recurrence occur within 20 mm distance from the surgical margin. Moreover, 56% occur within 10 mm of the presurgical margin of the primary tumor. A subsequent study of patients diagnosed with grade IV 4 astrocytomas treated with radiotherapy for the primary tumor showed that all recurrences were within 20 mm and that 10% had multifocal recurrences (16).

The rationale for extending the target area with the 20 mm margin is the prediction and expectation of tumor recurrence within this area. The margin is meant to cover pre-morphologic disease and non-enhancing malignant tissue that conventional imaging and interpretation methods fall short in specifying the spatial location. The consequence of poorly demarcated target area is damage to healthy brain tissue, such as radionecrosis with potential long-term neurocognitive deficits. An example of the size relation between the visible malignant tissue and the clinical target volume (CTV) is given in figure 1.

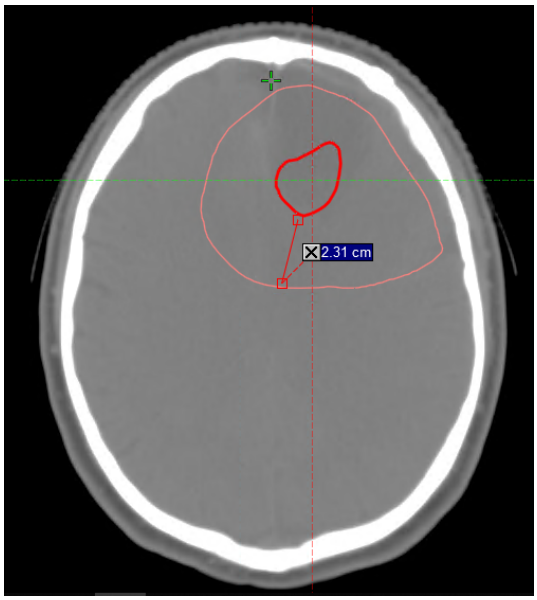


Figure 1 Schematic CT- treatment plan image showing the relationship between morphologically visible pathology (inner red circle) and safety margins including suspect areas and added margins (outer pink circle).The axial measurement shows a distance of 2,31cm from the visible tumor area (Gross Target Volume) to the area which covers the suspected infiltration (Clinical Target Volume). The CT-image is used due to its voxel-vise quantitative information in terms of attenuation coefficient, of which the dose-estimates in the treatment plan is calculated upon. Co-registered MRI is used for soft-tissue visualization support purposes. Illustration: Marianne H Hannisdal (in-study patient).

Traditionally, radiotherapy has been a profession characterized by large, open-field therapy volumes, often simulated on 2D-imaging, using large margins. Later years there has been a dramatic progress in personalized, high-precision radiation therapy planning and treatment conducting possibilities. However, imaging and visualization conditions in radiotherapy still have great potential for improvement, and are still not at the same level as for diagnostic radiology. At Haukeland University Hospital (HUH) the image visualization software used for radiotherapy target delineation (Eclipse Aria Oncology Information Systems (Varian, California, USA)) has limitations that is well known in the community. One example is lack of the gamma display function, which is a standard grey-scale shift tool in visualization software for radiologic interpretation purposes (17)[1]. Additionally, the lack of computer screens with diagnostic-image quality for oncologists, is in contrast to standard radiologist equipment. Consequently, the suboptimal linear gray-scale presentation (and more) makes the image interpretation and delineation task even harder for the radiation oncologists. The background for this study is a need to exploit the full potential in MRI images and detect the full in-image representation of glioblastoma for radiotherapy purposes.

[1] Internal HUH-reference, enclosed in appendix I.

The International Society of Radiographers and Radiological Technologists (ISRRT) and European Federation of Radiographer Societies (EFRS) has encouraged the role of Artificial intelligence (AI) in radiotherapy treatment planning in a joint statement in 2020:

> *Radiographers/radiological technologists working in radiation therapy should take advantage of AI technology in order to improve planning, deliver consistently high quality and personalized planning processes for patients and radiation therapy treatment, including organ(s) at risk identification, tumor segmentation, image matching and dose stratification*. (18)

The incorporation of AI and machine learning (ML) in MR neuro-oncology imaging could be feasible as a decision support tool for the oncologist in non-biased delineation of the tumor volume, as algorithms have a potential to distinguish voxel-vise grayscale nuances and not limited by a definite grayscale-range perception ability like humans are. Algorithms have the ability to process information across a spectrum of image sequences simultaneously, thereby extracting characteristic multispectral image features. The feasibility of AI as a method for interpretation of MRI images is given by the very nature of MRI images themselves: they are not merely pictures, they are datasets.

Moreover, the Norwegian Directorate of Health has currently sent new guidelines out for consideration, where they emphasize that large radiotherapy target volumes should be avoided as large volume radiotherapy induce increased risk of radiation induces neurotoxic damage (19). This is based on studies showing safe reduction from 20 mm to 10 mm in treating glioblastoma, without increased risk of recurrence (1-3).

This study makes use of quantitative radiological characteristics in multispectral magnetic resonance imaging (MRI), also called quantitative radiological biomarkers. The study aims to investigate if an artificial intelligence machine learning model for segmenting malignant tissue from multispectral MRI is geometrically correlated to the manual delineations, and thereby potentially be feasible as an oncologist support method to segment true malignant tissue in multispectral MRI with high specificity. The target margin used for radiotherapy serves as a representation of radiologic uncertainty, but it also serves as a prediction of recurrence. This study also aims to investigate if the machine learning model has longitudinal predictive value in terms of including the future tissue site of recurrence better than the standard clinical margin. For radiation treatment planning purposes, improving these aspects

may lead to better tumor control as well as improved quality of life by sparing normal tissue in accordance to the As Low As Reasonable Achievable (ALARA)-principle.

2 Theoretical aspects

2.1 Gross anatomy of the brain

The human cerebrum consists of subcortical white matter containing axon nerve fibers, and cortical grey matter that consists of the cell bodies or soma of the nerve cells. The cortex is divided into several parcellations in which nerve cells have dedicated tasks (20), as exemplified in the cortical homunculus-figure below. The spatial location of a brain tumor as well as the spatial extent of radiation treatment have direct impact on patient symptoms and functionality, in accordance to the regional nerve cells dedicated tasks.

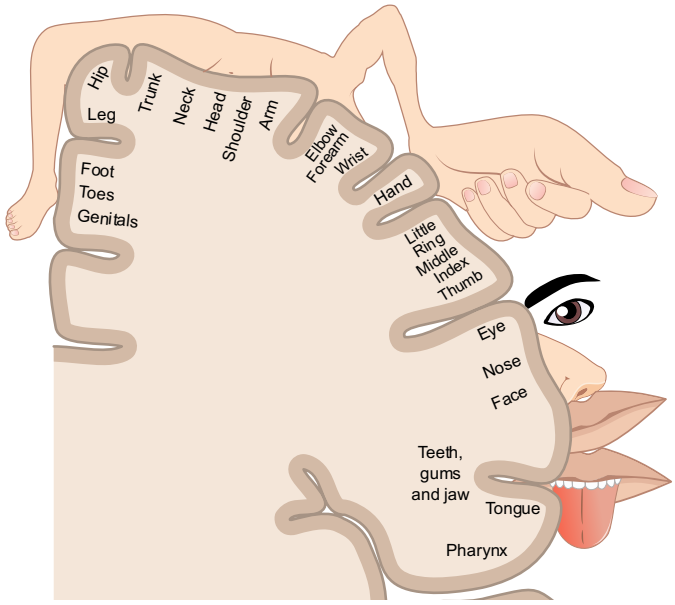


Figure 2 Example showing a cortical homunculus. This illustrates a neurological map of the areas controlled by the respective parts of the brain cortex. Illustration: Wikipedia commons (21)

The subcortical white matter, that consist of the nerve cells` myelin-covered axons, makes up the communicative "freeway" between the nerve cell bodies in the cortex and the nerve cells in the rest of the body.

Four different types of glial cells in the brain surround and support the nerve cells, namely, oligodendrocytes, astrocytes, ependymal cells and microglial cells. These glial cells outnumber the nerve cells in the brain, and provide the nerve cells with their environmental and existential needs (22). It is believed that astrocytes and oligodendrocyte cells as well as neural stem cells, may represent cells of origin for GBM (23).

The brain is supplied with blood from the internal carotid arteries and the vertebral arteries, and is protected from the influence of the diverse substances in the main blood stream by the blood-brain-barrier (20).

A comprehensive anatomical and physiological description of the human brain is beyond the scope of this study. The brain defines who we are, thus, malignant disease herein has destructive potential for cognitive, emotional, behavioral and motoric functionality. The symptoms the patient experiences are mainly dependent on the location, size and extent of the tumor. Figure 3 shows an image of the human brain.

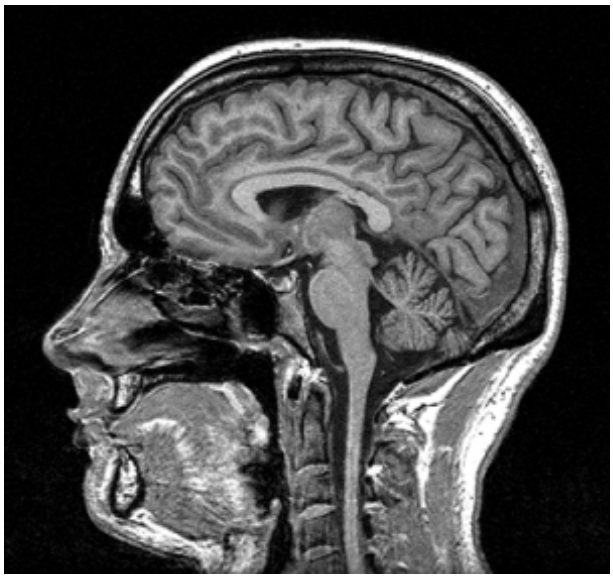


Figure 3 The human brain, sagittal view T1-weighted image showing cerebrum, corpus callosum, thalamus, pons, cerebellum, brainstem and proximal spinal cord . Illustration: Marianne Hannisdal (in-study patient)

2.2 Glioblastoma

Glioblastoma is the most malignant brain neoplasm, assigned the highest malignancy grade IV by the World Health Organisation (WHO) (24, 25). It was formerly known as Glioblastoma Multiforme (GBM), reflective of its morphological complexity, intra- and intersubjective genetic heterogeneity (24). Precisely, the heterogeneity in GBM extends from tissue- to cellular- and molecular level, and GBM can arise as a primary neoplasm (without prior evidence of a less malignant lesion) or as a secondary malignancy (evolving from lower grade gliomas) (25). A hallmark feature of GBM is necrosis and microvascular proliferation, which are defined by WHO classification system (26), see table below. This classification

was upgraded to incorporate molecular genetic characterization to the histological grading in order to define the myriad of tumor entities. The molecular classification is defined by whether or not the isocitrate dehydrogenase genes (IDH1/2) mutant, not mutated (wildtype) or not otherwise stated (NOS), or indeed, whether they represent diffuse midline glioma, H3 K27M-mutant among other molecular entities (25).

Table I WHO classification of Diffuse astrocytic and oligodendroglial tumors

WHO grade	Type
II	**Diffuse astrocytoma** ○ IDH-mutant ▪ gemistocytic astrocytoma ○ *IDH-wildtype* ○ NOS **Oligoastrocytoma** NOS **Oligodendroglioma** ○ IDH-mutant, 1p19q co-deleted ○ oligodendroglioma NOS
III	**Anaplastic astrocytoma** ○ IDH-mutant ○ *IDH-wildtype* ○ NOS **Anaplastic oligoastrocytoma** NOS **Anaplastic oligodendroglioma** ○ IDH-mutant, 1p19q co-deleted ○ anaplastic oligodendroglioma NOS
IV	**Glioblastoma** ○ IDH wildtype ▪ giant cell glioblastoma ▪ gliosarcoma ▪ *epithelioid glioblastoma* ○ IDH mutant ○ NOS **Diffuse midline glioma, H3K27M-mutant**

The main hallmark features of different grade gliomas, revised definitions by WHO in 2016 (27, 28), including isocitrate dehydrogenase genes (IDH1/2), 1p/19q codeletion (a genetic loss event), and Histone H3 K27M, which is a mutation in the H3F3A gene, encoding for histone H3.3
NOS= not otherwise stated

The risk factors for developing gliomas has not been fully explained, but exposure to ionizing radiation seems to be the main risk factor, along with increasing age (29). Genetic variations have also been weakly associated with glioma (30).

In adults, GBM tends to occur most often in the ages 50-70 years, with a slight overweight of men (31), but can occur at all ages (26). GBM occurs most often in the four lobes of the brain; frontal (25%), temporal (20%), parietal (13%) and occipital (3%), whereof the remaining 39% cases occurs in the brainstem, cerebellum and spinal cord (32). GBM is multifocal in 13 % of cases, meaning there is more than two lesions. This can occur as disseminations to meninges, diffuse disease (no clear borderline between lesions) or satellite nodules that are clearly noncontiguous with the primary lesion. Further, it is radiologically well established that non-enhancing tumor infiltration on microscopic level extends beyond the visual pathologic expression on anatomical MRI-images (26). The radiological expression of glioblastoma will be furtherly described in section 2.3.4.

The cell biological and pathophysiological details of GBM is beyond the scope of this thesis, but some main features are described in the following section. Glioblastoma initiation is hypothesized to take place in the neural stem cells or mature glial cell through initial mutations in tumor suppressor genes and protooncogenes that destabilize the genome and confer cell proliferative advantage (33). Further on, tumor micro-environment play an important role in pathological impact as tumor-associated parenchymal cells (vascular, microglia, peripheral immune cells and neural precursor cells) directly interact with GBM cells (33). Advanced interactions triggered by exponential increase in tumor nutrient and oxygen demand facilitates tumor growth and simultaneously co-opt the integrity of normal vasculature. This deprives normal vascularization in neighboring tissue, compromising normal cell function and subsequently, making this neighboring tissue more vulnerable to tumor invasion. Such dismantling of normal vascularization integrity further facilitates tumor growth in a diffuse and poorly delineated way, while simultaneously increasing treatment resistance by inducing hypoxia (34). All patients experience disease progression, and for approximately 70% of individuals, recurrence occurs within the first year after diagnosis (26). Such rapid recurrence is made possible by rapid cancer cell invasion to distant sites within the brain parenchyma. Tumor cells and glioma stem cells, are able to migrate to different brain regions on pathologically fenestrated blood vessels, allowing colonization of healthy brain tissue (34). This tumor cell invasion subsequently causes new

neovascularization at the new sites of renewed tumor expansive growth , that induces vasogenic edema and increased intracranial pressure which further promote malignant progression..

2.3 Imaging for treatment planning

Imaging used in treatment planning include both plan-dedicated imaging acquired with RT-patient positioning, as well as diagnostic imaging co-registered onto the plan-dedicated image material.

In order to administer safe radiation therapy, the dose to be delivered must be planned and calculated in detail in order to deliver the optimal therapeutic doses to the target volume while protecting organs at risk. Treatment planning is traditionally based on Computer Tomography (CT) images, however, MRI has superior soft-tissue resolution compared to that of CT. MRI is thus preferentially used as a support tool for target volume delineation by the oncologist and for organs-at-risk (OAR) delineation by the medical dosimetrist. CT and MRI imaging modalities provide personalized anatomical maps for dose-calculation and tumor visualization purposes, respectively, on which one can plan how to deliver highly conformal doses to a 3D target volume (35). The combination of different image modalities is called multimodal images.

Positron Emission Tomography (PET) scan can also provide additional imaging information for diagnosis and treatment planning. The most commonly used PET tracer is the 2-deoxy-2-(18F) fluoro-D-glucose (^{18}FDG), whose function is based on glycolytic metabolism. Newer amino acid tracers like hybrid ^{11}C-methyl-L-methionine (^{11}C-MET) PET/MRI and O-(2-(18)F- fluoroethyl)-L-tyrosine (^{18}F-FET) have demonstrated significant accuracy in imaging metabolically active tumor compartments (36). PET and PET-MR are however, very limited resources in our region, thus their in-depth description is beyond the scope of this study.

The imaging information provided by CT and MRI for treatment planning is the basis for target and OAR delineation. Hence, the accuracy of the treatment plan is limited to the sensitivity and specificity of the diagnostic imaging modality.

2.3.1 Computer Tomography

Computer Tomography (CT) is the primary imaging modality that is utilized in radiation therapy planning as the various tissues' Hounsfield Units (HU) presents a well-documented

and quantitative measure of the tissues' ability to absorb, hence attenuate the energy from the radiation. HU is a description of the linear attenuation coefficient, which quantifies how much a matter attenuates the radiation, and the gray scale is calibrated relative to the attenuation of air (-1000 HU) and water (0 HU). HU-values in the body's various anatomical tissues is well described, e.g. fat: ~-80 HU, muscle: ~30 HU, and all such values can be extrapolated for use in dose calculation (37). The pixel value is expressed by Hounsfield Unit (HU), and is found by the linear attenuation coefficient μ_x relative to water μ_w for every pixel in the volume by this relation (35):

$$CT\ pixel\ value\ (HU) = (\mu_x - \mu_w)/\mu_w$$

Thus, CT images provide important dose-attenuation information despite poor soft-tissue contrast.

2.3.2 Magnetic Resonance Imaging

MRI is the main diagnostic imaging tool for GBM, and diagnostic MRI are often co-registered onto the plan-dedicated CT. The Norwegian Directory of Health recommend that MRI used for radiation therapy target delineation should not be older than four weeks (19), while ESTRO recommend no older than two weeks (38).

MRI uses powerful magnets, commonly within a range of 1.5-7 Tesla to create hydrogen nuclear spins with a precession frequency according to the applied radiofrequency pulse and the tissues intrinsic molecular structure within each voxel. The product of MRI is a spatial map of the net magnetization in an image matrix (39), represented by greyscale in an image. A set of image parameters such as (but not limited to) repetition times (TR), echo times (TE), flip angle, and (if applicable) inversion times (TI) provide specific image weightings often called MR-sequences. The combination of various sequences is called multiparametric, or multispectral MRI. The minimum consensus recommended sequences for both 1.5 T and 3 T MR systems, according to the UCLA Neuro-Oncology Program and UCLA Brain Tumor Imaging Laboratory (BTIL) (40), include:

a) parameter-matched precontrast and postcontrast inversion recovery-prepared, isotropic 3D T1-weighted gradient-recalled echo

b) axial 2D T2-weighted turbo spin-echo (TSE) acquired after contrast injection and before postcontrast 3D T1-weighted images to control timing of images after contrast administration

c) precontrast axial 2D T2-weighted fluid-attenuated inversion recovery (FLAIR)

d) precontrast, axial 2D, 3-directional diffusion-weighted images,

whereof a), b) and c) is the most commonly used image series for brain tumor large scale machine learning purposes, according to both the widely known Brain Tumor Segmentation (BRATS) challenges as well as the HD-GLIO-algorithm used in this study. The latter will be more thoroughly explained in section 2.6.3. In the following section I will give a brief introduction to the four MRI image sequences used in our study.

2.3.3 MRI image sequences used in this study

T1-weighted images

T1 relaxation time, also called T1 recovery, is defined as the time it takes a tissue to recover 63 % of its longditunal magnetization, as illustrated in the figure below. This takes place as the excited spins deposit energy to the surrounding tissue, a process called spin-lattice relaxation (41). The different tissues have different microstructural compositions, quantities and concentrations of water, as well as different types of macromolecules and iron-composites, which all affect relaxation characteristics (42, 43). High concentrations of water in tissues increase their T1-time, while fat and iron reduce the T1-time. Thus, myelin-rich white matter has significantly shorter T1-time than grey matter that contains more water, as does cerebrospinal fluid (CSF) which has longer T1-time. The difference in relaxation times between two tissues is illustrated in the figure 4 below, where tissue-*a* is representative of fat and tissue-*b* water-rich tissue.

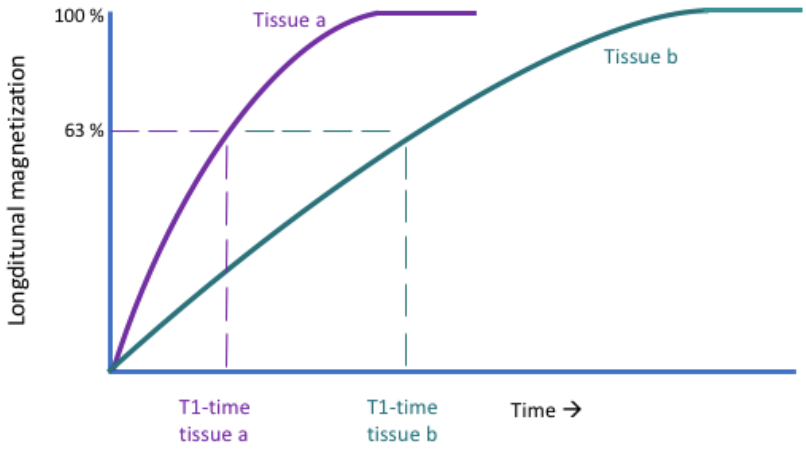


Figure 4 T1 recovery curve for two different tissues, illustrating how tissue a and b spend different amount of time to recover 63 % of the longditunal magnetization, and will therefore induce different voltages at signal readout. Illustration: Marianne H. Hannisdal

T1-weighted images with gadolinium contrast enhancement

Image contrast and hence image information may be increased by the use of contrast agents such as gadolinium chelates. The most commonly used agents are extravascular and extracellular agents. In a healthy brain, the contrast agent does not cross the intact blood-brain-barrier (BBB) and stay intravascular, but in tumors the contrast agent crosses, and are able to depict, a disrupted BBB. Gadolinium contain paramagnetic material with unpaired electrons that changes the tissues magnetic susceptibility and shortens the T1-relaxation time (35). Furthermore, contrast enhancement denotes neovascularization in the interstitial tissue and remaining neoplastic tissues, often seen in the outer part of the macroscopic tumor, or tumor bed (44, 45). This feature has become a radiologic biomarker, as vessels in the brain are not permeable for gadolinium-based contrast agent due to the BBB unless a pathological situation is present. The physiological permeability of the blood vessels' endothelial layer can be measured using dynamic contrast enhanced imaging (46). A T1-weighed image, and a contrast enhanced T1-weighted image (CET1) is exemplified in figure 5 below.

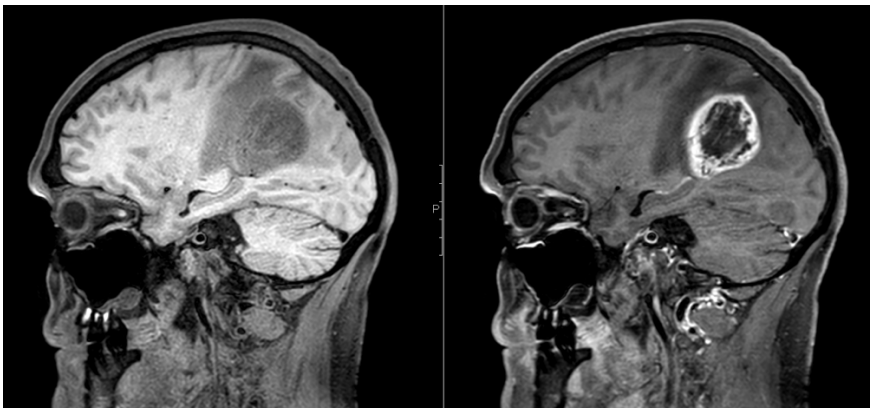


Figure 5 Right: sagittal view T1-weighted image. Left: sagittal view gadolinium enhanced T1-weighted image. Both images show a glioblastoma with peritumoral edema, and both uptakes have TE: 29 and TR: 600. The figure illustrates how the same in-patient-slice show tissue contrast changes with the use of gadolinium, keeping all other image parameters constant between uptakes. Illustration: Marianne Hannisdal (in-study patient)

T2-weighted imaging

T2 relaxation, also called T2 decay, is a parameter that denotes the tissue-specific loss of transverse magnetization as a result of energy-exchange with neighboring protons; spin-spin-relaxation. After the RF-pulse has excited the protons´ net magnetization to the transversal plane, the transversal magnetization, derived by the spins coherent precession, will decay in an exponential-like curve (46). The time it takes for the respective tissues to lose 63 % of their in-phase coherence by spin-spin-relaxation is called T2-time, which is a tissue-specific time constant, as illustrated in figure 6.

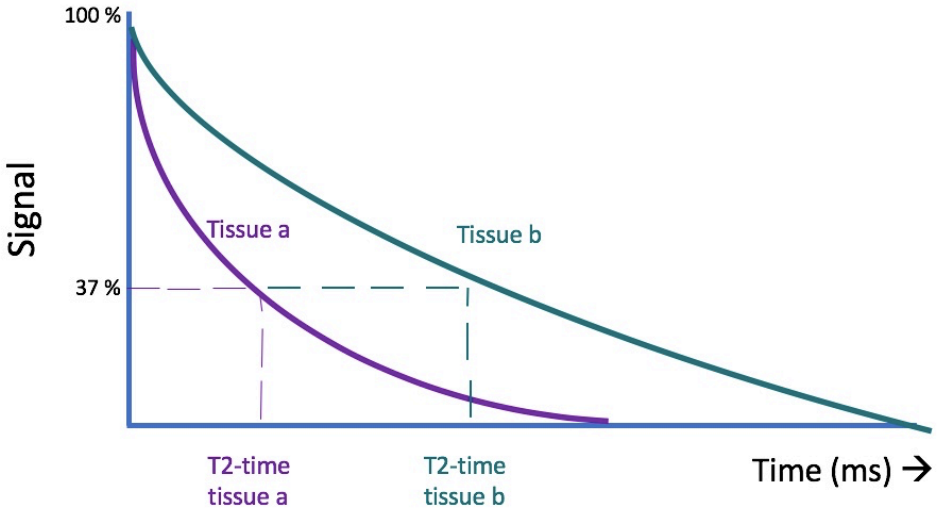


Figure 6 T2-relaxation. The figure shows how tissue a and tissue b spend different amount of time to lose 63 % of the transversal magnetization. Illustration: Marianne H. Hannisdal

The underlying source of T2-time heterogeneity across tissues is, like for T1, due to intrinsic characteristics in tissue microstructure. As previously explained, fat exhibits a molecular tumble rate of protons that exchanges energy more rapidly than water. However, the magnitude of transverse magnetization in water is much greater than in fat, leaving the signal from water with an amplitude superior to fat. Consequently, water will appear hyper-intense on the grey-scale while fat will appear dark or hyperdense in the T2-weighted image (41). A T2-weighted image is exemplified in figure 7.

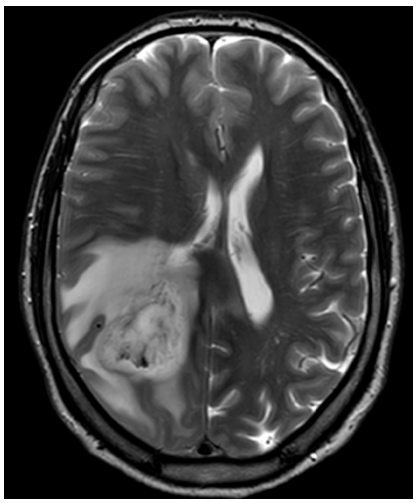


Figure 7 Axial view of a T2-weighted image showing a glioblastoma with peritumoral edema, TE: 90 TR: 5794. Illustration: Marianne Hannisdal (in-study patient)

T2* is also a type of T2-weightening which, in addition to T2 spin-spin-relaxation, denotes tissue- and microenvironmental-specific dephasing differences due to field inhomogeneities (41). The field inhomogeneities is caused by magnet field imperfections, chemical shift, macroscopic susceptibilities, and presence of para- or ferromagnetic substances (46). However, the T2*-decay happens much faster than T2, so one has to use a short echo-time (TE) to capture the effect, as illustrated in figure 8 below. T2* effect is therefore not present in the image-series used in this study.

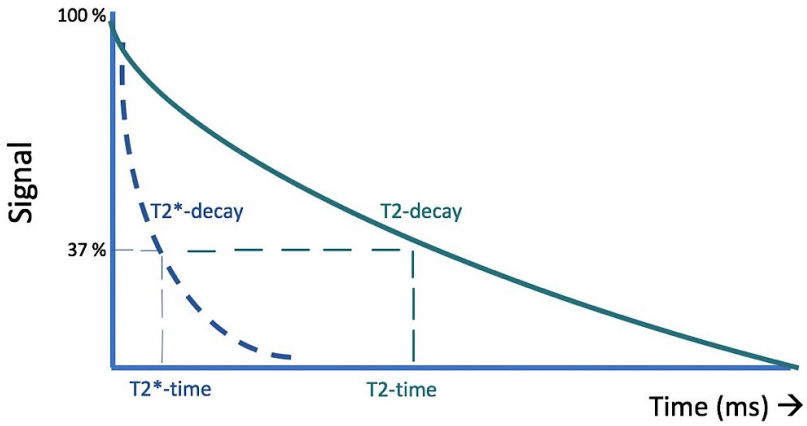


Figure 8 T2-decay (stapled blue line) and T2-decay (green line) for a tissue. The figure illustrates how T2* happens faster than T2-decay as the magnetic inhomogeneities contributes to dephasing as an addition to spin-spin-relaxation. Illustration: Marianne H. Hannisdal*

Fluid-Attenuated Inversion Recovery

Fluid-attenuated inversion recovery (FLAIR) is a T2-weighted spin-echo pulse with an RF inverting pulse that provides heavily T2-weighetd images with suppressed signal from fluid (47). The inversion pulse is timed to null out the magnetization from cerebrospinal fluid (CSF). Nulling the signal from CSF is beneficial for visualizing pathology adjacent to CSF without any hyperintense signals from CSF contaminating the visual perception. FLAIR provides a hyperintense signal from peritumoral edema, which for treatment-naive patients, can be related to tumor infiltration (48). The solid tumor compartment is often not represented by the same hyperintense FLAIR-signal. FLAIR-series can thereby provide differentiation between edema and tumor (49). A FLAIR image is exemplified in figure 9.

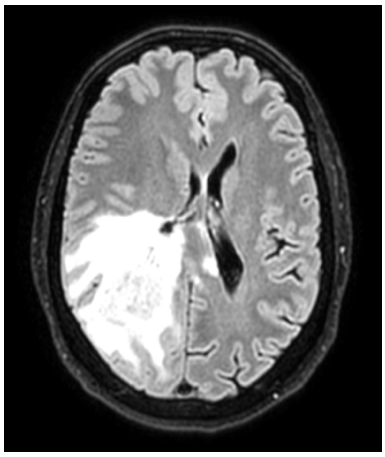


Figure 9 Axial view FLAIR image showing a glioblastoma with peritumoral edema. The image was acquired with TE: 293, TR: 4800, TI: 1660. Illustration: Marianne Hannisdal (in-study patient)

Other MRI sequences used for GBM characterization

There are numerous image sequences available, some of those utilized in evaluation of MRI-biomarker for brain tumors will be briefly mentioned here. Diffusion-weighted imaging (DWI) is a much used technique that measures the random thermal motion of water molecules in x-,y-, and z-direction in extracellular space (50). In white matter, the diffusion is directional as the lipid membranes in myelin form structured boundaries, whereas in grey matter the diffusion is more isotropic. In glioblastoma and generally in malignant tumors, diffusion is restricted due to increased tumor cellularity (51). The strength and duration of the gradients is denoted in B-values, at least 3 B-values are typically used in a DWI-acquisition. As the signal intensity can be influenced by T2-shine-through artefact, the DWI images are usually interpreted taking the apparent diffusion coefficient (ADC)-map into account (52). This is calculated from voxel-wise signal intensities across the range of the applied B-values (50). An example of a diffusion-weighted image and ADC-map is given in figure 10.

