FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Vergleichende visuelle Analyse in einer Kohorte von Brustkrebspatienten

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Nikolaus Karall, BSc
Matrikelnummer 00625631

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Eduard Gröller
Mitwirkung: Univ.Ass Renata Raidou, MSc PhD

Wien, 3. Mai 2018

Nikolaus Karall	Eduard Gröller

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Comparative Visual Analytics in a Cohort of Breast Cancer Patients

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Medical Informatics

by

Nikolaus Karall, BSc
Registration Number 00625631

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Eduard Gröller
Assistance: Univ.Ass Renata Raidou, MSc PhD

Vienna, 3rd May, 2018

Nikolaus Karall	Eduard Gröller

Erklärung zur Verfassung der Arbeit

Nikolaus Karall, BSc
Steingasse 15/8, 1030 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 3. Mai 2018

Nikolaus Karall

Danksagung

Zu allererst möchte ich mich beim Team der Research Division of Computer Graphics der Technischen Universität Wien für das Ermöglichen dieser Diplomarbeit bedanken, besonders bei Meister Eduard Gröller und Renata Raidou, die stets sehr hilfreich und inspirierend war.

Ich möchte mich weiters bei allen bedanken, die mich während meiner gesamten Studienzeit unterstützt haben, in erster Linie bei meinen Eltern Andrea und Martin Karall, die immer da waren, wenn ich Rat gesucht habe und die mich in jeglicher Hinsicht unterstützt haben. Viel Dank ergeht auch an meine Partnerin, Caroline Schuster, die mich mit vielen Ideen inspiriert hat und immer für mich da war.

Außerdem möchte ich mich bei Caroline Radolf bedanken, die diese Arbeit korrekturgelesen hat, und beim Forschungsinstitut für Molekulare Pathologie in Wien, besonders bei Peter Steinlein, Uta Möhle-Steinlein und dem Bioinformatik Team, die mich deren Equipment benutzen ließen.

Acknowledgements

First of all I want to thank the Research Division of Computer Graphics of the Technical University Vienna for making this thesis possible, especially Meister Eduard Gröller and Renata Raidou, who was always very helpful and inspiring.

I also want to thank all the people who supported me during the whole time of my studies, foremost my parents, Andrea and Martin Karall, who always were there when I needed advice and supported me in every possible way. Many thanks to my partner, Caroline Schuster, who inspired me with great ideas and was always there for me.

I want to thank Caroline Radolf for proofreading this thesis, and the Research Institute of Molecular Pathology in Vienna, especially Peter Steinlein, Uta Möhle-Steinlein, Uta Möhle-Steinlein and the Bioinformatics team, who let me use their equipment.

Kurzfassung

Die am meisten verbreitete Krebsart in der weiblichen Bevölkerung in Industrieländern ist Brustkrebs. Um die Sterblichkeit unter betroffenen Frauen signifikant reduzieren zu können, ist eine frühe Diagnose wesentlich und Behandlungsstrategien müssen sorgsam ausgewählt werden. Klinische Forscher und Forscherinnen, die eine geeignete chemotherapeutische Behandlungsmethode auswählen möchten, müssen den Verlauf der Krankheit während und nach der Behandlung analysieren und dabei verstehen, wie unterschiedliche Gruppen von Patienten und Patientinnen auf die jeweils gewählte Behandlungsmethode reagieren. Dies ist aktuell aufgrund der Vielzahl an involvierten (bildgebenden und nicht bildgebenden) Daten eine schwierige Aufgabe, daher sind adäquate Visualisierungen erforderlich. Das Ziel dieser Arbeit ist es, klinische Forscher und Forscherinnen, die an der Analyse des Verlaufs einer Chemotherapie arbeiten, beim Verstehen und Erforschen der Menge an vorhandenen Daten zu unterstützen.

Diese Arbeit präsentiert ein web-basiertes System, das drei Aufgaben zum Erforschen und Analysieren von bildgebenden und nicht bildgebenden Daten von Brustkrebspatientinnen in einer Kohorte, zur Verfügung stellt. Eine Funktionalität für eine Nachfolgestudie einer einzelnen Patientin, eine Funktionalität zum Vergleichen von zwei verschiedenen Patientinnen und eine Funktionalität zum Vergleichen von Patientinnengruppen werden zum einfachen Erforschen und Analysieren der vorhandenen multivariaten Daten bereitgestellt. Zu Beginn wurden die bildgebenden und nicht bildgebenden Daten einigen Vorverarbeitungsschritten, wie beispielsweise Registrierung, Segmentierung und Berechnung von Tumor-Wahrscheinlichkeitskarten, unterzogen. Anschließend wurden sorgfältig einige verknüpfte Ansichten entworfen und implementiert, in denen interaktive Darstellungen verschiedener Aspekte der Daten präsentiert werden, mit deren Hilfe klinische Forscher und Forscherinnen die vorhandenen Kohorten-Daten verstehen und analysieren können. Zum Demonstrieren der Ergebnisse wurden einige Anwendungsfälle durchgeführt und präsentiert, die die Funktionalität veranschaulichen und die Bedeutung der visuellen Analyse-Anwendung zeigen. Durch die Verwendung dieses Systems können klinische Forscher und Forscherinnen optisch die Vielzahl von bildgebenden und nicht bildgebenden Daten einer Patientin erforschen und analysieren, und diese mit anderen Patientinnen in der Kohorte vergleichen, was zuvor mit vorhandenen exploratorischen Anwendungen nicht möglich war.

Abstract

The most common cancer among the female population in the economically developed world is breast cancer. To significantly reduce the mortality among affected women, an early diagnosis is essential, and also treatment strategies need to be selected carefully. Clinical researchers working on the selection of chemotherapy treatment need to analyze the progress of the disease during and after treatment and to understand how different groups of patients respond to selected treatments. Currently this is a difficult task because of the multitude of involved (imaging and non-imaging) data, for which adequate visualizations are required. The aim of this work is to help clinical researchers working on the analysis of the progress of chemotherapy to understand and explore the multitude of data they have.

This thesis introduces a web-based framework realizing three tasks of exploring and analyzing imaging and non-imaging data of breast cancer patients in a cohort. A functionality for single patient follow-up studies (intra-patient study), a functionality to compare two different patients (pairwise inter-patient study) and a functionality to compare groups of patients (groupwise inter-patient study) are provided to enable an easier exploration and analysis of the available multivariate cohort data. To begin with, the imaging and non-imaging data underwent some preprocessing steps, such as registration, segmentation and calculation of tumor probability maps, to make them comparable. Afterwards, we carefully designed and implemented several multiple linked views, where interactive representations show distinct aspects of the data from which the clinical researcher can understand and analyze the available cohort data. A number of use cases to demonstrate the results that can be achieved with the provided framework are performed and they illustrate the functionality and also the importance of the designed and implemented visual analytics framework. Using this framework, clinical researchers are able to visually explore and analyze the multitude of both imaging and non-imaging data of a patient and compare patients within a cohort, which was not possible before with any available exploratory tools.

Contents

CHAPTER 1

Introduction

1.1 Motivation and problem definition

The most common cancer among the female population in the economically developed world is breast cancer [JBC+11]. To significantly reduce the mortality among affected women, an early diagnosis is essential. One of the modalities in breast cancer detection with high sensitivity is Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) [Tur09]. To treat breast cancer, a usual method involves neoadjuvant chemotherapy, which is administered before surgery in order to shrink the tumor so that a less extensive surgery will be required. Different combinations of available chemotherapy drugs can be administered to patients and a lot of trials are taking place, with the purpose to understand the effect of these strategies on different patients and the response to the selected therapy method.

For the purpose of this thesis, a population of longitudinal DCE-MRI studies of 64 patients undergoing neoadjuvant chemotherapy for invasive breast cancer was available from *The Cancer Imaging Archive (TCIA)* [NH16]. All patients had confirmed breast cancer diagnosis based on histopathology of biopsy or surgical excision, and none had prior treatment with chemotherapy, surgery, or radiation. All patients received preoperative chemotherapy in four cycles administered every three weeks (Anthracycline, see Figure 1.1). A subset of 17 patients received an additional weekly treatment after the first four cycles were completed (Taxane, shown in Figure 1.1). Three DCE-MRI scans were scheduled for all patients: MRI_1 before treatment, MRI_2 after one cycle of chemotherapy and MRI_3 after completion of Anthracycline treatment and before surgery or further Taxane treatment. Patients receiving Taxane were scheduled for an additional MRI scan (MRI_4) after completion of all chemotherapy treatment and before surgery. The outcome of the treatment was assessed for each patient at 6-month or 1-year intervals following surgery. Other clinical and endpoint data include patient information and lesion

characteristics (pre-treatment tumor size, histological type, pathologic size, and lymph node involvement) [NH16, CVS+13].

Clinical researchers working on the selection of chemotherapy treatment need to analyze the progress of the disease during and after treatment and to understand how different groups of patients respond to selected treatments. Currently this is a difficult task because of the multitude of longitudinal data (inter-patient data, multitude of patients in cohort studies), the latitudinal data (follow-up time data, multiple time-points of check-up for each patient) and the additional available non-imaging demographic data (e.g. race, breast cancer laterality, age, etc.). Adequate visualization strategies to explore and analyze these available data are required.

Figure 1.1: University of California, San Francisco (UCSF) Pilot neoadjuvant chemotherapy (NACT) study schema. Figure taken from the work of David Newitt et al. [NH16]

1.2 Aim of the work

The aim of this thesis is to help clinical researchers working on the analysis of the progress of chemotherapy to understand and explore the multitude of data they have. We are aiming at answering the following research questions:

- How can differences between patients that were treated in differing ways be explored?
- How can conclusions be drawn while comparing imaging and non-imaging data of patients or groups?

Based on imaging data and demographic data of the population, the outcome of this project should be a framework to explore the images of the different DCE-MRI studies, as well as a representation of the information about a single patient or a group of patients.

As the cohort in the provided data consists of two major groups, one having their treatment stopped after the Anthracycline therapy, the other having received Taxane as additional treatment after MRI_3, the design and specific visualizations is based on the assumption that a user wants to compare either two patients - each out of one of these

groups - or two groups. If more would need to be explored, the conceptual choices of this project would have to be adapted.

For the exploration and analysis of the available data the following three tasks should be realized:

1. **Functionality for single patient follow-up studies** (intra-patient study)
   The functionality for allowing researchers to explore and analyze follow-up data of one patient is needed. The framework should allow a clinical researcher to spot differences in tumor data at diverse time points. Additionally, the non-imaging demographic data should be displayed in interactive graphs. Multiple linked views should be employed.

2. **Functionality to compare two different patients** (pairwise inter-patient study)
   This should assist researchers to point out and compare data of two different patients in the cohort as a preliminary step to understanding the efficiency of the treatment, to be used as a basis for group comparison. Multiple linked views based on the imaging and non-imaging data of the patients will be made available for the users.

3. **Functionality to compare groups of patients** (groupwise inter-patient study)
   Finally a functionality to understand how two different groups of patients respond to selected treatments was implemented, allowing the researcher to compare groups of patients by selecting certain criteria. Adequate comparative visualization strategies will be employed to allow a clinical researcher to explore both imaging and non-imaging data of the two groups of patients.

1.3 Methodological approach

To achieve the aforementioned goals several steps need to be done. The provided image data already had segmentation files containing information about regions affected by the tumor. Nevertheless the image data had to undergo some preprocessing steps to make them comparable, through registration.

The main part of this project is the design and implementation of a visual analytics framework that enables intra-patient follow-up exploration, and also inter-patient exploration (pairwise and groupwise). The framework supports the exploration and analysis of both imaging and non-imaging cohort data.

The outcome is a web-based framework using the JavaScript libraries *Papaya* [RU] to present the image data and tumor segmentations, and *D3.js* [Bos] to make visual analytics possible. For the first part (intra-patient study) the researcher should be able to see the progress of the disease over time during the treatment. Therefore the parts affected by cancer recognized by the different MRI scans were highlighted by a heat map projected onto an MRI record. The second part (pairwise inter-patient study) should

be used assisting the researcher to point out and compare the status of the disease and furthermore indicate how the efficiency of the treatment differs in different patients. In the third part (groupwise inter-patient study) different groups of patients should be explored by the researcher, based on non-imaging demographic data or features from imaging data in the cohort.

Although each patient is anonymized, a lot of demographic data is available for each patient, such as age, race, and breast cancer laterality, but also the kind and results of their therapy. Beside the presentation of the MRI images a visualization of the available non-imaging demographic data is shown in the framework. To achieve this, suitable methods of information visualization were selected and implemented to let the researcher easily recognize similarities or differences in non-imaging characteristics of the patients.

1.4 Requirements

The framework should help a researcher to browse through images of a patient or a group of patients in a simple and intuitive way. The tumor, its variation through time and figures describing its behavior should be visualized to make it possible to recognize all necessary information. Furthermore the non-imaging demographic data of the patient or group of patients should be processed and analyzed easily and prepared to be compared to data of other patients or groups.

A clinical researcher should be able to handle the multitude of the available data and observe both imaging and non-imaging data via multiple linked views in an highly interactive environment. Easy sense-making and analysis of these data should be made possible.

1.5 Contributions

A method for the exploration of the treatment progress of breast cancer in a cohort of patients, which is currently not available, is required. The framework developed in this project provides a web-based visualization system for the exploration and analysis of breast cancer imaging and non-imaging cohort data, as well as the data-driven exploration of the breast cancer therapy, based on these data. Our intended user (a clinical researcher) is able to browse through MRI images of different patients and compare them to each other. The user can perform his exploration and analysis either on a single patient basis, to explore imaging and demographic data of the selected patient, or of two different patients, to compare for instance their tumor properties and treatment results. A third option is to compare two different groups of patients, which were generated by selecting certain features, to obtain understandings of similarities or differences in treatment or non-imaging demographic data and their effect on treatment results.

1.6 Structure of the work

In this thesis **Chapter 1** contains information about problem definition and gives a short overview over the provided data. Furthermore a brief summary of the functionality of the framework and the methods used to achieve this functionality will be given. **Chapter 2** describes the clinical background of the provided data, including information about the procedure that was used, the different MRI scans during the data collection and finally the chapter offers a description of the imaging and non-imaging demographic data of the cohort. **Chapter 3** reviews related work about visual analytics, tumor visualization, especially in breast cancer, and comparative data visualization of both imaging and non-imaging data. In **Chapter 4** the methods that were employed during the design and implementation process are described, including information about methods that were used or created. **Chapter 5** contains details about the implementation of the framework. The results are discussed in **Chapter 6**, where interesting cases that can be investigated with the framework are presented. Finally **Chapter 7** gives an outlook and directions for future work in this field and summarizes the information and conclusions that were derived by this project.

CHAPTER 2

Clinical background

2.1 Breast cancer

Breast cancer is the most common cancer among the female population in the economically developed world. Yearly 23% of new cancer cases are concerned with breast cancer [JBC+11]. According to the *International Classification of Diseases* [Org16], breast cancer is classified as *C50: Malignant neoplasm of breast*. The main risk factors for breast cancer are age (the higher the age, the higher the risk), geographical location (higher risk in developed countries), family history (high risk if a first degree relative is affected by breast cancer), age at first full-term pregnancy (higher risk when age >30), presence of the genes *BRCA1* or *BRCA2* and previous benign diseases [MSD00, KB88].

2.1.1 Classification of breast cancer

Several classifications exist to categorize appearances of breast cancer:

Histological

According to Sunil Lakhani et al. [Lak12] from the World Health Organization (WHO) breast cancer can be classified by its histology. The most frequent types are *Invasive ductal carcinoma, not otherwise specified* (ID, since the fourth edition of the *WHO Classification of Tumours of the Breast* [Lak12] in 2012 this type should be called *Invasive carcinoma of no special type*), *Invasive lobular carcinoma (IL)*, *Tubular carcinoma*, *Cribriform carcinoma*, and *Mucinous carcinoma*. A brief overview of this classification is given by Hans-Peter Sinn and Hans Kreipe [SK13].

TNM-Classification

The TNM classification (shown in Figure 2.1) by Mary Gospodarowicz [GBW17] describes the tumor by the size of the primary or original tumor (T), the number of involved

nearby lymph nodes (N) and the existence of metastasis (M).

Primary tumour (T)	
TX	Primary tumour cannot be evaluated
T0	No evidence of primary tumour
Tis	Carcinoma in situ (CIS; abnormal cells are present but have not spread to neighboring tissue; although not cancer, CIS may become cancer and is sometimes called preinvasive cancer)
T1, T2, T3, T4	Size and/or extent of the primary tumour
Regional lymph nodes (N)	
NX	Regional lymph nodes cannot be evaluated
N0	No regional lymph node involvement
N1, N2, N3	Involvement of regional lymph nodes (number of lymph nodes and/or extent of spread)
Distant metastasis (M)	
MX	Distant metastasis cannot be evaluated
M0	No distant metastasis
M1	Distant metastasis is present

Figure 2.1: TNM classification of tumors. Figure taken from the work of Tony Blakely et al. [BCS12]

Staging

Using the aforementioned TNM classification, the following table (Table 2.1) can be applied according to the Union for International Cancer Control (UICC) respectively the American Joint Committee on Cancer (AJCC) [GBW17]:

Stage	T	N	M
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IIA	T0, T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0, T1, T2	N2	M0
	T3	N1, N2	M0
Stage IIIB	T4	N0, N1, N2	M0
Stage IIIC	any T	N3	M0
Stage IV	any T	any N	M1

Table 2.1: Staging of breast cancer. Table generated according to the work of Mary Gospodarowicz et al. [GBW17]

Receptor status

The existence of three important receptors on the surface of breast cancer cells is another way to classify a tumor: *estrogen receptor (ER)*, *progesterone receptor (PR)*, and *human epidermal growth factor receptor 2 (HER2)*. Each of this receptors either exists (positive) or not (negative) [OEGM09, KAFA14].

2.2 Diagnosis of breast cancer

One of the common imaging diagnostics methods for breast cancer is *Magnetic Resonance Imaging (MRI)* - sometimes in addition with contrast agent injected, so called *Dynamic contrast-enhanced MRI (DCE-MRI)* [Tur09]. Other techniques are *Ultrasound* and *Mammography* [KSL+05], as well as *Positron Emission Tomography/Computed Tomography (PET/CT)* [YLPM+08].

Without using imaging diagnostics, breast cancer can be diagnosed by *clinical examination*. Sometimes breast cancer can be perceivable by visual inspection or palpation. Debbie Saslow et al. [SHO+04] provide practical recommendations for clinical examination.

2.3 Treatment of breast cancer

Depending on different factors, such as age of the patient, stage of the tumor, etc., various kinds of therapy are common practice:

- **Surgery**
  In surgery the tumor, usually conjoined with surrounding tissue, is physically removed from the body. According to the size, the standard types of tumor removal are *lumpectomy*, where only a small part of the breast is removed, *quadrantectomy*, where one quarter of the breast is removed and finally *mastectomy*, the removal of the whole breast [FAB+02].

- **Chemotherapy**
  In addition to surgery, patients often undergo a medication therapy - *chemotherapy*, *hormone blocking therapy*, or *monoclonal antibody therapy*. If the medication therapy is applied *after* surgery to remove possible residues of the tumor, it is called an *adjuvant therapy*. Medication therapies applied *before* surgery (to reduce the size of the tissue to be removed) are so called *neoadjuvant therapies*. In chemotherapy (mainly used for stage II to stage IV tumors) common regimes are *Anthracycline (AC)*, *Taxane* in addition to Anthracycline *(CAT)*, and a combination of *Cyclophosphamide, Methotrexate, and 5-Fluorouracil (CMF)* [HASH10]. The patients in the cohort used for this thesis were all undergoing neoadjuvant chemotherapy, having either only Anthracycline or both Anthracycline and Taxane applied [NH16].

- **Radiation**
  Also to remove possible microscopical tumor cells that remained after surgery it

is usual to apply a radiation therapy [BVD+08]. Radiation therapy consists of radiating tumors with high irradiation doses without harming the healthy adjacent tissues.

2.4 Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI)

Lindsay Turnbull [Tur09] states that "Dynamic contrast-enhanced MRI (DCE-MRI) is an evolving tool for determining breast disease." Contrast agent is administered intravenously to acquire functional information. Using this technique, neoangiogenically induced vascular changes can be examined, which are related to the vascular density in the lesion, according to comparative studies by Michael Knopp et al. [KWS+99], Paul Tofts et al. [TB94] and Gunnar Brix et al. [BSH+97].

To find a rapid way to analyze functional MRI data, Heiko Alfke et al. [AKM+04] created the Software *DynaVision*, based on pharmacokinetic modelling, that allows the user an analysis of the given data within 15 minutes. To determine the region of interest (ROI) in Dynamic contrast-enhanced MRI a common method is to compute the signal enhancement ratio (SER) from pre-contrast and post-contrast sequences. Shandong Wu et al. [WKB+15] describe that changes in signal intensity caused by contrast uptake over time, as well as contrast enhancement kinetics are measured by SER.

As stated in Section 2.2 one way of diagnosis of breast cancer is Dynamic contrast-enhanced MRI (DCE-MRI). A contrast agent gets injected into a tissue and signal intensities are recorded [POM+09]. When a contrast agent was injected, different tissues show different uptake properties. Especially in tumors a rapid enhancement followed by an early wash-out of contrast agent can be seen, contrary to healthy tissue (see Figure 2.2) [Rai17].

According to Renata Raidou et al. [Rai17] and Bernhard Preim et al. [POM+09] several values can be measured or derived based on these enhancement curves:

- wash-in velocity of enhancement (flow of contrast agent entering the blood plasma)
- wash-out velocity of enhancement (flow of contrast agent being diffused out of the blood plasma)
- peak (maximum enhancement intensity)
- time to peak (time between arrival of contrast agent in the blood plasma and the peak)
- area under the curve (integral)

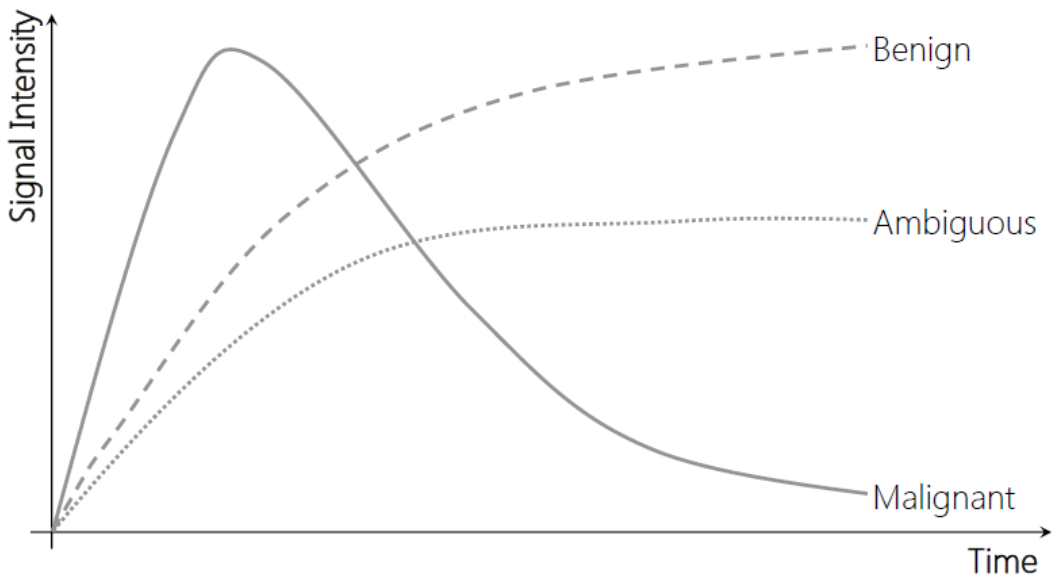


Figure 2.2: Signal enhancement curve for benign, ambiguous, and malignant tissue. Figure taken from the work of Renata Raidou et al. [Rai17]

2.5 Procedure & available data

The procedure of data acquisition in the underlying study of this thesis was to schedule three DCE-MRI scans for each patient and add one extra scan for the subset of 17 patients that additionally received Taxane after completing their Anthracycline treatment (see Figure 1.1). All patients in the cohort had stage II or stage III tumors according to Mary Gospodarowicz et al.'s [GBW17] TNM classification. To determine the region of interest, the signal enhancement ratio method as mentioned above was applied [NH16].

