UNIVERSITAT DE BARCELONA

Development and validation of cancer markers based on multiparametric MRI and machine learning

Alonso Garcia Ruiz

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

VHIO R
Vall d'Hebron
Institute of Oncology

Development and validation of cancer markers based on multiparametric MRI and machine learning

Doctoral thesis dissertation presented by

Alonso García Ruiz

Performed at the Radiomics Group, Vall d'Hebron Institute of Oncology

Directors: Raquel Pérez López and Carles Majós Torró
Tutor: Roser Sala Llonch

Doctoral Program in Biomedicine
Facultat de Medicina i Ciències de la Salut. Universitat de Barcelona

INDEX

LIST OF FIGURES

FIGURE

LIST OF ABBREVIATIONS

ADC Apparent Diffusion Coefficient

AI Artifical Intelligence

AUC Area Under the Curve

CI Confidence Interval

C-Index Concordance Index

CNN Convolutional Neural Networks

CR Complete Response

CT Computed Tomography

ctDNA circulating tumour DNA

DCE Dynamic Contrast Enhanced

DNA Deoxyribonucleic Acid

DSC Dynamic Susceptibility Contrast

DTI Diffusion Tensor Imaging

DWI Diffusion-Weighted Imaging

FF Fat Fraction

FN False Negative

FP False Positive

FPR False Positive Rate

GBM Glioblastoma Multiforme

HR Hazard Ratio

IBSI Image Biomarker Standardization Initiative

IP In Phase

IQR Inter-Quartile Range

IVIM Intra-Voxel Incoherent Motion

KM Kaplan-Meier

mCRPC metastatic Castration-Resistant Prostate Cancer

mHNPC metastatic Hormone-Naïve Prostate Cancer

MRI Magnetic Resonance Imaging

OOP Out Of Phase

OS Overall Survival

PCNSL Primary Central Nervous System Lymphoma

PD Progressive Disease

PET Positron Emission Tomography

PFS Progression-Free Survival

PR Partial Response

PSA Prostate Specific Antigen

PSR Percentage of Signal Recovery

RANO Response Asessment in Neuro-Oncology

rCBV relative Cerebral Blood Volume

RECIST Response Evaluation Criteria in Solid Tumours

RF Radiofrequency

ROC Receiver Operating Characteristic

ROI Region Of Interest

SD Stable Disease

SPECT Single Photon Emission Computed Tomography

SVM Support Vector Machine

TE Echo Time

TN True Negative

TP True Positive

TPR True Positive Rate

TR Repetition Time

TTP Time To Progression

US Ultrasound

CHAPTER 1

Global introduction

1.1 Motivation

Generally speaking, cancer can be described as a malignant growth stemming from malfunctioning or damaged cells, that prevents the cell from repairing itself while also preventing the normally-occurring cell death and promoting cell division, which leads to an uncontrolled growth of rogue cells that can eventually spread and invade other healthy organs, a process called metastasis. Among the causes of cancer, ageing, environmental agents and certain personal habits, like sedentarism, are known to increase cancer risk, but the root cause of the disease is usually unknown, which poses difficulties in preventing and treating it [1].

The characteristics of each tumour depend on the primary organ where the cells come from, the disrupted genes and further mutations it may develop, its microenvironment, the stage and aggressiveness, the resistance to previous treatments received and many other factors, which depicts the complexity of this disease [2]. The last decades have been a revolution in the understanding of cell biology and genetic alterations, which has led to better cancer diagnosis and therapies, overall increasing the survival rates. New therapeutic approaches guided by such biological mechanisms, such as targeted therapies and immunotherapy [3], have emerged towards personalized medicine, the ideal paradigm in which the specific condition of each patient is known and would be treated accordingly with specific treatments. But the advent of new treatments implies facing new challenges as well.

For example, the guidelines for the evaluation of cancer treatment from radiological images were updated to a newer version to include the assessment under immunotherapy [4]. Since then, atypical patterns of response have been identified following such guidelines on medical images, such as pseudoprogression and hyperprogression [5–7]. That warrants the further investigation and revision of evaluation methods on radiology for newer treatments,

in which the doctorand has participated. Unfortunately, many novel therapeutic strategies appear to be beneficial to only part of the patients [8, 9], with newer paths for cancer to acquire resistance.

This has prompted a race to understand cancer evolution under certain treatments, so that resistance mechanisms can be overcome. In fact, cancer is still one of the leading causes of death worldwide, with increasing incidence rates amounting to roughly 20 million new cases in 2022 and an estimate of 35 million new cases for the year 2050 [10]. That highlights the importance of devising and improving tools for cancer research, especially in the identification of responding patients to specific treatments, that translate to better healthcare.

The so-called "omics" approaches, that is, the analysis of large arrays of variables, have brought the study of biology to new levels, with genomics, transcriptomics, proteomics and metabolomics to process the genome, RNA sequences, proteins and biochemistry profiles, respectively. In cancer, the combination of those into a multi-omic study is hoped to help characterizing and classifying tumours better, and reveal new biological cell pathways to create new drugs or improve current ones [2]. In recent years, advances in image processing techniques applied to radiological images were proposed as a new "omics", radiomics, to take into account multiple metrics derived from images. The further development of those methods into deep learning and artificial intelligence have been postulated as tools that can revolutionize medical imaging, specifically enhancing our ability to understand, detect early, and treat cancer.

1.1.1 Medical imaging in the management of cancer

Medical imaging techniques are able to provide anatomical and functional information of the human body without invasive interventions, which support medical decisions in a rapid and convenient way. As such, medical images play a key role in various medical procedures, from routine inspections, diagnosis of traumatic injuries and guidance of surgical or therapeutical procedures to the detection and monitoring of diseases. In the management of cancer disease, medical images are involved in all steps of the process, namely: screening of the otherwise healthy population for early detection of cancer; diagnosis and estimating the extent of the disease after cancer has been detected; therapeutical procedures, in which imaging usually plays a support role; and monitoring during therapy with regular follow-up checks [11], as pictured in Figure **1**.

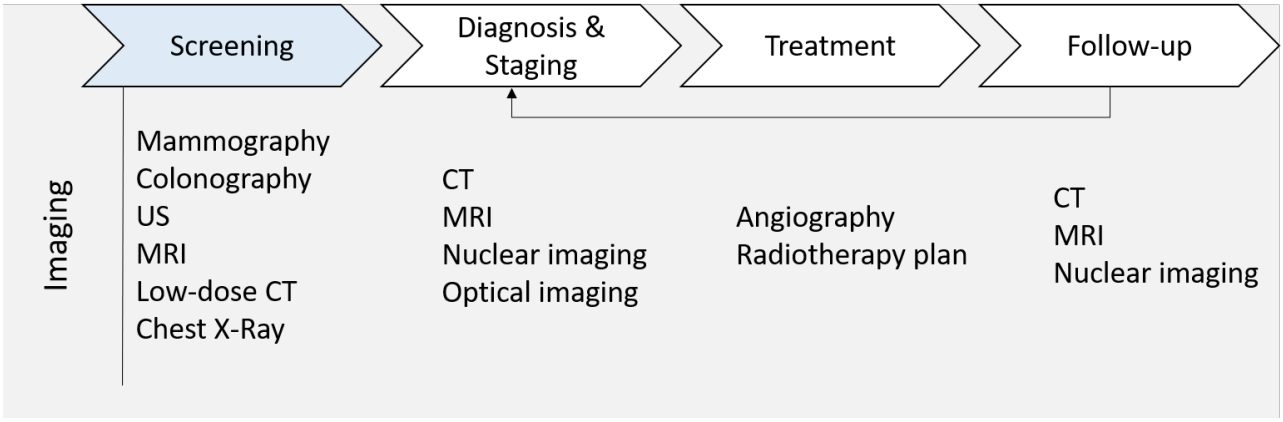


Figure 1: Summary of imaging modalities and applications within the different tasks in the management of cancer. US: ultrasound. MRI: magnetic resonance imaging. CT: computed tomography. Figure adapted from [12].

Imaging techniques

Markers based on medical imaging have the potential to capture unique information, such as the spatial distribution, metabolic or functional activity of cancer inside the body at different times along treatment. Depending on the mechanism of image generation, different types of imaging or modalities can be identified, including computed tomography (CT), nuclear imaging modalities and magnetic resonance imaging (MRI).

CT is an ionizing modality that relies on the attenuation of X-rays traversing through the body to obtain an image, giving contrast between the tissues by their capacity to absorb the rays. Multiple acquisitions at different angles will then be reconstructed into a 3-dimensional volume. The main advantages of CT are its relatively low cost to operate and maintain, its fast speed, its flexible field-of-view and its high spatial resolution. The values in a CT scan are standardised to Houndsfield units, which allows to quantify and compare between scans, an important aspect from an analytical point of view. Among the negative points of CT scans are exposing the patients to ionizing radiation and the limited tissue contrast, especially within soft tissues or specific conditions. CT is defined as a standard-of-care imaging protocol in many clinical situations, including screening and follow-up of patients during cancer treatment, particularly when dealing with metastatic disease, as CT allows to image the whole body in a fast way.

Nuclear imaging is a whole medical discipline on its own, based on capturing the radioactivity of a tracer with affinity for a certain tissue or body function. It encompasses two

important modalities for cancer imaging: positron emission tomography (PET) and single photon emission computed tomography (SPECT), with PET being preferred for its higher spatial resolution. The basic working principle is through the use of a tracer injected in the bloodstream, composed of a radionuclide attached to a specific ligand molecule that will bind target body tissues, usually tumoral cells expressing a given protein or demanding a high supply of glucose. During a window of time, the decay of the radionuclide can be detected through emitted photons (SPECT) or particles (PET). Standardised measurements can be estimated as uptake of the metabolite molecule or activity measured at the region of interest, relative to the tissue and patient characteristics. The sensitivity and specificity of nuclear imaging greatly vary on the tracer and the affinity towards the desired target, which in many cases depends on the histological type of cancer.

Nuclear magnetic resonance refers to the capacity of certain atomic nuclei to absorb energy of a specific frequency when in a strong magnetic field, and contrast between tissues can be captured from the subsequent relaxation processes. In this thesis, special focus will be given to MRI techniques, which will be further elaborated in following sections.

Although the imaging technique of choice would depend on the type of cancer, images typically acquired for the evaluation of cancer are anatomical CT or MRI, every two or three months for assessing treatment response [11], with nuclear medicine being more common in routine clinical use. Each modality provides different and complementary information and data from multiple scanners or multiple contrasts would ideally be acquired; however acquiring multi-modal image data needs to be cost-effective and practical to be implemented in standard clinical healthcare. Therefore, prior experience in clinical studies is warranted to build guidelines and ultimately regulations and protocols to perform the most informative scans in each scenario.

Screening

Routine screening of the population, especially of those individuals susceptible of higher risk of cancer like those with Li-Fraumeni syndrome [13] or BRCA gene mutations [14], the elderly or regular smokers [15], is critical for the early detection of the disease. An early detection facilitates the therapy and the surgical removal of tumours, effectively increasing the survival rate [11]. The screening procedure depends on the cancer type, but may take into account multiple sources of information: imaging evidence, specific marker levels in blood tests and biopsy results. An example of a screening imaging technique for the detection of breast cancer

is mammography, which is widely implemented in many countries, significantly reducing the mortality [16]. However, a number of cancers are still mainly detected in an advanced state, such as pancreatic or ovarian cancer, and other times, the screening procedure is a long process involving biopsies and histopathology assays, like for prostate or colorectal cancer [17]. Because of that, cost-effective imaging techniques can make a difference in preventing cancer deaths, as well as future potential markers in the blood, such as circulating tumour DNA, cells or metabolites.

Diagnosis

When cancer is suspected from symptomatology or imaging evidence, a formal diagnosis should be carried out. The type of cancer and its degree of malignancy must be assessed to determine the severity of the disease and the next steps. For that, medical imaging is essential in locating the disease within the body and they are often the first source of information, since image acquisition and evaluation is relatively fast. Images can provide enough information for an initial diagnosis [18, 19], and also guide biopsies, to confirm it via histological analysis of the tissue sample. Because biopsies are invasive procedures, with associated risks and longer evaluation times, a large research effort is focused on obtaining accurate and reliable diagnostic markers from medical images imaging [20]. There are a number of potential benefits from imaging at this stage, such as reducing the number of unnecessary surgeries and providing complementary diagnostic and prognostic information, as seen in Figure **2**.

Staging, grading and prognosis

Following diagnosis, the treatment for a certain type of cancer may depend upon how developed and how spread is the disease within the body. Staging and grading refer to established procedures that categorize the disease in a scale from benign or less aggressive to a high-risk malignancy, based on imaging, histopathological, molecular and genetic evidences [11, 21]. All of the information known about the type of cancer and the patient's condition is combined for a prognostic assessment. That provides the expected likelihood of survival or disease relapse and will determine the treatment options and whether careful monitoring is necessary.

Restaging and evaluation of the prognosis may be repeated when different therapies or procedures take place, such as surgical removal of cancer tissue and the response to specific treatments. For those reasons, finding prognostic factors is a primordial research focus to allow not only assessing the risks of diseases and choosing the best therapy option, but also

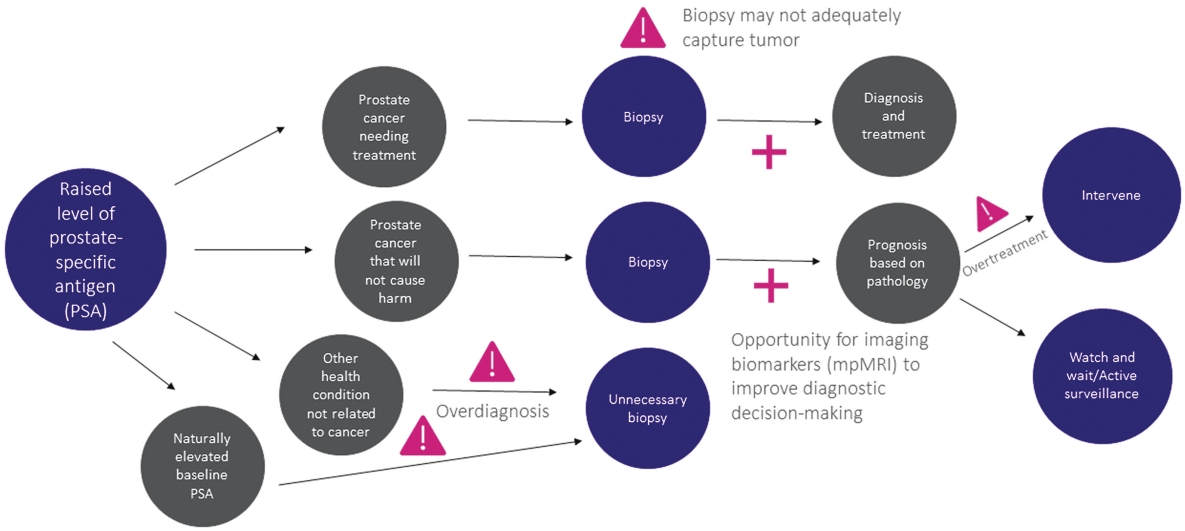


Figure 2: Current procedures when detecting and diagnosing prostate cancer. Limitations are highlighted with a warning sign. Potential improvements from imaging are marked with a plus sign. Figure reproduced from [17].

establishing reference values of mortality. That itself enables identifying high-risk subpopulations and validating therapeutic methods and drugs. To those ends, multiple sources of data are necessary to evaluate the full extent of the disease and medical images provide the ideal means of analysing cancer inside the body that can inform the prognosis.

Monitoring of treatment response

Once a patient starts receiving treatment, the evolution of the patient's condition must be carefully and frequently monitored to determine if it is responding, that is, if the therapy is still effective. Follow-up checks with acquisition of medical images are a standard practice for patients with advanced cancer or receiving some therapy prior to the main treatment (i.e., neoadjuvant treatment), as opposed to the total resection of cancer, for which other markers must be used.

The variables used for assessing response include any symptoms associated with the disease, blood test analyses, possible drug secondary effects, and mainly biomarkers of response specific to the cancer type. A biomarker of response can be defined as a quantifiable parameter related to the underlying biophysiological properties and conditions of a disease that is able to capture and indicate changes induced by a successful treatment [22]. Therefore, established response biomarkers are usually the primary guidance to patient monitoring during treatment, and essential in testing the effectiveness of new drugs under development.

The Response Evaluation Criteria In Solid Tumours (RECIST) version 1.1 [4], based on medical imaging, is the standard method for evaluating treatment response in clinical trials and patient care follow-up. It establishes four categories of response based on measuring the change in diameter of certain lesions during treatment, and its limitations will be presented in the following section.

1.1.2 Current clinical imaging markers and limitations

In the clinical setting, medical images are visually inspected and assessed by radiologists, receiving specialized training during years of career experience. In the context of cancer, established imaging biomarkers are based on visual assessments [23] and manual annotations are the reference method to evaluate the effectiveness of cancer treatments. The specific markers from images and the steps and rules to apply them are collected and integrated into different guidelines defined for specific cancer types or subtypes, based on prior evidence.

Some prominent guidelines for response assessment are the RECIST criteria [4]; the Response Assessment in Neuro-Oncology (RANO) cancer in the central nervous system [21]; and the Positron Emission Tomography Response Criteria in Solid Tumours (PERCIST) for evaluating PET images [24]. Other derived guidelines provide additional measurements for limiting cases, such as the METastasis Reporting and Data System for Prostate Cancer (MET-RADS-P), for evaluating advanced prostate cancer [25].

A common property of those guidelines is that they all rely on the manual delineation or measurement of the tumour size in order to evaluate the response based on changes in size during treatment. For example, the RECIST criteria establishes a defined number of target lesions measured at the start of treatment and at every follow-up. For soft tissue lesions, the diameter along the longest axis is measured in the axial plane, of up to two per organ and up to five in total. There are some exceptions to these rules, as for instance regarding lymph nodes, and some lesions are considered non-measurable, like lesions smaller than 10 mm and osteoblastic bone lesions as evaluated on CT scans [24, 25]. The change in number of lesions and in size of measurable target lesions will determine the response of the patient in four categories, from complete response to progressive disease, as indicated in Figure **3**.

Treatment response assessment after surgical removal of cancer is only performed by imaging if there is remaining tumour. Such is the case of lesions in the brain, for which only

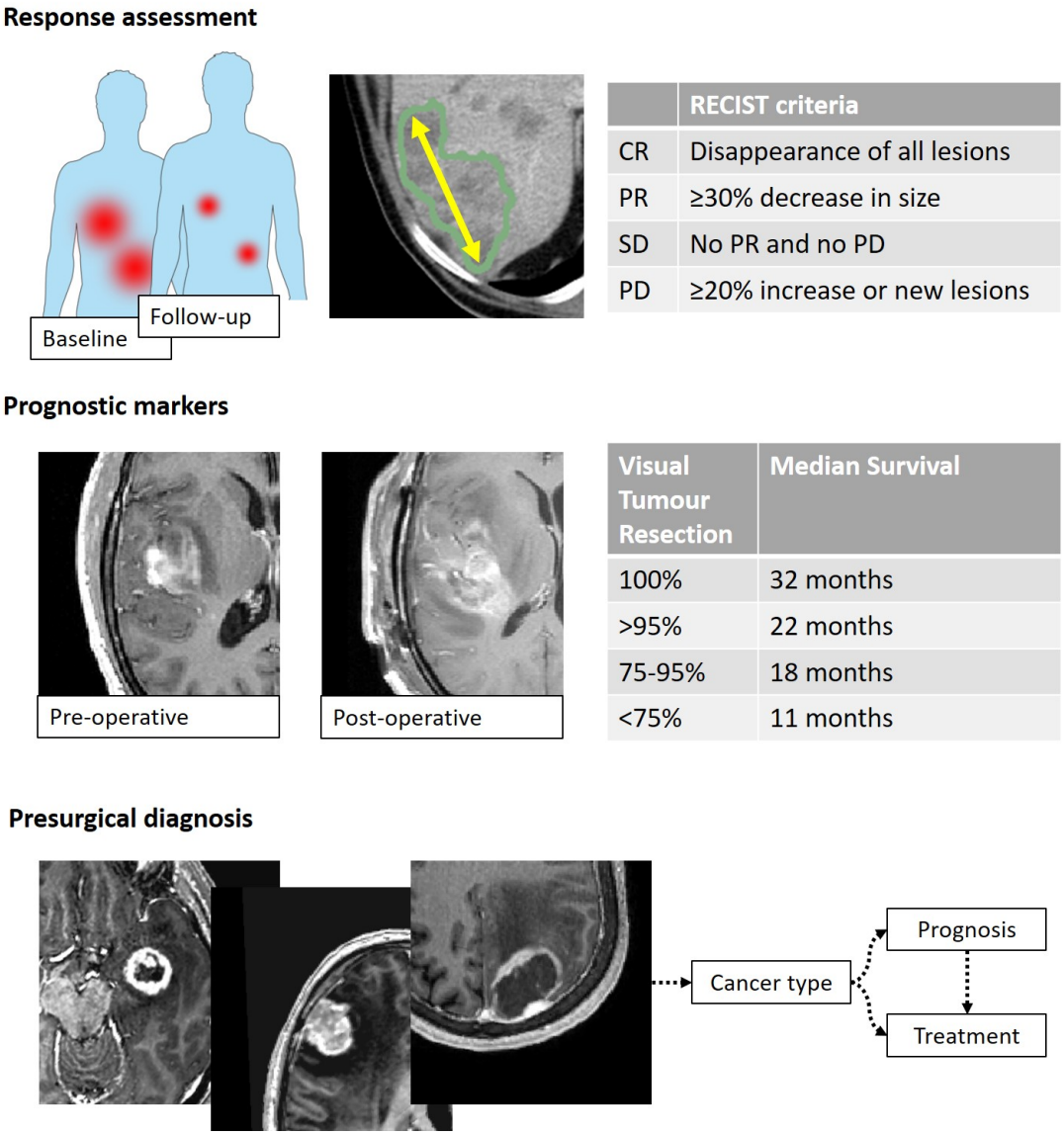


Figure 3: Examples of manual and visual assessments of radiological images for monitoring response to treatment, estimating prognosis and diagnosing cancer disease. The blue human shape was used under license from Servier.

partial resection may be achieved in order to preserve important brain regions, necessary for life and normal functioning. In that case, the percentage of tumour resection and measures or patterns such as the thickness of residual tumour [26–29] are useful prognostic factors to guide further medical decisions (Figure **3**). However, they are based on a subjective visual evaluation of images and therefore subject to higher variability and lower interpretability and reproducibility. To that end, some studies also suggested quantifying the volume of remaining disease [30].

Visual inspection of medical imaging can facilitate non-invasive cancer diagnosis, supporting medical decisions. Initial diagnosis based on radiological images can guide surgical procedures and accelerate the access to therapy [18]. However, the visual interpretation of medical images can be challenging in some cases, when malignancies have a similar appearance [31], as seen in Figure **3**. Those situations may require the consensus of a board of expert radiologists and, ultimately, undergoing surgery to obtain a tissue sample and determine the nature of the cancer in question.

In effect, there are a number of limitations inherent to current manual and visual approaches. The lesions selected for monitoring may be different depending on which physician evaluates the patient, which introduces a source of variability. In addition, the diameter and selected lesions alone may not be completely representative of the disease and some lesions cannot be currently evaluated by RECIST, such as osteoblastic bone tumours. Estimating the percent of tumour resected after surgery and finding patterns on medical images to diagnose cancer depend on the visual interpretation of radiologists, therefore suffering from variability and a degree of subjectivity. Those aspects hinder the generalization and validation of techniques for cancer management across studies, making it difficult to find reproducible, consistent and robust methods that can become globally useful metrics and markers of response, ultimately leaping from research to real clinical applications and advancing patient healthcare.

Medical images have the potential to characterize cancer, or at the very least, provide complementary information to support medical decisions, in an objective way. The computational analysis of medical images allows for quantitative measures of the whole tumour volume, including its intensity, texture and morphology. Machine learning algorithms and statistical models have been previously shown to relate imaging features with the underlying tumour biology [32] for a better characterization of cancer. Furthermore, mathematical models based on neural networks have recently experienced a surging development in image processing tasks, opening new opportunities for computational tools in medical scenarios.

Given the need for a better characterization of cancer from medical imaging, in this work we defined three objectives to explore potential markers derived from MRI for three main applications in the management of cancer patients, the estimation of prognosis, the diagnosis of cancer and the monitoring of treatment effectiveness.

1.2 Development of new imaging markers

To find novel markers of clinical value, a measurement, metric or feature must be proposed based on previously established biomarkers or formulating a hypothesis on the working principle of the proposed feature [23]. Next, a study should be designed to prove the hypothesis or the improvement compared to other markers. In the study design, a population of reference should be carefully chosen so that no biases would later affect the results, along with an estimated number of data samples, in this case patients' data, that enables enough statistical power.

Data must then be gathered, organized, cleaned and pre-processed, taking note of the criteria for inclusion and exclusion of patients in the study. When the data is ready, the desired feature is extracted or computed, depending on the initial formulation, and then analysed. The analysis can include observing the proposed feature value distribution and variability, looking for possible unaccounted biases or trends.

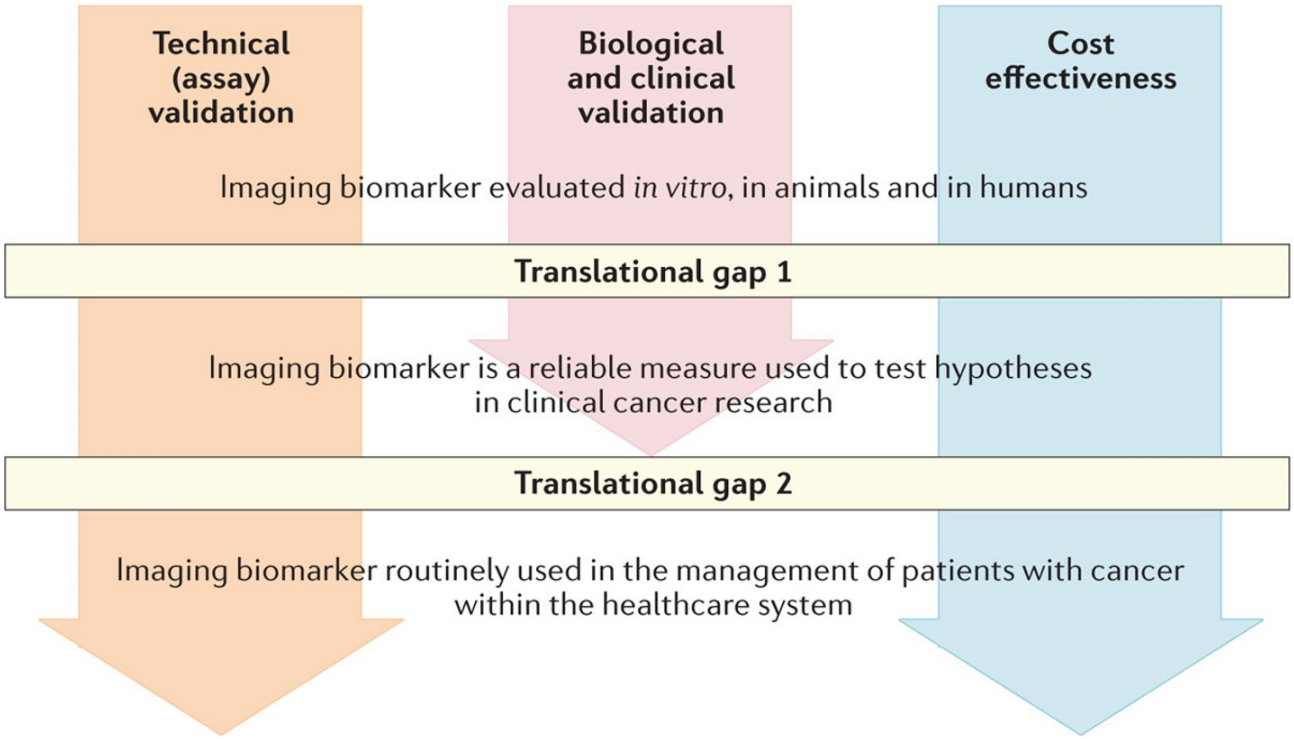


Figure 4: Process and challenges in developing imaging biomarkers, where they must go through a validation process (gap 1) and a feasibility process (gap 2) to finally reach clinical use. Figure reproduced from [33].

Finally, it must be demonstrated that the new feature performs the task for which it was conceived. For that, it can be compared to existing markers and it can also be statistically associated with clinical variables, depending on the hypothesized application of the feature. For that, the existence of a ground truth or, alternatively, a gold standard is of high relevance to compare against a reference. Eventually, for a biomarker to reach clinical use, it must be technically, clinically and practically feasible, as shown in Figure **4**.

1.2.1 Clinical outcomes and validation process

In the process of validating new features as biomarkers, clinical outcomes can be used in several ways, acting as reference metrics or as the output of regression models. For instance, in a clinical application, we may be interested in relating measured features like size and intensity of a cancer lesion of a patient with the well-being of the patient or the effectiveness of a given treatment. Survival times are well-established and generalized clinical endpoints used to validate the effectiveness of new drugs within clinical trials, and therefore are key for the discovery of biomarkers. Some of them include:

- Overall survival (OS): for each patient, it is defined as the time elapsed from the start of a specific treatment to the time of death or, alternatively, to the last recorded date known for the patient (in that case, the data is considered "censored"). The OS is an absolute measure of a patient's well-being, meaning that it is not specific to one treatment, given that one patient may receive multiple treatments that influence the OS, and it does not make assumptions about the cause of death of a patient, which may or may not be directly caused by the disease. However that may seem inaccurate, it is also absolute and unambiguous to implement, which is why OS is considered the main endpoint to evaluate and approve new drugs presented to the major regulatory agencies. However, collecting the OS usually takes a relatively long time and hence not always possible to use.

- Progression-free survival (PFS): it is defined as the time from the start of a specific treatment to the moment when clinical disease progression is detected, or death. If none of the events happens, then the last date is used as censored data. Compared to OS, the disease progression can be more difficult to assess and there is some degree of inter-observer variability. The PFS is a very common endpoint variable in comparing therapies and evaluating markers with predictive value of response (before receiving treatment). Because it captures the moment when the disease relapses or worsens within a line of therapy, it relates closer to the effectiveness of that specific drug or

procedures than OS. But because it also takes into account deaths, it is ultimately a proxy for OS, in case the latter is unavailable.

- Time to progression (TTP): it is defined as the time elapsed from the start of a specific treatment to the moment when clinical disease progression is detected, with deaths accounting as censored data. It is regarded as similar to PFS, but this endpoint assumes the disease progression as the only factor to determine treatment effectiveness, as deaths might not be necessarily caused by the disease under study. Because time to death is considered the most reliable reference in cancer, PFS as a proxy for OS is usually preferred to TTP, though arguably, TTP may be less affected by events unrelated to the disease, especially in smaller cohorts.

Regarding the radiological assessment of medical images, the RECIST v1.1 criteria is the reference guideline to study changes in the size of specific target lesions to determine the response to the treatment in cancer patients. Four categories are used to stratify patients according to the degree of response, which can be used as a clinical outcome variable:

- Complete response (CR) when all lesions completely heal and disappear.
- Partial response (PR) with at least a 30% decrease in sum of the longest diameter of the lesions from start of treatment (baseline) and no new lesions appear.
- Progressive disease (PD) is defined as an increase of at least 20% in the total sum of the longest diameters compared to the lowest measurement (nadir), with an absolute increase of at least 5 mm or the development of new lesions.
- Stable disease (SD) if changes do not meet the criteria of other categories.

Following those guidelines, a binary stratification can be made such as responding patients (CR and PR) and non-responding patients (SD and PD).

Additionally, the histological analysis of a tumour sample from biopsies is often considered the gold standard technique in the diagnostic process, which determines the specific classification of the cancer with high accuracy. The type of pathology identified, such as tumour type, grade or molecular characteristics, can be used as a label for classification models. Similarly, measures of the healthy and tumour regions, such as cellularity and the presence of certain biological structures, such as blood vessels from biopsy samples, can be used to further explain and validate the results obtained from medical images.

1.2.2 Hand-crafted features and radiomics

Image processing and computer vision problems involve tasks such as detection of specific objects, segmentation of structures and quantification of certain characteristics of objects from an image (features), allowing to automate the tasks or make them as convenient as possible in a vast range of applications involving images, e.g., satellite imagery, manufacturing quality inspection, self-driving cars and microscopy image analysis. So-called expert systems that support medical procedures have already been implemented in the radiology field, such as computer-aided diagnosis, surgery and radiation therapy tools.

Traditionally, solutions to specific tasks could be solved by means of hand-crafted features. Those can be regarded as manually designed filters that capture a specific pattern in an image or as an ad-hoc combination of image operations to detect or measure a specific object in images. Some common processing methods include image morphological operations, connected component analysis, convolutions with filters and transformations. Those methods provide derived maps and segmentations, but also extract intensity and spatial information, namely intensity distribution, texture and shape features.

Hand-crafted features are easy to interpret because their formulation follows a logic to capture a given trend or pattern in an image. However, that also means that they may be limited when trying to describe more complex structures. Additionally, the extraction of features usually depends on some manual parameters that should be set in accordance to the characteristics of the input images, such as resolution and noise, and therefore careful data inspection and pre-processing may be needed.