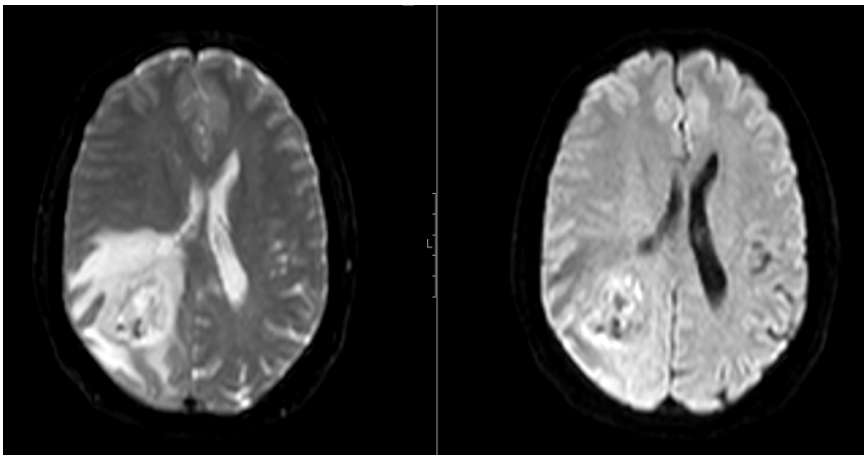


Figure 10 Left: axial diffusion weighted image showing a glioblastoma with peritumoral edema. Right: ADC calculation of the same slice.. Illustration: Marianne Hannisdal (in-study patient)

Diffusion tensor imaging (DTI) is a more advanced diffusion image series that denotes the anisotropic direction of the structural water motility from at least six non-collinear directions (52, 53) and useful for visualizing tensors depicting white matter fiber tracts. The direction of motility is indicated in color-code, whereas one color refers to left-right, another to cranio-caudal direction etc., as exemplified in figure 11 below. White matter tracts are frequently infiltrated or disrupted by glioblastoma (54). In grey matter, the water motility has more of a isotropic character, thus the signal is nulled.

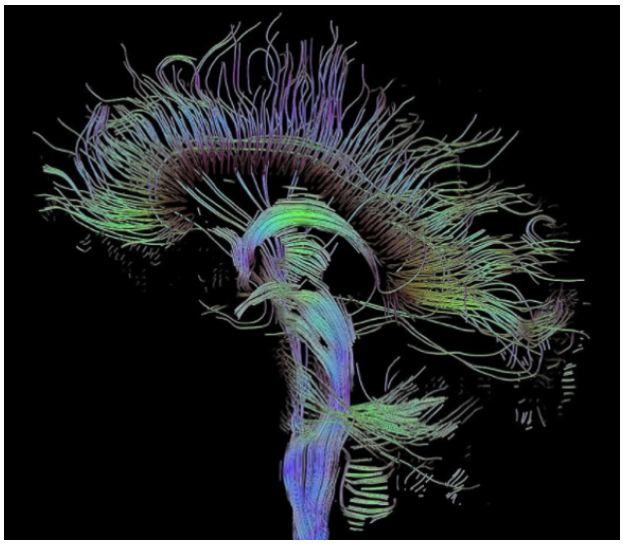


Figure 11 Diffusion tensor imaging, sagittal view, depicting reconstructed fiber tracts that run through the mid-sagittal plane. Illustration: courtesy of Thomas Schultz (55)

MRI spectroscopy (MRS) is another technique, where one does not acquire an image, but a quantitative chart measuring concentration of various metabolites that are present in a tissue. The method is based upon electron distribution within the atom, which cause the nuclei in different molecules to undergo a slightly different magnetic field. The result is somewhat different resonant frequencies, which ultimately results in a slightly different signal (56). Typical metabolites that are differentially present in peritumoral areas of glioblastoma and used for MRS diagnostic purposes are N-acetyl aspartate (NAA) and choline (Cho). Decrease in NAA levels is related to neuronal viability, and increased levels of Cho is related to cellular membrane turnover. These metabolites are therefore used as brain tumor biomarkers in spectroscopy (57). MRS is not part of the routine clinical imaging protocol.

Other parameters affecting image contrast

Magnetic susceptibility not only affects the image contrast in T2* weighted images but may also impact image quality. Macroscopically, large differences in magnetic susceptibility between bordering tissues can cause geometrical distortions, as is often present in sinus-near areas. However, microscopical susceptibility can also affect the image, especially when imaging small structures. The brain is structurally complex, and even with a sub millimeter isotropic voxel dimension, the bulk MR-signal from a voxel will be a combination or a weighted average of the tissue-types within the voxel. Such partial volume effect decrease spatial resolution, causing smudging in homogenous tissues as well as in tissue borders, and chemical shift exacerbates this (58). Chemical shift is as a result of the resonance frequency

shift between fat and water. It originates from molecular diamagnetic susceptibility, and will induce a small error in phase-frequency mapping, resulting in signal read-out error known as chemical shift (41). The effect of chemical shift can be reduced with increased bandwidth, inverting phase-direction and more, however, it can never be eliminated in full as it originates from an intrinsic feature in the tissues' preconditions for resonance.

2.3.4 The pathophysiology of glioblastoma related to MRI

GBM has a heterogenous appearance in MRI, commonly consisting of necrosis, hemorrhage, cysts and calcification (6). The GBM central necrotic core is often surrounded by a peripheral ring of contrast-enhancing tumor representing high angiogenic activity on T1-images. The solid tumor is often surrounded by an area of poorly circumscribed, vasogenic edema and infiltrative neoplastic tissue, represented by hyperintense T2/FLAIR-signal, as demonstrated in figure 12 below. This surrounding infiltrative component is a mixture of tumor and reactive normal cells where contrast enhancement does not occur as the tumor cells have not reached morphologic manifest with functioning angiographic structure (6).

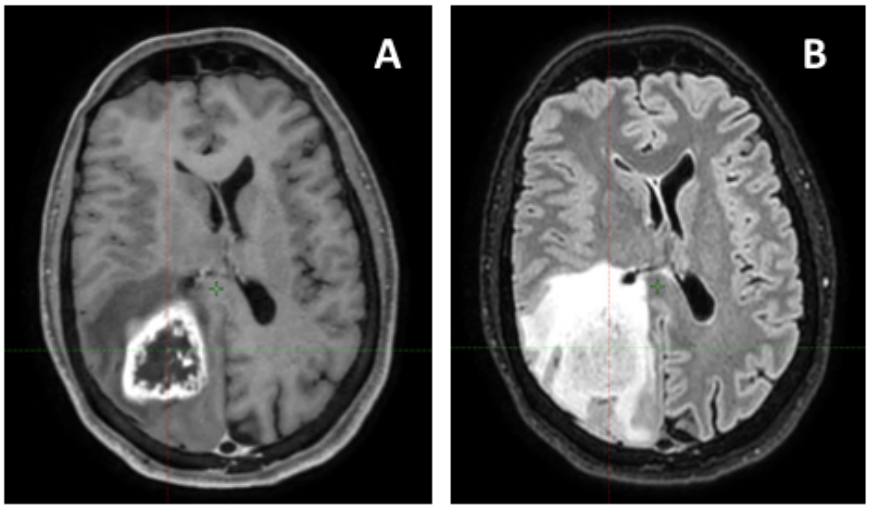


Figure 12 Example of a contrast-enhanced T1 (A) and FLAIR (B) series depicting a glioblastoma. Illustration: Marianne Hannisdal (in-study patient)

Neoplastic angiogenetic tissue can be distinguished from normal tissue by its characteristic pathological architecture of the vasculature. The disrupted blood-brain-barrier is permeable, allowing leaked contrast agent to enhance the signal in tumor cells, which is a well-established radiological feature in treatment-naïve patients. Furthermore, the GBM-cells have a higher nuclear-to-cytoplasm ratio than other, low grade gliomas such as astrocytoma, which

can be distinguished by lower T2 signal as well as mildly restricted diffusion on DWI (6). GBM can be discriminated from other high-grade gliomas by the presence of necrosis according to WHO classification system, in addition to other histological characteristics (28, 59).

Glioblastoma tissue boundary are hard to distinguish from surrounding normal tissue due to its highly infiltrative cellular invasion (60). In fact, this image-diffuse invasion into the surrounding healthy tissue has been stated as the most significant characteristic of glioblastoma (61). Biological chain-reactions makes the differentiation hard, studies have showed that macrophages and microglial cells that form part of the body's immune defence, infiltrate upto 40 % of the non-enhancing glioblastoma tumor mass. Their presence contribute to a poorer differentiation of malignant cells on both T1 and FLAIR images (44). However, both macrophages and microglial cells have the ability to phagocytose tissue or cellular components identified as damaged or threatening (62). Phagocytosis of red blood cells present in the highly angiogenic tumor vasculature leaves hemosiderin residues that heavily affect the tissue's ferromagnetic abilities (44, 63).

Pseudoprogression is defined as "*radiographic evidence of disease progression, typically within 3 to 6 months posttreatment, followed by spontaneous resolution or improvement without additional treatment*" (64). It is termed is termed *pseudo* (false) progression because it presents as a tumor progression on T1 weighted images. Pseudoprogression is mainly caused by radiation necrosis, which is a consequence of damage of normal tissue , followed by inflammation upregulation of vascular endothelial growth factor (VEGF). VEGF is a signal protein that stimulates formation of blood vessels, and was formerly called vascular permeability factor (VPF) (65). Upregulation of VEGF increases in vessel permeability and edema. Studies showed that patients with an unmethylated *MGMT* promoter infrequently present with pseudoprogression, presumably because the phenomenon is related to radiosensitivity, indicating response to treatment. Increased radiation doses correlates with the increased probability of pseudoprogression (64), this effect is compliant to the "normal tissue complication probability", as will be explained in section 2.4.1. Pseudoprogression can also be suspected after tumor surgical resection, in the form of new contrast enhancement associated with subacute ischemia, and is therefore why it is recommended that postoperative imaging is acquired within 48 hours to avoid signal contamination.

Progression assessment guidelines

Several guidelines guidelines are available for assessment of tumor progression by radiology, the most recent being the Response Assessment in Neuro Oncology (RANO) guidelines that came in 2010, and was modified in 2017 (64). The RANO guidelines take treatment and clinical history into account in relation to the radiological measurements. RANO criteria were upgraded in 2017 with an aim to increase focus on distinguishing *pseudo* progression from true tumor progression, in particular after immunotherapy According to RANO, new contrast enhancing area within the radiation field cannot be diagnosed as true progression within the first 12 weeks after radiotherapy, as pseudoprogression stands as an undistinguishable differential diagnosis.

2.4 Brief introduction to the physics and radiobiology in radiotherapy

Radiotherapy (RT) can be administered with various techniques but I will focus on external beam therapy with megavolt (MV) photons in this thesis. RT utilizes ionizing radiation to treat medical conditions, primarily oncological lesions.. The unit for dose is Gray (Gy) and is defined as absorbed mean energy in joule per kilogram (J/kg). 1 Gy means that 1 kg of the irradiated tissue has absorbed 1 joule radiation energy.

The energy from MV photons is deposited in radiated tissue by interactions that cause ionizations and scattered electrons. The scattered electrons and free radicals have the potential to cause damage to cellular DNA, which ultimately induces a range of destructive processes such as apoptosis, mitotic catastrophe, radiation induced senescence, autophagy and necrosis (66). These effects can deprive tumor cells of their reproductive abilities and lead to cell death Radiation can also induce tumor cell necrosis due to enhanced inflammatory responses (67).

2.4.1 Therapeutic effect

The objective of radiation therapy is to obtain tumor control without causing unacceptable toxicity to normal tissue. Both tumor control probability (TCP) and normal tissue complication probability (NTCP) increase with dose and irradiated volume (68). This evaluation of “risk-benefit” of the radiotherapy treatment is denoted as “the therapeutic window” and relates to the ratio between maximum radiation dose by which tumor cells are

controlled and the minimum radiation dose by which cells in normal tissues suffer acceptable complications (68), as illustrated in figure 13.

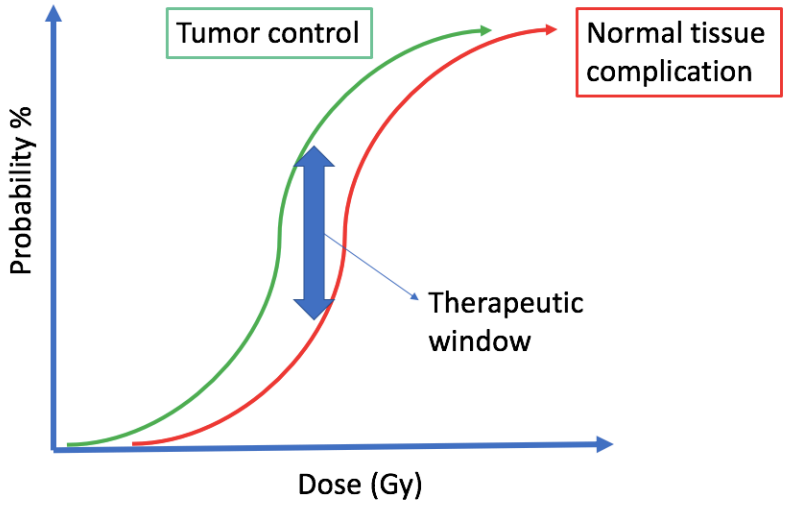


Figure 13 The therapeutic window in radiation therapy; the beneficial ratio between tumor control probability and normal tissue complication probability, as a function of increasing dose. Illustration: Marianne H. Hannisdal

2.4.2 Fractionation

The standard RT course for treatment of GBM is a total dose of 60 Gy administered as 5 fractions of 2 Gy per week, concomitantly with oral 75mg/m2 temozolomide chemotherapy for 6 weeks (8, 69). Dose fractionation defines how the total radiation dose is divided, where normally one (or more) fractions are given each day. The fractionation has four main goals; it allows normal cells to repair their damaged DNA and cellular repopulation in normal tissue between fractions. It also allows tumor cell reoxygenation that is required for generation of free radicals, and redistribution of cell cycle kinetics which renders them susceptible to the radiation effect (35).

GBM is notoriously radioresistant (69), that is, the tumor cells responds poorly to radiation, and hence the therapeutic window is narrowed. Current fractionation-schemes are limited by complications to normal tissue weighted up against probability of tumor growth control. The use of large margins to encompass all the GBM cells inevitably result in complications due to damage to normal tissue compartments. The use of margins will be more closely explained in section 2.5.2.

2.5 Radiotherapy of glioblastoma

2.5.1 The treatment planning process

When an oncologist referrers a patient to radiotherapy a defined cascade of processes are initiated, whereof the main steps are illustrated in the flow-chart in figure 14.

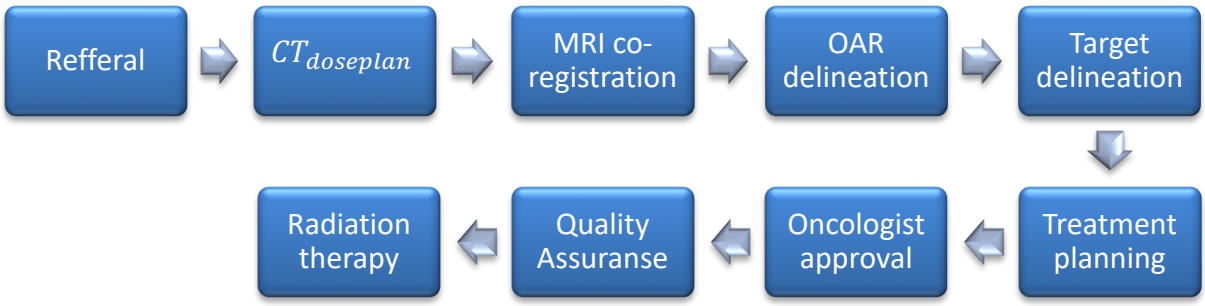


Figure 14 Flow-chart showing the main steps of a radiation treatment process from referral to therapeutic execution

$CT_{doseplan}$: simulation and fixation

The technology of $CT_{doseplan}$ is the same as for a diagnostic CT-scanner, only with a flat table top for positioning-reproducibility purposes, and an extra-large gantry, so-called "big-bore", to make room for space demanding patient fixation devices. The images taken on $CT_{doseplan}$ simulates the treatment position and hereby sets the standard for patient positioning for the entire remaining treatment time. Glioblastoma patients are fixated by a heated thermal head mask that is molded directly onto the patients` skin, as well as being pinned to the table top, as illustrated in figure 15 below. The mask leaves very little room for patient movement, and ensures a near-perfect patient-positioning throughout the 30 treatment fractions carried out over 6 weeks.

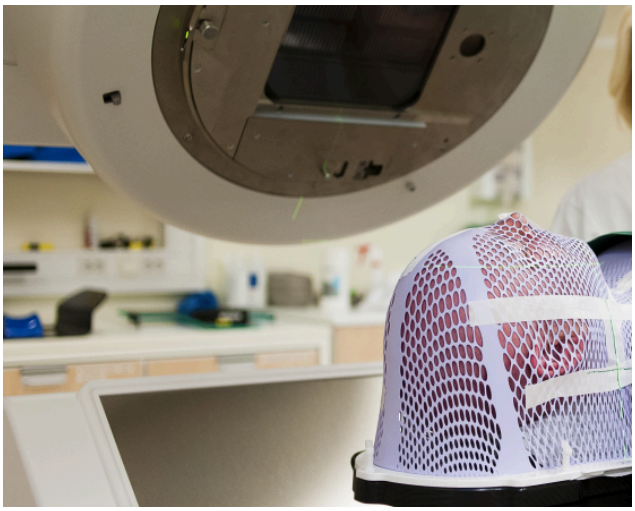


Figure 15 Patient fixated with thermal mask, situated on the treatment table top. Illustration: curtesy of J. F. Frantzen, UNN

Coordinate systems

When handling images across modalities, across MRI-sequences, and across world space and the image space, several coordinate systems come into use. For radiotherapy and in-scanner, one needs to relate the in-patient treatment volume to a physical environment and coordinate system, the transversal, the vertical and the longditunal plane are referred to as the x,y,z dimensions, respectively, where coordinates are points in R^3.

Image space coordinates which does not refer to the physical space, the transversal, vertical and longditunal plane are referred to as k,i,j, respectively, where coordinates are integer-valued voxel locations. The use or denotation of these coordinates thereby distinguishes the situation of a voxel in image space and a point in physical space, as often is the case in image analysis.

Co-registration

Co-registration, also called image registration or spatial alignment refers to the fusion of two sets of images into the same coordinate system, meaning they are visualized overlaid (aligned) in the same space. The registration thereby hold a second layer of voxel-position metadata. The registration can be done manually by rotating and moving one image to visually match the other, or automatically, using an algorithm that adjusts and rotates the image according to e.g. bone-structures. Manual adjustments can also be made on the automatic registration. The rigid co-registration can be done with up to 6 degrees of freedom, including translations and rotations.

There are two types of co-registration methods; affine and deformable registration. Rigid registration, the simplest in the affine category, is the most straight-forward method as it simply translates and rotates all voxels in the same way, keeping the relation between image voxels constant, as illustrated in figure 16.

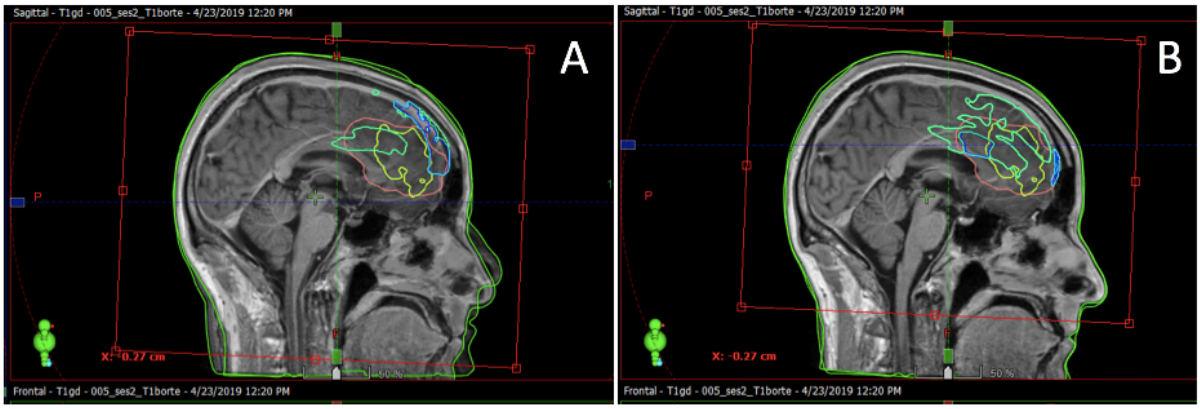


Figure 16 Co-registration of two MRI-series taken at different time-points. A) shows the images superimposed with their original positioning, the green outline is the body-contour of the patient. B) shows the same images superimposed after rigid co-registration. Illustration: Marianne Hannisdal (in-study patient)

Deformable image registration (DIR) attempts to correct for changes in the patient between scans, such as body weight changes, patient positioning, or displacements in soft tissue because of patient motion, or organ changes. This is done by making an elastic grid-map between landmarks in one image series and the corresponding landmarks in the second image series (70). For patients that have undergone surgery, where brain volume elements have been removed from one series to another, or when severe brain atrophy has occurred between series as a result of either radiotherapy or evolving pathology, deformable registration is not recommended in current clinical practice.

For glioblastoma patients, it is not standard of care to acquire dedicated MR images for treatment planning purposes, so-called $MRI_{doseplan}$, instead diagnostic MR-images are co-registered rigidly onto the $CT_{doseplan}$-images. This is mainly due to limited MRI resources, but also partly because (i) nearly all GBM-patients have very recent pre- and post-surgery MRI prior to radiotherapy, and (ii) because the cranium fixates the brain so well that the conditions for co-registration is very good. An example of an image registration between a $CT_{doseplan}$ and a $MR_{post-surgery}$ is given in figure 17.

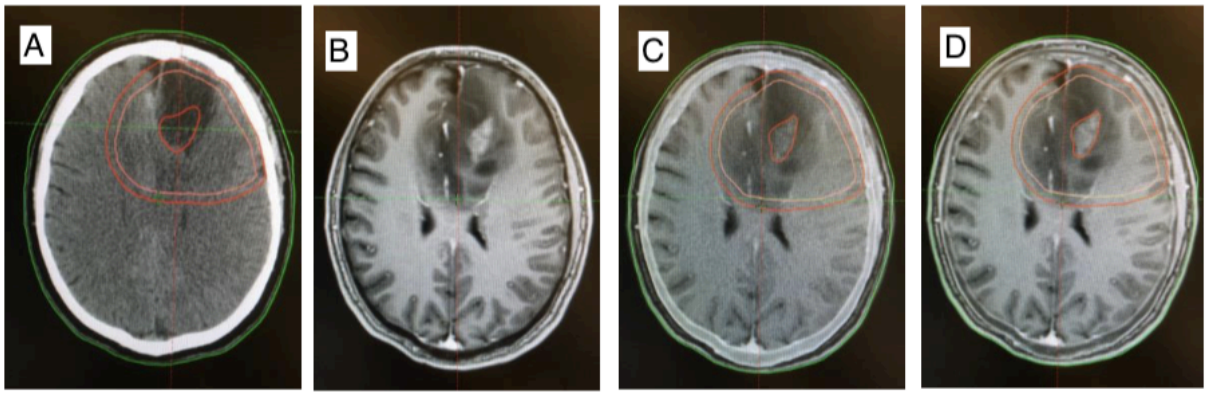


Figure 17 Co-registered CT- and MR-image. A): $CT_{doseplan}$ B): $MR_{post-surgey}$ C): A and B co-registered weighted 50%-50%, D): A and B co-registered weighted 25% and 75%, respectively. The red circles indicate the planning volumes delineated by the oncologist. Illustration: Marianne Hannisdal (in-study patient)

Organs at risk

To ensure that critical normal tissue does not get radiation doses above tolerance levels we define *Organ At Risk* (OAR) volumes. OAR are delineated by medical dosimetrists before the oncologists perform target delineation, and OAR-structures are viewed and approved by the oncologists. Figure 18 shows manual digital pencil delineation on an electronic drawing board.

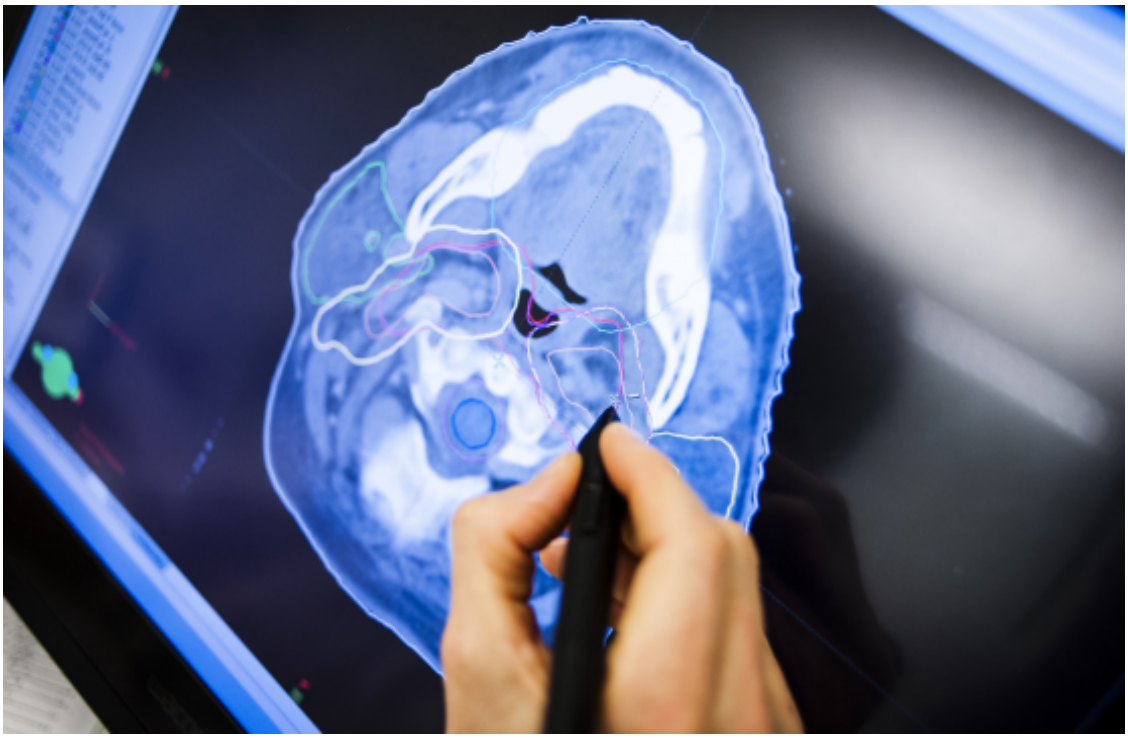


Figure 18 Manually delineation of organ-at-risk-structures by medical dosimetrists, often done on electronic drawing-boards, as shown here. Illustration: curtesy of J. F. Frantzen, UNN.

Target delineation

Target delineation is made on $CT_{doseplan}$, but soft-tissue visualization is supported by the co-registered MR-images. The oncologist will often make the delineations by visual interpretation of the contrast-enhanced T1-weighted images, for proper distinguishing of any remaining tumor fragments, and the T2FLAIR to distinguish the extent of the peritumoral edema. The oncologists delineate manually based on the available image material, and treatment margins are calculated from this, as will be explained in the next section.

2.5.2 Volumes in radiotherapy

Target volumes for radiotherapy is delineated on according to definitions from International Commission on Radiation Units and Measurements (ICRU) report 83 (71). There are variations in target volumes sizes according to various cancer types and their anatomical locations. In the following section I will present those relevant in treating glioblastoma. The relevant volumes and their geometrical relations are visualized in figure 19.

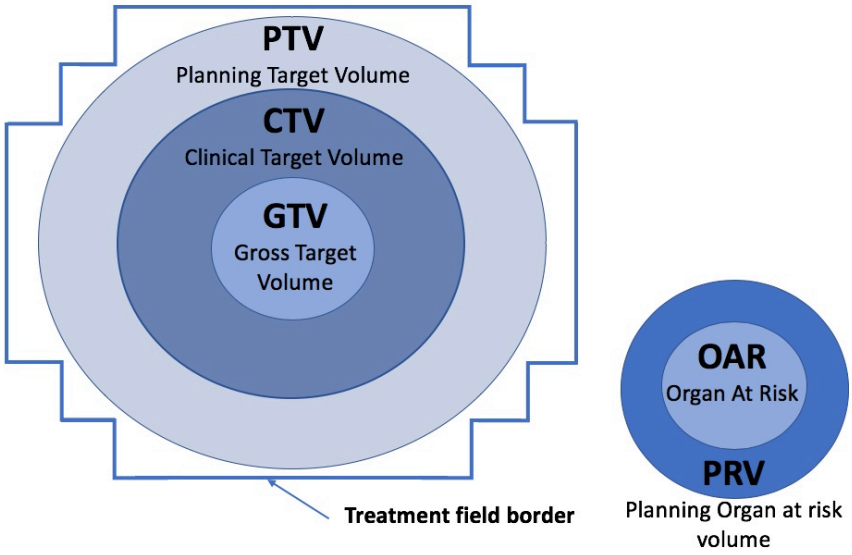


Figure 19 Target volumes used in radiotherapy and their relation to each other within the treatment field border. Gross Tumor Volume (GTV) in the center, Clinical Target Volume (CTV) surrounding GTV, and Planning Target Volume (PTV) surrounding CTV. Organ at risk (OAR) and Planning Organ at Risk Volume (PRV) on the side of target volumes. Illustration: Marianne Hannisdal

Gross Target Volume

The macroscopic tumor is defined as the *Gross Target Volume* and can be the primary tumor (GTVp) or recurrent tumor (GTVr). The GTV delineated by the oncologist for the radiation

treatment plan is the product of the oncologists qualitative interpretation of the malignant tissue`s extent, a subjective decision that is made from the sum of the available radiological information.

Clinical Target Volume

The area around the GTV with high probability of microscopic tumor infiltration, so-called pre-morphological malignant disease, is defined as the *Clinical Target Volume* (CTV)(38). The CTV is very often delineated as a volumetric margin where size is based on studies on what anatomical pathways the various cancer types tends to take, and how aggressive or infiltrative the cancer type tends to be. Glioblastoma is highly-infiltrative and existing contouring guidelines from RTOG, EORTC and ESTRO-ACROP advise the use of 20 mm margin around the entire GTV (12). Additionally, clinical considerations by the oncologist in respect of localization, natural barriers, radiological expression etc. subjectively defines the final CTV.

A note on CTV-delineation: As of September 2020, the Norwegian Directorate of Health has sent new guidelines for consideration to the Norwegian medical community and regional Health Trusts, that states:

> *Large target volumes should be avoided as radiotherapy up to 60 Gy targeting large volumes induce increased risk of radiation induces neurotoxic damage. Recurrence pattern studies of glioblastoma show safe GTV to CTV margin reduction to 10 mm without increase in edge-recurrences or out-target recurrence. This will reduce the radiation exposure of the surrounding normal tissue and thereby reduce the risk of late effects.* (19)p.27 [own translation]).

These guidelines are under consideration and are not published yet.

Planning Target Volume

To take treatment delivery-related uncertainties into account we add another margin around all other volumes called *Planning Target Volume* (PTV). This is purely a geometric margin to ensure that no daily variations in set-up, patient positioning, mechanical delivery techniques etc. compromises the planned dose-delivery to the CTV. For Glioblastoma patients positioned with fitted mask the PTV-margin is typically 3-5 mm.

