Advanced Magnetic Resonance Imaging techniques in soft tissue sarcoma studies – Modelling of quantitative MRI parameters

Author: Georgios C. Manikis

Supervisors:
Apostolos Karantanas
Thomas G. Maris
Kostas Marias

A thesis submitted for the degree of Doctor of Philosophy in collaboration with the:

Computational Biomedicine Laboratory, Institute of Computer Science (ICS), FORTH
Department of Radiology, School of Medicine, University of Crete

Monday, 5 April 2021

Νεώτερες τεχνικές απεικόνισης μαγνητικού συντονισμού στη μελέτη όγκων μαλακών μορίων - Μοντελοποίηση δεικτών q-MRI

Συγγραφέας: Γεώργιος Χ. Μανίκης
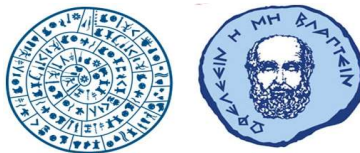
Επιβλέποντες:
Απόστολος Καραντάνας
Θωμάς Γ. Μαρής
Κωνσταντίνος Μαριάς

Μια διατριβή για την κατάκτηση του τίτλου διδακτορικών σπουδών σε συνεργασία με το:

Εργαστήριο Υπολογιστικής Βιο-Ιατρικής, Ινστιτούτο Πληροφορικής (ΙΠ), Ι.Τ.Ε.
Εργαστήριο Ιατρικής Απεικόνισης, ΠΑΓΝΗ

Δευτέρα, 5 Απριλίου 2021

Acknowledgments

This journey would not have been possible without the support of my family, professors and friends.

I would like to thank my supervisor Professor A. Karantanas for his guidance, continuous encouragement and kind advices throughout my PhD research studies. I would like to thank him for always providing the most conducive conditions to work despite his overloaded work schedule. I feel grateful to Professor Thomas Maris for his immense support as a supervisor and his valuable insight and guidance throughout all aspects of MRI analysis. I am also thankful to Professors I. Souglakos, A. Koutsopoulos S. Agelaki and Eelco de Bree for agreeing to be a part of the jury for the thesis defense. Thank you for your interest in improving our work with your insights. I am deeply grateful to Professor Kostas Marias for kindly supporting me not only with his advice but also by sharing his way of addressing tribulations. His clarity of vision on research, critical thinking ability and dexterity in seeing through ideas from conception to realization has immensely helped me in learning how to perform research.

I am grateful to have shared time with Dr E. Lagoudaki at her lab and watching the amount of effort that she allocates for each task. She has been very patient in explaining the complexities of identification of the histopathologic specific features in simpler terms and in helping me understand basic aspects and for provoking my interest in her field of expertise. My sincere thanks also go to Dr N. Papanikolaou for having played a very significant role to gaining the background knowledge and experience in quantitative MRI. I thank all the members of our team, K. Nikiforaki, L. Kontopodis , G. Ioannidis, I. Karatzanis, E. Trivizakis, N. Tsiknakis and K. Spanakis for working together and for sharing a very pleasant and fruitful collaboration and for creating an enviable working environment. They have all been very generous to me.

Acknowledgement would be incomplete without thanking my parents and family. To my parents Christos and Maria and my brother Stamatis, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams. I do not know how to thank you enough for providing me with the opportunity to be where I am today. This thesis is dedicated to my family Fotini and Marios. I love you more than anything.

Νεότερες τεχνικές απεικόνισης μαγνητικού συντονισμού στη μελέτη όγκων μαλακών μορίων - Μοντελοποίηση δεικτών q-MRI.

Σύνοψη

Οι όγκοι των μαλακών μορίων περιλαμβάνουν ένα ευρύ φάσμα μεσεγχυματικών νεοπλασμάτων, συμπεριλαμβανομένων περισσότερων από εκατό διαφορετικών υποτύπων. Η προεγχειρητική διάγνωση που καθορίζεται συνήθως από τη βιοψία με κόπτουσα βελόνα και των επακόλουθων ιστοπαθολογικών εξετάσεων είναι απαραίτητη για την αξιολόγηση του ιστολογικού υποτύπου και της βιολογικής συμπεριφοράς της βλάβης (καλοήθης ή κακοήθης, βαθμός κακοήθειας) προκειμένου να καθοριστεί η βέλτιστη θεραπεία. Στην περίπτωση των σαρκωμάτων, δηλαδή εκείνων με κακοήθη συμπεριφορά, η ευρεία εκτομή του όγκου μαζί με ένα χείλος γειτονικού υγιούς ιστού είναι η χειρουργική θεραπεία επιλογής για τη μείωση του κινδύνου τοπικής υποτροπής. Για την σταδιοποίηση του όγκου, χρησιμοποιείται ευρέως το σύστημα της Γαλλικής Ομοσπονδίας Κέντρων Καρκίνου Sarcoma Group (FNCLCC). Ωστόσο, το θεραπευτικό αποτέλεσμα μπορεί να επηρεαστεί από τον τρόπο λήψης της βιοψίας καθώς αυτή σχετίζεται με επιπλοκές που μπορεί να οδηγήσουν σε νοσηρότητα, εσφαλμένη διάγνωση και στο μη βέλτιστο σχήμα της θεραπείας. Επομένως, η αξιόπιστη προεγχειρητική διάγνωση με μια μη επεμβατική μέθοδο όπως η ιατρική απεικόνιση καθίσταται αναγκαία.

Η απεικόνιση μαγνητικού συντονισμού (ΑΜΣ) αποτελεί την βέλτιστη επιλογή για την προεγχειρητική απεικόνιση των όγκων μαλακών μορίων καθώς παρέχει εξαιρετική ευκρίνεια σήματος και μπορεί να κατευθύνει τη λήψη βιοψίας σε κρίσιμες υποπεριοχές του όγκου. Συγκεκριμένα, η απεικόνιση σταθμισμένης διάχυσης (ΑΣΔ) μπορεί να αναδείξει μη επεμβατικά δομικές και λειτουργικές ιδιότητες του ιστού όπως η κυτταροβρίθεια, αγγειοβρίθεια και η πολυπλοκότητα του ιστού. Ο φαινόμενος συντελεστής διάχυσης (ΦΣΔ) ήταν η πρώτη παράμετρος ΑΣΔ για τον ποσοτικό προσδιορισμό της κυτταροβρίθειας του ιστού. Στη συνέχεια, προτάθηκε το μοντέλο Intravoxel Incoherent Motion (IVIM) το οποίο εισήγαγε μια δι-εκθετική αναπαράσταση της εξασθένησης του σήματος, καθώς και τα μη-διαμερισματικά μοντέλα stretched-exponential και diffusion kurtosis για τον ποσοτικό προσδιορισμό της μικροδομικής ετερογένειας του όγκου και της πολυπλοκότητας των ιστών. Παράλληλα με την ποσοτική ανάλυση της ΑΜΣ, η ραδιομική (radiomics) δημιούργησε τα τελευταία χρόνια μια ώθηση στην υπολογιστική ιατρική απεικόνιση ενδυναμώνοντας σημαντικά την προγνωστική μοντελοποίηση και ποσοτική ανάλυση των απεικονιστικών δεδομένων. Σημαντικά στοιχεία της αποτελούν η ανάλυση υφής της εικόνας και η εξαγωγή πολυδιάστατων χαρακτηριστικών από πολύπλοκα μοτίβα των διαγνωστικών εικόνων που σπάνια μπορεί να δει το ανθρώπινο μάτι καθώς και η δημιουργία προγνωστικών μοντέλων από τα χαρακτηριστικά αυτά για την ακριβή διάγνωση, σχεδιασμό της θεραπείας και παρακολούθησης των ασθενών.

Σύμφωνα με την διεθνή βιβλιογραφία, ένας σημαντικός αριθμός μελετών έχει διερευνήσει μη-επεμβατικά τον χαρακτηρισμό του μικροπεριβάλλοντος των σαρκωμάτων, την διαφοροποίηση της ταξινόμησης των όγκων και την αξιολόγηση της ανταπόκρισης της θεραπείας με μεμονωμένα μοντέλα ΑΣΔ, υποθέτοντας ότι ένα μοναδικό μοντέλο μπορεί να χαρακτηρίσει εξ' ολοκλήρου το μικροπεριβάλλον και την πολυπλοκότητα του όγκου. Ωστόσο, η υπόθεση αυτή αντικρούεται στην ετερογενή φύση του όγκου καθώς και σε σύγχρονες αναφορές οι οποίες φανερώνουν την αναποτελεσματικότητα μεμονωμένων μοντέλων ΑΣΔ στην καταγραφή και ποσοτικοποίηση των συνολικών λειτουργικών και ανατομικών ιδιοτήτων διαφόρων τύπου όγκων, καταλήγοντας σε εσφαλμένες τιμές παραμέτρων από τα ΑΣΔ μοντέλα και σε στατιστική επεξεργασία χωρίς πρότερο έλεγχο της επιλογής κατάλληλου μοντέλου για κάθε υποπεριοχή της βλάβης. Από την άλλη πλευρά, αν και η ραδιομική ανάλυση έχει μελετηθεί εκτενώς σε διάφορους τύπους όγκων, ελάχιστες μελέτες έχουν αναδείξει το ρόλο της ραδιομικής στην ανάλυση των σαρκωμάτων. Επιπρόσθετα, υπάρχουν αμφιβολίες αναφορικά με τον τρόπο επιλογής των σημαντικών ραδιομικών χαρακτηριστικών (ραδιομική υπογραφή), ιδίως όσον αφορά την συνοχή και επαναληψιμότητα της απόδοσής τους καθώς και στον καθορισμό ενός ισχυρού και διαφανούς πλαισίου εξέτασης και επικύρωσης των ραδιομικών αποτελεσμάτων.

Με βάση τα προαναφερθέντα, η παρούσα διατριβή εστίασε στην ποσοτικοποίηση των δεδομένων ΑΣΔ και την ραδιομική ανάλυση για τον χαρακτηρισμό των σαρκωμάτων και τον αυτοματοποιημένο διαχωρισμό της σταδιοποίησής τους με την εργασία συνεδρίου Ι να παρέχει ένα ολοκληρωμένο πλαίσιο ανάλυσης για την ποσοτικοποίηση ΑΣΔ από πολλαπλά μοντέλα. Η ανάπτυξη και αξιολόγηση των μοντέλων ΑΣΔ πραγματοποιήθηκε σύμφωνα με το μαθηματικό πλαίσιο όπως αυτό περιγράφεται στο Κεφάλαιο Ι και της στατιστικής ανάλυσης από το Κεφάλαιο ΙΙ. Τα άρθρα ΙΙΙ και IV εισήγαγαν ένα πλαίσιο στατιστικής ανάλυσης για την αξιολόγηση των μοντέλων ΑΣΔ και στην δημιουργία ψευδοχρωματικών χαρτών ταξινόμησης που αντικατοπτρίζουν την επιλογή του βέλτιστου μοντέλου ΑΣΔ για κάθε εικονοστοιχείο. Οι παραπάνω δημοσιεύσεις οδήγησαν στην ανάπτυξη υβριδικών παραμέτρων ΑΣΔ από πολλαπλά μοντέλα για τον ποσοτικό χαρακτηρισμό της βλάβης σε διάφορους τύπους σαρκωμάτων. Τα αποτελέσματα (άρθρο Ι) επικυρώθηκαν με ιστοπαθολογική εξέταση των χειρουργικών δειγμάτων, αποδίδοντας νέες παραμέτρους υψηλής διακριτικής ισχύος για την διαφοροποίηση της σταδιοποίησης του όγκου. Παράλληλα, εξετάστηκε η εφαρμογή της ραδιομικής ανάλυσης σε δεδομένα ΑΜΣ Τ2 υψηλής ανάλυσης για την ανάπτυξη μοντέλων μηχανικής μάθησης και την αυτοματοποιημένη διαφοροποίηση της σταδιοποίησης των σαρκωμάτων. Τα αποτελέσματα παρουσιάζονται στο άρθρο ΙΙ, μετά από διεξοδική έρευνα που προηγήθηκε και δημοσιεύθηκε στην εργασία συνεδρίου ΙΙ για την αξιολόγηση της γενικευμένης απόδοσης και επαναληψιμότητας των μοντέλων μηχανικής μάθησης, καθώς και της επιλογής των σημαντικών ραδιομικών χαρακτηριστικών που οδηγούν στην βέλτιστη πρόβλεψη της απόκρισης στη θεραπεία του καρκίνου. Συμπερασματικά, οι στόχοι της διατριβής συνοψίζονται στην

ποσοτικοποίηση των εικόνων ΑΣΔ και της ραδιομικής ανάλυσης για τον μη επεμβατικό χαρακτηρισμό των μορφολογικών και λειτουργικών χαρακτηρισμών των σαρκωμάτων και την αυτοματοποιημένη διαφοροποίηση του βαθμού σταδιοποίησης της βλάβης.

Advanced Magnetic Resonance Imaging techniques in soft tissue sarcoma studies – Modelling of quantitative MRI parameters.

Abstract

Soft tissue tumors comprise a broad spectrum of mesenchymal neoplasms, including over a hundred different subtypes. Preoperative diagnosis routinely established by core needle biopsy and subsequent histopathologic examination is essential for assessment of histological subtype and biological behavior (benign or malignant, malignancy grade) in order to determine the optimal treatment. In the case of soft tissue sarcomas (STSs), i.e. those with malignant behavior, wide excision of the tumor together with a rim of adjacent healthy tissue is the surgical treatment of choice to reduce the risk of local recurrence. To assess STSs aggressiveness, the French Federation of Cancer Centers Sarcoma Group (FNCLCC) grading system is widely used. However, preoperative radiotherapy or chemotherapy may be indicated since the performance of preoperative core needle or open biopsy is an invasive procedure associated with complications that may lead to morbidity, misdiagnosis and alteration to less optimal treatment. Therefore, reliable preoperative diagnosis by a non-invasive method like imaging would be of immense value.

Preoperative imaging of STSs is optimally performed by magnetic resonance imaging (MRI), as it provides supreme soft tissue contrast and may direct core or open biopsies to be taken at those most representative sites in heterogeneous tumors. Specifically, diffusion weighted imaging (DWI) has the potential to reveal insights into the structural and functional tissue properties such as cellularity, neovascularity and tissue integrity. To capture, non-invasively, these properties, apparent diffusion coefficient (ADC) was the first diffusion-related imaging parameter to quantify tissue cellularity. Next, the intravoxel incoherent model (IVIM) model was proposed, introducing a bi-exponential representation of the signal attenuation. In addition, the non-compartmentalized models including the stretched-exponential and the diffusion kurtosis model were developed to quantify tumor microstructural heterogeneity and tissue complexity. Apart from the quantitative MRI analysis, radiomics created an unprecedented momentum in computational medical imaging over the last years by significantly advancing and empowering correlational and predictive quantitative studies in numerous clinical applications. An important element of this exciting field of research is multiscale texture analysis yielding the extraction of high-throughput quantitative features from complex patterns of the diagnostic images that can rarely be seen by the human eye, subsequently used as highly informative and non-invasive imaging features for precise diagnosis, therapy planning and disease monitoring.

On one hand, a significant number of DWI studies, developed from single DWI models, has been conducted to characterize STSs microenvironment, differentiate soft tissue tumor grading and assess treatment response, assuming that single models can solely characterize the overall

tissue microenvironment. However, this assumption is in contradiction to the heterogeneous nature of the tumor where recent reports claimed that single models fail to appropriately capture regional functional and anatomical tumor properties, concluding to incorrect diffusion parameter values and statistics with no prior examination of models' applicability. On the other hand, although radiomics has been extensively studied in many anatomical areas, to the best of our knowledge, few studies examined the role of radiomics in STSs grading. Additionally, several concerns exist regarding the plethora of radiomics features used in the literature especially regarding their performance consistency across studies and the lack of a robust and transparent framework for the validation of the radiomic results.

Motivated by the aforementioned observations, conference paper I provides a comprehensive analysis framework for DWI quantification from multiple models where most of its functionalities were further used in the pre-processing part of the radiomic analysis pipeline. DWI model development and validation was performed according to the mathematical models and the statistical analysis framework reported in Book Chapters I and II. Paper III and IV introduce a statistical analysis framework in which suitability of several DWI models was examined across all tumor pixels and a classification map was generated reflecting DWI model preference on a pixel-by-pixel-basis. These publications have set the basis to develop hybrid diffusion parameters from multiple models to differentiate low from high STSs grading. The results, published in paper I, were validated by histopathological examination of the surgical specimens, yielding to novel parameters of high discriminatory power. A secondary goal was considered in this thesis in order to examine the application of radiomics and the use of high-resolution T2-MRI in the differentiation of the STSs staging. This is outlined in paper II, following a thorough investigation published in conference Paper II to assess the generalization performance and the intra-observer agreement of radiomic models as well as the relative importance of radiomics features in predicting cancer therapy response. The goals of this thesis were set towards the use of DWI quantification and radiomics into the non-invasive STSs characterization and grading differentiation.

The thesis is built on the following papers:

Original Papers

I. Manikis GC, Nikiforaki K, Lagoudaki E, de Bree E, Maris TG, Marias K, Karantanas A. Differentiating low from high grade soft tissue sarcomas using post-processed quantitative imaging biomarkers derived from multiple DWI models, European Journal of Radiology, Volume 138, 109660.

II. Manikis GC, Nikiforaki K, Marias K, Lagoudaki E, de Bree E, Maris TG, Karantanas A. 2019. T2-based MRI radiomic features for discriminating tumor grading in soft tissues sarcomas. Hell J Radiol; 4(3): 11-17.

III. Manikis G, Marias K, Lambregts DM, Nikiforaki K, Heeswijk MV, Bakers F CH, Beets-Tan RG, Papanikolaou N. 2017. Diffusion weighted imaging in patients with rectal cancer: Comparison between Gaussian and non-Gaussian models. PLoS One, 12(9), e0184197, Public Library of Science.

IV. Kartalis N, Manikis GC, Loizou L, Albiin N, Zöllner FG, Del Chiaro M, Marias K, Papanikolaou N. 2016. Diffusion-weighted MR imaging of pancreatic cancer: A comparison of mono-exponential, bi-exponential and non-Gaussian kurtosis models. European Journal of Radiology Open, Volume 3, Pages 79-85.
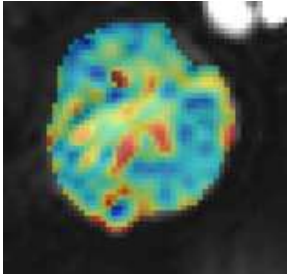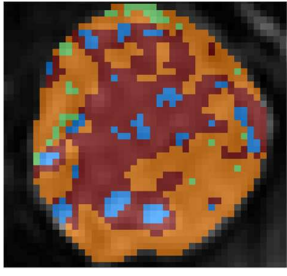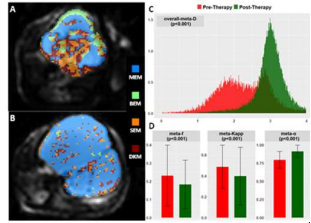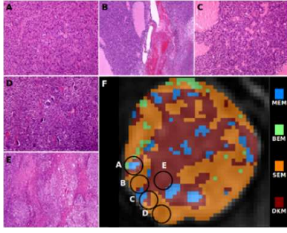
Book Chapters

I. Manikis G.C., Kontopodis E., Nikiforaki K., Marias K., Papanikolaou N. 2017. Imaging Biomarker Model-Based Analysis. In: Martí-Bonmatí L., Alberich-Bayarri A. (eds) Imaging Biomarkers. Springer, Cham.

II. Manikis GC, Papanikolaou N, Matos C. 2016. Validating the imaging biomarker: the proof of efficacy and effectiveness. Imaging Biomarkers: Development and Clinical Integration, Springer International Publishing, pp.115-122.
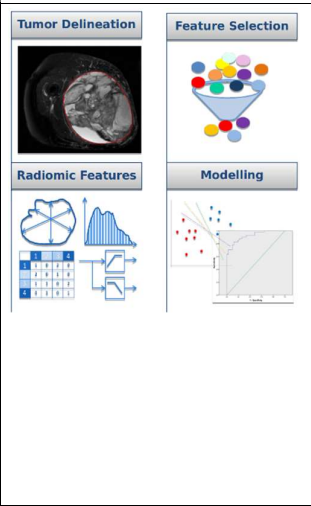
Conference Papers

I. Manikis GC, Nikiforaki K, Papanikolaou N, Marias K. 2016. Diffusion Modelling Tool (DMT) for the analysis of Diffusion Weighted Imaging (DWI) Magnetic Resonance Imaging (MRI) data. Computer Graphics International (CGI), the 33th Annual Conference, pp. 97-100.

II. Manikis GC, Venianaki M, Skepasianos I, Papadakis G, Maris T, Agelaki S, Karantanas K, Marias K. 2019. Scale-space DCE-MRI radiomics analysis based on Gabor filters for

predicting breast cancer therapy response. 19th IEEE International Conference on Bioinformatics and Bioengineering (BIBE).

Thesis at a glance

Question	Method	Result	Figure	Conclusion
Impact of DWI quantification in STSs	Four DWI models were implemented using in-house software	Multiple parametric maps were calculated and performance was assessed using several goodness-of-fit metrics		STSs cellularity, vascularization, complexity and heterogeneity can be quantified from DWI analysis
Most suitable DWI model to provide robustly spatial information of the STSs microarchitecture	Model selection using Akaike Weights was applied to four DWI models, yielding a classification map of model suitability across all tumor pixels	A composite diffusion model (CDM) classification map was generated depicting the most suitable model for each pixel within the tumor ROIs		Single DWI models fail to solely capture regional functional and anatomical STSs properties, concluding to incorrect DWI parameter values
Quantification of therapy induced changes using DWI analysis and the CDM map	Pre and post-therapy CDM maps were generated from patients treated with a drug combination of tumor necrosis factor antagonist (TNF-a) and melphalan with isolated limb perfusion (ILP)	A comparison between pre and post-therapy CDM maps revealed a significant decrease of the neo-vasculature and highly heterogeneous parts of the tumor		Quantitative MRI using DWI analysis from multiple models can potentially yield to non-invasive characterization of the treatment performance.
Histopathological validation of model suitability results in STSs	Histological images illustrating areas of central DWI tumor slice were registered to corresponding areas of the CDM maps	An excellent association was observed between histology and model suitability results		DWI multiple models and the CDM map can characterize non-invasively STSs microstructure

Assessment of DWI quantification and model selection into STSs grading differentiation	A statistical analysis was performed on multiple DWI parametric maps from 28 STSs patients, all calculated before and after model selection.	Only 3 histogram metrics, all derived after model selection were statistically significant in differentiating low from high-grade STSs achieving an AUC higher than 89%		Hybrid parameters acquired from DWI analysis and model selection are statistically significant and discriminatory factors in STSs grading
Investigate the feasibility of T2-based MRI radiomic features for discriminating tumor grading in STS	Radiomic-based machine learning models were developed using 1165 features from 22 STSs patients	The proposed predictive model achieved an AUC of 88.4% based on the 13 most significant radiomic features to discriminate low from high-grade STSs		Radiomics and machine learning are powerful discriminatory techniques for differentiating STSs grading

Acronyms

STSs: Soft Tissue Sarcomas

WHO: World Health Organization

FNCLCC: French Federation of Cancer Centers Sarcoma Group

NIH: National Institutes of Health

AJCC: American Joint Committee on Cancer

MRI: Magnetic Resonance Imaging

HR: High Resolution

DWI: Diffusion Weighted Imaging

DCE: Dynamic Contrast Enhanced

ADC: Apparent Diffusion Coefficient

IVIM: Intravoxel Incoherent Model

ROIs: Regions of Interest

ML: Machine Learning

DSS: Decision Support System

AUC: Area Under the Curve

OS: Overall Survival

RF: Random Forest

ASPS: Alveolar Soft Part Sarcoma

DDLS: Dedifferentiated Liposarcoma

UPS: Undifferentiated Pleomorphic Sarcoma

SCML: Spindle Cell Myxoid Liposarcoma

ILP: Isolated Limb Perfusion

EPI: Echo Planar Imaging

MEM: Mono-exponential Model

BEM: Bi-exponential Model

SEM: Stretched Exponential Model

DKM: Diffusion Kurtosis Model

Dslow: Slow Diffusion Coefficient

Dfast: Fast Component of Diffusion

f: Fraction of Fast Component

DDC: Distributed Diffusion Coefficient

α: Stretching Parameter

Kapp: Diffusional Kurtosis

DTI: Diffusion Tensor Imaging

Dapp: Apparent Diffusion

NLLS: Nonlinear Least Squares

R^2: R-square

SSE: Residual Sum of Squares

SST: Total Sum of Squares

RMSE: Root Mean Squared Error

AIC: Akaike Information Criteria

AICc: Corrected Akaike Information Criteria

AW: Akaike Weights

ΔAICc: AICc scores differences

GUI: Graphical User Interface

SNR: Signal to Noise Ratio

CDM: Composite Diffusion Model

ROC: Receiver Operator Characteristic

NPV: Negative Predictive

PPV: Positive Predictive Values

CI: Confidence Interval

GLCM: Grey-Level Co-occurrence Matrix

GLRLM: Gray Level Run Length Matrix

GLSZM: Gray Level Size Zone Matrix

mRMR: Minimum Redundancy Maximal Relevance

SMOTE: Synthetic Minority Oversampling

CV: Cross Validation

Table of Contents

Chapter 1 Introduction

1.1 Epidemiology

Soft tissue sarcomas (STSs) are neoplasms arising from the mesoderm derived tissues such as muscle, fat and connective tissue, thus constituting a broad and heterogeneous category of space occupying lesions. STSs are relatively rare and constitute less than 1.5% of all cancers with an annual incidence of about 6 per 100,000 persons. More than 50 different STSs subtypes have been defined by the World Health Organization (WHO), often associated with distinct radiological phenotype, different tumor biology and clinical outcome [1]. WHO divides tumor into benign, low grade (locally aggressive), intermediate grade (rarely metastasizing) and malignant. Histopathologic type, grade and tumor size and depth are determinant factors for soft tumor staging and therefore provide significant prognostic information. Core needle biopsy has been established for preoperative tumor grading in an attempt to classify tumors as high or low grade in order to contribute to the most appropriate therapeutic scheme [2]. Although histopathologic assessment of biopsy samples is the gold standard method for accurate tumor characterization and grading, it might often be subject to sampling errors underestimating thus tumor grade and misguiding therapeutic approach. Moreover, biopsy is an invasive procedure that can provoke several undesirable effects such as bleeding, pain, wound infection or breakdown and spillage of tumor cells. Noninvasive tumor characterization at the early stage of imaging is therefore of utmost importance to ensure the choice of the most appropriate therapeutic plan and minimize patient discomfort.

1.2 MRI acquisitions in STSs

Magnetic resonance imaging (MRI) has emerged as the imaging modality of choice for identification, staging, and monitoring of the response to therapy in patients with a suspicion for a malignant soft tissue mass as it provides supreme soft tissue contrast at multiple planes [3]. MRI reveals not only the anatomical extent of the tumor and the possible involvement of adjacent structures, but also provides functional information indicative of properties related to the degree of malignancy such as cellularity, vascularity or tissue integrity. Significant findings from a recent study revealed that an underestimation of STS grading using exclusively sample biopsy was evident in 40% of the enrolled patients, suggesting the inclusion of necrotic areas during biopsy, and the use of baseline MRI exams as a complementary prognostic approach to sample biopsy [4]. Conventional T1 and fat suppressed T2 sequences in coronal and axial planes are usually used for the localization and the complete coverage of the lesion. However, conventional MR imaging sequences have not equal power to biopsy in differentiating high from low soft tissue neoplasms as they exhibit a significant number of overlapping radiological features [5].

To this extend, since the radiological appearance of STSs may have a non-specific behavior in conventional MRI, high resolution (HR) MRI and advanced MRI protocols, now play a significant role towards this direction, providing a sensitive probe for the characterization of the disease at the early stage and serving as a valuable tool for all subsequent procedures, including biopsy. High resolution T2-weighted MRI acquired using a short-time protocol has been proved as a promising technique in STSs, yielding complete tumor coverage in high resolution and multiple possible planes, robustness to artifacts, low requirements in hardware and less patient discomfort. Diffusion Weighted Imaging (DWI) is an indispensable part of routine oncologic protocols in the clinical setting utilized to reveal insights into the structural and functional tissue properties. Dynamic Contrast Enhanced (DCE) imaging technique depict tissue perfusion and the microvascular environment. Since DCE-MRI is acquired after intravenous administration of specific contrast agents and the focus of this study was to explore non-invasively the use of MRI in the characterization of STSs and the differentiation of STSs grading, this technique was well outside the range of this study. However, a thorough description can be found in one of our publications [6].

1.3 Translating MRI into STSs features

1.3.1 Quantitative MRI using advanced DWI protocols

To investigate the role of advanced MRI towards the quantification of STSs characteristics, different mathematical models have been developed. In case of DWI, Apparent Diffusion Coefficient (ADC), derived when signal attenuation is expressed as a simple mono-exponential decay, was the first and most extensively used quantitative parameter in cancer imaging applications to describe cellular density [7]. However, in malignant areas where tissue neovascularity might be also apparent due to the angiogenesis effect, the mono-exponential model fails to capture the induced deviation in the signal decay caused by this effect and the Intravoxel Incoherent Model (IVIM) model was then proposed, introducing a bi-exponential representation of the signal attenuation occurred by a fast and a slow diffusion component [8]. Another aspect in cancer imaging is the complex and heterogeneous tissue microstructure of many tumors, resulting to restrictions in water molecule diffusion and deviation of its distribution from the Gaussian behavior. To quantify these specific characteristics, non-compartmentalized models including the stretched-exponential [9] and the diffusion kurtosis model [10] have been developed to reflect microstructural heterogeneity and tissue complexity. Research efforts were directed towards the quantification of the DWI in STSs studies. Hong et al suggested DWI along with conventional MRI analysis to assess tumor margin infiltration in STSs [11], while a recent study showed increased performance in differentiating benign from malignant soft tissue tumors when quantitative analysis using the mono-exponential model was

performed [12]. More advanced models including the IVIM and the diffusion kurtosis model were applied and their corresponding parameters exhibited significant results in discriminating STSs from vascular anomalies [13]. Radiotherapy response of STSs using the histopathology of the surgical specimen and summary statistics obtained from the ADC and the IVIM-related parameters yielded an increased median ADC after treatment and a correlation between ADC and tissue cellularity [14].

1.3.2 MRI-based radiomics and machine learning in STSs

Recent advances in medical imaging and computational modelling techniques have drastically altered the value of imaging data from qualitative representation of the disease to a significant tool allowing disease characterization, precise diagnosis, therapy planning and disease monitoring. On one hand, the emerging field of radiomics has enabled the conversion of routine radiological images into high-throughput quantitative data, describing non-intuitive properties of the tumor phenotype and microenvironment [15]. Within this context, handcrafted imaging features related to intensity, shape, volumetric, texture and wavelet characteristics are calculated from delineated regions of interest (ROIs), further comprising a representative high-throughput feature vector of the tumor volume. When radiomics is embedded with machine learning (ML) algorithms, automated decision support systems (DSS) are developed for the clinicians, providing advanced analytics and insights regarding future events of the disease. Corino et al. presented a radiomics analysis based on the ADC generated maps from diffusion-weighted MRI to distinguish intermediate from high-grade soft tissue sarcomas [16]. The performed analysis derived 64 imaging features and when applied to 19 patients achieved an area under the curve (AUC) of 0.85±0.16 and 0.87±0.34 using the validation and test set respectively. Another study explored the association between STS patients' overall survival (OS) and T1-weighted (T1w) contrast-enhanced MRI and found that the extracted radiomic features can be promising predictors of OS [17]. The proposed radiomics model was trained using 165 patients and performance was assessed using external validation (independent cohort comprising of 61 patients). Crombe and co-workers investigated the role of T2-based MRI delta-radiomics in predicting response of high-grade STS patients to neoadjuvant chemotherapy [18]. A limited number of radiomic features was calculated (33 features) and best predictive performance achieved from 3 top-ranked features (accuracy of 74.6%). Another study conducted a multicentric radiomic study and developed a nomogram model to differentiate low-from high-grade STSs [19]. Three radiomic models were implemented and the highest AUC that was reported from the nomogram model was 0.78% using an independent validation set. Multiple ML models were applied to conventional MRI data from 113 patients and radiomic data were divided into the training set (80 patients) and the validation set (33 patients). The performance of the examined models was assessed using a single validation set and an AUC of

0.96 was reported after radiomic feature selection and random forest (RF) classification. Another similar to our study performed a radiomic analysis on fat-suppressed T2-weighted (T2w) MRI on a 3.0T scanner from 35 pathologically diagnosed STS patients, identified 5 radiomic features that best discriminate low from high-histopathological grades. The provided model obtained an AUC of 0.92±0.07 using a 5-fold cross-validation. However, model performance was calculated within the algorithm training from the average accuracy of the 5-folds [20].

1.3.3 Unmet needs in STSs MRI analysis

Despite the aforementioned promising findings, in most of the quantitative studies, cellularity, neovascularity and microstructural complexity were quantified from summary statistics on single model parameters assuming that single models can solely characterize the overall tissue microenvironment. However, since DWI analysis with different models underpin different aspects of the tissue properties and tumor heterogeneity is evident in STSs, single models may fail to appropriately capture regional functional and anatomical tumor properties, concluding to diffusion parameters and statistics with no prior examination of models' applicability.

The aforementioned limitations motivated a number of studies to assess model suitability in a number of different cancers. Authors from [21] conducted a DWI analysis for antiandrogen treatment response in prostate cancer bone metastasis using three diffusion models. Model selection was performed across four different time points and stretched exponential was favored as the most suitable model to monitor response to treatment. A recent study investigated the role of diffusion related parameters from two models in radiotherapy response using pre- and post-treatment DWI data [22]. Spatial variations of the suitability of the models were prominent between pre- and post-therapy, indicating changes in tumor microstructure and the presence of different tumor subregions. Results from [23] reported that when brain tumor heterogeneity is ignored, single pharmacokinetic model analysis applied to Dynamic Contrast Enhanced MRI yielded to incorrect model parameters in over 35% of the tumor pixels. Additionally, when fitting quality of several models was examined, more than three models were required to quantify robustly the biological properties of the whole tumor, suggesting that pixel-based values in the parametric maps should be estimated according to the model selected for each particular pixel. Although some recent publications highlight the necessity to embed model selection in the quantification of the whole tumor microenvironment and the parametric map calculation, to the best of our knowledge, the potential benefits these approaches offer to discriminate patients into relative disease groups are still understudied. To this extend, none of the studies in DWI has explored the synergistic effect of quantifying DWI from multiple models, generating single parametric maps from several models using model selection techniques and

perform statistical analysis on these maps to differentiate patient groups of different cancers, including STSs.

Towards this direction, the main goal from this study was to present a novel MRI-guided computational framework, using DWI from multiple b-values, to quantify images on a pixel level using 4 different models, statistically identify and select models that best characterize particular STSs subregions, classify these regions into models and create corresponding pseudo-color classification maps, generate new parametric maps (meta-maps) based on the relative information content of each examined model as displayed in the classification map, and perform histogram analysis on all derived maps to differentiate low from high grading of STSs. The presented workflow was verified histopathologically using the surgical specimens and biopsy, yielding to novel parametric meta-maps of high discriminatory power.

Although radiomics has been extensively studied in many anatomical areas including head and neck, breast, liver and lung cancer, to the best of our knowledge, few studies examined the role of radiomics in STSs grading [24][25]. Additionally, major challenges towards translating radiomics into the clinical practice are usually arise due to the lack of a robust and transparent framework for the validation of the ML results [26]. Even if a significant number of studies reports good performance in their results, overfitting problems, feature stability and reproducibility as well as assessment of the ML performance using the so-called internal validation set are usually ignored or underestimated when designing the ML analysis framework. It is notable that a recent review reported an average of only 6% of the 516 published models that were developed using a proper validation schema [27]. Motivated by these observations, a secondary goal was considered in this study in order to examine the application of radiomics in the differentiation of the STSs staging under a careful design of the algorithms' validation based on nested cross-validation comprising multiple independent validation sets for ML performance assessment. To this end, high resolution fat-suppressed T2-weighted images were used for the analysis as they provide a rich spatial resolution, thus a more detailed representation of the disease.