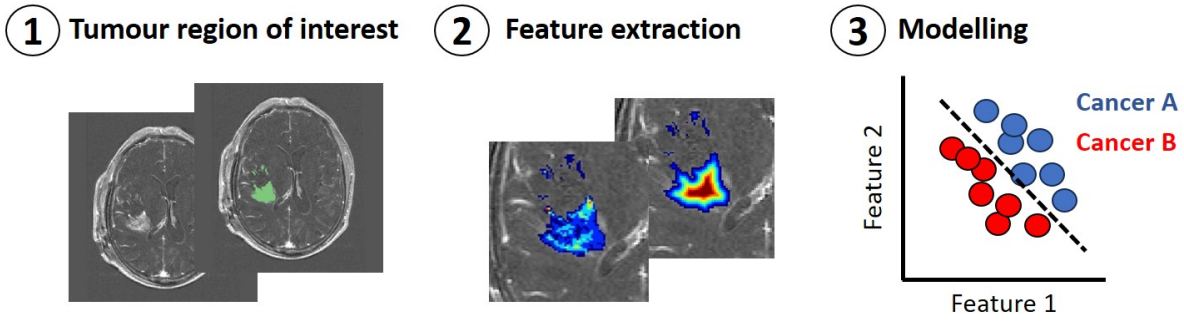


Figure 5: Representation of a radiomics pipeline, with the extraction of intensity and texture features from the region of interest and the final modelling or machine learning for diagnosis or associations with patient survival.

Traditionally, interpreting medical images involves describing a pathology's appearance,

shape, and size. Radiomics takes this concept a step further [34]. It involves extracting and analysing a large number of these features from radiological images, similar to other fields that deal with extensive and personalised data, such as gene analysis in genomics. Radiomics variables typically include first-order statistics of the image intensity distribution, second-order or texture features, shape features such as sphericity and additional features from filters like wavelets. Those features are typically used in conjunction with machine learning to establish relationships with clinical outcomes, existing biomarkers or other tumour features from histological samples of molecular characterization [35]. An example of that is shown in Figure **5**.

Studies including radiomics features have been mainly carried out on CT scans, as this is the most commonly performed imaging technique in cancer patients both for diagnosis and follow-up. Initially, there were controversial results regarding their robustness [36]. Nevertheless, several efforts have been pursued to improve reproducibility, such as the Image Biomarker Standardization Initiative (IBSI) [37] guidelines and some studies dedicated to the application of methods to minimise the variability of radiomics features were performed to improve their reproducibility [38–41]. In the study by Ligero et al. [42], where the doctorand participated, the robustness of the features was improved by means of image pre-processing and batch effect correction methods, a significant contribution to the field.

Radiomics features have been shown to correlate with the underlying tumour biology [32, 40, 41, 43] and, in recent studies, they have been associated with clinical outcomes of patients [44, 45], highlighting the potential of radiomics to capture information from medical images for clinical purposes.

1.2.3 Learned representations and artificial intelligence

In recent years, there has been a rapid development of models that have achieved impressive performance on high-level tasks, such as language comprehension and image segmentation, which readily prompted the label "artificial intelligence". Artificial intelligence or AI can be broadly defined as the ability of an automated system to perform a task at the level of a human, with machine learning regarded as a subfield of AI. A common aspect of recent successful AI models is that they are based on neural networks, referring to a type of mathematical model that resembles brain neuron connections. Neural networks used in AI are often referred to as deep learning models due to their complex, layered architecture. These are characterised by multiple nodes, hosting functions, and weighted connections between

them through chained layers of nodes (Figure **6**).

Although the conception of the perceptron and neural networks dates from circa 1960 [46], only recently those models started showing human-like performance on tasks like image recognition, for which the availability of large datasets and ever more powerful computer resources and the development of training optimization strategies may have been key. As of today, neural networks are witnessing a revolution in their development and application.

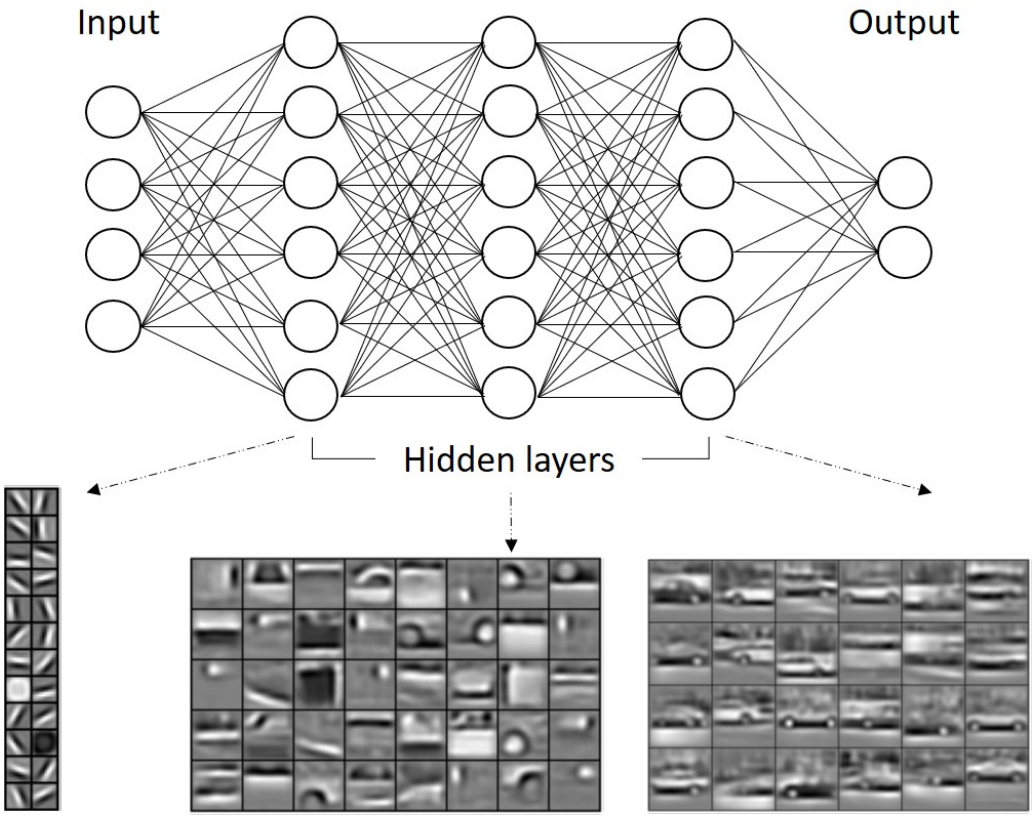


Figure 6: Graphic example of a neural network showing the interconnected nodes from the input to the output. In the hidden layers, deep CNNs can learn complex object shapes. Figure adapted from [47].

Neural networks are fitted end-to-end iteratively usually via the gradient direction that minimizes the error or loss function. Even though deep learning is used to perform a regression or classification just like machine learning methods, there are some key differences: firstly, neural networks can self-configure to obtain relevant features from the input, instead of using hand-crafted features; secondly, non-linear functions can be introduced in the nodes, such as the step or arc-tangent functions; and thirdly, neural networks can learn complex representations of the input.

The concept of deep learning may refer to the type of features found: in the first layers, low-level features are detected, such as edges, texture or simple shapes, while in deeper layers the model can combine previous features into more detailed and specific shapes, or even recognizable objects [47], as depicted in Figure **6**.

One particularly successful type of deep learning architecture for image analysis is the Convolutional Neural Network (CNN). CNNs include convolution layers in their architecture, which can be regarded as self-configuring convolution filters, and have demonstrated high performance for varying image processing tasks, like detection, segmentation, image synthesis and style transfer, but also applied to medical problems successfully [48, 49]. On generative text-to-image tasks, diffusion models have shown greater performance than CNN-based models. Similarly, a recent type of network unit called transformer has surpassed older architectures for natural language processing. Although transformers were originally thought for text inputs, vision transformers have been implemented by splitting an image into a sequence of patches. This has resulted in promising tools for the automated analysis of digitalized histological data [50–52]. However, for pure image tasks such as segmentation, convolutional layers still play a major role [53].

Possibly one of the most severe setbacks of deep learning architectures is interpreting what the model is inferring from the input and explaining the learned representation. That has prompted researchers to look for methods that enable a better understanding of the relation between the inputs and outputs of deep models, a topic coined as "explainability" [54]. The usual way to evaluate the models focus on how important the parts of the input are with respect to the output, called attribution. However, complete and interpretable explanations of what the model captures are still largely missing and remains a topic of active research [55]. This is a current necessity to build the confidence of researchers, physicians and patients for the potential implementation of AI in medical scenarios.

1.2.4 Machine learning methods and survival analyses

Machine learning can be broadly seen as a set of methods and algorithms that enable feature extraction, transformation, reduction and selection, and modelling of the features for regression or classification tasks. A common way to categorize machine learning methods is into supervised and unsupervised approaches. Unsupervised methods refer to those algorithms that adjust to the input data, without specific labels assigned, so that patterns can be cap-

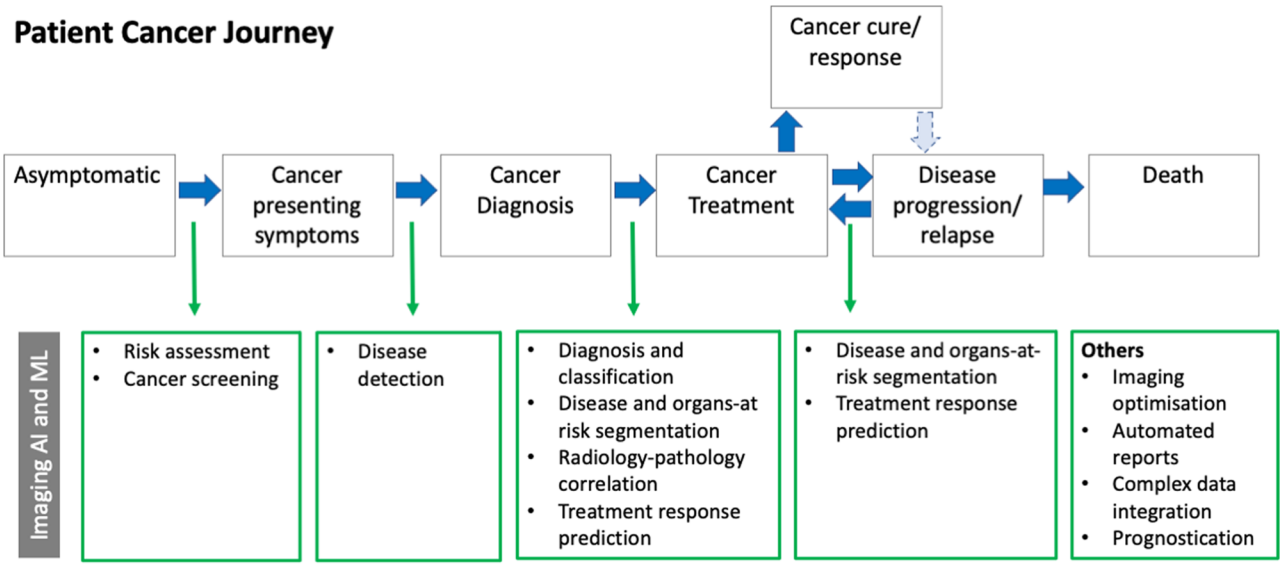


Figure 7: Schematic of the potential applications of machine learning and artificial intelligence in medical imaging for all phases of cancer management. Figure reproduced from [56].

tured for, e.g., clustering, dimension reduction and visualization. In contrast, if the data used as input has a known associated label or outcome that we would like to predict, that is referred to as supervised learning. These include regression and classification methods, which will be used in this work to find associations of potential biomarkers with clinical variables, as seen in Figure **7**. Specific methods will be further discussed in the next section.

A simple example would be to fit a model such as a linear regression, in which the output is defined as a linear combination of each input multiplied by an unknown weight, and the fitting of the weights or training is achieved following a function such as the difference between the estimated output and the actual true value (the error, loss or cost function, depending on the exact definition). In this process, a subset of data is commonly used for training, and another subset is withheld for testing the performance of the fitted model on unseen data. Some common models will be introduced next.

Logistic regression

The logistic regression is one of the called generalized linear models that performs a linear combination of input features to relate them to an output variable in binary form or a categorical 2-class variable, which would be unsuitable to fit using a linear regression. The logistic regression employs the logarithm of the odds (log-odds), based on the probability for the binary output variable. In Equation 1, p is the probability for the output variable, given

the linear combination of coefficients β and n input variables x, and solving for p reveals a logistic or sigmoid function.

$$ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{1}$$

Random Forests

They are a combination of decision trees. In a decision tree, an input passes through a series of nodes (leaves of a tree) where a binary decision is made (branching) towards predicting an output, with the optimal condition and threshold for decisions determined during training. Random forests combine multiple trees with random subpopulations and parameters to make the final result from an ensemble, making it more robust than single trees. They also provide a measure of importance for input features, which is helpful as feature selection. Although not as simple as a single decision tree, random forests can be still be interpreted or visualized.

Support Vector Machines

Support vector machines (SVM) consider each sample as a vector of N input features in an N-dimensional space, mapping them using a defined kernel, and finding the decision boundary or hyperplane with the maximum margin between the classes given for supervised classification. That is done using the closest vectors to the boundary for each class (called support vectors). This algorithm is therefore useful when the problem requires finding the largest distance between classes and, since the data can be mapped in different ways by defining the kernel, it can handle non-linear relationships.

Kaplan-Meier survival

The Kaplan-Meier (KM) method estimates the probability of survival as a function of time within a population, allowing to compare the survival with respect to an event (progression or death) among different populations. The survival probability $\widehat{S}$ at a given time t is given as the product of the chance of surviving at time t with the chances at any previous times $t_i < t$ (Equation 2).

$$\widehat{S} = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{2}$$

Where d_i is the number of events and n_i the number of surviving individuals at time t_i. A well-known feature of the Kaplan-Meier analysis is that the probability can be plotted against

time, effectively obtaining survival curves representing the time elapsed for specific events within a population or populations, useful to visually compare the difference in survivals, usually coupled with the log-rank statistical test.

Hazard funtion and hazard ratio

In contrast to the survival function, the hazard function or rate estimates the risk of a progression or death at time t, i.e. it gives the number of expected events. The survival function and the hazard function are related to each other, and the hazard can be obtained by taking the negative logarithm of the survival function. The hazard function formulation is given in Equation 3 with the same notation used for survival in Equation 2.

$$\widehat{H} = \sum_{i:t_i \leq t} (\frac{d_i}{n_i}) \tag{3}$$

The hazard ratio (HR) is a common parameter used as measure of the biomarkers to capture prognosis, and it is defined as the hazard rate of the population under study relative to the rate of a population of reference or control, making it independent of time. E.g., for an HR of 2, that population has twice the risk than the reference, and for an HR of 0.5, half the risk (an HR equal or close to 1 would mean no difference).

Cox regression

Also known as the Cox proportional hazards model, it defines a hazard function h(t) as a function of input features x_i, with coefficients β_i, with respect to a baseline hazard h_0, which corresponds to the model with nulled features (Equation 4).

$$h(t) = h_0(t)e^{\beta_1 x_1 + \cdots + \beta_n x_n} \tag{4}$$

From the Cox model, the coefficients are usually reported as e^{β_i}, which effectively provides the estimated HRs for each of the input features, allowing to directly associate each feature with a relative risk of progression or death. If $\beta_i = 0$, then $e^{\beta_i} = HR_i = 1$ and the feature x_i would not contribute as a hazard to the model. It must be noted that the e^{β_i} in the Cox model depends on the units of x_i, therefore attention should be given to the units and the confidence intervals when interpreting the HRs.

1.2.5 Measurements of performance

In the development of new methods, determining their performance is at the core of the process, as it will be the basis for its potential implementation. In the case of markers of prognosis, the performance can be regarding its ability, at the baseline of treatment, to identify patients that will experience lower and higher survival rates or, in other words, the survival risks associated with the numerical value of the marker. To measure response to therapy, survival data can also be used as outcome, but this time measuring after the treatment has started to see if the changes or values are associated with different survival rates. In case of diagnostic purposes, the marker should be able to accurately differentiate among pathologies, according to the ground truth, usually given by histological analysis from biopsies.

Common data distribution, correlation and statistical tests

The distribution of variables can be explored as a preliminary analysis and to further consider data assumptions and transformations. However, trends and differences among populations discovered by simple stratification of patients can already be relevant for medical applications. When considering the means between two groups, a common statistical test using the t-distribution is the Student's t-test, with the null hypothesis being that the two samples have identical means. It assumes normally distributed data and equal variances, so an alternative is the Welch's t-test for unequal variances or small populations where a normal distribution cannot be assumed.

When establishing associations between two continuous variables, correlation analyses are helpful in providing the strength of the association and further proof for cause-effect relationships, in specific cases. They test the null hypothesis that there is no relationship between the two variables. Pearson's correlation tests the linear relationship between two variables and assumes normal data and equal variance, whereas Spearman and Kendall correlations use ranks or ordered data to establish a relationship, handling non-normal data.

Classification metrics

When predicting a binary classification, there are four possible cases: a true positive (TP), when the predicted value and the truth are positive; a true negative (TN), when the predicted value and the truth are negative; a false positive (FP), when the predicted value is positive and the truth is negative; and a false negative (FN), when the predicted value is negative and the truth is positive. These values can be conveniently summarized in what is

called the confusion matrix or table. From those, some derived metrics provide useful information regarding classification problems. The precision or positive predictive value gives the predicted TP relative to the all predicted as positive:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

The sensitivity or true positive rate (TPR) gives the fraction of the predicted TP relative to all the actual positives:

$$sensitivity = TPR = \frac{TP}{TP + FN} \tag{6}$$

The specificity, on the other hand, provides the fraction of predicted TN with respect to all actual negatives. It is related to the false positive rate (FPR) as FPR=1-specificity.

$$specificity = \frac{TN}{TN + FP} \tag{7}$$

These metrics can be reported together or combined, given that a trade-off is usually reached between sensitivity and specificity for imperfect classifiers. The accuracy calculates the ratio between all correctly predicted samples over the total samples, giving a general performance overview:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Classification models are typically trained to perform in a balanced way in the trade-off between specificity and sensitivity, and for that the Youden's index is an example of the threshold value used for classification that would give the maximum sum of specificity and sensitivity. However, in specific scenarios, one of those may be more relevant to maximize because of real-life consequences; for instance, in the detection of cancer, we may prefer a high sensitivity over specificity, even if the number of false positives is high, so as not to leave any cancer patient undetected.

Area under the ROC curve and concordance index

The Receiver Operating Characteristic (ROC) curve is a visual interpretation of the performance of a binary classifier, with the TPR on the y-axis against the FPR on the x-axis. For each possible value of a feature or variable that can be used for classification, the TPR and FPR are drawn in the graph in a monotonically increasing curve. A perfect classifier would show a 100% TPR and 0% FPR, with the maximum area under the ROC curve value of 1 (area relative to the area of the perfect square under the curve). For imperfect classifiers, the

ROC curve shows the trade-off between TPR and FPR (or alternatively, between specificity and sensitivity) for every possible value to apply the binary classification and the area under the ROC curve (ROC-AUC) would remain below 1, with the reference AUC of 0.5 for a random binary classifier.

The concordance index or C-Index can be seen as a generalization of the AUC-ROC for survival models with censored data, and measures the ability of the model to correctly rank the patients in the correct order of survival, given the risk scores [57].

1.3 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) stands out from other imaging techniques in that it uses non-ionizing radiation, innocuous for patients, and it offers a wide range and flexibility for imaging contrasts that makes it possible to image virtually any body part, especially soft tissue, given high water and fat content. Those different contrast images can be manually configured in the form of "pulse sequences". To better understand its potential and limitations, a brief introduction will be given next.

1.3.1 Brief introduction to MRI

Descriptions of nuclear magnetic resonance can be found in the literature regarding the quantum mechanics point of view, helpful in understanding the properties of a single nucleus, as well as classical mechanics descriptions, convenient considering a region of nuclei with similar characteristics. In this synthesis, only an overview will be given as a gentle approach to the topic. Magnetic resonance refers generally to a physical phenomenon, which can be used for spectroscopy and imaging, the latter referred to as MRI. Magnetic resonance can be observed in nuclei with spin (non-zero spin, such as H1 protons), a property analogous to the angular momentum and that gives the nuclei a magnetic moment [58]. Magnetic resonance is usually tuned for protons because of their high natural abundance, especially in water molecules in the human body; therefore, protons will be considered as the magnetic resonance target from here on.

In a magnetic resonance experiment, first the spins are exposed to an external magnetic field B_0 which causes them to align with it, creating a net magnetization M_0 in the same direction, given a slightly higher number of spins in a low energy state (oriented in parallel) than high energy state (antiparallel) in equilibrium; the number in excess depending on magnetic field strength, number of spins ρ and temperature T, following a Boltzmann distribution (Equation 9). The spins precess about the external magnetic field axis with

a characteristic frequency ω_0 associated to the energy difference between the two states, called resonance or Larmor frequency, and that is specific of the nuclei and dependent on the external magnetic field strength (Equation 10, where γ is the gyromagnetic ratio).

$$M_0 \propto \frac{\rho\gamma^2 B_0}{T} \tag{9}$$

$$\omega_0 = \gamma B_0 \tag{10}$$

An oscillating magnetic field in the form of an electromagnetic pulse applied for a given time provides energy and perturbs the spins from the equilibrium (excitation). For this to happen, the pulse and the protons must be in resonance, that is, at the same Larmor frequency. Because it is in the radio range for protons, the pulse is also called radiofrequency or RF pulse. In Equation 11, the rate of change in magnetization M can be described as the cross product with the perceived magnetic field B, given by both the external field and the RF pulse.

$$\frac{d\mathbf{M}}{dt} = \mathbf{M} \times \gamma\mathbf{B} \tag{11}$$

Assuming the external magnetic field follows the z direction, the observed net magnetization will spiral down to a given angle, typically set at 90^{o} to the transversal plane, on which the transversal component of the magnetization, M_{xy} can be measured. The magnetization cannot be directly measured in the longitudinal direction because of the presence of the external magnetic field. In Equation 12, the M_{xy} component is treated as a complex signal for convenience, as the spins rotate in the xy plane.

$$M_{xy}(t) = M_0 e^{-i\omega_0 t} \tag{12}$$

When the pulse stops, the magnetization returns to the equilibrium state undergoing relaxation, during which a coil can be used to detect the induced current from the transversal component of the magnetization. Relaxation can be described according to two independent processes at different rates.

The longitudinal (T1 or spin-lattice) relaxation, describing the recovery of the longitudinal component M_z of the net magnetization to its equilibrium value, due to spins returning to a lower energy state. This is associated with an exchange of energy between the spins and surrounding media or material, hence the name spin-lattice, which depends on the nature of the sample. The rate at which the signal is recovered can be described by the time constant T1 (Equation 13).

$$M_z(t) = M_z(0) + M_0(1 - e^{-\frac{t}{T_1}}) \tag{13}$$

The transversal (T2 or spin-spin) relaxation describes the decay of the transversal magnetization component M_{xy}, called spin-spin because of local interactions with magnetic fields of surrounding spins. The magnitude of the transversal component is given by the sum of many individual spin magnetizations precessing in phase. Given the thermal random walk of molecules, a spin will experience local variations in the magnetic field, eventually leading to dephasing. The magnetization of spins oriented in opposite directions will cancel out and the transversal signal is eventually lost.

The rate at which the decay in magnetization occurs is given by the time constant T2 (Equation 14). The (faster) transversal relaxation with additional spin dephasing caused by magnetic inhomogeneities is referred to as T2-star (T2*).

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}} \tag{14}$$

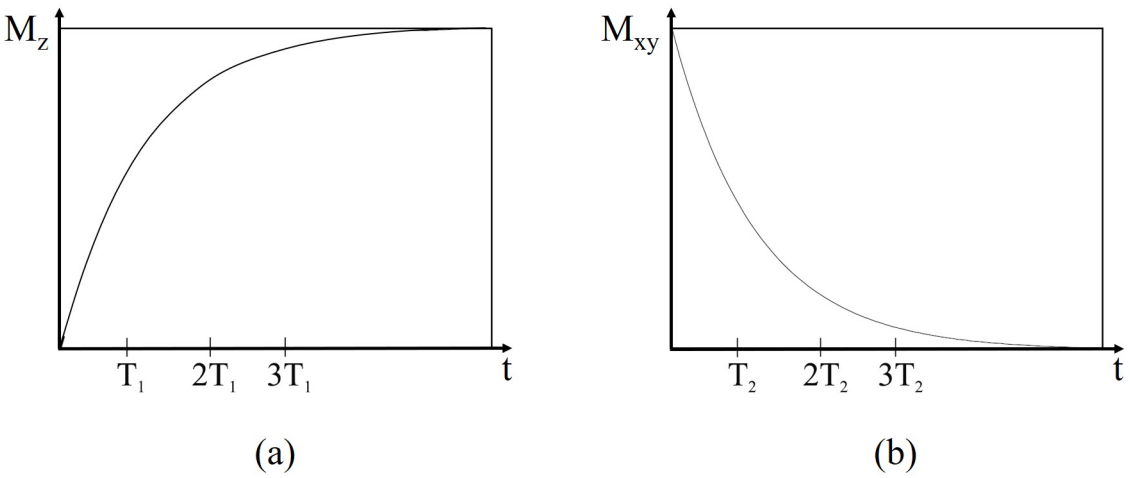


Figure 8: In this figure, the behaviour of the magnetization components is depicted during the relaxation processes, after the magnetization is tipped to the transversal plane during excitation, assuming an external magnetic field in the z direction. In (a), the recovery of the longitudinal magnetization M_z towards the equilibrium value M_0, due to spin-lattice interactions. The rate is characterised by the time constant T1. In (b), the loss of transversal magnetization, due to spin-spin interactions. The rate is characterised by the time constant T2. Figure adapted from [58].

Longitudinal and transverse relaxation processes are described by time constants T1 and T2, respectively, which define the time it takes to recover and lose, respectively, a portion

of magnetization and that depend on each compound or body tissue sampled (Figure **8**). Therefore, by choosing the time we let the longitudinal magnetization recover (repetition time or TR) and at which time to induce the magnetic resonance signal that will be measured in the transverse plane (echo time or TE), we can obtain images showing higher or lower signal difference based on the different T1 and T2 relaxation times of different tissues (giving so-called T1-weighted [T1w] and T2-weighted [T2w] images, while proton density aims to measure maximal signal from the number of spins in each region).

So far, it has been described where the measured signal comes from, but not how it becomes an actual image. In the receiving coil, an oscillating current is generated proportional to the number of spins, by the oscillating magnetic field caused by precession during relaxation from all the excited sample together. By applying a gradient field over the background magnetic field that changes the precession frequency with spatial location along the magnetic field axis, we can effectively select a plane that will be excited at a time.

Similarly, spatial gradients that slightly change the frequency and phase of precession on the other axes within the plane can be used. In Equation 15, the received signal $S_{receive}$ is described as a sum of the transveral magnetization component m_{xy} for every position x,y in a plane, which depends on the phase ϕ, which varies with time t.

$$S_{receive}(t) = \iint_{x,y} m_{xy}(x,y)e^{-\phi(x,y,t)}dxdy \tag{15}$$

$$\phi = \int_0^t \Delta\omega(x,y,\tau)d\tau \tag{16}$$

$$\omega(x,y,t) = \gamma(B_0 + G_x(t)x + G_y(t)y) \tag{17}$$

The frequency can be seen as the time rate of change of phase, or conversely, phase as the integral of frequency (Equation 16). By applying gradients that vary the magnetic field with spatial position, for instance $G_x = \frac{\partial B_x}{\partial x}$ in the x direction, we induce a change in frequency related to each x position, and similarly in the y direction (Equation 17); if we let the gradient over time, we keep changing the phase (Equation 18), thereby encoding the signal with varying frequencies ans phases that correspond to spatial locations.

$$\phi = \gamma \int_0^t \left[B_0 + G_x(\tau)x + G_y(\tau)y\right] d\tau = \omega_0 t + \gamma \left[\int_0^t G_x(\tau)d\tau\right] x + \gamma \left[\int_0^t G_y(\tau)d\tau\right] y \tag{18}$$

If we let $k_x = \frac{\gamma}{2\pi}\int_0^t G_x(\tau)d\tau$, then along time t we are moving in the k_x direction (Equation 19) in the frequency domain. The signal in the receiver is then modulated with this information, from which the frequency components can be extracted. These are arranged in a matrix called k-space, that corresponds to the Fourier transform of the desired image in the spatial domain. Therefore, an image can be obtained by applying the inverse Fourier transform.

$$S_{receive}(t) = \iint\limits_{x,y} m_{xy}(x,y)e^{-i\omega_0 t}e^{-i2\pi[k_x x + k_y y]}dxdy \tag{19}$$

There are multiple steps involved in the generation of images from the frequency information, with different approches specific of the acquisition, such as fast dynamic images that acquire only a portion of the k-space to save time and reconstruction algorithms that perform filtering and other signal processing. Those go beyond the scope of this work.

But it is important to note that there exist numerous MRI pulse-sequences and derived parametric maps, with some of them exclusively used for research purposes or highly specialized areas of study. This is in part because research MRI scanners and software allow a higher degree of customization of pulse-sequences than the clinical counterparts.

One drawback of MRI in general is that the measured intensity, although proportional to the number of spins, cannot be directly interpreted with a physical meaning. Furthermore, each scanner has a specific calibration, field homogeneity and electronics, where the induced signal is amplified, filtered and transformed, making it difficult to compare and generalize image intensities among scans.

A common approach to circumvent that issue is to consider the intensity relative to some other signal, usually from a given tissue of reference, or deriving quantitative maps. For some MR images, it is possible to derive interpretable metrics that reflect body tissue properties of interest and that allow their quantification and comparison. Those sequences deserve special attention, as they are powerful tools in the management of cancer. These sequences include the Dixon water and fat separation, perfusion and diffusion MRI and will be introduced next.

1.3.2 Anatomical images

The most basic anatomical sequences may be T1w and T2w images, but still essential for radiological interpretation and for building more complex MRI experiments. Anatomical

images allow to identify known structures of the body, as well as locate injuries or lesions, and provide a reference to spatially align other sequences and, consequently, these sequences are always at the core of any MRI study.

Pathological tissues can often be seen in anatomical images as regions with abnormal intensity values relative to their healthy counterparts, sometimes with a specific pattern, shape or appearance that help radiologists identify and diagnose them.

An intra-venous contrast agent can also be employed to obtain additional information. Contrast products like gadolinium show paramagnetic properties which will favour the spin relaxation processes when passing though the organ vasculature under study. This way, with respect to the same T1w, contrast-enhanced T1w will show hyperintense areas where the contrast has infiltrated, within vessels and extravascular spaces. That information can be very valuable to identify structures and characterize cancer based on the vascularity, i.e. highly irrigated tumours and abnormal vessels may be a cancer phenotype associated with higher aggressiveness.

1.3.3 Dixon technique

The contribution of both water and fat to MRI signal is an important aspect to consider in the imaging of organs and diseases, including cancer. In the interpretation of MRI of protons, it is assumed that a majority of signal comes from water. Therefore, the signal arising from fat, if not properly taken into account, can confound or significantly hinder the detection of lesions or the differentiation between benign and malignant tissues. To this end, there exist several common methods of fat signal suppression, such as inverse recovery and fat saturation techniques. But since malignant tissues do not generally accumulate fat, acquiring the information on the fat content could also be valuable, especially in organs or specific conditions that contain adipose tissues.

The sequences based on the Dixon method [59] take advantage of the chemical shift between protons in water and fat molecules. Chemical shift refers to the slight differences in the resonance frequency of nuclei in different molecules caused by electron clouds of the molecule atoms responding to the external magnetic field with opposing local magnetic fields, an effect called shielding. Those differences will depend on the molecule where the nucleus resides, making it possible to detect different compounds in a sample by means of measuring the frequencies, forming the basis of nuclear magnetic spectroscopy.

For imaging of the body, most protons reside in either water or fat molecules. Knowing the difference in frequency between fat and water molecules, Dixon methods are based on acquiring images at precisely the times when the spins of water are in phase (IP) with those of fat and when they are completely out of phase (OOP), then theoretically pure water signal W and fat signal F images can be derived from the magnitude and phase of those complex images (Equation 20 and Equation 21).

$$IP = |W + F| \tag{20}$$

$$OOP = |W - F| \tag{21}$$

The Dixon methods provide a means to separate signals for water and fat and calculate water and fat fraction maps, which is advantageous for diseases or conditions involving fat in tissues such as the liver, muscle and bone, but also pancreas, spleen, testes and others [60]. The main challenge when separating water and fat components is the water-fat swap artifacts as a consequence of local field inhomogeneities. Modern Dixon techniques with additional acquisitions, phase unwrapping methods and even deep learning methods have been proposed to solve this issue [61].

1.3.4 Perfusion MRI

The main focus of study of perfusion techniques is the properties of the vasculature in the tissues. The most widely used sequences are dynamic susceptibility contrast (DSC) and dynamic contrast enhanced (DCE), both consisting of a dynamic acquisition during the passage of a contrast agent dose, called bolus. In DCE, T1w images are acquired over a few minutes, where slower permeability of the contrast into tissues can be observed. In DSC, T2*w images are acquired over a short time, focusing on the one-time passage of the bolus through the vessels.