Organs At Risk

If the target volume is located near an OAR, the oncologist decides whether the target volumes should be reduced in that area to avoid dose to OAR. The Planned organ at Risk Volume (PRV) is a geometrical uncertainty margin around OAR accounting for daily variations in positioning and dose-delivery, as well as in-patient changes such as edema/swelling. The figure 20 illustrates how some of these volumes are presented for radiotherapy planning of glioblastoma in CT-images at Haukeland University Hospital (HUH).

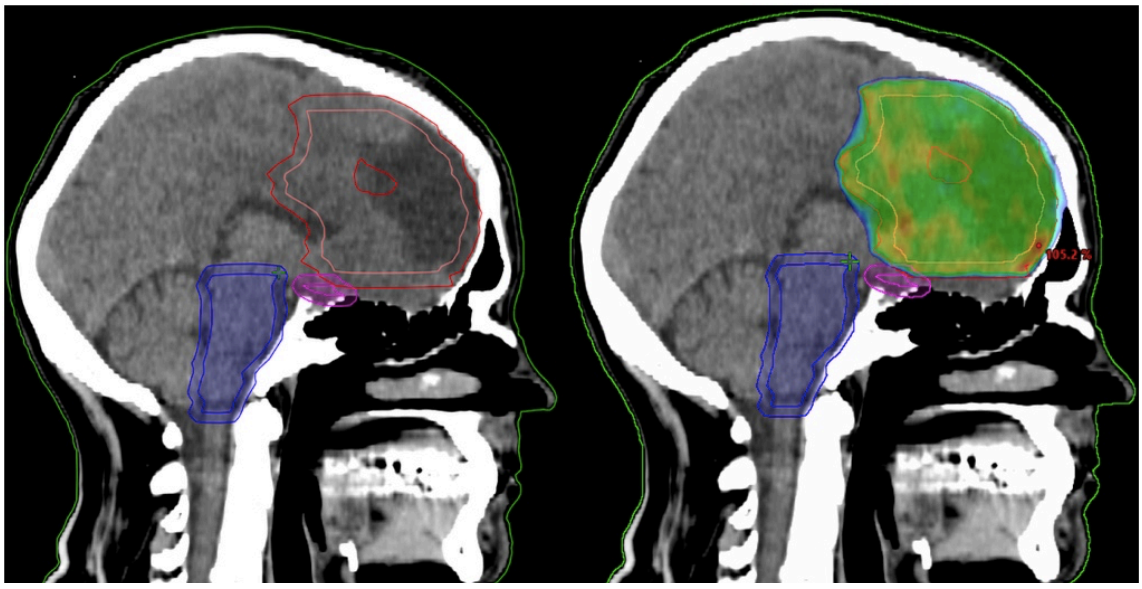


Figure 20 Example showing volumes in radiotherapy on a sagittal CT-image, with dose (right) and without dose (left). The Gross Target Volume (GTV) is the smallest inner circle in bright red in the frontal lobe. The Clinical Target Volume (CTV) shown in a pale pink line is the much larger volume, surrounding GTV with a margin of 20 mm. The Planning Target Volume (PTV) is the outer bright red line, surrounding CTV with a margin of 5 mm. The brainstem and optic chiasma are delineated as Organs At Risk (OAR) in blue and pink, respectively. There is also a 3mm margin surrounding the OAR, called the Planned organ at Risk Volume (PRV) with corresponding color. Illustration: Marianne H Hannisdal (in-study patient).

2.6 Imaging in a quantitative perspective

The quantitative perspective in imaging refers to the evaluation and analysis of images in an objective and quantitative manner, as opposed to qualitative evaluation that comprises visual interpretation.

2.6.1 Image quality

In MRI there are objective image quality measures. Signal-to-noise ratio (SNR) is used to evaluate the general amount of signal in relation to noise in an image, and contrast-to-noise

ratio (CNR) is used to evaluate the difference in signal between two tissue-types in relation to noise. Low spatial frequency, *i.e.* loss of contrast between two adjacent tissues in an image is caused by insufficient resolution so that more than one tissue type occupies the same voxel, causing partial volume effect (58).

Sensitivity in radiological imaging is defined as the method's ability to detect abnormal tissue or disease processes in a sick patient (true positive), and specificity is defined as the method's ability to *not* detect abnormal tissue or disease processes in a patient who is *not* sick (true negative) (72). Together, these terms define the method's ability to distinguish true disease from no disease. A limitation to sensitivity and specificity is that they are binary measures; true/false or sick/normal, which is less applicable when comparing continuous variables or states, such as the pre-morphological to morphological evolution of a cancer in tissue.

2.6.2 Multispectral MRI and radiological biomarkers

Radiological biomarkers are also called imaging biomarkers, they can be qualitative as well as quantitative, but in this thesis this I will focus only on quantitative biomarkers.

Kessler et al (2015) defines a quantitative imaging biomarker as: *«an objective characteristic derived from an in vivo image measured on a ratio or interval scale as an indicator of normal biological processes, pathogenic processes or a response to a therapeutic intervention"* (73).

The term quantitative imaging is defined by Sullivan et al., (2015) as "*the extraction of quantifiable features from medical images for the assessment of normal [findings] or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal [findings]*" (74).

In the present study, we use radiological biomarkers in terms of voxel signal intensities from the multispectral recordings as vector-valued input (features) to our deep learning segmentation procedure. In general, radiological biomarkers aim to reveal correlations between radiological features and histological findings (75). Established radiological hallmarks state that neoplastic cells differ distinctly from normal tissue, just as glioblastoma is radiologically distinct from lower grade gliomas (6). A range of quantitative characteristics can be extracted from the MR imaging data, for example, contrast-enhanced T1 provide information about signal intensity variation in relation to intra-tumor heterogeneity, while diffusion-weighted images can give information related to tumor cellularity (76). It is precisely because visual interpretation of qualitative radiological changes can be challenging,

that quantitative radiological biomarkers and statistical interpretation methods can add diagnostic value. As the technical improvements allow the radiological characteristics to be analyzed and systemized in a quantitative matter, the radiological discipline is also shifting towards a more computational paradigm. To be able to make use of quantitative information in the image datasets, we need proper tools to help us analyze the information in a systematic way. This brings us to statistical interpretation methods within artificial intelligence (AI) and machine learning (ML).

2.6.3 Artificial intelligence in analysis of radiological images

Intelligence can be defined as *"a general mental ability for reasoning, problem solving, and learning"* (77). Textbooks define artificial intelligence as "*the study and design of intelligent agents, where an intelligent agent is a system that perceives its environment and takes actions which maximizes its chances of success*" (78).

Medical imaging can benefit from using artificial intelligence, as has been stated by the ISRRT and EFRS:

> *Adoption of AI in medical imaging and radiation therapy requires radiographers and radiological technologists to adapt their imaging and treatment practices to ensure new technology is being implemented, used and regulated appropriately, based on high quality research evidence, maximizing benefits to their patient (18).*

Machine learning (ML) is a method in which computers are able to make generalized predictions on new and unseen data based on learning from experience (79). A myriad of terminologies are used in artificial intelligence and numerous machine learning techniques are available. A comprehensive overview of artificial intelligence is beyond the scope of this study, but I will present some key elements to computational image analysis, along with the most common algorithm types. In figure 21 major terms are presented in their relation to another, in which the essence is that machine learning is a type of artificial intelligence, and the various algorithms are different types of machine learning. Take note that machine learning is not limited to the listed algorithms, and artificial intelligence is not limited to machine learning.

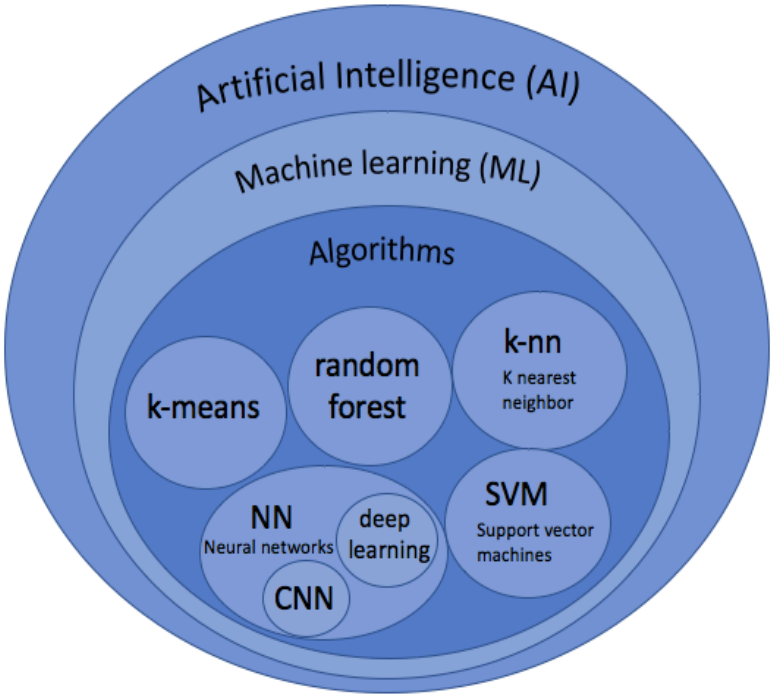


Figure 21 The relations between Artificial Intelligence, Machine Learning and algorithms. Illustration: Marianne Hannisdal

There is a collective hope that machine learning can learn to detect and distinguish malignant tissue from normal tissue better than the human eye. This is simply because the intelligent human brain has the capacity to evaluate only a limited range of grey shades - and a limited number of MR-sequences - simultaneously, whereas artificial intelligence does not have the same limitations. A machine-learning algorithm has the ability to evaluate the same voxel represented by numerous MR-image-weightings i.e. vector-valued voxels. Based on this, a "signature" prediction of what type of tissue the voxel consists of is made and subsequently used for segmentation purposes. The basic principles of segmentation of cerebral MR-images is to divide the image into meaningful, homogeneous and non-overlapping regions of similar features like signal intensity, depth, color or texture (80). Segments are also referred to as classes, whereas to classify means to assign every voxel a tissue type, where the tissue types is pre-defined in specific classes (80). Classification results from algorithms can be visualized by using *feature space,* a three- or higher dimensional space where the MRI-sequence-specific features from the respective classes are presented as data points, as illustrated in figure 22.

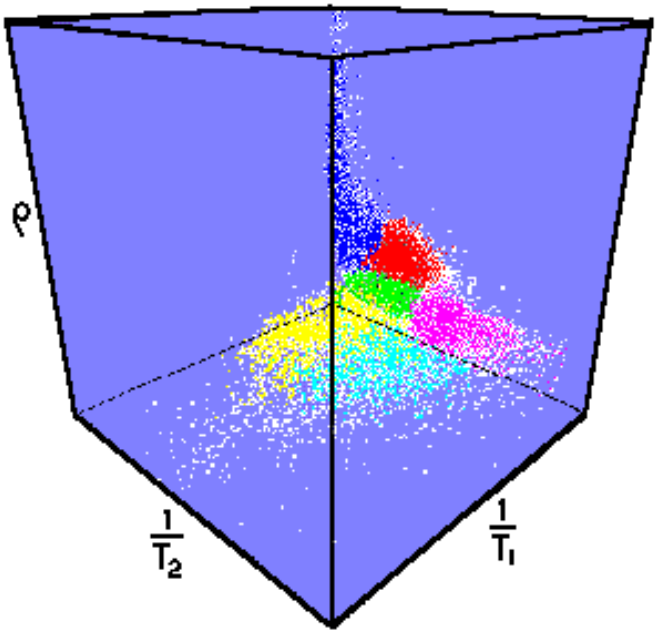


Figure 22 Feature space, illustrating how signal intensities from three different weightings are presented as data points, grouping the respective tissue types using color-coding. Illustration: courtesy of Arvid Lundervold

In feature space, the data points can be labelled, color-coded, and recognized as tissue classes. When this information is associated to voxel locations and mapped back into anatomical image space, the result is an interpretable visualization of the various classes, and their spatial extent are seen as an image *segmentation*. Figure 23 shows segmentation of various tissue types and structures in the brain.

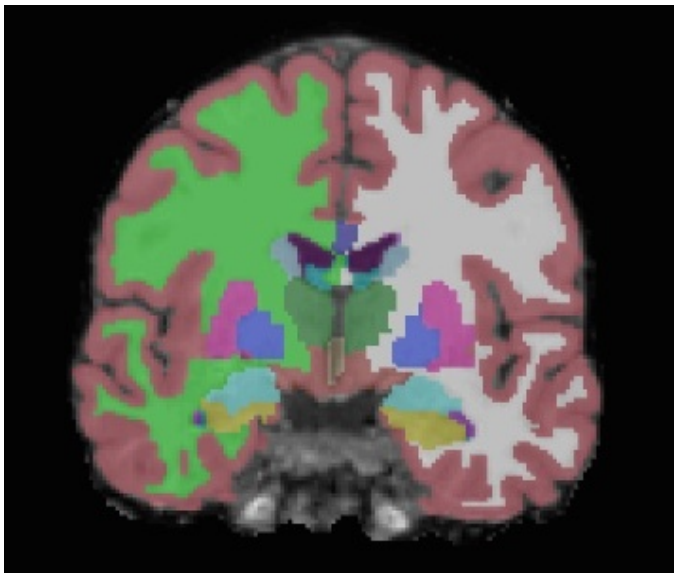


Figure 23 Coronal view of skull-stripped and segmented T1-weighted image of in-study patient, the various colors represent different parts of the brain, segmented with the Freesurfer software. Outer pink

segment is the grey matter in the cerebral cortex, bright green is white matter in right hemisphere, white segment is white matter in left hemisphere. We also see bilateral segmentation of the putamen (bright pink), the pallidum (dark blue), the thalamus (dark green), the amygdala (light blue) and the hippocampus (yellow). Illustration: Marianne Hannisdal

In figure 24 below, is a schematic representation of how machine learning segments a tumor into four classes, based on contrast enhanced T1, T2, and FLAIR- weighted sequences.

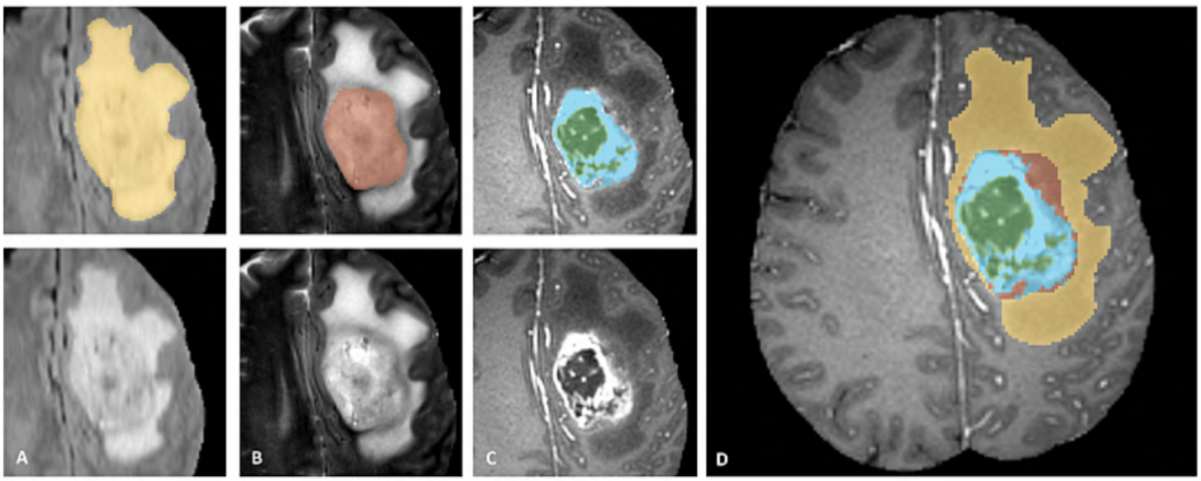


Figure 24 Glioblastoma segmented based on radiological biomarkers. A: T2FLAIR defines the extent of edema (yellow), B: T2 adds information about the extent of tumor (red), and C: contrast enhanced T1 defines the contrast-enhancing part of the tumor (blue), leaving the necrotic core green. D: all the segments presented together. Illustration: adopted from Menze et al., 2015 (81)

The image presented by shades of gray in a grid of pixels for human visualization and perception purposes, is in fact a matrix of numbers referring to the MRI-derived signal intensities, as illustrated in figure 25 below. An MRI image file consists of a data matrix, computationally read into a memory block (e.g. a Numpy array) as voxel-wise signal intensities. In addition the file contains accompanying metadata related to the acquisition, as illustrated in the figure below. The metadata keep track of each voxels in-patient position by image coordinates (i-j-k) in relation to in-scanner world coordinates (x-y-z), amongst many other things.

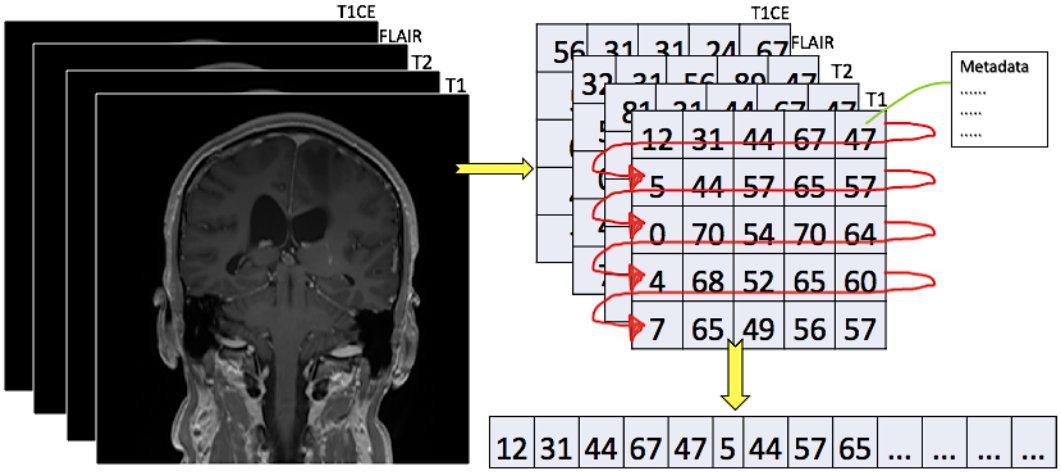


Figure 25 Computational processing of multispectral MRI including a T1-weighted sequence with contrast enhancement (T1CE), Fluid Attenuated Inversion Recovery (FLAIR), T2 and T1. The voxels have numerical values, and are computationally represented as numerical arrays. Each image contains metadata which holds information about the spatial coordinate transform between image space and scanner space in addition to sequence-specific information. Illustration: Marianne Hannisdal (MRI image from in-study patient)

The term radiological biomarker can in the simplest case refer to a voxel's appearance (i.e. signal intensity) on a single sequence, indicating either normal biological tissue, a pathological process or a response to a therapeutic intervention. However, the more expressive multispectral biomarkers refer to the voxels signature profile across the spectrum of available sequences, as illustrated in figure 26 below.

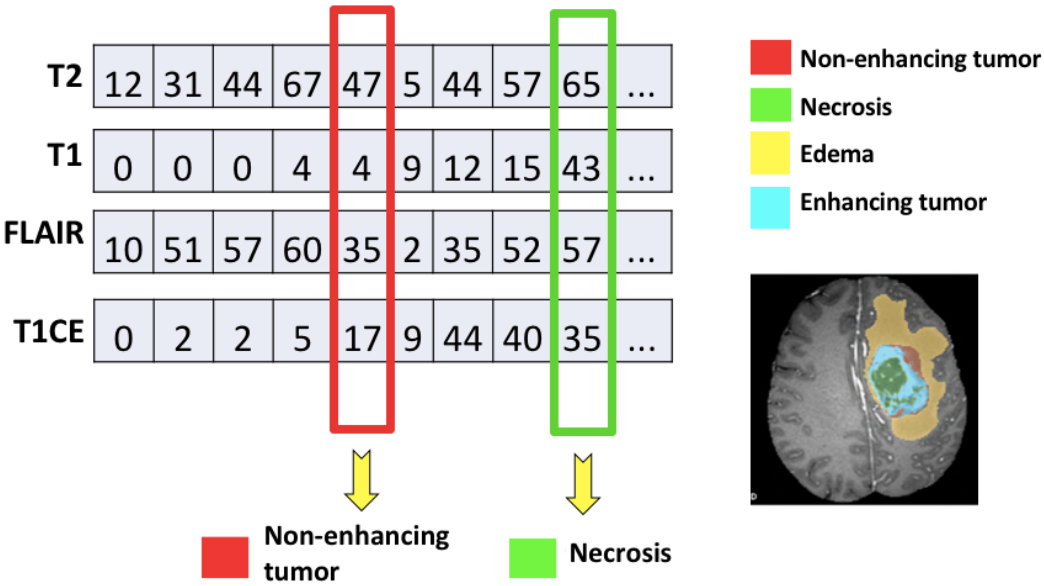


Figure 26 Multispectral MRI including T1 with contrast enhancement (T1CE), Fluid Attenuated Inversion Recovery (FLAIR), T2 and T1. Multispectral profiling for classification of tumor compartments based on computational reading of image data. Each sequence is computationally read as a component of a voxel-based vector. The multispectral combination of these values represents a characteristic numerical profile, labelled as e.g. a specific tissue-class. Illustration: Marianne Hannisdal, MRI-image with segments adapted from Menze et al., (2015).

When using appropriate and relevant sequences, one adds information by contributing with sequence-specific markers for the respective tissue classes. One thereby adds discriminative information, hence increase the specificity of the class-specific profile. This enables the potential to extract information of detailed pathological processes from multispectral MRI-images. Moreover, when including MRI-sequences sensitive to metabolic processes, the potential lies to classify voxels with pathological expressions before a solid morphological manifestation has taken place, utilizing predictions of tumor infiltration areas.

Machine learning algorithms

It is of great importance to use the most appropriate machine learning algorithm, or set of these, suitable to the task and the nature of the available data material. There are two main categories of machine learning methods: supervised and unsupervised.

The term *supervised learning* in our context, means that the user pre-defines tissue-classes that hold certain characteristics. For multispectral images, this is done by using a training mask of manually labeled (painted) voxels in an appropriate MRI-sequence, allocating the voxel's features across all sequences to one of the pre-defined classes. This results in supervised learning or training of the classifier. In the inference part of supervised learning, the trained algorithm can be applied to previously unseen voxels, measured with the same MRI sequence combination, to predict their class association. Neural Network is a much used supervised algorithm that takes use of large amounts of training data, as will be further explained in the next section. Other examples of supervised learning algorithms are random forest (RF) support vector machines (SVM), or K nearest neighbor (KNN). RF consists of an ensemble of decision trees, where every voxel is classified through the algorithm architecture, sorted by one sole feature at the time. This one-by-one feature is simple, iterative and effective, but leaves the algorithm with low repeatability (82). SVM has another supervised approach; it splits the image into classes by finding a 2D or 3D hyperplane with the greatest distance from the nearest data point on each side, which best classifies the data. This is done with a kernel function, taking the data points into a high-dimensional space for separation (83). KNN classification is a very simple supervised classification algorithm, that classifies unknown data points by evaluating its features up against neighboring data points. The term *neighboring* refers in this case to resembling features in image space. KNN can however, be

ineffective to large, high-dimensional data like multispectral MRI, and sensitive to noise and outliers (84).

Contrary to supervised learning, unsupervised learning is not based on pre-defined features and is agnostic to the risk of manual mislabeling which can compromise a training process. By its nature, unsupervised learning has the potential to uncover hidden or subtle structures in sets of images, or find sub-populations in cohorts (84). By making the use of image features like intensity or texture, the unsupervised algorithm will typically group pixels or voxels that contain homogeneous attributes together in classes. In K-means, one of the most common unsupervised classification algorithms the number of clusters (=K), a so-called hyperparameter is decided by the user. There are, however, heuristic methods that are able to propose this K to the user for a given dataset. Unsupervised learning can be useful as an explorative method and be able to "dissect" the (high-dimensional) data in complicated cases (83).

Deep learning

Artificial Neural Networks are often referred to as deep learning, or deep neural networks, and are the current state-of-the-art machine learning models used in image analysis (79). Neural networks (NN) is a type of algorithm in which the structure of the processing path is organized by imitating the neurons communicative structure in the human brain. Nielsen et al., (2015), describe NN as "*a beautiful biologically-inspired programming-paradigm, where the machine learns from observational data*". The algorithm is mainly used for supervised learning and it learns to recognize key features by training on labeled data. Its learning feature of holding experience-based and supervised-confirmed knowledge as a reference when evaluating new and unseen data is often referred to as a resemblance of a living being (85). A key strength of deep learning is that the voxels' features are evaluated by the voxels' own circumstances. In multispectral MRI-context this means that a voxel's appearance in one sequence is, as a feature, evaluated by the coinciding voxels at the same in-patient coordinates from other sequences. The architectural structure of the algorithm is illustrated in figure 27 below.

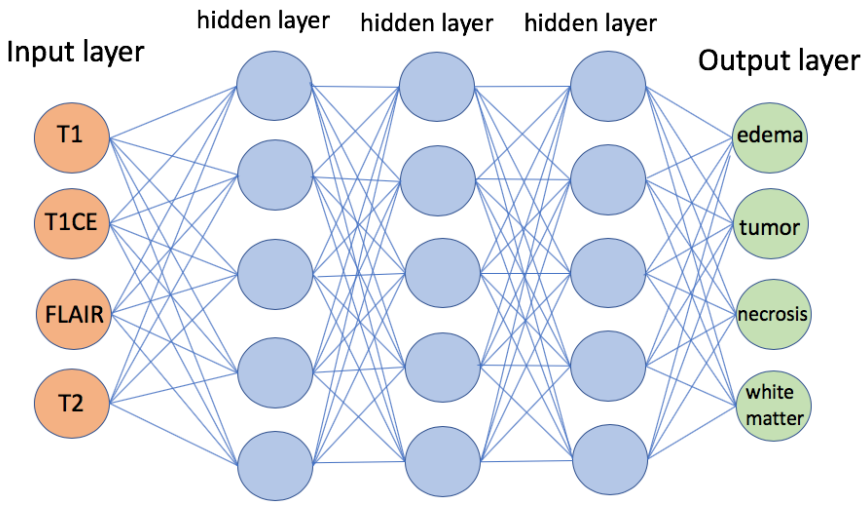


Figure 27 Deep Learning Neural network. In this model the input is the MRI sequences, output is the tissue classes. The nodes in between input and output are the hidden layers, representing the feature extracting. The connections between the nodes describes how the model incorporates the surrounding features, and relations between them, into account - much like a human tries to do when interpreting a multispectral MRI. Illustration: Marianne Hannisdal

One of the main advantages of deep learning models is that it holds feature learning, meaning the model itself determines which features are best feasible for characterization, e.g. if a circle is best characterized by counting edges or by other features (86). In multispectral MRI-contexts, this means that by supervising the algorithm, the user sets the constraints the weights in the neural network by labeling confirmed tumor-parts, and then the algorithm calculates the weights and make hierarchical feature representations of the input-output relationships suitable for future predictions. The name "deep learning" (DL) in neural network-algorithms refers to the algorithm being trained hierarchically from features in a dataset, structured step-wise by many – hence "deep" – neural layers. Convolutional neural networks (CNN) is a branch of NN that employs the mathematical operation convolution (87) that is often used in U-Net architecture consisting of a contracting path on one side, and an expansive path on the other side. This will be further illustrated in the next section.

The drawback of all deep learning type algorithms is that they need large and relevant datasets to be trained upon, a resource that is not always available. On the other hand, the main advantage of the algorithm is that it is very robust because it is trained on large datasets, adding validity and feasibility.

In-study algorithm: *HD-GLIO*

The machine-learning algorithm used in this study is a deep learning convolutional neural network U-Net. The network architecture is conceptually made of a pooling contracting path

and an up-convoluting expansive path, which makes the characteristic U-shaped form, as illustrated in figure 28 below. The contracting path is aggregating semantic information at the cost of decreased spatial information, which the expansive path then decodes and reconstruct, retrieving spatial information including the extracted semantic information.

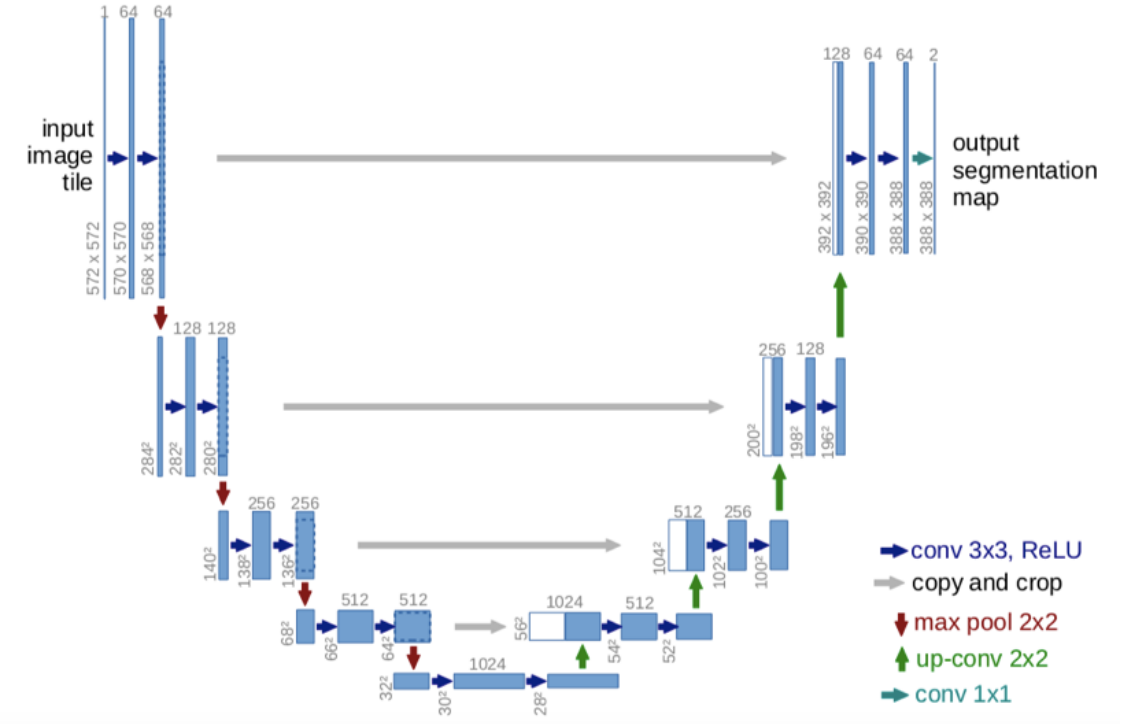


Figure 28 Illustration of a U-net architecture, U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Figure adapted from Ronneberger et al (2015) (88)

The U-net architecture makes the neural networking across the deep layers efficient, providing a processing time of approximately 6 minutes per MRI-examination.