The acquisition matrix for the MRI scans was 256 x 192 x 60 with a section thickness of 2 mm. A minimum of three time points were acquired during each MRI protocol:

- pre-contrast time point (t_0),
- early post-contrast time point (t_1), approximately 2.5 minutes from the start of the contrast injection,
- late post-contrast time point (t_2), approximately 7.5 minutes from the start of the contrast injection.

A minimum of two and a maximum of four MRI scans are available for each patient in the cohort. The first MRI scan (MRI_1) was scheduled before the treatment and was carried out for every patient. After one cycle of chemotherapy with Anthracycline 49 patients

were scanned again (MRI_2) and most of the patients (61 out of 64) were undergoing the third MRI scan (MRI_3) after completion of the chemotherapy treatment and before surgery or further treatment.

After MRI_3 a subset of 17 patients were additionally treated with Taxane. These patients were scheduled for an additional MRI exam (MRI_4) after completing this extra treatment and before having surgery.

The following imaging data was collected out of each MRI scan:

- volume (in cc) of the tumor
- longest diameter of the tumor: LD (in cm)
- Additionally the percentage of tumor volume change between MRI_1 and the last available MRI scan was calculated for each patient.

For each patient, a set of non-imaging demographic data was collected. These data can be divided into three categories:

- **demographic data**
  - age (in years with an accuracy of four decimal places)
  - breast cancer laterality (left or right)
  - race (african-american, asian, caucasian, hispanic or other/not given)
- **tumor-concerning data**
  - estrogen receptor status: ER (positive or negative)
  - progestereone receptor status: PR (positive or negative)
  - human epidermal growth factor receptor-2: HER2 (positive or negative)
  - type of cancer (according to Sunil R. Lakhani [Lak12])
  - clinical assessed size of tumor pre-chemotherapy
  - clinical assessed size of tumor post-chemotherapy
  - pathological size at endpoint (in cm)
- **treatment-concerning data**
  - chemotherapy: Anthracycline, Taxane, or Epirubicin 100 mg/m^2 with 5-Fluorouracil 500 mg/m^2 and Cyclophosphamide 500 mg/m^2 (FEC100)
  - type of surgery: mastectomy, lumpectomy, or quadrantectomy
  - lymph node involvement: LN (positive nodes found after surgery?)
  - follow-up status at last exam (recurrence or no recurrence)

- clinical response: no evidence of disease (NED), $>1/3$ decrease clinical longest diameter LD, $<1/3$ decrease LD or steady disease (SD)/progressive disease (PD)
- time from surgery to recurrence or last follow-up: disease-free survival (DFS) time (in weeks)
- type of recurrence if existent (metastasis or local recurrence)

CHAPTER 3

Related work

This thesis combines approaches of different topics in visualization. Related work of each topic will be discussed in this chapter. First, work concerning *visual analytics* will be considered. Next various concepts for *tumor visualization* will be reviewed to present different approaches that have been introduced concerning this topic. After that we will review related work of *comparative visualization* techniques showing concepts to optically compare various imaging data. Finally methods for the *visualization of non-imaging data* will be presented.

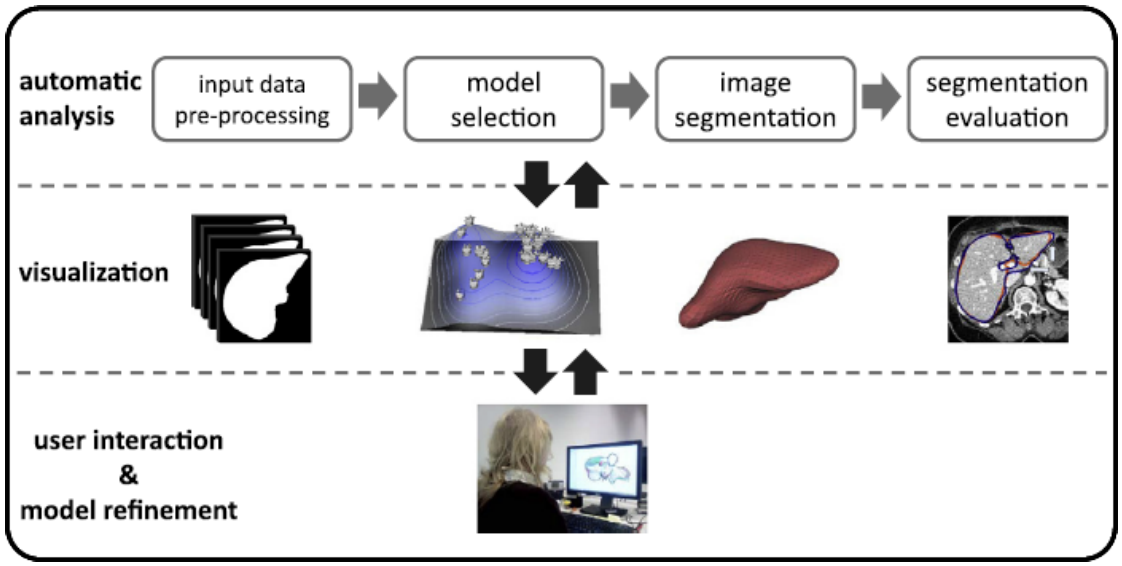


Figure 3.1: Model-based image segmentation, combining interactive visualization with data analysis. Figure taken from the work of Tatiana von Landesberger et al. [vLBK+13]

3.1 Visual analytics

According to Tatiana von Landesberger et al. [vLBK+13] visual analytics combines automated analysis techniques with interactive visualizations to provide the possibility to effectively analyze heterogeneous data sets. Based on *statistical shape models* (SSM) they introduce an approach supporting four stages of model-based image segmentation: *input data preprocessing*, *model selection*, *image segmentation*, and *segmentation evaluation* (See Figure 3.1).

3.2 Tumor visualization

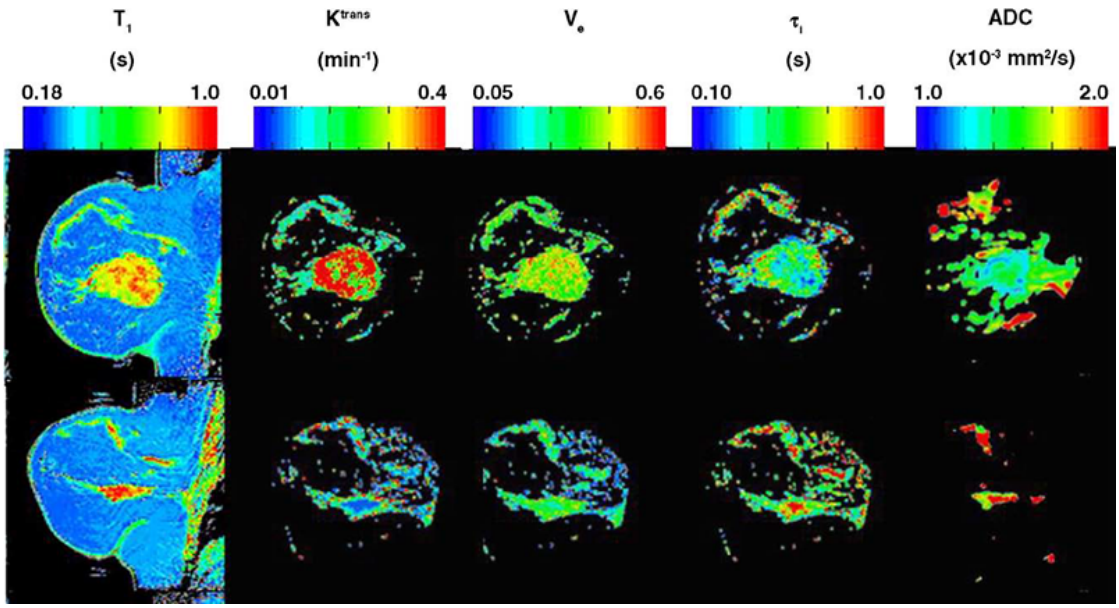
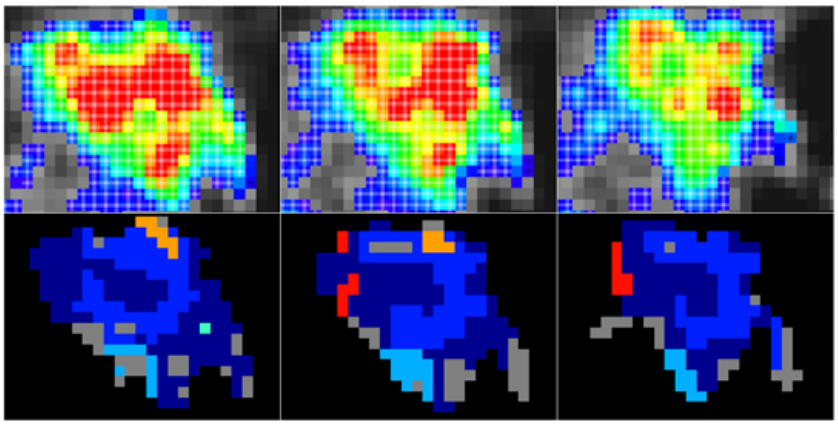
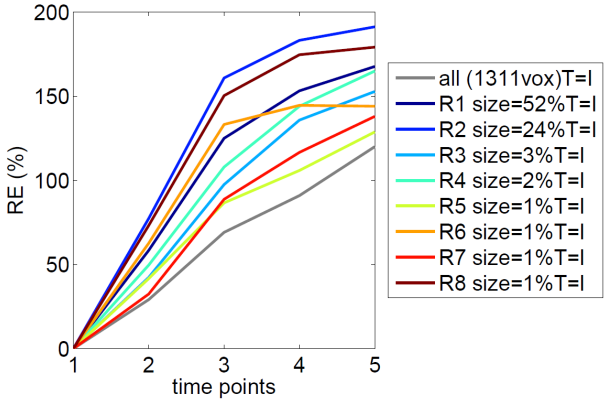


Figure 3.2: Parameter maps of tumor before and after treatment. Figure taken from the work of Thomas Yankeelov et al. [YLC+07]

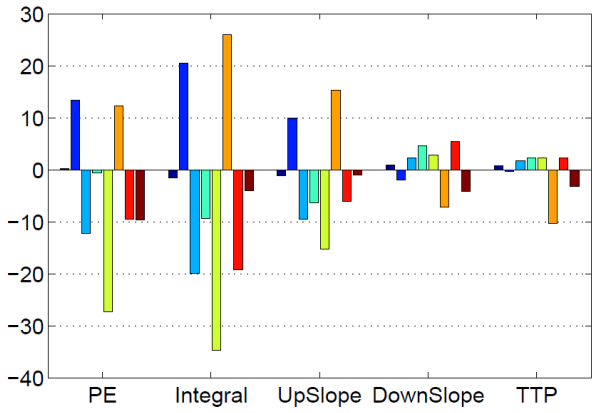
Dynamic contrast-enhanced MRI (as described in Section 2.4) and diffusion-weighted MRI (DW-MRI), are examples for specialized types of conventional MRI, which have shown to be more successful than conventional MRI [YLC+07]. In DCE-MRI contrast agent is applied and MRI scans are acquired before, during, and after the contrast agent passes through the tissue. In DW-MRI water diffusion changes, caused by tumor cells exposed to chemotherapy and radiation, are detected by measuring the apparent diffusion coefficient (ADC). Thomas Yankeelov et al. [YLC+07] used this method to monitor treatment response in breast cancer. They calculated tumor volumes out of qualitative T_1 and DCE-MRI parameter maps and derived pharmacokinetic parameter maps to detect changes in the tumor before and after treatment (see Figure 3.2). The approach of measuring perfusion data in tumors (*tumor perfusion*) and displaying the contrast-agent enhancement-curve is also discussed by Bernhard Preim et al. [PB13].



(a) Glyph-based overview visualization



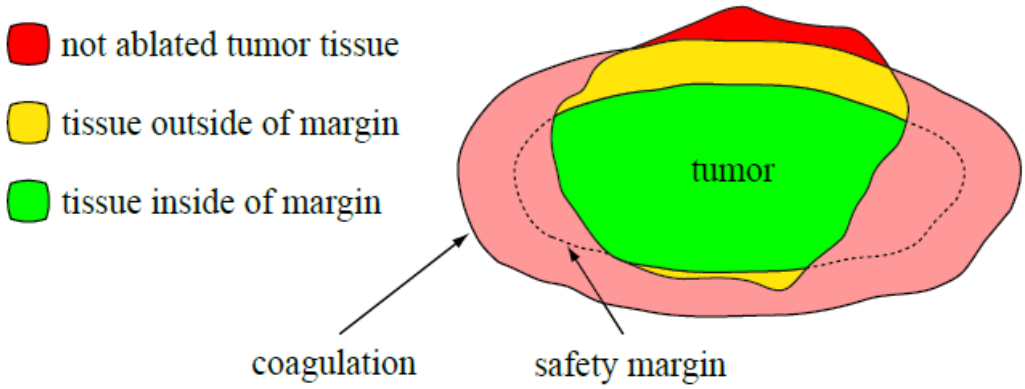
(b) Quantitative diagram (relative enhancement curve)



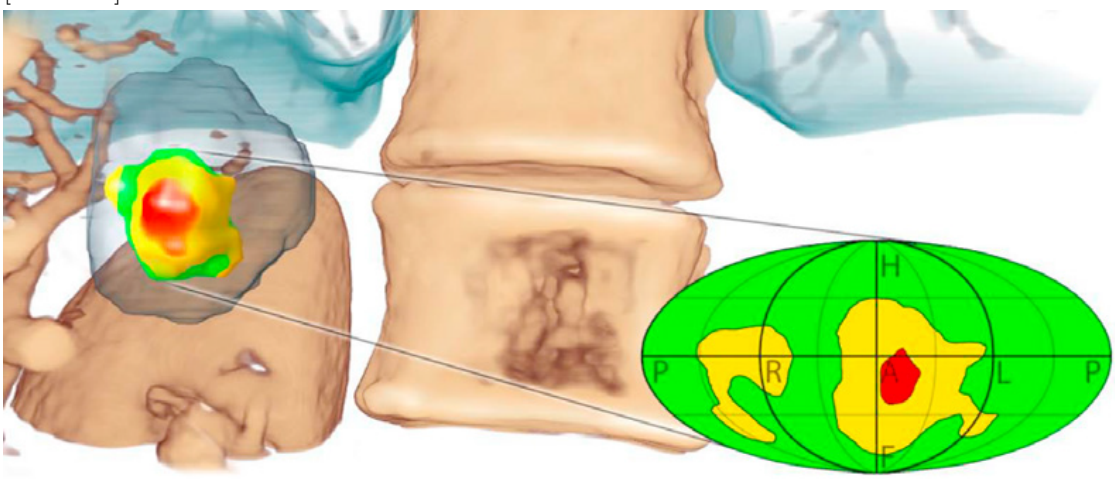
(c) Qualitative diagram (change diagram)

Figure 3.3: Glyph-based overview visualization (a), quantitative diagram (relative enhancement curve) (b), and qualitative diagram (change diagram) (c) for tumor visualization. Figures taken from the work of Sylvia Glaßer et al. [GPTP10]

Sylvia Glaßer et al. [GPTP10] address monitor treatment response in breast cancer as an applicative case of breast cancer exploration and analysis. They present a region merging method for breast cancer tumors to divide them in different regions having similar perfusion characteristics. To visualize the results they use a glyph-based overview visualization, as well as quantitative, and qualitative diagrams, as shown in Figure 3.3.



(a) Color scheme of coagulation zones. Figure taken from the work of Christian Rieder et al. [RWS+10]



(b) 3D visualization of the tumor using tumor maps. Figure taken from the work of Bernhard Preim et al. [PB13]

Figure 3.4: Tumor maps visualization methods

Another approach of tumor visualization is summarized by Bernhard Preim et al. [PB13]: *Tumor maps*, introduced by Christian Rieder et al. [RWS+10], is an approach of displaying the tumor surfaces in 3D. Considering the coagulation (ablation) volume and its safety margin, the segmented tumor is divided into three parts: tumor tissue lying inside the coagulation safety margin, tumor tissue lying inside the coagulation volume but outside

the safety margin, and tumor tissue lying outside the coagulation volume (and therefore it will not be ablated). Examples of this tumor visualization approach can be seen in Figure 3.4.

Ernesto Coto et al. [CGB+05] introduce the *MammoExplorer: An Advanced CAD Application for Breast DCE-MRI*, which combines advanced interaction, segmentation and visualization approaches. Based on DCE-MRI data with two timesteps (pre-contrast and post-contrast timestep), tumor volume data is calculated and displayed, together with enhancement scatterplots and a time-signal curve (see Figure 3.5).

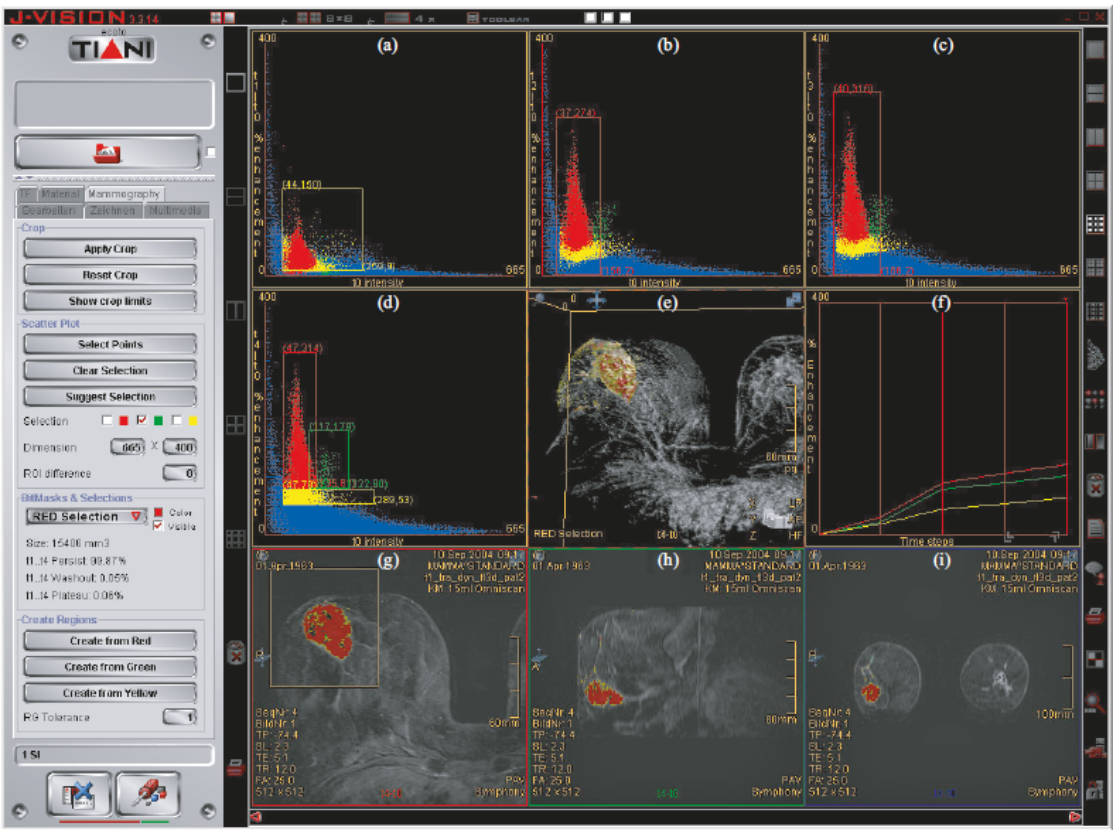
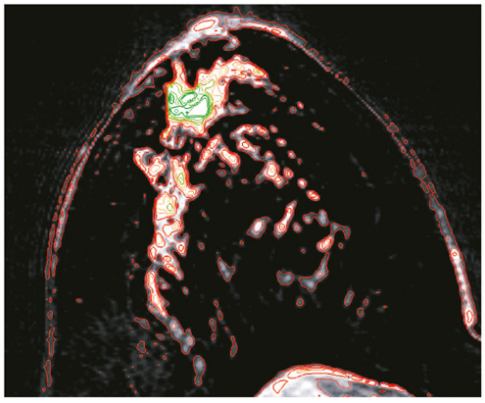
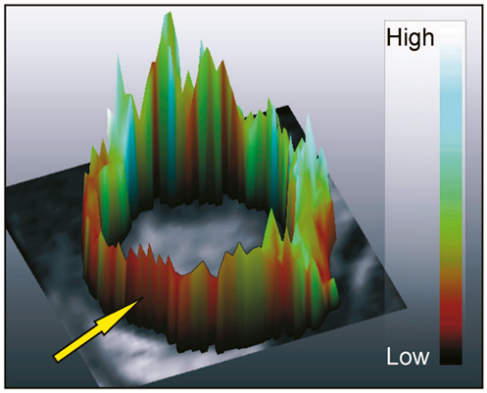


Figure 3.5: Example of *MammoExplorer*. Figure taken from the work of Ernesto Coto et al. [CGB+05]

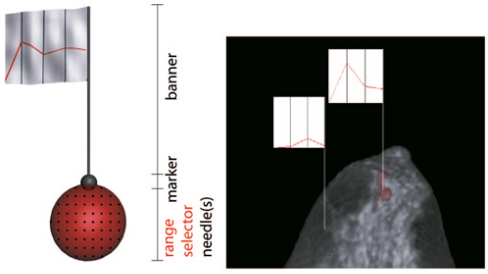
Bernhard Preim et al. [POM+09] give an overview of *basic visualization techniques* and *advanced visualization techniques* for the visualization of perfusion data. As basic techniques they list *cine-movies*, providing the possibility to browse through images like in a movie, *subtraction images*, where the difference in images between two time points is calculated and displayed, and *parameter maps*, where color maps are used to encode values of perfusion parameters. As advanced visualization techniques they present *multiparameter visualization*, a method where various precomputed parameter volumes are combined in one image to enable simultaneous visualization. Further *integrating*



(a) Multiparameter visualization. Image is property of Olaf Konrad, MeVis Research, Bremen. Figure taken from the work of Bernhard Preim et al. [POM+09]



(b) Integrating dynamic information and morphology. Image is property of Stefan Miller, University Hospital, Tübingen. Figure taken from the work of Bernhard Preim et al. [POM+09]



(c) Probing and annotating of perfusion data. Figure taken from the work of Matej Mlejnek et al. [MEV+06]

Figure 3.6: Examples for *multiparameter visualization* (a), *integrating dynamic information and morphology* (b), and *probing and annotating of perfusion data* (c).

dynamic information and morphology using the concept of displaying imaging data together with a visualization of information derived from the imaging data, and finally *probing and annotating of perfusion data*, where visualizations are rendered outside of their spatial location and displayed at certain positions (examples see Figure 3.6).