The dynamic sequence can then be mathematically modelled, estimating measures of the blood flow and blood volume, among others from DSC; and permeability, vascular and tissue fractions from DCE. Specifically, signal S from a DSC experiment is proportional to the T2* decay as follows:

$$S \propto e^{-\frac{TE}{T2^*}} \tag{22}$$

With $\frac{1}{T2^*} = R2^*$ representing the relaxation rate. In a dynamic acquisition, a signal S(t)

is acquired at discrete time points t during the passing of the contrast agent through the tissues, which favours relaxation. The concentration of contrast agent C(t) can be assumed to be proportional to the change in relaxation rate $\Delta R2^*(t)$, which can be obtained as follows:

$$C_t \propto \Delta R2_t^* = R2_t^* - R2_0^* = -\frac{TE}{T2^*} ln \frac{S_t}{S_0} \tag{23}$$

The curve of $\Delta R2^*$ over time shows a smooth, momentary loss of signal corresponding to the passage of contrast (fig:dsc). The relative cerebral blood volume (rCBV) would be proportional to the area under the $\Delta R2^*$ curve. The percentage of signal recovery (PSR) can be defined as the ratio of the recovered signal after contrast has passed over the baseline signal, as follows

$$PSR = \frac{S_r - S_{min}}{S_b - S_{min}} * 100 \tag{24}$$

Where S_{min} is the minimum peak point reached by the perfusion curve, ${}_Sb$ is the baseline perfusion signal before contrast bolus arrival and S_r is the recovered signal after bolus passage. The rCBV and PSR are both common parameters in the study of DSC perfusion MRI in cancer [62–64].

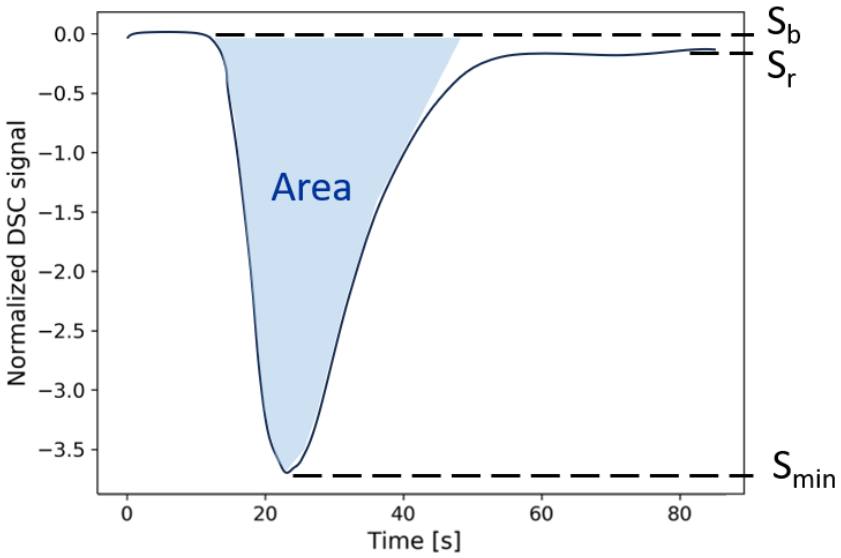


Figure 9: Dynamic T2*w signal from DSC MRI, with intensity normalized to the white matter and time in seconds. The contrast medium causes a temporary loss of signal. The area is proportional to the blood volume in the vasculature and the recovered signal Sr relative to the minimum and baseline signals Smin and Sb are related to vascular permeability and leakage.

Given that one important mechanism for the development and growth of malignant cells is the generation of new, distinct vasculature and the disruption of normal vessels and blood

flow, perfusion MRI is a commonly acquired sequence, especially for breast cancer and brain pathologies. The perfusion-derived parameters have shown to have diagnostic value and to differentiate between different tumour grades and types. In general, increased blood flow and volume are indicative of higher angiogenesis, which would point towards malignant processes, while a reduction in those could be a sign of working treatment [65].

1.3.5 Diffusion sequences

Diffusion MRI takes advantage of the sensitivity of MRI to interactions and motion of water protons at the molecular level. Similar to the spatial localization gradients of MRI, a diffusion gradient can be used to label water protons that will experience Brownian motion within different compartments of the living tissues (e.g.: within cells, in the extracellular matrix, along axons, etcetera) for a given time. After that time, an opposite gradient would reverse the effects of the first gradient should the molecules be completely static, except the protons experience random walk motion. Some water molecules in a confined space would travel a small overall distance and therefore most of the signal can be recovered in that case. Others, however, are less restricted by obstacles and can diffuse further away, which means a larger general dephasing of spins and an overall signal loss in that imaged region. Given that the diffusivity of the water causes a loss of signal, diffusion MRI enables estimating the diffusivity within the body regions, with units of area over time. Furthermore, diffusion data can be acquired sensitive to different directions, giving rise to diffusion tensor imaging (DTI), mainly explored in the central nervous system for neural tractography.

That loss of signal will be increased by the magnitude, duration and spacing of the diffusion gradients, yielding diffusion-weighted imaging (DWI). The parameters defining the gradients along with the gyromagnetic ratio are combined for convenience into a single factor called the b-value b, with inverse units to those of the ADC [66]. Relative to the baseline signal S_0, the signal decay due to diffusion as a function of the b-value is proportional to the negative exponential of the decay rate given by diffusivity D (Equation 25).

$$S = S_0 e^{-bD} \tag{25}$$

The diffusivity is typically averaged over three orthogonal diffusion directions as a total apparent diffusion coefficient (ADC), as seen in Figure 10. For this model, acquiring at least two images with different b-value allows to find the ADC, which holds for a range of $b \sim [50, 1000] s/mm^2$. In very short and very long diffusion time regimes, additional effects may be taken into account that deviate from a pure exponential, such as vascular pseudo-

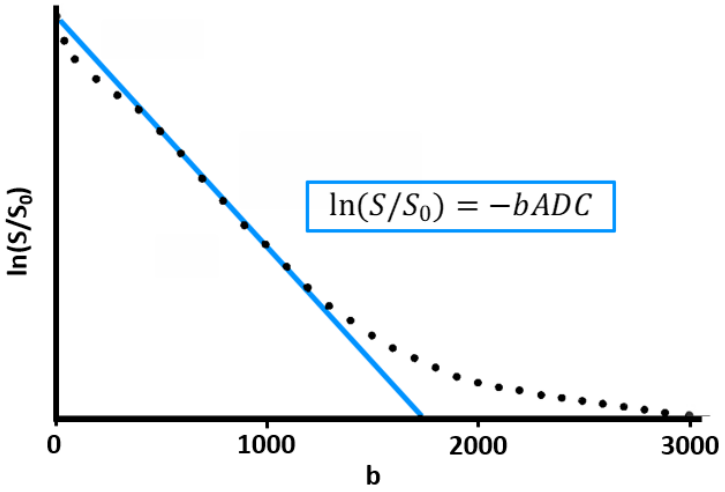


Figure 10: The decay in the diffusion T2w signal, in logarithmic scale, relative to no diffusion weighted, with respect to the b-value. The decay can be approximated as a mono-exponential with the rate given by the ADC.

diffusion and non-Gaussian diffusion. Moreover, advanced diffusion experiments allow for the modelling of tissue diffusivity to enable the association of biological properties to the acquired MRI signals. Assuming a certain tissue microstructure, the signal can be modelled as a mixture of a number of compartments that have distinct water diffusivity, such as vascular, intracellular and extracellular components, e.g., intravoxel incoherent motion (IVIM) [67]. Those methods remain mostly investigational, although have shown promising results for clinical applications [68–71].

In particular, the ADC has been extensively explored, especially in cancer disease, as it is simple to obtain and provides an estimate of the water diffusivity within the tissues. A very common trait of many tumours is that the cellularity increases as a result of a high growth rate of malignant cells. Therefore, the diffusivity is expected to be lower in areas with highly packed cells and so it is for the ADC, generally regarded as a proxy for cellularity [72]. The role of the ADC in cancer comes across multiple organs of the body and spans all main clinical imaging tasks, namely, diagnosis, differentiation, treatment response prediction, monitoring and prognosis and DWI is routinely acquired in neurological applications, as it helps characterizing tissues [35, 73]. However, the ADC is a global estimate of diffusivity, which is sensitive not only to cellularity, but also to other tissue properties, i.e. cells or regions with different intrinsic diffusivity, like oedema, haemorrhage, necrosis, fat, different cell sizes and the specific structural arrangement of the tissues [74–76].

1.3.6 Multi-parametric MRI and the special role of neuroradiology

Despite the fact that MRI allows acquiring multiple contrasts, multiparametric MRI may not be possible to apply in every case. Besides possible redundant information, the total scan time is a crucial factor when designing MRI sequences and acquisition protocols for the clinical practice. An MRI scan of a joint may take ten minutes, while a whole-body scan may take sixty minutes. Thus, scan time should be minimized whenever possible to prevent patient discomfort, maximize data acquisition and efficiency. That is to say, MRI studies usually include just the specific sequences that have been shown useful for a given exploration. Likewise for cancer imaging, because the disease can manifest in a wide range of forms and body organs, different acquisitions and spatial and time constraints would apply in every case.

It is important to mention here that there may be an exception to the rule. Imaging of the brain typically comprises multiple acquisitions, including T1w, with and without contrast, T2w, proton density, perfusion and diffusion quantitative sequences, among others. The importance of the brain for the function of the human body, its differentiated soft tissues and functional areas, its contained size and regular shape all enable acquiring rich datasets where MRI shines, allowing the subsequent development of studies and methods that pave the way for their use in other body parts. Such is the position of brain imaging that it stands in its own medical field of neuroradiology.

That implies that many of the common MRI techniques and methods of acquisition, processing and analysis had originally been developed for brain applications, which later got adapted to other body regions. Some examples include registration [77, 78], segmentation [79–81] and imaging and modelling [67, 82–85] methods. As many of the mentioned acquisitions are common on multiple body regions, the latest methods developed on brain today could be applied in the imaging of the body tomorrow. Even though imaging of each body organ is ultimately specialized to specific MRI sequences that optimize the scan time, each and every sequence can be a valuable source of information. Multiparametric MRI has shown the potential to build supporting and processing tools [86], as well as better evaluating cancer disease in clinical guidelines incorporating multi-modal images [21, 25], pointing at the importance of acquiring and exploring comprehensive datasets for advancing MRI imaging in healthcare.

CHAPTER 2

Hypothesis and Objectives

2.1 Hypothesis

The clinical management of cancer disease relies on radiological images for critical tasks like diagnosis, prognosis and assessment of response to the treatment. The current reference methods of image analysis involve visual inspection, manual annotations and a degree of subjectivity which inherently limit the characterization and evaluation of cancer for those three tasks.

The development of computerized features from medical images has been an active research topic for years, especially within the field of neuroradiology, showing the potential of MRI for characterizing cancer disease. However, further development and validation of novel imaging markers are still needed to address the multiple clinical needs related to a complex disease.

For this work, it was hypothesized that MRI-derived markers could improve over current assessments of the medical images in three identified scenarios in the management of cancer disease, specifically:

- The residual tumour after surgery can be objectively quantified from MRI and be useful for estimating prognosis in patients with malignant primary brain tumours (i.e., brain glioblastoma). The time elapsed from surgery to MRI can affect or confound the estimation of residual tumour.

- The use of the complete dynamic profile of perfusion MRI data can be used for the differential diagnosis of brain malignancies, and it can improve over existing methods.

- The metrics derived from whole-body MRI in bone metastases associate with the response of the patients to the treatment. The MRI-derived metrics of bone metastases can capture underlying biology traits of tumours.

2.2 Objectives

To test the hypotheses, three objectives were defined:

- **Objective 1:** To extract computerized metrics from post-surgical MRI and study their prognostic value in patients with brain glioblastoma, and to explore the effect of the time elapsed from surgery to the MRI scan on the prognosis evaluation.
- **Objective 2:** To develop and validate a method to process all time-points from dynamic perfusion MRI for brain tumour diagnosis, differentiating between the three most common brain malignancies and comparing the performance with existing metrics.
- **Objective 3:** To study MRI-derived data during treatment of bone metastases for the evaluation of response, exploring the combination of imaging features and their relation to histological evidence.

CHAPTER 3

Supervisor Report

Dr. Raquel Pérez López, Principal Investigator of the Radiomics Group, Vall d'Hebron Institute of Oncology; and Dr. Carles Majós Torró, Medical Doctor in the Department of Radiology, University Hospital of Bellvitge, as directors of the doctoral thesis entitled "*Development and validation of cancer markers based on multiparametric MRI and machine learning*" and, in accordance with the provisions of art. 35 of the Doctorate regulation at the University of Barcelona, state the following:

REPORT

In this thesis, Alonso Garcia Ruiz has focused on enhancing cancer disease assessment through the application of machine learning methods to leverage MRI for improved diagnosis and treatment response evaluation in cancer patients.

The hypothesis posits that current prognosis, diagnosis, and treatment response determination can benefit from recent developments in quantifiable metrics extracted from medical images. To explore this, three comprehensive studies were conducted, proposing novel methods to enhance brain tumour diagnosis, evaluate residual brain tumours post-surgery, and improve response assessment in patients with advanced cancer and bone metastases. The results demonstrate significant contributions to the field, particularly from clinical and practical perspectives.

These studies have been published in high-impact scientific journals, including one quartile 1 journal (Scientific Reports, IF 3.8) for Study 1 and two decile 1 journals (Cell Report Medicine, IF 11.7, and European Urology, IF 25.3) for Studies 2 and 3, meeting the University of Barcelona's criteria for thesis submission as a compilation of scientific articles.

In all studies, Alonso appears as first author, although he declares equal contributions with fellow authors in Study 1 and Study 2. In Study 1, Alonso was responsible for data inspection, image processing, feature extraction, definition of new features, statistical analyses and manuscript drafting. In Study 2, Alonso was in charge of data collection from public databases, data inspection, image processing, building, training and testing of classification models, statistical analyses and manuscript drafting.

None of the studies in this thesis were used implicitly or explicitly in any other thesis.

Throughout these investigations, Alonso Garcia Ruiz has shown adeptness in overcoming challenges and adaptability in implementing new methodologies, such as deep learning techniques. His research process has been marked by thorough literature review, innovative technique development, critical thinking, and rigorous methodology.

This thesis represents the achievement of primary objectives, yet Alonso Garcia Ruiz has also contributed to other activities, co-authoring additional publications and presenting at conferences, demonstrating interdisciplinary interest and a continuous pursuit of knowledge. Considering the aforementioned achievements, we hereby express our approval to be considered for a PhD in Biomedicine from the University of Barcelona.

Signed

Raquel Pérez López

Carles Majós Torró

CHAPTER 4

Articles

4.1 Study 1

Precise enhancement quantification in post-operative MRI as an indicator of residual tumor impact is associated with survival in patients with glioblastoma

Scientific Reports (2021), Vol 11, Page 95. **Impact Factor (quartile)**: 3.8 (Q1)

Authors and affiliations: Alonso Garcia-Ruiz[1], Pablo Naval-Baudin[2], Marta Ligero[1], Albert Pons-Escoda[2], Jordi Bruna[3], Gerard Plans[3], Nahum Calvo[2], Monica Cos[2], Carles Majós[2,3], Raquel Perez-Lopez[1,4].

1. Radiomics Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain.
2. Department of Radiology, Institut de Diagnòstic Per La Imatge (IDI), Bellvitge University Hospital, Barcelona, Spain.
3. Neuro Oncology Unit, Institut d'Investigacio Biomedica de Bellvitge (IDIBELL), Barcelona, Spain.
4. Department of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain.

This study corresponds to **objective 1** of the thesis. This work aimed to find markers of prognostic value in the post-surgery MRI as indicators of remaining tumour in patients with brain glioblastoma. The study proposed a new metric and explored radiomics and multivariate regression models to associate them to patient survival. Additionally, the effect of the time elapsed between surgery to the MRI scan was also investigated.

The PhD candidate was in charge of data inspection, image processing, feature extraction, definition of new features, statistical analyses and manuscript drafting.

scientific reports

Check for updates

OPEN

Precise enhancement quantification in post-operative MRI as an indicator of residual tumor impact is associated with survival in patients with glioblastoma

Alonso Garcia-Ruiz[1,7], Pablo Naval-Baudin[2,7], Marta Ligero[1], Albert Pons-Escoda[2,3], Jordi Bruna[3,4], Gerard Plans[3,5], Nahum Calvo[2], Monica Cos[2], Carles Majós[2,3,8] & Raquel Perez-Lopez[1,6,8]✉

Glioblastoma is the most common primary brain tumor. Standard therapy consists of maximum safe resection combined with adjuvant radiochemotherapy followed by chemotherapy with temozolomide, however prognosis is extremely poor. Assessment of the residual tumor after surgery and patient stratification into prognostic groups (i.e., by tumor volume) is currently hindered by the subjective evaluation of residual enhancement in medical images (magnetic resonance imaging [MRI]). Furthermore, objective evidence defining the optimal time to acquire the images is lacking. We analyzed 144 patients with glioblastoma, objectively quantified the enhancing residual tumor through computational image analysis and assessed the correlation with survival. Pathological enhancement thickness on post-surgical MRI correlated with survival (hazard ratio: 1.98, $p < 0.001$). The prognostic value of several imaging and clinical variables was analyzed individually and combined (radiomics AUC 0.71, $p = 0.07$; combined AUC 0.72, $p < 0.001$). Residual enhancement thickness and radiomics complemented clinical data for prognosis stratification in patients with glioblastoma. Significant results were only obtained for scans performed between 24 and 72 h after surgery, raising the possibility of confounding non-tumor enhancement in very early post-surgery MRI. Regarding the extent of resection, and in agreement with recent studies, the association between the measured tumor remnant and survival supports maximal safe resection whenever possible.

Abbreviations

AUC	Area under the curve
CI	Confidence interval
DSC	Dynamic susceptibility contrast
EPMR	Early post-operative magnetic resonance
HR	Hazard ratio
IDH	Isocitrate dehydrogenase
KPS	Karnofsky performance status
LR	Likelihood ratio

[1]Radiomics Group, Vall d'Hebron Institute of Oncology (VHIO), 117 Natzaret, 08035 Barcelona, Spain. [2]Department of Radiology, Institut de Diagnòstic Per La Imatge (IDI), Bellvitge University Hospital, Barcelona, Spain. [3]Neuro-Oncology Unit, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Barcelona, Spain. [4]Department of Neurology, Bellvitge University Hospital, Barcelona, Spain. [5]Department of Neurosurgery, Bellvitge University Hospital, Barcelona, Spain. [6]Department of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain. [7]These authors contributed equally: Alonso Garcia-Ruiz and Pablo Naval-Baudin. [8]These authors jointly supervised this work: Carles Majós and Raquel Perez-Lopez. ✉email: rperez@vhio.net

PSR	Percentage of signal recovery
rCBV	Relative cerebral blood volume
ROI	Region of interest
OS	Overall survival
VOI	Volume of interest
W	Weighted

Glioblastoma is the most common primary brain tumor[1]. Although some treatments prolong survival, the prognosis of patients with glioblastoma is very poor, e.g. 2-year survival of 26.5%[2]. Standard of care therapy consists of maximum safe resection combined with adjuvant radiochemotherapy followed by chemotherapy with temozolomide. The extent of tumor resection is a relevant prognostic factor in this patient population[3–9]. Albert et al.established that the extent of resection should be evaluated with a magnetic resonance imaging (MRI) scan performed shortly after surgery, as inflammatory reparative changes can result in benign non-tumor contrast enhancement that can be misinterpreted as tumor remnants[10]. This is widely accepted in clinical practice, and most guidelines recommend performing early postoperative magnetic resonance (EPMR) to evaluate the extent of resection. Some guidelines suggest performing the EPMR scan within the first 72 h after surgery, while others are more restrictive and state with 48 h[11–13]. However, these recommendations rely on the experience and opinion of experts, rather than on formal clinical evidence. To the best of our knowledge, an in-depth evaluation of the optimal time range to perform the EPMR scan has not yet been described. An objective numerical assessment of the residual tumor enhancement to support this would allow identification of the optimal timing to evaluate the enhancement in the EPMR, and thus, the timing at which it is most informative for patient prognosis.

There is no consensus about the extent of resection to stratify prognostic groups. Some studies suggest an "all-or-nothing" approach[8] in which complete resection should be achieved to improve patient survival. However, others suggest that it is possible to stratify several prognostic groups according to ranges of tumor resection[3–7,9,10,14,15], namely < 75%, 75–95%, 95–100% and total resection.

An objective numerical evaluation of the residual tumor would allow analysis of a continuous range of values instead of subjective qualitative assessment, and help identify the best approach in regards to therapy and prognosis, while also being readily applicable and comparable between centers and studies. In this study, we focus on the enhancing residual tumor only. While there may be also non-enhancing residual tumor, a recent study has demonstrated an association between the post-operative enhancement volume and survival[16]. Therefore, it may be possible to identify additional prognosis value using more extensive image analysis of the enhancing tumor remnant in the post-surgical MRI. To this end, we focused on the post-surgical MRI and investigated radiomics and perfusion imaging features as an alternative evaluation to the extent of resection, which requires paired pre- and post-surgical MRI.

Image processing and radiomics extract quantifiable features from body tissues using characterization algorithms on the image spatial data. Radiomics is a lively topic of research that is shaping medical image assessment and interpretation, providing crucial information regarding tumor biology including glioblastomas[17,18].

In this study, our goals were to develop a tool to facilitate the quantification of enhancing post-surgery residual tumor in patients with glioblastoma and analyze whether precise quantification of the area of contrast enhancement by image processing and radiomics could improve the patient´s prognostic evaluation. With this objective data, we also assessed the time range from surgery to the EPMR scan for an optimal association of residual tumor and overall survival (OS), to provide quantitative evidence for the optimal time range to perform the EPMR scan after surgery.

Results

A total of 144 patients were included in the study (92 [64%] men, 52 [36%] women); median age of all patients was 59 years (range 20–77 years). Population demographics are described in Supplementary Table S1.

Correlation of post-operative enhancement with survival. The enhancement thickness was calculated as the 3D distance transform of the segmented tumor, and mean and maximum values were determined. The mean and maximum enhancement thickness correlated with OS with a hazard ratio (HR) of 1.98 (95% CI 1.36–2.90, $p < 0.001$) and 1.11 (95% CI 1.05–1.17, $p < 0.001$), respectively (Table 1). Mean and maximum thickness also correlated significantly with progression-free survival (PFS) with HR of 2.01 (95% CI 1.35–3.12, $p < 0.001$) and 1.08 (95% CI 1.03–1.14, $p = 0.004$), respectively.

The survival rate of patients with a mean thickness of 1.4 mm or more was 47% after 12 months, whereas survival of patients with a mean thickness less than 1.4 mm was 78% ($p < 0.001$, Fig. 1, left panel). Similarly, the survival rate of patients with a maximum thickness of 8 mm or more was 55% after 12 months, whereas survival of patients with maximum thickness of less than 8 mm was 80% after 12 months ($p = 0.0012$, Fig. 1, right panel).

The volume of enhancement was also associated with OS, with an HR of 1.04 (95% CI 1.01–1.07, $p = 0.009$), however there was no significant association with PFS (HR 1.02, 95% CI 0.99–1.05, $p = 0.19$). A previously reported threshold of 12 mL[16] resulted in unbalanced subpopulations: patients with enhancement volume greater than 12 mL (n = 10) had a survival rate of 20% at 12 months, while patients with volume less than 12 mL (n = 134) had a survival rate of 75% at 12 months (Supplementary Fig. S1). Despite the different survival rates, Kaplan–Meier analysis did not yield a significant difference ($p = 0.1$).

Survival prediction for time between surgery and the EPMR scan. We analyzed the population in terms of the time elapsed between surgery and the EPMR scan according to the following groups: within 24 h (n = 26), from 24 up to 48 h (n = 51), from 48 up to 72 h (n = 42), and 72 h or more (n = 25). The mean and

Variable	Group	Distribution			Cox regression		
		n	Mean	SD	HR	95% CI	p-value
Mean thickness (mm)	All	144	1.18	0.44	1.98	1.36–2.90	< 0.001
	< 24 h	26	1.25	0.57	1.19	0.59–2.39	0.44
	24–48 h	51	1.18	0.39	3.00	1.24–7.24	0.03
	48–72 h	42	1.16	0.44	3.30	1.30–8.40	0.03
	≥ 72 h	25	1.21	0.51	2.40	0.74–7.72	0.16
Maximum thickness (mm)	All	144	7.16	3.27	1.11	1.05–1.17	< 0.001
	< 24 h	26	6.98	3.54	1.02	0.91–1.14	0.56
	24–48 h	51	7.41	3.23	1.21	1.08–1.36	0.005
	48–72 h	42	6.92	3.05	1.19	1.04–1.36	0.02
	≥ 72 h	25	7.16	3.81	1.08	0.96–1.21	0.21

Table 1. Univariate Cox model of the enhancement thickness (mean and maximum) for overall survival, for the entire population and by subgroup according to time elapsed between surgery and the MRI. P-values were adjusted for multiple test comparison. *N* number of observations, *SD* standard deviation, *mm* millimetres, *h* hour, *HR* hazard ratio, *CI* confidence interval, *LR* likelihood ratio.

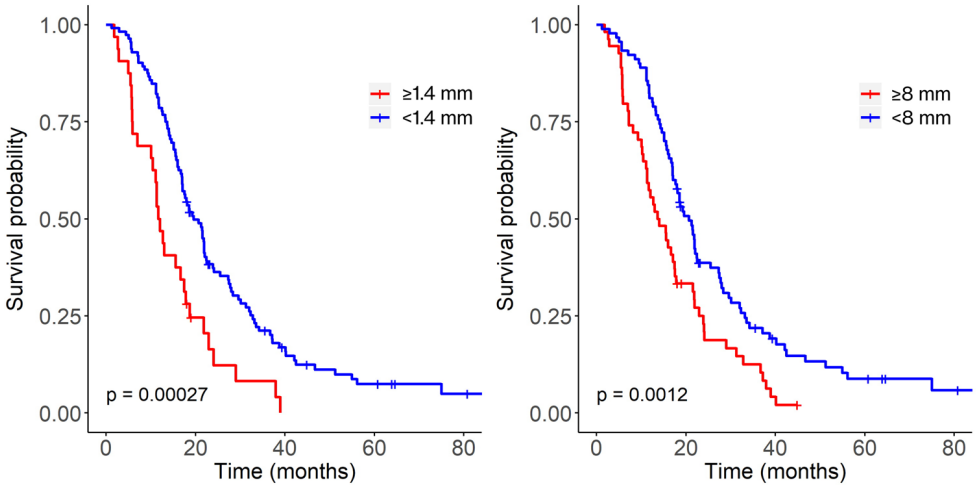


Figure 1. Kaplan–Meier survival curves for the enhancement thickness analysis, according to high and low mean (left) and maximum (right) thickness. Censored data are indicated with tick marks.

maximum thickness of enhancement were significantly associated with survival in patients with an EPMR scan performed between 24 and 72 h after surgery ($p < 0.05$, adjusted for multiple test comparisons, Table 1). There was no association between thickness and OS when the EPMR was performed within 24 h of surgery or more than 72 h after. In Supplementary Fig. S2, the HR and the confidence intervals of each subgroup are shown for comparison, and the Kaplan–Meier curves are shown in Supplementary Fig. S3.

Prognostic value of perfusion sub-analysis. Dynamic susceptibility contrast (DSC) data were available for 113 patients (113/144, 78%). When the entire population was analyzed, patients with a relative cerebral blood volume in the 99th percentile (rCBV-99) above 8.26 had worse survival ($p = 0.05$, Supplementary Table S2), with a survival rate of 50% after 12 months, compared to a 75% rate in patients below this threshold (Supplementary Fig. S4). A maximum percentage of signal recovery (PSR) above 112% defined patients with lower survival ($p < 0.001$, Supplementary Table S2), with a 50% survival rate after 12 months, compared to an 80% rate for patients below the threshold (Supplementary Fig. S4).

We also analyzed the perfusion data in terms of the time between surgery and the EPMR scan, according to within the first 24 h (n = 22), between 24 and 48 h (n = 37), between 48 and 72 h (n = 33) and 72 h or later (n = 21). rCBV-99 predicted survival when the EPMR scan was performed 24 to 48 h post-surgery ($p = 0.004$ adjusted for multiple test comparisons, Supplementary Table S2); for patients with a rCBV-99 above 7.93, the survival rate was 20% after 12 months, compared to 70% for those below this threshold (Supplementary Fig. S4). The maximum PSR predicted survival when the EPMR scan was performed 24 to 72 h post-surgery ($p < 0.001$, adjusted for multiple test comparisons, Supplementary Table S2), with a survival rate of 20% after 12 months for patients above a 111% threshold and a rate of 80% for patients below this threshold (Supplementary Fig. S4). Neither the rCBV-99 nor PSR had prognostic value when the EPMR scan was performed within 24 h of surgery. When

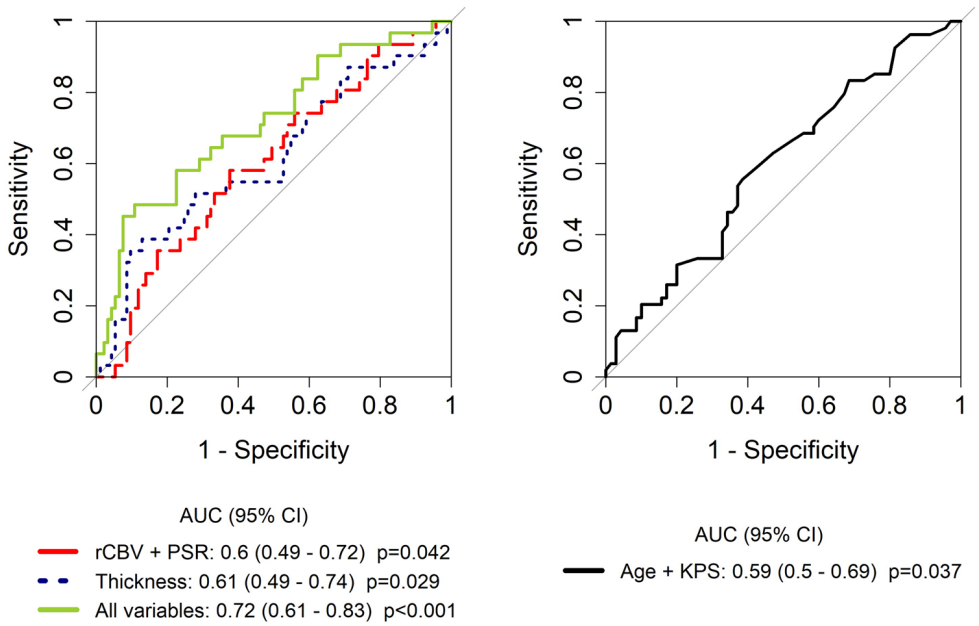


Figure 2. ROC of different prognostic models: perfusion variables (99th percentile cerebral blood volume [rCBV], maximum percentage of signal recovery [PSR]), mean thickness of enhancement and all variables (perfusion, thickness, age, Karnofsky performance status [KPS] and radiomics) together for predicting survival (left panel). The model with clinical variables age and KPS is shown in the right panel for clarity.

the scan was performed more than 72 h after surgery, the rCBV-99 showed significantly different survival rates ($p = 0.011$), albeit with unbalanced subgroups.

Prognostic value of the radiomics signature. The population with measurable enhancement segmentation (129/144, 90%; see "Methods" for details) was analyzed to determine the prognosis of the clinical endpoint of OS ≥ 2 years (long; 36/129, 28%) or OS < 2 years (short; 93/129, 72%). The population was randomly split into training (92/129, 70%) and test (37/129, 30%) sets, with balanced distribution of survival in both sets (26/92, 28% of long survivors in the training; 10/37, 27% of long survivors in the test set) (Supplementary Table S3).

Twelve radiomics variables were selected in the training set using minimum-redundancy-maximum-relevance and stepwise regression (Supplementary Table S4). This radiomics signature predicted short and long survival groups with an AUC of 0.73 (0.60–0.86 $p < 0.001$) in the training set and an AUC of 0.71 (0.55–0.88 $p = 0.01$) in the test set. Receiver operating characteristic (ROC) curves and Kaplan–Meier curves of the training and test sets are presented in Supplementary Fig. S5.

Combining quantitative imaging and clinical data for predicting survival. The multivariate logistic model including all imaging features (mean enhancement thickness, DSC, radiomics) and clinical data (age, postoperative KPS) yielded the highest prognostic capacity for predicting long and short survival (AUC 0.72, 95% CI 0.61–0.83, $p < 0.001$, Fig. 2). In the multivariate Cox model including all of these parameters, the mean thickness and age were retained as independent prognostic factors (Supplementary Table S5).

Discussion

Radiological visual assessment of the remaining tumor after surgery is the standard of care in oncology clinical practice, and should be performed with an MRI rapidly after the surgical procedure. The extent of resection is a well-known prognostic factor in glioblastomas[4,5,8], though inter-reader reproducibility is limited and hinders comparisons between centers and studies.