1.4 Thesis Roadmap

This doctoral thesis comprises seven chapters in total. Given the main epidemiology behind soft tissue sarcomas, the role of advanced MRI acquisition protocols and post-processing techniques comprising DWI quantification, radiomics and machine learning towards the disease characterization and patients' stratification (Chapter 1), Chapter 2 gives a brief overview of the soft tissue sarcomas and their different subtypes in terms of the histopathological findings encountered during patient enrollment. Chapter 3 moves into the quantification of the diffusion signal attenuation using several mathematical models, describes the fitting techniques to transform qualitative DWI results into a quantitative representation of the images, outlines the

necessity to assess fitting quality using different statistical metrics and presents a step-by-step workflow for the DWI analysis through an in-house software implemented for this purpose. Chapter 4 presents major aspects in the field of radiomics and machine learning and chapter 5 and 6 address specific clinical questions raised in STSs studies and present the results and discussion related to them. Chapter 7 composes the final concluding aspect of this work as an epilogue.

Chapter 2 Soft Tissue Sarcomas (STSs)

STSs are neoplasms arising from the mesoderm derived tissues such as muscle, fat and connective tissue, thus constituting a broad and heterogeneous category of space occupying lesions. More than 50 different STSs subtypes have been defined by the World Health Organization (WHO), which are often associated with distinct radiological phenotype, different tumor biology and clinical outcome [1]. WHO divides tumor into benign, low grade (locally aggressive), intermediate grade (rarely metastasizing) and malignant. Histopathologic type, grade and tumor size and depth are determinant factors for soft tumor staging and therefore provide significant prognostic information. In the case of a malignant soft tissue tumors (sarcomas), wide excision of the tumor together with a rim of adjacent healthy structures is the surgical treatment of choice to reduce the risk of local recurrence. Grading is based on the analysis of the degree of cell differentiation, histopathologic subtype, mitotic activity and presence of necrosis. The most widely used systems for grading are the three-tiered ones suggested by the French Federation of Cancer Centers Sarcoma Group (FNCLCC) [28] (Table 1).

Table 1 FNCLCC Scoring system for soft tissue sarcomas

Factor	FNCLCC Scoring system
Differentiation	Score 1 (close similarity to normal cells) to score 3 (abnormal cell morphology)
Mitotic Count	Score 1 (low mitotic activity) to score 3 (high mitotic activity)
Extent of Necrosis	Score 0 (very little dying tissue) to score 2 (larger area of dying tissue)

Staging in practice though, has functioned through a two-stage system classifying tumors as low or high grade [29]. Core needle biopsy has been established for preoperative tumor grading in an attempt to classify tumors as high or low grade in order to contribute to the most appropriate therapeutic scheme. Although histopathologic assessment of biopsy samples is the gold standard method for accurate tumor characterization and grading, it might often be subject to sampling errors underestimating thus tumor grade and misguiding therapeutic approach [2]. Moreover, biopsy is an invasive procedure that can provoke several undesirable effects such as bleeding, pain, wound infection or breakdown and spillage of tumor cells. Noninvasive tumor

characterization at the early stage of imaging is therefore of utmost importance to ensure the choice of the most appropriate therapeutic plan and minimize patient discomfort. The following sections give a thorough description and an illustrative overview from MRI and histology imaging of the most common STSs types. The diagnostic pipeline workflow followed from this thesis is displayed in Figure 1.
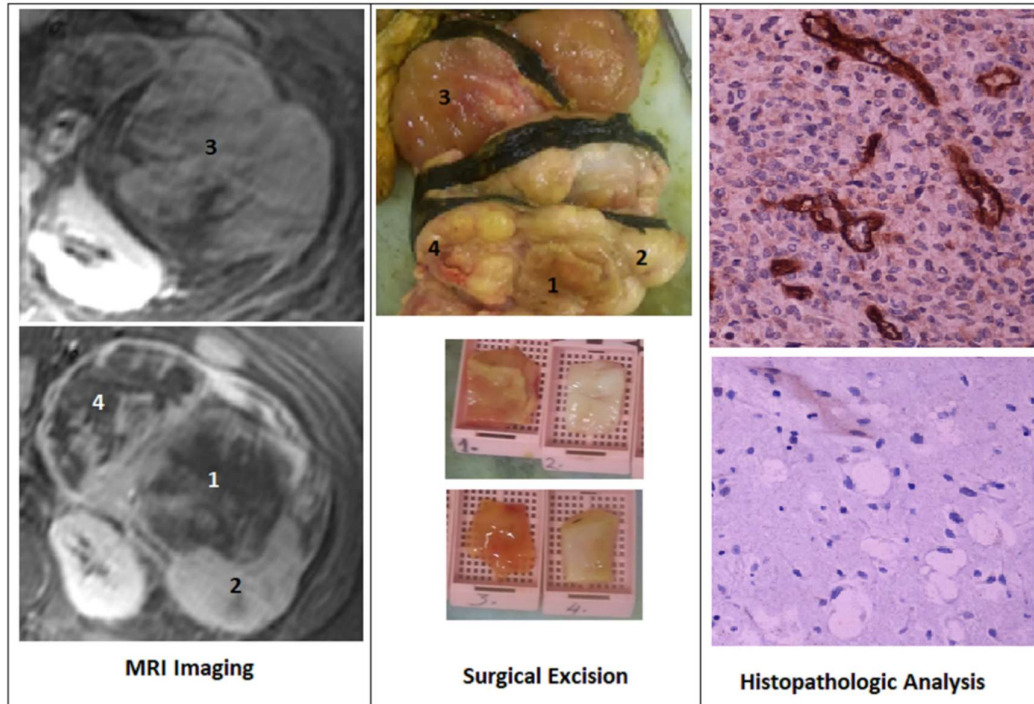
Figure 1 Diagnosis Workflow Steps followed during this study (images from a dedifferentiated liposarcoma) Specimen 1: Necrosis, 2: poorly differentiated, 3: well differentiated, 4: moderately differentiated liposarcoma.

2.1 Dedifferentiated Liposarcoma

Dedifferentiated liposarcomas (DDLS) are malignant adipocytic tumors showing transition from atypical lipomatous tumor/well-differentiated liposarcoma (ALT/WDL) to a nonlipogenic sarcoma of variable histological grade. DDLS are highly aggressive tumors and most of them underlie chromosomal alterations (12q14–15 amplification involving *MDM2* gene) to well differentiated liposarcomas. On MRI, ALT/WDL present as predominantly lipomatous tumors with thick septa (>2 mm), globular and/or nodular areas, and/or associated masses. The nonlipomatous component has low signal on T1-weighted MRI and intermediate to high signal intensities on T2-weighted MRI due to histologic variations in the dedifferentiated region of the tumor mass. The thick septa or nodules enhance markedly after contrast administration. Figure

2 depicts a dedifferentiated liposarcoma case from our studied cohort using multi-sequential MRI examination.

Figure 2 MR imaging of a dedifferentiated liposarcoma. From left to right: T2 fat suppressed, DWI at b zero and T1 fat suppressed after contrast medium administration.

Indicative histology results depicted in the figures below (Figures 3-7) underline the highly heterogeneous structure of a dedifferentiated liposarcoma. The histology images were acquired from the true axial tumor plane in the middle of the tumor length, corresponding to the central tumor MRI slice of the patient, showing different tumor subregions.
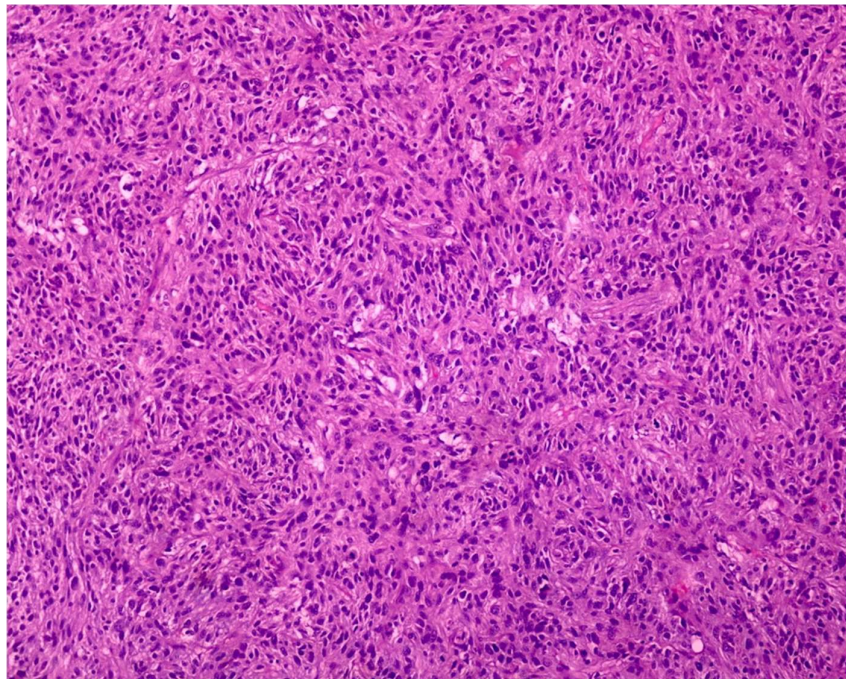
Figure 3 Microphotograph reveals well-structured areas of high cellularity. It illustrates an area of the dedifferentiated non-lipogenic component of the tumor with an intermediate grade of malignancy consisting of relatively monomorphic tumor cells with cytologic features of moderate dysplasia a solid architectural pattern, an increased vascularity.
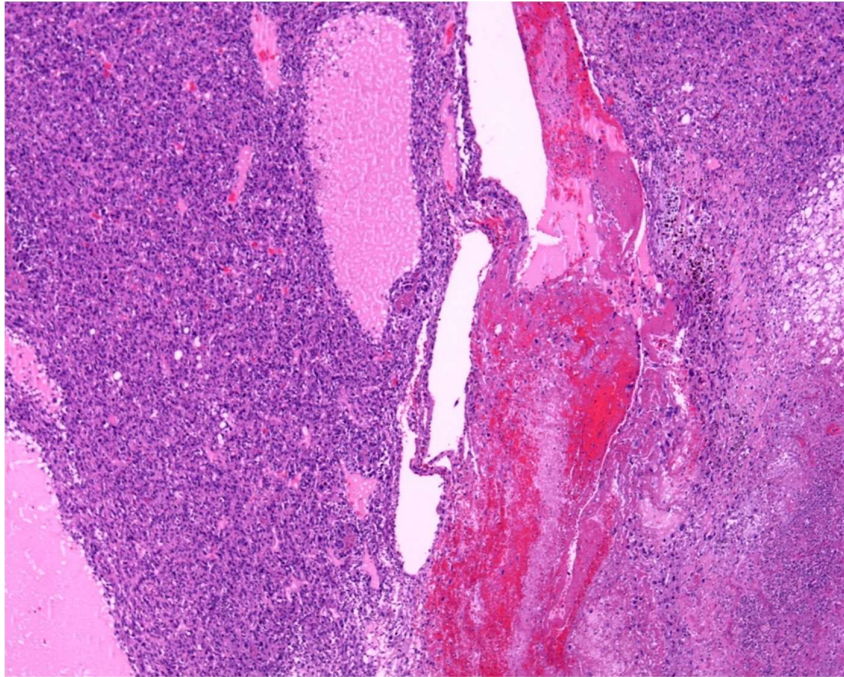
Figure 4 Microphotograph of a highly heterogeneous architecturally and cytologically area of the dedifferentiated non-lipogenic component with solid regions consisting of pleomorphic anaplastic tumor cells with high mitotic rate, plentiful abnormal mitoses, cystic and hemorrhagic cavities, a small area of necrosis. In the area is also included a distinct small area of the well-differentiated lipogenic component (middle right).
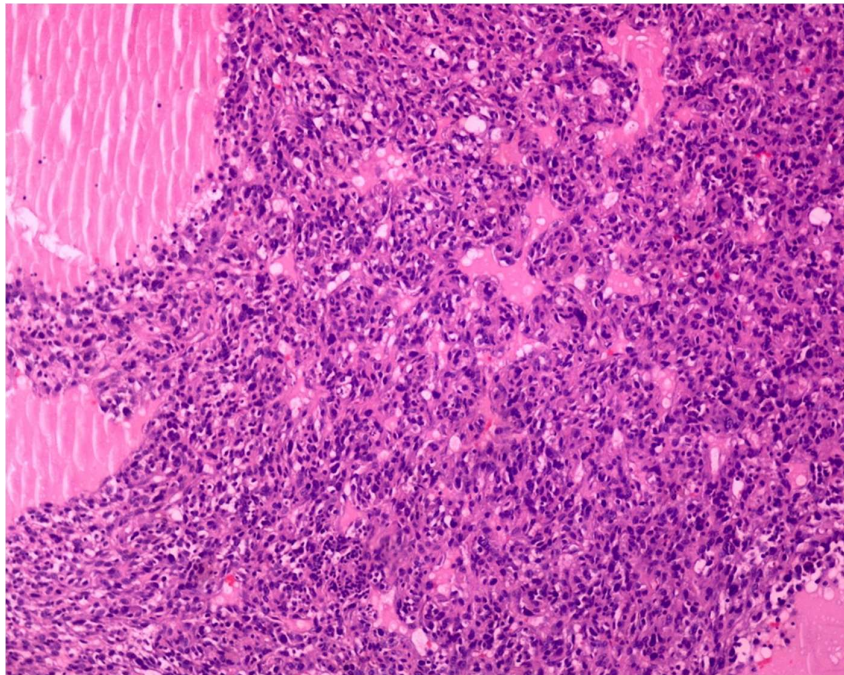
Figure 5 A high-grade cellular area of the dedifferentiated non-lipogenic component with high vascularity, and both solid and cystic architecture.
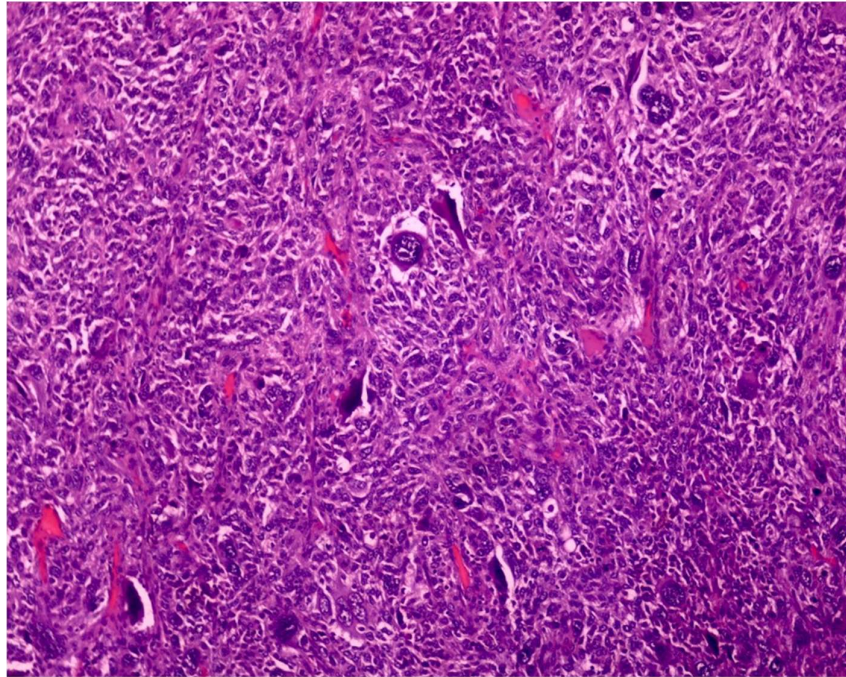
Figure 6 The most cellular and highest-grade area of the dedifferentiated non-lipogenic component consisting of a highly pleomorphic tumor cell population including gigantic and multinucleated tumor cells with nuclear features of extreme anaplasia, very high cellularity and increased vascularity with abnormally structured blood vessels.
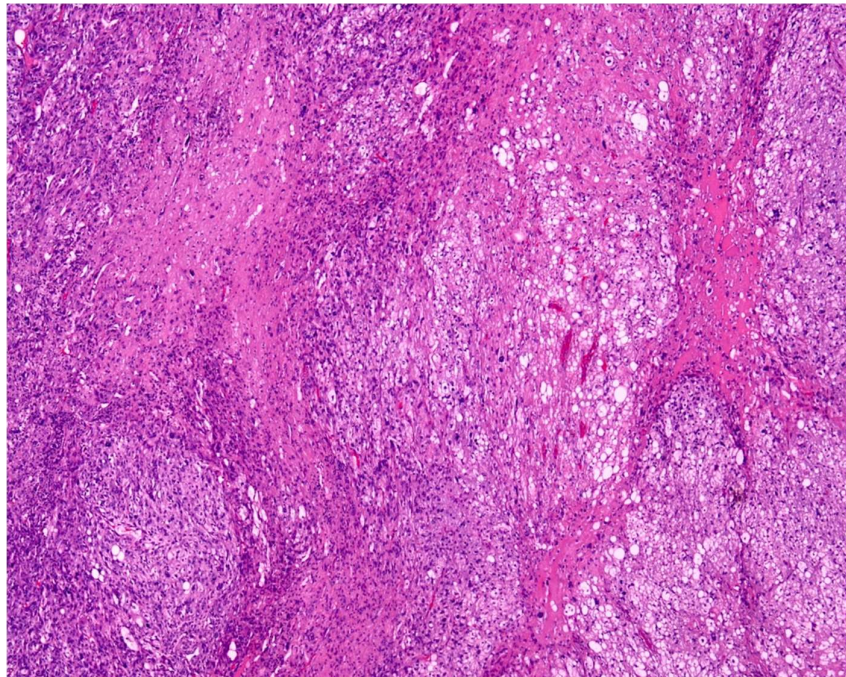
Figure 7 An architecturally heterogeneous and complex area of the tumor highlighting areas of transition between the relatively "well-differentiated" lipogenic and the dedifferentiated high-grade non-lipogenic tumor components, regions of low and high cellularity, regions of dilated and ectatic abnormal blood vessels.

2.2 Pleomorphic Liposarcoma

Pleomorphic liposarcoma (PLS) is a rare subtype of liposarcoma (5%) characterized as a fast-growing tumor, highly metastatic (more than 50% the risk of metastasis, primarily to the lungs)

and with increased mortality rate. PLS is characterized by a number of pleomorphic lipoblasts that are frequently very large and contain irregular, hyperchromatic, scalloped nuclei, with prominent nucleoli and multi-vacuolated cytoplasm. Genetically, pleomorphic liposarcomas have complex caryotypes with a high frequency of p53 mutations. Treatment involves surgery and preoperative chemotherapy in cases where PLS tumors are larger than 5-8 cm. PLS appears as a non-specific soft-tissue mass in MRI where the non-lipomatous areas have intermediate signal on T1-weighted MRI and intermediate to high signal on T2-weighted (indicative PLS in Figure 8).

Figure 8 MR imaging of a pleomorphic liposarcoma. From left to right: T2 fat suppressed, DWI at b zero and T1 fat suppressed after contrast medium administration.

Histologically PLS contains a variable number of pleomorphic lipoblasts, with hemorrhage and necrosis commonly observed (Figures 9-11).
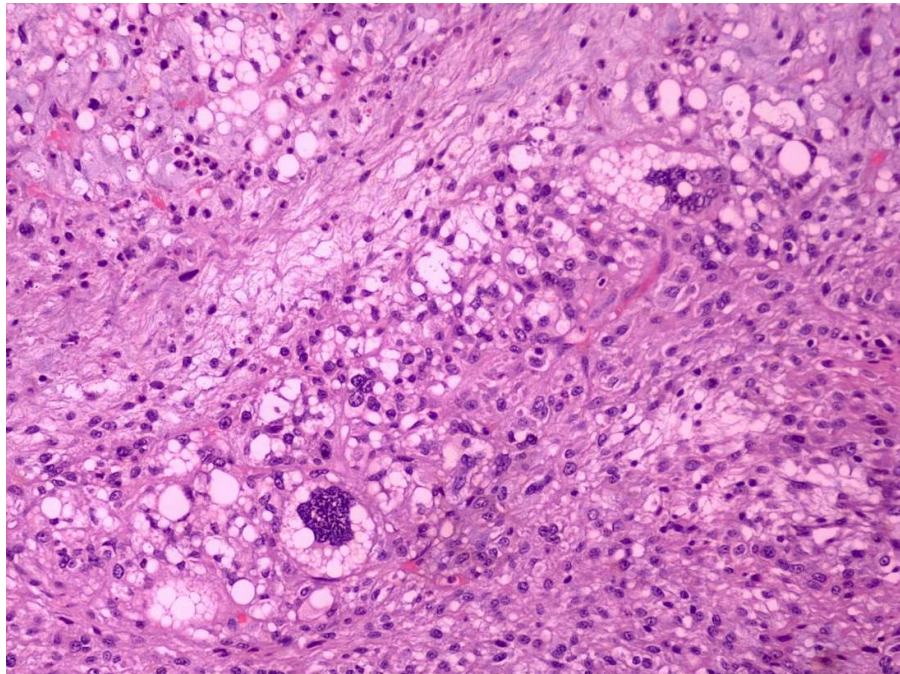
Figure 9 Pleomorphic lipoblasts in a pleomorphic liposarcoma case.
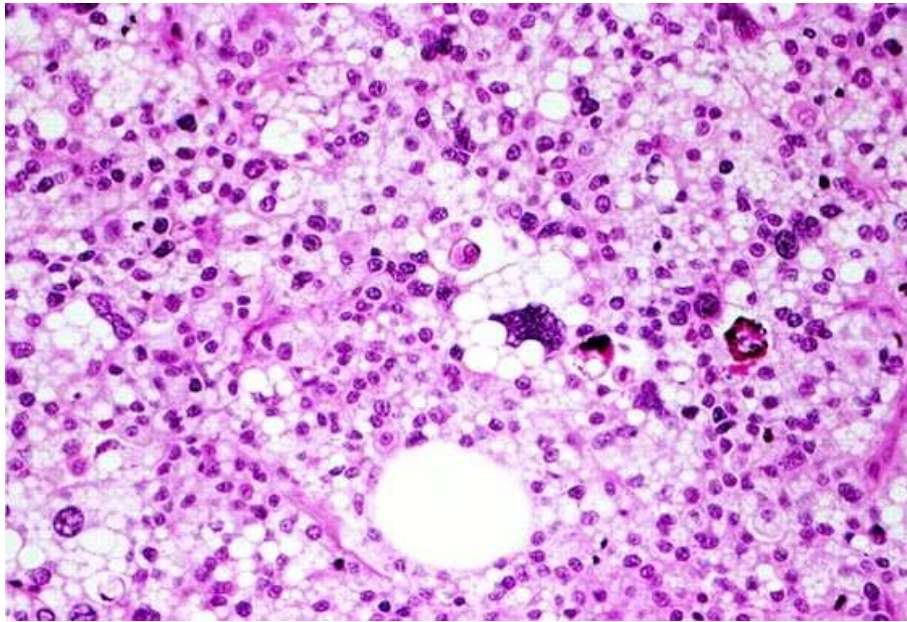
Figure 10 A pleomorphic lipoblast in a tumor composed predominantly of cells with an epithelioid morphology. Image taken from [30].
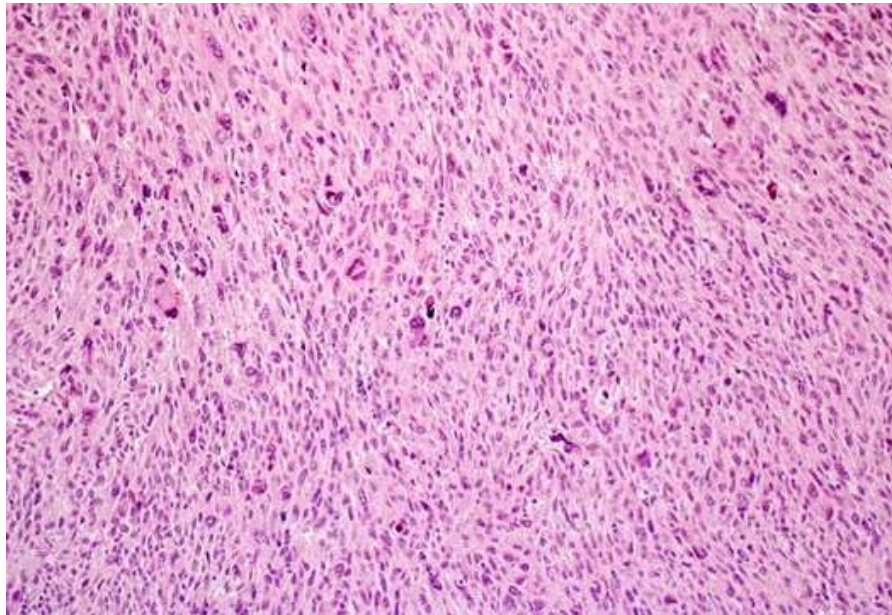
Figure 11 Malignant fibrous histiocytoma-like area of a pleomorphic liposarcoma. Pleomorphic lipoblasts were only focally present in this tumor. Image taken from [30].

2.3 Well-differentiated liposarcoma

Well differentiated liposarcoma (WDL) is a low-grade and slow-growing malignancy. WDL is the most common type in liposarcomas (approximately 50% of all liposarcomas), usually arises in the retroperitoneum and the limbs and rarely metastasize. However, it tends to recur locally and has the potential for de-differentiation and conversion to a higher grade liposarcoma. WDL has predominantly adipose multi-lobulated and well circumscribed masses containing non-

lipomatous components. Fat cells vary in size and there are also scattered lipoblasts and thick-walled blood vessels. Non lipomatous components are most often seen as prominent thick septa within the adipocytic part, characterized by limited morbidity and lack of significant potential for dedifferentiation. Surgical excision is the treatment of choice. A well-differentiated liposarcoma as displayed by a multi-sequential MRI examination comprising T2, DWI and T1 fat suppressed after contrast medium administration is given in Figure 12. Histologically common characteristics of well-differentiated liposarcomas are depicted in Figures 13-15.
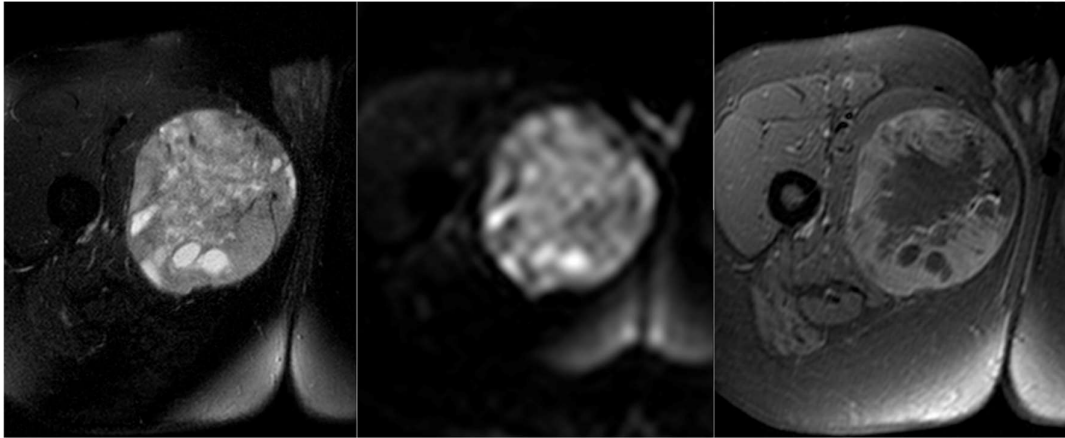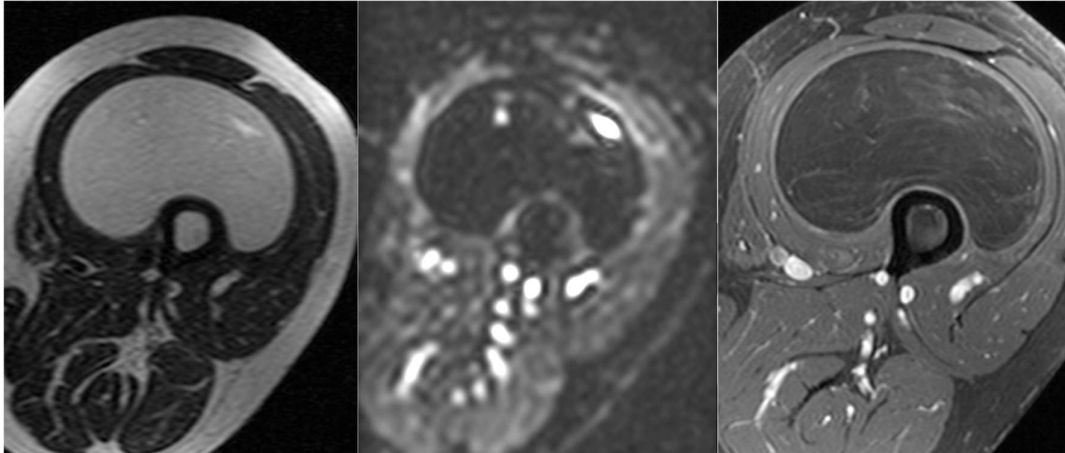

Figure 12 MR imaging of a well-differentiated liposarcoma. From left to right: T2, DWI at b zero and T1 fat suppressed after contrast medium administration.
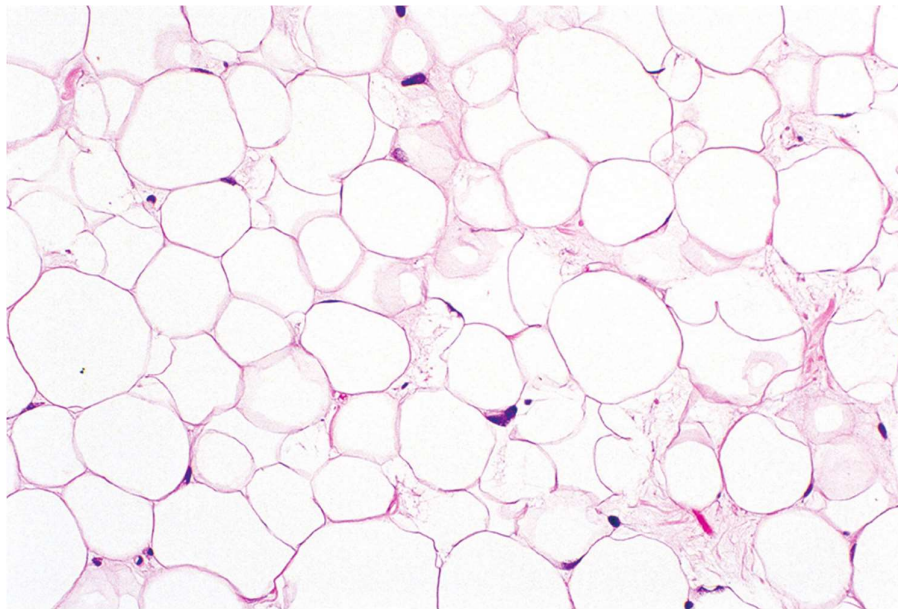
Figure 13 Well-differentiated liposarcoma showing only a rare atypical stromal cell amid a mature lipomatous backdrop. Image from [31].
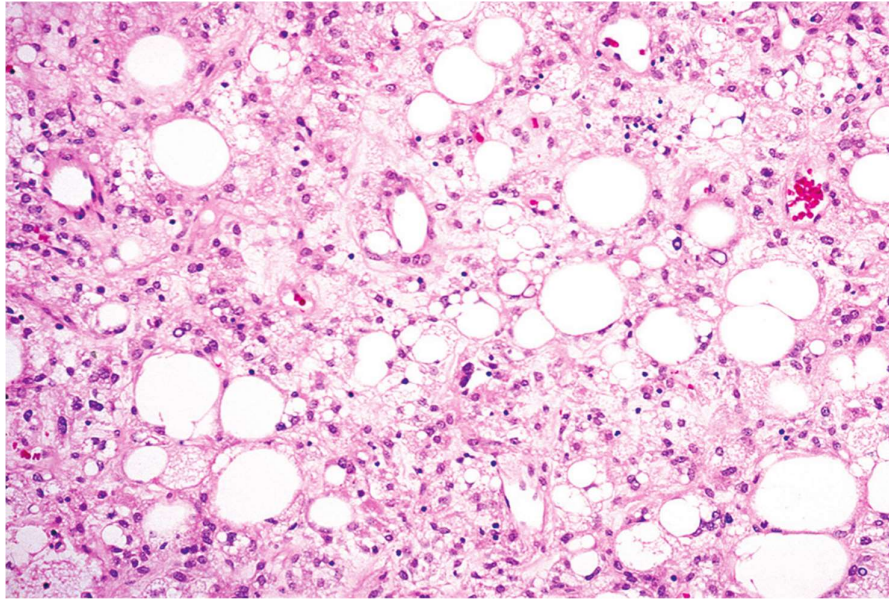
Figure 14 Well-differentiated liposarcoma with numerous lipoblasts. Image from [31].
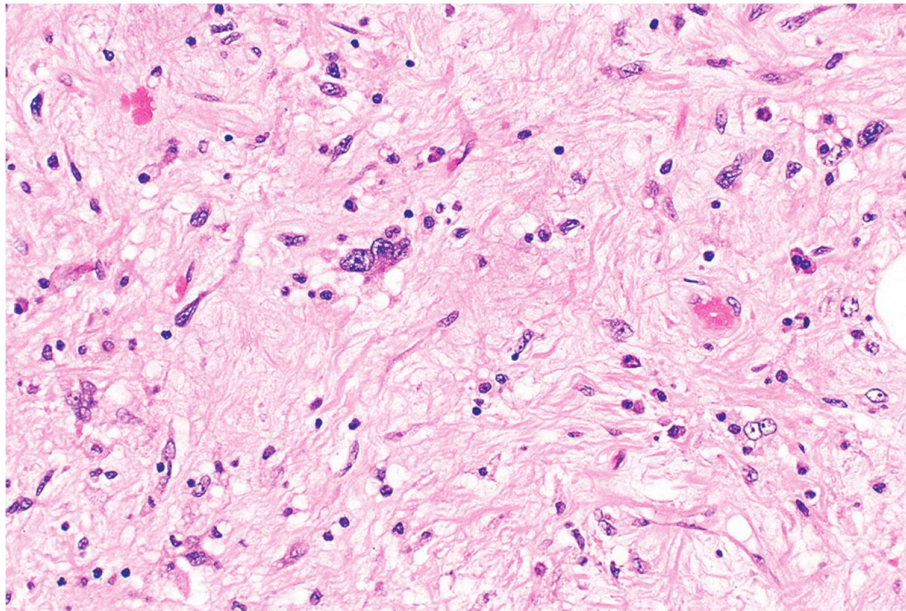
Figure 15 Nonlipogenic zone in a well-differentiated liposarcoma. Image from [31].