Renata Raidou et al. [RBV+14] introduce an application called *iCoCooN* (integrated cobweb charts and parallel coordinate plots for visual analysis), an interactive tool to visually explore and analyze DCE-MRI data parameter values. The application combines parallel coordinate plots (PCP) and cobweb charts (CC) to enable the synchronous exploration of different significant attributes of the existing data, and to make the visual analysis of DCE-MRI modeling variations possible. An example can be seen in Figure 3.7.

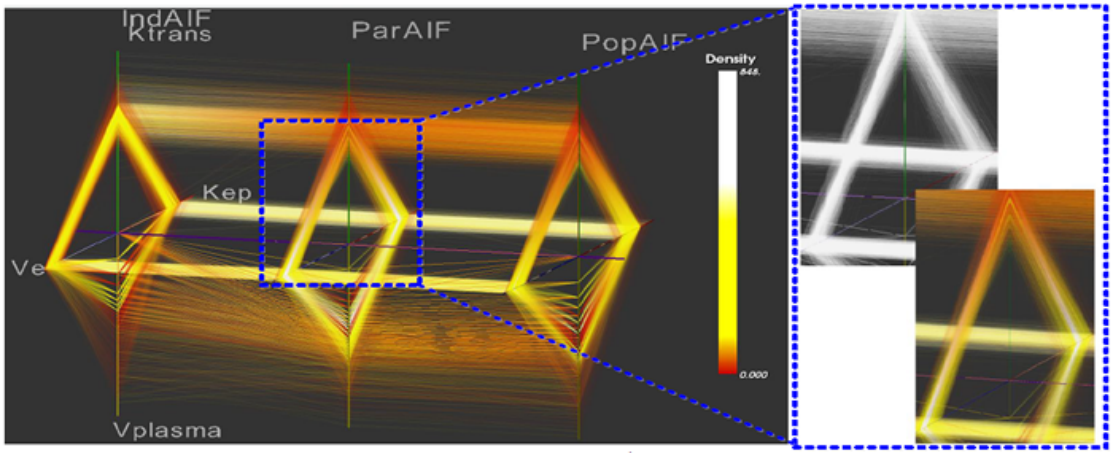


Figure 3.7: Example for an implementation of *iCoCoon*. Figure taken from the work of Renata Raidou et al. [RBV+14]

3.3 Comparative visualization

Christian Tominski et al. [TFJ12] explain that comparative visualization is deduced from the process how humans do comparison on paper: 1) identify information to be compared, 2) arrange this information, and 3) view the result. Kyungyoon Kim et al. [KCK17] state that four fundamental approaches can be defined for comparative data visualization, both when comparing only two data instances and when comparing multiple instances (see Figure 3.8):

- In *juxtaposition* (Figure 3.8 a and e) multiple data instances are simply displayed side-by-side. This approach is often used to compare instances at an overview level (an example can be seen in Figure 3.9a). In many cases interactivity is added, like it is done in linked camera views or linked interactive highlighting. Examples for interactive juxtaposition approaches are shown by Woodring and Shen [WS06] or Johanna Schmidt et al. [SPA+14].

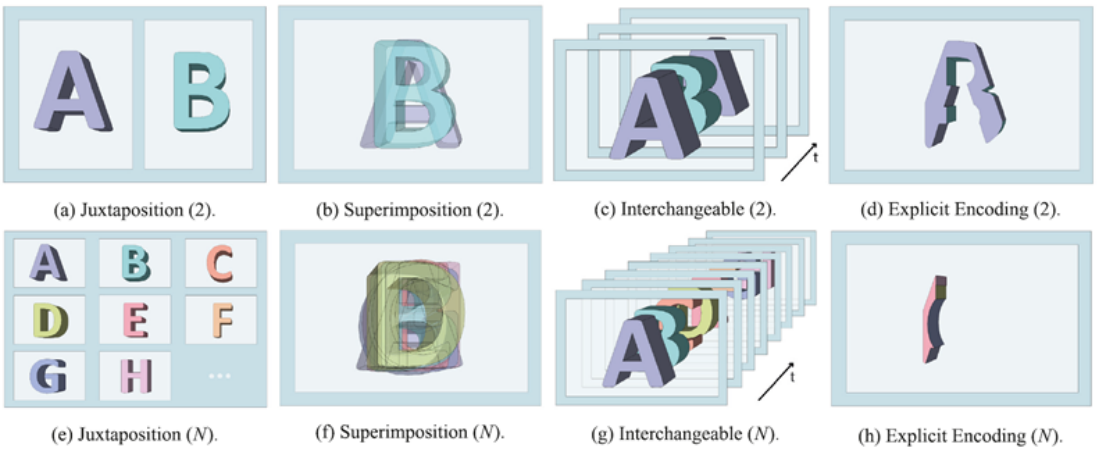


Figure 3.8: Approaches for comparative visualization. Figure taken from the work of Kyungyoon Kim et al. [KCK17]

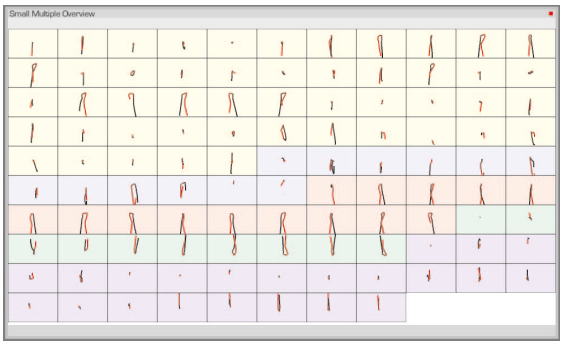
- *Superimposition* (Figure 3.8 b and f) - sometimes called overlay or superposition - shows data instances at the same time in the same place. Often methods like filtering or the use of opacity is used with this approach. Kyungyoon Kim et al. state that the maximum number of data instances to effectively analyze cannot be determined by simple rules. Examples for superimpositions are shown by Penny Rheingans [Rhe96] or Victoria Interrante et al. [IFP97].

- In the *interchangeable* approach (Figure 3.8 c and g) - *temporal superimposition* - only a single data instance is displayed at a time, but the display changes over time, either automatically with animations or with user interaction. Examples for the interchangeable approach are shown by Kyungyoon Kim et al. [KJK+15], who used short animated transitions between the data instances, and Madhura N. Phadke et al. [PPA+12], who smoothly changed the opacity and size of the instances.

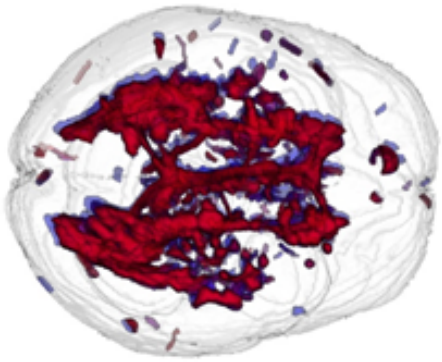
- The last approach, *explicit encoding* (Figure 3.8 d and h), calculates a composition of multiple data instances (e.g. the difference) and displays the result. It is the most common of the mentioned approaches, but does not preserve all data of the several instances. An example for explicit encoding is the *YMCA - Your Mesh Comparison Application* by Johanna Schmidt et al. [SPA+14] (see Figure 3.9c).

However, Kyungyoon Kim et al. [KCK17] also believe that combinations of these approaches, so called *hybrid approaches*, will be the most suitable ones in future work, because they see a hybrid approach as the only one to do an excellent job in comparative data visualization. An example of a hybrid approach can be seen in Figure 3.10.

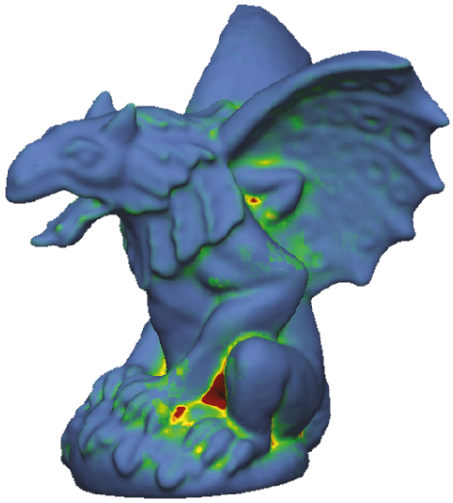
Michael Gleicher et al. [GAW+11] give an overview and comparison of three of these approaches: *juxtaposition*, *superimposition*, and *explicit encoding*. They state that each of



(a) Example for juxtaposition approach. Figure taken from the work of Daniel Keefe et al. [KERC09]



(b) Example for superimposition approach. Figure taken from the work of Melanie Tory et al. [TMA01]



(c) Example for explicit encoding approach. Figure taken from the work of Johanna Schmidt et al. [SPA+14]

Figure 3.9: Examples for juxtaposition (a), superimposition (b), and explicit encoding (c)

these approaches has both strengths and weaknesses and therefore provide a survey for an easy comparison of these strategies. They explored a number of information visualization designs in literature and created a map of the design space, showing the observed application domains fitting to the three approaches or their intermediary categories (see Figure 3.11).

Another approach of comparative visualization is introduced by Johanna Schmidt et al. [SGB13]. They state that side-by-side image comparison or multiple views is limited by human perceptual capacity and limited screen space and therefore do not scale well for a larger number of images. They introduce *VAICo* (Visual Analysis for Image

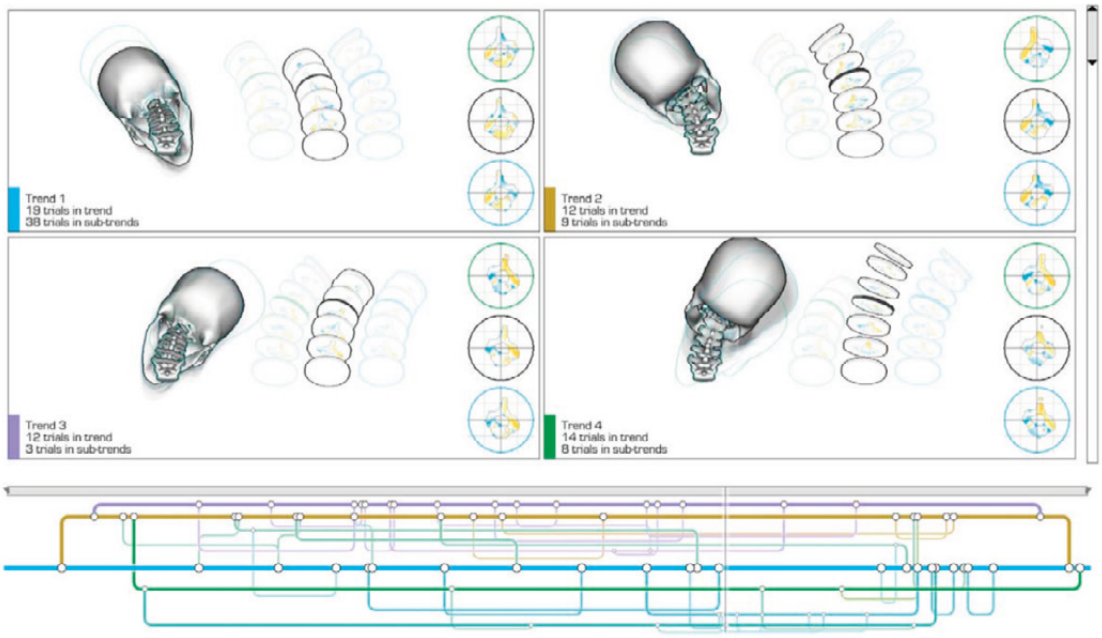


Figure 3.10: Example of a hybrid approach: Trend-centric motion visualization. Figure taken from the work of David Schroeder et al. [SKK+14]

Comparison), an approach for large sets of images, which enables visualizing differences and similarities for detailed analysis, but still preserves contextual information. The three main features of this method are *scalability* (VAICo allows comparison of large sets of images), *focus+context* (an overview of image differences is provided) and *flexibility* (no specific type of image is needed for this approach).

An approach closer to the field of comparative visualization of medical images is provided by Muhammad Malik et al. [MHG10]. They introduce two techniques to compare visualizations of 3D X-ray computed tomography (3DCT) data: The *multi-image view* enables viewing and comparing the whole dataset, the *edge explorer* allows edge comparison in a dataset series. They provide a system that uses algorithms to handle the large size of the datasets in a fast way and does an automatic scaling according to the number of active datasets.

3.4 Visualization of non-imaging data

Martijn Steenwijk et al. [SMvB+10] state that research in medicine is usually hypothesis driven and only a few measurements are analyzed at one time point. Through visual analytics it becomes possible to analyze datasets with multiple parameters and time points. They introduce a tool called *Papilio* consisting of *PrePap* for preprocessing and *VisPap* for visualization, which makes it possible to explore both imaging and non-imaging

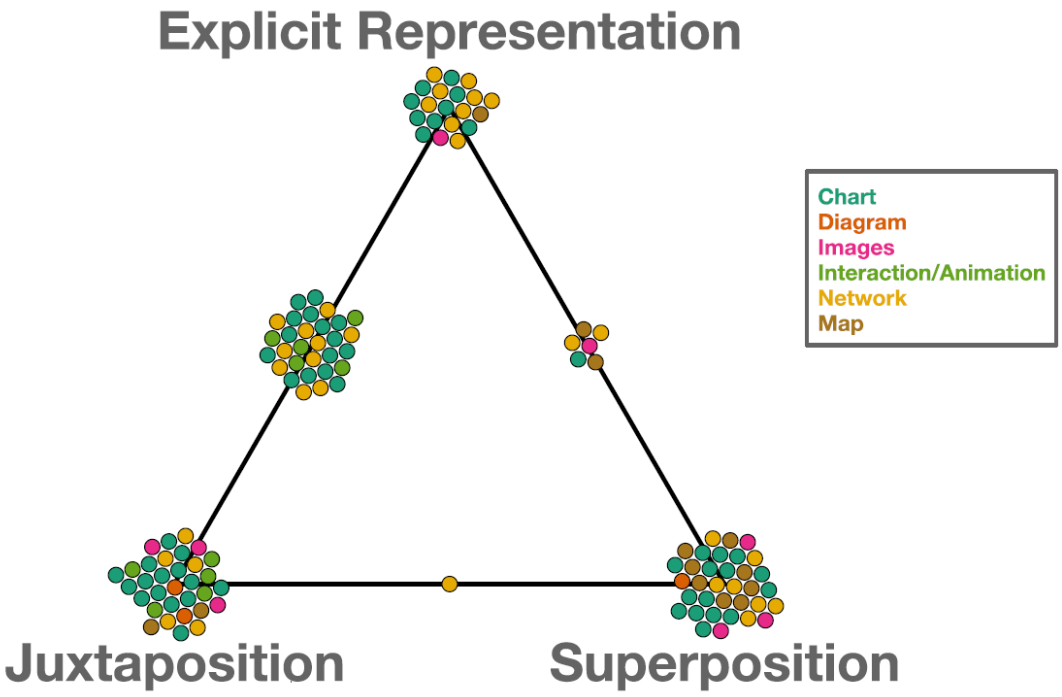


Figure 3.11: Design space showing application domains fitting to the approaches *juxtaposition*, *superposition*, and *explicit representation* or their intermediary categories. Figure taken from the work of Michael Gleicher et al. [GAW+11]

data of a cohort.

A way to visually explore and analyze heterogeneous data in a cohort of patients is described by Paolo Angelelli et al. [AOH+14]. They introduce model constructing data-cubes, where categorical values are stored as dimensions, and numerical values are stored as measures. With the help of these data-cubes a visualization tool was implemented, which can be seen in Figure 3.12.

The usual way to analyze non-imaging data in a cohort, is to manually extract them from the dataset and apply statistical operations to make mathematical plots possible [AOH+14]. Chris Stolte et al. [STH02] introduce *Polaris*, a tool that makes it possible to visualize data based on large multidimensional databases (so called *OLAP cubes* - On-Line Analytical Processing). They state that the user interface of Polaris and the fast generation of various displays make it possible to explore and analyze the data easily. An example of data visualization with Polaris is shown in Figure 3.13.

This chapter gives an overview of related work concerning visual analysis, tumor visualization, comparative visualization, and visualization of non-imaging data. More related work will be discussed in the coming chapters, where we will address the design of the visualization strategies we used in this thesis.

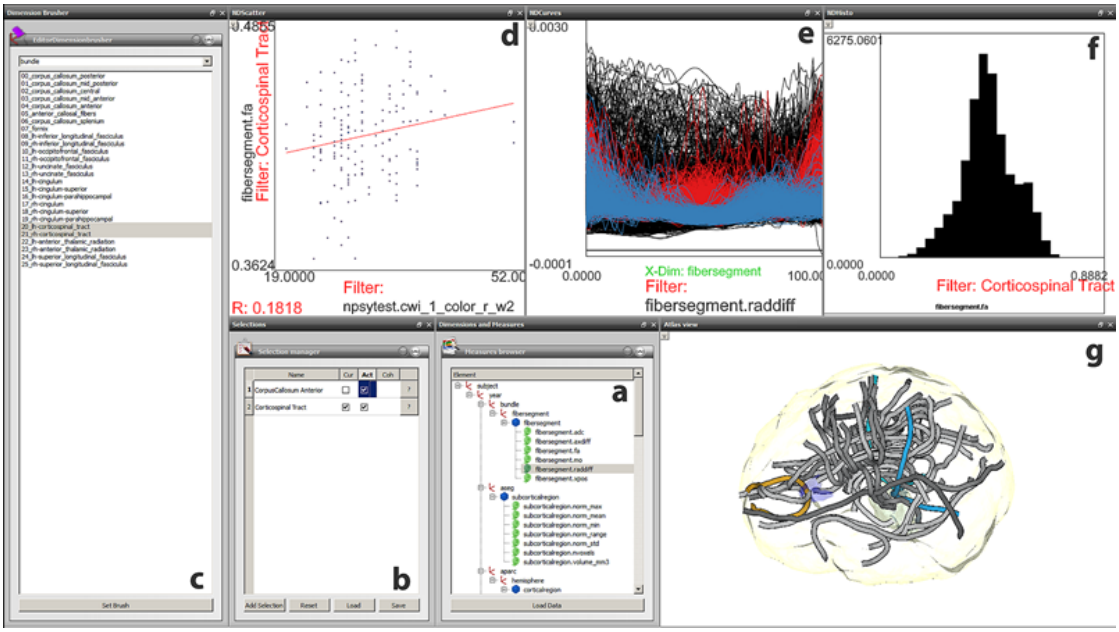


Figure 3.12: Prototype of a model using data-cubes to provide visual analytics. Figure taken from Paolo Angelelli et al. [AOH+14]

In this thesis we present a combination of different topics, which were not combined before. We use tumor visualization techniques, as well as visualizations of non-imaging data through a comparative approach to make visual analysis of patient data within a cohort possible. This method will be described in the following chapters.

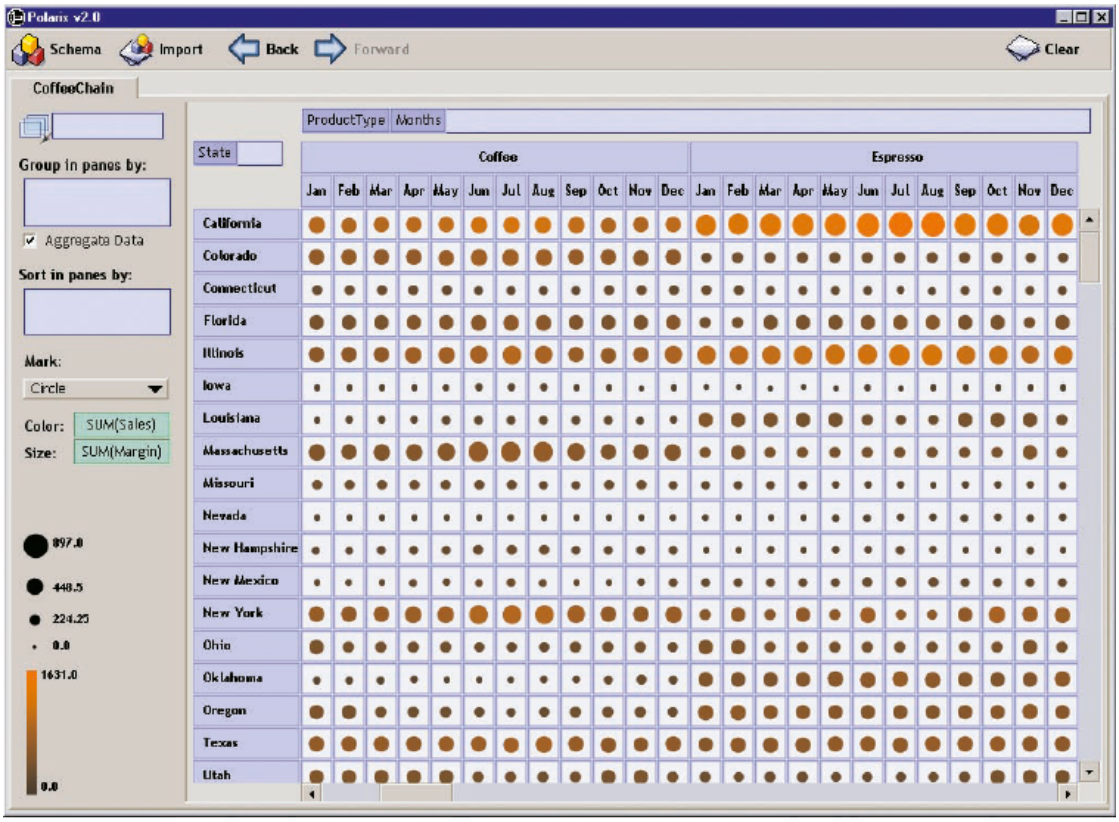


Figure 3.13: Example of data visualization in Polaris. Figure taken from Chris Stolte et al. [STH02]

CHAPTER 4

Design of a visual analytics system in a cohort of breast cancer patients

4.1 Workflow

The scope of the framework was to help clinical researchers analyze both imaging and non-imaging data of the available cohort data. Namely, each individual patient should be explored and also functionality to compare different patients should be provided. Additionally the possibility to compare two groups of patients should be provided. The researcher should be able to select the desired patient, patients or groups in an easy and intuitive way by arranging various criteria. On the one hand, the presented solution should display to the user the preprocessed imaging data, so an analysis and assessment of the available imaging data is made possible. On the other hand, information derived from the non-imaging data should be displayed in appropriate graphs and eventually combined with information from the imaging domain. We decided to limit the comparison to two patients or two groups, because the main scope of the cohort data was to compare only the available treatment procedures - those *without* Taxane to those *including* Taxane. For a more generalized application, the hereby proposed visualization strategies should be adapted adequately.

To achieve this, several steps had to be done. As shown in Figure 4.1, the imaging data of a patient had to undergo a preprocessing procedure, which involves registration and segmentation of the tumor data. From the results, suitable visualizations had to be designed. The non-imaging demographic data had to be structured by their affiliation to patient, tumor, or treatment and in some cases arranged into groups (i.e. age, volume change, etc.), to enable easier querying. Afterwards all imaging and non-imaging data had

to be visualized adequately by visualization tools or graphs and displayed comparatively to each other, to facilitate data exploration and analysis. Due to the multitude of information that we needed to visualize, we followed the well-known concept of multiple linked views. In this concept, multiple views are generated through interaction of the user, and are linked together to make further exploration possible [Rob05].