In this study, we identified a method to facilitate the quantification of the remaining tumor using a processing pipeline of multi-sequence MRI scans. By automatic registration and subtraction of T1w and T1wC images, it is possible to optimally isolate the enhancing areas from confounding post-operative changes. This method was also conceived by Ellingson et al.[16], although they reported only the enhancement volume and did not analyze the influence of time to the EPMR, to measure post-surgery residual enhancing tumor. Moreover, we analyzed and internally validated a multivariate prognostic model including quantified residual tumor, perfusion, radiomics and clinical variables in an effort to improve the prognostic performance of residual tumor enhancement.

In our population, the quantification of the residual enhancement thickness after surgery shows a continuous association with OS and PFS (HR of 1.98 and 2.01 respectively, for mean enhancement thickness) that outperforms the subjective assessment of a dichotomous thin-or-thick tumor remnant[3]. Our findings also showed that the enhancement thickness had a stronger association with survival compared to the enhancement volume, suggesting the value of further investigation of this metric.

The combination of the enhancement thickness, perfusion, radiomics and clinical data into a predictive model showed slightly better performance (AUC 0.72) for distinguishing patients with short and long survival compared to thickness alone, compared to the univariate perfusion or the clinical models. When analyzed in separate groups, age and KPS yielded the lowest AUC (0.59), followed by rCBV and PSR (0.60), thickness (0.61) and radiomics (0.71). Therefore, in our population, radiomics variables performed better when applied in a dichotomous patient longer-or-shorter survival model, while the other variables added only marginal value. However, when correlating with specific patient survival, enhancement thickness alone demonstrated meaningful prognostic value, with higher mean thickness (as a continuous variable) associated with poorer survival (HR of 1.98 [95% CI 1.36–2.90, $p < 0.001$]). Both the enhancement thickness and the radiomics signature can be automatically calculated from the enhancement mask. Although further work is needed to facilitate the implementation of these assays in clinical practice, these results show the potential application of quantitative data from EPMR T1w and T1wC for supporting medical decisions.

We also explored the impact of the time between surgery and the MRI scan for quantifying the residual tumor. According to the National Comprehensive Cancer Network guideline, the early post-surgery MRI scan should be performed within the first 72 h after surgery[19]. Other guidelines are more restrictive and suggest that the scan be performed within the first 24 to 48 h after surgery[12,13,20]. The rationale for earlier imaging is that after 72 h post-surgery, inflammatory reparative changes result in benign non-tumor contrast enhancement that could be misinterpreted as tumor remnant[10]. Correspondingly, in our study we found no significant association between the measured enhancement and OS in the group of patients with EPMR scans performed more than 72 h after surgery.

Interestingly, we found no association between the enhancement thickness and survival when the MRI was acquired within the first 24 h after surgery. Although the sample size of this subpopulation was small ($n = 26$ patients), there was no visible trend towards the strong association found for the next time range. We hypothesize that immediately after surgery, intra-cavity transudation leads to non-tumor enhancement. Accordingly, Smets et al. showed that contrast transudation could appear in the very early postoperative MRI scans of glioblastomas[6]. Thereafter, the tissues may repair so that the enhancement corresponds to the real tumor remnants. Nevertheless, other possible explanations for this lack of association include the presence of surgical products that may affect MRIs performed soon after surgery.

We further explored whether features describing tissue vasculature from the DSC would differentiate the enhancement corresponding to post-surgery residual tumor and inflammation. Interestingly, 99th percentile rCBV and maximum PSR are indicators of bad prognosis when the residual tumor was assessed in an EPMR performed more than 24 h post-surgery. Accordingly, associations between recurrent tumor and higher rCBV values have been previously reported[21]. However, this correlation is strongest between 24 to 48 h but was not present for scans acquired before 24 h or after 72 h.

Taken together, these findings suggest that the enhancement thickness is representative of the remaining tumor when the MRI scan is acquired from 24 to 72 h after surgery, however there may be some confounders (i.e., non-tumor contrast enhancement) when the EPMR is acquired too early (during the first day after surgery) or too late (more than 4 days after surgery). Further studies are needed to better ascertain the importance of the number of days after surgery in measuring the remaining tumor tissue and elucidating the biological explanation behind such a phenomenon.

The prognostic value of the extent of resection is widely accepted by the scientific community. Nevertheless, there is no consensus about the threshold of the extent of resection that can influence survival. Some authors have suggested an "all-or-nothing" approach[8]. In this setting, no debulking surgery should be pursued when the tumor involves critical brain areas. Other studies however, suggest that a smaller extent of tumor resection could also positively impact patient survival[3–7]. In agreement with this, we found a correlation between residual tumor extent and patient survival. This supports maximal safe tumor resection with the knowledge that it could improve patient survival even in cases of partial tumor resection.

In our population, the radiomics analysis of the enhancement area including texture features, allows for identification of patients with long and short survival. As previously shown, radiomics features inform tumor heterogeneity and are indicators of tumor aggressiveness in glioblastoma[22–24]. We have explored several other variables to evaluate their clinical utility. While all quantitative imaging data (radiomics, thickness, DSC) provide relevant information that complements clinical data (age and KPS), we found the enhancing tumor thickness and radiomics to have high prognostic value and only T1w and T1wC are required to calculate them, thus simplifying the implementation.

This study has a number of potential limitations. Firstly, as a retrospective study, isocitrate dehydrogenase (*IDH)* mutation status data were available for 54 patients, only one of whom (2%) had an *IDH* mutation. This low incidence rate may be explained by the relatively old age of our population, as *IDH*-mutant glioblastoma has been reported to correlate with younger age[25]. In addition, only glioblastomas treated with maximal safe resection were included in our study. Newly diagnosed *IDH*-mutant glioblastoma tend to present with a more diffuse infiltrative pattern than *IDH*-wildtype glioblastoma, likely leading to a biopsy instead of a maximal safe resection, and *IDH*-mutant glioblastomas progressing after a previous known *IDH*-mutant low-grade astrocytoma[26] were also not included in our study. Therefore, there may be confounders in patients for whom *IDH* status is missing, although as the rate of *IDH* mutant in glioblastoma is 10% in the general population[27] compared to approximately 2% in our population, we believe this did not significantly affect our results.

Secondly, in this analysis we focus on the post-surgical residual enhancing tumor and certain drugs may be administered perioperatively that can affect the enhancement on post-surgery T1wC scans (i.e., glucocorticoids or antiangiogenic drugs such as bevacizumab). We confirmed that bevacizumab was not administered prior to EPMR in any patients. Glucocorticoids might have been administered in cases with extensive edema, which could

have influenced the characteristics of post-contrast MRI. Nevertheless, glucocorticoids are avoided as much as possible in our hospital to prevent this drug from affecting the histological interpretation of biopsy samples.

Thirdly, given the observational retrospective nature of this study, standardization of all the image acquisition parameters was not possible. Therefore, all images were homogenously normalized in intensity and resampled to minimize variability. Additionally, DSC scans without contrast preload and acquisition at 1.5 T may result in leakage effects, hence leakage correction methods were applied to avoid confounders. Additionally, the method used requires input (semi-automatic segmentation) from an experienced radiologist who supervises the enhancement mask. This might induce some segmentation variability. However, the thickness is calculated as the mean value of the distance transform, and thus it is possible that an average value smooths out small differences in segmentations, making it more robust. Lastly, we performed an internal validation, though applying the prognostic models to an external dataset would confirm its performance.

It is noteworthy to mention that there may still be non-enhancing residual tumor, which is out of the scope of the current analysis. Therefore, as future work, combining other sequences such as T2w and FLAIR (Fluid-Attenuated Inversion Recovery) may be useful to delineate additional areas of residual tumor with a potential role in patient prognosis[28,29], although at the expense of a more complex pipeline.

Additionally, novel MRI techniques using chemical exchange saturation transfer (CEST) imaging have been reported to correlate with response to treatment and survival. With CEST MRI, pulses off the water resonance frequency are absorbed by proteins that then transfer the energy to water through chemical exchange, lowering the signal. Derived signal maps such as amide proton transfer (APT) and nuclear Overhauser effect (NOE) are both reported to correlate with survival in gliomas and glioblastomas[30–32], even in a sub-cohort of *IDH*-wildtype patients. While the number of patients evaluated in these studies is small, they showed similar HRs as those reported in our study. This highlights the potential of CEST MRI and its radiomics analysis, and merits further research. In combination with the relevant variables proposed in this study, a thorough imaging signature could provide major insight into patient response.

In conclusion, in patients with glioblastoma multiforme, objective quantification of the area of enhancement in the tumor bed after surgery, in an EPMR scan performed between 24 and 72 h after surgery provides relevant information on the remaining tumor and relevant prognostic information. The MRI processing pipeline defined in this study is an easy and rapid method for a more accurate evaluation of the post-surgery residual tumor, that could be implemented in routine clinical practice.

Methods

Ethical approval. The Research Ethics Committee of the Hospital Universitari de Bellvitge revised and approved this study, in compliance with applicable regulatory requirements and the International Conference on Harmonization Guidelines on Good Clinical Practice (ICH GCP). For this retrospective study, informed consent was waived by the Research Ethics Committee of the Hospital Universitari de Bellvitge.

Study population and design. We evaluated data from 160 consecutive patients with primary glioblastoma multiforme who underwent maximum safe resection surgery between February 2009 and December 2017 at the Hospital Universitari de Bellvitge, Spain, and who had an early post-surgery MRI scan (i.e., within the first week after surgery). Data on isocitrate dehydrogenase (*IDH*) 1/2 mutation status were available in 54 out of 160 patients (33%)[33]. Only one patient (1/54, 2%) had an identified *IDH* mutation. To be eligible for the analysis, patients had to have been treated with adjuvant radiotherapy plus concomitant and post-radiotherapy temozolomide (i.e., Stupp protocol)[2] (144/160, 90%).

MRI scan protocol and image pre-processing. The earliest post-surgery MRI scan per patient was collected, including T1-weighted before (T1w) and after intravenous contrast administration (T1wC), and dynamic susceptibility contrast (DSC) images. Approximately half of the MRI scans were acquired with a 1.5 T MR magnet Intera (Philips Healthcare, Best, The Netherlands) and half with a 1.5 T MR magnet Achieva (Philips Healthcare, Best, The Netherlands) with the following parameters for T1w: spin-echo; TE 15 ms; TR 540 ms; flip angle 90°; matrix size 256×256 mm^2; slice thickness 5 mm.

A single MRI scan was performed with a 3 T MR unit (Achieva; Philips Healthcare, Best, The Netherlands), gradient-echo; TE 4 ms; TR 9 ms; flip angle 8°; matrix size 672×672 mm^2; slice thickness 1 mm.

Regarding DSC, all acquisitions were gradient-echo without bolus preload; a total of 10 baseline points were collected before the pass of contrast with the following protocol: in the 1.5 T Intera MR unit (scans from 2009 to 2015) PRESTO sequence; TE 25 ms; TR 17 ms; flip angle 7°; matrix size 128×128 mm^2; slice thickness 3.5 mm; dynamic acquisitions 40. In the 1.5 T Achieva MR unit (from 2015 to 2917) TE 40 ms; TR 1522–1642 ms; flip angle 75°; matrix size 128×128 mm^2; slice thickness, 4–5 mm; dynamic acquisitions 60.

A single MRI scan was performed with a 3 T MR unit (Achieva; Philips Healthcare, Best, The Netherlands), echo planar sequence; TE 40 ms; TR 1618 ms; flip angle 75°; matrix size 128×128 mm^2; slice thickness 4 mm; dynamic acquisitions 40.

The pre-processing of images for homogenization included N4 bias field correction[34] and intensity normalization with Nyul's method as adapted by Shah[35,36].

Enhancement mask definition and thickness quantification. The T1w and T1wC images with different slice spacing (1/144 = 1% of the scans) were resampled to 5 mm for consistency and robustness. The T1w was registered to the T1wC with the rigid transformation tool of 3DSlicer v4.10[37]. The T1w signal intensity was normalized to that of the T1Cw[28]. A 3D enhancement map was obtained by T1w-T1wC subtraction. The area of enhancement within and around the tumor bed was delineated by an experienced radiologist (P.N.B), blinded

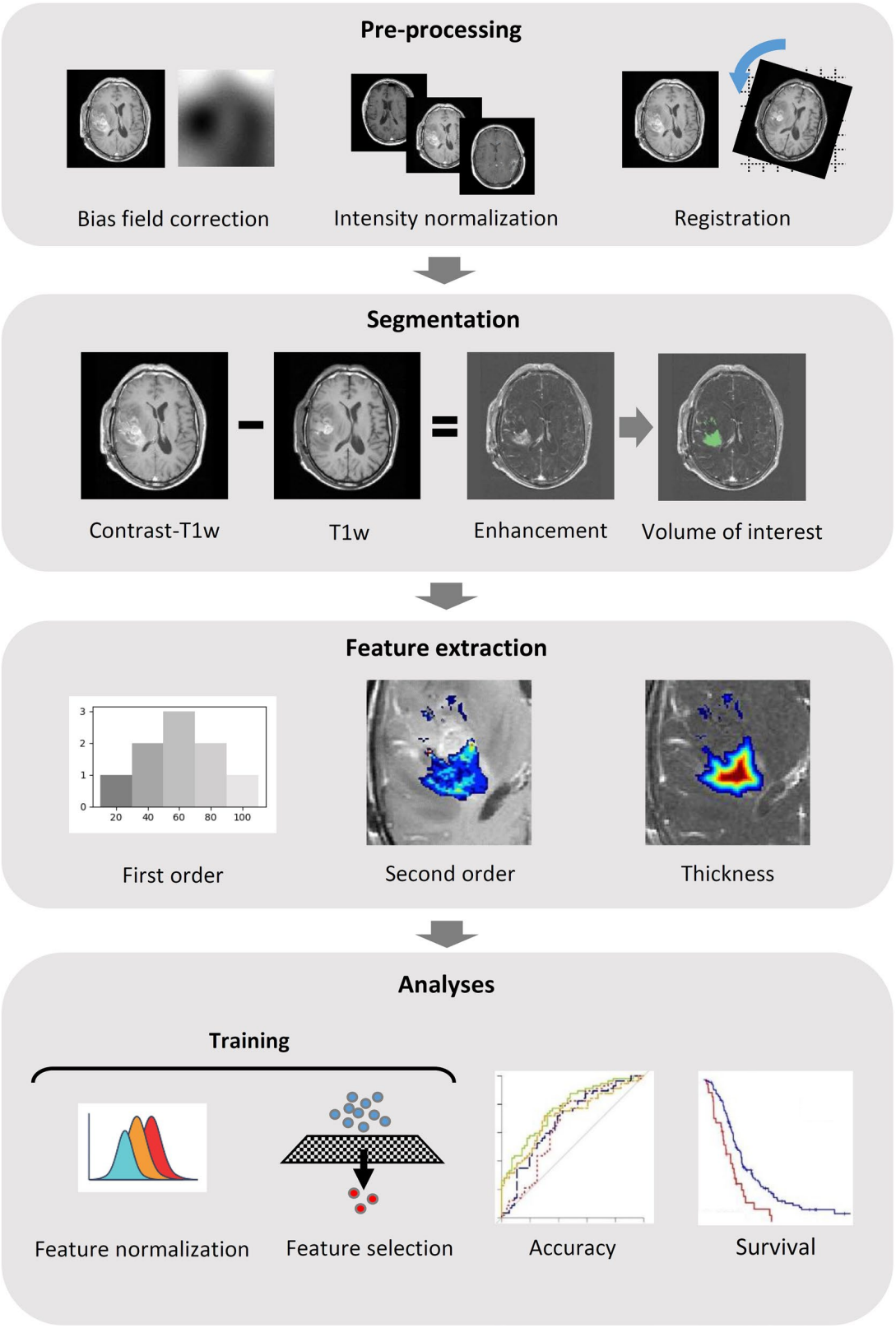


Figure 3. Pipeline of the applied methods. Registration is performed with the contrast-T1w scan as the reference image and the radiomics features are extracted from the contrast-T1w. The thickness is calculated from the 3D distance transform of the volume of interest, depicted as a colormap.

to the clinical outcome, using the 3DSlicer semi-automatic delineation tools (thresholding and morphological operations).

The 3D distance transform of the volume of interest (VOI) was then calculated. The distance transform provides the Euclidean distance in millimeters of every voxel within the VOI to its nearest VOI boundary[38]. We calculated the mean and the maximum values of the 3D distance transform as measures of the enhancement mask thickness for each patient (Fig. 3), as well as the total volume of the VOI.

Perfusion MRI sub-analysis. The DSC temporal volumes were processed with the DSCMRIAnalysis module of 3DSlicer[37,39], the curves were leakage-corrected with the Boxerman–Schmainda–Weiskoff method[40] and the rCBV were obtained. The DSC volumes were then registered to the T1wC images for co-alignment with the enhancement masks. The temporal curves of the DSC were first analyzed to remove noisy and low-signal curves from the mask. For this purpose only, the curves were low-pass filtered, normalized between − 1 and 0, and then those that reflected the bolus arrival[41,42] were selected by resemblance to a Gaussian distribution (Supplementary Fig. S6). Once the DSC curves had been filtered, the original non-normalized DSC values were analyzed.

Relative cerebral blood volume. An ROI within the white matter contralateral to the tumor was delineated in each MRI and the relative cerebral blood volume (rCBV) map was normalized to this ROI. To retrieve the tumor hot-spot, we calculated the 99th percentile of the rCBV (rCBV-99) and the maximum percentage of signal recovery (PSR) from the enhancement mask selected curves. The rCBV-99 and not the maximum was calculated to avoid extreme outlier values of the rCBV. The PSR was calculated as described by Cha et al.[42].

Radiomics feature extraction and robustness. Radiomics extraction was performed with Pyradiomics v2.1.2 for Python[43], with image resampling of 1 mm isotropic voxels and fixed binarization to 10 levels of bin-width, as has been suggested to maximize radiomics reproducibility[44–46]. Ninety-four radiomics features were extracted from the enhancement mask applied to the T1wC images, including first-order (n = 19) and second-order from texture Grey-Level Co-occurrence Matrix (n = 24), Grey-Level Run-Length Matrix (n = 16), Grey-Level Size-Zone Matrix (n = 16), Neighboring-Grey-Tone Difference Matrix (n = 5) and Grey-Level Dependence Matrix (n = 14) (Supplementary Table S6). Feature description and compliance with image biomarker standardization initiative guidelines is publicly available in the package documentation[43]. We also studied the variability of radiomics when changing the extraction parameters voxel size and bin width, and when the most robust features (i.e., < 20% coefficient of variation) were selected for further analysis (54 out of 94 [57%] radiomics features, see Supplementary Methods).

Radiomics feature selection and analysis. Patients with measurable enhancement segmentation (129/144, 90%) were eligible for this analysis. For clarification, these patients may present gross total resection, however benign reactive enhancement has been reported by some authors up to 72 h after surgery[3,6,16,47]. The patients were split into 'training' (70%) and 'test' (30%) sets for internal validation, with balanced survival distribution in both sets. Patient characteristics of the sets are presented in Supplementary Table S3. The radiomics features were normalized according to the mean value and standard deviation of the training set. From the 54 robust features of the previous section, further selection was performed in the training set by minimum-redundancy–maximum-relevance[48,49] plus stepwise regression. The logistic regression model with the selected variables was then internally validated in the test set.

Statistical analysis. Log-rank analysis on Kaplan–Meier data and Cox proportional hazards regression models were used to evaluate the association of the analyzed variables with survival.

For dichotomized survival analysis, the population was split according to lower or higher than 2-year overall survival (< 2 years, ≥ 2 years respectively, Supplementary Table S1). Censored patients with a survival shorter than the defined endpoint were excluded. These survival groups were defined for logistic regression models and to obtain the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. A threshold for each independent variable was calculated from the regression models from the maximum sum of sensitivity and specificity (Youden's index). The postoperative Karnofsky performance status (KPS) was dichotomized as < 90 and ≥ 90.

To assess how the time from surgery to the EPMR scan affects the residual tumor quantification, sub-analyses of survival prediction were performed based on the time elapsed between surgery and the EPMR scan. The patients were split into four subpopulations according to the time between surgery and the MRI scan (< 24, 24 to < 48, 48 to < 72 and ≥ 72 h). Multiple comparison tests were adjusted with the Benjamini–Hochberg method, and p-values were considered statistically significant below 0.05.

The thickness analysis was done in-house with Matlab R2015a (Mathworks). Radiomics selection and statistical analysis were performed with R Statistical Software v3.5.1[50].

Received: 12 March 2020; Accepted: 9 December 2020
Published online: 12 January 2021

References

1. Ostrom, Q. T. *et al.* CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. *Neuro Oncol.* **20**, 1–86. https://doi.org/10.1093/neuonc/noy131 (2018).

2. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996. https://doi.org/10.1056/NEJMoa043330 (2005).
3. Majos, C. *et al.* Early post-operative magnetic resonance imaging in glioblastoma: Correlation among radiological findings and overall survival in 60 patients. *Eur. Radiol.* **26**, 1048–1055. https://doi.org/10.1007/s00330-015-3914-x (2016).
4. Kuhnt, D. *et al.* Correlation of the extent of tumor volume resection and patient survival in surgery of glioblastoma multiforme with high-field intraoperative MRI guidance. *Neuro Oncol.* **13**, 1339–1348. https://doi.org/10.1093/neuonc/nor133 (2011).
5. Bloch, O. *et al.* Impact of extent of resection for recurrent glioblastoma on overall survival: Clinical article. *J. Neurosurg.* **117**, 1032–1038. https://doi.org/10.3171/2012.9.JNS12504 (2012).
6. Smets, T., Lawson, T. M., Grandin, C., Jankovski, A. & Raftopoulos, C. Immediate post-operative MRI suggestive of the site and timing of glioblastoma recurrence after gross total resection: A retrospective longitudinal preliminary study. *Eur. Radiol.* **23**, 1467–1477. https://doi.org/10.1007/s00330-012-2762-1 (2013).
7. Krivoshapkin, A. L. *et al.* Automated volumetric analysis of postoperative magnetic resonance imaging predicts survival in patients with glioblastoma. *World Neurosurg.* https://doi.org/10.1016/j.wneu.2019.03.142 (2019).
8. Lacroix, M. *et al.* A multivariate analysis of 416 patients with glioblastoma multiforme: Prognosis, extent of resection, and survival. *J. Neurosurg.* **95**, 190–198. https://doi.org/10.3171/jns.2001.95.2.0190 (2001).
9. Ellingson, B. M. *et al.* Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. *Neuro Oncol.* **20**, 1240–1250. https://doi.org/10.1093/neuonc/noy053 (2018).
10. Albert, F. K., Forsting, M., Sartor, K., Adams, H. P. & Kunze, S. Early postoperative magnetic resonance imaging after resection of malignant glioma: Objective evaluation of residual tumor and its influence on regrowth and prognosis. *Neurosurgery* **34**, 45–60 (1994).
11. Stupp, R. *et al.* High-grade glioma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol* **25**(Suppl 3), 93–101. https://doi.org/10.1093/annonc/mdu050 (2014).
12. Wen, P. Y. *et al.* Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. *J. Clin. Oncol.* **28**, 1963–1972. https://doi.org/10.1200/JCO.2009.26.3541 (2010).
13. Martinez-Garcia, M. *et al.* SEOM clinical guidelines for diagnosis and treatment of glioblastoma (2017). *Clin. Transl. Oncol.* **20**, 22–28. https://doi.org/10.1007/s12094-017-1763-6 (2018).
14. Allahdini, F., Amirjamshidi, A., Reza-Zarei, M. & Abdollahi, M. Evaluating the prognostic factors effective on the outcome of patients with glioblastoma multiformis: Does maximal resection of the tumor lengthen the median survival?. *World Neurosurg.* **73**, 128–134. https://doi.org/10.1016/j.wneu.2009.06.001 (2010).
15. Sanai, N., Polley, M. Y., McDermott, M. W., Parsa, A. T. & Berger, M. S. An extent of resection threshold for newly diagnosed glioblastomas. *J. Neurosurg.* **115**, 3–8. https://doi.org/10.3171/2011.2.JNS10998 (2011).
16. Ellingson, B. M. *et al.* Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. *Neuro-Oncology* **20**, 1240–1250. https://doi.org/10.1093/neuonc/noy053 (2018).
17. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006. https://doi.org/10.1038/ncomms5006 (2014).
18. Lambin, P. *et al.* Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762. https://doi.org/10.1038/nrclinonc.2017.141 (2017).
19. National Comprehensive Cancer Network. Central Nervous System Cancers (Version 1.2019). NCCN Clin Pract Guidel Oncol 2019. https://www.nccn.org/professionals/physician_gls/ (accessed April 22, 2019).
20. Stupp, R. *et al.* High-grade glioma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **25**(Suppl 3), 93–101. https://doi.org/10.1093/annonc/mdu050 (2014).
21. Bisdas, S. *et al.* Cerebral blood volume measurements by perfusion-weighted MR imaging in gliomas: Ready for prime time in predicting short-term outcome and recurrent disease?. *AJNR Am. J. Neuroradiol.* **30**, 681–688. https://doi.org/10.3174/ajnr.A1465 (2009).
22. Li, Q. *et al.* A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci. Rep.* **7**, 14331. https://doi.org/10.1038/s41598-017-14753-7 (2017).
23. Su, C. *et al.* Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. *Eur. Radiol.* **29**, 1986–1996. https://doi.org/10.1007/s00330-018-5704-8 (2019).
24. Ditmer, A. *et al.* Diagnostic accuracy of MRI texture analysis for grading gliomas. *J. Neurooncol.* **140**, 583–589. https://doi.org/10.1007/s11060-018-2984-4 (2018).
25. Robinson, C. & Kleinschmidt-DeMasters, B. K. IDH1-mutation in diffuse gliomas in persons age 55 years and over. *J. Neuropathol. Exp. Neurol.* **76**, 151–154. https://doi.org/10.1093/jnen/nlw112 (2017).
26. Ohgaki, H. & Kleihues, P. The definition of primary and secondary glioblastoma. *Clin. Cancer Res.* **19**, 764–772. https://doi.org/10.1158/1078-0432.CCR-12-3002 (2013).
27. Louis, D. N. *et al.* The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **131**, 803–820. https://doi.org/10.1007/s00401-016-1545-1 (2016).
28. Rathore, S. *et al.* Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci. Rep.* **8**, 5087. https://doi.org/10.1038/s41598-018-22739-2 (2018).
29. Porz, N. *et al.* Multi-modal glioblastoma segmentation: Man versus machine. *PLoS ONE* **9**, e96873. https://doi.org/10.1371/journal.pone.0096873 (2014).
30. Meissner, J. E. *et al.* Early response assessment of glioma patients to definitive chemoradiotherapy using chemical exchange saturation transfer imaging at 7 T. *J. Magn. Reson. Imaging* **50**, 1268–1277. https://doi.org/10.1002/jmri.26702 (2019).
31. Paech, D. *et al.* Relaxation-compensated amide proton transfer (APT) MRI signal intensity is associated with survival and progression in high-grade glioma patients. *Eur. Radiol.* **29**, 4957–4967. https://doi.org/10.1007/s00330-019-06066-2 (2019).
32. Mehrabian, H., Myrehaug, S., Soliman, H., Sahgal, A. & Stanisz, G. J. Evaluation of glioblastoma response to therapy with chemical exchange saturation transfer. *Int. J. Radiat. Oncol. Biol. Phys.* **101**, 713–723. https://doi.org/10.1016/j.ijrobp.2018.03.057 (2018).
33. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773. https://doi.org/10.1056/NEJMoa0808710 (2009).
34. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908 (2010).
35. Nyul, L. G. & Udupa, J. K. On standardizing the MR image intensity scale. *Magn. Reson. Med.* **42**, 1072–1081. https://doi.org/10.1002/(sici)1522-2594(199912)42:6%3c1072::Aid-mrm11%3e3.0.Co;2-m (1999).
36. Shah, M. *et al.* Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Med. Image Anal.* **15**, 267–282. https://doi.org/10.1016/j.media.2010.12.003 (2011).
37. Fedorov, A. *et al.* 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imaging* **30**, 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001 (2012).
38. Grevera, G.J. in Deformable *Models* Ch. Chapter 2, 33–60 (2007).
39. Schmainda, K. M. *et al.* Multisite concordance of DSC-MRI analysis for brain tumors: Results of a national cancer institute quantitative imaging network collaborative project. *AJNR Am. J. Neuroradiol.* **39**, 1008–1016. https://doi.org/10.3174/ajnr.A5675 (2018).
40. Boxerman, J. L., Schmainda, K. M. & Weisskoff, R. M. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *Am. J. Neuroradiol.* **27**, 859–867 (2006).

41. Lupo, J. M., Cha, S., Chang, S. M. & Nelson, S. J. Dynamic susceptibility-weighted perfusion imaging of high-grade gliomas: Characterization of spatial heterogeneity. *AJNR Am. J. Neuroradiol.* **26**, 1446–1454 (2005).
42. Cha, S. *et al.* Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *AJNR Am. J. Neuroradiol.* **28**, 1078–1084. https://doi.org/10.3174/ajnr.A0484 (2007).
43. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339 (2017).
44. Molina, D. *et al.* Lack of robustness of textural measures obtained from 3D brain tumor MRIs impose a need for standardization. *PLoS ONE* **12**, e0178843. https://doi.org/10.1371/journal.pone.0178843 (2017).
45. Goya-Outi, J. *et al.* Computation of reliable textural indices from multimodal brain MRI: Suggestions based on a study of patients with diffuse intrinsic pontine glioma. *Phys. Med. Biol.* **63**, 105003. https://doi.org/10.1088/1361-6560/aabd21 (2018).
46. Duron, L. *et al.* Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS ONE* **14**, e0213459. https://doi.org/10.1371/journal.pone.0213459 (2019).
47. Bette, S. *et al.* Patterns and time dependence of unspecific enhancement in postoperative magnetic resonance imaging after glioblastoma resection. *World Neurosurg.* **90**, 440–447. https://doi.org/10.1016/j.wneu.2016.03.031 (2016).
48. De Jay, N. *et al.* mRMRe: An R package for parallelized mRMR ensemble feature selection. *Bioinformatics* **29**, 2365–2368. https://doi.org/10.1093/bioinformatics/btt383 (2013).
49. Parmar, C. *et al.* Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front. Oncol.* **5**, 272. https://doi.org/10.3389/fonc.2015.00272 (2015).
50. Core Team R. *R: A Language and Environment for Statistical Computing* (R Foundation for statistical computing, Vienna, 2013).

Acknowledgements

This work was supported by the Fundacio La Caixa. R.P.L is supported by a Prostate Cancer Foundation Young Investigator Award, CRIS Foundation Talent Award (TALENT-05), Fero Foundation, and the Spanish Ministry of Health FIS Program (Instituto de Salud Carlos III-Investigación en Salud PI18/01395). Mr Guillermo Villacampa Javierre kindly provided statistical advice for this manuscript.

Author contributions

All of the authors substantially contributed to the presented work in many aspects. Furthermore, we consider A.G.R. and P.N.B. to be joint first authors for their equal efforts in this study, as well as R.P.L. and C.M. to be joint senior authors for their decisive contributions to all aspects of this study. Other contributions made by non-authors were properly acknowledged within the manuscript with their consent. As a summary of author participations: C.M., R.P.L., A.G.R, P.N., A.P. and N.C. together conceived the study design and development. C.M., J.B., G.P. and M.C. were in charge of the medical data acquisition. P.N., C.M. and N.C. were responsible of defining the criteria for data collection into a structured database. P.N., C.M., R.P.L. and A.G.R. worked in the segmentation of the images. A.G.R. and M.L. carried out the processing of data and performed the analyses, with the help of C.M., P.N. and A.P. in the interpretation and validation of results. A.G.R, R.P.L, C.M. and P.N. worked in the manuscript drafting. R.P.L. and C.M. critically revised the written works. Lastly, all authors reviewed and approved the manuscript for its publication. Each and every one of the authors certifies the integrity of the presented work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-79829-3.

Correspondence and requests for materials should be addressed to R.P.-L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

4.2 Study 2

An accessible deep learning tool for voxel-wise classification of brain malignancies from perfusion MRI

Cell Reports Medicine (2024), Vol 5, Issue 3, Page 101464. **Impact Factor (quartile)**: 11.7 (Q1)

Authors and affiliations: Alonso Garcia-Ruiz[1], Albert Pons-Escoda[2,3], Francesco Grussu[1], Pablo Naval-Baudin[2], Camilo Monreal-Aguero[1], Gretchen Hermann[4], Roshan Karunamuni[4], Marta Ligero[1], Antonio Lopez-Rueda[5], Laura Oleaga[5], Alvaro Berbis[6], Alberto Cabrera-Zubizarreta[7], Teodoro Martin-Noguerol[7], Antonio Luna-Alcalá[7], Tyler M. Seibert[4,8,9], Carlos Majos[2,3], Raquel Perez-Lopez[1,10].

1. Radiomics Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain.
2. Department of Radiology, Bellvitge University Hospital, Barcelona, Spain.
3. Neuro Oncology Unit, Institut d'Investigacio Biomedica de Bellvitge (IDIBELL), Barcelona, Spain.
4. Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, California, United States of America.
5. Department of Radiology, Hospital Clínic de Barcelona, Spain.
6. Department of Radiology, HT Medica, Hospital San Juan de Dios, Cordoba, Spain
7. Department of Radiology, HT Medica, Jaen, Spain
8. Department of Radiology, University of California, San Diego, La Jolla, California, United States of America.
9. Department of Bioengineering, University of California, San Diego, La Jolla, California, United States of America.
10. Department of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain.