2.4 Myxoid Liposarcoma

Myxoid liposarcoma (MLS) is the second most frequently encountered entity after well differentiated liposarcomas. MLS is characterized by a mixed composition of myxoid and round cell components and is characterized by increased cellularity, higher mitotic activity and necrosis (Figures 16-18). A smooth transition is evident between myxoid and round cell components and mature adipose tissue areas occupy a small percentage of the tumor. In general, MLS exhibit low signal on T1-weighted and high signal on T2-weighted but MRI spectrum abnormalities can occur due to the fat and myxoid component, degree of cellularity and vascularity and the presence of necrosis (Figure 19). Contrast enhanced MRI is required to distinguish MLS from other benign

tumors as it sometimes mimic cystic tumors as the extracellular mucoid and myxoid compartments attract water from the blood supply which is trapped inside the lesion.
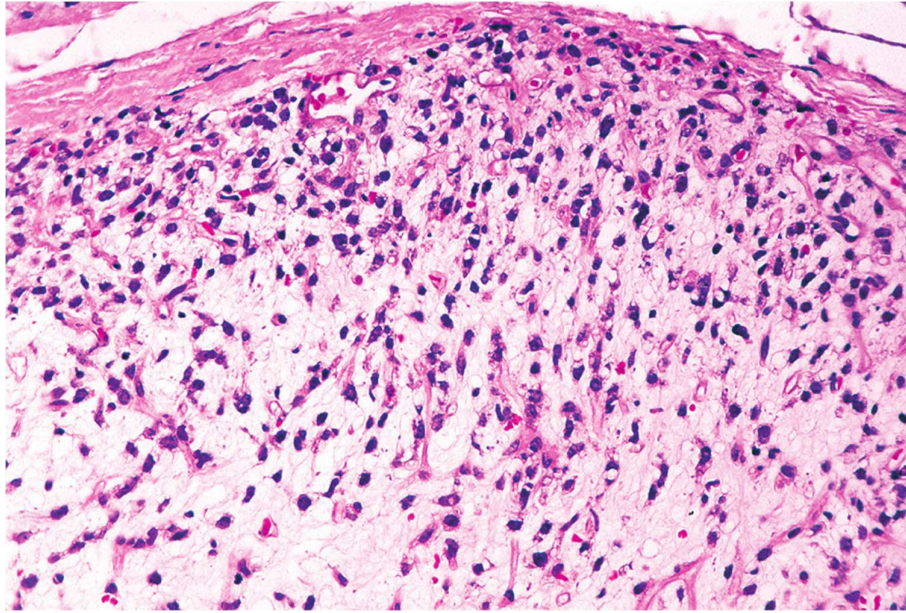
Figure 16 Enhanced cellularity at the periphery of nodules in a myxoid liposarcoma. Image taken from [31].
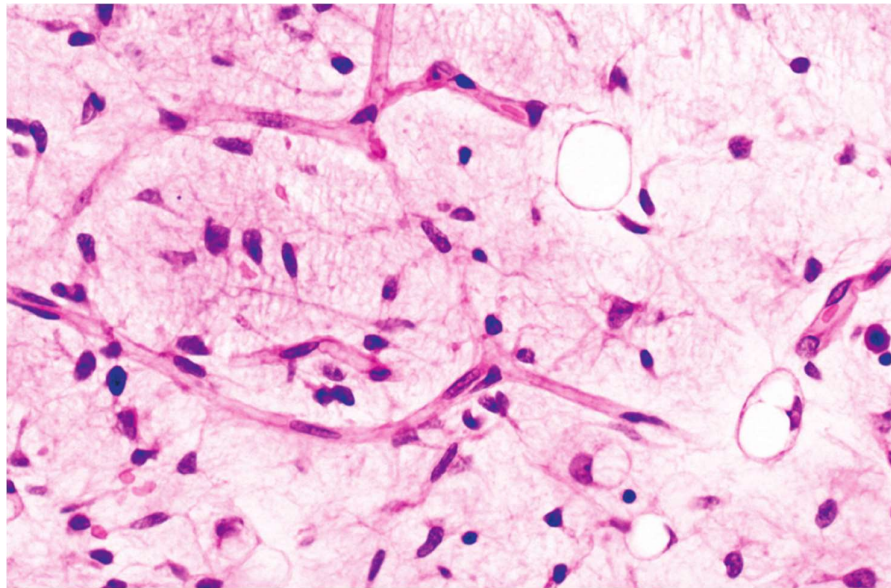
Figure 17 Myxoid liposarcoma with arborizing vasculature and lipoblasts at varying stages. Image from [31].
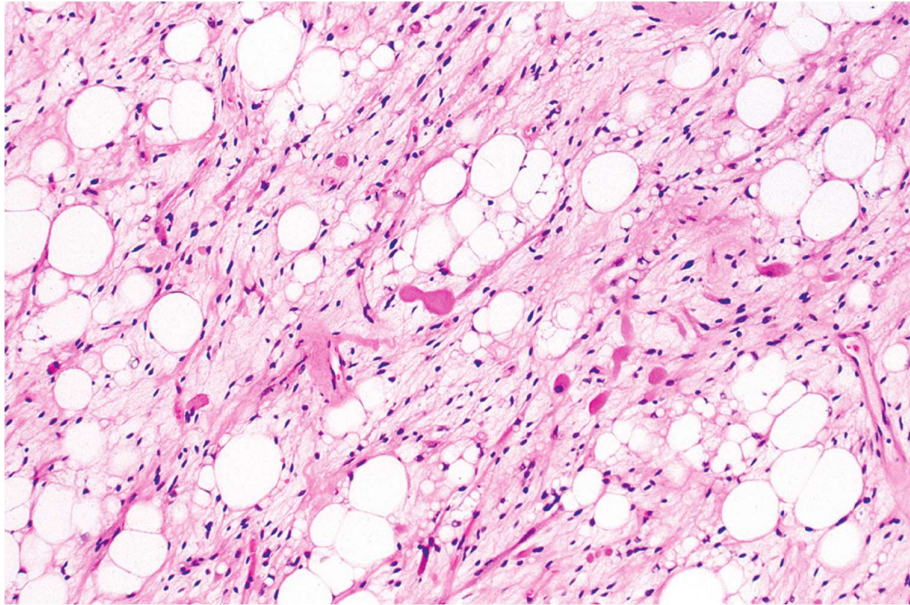
Figure 18 Unusual myxoid liposarcoma with rhabdomyoblastic differentiation. Image taken from [31].
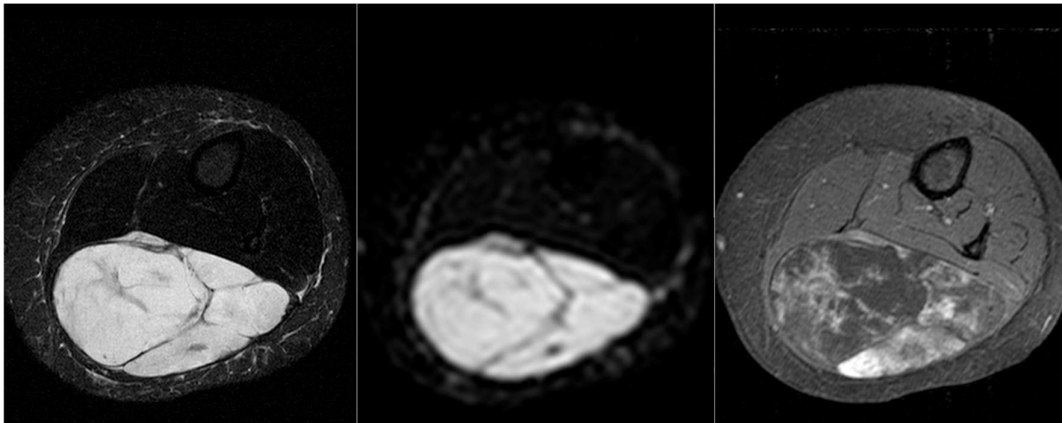
Figure 19 MRI of a myxoid liposarcoma. From left to right: T2 fat suppressed, DWI at b zero and T1 fat suppressed after contrast medium administration.

2.5 Leiomyosarcoma

Leiomyosarcoma is a rare malignant neoplasm (account for 5% to 10% of STSs) that affect smooth muscles, mostly located to the retroperitoneum and can also affect large blood vessels. Three different types of leiomyosarcoma exist, including the conventional/spindle cell type, the myxoid and the epithelioid. Leiomyosarcoma is classified as an STS of poor prognosis and limited therapeutic options and typically are large tumors with a mean diameter of 10cm. Leiomyosarcomas have a fascicular growth histological pattern where tumor cells merge with blood vessels. The tumours are usually compactly cellular and hemorrhagic, but fibrosis or myxoid change may be present. Occasionally, leiomyosarcoma comprises poorly differentiated pleomorphic areas (dedifferentiated leiomyosarcoma) and hypocellular zones and areas of necrosis can be found frequently in larger leiomyosarcoma tumors. In MRI, leiomyosarcoma

exhibits an isointense to muscle T1 signal and mostly a hyperintense T2 fat suppressed signal. Multiple MR images of a leiomyosarcoma are given in the following figure.
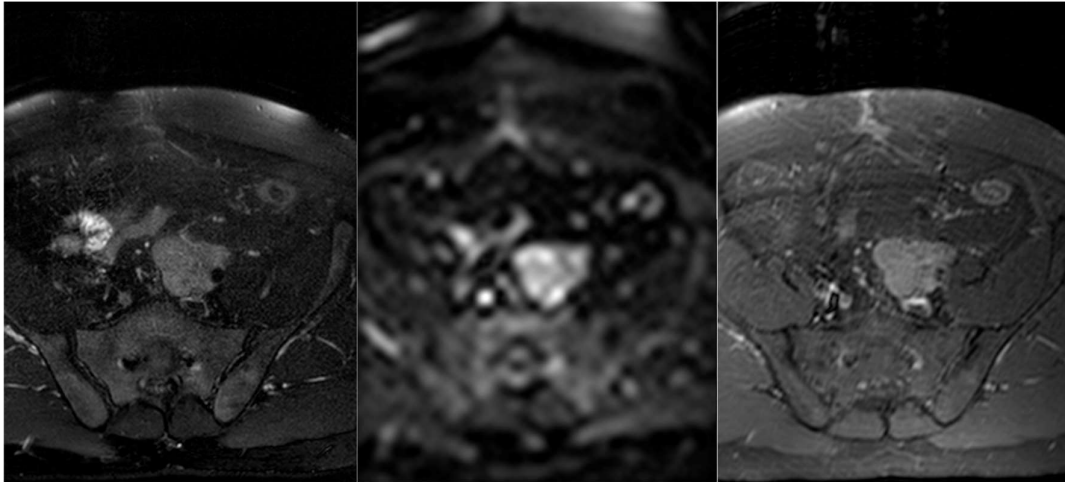
Figure 20 MRI of a leiomyosarcoma. From left to right: T2 fat suppressed, DWI at b zero and T1 fat suppressed after contrast medium administration.

Leiomyosarcoma cases as illustrated from histopathological images (Figures 21-23).
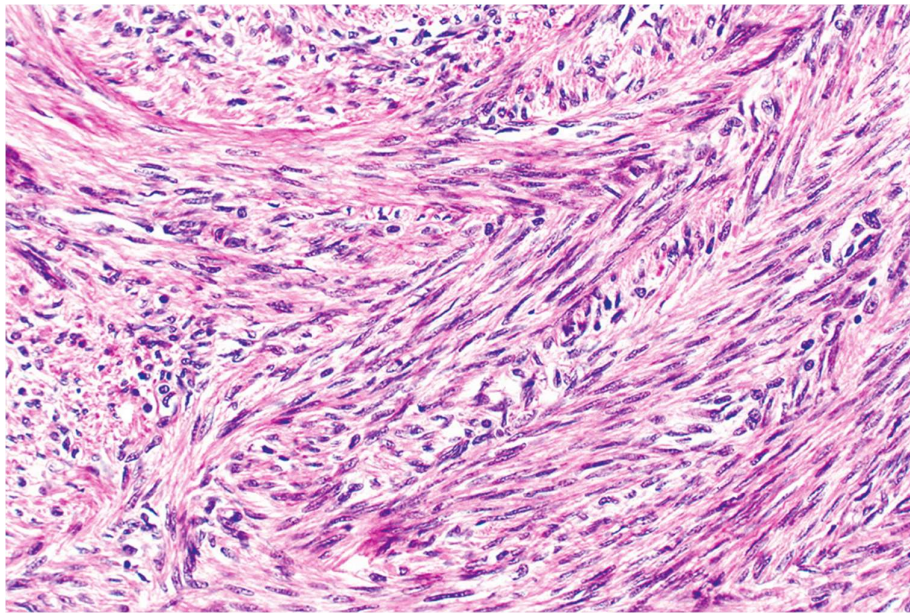
Figure 21 Well-differentiated leiomyosarcoma with a fascicular growth pattern. Image from [31].
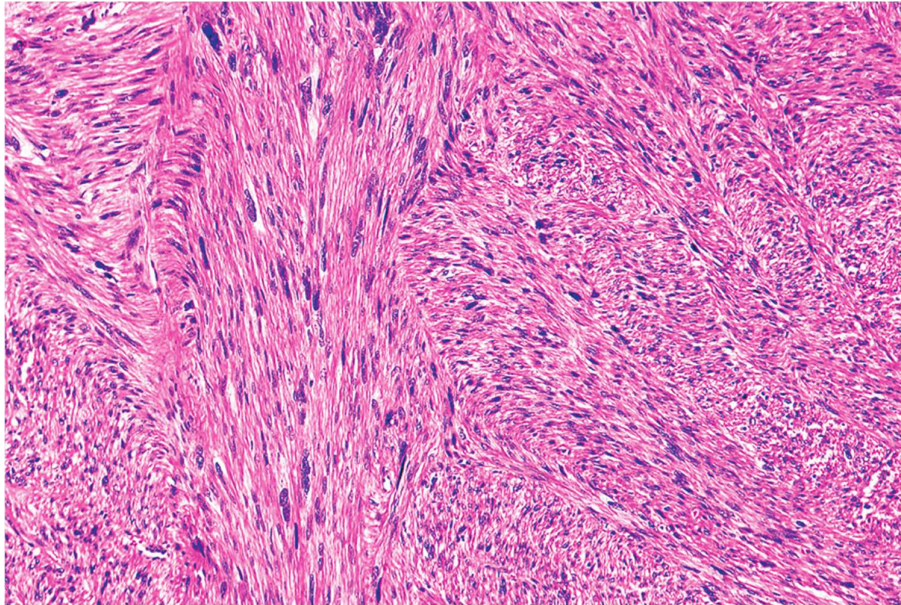
Figure 22 Moderately differentiated leiomyosarcoma composed of deeply eosinophilic fascicles intersecting at right angles. Image from [31].
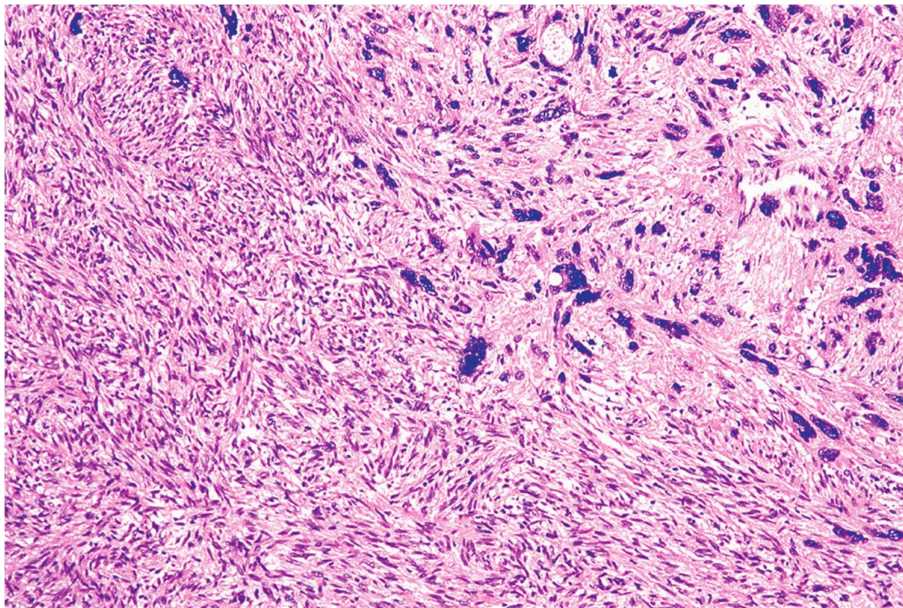
Figure 23 Leiomyosarcoma with pleomorphic areas resembling undifferentiated pleomorphic sarcoma [31].

2.6 Alveolar soft part sarcoma

Alveolar soft part sarcoma (ASPS) is a rare sarcoma of uncertain origin (approximately 0.2% to 1% of all STSs) that comes from different anatomical areas including the muscles, fat and nerves and typically occurs in young patients. The prognosis of ASPS is poor and although other high-grade STSs rarely metastasize to the brain, ASPS is characterized by a late metastatic process including the brain in up to 19% of the cases. Surgery is the treatment of choice while traditional chemotherapy and radiotherapy have failed to provide significant results in patients' survival. ASPS is presented as a large and highly vascularized mass in MRI with an isointense or slightly hyperintense signal on T1-weighted images, hyperintense T2 signal, enhanced signal in contrast

enhanced MRI due to the generated neovascularization network in the solid part of the tumor and high signals on DWI (Figure 24).
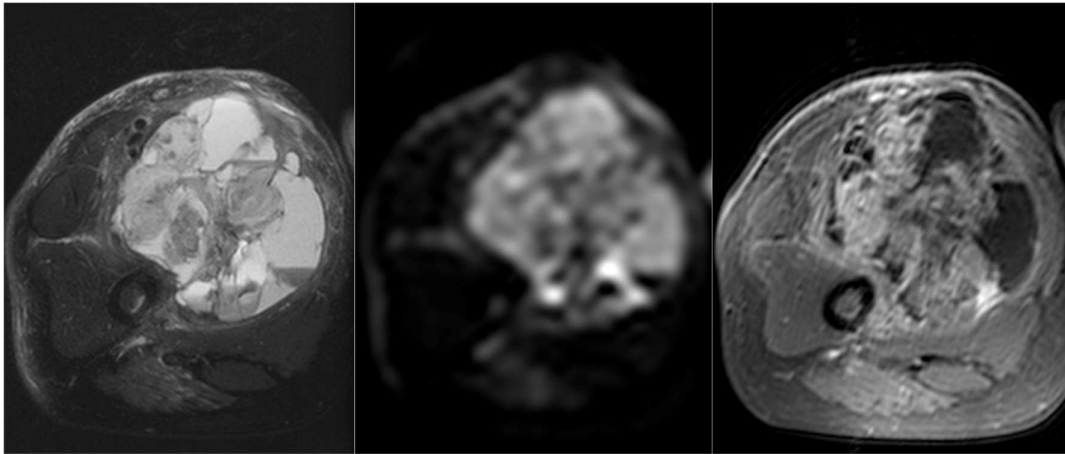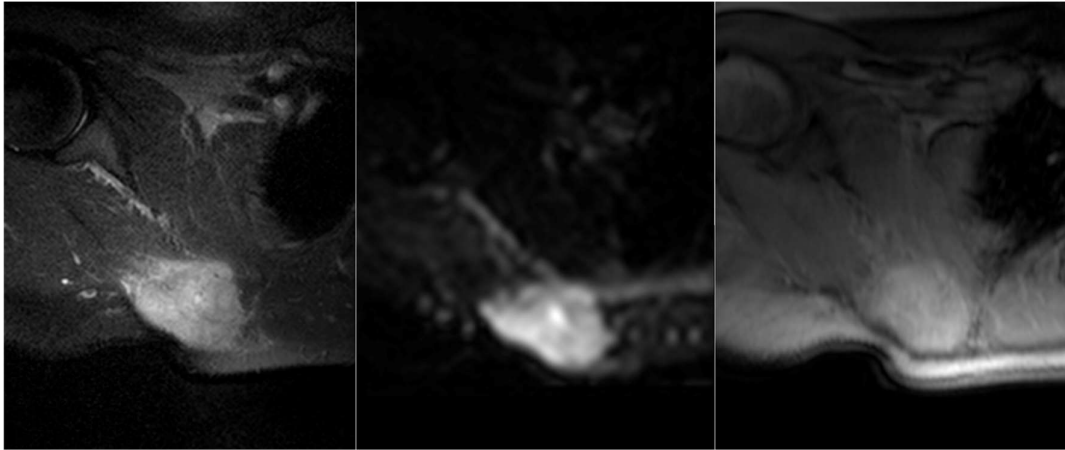

Figure 24 MRI of an alveolar soft part sarcoma. From left to right: T2 fat suppressed, DWI at b zero and T1 fat suppressed after contrast medium administration.

Histological findings report large eosinophilic tumor cells with an intravascular tumor extension in most of the cases. Histology results from different subregions of an alveolar soft part sarcoma are illustrated in the figures below.
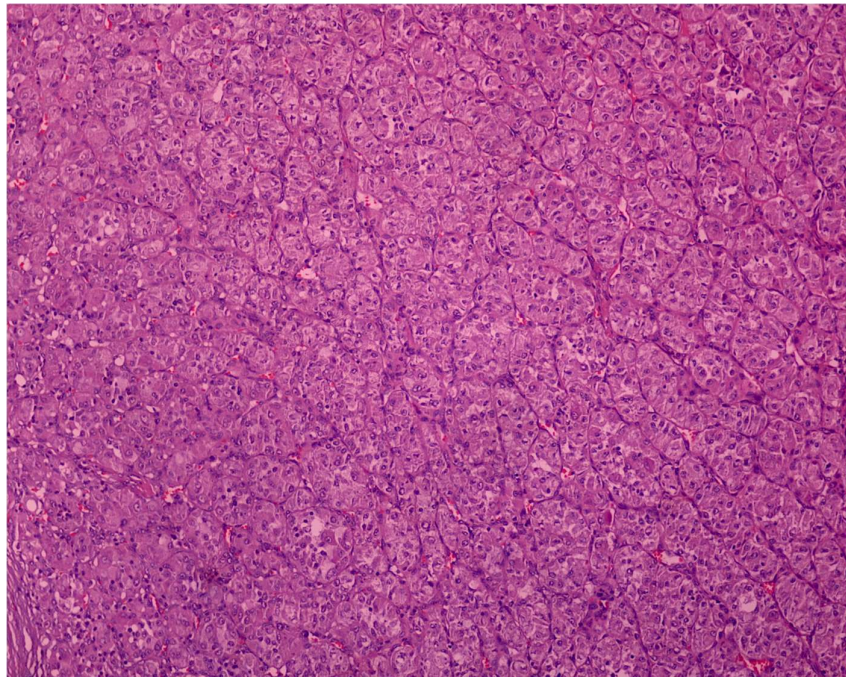
Figure 25 An alveolar soft part sarcoma characterized by the orderly structured well- defined cellular tumor compartments.
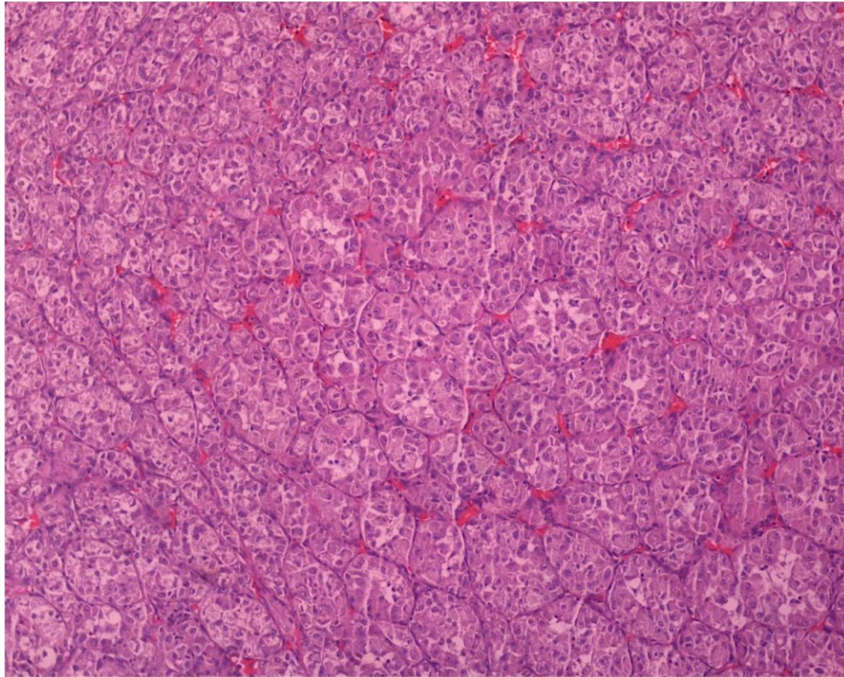
Figure 26 Histology image from an alveolar soft part sarcoma illustrating a tumor area consisting exclusively of geometric honeycomb-like cellular compartments outlined by highly vascularized fibrous septa.
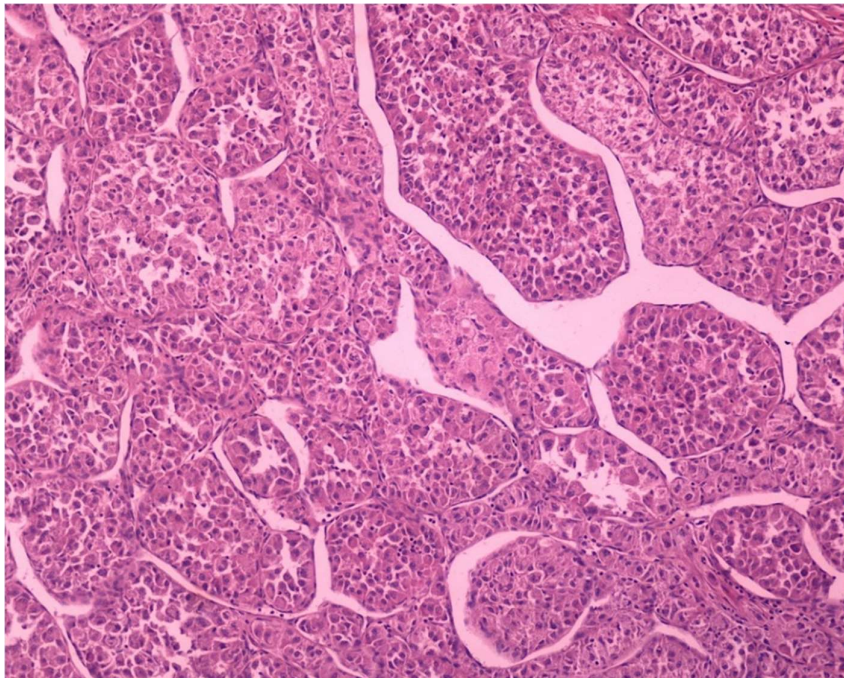
Figure 27 Less monomorphic but still distinct cellular compartments of various sizes and shapes with some variations in cellularity divided by fibrous septa with ectatic (vascular) spaces from an alveolar soft part sarcoma.
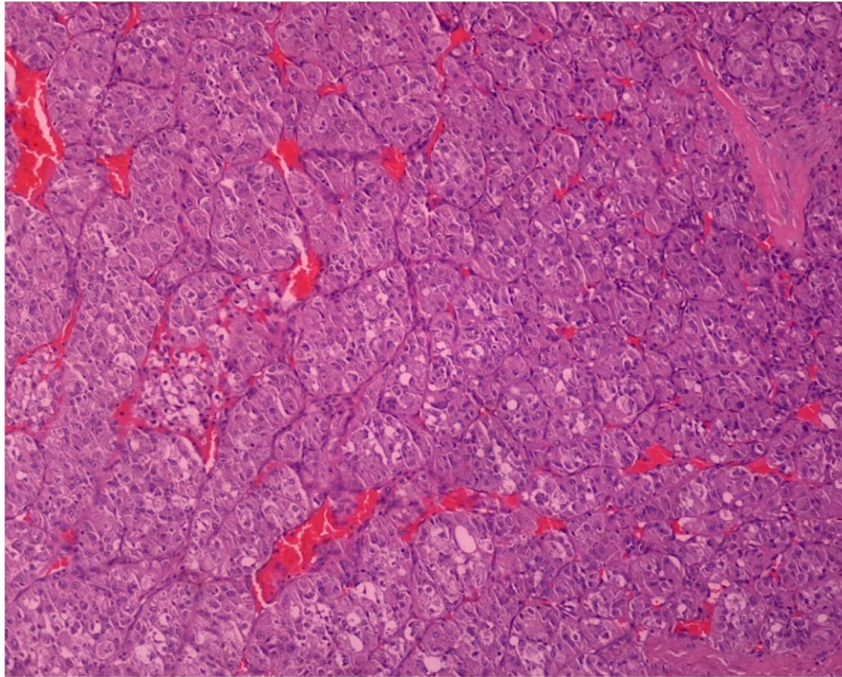
Figure 28 An alveolar soft part sarcoma with geometric honeycomb-like cellular compartments divided by highly vascularized fibrous septa containing congested variably sized blood vessels.
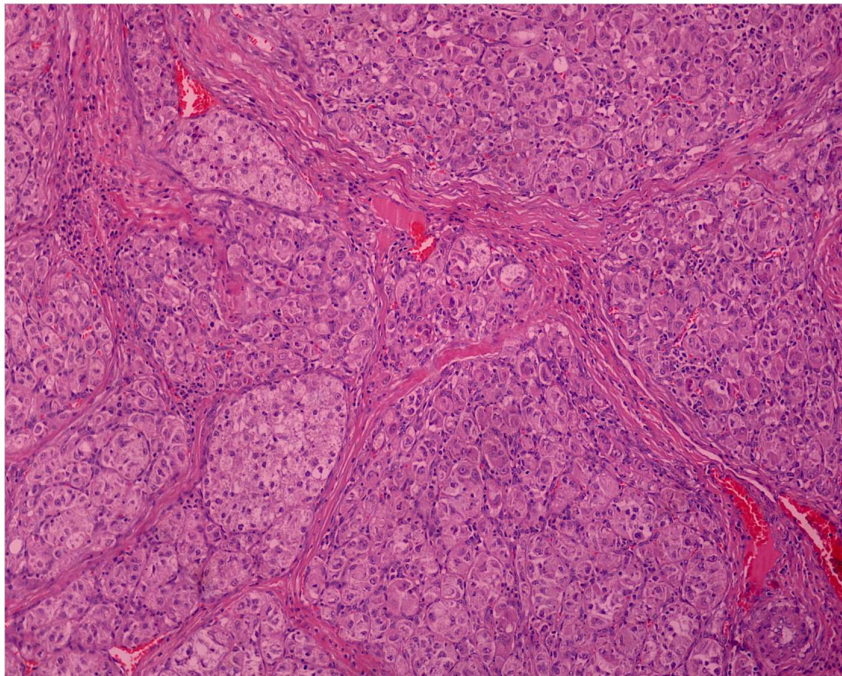
Figure 29 A relatively complex architecturally area of an alveolar soft part sarcoma exhibiting multiple variably sized tumor cell nodules with uneven cellularity outlined by fibrous septa of variable thickness comprising unevenly distributed variably sized and focally congested blood vessels.

Chapter 3 Advanced MRI

This chapter gives a general overview of all developed DWI models and their principles, describes the most widely used techniques to assess model fitting quality, highlights common representations of the DWI quantitative results and concludes to the design of a dedicated

software for DWI quantification further used as a pre-processing tool before the implemented radiomic analysis (Chapter 4). It is highly noted that, as reported in the previous chapters, this study focused exclusively on non-invasive quantitative MRI techniques, thus experiments conducted on DCE-MRI data was beyond the scores of the current thesis. However, DCE-MRI analysis and the current chapter are parts of one of our publications in this field, found in [6].

3.1 Diffusion Weighted Imaging (DWI)

Diffusion is the process of random motion of water molecules in a free medium. For human tissues, water molecules can move inside cells, in the extracellular space, and also in the vessels. A DWI, sensitized to microscopic water mobility by means of strong gradient pulses can be utilized to provide insights in the complexity of the tissue microarchitecture. The amplitude and duration of the diffusion gradients is represented by the "b-value" (measured in s/mm^2), an index used to control the sensitivity of DWI contrast to water mobility. A major requirement in diffusion imaging is to select ultrafast pulse sequences that may freeze macroscopic motion in the form of respiration, peristalsis or patient motion in general. For that reason, Echo Planar Imaging (EPI) sequences modified with the addition of two identical strong diffusion gradients are routinely used to provide diffusion images.

3.1.1 The mono-exponential model (MEM)

ADC [7] was the first and most widely used quantitative biomarker associated to cellular density and the extracellular space fraction [32] with "apparent" giving away a reluctance to use it literally as the distance travelled from the water molecule in a certain time in certain surroundings. The simplest model assumes exponential signal decay where exponential coefficient correlates with the product of b*ADC for each tissue type. For a given b-value, the signal intensity of the Diffusion Weighted Image (DWI) depends on the ADC of tissue in each individual pixel. ADC can be estimated with acknowledgment of two or more measurements on different b-values (one with a b-value of zero and at least one with a higher b-value) as described in the following mono-exponential equation:

$$\frac{S_b}{S_0} = exp(-b \times ADC)$$

where S_b and S_0 denote the acquired diffusion signal for a particular degree of diffusion weighting b (s/mm^2) and the signal without diffusion sensitization, respectively. ADC is the apparent diffusion coefficient (mm^2/s), associated to tissue cellularity. The figure below depicts an ADC map calculated from a patient with myxoid liposarcoma.
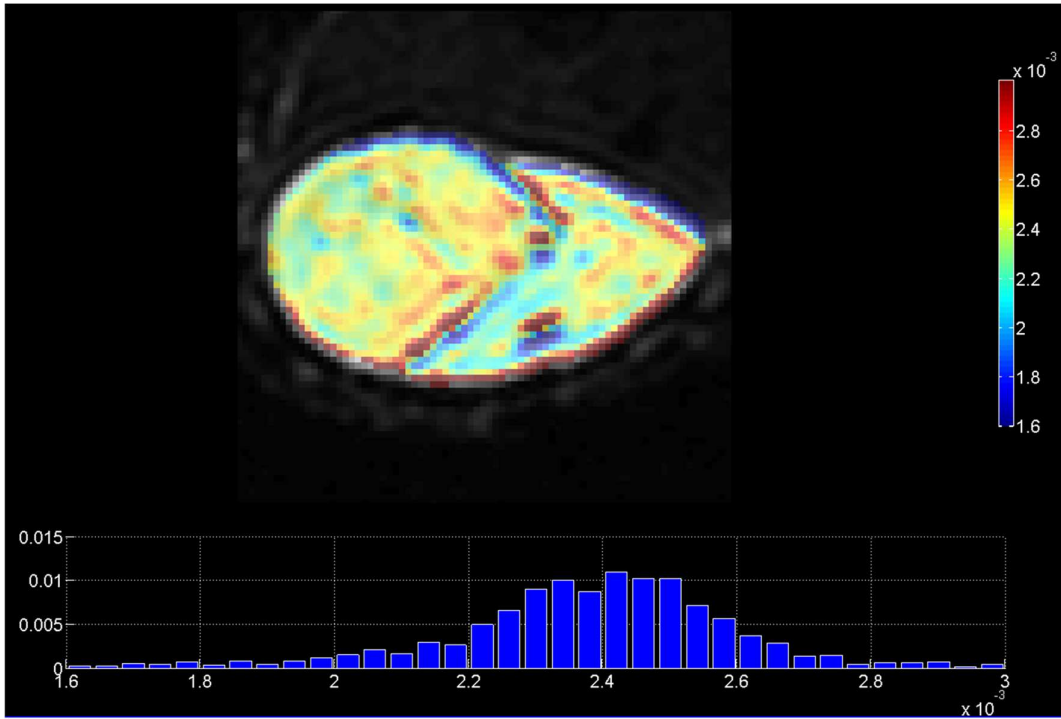
Figure 30 Apparent Diffusion Coefficient (ADC) map and its corresponding histogram of a patient with myxoid liposarcoma of Grade 1.