Kickingeder Isensee et al (2019) developed the in-study algorithm as a joint project between the Department of Neuroradiology at the Heidelberg University Hospital, Germany and the Division of Medical Image Computing at the German Cancer Research Center (DKFZ) Heidelberg, Germany, and the developers named this algorithm *HD-GLIO (89, 90)*. When developing HD-GLIO the authors used 3220 MRI examinations from 1450 brain tumor patients, of which 80% was used for training the algorithm and 20% was used for testing and validating the algorithm. All 3220 MRI examinations included pre- and postcontrast T1-weighted, T2-weighted and FLAIR sequences, and all sequences were brain extracted (by their HD-BET) and co-registered. All MRI examinations did also have corresponding targeted

"ground-truth" tumor segmentation masks (i.e. enhancing core and non-enhancing tumor). This data included the following three datasets:

1. Heidelberg training dataset and Heidelberg test dataset; a single-institutional retrospective dataset with 694 MRI examinations from 495 patients acquired at the Department of Neuroradiology, Heidelberg University Hospital, Germany (90)

2. EORTC-26101 test dataset; a multicenter clinical trial dataset with 2034 MRI examinations from 532 patients acquired across 34 institutions in Europe (90)

3. A single-institutional retrospective dataset with 492 MRI examinations from 423 patients (80% glial brain tumors, 20% other histological entities) undergoing routine MRI at different stages of the disease (including 79 early postoperative MRI scans acquired <72h after surgery) at the Department of Neuroradiology, Heidelberg University Hospital, Germany (90)

The output of HD-GLIO contains two segments: enhancing core (EC), and non-enhancing tumor (NE). The labeling procedure for these two segments, which is critical for the defined task as well as the quality of the training and thus the performance of the classifier, was performed by a panel of experienced neuroradiologist.

The training labeling of the EC compartment of HD-GLIO included volumetric delineation of contrast-enhancing tumor, segmentation of the angiogenic part of the tumor. The labeling of EC in HD-GLIO is in compliance with the clinical delineation practice of GTV in radiotherapy.

The training labeling of NE compartment in HD-GLIO included T2FLAIR hyperintense abnormalities, excluding the contrast-enhancing and necrotic portion of the tumor, resection cavity, and obvious leukoaraiosis[2] (90). Note that the training labeling of NE in HD-GLIO differs from the clinical manual low specificity delineation practice of CTV in radiotherapy.

2.6.4 Volumetric evaluation – DICE

The standard tool for reporting the overlap performance of a segmentation in relation to ground truth is the Dice-similarity-coefficient (81), often referred to as Dice-coefficient, or DC. The Dice-coefficient is a calculation related to spatial overlap between two binary

[2] White matter changes, not related to glioblastoma

segments, whereas similarity is a metric that reflects the strength of relationship between two data objects. Dice-coefficient, typically referring to a "ground truth" segmentation versus a "predicted" segmentation, uses the spatial extension of true positive (TP), false negative (FN) and false positive (FP) regions, as a statistical tool that provides a relative measure of degree of overlap between two structures, normalized by the sum of their volumes. TP is used as a sensitivity measure that quantifies the portion of voxels that are identified as positive by both the ground truth and the segmentation being evaluated (91). Dice-coefficient is expressed by a number between 1 and 0, where 1 is a perfect overlap, and 0 is no overlap in this relation (92):

$$Dice\ Coefficient = (2 * TP)/(FN + (2 * TP) + FP)$$

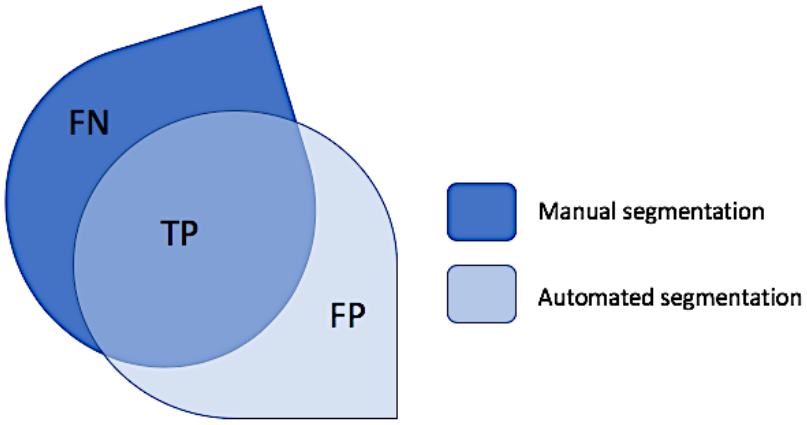


Figure 29: Dice-coefficient. The figure shows the relations between to segments; manual segmentation (dark blue) and automated segmentation (light blue), where the manual segmentation is set to be gold standard. The Dice-coefficient calculates the similarity between them: TP is the true positive volume, detected in both segments, FN (false negative) is the volume falsely labeled as negative by the automated segmentation, in relation to the gold standard. FP is the false positive volume which is falsely labeled positive in relation to the gold standard. Illustration: Marianne H Hannisdal

Geometrically, this can also be stated as DC = 2 |M ∩ A| / (|M| + |A|) where |M ∩ A| is the number of voxels in the intersection of the Manual segmentation ("ground truth") and the Automated segmentation ("predicted"), and |M|, |A| are the number of voxels in the Manually delineated region and the Automatically predicted region, respectively.

Dice-coefficient is a well-known and much used tool when comparing volumes in context of radiotherapy volumes, *e.g.* for inter-observer delineation variation purposes (92). It can be used to measure how well a machine-learning-derived segment matches the manually delineated segment, defining the current clinical practice "gold standard" or "ground truth". There lies a potential methodological weakness in defining one method gold standard, in case of a hypothetical situation where pathology detected with gold standard does not in fact

harmonize with the true morphologic extent. It is still good common practice that makes sure new methodology research always build on earlier stated truth.

Other relevant metrics for evaluating volumetric similarity which was considered for this thesis include the Jaccard index (*JAC*), which defines similarity between two volumes as the intersection between them divided by their union. In addition, the *ROC* curve (Receiver Operating Characteristic) could be eligible, which is the plot of the true positive rate (*TPR*) against the false positive rate (*FPR*). The Hausdorff Distance (*HD*) which essentially calculates the locally maximum distance between two crisp volumes, and the Cohen Kappa Coefficient (*KAP*) which is a measure of agreement between two samples, are other examples of metrics that could be used (91). The Dice-coefficient still remains the most commonly used metric tool for 3D medical imaging analysis (81) and was found best suitable for the purpose in this thesis.

2.7 Previous research

Although previous studies comparing manual tumor delineation versus machine-learning derived tumor segments exist, the best of my knowledge, no studies comparing clinical radiotherapy treatment volumes to machine-learning derived tumor segments exists. There has been performed some studies in context of MRI, biomarkers and radiotherapy, mainly focusing on feasibility of quantitative assessment in multispectral MRI, in the hope of increasing specificity for margin reduction purposes. Baumann et al. (2016) states that biomarkers have unexplored potential to predict extent of subclinical infiltration of tumor cells, and to help define the clinical target volume. Further, they write that by integrating biomarkers in the treatment planning process one could increase the degree of personalized radiation therapy beyond already established practice (93). Andreassen et al. (2018) reports in their study that multispectral imaging enables more precise therapy planning, and suggest that biomarkers be used to identify tumor sub-volumes with high risk of radio-resistance, which in turn can provide rationale for an even higher-precision treatment regime (94). J. Y. Lim and Leech (2016) found that automated definition of tumor and OAR in the head and neck-region defined by artificial intelligence has great potential in terms of saving time and recourses (95).

The use of advanced imaging techniques in relation to target delineation for radiation therapy has shown promising potential. Lopez et al. (2017) demonstrated how MR-spectroscopy has the potential to reveal relevant pre-morphologic information in the terms of tumor metabolism indicators. In their study, they found that spectroscopic mapping of NAA was highly

correlated to the area included in the manually delineated target volume, while the areas containing elevated Choline-values added new radiologic information, which was highly correlated to histologically confirmed tumor extension. Rahmat et al., (2020) investigated another advanced MRI-technique: DTI, in which anisotropic water motility aligned with white matter fibers has been suggested as a fiducial biomarker for the infiltrative component of the tumor, quantified by anisotropic diffusion tensor imaging. In their study, they showed a personalized reduction in CTV-margins up to 40% when delineating target volume according to these findings (53). These image acquisitions, however, are not part of standard imaging procedures. As for cost-efficiency, Rathore et al. (2018) showed in his study that machine learning with the use of already standardized anatomical MRI-sequences is sufficient, both for determining anatomical tumor borders, and for indicating postoperative tumor infiltration in the peritumoral edema (96). Their study suggested that the machine learning method was applicable to predict areas with high risk of recurrence and suggest clinical implementation of non-uniform treatment planning based on their method.

Several studies have investigated the safety and efficacy of non-individualized CTV margin reduction on glioblastoma patients. These studies compared patient outcomes of administrating radiotherapy with 10 mm GTV to CTV margin, according to Adult Brain Tumor Consortium guidelines (ABTC), to patient outcomes administered radiotherapy with conventional 20 mm margin (1, 3, 19, 97-99). The studies found that wider margins had little impact on the recurrence patterns, suggesting clinical implementation of limited-margin radiotherapy. Findings also included that a reduction of CTV-margin is associated with limiting negative immune-effects as well as limiting normal tissue complication, both of which associated with better survival (1-3, 98). As for organs at risk, several studies suggest that as we work to prolong life expectancy for these patients, the clinicians must also focus on post-treatment quality of life as to limiting radiation induces neurotoxity. Another aspect is increased feasibility of re-irradiation after relapse, because of better adherence to OAR constraints (1, 100). This means a thoroughly considered evaluation of doses allowed to critical structures, *e.g.* the hippocampus, which plays an important role in memory and cognitive information processing (100). As dose in OAR is closely related to prescribed dose and distance to the target volume, it is consequently affected by the CTVs extent and margin-size (100). Further, Ali et al. (2014) argues that larger margins pose greater risk of complications to normal tissue and the lower specificity and sensitivity in the image material, the larger margins must be used. The authors suggest that by constantly improving the imaging techniques, one can aim to eliminate delineation insecurities and thereby move to

authorize smaller margins without cost of therapeutic ratio (100). The always-present uncertainty associated with human imaging-interpretation can also expected to be reduced by using machine learning (98).

Some of the aforementioned studies make base for the new guidelines from the Norwegian Directorate of Health which currently is under consideration, emphasizing limitation of radiation induced neurotoxic damage due to large CTV margins (19).

The aforementioned studies does not take MGMT-status into consideration specifically. Brandes et al. (2009) have investigated recurrence pattern in relation to MGMT-status, they found that from the 95 in-study patients, 85% of those with unmethylated MGMT recurred within or at the CTV-margin, compared with 57% of patients with methylated MGMT. For all patients, recurrences inside the overall target volume occurred significantly faster than those outside (101).

The use of quantitative radiological biomarkers and machine learning for automated segmentation purposes have been investigated in some recent radiotherapy clinical studies as well as for the Multimodal Brain Tumor Segmentation (BraTS)-Challenge. The BraTS-challenge was created in 2012 and is an ongoing annual hackathon[3]-contest where researchers compete about making the best machine learning segmentation algorithm. The algorithms are trained on a dataset of preoperative brain tumor MR-images, where results are compared against a histologically confirmed ground truth, quantified by Dice similarity coefficient. This has been extensively used to demonstrate the efficacy of deep learning applications in segmenting glioblastomas (81). Wu et al., (2019) used the BraTS-dataset to develope an automatic segmentation method based on T2 sequence, by using a combination several machine-learning algorithms. They achieved a Dice-coefficient of 84,9% and suggest the use of this segmentation method on both high-grade and low-grade gliomas (82). The use of multiple machine learning models is also reported by Feng et al., (2020) who present an automatic brain tumor segmentation using a deep learning u-net algorithm with a Dice score of 79%, compared to BraTS-ground truth (102). The Cancer Imaging Archive (TCIA) dataset is another publicly available source of Big Data eligible for machine-learning training and validation purposes. Bakas et al (2017) used the TCIA-dataset for developing their "GLISTRboost" algorithm, which won the BraTS-challenge in 2015 with a whole-tumor

[3] an event, typically lasting several days, in which a large number of people meet to engage in collaborative computer programming

similarity Dice-coefficient of 95%. The authors advise image based phenotyping through machine learning algorithms for radiotherapy planning purposes in their study, which was published in Nature (103).

Kickingereder, Isensee et al (2019) developed the HD-GLIO-algorithm, and aimed to engineer a framework relying on artificial neural networks for fully automated quantitative analysis of tumor burden in MRI. They used three datasets with 3220 MRI examination in total, including pre- and postcontrast T1, T2 and T2FLAIR. With this multispectral MRI-biomarker approach, they achieved a mean Dice-coefficient of 90% for enhancing tumor, and 93% for non-enhancing tumor (90). The source code is publicly available on the code hosting platform Github (https://github.com/MIC-DKFZ/HD-BET) and is the code used in this study.

3 Aim and research questions

3.1 Aim

The overall aim of this study is to investigate whether artificial intelligence with machine learning can be useful as a decision support tool for the treating oncologist, as a method in analyzing multispectral magnetic resonance imaging (MRI) of glioblastoma for treatment planning purposes. The project aims to compare the dual-compartment clinical manually delineated tumor volumes (primary GTV and CTV) in each dataset to the dual-compartment machine learning-derived tumor volumes (primary EC and NE), as illustrated in figure 30. The current clinical practice manual delineated tumor volumes are set to be ground truth, as is common practice in method studies and methodology development. If the machine learning method is proven highly correlated in terms of spatial overlap, the potential lies in using machine learning as an oncologist support tool for both more objective segmentation of the tumor volume, and for more quantitative segmentation of true malignant tissue. This is especially important with regards to the non-enhancing tumor compartment which is hard to differentiate qualitatively. The recent guidelines from the Norwegian Department of Health, stating that CTV-margins can be safely reduced to 10 mm for reduction of radiation induced neurotoxic damage for diffuse high-grade gliomas also adds to the gain of using a high-specificity tumor delineation approach (19).

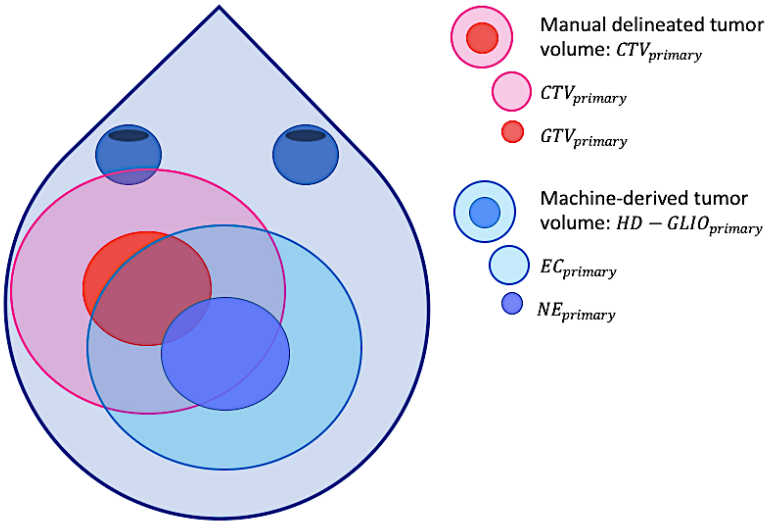


Figure 30 Illustration of tumor volumes in comparison. Red segments represent manual delineated tumor volumes; blue segments represent machine-derived tumor volumes. The purpose of this study is to analyze the degree of similarity/overlap between the volumes. Illustration: Marianne Hannisdal

The dual-compartments approach complies with radiotherapy planning in clinical practice and includes the central enhancing tumor volume representing the "gross" tumor compartment, and the surrounding non-enhancing tumor area representing the "clinical" infiltrative tumor compartment.

Additionally, the project aims to measure the similarity overlap between the site of recurrence to the primary manually- and the machine-derived tumor volume, respectively, as illustrated in figure 31. If the machine learning method perform better in predicting recurrence site (i.e. including site of recurrence in volume), the method could potentially have predictive value that could be taken into account when determining in-cranial tumor treatment site. Also, the CTV-margin is meant to include tissue with high probability for recurrence, like a prediction. Therefore, it is interesting to quantify to what degree it actually does include the site of tumor recurrence.

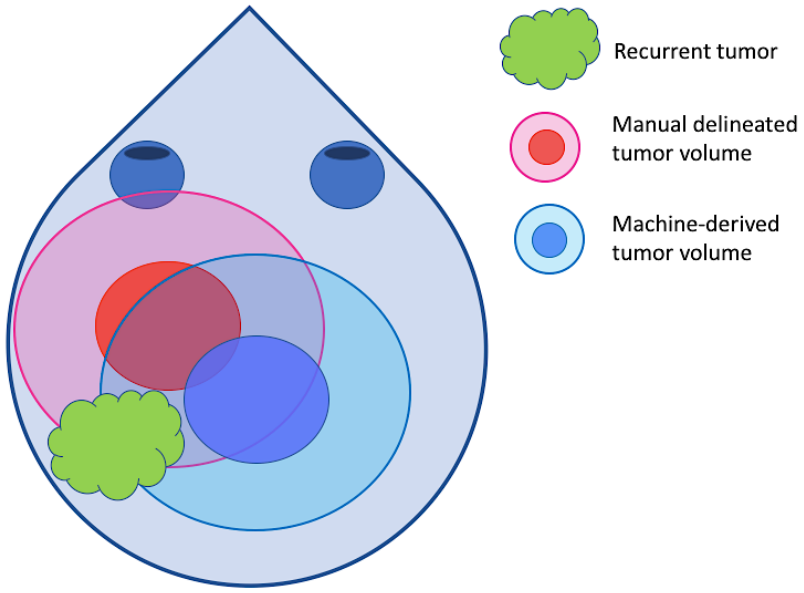


Figure 31 Illustration of tumor volumes in comparison with recurrent tumor. Red segments represent manual delineated tumor volumes, blue segments represent machine-derived tumor volumes, green segment represent recurrent tumor. The purpose of this study is to analyze which method has the largest degree of overlap with the tissue where the recurrent tumor evolves. Illustration: Marianne Hannisdal

Two MRI-examinations for each patient will be analyzed: $MRI_{primary}$ and $MRI_{recurrent}$. Volumes used in this study are named according to their tumor compartment and MRI examination. The clinical tumor volumes GTV and CTV correspond to the HD-GLO-volumes EC and NE, respectively.

3.2 Research questions

Q1.) For radiation therapy patients with primary and recurrent glioblastoma: to what extent is there a spatial overlap between machine-learning-derived tumor volumes, based on quantitative radiological biomarkers derived from multispectral MRI images, and standard manual oncologist-delineated target volumes, in terms of Dice-coefficients:

a) Dice-coefficient between $HD - GLIO_{EC\_primary}$ and $GTV_{primary}$

b) Dice-coefficient between $HD - GLIO_{NE\_primary}$ and $CTV_{primary}$

c) Dice-coefficient between $HD - GLIO_{EC\_recurrent}$ and $GTV_{recurrent}$

Q2.) For radiation therapy patients with primary and recurrent glioblastoma: to what extent is there a discrepancy in the inclusion of the recurrent tumor site in the primary target volume, when using machine-learning-derived primary tumor volumes, compared to standard manual oncologist-delineated target volumes, measured by:

a) Dice-coefficient between the localization of recurrent glioblastoma, $GTV_{recurrent}$, and $HD - GLIO_{primary}$

and

b) Dice-coefficient between the localization of recurrent glioblastoma, $GTV_{recurrent}$, and $CTV_{primary}$

3.3 Hypothesis

Based on the research questions raised above, I will approach my analyzing of the data material on the following hypothesis:

$Q1$ H_1: There is a significant spatial overlap between machine-learning-derived tumor volumes, based on quantitative radiological biomarkers derived from multispectral MRI images, and standard manual oncologist-delineated target volumes, in terms of Dice-coefficients.

$Q2\ H_1$: There is a significant discrepancy in the inclusion of the recurrent tumor site in the primary target volume, when using machine-learning-derived primary tumor volumes, compared to standard manual oncologist-delineated target volumes.

These stated hypothesis is accompanied by the following null-hypothesis:

$Q1\ H_0$:There is no significant spatial overlap between machine-learning-derived tumor volumes, based on quantitative radiological biomarkers derived from multispectral MRI images, and standard manual oncologist-delineated target volumes, in terms of Dice-coefficients.

$Q2\ H_0$:There is no significant discrepancy in the inclusion of the recurrent tumor site in the primary target volume, when using machine-learning-derived primary tumor volumes, compared to standard manual oncologist-delineated target volumes.

4 Methods

This study is an independent part of a larger study (https://clinicaltrials.gov/ct2/show/NCT03643549) for recurrent GBM-patients with unmethylated O6 methylguanine DNA methyltransferase *(MGMT*) promoter, called the "BORTEM-17-study", phase 1B (104). *MGMT* promoter methylation is an epigenetic biomarker that is both prognostic and predictive of the tumor's response to standard temozolomide chemotherapy, where patients harboring unmethylated *MGMT* have a poorer life-expectancy than patients with methylated *MGMT* promoter. The BORTEM-17 phase 1B study seeks to investigate the benefit and feasibility of pretreatment with the drug *bortezomib* in order to deplete levels of the *MGMT* DNA repair enzyme prior to temozolomide. The BORTEM-17 study also aims to identify radiological biomarkers that can improve radiographical volumetric treatment control of recurrent GBM, in which my study is an integral part of the overall objective.

4.1 Research Design

As the character of the empiric material in this study is numbers in terms of volumetric sizes and statistical Dice-coefficient measurements, a quantitative experimental research design was chosen. This study is a prospective, methodological study with the major aim to evaluate the benefit of a machine-learning approach as a method, in relation to the current standard method. A methodological study is designed to refine or develop methods of obtaining, organizing or analyzing data (105).

When it comes to blinding, this study`s quantitative character leaves little room for observer interpretation, as the calculated volumetric sizes are empirical numbers. The qualitative target delineation $CTV_{primary}$, made by the oncologists as a part of the clinical treatment plan was performed prior to start of the study; hence, the machine-derived segment could not induce any bias to the oncologist image interpretation. Further on, the delineation of $GTV_{recurrent}$ was also executed prior to deriving the $HD - GLIO_{recurrent}$, so no interpretation bias could have taken place.

4.2 Patient inclusion and exclusion criteria

Ten eligible patients with recurrent GBM and unmethylated *MGMT* promoter-status were included prospectively to the BORTEM-17 study phase 1B between 2018-2019, according to the BORTEM-17 inclusion criteria. The full inclusion criteria are listed in appendix A. Below is listed the major relevant inclusion criteria for this current study:

1. Histologically confirmed recurrent or progressed WHO grade IV intracranial glioblastoma (GBM)
2. Cranial MRI or contrast CT scan showing tumor relapse ≥ 12 weeks since radiation treatment
3. Measurable recurrent tumor
4. MRI evidence of recurrence within 14 days prior to enrolment
5. Available T1-, T2-, T1 with gadolinium enhancement- and T2FLAIR-image series with adequate image quality
6. Must be > 18 years old
7. Written informed consent for study participation
8. Estimated GFR ≥ 60
9. Negative pregnancy test no longer than 14 days prior to enrollment

In addition to the listed BORTEM-inclusion criteria above, I added the following inclusion criteria for my independent study in order for the dataset to be applicable for HD-GLIO analysis:

10. Available pre-radiotherapy MRI-examination including T2, T2FLAIR, and pre- and postcontrast T1

The participants were also selected according to the following exclusion criteria:

1. Serious medical or psychiatric illness that would interfere with study participation
2. Unable to lay still for MRI imaging
3. Another ongoing experimental therapy

4.3 Data Collection

As this study is a part of the BORTEM17- phase 1B study, the candidate was granted access to the radiological material of the 10 patients, according to the consent form on which all participants signed (Appendix H). All 10 patients had been previously received radiation therapy, whereof n=6 received radiation therapy at Haukeland University Hospital (HUH), n=2 received radiation therapy at Oslo University Hospital (OUH), and n=2 had received radiation therapy at Stavanger University Hospital (SUH).

4.3.1 Data exploration

Available image series were explored in Sectra Picture Archiving and Communication System (PACS) (Sectra AB, Linköping, Sweden). Image series from SUH were available for exploration in HUH PACS, but image series from OUH had to be requested by phone, put manually on a safe server, and imported manually into HUS image systems. This was done according to the clinical protocol as described in appendix C. The available image series for each of the ten patients is marked with +, while missing image series is marked with not available "*na"* in Table II below. The full data exploration details regarding image resolution, scanner type and field strength etc., are given in appendix B.

Table II Pre-radiotherapy image series available for the in-study patients

Imaging site	Patient	T1	T1 gadolineum	T2	T2FLAIR
HUH	1 Preoperative	+	+	+	+
	1 Postoperative	+	+	+	na
HUH	2 Preoperative	na	+	na	+
	2 Postoperative	+	+	+	na
OUS	3 Preoperative	+	+	+	na
	3 Postoperative	+	+	na	na
OUS	4 Preoperative	na	na	na	na
	4 Postoperative	na	+	na	+
HUH	5 Preoperative	+	+	+	+
	5 Postoperative	+	+	+	na
HUH	6 Preoperative	+	+	+	+
	6 Postoperative	na	na	na	na
OUS	7 Preoperative	na	na	na	na
	7 Postoperative	+	+	+	na
HUH	8 Preoperative	+	+	+	+
	8 Postoperative	na	na	na	na
SUH	9 Preoperative	+	+	+	+
	9 Postoperative	+	+	+	+
SUH	10 Preoperative	+	+	+	+
	10 Postoperative	na	na	na	na

Table showing available MRI-sequences marked with "+", and missing image sequences marked "na"(not available).

All patients in the study had had both surgery and radiation therapy for their primary GBM, however, not all MRI examinations were available due to different imaging sites and different imaging archive systems. Even though efforts were made to import imaging series from OUS, T2FLAIR-series were not available on any of them. Of the 10 patients available, only patient 9 had full preoperative and postoperative MRI. A total of 6 patients had full preoperative MRI including T2, T2FLAIR, and pre-and post-contrast T1, hereafter referred to as $MRI_{primary}$. Of these 6, n=4 had received radiotherapy at HUH, and n=2 had received radiotherapy at SUH. These 6 eligible patients were assigned patient numbers 1,5,6,8,9,10 according to the BORTEM-17 study-enrollment.

After exploring the data, only the patients with full sets of preoperative image series, i.e. patients 1, 5, 6, 8, 9, and 10 were included for further analyses. Patient 2, 3, 4, and 7 were excluded from the study due to missing image series. Take note that preoperative in this context does not necessarily mean that patients are surgically naïve.

4.3.2 Acquiring $CTV_{primary}$

$CTV_{primary}$ is used as ground truth in respect of primary tumor, as it is the segment on which patients have received radiation therapy according to current clinical practice.

The six eligible patients were identified in the treatment planning system Eclipse Aria Oncology Information Systems. Four patients had received treatment at Haukeland University Hospital, two patients had received treatment in Stavanger University Hospital (SUH). Patient data from SUH was collected via request in Eclipse, according to protocol described in the *Elektronisk Kvalitetssystem* (EK) (protocol enclosed in appendix C). The $MRI_{primary}$ was imported as DICOM-sequences from Sectra PACS for all six patients.

The $MRI_{primary}$ was co-registered to the $CT_{doseplan}$ for all six patients. The co-registration match was made on T1, rigidly and automatically performed on "Bone" (facial and scull structures), supervised and manually corrected if needed, according normal clinical practice. The co-registration was done to provide a "bridge" between image coordinates in CT and MRI, which is necessary in order to copy the GTV- and CTV-structures from CT onto MRI. The copied structures were thereby given MRI-relative *i,j,k*-coordinates in image space.

When the image material data collection was finished, all 6 patients were duplicated into de-identified study patients by three steps in Eclipse:

1. "export aria anonymous"-function, putting them in the "anonymous"-folder
2. they were moved from the "anonymous"-folder to the "DICOM-import"-folder
3. imported as øzz-study patients, according to EK-procedures (procedure enclosed in appendix D) and named "øzz_pt01" - "øzz_pt10", according to their original inclusion number in the BORTEM study.

4.3.3 Acquiring $HD-GLIO_{primary}$

The $HD-GLIO_{primary}$-volumes were acquired from the HD-GLIO brain tumor segmentation tool (89, 90). This was done by three main steps; preparing the dataset, running the script (full script enclosed in appendix E), and preparing the output-segments for comparison, as will be explained in the following sections.

Preparing the clinical MRI-images ($MRI_{primary}$) for analysis

After the de-identification duplication described in section 4.3.2, the now de-identified $MRI_{primary}$ was exported to a portable hard drive in DICOM-format. The portable hard drive was kept in a locked locker on hospital grounds when not in use, according to research policy at HUH. The files were prepared for analysis with HD-GLIO by the following steps:

1. the de-identified DICOM-sequences were converted into Neuroimaging Informatics Technology Initiative (NIFTI)-files, according to HD-GLIO standards This was performed using ITK-snap(106) version 3.8.0 by loading the DICOM image series and saving them as NIFTI image series.
2. The NIFTI files were then compressed into nii.gz zipped format by using the [gzip filename.nii.gz] command in MacOS terminal (version 2.8.3)
3. The compressed NIFTI-filename was named according to HD-GLIO standards; T1CE, T1, T2, FLAIR with patient-specific numbering prefixes.

Running the script

4. Installing HD-GLIO in Python (version 3) with [pip install hd_glio]

5. HD-GLIO was then run on $MRI_{primary}$, on a DELL Precision 7540 laptop with Ubuntu 18.04 and NVIDIA Quadro RTX 3000 6GB GPU, using the T1CE, T1, T2 and FLAIR-series on each patient as input. This step was run by my co-supervisor Arvid Lundervold at the Mohn Medical Visualisation Senter (MMIV), as I did not have access to a machine with the required NVIDIA GPU on minimum 6 GB VRAM

Output preparation

6. Output was imported to 3Dslicer (version 4.11) in compressed NIFTI-format (nii.gz) and converted to synthetic CT in DICOM-format, using the "create DICOM series"-module, providing the EC-segment with HU-value=2 and NE-segment with HU-value=1.

7. The output DICOM-series was imported to the corresponding de-identified study patient in Eclipse. The output DICOM series holds the same image-coordinates as the $MRI_{primary}$, and is therefore automatically registered to the anatomical image series from which it was first derived.