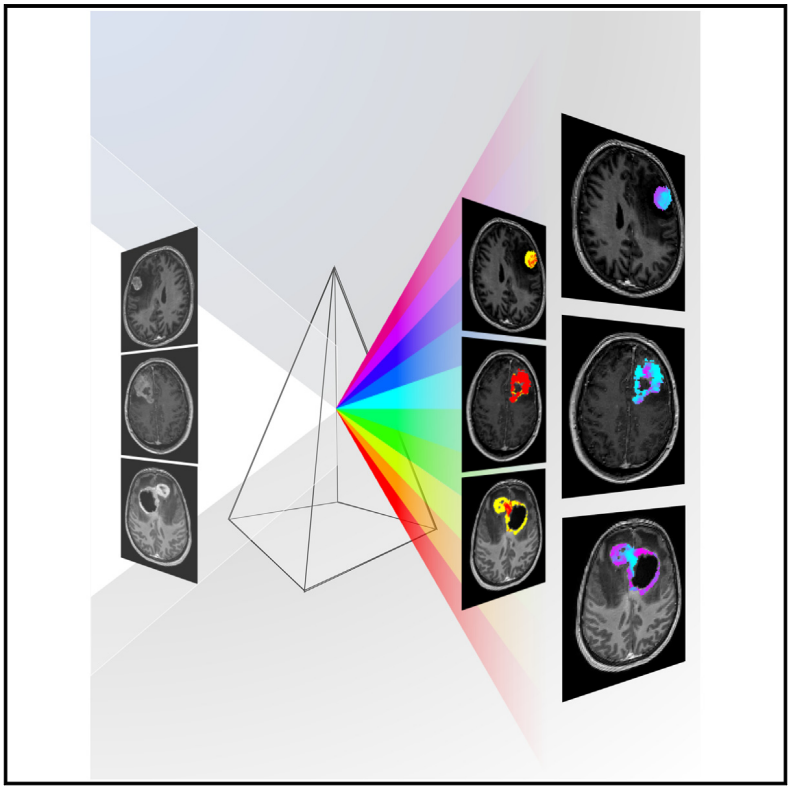
This study corresponds to **objective 2** of the thesis. This study proposed a novel approach of analysing DSC perfusion MRI for the diagnosis of the three most common brain malignancies in adults. The method proposed a voxel-wise normalization of DSC data to train a CNN-based classification model. Additional efforts were made to validate the performance of the model and to deploy and use it in an easy way. The PhD candidate was responsible for data collection from public databases, data inspection, image processing, building, training and testing of classification models, statistical analyses and manuscript drafting.

Cell Reports
Medicine

Article

An accessible deep learning tool for voxel-wise classification of brain malignancies from perfusion MRI

Graphical abstract



Highlights

- Voxel-wise approach enables training neural networks with limited patient cohorts
- DISCERN facilitates brain tumor classification to aid medical decisions
- DISCERN is a user-friendly tool designed for use with clinical perfusion MRI
- DISCERN enables accurate brain tumor diagnosis, surpassing conventional metrics

Authors

Alonso Garcia-Ruiz, Albert Pons-Escoda, Francesco Grussu, ..., Tyler M. Seibert, Carlos Majos, Raquel Perez-Lopez

Correspondence

rperez@vhio.net

In brief

Diagnosing brain tumors can be challenging, even with multiparametric MRI that includes vascular tumor evaluation through perfusion imaging. In this work, Garcia-Ruiz et al. developed a deep learning-based tool that leverages the full spatial and temporal information of perfusion MRI, outperforming conventional methods.

Garcia-Ruiz et al., 2024, Cell Reports Medicine *5*, 101464
March 19, 2024 © 2024 The Authors.
https://doi.org/10.1016/j.xcrm.2024.101464

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

An accessible deep learning tool for voxel-wise classification of brain malignancies from perfusion MRI

Alonso Garcia-Ruiz,[1,10] Albert Pons-Escoda,[2,3,10] Francesco Grussu,[1,10] Pablo Naval-Baudin,[2] Camilo Monreal-Aguero,[1] Gretchen Hermann,[4] Roshan Karunamuni,[4] Marta Ligero,[1] Antonio Lopez-Rueda,[5] Laura Oleaga,[5] M. Álvaro Berbís,[6] Alberto Cabrera-Zubizarreta,[7] Teodoro Martin-Noguerol,[7] Antonio Luna,[7] Tyler M. Seibert,[4,8,9] Carlos Majos,[2,3] and Raquel Perez-Lopez[1,11,*]

[1]Radiomics Group, Vall d'Hebron Institute of Oncology (VHIO), 08035 Barcelona, Spain
[2]Radiology Department, Bellvitge University Hospital, 08907 Barcelona, Spain
[3]Neuro-Oncology Unit, Institut d'Investigacio Biomedica de Bellvitge (IDIBELL), 08907 Barcelona, Spain
[4]Radiation Medicine Department and Applied Sciences, University of California, San Diego, La Jolla, CA 92093, USA
[5]Radiology Department, Hospital Clínic de Barcelona, 08036 Barcelona, Spain
[6]Radiology Department, HT Medica, Hospital San Juan de Dios, 14012 Cordoba, Spain
[7]Radiology Department, HT Medica, 23008 Jaen, Spain
[8]Radiology Department, University of California, San Diego, La Jolla, CA 92093, USA
[9]Bioengineering Department, University of California, San Diego, La Jolla, CA 92093, USA
[10]These authors contributed equally
[11]Lead contact
*Correspondence: rperez@vhio.net
https://doi.org/10.1016/j.xcrm.2024.101464

SUMMARY

Noninvasive differential diagnosis of brain tumors is currently based on the assessment of magnetic resonance imaging (MRI) coupled with dynamic susceptibility contrast (DSC). However, a definitive diagnosis often requires neurosurgical interventions that compromise patients' quality of life. We apply deep learning on DSC images from histology-confirmed patients with glioblastoma, metastasis, or lymphoma. The convolutional neural network trained on ~50,000 voxels from 40 patients provides intratumor probability maps that yield clinical-grade diagnosis. Performance is tested in 400 additional cases and an external validation cohort of 128 patients. The tool reaches a three-way accuracy of 0.78, superior to the conventional MRI metrics cerebral blood volume (0.55) and percentage of signal recovery (0.59), showing high value as a support diagnostic tool. Our open-access software, Diagnosis In Susceptibility Contrast Enhancing Regions for Neuro-oncology (DISCERN), demonstrates its potential in aiding medical decisions for brain tumor diagnosis using standard-of-care MRI.

INTRODUCTION

Differential diagnosis between the most common brain malignancies (i.e., glioblastoma multiforme [GBM], brain metastasis from solid tumors, and primary CNS lymphoma [PCNSL]) represents a clinical unmet need because each of these entities requires a distinct therapeutic approach.[1–3] Although pathology evaluation of tumor samples remains the gold standard for diagnosis, it requires invasive neurosurgical procedures, with a significant risk of complications, and eventually can be confounded by the use of prior medication, such as steroids.[4,5]

To overcome the need for surgery, magnetic resonance imaging (MRI) with intravenous contrast injection is being used as a noninvasive support system for differential diagnosis of brain malignancies. GBM, brain metastasis, and PCNSL represent up to 70% of all malignant brain tumors and more than 80% of contrast-enhancing tumors within the brain.[6] Nevertheless, the enhancing patterns on imaging exhibit a high degree of similarity across these tumor types, making differential diagnosis challenging even for experienced neuroradiologists.[7–9]

The noninvasive characterization of brain tumors on MRI has been an active subject of study for years,[10,11] gaining renewed interest with the application of recent machine learning techniques to imaging data. Among the existing literature, some studies have focused on differentiating GBM from solitary brain metastasis, either with anatomical[12–17] or functional imaging,[10,18–23] whereas other works have concentrated on the identification of PCNSL.[24–27] Of particular significance is dynamic susceptibility contrast (DSC) perfusion MRI, which enables the visualization of vascular characteristics, including vascular density and permeability, and is proving to be valuable for brain tumor diagnosis.[9,10,18,20,22,27–37]

DSC is a quantitative MRI technique that consists of a temporal T2*-weighted acquisition during the administration of a vascular

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

contrast bolus. The contrast agent causes an initial decrease in the T2*-weighted signal intensity, followed by the signal recovery during washout. In DSC, every voxel in the image yields a unique dynamic curve that describes the temporal evolution of the T2*-weighted signal intensity and reflects local tissue vascular properties. The standard approach to analyze DSC is to derive metrics such as the relative cerebral blood volume (rCBV) and the percentage of signal recovery (PSR), both of which simplify the dynamic signal. The rCBV relates to the tumor vascular density with respect to normal tissue, and the PSR reflects the vascular permeability.[38] Both parameters remain the main focus of DSC analyses for tasks such as tumor type differentiation, grade stratification, and treatment response assessment.[28,39,40] However, among diverse clinically used DSC protocols, the performance of these parameters differs greatly,[38,41,42] which limits its use in routine clinical practice. Although recent multidisciplinary efforts have been made in the community to agree on a common procedure,[43–45] a global standardized DSC workflow is still lacking, regarding variations in the contrast preload settings, imaging parameters, and processing methods, all of which pose additional challenges to the generalizability of the technique and the establishment of reference rCBV/PSR values.

Regarding the use of DSC in aiding brain tumor diagnosis, it is worth noting that, with a few exceptions,[16,21,46] most studies to date have been designed to discriminate between two tumor types or pairs of malignancies. Furthermore, even fewer analyses can be found over large populations or validating external data, thus limiting the generalizability and clinical utility of the presented approaches.

DSC curve normalization and voxel-by-voxel analyses of the full dynamic range can overcome the limitations of conventional metrics and unlock the potential of DSC as a tool for differential diagnosis among the most common brain malignancies. Moreover, the application of deep learning techniques may enable new strategies of analysis and inference for dynamic data. In one of the first works exploring deep learning in DSC data,[47] the authors described an end-to-end pipeline to obtain model-free perfusion metrics from the raw data. They used one-dimensional (1D) convolutional neural networks (CNNs) to characterize the dynamic DSC data of individual voxels. Park et al.[34] developed an autoencoder and a clustering strategy to distinguish different brain areas, including pathologies, from the 1D dynamic data.

In this work, we describe the development and validation of an innovative, comprehensive framework for differential diagnosis of GBM, brain metastasis, and PCNSL, taking advantage of all of the time points of the normalized DSC (nDSC) data. The proposed Diagnosis In Susceptibility Contrast Enhancing Regions for Neuro-oncology (DISCERN) app provides voxel-by-voxel signatures of tumor type and is based on training 1D deep CNNs, with only a small number of pilot scans for a given DSC protocol. In the present study, we demonstrate the feasibility and accuracy of the method and show its superior performance compared to classifiers based on conventional DSC metrics. In addition, our method exhibits on par or higher diagnostic capabilities in comparison to those of expert neuroradiologists. The potential of DISCERN is to aid radiologists in interpreting brain MRI data, thereby enhancing the diagnostic capacity of experienced neuroradiologists and allowing less experienced radiologists to achieve a higher level of proficiency. To facilitate ease of use, we have designed a user-friendly interface with the ultimate goal of minimizing the need for invasive brain biopsies and guiding the selection of optimal treatment strategies in clinical practice.

RESULTS

Population demographics

In this multicenter, retrospective study, we analyzed MRI data from a total of 568 patients with biopsy-confirmed GBM, brain metastasis, or PCNSL. The classification model was developed and tested with 440 patients from a single center, and additional independent cohorts with varying imaging protocols were processed for external validation (Figure 1A). Further information about the study cohorts and classification results can be found in the STAR Methods and in Table S1. No statistically significant differences ($p > 0.05$) in terms of age and gender were observed between the three tumor types.

Development of a CNN for brain tumor classification

We trained our CNN classifier on a development cohort in which patients were randomly split into training and test sets. For the training set, we included 20 patients with PCNSL and 20 non-PCNSL patients (10 with GBM and 10 with metastasis). This provides a comparable number of voxels for each tumor type and each binary classification (i.e., PCSNL vs. non-PCNSL; GBM vs. metastasis for the non-PCNSL cases). The test set consisted of 25 patients with PCNSL, 85 with metastasis, and 290 with GBM (Figure 1). Approximately 50,000 nDSC curves from voxels of the enhancing region in the training group were used to train the classifier. Each nDSC curve corresponds to a specific spatial voxel of the enhancing tumor.

DISCERN outperforms standard classifiers for brain tumor diagnosis

Following a hierarchical classification approach, our CNN method successfully achieved three-way tumoral classification, outperforming the traditional perfusion metrics (i.e., rCBV and PSR) and standing out from simpler binary classifiers. Specifically, for the task of PCNSL diagnosis, DISCERN achieved superior performance, with an accuracy of 0.94 (95% confidence interval [CI]: 0.93–0.94), and mean rCBV and mean PSR classified patients with accuracies of 0.72 (95% CI: 0.70–0.74) and 0.84 (95% CI: 0.83–0.85), respectively. In a second step, patients not classified as PCNSL were categorized as GBM or brain metastasis. DISCERN differentiated GBM from metastases, with an accuracy of 0.81 (95% CI: 0.79–0.82). By contrast, the performance of standard DSC-derived metrics was markedly lower: rCBV classification achieved an accuracy of 0.69 (95% CI: 0.67–0.71) and a mean PSR of 0.65 (95% CI: 0.63–0.67). In Figure 2, the area under the receiver operating characteristic (AUC ROC) curves of the binary classifiers and 3-way average ROC curves are shown for both the DISCERN classifier and conventional rCBV/PSR. For PCNSL vs. non-PCNSL, the CNN provided a significantly higher AUC than rCBV (DeLong test against rCBV: $p = 0.0019$, against PSR, $p = 0.4615$). For GBM vs. metastasis, the CNN resulted in a higher AUC than rCBV and also than PSR (against rCBV: $p = 0.0049$, against PSR, $p < 0.001$).

CellPress
OPEN ACCESS

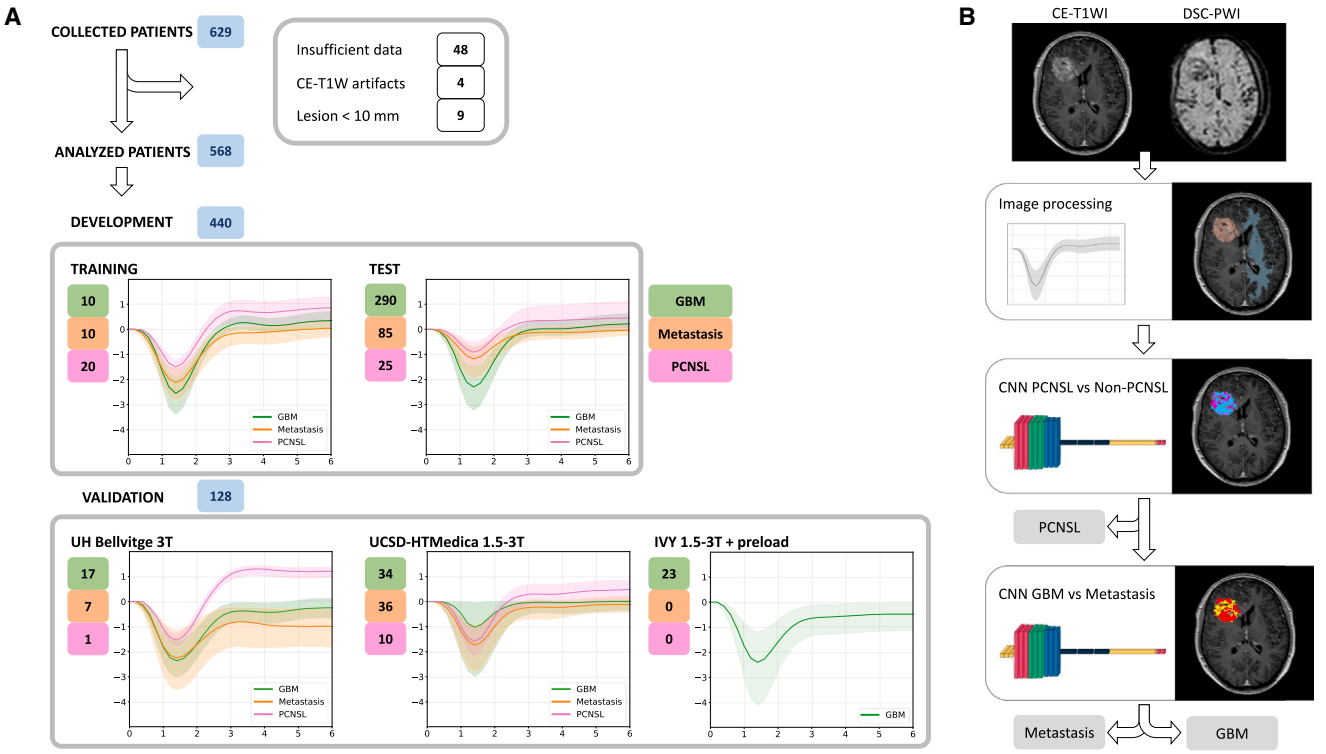


Figure 1. Summary of the population and study design

(A) Collected, excluded, and included data for analysis and further split into the development cohort and external validation cohorts. The number of patients for each tumor type and the respective nDSC perfusion distribution are shown for each cohort.

(B) Processing pipeline of the DISCERN app for 3-way tumor classification of DSC data. At the top, the input images of CE-T1WI for automated region of interest selection and DSC for classification are provided. DSC curves are then extracted voxel-wise from the enhancing tumor and normal white matter in the contralateral hemisphere. The dynamic DSC signals from the enhancing tumor are normalized (nDSC) to the white matter. Every nDSC is classified by 2 sequential CNNs, obtaining a probability map and an overall tumor classification. CE-T1WI, contrast-enhanced T1-weighted imaging.

Lastly, we mimicked a real-world clinical scenario in which our diagnostic support system is confronted with brain lesions comprising the three most common malignancies, in this case represented by the internal test dataset (unseen in the training). In this setting, DISCERN achieved an accuracy of 0.78 (95% CI: 0.76–0.79), which is substantially better than the three-way accuracy achieved using mean rCBV, 0.59 (95% CI: 0.57–0.60) and mean PSR, 0.55 (95% CI: 0.53–0.56). Furthermore, the combination of rCBV and PSR into a logistic regression model also yielded poor performance (Table S2). When validating DISCERN, the tool obtained a three-way classification accuracy of 0.71 evaluating 80 scans of patients from external centers, 0.72 on 25 cases from a 3T scanner (same center as the 1.5T development data), and 0.78 on 23 patients with GBM from the Ivy GAP public dataset (Table S3). These data underscore the potential of DISCERN for differentiating among the three most common clinical diagnostic challenges in patients with enhancing brain lesions.

Voxel-wise explainable representation of the CNN decision process

DISCERN provides spatial probability maps of tumor classification, which are then used to obtain a voxel proportion and a patient classification label. In Figure 2A, we present three examples per tumor type of the voxel-wise probability maps according to the DISCERN classifier. The probability maps are shown overlaid onto the CE-T1W MRI for anatomical references. Overall, the tumor type probability maps are smooth and identify the tumor type with high confidence in most voxels, even when intratumor signal heterogeneity is seen in the contrast-enhanced T1W scan. Voxels exhibiting a high probability of belonging to the incorrect tumor class tend to be located either in the boundary of the enhancing area or around necrotic intratumoral spots. This potentially reflects partial volume (i.e., inclusion of signal from tumor and nontumor areas within a voxel) or intratumoral heterogeneity.

Visual interpretation of the CNN classification

We further sought to implement Class Activation Mapping (CAM) to provide a visual explanation of the DISCERN classification network. The ScoreCAM[48] method yields a normalized score of the contribution of every input to the final classification of a CNN. This allows us to identify the most discriminative time points for nDSC differentiation. ScoreCAM spatial maps were obtained for each binary classification (Figure 3A). The CNN focuses mostly on the bolus passage to classify the central tumor region (middle row for PCNSL vs. non-PCNSL and lower row for GBM vs. metastasis

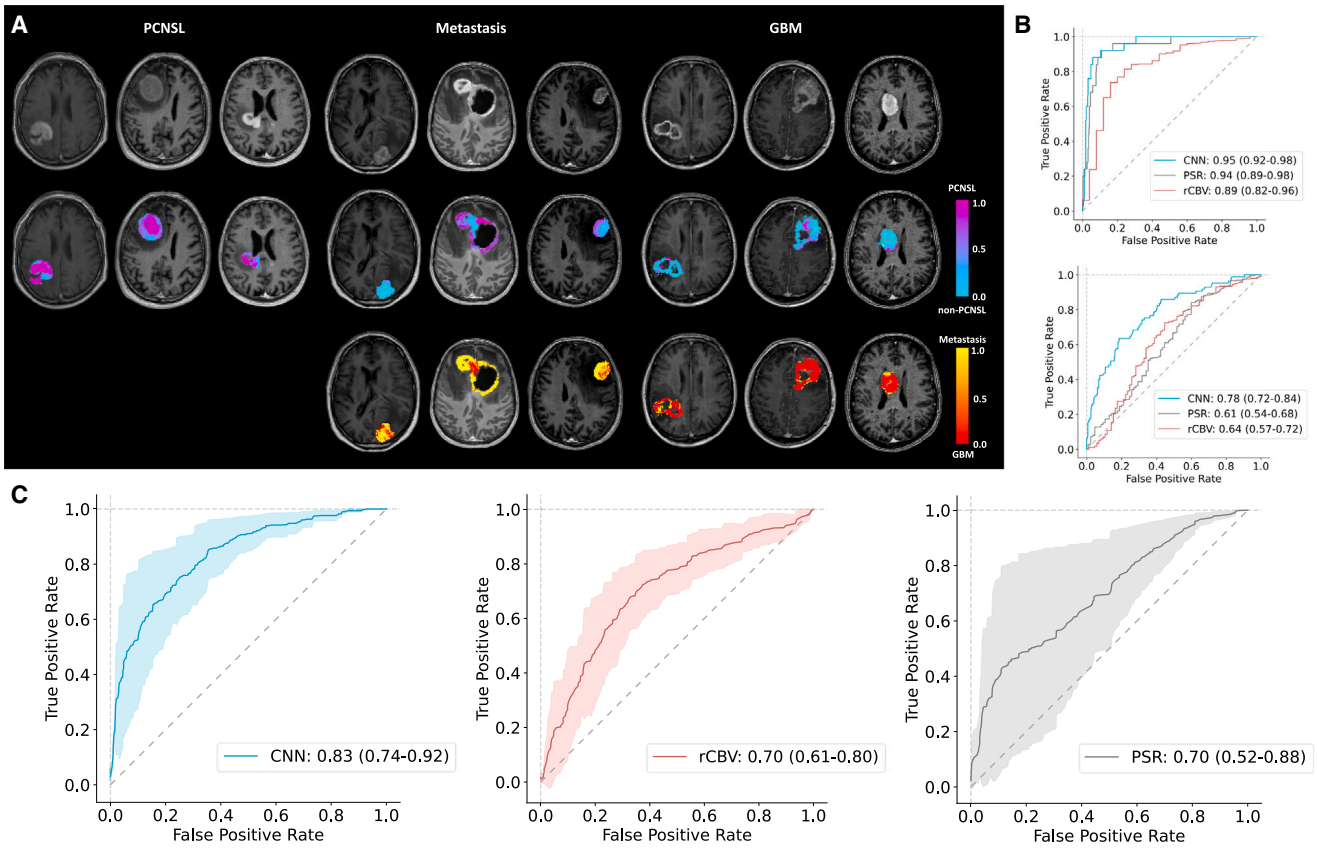


Figure 2. Probability maps and diagnostic performance of DISCERN

(A) Nine cases (3 per tumor type) correctly classified by DISCERN are shown, from left to right: PCNSL, metastasis, and GBM. In the upper row, a representative 2D slice of the CE-T1WI registered to the DSC with overlaid probability maps for PCNSL vs. non-PCNSL (center row) and for GBM vs. metastasis (lower row) of non-PCNSL cases.

(B) ROC curves for binary classifiers PCNSL vs. non-PCNSL (top) and GBM vs. metastasis (bottom) for the proposed CNN, rCBV, and PSR. The CNN provided significantly higher AUC than rCBV for PCNSL vs. non-PCNSL and higher than rCBV and PSR for GBM vs. metastasis.

(C) Three-class ROC curves showing mean and SD of 2-class combinations, from left to right: the proposed CNN, rCBV, and PSR.

differentiation in Figure 3A). In contrast, the bolus passage seems less important for some voxels in surrounding regions. This suggests that the CNN effectively considered the bolus passage as a discerning characteristic, but also that it provides additional tissue perfusion differences compared to the raw perfusion signal (top row in Figure 3A).

The average ScoreCAM values per tumor type and per CNN classifier can be found in Figure 3B (upper row for PCNSL vs. non-PCNSL and lower row for GBM vs. metastasis differentiation). Overall, the sharper signal changes of the nDSC (i.e., steep slopes during contrast arrival and washout) have a higher contribution score. This is especially true for GBM, with greater differences in these time points with respect to the other two tumor types (average nDSC shown in black in Figure 3B). For PCNSL and metastasis, the last part of the signal is also considered important, which can be expected given the overall higher signal magnitude reached in these cases. Importantly, applying 1D CNNs over nDSC signals allowed analysis of the local changes of the signal over time. In this regard, methods that only consider the signal magnitude of specific time points, such as PSR, or a derived measurement such as rCBV, may overlook local nDSC changes occurring over time that reflect specific physiological traits of the tumor.

A user-friendly app to aid brain tumor diagnosis

The DISCERN app was successfully implemented at the participating institutions for validating the tool in external cohorts, as illustrated in Figure 4 and Table S3. The tool requires approximately 2 min to process a new case and provides a classification outcome, in the form of (1) voxel-wise tumor type probability maps and (2) patient-wise tumor type. In addition, it shows the average nDSC for the enhancing tumor and white matter, as well as a visualization of the segmentation for the user to safely check the process. The mask can be automatically segmented from the enhancing tumor by DISCERN or it can be provided by the user. The DISCERN app provides a classification label with balanced sensitivity and specificity (Youden's index) by default, but a given clinical scenario may require a different classification threshold. To that end, sensitivities and specificities for every threshold are displayed, and the default settings can be changed.

In the benchmark study assessing the diagnostic efficacy of our tool in comparison to two neuroradiologists, notable

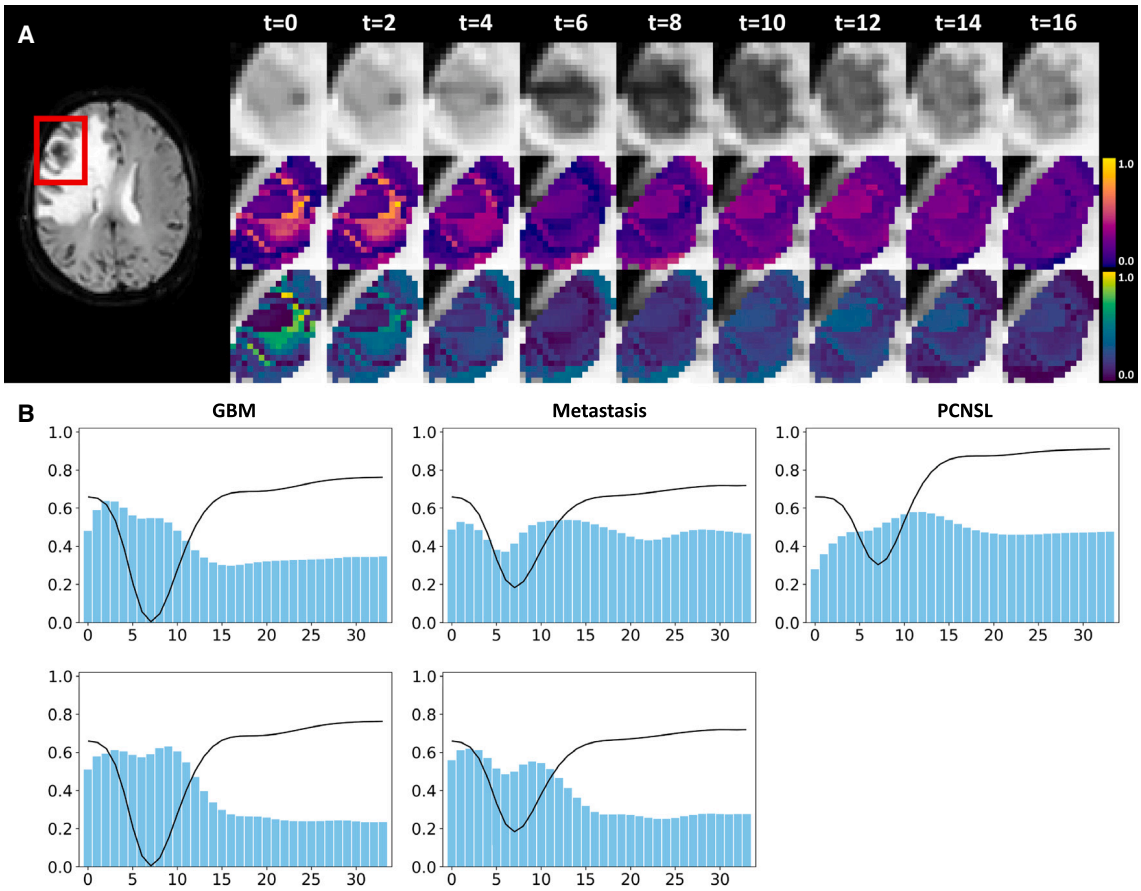


Figure 3. Visual interpretation of the CNN classification

(A) ScoreCAM spatial maps to further understand what the most discriminative nDSC time points for classification per voxel are. We show here a representative case of a metastasis in a 2D slice of the DSC (red box at leftmost). In the upper row, consecutive DSC dynamic time points, zoomed in on the lesion. In the center row, spatial importance score maps obtained with ScoreCAM for PCNSL vs. non-PCNSL and for GBM vs. metastasis in the lower row; the score was scaled to sum 1 over all time points in each voxel to observe relative importance in space.

(B) The average importance of each time point obtained from ScoreCAM that contributes to the tumor classification of nDSC curves, for PCNSL vs. non-PCNSL (upper row) and GBM vs. metastasis (lower row) differentiation; average tumor type nDSC in the training set is overlaid as a black solid line.

distinctions were observed between the senior and junior radiologists, achieving accuracies of 0.80 and 0.40, respectively (it is worth noting that a random chance accuracy is 0.33 in the three-way classification scenario). Within this subset, our tool demonstrated a commendable performance with an accuracy of 0.73, effectively identifying all of the instances of PCNSL by relying solely on perfusion-based information. Furthermore, in cases in which the radiologists exhibited elevated levels of uncertainty (16 out of 30 cases), our tool accurately diagnosed 11 out of these 16 instances (Figure S1). These results underscore the potential of the tool to enhance diagnostic accuracy and reliability, particularly in scenarios characterized by increased diagnostic complexity.

DISCUSSION

We present a voxel-wise method for analyzing perfusion scans with CNNs and improve brain cancer diagnosis, built upon prior DSC signal normalization.[49] By applying this method, we were able to surpass the performance of previous models for noninvasive differential diagnosis of the most frequent malignant brain tumors (i.e., GBM, metastasis, and PCNSL, representing up to 70% of all malignant tumors in the brain[6]), which is critical to define an optimal treatment approach. Notably, our method showcases superior diagnostic capabilities compared to those of neuroradiologists. The potential of DISCERN is to assist radiologists in interpreting brain MRI data, amplifying the diagnostic proficiency of expert neuroradiologists and enabling less experienced counterparts to attain a heightened level of expertise.