3.1.2 The bi-exponential model (BEM)

After the introduction of the mono-exponential decay, a more complex model was proposed in order to add sensitivity to the arbitrary motion related to micro-capillary perfusion, which induced deviation from the initially assumed decay. The previously described mono-exponential decay in many cases failed to estimate fast decaying signal appearing in the low b-values area, attributed to faster water motion mimicking diffusion process in tissue, most probably to arbitrary oriented flow in the micro-capillary network. A more complex model, Intravoxel Incoherent Motion (IVIM) [8], was proposed by adding additional acquisitions at low b values to sample the signal in the tissue where the faster decaying component contributes significantly to the overall diffusion signal. The IVIM model assumes that tissue is primarily characterized by two distinct compartments (an intravascular and an interstitial space) with negligible water exchange between them, where the DW signal of each pixel can be expressed from the following bi-exponential equation:

$$\frac{S_b}{S_0} = f \times exp\left(-b \times D_{fast}\right) + (1 - f) \times exp(-b \times D_{slow})$$

Similarly, to the mono-exponential fit equation, S_b is the measured signal intensity of the diffusion-weighted image with a gradient factor attenuation b (s/mm^2), and S_o is the measured signal intensity in the absence of diffusion weighting. The pure diffusion activity is given by the slow diffusion coefficient (D_{slow}) (Figure 31). This new model attempted to measure the pure

diffusion signal contamination with the added term of "pseudo-diffusion" or the fast component of diffusion (D_{fast}), representing signal loss resulting from other processes, most likely micro-perfusion of blood nutrients at capillary level. D_{fast} is associated with blood velocity and vessel geometry. Parameter f is the unitless fraction of fast component (f) (Figure 32).
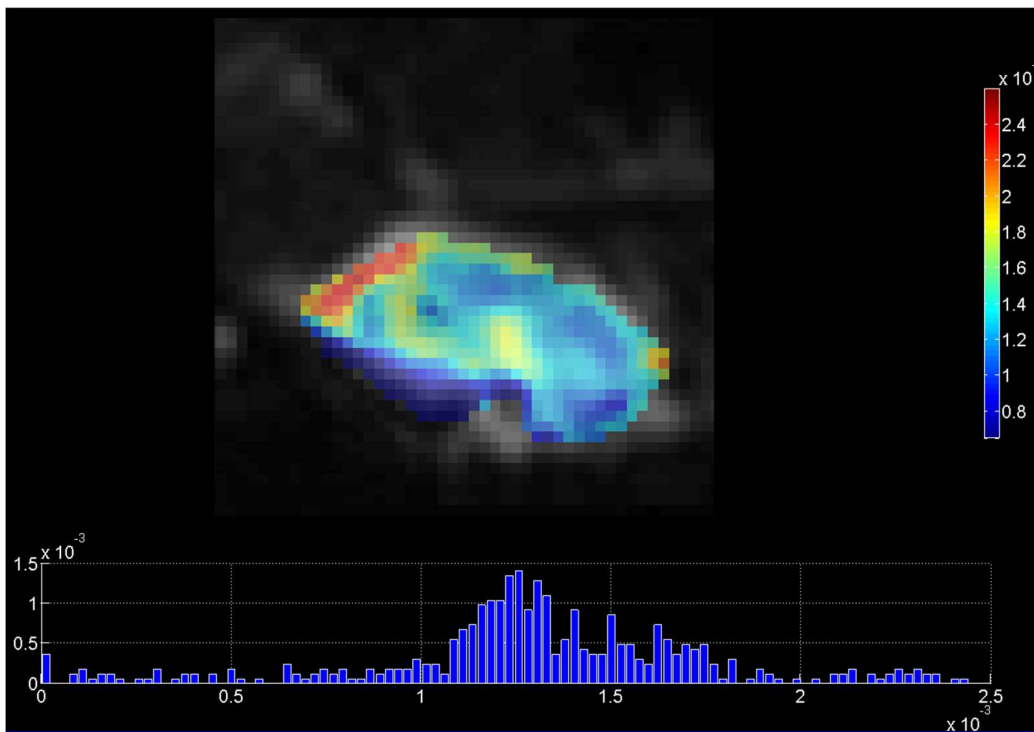
Figure 31 Parametric map of the slow diffusion coefficient (D_{slow}) and its corresponding histogram of a patient with alveolar soft part sarcoma (ASPS) of Grade 3.
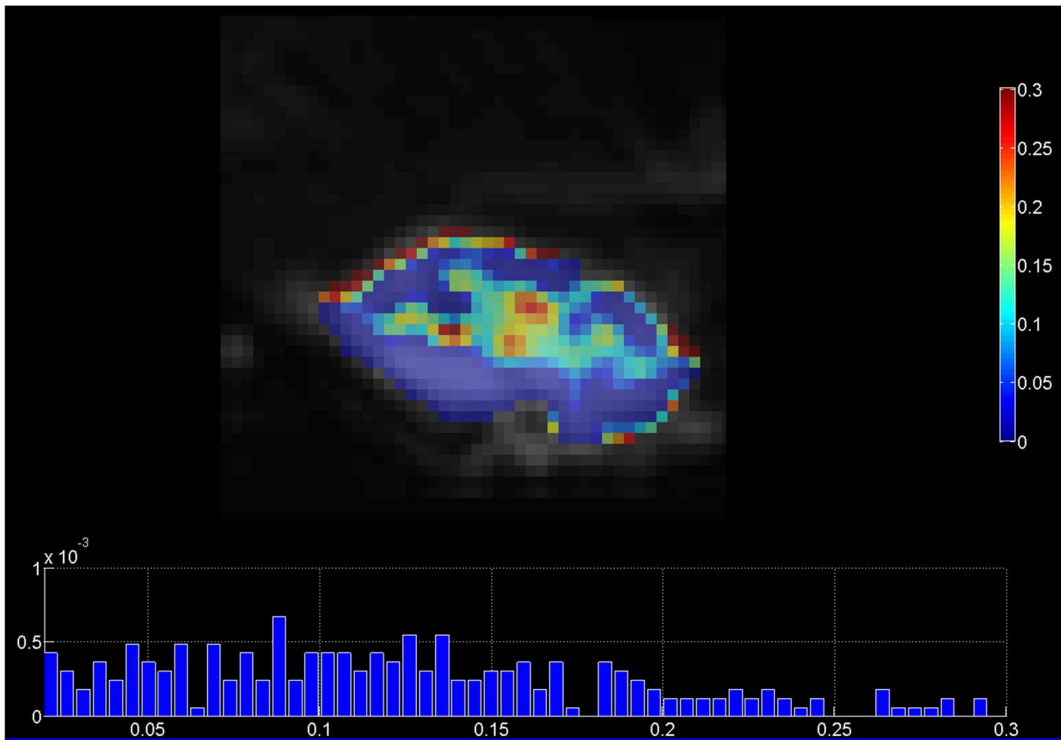
Figure 32 Parametric map of the fraction of fast component (f) and its corresponding histogram of a patient with alveolar soft part sarcoma (ASPS) of Grade 3. High values of f are depicted in the central part of the tumor indicative to the vasculature characteristics of the ASPS.

3.1.3 The stretched exponential model (SEM)

Intravoxel heterogeneity in the distribution of diffusion coefficients because of heterogeneity in fluid viscosity or diffusive restrictions has been quantified in the stretched exponential model [9]. The existence of multiple pools rather than only two inside a ROI or pixel, together with proton exchange between pools, have been pinpointed as the reason of mismatch observed between expected volume fractions and fitted results from DWI data [33]. The proposed model assumes continuous distribution of sources decaying at different rates without any restriction in the number of participating sources and thus the signal attenuation can be attributed to the aggregation of a number of uncoupled decay processes, such that:

$$\frac{S_b}{S_0} = exp(-(b \times DDC)^a)$$

where α is the stretching parameter and a measure of deviation of the signal decay from the simple mono-exponential behavior. Range of α is from 0 to 1. Lower values of parameter α would imply presence of multiple compartments within the ROI, while at the upper limit (α = 1) the model coincides with a simple exponential decay of a homogeneous sample. DDC is the distributed diffusion coefficient (mm^2/s) which is equivalent to the ADC when index α equals to 1. Figures 33-34 present SEM analysis results from a dedifferentiated liposarcoma case.
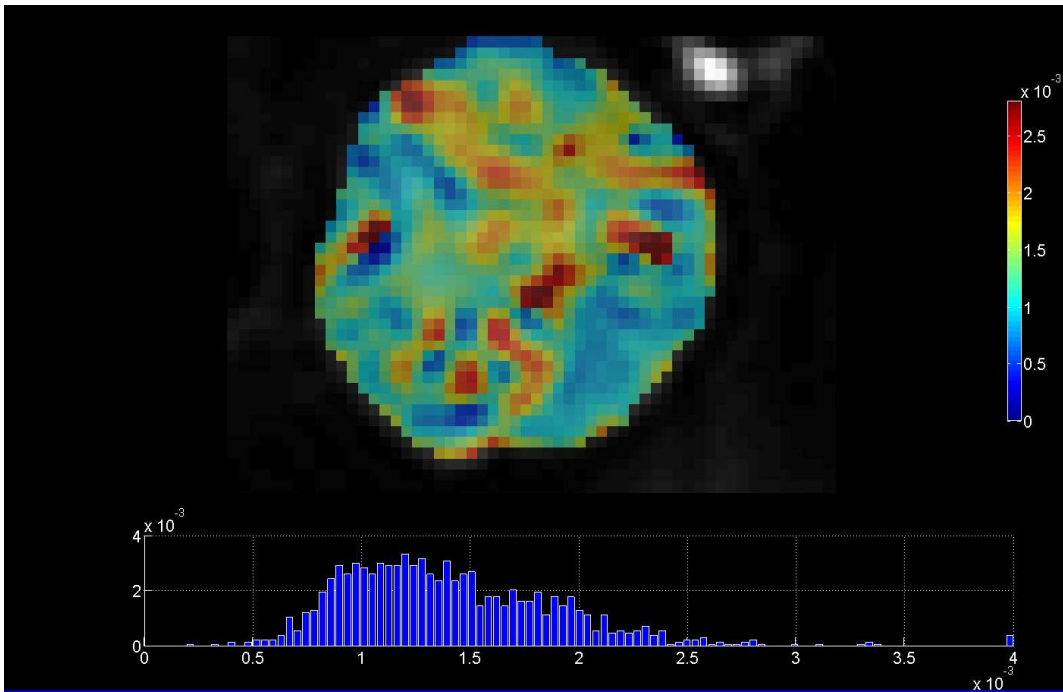
Figure 33 Distributed diffusion coefficient (DDC) map and related histogram from a patient with a dedifferentiated liposarcoma (DDLS) of Grade 3.
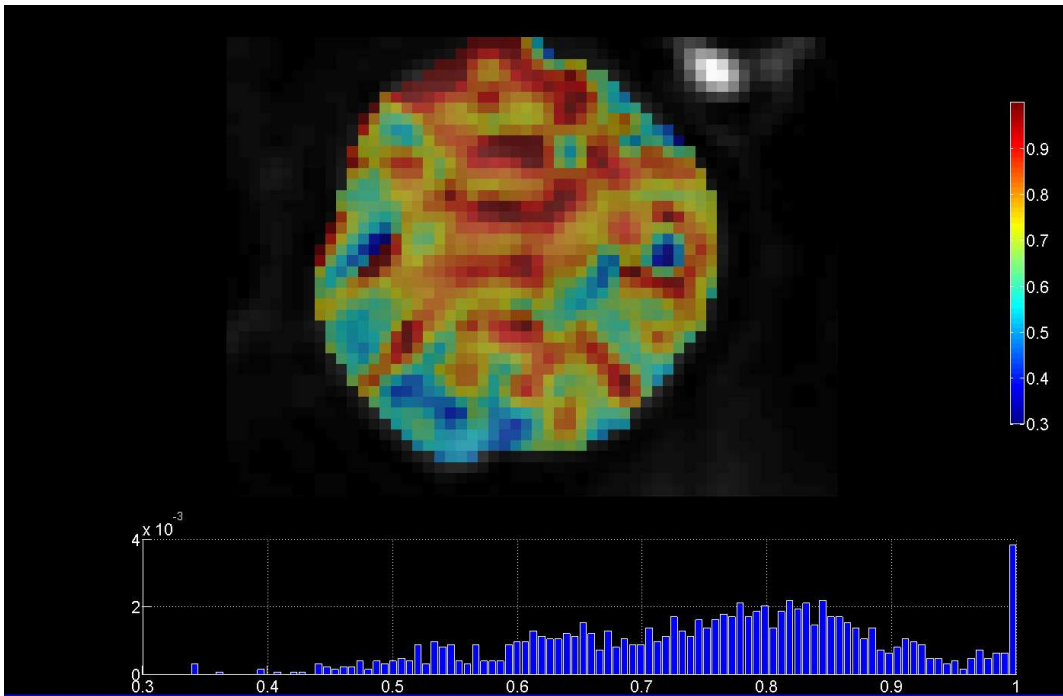
Figure 34 Parametric map of the stretching parameter α revealing highly heterogeneous parts of a dedifferentiated liposarcoma (DDLS) of Grade 3. Corresponding histogram of the tumor slice is also illustrated.

3.1.4 The diffusion kurtosis model (DKM)

Both the mono-exponential and the bi-exponential models rely on the assumption that water mobility follows a random, unrestricted pattern, which can be considered as a Gaussian

displacement distribution. However, in biological tissues the presence of physical barriers like cell membranes or compartments (intracellular and extracellular spaces) restrict the Brownian motion of water [10]. When incorporating the assumption of a restrictive environment, the displacement probability distribution for the water molecules deviates from the Gaussian shape and the degree of this deviation is quantified by kurtosis. Kurtosis is a dimensionless metric expressing the difference of an arbitrary distribution from a Gaussian with the same variance in terms of more or less weight on the center and tails:

$$K_{app} = \frac{M_4}{M_2^2} - 3$$

where M_n is the nth moment of the arbitrary distribution. Similarly, to the ADC, the diffusional kurtosis (K_{app}) is not specific for any tissue property and thus its interpretation in terms of tissue structure is not always well defined. For example, changes in K_{app} might be the overall result of more than one complex processes in tissue and is unable to identify the precise biological mechanisms behind this change (Figure 35). Several models have been proposed in order to study the utility of kurtosis parameters in clinical practice. Most widespread model for kurtosis stems from the expansion of Diffusion Tensor Imaging (DTI) approach where the exponential decay of the signal is analyzed by Taylor series:

$$\ln(S_b) = \ln S_0 - b * D_{app} + O(b^2) + \cdots$$

With the introduction of high (b>1000) values in clinical practice the contribution from the second order term cannot be considered negligible as when employed for DTI calculations and the above expression can be rewritten as:

$$\frac{S_b}{S_0} = exp\left(-b \times D_{app} + \frac{1}{6} \times b^2 \times D_{app}^2 * K_{app}\right)$$

Similarly, when b exceeds a certain upper limit, the latter expression may also suffer from systematic errors in the calculation of D_{app} and K_{app} from the omission of even higher order terms. Apparent kurtosis (K_{app}, unitless) reflects the deviation of water motion from the Gaussian distribution, and apparent diffusion (D_{app}, in mm^2/s) equals to ADC when K_{app} equals to 0 (Figure 36).
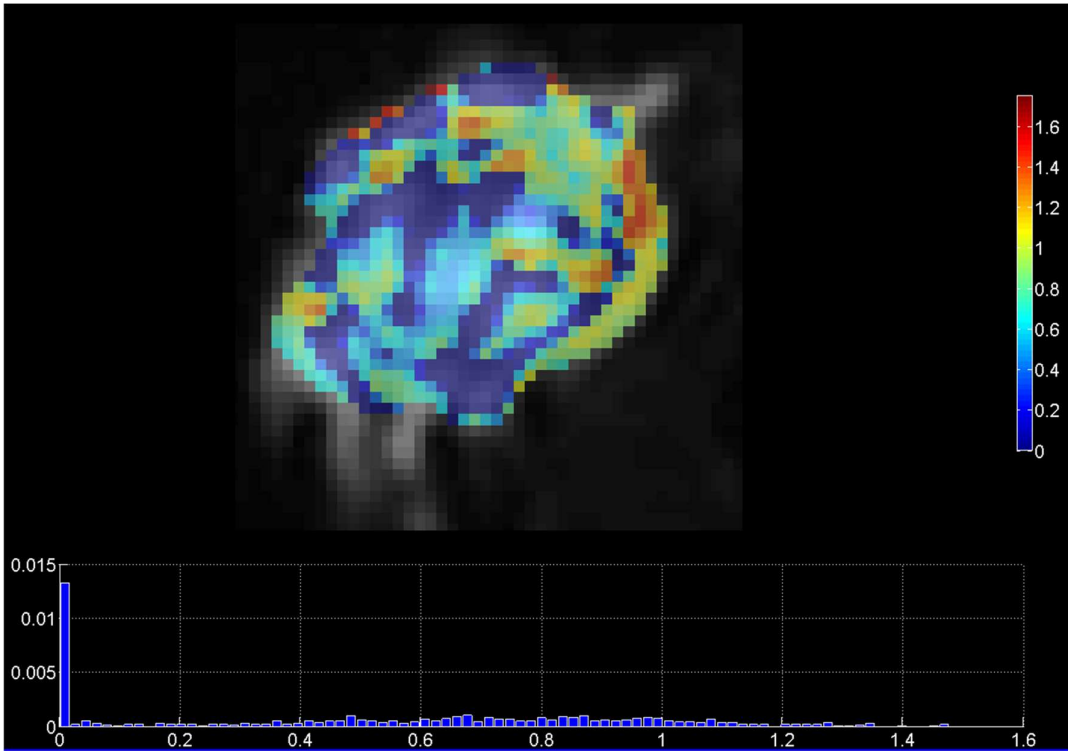
Figure 35 A patient with pleomorphic undifferentiated sarcoma. Apparent kurtosis (K_{app}) map and related histogram illustrate areas of high tissue complexity (high K_{app} values).
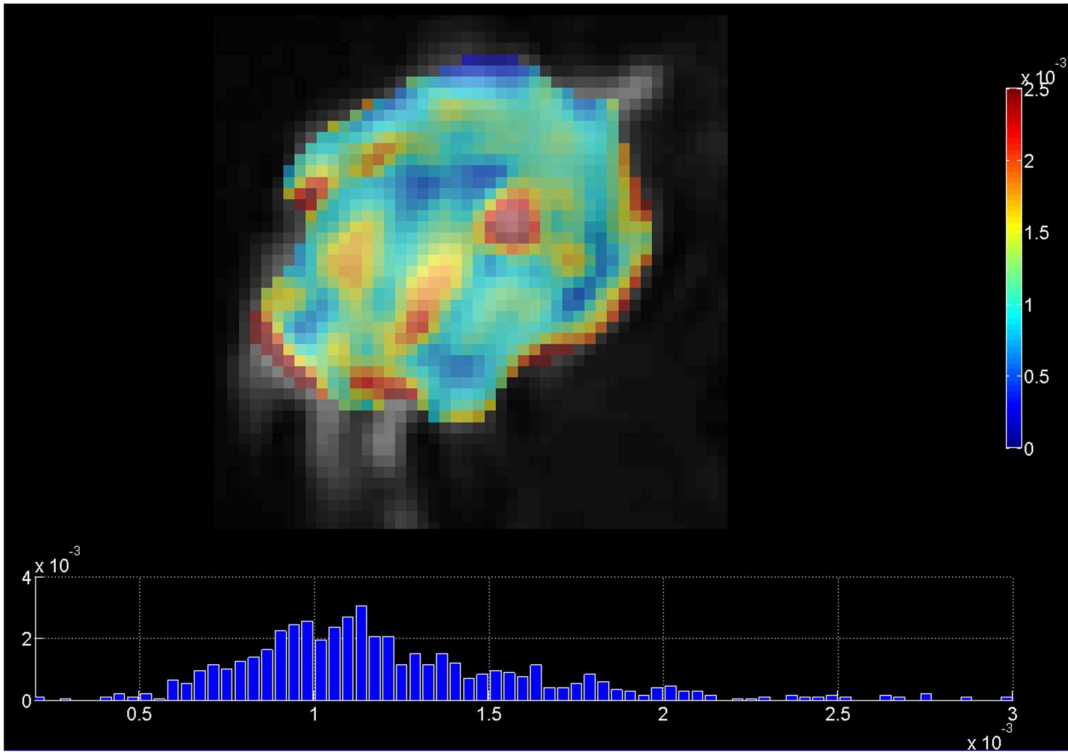
Figure 36 Apparent diffusion (D_{app}) map and corresponding histogram of a patient diagnosed with a pleomorphic undifferentiated sarcoma.

3.2 Fitting the diffusion signal attenuation

Several mathematical models have been proposed to quantify the DWI signal decay into diffusion biomarkers. In contradiction to the mono-exponential model and the single ADC biomarker calculated, a more complex mathematical framework is required for fitting the diffusion signal according to the IVIM and the non-Gaussian models. In case of the IVIM and the extended non-Gaussian IVIM model, two main categories are presented in the literature; a) complete fitting methods for calculating simultaneously all the biomarkers using nonlinear regression models and b) partial fitting methods that provide biomarkers in a more simplified way based on observations related to the behavior of the true-diffusion and micro-perfusion effects in the b-value range used.

3.2.1 Complete fitting

In order to extract multiple biomarkers from the IVIM and the non-Gaussian models, nonlinear least squares (NLLS) are widely used. The NLLS fitting technique is based on the Levenberg-Marquardt algorithm [34]. NLLS are minimization problems in mathematics that given initial, lower, and upper bounds for the estimated parameters (i.e. D_{slow}, D_{fast} and f in case of the IVIM model) approximate the diffusion model by a linear one and iteratively refine the values of the parameters to reach their optimal values. The initial point is of crucial significance for the convergence of the algorithm and a lot of attention has been paid for assessing the precision and uncertainty in the estimation of the diffusion biomarkers as reported in [35].

3.2.2 Partial fitting

Many studies in the literature use partial fitting for calculating IVIM related biomarkers. Every partial fitting method relies on the fact that, as stated in the IVIM theory, D_{fast} is roughly one order of magnitude greater than D_{slow} [36] at high b-values (b>200 s/mm^2) and therefore the micro-perfusion term in the IVIM model can be neglected. According to [37] the micro-perfusion effect in high b-values is eliminated and the IVIM bi-exponential fit equation is simplified to the following mono-exponential where D can be obtained linearly using least-squares regression.

$$\frac{S_b}{S_0} = exp(-b \times ADC)$$

The fitted curve from the mono-exponential model is then extrapolated at b=0 and the ratio between the x-intercept and the DW-MRI data at b=0 gives an estimation of the perfusion fraction f. Biomarkers D_{slow} and f, are then substituted into the IVIM bi-exponential equation and nonlinear least squares are applied to the entire b-value range for calculating D_{fast}. Alternatives can be also found in [38].

3.3 Evaluation and representation of the fitted results

This section focuses on the assessment of the fitting performance when a mathematical model is applied to the signal attenuation of the diffusion signal. This is an important step in the quantitative MRI in general since fitting quality highly influences the model parameter values calculation and therefore the representation of the tissue characteristics as depicted from each corresponding parameter. Another important aspect in evaluating fitting performance lies on comparative model selection studies where statistical metrics are calculated in order to find the most suitable model that best quantifies the examined diffusion signal. This was a major component of our analysis pipeline in differentiating STSs grading and is described thoroughly in the following chapters. Finally yet importantly, the potential ability of a model parameter to act as an indicator of a biological process and for monitoring the response to therapy can be severely influenced by the lack of reproducibility and repeatability. Although repeatability and reproducibility of the derived parameters was out of scope of this study, a thorough overview of these tests is reported in our of our publication reports [39].

3.3.1 Model fitting techniques

Assessing how well models fit the real diffusion signal in every b-value is a crucial step in the analysis. Without the knowledge of the trues values of the estimated biomarkers, an accurate fit of the diffusion data can significantly provide relative confidence in the results. Several statistical measures from the regression analysis can be used for evaluating models' goodness of fit to the diffusion data. R-square (R^2) is one of the most commonly used statistical measure for assessing the goodness of fit and is given by:

$$R^2 = 1 - \frac{SSE}{SST}$$

where SSE is the residual sum of squares and SST is the total sum of squares. R^2 values range from 0 to 1 with 1 indicating a perfect fit and 0 a total dissimilarity. Alternatively, adjusted R^2 can be used since it captures the number of the b-values used in the analysis (n), as well as, the number of the parameters provided by each model (K).

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2) \times (N - 1)}{N - K - 1} \right]$$

Another statistical measure rely on the Root Mean Squared Error (RMSE) that estimates the differences in signal intensity between the real and the modelled diffusion data in every b-value. However, as reported in the literature, such metrics mainly rely on the measurement of the absolute distance between the fitted curve and the acquired diffusion signal, thus tend to favour the most complex models [40]. Statistically, a complex model (i.e. BEM) will fit better the data than a simple model (i.e. MEM) causing overfitting and consequently false model selection in

many cases. For this reason, the aforementioned goodness-of-fit metrics mostly contribute to assess the proximity of the fitted curve to the experimental data for each model, and act as a thresholding criterion for excluding pixels that are noise and motion contaminated.

The core statistics in DWI studies where multiple models contribute to the diffusion signal fitting and a decision of the most suitable model for a particular signal needs to be taken, Akaike Information Criteria (AIC) [41], its corrected approach (AICc) [42], and the F-test statistics (F-ratio) [43] are suggested (Figure 37). The corrected AIC is a more appropriate metric than the standard AIC in studies where there is a relatively large number of estimated model parameters compared to the number of b-values. AICc is given by the following equation:

$$AICc = AIC + \frac{2 \times (K + 1) \times (K + 2)}{N - K - 2}$$

AIC is given according to:

$$AIC = N \times ln\left(\frac{RSS}{N}\right) + 2 \times (K + 1)$$

where RSS is the sum of squares of residuals and low values for AIC and AICc signify a good model fitting.
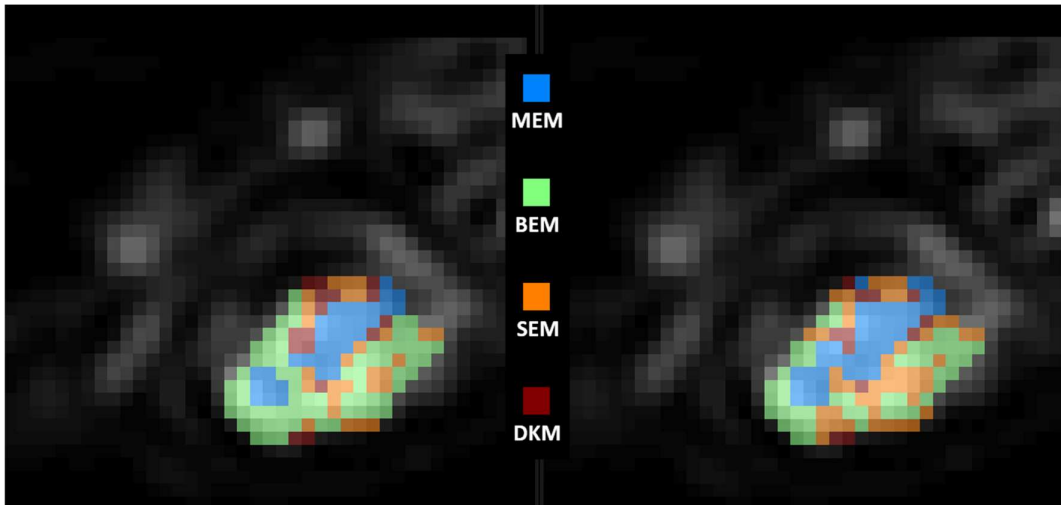
Figure 37 A pseudo-colorized classification map depicting model suitability of four examined models across pixels within a myxoid liposarcoma. Left: classification map based on the AIC criterion. Right: AICc-based model suitability map.

However, representing Akaike criterion scores with probability measures is reported to be more beneficial in model selection studies since it facilitates a direct interpretation of the relative merits of the investigated models [44]. To this end, Akaike Weights (AW) are calculated for each model in the examined set using AICc scores differences (ΔAICc) between all models and the model with the lowest AICc at a given pixel (i).

$$\Delta AICc_i = AICc_i - minAICc$$

AICc scores of all models are then transformed into conditional probabilities (varying from 0 to 1) using the following equation where R is the number of models in the examined set and the sum of AW being equal to 1. An analytical representation of the proposed model suitability procedure is well outside the range of this study and a thorough description can be found in [45].

$$AW_i = \frac{exp\left(-\frac{\Delta AICc_i}{2}\right)}{\sum_{r=1}^{R} exp\left(-\frac{\Delta AICc_r}{2}\right)}$$

The last metric for model selection relied on a hypothesis test using F-ratio with a 5% level of significance and is calculated based on the following equation:

$$F = \frac{(SSE_1 - SSE_2)/SSE_2}{(DF_1 - DF_2)/DF_2}$$

where DF is the degree of freedom given by the number of the b-values minus the number of model parameters, and subscripts 1 and 2 present the simpler and the more complex examined models respectively. F-ratio indicates a pairwise comparison between two candidate models for best fitting, choosing the more complex model (i.e. with subscript 2) in case its p-value is less than the one from the F-table with a 5% level of significance. Multiple pairwise comparisons are then conducted between the candidate models until the best model is determined. Wilcoxon signed-rank and Dunn's non-parametric statistical tests are used to disclose any significant differences between all examined models.

3.3.2 Qualitative and quantitative data presentation

Pixel based calculation of DW related parameters has the advantage of a detailed view over the area of interest compared to simple ROI techniques [46], but the size of data is sometimes overwhelming and difficult to handle. In order to gain an immediate insight into the complete information contained in the calculated indices, parametric maps and related histograms are produced (Figure 38). The quantification of specific histogram metrics may serve as the ground for statistical interpretation of the results and possibly indicate well-aimed biomarkers for certain pathology. Histogram metrics in some cases are able to discriminate two groups that simple ROI methods fail to [47]. Most histogram analyses use descriptive statistics such as the mean, standard deviation, quartiles, minimum, maximum, skewness and kurtosis to characterize and compare distributions of the diffusion biomarkers in examined ROIs in a quantitative manner.
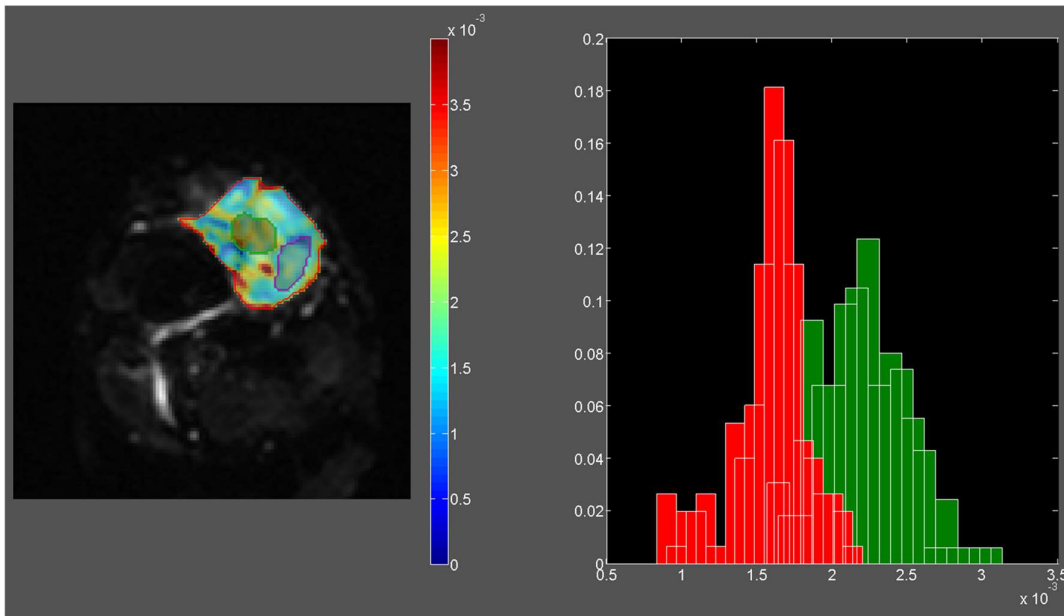
Figure 38 ADC histogram analysis performed on a pleomorphic liposarcoma. ROIs on an overlaid ADC map to the tumor DWI slice (left part). ROIs were drawn on two different tumor subregions reflecting areas of low (red annotation) and high cellularity (green annotation) as depicted in the histograms on the right part of the figure.

3.3 Implemented Software

Quantification of DWI requires a pipeline process comprising several steps of analysis after images are acquired from the MRI vendor. Within the scope of this thesis, a software platform was developed and published in [48] deriving a number of functionalities made available via a graphical user interface (GUI) for empowering users to read, analyze, visualize and quantify the DWI data into diffusion related imaging biomarkers. The main idea behind this tool was to assist clinicians with a simple and comprehensive processing tool for DWI preparation (e.g. tumor annotation) with no restrictions in imaging data and meta-data sharing across multiple software platforms. The presented software decomposed a highly complex analysis to a guided step-by-step procedure, meaning from data import to their quantitative interpretation. The tool is addressed to an experienced user who wants to structure a composite analysis with multiple facets, but is equally appealing to a clinical user who wants to benefit from a robust and simple workflow in order to employ with ease medical image analysis in DWI data. To achieve this goal, the demanding programming aspects and the complexities of the computational infrastructure in all steps of the analysis were covered by a pipeline process in which GUI buttons were linked to all analysis steps.

DMT was developed natively in MATLAB software platform, packaged as an executable file running under Windows and was structured in a simple five-step workflow underlined in the document below. To further support the pre-processing steps required for the radiomic analysis workflow (Chapter 4), an interconnection between the tool and in-house software developed for radiomics was established and meta-imaging files of any kind of MRI sequence, containing

the annotated ROIs, were directly transferred for radiomic feature extraction and ML analysis. The following subsection describes the main functionalities of the platform through a pipeline analysis mostly used in DWI quantification.

3.3.1 Region of Interest (ROI) delineation

Once DWI data are imported in the software a pop-up window permits interaction with the user and ROI delineations are performed either manually by contour drawing within the DWI slices or by importing the ROIs that are stored locally as meta-imaging files (i.e. NIFTI format). Manual drawing is assisted with tools for: a) adjusting the levels of the image, and b) navigation through the b-value acquisitions and the slices of the data to best determine the site of interest for analysis. Alternatively, ROI delineation can be performed under an automated process in which the platform imports the associated to the ROIs medical file and place the delineated contours within the slices (Figure 39).
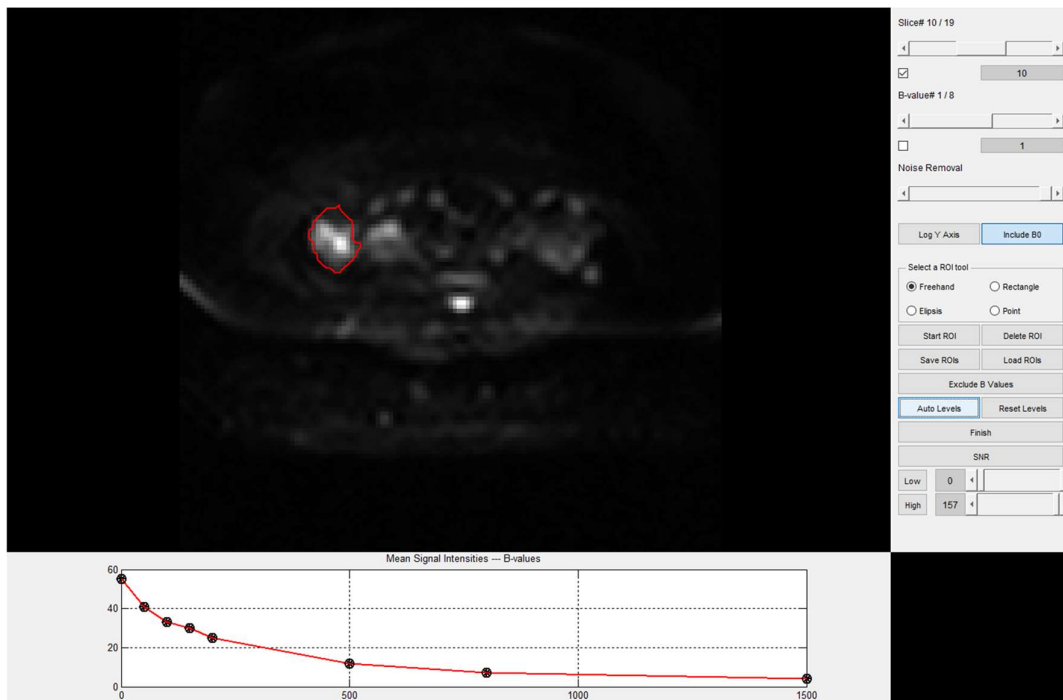
Figure 39 ROI delineation using freehand drawing contours. The mean exponentially decaying curve of the selected ROIs (red curve on bottom of the image) is displayed automatically as a function of the b-values. Settings are depicted on the right part of the window.