8. Three volumes were derived from the HD-GLIO output segments in Eclipse, name of structure was followed by the patients enrollment number in BORTEM-17 (01-10):

    a. Enhancing Core $EC_{primary}$

    b. Non-enhancing tumor $NE_{primary}$

    c. Fusion of the two volumes above: $HD - GLIO_{primary}$ by bolean operation

Fitting the data to the model

9. For patients 1,5,6,9,10: two additional segments were acquired by postprocessing command "fill all cavities":

    a. $EC_{primary}$ including central lesion

    b. $HD - GLIO_{primary}$ including central lesion

The purpose of these two additional segments described under 9 a-b) was more realistic volumetric comparison grounds with $CTV_{primary}$, as the nature of a CTV always includes all

cavities and radiotherapy is not given with sparing cavities. When volumes including cavities are used in the results-section, this will be explicitly given in the text. “Cavities” only include true *islands*, meaning a cavity had to be surrounded by original segmentation on all sides to be defined as a cavity. Example of including a cavity in the volume is illustrated in figure 32 as follows:

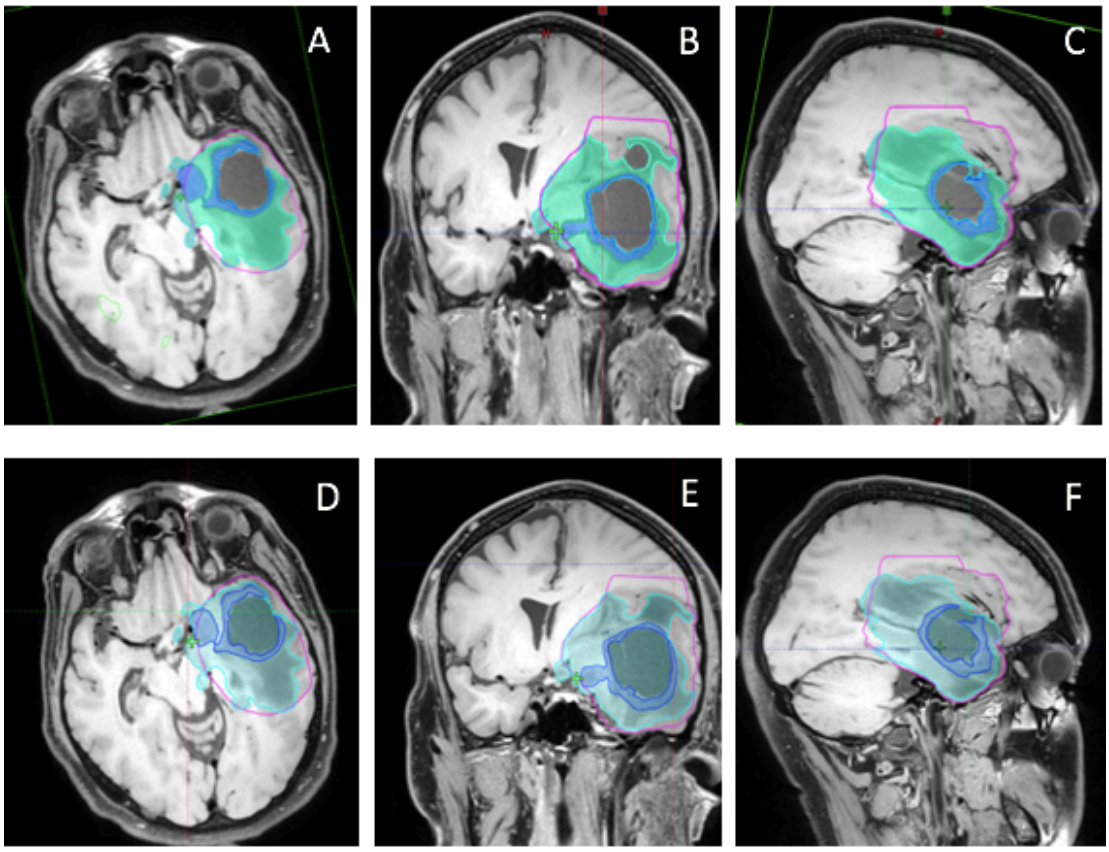


Figure 32 Illustration of segments. In A, B and C on the top row is axial, coronal and sagittal view, respectively of ECprimary=Blue, NEprimary=Green. Pink line is the CTV-outline. On the lower row D, E and F shows is axial, coronal and sagittal view, respectively, with an addition superimposing segment, in which central lesions are included in the segment. Illustration: Marianne H Hannisdal (in-study patient)

4.3.4 Acquiring $GTV_{recurrent}$

The $MRI_{recurrent}$ datasets were obtained directly from the BORTEM-17 crew in de-identified DICOM format. The recurrent glioblastoma $GTV_{recurrent}$ as it occurred on the first MRI-examination after inclusion in the BORTEM-17 study $MRI_{recurrent}$, was defined and volumetrically quantified on a contrast enhanced T1-weighted high-resolution 3D MP-RAGE (Magnetization Prepared - RApid Gradient Echo)-sequence with isotropic voxel size 1 mm^3. This was performed using manual labeling in ITK-SNAP version 3.6.0 (106) by in-painting slice-by-slice on axial slices. The delineation of $GTV_{recurrent}$ included all contrast

enhanced areas of the tumor, excluding the surgical resection cavity (if any) but included any residual enhancing tumor situated along the resection cavity boundary. The delineation was supervised and approved by oncologist at HUH. The window level (WL) and window width (WW) of all images was kept constant and unchanged throughout the course of the delineation process.

4.3.5 Acquiring $HD - GLIO_{recurrent}$

I preformed the exact same steps on the $MRI_{recurrent}$ datasets as for the $MRI_{primary}$ datasets, according to the steps numerically listed in section 4.3.3 "acquiring $HD - GLIO_{primary}$".

4.4 Variables and analysis

The relationships between the output volumes of manual delineation as well as machine-derived volumes are presented as follows.

Manually delineated volumes

The $CTV_{primary}$ includes two volumes[4]: the full $CTV_{primary}$ and the inner central $GTV_{primary}$, according to clinical practice as explained in section 2.5.2.

The $GTV_{recurrent}$ is a single volume including the enhancing part of the recurrent tumor. The relation between the volumes in illustrated in figure 33 below.

[4] Pt 10 did not have GTV in the clinical plan, this is an oncologist-dependent variable

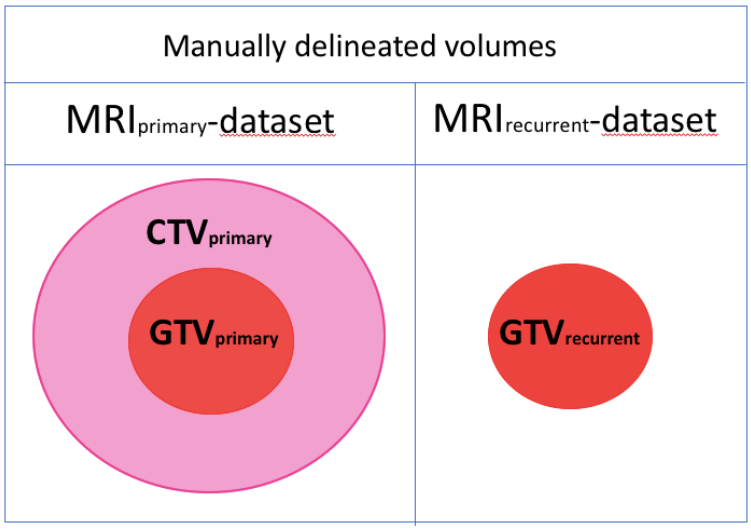


Figure 33 Illustration of the relationships between the manually delineated volumes. Colors are in identical to those used for volume visualization in following figures. Illustration: Marianne H Hannisdal

Machine learning-derived volumes

The $HD-GLIO_{primary}$ include three volumes; the enhancing core (EC), the non-enhancing tumor (NE), and the sum of (EC+NE) + cavitation, as described in section 4.3.3, number 8d.

The $HD-GLIO_{recurrent}$ include two volumes; the enhancing core (EC) and the non-enhancing tumor (NE). The relationships between them are given in figure 34.

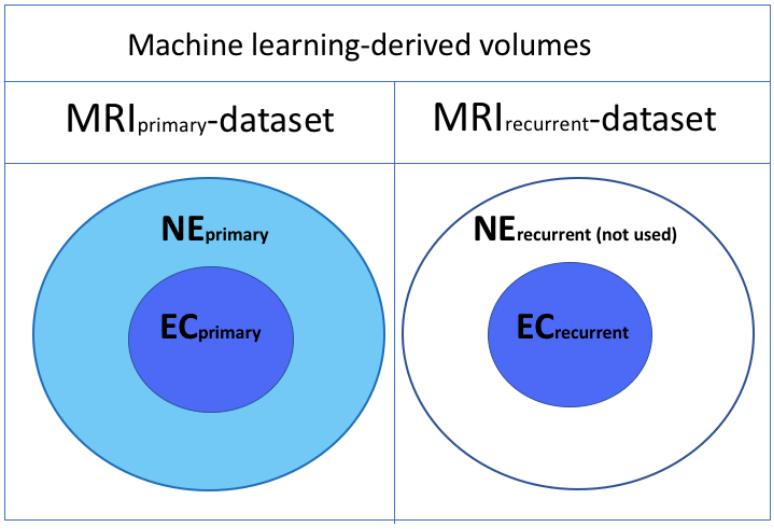


Figure 34 Illustration of the relationships between the machine learning-derived volumes. Colors are identical to those used for volume visualization in following figures. Illustration: Marianne H Hannisdal

The relationship between datasets and volumes in total is presented in table III as follows.

Table III Datasets with adjacent volumes used for the analysis

		Datasets			
		Primary tumor $MRI_{primary}$		Recurrent tumor $MRI_{recurrent}$	
Tumor volumes	Clinical tumor volume	$CTV_{primary}$		$GTV_{recurrent}$	
		$GTV_{primary}$	$CTV_{primary}$	$GTV_{recurrent}$	
	Machine-derived tumor volume (HD-GLIO-algorithm)	$HD-GLIO_{primary}$		$HD-GLIO_{recurrent}$	
		$EC_{primary}$	$NE_{primary}$	$EC_{recurrent}$	$NE_{recurrent}$

The $CTV_{primary}$ includes two volumes: a Gross Target volume (GTV) and a Clinical target volume (CTV). The $GTV_{recurrent}$ is a single volume. The $HD-GLIO_{primary}$ include two volumes: Enhancing Core (EC) and NonEnhancing tumor (NE). The $HD-GLIO_{recurrent}$ include two volumes: EC and NE, of which only NE is used in this study.

4.4.1 Analyzing the volumes

The Dice-coefficients and volumetric sizes results were acquired from the clinical treatment planning system Eclipse, using the "statistics"-tool in the "SmartSegmentation" module. Manual delineations representing current clinical practice was defined ground truth.

4.4.2 Analyzing the Dice-coefficient results

All results were analyzed in Prism 8 for macOS (GraphPad Software, version 8.4.3 (471) La Jolla, California, www.graphpad.com). The Dice-Coefficient results was analyzed for variance by Friedman's test, as well as successively analyzed for outliers by Grubbs test, and analyzed for statistical significance by Wilcoxon test for non-parametric[5] data. The Grubbs test presumes normal distributed data, so for this non-parametric data the result was carefully evaluated manually, confirming the Grubbs results. The alpha value is set to 5% for all

[5] Not inhabiting gaussian distribution

statistical analyses, and sets the standard for how extreme the data must be before the research question is answered with a statistically significant result. This gives a level of significance of 95%, meaning the probability of the observed data has occurred due to an actual measured effect is at least 95%, when the p-value is below 0.05.

4.5 Methodological ethical concerns

The BORTEM-17 study within which this study is an independent part of, has been approved by the Regional Ethics Committee in Western Norway, reference number 2017/2084/REK vest (Approval letter attached in appendix F). The need for data protection impact assessment (DIPA) evaluation for this thesis was discussed with the supervisors of this thesis and found to be not necessary. The BORTEM-17 study was reported to the official Data Protection Officer (DPO) of HUH (Rapport enclosed in appendix G), and this independent study was also discussed with the HUH DPO to ensure all parts of the methodological approach was ethically substantial, and that all use of patient data was approved through the BORTEM-17 study approval and patient information.

The patients were presented with both oral and written information of the study (Patient information letter attached in appendix H), including information of use of radiological data, before asked to give informed consent. The patients did not undergo any additional exams or additional image sessions, as my study only took use of data material already available. The study results have no influence on the participants' cancer treatment as it only retrospectively takes into account the data, investigating the hypothetical volumetric outcome using HD-GLIO.

All analysis have taken place on de-identified datasets, but the identification key has been available in a locked cabinet.

I chose to keep the original enrollment numbering according to the BORTEM-17 study to ease further analysis of the data material. However, the patients are aware of their enrollment number, and a situation could occur where they, or their next of kin, read about their results in published rapports. Facial features were masked on 3D-illustrations to prevent recognition. Still, the results describe to some detail individual patient disease- and treatment course, which for some, could be problematic in respect of retrospective treatment analysis. The fact that these patients have undergone aggressive treatment to match the aggressiveness of their

tumors does not come without personal cost for the individual patients. In the same time, these patients are well informed that they are part of a research project regarding tumor treatment, and have given their consent to their disease and treatment can undergo thorough investigation. They are also informed that the results of the research project will be published.

4.6 Reliability and validity

Validity is of the essence in good research. Validity is an expression of whether or not the findings in the study are unbiased and well grounded, true for the studied population (internal validity) and ultimately, if the results are transferable onto other populations (external validity). Reliability refers to the accuracy and consistency of measured variables (105).

The BORTEM-17 study that this thesis is an individual part of was a phase 1 study with 10 participants. As four of these had to be excluded in this study due to missing $MRI_{primary}$ image series, the sample size (n=6) is low. However, there are three separate volume-comparisons for investigating Q1 regarding correlation, giving an effective sample size of n=18, hence increasing statistical power. For investigating Q2 regarding the attempt of isolating the predictive value of the methods in comparison, there is an actual sample size of six patients, but as there are two longitudinal volume comparisons, the overall effective sample size of 12. The longditunal nature of the study therefore increases the statistical power, hence increases the validity of the study. However, the reliability in terms of grounds for generalization to a bigger population is still somewhat limited.

When comparing HD-GLIO segments derived from $MRI_{primary}$ to the ground truth recurrent tumor $GTV_{recurrent}$ as it appears on the $MRI_{recurrent}$ dataset for Q2 investigation, there lies a methodological insecurity in tissue localization shift. Potential factors inducing tissue localization shift include radiotherapy-related atrophy, pseudoprogression, and recurring tumor mass volume compressing effect. E.g. if the ventricles have doubled in size from one imaging timepoint to the other, the tissue inhabiting the primary tumor-cells could have been displaced, compressed or for some other reason shifted their in-cranial position. The idea of doing longitudinal segment comparison is to investigate what has happened with the tissue a specific in-image *i, j, k* -localization over time, and does not take into account that the tissue in this voxel may have been replaced by neighboring tissue due to tissue localization shift. This theoretical methodological insecurity lies inherent in using rigid co-registration;

however, the alternative of using deformable registration would only induce more methodological insecurities.

There lies an inherent methodological weakness in comparing a machine learning-derived segment of non-enhancing tumor with theoretically high specificity, to a ground truth manually delineated segment in which a large insecurity margin, hence low specificity. The intrinsic low specificity in the ground truth can make it hard to isolate an effect in respect of accuracy in the comparing method. Because of this, the true positive part of the Dice similarity coefficient could be investigated as an indication of specificity, along with the Dice similarity coefficient score in general.

5 Results

5.1 Patient characteristics

Of the ten patients with recurrent glioblastoma in BORTEM-17 phase 1 study, patient number 1, 5, 6, 8, 9 and 10 had eligible image material available to perform analysis with HD-GLIO according to minimum requirements. Mean age at treatment for primary GBM was 44,4 years (median=48). Gender distribution was male n=8, female n=2. Patient- and treatment characteristics, in addition to structure volumetrics are given in the table below.

Table IV In-study-patients characteristics in MRIprimary

P	*Tumor site**	*Age **	*Treatment site*	*Gender*	*GTV-size*	*CTV-size*	*EC size*	*HD-GLIO size*	*Days to MRIrecurrent****
1	FL	44	HUH	M	6.5	205.7	3.1	222.0	1294
2	TL-R	52	HUH	M	45.2	107.0	--	--	--
3	LH	52	OUS	M	60.1	103.9	--	--	--
4	RH	39	OUS	F	72.7	263.8	--	--	--
5	FL	33	HUH	M	26.0	163.8	0.5	70.5	491
6	PL	34	HUH	F	7.8	223.4	2.4	21.2	312
7	RH	25	OUS	M	241.5	507.1	--	--	--
8	FL	56	HUH	M	20.5	201.1	0.0	16.6	515
9	RH	52	SUH	M	45.0	299.9	25.4	132.7	239
10	LH	57	SUH	M	--[6]	236.6	16.1	163.5	340

*P=Patient number according to enrollment. FL=Frontal Lobe, TL=Temporal Lobe, R=Right side, RH=Right Hemisphere, PL=Parental Lobe, LH=Left Hemisphere. M=male, F=female. nf=not found in patient journal, patient treated outside HUH. All volumetric sizes is given in cm3. *as listed in patient journal. **at time of enrollment. ***Number of days between MRIprimary and MRIrecurrent*

All patients were recurrence-free for at least 12 weeks between radiotherapy and enrollment to the study.

[6] GTV was not delineated

5.2 Q1: Correlation between clinical and machine learning-derived tumor volumes

5.2.1 Primary tumor site segmentation results in $MRI_{primary}$

One aim of the study was to compare the manually delineated volumes to the machine-derived volumes of the primary MRI-examination. The volumes in $CTV_{primary}$ and $HD-GLIO_{primary}$ that are compared with Dice-coefficient are :

Q1-A: $GTV_{primary}$ (red) versus $EC_{primary}$ (dark blue)

Q1-B: $CTV_{primary}$ (pink) versus total $HD-GLIO_{primary}$ (cyan)

Due to the heterogenous patient material, image material and small sample size, the segmentation results for are presented for each patient to fully illustrate the results, in the following figures:

Patient 1

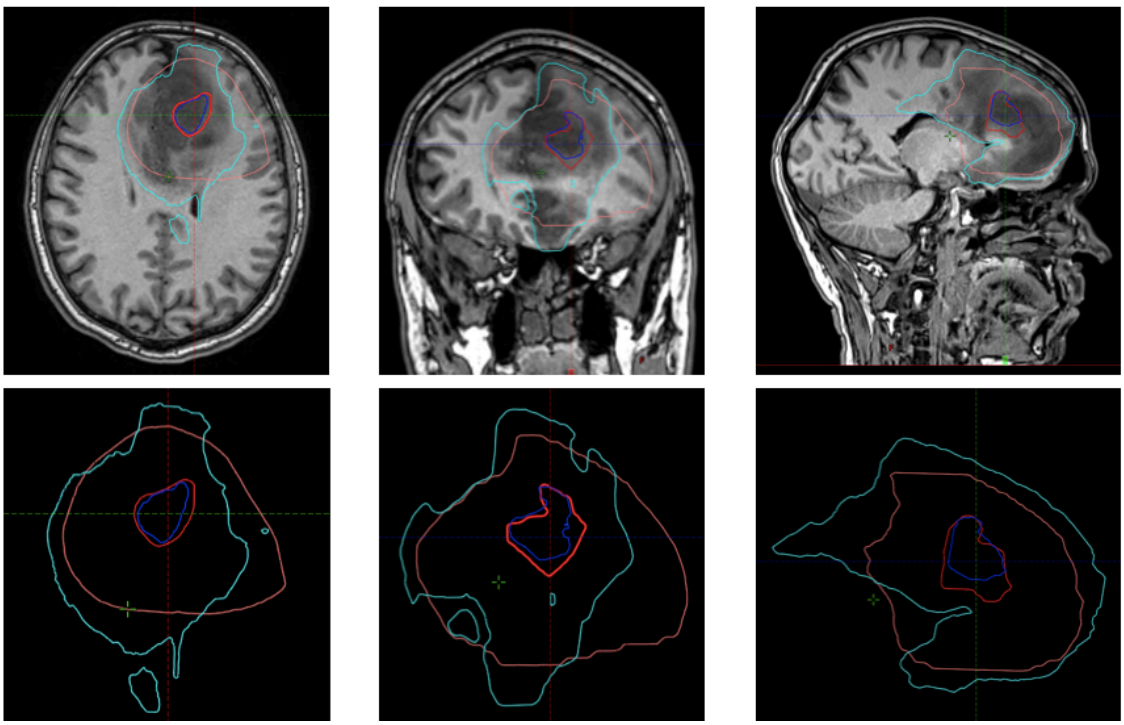


Figure 35 Segments in Patient 1: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their original anatomical background, bottom row: segments only. Red line= Gross Target Volume (GTV), Blue line=Enhancing Core (EC) compartment of $HD-GLIO_{primary}$, Pink line=Clinical Target Volume (CTV) Light blue=total $HD-GLIO_{primary}$. Illustration: Marianne H Hannisdal

Patient 5

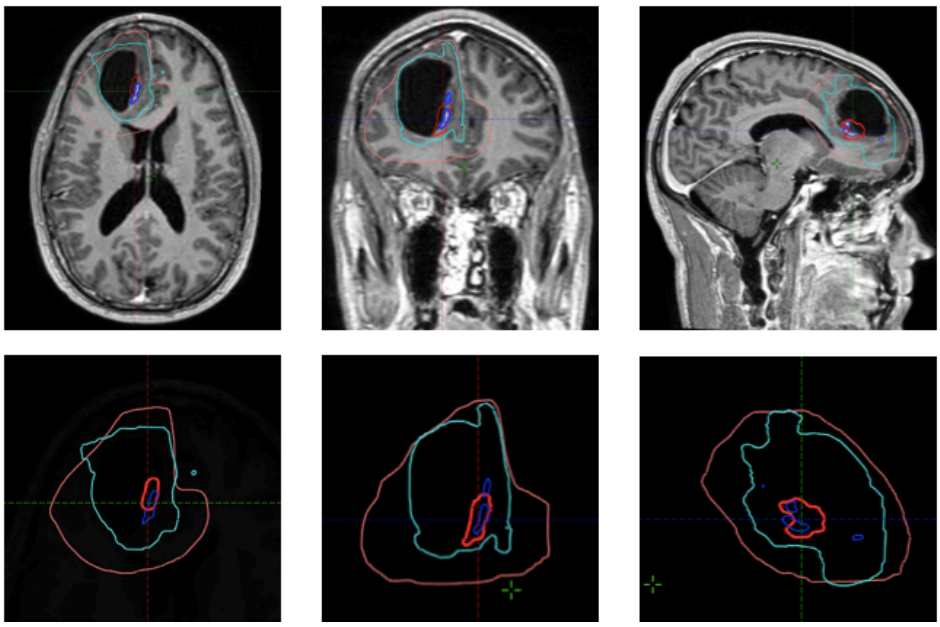


Figure 36 Segments in Patient 5: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their original anatomical background, bottom row: segments only. Red line= Gross Target Volume (GTV), Blue line=Enhancing Core (EC) compartment of $HD-GLIO_{primary}$, Pink line=Clinical Target Volume (CTV) Light blue=total $HD-GLIO_{primary}$ Illustration: Marianne H Hannisdal

Patient 6

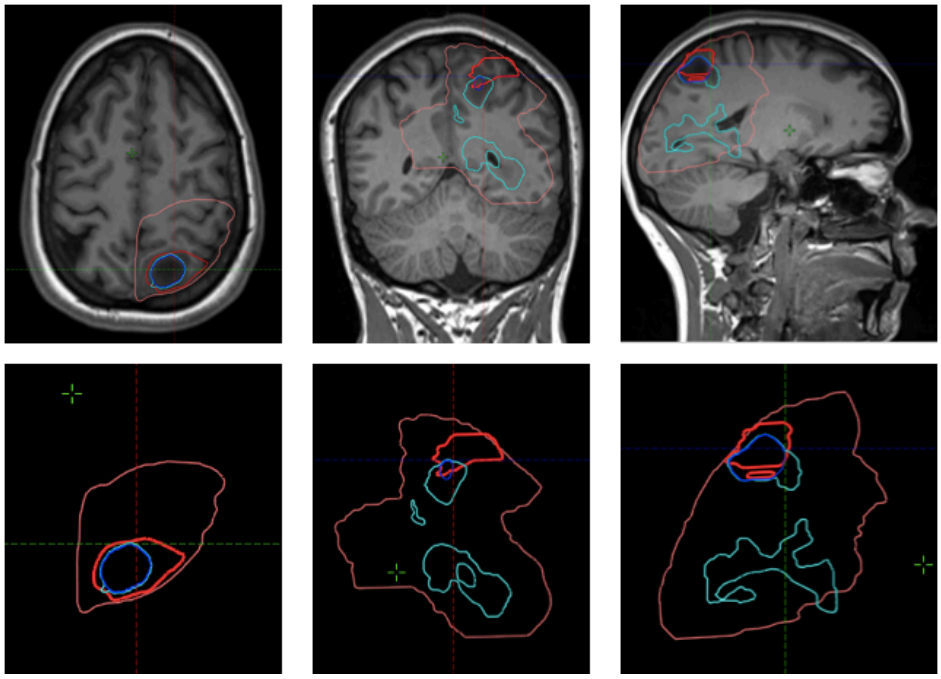


Figure 37 segments in Patient 6: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their original anatomical background, bottom row: segments only. Red line= Gross Target Volume (GTV), Blue line=Enhancing Core (EC) compartment of $HD-GLIO_{primary}$, Pink line=Clinical Target Volume (CTV) Light blue=total $HD-GLIO_{primary}$ Illustration: Marianne H Hannisdal

Patient 8

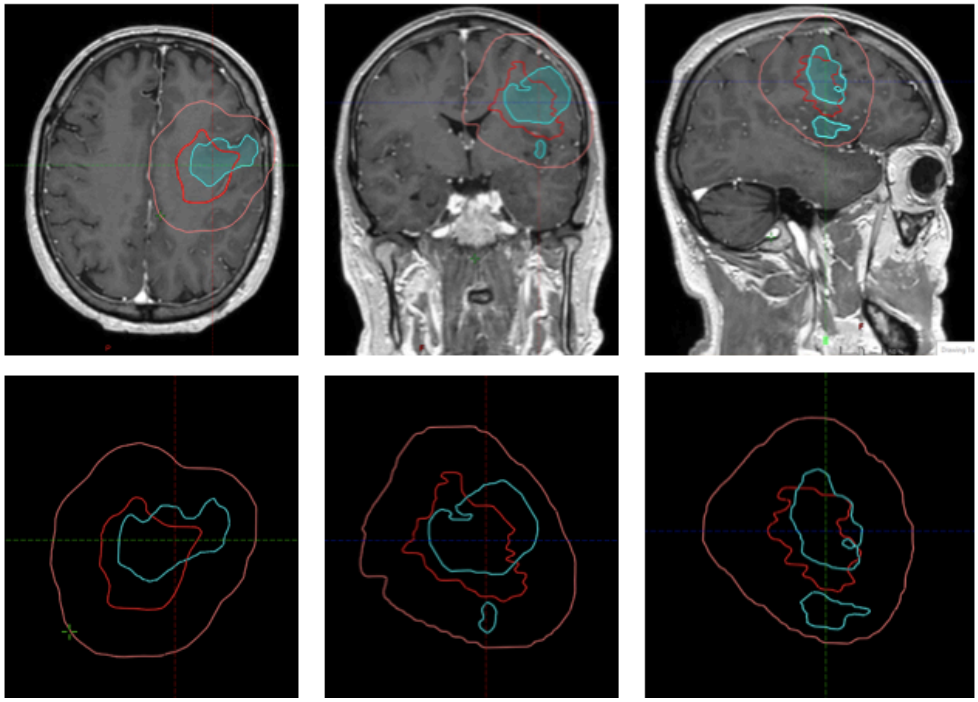


Figure 38 segments in Patient 8: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their original anatomical background, bottom row: segments only. Red line= Gross Target Volume (GTV), Blue line=Enhancing Core (EC) compartment of $HD-GLIO_{primary}$, Pink line=Clinical Target Volume (CTV) Light blue=total $HD-GLIO_{primary}$ Illustration: Marianne H Hannisdal

Patient 8 had no segmented Enhancing Core by HD-GLIO.

Patient 9

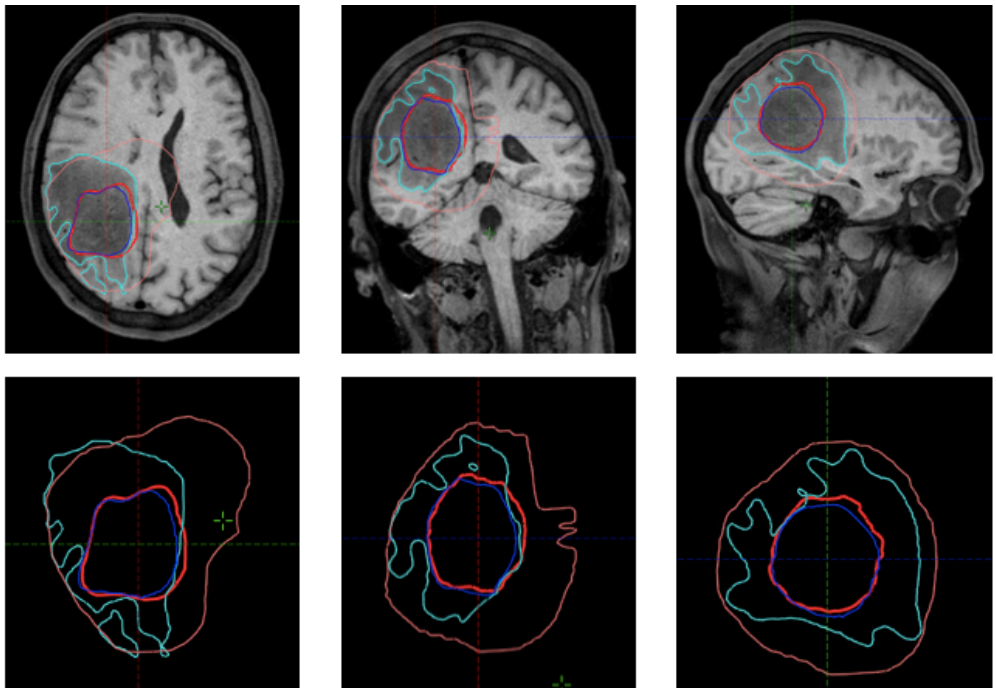


Figure 39 segments relationships in Patient 9: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their

original anatomical background, bottom row: segments only. Red line= Gross Target Volume (GTV), Blue line=Enhancing Core (EC) compartment of $HD - GLIO_{primary}$, Pink line=Clinical Target Volume (CTV) Light blue=total $HD - GLIO_{primary}$ Illustration: Marianne H Hannisdal

Patient 10

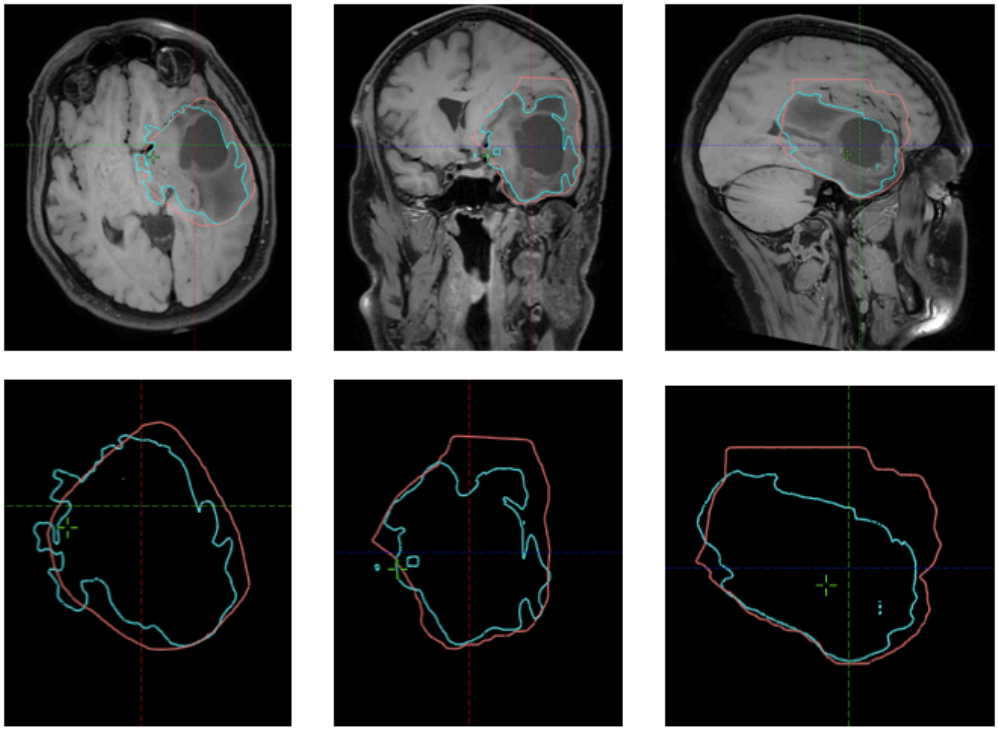


Figure 40 segments in Patient 10: Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: manual and HD-GLIO segments on their original anatomical background, bottom row: segments only. Pink line=Clinical Target Volume (CTV) Light blue=total $HD - GLIO_{primary}$ Illustration: Marianne H Hannisdal

Patient 10 had no GTV delineated by oncologist, therefore Enhancing Core compartment of HD-GLIO is also excluded from the figure 40 above.