Our deep learning framework takes advantage of the large amount of information provided by the thousands of voxel-wise nDSC signals available in each individual DSC MRI scan,[47,50] and achieves optimal performance through training with a limited number of scans from a few patients at fixed DSC protocol (on the order of 30–40 cases). Our approach is particularly appealing for medical imaging applications, in which the design of robust deep learning methods is challenged by the limited number of scans available. In addition, our method

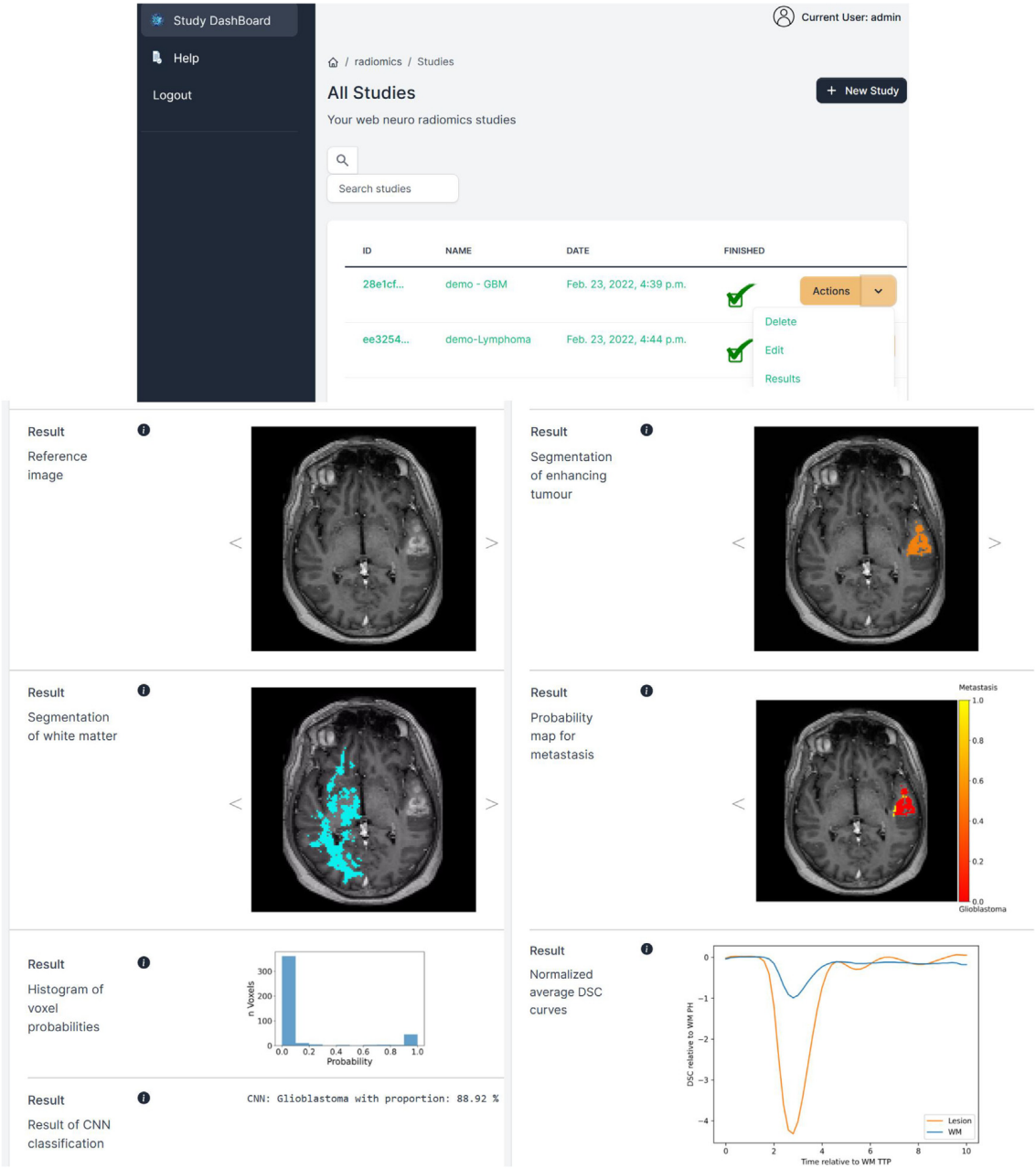


Figure 4. Implementation of the app with an easy-to-use interface

The image illustrates the web interface of the DISCERN app Docker, used to infer results for external validation cohorts. In the uppermost tile, the user dashboard shows the ongoing and finished studies in which to run the pipeline. The next tiles show the results, namely the reference image used for segmentation and respective enhancing tumor and white matter regions, the average nDSC curves of those regions, the tumor probability map and distribution, and a final classification result.

distinguishes between tumor types in a three-way classification task. This can be of particular relevance as a support tool for differential diagnosis in clinical practice, and is a considerable step forward as compared to the current literature, which is dominated by binary classification studies.[11–14,28]

Although three other published works[16,21,46] have included different kinds of three-way classifications among the most common brain tumors (GBM, metastases, and lymphoma), they show some limitations. Liu et al.[16] used the three-way classification as a first step to select the metrics for further pairwise classifications. They considered T1w and T2w anatomical images, with DSC regarded as a future step to improve their results. Tariciotti et al.[46] developed a three-way classification with multiparametric MRI data, using a 2D ResNet model, which would only take advantage of spatial information, but not of dynamic temporal information of DSC, and the reported performance of

the model was only moderate. Wang et al.[21] reported a two-step classification scheme similar to ours, and using only conventional DSC metrics, with the main limitation being the lack of validation in external cohorts.

In contrast, we present the largest dataset reported in this context. The voxel-by-voxel approach to the classification of DSC data takes advantage of inherently redundant information. Furthermore, the 1D CNN takes into account the changes in the temporal profile of the bolus passage, which other methods ignore or simplify. It also produces probability maps, facilitating the visual inspection of the spatial distribution of the classification. Finally, we have validated the results in external cohorts, demonstrating the generalizability and potential utility of our findings.

A noninvasive diagnostic support tool is especially relevant when considering PCNSL among potential diagnoses. Corticosteroids are a reasonable therapeutic option for mitigating neurological symptoms secondary to edema in patients with malignant brain tumors. However, early stereotactic biopsy before corticosteroid administration is mandatory when a brain PCNSL is suspected, because medication with steroids can alter the histological pattern of PCNSL.[5] Moreover, PCNSL is highly sensitive to chemoradiotherapy instead of resection, which is contraindicated, as opposed to GBM or metastasis. Therefore, a reliable characterization of the tumor type by imaging is critical to devise the appropriate management of patients.

The DISCERN app provides voxel-wise tumor type probability maps, which are then used to obtain a voxel proportion and a patient classification label. The default Youden's index (tradeoff between sensitivity and specificity) can be changed to the needs of different clinical scenarios. For instance, some medical cases may require a very high specificity for suspected GBM and metastases with respect to PCNSL and, if all evidence supports it, then an additional intervention for a biopsy could be prevented. Therefore, the voxel proportion can be adjusted in the app to match the user's needs.

The presented method successfully achieved three-way tumoral classification, outperforming the traditional perfusion metrics, and standing out from simpler binary classifiers. When tested, our method performed with accuracies of 0.94 for PCNSL identification, 0.81 for differentiation of GBM from metastasis, and 0.78 for three-way classification.

Of note, the DSC protocol used for model development did not include contrast preload. Contrast preload is a common approach described in the literature to achieve a better estimate of the rCBV.[42] However, preload can be undesirable for a number of reasons. First, it delivers a higher contrast dose to the patient. Second, it can introduce potential variability sources, affecting the nDSC signal morphology. As countermeasures, leakage correction and acquisition parameters that minimize the T1 effect, such as low flip angle, have been shown to effectively yield reliable rCBV estimates without preload.[42,44] In this study, we only had access to retrospective data with high flip angle; consequently, rCBV was estimated with leakage correction, obtaining results comparable to those of PSR. The combination of both rCBV and PSR in logistic regression was explored for completeness, but it did not improve the results of the individual parameters. It is noteworthy that even though the DSC protocol used a nonoptimal flip angle for PSR or rCVB quantification, DISCERN achieved a good level of discriminatory performance. This was further confirmed by validating the results using heterogeneous external validation data, indicating a high level of robustness in the method.

A key feature of our CNN approach is the computation of voxel-wise spatial representations of perfusion curve characteristics in the form of maps describing the probability of a voxel to belong to a specific tumor type. Such spatial probability maps provide an explainable representation of the CNN decision process and may enable further studies of intratumor heterogeneity, making them an appealing tool for integrative multi-omics research and also of potential clinical interest to plan surgical procedures. Few studies have applied deep learning to voxel-wise DSC signals in neuro-oncological applications. A recent study[34] used a deep autoencoder to derive a set of five descriptors of DSC that could differentiate between pairs of tumor types. However, the reconstructed time signals from such a minimal set of descriptors, in contrast to the original signal, produce a smooth perfusion curve morphology, which may omit relevant details for diagnostic applications.

In conclusion, the presented CNN framework for three-class brain tumor classification based on voxel-wise DSC signal analysis is feasible and outperforms classifiers built on conventional rCBV and PSR metrics. The method can be trained using a limited number of scans, which most centers are likely to have available, with notable generalization to external data. In addition, it provides voxel-wise maps of tumor type signatures that could be useful to visualize the CNN classification process and for tumor spatial characterization. As a way to make this tool more accessible and eventually make an impact in clinical practice, the proposed method has been implemented on the user-friendly DISCERN application, which is made freely accessible at http://84.88.64.102:5000/discern-app, to enhance study reproducibility and accelerate its adoption in future clinical studies.

Limitations of the study

The diagnostic tool DISCERN was trained with perfusion MRI data from scans without preload contrast injection, which may limit its performance on preloaded MRI scans. Tests on all eligible external 3T scans with a contrast preload of 23 patients with GBM from the IvyGAP[51] dataset yielded 18 cases correctly classified as GBM (0.78 accuracy). Further testing in preload MRI scans should be performed to explore the generalizability of the results in this context.

The study used automatically segmented regions of interest, revised by an experienced neuroradiologist. However, the variability in segmentations among different neuroradiologists has not been explored. The proposed future segmentation methods aim to minimize manual input, but this aspect requires further development.

The training data were obtained from only 40 patients, which may introduce a bias toward the inherent characteristics of this subpopulation. To account for patient and scanner biases, we normalized the signals to those of white matter and we trained with an equal number of patients for every malignancy. As a limitation of the retrospective nature of this study, older diagnostic

standards were in place and isocitrate dehydrogenase mutation status was missing from this study, which would provide a cleaner glioblastoma cohort.

The current algorithm is limited to certain MRI sequences (T1-weighted and perfusion MRI) and does not yet incorporate other potentially useful image data, such as diffusion MRI,[52] which may offer more detailed insights into tumor microstructure and potentially improve the performance of the tool.

DISCERN is user-friendly and can classify three common brain tumors, but its application to other tumor types is still under development. In addition, although the framework shows promise, it requires further clinical qualification and approval for use as a medical diagnostic tool.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cohort and clinical characteristics
- METHOD DETAILS
  - Patient eligibility
  - Imaging acquisition parameters
  - Data pre-processing
  - Conventional DSC-PWI metrics
  - CNN architecture
  - Classification scheme
  - Classification performance
  - CNN interpretation
  - Utility in aiding medical decisions for brain tumor diagnosis
  - Development of the online app DISCERN
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2024.101464.

ACKNOWLEDGMENTS

This project was supported by "La Caixa" Foundation (RTI2018-095209-B-C21) and the Spanish Ministry of Science and Innovation (FIS-G64384969). R.P.-L. is supported by the Prostate Cancer Foundation Young Investigator Award, the FERO Foundation, the CRIS Foundation Talent Award (TALENT19-05), the Instituto de Salud Carlos III Investigacion en Salud (PI21/01019), and the Asociacion Espanola Contra el Cancer (PRYCO211023-SERR). F.G. was funded by the Government of Catalonia (Beatriu de Pinos 2020 00117 BP) and by the Fundacio LaCaixa (ID 100010434, code LCF/BQ/PR22/11920010). C.M. and A.P.-E. acknowledge support from the Instituto de Salud Carlos III-Investigación en Salud (PI20/00360). We would like to express our sincere appreciation to Javier Carmona for his valuable support and assistance in reviewing the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, C.M., A.P.-E., and R.P.-L. Methodology, A.G.-R., F.G., and A.P.-E. Software, C.M.-A. and R.K. Validation, G.H., R.K., T.M.S., A.L.-R., L.O., M.A.B., A.C.-Z., T.M.-N., A.L., and T.M.S. Formal analysis, A.G.-R. Investigation, A.G.-R., L.O., and A.C.-Z. Resources, C.M., A.P.-E., M.L., G.H., T.M.S., A.L.-R., L.O., M.A.B., A.C.-Z., T.M.-N., and A.L. Data curation, P.N.-B., A.P.-E., G.H., and M.A.B. Writing – original draft, A.G.-R. Writing – review & editing, R.P.-L., C.M., A.P.-E., P.N.-B., F.G., T.M.S., and A.G.-R. Funding acquisition, R.P.-L., C.M., and A.P.-E. Supervision, R.P.-L., C.M., and F.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 30, 2023
Revised: November 16, 2023
Accepted: February 15, 2024
Published: March 11, 2024

REFERENCES

1. Young, R.M., Jamshidi, A., Davis, G., and Sherman, J.H. (2015). Current trends in the surgical management and treatment of adult glioblastoma. Ann. Transl. Med. *3*, 121.
2. Hatiboglu, M.A., Wildrick, D.M., and Sawaya, R. (2013). The role of surgical resection in patients with brain metastases. Ecancermedicalscience *7*, 308.
3. Hoang-Xuan, K., Bessell, E., Bromberg, J., Hottinger, A.F., Preusser, M., Rudà, R., Schlegel, U., Siegal, T., Soussain, C., Abacioglu, U., et al. (2015). Diagnosis and treatment of primary CNS lymphoma in immunocompetent patients: guidelines from the European Association for Neuro-Oncology. Lancet Oncol. *16*, e322–e332.
4. Dammers, R., Haitsma, I.K., Schouten, J.W., Kros, J.M., Avezaat, C.J.J., and Vincent, A.J.P.E. (2008). Safety and efficacy of frameless and frame-based intracranial biopsy techniques. Acta Neurochir. *150*, 23–29.
5. Chiavazza, C., Pellerino, A., Ferrio, F., Cistaro, A., Soffietti, R., and Rudà, R. (2018). Primary CNS Lymphomas: Challenges in Diagnosis and Monitoring. BioMed Res. Int. *2018*, 3606970.
6. Miller, K.D., Ostrom, Q.T., Kruchko, C., Patil, N., Tihan, T., Cioffi, G., Fuchs, H.E., Waite, K.A., Jemal, A., Siegel, R.L., and Barnholtz-Sloan, J.S. (2021). Brain and other central nervous system tumor statistics. CA A Cancer J. Clin. *71*, 381–406.
7. Leung, D., Han, X., Mikkelsen, T., and Nabors, L.B. (2014). Role of MRI in primary brain tumor evaluation. J. Natl. Compr. Cancer Netw. *12*, 1561–1568.
8. Arita, K., Miwa, M., Bohara, M., Moinuddin, F.M., Kamimura, K., and Yoshimoto, K. (2020). Precision of preoperative diagnosis in patients with brain tumor - A prospective study based on "top three list" of differential diagnosis for 1061 patients. Surg. Neurol. Int. *11*, 55.
9. Chakravorty, A., Steel, T., and Chaganti, J. (2015). Accuracy of percentage of signal intensity recovery and relative cerebral blood volume derived from dynamic susceptibility-weighted, contrast-enhanced MRI in the preoperative diagnosis of cerebral tumours. NeuroRadiol. J. *28*, 574–583.
10. Law, M., Cha, S., Knopp, E.A., Johnson, G., Arnett, J., and Litt, A.W. (2002). High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology *222*, 715–721.
11. Fordham, A.J., Hacherl, C.C., Patel, N., Jones, K., Myers, B., Abraham, M., and Gendreau, J. (2021). Differentiating Glioblastomas from Solitary Brain Metastases: An Update on the Current Literature of Advanced Imaging Modalities. Cancers *13*, 2960.

12. Artzi, M., Bressler, I., and Ben Bashat, D. (2019). Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. J. Magn. Reson. Imag. *50*, 519–528.

13. Bae, S., An, C., Ahn, S.S., Kim, H., Han, K., Kim, S.W., Park, J.E., Kim, H.S., and Lee, S.K. (2020). Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. Sci. Rep. *10*, 12110.

14. Qian, Z., Li, Y., Wang, Y., Li, L., Li, R., Wang, K., Li, S., Tang, K., Zhang, C., Fan, X., et al. (2019). Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. Cancer Lett. *451*, 128–135.

15. Maurer, M.H., Synowitz, M., Badakshi, H., Lohkamp, L.N., Wüstefeld, J., Schäfer, M.L., and Wiener, E. (2013). Glioblastoma multiforme versus solitary supratentorial brain metastasis: differentiation based on morphology and magnetic resonance signal characteristics. Röfo *185*, 235–240.

16. Liu, D., Chen, J., Ge, H., Hu, X., Yang, K., Liu, Y., Hu, G., Luo, B., Yan, Z., Song, K., et al. (2022). Differentiation of malignant brain tumor types using intratumoral and peritumoral radiomic features. Front. Oncol. *12*, 848846.

17. Liu, Y., Xu, X., Yin, L., Zhang, X., Li, L., and Lu, H. (2017). Relationship between Glioblastoma Heterogeneity and Survival Time: An MR Imaging Texture Analysis. AJNR. Am. J. Neuroradiol. *38*, 1695–1701.

18. Lee, E.J., Ahn, K.J., Lee, E.K., Lee, Y.S., and Kim, D.B. (2013). Potential role of advanced MRI techniques for the peritumoural region in differentiating glioblastoma multiforme and solitary metastatic lesions. Clin. Radiol. *68*, e689–e697.

19. Yang, G., Jones, T.L., Barrick, T.R., and Howe, F.A. (2014). Discrimination between glioblastoma multiforme and solitary metastasis using morphological features derived from the p:q tensor decomposition of diffusion tensor imaging. NMR Biomed. *27*, 1103–1111.

20. Bauer, A.H., Erly, W., Moser, F.G., Maya, M., and Nael, K. (2015). Differentiation of solitary brain metastasis from glioblastoma multiforme: a predictive multiparametric approach using combined MR diffusion and perfusion. Neuroradiology *57*, 697–703.

21. Wang, S., Kim, S., Chawla, S., Wolf, R.L., Knipp, D.E., Vossough, A., O'Rourke, D.M., Judy, K.D., Poptani, H., and Melhem, E.R. (2011). Differentiation between glioblastomas, solitary brain metastases, and primary cerebral lymphomas using diffusion tensor and dynamic susceptibility contrast-enhanced MR imaging. AJNR. Am. J. Neuroradiol. *32*, 507–514.

22. Neska-Matuszewska, M., Bladowska, J., Sąsiadek, M., and Zimny, A. (2018). Differentiation of glioblastoma multiforme, metastases and primary central nervous system lymphomas using multiparametric perfusion and diffusion MR imaging of a tumor core and a peritumoral zone-Searching for a practical approach. PLoS One *13*, e0191341.

23. Swinburne, N.C., Schefflein, J., Sakai, Y., Oermann, E.K., Titano, J.J., Chen, I., Tadayon, S., Aggarwal, A., Doshi, A., and Nael, K. (2019). Machine learning for semi-automated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. Ann. Transl. Med. *7*, 232.

24. Priya, S., Ward, C., Locke, T., Soni, N., Maheshwarappa, R.P., Monga, V., Agarwal, A., and Bathla, G. (2021). Glioblastoma and primary central nervous system lymphoma: differentiation using MRI derived first-order texture analysis - a machine learning study. NeuroRadiol. J. *34*, 320–328.

25. Alcaide-Leon, P., Dufort, P., Geraldo, A.F., Alshafai, L., Maralani, P.J., Spears, J., and Bharatha, A. (2017). Differentiation of Enhancing Glioma and Primary Central Nervous System Lymphoma by Texture-Based Machine Learning. AJNR. Am. J. Neuroradiol. *38*, 1145–1150.

26. Ahn, S.J., Shin, H.J., Chang, J.H., and Lee, S.K. (2014). Differentiation between primary cerebral lymphoma and glioblastoma using the apparent diffusion coefficient: comparison of three different ROI methods. PLoS One *9*, e112948.

27. Toh, C.H., Wei, K.C., Chang, C.N., Ng, S.H., and Wong, H.F. (2013). Differentiation of primary central nervous system lymphomas and glioblastomas: comparisons of diagnostic performance of dynamic susceptibility contrast-enhanced perfusion MR imaging without and with contrast-leakage correction. AJNR. Am. J. Neuroradiol. *34*, 1145–1149.

28. Cha, S., Lupo, J.M., Chen, M.H., Lamborn, K.R., McDermott, M.W., Berger, M.S., Nelson, S.J., and Dillon, W.P. (2007). Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. AJNR. Am. J. Neuroradiol. *28*, 1078–1084.

29. Lee, M.D., Baird, G.L., Bell, L.C., Quarles, C.C., and Boxerman, J.L. (2019). Utility of Percentage Signal Recovery and Baseline Signal in DSC-MRI Optimized for Relative CBV Measurement for Differentiating Glioblastoma, Lymphoma, Metastasis, and Meningioma. AJNR. Am. J. Neuroradiol. *40*, 1445–1450.

30. Barajas, R.F., Jr., Chang, J.S., Segal, M.R., Parsa, A.T., McDermott, M.W., Berger, M.S., and Cha, S. (2009). Differentiation of recurrent glioblastoma multiforme from radiation necrosis after external beam radiation therapy with dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. Radiology *253*, 486–496.

31. Hu, L.S., Baxter, L.C., Pinnaduwage, D.S., Paine, T.L., Karis, J.P., Feuerstein, B.G., Schmainda, K.M., Dueck, A.C., Debbins, J., Smith, K.A., et al. (2010). Optimized preload leakage-correction methods to improve the diagnostic accuracy of dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging in posttreatment gliomas. AJNR. Am. J. Neuroradiol. *31*, 40–48.

32. Kim, Y.E., Choi, S.H., Lee, S.T., Kim, T.M., Park, C.K., Park, S.H., and Kim, I.H. (2017). Differentiation between Glioblastoma and Primary Central Nervous System Lymphoma Using Dynamic Susceptibility Contrast-Enhanced Perfusion MR Imaging: Comparison Study of the Manual versus Semiautomatic Segmentation Method. Investig. Magn. Reson. Imaging *21*, 9.

33. Mangla, R., Kolar, B., Zhu, T., Zhong, J., Almast, J., and Ekholm, S. (2011). Percentage signal recovery derived from MR dynamic susceptibility contrast imaging is useful to differentiate common enhancing malignant lesions of the brain. AJNR. Am. J. Neuroradiol. *32*, 1004–1010.

34. Park, J.E., Kim, H.S., Lee, J., Cheong, E.N., Shin, I., Ahn, S.S., and Shim, W.H. (2020). Deep-learned time-signal intensity pattern analysis using an autoencoder captures magnetic resonance perfusion heterogeneity for brain tumor differentiation. Sci. Rep. *10*, 21485.

35. Pons-Escoda, A., Garcia-Ruiz, A., Naval-Baudin, P., Grussu, F., Fernandez, J.J.S., Simo, A.C., Sarro, N.V., Fernandez-Coello, A., Bruna, J., Cos, M., et al. (2022). Voxel-level analysis of normalized DSC-PWI time-intensity curves: a potential generalizable approach and its proof of concept in discriminating glioblastoma and metastasis. Eur. Radiol. *32*, 3705–3715.

36. Surendra, K.L., Patwari, S., Agrawal, S., Chadaga, H., and Nagadi, A. (2020). Percentage signal intensity recovery: A step ahead of rCBV in DSC MR perfusion imaging for the differentiation of common neoplasms of brain. Indian J. Cancer *57*, 36–43.

37. Zhang, J., Liu, H., and Tong, H. (2017). Clinical Applications of Contrast-Enhanced Perfusion MRI Techniques in Gliomas: Recent Advances and Current Challenges. Contrast Media Mol. Imaging *2017*, 7064120.

38. Bell, L.C., Hu, L.S., Stokes, A.M., McGee, S.C., Baxter, L.C., and Quarles, C.C. (2017). Characterizing the Influence of Preload Dosing on Percent Signal Recovery (PSR) and Cerebral Blood Volume (CBV) Measurements in a Patient Population With High-Grade Glioma Using Dynamic Susceptibility Contrast MRI. Tomography *3*, 89–95.

39. Bell, L.C., Semmineh, N., An, H., Eldeniz, C., Wahl, R., Schmainda, K.M., Prah, M.A., Erickson, B.J., Korfiatis, P., Wu, C., et al. (2020). Evaluating the Use of rCBV as a Tumor Grade and Treatment Response Classifier Across NCI Quantitative Imaging Network Sites: Part II of the DSC-MRI Digital Reference Object (DRO) Challenge. Tomography *6*, 203–208.

40. Fu, R., Szidonya, L., Barajas, R.F., Jr., Ambady, P., Varallyay, C., and Neuwelt, E.A. (2022). Diagnostic performance of DSC perfusion MRI to

distinguish tumor progression and treatment-related changes: a systematic review and meta-analysis. Neurooncol. Adv. *4*, vdac027.

41. Boxerman, J.L., Paulson, E.S., Prah, M.A., and Schmainda, K.M. (2013). The effect of pulse sequence parameters and contrast agent dose on percentage signal recovery in DSC-MRI: implications for clinical applications. AJNR. Am. J. Neuroradiol. *34*, 1364–1369.
42. Paulson, E.S., and Schmainda, K.M. (2008). Comparison of dynamic susceptibility-weighted contrast-enhanced MR methods: recommendations for measuring relative cerebral blood volume in brain tumors. Radiology *249*, 601–613.
43. Boxerman, J.L., Quarles, C.C., Hu, L.S., Erickson, B.J., Gerstner, E.R., Smits, M., Kaufmann, T.J., Barboriak, D.P., Huang, R.H., Wick, W., et al. (2020). Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. Neuro Oncol. *22*, 1262–1275.
44. Schmainda, K.M., Prah, M.A., Hu, L.S., Quarles, C.C., Semmineh, N., Rand, S.D., Connelly, J.M., Anderies, B., Zhou, Y., Liu, Y., et al. (2019). Moving Toward a Consensus DSC-MRI Protocol: Validation of a Low-Flip Angle Single-Dose Option as a Reference Standard for Brain Tumors. AJNR. Am. J. Neuroradiol. *40*, 626–633.
45. Open Science Initiative for Perfusion Imaging (OSIPI). (2023). OSIPI/osipi.github.io. GitHub. https://github.com/OSIPI.
46. Tariciotti, L., Ferlito, D., Caccavella, V.M., Di Cristofori, A., Fiore, G., Remore, L.G., Giordano, M., Remoli, G., Bertani, G., Borsa, S., et al. (2022). A Deep Learning Model for Preoperative Differentiation of Glioblastoma, Brain Metastasis, and Primary Central Nervous System Lymphoma: An External Validation Study. NeuroSci *4*, 18–30.
47. Hess A, Meier R, Kaesmacher J Jung, S., Scalzo, F., Liebeskind, D., Wiest, R. and McKinley, R. Synthetic Perfusion Maps: Imaging Perfusion Deficits in DSC-MRI with Deep Learning. InBrainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4 2019 (pp. 447-455). Springer International Publishing.
48. Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. In Paper presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
49. Pons-Escoda, A., Garcia-Ruiz, A., Naval-Baudin, P., Cos, M., Vidal, N., Plans, G., Bruna, J., Perez-Lopez, R., and Majos, C. (2020). Presurgical Identification of Primary Central Nervous System Lymphoma with Normalized Time-Intensity Curve: A Pilot Study of a New Method to Analyze DSC-PWI. AJNR. Am. J. Neuroradiol. *41*, 1816–1824.
50. Grussu, F., Blumberg, S.B., Battiston, M., Kakkar, L.S., Lin, H., Ianuş, A., Schneider, T., Singh, S., Bourne, R., Punwani, S., et al. (2021). Feasibility of Data-Driven, Model-Free Quantitative MRI Protocol Design: Application to Brain and Prostate Diffusion-Relaxation Imaging. Front. Physiol. *9*.
51. Shah, N., Feng, X., Lankerovich, M., Puchalski, R.B., and Keogh, B. (2016). Data from Ivy Glioblastoma Atlas Project (IvyGAP): The Cancer Imaging Archive.
52. Nilsson, M., Englund, E., Szczepankiewicz, F., van Westen, D., and Sundgren, P.C. (2018). Imaging brain tumour microstructure. Neuroimage *182*, 232–250.
53. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A., Lankerovich, M., Bertagnolli, D., Bickley, K., et al. (2018). An anatomic transcriptional atlas of human glioblastoma. Science *360*, 660–663.
54. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digit. Imag. *26*, 1045–1057.
55. Data from the Multi-Institutional Paired Expert Segmentations and Radiomic Features of the Ivy GAP Dataset. The Cancer Imaging Archive (TCIA), (2020). https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70222827.
56. Pati, S., Verma, R., Akbari, H., Bilello, M., Hill, V.B., Sako, C., Correa, R., Beig, N., Venet, L., Thakur, S., et al. (2020). Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. Med. Phys. *47*, 6039–6052.
57. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn. Reson. Imaging *30*, 1323–1341.
58. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., and Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. *114*, 97–109.
59. van Gelderen, P., Duyn, J.H., Ramsey, N.F., Liu, G., and Moonen, C.T.W. (2012). The PRESTO technique for fMRI. Neuroimage *62*, 676–681.
60. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: improved N3 bias correction. IEEE Trans. Med. Imag. *29*, 1310–1320.
61. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., and Wells, W.M. (2007). A hierarchical algorithm for MR brain image parcellation. IEEE Trans. Med. Imag. *26*, 1201–1212.

4.3 Study 3

Whole-body Magnetic Resonance Imaging as a Treatment Response Biomarker in Castration-resistant Prostate Cancer with Bone Metastases: The iPROMET Clinical Trial

European Urology (2024).**Impact Factor (quartile)**: 25.3 (Q1)

Authors and affiliations: Alonso Garcia-Ruiz[1], Carlos Macarro[1], Francesca Zacchi[1,2,3], Rafael Morales-Barrera[1,2], Francesco Grussu[1], Irene Casanova-Salas[1], Francesco Sanguedolce[4,5], Macarena Gonzalez[1,2], Pablo Cresta-Morgado[1,2], Matias de Albert[2], Josep Garcia-Bennett[6], David Marmolejo[1,2], Jacques Planas[2], Sarai Roche[2], Richard Mast[2], Christina Zatse[1], Josep M Piulats[6,7], Bernardo Herrera-Imbroda[8], Lucas Regis[2], Laura Agundez[1], David Olmos[9], Nahum Calvo[6], Manuel Escobar[2], Joan Carles[1,2], Joaquin Mateo[1,2] and Raquel Perez-Lopez[1].

1. Vall d'Hebron Institute of Oncology, Barcelona, Spain.
2. Vall d'Hebron University Hospital, Barcelona, Spain.
3. Section of Innovation Biomedicine-Oncology Area, Department of Engineering for Innovation Medicine (DIMI), University of Verona and University and Hospital Trust (AOUI) of Verona, Verona, Italy.
4. Fundació Puigvert, Institut de Recerca Sant Pau, Barcelona, Spain.
5. Department of Medicine, Surgery and Pharmacy, Universitá degli Studi di Sassari, Italy.
6. Bellvitge University Hospital, L'Hospitalet del Llobregat, Spain.
7. Institut Catala d'Oncologia, L'Hospitalet del Llobregat, Spain.
8. Hospital Universitario Virgen de la Victoria and Instituto de Investigación Biomédica de Málaga IBIMA-Plataforma Bionand, Malaga, Spain.
9. Hospital Doce de Octubre and Instituto de Investigacion i+12, Madrid, Spain.

This study corresponds to **objective 3** of the thesis. In this work, markers derived from whole-body MRI of patients undergoing treatment were extracted and analysed to look for associations to response to therapy. The project aimed at supporting MRI as the preferred imaging modality for the evaluation of response in bone metastases. The PhD candidate was in charge of data collection, data management and inspection, image processing, feature extraction, statistical analyses and manuscript drafting.

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Research Letter

Whole-body Magnetic Resonance Imaging as a Treatment Response Biomarker in Castration-resistant Prostate Cancer with Bone Metastases: The iPROMET Clinical Trial

Alonso Garcia-Ruiz[a], *Carlos Macarro*[a], *Francesca Zacchi*[a,b,c], *Rafael Morales-Barrera*[a,b], *Francesco Grussu*[a], *Irene Casanova-Salas*[a], *Francesco Sanguedolce*[d,e], *Macarena Gonzalez*[a,b], *Pablo Cresta-Morgado*[a,b], *Matias de Albert*[b], *Josep Garcia-Bennett*[f], *David Marmolejo*[a,b], *Jacques Planas*[b], *Sarai Roche*[b], *Richard Mast*[b], *Christina Zatse*[a], *Josep M Piulats*[f,g], *Bernardo Herrera-Imbroda*[h], *Lucas Regis*[b], *Laura Agundez*[a], *David Olmos*[i], *Nahum Calvo*[f], *Manuel Escobar*[b], *Joan Carles*[a,b], *Joaquin Mateo*[a,b,*], *Raquel Perez-Lopez*[a,*]

Evaluation of bone metastasis in metastatic prostate cancer (mPC) remains a clinical challenge, as computed tomography and bone scans are unable to capture the response of these metastases to therapy. This is relevant for patient management and drug development, as radiographic response is a common intermediate endpoint in phase 1/2 clinical trials [1].

We and others previously reported that whole-body magnetic resonance imaging (WB-MRI) including diffusion-weighted (DW) imaging allows quantitative assessment of bone metastasis; specific guidelines for interpreting WB-MRI in clinical practice have now been proposed [2]. DW-MRI provides information on tissue cellularity, primarily via the apparent diffusion coefficient (ADC), and on fat content (fat fraction, FF) [3]; as the fatty bone-marrow repopulates the metastatic niche on killing of tumour cells, FF could be relevant for response evaluation [4].

We conducted a multicentre clinical trial of WB-MRI in patients with mPC (iPROMET; NCT05078151). Patients with mPC and bone metastases starting standard-of-care systemic therapies (androgen receptor signalling inhibitors [ARSIs]: abiraterone acetate, enzalutamide, apalutamide, or taxane-based chemotherapy) underwent WB-MRI at treatment baseline, after 4 and 12 wk of treatment, and, when possible, at disease progression. We assessed the correlation of imaging biomarkers to response and time to progression (TTP). Response was defined as either a Response Evaluation Criteria in Solid Tumours (RECIST) v1.1 response and/or a 50% decrease in prostate-specific antigen (PSA_{50}) from baseline. TTP was defined as the time from treatment initiation to radiographic progression or unequivocal clinical progression triggering treatment discontinuation. Further details on the study design are available in the Supplementary material.