3.3.2 Image preprocessing and quality assessment

Medical images and particularly DWI are artifact prone data mainly influenced during the acquisition process. The attenuation in the diffusion data, especially when the b-value increases, yields to images with low Signal to Noise Ratio (SNR) and consequently to poor fitting success. DWI data acquired at high b-values provide insight into the tissue microstructure, thus excluding

such data from the analysis will hamper the obtained quantitative and structural information of the analyzed tissue. Spatially varying noise levels are inherent in the DWI data and advanced image denoising techniques are required for noise reduction. The platform is equipped with fully automated filters [49] that are applied iteratively to multiple 3D DWI data acquired at different b-values and reduce noise without affecting signal abundant areas. A common metric to assess the quality of the diffusion signal is given from the SNR of every pixel within the examine ROI, yielding to SNR parametric maps as depicted in Figure 40 and given by the following formula when images obtained with parallel imaging techniques are considered [50].

$$SNR = \frac{SI_{tumor}}{\sqrt{\frac{2}{4-\pi}} sd_{air}}$$

where SI_{tumor} was the signal intensity of each tumoral pixel and sd_{air} the standard deviation of a region of interest drown in the air near the anterior abdominal wall.
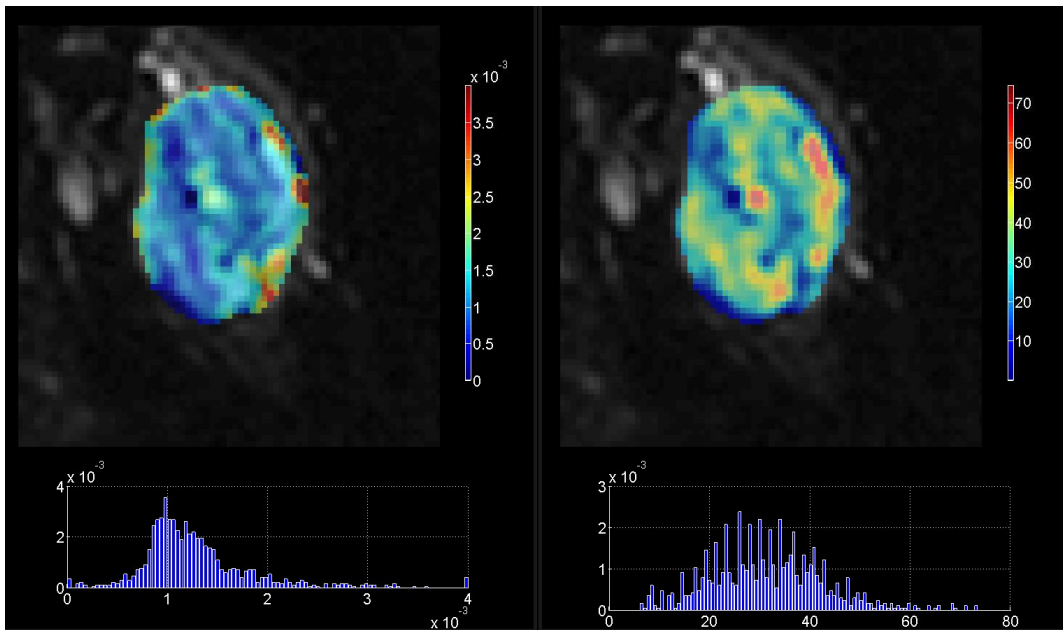
Figure 40 A pleomorphic liposarcoma case. Left: ADC map and associated histogram. Right: Calculated SNR map using the diffusion signal acquired at zero b-value

3.3.3. Quantification of the diffusion signal

This section describes the core DWI analysis, comprising the pixel-by-pixel quantification of the diffusion signal attenuation using the aforementioned models. Parameters of each model are calculated, using mathematical curve fitting methods, for further analysis and visualization. The platform uses fitting techniques relying on NLLS minimization problems in mathematics that given initial, lower, and upper bounds for the estimated parameters (i.e. ADC) approximate the diffusion model by a linear one and iteratively refine the values of the parameters to reach their

optimal values. The platform supports statistical analysis to determine the fitting accuracy of each model on the diffusion data using the statistical metrics outlined in Section 3.3.1.

3.3.4 Visualization of the results

DWI analysis visualization is a significant step in the pipeline process as it gives a through qualitative representation of the results. To this end, within the platform, the visualization screen can show the anatomy of the examined tissue with an overlay of the parametric map of the selected parameter, the SNR map and the goodness-of-fit metric. The corresponding histogram can be shown in the lower part of the image and the user is able to load previous analyses of the same patient or of another person. The delineation of a smaller ROI inside the initial is allowed in order to focus on a specific area, which may be presented with specific characteristics and attract clinical interest. For that specific ROI, the signal drop is plotted both on normal and logarithmic scale and the mean value of all related metrics is shown (Figure 41). Once the analysis is over, the user can export the calculated values of each parametric map or a set of statistical metrics (minimum, maximum, mean, standard deviation, median, skewness, kurtosis, variance, 5%, 30%, 70% and 95% percentiles) from the corresponding histogram in a format of his choice (DICOM, text or a csv file). Lastly, the user can recall results from previous analyses and perform comparison between different studies in a single screen frame. This way, the platform can be a tool for longitudinal evaluation of disease progression or therapeutic outcome for a single patient, or may facilitate the clinician to perform group studies of specific patient cohorts (Figure 42).
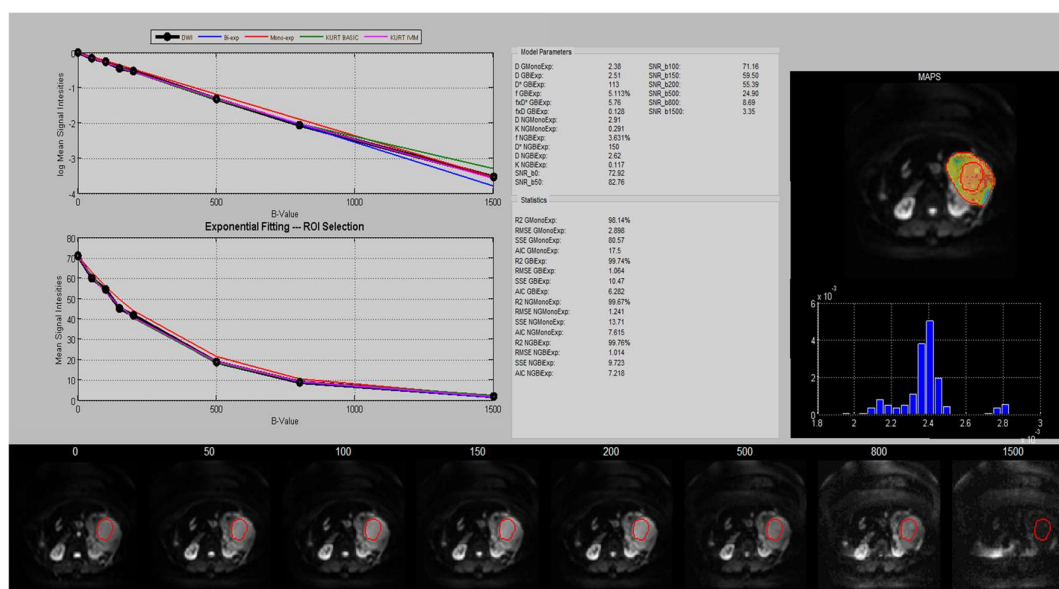
Figure 41 Fitting accuracy represented in normal and logarithmic scale for all selected models and quantitative assessment of their parameters and the relevant statistical metrics. The results may stem from the initial ROI, a selected subpart of it or a single pixel within it.
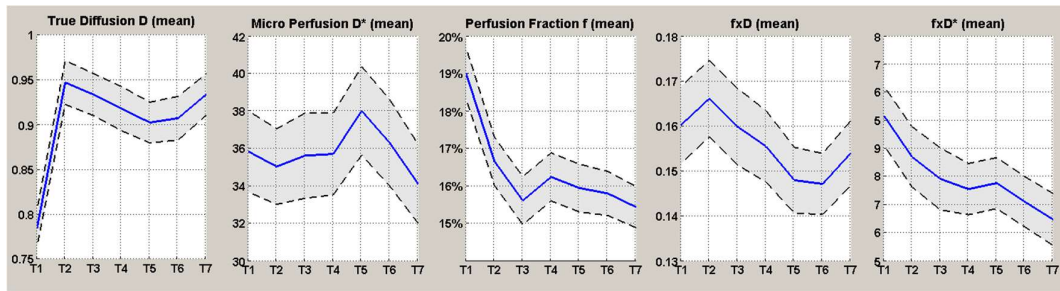
Figure 42 Longitudinal mean value changes of the IVIM parameters with 95% confidence intervals, calculated at seven distinct time points (T1 to T7). Image taken from one of our e-posters published in the European Society of Radiology (ECR) https://dx.doi.org/10.1594/ecr2017/C-2835.

Chapter 4 Radiomics

4.1 Introduction

Radiomics has recently gained a lot of attention in medical imaging as a promising technique to extract high-throughput quantitative features from complex patterns of the diagnostic images (e.g. MRI) that can rarely be seen by the human eye [51]. Radiomics enables digital decoding of images into quantitative features using histogram analysis, shape characteristics and image texture analysis which is a standard image processing technique to define local variations in the signal intensity of the image, identify image patterns and quantify image characteristics into high-throughput vectors of hand-crafted quantitative features. Main radiomics objective is to shift from visual interpretation of medical images to the prediction of a clinical outcome when machine learning is embedded to the radiomic analysis, allowing a more precise diagnosis, therapy planning and disease monitoring.

4.2 Analysis pipeline

A radiomic analysis pipeline comprises several steps including the identification of the clinical question, data collection, image acquisition protocol optimization and standardization, image pre-processing, delineation of the examined tissue, radiomic feature extraction, feature selection/reduction and machine learning model development and validation (Figure 43). The clinical question should be precise and clear enough in order to address an unmet clinical need in cancer management. Data collection issues including sample size, data quality and diversity are important factors for accurate, reproducible and repeatable radiomic results. In general, machine learning algorithms are "data hungry" approaches. However, a specific number regarding the sample size required for model development cannot be easily defined since this depends on the complexity of the model, the validation approach and the amount of information that can be extracted from the images (i.e. high-dimensional data). As a rule of thumb, an 1:10 ratio between the imaging features and the samples is recommended for radiomics-based machine learning models to avoid overfitting. Data homogeneity in the

examined cohort is suggested in radiomics since it decreases variability on both data and analysis results. Nonetheless, data heterogeneity and the design of multicenter radiomic studies (train with data from one center and validate externally with data from other institutes) is inevitable to meet real-life conditions and increase generalization performance of the models. Another significant issue in imaging studies is the development of standardized and optimized acquisition protocols. Although data heterogeneity is recommended as it potentially increases the generalizability of the developed models, differences in the acquisition protocol and vendor parameters yield spatial and contrast differences in the acquired images, further affect the calculated radiomic features. Therefore, standardization of image acquisition and of the radiomic features is an important step to maintain radiomics integrity.
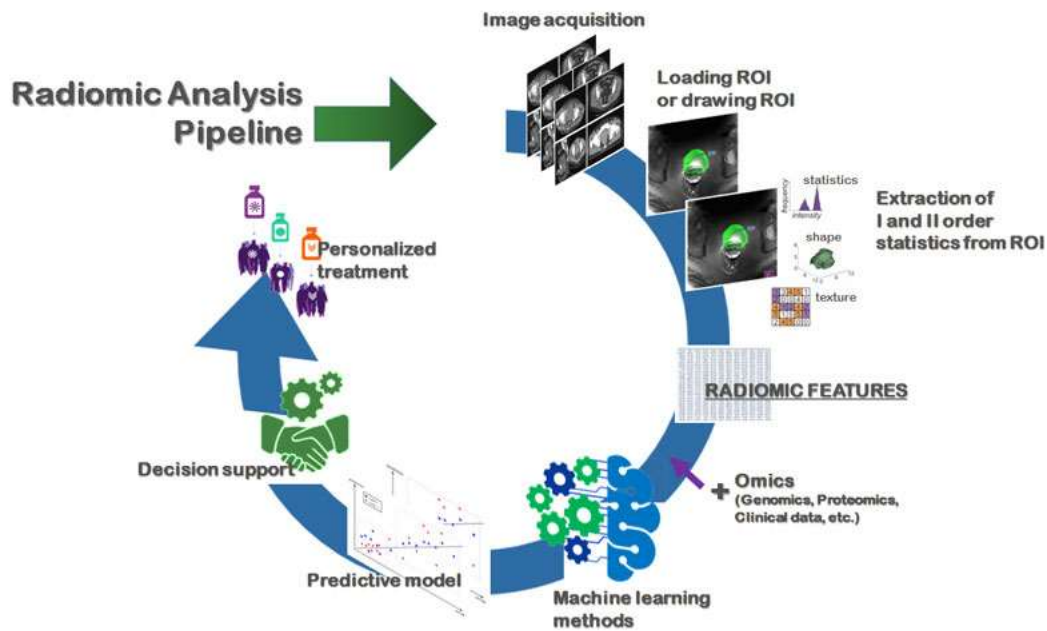
Figure 43 Radiomic analysis pipeline supporting precision medicine decision making. Image taken from the European Research Consortium for Informatics and Mathematics (ERCIM) news.

Tumor delineation is a crucial part in the radiomic studies that can be performed either on a two-dimensional basis (i.e., on a 2D MRI slice) or on the entire tumor volume (three-dimensional using multiple MRI slices). Regions of interest (ROIs) can be drawn manually, semi- or fully automatically using a variety of existed software solutions (e.g., deep learning models for tumor segmentation). Pre-processing of the examined images is also an urgent step in the analysis pipeline and generally includes pixels outlier filtering, noise reduction or elimination, and discretization of the single intensity of all pixels within the examined ROIs. A thorough presentation of the radiomic analysis pipeline can be found in [52].

4.2.1 Radiomic feature extraction

Radiomic feature extraction involves the calculation of high-dimensional handcrafted imaging features once specific regions of the examined tissue are identified and segmented. These features mainly include descriptors of the signal intensity of an image, size and shape-based characteristics, features that identify potential relationships between the pixels of an image, and texture features calculated using image transformation methods (e.g. wavelet transform, Gabor filters, etc.). A broad categorization classifies the radiomic features into morphological, first and higher-order texture features. Morphological features are related to the 2-dimensional and 3-dimensional size and shape of the examined area given by a single ROI (e.g. from an MRI slice) or several ROIs across multiple ROIs across all axis (e.g. 3D tumor volume from multiple MRI slices). They assess the geometrical characteristics of the annotated area and, among others, they include surface to volume ratio, sphericity, maximum 2D and 3D diameter, flatness, perimeter, compactness, etc. First-order texture features are histogram-based features that describe the distribution of the signal intensity of the pixels within the examined ROIs. Among others, these include the mean, median, entropy, uniformity, interquartile range, variance, skewness (asymmetry) and kurtosis (flatness) of the histogram of values. Higher-order texture features (see following section) are widely used in radiomics and describe spatial relationships and statistical correlations between the pixels (inter-pixel relationships) within the ROI(s). Multi-resolution transformations of the image using wavelets and Gabor filtering which is a multi-scale technique able to quantitatively characterize texture at different scales and orientations are transform-based techniques that are frequently used in radiomics in order to expand the number of the acquired features and obtain more textural information across the different scales of the image. The underlying concept behind multi-scale/resolution texture techniques is that important information about the image structure is contained at a number of different scales and not only at one scale. Therefore, scale can be considered as a continuous variable rather than a simple parameter, and features may be extracted at multiple scales, therefore allowing for a multi-scale representation. A comprehensive radiomic analysis using Gabor filtering to predict breast cancer therapy response can be found in one of our publications [53].
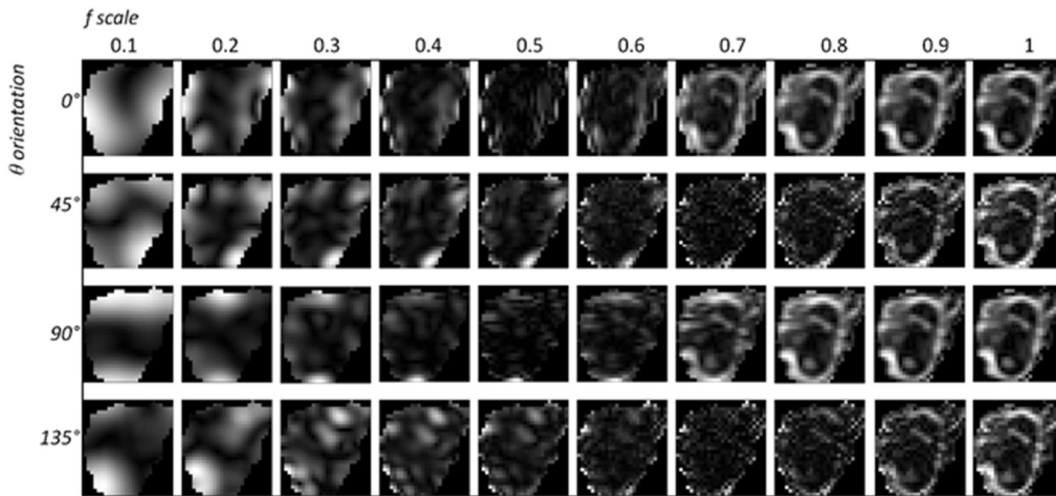
Figure 44 . Gabor scale-orientation filtering illustrated for orientations (ϑ)={0^0, 45^0, 90^0, 135^0} and for scales {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}. After applying the Gabor filter bank to the selected ROI of the image the filter responses lead to 40 different scale-orientation image representations.

4.2.1.1 Higher-order texture features

Grayscale co-occurrence matrix (GLCM)

Grayscale co-occurrence matrix (GLCM) has proved to be a robust method for extracting texture features from images. Haralick has defined fourteen features from GLCM, for the extraction of the image texture characteristics. Considering a GLCM of size $N_g \times N_g$ and the description of second-order joint probability function of the image is defined as $P(i, j \mid \delta, \theta)$ where *i,j* represents the times of a certain combination that occur in the image which are separated by a δ distance of pixels with angle ϑ. Indicative texture analysis results on whole MRI slice of a high and low-grade STS are illustrated in Figure 45 and 46, respectively.
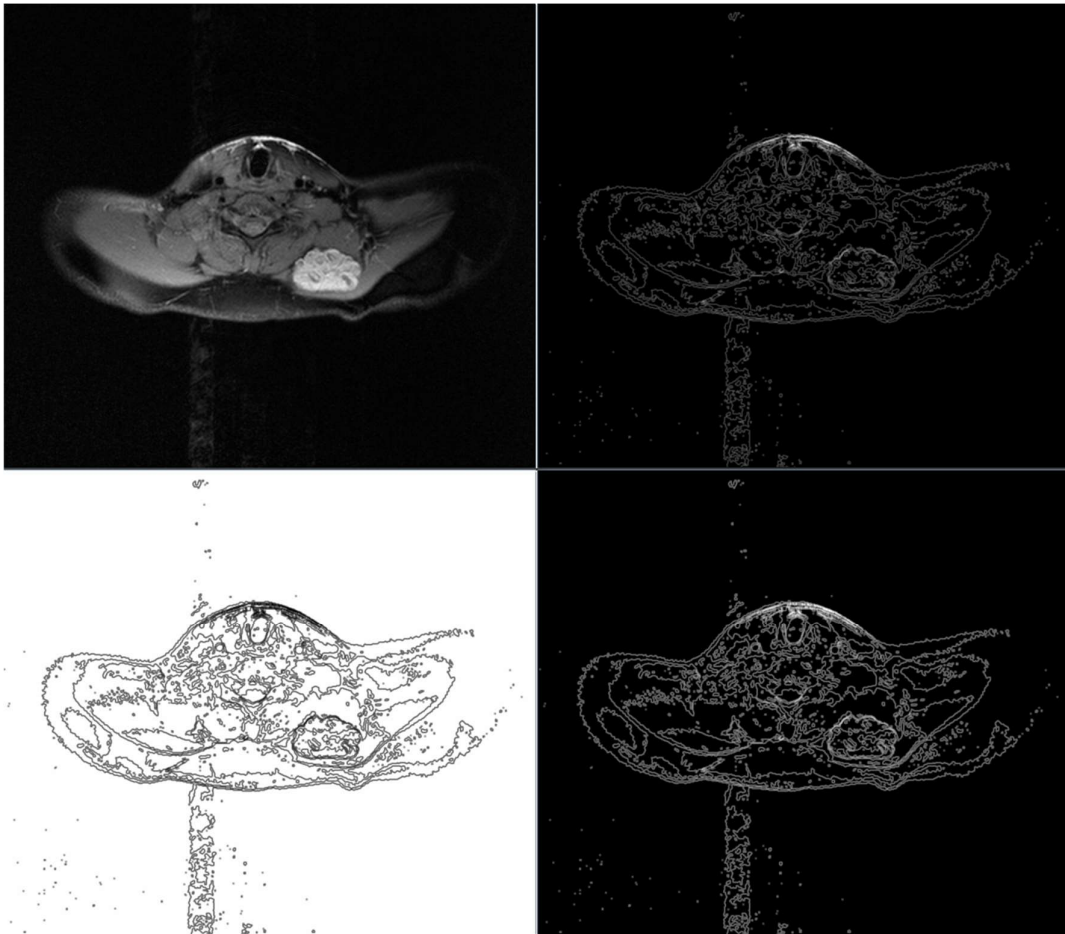
Figure 45 Whole image texture analysis on a high-grade STS (alveolar soft part sarcoma). Top left: the original T2w MRI. Top right: Calculated image of the difference average. Bottom left: Calculated energy. Bottom right: Entropy of the MR image.
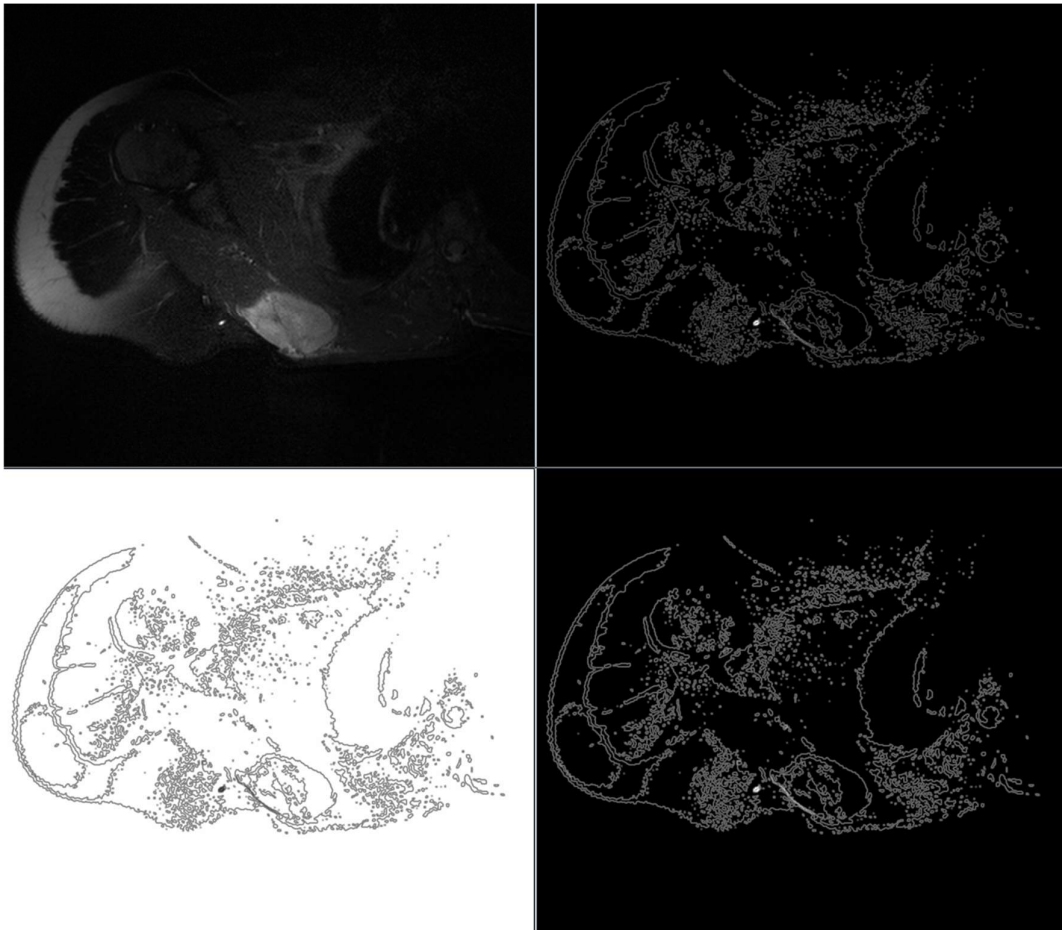
Figure 46 Whole image texture analysis on a low-grade STS (myxoid tumor). Top left: the original T2w MRI. Top right: Calculated image of the difference average. Bottom left: Calculated energy. Bottom right: Entropy of the MR image.

Gray-Level Size-Zone Matrix (GLSZM)

An additional statistical texture descriptor is the Gray-Level Size-Zone Matrix. Similarly to GLCM, GLSZM calculates the pixel intensities of an image. GLSZM considers the relationship between same pixel intensities and areas. While GLCM uses one direction, GLSZM is calculated in all directions. To calculate a GLSZM one has to measure the probabilities of different sized voxels with certain intensities. Considering the following matrix of pixel intensities of an image, the calculation of GLSZM requires to count the size of the relationship between same intensity pixels. The connection which defines a relationship is same intensity value with a pixel offset regardless the direction. A simpler approach/explanation would be to "follow" a single intensity of a pixel and count how many times it occurred in our path. A simple visualization is shown below.

Figure 47 Example of filling the GLSZM for an image of size 4x4 including 4 gray levels.

GLSZM is particularly effective in characterizing texture homogeneity.

Gray Level Run Length Matrix (GLRLM)

A Gray Level Run Length Matrix (GLRLM) quantifies gray level runs of consecutive pixels with the same gray level value, which are defined as the number of pixels. In a GLRLM the y axis represents the number of runs of the gray levels and the x axis the occurrences in the image. GLRLM is calculating for a single angle, however, one can calculate for several angles depending on the application. As with the former matrices, more angles can provide better textural information. One can deduct several textural features from a GLRLM. Considering N_p is the number of voxels in the image, N_r the number of discreet run lengths, $N_r(\theta)$ the number of run along angle θ, N_g the number of discreet intensity values and $p(i, j|\theta)$ the run length matrix along an angle.

4.2.2 Radiomic feature selection/reduction

Since radiomics provides high-throughput imaging features to quantify non-invasively tumor phenotype, feature selection has been proved to be a major step in the radiomic analysis pipeline as it contributes significantly in the identification of relevant, non-redundant and stable radiomic features, yielding most likely to increased machine learning models' performance, decreased time required for model training and reduction of model overfitting [54]. Most of the machine learning algorithms used in radiomics are sensitive to the so-called curse of dimensionality issues, thus an appropriate balance between the number of the examined images and the radiomic features exported from these images is favorable to avoid a sparsely populated examined feature space that makes the machine learning process to become more difficult. To this end, several techniques have been developed to find the optimal feature subset from a large pool of radiomic features according to a quantitative evaluation metric and are broadly classified into four main categories: a) filter methods, b) wrapper methods, c) embedded methods and d) hybrid methods [55].

Filter methods are simple, effective and computationally inexpensive feature selection techniques that are independent of any employed model and applied prior to any machine learning analysis. Within filtering feature selection, features are either ranked in a univariate way (single and independent features) or selected as feature subsets (i.e. multivariate) based

primarily on statistical measures. In most of the univariate filtering methods, once features are selected and ranked, a non-automated process needs to be applied in order to select the optimal number of selected features further used as input to the machine learning models. Among other filter selection techniques, correlation-based, statistical, and ranking filter methods are widely used in radiomics. Correlation between radiomic features is examined in a pairwise fashion where a linear relationship between two features is defined and quantified. Pearson and spearman's rank correlation coefficients are fast and easy to implement techniques to examine whether two features are highly correlated among themselves and therefore keep one of the redundant features for further processing. Statistical and ranking filter selection is mainly relied on mutual information, information gain and pure statistical tests (i.e. ANOVA, chi-square and wilcoxon rank sum test).

Unlike the filter techniques, wrapping methods use a certain subset of features and a machine learning model to assess their predictive performance. This is a greedy and computational expensive process since various feature subsets are generated and examined in terms of their predictive performance until the optimal subset (i.e. best possible combination of features) is found. Stopping criteria in searching the optimal subset needs to be defined by the machine learning model prior to wrapping (e.g. select the feature subset one step before the model performance starts to decrease) and selection of the feature subsets is made by forward feature selection, backward feature elimination and recursive feature elimination. Forward selection typically starts with no features and subsequently adds one or more features to the set at a next step. In backward elimination, the entire feature set initially participates in the modelling phase where one or more least significant features are iteratively removed from the examined set till there is no improvement observed in the model performance. Recursive feature elimination is a greedy optimization algorithm which aims to find the best performing feature subset by creating models repeatedly and storing features from each iteration that yields the best or the worst model performance. At the end of the process, until all features are examined exhaustively, features are ranked according to their elimination order. Wrapper techniques are more computationally demanding than filtering, especially in radiomic studies where data are highly dimensional. Additionally, wrapper techniques are machine learning model dependent and thus biased towards the model chosen to assess their significance in the prediction. Last, embedded feature selection combines both filtering and wrapper techniques and is embedded into the design of the machine learning model in order to reduce the computational time required for identifying and selecting the most appropriate subset of features that yields to the best performance.

4.2.3 Machine learning model development and validation

Radiomics aims to construct a machine learning model able to predict specific clinical questions using a set of redundant and stable radiomic features after feature selection. According to the different levels of the available clinical information and whether the clinical result is based on a discrete (e.g. healthy vs malignant) or a continuous (e.g. time of relapse) outcome, several machine learning models can be deployed, typically grouped into supervised, semi-supervised, and unsupervised learning techniques. Unsupervised techniques (e.g. clustering) explore whether potential associations and groups exist between the radiomic data without knowing a priory any clinical information about the addressed question (i.e. unlabeled data). In semi-supervised learning, models are developed using both unlabeled and labeled radiomic data from a specific patient cohort. Supervised learning typically includes two steps of analysis: a) the training and b) the testing phase. In the training phase, a large labeled dataset, in terms of the recruited patient cohort, is used for building and fine-tuning the model while a smaller dataset (ideally from a different data provider/institution) validates the model performance. Supervised learning models that aim to predict a continuous clinical outcome are called regression techniques whereas models that predict discrete outcomes (i.e. categorical outcome) are known as classifiers. Widely used regression techniques include linear, logistic, LASSO-logistic and elastic-net regression. Classifiers, according to their architecture design, can be further grouped into tree-, distance- and rule-based machine learning models. Kernel-based classification approaches including the support vector machines (SVMs) and their alternatives are common distance-based techniques in radiomic analysis with generally a good predictive performance. Rule-based classifiers are developed using various if-else rules coming from the data used for training whereas tree-based techniques use a series of yes/no questions in order to determine the predictive outcome. Decision trees are the most common tree-like classifiers comprising a series of sequential decisions whereas random forests classifiers are developed using multiple combinations of decision trees to construct the final decision.

In recent years, the ensemble classifier techniques are rapidly growing and enjoying a lot of attention in radiomic studies due to their potential to greatly increase prediction accuracy of a learning system. These techniques generally work by means of firstly generating an ensemble of base classifiers via applying a given base learning algorithm to different permutated training sets, and then the outputs from each ensemble member are combined in a suitable way to create the prediction of the ensemble classifier. The combination is often performed by voting for the most popular class and examples of these techniques, among others, include bagging, adaboost and random forests.

Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)   Deep Learning
Stacked Auto-Encoders
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)   Ensemble
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Radial Basis Function Network (RBFN)
Perceptron   Neural Networks
Back–Propagation
Hopfield Network
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)   Regularization
Elastic Net
Least Angle Regression (LARS)
Cubist
One Rule (OneR)   Rule System
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression   Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression

Machine Learning Algorithms

Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)   Bayesian
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)   Decision Tree
Decision Stump
Conditional Decision Trees
MS
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis   Dimensionality Reduction
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)   Instance Based
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
k–Means
k–Medians   Clustering
Expectation Maximization
Hierarchical Clustering

Figure 48 Common machine learning model categories. Image from [56].

Performance metrics

A crucial term for evaluation of the machine learning models is the classification error. However, in many applications distinctions among different types of errors turn out to be important. In order to distinguish among error types, a confusion matrix (see following Table) can be used to lay out the different errors. In case of a binary classification problem, a classifier predicts the occurrence (Class Positive) or non-occurrence (Class Negative) of a single event or hypothesis.

Table 2 Confusion matrix for a binary classification problem.

	True Class	
Predicted Class	Class Positive	Class Negative
Prediction Positive	True Positives (tp)	False Positives (fp)
Prediction Negative	False Negatives (fn)	True Negatives (tn)

Common metrics for evaluation of the classification performance, calculated from the confusion matrix, are given below. Apart from the accuracy score, which is calculated as the ratio of correctly classified samples by the total number of samples, showing the overall accuracy of the model, other metrics of classification performance are also used.

$$ACC = \frac{tp + tn}{tp + tn + fp + fn}$$

Precision (positive predictive value) can be interpreted as the ability of the classifier not to label as positive a sample that is negative and is calculated as:

$$PPV = \frac{tp}{tp + fp}$$

The sensitivity or Recall (true positive rate) of a test is its ability to determine the patient cases correctly. This is also obvious through its calculation:

$$TPR = \frac{tp}{tp + fn}$$

Specificity (true positive rate) represents the ability of a method to determine the negative (patient) cases correctly. In simpler terms, a higher value of specificity will allow for less false positives, or healthy subjects classified as patients.

$$TNR = \frac{tn}{tn + fp}$$

The F1 score metric can be thought of as a combination of Recall and Precision as it is calculated using these two metrics:

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} = \frac{2tp}{tp + fp + fn}$$

Sensitivity, specificity and accuracy describe the true performance with clarity, but failed to provide a compound measure for the classification performance. This measure is given through Receiving Operating Characteristic (ROC) analysis. For a two-class classification problem ROC curve is a graphical plot of the sensitivity vs. 1-specificity as the discrimination threshold of the classifier is varied (Figure 49).
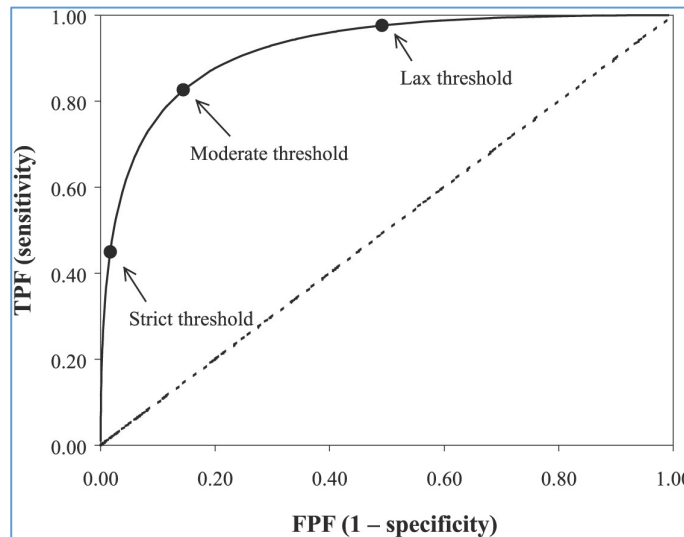
Figure 49 A typical ROC curve, showing three possible operating thresholds.

While the ROC curve contains most of the information about the accuracy of a classifier through several values of thresholds, it is sometimes desirable to produce quantitative summary measures of the ROC curve. The most commonly used quantitative measure is the area under the ROC curve (AUC). AUC is a portion of the area of the unit square, ranging between 0 and 1, and is equivalent to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

Validation

In a machine learning analysis pipeline, a typical task is not only to learn a model for the available data but also to learn a model with good generalization in its predictive performance. Such a model may demonstrate adequate prediction capability on the training data and on future unseen data (e.g. using an independent group of patients for validating externally model performance). Cross validation is a procedure for estimating the generalization performance in this context in a way to protect the classification model against over-fitting. No matter how sophisticated and powerful algorithms for classification are developed, if no reliable performance estimates are obtained, no reliable decisions can be made based on model results. Basic forms in cross-validation are the k-fold and the leave-one-out cross-validation. In k-fold cross-validation the data is first partitioned into k equally (or nearly equally) sized folds. Subsequently k iterations of training and validation are performed such that, within iterations, a different fold of the data is held-out for validation while the remaining k-1 folds are used for learning. If k equals the sample size, this is called the leave-one-out. In case of stratified k-fold cross-validation, the data are stratified prior to being split into k folds in order to ensure that each fold is a good representative of the whole. Nested cross-validation techniques are used both for the evaluation of the model's performance in various stages, to select the best model from a large grid of models as well as to optimize the model parameters. In this case, feature selection, model selection and estimator hyper-parameter tuning and testing are all contained within the bounds of the iterations of a single CV method. The generalization performance of the proposed radiomic analysis reported in Chapter 6 was relied on the nested cross-validation depicted in the following figure to eliminate random or cherry-picked train-test combinations and provide a robust subgroup of radiomic features after feature selection.
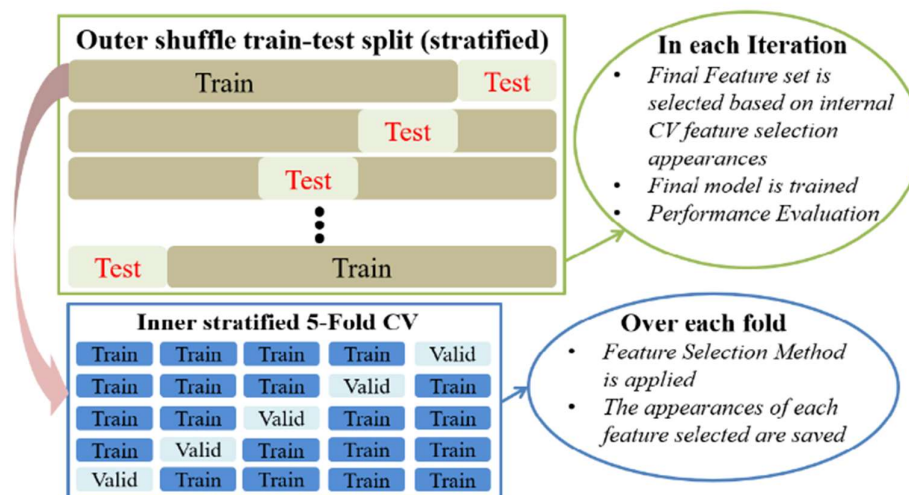
Figure 50 The proposed validation schema.