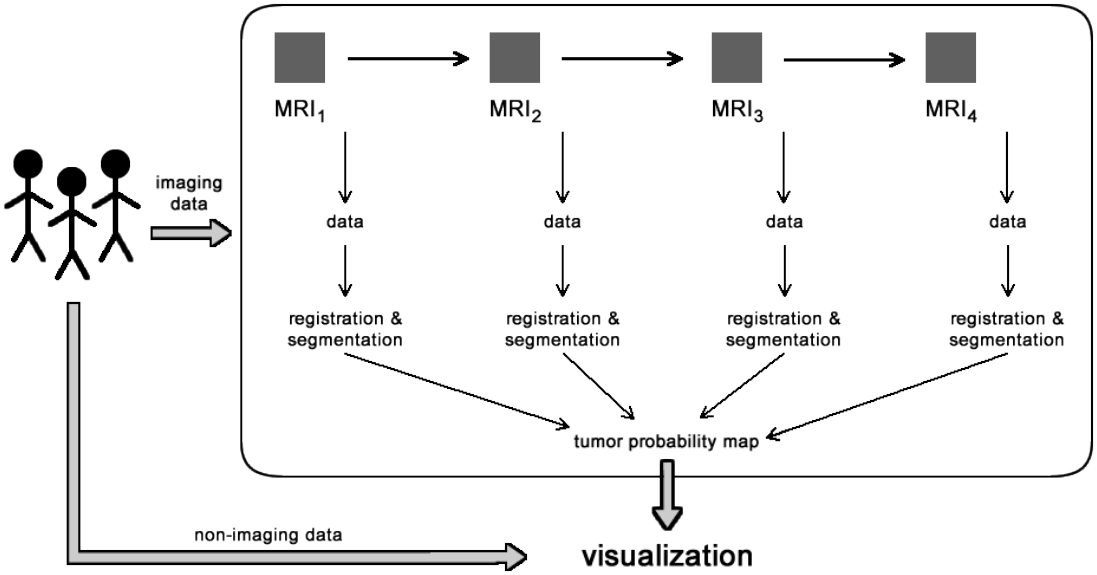


Figure 4.1: Data acquisition and processing workflow of the data available for the purpose of this thesis.

4.2 Preprocessing

At first we reduce the provided imaging data to only the relevant parts we need for this thesis by removing unnecessary data, such as breast tissue segmentations without tumor data, clipped MRI data, or additional mammography data, which were out of the scope of this thesis. The remaining relevant MRI data first undergo a registration process.

"Image registration is the process of aligning images so that corresponding features from different image data can easily be related" [HH01]. To make different MRI scans within the follow-up study of one patient and different MRI scans also from different patients comparable, registration is necessary. The predominant goal of registration in our case is to adapt the shapes of the different breasts so regions with high probability of tumor appearance can be marked. The reasons for that will be discussed in Section 4.4. As there is no hard tissue such as bones in a breast, a rotation and scaling of the breasts are possible and reasonable. We are aware that there is a deformation during the registration process, which can bring uncertainty into an analysis. However, this is considered to be out of the scope of this thesis.

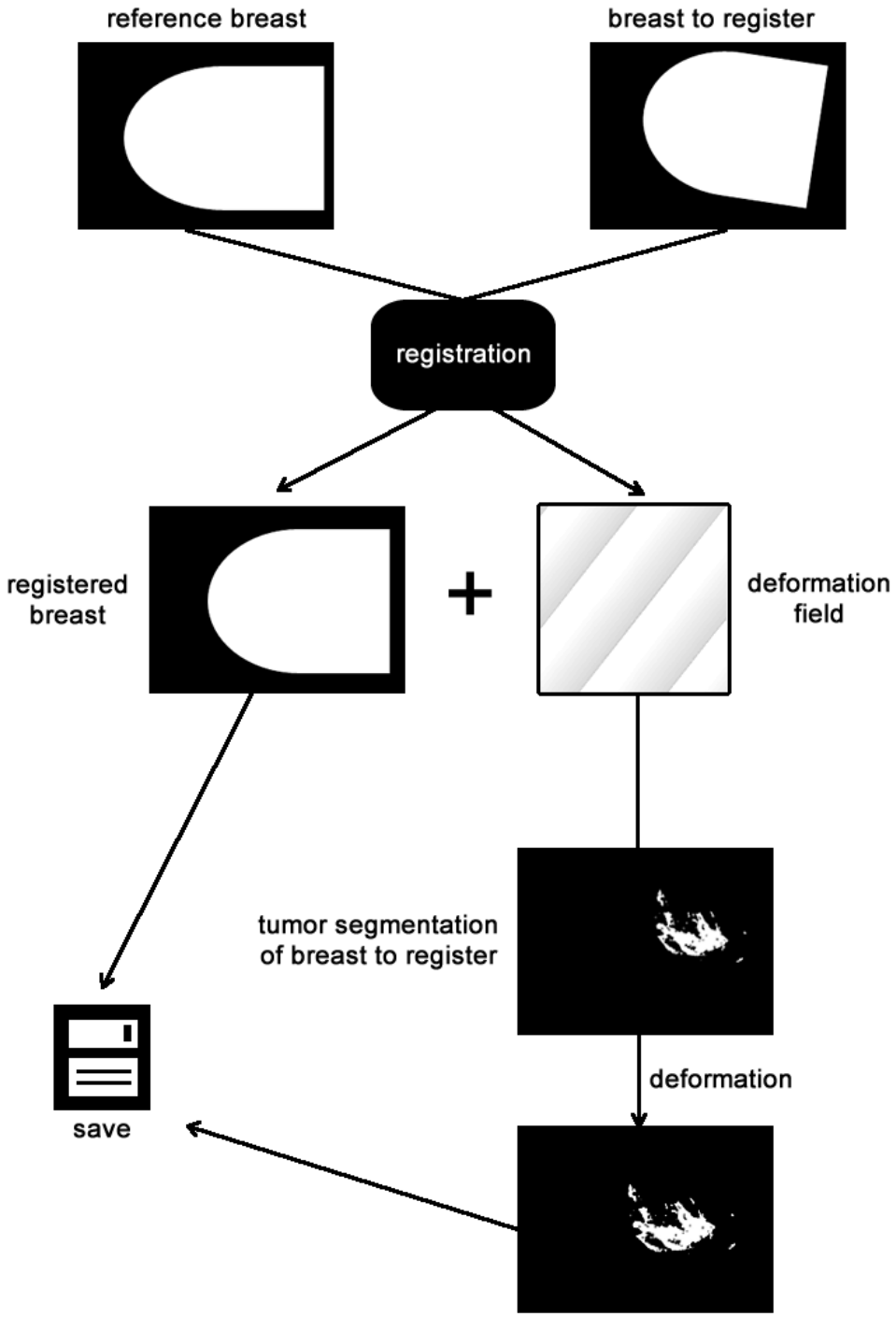


Figure 4.2: Scheme of MRI registration and applying the deformation field onto the segmentation.

Registration algorithms providing these features are used. In most of the cases registration with *affine transformation* and *nearest neighbor* as interpolation algorithm leads to suitable results, nevertheless this procedure does not succeed for all breast scans. In some cases affine transformation deforms the input image too much and only parts of the breast are visible in the outcome, so some of them have to be registered with *rigid transformations*. While a rigid transformation only provides rotation and translation of the image to be registered, an affine transformation additionally includes anisotropic scaling and shearing [PB13]. Visual assessment of the registration results is used to decide which outcome is the best, but a more quantitative assessment would be desirable in the future.

Segmentation data of the tumors were already provided [NH16], but since the segmentations are based on the original MRI scans, we have to adapt them according to the deformation that results from the MRI scan registration. The output of the registration procedure is the registered MRI scan as well as a deformation field, which we then apply on the corresponding segmentation. The deformation field contains information about displacements of the concerning voxels in the deformed MRI scan. A scheme of the registration process and the process of applying the deformation field onto the segmentation is shown in Figure 4.2.

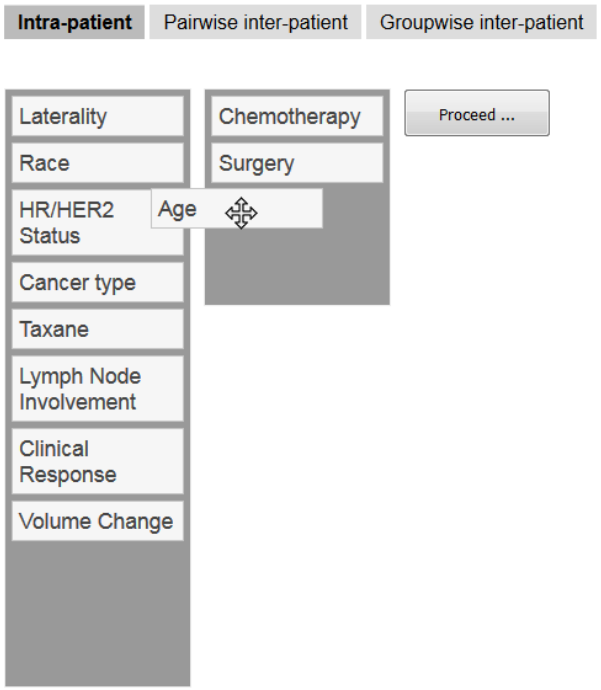


Figure 4.3: Selection of criteria to aid the users choose patient, patients, or groups

Also some non-imaging demographic data have to undergo some preprocessing steps. We divide these data into three groups, based on their belonging to *demographic data*,

tumor-concerning data, or *treatment-concerning data*, according to Section 2.5. This is done to enable easier querying of the data of the population. Furthermore not all data are divided into categories or have too many different values (such as age or cancer type), so they have to be put in groups (like *age 25-35*, *age 35-45*, ...) to make them easily comparable. This binning is done to reduce minor observation errors and to reduce a big number of categorical values within one property [vBH89].

4.3 Initiation of the data exploration and analysis

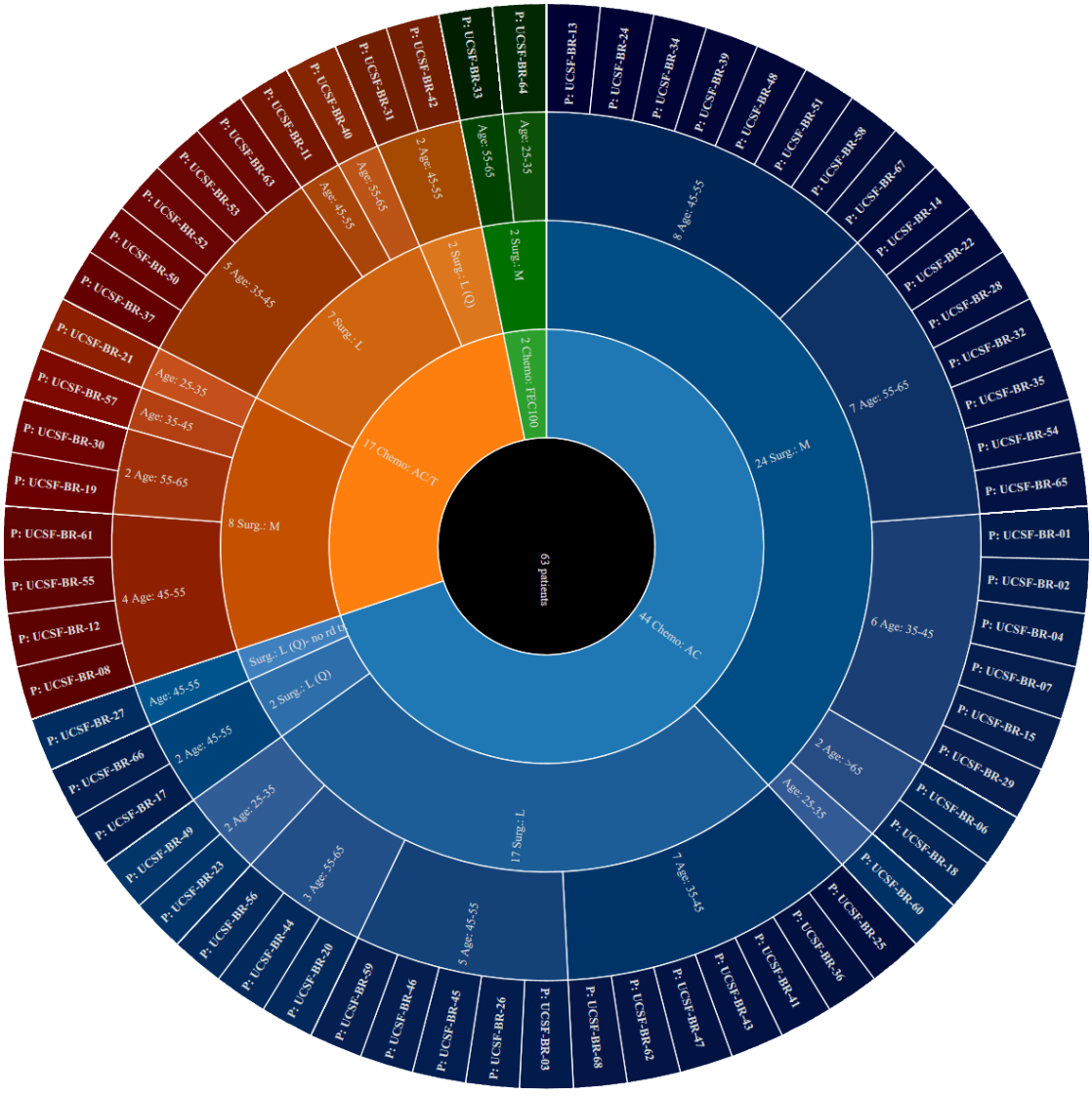


Figure 4.4: Interactive wheel showing subset of patients or groups

Since there are lots of criteria to select a patient for an intra-patient study, or to select two patients or groups for inter-patient studies, we created a visual representation to simplify this procedure for a researcher. After having selected certain criteria (as shown in Figure 4.3) an interactive wheel is created to visually provide an overview on the desired subset of patients or groups (see Figure 4.4), from which a researcher can now choose a patient (or two patients or groups, depending on the task) to proceed with. The center of this wheel represents the total cohort, the segments of the circle represent the selected criteria in the order first criterion closest to the center. In case the researcher has decided to use the intra-patient or the pairwise inter-patient study, the patients are added on the edge of the wheel. To easily recognize which major category (first selection criterion) the subordinated groups and patients belong to, these groups and patients are marked with a similar background color. The specification of this coloring is done by the use of the *HCL* color space, which is included in the D3 JavaScript library.

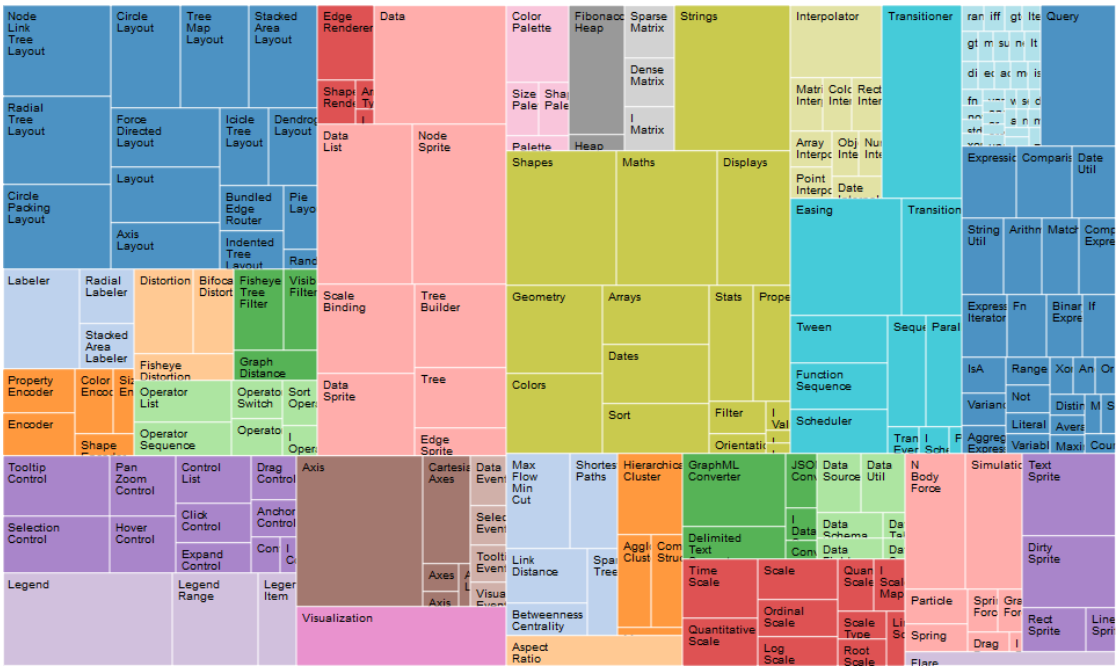


Figure 4.5: Example of a Treemap representation. Figure taken from the work of Mike Bostock [Bos17c]

We decided to use this way of choosing a patient, patients or groups, because of its intuitivity and interactivity. A user quickly recognizes the correlation between various objects either by their position or their color. Furthermore it is possible to hide a group, which may not be needed, just by enlarging the desired group. Therefore a segment simply has to be clicked on and it becomes the center of the circle. Different approaches would have been a *Treemap* [SW01] (see Figure 4.5), or a simple file tree (see Figure 4.6) but John Stasko et al. [SCGM00] stated that circular layout methods, such as the wheel used in this thesis, provide an easier perceivable view of the subcategories. Also the

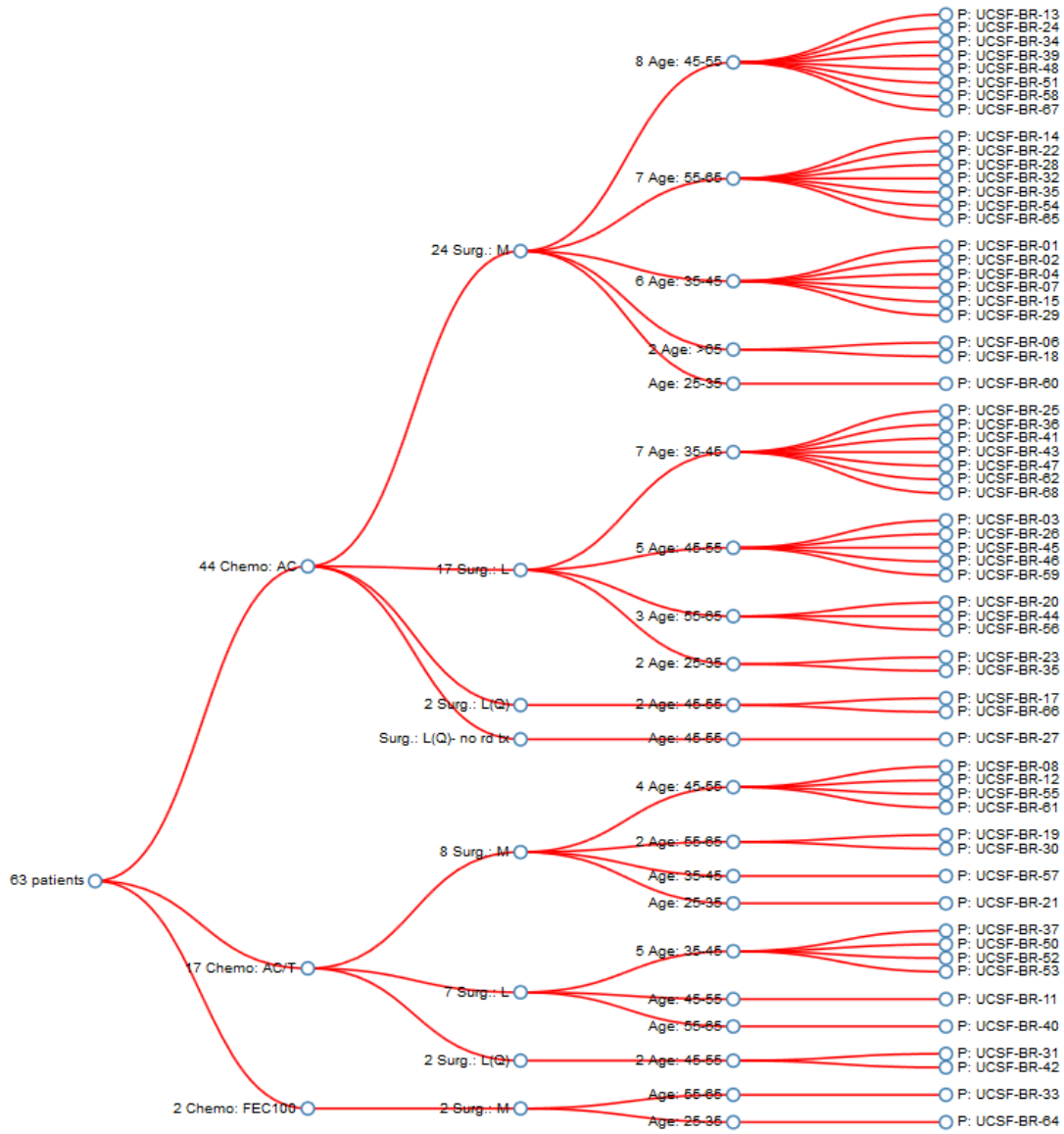


Figure 4.6: Example of a file tree representation. The figure represents the selection of patients, that were used in the interactive wheel (Figure 4.4), represented as a file tree based on the work *Simple tree* of David Richard [Ric17]

size of various groups and their relation to a parent group are easier perceivable in the selection wheel than in a file tree. Furthermore a file tree is not scalable for a higher number of patients in the cohort and will create clutter on the screen to list all the attributes one by one.

From this point on, the intended user can follow different pathways, based on the three main tasks that clinical researchers would be interested in performing.

1. **intra-patient study**, providing functionality for single patient follow-up studies,
2. **pairwise inter-patient study**, providing the functionality to compare two different patients and finally
3. **groupwise inter-patient study**, providing the functionality to compare groups of patients.

The three functionalities for fulfilling the above mentioned exploratory tasks are described in the following sections.