5.2.2 Dice-coefficient results in $MRI_{primary}$

First, the machine learning-derived enhancing core $EC_{primary}$ including cavities were compared to the ground truth manually delineated gross target volume $GTV_{primary}$ by the oncologist. Secondly, the total machine learning-derived tumor volume $HD - GLIO_{primary}$ including both EC and NE-volumes and cavities, were compared to the ground truth manually delineated clinical target volume, $CTV_{primary}$. Dice similarity coefficient results and true positive results of these comparisons is presented in table V below.

Table V Dice-coefficient and True Positive results in $MRI_{primary}$

	A $GTV_{primary}$ vs. $EC_{primary}$		B $CTV_{primary}$ vs. $HD-GLIO_{primary}$	
	Dice-score	True Positive	Dice-score	True Positive
Pt 1	0.70	0.94	0.67	~~0.64~~
Pt 5	0.34	0.33	0.57	0.94
Pt 7	0.54	0.80	0.17	0.98
Pt 8	--	--	0.15	1.00
Pt 9	0.87	0.92	0.60	0.98
Pt 10	--	--	0.77	0.93
Mean	**0.49**	**0.75**	**0.49**	**0.97**
median	**0.54**	**0.80**	**0.58**	**0.98**
SD	**0.34**	**0.29**	**0.27**	**0.03**
SEM	**0.15**	**0.14**	**0.11**	**0.05**
p-value	**0.125**	**0.12**	**0.03**	**0.01**

All volumetric sizes is given in cm3.
–not taken into account. Patient 8 had no Enhancing Core (EC) segmented by HD-GLIO. Patient 10 did not have GTVprimary delineated by oncologist. True positive on Patient 1 in dataset B was removed as an outlier by Grubbs test.

Grubbs outlier test was performed as described in the method section, and True Positive result on patient 1 in dataset B was excluded from further analysis.

5.2.3 Recurrent tumor site segmentation results in $MRI_{recurrent}$

Majority of GBM patients experience recurrence of tumor after treatment, and another aim of this study was to compare the similarity in detection of spatial location of the recurrent tumor between manual and machine-derived segmentation. The volumes in $MRI_{recurrent}$ that were compared with Dice-coefficient are:

Q1-C: $GTV_{recurrent}$ (magenta) versus $EC_{recurrent}$ (dark blue)

Patient 5

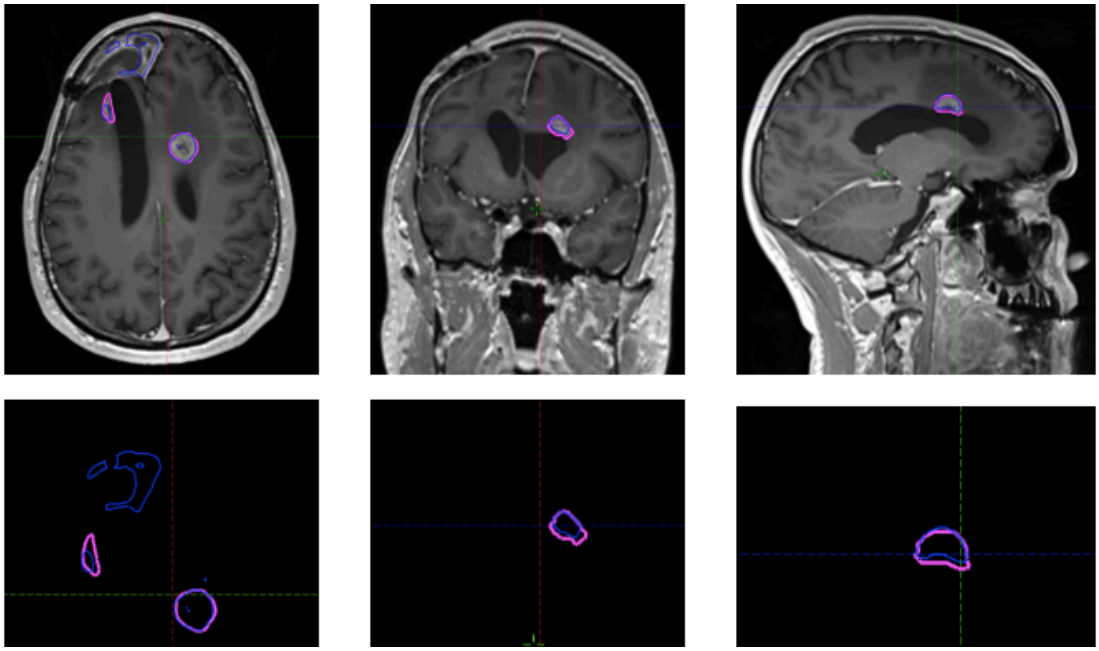


Figure 41 Patient 5: Segments in $MRI_{recurrent}$. Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: segments on their original anatomical background, bottom row: segments only. Gross Target Volume (GTV) (magenta) and Enhancing Core (blue). HD-GLIO detects the same two lesions that is manually delineated, but HD-GLIO also segments a lesion in the frontal lobe that is clinically described as post-operative changes in relation to the meninges (ref: patient journal). Illustration: Marianne H Hannisdal

Patient 6

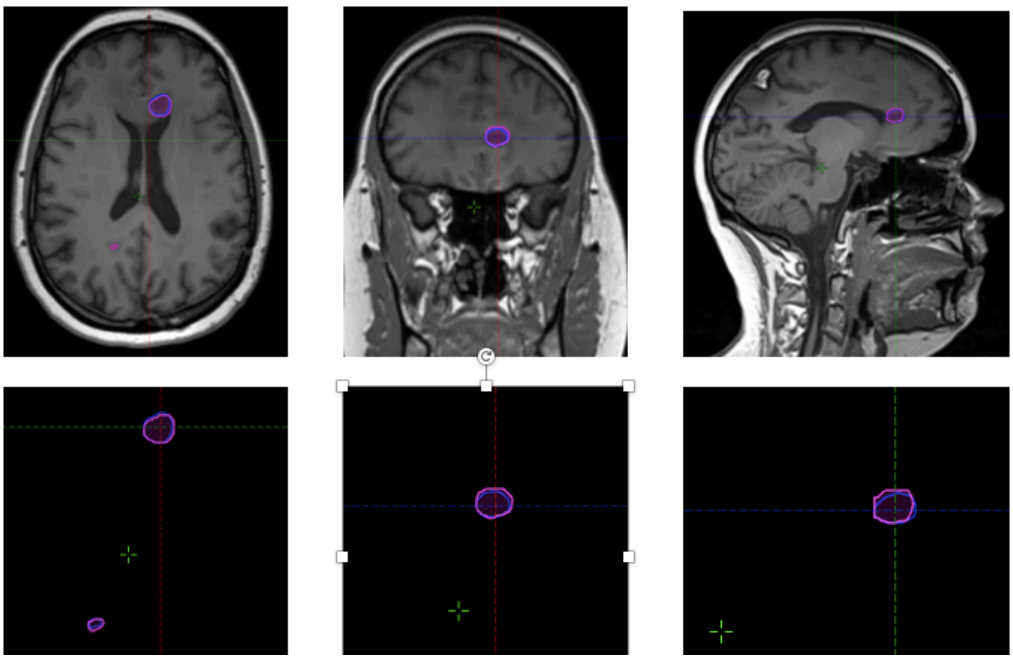


Figure 42 Patient 6: Segments in $MRI_{recurrent}$. Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: segments on their original anatomical background, bottom row: segments only. Gross Target Volume (GTV) (magenta) and Enhancing Core (blue) Illustration: Marianne H Hannisdal

Patient 8

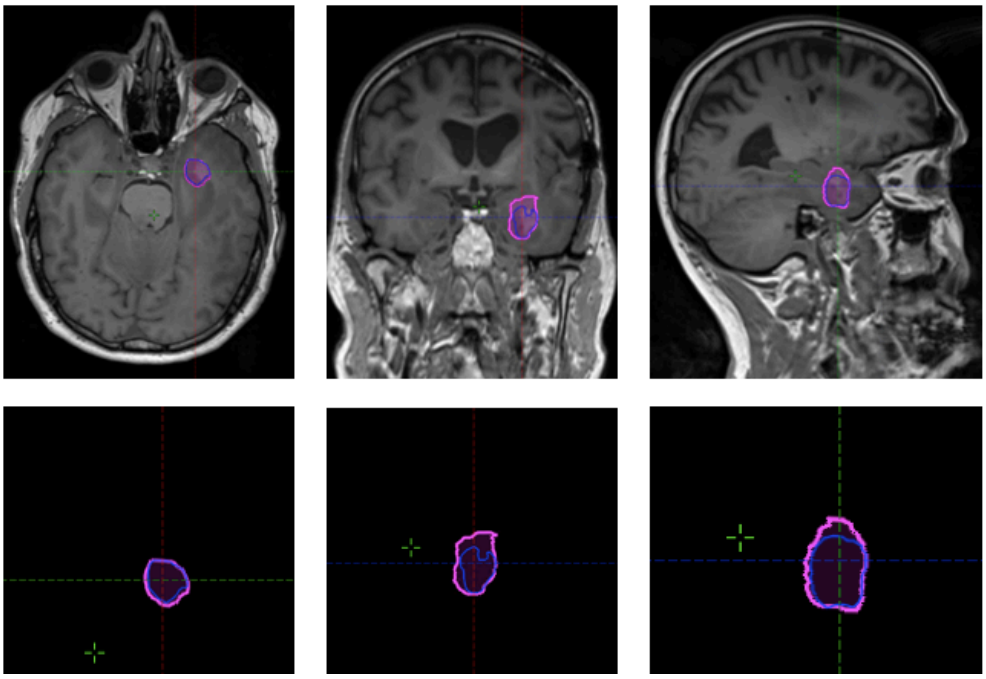


Figure 43 Patient 8: Segments in $MRI_{recurrent}$. Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: segments on their original anatomical background, bottom row: segments only. Gross Target Volume (GTV) (magenta) and Enhancing Core (blue). Illustration: Marianne H Hannisdal

Patient 9

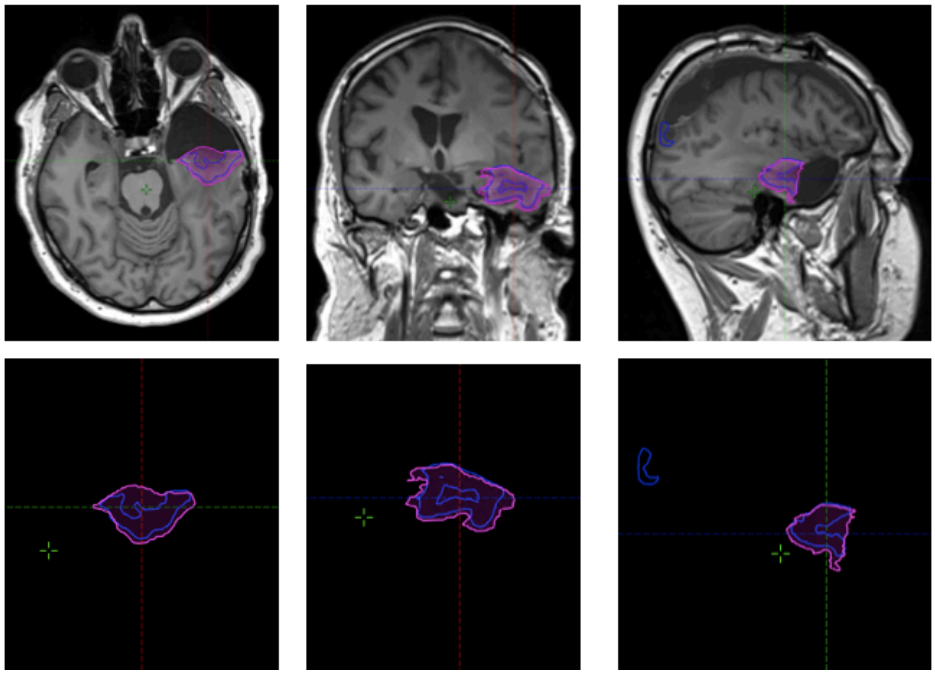


Figure 44 Patient 9: Segments in $MRI_{recurrent}$. Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: segments on their original anatomical background, bottom row: segments only. Gross Target Volume (GTV) (magenta) and Enhancing Core (blue). Illustration: Marianne H Hannisdal

Patient 10

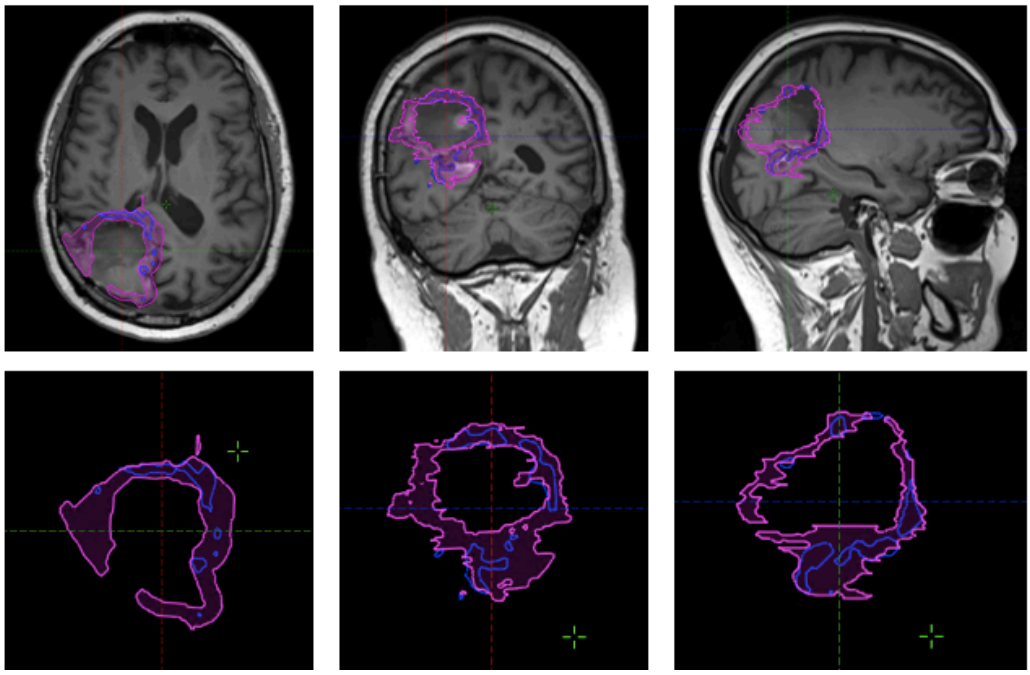


Figure 45 Patient 10: Segments in $MRI_{recurrent}$. Axial view (left -top and bottom), coronary view (middle – top and bottom), sagittal view (right – top and bottom). Top row: segments on their original anatomical background, bottom row: segments only. Gross Target Volume (GTV) (magenta) and Enhancing Core (blue). Illustration: Marianne H Hannisdal

5.2.4 Dice-coefficient similarity between $GTV_{recurrent}$ and $EC_{recurrent}$

The machine learning-derived enhancing core $EC_{recurrent}$ compartment of $HD-GLIO_{recurrent}$, was compared to the ground truth manually delineated enhancing tumor volume, $GTV_{recurrent}$.

Volume sizes, Dice-coefficient similarity and True Positive results is presented in table VII.

Table VI Volumetric sizes and Dice-coefficient similarity between $GTV_{recurrent}$ and $EC_{recurrent}$

P	$GTV_{recurrent}$ *size*	$EC_{recurrent}$ *size*	*Dice-coefficient*	*True Positive*
1	18.5	--	--	--
5	2.5	13.9	0.15	~~0.05~~
6	0.9	0.7	0.79	1.00
8	3.5	1.9	0.71	1.00
9	10.7	8.0	0.67	0.80
10	42.6	12.2	0.37	0.87
mean	**13.1**	**6.1**	**0.53**	**0.92**
median	**7.1**	**5.0**	**0.67**	**0.94**

All volumetric sizes is given in cm3. -- missing/not taken into account. –True positive of patient 1 was found to be an outlier by Grubbs test, and excluded from further analysis.

Grubbs outlier analysis was performed as described in the method section, and True Positive result on patient 5 was excluded from the dataset.

5.2.5 Q1 Overview and statistical analysis:

Dice-coefficient similarities between manual delineation and machine-derived segmentation on the two MRI-datasets are listed in the table below, along with statistical calculations performed in Prism 8. The sample size *n* is 16: there are six patients undergoing three separate measures (n=6*3=18), but patient 1 is missing $EC_{recurrent}$, and patient 10 is missing $GTV_{primary}$. The A, B, C datasets (A=$GTV_{primary}$ *vs.* $EC_{primary}$, B=$CTV_{primary}$ *vs.* $HD - GLIO_{primary}$ and C=$GTV_{recurrent}$ vs. $EC_{recurrent}$) was successively analyzed for outliers with the Grubbs test, resultingly True Positive-value of patient 1 in dataset B, and True Positive value of patient 5 in dataset C was identified as an outlier. These outlier values are excluded from further analysis.

The Dice-coefficient scores and True Positive scores of the segments in comparison are given group-wise in figure 46 and 47, respectively.

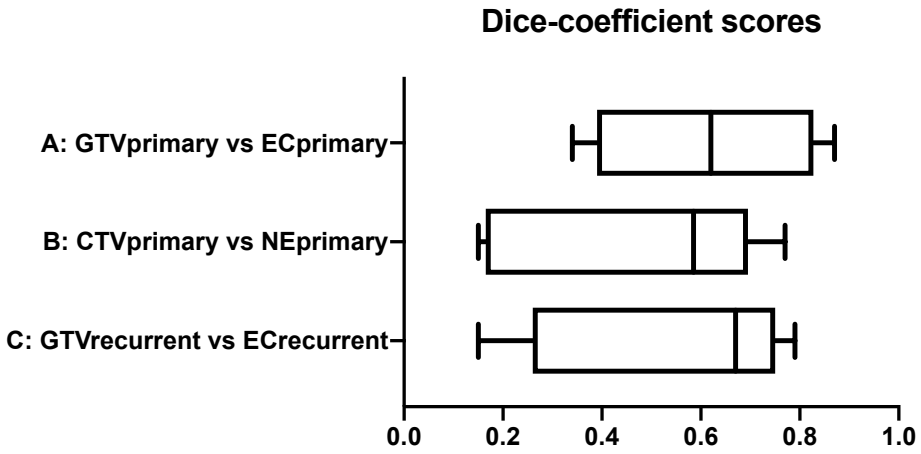


Figure 46 Box-plot of Dice-coefficient group scores for the segments in comparison, respectively. A: mean= 0.61, median 0.62 range=0.53 (0.34-0.87). B: mean=0.49, median=0.58, range=0.62 (0.15-0.77). C: mean=0.54, median=0.67, range=0.64 (0.15-0.79).

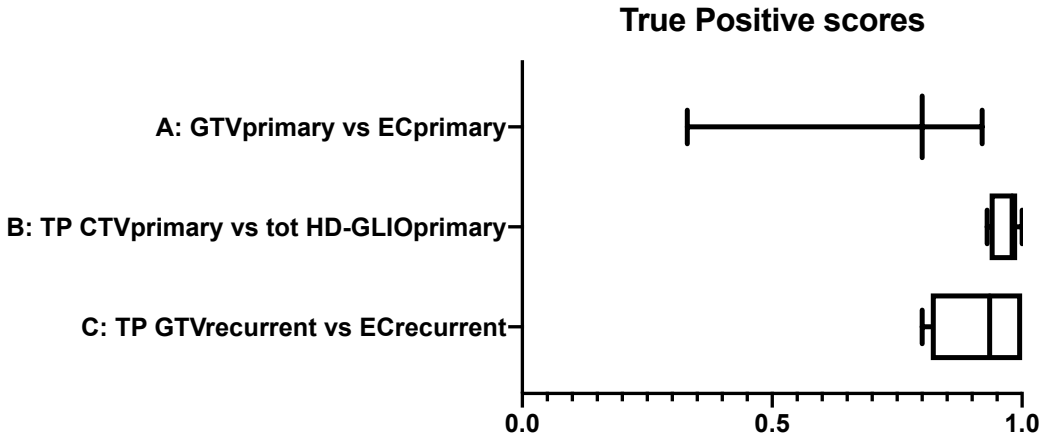


Figure 47 Box-plot of True Positive group scores for the segments in comparison, respectively. A: mean=0.68, median=0.80, range=0.59 (0.33-0.92). B: mean=0.97, median= 0.98, range 0.07 (0.93-1.0). C: mean=0.92, median=0.95, range=0.2 (0.80-1.0).

The results are also given by patient, group and total in table VIII, below.

Table VII Dice-coefficient results and True Positive scores between volumes in comparison

	A $GTV_{primary}$ *vs.* $EC_{primary}$		*B* $CTV_{primary}$ *vs.* $HD - GLIO_{primary}$		C $GTV_{recurrent}$ *vs.* $EC_{recurrent}$		Overall Dice	Overall True Positive
	Dice-score	True Positive	Dice-score	True Positive	Dice-score	True Positive		
Pt 1	0.70	0.94	0.67	~~0.64~~	--	--		
Pt 5	0.34	0.33	0.57	0.94	0.15	~~0.05~~		
Pt 7	0.54	0.80	0.17	0.98	0.79	1.00		
Pt 8	--	--	0.15	1.00	0.71	1.00		
Pt 9	0.87	0.92	0.60	0.98	0.67	0.80		
Pt 10	--	--	0.77	0.93	0.37	0.87		
Mean	**0.61**	**0.75**	**0.49**	**0.97**	**0.54**	**0.92**	**0.54**	**0.88**
median	**0.62**	**0.86**	**0.59**	**0.98**	**0.67**	**0.94**	**0.60**	**0.94**
SD	**0.23**	**0.29**	**0.26**	**0.03**	**0.33**	**0.10**	**0.24**	**0.18**
SEM	**0.11**	**0.14**	**0.11**	**0.01**	**0.13**	**0.05**	**0.06**	**0.05**
p-value	**0.12**	**0.12**	**0.03**	**0.06**	**0.06**	**0.12**	**<0.001**	**<0.001**

Overall values are calculated from all results presented in the table. SD=standard deviation, SEM=Standard Error of Mean. P value is calculated with Wilcoxon signed-rank test (two tailed, with a theoretical median=0). ~~—~~Excluded as an outlier by Grubbs test.

Because the Dice-coefficient is a statistical measure of similarity, the nature of the results does not exhibit a Gaussian distribution and must therefore be analyzed with non-parametric tests. The three datasets were scrutinized for variance by Friedman test, which is a one-way repeated measures analysis of variance by ranks. The Friedman test result was p>0.99 for difference, hence no significant difference between groups (p<0.001). This means the three groups can be analyzed as one dataset with 16 samples in terms of similarity between manual oncologist delineation and machine-derived segmentation. The n=16 Dice-coefficient results was then analyzed for statistical significance with the Wilcoxon signed-rank test. The test was made with a hypothetical median of 0, representing the theoretical expectation of the null-hypothesis. The discrepancy in theoretical and actual median was 0.60 (p <0.001). As the actual median of 0.60 is significantly larger than the hypothetical median of 0, the correlation between manual and machine-learning-derived segments are statistically significant.

The True Positive data is descriptive as to where the machine-derived segment is situated with relation to the ground truth. The Friedman test was also used on the ranked True Positive datasets, and was not found significantly different (p=0.58), meaning the True Positive results can also be viewed as one group.

Because the Dice-coefficient is normalized to the sum of the volumes in comparison, it inherently lies in the model that large indifferences between the volumes sizes will affect the Dice-coefficient score. The impact of this factor has been analyzed for the specific population in this study. In this respect, *size* here refers to the 3-dimentional volumetric size in cm^3. The "*Relation-In-Size*" factor was found by: $HD - GLO\ size/ground\ truth\ size$, which gave an output of a number where 1 is equal size, 0.5 means HD-GLIO is half the size of ground truth, 2 means HD-GLIO is double the size of ground truth, and 0 means one of the sizes were zero. Relevant values are plotted in table VIII below.

Table VIII Volumetric sizes and their relation in size

	Patient number	Manual size (cm3)	HD-GLIO size (cm3)	*Relation-in-size* factor	Dice-coefficient
A: $GTV_{primary}$ *vs* $EC_{primary}$	1	*6.5*	*4.0*	0.62	*0.70*
	5	*1.2*	0.6	0.50	0.34
	6	*8.1*	3.6	0.44	0.54
	8	*20.5*	0.0	0.00	0.00
	9	*45.5*	38.3	0.84	0.87
B: $CTV_{primary}$ *vs* $HD-GLIO_{primary}$	1	*206.4*	*222.0*	1.08	*0.67*
	5	*164.4*	70.5	0.43	0.57
	6	*224.3*	21.2	0.09	0.17
	8	*201.1*	16.6	0.08	0.14
	9	*300.6*	132.7	0.44	0.60
	10	*237.7*	163.5	0.69	0.77
C: $GTV_{recurrent}$ *vs* $EC_{recurrent}$	1	*18.5*	0.0	0.00	0.00
	5*	~~*2.5*~~	~~*13.9*~~	~~5.56~~	~~0.15~~
	6	*0.9*	0.7	0.78	0.79
	8	*3.5*	1.9	0.54	0.71
	9	*10.7*	8.0	0.75	0.67
	10	*42.6*	12.2	0.29	0.37
Mean		*93.3*	43.5		
Median		*31.6*	10.1		

*Volumetric sizes of manual- and machine learning-derived tumor volumes given in cm3, and their relation-in-size factor between volumes in comparison.*patient 5 was ruled out of the further analysis as an outlier by Grubbs test.*

In order to analyze how the relation-in-size factor is correlated with the Dice-coefficient results for our patients, the correlation coefficient was calculated. First, the data was searched for outliers using the Grubbs test and as a result patient 5 in dataset C (recurrent tumor volume comparison) was excluded as an outlier based on its relation-in-size factor. As the Dice-coefficient is non-parametric, but the relation-in-size factor is based on size which is

considered parametric exhibiting gaussian distribution, both Pearson (parametric) and Spearman (non-parametric) coefficient was calculated. The two correlation methods gave the same result: r=0.89 (p=0.001(two-tailed)). Both correlation coefficient values range between -1.0 and 1.0, where a correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation above 0.7 is considered strong, both positive and negative (107).

A simple linear regression was calculated to investigate the linear relationship of how strongly the Dice-coefficient result is dependent on the relation-in-size factor for our patients: R^2=0.80 (p<0.001). Also here, excluding patient 5 in dataset C (recurrent tumor volume comparison) as an outlier as earlier described. All plots and linear trendline and are shown in figure 48 below.

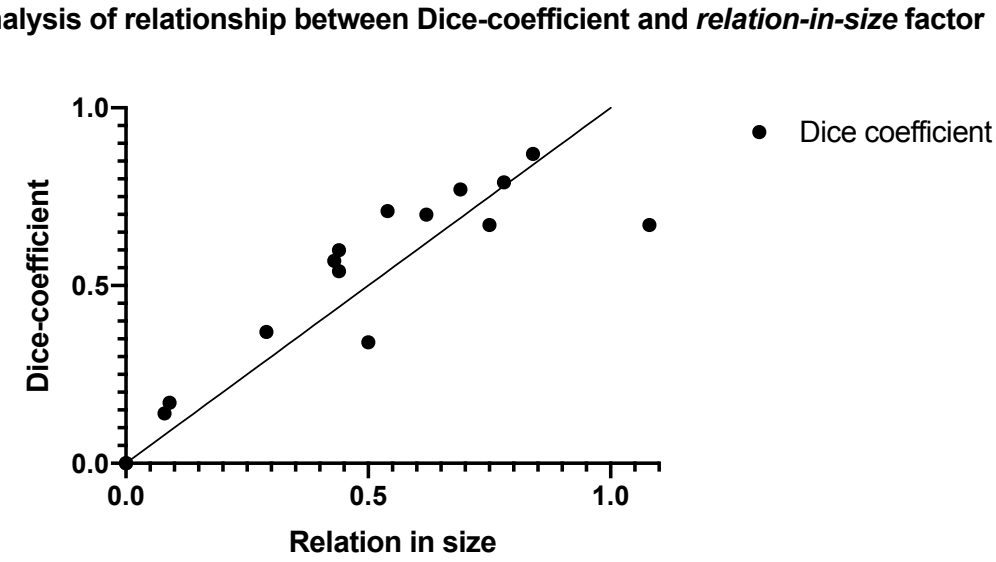


Figure 48 Graph showing the linear relationship between Dice-coefficient score (dependent variable) and relation-in-size factor (independent) between compared volumes; R^2=0.80 (p<0.001). Calculations were made excluding patient 5 in dataset C (recurrent tumor volume comparison) as an outlier.

For our patients, we therefore see a strong positive relationship between Dice-coefficient and relation in size between the compared segments. This means that large differences in volumetric size cause low Dice-coefficient scores. The probability of this observed difference having occurred by random chance (p-value) is 0.000003, calculated by t-test using t-distribution with n-2 degrees of freedom. The observed strong correlation findings between Dice-score and size-relation are therefore statistically significant.

5.3 Q2: Predictive value – primary inclusion of the recurrent tumor site

One aim of this study was to investigate if the machine learning model have longitudinal predictive value in terms of including the future tissue site of recurrence better than the standard clinical margin.

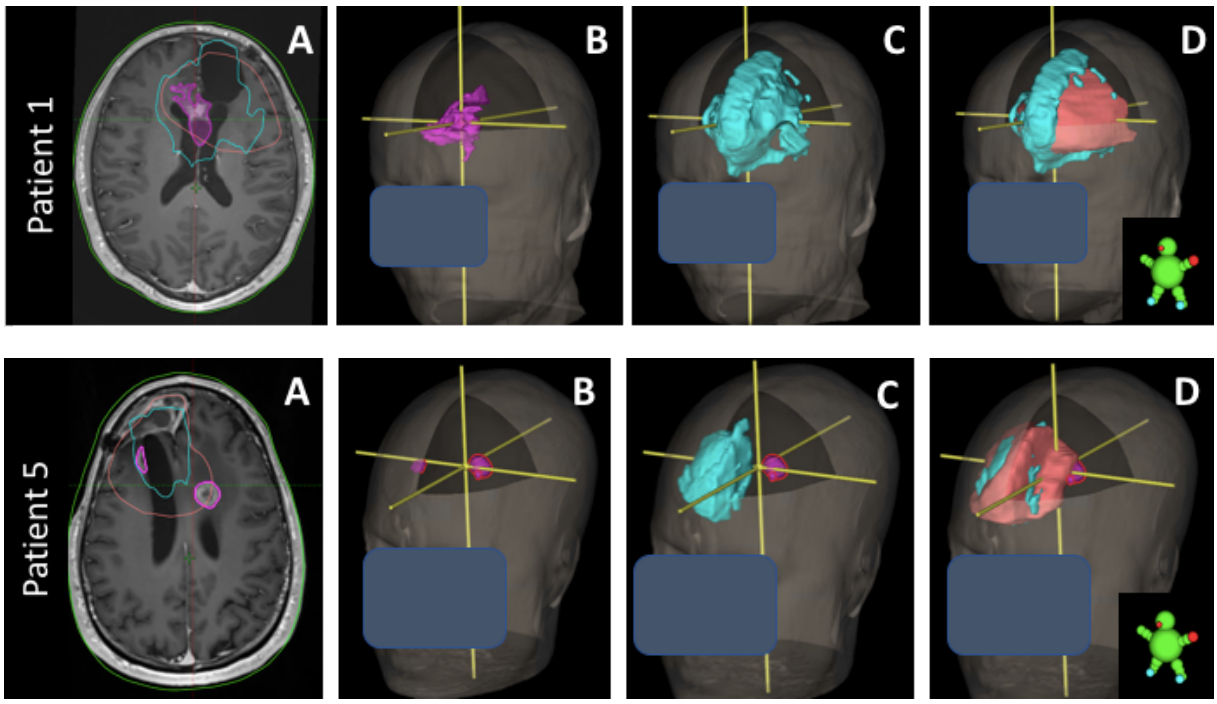
5.3.1 Q2 - Segmentation results

The volumes that are compared across $MRI_{primary}$ and $MRI_{recurrent}$ with Dice-coefficient are: $GTV_{recurrent}$ (magenta) versus

A. $CTV_{primary}$ (pink) , and

B. total $HD - GLIO_{primary}$ (cyan)

Illustration of the spatial localization of the recurrent tumor ($GTV_{recurrent}$) in relation to $CTV_{primary}$ and $HD - GLIO_{primary}$ for all patients is given in figure 49 below.