Chapter 5 Application of the DWI quantification into the differentiation of STSs grading

5.1 Scope of the study

The presented study reports the core analysis results of the thesis aiming to investigate and histopathologically validate the role of model selection statistical techniques in the design of novel parametric meta-maps towards the discrimination of low- from high-grade soft tissue sarcomas (STSs) using multiple Diffusion Weighted Imaging (DWI) models. The histologic analysis on several tumor subregions confirmed model suitability results on these areas. Among all examined histogram metrics, only 3 metrics all derived from the meta-maps were found to be statistically significant in differentiating low- from high-grade STSs with an AUC higher than 89%. DWI data of 28 patients with soft tissue malignancy were quantified using the mono-exponential, bi-exponential, stretched-exponential and the diffusion kurtosis model outlined in Chapter 3. Akaike Weights (AW) were calculated from the corrected Akaike Information Criteria (AICc) to select the most suitable model for every pixel within the tumor volume (details can be found in Chapter 3). All parametric maps were calculated using the aforementioned software and pseudo-colorized composite diffusion model (CDM) maps were then generated to depict model suitability, hypothesizing that every single model underpins different tissue properties of cellularity, neovascularization, complexity and heterogeneity and cannot solely characterize the whole tumor. Single model parametric maps were turned into meta-maps using the CDM map and a histological validation of the model suitability results was conducted on several subregions of different tumors. Histogram analysis on all maps and meta-maps was performed, Mann-Whitney U test was conducted, p-values were adjusted for multiple comparisons and performance of all statistically significant metrics was evaluated using the Receiver Operator Characteristic (ROC) analysis.

5.2 Material and Methods

5.2.1 Patient population

This study was approved by the local's ethics committee. From July 2015 to February 2019, 31 patients with suspicion of a malignant soft tissue tumor underwent preoperative MRI examination. Data were anonymized from the MRI vendor and within 14 days from the acquisition, immediately after surgical excision, specimens were transferred to the pathology laboratory for histopathologic evaluation and grading. Three patients were excluded due to therapeutic intervention prior to surgery and 28 eligible patients (figure below) were grouped according to the FNCLCC Histologic Grade (except for alveolar soft part sarcoma).

Low-grade STSs	High-grade STSs
well differentiated liposarcoma (4 patients)	dedifferentiated liposarcoma (4 patients)
myxoid liposarcoma (5 patients)	pleomorphic liposarcoma (3 patients)
	Ewing sarcoma (1 patient)
	leiomyosarcoma (3 patients)
	spindle cell myxoid liposarcoma (3 patients)
	undifferentiated pleomorphic sarcoma (3 patients)
	malignant peripheral nerve sheath tumor (1 patient)
	alveolar soft part sarcoma (1 patient)

Figure 51 The examined cohort.

A parallel preliminary study was also conducted using two additional patients diagnosed with undifferentiated pleomorphic sarcoma (UPS) from an ultrasound-guided core needle biopsy and spindle cell myxoid liposarcoma (SCML) when an open biopsy was taken from the dorsal site of the tumor. A drug combination of tumor necrosis factor antagonist (TNF-a) and melphalan with isolated limb perfusion (ILP) was administered as it has shown great anti-neovascularization effects in STSs [57] and both patients underwent a baseline pre-therapy and a secondary MRI examination 2 months after the end of the treatment. Patient diagnosed with UPS also received a surgical excision of the mass after chemotherapy, while only chemotherapy was considered adequate for the patient with SCML. The goal from this study was to investigate whether vascularization, cellularity and tissue morphology changes induced by ILP with chemotherapy and TNF-a could be assessed and monitored non-invasively from pixel level variations of DWI model suitability within the tumor.

5.2.2 Histopathologic Analysis

Each fresh surgical specimen was oriented according to its anatomical axial plane based on suture markings precisely placed by the surgeon, thoroughly inked at its outer surface with permanent ink. Tumor distances were measured by a trained pathologist which consequently sliced tumors in parallel consecutive thin slices 2-3 cm thick from its upper to its lower limit, hence the preformed cuttings following the true axial tumor plane. The axial diameter of the tumor mass was measured to identify the central tumor slice which subsequently was divided in a grid-manner in orthogonal slabs each of which was placed in a labelled cassette. Standard procedures for fixation, dehydration clearing and waxing were followed. 4μm thick sections of

each of the FFPE (Formalin-fixed paraffin-embedded) tumor slab was placed into glass slides, stained for H/E and examined microscopically (Nikon Eclipse E200) to characterize areas according to cell type, tumor differentiation, cellular atypia, cellularity, mitotic activity, vascularity and presence of necrosis.

5.2.3 MR examination protocol

All imaging data were acquired at a 1.5T scanner (Vision/Sonata hybrid System, Siemens, Erlangen, Germany, Gradient Strength: 45mT m^{-1}, Slew Rate: 200mT m^{-1} s^{-1}) in the local university hospital. The protocol consisted of dual PD to T2 echo (TE1/TE2/TR : 13/80/ 3250ms, NEX:1) weighted sequence in axial and coronal planes with slice thickness 4mm (20% interslice gap) to ensure complete lesion coverage and T1 TSE (TE/TR: 13/498ms) sequences which was repeated after contrast medium administration. The number of slices and field of view were adjusted among patients to cover the whole extent of the lesion ranging from 12-20 slices and 00x200mm (frequency phase matrix: 320x289) or 400x400mm (384x320) respectively. Also, the coil differed among lesions, depending on the anatomy of interest and the optimal signal to noise ratio (SNR). Spectral fat suppression was used to create fluid sensitive images with increased lesion conspicuity. Acquisition time was 2min and 49s, regardless the variance in the number of slices. The EPI DWI acquisition comprised 8 different b-values (0, 50, 100,150, 200, 500, 800, 1500 s/mm²) with TE/TR: 100/2900ms and slice positions / FOV / interslice gap were identical to the dual echo acquisition. The matrix size was 128x128 or 92x92 for FOV of 400x400mm or 200x200mm respectively.

5.2.4 Image processing

For each patient, regions of interest (ROIs) on the outer edge of the lesion were traced manually slice by slice on the diffusion images acquired at zero b-value with anatomical reference to the co-registered PD and T2 dual echo images. Whole tumor volumes were determined from the corresponding 2D slices by an expert with 12 years of experience, reviewed by a senior radiologist with 34 years of experience in musculoskeletal MRI (both were blinded to the histologic results) and copied to the diffusion images at the same slice location. Disagreement between the two readers was resolved by consensus reading and refined annotations were provided. Whole tumor volumes were digitally transferred to a personal computer for post-processing using the developed software from Chapter 3. Diffusion signals within ROIs were analyzed on a pixel-by-pixel basis and several parametric maps (exemplary images are shown in Figure 52) were calculated using the mathematical models presented in Chapter 3. For the reader's convenience the equation of the four examined models are also given below:

1) The mono-exponential model (MEM):

$$\frac{S_b}{S_0} = exp(-b \times ADC)$$

where S_b and S_0 denote the acquired diffusion signal for a particular degree of diffusion weighting b (s/mm^2) and the signal without diffusion sensitization, respectively. ADC is the apparent diffusion coefficient (mm^2/s), associated to tissue cellularity.

2) The bi-exponential model (BEM):

$$\frac{S_b}{S_0} = f \times exp(-b \times D_{fast}) + (1-f) \times exp(-b \times D_{slow})$$

where three parameters including the fast component of diffusion (D_{fast}, in mm^2/s) related to incoherent microcirculation, slow diffusion coefficient (D_{slow}, in mm^2/s) reflecting pure molecular diffusivity, and the unitless fraction of fast component (f) were calculated using the above, extended to the MEM, form.

3) The stretched-exponential model (SEM):

$$\frac{S_b}{S_0} = exp(-(b \times DDC)^a)$$

where index α, varying from 0 to 1, corresponds to water molecular diffusion heterogeneity (unitless), and DDC is the distributed diffusion coefficient (mm^2/s) which is equivalent to the ADC when index a equals to 1.

4) The diffusion kurtosis model (DKM):

$$\frac{S_b}{S_0} = exp\left(-b \times D_{app} + \frac{1}{6} \times b^2 \times D_{app}^2 \times K_{app}\right)$$

where calculated apparent kurtosis (K_{app}, unitless) reflects the deviation of water motion from the Gaussian distribution, and apparent diffusion (D_{app}, in mm^2/s) equals to ADC when K_{app} equals to 0.

In this study, mono-exponential, bi-exponential, stretched-exponential, and the diffusion kurtosis model were referred as MEM, BEM, SEM and DKM, respectively. DWI data were fitted nonlinearly using the Levenberg-Marquardt minimization algorithm and corresponding imaging parameters from the four models were derived using the following constraints in their initialization values:

1) MEM model: ADC from 0.1 ($10^{-3}mm^2/s$) to 4.0 ($10^{-3}mm^2/s$) with an initial value of 1.5 ($10^{-3}mm^2/s$).
2) BEM model: D_{slow} from 0.1 ($10^{-3}mm^2/s$) to 4.0 ($10^{-3}mm^2/s$) with an initial value of 1.5 ($10^{-3}mm^2/s$), D_{fast} from 10 ($10^{-3}mm^2/s$) to 300 ($10^{-3}mm^2/s$) with an initial value of 50 ($10^{-3}mm^2/s$), and f from 0.05 to 0.9 with an initial value of 0.2.

3) SEM model: DDC from 0.1 (10^{-3}mm^2/s) to 4.0 (10^{-3}mm^2/s) with an initial value of 1.5 (10^{-3}mm^2/s), and α from 0 to 1 with an initial value of 0.5.

4) DKM model: D$_{app}$ from 0.1 (10^{-3} mm^2/s) to 4.0 (10^{-3} mm^2/s) with an initial value of 1.5 (10^{-3} mm^2/s), and K$_{app}$ from 0 to 2.5 with an initial value of 1.
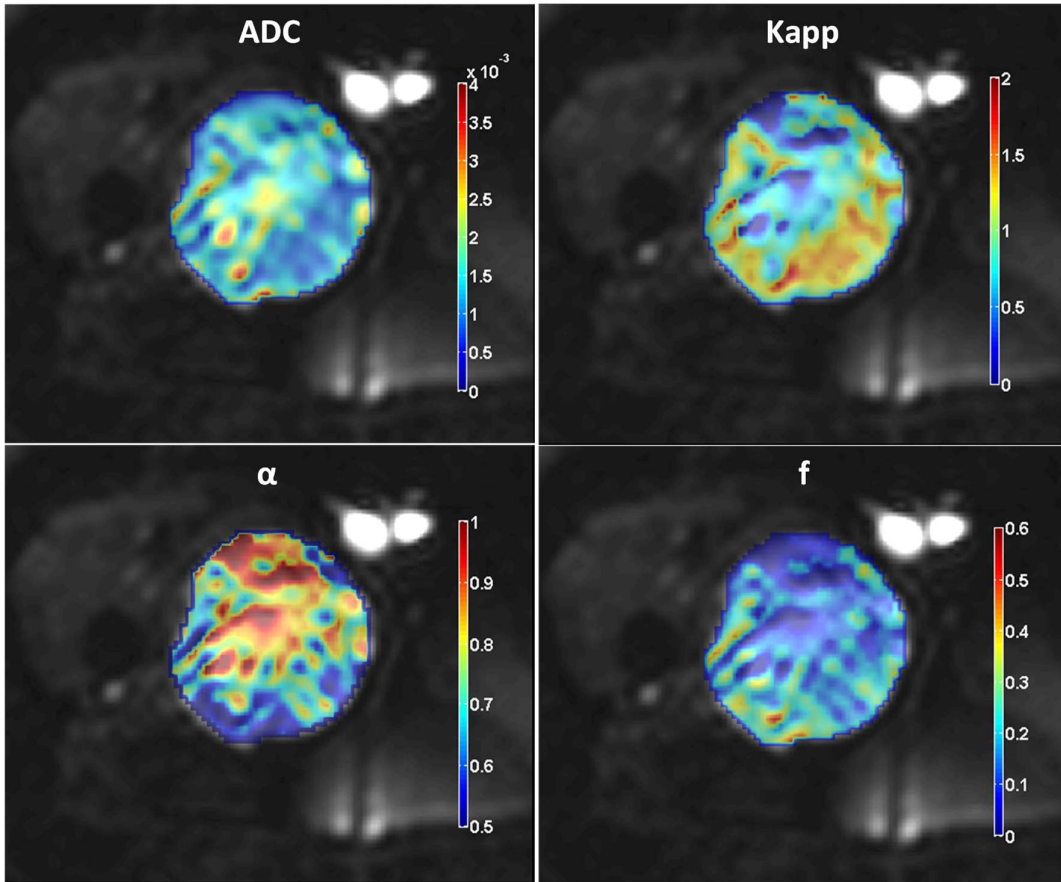
Figure 52 An undifferentiated pleomorphic sarcoma (UPS) case. From top-left to bottom-right: parametric maps of ADC, Kapp, α, and f derived from the MEM, DKM, SEM, and BEM, respectively.

5.2.5 Model Suitability

Small size corrected Akaike Information Criteria (AIC), denoted as AICc, was initially calculated to assess quantitatively how a model fits diffusion signal attenuation and select the most suitable model for each pixel on the basis of a trade-off between fitting performance and complexity (i.e. number of estimated model parameters). AICc, an adjustment to AIC used to compensate for the relatively large number of the estimated model parameters (K equals to 1, 3, 2 and 2 for the MEM, BEM, SEM and DKM, respectively) compared to the number of b-values (N=8), was performed for all pixels within ROIs and models were ranked with lower AICc score implying a better score.

$$AICc = AIC + \frac{2 \times (K + 1) \times (K + 2)}{N - K - 2}$$

AIC is given according to:

$$AIC = N \times ln\left(\frac{RSS}{N}\right) + 2 \times (K + 1)$$

where RSS is the sum of squares of residuals. However, representing Akaike criterion scores with probability measures is reported to be more beneficial in model selection studies since it facilitates a direct interpretation of the relative merits of the investigated models [44]. To this end, Akaike Weights (AW) were calculated for each model in the examined set using AICc scores differences (ΔAICc) between all models and the model with the lowest AICc at a given pixel.

$$\Delta AICc_i = AICc_i - minAICc$$

AICc scores of all models were then transformed into conditional probabilities (varying from 0 to 1) using the following equation where R is the number of models in the examined set and the sum of AW being equal to 1. An analytical representation of the proposed model suitability procedure can be found in [45].

$$AW_i = \frac{exp\left(-\frac{\Delta AICc_i}{2}\right)}{\sum_{r=1}^{R} exp\left(-\frac{\Delta AICc_r}{2}\right)}$$

5.2.6 Design of the parameters meta-map

Pixel-based model probabilities were further used to post-process every single model parametric map. The underlying hypothesis was that since different pixels were assigned to different models, single model parametric maps could not represent whole tumor characteristics, thus resulting to incorrect values for each specific diffusion-related property (e.g. cellularity). Therefore, post-processing of all acquired maps was performed using the AW, yielding post-processed maps that only exhibit values from pixels that correspond to each selected model (e.g. from the initial ADC map select only these pixels where MEM was selected as the most suitable model). The CDM map design was focused on two directions. On one hand, all derived parametric maps were optimized since only pixels that favored the selection of the corresponding models participated in the analysis (herein meta-maps). This way, e.g. meta-Kapp (exclusive to DKM) was computed only from the pixels that favor the DKM instead of the Kapp map using all pixels within the examined ROI. On the other hand, meta-maps underlying the same tissue properties (e.g. meta-ADC, meta-D$_{slow}$, meta-DDC and meta-D$_{app}$) were fused into a single meta-map (overall-meta-D), reflecting cellularity of the whole ROI (see below figure).

Precisely, the design of all post-processed maps, herein called meta-maps, comprises four distinct phases outlined below. At the first step, once model weights had been assigned, a composite diffusion model (CDM) map was generated based on a simple voting scheme where

a model with the highest AW was selected as the most suitable model for each examined pixel within ROIs. Next, the composite map was overlaid on the diffusion images and subregions within ROIs were annotated and classified using a particular pseudo-color linked to the most suitable model. Step 3, focused on the development of the following meta-maps: a) ADC meta-map from MEM, b) D_{slow}, D_{fast} and f meta-maps from BEM, c) DDC and α meta-maps from SEM, and d) D_{app} and K_{app} meta-maps from DKM. For the reader's convenience, all meta-maps were considered as meta-ADC, meta-D_{slow}, meta-D_{fast}, meta-f, meta-DDC, meta-α, meta-D_{app}, and meta-K_{app}. Subsequently, at Step 4, tissue cellularity of the whole ROI was displayed by a unique meta-map for D (overall-meta-D), generated by incorporating all subregions from meta-ADC, meta-D_{slow}, meta-DDC and meta-D_{app}.

An indicative example is given in the following figure where maps related to tissue cellularity were initially calculated from all models and for the entire ROI (1st column in the schematic pipeline), the CDM map was masked individually to all maps (2nd column), meta-maps were generated exclusively from pixels assigned to the most suitable model (3rd column) and fused to compose the overall-meta-D depicted at the final step of the pipeline process (4th column).
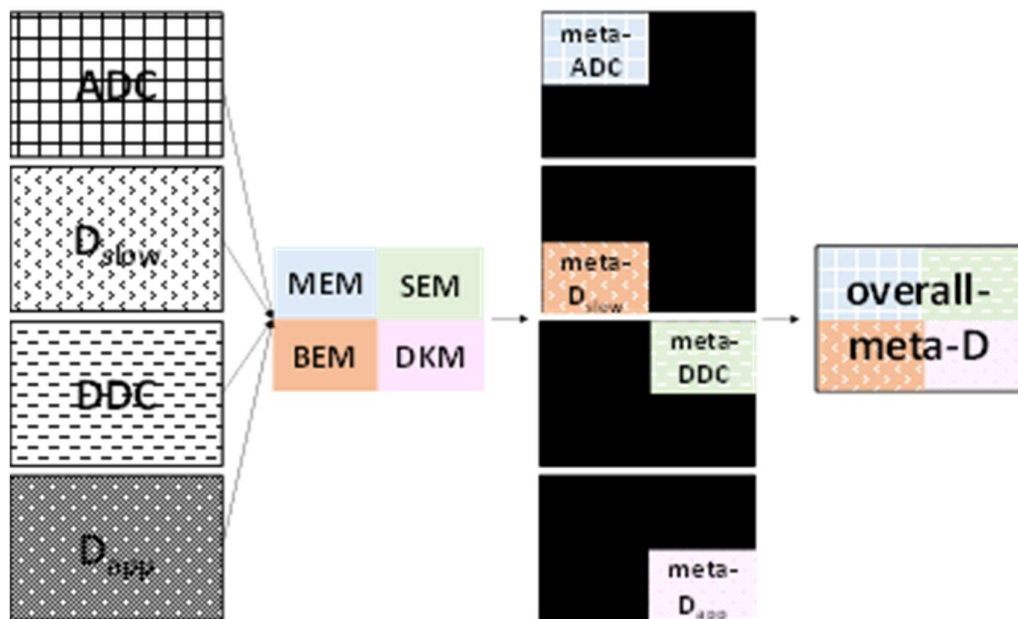
Figure 53 An illustrative example of the parameter meta-map design displayed through a pixel-based 4-step pipeline process. Step 1: Assuming that each rectangle on the left part of the figure represents whole slice tumor delineation, ADC, Dslow, DDC and Dapp maps were initially derived using the 4 models. Step 2: Akaike Weights were calculated, yielding the composite diffusion model (CDM) map for this particular slice. Step 3: CDM was masked individually to each map, exhibiting only these pixels where the MEM, BEM, SEM, and DKM was the most suitable for these particular pixels, respectively. Black regions include all rejected pixels. Subsequently, meta-maps of ADC, Dslow, DDC and Dapp were calculated. Step 4: overall-meta-D was then generated, by a weighted sum of all calculated meta-maps from step 3.

5.2.7 Ground-truth validation

Prior to the core analysis, validation of model suitability was carried out and parametric maps from the four examined models were generated from several anatomical areas of known tissue properties from pathology. Specifically, ROIs across multiple slices were drawn on the healthy bladder, gallbladder and vasculature areas including the common iliac arteries. From one hand, healthy bladder and gallbladder store the urine and the bile respectively, therefore a single compartment of liquids is prominent and the MEM was expected to dominate in both areas. On the other hand, clearly defined vasculature areas were annotated from multiple patients and the BEM model, reflecting vascularization, was assumed to be the model that best characterizes these particular areas.

5.2.8 Summary Statistics and parameters performance

Statistical analysis was conducted in R (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria) and data were expressed as mean (standard deviation). Prior to any analysis, goodness-of-fit of all models was evaluated using the bias corrected adjusted coefficient of determination (adjusted-R2) and pixels having an adjusted-R2 of less than 0.7 for all models were excluded from further processing, indicating noisy acquired signals, artifacts and failure of the fitting. An exploratory whole-tumor histogram analysis was performed for each calculated parameter summarized in Table 3, yielding several metrics including the mean, median, skewness, kurtosis, entropy, and interquartile range (IQR) which quantifies the spread of data between the 75th and 25th percentiles of the histogram. To this end, 102 histogram metrics were produced by each patient further used to differentiate low- from high-grade STSs and normality of their distribution was evaluated using Shapiro-Wilk test. When assumption was satisfied metrics were compared between low- and high-grade STS using independent sample t-test; otherwise a Mann-Whitney U test was conducted to disclose differences between the two groups. STSs of grade 1 were classified as low-grade tumors whereas grades 2 and 3 were classified as high-grade [58]. Results with p-values less than 0.0005 (i.e. 0.05/102 where 102 was the number of the examined parameters) were declared statistically significant incorporating correction to account for multiple comparisons using the Bonferroni adjustment. Prior to any analysis, goodness-of-fit of all models was evaluated using the bias corrected adjusted coefficient of determination (adjusted-R^2) and pixels having an adjusted-R^2 of less than 0.7 for all models were excluded from further processing, indicating noisy acquired signals and artifacts. Subsequently, Receiver Operator Characteristic (ROC) analysis was used to assess the performance of each statistically significant parameter. Quantitative metrics including the Area Under the Curve (AUC), sensitivity, specificity, negative and positive predictive values (NPV and PPV), accuracy and the optimal cutoff value of each ROC curve were calculated from ROC analysis. The optimal cutoff was defined based on the Youden index and 95% confidence interval (CI) for each metric was estimated using 1000 bootstrap resamples.

Table 3 Calculated parametric maps and related meta-maps from the four examined models

Parameters	Corresponding map
ADC	apparent diffusion coefficient from MEM
meta-ADC	ADC and the CDM
D_{slow}	slow diffusion coefficient from BEM
meta-D_{slow}	D_{slow} and the CDM
DDC	distributed diffusion coefficient from SEM
meta-DDC	DDC and the CDM
D_{app}	apparent diffusion from DKM
meta-D_{app}	D_{app} and the CDM
overall-meta-D	overall diffusion from meta-ADC, meta-D_{slow}, meta-DDC and D_{app}
D_{fast}	fast diffusion coefficient from BEM
meta-D_{fast}	D_{fast} and the CDM
f	fraction of fast component from BEM
meta-f	f and the CDM
α	index α from SEM
meta-α	α and the CDM
K_{app}	apparent kurtosis from DKM
meta-K_{app}	K_{app} and the CDM

5.3 Results

5.3.1 Confirming the necessity of the proposed CDM approach

Prior to the core analysis, fitted pixels within ROIs where all models failed to achieve an adjusted-R^2 higher than 0.7 were excluded from the analysis (Table 4).

Table 4 Pixel-wise adjusted-R^2 for each of the four examined models expressed as mean (sd). Percentage of pixels within tumor volumes and other anatomical areas having an adjusted-R^2 value below the defined threshold (adjusted-R^2<0.7) are given in brackets. Pixel percentages of the examined regions where all models achieved an adjusted-R^2 less than 0.7 and therefore were excluded from further analysis are reported in the last column.

Grade	Patient	Adjusted-R^2: mean (sd) [%of pixels below threshold]				%of excluded pixels
		MEM	BEM	SEM	DKM	
Low	P1	0.92 (0.15) [8.21]	0.93 (0.17) [8.07]	0.94 (0.15) [6.70]	0.92 (0.17) [9.17]	5.79
	P2	0.91 (0.16) [6.89]	0.97 (0.07) [0.90]	0.98 (0.06) [0.60]	0.94 (0.13) [3.74]	0.60
	P3	0.99 (0.04) [0.44]	0.99 (0.05) [0.75]	0.99 (0.04) [0.52]	0.99 (0.04) [0.53]	0.42
	P4	0.98 (0.05) [0.59]	0.99 (0.06) [0.90]	0.99 (0.05) [0.67]	0.99 (0.05) [0.71]	0.56
	P5	0.99 (0.01) [0]	0.99 (0.01) [0]	1 (0) 0	0.99 (0.01) [0]	0
	P6	0.95 (0.07) [1.50]	0.95 (0.08) [0.92]	0.96 (0.06) [0.75]	0.95 (0.08) [1.27]	0.70

	P7	0.98 (0.07) [1.24]	0.99 (0.07) [1.10]	0.99 (0.06) [0.88]	0.98 (0.08) [1.44]	0.75
	P8	0.99 (0.05) [0.54]	0.99 (0.06) [0.75]	0.99 (0.05) [0.54]	0.99 (0.06) [0.63]	0.45
	P9	0.98 (0.06) [0.75]	0.99 (0.04) [0.17]	0.99 (0.03) [0.17]	0.99 (0.05) [0.54]	0.10
	P10	0.97 (0.09) [1.59]	0.97 (0.1) [1.86]	0.97 (0.09) [1.46]	0.97 (0.09) [1.77]	1.27
	P11	0.97 (0.07) [1.30]	0.98 (0.09) [1.47]	0.98 (0.08) [1.20]	0.97 (0.08) [1.47]	1.02
	P12	0.92 (0.16) [5.24]	0.94 (0.19) [6.31]	0.94 (0.17) [5.54]	0.92 (0.17) [6.05]	4.91
	P13	0.96 (0.12) [2.99]	0.98 (0.09) [1.42]	0.98 (0.09) [1.30]	0.96 (0.12) [2.42]	1.05
	P14	0.96 (0.04) [0.14]	0.98 (0.02) [0]	0.98 (0.01) [0]	0.97 (0.03) [0]	0
	P15	0.97 (0.09) [1.54]	0.99 (0.06) [0.55]	0.99 (0.06) [0.53]	0.98 (0.08) [1.27]	0.42
	P16	0.87 (0.18) [11.75]	0.98 (0.05) [0.39]	0.97 (0.05) [0.78]	0.9 (0.17) [8.54]	0.19
	P17	0.92 (0.13) [6.09]	0.99 (0.04) [0.22]	0.99 (0.03) [0.19]	0.96 (0.08) [1.69]	0.16
	P20	0.79 (0.2) [22.43]	0.92 (0.12) [3.58]	0.89 (0.13) [6.88]	0.76 (0.24) [26.20]	1.92
High	P21	0.96 (0.12) [2.70]	0.99 (0.08) [0.90]	0.98 (0.08) [1.17]	0.97 (0.12) [2.62]	0.72
	P22	0.94 (0.09) [2.63]	0.99 (0.02) [0.12]	0.99 (0.03) [0.14]	0.96 (0.08) [1.56]	0.05
	P23	0.91 (0.13) [6.71]	0.97 (0.09) [2.14]	0.97 (0.08) [1.58]	0.94 (0.12) [4.18]	1.17
	P24	0.87 (0.19) [10.27]	0.93 (0.16) [5.68]	0.94 (0.14) [4.41]	0.9 (0.18) [7.47]	3.73
	P25	0.98 (0.05) [0.47]	0.99 (0.06) [0.57]	0.99 (0.04) [0.47]	0.98 (0.05) [0.66]	0.28
	P26	0.96 (0.06) [0.87]	0.98 (0.07) [0.96]	0.99 (0.06) [0.65]	0.98 (0.06) [0.65]	0.48
	P27	0.98 (0.05) [0.66]	0.98 (0.03) [0.19]	0.99 (0.02) [0.16]	0.98 (0.04) [0.38]	0.07
	P28	0.99 (0.05) [0.35]	0.99 (0.05) [0.39]	0.99 (0.05) [0.33]	0.99 (0.05) [0.41]	0.29
	P29	0.96 (0.09) [2.18]	0.99 (0.03) [0.13]	0.99 (0.03) [0.10]	0.97 (0.07) [1.08]	0.05
	P30	0.96 (0.06) [0.98]	0.98 (0.03) [0.14]	0.98 (0.03) [0.13]	0.96 (0.06) [1.25]	0.11
Iliac Arteries		0.02 (0.08) [100]	0.99 (0) [0]	0.10 (0.20) [98.33]	0.01 (0.05) [100]	0
Gallbladder		0.97 (0.02) [0]	0.97 (0.03) [0]	0.97 (0.02) [0]	0.97 (0.02) [0]	0
Bladder		0.99 (0.02) [0]	0.99 (0.01) [0]	0.99 (0.02) [0]	0.99 (0.02) [0]	0

Known Anatomical sites

Model suitability was next validated using multiple annotations from anatomical areas of known tissue properties from pathology. More than 96% and 98% of the bladder and gallbladder pixels were assigned to the MEM, respectively. The BEM was favored from all pixels within the iliac arteries. High meta-ADC (bladder: 3.35 (0.33) x10^{-3}mm^2/s and gallbladder: 3.37 (0.57) x10^{-3}mm^2/s) and meta-f values (iliac arteries: 0.87 (0.09)) were detected, as theoretically expected, since a single compartment of liquids is prominent in the bladder and gallbladder, whereas arteries comprise vasculature areas (Figure 54 and Table 5).
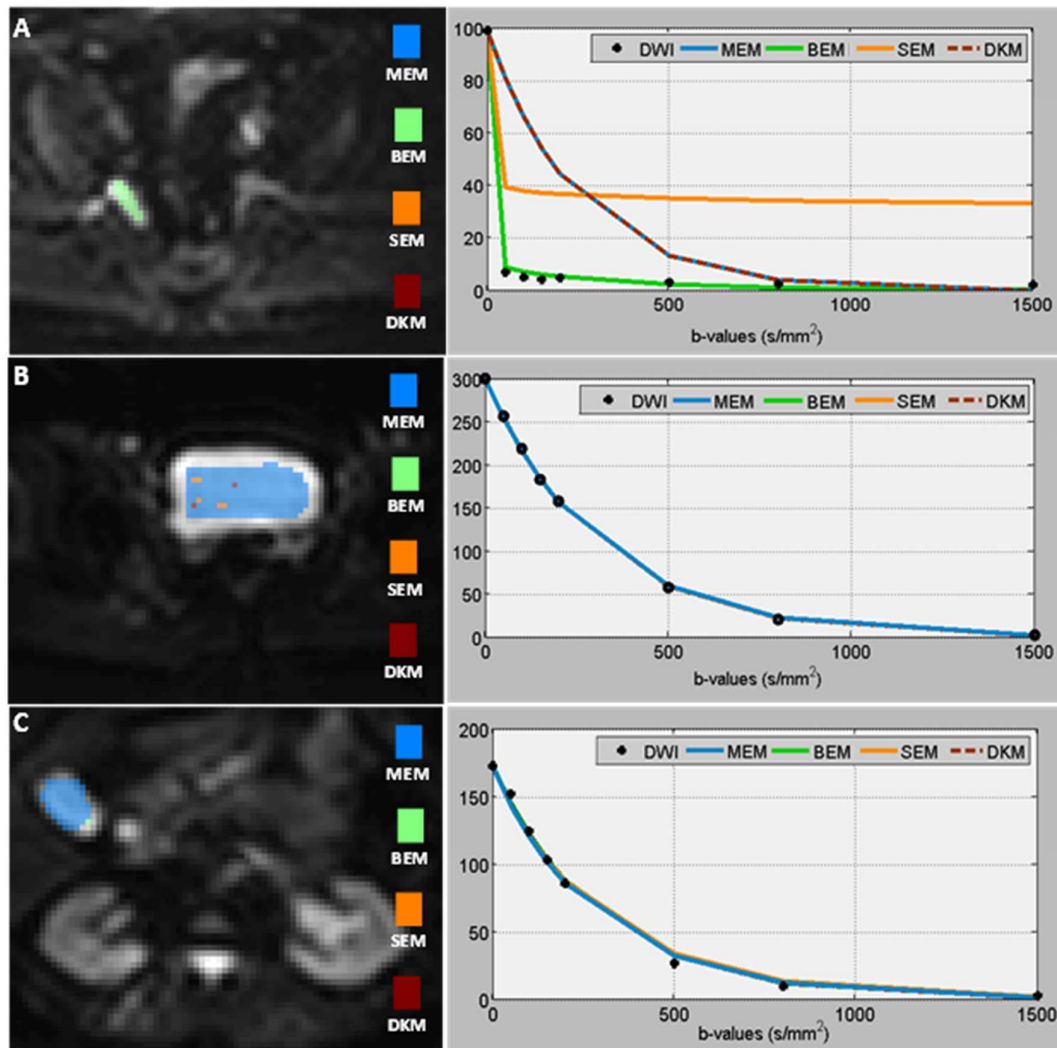
Figure 54 Validating model preference using indicative anatomical areas of known tissue properties from pathology. (Left): The composite diffusion model (CDM) map, overlaid to the right internal iliac artery (A), a bladder (B) and a gallbladder (C). (Right): Calculated mean signal intensities and related fitted curves from all models when applied on a pixel-by-pixel basis to the iliac artery (A), the bladder (B) and the gallbladder (C).

Table 5 Model preference in pixel percentages within the bladder, gallbladder and the iliac arteries. Meta-f and meta-ADC were also displayed as mean (sd).

Area	MEM (%)	BEM (%)	DKM (%)	SEM (%)	Parameter

Bladder	96.33	0.40	1.83	1.44	meta-ADC [x10^{-3} mm^2/s]: 3.35 (0.33)
Gallbladder	98.57	0.94	0	0.49	meta-ADC [x10^{-3} mm^2/s]: 3.37 (0.57)
Arteries	0	100	0	0	meta-f: 0.87 (0.09)

Pre and post-therapy induced changes

The purpose of this analysis was to examine whether therapy induced changes in the tumor could be quantified non-invasively using the proposed technique. To this end, whole tumor model suitability changes associated to treatment in patients diagnosed with undifferentiated pleomorphic sarcoma (UPS) and spindle cell myxoid liposarcoma (SCML) were investigated and the corresponding results are summarized in Table 6. According to the results, MEM was the model of choice in post-therapy DWI from both tumor types. Especially for the UPS case, post-treatment CDM maps reported MEM suitability in the majority of the pixels within the tumor (83.07%), indicating a homogeneous non-vascularized tissue microenvironment almost fully characterized by this model. Chemotherapy treatment caused an increase of 49.02% (from 34.05% pre-therapy pixels to 83.07% post-therapy pixels within the tumor) and 19.25% (from 33.65% pre-therapy pixels to 52.9% post-therapy pixels) of the UPS and the SCML pixels that voted for the MEM when single-slice CDM maps were overlaid to each relative diffusion image, respectively. On the contrary, BEM was the model of choice in 13.98% and 28.64% of the pre-treatment UPS and SCML pixels, but model's preference was restricted after chemotherapy (3.11% and 18.19% respectively). Similarly, SEM and DKM were most suitable for characterizing 29.53% and 22.44% of the total tumor area in UPS before treatment. However, their applicability in post-treatment DWI was significantly limited to an overall area of 5.53% and 8.29%, respectively. Pre- and post-treatment CDM maps also demonstrated restrictions to pixel percentages best characterized by the SEM in a patient with SCML. DKM almost equally contributed to pre- and post-therapy CDM maps in case of the SCML patient. According to the generated pre- and post-therapy CDM maps, in both cases, a drug combination of tumor necrosis factor antagonist (TNF-a) and melphalan with isolated limb perfusion (ILP) targeted neovasculature and highly heterogeneous parts of the tumor favored initially to the BEM and the SEM models (pre-therapy), yielding post-therapy composite where MEM dominated their areas (especially in the UPS case). In other words, heterogenous and vasculature subregions were replaced by areas with pure cellularity.