A total of 88 patients participated in the study, of whom 74 received therapy and underwent at least one follow-up MRI examination. Here we report results for all 55 evaluable patients with metastatic castration-resistant prostate cancer (mCRPC; 52 ARSI; 3 chemotherapy). Progression events were observed in 36/55 patients (median follow-up 11.13 mo), with a median TTP of 8.28 mo. The combined RECIST/PSA_{50} response rate was 41/55 (75%).

At week 12, median FF on therapy was higher for responding patients (17.7, interquartile range [IQR] 13.2–22.1) than for nonresponding patients (10.34, IQR 8.4–16.8; $p < 0.001$). For bone metastatic volume on DWI at week 12, a relative decrease was observed for responding patients (median −24.0%, IQR −47.0% to +1.8%) and a relative increase for nonresponding patients (median +12.4%, IQR −4.3% to +26.3%; $p = 0.024$). An ADC increase >35% (predefined in the study) on therapy was observed in three cases, all of whom were responding patients. However, the response rate was also high (36/48, 75%) among patients who did not achieve a >35% increase in ADC.

Both absolute bone metastatic volume (hazard ratio [HR] 1.004; $p = 0.02$) and FF (HR 0.92; $p < 0.005$) at week 12 correlated with TTP. No significant associations for absolute ADC values were observed. Combining volume, FF, and ADC into multivariable models resulted in a concordance

https://doi.org/10.1016/j.eururo.2024.02.016
0302-2838/© 2024 European Association of Urology. Published by Elsevier B.V. All rights reserved.

Please cite this article as: A. Garcia-Ruiz, C. Macarro, F. Zacchi et al., Whole-body Magnetic Resonance Imaging as a Treatment Response Biomarker in Castration-resistant Prostate Cancer with Bone Metastases: The iPROMET Clinical Trial, Eur Urol (2024), https://doi.org/10.1016/j.eururo.2024.02.016

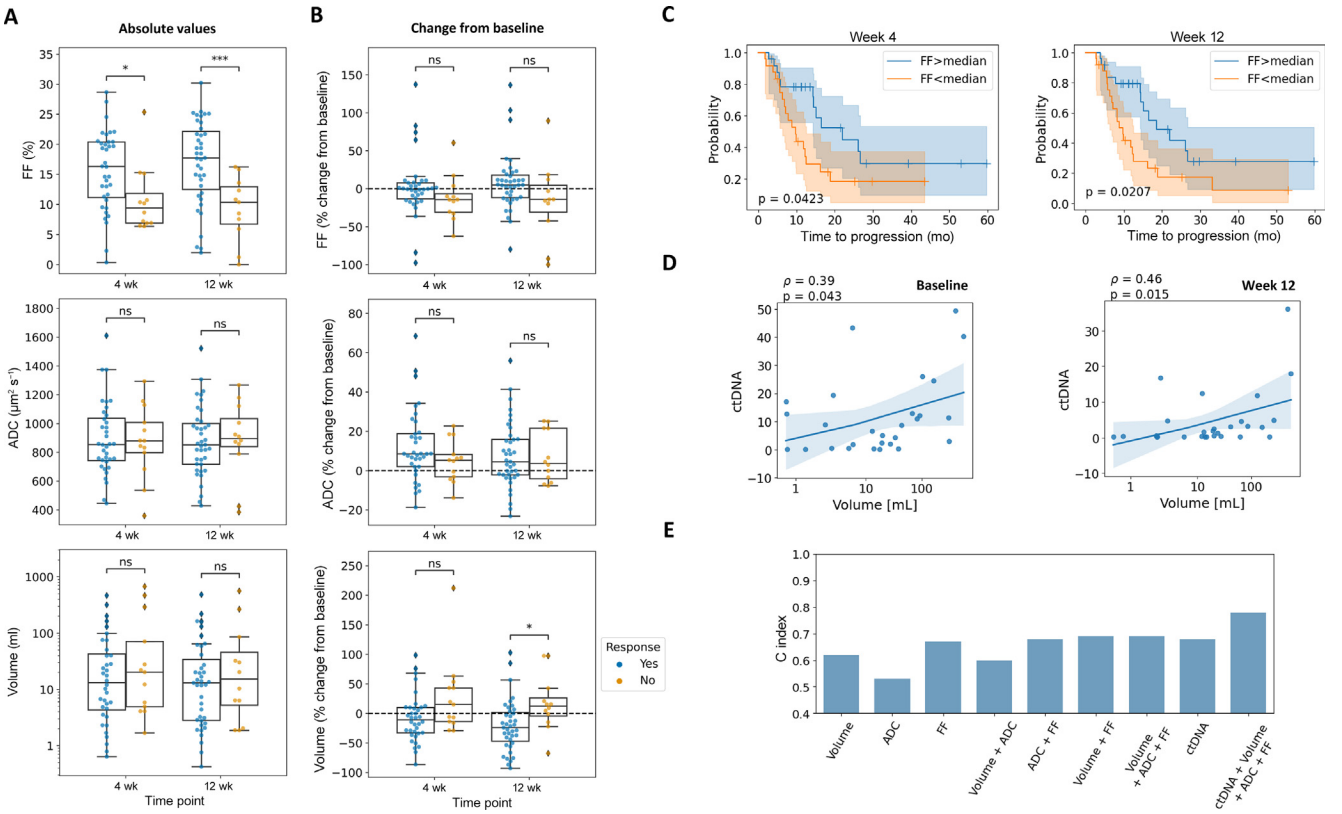


Fig. 1 – Differences in imaging-derived metrics between responding (blue) and nonresponding (yellow) patients. (A) Absolute values and (B) relative change in the volume of bone disease, apparent diffusion coefficient (ADC), and fat fraction (FF) on magnetic resonance imaging (MRI) at 4 and 12 wk after treatment initiation (Welch's t test: ns = not significant; * $p \le 0.05$; *** $p \le 0.001$). (C) Kaplan-Meier curves for time to progression for patients stratified by the median FF value for the population at 4-wk and 12-wk follow-up. The difference is shown as the *p* value from log-rank tests. (D) Spearman's rank correlation between the circulating tumour (ctDNA) fraction and the volume of bone disease via whole-body MRI. (E) Concordance index (C index) from Cox regression models with time to progression as the outcome.

index (C index) ranging from 0.60 (volume + ADC) to 0.69 (volume + ADC + FF). According to these models, higher bone-disease volume (HR 1.003, 95% confidence interval [CI] 1.000–1.006), lower FF (HR 0.899, 95% CI 0.841–0.960), and lower ADC (HR 0.999, 95% CI 0.997–1.001) at week 12 were associated with a higher risk of progression.

Blood samples for assessment of circulating tumour DNA (ctDNA) were collected from 31 patients. The ctDNA fraction was positively correlated with the volume of bone disease at baseline (ρ = 0.39, p = 0.043) and at week 12 (ρ = 0.46, p = 0.015). On multivariable analysis, addition of ctDNA to the imaging variables increased the C index from 0.68 to 0.78 (Fig. 1E).

The key findings from our study are the association between decreasing volume of bone metastasis, and higher FF, with response to therapy. Prior WB-MRI studies focused on ADC [5]. In this trial, some responding patients presented ADC increased in some responding patients and decreased in others; the ADC association with outcome improved after adjusting for FF at the individual-patient level. We hypothesise that low ADC at follow-up in some responding patients was because of an increase in fat infiltration in responding bones; hence, we conclude that fat infiltration should be considered when interpreting ADC results. This is further supported by the observation that in ten bone metastatic biopsies acquired in the study, MRI-derived FF correlated with pathology evaluation of the adipose fraction (R = 0.809; p = 0.0003).

We acknowledge that the use of multiple MRI scanners is a limitation of this analysis; common imaging protocols were defined across participating sites to control for potential bias. Another limitation is the sample size, with 55 evaluable mCRPC patients; arguably, this still represents the largest multicentre prospective trial of WB-MRI in mCRPC. Future studies that include other functional imaging techniques, such as prostate-specific membrane antigen positron emission tomography, would help in defining the optimal use of imaging tools for precise monitoring of mPC.

In conclusion, our study demonstrates that MRI-derived biomarkers are valuable indicators of patient benefit in mCRPC. Multivariable combinations of imaging variables improved outcome predictions. Moreover, we showed that multi-omics approaches combining MRI and ctDNA hold promise for monitoring patients with mCRPC.

***Conflicts of interest*:** Joaquin Mateo has served as an advisor for AstraZeneca, Amunix/Sanofi, Daichii-Sankyo, Janssen, MSD; Pfizer, and Roche; is a scientific board member for Nuage Therapeutics; is involved as an investigator in several company-sponsored clinical trials; and is principal investigator for institutional research funds from AstraZeneca, Amgen, and Pfizer, none of them related to this work. Raquel Perez-Lopez has received research funding from AstraZeneca and Roche, and serves on a

Please cite this article as: A. Garcia-Ruiz, C. Macarro, F. Zacchi et al., Whole-body Magnetic Resonance Imaging as a Treatment Response Biomarker in Castration-resistant Prostate Cancer with Bone Metastases: The iPROMET Clinical Trial, Eur Urol (2024), https://doi.org/10.1016/j.eururo.2024.02.016

steering committee for a clinical trial sponsored by Roche, not related to this work. The remaining authors have nothing to disclose.

***Acknowledgments*:** This project was supported by a Prostate Cancer Foundation Young Investigator Award to Raquel Perez-Lopez, and grants from Fundacio La Marato de TV3 (grant agreement 201934-30)and Instituto de Salud Carlos III (PI18/01395). Raquel Perez-Lopez and Joaquin Mateo are also supported by a CRIS Foundation Talent Award (TALENT19-05 and TALENT20-10), Asociacion Espanola Contra el Cancer (PRYCO211023SERR and LABAE20019MATE) and the FERO Foundation. Francesco Grussu was supported by AGAUR (Beatriu de Pinos Programme 2020-BP-00117). We acknowledge support from Fundacio LaCaixa (ID 100010434, grant LCF/BQ/PR22/11920010 to Francesco Grussu and grant LCF/BQ/PI20/11760033 to Irene Casanova-Salas). Vall d'Hebron Institut d'Oncologia (VHIO) authors would like to acknowledge the Spanish State Agency for Research (Agencia Estatal de Investigación) for financial support as a Centre of Excellence Severo Ochoa (CEX2020-001024-S/AEI/10.13039/501100011033), the Cellex Foundation for providing research facilities and equipment, and the Generalitat de Catalunya CERCA Programme for their support of this research. We would like to acknowledge all the patients who participate in this and other clinical research projects for their contribution to science. Victor Navarro, biostatistician at VHIO, kindly provided advice on statistical analyses.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eururo.2024.02.016.

References

[1] Scher HI, Morris MJ, Stadler WM, et al. Trial design and objectives for castration-resistant prostate cancer: updated recommendations from the Prostate Cancer Clinical Trials Working Group 3. J Clin Oncol 2016;34:1402–18.

[2] Padhani AR, Lecouvet FE, Tunariu N, et al. Metastasis Reporting and Data System for Prostate Cancer: practical guidelines for acquisition, interpretation, and reporting of whole-body magnetic resonance imaging-based evaluations of multiorgan involvement in advanced prostate cancer. Eur Urol 2017;71:81–92.

[3] Perez-Lopez R, Nava Rodrigues D, Figueiredo I, et al. Multiparametric magnetic resonance imaging of prostate cancer bone disease: correlation with bone biopsy histological and molecular features. Invest Radiol 2018;53:96–102.

[4] Latifoltojar A, Hall-Craggs M, Rabin N, et al. Whole body magnetic resonance imaging in newly diagnosed multiple myeloma: early changes in lesional signal fat fraction predict disease response. Br J Haematol 2017;176:222–33.

[5] Perez-Lopez R, Mateo J, Mossop H, et al. Diffusion-weighted Imaging as a treatment response biomarker for evaluating bone metastases in prostate cancer: a pilot study. Radiology 2017;283:168–77.

[a] *Vall d'Hebron Institute of Oncology, Barcelona, Spain*
[b] *Vall d'Hebron University Hospital, Barcelona, Spain*
[c] *Section of Innovation Biomedicine-Oncology, Department of Engineering for Innovation Medicine, University of Verona and University and Hospital Trust of Verona, Verona, Italy*
[d] *Fundació Puigvert, Institut de Recerca Sant Pau, Barcelona, Spain*
[e] *Department of Medicine, Surgery and Pharmacy, Universitá degli Studi di Sassari, Sassari, Italy*
[f] *Bellvitge University Hospital, L'Hospitalet del Llobregat, Spain*
[g] *Institut Catala d'Oncologia, L'Hospitalet del Llobregat, Spain*
[h] *Hospital Universitario Virgen de la Victoria and Instituto de Investigación Biomédica de Málaga-Plataforma Bionand, Malaga, Spain*
[i] *Hospital Doce de Octubre and Instituto de Investigacion i+12, Madrid, Spain*

*Corresponding authors. Vall d'Hebron Institute of Oncology, 115–117 Natzaret, Cellex Center, 08035 Barcelona, Spain. Tel. +34 932543450.
E-mail address: jmateo@vhio.net (J. Mateo), rperez@vhio.net (R. Perez-Lopez).

February 26, 2024

CHAPTER 5

Global discussion

Medical image analysis has experienced major developmental breakthroughs from the digitalization of data and the implementation of imaging archiving systems to the late methods of analysis, feature extraction and deep learning. However, the guidelines for major clinical steps within the management of cancer like diagnosis, prognosis and treatment monitoring, still rely on subjective assessments of the images, based on visual cues and annotations.

With cancer rates becoming ever more concerning, radiomics computerized features, machine and deep learning techniques are hoped to help characterize the disease, support medical decisions and automate tasks, providing better reproducibility and generalization of methods that can escalate to eventually improve the management of cancer patients. Among the different imaging modalities, MRI offers a wide range of contrasts that can be obtained from the same scanner, a unique feature that no other modality has. Besides anatomical information, MRI can also acquire sequences that relate to biophysical properties of the body tissues, such as vasculature and tissue microstructure by means of perfusion and diffusion MRI. Furthermore, metabolic and molecular information can also be obtained with special techniques [83–85, 87]. Metrics derived from MRI have already shown value for various medical applications and there is great potential for novel biomarker development.

In this thesis, three specific clinical needs were identified where MRI-derived markers could help in the characterization of cancer to support medical decisions, in the context of cancer diagnosis, prognosis and evaluation of treatment. Novel approaches were proposed to improve over previously described methods, validating the results by comparing against established reference methods and looking for further evidence from biological samples.

In **Study 1**, metrics quantifying the tumour remnant were defined and associated with patient survival. The study aimed at providing quantifiable and objective metrics, as opposed to prior visual assessments. In addition, the optimal timing of the MRI scan after

surgery was observed to be within 24 hours and 72 hours for prognosis assessment. The ability to identify patients with longer and shorter survival can impact therapy options and monitoring, and further improving and developing new biomarkers.

In **Study 2**, neural networks were explored in DSC MRI to differentiate among three types of cancer. Those are the most common brain malignancies diagnosed in adults and the differential diagnosis on medical images is a complex task of high relevance for surgery and treatment options. The method was compared against previously described metrics, against neuroradiologists with different experience and validated in external data. The implementation was also publicly released.

In **Study 3**, the focus is on the monitoring of response during treatment for bone metastases using MRI, that otherwise cannot be evaluated with CT by RECIST [4]. The strong association of the fat fraction points at its potential use as a quantifiable biomarker. Although the MET-RADS-P guideline [25] covers bone metastases, only the ADC is quantified, with Dixon MRI relegated for visual inspection only. With the results obtained, it is hoped that further studies explore the FF and replicate the results, so that the guideline can be updated to take it into account.

Given the heterogeneity and complexity of cancer disease, the management of cancer has been switching latey from a global view to specific solutions unique to each tumour. For a long time, cytotoxic drugs of broad spectrum, such as traditional chemotherapy, were the only treatment. Such approach may be overly aggressive, in the sense that both tumoral and healthy cells would be killed, resulting in high toxicity and adverse side effects. The discovery and understanding of the biological mechanisms of the cells, by the analysis of genes, proteins, but also of histological assays of the tissues and radiological images, have made it possible to start differentiating between types of tumours and developing better, specific solutions.

In the same way, the studies described in this thesis focus in the practical implementation of solutions for three clinical scenarios that correspond to real needs in cancer management. By tackling one specific problem at a time, the proposed methods are meant to become a useful contribution in the characterization and assessment of cancer. Notably, the development of imaging biomarkers follows common steps, even when there may be differences in the acquired images or the body region under study. Some common procedures include the definition or extraction of features, the association to the patient well-being and the validation of the proposed methods.

Medical imaging biomarkers have demostrated their value in critical steps like screening and follow-up, currenly described in guidelines of reference. Medical image analysis is a very active research discipline, recently boosted with solutions based on neural networks and deep learning, that is hoped to bring better, faster biomarkers for cancer disease.

5.1 Bridging the gap between research and clinical practice

While it was shown that the development of imaging markers for cancer management is possible, there are still multiple factors to consider to bring imaging-derived features to be used in medical settings. The establishment of novel markers requires a long process of validation and implementation steps, facing multiple challenges. That makes it difficult for researchers to pursue and for the novel methods to achieve their intended purpose. Some of those challenges and proposed solutions in the included studies will be discussed next.

5.1.1 The need for collaborative research

Among the considerations for novel biomarkers to be widely accepted, a key aspect is the validation. A comprehensive, multi-centric and large-scale validation is necessary to show how the biomarker works, how robust it is and how well it generalizes. That would build confidence in other researchers and companies to include the biomarker in clinical trials and, as a result, propelling its use, popularity and development. In this sense, a major factor for the recognition of new biomarkers may be the visibility of the proposed method and how many people it reaches. As a step towards higher visibility and impact, the implementation of **Study 2** was made publicly available. However, visibility also relates to the specific use-case and the rate of occurrence of the disease in the general population. As an example, a rare disease may remain a niche topic, as fewer researchers would be working on it. Paradoxically, a very crowded research topic may be detrimental to potential biomarkers, as many different approaches may be proposed for the same purpose, a long-standing problem of research that was visibly criticized lately [88].

Specifically for the prognosis assessment based on the remaining tumour after surgery, another study was developed in parallel to **Study 1** [30], proposing a different metric, which was not replicated when applied to the cohort in **Study 1**. Whether that was related to methodological differences or population differences, collaborative efforts between research groups could bring together many benefits. A single study with a unique methodology, a

larger dataset that may better represent the general population, and a higher visibility and impact achieved. In view of that, multiple institutions were involved to participate in **Study 2**, especially regarding the validation of the method.

However, dealing with larger datasets is a double-edged sword, as it usually implies higher heterogeneity in the imaging acquisition and inherent demographic differences in populations, which demands extensive data curation and methods handling missing data, data imbalance, biases and confounding factors. That complexity rises exponentially when considering multi-modality images and other sources of clinical information, a setback that was apparent in **Study 3**, with the added complexity of whole-body images. Methods for the homogenization of images, registration of images and normalization of values are known to facilitate the analysis and mitigate potential issues, but there is still a need for better, automated methods.

More and more, publicly available datasets in repositories or challenges [89, 90], are being published that provide anonymized curated data ready for use, enabling a better development and validation of tools and markers for medical purposes. As such, public datasets were explored and used in **Study 2** for testing and validation. Challenges also support competitions for developing the best algorithm for a specific task, sometimes yielding significant research advances. An advantage of such contests is that holding the methods under a single contest allows a fair comparison, something often difficult to do from single, isolated studies.

5.1.2 Implementation and cost-effectiveness

A key aspect on the deployment of novel imaging biomarkers is the technical knowledge. The implementation of a new system would need to comply with regulatory standards of data privacy and safety of use. It may need to communicate with other systems in the hospital or provide the means to access data easily [33]. It is possible that the hospital staff needs training and that an expert maintains the system. In the case of medical images, that means dealing with the DICOM (Digital Imaging and Communication in Medicine) standard, which oftentimes is not met by different vendors and the overall complexity makes the implementation unattractive.

Fortunately, existing software such as Docker enables building a self-contained application with all necessary packages, simplifying a part of the process. This was explored in the implementation of the methods from **Study 2** for use in external centres. Still, a dedicated

application needs continuous maintenance during real deployment, which is another difficulty in bringing new solutions forward.

Besides that, medical images may pose additional challenges, such as handling 3D or higher dimensional data, as well as relatively large image volumes. This was an issue in **Study 3** with whole-body images, where processing steps such as registration required powerful computing resources. Arguably, affordable computing power is available nowadays in many shapes, including cloud services. Other approaches to handle large images include the analysis by patches or small regions, so it is expected that intensive processing would not be problematic in any case.

5.2 Inherent limitations of reference valuesground truth

Current methods to evaluate the response to treatment or the characterization of cancer are imperfect. That also means that comparing novel markers with the current references may not completely reflect their potential. In particular, the definition of response by RECIST is based on the manual measurements of the tumour diameter of a few lesions only, which may not completely represent the disease burden.

The validation against the established markers and outcomes proved challenging to apply in **Study 3**, where MRI was supposed to provide response information from lesions that were not possible to measure from current guidelines on CT imaging. In fact, two criteria are used in prostate cancer patients, the presence of prostate-specific antigen (PSA) [91] and the standard radiological criteria [4], which does not cover osteoblastic metastases, the subject of **Study 3**. That means that, for a number of patients, only the PSA could be regarded as reference of response to validate the proposed markers. The PSA, though an established biomarker in prostate cancer, is very variable from patient to patient and the percentage of relative change of the PSA has been proposed with different cut-off values [92]. Moreover, aggressive prostate cancers, like those undergoing neuroendocrine transformation, may not express PSA. This limits the usefulness of PSA dynamics as a response biomarker in these often-metastatic tumours.

Alternatively, survival information is an established outcome in the development of biomarkers. However, survival outcomes are not free from limitations, as mentioned in the introduction. The OS is arguably the most unambiguous of them all by definition, although it does not take into account deaths unrelated to the disease, which in smaller populations may introduce outliers. Additionally, it usually takes a longer time to collect. The PFS and

TTP both need an evaluation of the tumour progression, which also suffers from limitations, as described. Ideally, both OS and PFS would be used as reference, as in **Study 1**. In **Study 3**, a prospective study with more recent data, OS was not available for all patients. Additionally, as it was a smaller population, TTP was used in **Study 3** and no conflicts were detected among different criteria for patient response.

Given those issues, relating imaging markers to results from biopsy analysis or histological features is a valuable addition in the validation of new methods. In **Study 1** and **Study 2**, patient diagnosis was confirmed by histological analysis, the gold standard technique, and relations to the tissue properties with imaging markers were explored in **Study 3** to support the hypothesis and results. Even so, biopsy samples also suffer from limitations. The obtained sample may be only a small portion of the affected region, so there is the risk that the sample is not completely representative of the disease. Furthermore, the diagnostic rate of biopsies may depend on the experience of the physician [93], the location of the tumour and other uncontrollable variables, which further supports the discovery of non-invasive imaging biomarkers.

Additionally, it must be noted the role of manual annotations. Current imaging markers are based on the delineation of areas of interest by an expert radiologist, which are considered the ground truth. This step is crucial in the development of imaging biomarkers and a considerable bottleneck in any practical implementation of computerized imaging features, besides introducing inter-observer variability. In the case of **Study 1 and 2**, it was possible to identify the region of interest for most patients via simple image processing, which then were inspected for any required corrections. However, this was not possible to perform for **Study 3**. Newer methods of automated segmentation are showing impressive results on medical images [49, 94, 95] and could be a major breakthrough for research and medical purposes.

5.3 Prospects of artificial intelligence in medicine

The rise of models based on neural networks and transformers have brought impressive feats on a number of applications, some of them with a medical focus. Their flexibility with the ability to capture complex representations from virtually any input, and an end-to-end pipeline of feature extraction and model fitting make them a very attractive approach for current research. In fact, the rationale for using convolutional neural networks in **Study 2** was that it was deemed suitable to extract as much information from the dynamic data as

possible, contrasting with previous methods.

In the revolution that AI is experiencing, it may seem as if these methods were the ultimate answer to any question or task. However, there are still limitations to address, especially when deployed for sensitive applications such as healthcare services.

5.3.1 Data availability

Because the models rely on many layers, and therefore many parameters to fit, training of such models generally needs large pools of data, which also helps prevent overfitting and make the model generalize better. A small data size is often a challenge in medical imaging studies, where data is inherently limited by the available digital data collected and by the incidence rate of the disease or condition under study.

Multicentric studies may be able to include more patients, but such collaborations require a considerable effort in terms of coordination, logistics, ethical permissions and access to technical knowledge. In this regard, open anonymized datasets are ever growing and becoming a powerful leverage for model development and validation, as well as for the creation of contests, such as the MICCAI challenges, where groups or individuals compete to develop the best model for a predefined task [89]. When the number of images available is low, transfer learning can be used, which refers to the retraining or fine-tuning of an already trained model but on different types of images. Other approaches consist on taking advantage of redundant data in single scans, such as the voxel-wise method used in **Study 2**, which enabled obtaining thousands of 1D data from a few patients, but also splitting the image in smaller patches that can serve as input samples, such as the works developed for segmentation and on histological images [50, 95].

5.3.2 Explainability

A proper understanding of the learned representations from deep learning is still a work in progress that could boost the confidence of users and the implementation of these models, especially in critical applications like healthcare. To address this, a method of explainability was used in **Study 2**, showing the relevant information of the perfusion dynamics to achieve a diagnosis. Moreover, the model was tested on different datasets to demonstrate its generalisation. Arguably, comprehensive testing of a model under different conditions and inputs, and further validation on other datasets or against established biomarkers can alternatively indicate the safe use of a model, something that can be missing or disregarded in published

methods [88, 96].

In contrast, hand-crafted features have fixed definitions corresponding to image patterns and linear models can be readily understood as a weighted sum of their elements. That was the case for the metrics proposed in **Study 1**. On specific imaging acquisitions, models based on the biophysical mechanisms that influence the measured signal provide a natural way of linking the imaging data with biological traits, as well as inherently knowing the assumptions and limitations of the model. That was the case of the ADC and FF in **Study 3**. In biomarker development, those are very desirable characteristics and, ideally, the best aspects from all techniques could be eventually integrated in the future, that is, flexible self-tuned models that can still be parametrized or trained with constraints based on biophysical or biological evidence.

5.3.3 Multi-modal data

Interestingly, deep learning intra-model comparisons have been used to shed some light into the learned representations [97, 98], finding similarities between the representations of models, especially larger ones. Some studies have explored the multi-modality aspect of model representations [99] and there is the hypothesis that models are converging towards a unified representation space the more information they include [100], explaining the improved performance and the benefit of multi-modal approaches. In fact, foundational models have emerged as multi-purpose methods capable of wide generalization [52, 94].

Multi-modal and multidisciplinary approaches have long been regarded as one of the keys to integrate information for cancer characterization, and actually guidelines take into account multiple sources of data. The combination of metrics including clinical variables, texture features, multiparametric MRI and others was explored in **all three studies** of the thesis, whenever possible. But computerized methods still struggle in this aspect, possibly lacking a unified effort of common methodology and data sharing. Looking at the rapid development of AI models, it is possible that the community effort in improving AI would indirectly help in the development of conventional models via the sharing of datasets, trained representations and novel methodologies.

5.3.4 Bias and privacy concerns

The potential of AI applications in clinical settings is undeniable. However, several key concerns require attention before widespread implementation. These concerns include both

potential bias within the models and the privacy of patient data. Biases in the training data, such as imbalances in ethnicity and sex of patients in medical data can make the models be less sensitive to specific subpopulations, which warrants the use of techniques of bias mitigation as well as comprehensive testing and validation of models [101, 102]. In this sense, statistical tests can be used to evaluate if sex and cancer type were balanced in defined populations, as in **Study 2**.

As for privacy, building such models and bringing them forward as approved medical tools require sharing data, often involving private companies that provide technical expertise, which has triggered concerns for potential data breaching or misuse of sensitive and confidential data [103]. For **all studies** presented in this work, appropriate communication links, anonymization, storage and management of data were used to ensure data safety and privacy. Still, there is active research focusing on anonymizing, deidentification and sharing techniques, such as encryption and federated learning, as well as attempting to reconstruct data from trained models [104]. Although bias and privacy issues have been long considered in the sharing and analysis of medical data, the use of large datasets and data sharing for AI implementations must come with the responsibility to protect confidential information.

In view of the aforementioned points, it would be expected that current AI methods dominate in a number of tasks, but remain heavily application-dependent. Until now, AI methods shine with large training datasets and for relatively simple, repetitive tasks. Furthermore, many research questions require a level of understanding of the physical or biological assumptions made, for which explicit mathematical modelling may be preferred, at least until the interpretation of neural networks is improved. Still, overly-complex problems such as predicting protein folding and new protein structures may be readily attainable with AI [105], which points at the dilemma of trading interpretability for blind performance.

In the medical context, any device, software or biomarker must pass several validation steps for its safe, ethical and legal use, which on the one hand makes the development harder to attain, but on the other hand guarantees its safety for medical purposes. Nevertheless, multiple AI-related and machine learning methods have already been approved for clinical use [106]. All in all, neural network models have come to stay and evolve. It would seem that those disciplines that can access large enough datasets would probably benefit from AI models in the short term. For more niche fields, hybrid models or alternative approaches may still enable the use of these novel methods.

CHAPTER 6

Conclusions

1. Computerized metrics from MRI allow to quantify the remaining enhancing tumour after surgery in patients with glioblastoma multiforme, especially in the scans acquired between 24 and 72 hours after surgery. The proposed metrics showed associations with the survival of the patients, demonstrating prognostic value for clinical applications. The use of quantifiable features can improve over the visual assessment of images and help in the reproducibility and generalization of methods.

2. The differential diagnosis of glioblastoma multiforme, solitary brain metastasis and primary central nervous system lymphoma can be performed by a model based on convolutional neural networks on normalized DSC MRI data. The voxel-by-voxel approach allows training such models with few MRI scans. The proposed method performed with remarkable accuracy as compared with conventional parameters and with neuroradiologists, as well as in external data. The method provides probability maps of the tumour diagnosis voxel by voxel, potentially serving as a computer-aided diagnostic tool for supporting medical decisions.

3. The volume of disease, ADC and FF of bone metastases derived from whole-body diffusion-weighted MRI were associated with response to treatment and with survival of patients with advanced prostate cancer. The combination of imaging and clinical parameters improved the accuracy of survival regression models, suggesting the multimodality approaches for the development of biomarkers. The fat content in bone metastases computed by means of the FF showed a strong association with survival. The role of the fat in the healing bone may play a major role in the interpretation of the ADC and should be taken into special consideration for the assessment of patients with bone metastases.

Bibliography

[1] A. Jassim, E. P. Rahrmann, B. D. Simons, *et al.*, "Cancers make their own luck: Theories of cancer origins," *Nat Rev Cancer*, vol. 23, no. 10, pp. 710–724, 2023, Type: Journal Article, ISSN: 1474-1768 (Electronic) 1474-175X (Linking). DOI: 10.1038/s41568-023-00602-5.

[2] C. Swanton, E. Bernard, C. Abbosh, *et al.*, "Embracing cancer complexity: Hallmarks of systemic disease," *Cell*, vol. 187, no. 7, pp. 1589–1616, 2024, Type: Journal Article, ISSN: 1097-4172 (Electronic) 0092-8674 (Linking). DOI: 10.1016/j.cell.2024.02.009.

[3] F. S. Hodi, S. J. O'Day, D. F. McDermott, *et al.*, "Improved Survival with Ipilimumab in Patients with Metastatic Melanoma," *New England Journal of Medicine*, vol. 363, no. 8, pp. 711–723, Aug. 2010, ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa1003466.

[4] E. A. Eisenhauer, P. Therasse, J. Bogaerts, *et al.*, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *Eur J Cancer*, vol. 45, no. 2, pp. 228–47, 2009, Type: Journal Article, ISSN: 1879-0852 (Electronic) 0959-8049 (Linking). DOI: 10.1016/j.ejca.2008.10.026.

[5] I. Matos Garcia, A. García Ruiz, J. Martin-Liberal, *et al.*, "Refining criteria of hyperprogression (HPD) with immune checkpoint inhibitors (ICIs) to improve clinical applicability," *Annals of Oncology*, vol. 29, p. viii653, Oct. 2018, ISSN: 09237534. DOI: 10.1093/annonc/mdy303.011.

[6] C. Suárez, R. Morales-Barrera, A. Garcia-Ruiz, *et al.*, "Hyperprogressive disease in patients with metastatic genitourinary tumors treated with immune checkpoint inhibitors," vol. 37, no. 7_suppl, pp. 448–448, 2019. DOI: 10.1200/JCO.2019.37.7_suppl.448.