4.4 Intra-patient study

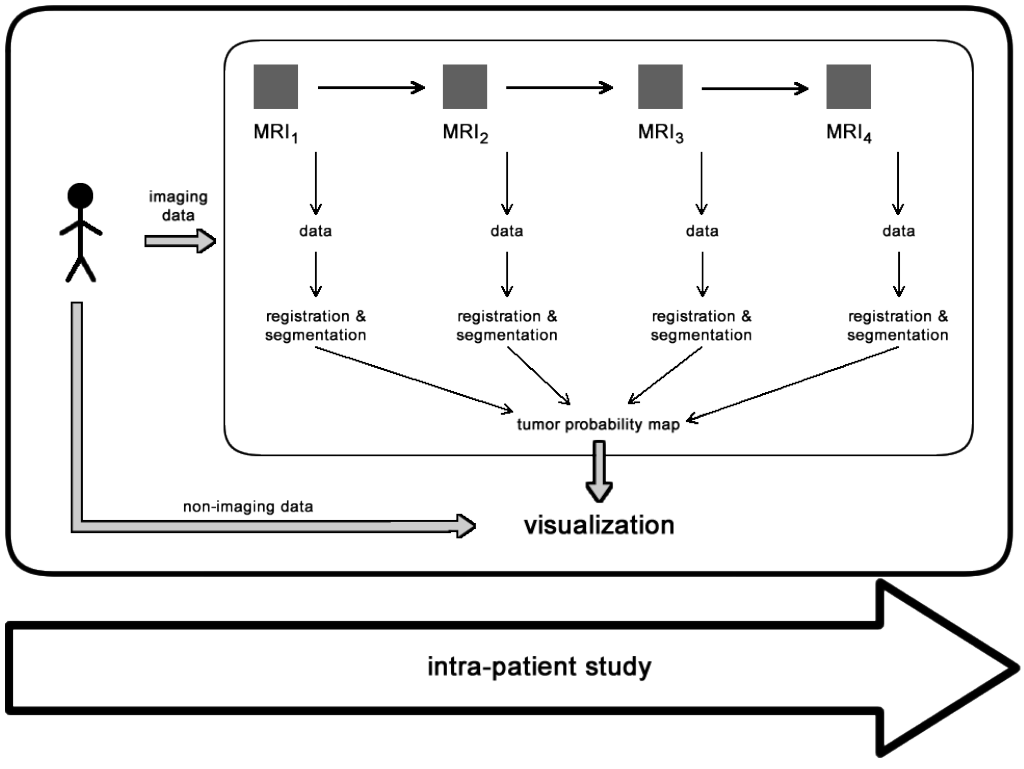


Figure 4.7: Workflow of intra-patient follow-up study

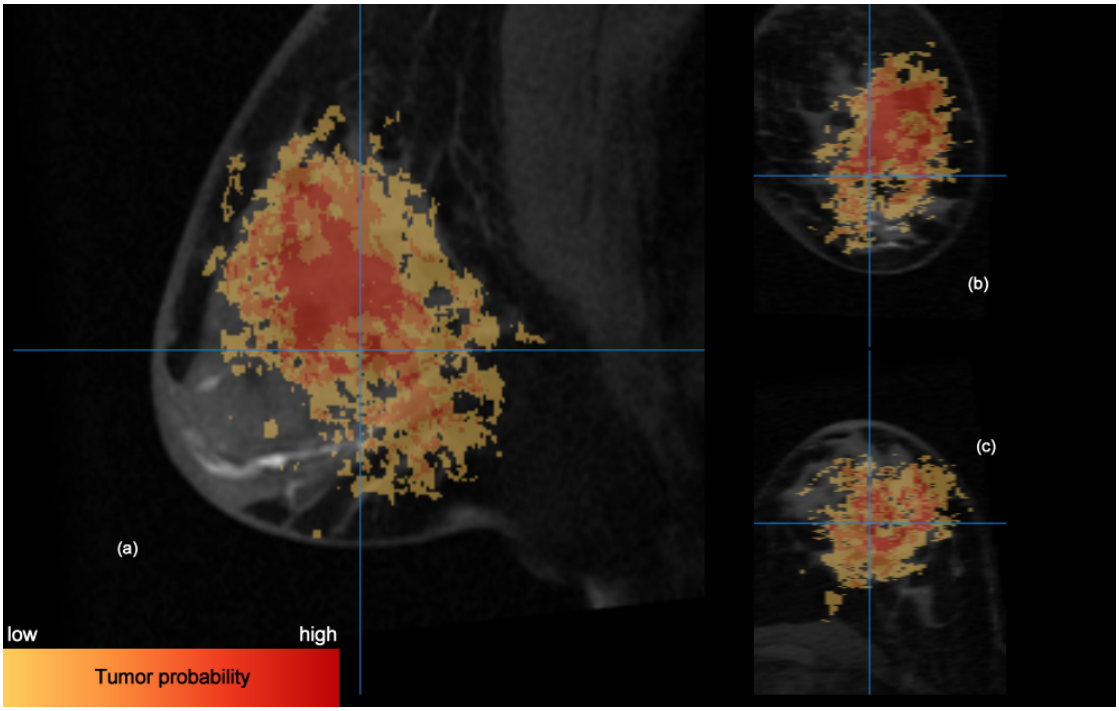


Figure 4.8: Example: Breast MRI scan with probability map, showing the sagittal view (a), the coronal view (b), and the axial view (c).

The intra-patient study allows a researcher to explore and analyze follow-up data of one patient. The researcher should be able to see imaging-data across time and how the tumor developed within the patient, as well as non-imaging data of the patient. For the imaging data, different approaches could have been used, but for a better visualization to query and analyze, a probability map of the tumor was selected due to the fact that it is commonly used in clinical practice [HBG02]. The probability map was calculated from the adjusted segmentation data we received from the registration process, and projected onto the MRI data of the selected patient (see Figure 4.8). To calculate the probability map, pixels of the individually aligned segmentations have been added up, this means that the higher the obtained value, the higher the probability of tumor localization at a specific pixel. A scheme of this process can be seen in Figure 4.9.

Secondary visualizations of the non-imaging data are displayed, to let the researcher visually explore various information about the patient in an easy way.

- The **progress of the tumor volume and longest diameter** (LD) should be recognized in a chart. Since only a part of the patients in the cohort was scheduled for all four MRI scans, and a lot of patients skipped the second or third MRI scan, interpolation would be needed when using line graphs, which would distort the visualization. Also, line graphs give the impression of continuity, which would

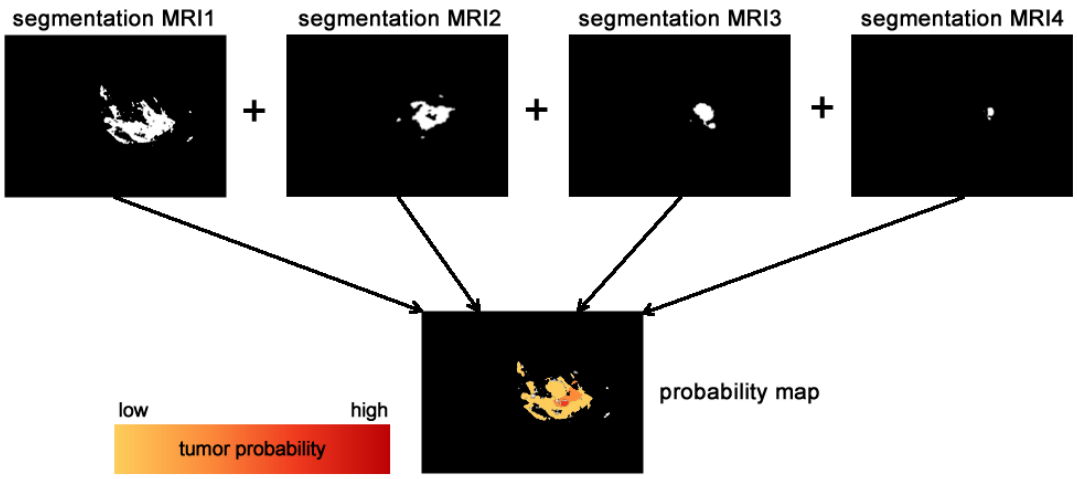


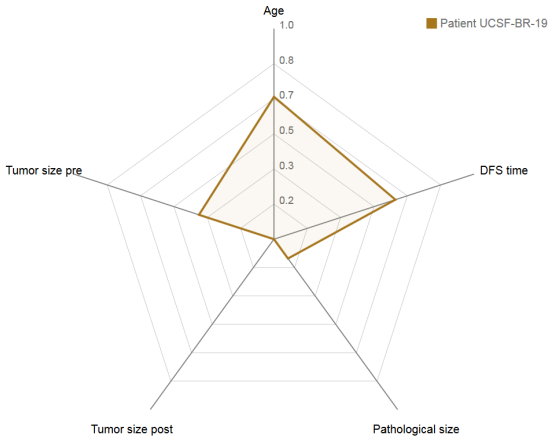
Figure 4.9: Scheme for calculation of the probability map.

be deceptive in our case, since the tumor volume and LD are only measured at discrete time points. Line graphs would only be suitable when showing the data of continuous measurements with lots of time points. This issue also concerns a *Streamgraph* [Tur18], which is why we considered graphs like a *Violin plot* [Sie17], or a simple *Bar chart* [Bos18a] as most applicable in our situation. We decided to show this by a **Bar chart** (Figure 4.10a). Since this is a very common way of presenting data over time, a researcher will intuitively recognize a growth or decrease of the tumor volume or the LD. For easier discrimination of volume and LD we use two different colors. Examples for a Streamgraph or a Violin plot can be seen in Section 4.7.

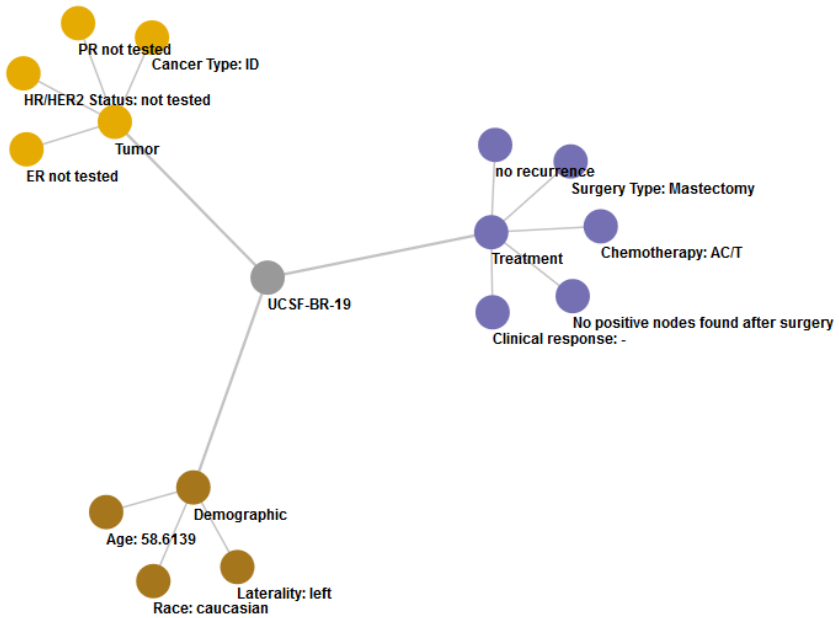
- Displaying **multivariate numerical data** in the cohort, such as age, the clinical assessed size of the tumor pre-chemo (*Tumor size pre*) and post-chemo (*Tumor size post*), the pathological endpoint size of the tumor (*Pathological size*), or the time from surgery to recurrence or last follow-up (*DFS time*), can be made accessible by the use of radial charts like *Radar charts* [Zho18] or *Radial boxplots* [Lin18]. We decided to use the **Radar chart** (Figure 4.10b) showing the mentioned multivariate properties according to their proportion to values of the total cohort. This is in preparation for the second and third part, because the comparison of these data to another patient or group can easily be done within one chart. An example for a Radial boxplot can be seen in Section 4.7.

- Furthermore we want to show the **division of the non-imaging demographic data** (as stated in Section 2.5) and their relation to each other. Suitable graphs for showing divisions and relations would be for instance a *Treemap* [Bos17c] (example see Figure 4.5), a *Mind map* [Stu17], or a *Force-directed graph* [Bos17b]. We adapted a **Force-directed graph** (Figure 4.10c) and added labels to show the relationships

(a) Bar chart for intra-patient follow-up study. The x-axis represents the scheduled MRI scans from 1 to 4, the y-axis represents the value of the volume at the upper part (dark yellow), and the value of the longest diameter (LD) at the lower part (yellow).



(b) Radar chart for intra-patient follow-up study. The various multivariate data instances are arranged clock-wise around the center, the y-axis represents the particular value normalized from 0 to 1 according to the value of the total cohort.



(c) Force-directed graph for intra-patient follow-up study. Properties belonging to one of the three divisions *demographic data*, *tumor-concerning data*, and *treatment-concerning data* are represented with similar colors and connections to their parent node.

Figure 4.10: The employed visualization charts for intra-patient follow-up study

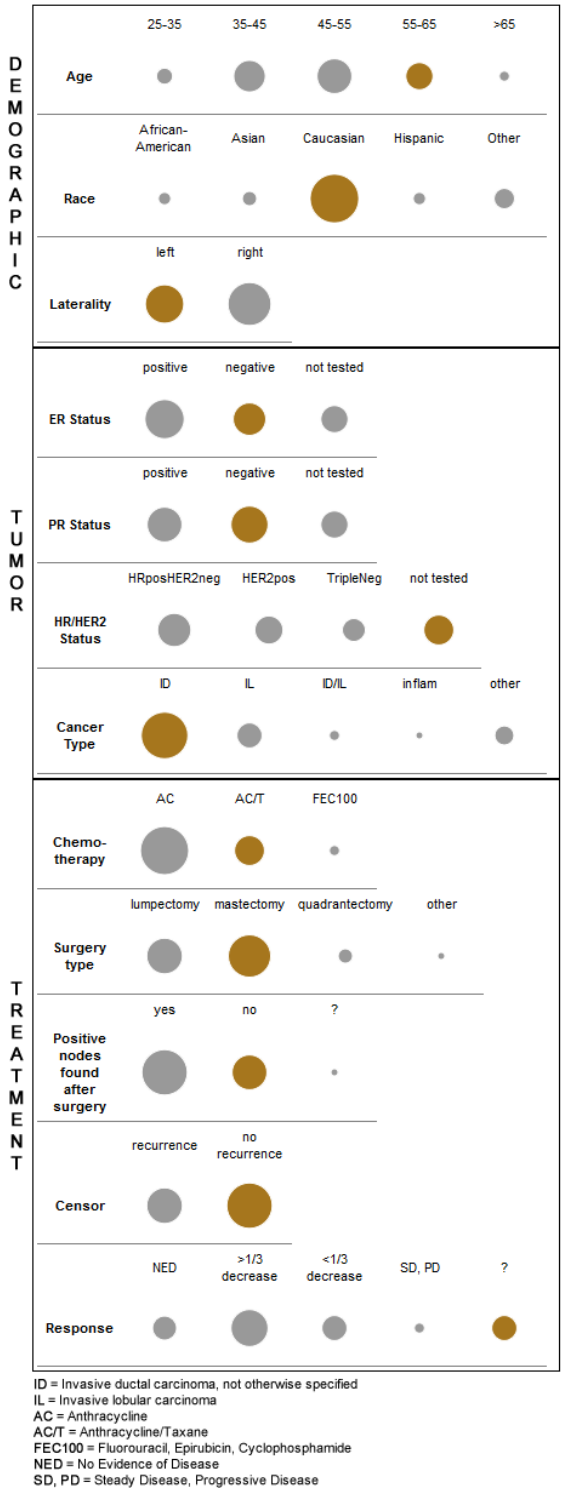


Figure 4.11: Spotmatrix for intra-patient follow-up study, grouped into three divisions according to *demographic data*, *tumor-concerning data*, and *treatment-concerning data*. The area of the circles represents the quantity in the cohort, current patient relevant properties are highlighted (dark yellow).

in the various non-imaging data in the cohort. We present each of the three groups we generated out of the non-imaging data - *demographic data*, *tumor-concerning data* and *treatment-concerning data* - with different colors (one distinct color for each group). We considered the Force-directed graph as more intuitive than the Treemap and the possibility to independently arrange the nodes was the main reason to choose it over the Mind map. An example for a Mind map can be seen in Section 4.7.

- To show the values of the various non-imaging data **in comparison to the total cohort**, it was necessary to add another visualization. The user should be able to identify the data of the patient compared to the cohort and to see the data distribution inside the cohort. This functionality would be given by several visualizations like a *Bubble chart* [Bos17a], a *Circle packing* chart [Bos18b], a visualization called *Donut multiples* [Bos18c] or a *Spotmatrix* [Nar]. In order to have the data clearly arranged, we decided against a Bubble chart or Circle packing. Also Donut multiples would not let the user recognize the distribution of the data at a glance, so we decided to use the **Spotmatrix** (see Figure 4.11), showing circles with an area according to their quantity in the cohort and highlighting the one that is coming from the selected patient. We decided to map these quantities to the area of circles instead of simple bars, because it allows centering both in horizontal and vertical direction and and the glyph-like representation allows easy detection of differences, even with a high number of attributes. Since an optical comparison when dealing with two patients or groups is easier if the visualized data is arranged as a matrix, we decided not to include the functionality to show different sized circles into the Force-directed graph. To be consistent in all three tasks, we already introduced this chart in the intra-patient study. Properties that apply to the current patient are highlighted (dark yellow), the other properties are shown in gray circles. Examples for a Bubble chart, a Circle packing chart, or Donut multiples can be seen in Section 4.7.

We point out that several visualization design choices were made with respect to all three tasks, i.e. we chose representations that would be suitable for all three tasks, to minimize the risk of overwhelming our intended users with a complex interface using a high number of different representations.

4.5 Pairwise inter-patient study

In the second part of the framework, the user can compare two different patients. The researcher should be able to compare both imaging and non-imaging demographic data of these patients, to recognize possible differences or commonalities. Therefore the shape of a standardized breast was created. This was done by applying a threshold onto an MRI scan, to only keep values concerning the breast tissue, and afterwards employing morphology operations *closing* (dilation followed by erosion) and *opening*

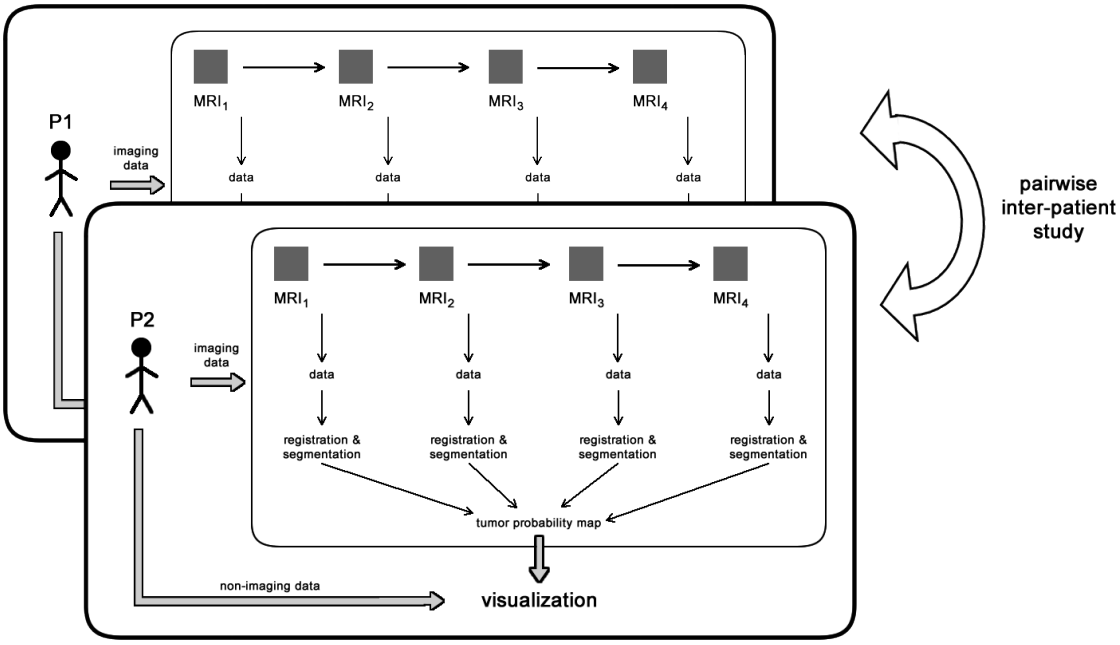


Figure 4.12: Workflow of pairwise inter-patient study

(erosion followed by dilation) [NT10] to remove artifacts that do not belong to breast tissue and close holes in the resulting shape (see Figure 4.13). The probability map of each patient was projected onto this standardized breast to let the researcher detect regions of high tumor probability (see Figure 4.14) - a method that Kyungyoon Kim et al. [KCK17] described as *superimposition*. We chose this way of displaying the tumor visualization over *juxtaposition* (side-by-side visualization), because with *superimposition* the researcher has the possibility to easily identify regions with higher tumor probability concerning both patients.

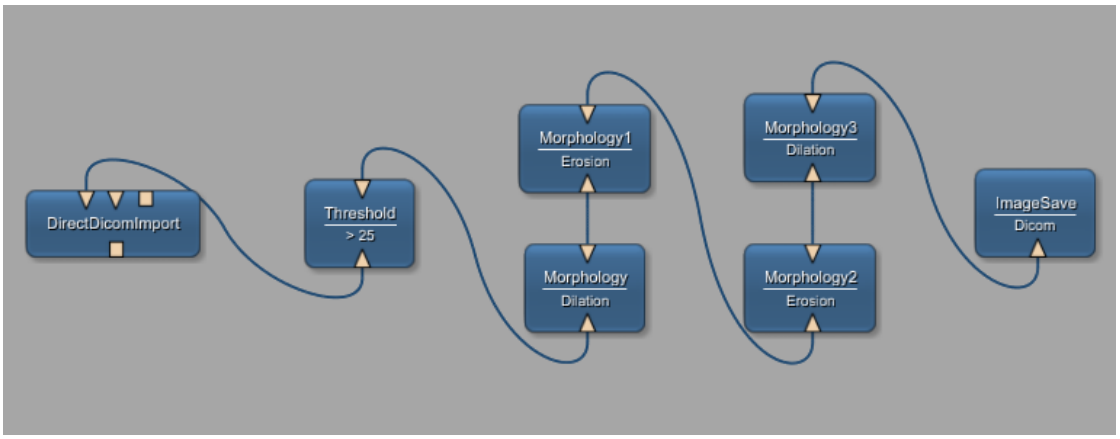


Figure 4.13: *MeVisLab* [AG] network used in order to create the standardized breast

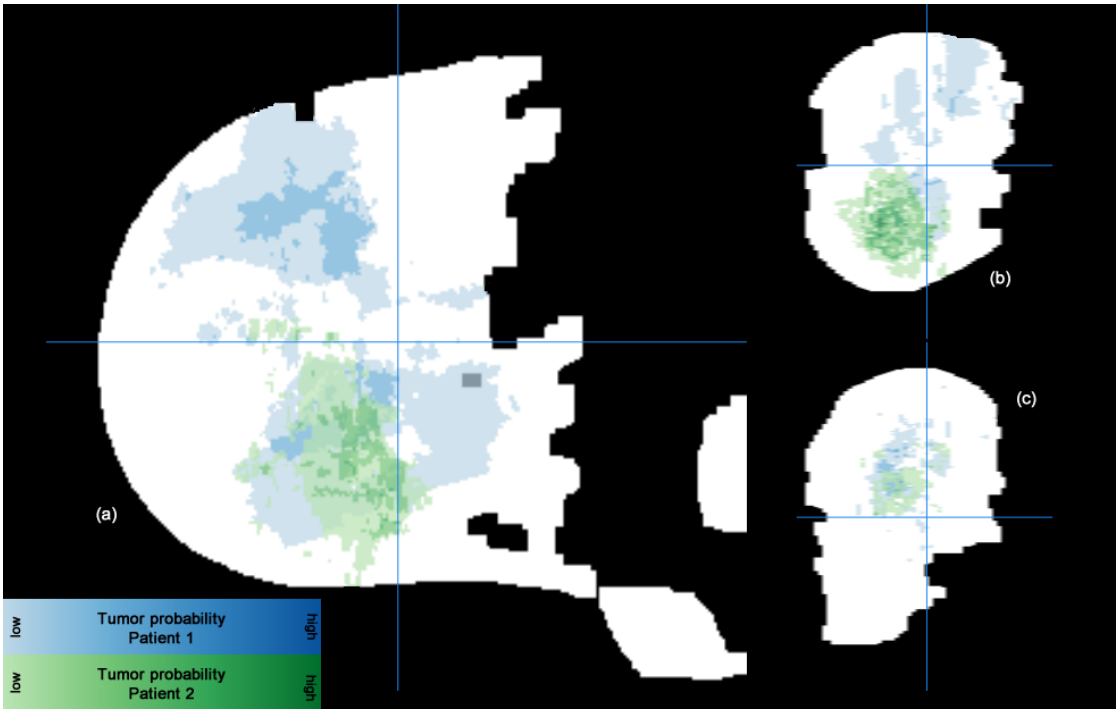
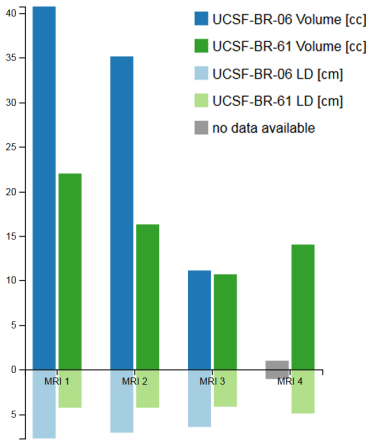


Figure 4.14: Example: Tumor probability map of two patients projected onto a standardized breast. One of the patients is shown in blue, the other in green. The sagittal view (a) is considered as the main view and therefore larger, the coronal view (b) and the axial view (c) are shown smaller on the right side.