Table 6 Before and after treatment pixel percentages of the whole tumor preferred by the 4 examined models in an undifferentiated pleomorphic sarcoma (UPS) and spindle cell myxoid liposarcoma (SCML) patient. Treatment induced changes were also reported as percentage changes of model preference between the two time points.

		UPS	SCML
MEM (%)			
	Before Treatment	34.05	33.65

	After Treatment	83.07	52.90
	Treatment Changes	+49.02	+19.25
BEM (%)			
	Before Treatment	13.98	28.64
	After Treatment	03.11	18.19
	Treatment Changes	-10.87	-10.45
SEM (%)			
	Before Treatment	29.53	29.77
	After Treatment	05.53	21.76
	Treatment Changes	-24	-08.01
DKM (%)			
	Before Treatment	22.44	07.94
	After Treatment	08.29	07.14
	Treatment Changes	-14.15	-0.80

Single slice model suitability changes in UPS and SCML areas are illustrated in figures 55 and 56, respectively. Taking into account the large variation in the shape and size of the mass preceding chemotherapy, an absolute measurement of the same slice position with reference to anatomical landmarks was not preferred as it corresponded to different relative position within the lesion. Therefore, pre- and post-therapy slices were chosen from the central position as defined by the middle of the maximum length of the lesion at the coronal slice at the corresponding imaging sequence. Figures 55A, 56A, 55B and 56B depict a slice-to-slice pre- and post-treatment comparison of the central part of the UPS and the SCML when the CDM map was overlaid on the diffusion image, respectively. In both tumor types, pre-therapy slice reveals multiple tumor subregions characterized by different models (Figures 55A and 56A). In case of the UPS, the pre-therapy slice (Figure 55A) shows a peripheral area with vasculature activity that best characterized by the BEM (green), an intermediate rounded part of the tumor that best described by the single MEM (blue color), and the most heterogeneous and complex part of the tumor (lower central part) as the SEM (orange) and the DKM (red) where the most applicable models in this particular area. A significant increase in the percentage of pixels that favored the MEM was apparent in the post-therapy slice of the UPS (Figure 55B), whereas BEM, SEM and DKM preference was notably decreased. Accordingly, pre-therapy slice of the SCML (Figure 56A) depicts a central area that mostly voted for MEM (blue), surrounded by a vasculature area according to the BEM (green) and subregions of SEM. Figure 56B underpins the anti-vascularization effect of the treatment, showing a significant decrease of pixels preferred by the BEM and increased pixels of MEM.

Pre- and post-therapy overall-meta-D histograms of the whole UPS and SCML are demonstrated in Figure 55C and Figure 56C showing a shift of overall-meta-D towards higher values. A decrease in meta-f and meta-K_{app} and higher meta-α values from the UPS patient were also apparent after therapy (Figure 55D). All differences were statistically significant (p<0.001). Whole tumor

changes in the parameters related to vascularization, complexity and heterogeneity were shown as bar plots in Figure 56D. In case of SCML tumor, all parameter differences except meta-α were found to be statistically significant (p<0.001). Whole tumor changes in the parameter values for both the UPS and SCML are also presented quantitatively in Table 7.
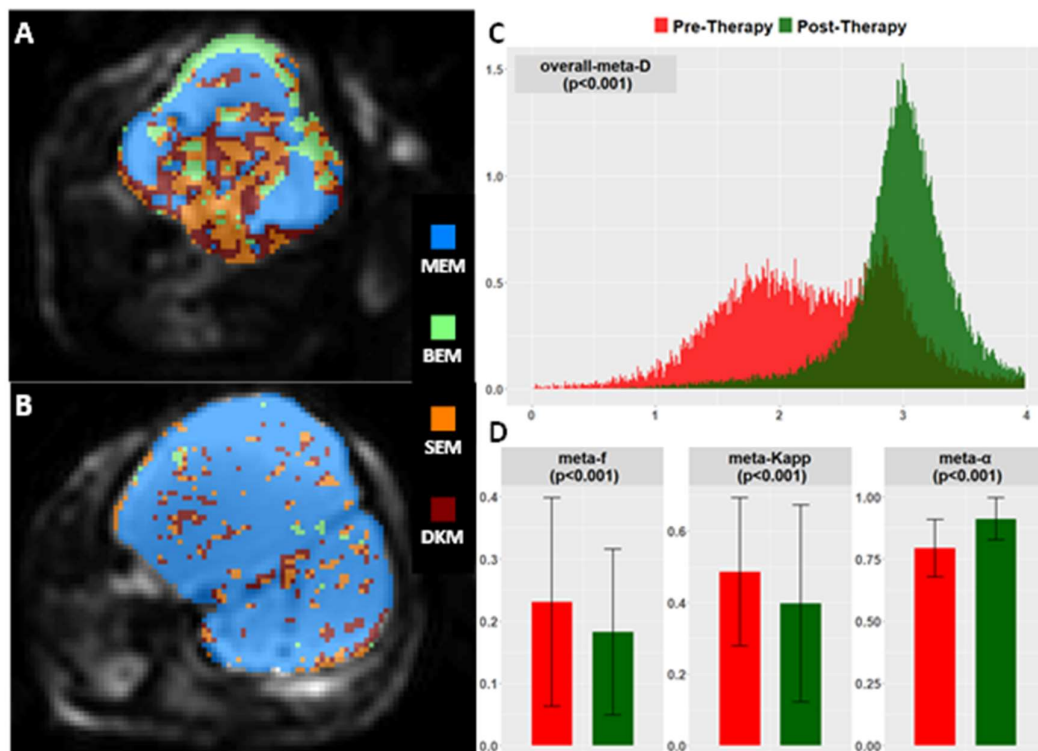
Figure 55 Slice-to-slice pre- and post-treatment comparison of a patient diagnosed with an undifferentiated pleomorphic sarcoma. A: central slice of a diffusion image before treatment overlaid with the calculated CDM map. B: Superimposed CDM map to the central slice of the tumor after treatment. C: pixel-based histograms demonstrating changes in the overall-meta-D values before (red) and after (green) treatment. D: Bar plots showing treatment induced changes in meta-f, meta-Kapp and meta-α. Note: In both A and B different colors are associated to different models depicting model preference across all pixels within the tumor.
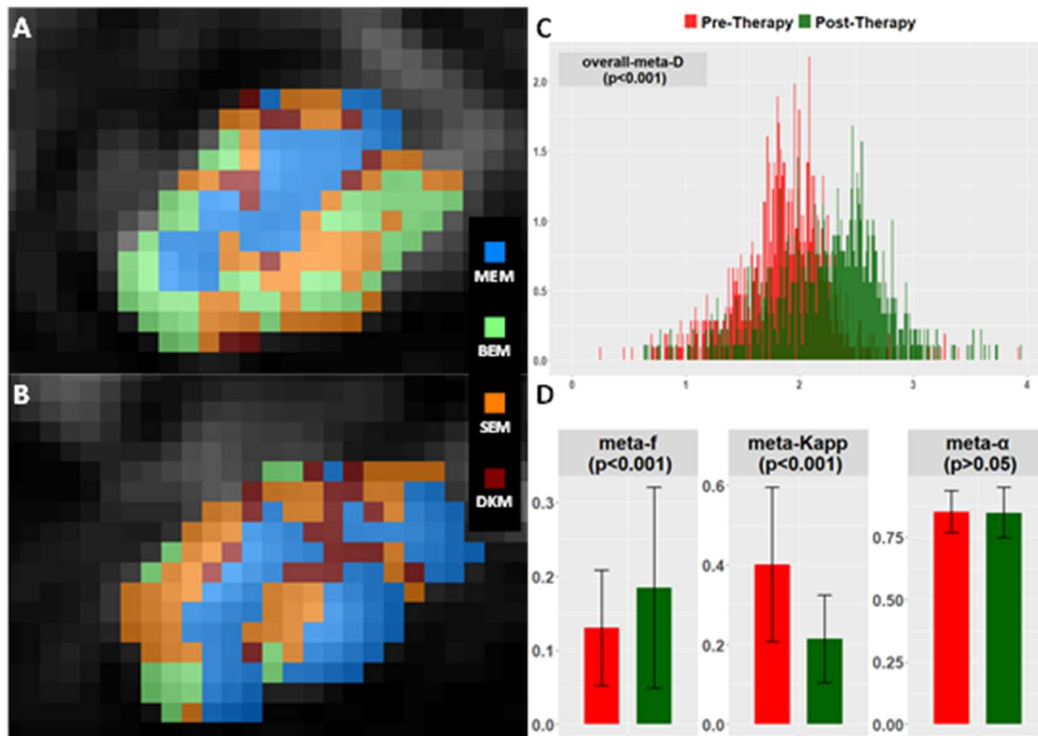
Figure 56 Slice-to-slice pre- and post-treatment comparison of a patient diagnosed with a spindle cell myxoid liposarcoma. A: central slice of a diffusion image before treatment overlaid with the calculated CDM map. B: Superimposed CDM map to the central slice of the tumor after treatment. C: pixel-based histograms demonstrating changes in the overall-meta-D values before (red) and after (green) treatment. D: Bar plots showing treatment induced changes in meta-f, meta-Kapp and meta-α. Note: In both A and B different colors are associated to different models depicting model preference across every pixel within the tumor.

Table 7 Pre, post- and treatment changes in overall-meta-D, meta-f, meta-Kapp and meta-α calculated from their corresponding meta-maps from 2 patients with undifferentiated pleomorphic sarcoma (UPS) and spindle cell myxoid liposarcoma (SCML), respectively.

		UPS	SCML
overall-meta-D ($\times 10^{-3}$ mm^2/s)			
	Before Treatment	2.280 (0.722)	1.863 (0.401)
	After Treatment	2.989 (0.462)	2.239 (0.509)
	Treatment Changes	+0.709 (-0,260)	+0,376 (+0.108)
meta-f			
	Before Treatment	0.231 (0.167)	0.130 (0.078)
	After Treatment	0.182 (0.133)	0.185 (0.135)
	Treatment Changes	-0.049 (-0.034)	+0.055 (+0.057)
meta-Kapp			
	Before Treatment	0.486 (0.207)	0.401 (0.193)

	After Treatment	0.397 (0.275)	0.214 (0.110)
	Treatment Changes	-0.089 (+0.068)	-0.187 (-0.083)
meta-α	Before Treatment	0.793 (0.115)	0.889 (0.112)
	After Treatment	0.911 (0.085)	0.848 (0.101)
	Treatment Changes	+0.118 (-0.030)	-0.041 (-0.011)

As a next validation step, whole tumor model selection for a dedifferentiated liposarcoma (DDLS) (highly aggressive with distinctive intra-tumoral heterogeneity) and an alveolar soft part sarcoma (ASPS) (highly vascularized of uncertain differentiation) is summarized in Table 8. The SEM was the most preferred model for the DDLS, accounting 52.04% of all tumor pixels whereas the BEM was selected for 40.15% of the ASPS pixels.

Table 8 Whole tumor model preference percentages of a dedifferentiated liposarcoma (DDLS) (highly aggressive with increased intra-tumoral heterogeneity) and an alveolar soft part sarcoma (ASPS) (rare and highly vascularized).

	MEM (%)	BEM (%)	SEM (%)	DKM (%)
DDLS	10.31	6.54	52.04	31.11
ASPS	40.08	40.15	18	1.77

To investigate and illustrate the prospective associations between histology and model suitability, a spatial alignment was obtained between the surgical specimen and the central DWI slice of the DDLS and the ASPS by dividing tumoral area into several localized ROIs using a virtual grid in the DWI matching the slabs of the central tumor slice. Furthermore, histopathology sequential microphotographs illustrating areas of interest of the slabs of the central tumor slice were taken. CDM map and five tumor subregions (black circles), associated to the corresponding areas of the tumor slabs (microphotographs 57A-57E), were overlaid to the central DWI tumor slice of the DDLS (Figure 57F). ROI A showed a mixed model suitability pattern, mostly favored to the MEM which was confirmed from microphotograph 57A, illustrating an area of the dedifferentiated non-lipogenic tumor component with an intermediate grade of malignancy consisting of relatively monomorphic tumor cells with cytologic features of moderate dysplasia, a solid architectural pattern and increased vascularity. SEM dominated ROI B, indicating high heterogeneity. This was evident from the histopathology of this area (microphotograph 57B), rendering a highly heterogeneous architecturally and cytologically solid-cystic dedifferentiated tumor area (pleomorphic anaplastic tumor cells with high mitotic rate, plentiful abnormal mitoses, cystic and hemorrhagic cavities, a small area of necrosis). ROI C also presented a mixed

pattern of three models (MEM, SEM and DKM). This area (microphotograph 57C) was defined histopathogically by cellularity, high grade morphology and vascularity. The lowest α values of the slice (0.476 (0.06)) were found in ROI D where the SEM was favored by all pixels. This result was in accordance with the histology of this subregion (microphotograph 57D), reporting the most cellular, highly anaplastic area of the dedifferentiated component, comprising a highly pleomorphic tumor cell population of gigantic and multinucleated cells with nuclear features of extreme anaplasia, very high cellularity and increased vascularity with abnormally structured blood vessels. Finally, pixels from ROI E voted for the DKM where this subregion (microphotograph 57E) was histologically characterized by cellularity and complexity with multiple foci of transitions of the lipogenic well-differentiated component in the solid dedifferentiated component.
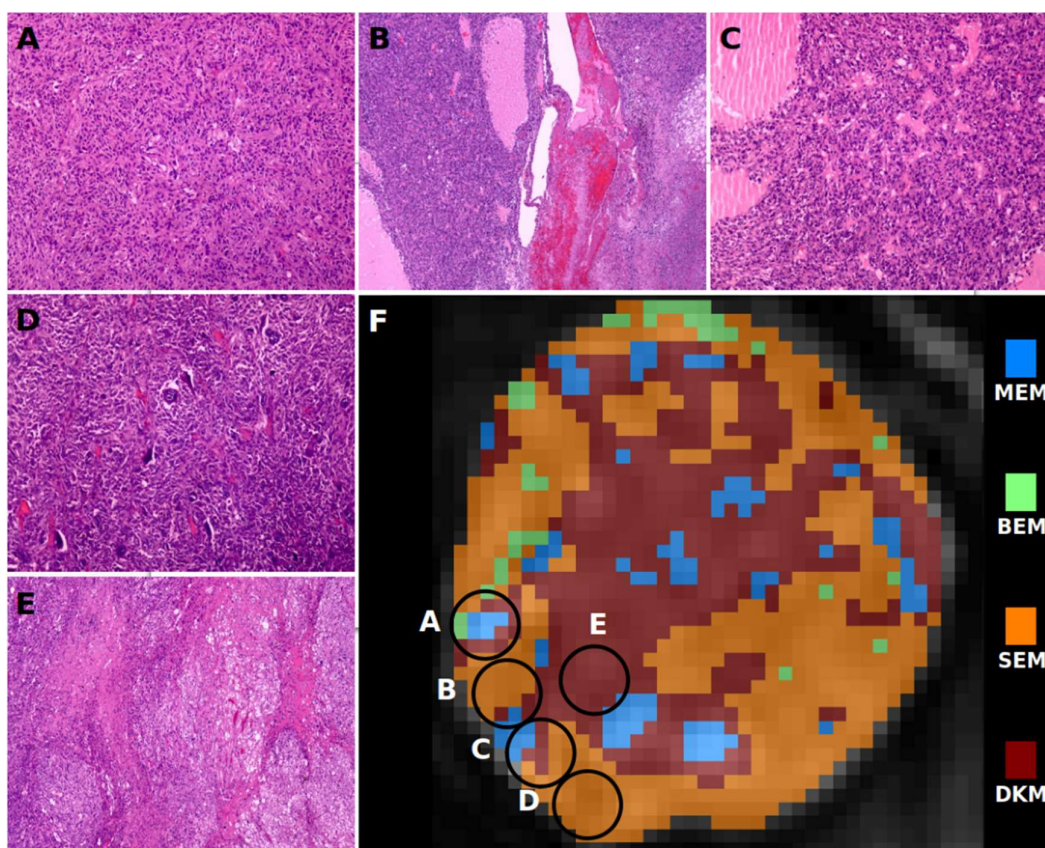
Figure 57 A dedifferentiated liposarcoma. From (A) to (E): Tumor exhibiting a complex highly heterogeneous architecture, cytologically, in the morphologies and feeding blood vessels in cellularity, presenting extreme variations in the degrees of differentiation. (F): The composite diffusion model (CDM) map, overlaid to the central DWI tumor slice. ROIs A-E were associated to the microphotographs of the surgical specimen. MEM: mono-exponential; BEM: bi-exponential; SEM: stretched-exponential; DKM: diffusion kurtosis model.

BEM was the most voted model in the central tumor slice of the ASPS, confirming the highly vascularized nature of the tumor (Figure 58F). Model preference from ROI A and C (MEM was favored by all pixels) was in accordance with histology of the areas of the tumor slabs (microphotographs 58A and 58C), which underpin the well-structured mostly cellular tumor

areas comprising well-defined compartments delineated by very thin focally dilated fibrous septa. Highly vascularized areas were noticed in regions from ROI B and D (BEM dominated) and also depicted in the microphotographs 58B and 58D where the histological examination revealed tumor areas with geometric honeycomb-like cellular compartments outlined by a rich network of thin-walled sinusoid like, regionally congested blood vessels. Finally, the SEM dominated region from ROI E where histology reported an area of the tumor characterized by a relatively complex nodular architecture consisting of multiple variously sized and variably cellular tumor lobules, delineated by thick fibrous septa with random focally dilated and congested blood vessels (microphotographs 58E).
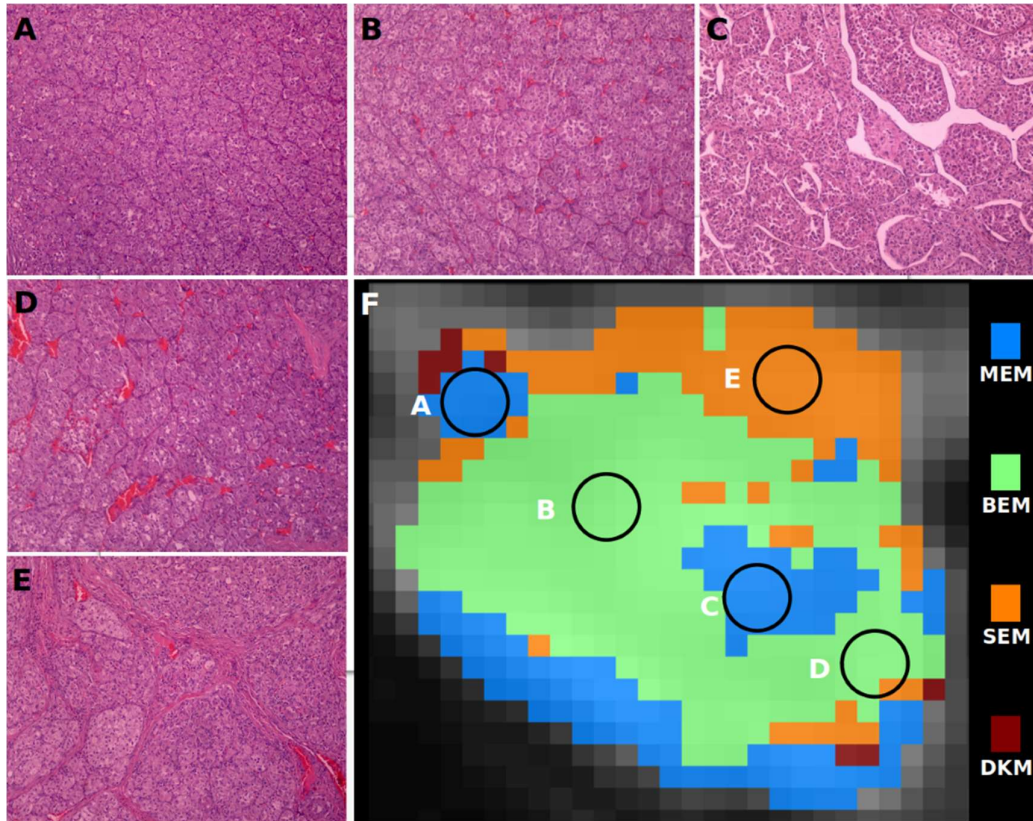
Figure 58 An alveolar soft part sarcoma. (A): a tumor area characterized by the orderly structured well- defined cellular tumor compartments, (B) a tumor area consisting exclusively of geometric honeycomb-like cellular compartments outlined by highly vascularized fibrous septa, (C): less monomorphic but still distinct cellular compartments of various sizes and shapes with some variations in cellularity divided by fibrous septa with ectatic (vascular) spaces, (D): tumor area with geometric honeycomb-like cellular compartments divided by highly vascularized fibrous septa containing congested variably sized blood vessels, (E): a relatively complex architecturally area of the tumor exhibiting multiple variably sized tumor cell nodules with uneven cellularity outlined by fibrous septa of variable thickness comprising unevenly distributed variably sized and focally congested blood vessels, (F): the composite diffusion model (CDM) map, overlaid to the central DWI slice of the tumor. ROIs A-E were associated to the microphotographs of the surgical specimen.

5.3.2 Assessing the statistical superiority of the CDM derived-metrics

Model preference for low- and high-grade STSs is summarized in Table 9, revealing contribution of all examined models in the characterization of the low- and high-grade STSs. The MEM was the most preferred model for both groups, favored by 38.23% (low-grade) and 36.51% (high-grade) of all pixels within the tumor ROIs. The BEM was mostly suitable for pixels within high-grade tumors (14.51%) compared to low-grade tumors (8.46%). AICc analysis yielded increased number of pixels from high-grade tumors classified to the SEM (29.03%) compared to pixels within low-grade tumors (24.78%). On the contrary, the DKM model was selected as the most suitable model in 28.53% and 19.95% of the total pixels for low-grade for high-grade STSs, respectively.

Table 9 Model preference percentages of low and high-grade STSs.

	MEM (%)	BEM (%)	SEM (%)	DKM (%)
Low (n=9)	38.23	14.51	24.78	28.53
High (n=19)	36.51	08.46	29.03	19.95

Table 10 presents the statistical analysis of the diffusion parameters calculated from the whole tumor. Statistical significance was only determined from three parameters out of the 102 examined metrics, all calculated after model selection using the CDM map. The analysis results indicated statistically significant differences between the two groups when the skewness of the meta-D_{slow} (corrected p<0.05) and the overall-meta-D (corrected p<0.05) as well as the IQR of the meta-D_{app} were calculated. Meta-D_{slow} and meta-overall-D were highly positively skewed in high-grade STSs (0.985 and 0.809) whereas a negatively skewed distribution was reported in low-grade patients (-0.467 and -0.796). According to the IQR metric, a higher variability in the histogram calculated from the meta-D_{app} of the high-grade group was computed compared to the meta-D_{app} histogram from the low-grade patients (0.345 vs 0.686).

Table 10 A comparative analysis of all acquired parameters from the four models calculated using parametric maps from single models (left columns) and the CDM map after model selection (right columns) between low- and high-grade STSs. Note: Statistically significant results are displayed in bold. P-values are corrected for multiple testing.

Model	Metrics	Parameters					
		ADC (x10⁻³ mm²/s)			meta-ADC (x10⁻³ mm²/s)		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
MEM	mean	2.450 (0.151)	1.939 (0.564)	0.241	2.149 (0.276)	1.679 (0.544)	1.000
	median	2.425 (0.200)	1.885 (0.703)	0.530	2.198 (0.272)	1.672 (0.604)	1.000
	skewness	0.087 (0.486)	0.945 (1.016)	0.361	-1.245 (1.227)	0.422 (0.789)	0.241
	kurtosis	3.157 (3.207)	2.914 (4.727)	1.000	4.553 (5.759)	1.840 (3.111)	1.000
	entropy	11.618 (1.890)	12.008 (2.578)	1.000	10.362 (1.890)	10.680 (2.224)	1.000

		D$_{slow}$ (x10^{-3} mm^2/s)			meta-D$_{slow}$ (x10^{-3} mm^2/s)		
	IQR	0.532 (0.288)	0.708 (0.289)	1.000	0.438 (0.230)	0.672 (0.427)	1.000
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	2.066 (0.279)	1.537 (0.424)	0.241	1.909 (0.396)	1.506 (0.504)	1.000
	median	2.112 (0.289)	1.495 (0.440)	0.100	1.963 (0.370)	1.467 (0.527)	0.907
	skewness	-1.057 (1.181)	0.686 (0.716)	0.078	**-0.467 (0.701)**	**0.985 (0.884)**	**0.028**
	kurtosis	4.202 (3.967)	2.378 (3.091)	1.000	0.907 (2.729)	3.070 (3.685)	1.000
	entropy	11.771 (1.782)	12.453 (1.933)	1.000	7.403 (2.461)	9.406 (2.050)	1.000
	IQR	0.400 (0.189)	0.531 (0.339)	1.000	0.450 (0.238)	0.516 (0.500)	1.000
		D$_{fast}$ (x10^{-3} mm^2/s)			meta-D$_{fast}$ (x10^{-3} mm^2/s)		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
BEM	mean	37.120 (17.208)	58.601 (41.086)	1.000	149.320 (94.925)	139.113 (43.314)	1.000
	median	10.001 (0.004)	29.200 (67.853)	1.000	155.633 (125.583)	104.947 (108.189)	1.000
	skewness	4.486 (3.359)	2.669 (1.720)	1.000	0.339 (2.151)	0.514 (0.812)	1.000
	kurtosis	29.489 (44.105)	8.212 (11.384)	1.000	2.631 (7.309)	-1.023 (1.879)	1.000
	entropy	5.081 (1.305)	6.137 (1.799)	1.000	5.081 (2.223)	7.138 (1.668)	1.000
	IQR	4.999 (6.912)	60.684 (105.428)	1.000	65.778 (109.088)	237.684 (78.609)	0.125
		f			meta-f		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	0.096 (0.045)	0.114 (0.069)	1.000	0.226 (0.150)	0.191 (0.100)	1.000
	median	0.066 (0.043)	0.095 (0.082)	1.000	0.205 (0.149)	0.165 (0.101)	1.000
	skewness	1.575 (0.911)	1.586 (1.070)	1.000	1.049 (0.466)	1.318 (0.745)	1.000
	kurtosis	3.992 (5.708)	4.855 (7.801)	1.000	1.235 (1.751)	2.698 (3.330)	1.000
	entropy	8.612 (1.454)	9.360 (2.356)	1.000	7.401 (2.459)	9.493 (2.044)	1.000
	IQR	0.127 (0.063)	0.128 (0.077)	1.000	0.129 (0.060)	0.125 (0.057)	1.000
		DDC (x10^{-3} mm^2/s)			meta-DDC (x10^{-3} mm^2/s)		
SEM		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	2.365 (0.194)	1.838 (0.562)	0.296	2.589 (0.291)	1.921 (0.575)	0.157

	median	2.373 (0.227)	1.806 (0.719)	0.296	2.571 (0.259)	1.850 (0.679)	0.100
	skewness	-0.460 (0.861)	0.930 (0.985)	0.078	0.201 (1.146)	0.909 (0.943)	1.000
	kurtosis	4.606 (4.549)	3.129 (4.713)	1.000	3.720 (4.747)	1.905 (2.908)	1.000
	entropy	11.683 (1.858)	12.113 (2.468)	1.000	9.711 (1.787)	10.435 (2.233)	1.000
	IQR	0.451 (0.214)	0.662 (0.318)	1.000	0.428 (0.297)	0.670 (0.329)	1.000

		α			meta-α		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	0.898 (0.057)	0.847 (0.081)	1.000	0.813 (0.091)	0.778 (0.082)	1.000
	median	0.927 (0.060)	0.868 (0.098)	1.000	0.827 (0.089)	0.794 (0.084)	1.000
	skewness	-1.700 (0.855)	-1.177 (0.934)	1.000	-1.334 (0.873)	-1.069 (0.591)	1.000
	kurtosis	4.274 (4.454)	2.622 (4.939)	1.000	4.541 (6.621)	2.171 (2.511)	1.000
	entropy	9.039 (1.565)	9.684 (2.350)	1.000	9.818 (1.667)	10.658 (1.902)	1.000
	IQR	0.138 (0.063)	0.186 (0.085)	1.000	0.106 (0.053)	0.130 (0.039)	1.000

		D_{app} (x10^{-3} mm^2/s)			meta-D_{app} (x10^{-3} mm^2/s)		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	2.656 (0.095)	2.166 (0.545)	0.100	2.679 (0.245)	2.167 (0.484)	0.125
	median	2.619 (0.136)	2.109 (0.662)	0.195	2.659 (0.236)	2.099 (0.565)	0.125
	skewness	0.104 (0.322)	0.622 (0.948)	1.000	0.067 (0.447)	0.624 (0.694)	1.000
	kurtosis	1.386 (2.193)	1.515 (3.304)	1.000	0.840 (1.237)	0.703 (1.646)	1.000
	entropy	11.388 (2.059)	11.799 (2.689)	1.000	9.754 (2.035)	9.243 (2.741)	1.000
DKM	IQR	0.745 (0.411)	0.830 (0.295)	1.000	**0.345 (0.176)**	**0.686 (0.234)**	**0.048**

		K_{app}			meta-K_{app}		
		Low (n=9)	High (n=19)	p-value	Low (n=9)	High (n=19)	p-value
	mean	0.329 (0.139)	0.468 (0.232)	1.000	0.420 (0.166)	0.557 (0.176)	1.000
	median	0.334 (0.165)	0.479 (0.306)	1.000	0.410 (0.165)	0.543 (0.180)	1.000
	skewness	0.276 (0.383)	0.258 (0.837)	1.000	0.562 (0.537)	0.506 (0.606)	1.000
	kurtosis	0.124 (1.770)	0.575 (1.623)	1.000	1.101 (1.776)	1.299 (2.774)	1.000
	entropy	9.454 (1.753)	9.424 (2.802)	1.000	9.768 (2.017)	9.309 (2.702)	1.000

				overall-meta-D (x10^{-3} mm^2/s)		
IQR	0.406 (0.128)	0.490 (0.168)	1.000	0.175 (0.058)	0.241 (0.069)	1.000

		overall-meta-D (x10^{-3} mm^2/s)		
		Low (n=9)	High (n=19)	p-value
Hybrid	mean	2.402 (0.179)	1.790 (0.468)	0.157
	median	2.427 (0.206)	1.728 (0.524)	0.195
	skewness	**-0.796 (0.983)**	**0.809 (0.905)**	**0.037**
	kurtosis	3.244 (3.470)	2.864 (6.281)	1.000
	entropy	11.709 (1.840)	12.347 (2.058)	1.000
	IQR	0.545 (0.282)	0.700 (0.390)	1.000

The diagnostic performance of the three statistically significant parameters is given in Table 11. Herein, high-grade STSs was considered the positive class and all performance metrics were given by the optimal cutoff of the ROC curve. All significant parameters after model selection differentiated low- from high-grade STSs with an AUC higher than 89%. Highest accuracy (88.9% with 95% CI of 88.2% and 89.6%) in grading STSs was obtained from the skewness of the overall-meta-D and the IQR of the meta-D_{app}. Among all histogram metrics, skewness of the overall-meta-D correctly classified all high-grade STSs (sensitivity of 100%), while the skewness of the meta-D_{slow} showed the highest specificity (88.9%) and a 95% CI of 68.4% - 100%. Accordingly, the skewness of the overall-meta-D had the highest negative predictive value (100%), while the highest positive predictive value was calculated from the skewness of the meta-D_{slow} (PPV = 93.8, 95% CI = 81.9% - 100%).

Table 11 Performance assessment of the three statistically significant parameters using several quantitative metrics from the ROC analysis. A 95% confidence interval (CI) for each metric was estimated using 1000 bootstrap resamples.

	Skewness		IQR
	overall-meta-D (x10^{-3} mm^2/s)	meta-D_{slow} (x10^{-3} mm^2/s)	meta-D_{app} (x10^{-3} mm^2/s)
Accuracy (%)	88.9 [88.2 - 89.6]	85.2 [84.3 - 86.1]	88.9 [88.2 - 89.6]
AUC (%)	90.1 [78.1 - 100]	90.7 [77.7 - 100]	89.5 [75.6 - 100]
Sensitivity (%)	100 [100 - 100]	83.3 [66.1 - 100]	94.4 [83.9 - 100]
Specificity (%)	66.7 [35.9 - 97.5]	88.9 [68.4 - 100]	77.8 [50.6 - 100]
NPV (%)	100 [100 - 100]	72.7 [46.4 - 99.0]	87.5 [64.6 - 100]
PPV (%)	85.7 [70.7 - 100]	93.8 [81.9 - 100]	89.5 [75.7 - 100]
Optimal Cutoff	-0.514	0.153	0.382

5.4 Discussion

In the present study, novel DWI parametric meta-maps incorporating four diffusion models and model selection were developed, providing insights into the characterization of the STSs microstructure, assess therapy induced changes using pre- and post-therapy diffusion MR images and the discrimination of STSs grading. A ground truth validation was initially conducted using areas of known tissue properties (bladder and the gallbladder and iliac arteries). Then, whole tumor model suitability changes associated to treatment received in patients diagnosed with undifferentiated pleomorphic sarcoma (UPS) and spindle cell myxoid liposarcoma (SCML) were investigated and patients treated with a drug combination of tumor necrosis factor antagonist (TNF-a) and melphalan with isolated limb perfusion (ILP) shown a significant reduce in neovasculature and heterogeneous parts of the tumor. Subsequently, a histopathological analysis was performed on a DDLS and an ASPS, two highly aggressive STSs with poor prognosis. DDLS is mostly characterized by major cellular heterogeneity and histomorphology diversity, while ASPS is a very well-structured distinctly compartmentalized highly vascular soft tissue sarcoma [31]. Tumor characteristics were disclosed in the calculated CDM maps of the whole tumor where suitability of the SEM, presenting tissue heterogeneity, exceeded half of the DDLS pixels (52.04%) whereas a vascularization activity was evident in 40.15% of the total pixels within the ASPS. Finally, several histopathology microphotographs representative of previously topographically well-defined areas of the tumor further aligned to the DWI were meticulously reviewed. An excellent association was observed between histology and model suitability results when tumor subregions of histopathologically proven cellularity, neovascularization activity, tissue heterogeneity and complexity were assigned to the MEM, BEM, SEM and the DKM, respectively.

Subsequently, the method was statistically validated in differentiating low from high-grade STSs. A histogram analysis was applied to all maps and meta-maps and results demonstrated that only the IQR of the meta-D_{app} and the skewness of the overall-meta-D and the meta-D_{slow} were statistically significant to differentiate STSs low from high grading, all derived after model selection. In detail, a positive skewness in the cellularity-based parameters of overall-meta-D and the meta-D_{slow} was observed in high-grade STSs compared to negative values in the low-grade group. This finding is in agreement with reports from studies in other cancer types, claiming that increased ADC skewness is associated with a more advanced cancer staging [59]. On the contrary, none of the parameters directly calculated from the traditional, single-model maps yielded an adjusted p-value<0.05. Last, a ROC analysis resulted to AUC values higher than 89%, high values of sensitivity and specificity and narrow 95% CI in most of the cases after conducting a 1000 bootstrap resample test.

Despite the promising results, the relatively small size of the low-grade STSs was the major limitation of this study. However, this is a common problem since STSs are rare tumors and the incidence of low-grade lower than that of high-grade tumors. Another limitation is related to the monocentric setup of our study in the sense that all imaging data were from a single vendor, possibly limiting generalizability and calling for more extensive, multi-centric studies. Additionally, investigating how whole tumor model suitability changes are associated to changes induced from STSs therapy treatment will be explored in the future. Preliminary findings from one of our radiomic studies on T2 images (Chapter 6) have shown promising results in discriminating STSs [60] and a future goal is to embed radiomics in the proposed analysis.

Chapter 6 Application of the DWI quantification into the differentiation of STSs grading

6.1 Scope of the study

The presented study reports aimed to develop an MRI-based radiomics analysis framework and investigate the feasibility of the calculated quantitative imaging features for differentiating low from high grade soft tissue sarcomas (STSs). A total of 22 patients (9 low grade and 13 high grade) who were pathologically diagnosed with soft tissue sarcomas were recruited for the analysis and corresponding T2-weighted MR images were acquired for further post-processing. Tumor delineations were manually traced slice by slice concluding to whole tumor annotated volumes from all enrolled patients. A total of 1165 high-throughput patient-specific quantitative imaging features were exported from each volume using radiomics and evaluated using random forest machine learning classifiers. The overall analysis framework was coupled with feature selection and oversampling techniques to address high-dimensionality dataset issues and the unbalanced ratio between the two examined groups. Validation was performed using repeated nested cross-validation to eliminate overfitting problems and assess the stability of the classification performance. The classifier, using the 13 most important radiomic features selected though training, yielded an accuracy of 0.808±0.135, an area under the receiver operating characteristic curve equal to 0.884±0.154, F1-score of 0.735±0.209, 0.760±0.295 and 0.680±0.304 for precision and recall respectively using independent test sets. To this end, radiomic features from routine MR imaging protocols can provide a strong discriminatory performance between low- and high-grade soft tissue sarcomas.