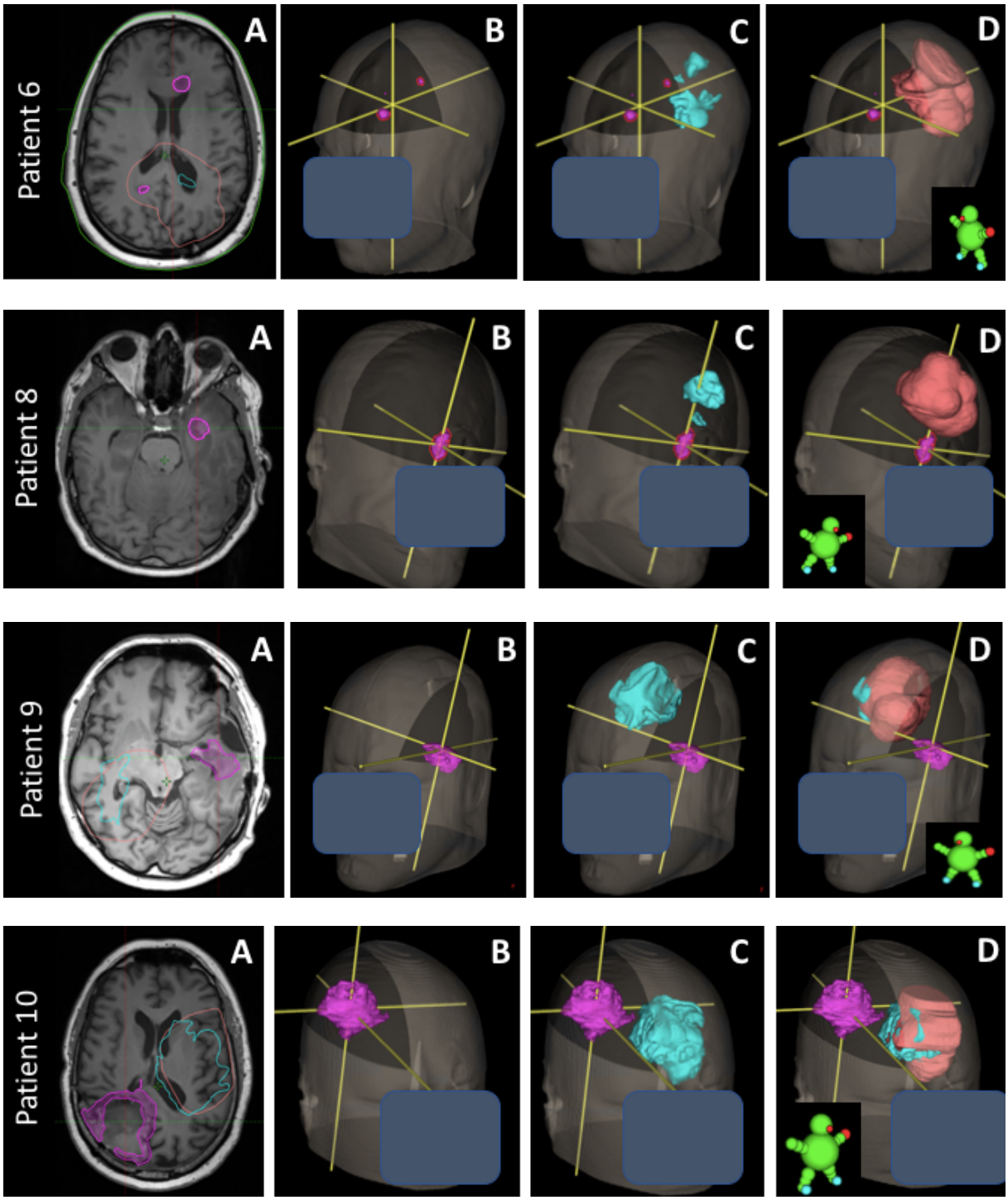


Figure 49 Site of recurrent tumor seen in relation to A: Axial view of recurrent tumor ($GTV_{recurrent}$) (magenta) as delineated on $MRI_{recurrent}$. The Clinical Target Volume ($CTV_{primary}$) (pink) and the total $HD - GLIO_{primary}$ (cyan) from $MRI_{primary}$ is not situated in the same level and is therefore not showing. B: 3D-renderd images of $GTV_{recurrent}$, adding (C): $HD - GLIO_{primary}$ and (D): $CTV_{primary}$, respectively. Illustration: Marianne H Hannisdal

5.3.2 Q2 - Predictive value: Similarity between $CTV_{primary}$ and $GTV_{recurrent}$

The manually delineated primary tumor site, $CTV_{primary}$ was compared to manually delineated localization of the recurrent tumor $GTV_{recurrent}$. The purpose of this was to quantify to what degree the CTV-margin in the primary radiotherapy treatment plan covered the site of recurrence. Overall, we could see that only patient 1 and 5 experienced recurrence within the boundaries of the CTV margin, patients 6,8,9,and 10 experienced recurrence in a completely different site. The similarity coefficient results are presented in table IX below:

Table IX Volumetric sizes and Dice-coefficient between $CTV_{primary}$ and $GTV_{recurrent}$

P	*$CTV_{primary}$ size*	*$GTV_{recurrent}$ size*	*Dice-coefficient*	*True positive*
1	206.4	18.0	0.13	0.79
5	164.4	2.5	0.02	0.01
6	224.3	0.9	0.00	0.00
8	225.7	3.5	0.00	0.00
9	300.6	10.7	0.00	0.00
10	237.7	42.6	0.00	0.00
mean	**226.52**	**13.03**	**0.03**	**0.13**
median	**225.00**	**7.10**	**0.00**	**0.00**

All volumetric sizes is given in cm3. The True positive is the volumetric share of GTVrecurrent situated inside CTVprimary

5.3.3 Q2 - Predictive value: Similarity between $HD - GLIO_{primary}$ and $GTV_{recurrent}$

The total machine-derived tumor volume of the primary tumor site $HD - GLIO_{primary}$ was compared to the ground truth manually delineated localization of the recurrent tumor $GTV_{recurrent}$. The purpose of this was to quantify to what degree the $HD - GLIO_{primary}$ included the site of recurrence. Overall, we could see that $HD - GLIO_{primary}$ only included the site of recurrence on patient 1.

The similarity coefficient results are presented in table X below:

Table X Volumetric sizes and Dice-coefficients between $HD - GLIO_{primary}$ and $GTV_{recurrent}$

P	$HD - GLIO_{primary}$	$GTV_{recurrent}$	*Dice-coefficient*	*True positive*
1	220.20	18.00	0.13	0.07
5	31.90	2.50	0.00	0.00
6	20.30	0.90	0.00	0.00
8	16.60	3.50	0.00	0.00
9	122.70	10.70	0.00	0.00
10	137.50	42.60	0.00	0.00
mean	**91.53**	**13.03**	**0.02**	**0.01**
Median	**77.30**	**7.10**	**0.00**	**0.00**

All volumetric sizes is given in cm3. True Positive is the share of $GTV_{recurrent}$ situated inside $HD - GLIO_{primary}$

5.3.4 Q2 - Overview and statistical analysis

Overall, we could see that both methods had poor longitudinal predictive value in terms of including the future tissue site of recurrence, as listed in table XI and shown in figure 55 below.

Table XI Dice-coefficient similarities and True Positive scores between volumes compared across longitudinal datasets

	A $CTV_{primary}$ *vs.* $GTV_{recurrent}$		*B* $HD-GLIO_{primary}$ *vs.* $GTV_{recurrent}$	
	Dice-score	True Positive	Dice-score	True Positive
Pt 1	0.13	0.79	0.13	0.07
Pt 5	0.02	0.01	0.00	0.00
Pt 7	0.00	0.00	0.00	0.00
Pt 8	0.00	0.00	0.00	0.00
Pt 9	0.00	0.00	0.00	0.00
Pt 10	0.00	0.00	0.00	0.00
Mean	**0.03**	**0.13**	**0.02**	**0.01**
median	**0.00**	**0.00**	**0.00**	**0.00**
SD	**0.05**	**0.32**	**0.05**	**0.03**
SEM	**0.02**	**0.13**	**0.02**	**0.01**
p-value	**0.50**		**0.99**	
discrepancy	**0.00**		**0.00**	

SD=standard deviation, SEM=Standard Error of Mean. P value is calculated with Wilcoxon signed-rank test (two tailed, with a theoretical median=0). Discrepancy is the calculated difference between the calculated median and the theoretical median in the hypothetical case of no difference.

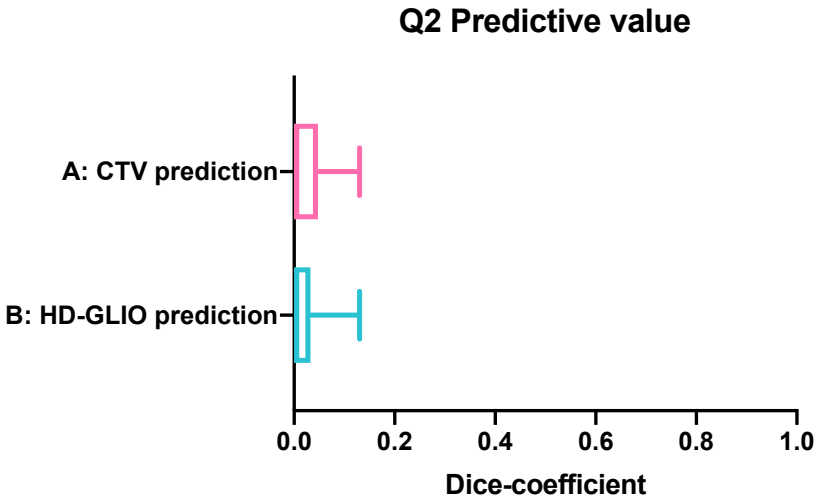


Figure 50 Box-plot of Dice-coefficient results comparing A: the CTVprimary to GTV recurrent, and B: HD-GLIOprimary to GTVrecurrent

The datasets were compared with a non-parametric Wilcoxon matched-paired signed rank test. The two sets of Dice-coefficient scores did not differ significantly, but these data are not statistically significant ($p>0.99$) so no further analysis was executed.

6 Discussion

In this small-scale study, the potential for using deep learning artificial intelligence in RT tumor volume determination was investigated. This was performed by comparing the manually delineated tumor volume and machine learning derived tumor volume in terms of similarity and predictive value for tumor recurrence. The applied machine learning method makes use of established quantitative radiological biomarkers from multispectral MRI.

Radiation therapy planning is personalized, aiming to administer a target specific as possible treatment , based on the best image material and analysis techniques available. Cost-efficiency, radiological geometric fidelity, and the balance between tumor control and inflicting damage to OAR remains the main challenges. Also, imaging and visualization conditions in radiotherapy fails to match than in diagnostic radiology. Current clinical practice in determining the target volume for glioblastoma radiotherapy is to use 20 mm isotropic margin from the radiologically visible pathology. Previous research suggests increased focus on preserving quality of life of the remaining life time. This implies that smaller margins without cost of therapeutic ratio with specified imaging and optimal image analyzing techniques are desirable (1-3, 98, 100). New national guidelines under consideration opens for margin reduction in radiotherapy of diffuse high-grade gliomas (19). The overall aim of this project was therefore to study the potential of machine learning to further improve radiation therapy. This, regarding (i) reduced dose to normal tissue and OAR and (ii) the narrow therapeutic window of GBM in radiation therapy.

6.1 Result discussion – correlation between HD-GLIO segmentation and manual delineations

The potential advantage of utilizing machine learning applied to target delineation lies in the elimination of inter-oncologist variation and an improvement in specificity, and hence a reduction of the amount of unnecessary normal tissue being included in the target area.

Dice-coefficient comparison between the oncologist`s manually delineated tumor volume and the machine learning derived tumor volume showed an overall mean concordance of 0.54 (median=0.60). The Dice-coefficient is a similarity score, so when speaking of 54% similarity one can imagine a scenario like illustrated in figure 51 below.

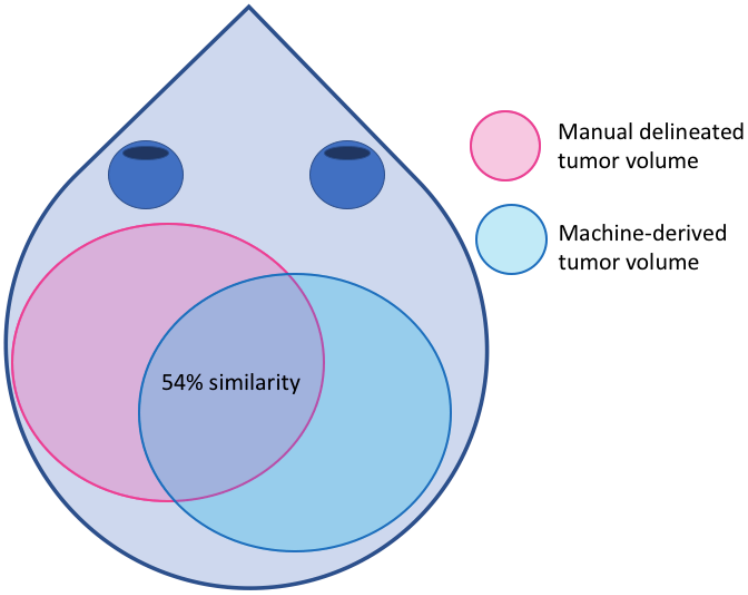


Figure 51 Example of a 0,54 Dice-coefficient similarity score situation when volumes are equally sized.

When we are comparing three manually delineated volumes to their corresponding machine-learning volumes, we found that the overall mean sizes of manual volumes were twice the size of machine learning-derived volumes, with an overall True Positive score of 0.88 (median=0.94). A general finding in this study is therefore a situation as illustrated in figure 52 below.

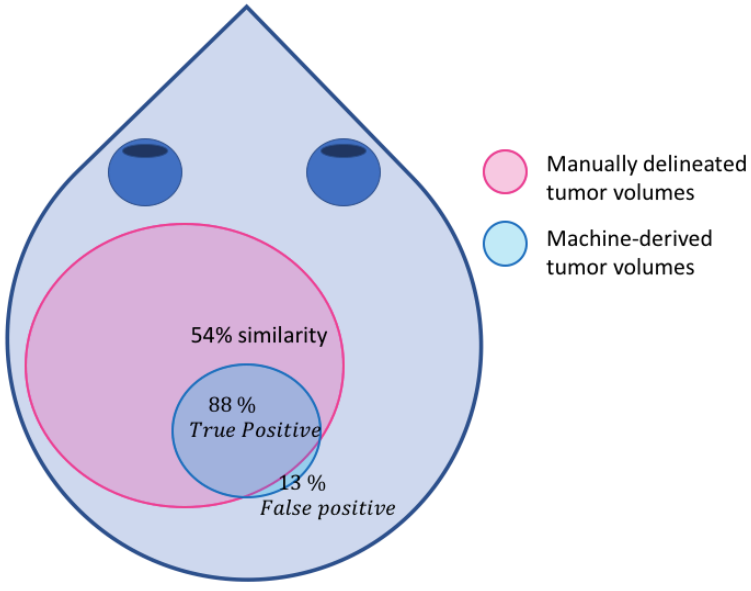


Figure 52 The realistic situation of the 0.54 Dice-coefficient similarity score for the data in this study, visualized with authentically congruent size relations. The overall results showed a mean of 88% True Positive, and 13%.False Positive.

So the Dice-coefficient related to research question 1 (Q1) of mean 0.54 (median=0.60) does not represent a scenario of 46% voxel occurrences for which the HD-GLIO detects malignant tissue in a different site than $CTV_{primary}$, but rather a scenario where HD-GLIO detects malignant tissue situated inside the CTV-margin, but in a smaller volumetric extent. This is consistent with our expectations as the very motivation for this study is the search for a tool that can discriminate true malignant tissue more precisely from normal tissue, and hence provide higher specificity. This is also interesting in light of new guidelines under consideration from the Norwegian Directory of Health, in which avoiding large margins are emphasized (19).

Even if the Dice-coefficient is the most frequently used measure of segmentation similarity or degree of overlap, it is not necessarily the overall best measure of volume correspondence. Other factors such as volumetric sizes and the *True Positive* overlapping part contribute to the descriptive value. If we had good reasons to be sensitive to local segmentation mismatch spatially characterized by "islands", "peninsulas" or "lagoons", the Hausdorff Distance (HD) metric could have been included for this study. However, HD will likely have given a disproportionally poor result on *e.g.* patient-one and patient-eight, who both had islands segmented by HD-GLIO. The results showed a strong positive relationship between Dice-coefficient and the calculated relation-in-volume factor for the patients in this study, meaning that the Dice-coefficient result is strongly affected by incongruences in size between the compared volumes. It is, however, important to keep in mind that this project did not presume the machine-derived segment to have the same morphologic extent as the manual volume. For dataset B (comparing $CTV_{primary}$ to $HD-GLIO_{primary}$ including cavities), this is especially important regarding the inherent disparity of the compared volumes, in terms of low specificity in the $CTV_{primary}$ volume, which is serving as a margin. The use of the CTV-margin as ground truth in this context will be further discussed in the Method Discussion *Validity and Reliability* section below.

For dataset A (comparing $GTV_{primary}$ versus $EC_{primary}$), the premises between comparisons are more alike, as they are both based on the gadolinium enhancing tumor tissue, which one could speculate is more straight forward in both qualitative and quantitative interpretation methods. Therefore it is interesting that the mean volumetric sizes differ between $GTV_{primary}$ and $EC_{primary}$ as much as they do; mean 16.36 cm3 (median 8.1) and 9.3 cm3 (median=3.6), respectively, so manual delineation includes a bigger volume than HD-GLIO. However, the True positive rate of mean=0.92 (median=0.94) means that $EC_{primary}$ is mainly

situated inside $GTV_{primary}$. There was not sufficient statistical power to ensure statistical significance in dataset A (p=0.12), so the trend from this dataset must be interperated in context with dataset B and C. The sparseness of the data will be further discussed in the method discussion section.

As for correlation between GTV and HD-GLIO in MRI-recurrent (dataset C), the volumetric trend is also that $EC_{recurrent}$ is smaller than $GTV_{recurrent}$. The mean Dice-coefficient is 53% (median=67%) but the mean True Positive result is 91% (median=93%). This indicates that even though $EC_{recurrent}$ does not detect as much malignant tissue as the manual delineation $GTV_{recurrent}$, the volume still had a tendency to be situated in the same site (p=0.06). These data were also not statistically powered when analyzed alone.

When analyzing Dice-coefficient results, it should be acknowledged that a low Dice-coefficient is not necessarily a negative indication. The CTV-margin contains normal tissue, which is not expected to be segmented as malignant tissue, on the contrary. Several margin-reducing studies have shown good results (1, 3, 19, 97-99) indicating that a better method is needed to specify the radiotherapy target and avoid inducing unnecessary neurocognitive side effects for the patient. An ideal study would perhaps be to compare two groups, where one group received treatment based on traditional manual delineation, and the other group received treatment based on segmentations made with HD-GLIO. This might enable the study of both treatment outcomes and side-effects between groups. However, due to limited knowledge at this time point, it would not be ethically responsible to pursue such a study, but with advanced knowledge through further studies and continued development of dynamic algorithms, a study of this character could be a goal to aim towards. Again, a specification of the radiotherapy target is substantiated by the new guidelines from the Norwegian Directorate of Health currently under consideration, emphasizes the limiting of radiation induced neurotoxic damage, hence avoiding unnecessary large margins (19).

6.2 Result discussion – longitudinal predictive value

When it comes to predictive value, neither $CTV_{primary}$ nor $HD - GLIO_{primary}$ had significant accuracy for inclusion of the site of the recurrent tumor. Again, large size differences was partially the cause of these low Dice-coefficient results, but there is no indication that $HD - GLIO_{primary}$ exhibits a predictive value higher than the inherent prediction within the large margin of CTV confers. However, the large CTV-margin did not

show significant predictive value either, which in one respect could mean that the radiotherapy treatment had worked well, terminating pre-morphologic tumor cells within the target area and preventing resurgent growth from occurring in this area. On the other hand, in the hypothetical case that the site of recurrence in any case would be situated distant from the primary site, the radiation of this large in-margin normal tissue area would be solitarily unbeneficial for the patients, only causing complications to the surrounding normal tissue. These data, however, did not have enough statistical power to be statistically significant ($p>0.99$).

It was surprising to see that the recurrent tumors occurred in sites so distant from the primary tumor site, which is not in compliance to what we expected based on previous research (11, 12, 61). As previously mentioned, tissue localization shift in the rigid co-registration process could to some degree present a methodological uncertainty, however, manual inspection revealed that this does not entirely account for the dispersed spread of tumor sites. The reason for unexpected site for recurrence in this study was not thoroughly elucidated but one explanation might be due to the highly selected patient group with specific characteristics determined by BORTEM-17 inclusion criteria. The model of the study was built on a preconception and the expectation that the recurrent tumor would come within the nearest 2-3 centimeters of the primary tumor site. This preconception founded on previous studies that established this to be true for 85% of all cases in glioblastoma with unmethylated *MGMT* promoter (101), and in roughly 60% of all glioblastomas in total (108). Also, the daily practice with making large-margin treatment plans is based on the consensus understanding that recurrence within the 2-3cm margin is the common scenario. Therefore, it was quite surprising to find that this model only fitted patient 1, whereas 79 % of the recurrent tumor site turned out to be situated inside the CTV-margin, giving a Dice similarity coefficient of 13 %. For patient 5 only 1 % of the recurrent tumor was situated inside the CTV-margin, giving a Dice similarity coefficient of 2 %, and as for patient 6,8,9 and 10 the recurrent tumor site was in a completely different lobe or hemisphere. As patients with unmethylated *MGMT* promoter respond poorerly to both radiotherapy and temozolomide (61, 64), this could give credence for the expection that recurrent tumors occur more rapidly than for patients with methylated *MGMT* promoter. Moreover, one of the inclusion criteria was that the patients had to be to free of recurrence at least 12 weeks after primary radiotherapy, a criterium that was set to be able to isolate the effect of bortezomib pretreatment in the BORTEM-17 study. However, this selection criteria could potentially have excluded the patients with the most aggressive disease progression.

Four patients were excluded in this study due to missing image material and the 6 patients in my independent study had a mean disease-free period of 76 weeks (median 59 weeks), including an outlier that had 185 disease-free weeks (3.5 years). Stupp et al., (2005) reports that median disease-free period for glioblastoma is 6.9 months (8), equivalent to approximately 30 weeks. Hence, the patients in this study had a disease-free period almost double of the reported. A study of recurrence patterns reported that only 10% of recurrent tumors occurred distant[7] to the primary tumor site (109), whereas 66% (4/6) of the patients in this study had sole distant recurrence. Also, it has been reported that recurrences situated outside the radiotherapy target volume occurs significantly slower than recurrences situated inside the radiotherapy target volume (101). This can indicate that the 12 weeks recurrence-free inclusion criteria could possibly have excluded patients with rapid recurrence, having a larger statistical chance of in-target volume regrowth. This could mean that the recurrence sites in this study might not be representative of the general population with recurrent glioblastoma. Also, the low number of patients makes it difficult to make general conclusions of whether this is a characteristic of this selected patient group or simply a coincidence. Nevertheless, the potential signal intensity changes due to underlying pathology were either not detectable by the applied method or they simply were not apparent. Hence, a larger, unselected patient group is required in order to isolate the ability of using HD-GLIO to predict site of future recurrence, or indeed , if there is an effect to be detected.

6.3 Method discussion

Much research has been done on radiological biomarkers in recent years, especially focusing on what type of information they can provide within the various MRI-sequences, and how they can help to decipher the true radiologic expression of tumor pathophysiology. However, the studies often include high-complexity and time-consuming MRI sequences, hence cost-efficiency evaluation of the resource often overweighs towards cost, compromising the clinical applicability. Rathore, Akbari et al., (2018) proposed the use of already standardized image sequences which are available for all scanners, and thereby multispectral MRI-examinations on all patients could be eligible for machine learning image analysis. In developing machine learning algorithms that make use of standardized protocols *e.g.* the BrATS Challenges and HD-GLIO, unnecessary complexity is eliminated, along with additional costs and time use in the MR-examination process. At a practical level, established

[7] Situated minimum 3 cm outside enhancing part of primary tumor, compliant to 3 cm outside GTV

timeframes for examination protocols and radiologic interpretation should be taken into account. Even if novel image acquisitions and high-dimensional functional uptakes is of utmost interest, it is not always available and clinically eligible. However, clinical protocols for brain tumor imaging should be updated to fit the minimum of established standards, including pre- and postcontrast T1, T2 and T2FLAIR. In this study 40% of the available patients had to be excluded due to missing imaging series.

The scientific ideal in hypothetical deductive research tradition is that truth is achieved by logical deduction, and that knowledge can be derived from objective observations and experiments. When comparing objective, empirical sizes like tumor volumes, differences between variables being compared should be brought to a minimum in the attempt of isolating an effect, hence rule out biases. In this respect, one can say that comparing a segment derived purely from image biomarkers on a single MRI-examination, to a manual volume delineation based on human interpretation of an image spectrum including both CT as well as various full- and fractional MRI-examinations - in addition to qualitative clinical assessments, this does not fulfill an aim of similar comparison terms. As a consequence, effects could be harder to isolate. Still, the essence of this study is to identify a potential effect of a quantitative tumor delineation approach, using artificial intelligence. Therefore, it is interesting to evaluate current clinical practice based on general tumor infiltration biology experienced by GBM patients as a group, up against novel methods proposed in research literature, segmenting the tumor infiltration site based on patient-specific imaging parameters and quantitative image-biomarker exploration.

When comparing volumetric sizes, the use of the Dice-coefficient is a standardized approach (81), which ensures good internal validity. The clinical target volume is by nature a margin in which the oncologists are well aware that some normal tissue is included. It is still a calculated risk, based on empirical studies stating that most recurrent tumor sites occur within this margin on group level. In this respect, there lies a prediction within this margin as to where the recurrence is expected to develop. However, when defining this the ground truth and comparing with Dice-coefficient, the whole CTV is defined “true positive” when we know it is not. The best measure, one could argue, might be to compare HD-GLIO segments to multiple radiologist tumor delineations, not including any treatment margin. This would give a better picture of the true correlation between manual interpretations of tumor extensiveness, to the machine-derived segmentation. Alternatively, in a hypothetical situation; performing a full slice-by-slice histological tumor boundary analysis directly after imaging,

defining the histological ground truth. Nevertheless, all methods have limitations, and a conventional histological analysis is another qualitative analysis, just as a manually delineated image-based volume is. Even the training material upon which HD-GLIO has gained its robustness and strength, is also delineated manually, and even if delineations are performed by experts one can never completely rule out observer-dependent methodological bias. Taking into account that all method has inherent methodological insecurities, any ground truth has its limitations. So, when measuring one method`s sensitivity and specificity up against a gold standard, there will always be a risk that pathological sites detected only by a novel method will falsely be interpreted as false-positive. Still - the use of current gold standard clinical practice as is common practice in method studies ensures we build new knowledge on top of pre-stated scientific truths.

The volume of recurrent tumor was delineated by me. I have five years of experience delineating cerebral structures for radiotherapy planning purposes. All delineations were approved by an experienced oncologist. However, increased validity and reliability may have been achieved if an experienced radiologist made the delineation. As machine-learning segmentation holds the potential to eliminate inter-observer variability, it would also have been useful to compare delineations from multiple radiologists, but this was not performed due to time and resource constrains. Recurrent tumor delineations were made solitarily on contrast enhancing tissue on T1 post contrast images, both because contrast enhancement is a well-established hallmark of cancer (44), but also because T2-signal can be affected by post-irradiation changes. All patients had received radiotherapy for their primary GBM, this makes it more complicated to rely on T2 or T2FLAIR images to quantify any tumor progression, as radiation treatment-related toxicity represent a common attribue to hyperintense signal in these MRI-series (110). For this group, hyperintense T2/FLAIR-signal cannot be solitarily related to tumor infiltration or peritumoral edema as a marker of neovascularization or vascular endothelial growth factor (VEGF) expression, as when prior radiotherapy is not an issue (48), even if GBM harboring unmethylated *MGMT* promoter might infrequently induce pseudoprogression compared to GBM bearing methylated *MGMT* promoter (64). The visual contrast-enhancing tumor part is therefore used as a surrogate to the full tumor burden, and therefore also only compared to the “Enhancing Core” segment ‘in HD-GLIO.

6.3.1 Validity

The limited statistical power due to low number of participants (n=6) is a major limitation of this study in terms of attaining sufficient statistical power to ensure low probability of making

a Type II error. Statistical power states the likelihood of detecting an effect when there is an effect to be detected (72). Conducting a Type II error is to conclude there is no effect when there in fact is one, hence, not rejecting a false null hypothesis. Conducting a Type I-error is to conclude one has detected an effect when there in fact is no effect to be detected, hence, rejecting a true null-hypothesis. A small sample size affects the ability to make generalizing conclusions, transferring the result upon the population in general, limiting the external validity of the study. However, I had multiple measurements on the same patients, so the effective sample size increased, which ultimatly increased the external validity. Furthermore, when working with limited clinical image material and the limited time-frame of the master study program, the need for pragmatism and somewhat settle for what is feasible and available comes into play. The important thing is still to perform the analyses and critical interpretation task in respect of the actual sample size, using suitable methods, and not the least - dedicate meaning to the results in accordance to their actual power. Another factor is that the trends from small scale studies can break grounds for larger studies, and small-scale methodological studies can act as feasibility testing and method-mining purposes.

When using statistical methods, any scientist should emphasize that statistically significant findings is not necessarily the same as a scientifically interesting findings. Statistical significance assure that relations observed in a sample are not simply due to chance, but the validity of such depends on correct use. The Wilcoxon test used in this study will inherently define very small discrepancies between theoretical median (representing the case of the null-hypothesis) and actual median as statistically significant. Nevertheless, statistical results and their importance weighting should always be evaluated with respect to the nature of the data material and in context with the nature of the research question.

The study population is heterogenous in terms of age but not in respect to gender, as 83% (n=5) of patients are men and 17% (n=1) women. To my knowledge, there are no gender specific radiological features and hence, there are no reason to suspect that lack of gender balance affects the internal or external validity of the study .