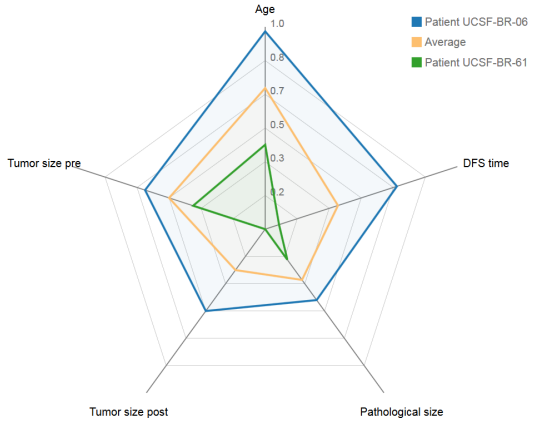
To display visualizations of non-imaging demographic data we used graphs based on the visualizations we introduced in the first part (Section 4.4). Since these graphs were initially designed to present single-patient data, they had to be adapted for this inter-patient study:

- The researcher should be able to compare the progress of the tumor and LD of both patients. Instead of displaying two Bar charts side-by-side, we designed a way to show the progress in only one visualization, so differences can be distinguished easier (see Figure 4.15a). For each MRI scan two bars are shown next to each other, each representing the value of one of the compared patients. The tumor volume data above the axis, the LD data below. Distinct colors are used for each of the patients.

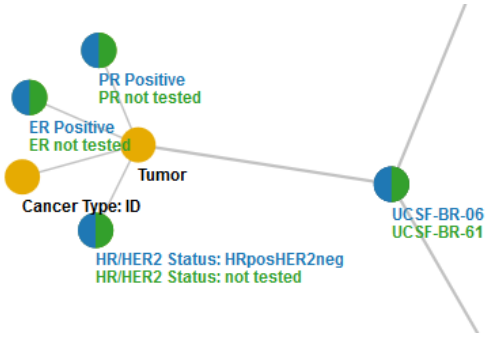
- To intuitively compare multivariate data of two patients, we extended the Radar chart, which is now showing three layers in one chart: The multivariate data of *the first patient*, the multivariate data of *the second patient* and additionally the *average* of these data (see Figure 4.15b). A researcher is able to detect differences



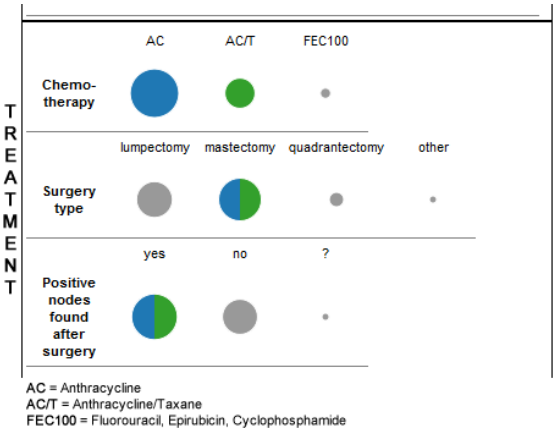
(a) Bar chart for pairwise inter-patient study. The x-axis represents the scheduled MRI scans from 1 to 4, the y-axis represents the value of the volume at the upper part (dark), and the value of the longest diameter (LD) at the lower part (light). For each MRI scan two bars are shown next to each other, each representing the value of one of the compared patients (colored in blue or green).



(b) Radar chart for pairwise inter-patient study. The various multivariate data instances are arranged clock-wise around the center, the y-axis represents the particular value normalized from 0 to 1 according to the value of the total cohort. Displayed are the values of both patients (blue and green) as well as the average (yellow). The area of each polygone is not shown in full opacity to avoid occlusion. Mouseover interaction highlights specific polygones.



(c) Extract of Force-directed graph for pairwise inter-patient study. Properties belonging to one of the three divisions *demographic data*, *tumor-concerning data* and *treatment-concerning data* are represented with connections to their parent node and shown in same colors. In case of different values in both patients, the node is split. If both patients share the same value, the node is not split and colored according to the data division (here yellow). Here only tumor data are shown.



(d) Extract of Spotmatrix for pairwise inter-patient study. The area of the circles represents the quantity in the cohort, the current patient relevant properties are highlighted. In case both patients share the same value, the node is split.

Figure 4.15: The employed visualization charts for pairwise inter-patient study

or commonalities of these multivariate data. Distinct colors are used for each of the patients.

- The comparison of the relation of the non-imaging demographic data is restricted to the values of these data. Thus we used the same graph as for the intra-patient study, but added a feature to explore differences between the two patients. As seen in Figure 4.15c a node is split in case of different values in both patients. If both patients share the same value, the node is not split and colored according to the data division *demographic data*, *tumor-concerning data*, or *treatment-concerning data*. This allows a researcher to spot differences in an easy and intuitive way.
- To additionally compare the values of the non-imaging demographic data to the cohort, we adapted the Spotmatrix, which we discussed in Section 4.4. Similar to the intra-patient study, the area of the circle represents the quantity of the value in the cohort. To show which values are concerned by the observed patients, the corresponding circles are highlighted. If both patients share the same value, the circle is split (as shown in Figure 4.15d). Again, we used distinct colors for each of the patients.

4.6 Groupwise inter-patient study

Finally, a functionality to compare different groups of patients is provided. The user should be able to select two groups of patients by arranging certain criteria. An example of selecting two groups would be to choose one group having received only Anthracycline as chemotherapy, while the other group additionally received Taxane. A clinical researcher should now be able to visually explore and analyze imaging and non-imaging data of these groups.

To compare the imaging data of the two groups, a summarized probability map of both groups projected onto the standardized breast would not let the researcher recognize differences between the groups, because the visualization would be cluttered. So we decided to use the concept of *juxtaposition* [KCK17], where visualization data is displayed side-by-side (see Figure 4.17). Using this method, the user can easily spot differences in the imaging data, since the underlying standardized breast is the same for both parts of the visualization.

For the visualization of the non-imaging demographic data we decided to again use adapted versions of the visualization graphs we introduced in Section 4.4 for the intra-patient study, like we did for the pairwise inter-patient study (see Section 4.5), except for the Bar chart, which we replaced with Boxplots:

- For the comparison of the tumor volume and LD of two groups, we decided to replace the Bar chart with **Boxplots** [Gru17], because we considered it as the best way to show the distribution of the values within a group. The chart displays the distribution of either the tumor volume or the LD of the selected groups side-by-side

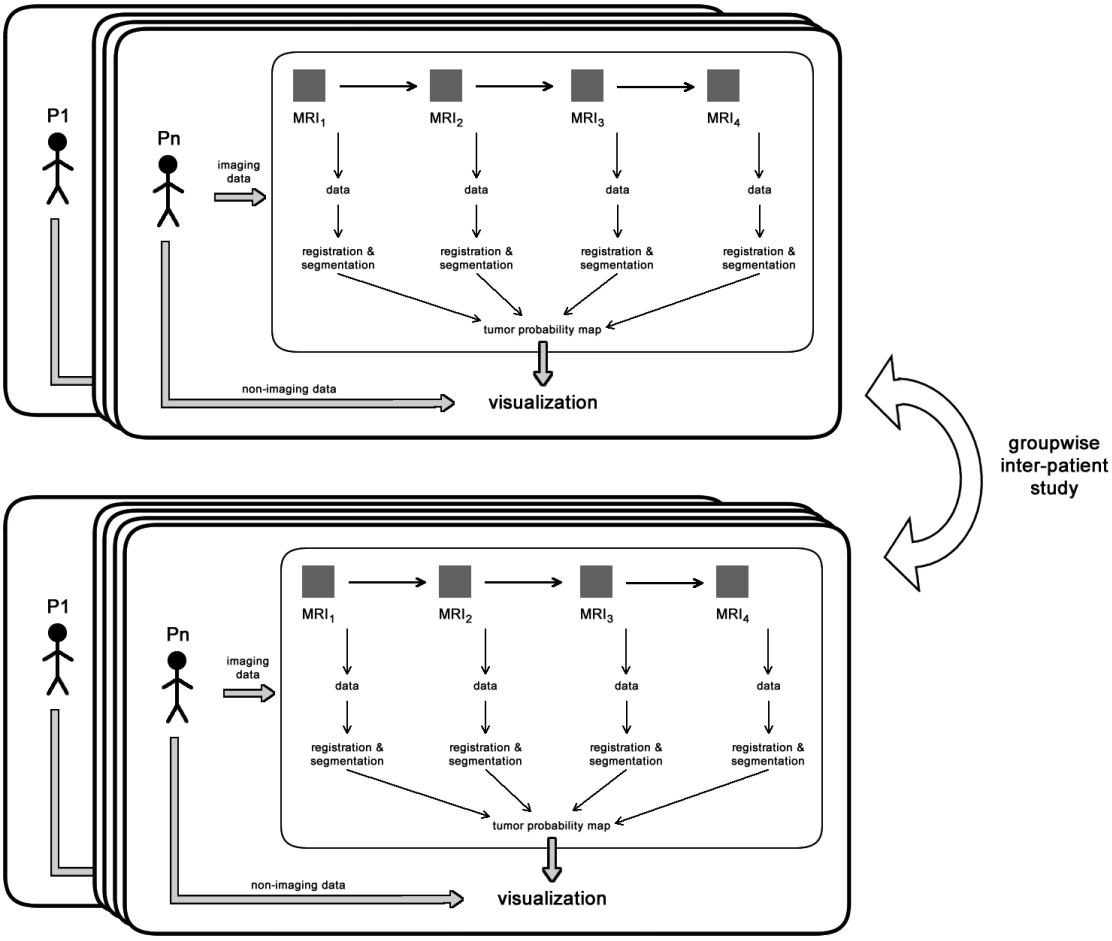


Figure 4.16: Workflow of groupwise inter-patient study

(see Figure 4.18a). The distribution of the tumor volume is shown in the upper chart, the distribution of the LD is shown in the lower chart. Regarding this chart, a researcher can easily spot differences in the tumor progress between the two groups. Different colors now denote different groups.

- To analyze the multivariate data of the two groups only by using one overlapping chart (*superimposition* [KCK17]), we adapted the Radar chart again. The user can now compare the maximum, average, and minimum of each multivariate property in both groups and in this way easily recognize differences or commonalities (see Figure 4.18b). These three values show the range of the concerning data. We chose the average as middle value over e.g. the median, because it is the best value to display for both large and small groups. The median or other values like interquartile range or variance would distort the graph in very small groups of only 2 or 3 patients.

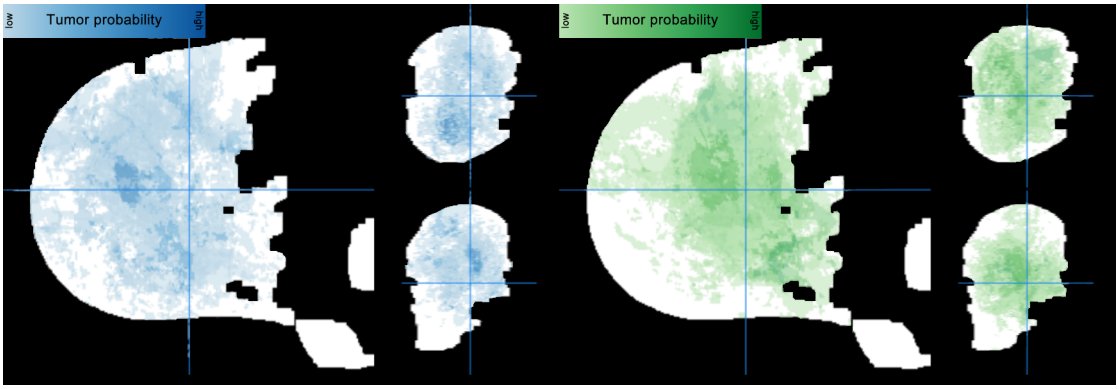
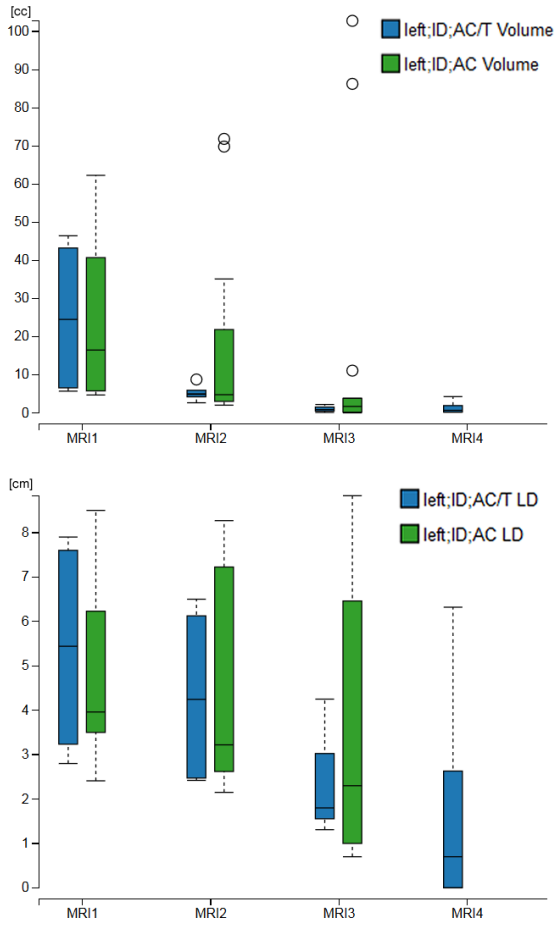


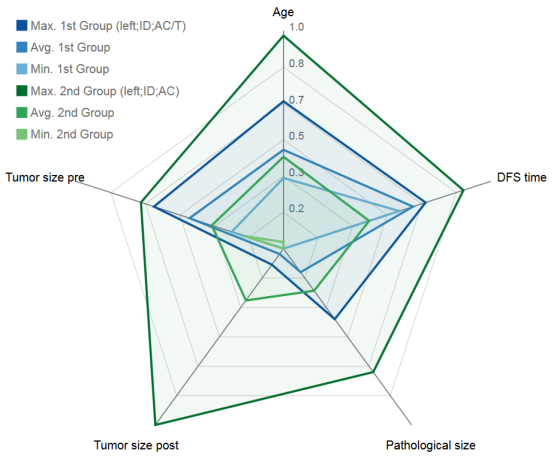
Figure 4.17: Example: Tumor probability map of two patient groups projected onto a standardized breast and displayed side-by-side. One group is shown in blue, the other in green.

- Concerning the relation of the non-imaging demographic data we wanted to show two aspects: the relation of the data to each other (like we did in the other two functionalities intra-patient study and pairwise inter-patient study), and additionally the distribution of these data within the two groups. An optimal design would be to show this information within one chart, so we extended the Force-directed graph with one more feature. Based on the idea of a *Waffle chart* [Gim17], we displayed the distribution of the non-imaging demographic data in each group with small colored squares (see Figure 4.19a), so a researcher can easily perceive differences in these data in the selected groups. Distinct colors correspond to the two distinct groups.

- Finally, a visualization to show the quantity of a property within one of the selected groups in comparison to the total cohort was desired. As seen in Figure 4.19b we used the *Spotmatrix* and added one more feature. Again the area of the circle represents the quantity of the property in the cohort. To represent the distribution of the various properties within the two groups, the circles were split into halves and we added luminance: the more luminous the concerned circle is displayed, the more patients are affected by this property. The left parts of the circles represent one group, the right parts the other. If only a part of the group is represented in a property, this is mapped to luminance. In this way the user can spot both the quantity of each property inside the cohort, and the distribution within both of the selected groups in only one visualization.

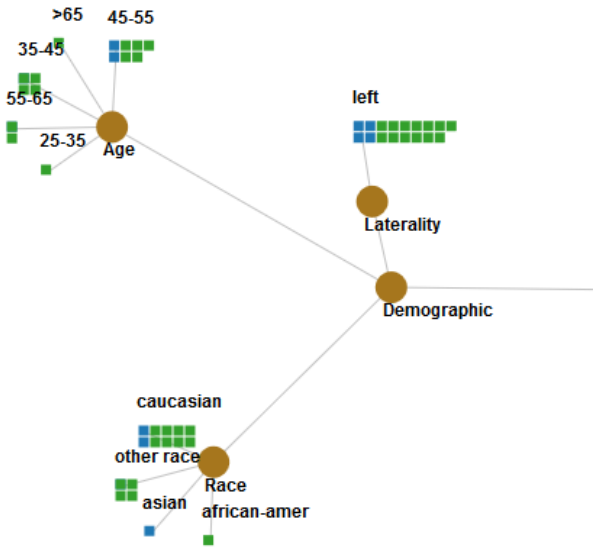


(a) Boxplots for groupwise inter-patient study. The x-axis represents the scheduled MRI scans from 1 to 4, the y-axis represents the value of the tumor volume in the upper chart and the value of the longest diameter (LD) in the lower chart. One of the regarded groups is shown in blue, the other in green.

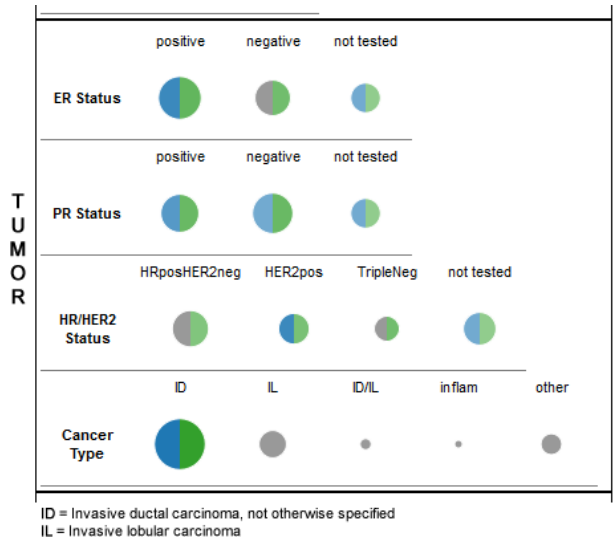


(b) Radar chart for groupwise inter-patient study. The various multivariate data instances are arranged clock-wise around the center, the y-axis represents the particular value normalized from 0 to 1 according to the value of the total cohort. For each regarded group the maximum, average, and minimum of each property are displayed.

Figure 4.18: The employed visualization charts *Boxplots* and *Radar chart* for groupwise inter-patient study



(a) Extract of Force-directed graph for groupwise inter-patient study. Properties belonging to one of the three divisions *demographic data*, *tumor-concerning data*, and *treatment-concerning data* are represented with similar colors and connections to their parent node. Here only demographic data are shown. Every property node is extended by a *Waffle chart* showing the quantity of the property within the considered groups.



(b) Extract of Spotmatrix for groupwise inter-patient study. The size of the circles represents the quantity in the cohort, each node is split to represent both groups. The luminance of a particular circle represents the quantity of the property within the group.

Figure 4.19: The employed visualization charts *Force-directed graph* and *Spotmatrix* for groupwise inter-patient study

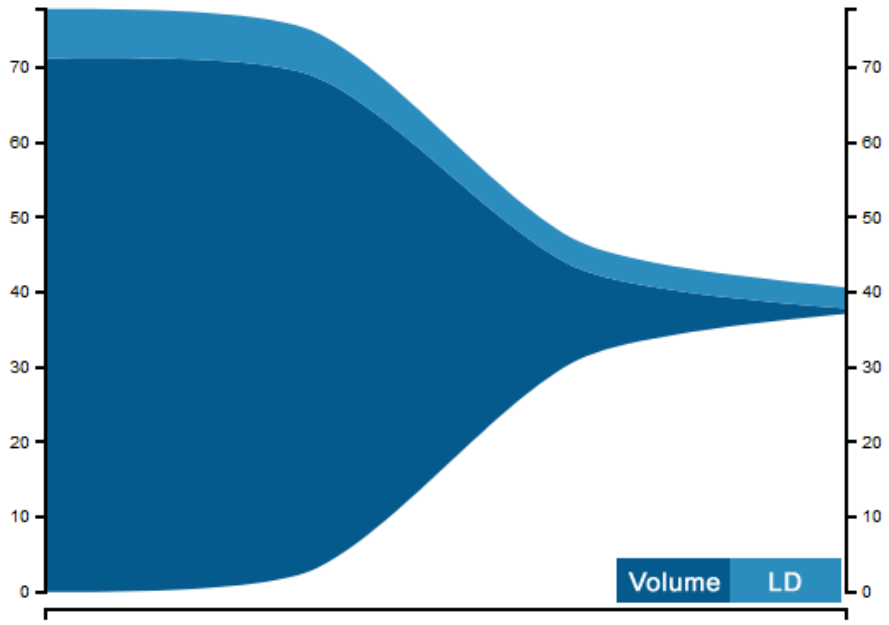
4.7 Conceptual design choices & alternatives

As stated in Section 4.4 for every visualization several approaches were considered. The decision to use the chosen attempts were based on several thoughts and analyzes of the advantages and disadvantages of each possibility. In the following, the feasible charts will be discussed, the solution we selected is marked bold.

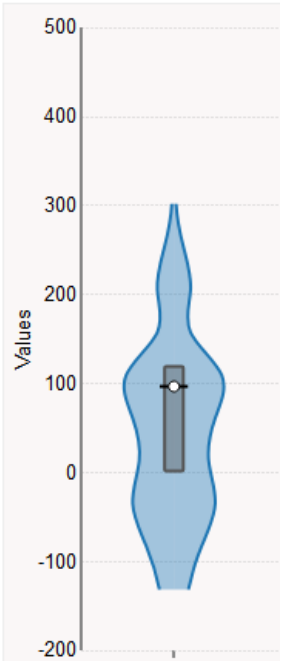
To display medical imaging data within the domain of a web-based interface, several solutions and libraries already exist, or are still in the process of development. The following approaches were explored and analyzed: The *AMI - Medical Imaging JavaScript ToolKit* [RBRH+17] uses Three.js [mrd17]. The disadvantage of this tool was that only the sagittal view was displayed, the interactivity was on the same level as the finally chosen Papaya tool. *Cornerstone - Browser-based imaging library in JavaScript/HTML5* [Haf17] is a tool, which is still in development and only provides low-level interactivity. The *DWV - DICOM Web Viewer* [ivm16] is a very simple tool to display DICOM [All] images, where the level of interactivity is very low and the possibility to project segmentation data is not available in this tool. ***Papaya*** - "a pure JavaScript medical research image viewer" [RU] has the advantage of displaying three different views (axial, coronal and sagittal) of the imaging data and was highly interactive to use.