6.2 Material and Methods

6.2.1 Study Population

Twenty-six patients with soft tissue tumors of variable degree of malignancy underwent MRI examination from July 2015 to February 2019. Exclusion criteria included compromised co-operation, claustrophobia, patients who underwent therapy prior to imaging or between

imaging and surgical excision, and tumors completely suppressed by fat saturation. Hence, a total of 22 patients were eligible for this study and corresponding images were anonymized and transferred to a local database for further post processing. The examination protocol was submitted and approved by the local ethics committee and all patients signed an informed consent for the use of their data for research purposes. All data were anonymized at the hospital premises. Within a short time-interval from radiological examination surgical excision took place and the specimen was transferred to the pathology department for histopathologic analysis to conclude on tissue type and grading. Grading was based on the FNCLCC system (grades 1-3). Group A tumors of grade 1 (low grade) comprised 9 patients with histopathologically proven well differentiated liposarcomas, myxoid liposarcomas, hibernoma (1 patient), desmoid tumor (1 patient) and angiolipoma (1 patient). Group B (high grading tumors of grade 2 and 3) was composed of 13 patients with poorly differentiated liposarcoma, pleomorphic liposarcomas, Ewing sarcoma, leiomyosarcoma and alveolar soft part sarcoma. Indicative MR images depicting low- and high-grade STSs are shown in the following figure.
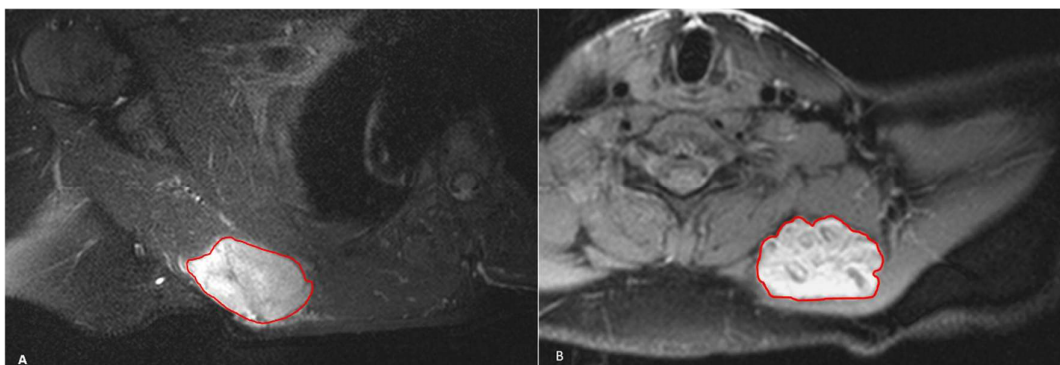
Figure 59 Axial fat suppressed T2w (TE=80ms) TSE MRI in two different patients. (a) Low-grade myxoid tumor from a male patient aged 26, and (b) a high-grade soft tissue tumor (alveolar soft part sarcoma) from a female aged 28. Both are located in the shoulder girdle area.

6.2.2 MR acquisition protocol

Imaging protocol performed at 1.5T scanner (Vision/Sonata hybrid System, Siemens, Erlangen, Germany, Gradient Strength: 45mT m^{-1}, Slew Rate: 200mT m^{-1} s^{-1}) included dual PD to T2w echo (TE1/TE2/TR: 13/80/3250ms, NEX:1) sequence in axial and coronal planes with slice thickness 4mm (20% interslice gap) to ensure complete lesion coverage as well as pre- and post-contrast T1w TSE (TE/TR: 13/498ms) sequences. The number of slices differed between patients depending on lesion size. Given the variable locations of the lesions, the coil selection differed between acquisitions to ensure complete lesion coverage at highest possible SNR. The field of view depended on the lesion size and location and was set to 200x200mm (frequency phase matrix: 320x289) or 400x400 mm (384x320). Spectral fat suppression was used to create fluid

sensitive images with increased lesion conspicuity. Acquisition time was 12min 49s, regardless the variance in the number of slices.

6.2.3 MRI post-processing

The overall framework was developed to address four major steps of radiomics analysis including tumor segmentation, calculation of high-dimensional quantitative imaging features, feature selection, and development of predictive models relying on machine learning techniques (Figure 60).
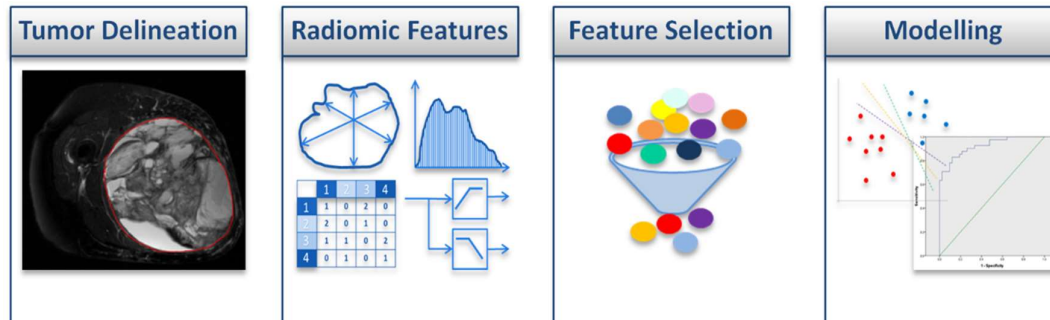
Figure 60 Conceptual overview of the proposed radiomics analysis framework comprising of the four major steps of tumor segmentation, feature extraction, feature selection and model development.

Initially, tumor delineations were manually traced slice by slice on T2w images by an MR physicist with 12 years of clinical experience. ROIs were re-examined by a senior radiologist with 34 years of experience in musculoskeletal MRI. Tumor delineations were performed using a modified version of our in-house developed software presented in Chapter 3, concluding to 22 whole tumor volumes from all enrolled patients. High-throughput patient-specific quantitative imaging features were calculated from all tumor volumes using image analysis techniques to derive a comprehensive spatial and functional view of the examined tissue areas based on intensity, shape and textural characteristics. Specifically, histogram analysis describing the spatial relationships between pixels was applied to each volume resulting to quantitative metrics including mean signal intensity (SI), standard deviation, median, skewness, kurtosis, variance, 10% and 90% percentiles, etc. Volumetric and shape-based features that capture the shape characteristics of the tumor were also calculated (e.g. volume, surface area, sphericity, spherical disproportion, maximum 3D diameter, etc.). Second-order statistics based on grey-level co-occurrence matrices (GLCM), Gray Level Run Length Matrix (GLRLM) and Gray Level Size Zone Matrix (GLSZM) were applied to all delineated tumor volumes providing relevant information about the inter-pixel relationships within each examined region. All the aforementioned techniques were extended to a multiresolution image scaling using wavelet decompositions of level 1 and 2 and the extracted radiomic features were exported across different scales and

frequency directions. A total of 1165 imaging features were calculated using Python software and the Pyradiomics library [61].

Prior to predictive modelling, preprocessing of the extracted radiomic features was employed including feature selection, feature scaling and oversampling. To reduce high-dimensionality of the provided radiomic signature, a univariate feature selection was initially performed and Spearman's rank correlation coefficient (rho) was calculated for each feature with respect to tumor grading. A correlation above 0.4 was considered as significant and the remaining, indicated by Spearman's correlation coefficient, features were normalized using RobustScaler using scikit-learn Python library [62]. RobustScaler was used instead of other widely used techniques (e.g. StandardScaler from scikit-learn) as it is robust to outliers and can operate on features that are not normally distributed. In this study all radiomic features were tested for normality using Shapiro-Wilk test and most of them failed to achieve a p-value higher than 5% indicating a non-normal distribution. A multivariate feature selection and ranking was then performed using minimum redundancy maximal relevance (mRMR) [63]. Feature selection and ranking was performed sequentially using a tradeoff for relevance and redundancy by calculating the mutual information (MI) between the radiomic features and the features with the corresponding outcome. The 100 most highly ranked radiomic features selected from mRMR were then forward as input to the predictive modelling phase. In the current study, an unbalanced ratio was evident between the two classes (9/22 patients with low-grade tumor). To tackle this issue, a synthetic minority oversampling (SMOTE) technique was conducted to increase the size of the minority class by introducing synthetic patients from the corresponding radiomic features [64]. The predictive modelling phase was based on ensemble techniques using Random Forest classifier from scikit-learn Python library. Random Forest (RF) classifier was chosen to discriminate low from high grade STSs as it is less prone to overfitting and generally performs well when applied to high-dimensional low sample size datasets [65].

To assess the generalization performance of the proposed radiomics analysis framework and eliminate any bias occurred during training, validation and testing of the technique, a repeated nested cross-validation (CV) schema was followed as outlined in Figure 61. Although hyperparameter optimization was out of scope of this study since the provided classifier was launched using its default parameters, a nested CV comprising of an inner stratified 3-fold CV and an outer stratified shuffle split (20% for testing and 80% for training) was used for selecting the optimal subset of radiomic features and avoid any overfitting issues. The overall preprocessing of the radiomic features was nested using the inner CV level and the chosen subset of features were finally used to calculate the predictive performance of the model at the outer shuffle split level. In more detail, at the outer level, the overall dataset was divided into training (80% of the overall patients) and testing set (20%) using stratified random sampling.

Subsequently, the training set from the outer level was further divided into 3 inner folds to define and evaluate the preprocessing phase. Each fold was acted as a validation set within the inner CV to evaluate the performance generalizability of the classifier when trained using the remaining inner folds. The whole nested schema was repeated 100 times to iterate through all possible combination of train, validation and test sets.
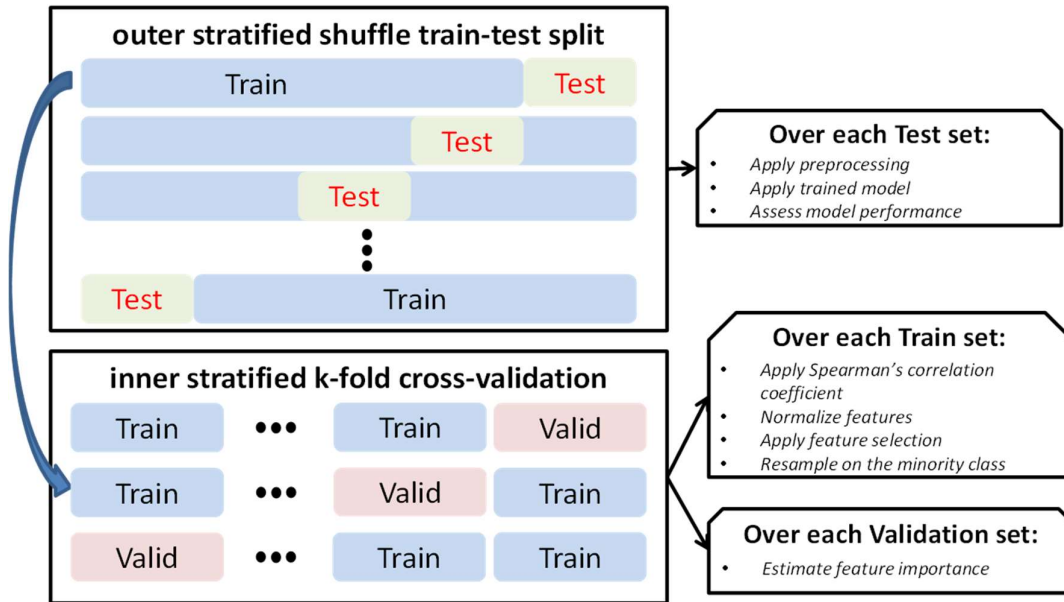
Figure 61 The proposed nested cross-validation schema to assess the generalization performance of the radiomics framework and to select candidate radiomic features for discriminating low from high grade STSs.

6.3 Results

A high-dimensional dataset comprising of 1165 radiomic features was calculated from 22 patients (9 of low-grade STS) and further analyzed using the proposed repeated nested cross-validation (CV) schema. A total of 300 independent training/validation iterations (3 inner folds x 100 outer iterations) were performed to assess the importance of each radiomic feature in discriminating low from high grade STSs and 957 out of 1165 features were reported from the mRMR method as candidate features for modelling. During the inner part of the repeated nested CV, feature importance was quantified using: a) the proportion of the number of times each of the 957 features was selected as input to the classifier out of the maximum number of iterations (300), and b) the feature importance as it is calculated from the RF using Gini impurity as a criterion. Indicative results from the 10 most important radiomic features in terms of a) and b) are presented in Table 12 and 13 respectively.

Table 12 The ten most important radiomic features according to their proportion of the number of times they appear in the classification process during the repeated nested cross-validation.

Radiomic Features	Proportion (%)

wavelet2-LLL_glrlm_LongRunLowGrayLevelEmphasis	80.25
wavelet2-LLL_glrlm_LowGrayLevelRunEmphasis	79.75
wavelet2-LLL_glrlm_ShortRunLowGrayLevelEmphasis	75.75
wavelet2-LLL_glszm_LowGrayLevelZoneEmphasis	71.75
wavelet2-LLH_firstorder_Skewness	65.5
wavelet-LLH_firstorder_Skewness	62.75
original_glrlm_LongRunLowGrayLevelEmphasis	62.5
wavelet2-LLL_glszm_SmallAreaLowGrayLevelEmphasis	61.25
original_glrlm_LowGrayLevelRunEmphasis	61
original_glrlm_ShortRunLowGrayLevelEmphasis	58.5

Table 13 The ten most important radiomic features according to Gini impurity criterion using a RF classifier.

Radiomic Features	Gini Impurity
original_shape_Elongation	0.036
wavelet2-LLL_glrlm_LongRunLowGrayLevelEmphasis	0.033
original_firstorder_Kurtosis	0.031
wavelet2-LLL_glrlm_LowGrayLevelRunEmphasis	0.029
wavelet2-LLL_glrlm_ShortRunLowGrayLevelEmphasis	0.028
wavelet2-LLL_glszm_LowGrayLevelZoneEmphasis	0.018
wavelet2-LLH_firstorder_Skewness	0.016
wavelet-LLH_firstorder_Skewness	0.016
original_glrlm_LongRunLowGrayLevelEmphasis	0.016
wavelet2-LLL_glszm_SmallAreaLowGrayLevelEmphasis	0.013

Concerning criterion a), a subset of radiomic features with proportion values above the 95% percentile of the proportions related histogram was selected as candidate biomarkers.

Accordingly, all features having an importance based on Gini impurity above a specific threshold (above 95% percentile) were also determined. To account both in radiomic feature stability as it is quantified using criterion a) and feature's classification importance as determined by the RF (criterion b) an intersection of the two distinct aforementioned subsets concluded to 13 most highly ranked group of radiomic features (Table below).

Table 14 The 13 most significant radiomic features based on both criteria as outlined from the proportion of the number of times they appear in the classification process and their feature importance level according to Gini impurity criterion from the RF classifier during the repeated nested cross-validation.

Radiomic Features	Proportion (%)	Gini Impurity
original_shape_Elongation	49.5	0.036
wavelet2-LLL_glrlm_LongRunLowGrayLevelEmphasis	80.25	0.033
original_firstorder_Kurtosis	54.5	0.031
wavelet2-LLL_glrlm_LowGrayLevelRunEmphasis	79.75	0.029
wavelet2-LLL_glrlm_ShortRunLowGrayLevelEmphasis	75.75	0.028
wavelet2-LLL_glszm_LowGrayLevelZoneEmphasis	71.75	0.018
wavelet2-LLH_firstorder_Skewness	65.5	0.016
wavelet-LLH_firstorder_Skewness	62.75	0.016
original_glrlm_LongRunLowGrayLevelEmphasis	62.5	0.016
wavelet2-LLL_glszm_SmallAreaLowGrayLevelEmphasis	61.25	0.013
original_glrlm_RunEntropy	48.75	0.011
original_glrlm_LowGrayLevelRunEmphasis	61	0.011
wavelet-HHL_glcm_InverseVariance	43.75	0.011

RF performance was evaluated comprehensively across the 100 outer stratified shuffle split iterations using accuracy, area under the curve (AUC), F1-score, precision and recall. The selected radiomic 'signature' from the inner part of the proposed CV schema comprising of 13 imaging features was examined in terms of its predictive performance using the unseen 100 testing sets. All metrics are reported as mean ± std where 'std' stands for standard deviation. The classifier achieved an accuracy of 0.808±0.135, an AUC equal to 0.884±0.154, F1-score of 0.735±0.209, 0.760±0.295 and 0.680±0.304 for precision and recall respectively.

6.4 Discussion

Our results suggested that the emerging field of radiomics offer a massive amount of quantitative imaging features from T2w MR images from which significant biomarkers for differentiating low from high grade soft tissue sarcomas can be identified after proper analysis. In the current study a total of 13 radiomic features were selected as the most significant imaging features that contribute to the best predictive performance when Random Forest classifier was used for classification. The proposed classifier achieved an AUC of 0.884 ± 0.154 using a repeated nested cross-validation schema comprising of 100 independent testing sets to assess its generalization performance. A comprehensive preprocessing phase including feature selection, feature scaling and oversampling was applied through the training phase to each independent training set and a subset of radiomic features was defined each time to serve as candidate biomarkers for differentiating low from high grade STSs. Two distinct criteria were followed to select the most important set of radiomic features that concluded to the best predictive performance. A limitation of our study was related to the rather limited size of patients recruited for the analysis. Limited sample sizes of high-dimensional imaging features are a general concern when performing radiomics analysis and overfitting problems might easily occur when data is not handled carefully during model training and testing. Additionally, RF classifier and mRMR were chosen by default as the most suitable methods for classification and feature selection respectively. A large parameter grid consisting of several classifiers and feature selection techniques which runs in parallel under a generalized repeated cross-validation framework might potentially yield to a better predictive performance. Lastly, advanced MRI acquisition protocols such as the Diffusion Weighted MRI when further analyzed using radiomics, can provide additional information to the corresponding anatomical information retrieved from the T2w images about the functional and morphological environment of the STSs.

Chapter 7 Conclusion

Considering the importance of accurate pre-operative and non-invasive characterization and the discrimination of soft tissue sarcomas into low- and high-grade tumors, this thesis presented a novel MRI-guided computational framework flexible enough to be applied both to advanced MRI protocols (e.g. DWI of multiple b-values) and high resolution fat-suppressed T2-weighted MR images. The overall scope of this work was to maximize the impact of pre-operative MRI for soft tissue sarcomas and to present a comprehensive quantitative MRI methodology and a radiomic analysis pipeline flexible enough to be deployed and in other fields of oncology. Regarding the use of advanced DWI, the main goal was to quantify diffusion images on a pixel level using 4 different mathematical models, spatially identify and select models that best characterize subregions of the tumor, generate new parametric maps (meta-maps) based on the relative information content of each examined model for the overall region of the tumor, and

perform a statistical analysis on all derived maps to differentiate low from high grading of STSs. A secondary aim was considered in order to examine the application of radiomics in the differentiation of the STSs staging based on anatomical high-resolution fat-suppressed T2-weighted MRI. The conclusions drawn from this study are summarized below:

7.1 Differentiating low from high grade soft tissue sarcomas using post-processed quantitative imaging biomarkers derived from multiple DWI models

An excellent association was observed between histology and the derived composite diffusion model (CDM) maps when tumor subregions of histopathologically proven cellularity, neovascularization activity, tissue heterogeneity and complexity were selected for validating the proposed technique. Additionally, therapy induced changes using a drug combination of tumor necrosis factor antagonist (TNF-a) and melphalan with isolated limb perfusion (ILP) were quantified non-invasively using the calculated pre- and post-therapy CDM maps. The aforementioned results indicated that the proposed analysis workflow can potentially assist pre-operative biopsy especially in heterogenous tumors and to assess non-invasively pre-operative diagnosis of tumor type and grade, which is essential for adequate treatment planning. Meta-parametric maps were subsequently calculated and a statistical analysis demonstrated that only the histogram metrics IQR and the skewness derived from the meta-maps were statistically significant to differentiate STSs low from high grading, showing AUC values higher than 89%, high values of sensitivity and specificity. To this end, a DWI quantification using multiple models and the design of composite parametric maps after model selection are suggested as statistically significant and discriminatory factors in STSs grading.

7.2 High-resolution fat-suppressed T2-based MRI radiomic features for discriminating tumor grading in soft tissues sarcomas

A radiomic analysis pipeline, comprising the calculation of high-dimensional handcrafted features from T2-MRI, a feature selection strategy to select a stable and non-redundant subset of radiomic features, the design of an ensemble machine learning model using random forests, and a careful validation and assessment of the generalization performance of the model using a nested cross-validation schema, resulted to an AUC of 0.884 ± 0.154 and to 13 radiomic features as the most significant imaging features that best predict tumor grading in soft tissue sarcoma. To this end, radiomic features from routine MR imaging protocols can provide a strong discriminatory performance between low- and high-grade soft tissue sarcomas.

References

[1]   V. Y. Jo and C. D. M. Fletcher, "WHO classification of soft tissue tumours: An update

based on the 2013 (4th) edition," *Pathology*, 2014, doi: 10.1097/PAT.0000000000000050.

[2]   A. Leithner, W. Maurer-Ertl, and R. Windhager, "Biopsy of bone and soft tissue tumours: Hints and hazards," *Recent Results in Cancer Research*. 2009, doi: 10.1007/978-3-540-77960-5_1.

[3]   S. M. Fisher, R. Joodi, A. J. Madhuranthakam, O. K. Öz, R. Sharma, and A. Chhabra, "Current utilities of imaging in grading musculoskeletal soft tissue sarcomas," *European Journal of Radiology*, vol. 85, no. 7. Elsevier Ireland Ltd, pp. 1336–1344, Jul. 01, 2016, doi: 10.1016/j.ejrad.2016.05.003.

[4]   A. Crombé *et al.*, "Soft-tissue sarcomas: Assessment of MRI features correlating with histologic grade and patient outcome," *Radiology*, 2019, doi: 10.1148/radiol.2019181659.

[5]   D. F. Papp, A. J. Khanna, E. F. McCarthy, J. A. Carrino, A. J. Farber, and F. J. Frassica, "Magnetic resonance imaging of soft-tissue tumors: Determinate and indeterminate lesions," 2007, doi: 10.2106/JBJS.G.00711.

[6]   G. C. Manikis, E. Kontopodis, K. Nikiforaki, K. Marias, and N. Papanikolaou, "Imaging Biomarker Model-Based Analysis," in *Imaging Biomarkers: Development and Clinical Integration*, L. Martí-Bonmatí and A. Alberich-Bayarri, Eds. Cham: Springer International Publishing, 2017, pp. 71–86.

[7]   A. R. Padhani *et al.*, "Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations.," *Neoplasia*, vol. 11, no. 2, pp. 102–125, 2009, doi: 10.1593/neo.81328.

[8]   D. Le Bihan, E. Breton, D. Lallemand, M.-L. Aubin, J. Vignaud, and M. Laval-Jeantet, "Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging," *Radiology*, vol. 168, no. 2, pp. 497–505, 1988, doi: 10.1148/radiology.168.2.3393671.

[9]   K. M. Bennett, K. M. Schmainda, R. Bennett, D. B. Rowe, H. Lu, and J. S. Hyde, "Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model," *Magn. Reson. Med.*, vol. 50, no. 4, pp. 727–734, 2003, doi: 10.1002/mrm.10581.

[10]   J. H. Jensen, J. a. Helpern, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging," *Magn. Reson. Med.*, vol. 53, no. 6, pp. 1432–1440, 2005, doi: 10.1002/mrm.20508.

[11]   J. H. Hong, W.-H. Jee, C.-K. Jung, J.-Y. Jung, S. H. Shin, and Y.-G. Chung, "Soft tissue sarcoma: adding diffusion-weighted imaging improves MR imaging evaluation of tumor margin infiltration," *Eur. Radiol.*, vol. 29, no. 5, pp. 2589–2597, 2019, doi: 10.1007/s00330-018-5817-0.

[12]   S. Y. Lee *et al.*, "Differentiation of malignant from benign soft tissue tumours: use of additive qualitative and quantitative diffusion-weighted MR imaging to standard MR imaging at 3.0 T," *Eur. Radiol.*, vol. 26, no. 3, pp. 743–754, Mar. 2016, doi: 10.1007/s00330-015-3878-x.

[13]   G. Wu, X. Liu, Y. Xiong, J. Ran, and X. Li, "Intravoxel incoherent motion and diffusion kurtosis imaging for discriminating soft tissue sarcoma from vascular anomalies," *Med. (United States)*, vol. 97, no. 50, Dec. 2018, doi: 10.1097/MD.0000000000013641.

[14]   J. M. Winfield *et al.*, "Utility of multi-parametric quantitative magnetic resonance imaging for characterization and radiotherapy response assessment in soft-tissue sarcomas and correlation with histopathology," *Front. Oncol.*, vol. 9, no. APR, 2019, doi: 10.3389/fonc.2019.00280.

[15]   P. Lambin *et al.*, "Radiomics: The bridge between medical imaging and personalized medicine," *Nature Reviews Clinical Oncology*. 2017, doi: 10.1038/nrclinonc.2017.141.

[16]   V. D. A. Corino *et al.*, "Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions," *J. Magn. Reson. Imaging*, vol. 47, no. 3, pp. 829–840, Mar. 2018, doi: 10.1002/jmri.25791.

[17]   M. B. Spraker *et al.*, "MRI Radiomic Features Are Independently Associated With Overall Survival in Soft Tissue Sarcoma," *Adv. Radiat. Oncol.*, vol. 4, no. 2, pp. 413–421, Apr. 2019, doi: 10.1016/j.adro.2019.02.003.

[18]   A. Crombé *et al.*, "T2-based MRI Delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy.," *J. Magn. Reson. Imaging*, vol. 50, no. 2, pp. 497–510, Aug. 2019, doi: 10.1002/jmri.26589.

[19]   J. C. Peeken *et al.*, "Tumor grading of soft tissue sarcomas using MRI-based radiomics," *EBioMedicine*, vol. 48, pp. 332–340, Oct. 2019, doi: 10.1016/j.ebiom.2019.08.059.

[20]   Y. Zhang *et al.*, "Soft Tissue Sarcomas: Preoperative Predictive Histopathological Grading Based on Radiomics of MRI," *Acad. Radiol.*, vol. 26, no. 9, pp. 1262–1268, Sep. 2019, doi: 10.1016/j.acra.2018.09.025.

[21]   C. Reischauer, R. Patzwahl, D. M. Koh, J. M. Froehlich, and A. Gutzeit, "Non-Mono-Exponential Analysis of Diffusion-Weighted Imaging for Treatment Monitoring in

Prostate Cancer Bone Metastases," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-06246-4.

[22]   D. J. McHugh *et al.*, "Diffusion model comparison identifies distinct tumor sub-regions and tracks treatment response," *Magn. Reson. Med.*, vol. 84, no. 3, pp. 1250–1263, Sep. 2020, doi: 10.1002/mrm.28196.

[23]   M. Inglese *et al.*, "Reliability of dynamic contrast-enhanced magnetic resonance imaging data in primary brain tumours: a comparison of Tofts and shutter speed models," *Neuroradiology*, vol. 61, no. 12, pp. 1375–1386, Dec. 2019, doi: 10.1007/s00234-019-02265-2.

[24]   S. Sanduleanu *et al.*, "Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score," *Radiotherapy and Oncology*, vol. 127, no. 3. Elsevier Ireland Ltd, pp. 349–360, Jun. 01, 2018, doi: 10.1016/j.radonc.2018.03.033.

[25]   J. Song, Y. Yin, H. Wang, Z. Chang, Z. Liu, and L. Cui, "A review of original articles published in the emerging field of radiomics," *European Journal of Radiology*, vol. 127. Elsevier Ireland Ltd, Jun. 01, 2020, doi: 10.1016/j.ejrad.2020.108991.

[26]   N. Papanikolaou, C. Matos, and D. M. Koh, "How to develop a meaningful radiomic signature for clinical use in oncologic patients," *Cancer Imaging*, vol. 20, no. 1. BioMed Central Ltd., May 01, 2020, doi: 10.1186/s40644-020-00311-4.

[27]   D. W. Kim, H. Y. Jang, K. W. Kim, Y. Shin, and S. H. Park, "Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers," *Korean J. Radiol.*, vol. 20, no. 3, pp. 405–410, Mar. 2019, doi: 10.3348/kjr.2019.0025.

[28]   L. Guillou *et al.*, "Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma," *J. Clin. Oncol.*, vol. 15, no. 1, pp. 350–362, 1997, doi: 10.1200/JCO.1997.15.1.350.

[29]   "AJCC Cancer Staging Handbook - From the AJCC Cancer Staging Manual | Stephen Edge | Springer." https://www.springer.com/gp/book/9780387884424 (accessed Sep. 10, 2020).

[30]   K. A. Downes, J. R. Goldblum, E. A. Montgomery, and C. Fisher, "Pleomorphic liposarcoma: A clinicopathologic analysis of 19 cases," *Mod. Pathol.*, vol. 14, no. 3, pp. 179–184, 2001, doi: 10.1038/modpathol.3880280.

[31]   "Enzinger and Weiss's Soft Tissue Tumors - 6th Edition."

https://www.elsevier.com/books/enzinger-and-weisss-soft-tissue-tumors/goldblum/978-0-323-08834-3 (accessed Sep. 10, 2020).

[32] A. C. Guo, T. J. Cummings, R. C. Dash, and J. M. Provenzale, "Lymphomas and high-grade astrocytomas: comparison of water diffusibility and histologic characteristics.," *Radiology*, vol. 224, no. 16, pp. 177–183, 2002, doi: 10.1148/radiol.2241010637.

[33] J. D. Quirk, G. L. Bretthorst, J. J. Neil, and J. J. H. Ackerman, "Measurement of Transmembrane Water Exchange Rate for the in situ Rat Brain," vol. 9, 2001.

[34] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Q. Appl. Math.*, vol. 2, no. 2, pp. 164–168, Jul. 1944, doi: 10.1090/qam/10666.

[35] M. Sasaki, M. Sumi, S. Eida, I. Katayama, Y. Hotokezaka, and T. Nakamura, "Simple and reliable determination of intravoxel incoherent motion parameters for the differential diagnosis of head and neck tumors.," *PLoS One*, vol. 9, no. 11, p. e112866, Jan. 2014, doi: 10.1371/journal.pone.0112866.

[36] D. Le Bihan, "Intravoxel incoherent motion perfusion MR imaging: a wake-up call.," *Radiology*, vol. 249, no. 3, pp. 748–52, Dec. 2008, doi: 10.1148/radiol.2493081301.

[37] J. Yuan *et al.*, "Non-Gaussian analysis of diffusion weighted imaging in head and neck at 3T: A pilot study in patients with nasopharyngeal carcinoma," *PLoS One*, vol. 9, no. 1, Jan. 2014, doi: 10.1371/journal.pone.0087024.

[38] S. Suo *et al.*, "Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods.," *J. Magn. Reson. Imaging*, pp. 1–9, 2014, doi: 10.1002/jmri.24799.

[39] G. C. Manikis, N. Papanikolaou, and C. Matos, "Validating the imaging biomarker: The proof of efficacy and effectiveness," in *Imaging Biomarkers: Development and Clinical Integration*, Springer International Publishing, 2016, pp. 115–122.

[40] A. N. Spiess and N. Neumeyer, "An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach," *BMC Pharmacol.*, vol. 10, no. 1, p. 6, Jun. 2010, doi: 10.1186/1471-2210-10-6.

[41] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," 1998.

[42] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, 1989, doi: 10.1093/biomet/76.2.297.

[43]   L. Sachs, "Angewandte Statistik: Anwendung statistischer Methoden, 8. Auflage.," *Biometrics*, 1998, doi: 10.2307/2533881.

[44]   E. J. Wagenmakers and S. Farrell, "AIC model selection using Akaike weights," *Psychonomic Bulletin and Review*, vol. 11, no. 1. Psychonomic Society Inc., pp. 192–196, 2004, doi: 10.3758/BF03206482.

[45]   P. Kletting and G. Glatting, "Model selection for time-activity curves: The corrected Akaike information criterion and the F-test," *Z. Med. Phys.*, vol. 19, no. 3, pp. 200–206, Aug. 2009, doi: 10.1016/j.zemedi.2009.05.003.

[46]   N. Just, "Improving tumour heterogeneity MRI assessment with histograms.," *Br. J. Cancer*, vol. 111, no. 12, pp. 2205–2213, 2014, doi: 10.1038/bjc.2014.512.

[47]   B. Gaing *et al.*, "Subtype differentiation of renal tumors using voxel-based histogram analysis of intravoxel incoherent motion parameters.," *Invest. Radiol.*, vol. 50, no. 3, pp. 144–52, Mar. 2015, doi: 10.1097/RLI.0000000000000111.

[48]   G. C. Manikis, K. Nikiforaki, N. Papanikolaou, and K. Marias, "Diffusion modelling tool (DMT) for the analysis of diffusion weighted imaging (DWI) magnetic resonance imaging (MRI) data," 2016, doi: 10.1145/2949035.2949060.

[49]   J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imaging*, vol. 31, no. 1, pp. 192–203, Jan. 2010, doi: 10.1002/jmri.22003.

[50]   O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg, "Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters," *J. Magn. Reson. Imaging*, vol. 26, no. 2, pp. 375–385, Aug. 2007, doi: 10.1002/jmri.20969.

[51]   R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016, doi: 10.1148/radiol.2015151169.

[52]   J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, "Radiomics in medical imaging—'how-to' guide and critical reflection," *Insights into Imaging*, vol. 11, no. 1. Springer, pp. 1–16, Dec. 01, 2020, doi: 10.1186/s13244-020-00887-2.

[53]   G. C. Manikis *et al.*, "Scale-Space DCE-MRI radiomics analysis based on gabor filters for predicting breast cancer therapy response," in *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019*, Oct. 2019, pp. 994–1001, doi: 10.1109/BIBE.2019.00185.

[54] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine Learning methods for Quantitative Radiomic Biomarkers," *Sci. Rep.*, vol. 5, Aug. 2015, doi: 10.1038/srep13087.

[55] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, Jul. 2015, pp. 1200–1205, doi: 10.1109/MIPRO.2015.7160458.

[56] "An Introduction to Radiomics: Capturing Tumour Biology in Space and Time | Papanikolaou | Hellenic Journal of Radiology." http://www.hjradiology.org/index.php/HJR/article/view/210/85 (accessed Nov. 17, 2020).

[57] D. Lienard, P. Ewalenko, J. J. Delmotte, N. Renard, and F. J. Lejeune, "High-dose recombinant tumor necrosis factor alpha in combination with interferon gamma and melphalan in isolation perfusion of the limbs for melanoma and sarcoma," *J. Clin. Oncol.*, vol. 10, no. 1, pp. 52–60, 1992, doi: 10.1200/JCO.1992.10.1.52.

[58] D. Reynoso *et al.*, "Neoadjuvant treatment of soft-tissue sarcoma: A multimodality approach," *J. Surg. Oncol.*, vol. 101, no. 4, pp. 327–333, Mar. 2010, doi: 10.1002/jso.21481.

[59] N.-E. Enkhbaatar *et al.*, "MR Imaging with Apparent Diffusion Coefficient Histogram Analysis: Evaluation of Locally Advanced Rectal Cancer after Chemotherapy and Radiation Therapy," *Radiology*, vol. 288, no. 1, pp. 129–137, Jul. 2018, doi: 10.1148/radiol.2018171804.

[60] A. H. Karantanas *et al.*, "T2-based MRI radiomic features for discriminating tumour grading in soft tissues sarcomas," Sep. 2019. doi: 10.36162/HJR.V4I3.301.

[61] J. J. M. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[62] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. Accessed: Nov. 03, 2020. [Online]. Available: http://scikit-learn.sourceforge.net.

[63] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proceedings of the 2003 IEEE Bioinformatics Conference, CSB 2003*, 2003, pp. 523–528, doi: 10.1109/CSB.2003.1227396.

[64]   N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," 2002.

[65]   S. Dutta and A. K. Ghosh, "On some transformations of high dimension, low sample size data for nearest neighbor classification," *Mach. Learn.*, vol. 102, no. 1, pp. 57–83, Jan. 2016, doi: 10.1007/s10994-015-5495-y.