[7] I. Matos, J. Martin-Liberal, A. Garcia-Ruiz, *et al.*, "Capturing Hyperprogressive Disease with Immune-Checkpoint Inhibitors Using RECIST 1.1 Criteria," *Clin Cancer*

Res, vol. 26, no. 8, pp. 1846–1855, Apr. 2020, Edition: 2019/11/24, ISSN: 1557-3265 (Electronic) 1078-0432 (Linking). DOI: 10.1158/1078-0432.CCR-19-2226.

[8] P. Sharma, S. Hu-Lieskovan, J. A. Wargo, *et al.*, "Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy," *Cell*, vol. 168, no. 4, pp. 707–723, Feb. 2017, ISSN: 00928674. DOI: 10.1016/j.cell.2017.01.017.

[9] A. Marusyk, M. Janiszewska, and K. Polyak, "Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance," *Cancer Cell*, vol. 37, no. 4, pp. 471–484, Apr. 2020, ISSN: 15356108. DOI: 10.1016/j.ccell.2020.03.007.

[10] F. Bray, M. Laversanne, H. Sung, *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 74, no. 3, pp. 229–263, 2024, Type: Journal Article, ISSN: 1542-4863 (Electronic) 0007-9235 (Linking). DOI: 10.3322/caac.21834.

[11] A. F. Shields and P. Price, "Role of Imaging in Cancer Treatment," in *In Vivo Imaging of Cancer Therapy*, A. F. Shields and P. Price, Eds., Totowa, NJ: Humana Press, 2007, pp. 1–12, ISBN: 978-1-58829-633-7 978-1-59745-341-7. DOI: 10.1007/978-1-59745-341-7_1.

[12] L. Fass, "Imaging and cancer: A review," *Mol Oncol*, vol. 2, no. 2, pp. 115–52, 2008, Type: Journal Article, ISSN: 1878-0261 (Electronic) 1574-7891 (Print) 1574-7891 (Linking). DOI: 10.1016/j.molonc.2008.04.001.

[13] C. P. Kratz, M. I. Achatz, L. Brugières, *et al.*, "Cancer Screening Recommendations for Individuals with Li-Fraumeni Syndrome," *Clinical Cancer Research*, vol. 23, no. 11, e38–e45, Jun. 2017, ISSN: 1078-0432, 1557-3265. DOI: 10.1158/1078-0432.CCR-17-0408.

[14] S. Paluch-Shimon, F. Cardoso, C. Sessa, *et al.*, "Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening," *Annals of Oncology*, vol. 27, pp. v103–v110, Sep. 2016, ISSN: 09237534. DOI: 10.1093/annonc/mdw327.

[15] A. M. D. Wolf, K. C. Oeffinger, T. Y.-C. Shih, *et al.*, "Screening for lung cancer: 2023 guideline update from the American Cancer Society," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 1, pp. 50–81, Jan. 2024, ISSN: 0007-9235, 1542-4863. DOI: 10.3322/caac.21811.

[16] F. Peintinger, "National Breast Screening Programs across Europe," *Breast Care (Basel)*, vol. 14, no. 6, pp. 354–358, 2019, Type: Journal Article, ISSN: 1661-3791 (Print) 1661-3805 (Electronic) 1661-3791 (Linking). DOI: 10.1159/000503715.

[17] D. Crosby, S. Bhatia, K. M. Brindle, *et al.*, "Early detection of cancer," *Science*, vol. 375, no. 6586, eaay9040, 2022, Type: Journal Article, ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aay9040.

[18] K. Arita, M. Miwa, M. Bohara, *et al.*, "Precision of preoperative diagnosis in patients with brain tumor - A prospective study based on "top three list" of differential diagnosis for 1061 patients," *Surg Neurol Int*, vol. 11, p. 55, 2020, Type: Journal Article, ISSN: 2229-5097 (Print) 2152-7806 (Linking). DOI: 10.25259/sni_5_2020.

[19] D. N. Louis, A. Perry, P. Wesseling, *et al.*, "The 2021 WHO Classification of Tumors of the Central Nervous System: A summary," *Neuro Oncol*, vol. 23, no. 8, pp. 1231–1251, 2021, Type: Journal Article, ISSN: 1523-5866 (Electronic) 1522-8517 (Print) 1522-8517 (Linking). DOI: 10.1093/neuonc/noab106.

[20] R. Di Bonaventura, N. Montano, M. Giordano, *et al.*, "Reassessing the Role of Brain Tumor Biopsy in the Era of Advanced Surgical, Molecular, and Imaging Techniques-A Single-Center Experience with Long-Term Follow-Up," *J Pers Med*, vol. 11, no. 9, 2021, Type: Journal Article, ISSN: 2075-4426 (Print) 2075-4426 (Electronic) 2075-4426 (Linking). DOI: 10.3390/jpm11090909.

[21] P. Y. Wen, M. van den Bent, G. Youssef, *et al.*, "RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults," *J Clin Oncol*, vol. 41, no. 33, pp. 5187–5199, 2023, Type: Journal Article, ISSN: 1527-7755 (Electronic) 0732-183X (Print) 0732-183X (Linking). DOI: 10.1200/JCO.23.01059.

[22] G. Biomarkers Definitions Working, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework," *Clin Pharmacol Ther*, vol. 69, no. 3, pp. 89–95, 2001, Type: Journal Article, ISSN: 0009-9236 (Print) 0009-9236 (Linking). DOI: 10.1067/mcp.2001.113989.

[23] S. Chauvie, L. N. Mazzoni, and J. O'Doherty, "A Review on the Use of Imaging Biomarkers in Oncology Clinical Trials: Quality Assurance Strategies for Technical Validation," *Tomography*, vol. 9, no. 5, pp. 1876–1902, 2023, Type: Journal Article, ISSN: 2379-139X (Electronic) 2379-1381 (Print) 2379-1381 (Linking). DOI: 10.3390/tomography9050149.

[24] T. Tirkes, M. A. Hollar, M. Tann, *et al.*, "Response criteria in oncologic imaging: Review of traditional and new criteria," *Radiographics*, vol. 33, no. 5, pp. 1323–41, 2013, Type: Journal Article, ISSN: 1527-1323 (Electronic) 0271-5333 (Linking). DOI: 10.1148/rg.335125214.

[25] A. R. Padhani, F. E. Lecouvet, N. Tunariu, *et al.*, "METastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-body Magnetic Resonance Imaging-based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer," *Eur Urol*, vol. 71, no. 1, pp. 81–92, 2017, Type: Journal Article, ISSN: 1873-7560 (Electronic) 0302-2838 (Print) 0302-2838 (Linking). DOI: 10.1016/j.eururo.2016.05.033.

[26] N. Sanai, M. Y. Polley, M. W. McDermott, *et al.*, "An extent of resection threshold for newly diagnosed glioblastomas," *J Neurosurg*, vol. 115, no. 1, pp. 3–8, 2011, Type: Journal Article, ISSN: 1933-0693 (Electronic) 0022-3085 (Linking). DOI: 10.3171/2011.2.Jns10998.

[27] C. Majos, M. Cos, S. Castaner, *et al.*, "Early post-operative magnetic resonance imaging in glioblastoma: Correlation among radiological findings and overall survival in 60 patients," *Eur Radiol*, vol. 26, no. 4, pp. 1048–55, 2016, Type: Journal Article, ISSN: 1432-1084 (Electronic) 0938-7994 (Linking). DOI: 10.1007/s00330-015-3914-x.

[28] N. A. Shonka and M. R. Aizenberg, "Extent of Resection in Glioblastoma," *J Oncol Pract*, vol. 13, no. 10, pp. 641–642, 2017, Type: Journal Article, ISSN: 1935-469X (Electronic) 1554-7477 (Linking). DOI: 10.1200/jop.2017.027599.

[29] A. W. Awad, M. Karsy, N. Sanai, *et al.*, "Impact of removed tumor volume and location on patient outcome in glioblastoma," *J Neurooncol*, vol. 135, no. 1, pp. 161–171, 2017, Type: Journal Article, ISSN: 1573-7373 (Electronic) 0167-594X (Linking). DOI: 10.1007/s11060-017-2562-1.

[30] B. M. Ellingson, L. E. Abrey, S. J. Nelson, *et al.*, "Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma," *Neuro-Oncology*, vol. 20, no. 9, pp. 1240–1250, 2018, Type: Journal Article, ISSN: 1522-8517 1523-5866. DOI: 10.1093/neuonc/noy053.

[31] D. Leung, X. Han, T. Mikkelsen, *et al.*, "Role of MRI in primary brain tumor evaluation," *J Natl Compr Canc Netw*, vol. 12, no. 11, pp. 1561–8, 2014, Type: Journal Article, ISSN: 1540-1413 (Electronic) 1540-1405 (Linking). DOI: 10.6004/jnccn.2014.0156.

[32] S. Rathore, H. Akbari, M. Rozycki, *et al.*, "Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1," *Sci Rep*, vol. 8, no. 1, p. 5087, 2018, Type:

Journal Article, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-018-22739-2.

[33] J. P. O'Connor, E. O. Aboagye, J. E. Adams, *et al.*, "Imaging biomarker roadmap for cancer studies," *Nat Rev Clin Oncol*, vol. 14, no. 3, pp. 169–186, 2017, Type: Journal Article, ISSN: 1759-4782 (Electronic) 1759-4774 (Print) 1759-4774 (Linking). DOI: 10.1038/nrclinonc.2016.162.

[34] P. Lambin, E. Rios-Velazquez, R. Leijenaar, *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur J Cancer*, vol. 48, no. 4, pp. 441–6, 2012, Type: Journal Article, ISSN: 1879-0852 (Electronic) 0959-8049 (Print) 0959-8049 (Linking). DOI: 10.1016/j.ejca.2011.11.036.

[35] C. Macarro, K. Bernatowicz, A. Garcia-Ruiz, *et al.*, "Enhancing Tumor Microstructural Quantification With Machine Learning and Diffusion-Relaxation MRI," *Journal of Magnetic Resonance Imaging*, jmri.29484, Jun. 2024, ISSN: 1053-1807, 1522-2586. DOI: 10.1002/jmri.29484.

[36] R. Berenguer, M. D. R. Pastor-Juan, J. Canales-Vazquez, *et al.*, "Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters," *Radiology*, vol. 288, no. 2, pp. 407–415, 2018, Type: Journal Article, ISSN: 1527-1315 (Electronic) 0033-8419 (Linking). DOI: 10.1148/radiol.2018172361.

[37] A. Zwanenburg, M. Vallieres, M. A. Abdalah, *et al.*, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020, Type: Journal Article, ISSN: 1527-1315 (Electronic) 0033-8419 (Print) 0033-8419 (Linking). DOI: 10.1148/radiol.2020191145.

[38] K. Bernatowicz, F. Grussu, M. Ligero, *et al.*, "Robust imaging habitat computation using voxel-wise radiomics features," *Sci Rep*, vol. 11, no. 1, p. 20 133, 2021, Type: Journal Article, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-021-99701-2.

[39] O. Prior, C. Macarro, V. Navarro, *et al.*, "Identification of Precise 3D CT Radiomics for Habitat Computation by Machine Learning in Cancer," *Radiol Artif Intell*, vol. 6, no. 2, e230118, 2024, Type: Journal Article, ISSN: 2638-6100 (Electronic) 2638-6100 (Linking). DOI: 10.1148/ryai.230118.

[40] A. Lecler, L. Duron, D. Balvay, *et al.*, "Combining Multiple Magnetic Resonance Imaging Sequences Provides Independent Reproducible Radiomics Features," *Sci Rep*, vol. 9, no. 1, p. 2068, 2019, Type: Journal Article, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-018-37984-8.

[41] A. Traverso, L. Wee, A. Dekker, *et al.*, "Repeatability and Reproducibility of Radiomic Features: A Systematic Review," *Int J Radiat Oncol Biol Phys*, vol. 102, no. 4, pp. 1143–1158, 2018, Type: Journal Article, ISSN: 1879-355X (Electronic) 0360-3016 (Linking). DOI: 10.1016/j.ijrobp.2018.05.053.

[42] M. Ligero, O. Jordi-Ollero, K. Bernatowicz, *et al.*, "Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis," *Eur Radiol*, vol. 31, no. 3, pp. 1460–1470, 2021, Type: Journal Article, ISSN: 1432-1084 (Electronic) 0938-7994 (Linking). DOI: 10.1007/s00330-020-07174-0.

[43] Y. Liu, X. Xu, L. Yin, *et al.*, "Relationship between Glioblastoma Heterogeneity and Survival Time: An MR Imaging Texture Analysis," *AJNR Am J Neuroradiol*, vol. 38, no. 9, pp. 1695–1701, 2017, Type: Journal Article, ISSN: 1936-959X (Electronic) 0195-6108 (Linking). DOI: 10.3174/ajnr.A5279.

[44] M. Ligero, A. Garcia-Ruiz, C. Viaplana, *et al.*, "A CT-based Radiomics Signature Is Associated with Response to Immune Checkpoint Inhibitors in Advanced Solid Tumors," *Radiology*, vol. 299, no. 1, pp. 109–119, Apr. 2021, ISSN: 0033-8419, 1527-1315. DOI: 10.1148/radiol.2021200928.

[45] M. Ligero, J. Hernando, E. Delgado, *et al.*, "Radiomics and outcome prediction to antiangiogenic treatment in advanced gastroenteropancreatic neuroendocrine tumours: Findings from the phase II TALENT trial," *BJC Reports*, vol. 1, no. 1, p. 9, Aug. 2023, ISSN: 2731-9377. DOI: 10.1038/s44276-023-00010-0.

[46] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol Rev*, vol. 65, no. 6, pp. 386–408, 1958, Type: Journal Article, ISSN: 0033-295X (Print) 0033-295X (Linking). DOI: 10.1037/h0042519.

[47] H. Lee, R. Grosse, R. Ranganath, *et al.*, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal Quebec Canada: ACM, Jun. 2009, pp. 609–616, ISBN: 978-1-60558-516-1. DOI: 10.1145/1553374.1553453.

[48] A. Esteva, B. Kuprel, R. A. Novoa, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, Type: Journal Article, ISSN: 1476-4687 (Electronic) 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/nature21056.

[49] J. Wasserthal, H. C. Breit, M. T. Meyer, *et al.*, "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images Multisite Concordance of DSC-MRI Analysis for Brain Tumors: Results of a National Cancer Institute Quantitative Imaging Network Collaborative Project," *Radiol Artif Intell*, vol. 5, no. 5, e230024, 2023, Type: Journal Article, ISSN: 2638-6100 (Electronic) 2638-6100 (Linking) 1936-959X (Electronic) 0195-6108 (Linking). DOI: 10.1148/ryai.23002410.3174/ajnr.A5675.

[50] X. Wang, S. Yang, J. Zhang, *et al.*, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Med Image Anal*, vol. 81, p. 102 559, 2022, Type: Journal Article, ISSN: 1361-8423 (Electronic) 1361-8415 (Linking). DOI: 10.1016/j.media.2022.102559.

[51] H. Xu, N. Usuyama, J. Bagga, *et al.*, "A whole-slide foundation model for digital pathology from real-world data," *Nature*, vol. 630, no. 8015, pp. 181–188, 2024, Type: Journal Article, ISSN: 1476-4687 (Electronic) 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/s41586-024-07441-w.

[52] R. J. Chen, T. Ding, M. Y. Lu, *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nat Med*, vol. 30, no. 3, pp. 850–862, 2024, Type: Journal Article, ISSN: 1546-170X (Electronic) 1078-8956 (Linking). DOI: 10.1038/s41591-024-02857-3.

[53] Z. Liu, H. Mao, C.-Y. Wu, *et al.*, "A ConvNet for the 2020s," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 11 966–11 976, ISBN: 978-1-66546-946-3. DOI: 10.1109/CVPR52688.2022.01167.

[54] H. Wang, Z. Wang, M. Du, *et al.*, "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Type: Conference Proceedings, IEEE Computer Society, 2020, pp. 111–119. DOI: 10.1109/cvprw50498.2020.00020.

[55] R. Achtibat, M. Dreyer, I. Eisenbraun, *et al.*, "From attribution maps to human-understandable explanations through Concept Relevance Propagation," *Nature Ma-*

chine Intelligence, vol. 5, no. 9, pp. 1006–1019, 2023, Type: Journal Article, ISSN: 2522-5839. DOI: 10.1038/s42256-023-00711-8.

[56] D. M. Koh, N. Papanikolaou, U. Bick, *et al.*, "Artificial intelligence and machine learning in cancer imaging," *Commun Med (Lond)*, vol. 2, p. 133, 2022, Type: Journal Article, ISSN: 2730-664X (Electronic) 2730-664X (Linking). DOI: 10.1038/s43856-022-00199-0.

[57] E. Longato, M. Vettoretti, and B. Di Camillo, "A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models," *J Biomed Inform*, vol. 108, p. 103 496, 2020, Type: Journal Article, ISSN: 1532-0480 (Electronic) 1532-0464 (Linking). DOI: 10.1016/j.jbi.2020.103496.

[58] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, *et al.*, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 1st ed. Wiley, Apr. 2014, ISBN: 978-0-471-72085-0 978-1-118-63395-3. DOI: 10.1002/9781118633953.

[59] W. T. Dixon, "Simple proton spectroscopic imaging," *Radiology*, vol. 153, no. 1, pp. 189–94, 1984, Type: Journal Article, ISSN: 0033-8419 (Print) 0033-8419 (Linking). DOI: 10.1148/radiology.153.1.6089263.

[60] I. S. Idilman, A. E. Yildiz, A. D. Karaosmanoglu, *et al.*, "Proton density fat fraction: Magnetic resonance imaging applications beyond the liver," *Diagn Interv Radiol*, vol. 28, no. 1, pp. 83–91, 2022, Type: Journal Article, ISSN: 1305-3612 (Electronic) 1305-3825 (Linking). DOI: 10.5152/dir.2021.21845.

[61] N. Basty, M. Thanaj, M. Cule, *et al.*, "Artifact-free fat-water separation in Dixon MRI using deep learning," *J Big Data*, vol. 10, no. 1, p. 4, 2023, Type: Journal Article, ISSN: 2196-1115 (Print) 2196-1115 (Electronic) 2196-1115 (Linking). DOI: 10.1186/s40537-022-00677-1.

[62] E. S. Paulson and K. M. Schmainda, "Comparison of dynamic susceptibility-weighted contrast-enhanced MR methods: Recommendations for measuring relative cerebral blood volume in brain tumors," *Radiology*, vol. 249, no. 2, pp. 601–13, 2008, Type: Journal Article, ISSN: 1527-1315 (Electronic) 0033-8419 (Linking). DOI: 10.1148/radiol.2492071659.

[63] R. Mangla, D. T. Ginat, S. Kamalian, *et al.*, "Correlation between progression free survival and dynamic susceptibility contrast MRI perfusion in WHO grade III glioma subtypes," *J Neurooncol*, vol. 116, no. 2, pp. 325–31, 2014, Type: Journal Article, ISSN: 1573-7373 (Electronic) 0167-594X (Linking). DOI: 10.1007/s11060-013-1298-9.

[64] L. C. Bell, N. Semmineh, H. An, *et al.*, "Evaluating the Use of rCBV as a Tumor Grade and Treatment Response Classifier Across NCI Quantitative Imaging Network Sites: Part II of the DSC-MRI Digital Reference Object (DRO) Challenge," *Tomography*, vol. 6, no. 2, pp. 203–208, 2020, Type: Journal Article, ISSN: 2379-139X (Electronic) 2379-1381 (Linking). DOI: 10.18383/j.tom.2020.00012.

[65] M. Essig, F. Giesel, M. Le-Huu, *et al.*, "Perfusion MRI in CNS disease: Current concepts," *Neuroradiology*, vol. 46 Suppl 2, s201–7, 2004, Type: Journal Article, ISSN: 1432-1920 (Electronic) 0028-3940 (Linking). DOI: 10.1007/s00234-004-1331-y.

[66] D. Sinnaeve, "The Stejskal–Tanner equation generalized for any gradient shape—an overview of most pulse sequences measuring free diffusion," *Concepts in Magnetic Resonance Part A*, vol. 40A, no. 2, pp. 39–65, 2012, Type: Journal Article, ISSN: 1546-6086 1552-5023. DOI: 10.1002/cmr.a.21223.

[67] D. Le Bihan, E. Breton, D. Lallemand, *et al.*, "Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging," *Radiology*, vol. 168, no. 2, pp. 497–505, 1988, Type: Journal Article, ISSN: 0033-8419 (Print) 0033-8419 (Linking). DOI: 10.1148/radiology.168.2.3393671.

[68] E. Panagiotaki, S. Walker-Samuel, B. Siow, *et al.*, "Noninvasive quantification of solid tumor microstructure using VERDICT MRI," *Cancer Res*, vol. 74, no. 7, pp. 1902–12, 2014, Type: Journal Article, ISSN: 1538-7445 (Electronic) 0008-5472 (Linking). DOI: 10.1158/0008-5472.Can-13-2511.

[69] C. Bailey, D. J. Collins, N. Tunariu, *et al.*, "Microstructure Characterization of Bone Metastases from Prostate Cancer with Diffusion MRI: Preliminary Findings," *Front Oncol*, vol. 8, p. 26, 2018, Type: Journal Article, ISSN: 2234-943X (Print) 2234-943X (Linking). DOI: 10.3389/fonc.2018.00026.

[70] R. L. Brunsing, N. M. Schenker-Ahmed, N. S. White, *et al.*, "Restriction spectrum imaging: An evolving imaging biomarker in prostate MRI," *J Magn Reson Imaging*, vol. 45, no. 2, pp. 323–336, 2017, Type: Journal Article, ISSN: 1522-2586 (Electronic) 1053-1807 (Print) 1053-1807 (Linking). DOI: 10.1002/jmri.25419.

[71] M. Palombo, V. Valindria, S. Singh, *et al.*, "Joint estimation of relaxation and diffusion tissue parameters for prostate cancer with relaxation-VERDICT MRI," *Sci Rep*, vol. 13, no. 1, p. 2999, 2023, Type: Journal Article, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-023-30182-1.

[72] N. S. White, C. McDonald, N. Farid, *et al.*, "Diffusion-weighted imaging in cancer: Physical foundations and applications of restriction spectrum imaging," *Cancer Res*, vol. 74, no. 17, pp. 4638–52, 2014, Type: Journal Article, ISSN: 1538-7445 (Electronic) 0008-5472 (Print) 0008-5472 (Linking). DOI: 10.1158/0008-5472.CAN-13-3534.

[73] P. Svolos, E. Kousi, E. Kapsalaki, *et al.*, "The role of diffusion and perfusion weighted imaging in the differential diagnosis of cerebral tumors: A review and future perspectives," *Cancer Imaging*, vol. 14, no. 1, p. 20, 2014, Type: Journal Article, ISSN: 1470-7330 (Electronic) 1740-5025 (Print) 1470-7330 (Linking). DOI: 10.1186/1470-7330-14-20.

[74] R. Bourne and E. Panagiotaki, "Limitations and Prospects for Diffusion-Weighted MRI of the Prostate," *Diagnostics (Basel)*, vol. 6, no. 2, 2016, Type: Journal Article, ISSN: 2075-4418 (Print) 2075-4418 (Electronic) 2075-4418 (Linking). DOI: 10.3390/diagnostics6020021.

[75] F. Grussu, K. Bernatowicz, I. Casanova-Salas, *et al.*, "Diffusion MRI signal cumulants and hepatocyte microstructure at fixed diffusion time: Insights from simulations, 9.4T imaging, and histology Multiparametric Magnetic Resonance Imaging of Prostate Cancer Bone Disease: Correlation With Bone Biopsy Histological and Molecular Features," *Magn Reson Med*, vol. 88, no. 1, pp. 365–379, 2022, Type: Journal Article, ISSN: 1522-2594 (Electronic) 0740-3194 (Linking) 1536-0210 (Electronic) 0020-9996 (Linking). DOI: 10.1002/mrm.2917410.1097/rli.0000000000000415.

[76] R. Perez-Lopez, D. Nava Rodrigues, I. Figueiredo, *et al.*, "Multiparametric Magnetic Resonance Imaging of Prostate Cancer Bone Disease: Correlation With Bone Biopsy Histological and Molecular Features," *Invest Radiol*, vol. 53, no. 2, pp. 96–102, 2018, Type: Journal Article, ISSN: 1536-0210 (Electronic) 0020-9996 (Linking). DOI: 10.1097/RLI.0000000000000415.

[77] H. Johnson, G. Harris, and K. Williams, "BRAINSFit: Mutual Information Registrations of Whole-Brain 3D Images, Using the Insight Toolkit," *The Insight Journal*, 2007, Type: Journal Article, ISSN: 2327-770X. DOI: 10.54294/hmb052.

[78] G. Balakrishnan, A. Zhao, M. R. Sabuncu, *et al.*, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," *IEEE Trans Med Imaging*, 2019, Type: Journal Article, ISSN: 1558-254X (Electronic) 0278-0062 (Linking). DOI: 10.1109/TMI.2019.2897538.

[79] K. M. Pohl, S. Bouix, M. Nakamura, *et al.*, "A hierarchical algorithm for MR brain image parcellation," *IEEE Trans Med Imaging*, vol. 26, no. 9, pp. 1201–12, 2007, Type: Journal Article, ISSN: 0278-0062 (Print) 0278-0062 (Linking). DOI: 10.1109/tmi.2007.901433.

[80] J. E. Iglesias, C. Y. Liu, P. M. Thompson, *et al.*, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Trans Med Imaging*, vol. 30, no. 9, pp. 1617–34, 2011, Type: Journal Article, ISSN: 1558-254X (Electronic) 0278-0062 (Linking). DOI: 10.1109/TMI.2011.2138152.

[81] R. Meier, U. Knecht, T. Loosli, *et al.*, "Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry," *Sci Rep*, vol. 6, p. 23 376, 2016, Type: Journal Article, ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/srep23376.

[82] E. T. McKinnon, J. A. Helpern, and J. H. Jensen, "Modeling white matter microstructure with fiber ball imaging," *Neuroimage*, vol. 176, pp. 11–21, 2018, Type: Journal Article, ISSN: 1095-9572 (Electronic) 1053-8119 (Linking). DOI: 10.1016/j.neuroimage.2018.04.025.

[83] H. M. De Feyter and R. A. De Graaf, "Deuterium metabolic imaging – Back to the future," *Journal of Magnetic Resonance*, vol. 326, p. 106 932, May 2021, ISSN: 10907807. DOI: 10.1016/j.jmr.2021.106932.

[84] H. Mehrabian, S. Myrehaug, H. Soliman, *et al.*, "Evaluation of Glioblastoma Response to Therapy With Chemical Exchange Saturation Transfer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 101, no. 3, pp. 713–723, Jul. 2018, ISSN: 03603016. DOI: 10.1016/j.ijrobp.2018.03.057.

[85] D. Paech, C. Dreher, S. Regnery, *et al.*, "Relaxation-compensated amide proton transfer (APT) MRI signal intensity is associated with survival and progression in high-grade glioma patients," *European Radiology*, vol. 29, no. 9, pp. 4957–4967, Sep. 2019, ISSN: 0938-7994, 1432-1084. DOI: 10.1007/s00330-019-06066-2.

[86] N. Porz, S. Bauer, A. Pica, *et al.*, "Multi-modal glioblastoma segmentation: Man versus machine," *PLoS One*, vol. 9, no. 5, e96873, 2014, Type: Journal Article, ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0096873.

[87] Z. J. Wang, M. A. Ohliger, P. E. Z. Larson, *et al.*, "Hyperpolarized ^{13}C MRI: State of the Art and Future Directions," *Radiology*, vol. 291, no. 2, pp. 273–284, May 2019, ISSN: 0033-8419, 1527-1315. DOI: 10.1148/radiol.2019182391.

[88] L. Wynants, B. Van Calster, G. S. Collins, *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, p. m1328, 2020, Type: Journal Article, ISSN: 1756-1833 (Electronic) 0959-8138 (Print) 0959-8138 (Linking). DOI: 10.1136/bmj.m1328.

[89] 3. Armato S. G., K. Drukker, and L. Hadjiiski, "AI in medical imaging grand challenges: Translation from competition to research benefit and patient care," *Br J Radiol*, vol. 96, no. 1150, p. 20 221 152, 2023, Type: Journal Article, ISSN: 1748-880X (Electronic) 0007-1285 (Print) 0007-1285 (Linking). DOI: 10.1259/bjr.20221152.

[90] K. Clark, B. Vendt, K. Smith, *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository," *J Digit Imaging*, vol. 26, no. 6, pp. 1045–57, 2013, Type: Journal Article, ISSN: 1618-727X (Electronic) 0897-1889 (Print) 0897-1889 (Linking). DOI: 10.1007/s10278-013-9622-7.

[91] H. I. Scher, M. J. Morris, W. M. Stadler, *et al.*, "Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3," *J Clin Oncol*, vol. 34, no. 12, pp. 1402–18, 2016, Type: Journal Article, ISSN: 1527-7755 (Electronic) 0732-183X (Print) 0732-183X (Linking). DOI: 10.1200/JCO.2015.64.2702.

[92] P. Rescigno, D. Lorente, D. Bianchini, *et al.*, "Prostate-specific Antigen Decline After 4 Weeks of Treatment with Abiraterone Acetate and Overall Survival in Patients with Metastatic Castration-resistant Prostate Cancer," *Eur Urol*, vol. 70, no. 5, pp. 724–731, 2016, Type: Journal Article, ISSN: 1873-7560 (Electronic) 0302-2838 (Linking). DOI: 10.1016/j.eururo.2016.02.055.

[93] B. L. Lau, K. Vijian, D. N. S. Liew, *et al.*, "Factors affecting diagnostic yield in stereotactic biopsy for brain lesions: A 5-year single-center series," *Neurosurg Rev*, vol. 45, no. 2, pp. 1473–1480, 2022, Type: Journal Article, ISSN: 1437-2320 (Electronic) 0344-5607 (Linking). DOI: 10.1007/s10143-021-01671-6.

[94] J. Ma, Y. He, F. Li, *et al.*, "Segment anything in medical images," *Nat Commun*, vol. 15, no. 1, p. 654, 2024, Type: Journal Article, ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-024-44824-z.

[95] F. Isensee, P. F. Jaeger, S. A. A. Kohl, *et al.*, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nat Methods*, vol. 18, no. 2, pp. 203–211, 2021, Type: Journal Article, ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/s41592-020-01008-z.

[96] M. Roberts, D. Driggs, M. Thorpe, *et al.*, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 199–217, 2021, Type: Journal Article, ISSN: 2522-5839. DOI: 10.1038/s42256-021-00307-0.

[97] S. Oron, T. Dekel, T. Xue, *et al.*, "Best-Buddies Similarity-Robust Template Matching Using Mutual Nearest Neighbors," *IEEE Trans Pattern Anal Mach Intell*, vol. 40, no. 8, pp. 1799–1813, 2018, Type: Journal Article, ISSN: 1939-3539 (Electronic) 0098-5589 (Linking). DOI: 10.1109/TPAMI.2017.2737424.

[98] K. Lenc and A. Vedaldi, "Understanding Image Representations by Measuring Their Equivariance and Equivalence," *Int J Comput Vis*, vol. 127, no. 5, pp. 456–476, 2019, Type: Journal Article, ISSN: 0920-5691 (Print) 0920-5691 (Linking). DOI: 10.1007/s11263-018-1098-y.

[99] J. Y. Koh, R. Salakhutdinov, and D. Fried, "Grounding Language Models to Images for Multimodal Inputs and Outputs," in *Proceedings of the 40th International Conference on Machine Learning*, A. Krause, E. Brunskill, K. Cho, *et al.*, Eds., ser. Proceedings of Machine Learning Research, vol. 202, PMLR, Jul. 2023, pp. 17 283–17 300.

[100] M. Huh, B. Cheung, T. Wang, *et al.*, *The Platonic Representation Hypothesis*, _eprint: 2405.07987, 2024. DOI: 10.48550/arXiv.2405.07987.

[101] M. Mittermaier, M. M. Raza, and J. C. Kvedar, "Bias in AI-based models for medical applications: Challenges and mitigation strategies," *NPJ Digit Med*, vol. 6, no. 1, p. 113, 2023, Type: Journal Article, ISSN: 2398-6352 (Electronic) 2398-6352 (Linking). DOI: 10.1038/s41746-023-00858-z.

[102] M. D. Abramoff, M. E. Tarver, N. Loyo-Berrios, *et al.*, "Considerations for addressing bias in artificial intelligence for health equity," *NPJ Digit Med*, vol. 6, no. 1, p. 170, 2023, Type: Journal Article, ISSN: 2398-6352 (Electronic) 2398-6352 (Linking). DOI: 10.1038/s41746-023-00913-9.

[103] S. M. Williamson and V. Prybutok, "Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare," *Applied Sciences*, vol. 14, no. 2, 2024, Type: Journal Article, ISSN: 2076-3417. DOI: 10.3390/app14020675.

[104] Q. Wang and D. Kurz, "Reconstructing Training Data from Diverse ML Models by Ensemble Inversion," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA: IEEE, Jan. 2022, pp. 3870–3878, ISBN: 978-1-66540-915-5. DOI: 10.1109/WACV51458.2022.00392.

[105] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[106] U. J. Muehlematter, P. Daniore, and K. N. Vokinger, "Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): A comparative analysis," *The Lancet Digital Health*, vol. 3, no. 3, e195–e203, Mar. 2021, ISSN: 25897500. DOI: 10.1016/S2589-7500(20)30292-2.