To explore changes in tumor volume and longest diameter (LD) both as follow-up data and as comparison, the following approaches were possible visualizations: A ***Bar chart*** [Bos18a] is a very common and therefore intuitive way of presenting data over time without the need of interpolating missing data. Furthermore a Bar chart provides the possibility to compare data to each other within the same time point. A *Streamgraph* [Tur18] (Figure 4.20a) is showing the progress of the data as a stream. Since only a few of the patients in the cohort were scheduled for all four MRI scans, interpolation of missing data is necessary, which would distort the visualization in our situation. A *Violin plot* [Sie17] (Figure 4.20b) is presenting the distribution of data, using the x-axis for a value assigned on the y-axis. A disadvantage of this visualization is the need of multiple values, which would only be possible for the third part groupwise inter-patient study, and only if the selected group consists of multiple patients. In the third part we decided to use **Boxplots** [Gru17], because this is a very common way of presenting distributed data and is intuitive to use for a researcher.

Multivariate data, such as age, the clinical assessed size of the tumor pre-chemo and post-chemo, the pathological endpoint size of the tumor, or the time from surgery to recurrence or last follow-up, can most suitably be visualized by radial plots. In a ***Radar chart*** [Zho18] the multivariate data is arranged in a circle and the values are plotted on normalized axes. The Radar chart also makes it possible to display data of multiple patients or groups, which is why we finally chose this visualization. A *Radial boxplot* [Lin18] (Figure 4.21) is displaying the data radially, where the x-axis is represented as a circle, and the y-axis extends from the center to the border. It also shows characteristics of a boxplot, such as outliers, and has the disadvantage that multiple values are necessary to suitably present data, which is why we chose the Radar chart to display multivariate data.



(a) Example of *Streamgraph*. Figure is a screenshot of a discarded implementation showing the development of tumor volume and longest diameter (LD) based on the work of William Turman [Tur18]



(b) Example of *Violin plot*. Figure taken from the work of Andrew Sielen [Sie17]

Figure 4.20: Distorted example graphs to visualize tumor volume and longest diameter (LD) data.

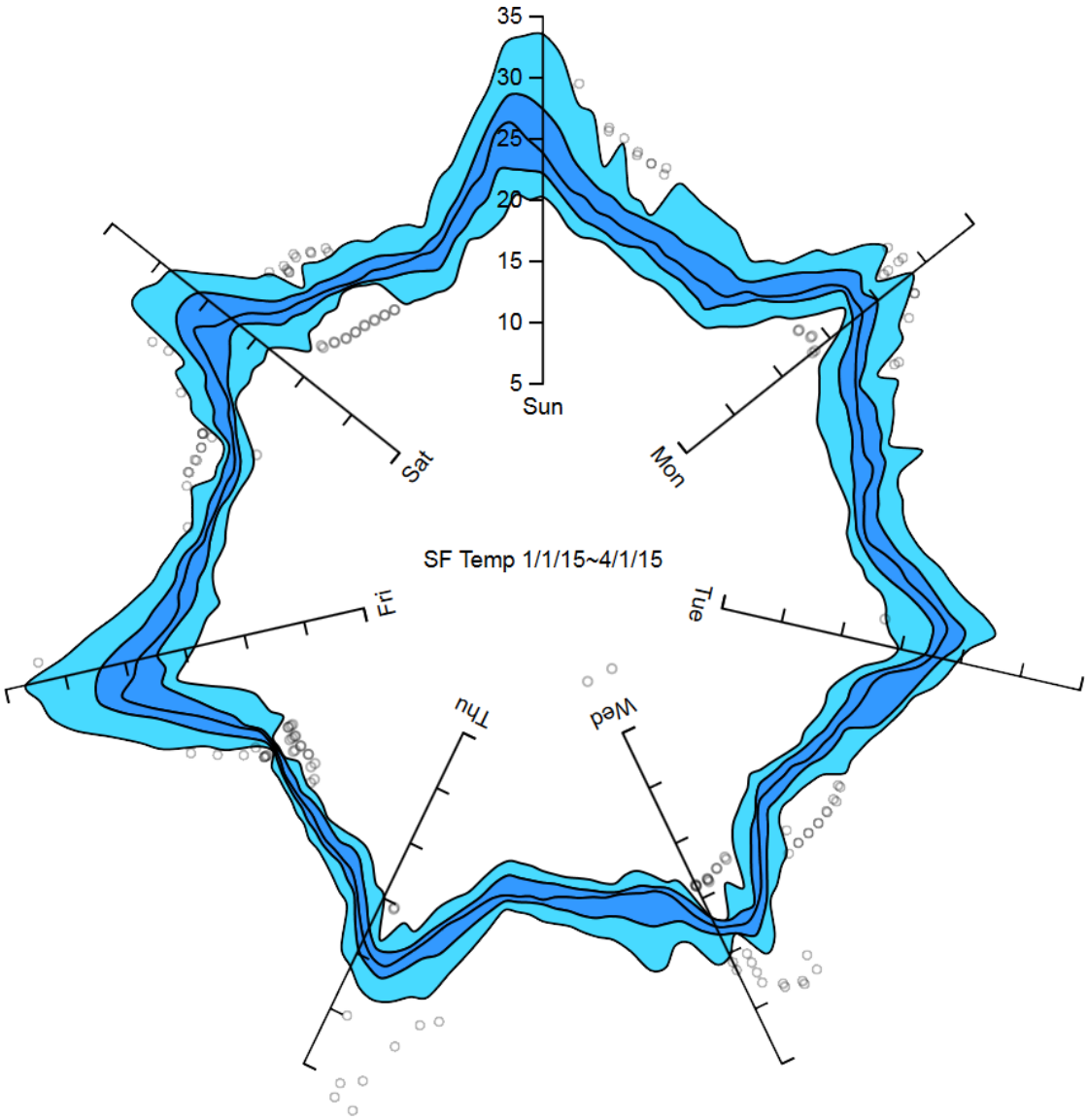


Figure 4.21: Example of *Radial boxplot*. Figure taken from the work of David Lin [Lin18]

A visualization of the division of non-imaging demographic data and their relation to each other can be shown in different ways. A ***Force-directed graph*** [Bos17b] is an easy and intuitive graph, with the ability to show differences and commonalities in the same visualization as the relation of the data. It shows the relation of the data as a network, where each property is represented by a node and the relation of these properties to each other is shown by a link. The advantage of this graph is the possibility to show even more information by adapting the size, color or shape of the nodes. A *Mind map* [Stu17] (Figure 4.22) is very similar to a Force-directed graph, but radially arranged. The

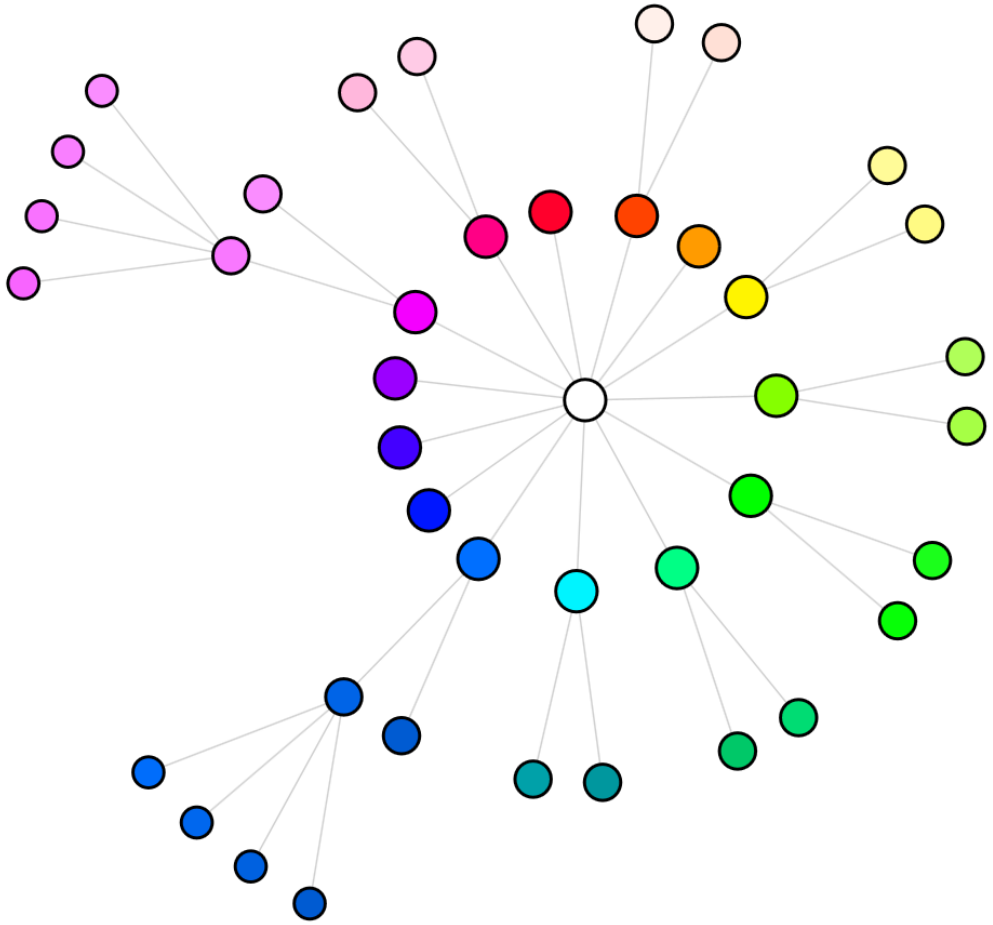


Figure 4.22: Example of *Mind map*. Figure taken from the work of DecemberCafe Studio [Stu17]

possibility to independently arrange the nodes of the graph, was the main reason why we chose the Force-directed graph over the Mind map. A *Treemap* [Bos17c] (Figure 4.5) is a diagram showing corresponding data with the same color. The intuitivity in the Treemap is at a lower level than in a Force-directed graph, and also the possibility to display more information about the data is limited in a Treemap.

Comparing the values of the various non-imaging demographic data to the cohort can be done by displaying visualizations using the following charts: A *Bubble chart* [Bos17a] (Figure 4.23) is showing the quantity of a property in the cohort by the size of the bubbles, it can be adapted to highlight the concerning properties through coloring. Nevertheless, the bubbles in this chart are not arranged clearly, which is why we did not choose this way of visualizing the desired information. A *Circle packing* [Bos18b] (Figure 4.24) is an adaption of a Bubble chart, where bubbles that belong to the same cluster are surrounded by a circle. Since the arranging of the bubbles does not fit to a grid, the perceptibility of

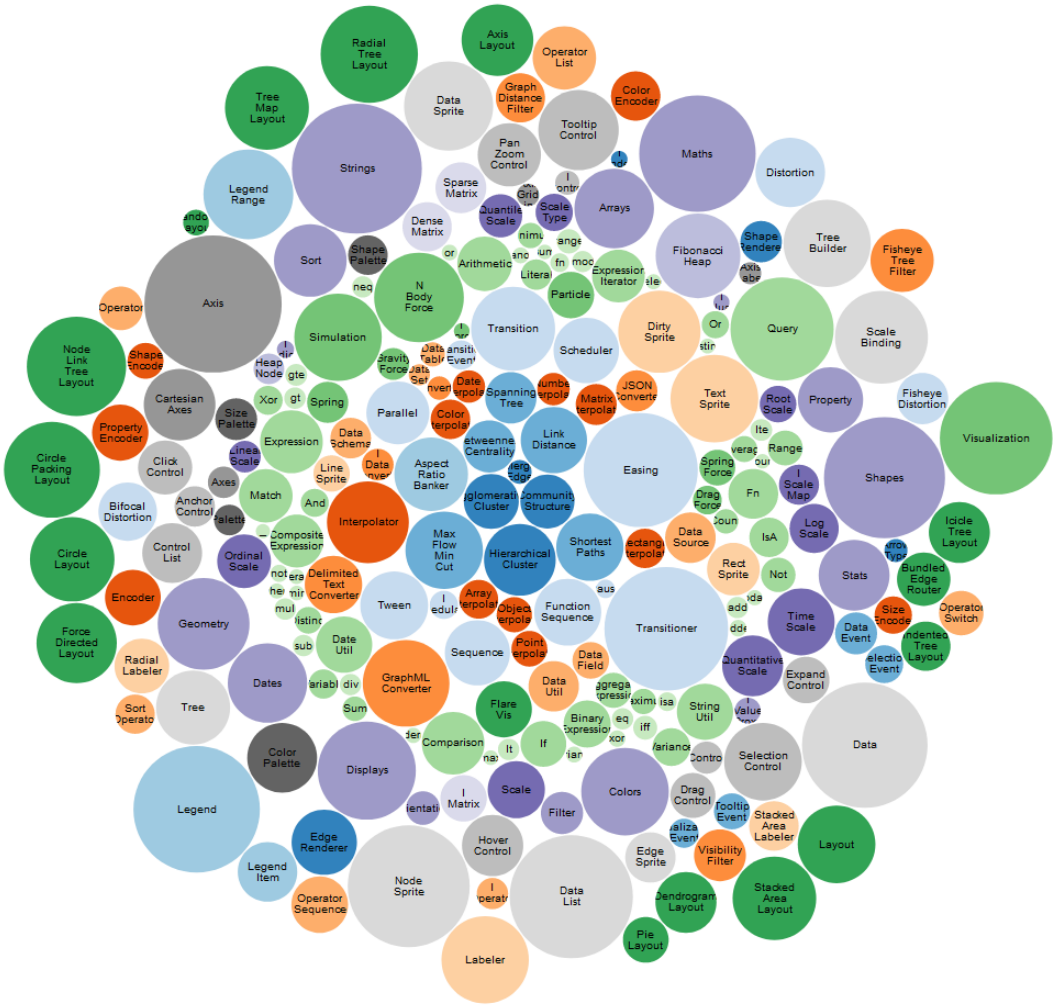


Figure 4.23: Example of *Bubble chart*. Figure taken from the work of Mike Bostock [Bos17a]

the data did not reach the level of the Spotmatrix, which is why we decided against this approach. *Donut multiples* [Bos18c] (Figure 4.25) consist of several pie charts displaying various data. This chart would not let the user recognize the distribution of the data at a glance and the comparability of data of different groups or patients is not implemented, which is why we chose the Spotmatrix over this approach. A ***Spotmatrix*** [Nar] shows the desired information arranged as a matrix, showing circles with a size according to the quantity in the cohort. Each circle represents a property of the data. The possibility to display even more information by adding several features to the circles, was a big advantage of the Spotmatrix. Thus this approach could be used in all three tasks.

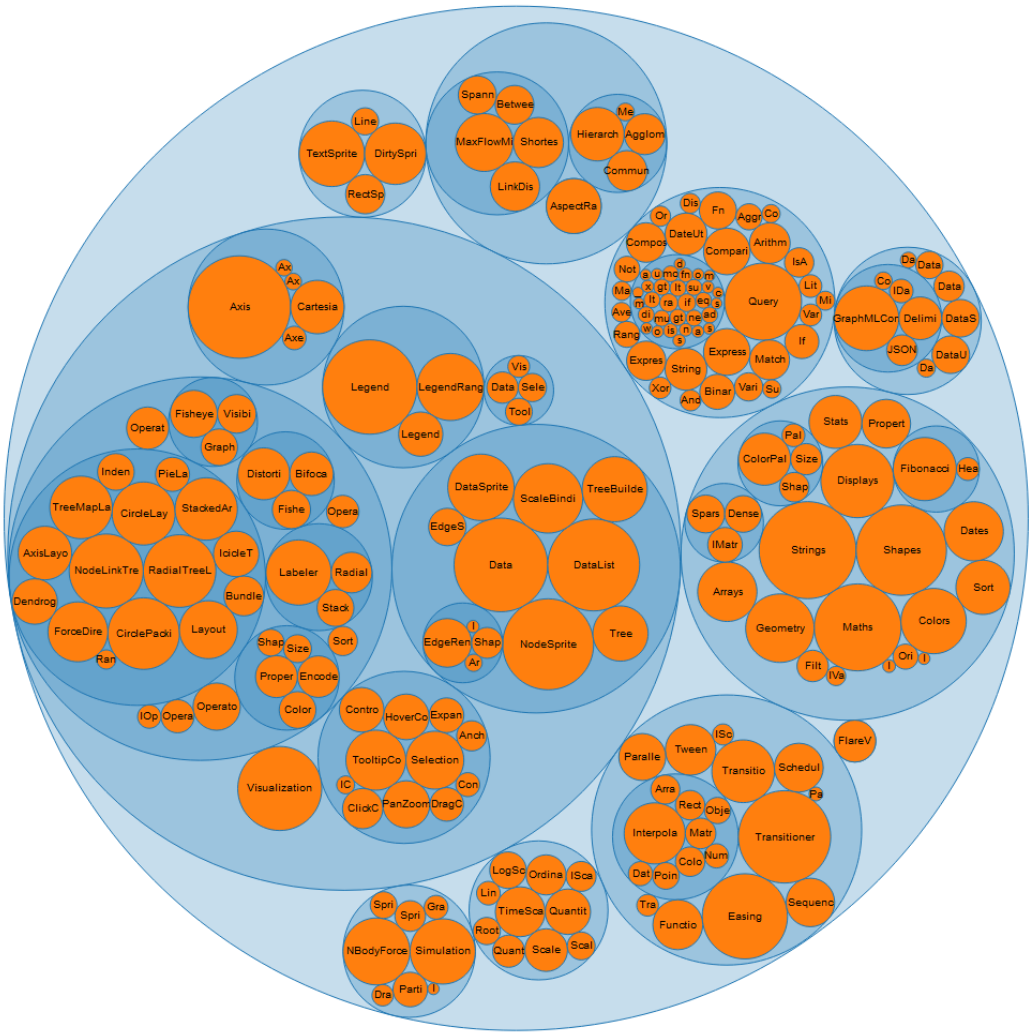


Figure 4.24: Example of *Circle packing*. Figure taken from the work of Mike Bostock [Bos18b]

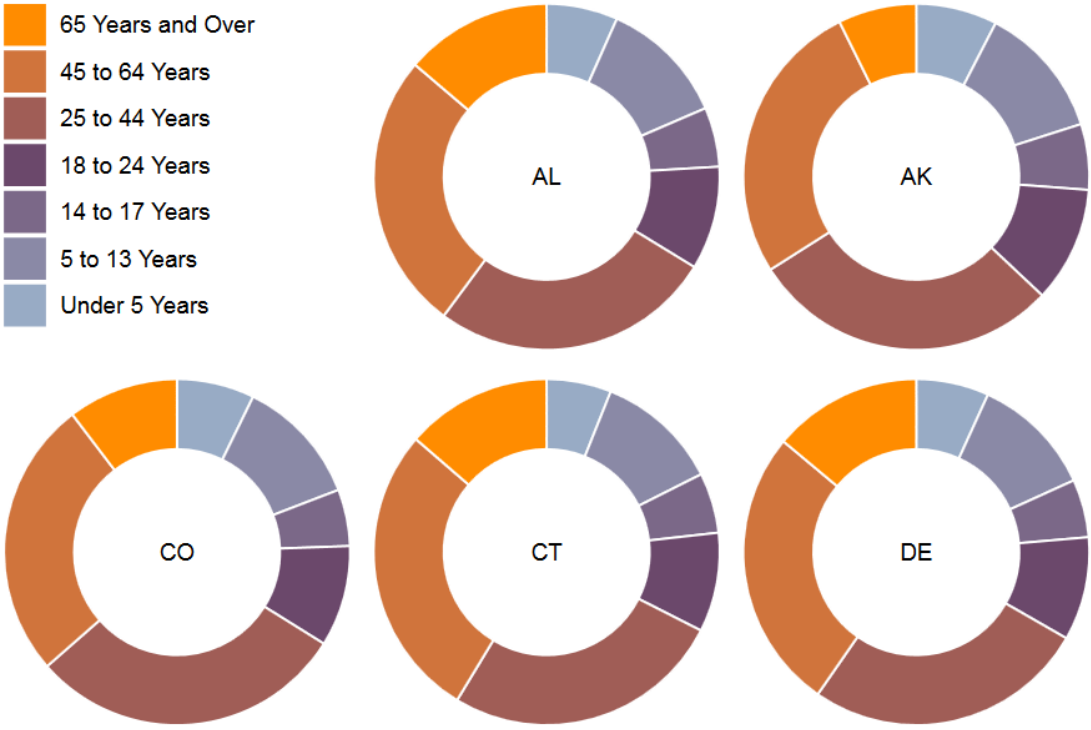


Figure 4.25: Example of *Donut multiples*. Figure taken from the work of Mike Bostock [Bos18c]

CHAPTER 5

Implementation

5.1 Database

To provide the possibility to query the existing imaging and non-imaging data, a database was built, which can be accessed by the framework in real time. We decided to use a database based on *MySQL* [Cor], set up the required tables and imported the provided non-imaging demographic data. An *Entity-Relationship-model* (ER-model) of the database can be seen in Figure 5.1. All the data referring directly to the patient (such as age, race, etc.) is stored in the table *patient*, data referring to an MRI scan (such as volume and LD) is stored in the table *mri*. Since there are two datasets for every MRI scan (MR images and segmentation) a *series* was created for each in the according table. The table *files* contains all DICOM-files [All] with reference to the related series. Based on this database all imaging and non-imaging demographic data can be accessed, all calculations can be done and all visualizations can be generated and displayed.

5.2 Web interface

The interface to connect the database with the website was realized by the programming language PHP [gro]. PHP provides all functionalities to access a MySQL database. Furthermore using PHP it is possible to create files, such as comma-separated-value-files (CSV files) or JavaScript Object Notation files (JSON files), which were needed as a basis for some D3 visualization charts.