The study population is homogenous in relation to *MGMT* promoter methylation, however, there has not been reported that unmethylation *MGMT* promoter influences image interpretation in terms or radiological features. Still, the patients included in this project are highly selected for certain characteristics in order to be eligible for inclusion in the BORTEM-17 study and these specific biological requirements may affect the results in my study. As previously stated, studies show that patients with unmethylated *MGMT* promoter

experience less pseudoprogression, presumably because pseudoprogression is a phenomenon related to radiosensitivity, indicating response to treatment (64). Furthermore, as GBM with unmethylated *MGMT* promoter is a particularly treatment-resistant, more infiltrative malignant tissue could perhaps be expected, compared to both methylated *MGMT*-promoter glioblastomas, as well as lower grade gliomas. For image analysis, more tumor cells and less signal-contamination from pseudoprogression could serve as a beneficial platform for both visual interpretation in manual delineation, as well as for machine-learning segmentation. If that is the case, this patient group could be especially suitable for detecting an effect if there is one, which could increase the internal validity of the study. Furthermore, when comparing treatment safety margins versus machine learning-derived segments for this group of *MGMT* unmethylated glioblastoma patients, there could be several biases. There is a possibility that since this group is the most treatment-resistant, this group have a definite need of large margins, meaning the calculated normal tissue complication probability and tumor control probability with 20 mm margins is the correct choice for this specific group. In that case, the actual group of glioblastoma patients that has the most spared brain-tissue to gain from reduced margins, could in fact be the patient group with methylated *MGMT*-promoter or even lower grade gliomas. If that is the case, this unmethylated group will have a smaller effect to be detected, possibly influencing the internal validity. The degree of influence by these variables can potentially be controlled by executing the same study on the methylated patient group as a control.

When using machine learning algorithms, it is important that the image material upon which the algorithm was trained, is representative for the heterogeneity in glioblastoma patients in respect of possible single- or multiple surgeries, radiotherapy treatment, as well as medications. Treatment naïve patients will have a different radiological signature than treated patients in terms of tumor cavitation, pseudo progression, and altered gadolinium leakage features across the blood-brain-barrier due to anti-angiogenic and steroidal medications, amongst many other things. So - as intelligent artificial intelligence can be -, it is still only as intelligent or robust as it has been given the chance to be by its training master, meaning there lays a definite limitation for any trained machine learning algorithm within the training material. However, the HD-GLIO algorithm used in this study has been trained upon multi-institutional MRI-examinations including both treated and treatment naïve patients with ground truth reference, including glioblastoma of both methylated and *MGMT*-unmethylated type, which adds to both reliability and validity. It is important in choosing type of algorithm that the nature of the algorithm needs to fit the nature of the data. Supervised learning is

beneficial to use when the features one wishes to extract are well-defined, but a bias that never could be ruled out in supervised learning is label-contamination; when a voxel is labeled wrongly by human misinterpretation or lack of human distinguishing abilities, and is thereby trained into further misinterpretation. Such potential label contamination-"noise" is made less prudent by using a large training material, in the same way as image noise is made less prudent by increasing the signal in MRI, thereby increasing SNR.

The HD-GLIO was found to be the most eligible algorithm for this study based on a thorough evaluation and elimination process. Other algorithms of both supervised and unsupervised character were considered. HD-GLIO was found to be superior to random forest (RF) since (i) RF would depend on de novo local labeling, (ii) HD-GLIO CNN is contextual and can evaluate relations between multiple features at the same time, and (iii) RF has multiple hyperparameters that could be hard to adjust. HD-GLIO, being a supervised classification method, was also found superior to K-means clustering, which is unsupervised (i.e cluster interpretation is subjective) and not robust to variations in voxel expression across image acquisitions and MRI-machines (106). K-nearest neighbour (KNN) classification was ruled out early since (i) it would depend on de novo local labeling, (ii) speed will be slow with large training set and feature dimension, (iii) the value of K (number of neighbors) is a hyperparameter that is challenging to estimate, (iv) does not perform well on imbalanced data (e.g. fraction of tumor-specific voxels are disproportionally small), (v) is sensitive to outliers, and (vi) not contextual as CNN, ignoring spatial relationships (cf. (84) for a more detailed discussion). In general, deep learning methods have shown to outperform traditional machine learning approaches to MRI analysis and pixel/voxel classification tasks (74) both regarding prediction accuracy and time consumption (in our case 6 min per MRI exam including preprocessing for which HD-GLIO had a run time of only 20 sec). The fact that HD-GLIO is developed in a state-of-the-art research environment at Heidelberg University Hospital and DKFZ by prominent computer scientists and a panel of experienced radiologists dealing with 3220 MRI examinations from 1450 brain tumor patients, leading to high-ranked publications (e.g. Lancet Oncology) of their algorithm and its performance, also adds credibility our methodological choice.

6.3.2 Reliability

Reliability refers to the accuracy and consistency of measuring variables (105). In this respect, it is relevant to consider if the image material used as base for the volumes in comparison is consistent. The image material used for clinical manual delineation of

CTVprimary by the oncologist was the sum of pre-and postoperative MRI-examinations in addition to CT. What image series most heavily weighted by the oncologist are not specified in any journal records to my findings. The image material used for analysis with HD-GLIO could be somewhat inconsistent as preoperative MRI does not rule out prior surgery. It was not possible to keep this variable constant across the included patients as some have had single surgical intervention, and some have had several surgical interventions. Also factors inflicting image quality, like image matrixes, was not constant across all MRI-examinations, as they were acquired at different sites and in different scanners (complete list is enclosed in appendix B). In sum, all these heterogenic factors may very well induce bias as to poor reliability, and potentially induce poor reproducibility. This means that it is possible that some unknown or random factors in the image material may affect the measurements, influencing the degree of effect measured in the study.

Generally, image quality may interfere with radiological interpretation and the outcome of machine learning. As for image quality, there are patient-related aspects regarding patient tolerability that could affect the quality, as well as acquisition dependent quality aspects like partial volume effect (PVE), SNR and CNR. Image series with suboptimal image acquisition have been attempted to be ruled out by visual inspection of the images, but still remains as a theoretical methodological insecurity factor. Variability between image acquisitions of the patients in the study is present in the form of different scanner types, different hospital sites, different field strengths etc. Nevertheless, the image-data on all six patients fulfilled HD-GLIO minimum requirements, which increases the robustness and applicability of the model.

The use of “cavity filled” volumes as described in methods section 4.3.3,9d) was performed in order to fit the data to the model, hence make it easier to isolate a potential effect. A central necrotic or cystic lesion is a diagnostic radiological marker of GBM, however, it is not consistent of malignant tissue but a secondary effect due to the strangulating and chaotic in-tumor dysfunctional vascularization structure. Since the model was made to compare overlap in machine learning-detected malignant tissue, which inherently will exclude a cystic or necrotic core - to a CTV containing both malignant tissue, surgical cavities and necrotic cores, there was a premises-shift between the two methods. If these cavities were not filled, if would induce a bias in the form of low Dice-coefficients and low True Positive-rates, which one could argue would not be representative for the actual segment volume overlap comparison. In literature, segment cavities are also described as one of four basic types of errors in

segmentation metrics (added regions, added background, inside holes and border holes) (111). This data-fitting was therefore performed with the intention of increasing the reliability.

In terms of reliability it is also of the essence to investigate if the image material contains other abnormalities than the glioblastoma-malignant tissue. Pathological co-conditions like strokes, degeneration, demyelization or other malignant conditions that is not directly correlated to the glioblastoma can potential present a variable inconsistency-bias if interpreted as glioblastoma-related. Of the total 6 patients, only one patient had a known additional condition; a cavernoma situated in his brainstem. However, this was discussed with the oncologist, and as the location of the cavernoma was in a definite different place than the glioblastoma, this bias could therefore easily be excluded by visually monitor that the machine learning segment did not include the cavernoma.

Reproducibility is an important aspect of reliability. The HD-GLIO algorithm is easily accessible, and therefore HD-GLIO -outputs are easily testable. Small difference can occur in outputs using different hardware to run the algorithm, this factor is expected to be minimal, but could have been quantified by running the algorithm on multiple computers for output comparison. It should be mentioned that HD-GLIO and deep learning algorithms in general, are not very transparent, meaning there is hard to know which features in the data material was most emphasized.

6.4 Ethical concerns – use of AI in medicine

Patient-related ethical concerns have been discussed in the method section.

In true hypothetical deductive research tradition, current standards should sometimes be set to the test, as it is the only way progress can take place. Still, ethical concerns regarding novel techniques in medical science is always important to take into consideration. When comparing human competence versus artificial intelligence in general, ethical concerns relates to the value of human perception, emotion and cognition, sensitivity to patient's personal wishes, and the high-dimensional "whole clinical picture" (qualitatively and quantitatively speaking) decision-making process. In our context, we must also take into consideration that the analysis and decision-making based on multi-spectral image data can be very complex and often a challenging task for any trained specialist.

Artificial intelligence, machine learning, and computer aided detection has during the recent years showed promising results as a decision support tool for radiologists in MRI, CT and more. The evolving of these technologies has also raised some important questions that is worth giving reflection. Some may argue that if humans are proven to be outperformed by artificial intelligence in more and more ways, the human side can be put in a degraded position where it no longer trust its own performance, thinking that AI could always do a better job. If artificial intelligence is taken into more use, there is also a risk that the analyzing process can be increasingly non-transparent, and as an extreme scenario: even that radiologist get less training within certain areas and thereby loose experience-based radiologic interpretation skills. On the other side, there has been an exponential increase in medical examinations latter years, combined with high resolution and advanced imaging techniques, producing a high-volume multi-dimensional data to be analyzed. Artificial intelligence could help make image interpretation more efficient by acting as a decision support tool, optimizing the utilization of radiologists and thereby both reducing costs and maximizing the capacity of image interpretation. If we were to embrace this new technology as a prolonged part of our own intelligence, it has the potential to act as a useful tool for us. To make best use of such a tool we should always make sure we understand the algorithms methodology, by carefully ensuring we stay up to date, seeking to encounter the algorithms statistical and logical approach towards solving human challenges. The joint statement from ISRRT and EFRS addresses that radiographers must take an active role in planning, development, implementation, use and validations of AI applications:

> *The optimal integration of AI into medical safety, clinical imaging and radiation therapy can only be achieved through appropriate education of the current and future workforce and the active engagement of radiographers and radiologic technologists in AI advancements going forwards* (18).

Therefore, artificial intelligence should not be seen as a threat to human intelligence, but as a product of such. AI can be extremely useful when used with the right training, in the right setting, and according to high professional standards and ethical principles.

6.5 Conclusion

For the patients in this study, there were a significant overlap in the detected spatial extent of malignant tissue using HD-GLIO machine learning algorithm based on multispectral MRI, compared to manual delineation. This with respect to tumor volume and location similarity

measured with the Dice-coefficient ($p<0.001$). Our $Q1\ H_1$ is thereby confirmed and $Q1\ H_0$ is rejected. Hence, HD-GLIO can potentially be feasible as an oncologist support tool for specific segmentation of enhancing core and non-enhancing tumor, and should therefore be tested more excessively on a larger group of patients.

However, when comparing the two methods ability to include the site of tumor recurrence, neither the clinical target volume nor the machine-derived tumor volume showed significant predictive value. There were no significant discrepancy in Dice-coefficient between recurrent tumor and machine-derived segment, compared to Dice-coefficient between recurrent tumor and manual target volume. Hence, $Q2\ H_0$ cannot be rejected, and $Q2\ H_1$ cannot be confirmed in this study. We suggest further studies, using the HD-GLIO tools (or similar) to a larger sample of patients and a broader spectrum of glioblastoma tumors.

6.5.1 Contribution of knowledge

This thesis contributes to the ongoing trend of using multispectral recordings and machine learning as a method to explore the full diagnostic potential of MR imaging technology. The purpose of this study has been improving specificity in target delineation for radiation therapy, in compliance to new guidelines under consideration from the Norwegian Directory of Heath (19). The study has been performed in compliance with the joint statement from ISRRT and EFRS saying that radiographers in radiation therapy should take advantage of AI technology in order to improve personalized planning, including tumor segmentation (18).

The project, being a small-scale exploration comparing the deep learning U-Net based algorithm HD-GLIO to manual contours performed by experienced clinicians, has contributed with some initial steps towards engineering an updated toolbox for oncologists. We suggest that this automated deep-learning segmentation approach should be further investigated when deciding patient-specific radiation therapy target volumes based on quantitative image markers, emphasizing the ALARA-principle.

This study has not contributed with knowledge of the potential predictive value of using HD-GLIO in respect of tumor recurrence site, other than indicating that this method design might not be fully eligible for this purpose. We suggest that this research question should be emphasized in another study in order to answer this research question.

7 Sources and literature

1. Buglione M, Pedretti S, Poliani PL, Liserre R, Gipponi S, Spena G, et al. Pattern of relapse of glioblastoma multiforme treated with radical radio-chemotherapy: Could a margin reduction be proposed? J Neurooncol. 2016;128(2):303-12.
2. Jansen EPM, Dewit LGH, van Herk M, Bartelink H. Target volumes in radiotherapy for high-grade malignant glioma of the brain. Radiother Oncol. 2000;56(2):151-6.
3. Gebhardt BJ, Dobelbower MC, Ennis WH, Bag AK, Markert JM, Fiveash JB. Patterns of failure for glioblastoma multiforme following limited-margin radiation and concurrent temozolomide. Radiat Oncol. 2014;9(1):130-.
4. Klepp O. Glioblastom. I store medisinske leksikon 2018 [Available from: https://sml.snl.no/glioblastom.
5. American Brain Tumor Association. Glioblastoma (GBM) 2020 [Available from: https://www.abta.org/tumor_types/glioblastoma-gbm/.
6. McCullough BJ, Henson JW. Chapter 54 - Neuroimaging Issues in Assessing Response to Brain Tumor Therapy: Elsevier Ltd; 2016. 667-80 p.
7. Kreftlex. Stereotaktisk vevsprøvetaking ved hjernesvulster: Kreftlex; 2020 [Available from: https://www.kreftlex.no/Sentralnervesystemet-hjernekreft/ProsedyreFolder/UTREDNING/StereotaktiskVevsprove?lg=ks&CancerType=CNS%20Intrakranial&containsFaq=False.
8. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med. 2005;352(10):987-96.
9. Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, Weller M, et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. N Engl J Med. 2005;352(10):997-1003.
10. Oberheim Bush NA, Cha S, Chang SM, Clarke JL. Chapter 55 - Pseudoprogression in Neuro-Oncology: Overview, Pathophysiology, and Interpretation: Elsevier Ltd; 2016. 681-95 p.
11. Zhao F, Li M, Kong L, Zhang G, Yu J. Delineation of radiation therapy target volumes for patients with postoperative glioblastoma: a review. Onco Targets Ther. 2016;9:3197-204.
12. Peeken JC, Molina-Romero M, Diehl C, Menze BH, Straube C, Meyer B, et al. Deep learning derived tumor infiltration maps for personalized target definition in Glioblastoma radiotherapy. [Amsterdam] :2019. p. 166-72.
13. Pirtoli L, Gravina GL, Giordano A. Radiobiology of Glioblastoma : Recent Advances and Related Pathobiology. Cham: Springer International Publishing : Imprint: Humana; 2016.
14. Walker MD, Green SB, Byar DP, Alexander E, Batzdorf U, Brooks WH, et al. Randomized Comparisons of Radiotherapy and Nitrosoureas for the Treatment of Malignant Glioma after Surgery. The New England journal of medicine. 1980;303(23):1323-9.
15. Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiother Oncol. 2018;127(1):49-61.
16. Liang BC, Thornton JAF, Sandler HM, Greenberg HS. Malignant astrocytomas: focal tumor recurrence after focal external beam radiation therapy. J Neurosurg. 1991;75(4):559-63.
17. Sikora MoN, Britt. Forskjell i bildekvalitet mellom ImPax og Eclipse systemer. In: Hospital HU, editor. Power Point Presentation ed: Haukeland University Hospital (intern documents); 2014.
18. The European Federation of Radiographer S. Artificial Intelligence and the Radiographer/Radiological Technologist Profession: A joint statement of the International

Society of Radiographers and Radiological Technologists and the European Federation of Radiographer Societies. Radiography (Lond). 2020;26(2):93-5.
19. Helsedirektoratet. Høring - Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og

oppfølging av hjernesvulster. Nasjonale handlingsprogram med faglige retningslinjer for kreft2020.
20. Jacob SW, Hurlen P, Lossow WJ, Francone CA, Dietrichs E. Anatomi og fysiologi. 2. utg. ed. Oslo: Universitetsforl.; 1985.
21. Wikipedia. Sensory Homunculus [Available from: https://commons.wikimedia.org/wiki/File:Sensory_Homunculus-en.svg.
22. LumenLearning. Glial Cells lumenlearning.com: OpenStax CNX; 2020 [Available from: http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@10.8.
23. Yao M, Li S, Wu X, Diao S, Zhang G, He H, et al. Cellular origin of glioblastoma and its implication in precision therapy. Cell Mol Immunol. 2018;15(8):737-9.
24. Kleihues P, Cavenee WK, International Agency for Research on C. Pathology and genetics of tumours of the nervous system. Lyon: IARC Press; 1997.
25. Louis D, Perry A, Reifenberger G, Deimling A, Figarella-Branger D, Cavenee W, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Pathology and Mechanisms of Neurological Disease. 2016;131(6):803-20.
26. Davis ME. Glioblastoma: Overview of Disease and Treatment. [Pittsburgh, Pa.] :2016. p. S2-S8.
27. Louis DN, International Agency for Research on C. WHO classification of tumours of the central nervous system. 4th Revised ed. ed. Lyon $b IARC Press2016.
28. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016;131(6):803-20.
29. Braganza MZ, Kitahara CM, Berrington de Gonzalez A, Inskip PD, Johnson KJ, Rajaraman P. Ionizing radiation and the risk of brain and central nervous system tumors: a systematic review. Neuro Oncol. 2012;14(11):1316-24.
30. Omuro A, DeAngelis LM. Glioblastoma and Other Malignant Gliomas: A Clinical Review. JAMA. 2013;310(17):1842-50.
31. Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. Neuro-Oncology. 2015;17(suppl4):iv1-iv62.
32. Nurses AAoN. Clinical practice guideline series update: care of the adult patient with a brain tumor. 2014;Journal of neuroscience nursing. Vol.46(6):367-8.
33. Goffart N, Kroonen J, Rogister B. Glioblastoma-initiating cells: relationship with neural stem cells and the micro-environment. Cancers (Basel). 2013;5(3):1049-71.
34. Kane JR. The Role of Brain Vasculature in Glioblastoma. Mol Neurobiol. 2019;56(9):6645-53.
35. Khan FM, Gibbons JP, Sperduto PW. Khan's Treatment Planning in Radiation Oncology. Philadelphia: Philadelphia: Wolters Kluwer Health; 2016.
36. Zikou A, Sioka C, Alexiou GA, Fotopoulos A, Voulgaris S, Argyropoulou MI. Radiation Necrosis, Pseudoprogression, Pseudoresponse, and Tumor Recurrence: Imaging Challenges for the Evaluation of Treated Gliomas. Contrast Media Mol Imaging. 2018;2018:1-6.
37. Schneider W, Bortfeld T, Schlegel W. Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. Phys Med Biol. 2000;45(2):459-78.

38. Niyazi M, Brada M, Chalmers AJ, Combs SE, Erridge SC, Fiorentino A, et al. ESTRO-ACROP guideline "target delineation of glioblastomas". Radiother Oncol. 2016;118(1):35-42.
39. Smith NB, Webb A. Introduction to medical imaging : physics, engineering and clinical applications. Cambridge: Cambridge University Press; 2011.
40. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol. 2015;17(9):1188-98.
41. Westbrook C, Roth CK, Talbot J. MRI in practice. 3rd ed. ed. Oxford: Blackwell; 2005.
42. Gelman N, Ewing JR, Gorell JM, Spickler EM, Solomon EG. Interregional variation of longitudinal relaxation rates in human brain at 3.0 T: relation to estimated iron and water contents. Magn Reson Med. 2001;45(1):71-9.
43. Rooney WD, Johnson G, Li X, Cohen ER, Kim SG, Ugurbil K, et al. Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. Magn Reson Med. 2007;57(2):308-18.
44. Newton HB. Handbook of Neuro-Oncology Neuroimaging. 2 ed: Elsevier Science; 2016.
45. Leeds Uo. Circulatory system: Cappilaries 2020 [Available from: https://www.histology.leeds.ac.uk/circulatory/capillaries.php
46. Cercignani M, Dowell NG, Tofts P. Quantitative MRI of the brain : principles of physical measurement. Boca Raton, Florida ;,London, England ;,New York: CRC Press; 2018.
47. Mechtler L. Neuroimaging in Neuro-Oncology. Neurologic Clinics. 2009;27(1):171-201.
48. Seidel C, Dörner N, Osswald M, Wick A, Platten M, Bendszus M, et al. Does age matter? - A MRI study on peritumoral edema in newly diagnosed primary glioblastoma. BMC cancer. 2011;11(1):127.
49. Westbrook C, Talbot J. MRI in Practice. Newark: Newark: John Wiley & Sons, Incorporated; 2018.
50. Westbrook C, Talbot J. MRI in practice. Hoboken, NJ: Wiley Blackwell; 2019.
51. Gupta A, Young RJ, Karimi S, Sood S, Zhang Z, Mo Q, et al. Isolated diffusion restriction precedes the development of enhancing tumor in a subset of patients with glioblastoma. AJNR American journal of neuroradiology. 2011;32(7):1301.
52. Abdulla S, Saada J, Johnson G, Jefferies S, Ajithkumar T. Tumour progression or pseudoprogression? A review of post-treatment radiological appearances of glioblastoma. Edinburgh :2015. p. 1299-312.
53. Rahmat R, Brochu F, Li C, Sinha R, Price SJ, Jena R. Semi-automated construction of patient individualised clinical target volumes for radiotherapy treatment of glioblastoma utilising diffusion tensor decomposition maps. The British journal of radiology. 2020;93(1108):20190441.
54. Itagiba V, Borges R, Jr LCH, Furtado A, Domingues R, Gasparetto E. Use of diffusion tensor magnetic resonance imaging in the assessment of patterns of white matter involvement in patients with brain tumors: Is it useful in the differential diagnosis? Radiologia Brasileira. 2010;43:362-8.
55. Schultz T. DTI-sagittal-fibers Wikipedia Commons2006 [cited 2006. Available from: https://commons.wikimedia.org/wiki/File:DTI-sagittal-fibers.jpg.
56. Gaillard MGAdVF. MR Spectroscopy Radiopedia2020 [Available from: https://radiopaedia.org/articles/mr-spectroscopy-1.
57. Gaillard DBF. Choline peak Radiopedia2020 [Available from: https://radiopaedia.org/articles/choline-peak?lang=us.

58. Tofts P. Quantitative MRI of the brain : measuring changes caused by disease. Chichester: Wiley; 2003.
59. Gladson CL, Prayson RA, Liu WM. The Pathobiology of Glioma Tumors. Annu Rev Pathol Mech Dis. 2010;5(1):33-50.
60. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017;35:18-31.
61. Zygogianni A, Protopapa M, Kougioumtzopoulou A, Simopoulou F, Nikoloudi S, Kouloulias V. From imaging to biology of glioblastoma: new clinical oncology perspectives to the problem of local recurrence. Clin Transl Oncol. 2018;20(8):989-1003.
62. Stuge TB. Fagocytter 2020 [updated 2020, 27 january. Available from: https://sml.snl.no/fagocytter.
63. Evensen SA. Hemosiderin Store medisinske leksikon2018 [20. februar 2018:[Available from: https://sml.snl.no/hemosiderin.
64. Strauss SB, Meng A, Ebani EJ, Chiang GC. Imaging Glioblastoma Posttreatment: Progression, Pseudoprogression, Pseudoresponse, Radiation Necrosis. 2019:1199-216.
65. Senger DR, Perruzzi CA, Feder J, Dvorak HF. A highly conserved vascular permeability factor secreted by a variety of human and rodent tumor cell lines. Cancer Res. 1986;46(11):5629.
66. Joiner MC, van Der Kogel A. Basic Clinical Radiobiology Fourth Edition. 4 ed2009.
67. Eriksson D, Stigbrand T. Radiation-induced cell death mechanisms. Tumor Markers, Tumor Targeting and Translational Cancer Research. 2010;31(4):363-72.
68. Chang DS, Lasley FD, Das IJ, Mendonca MS, Dynlacht JR. Basic Radiotherapy Physics and Biology. 2014 ed. Cham: Cham: Springer International Publishing; 2014.
69. Pedicini P FA, Simeon V, Tini P, Chiumento C, Pirtoli L, Salvatore M, Storto G. Reports Outline Glioblastomas Study Findings from Institute for Cancer Research and Treatment (IRCCS) (Clinical radiobiology of glioblastoma multiforme Estimation of tumor control probability from various radiotherapy fractionation schemes).(Clinical report). Clinical Trials Week. 2014:134.
70. Kovalchuk N, Niedermayr TR, Qureshi MM, Willins J, Truong MT. Comparison of Deformable Versus Rigid PET/CT Registration in Head-and-Neck Cancer (HNC) Patients Undergoing Postoperative Radiation Therapy (PORT). International Journal of Radiation Oncology, Biology, Physics. 2012;84(3):S511-S2.
71. International_Commission_on_Radiation_Units_and_Measurements. ICRU REPORT 83: Prescribing, Recording, and Reporting Photonbeam Intensity-Modulated Radiation Therapy (IMRT). 2010.
72. Aalen OO, Frigessi A. Statistiske metoder i medisin og helsefag. Oslo: Gyldendal akademisk; 2006.
73. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res. 2015;24(1):9-26.
74. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277(3):813-25.
75. Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. Med Phys. 2016;43(6):2835-44.
76. Kim Y, Cho HH, Kim ST, Park H, Nam D, Kong DS. Radiomics features to distinguish glioblastoma from primary central nervous system lymphoma on multi-parametric MRI. Neuroradiology. 2018;60(12):1297-305.
77. Colom R, Karama S, Jung RE, Haier RJ. Human intelligence and brain networks. Dialogues Clin Neurosci. 2010;12(4):489-501.

78. Seshadri DS. Artificial Intelligence and Agent Technology Made Easy. International Journal of Information Technology and Decision Making. 2016.
79. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29(2):102-27.
80. Despotović I, Goossens B, Philips W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Computational and Mathematical Methods in Medicine. 2015;2015.
81. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993-2024.
82. Wu Y, Zhao Z, Wu W, Lin Y, Wang M. Automatic glioma segmentation based on adaptive superpixel. BMC medical imaging. 2019;19(1):73.
83. Wadhwa A, Bhardwaj A, Singh Verma V. A review on brain tumor segmentation of MRI images. Magn Reson Imaging. 2019;61:247-59.
84. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroirnaging with scikit-learn. Front Neuroinformatics. 2014;8.
85. Nielsen MA. Neural Networks and Deep Learning. http://neuralnetworksanddeeplearning.com/: Determination Press; 2015 2018.12.03]. Available from: http://neuralnetworksanddeeplearning.com/.
86. Browlee J. What is Deep Learning https://machinelearningmastery.com/what-is-deep-learning/2016 [updated 2018 nov 28. Available from: https://machinelearningmastery.com/what-is-deep-learning/.
87. Wikipedia. Convolutional Neural Networks Wikipedia [updated 6. june 2020. Available from: https://en.wikipedia.org/wiki/Convolutional_neural_network.
88. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
89. Isensee F, Petersen J, Kohl S, Jaeger P, Maier-Hein K. nnU-Net: Breaking the Spell on Successful Medical Image Segmentation2019.
90. Kickingereder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol. 2019;20(5):728-40.
91. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging. 2015;15(1):29-.
92. Lim TY, Gillespie E, Murphy J, Moore KL. Clinically-oriented Contour Evaluation using Dosimetric Indices Generated from Automated Knowledge-based Planning. Int J Radiat Oncol Biol Phys. 2018.
93. Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer. 2016;16(4):234-49.
94. Andreassen CN, Eriksen JG, Jensen K, Hansen CR, Sorensen BS, Lassen P, et al. IMRT - Biomarkers for dose escalation, dose de-escalation and personalized medicine in radiotherapy for head and neck cancer. Oral Oncol. 2018;86:91-9.
95. Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. Acta Oncol. 2016;55(7):799-806.
96. Rathore S, Akbari H, Doshi J, Shukla G, Rozycki M, Bilello M, et al. Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning. J Med Imaging (Bellingham). 2018;5(2):021219.
97. Guram K, Smith M, Ginader T, Bodeker K, Pelland D, Pennington E, et al. Using smaller-than-standard radiation treatment margins does not change survival outcomes in patients with high-grade gliomas. Pract Radiat Oncol. 2018.
98. Mann J, Ramakrishna R, Magge R, Wernicke AG. Advances in Radiotherapy for Glioblastoma. Front Neurol. 2017;8:748.

99. Chang EL, Akyurek S, Avalos T, Rebueno N, Spicer C, Garcia J, et al. Evaluation of Peritumoral Edema in the Delineation of Radiotherapy Clinical Target Volumes for Glioblastoma. International Journal of Radiation Oncology, Biology, Physics. 2007;68(1):144-50.
100. Ali AN, Ogunleye T, Hardy CW, Shu HK, Curran WJ, Crocker IR. Improved hippocampal dose with reduced margin radiotherapy for glioblastoma multiforme. Radiat Oncol. 2014;9:20.
101. Brandes AA, Tosoni A, Franceschi E, Sotti G, Frezza G, Amistà P, et al. Recurrence pattern after temozolomide concomitant with and adjuvant to radiotherapy in newly diagnosed patients with glioblastoma: correlation With MGMT promoter methylation status. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1275.
102. Feng X, Tustison NJ, Patel SH, Meyer CH. Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features. Frontiers in computational neuroscience. 2020;14:25.
103. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. London :2017. p. 170117.
104. Rahman MA, Brekke J, Arnesen V, Hannisdal MH, Navarro AG, Waha A, et al. Sequential bortezomib and temozolomide treatment promotes immunological responses in glioblastoma patients with positive clinical outcomes: A phase 1B study. Immunity, Inflammation and Disease. 2020;8(3):342-59.
105. Polit DF, Beck CT. Nursing Research : generating and assessing evidence for nursing practice. 10th ed. ed. Philadelphia: Wolters Kluwer; 2017.
106. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116-28.
107. Hk R. Medical Statistics: For Beginners. Singapore: Singapore: Springer Singapore Pte. Limited; 2016.
108. Campos B, Olsen LR, Urup T, Poulsen HS. A comprehensive profile of recurrent glioblastoma. Oncogene. 2016;35(45):5819-25.
109. Bette S, Barz M, Huber T, Straube C, Schmidt-Graf F, Combs SE, et al. Retrospective Analysis of Radiological Recurrence Patterns in Glioblastoma, Their Prognostic Value And Association to Postoperative Infarct Volume. Sci Rep. 2018;8(1):4561-12.
110. Abbasi A, Westerlaan H, Holtman G, Aden K, van Laar P, van der Hoorn A. Incidence of Tumour Progression and Pseudoprogression in High-Grade Gliomas: a Systematic Review and Meta-Analysis. Official Journal of the German, Austrian, and Swiss Societies of Neuroradiology. 2018;28(3):401-11.
111. Shi R, Ngan KN, Li S. The Objective Evaluation of Image Object Segmentation Quality. Cham: Cham: Springer International Publishing. p. 470-9.

8 Appendix

The following files are enclosed:

Name	Referred to in chapter:	Regarding:
A	4.2	Full BORTEM-17 protocol with inclusion criteria
B	4.3.1	Data exploration: table of details in MRI-sequences
C	4.3.2	Electronic quality manual of HUH regarding transfer and images and treatment plan from SUS/OUS
D	4.3.2	Electronic quality manual of HUH regarding anonymization of treatment plan data
E	4.3.3	HD-GLIO script
F	4.5	REK approval of BORTEM-17 study
G	4.5	DPO rapport BORTEM-17 study
H	4.5	BORTEM-17 patient information and consent form
I	1	Internal HUH-reference regarding image quality in the Eclipse software