The representation of the imaging data in the web interface was realized with the library *Papaya* [RU], as described in Section 4.7, while the visualization of the non-imaging data was realized by various charts based on *D3.js* [Bos]. Additionally some interaction functionalities not concerning the visualization charts (such as smooth scrolling or sortable list views) were implemented by the use of the JavaScript library *jQuery* [jf]. Optical

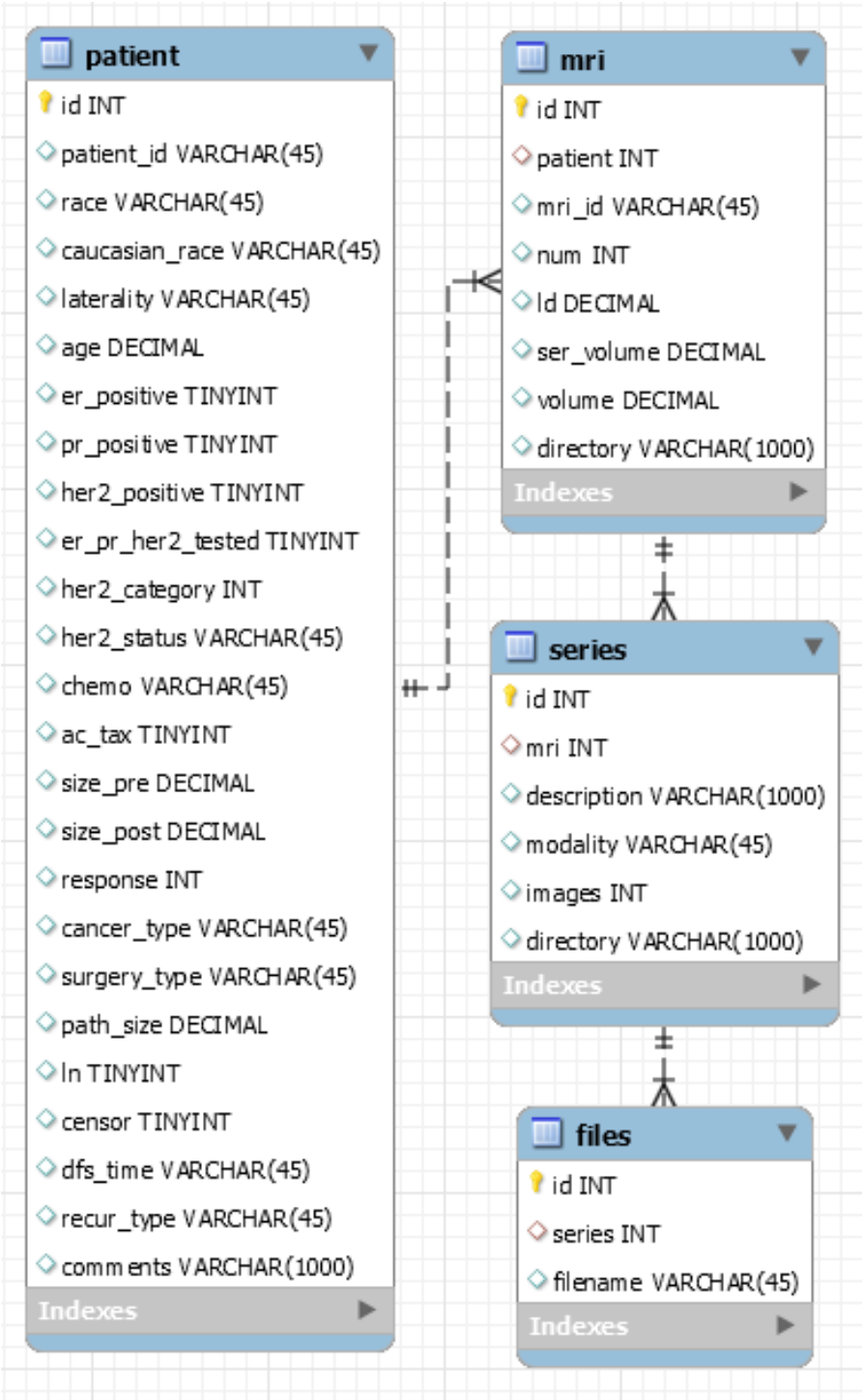


Figure 5.1: Entity-Relationship-model of the database

appearance attributes - i.e. font size, font family, background colors, etc. - were realized using *Cascading Style Sheets (CSS)* [Con]. The arrangement of the visualizations within the main interface was realized by *iFrames*, which means that for every graph a distinct page was generated and displayed (see Figure 5.2).

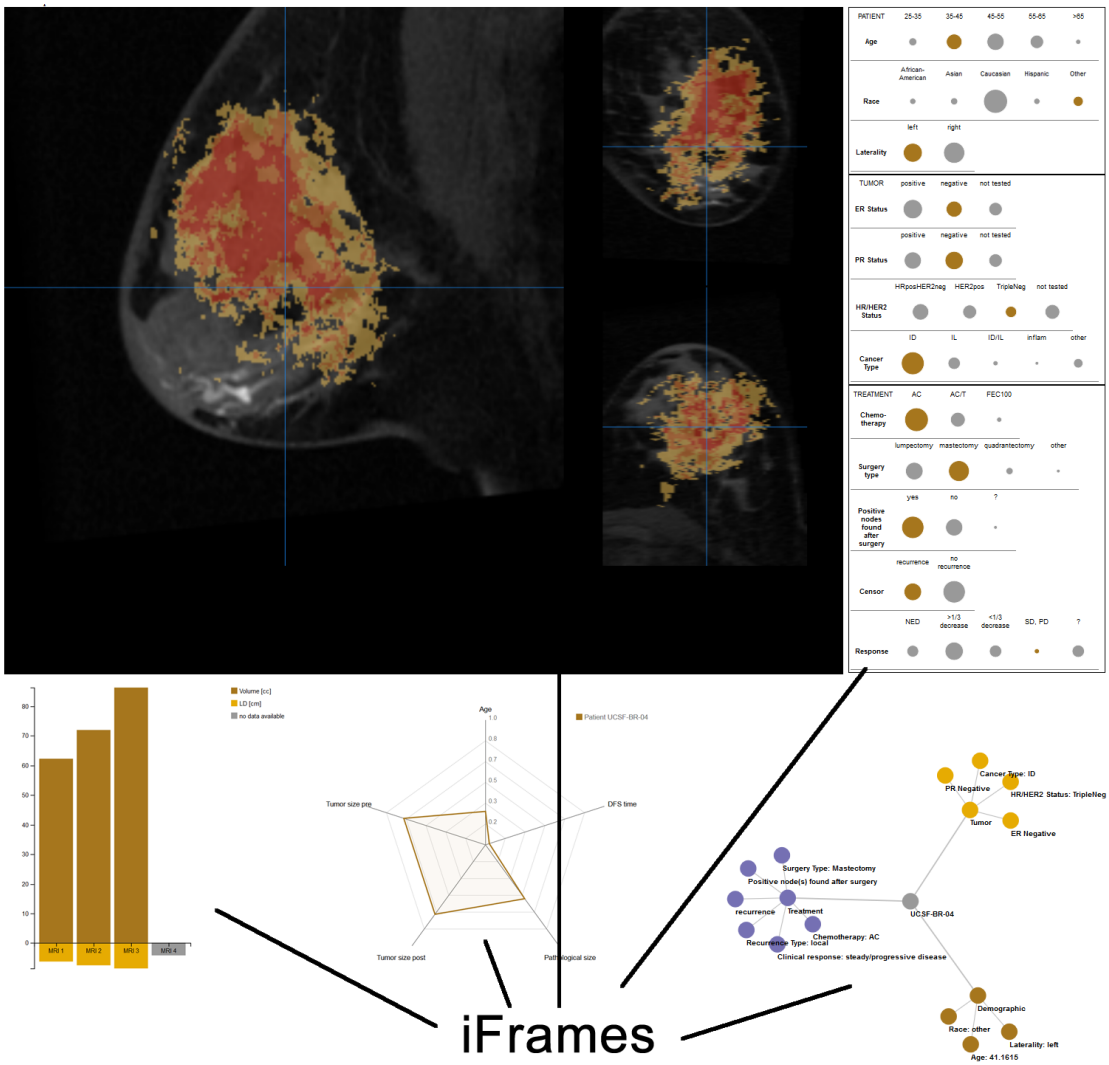


Figure 5.2: Screenshot of Web interface using iFrames

5.3 Imaging data

To display the imaging data we used the JavaScript library *Papaya* [RU]. Since this library provides a lot of configuration possibilities, we decided to adjust the settings as follows. A view of Papaya with standard configuration can be seen in Figure 5.3.

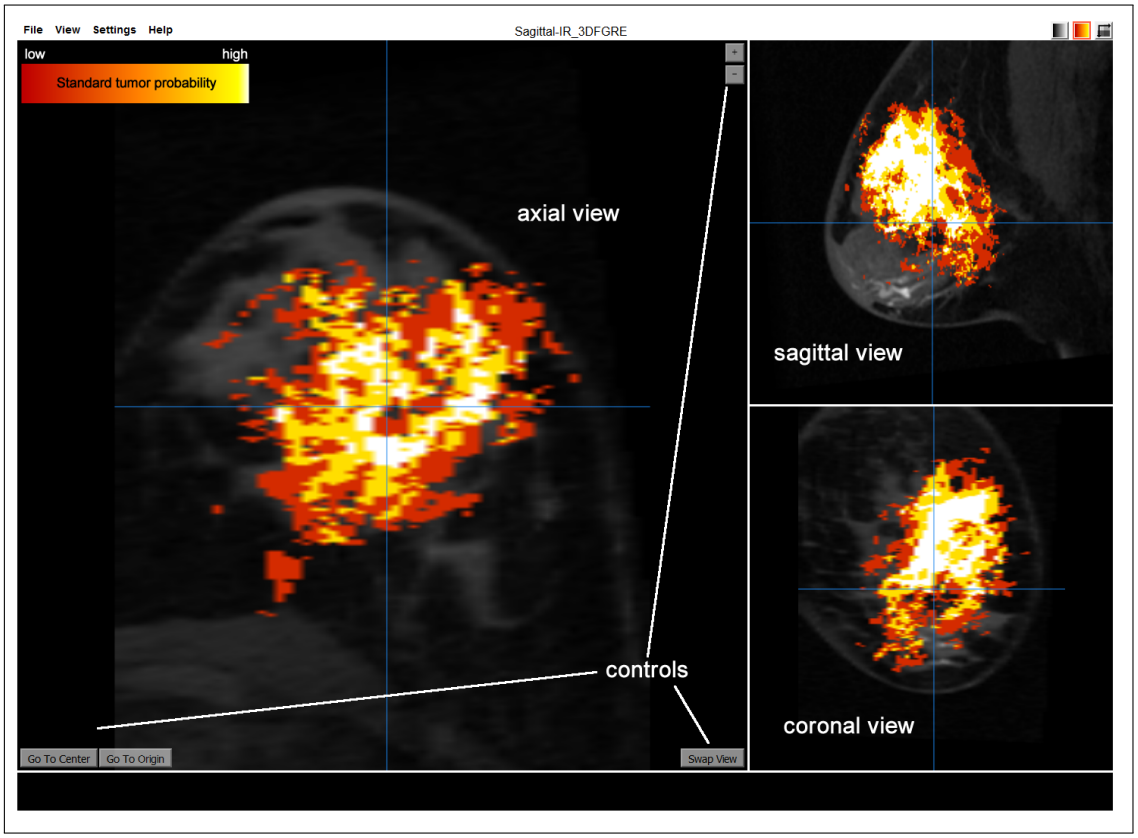
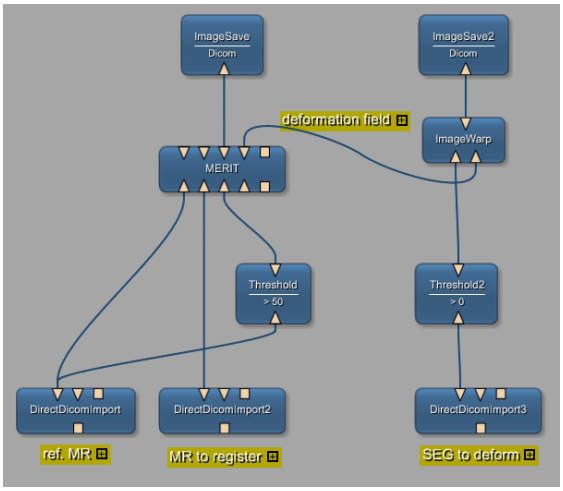


Figure 5.3: Screenshot of standard configuration view of *Papaya*, displaying a breast MRI and segmentation data of the cohort used in this thesis before changing the configuration as described.

- *Fullscreen Mode* - The visualization of the imaging data fills out the complete iFrame and no white bar is displayed at the top and the bottom (which would be the standard setting).
- *Main View: sagittal* - The most common use of displaying breast MRI, the sagittal view, is set to be shown on the left and large part. Other possibilities would be the *axial* (standard) and the *coronal* view, which are shown on the right (smaller parts).
- *Do not show controls* - In the standard settings a couple of controls are shown, which we avoid, because of their uselessness in our case.
- *Coloring* - The colors of the segmentation overlay were changed according to recommendations of *Colorbrewer.org* [HB03].
- *Opacity* - When using the pairwise inter-patient mode, the opacity of the overlaying segmentations is set to 70%, so the top-level segmentation does not completely

cover the displayed data below. In the groupwise inter-patient mode the opacity is calculated according to the number of patients in the group, to normalize the visualization. Nevertheless we set a minimum value of 50% to avoid segmentations to be too transparent ($opacity = \frac{0.5}{number\_of\_patients} + 0.5$).

For the preprocessing procedures of the imaging data, as described in Section 4.2, the tool *MeVisLab* [AG] was used. For the registration part and concurrently the adaption of the segmentation files according to the deformation field, the MeVisLab network shown in Figure 5.4a was employed. The deformation field, that was generated during the registration process, contains information about displacements of the concerning voxels in the deformed MRI scan. The deformation field in the example shown in Figure 5.4b, shows that the voxels in the concerned MRI scan are tilted and scaled.



(a) *MeVisLab* [AG] network to register MRI scans and apply deformation fields on segmentations.

(b) Example for deformation field

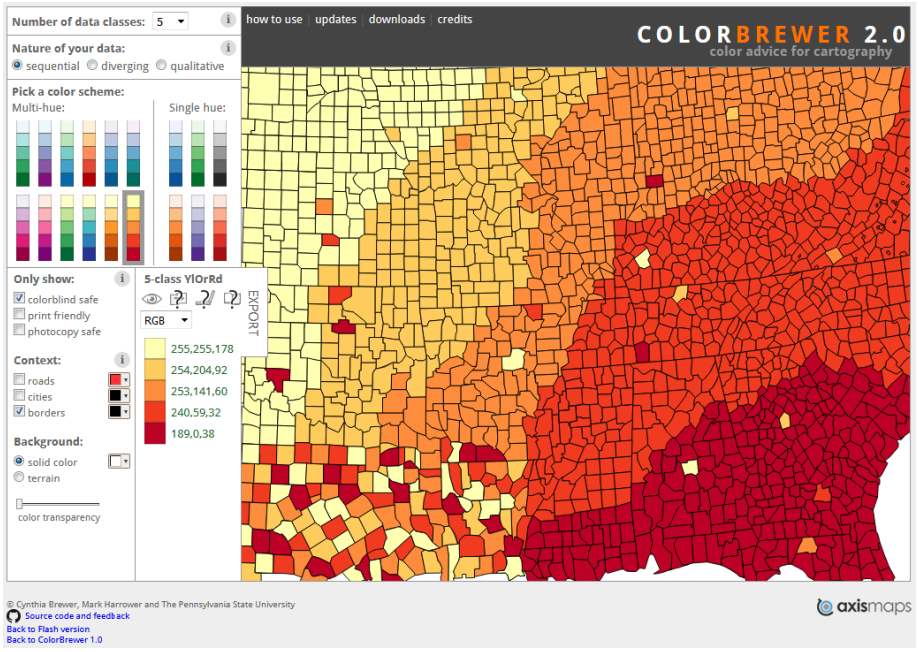
Figure 5.4: Screenshots of *MeVisLab* for doing registration and deformation.

5.4 Non-imaging demographic data

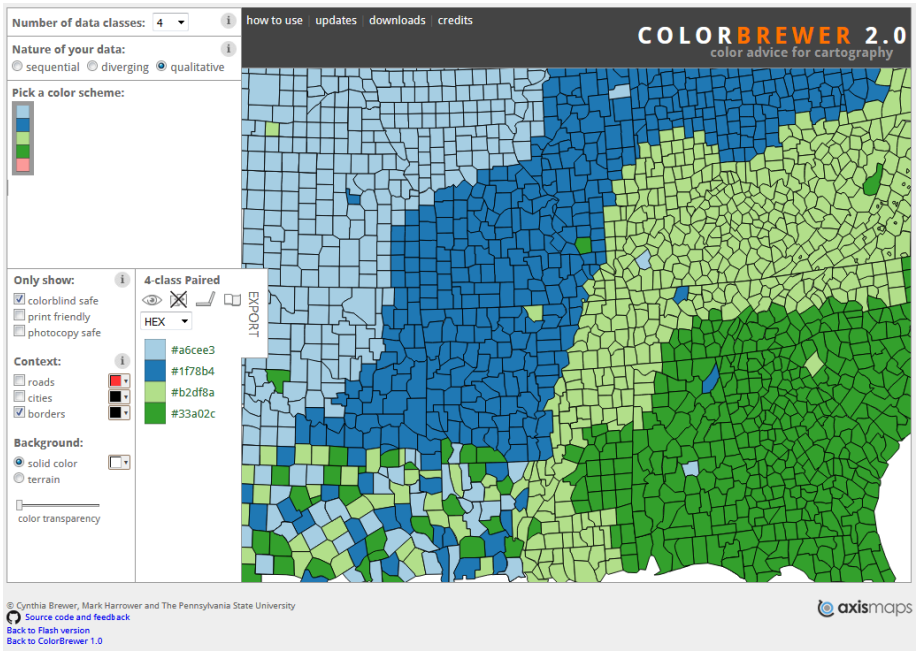
As described in Section 4.4 and Section 4.7 various visualizations based on D3.js [Bos] to display non-imaging demographic data were implemented. For this the relevant data were obtained out of the database and prepared to be processed by the charts. The coloring of the data, such as the highlighting of the data within the visualizations and the color-based differentiation of the patients or groups, was based on the tool *Colorbrewer.org* [HB03] (see example screenshots in Figure 5.5).

5.5 Connection of tools

As described, many different tools were used for the implementation of this thesis. The connection between these tools were mostly based on the database, which was designed and implemented first. With the possibility to query all the imaging and non-imaging data it was easier to dynamically generate the required source code files for displaying the various visualizations and charts. However, for image processing of DICOM files, such as registration and tumor probability map calculation (as explicitly described in Section 4.2), we considered *MeVisLab* [AG] as the best tool to do these preprocessing steps. Since MeVisLab provides the opportunity to export results of processing steps as DICOM files, it was no problem to do these steps with an external tool.



(a) Screenshot of Colorbrewer.org settings for the tumor probability map colors, that were used in the tumor visualization in the intra-patient study. URL: `http://colorbrewer2.org/?type=sequential&scheme=YlOrRd&n=5`



(b) Screenshot of Colorbrewer.org settings for the tumor probability map colors, that were used in the tumor visualization in the pairwise inter-patient study. URL: `http://colorbrewer2.org/?type=qualitative&scheme=Paired&n=4`

Figure 5.5: Screenshot examples of Colorbrewer.org [HB03] settings.

CHAPTER 6

Results

In this chapter we will show representative use cases for each task we implemented. We will show the workflow of how a researcher will achieve results with the framework and subsequently discuss these results. We will present results for intra-patient studies, pairwise inter-patient studies, and groupwise inter-patient studies and will depict and explore various charts to analyze them.

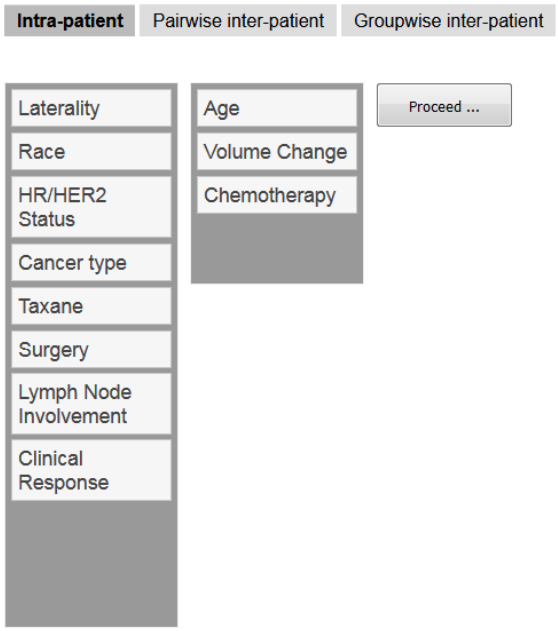


Figure 6.1: Selection of criteria to create divisions according to age, volume change and chemotherapy

6.1 Intra-patient study

In the first task, which is described in detail in Section 4.4, a clinical researcher has the possibility to explore and analyze follow-up data of one patient in the cohort and partially compare the patient to the cohort.

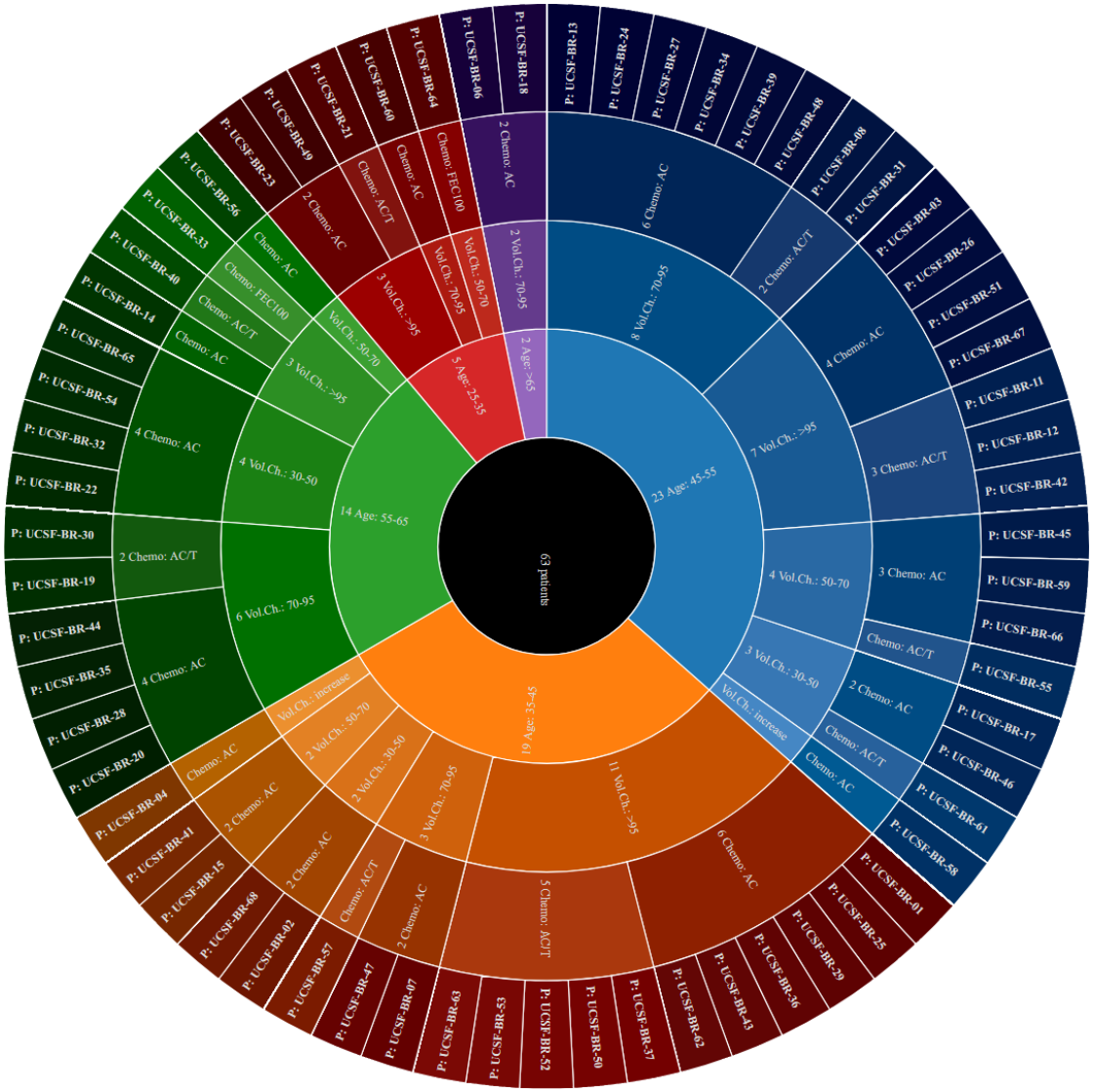


Figure 6.2: Interactive selection wheel showing cohort division by age, volume change, and chemotherapy for the intra-patient study.

6.1.1 Cohort division by age, volume change, and chemotherapy

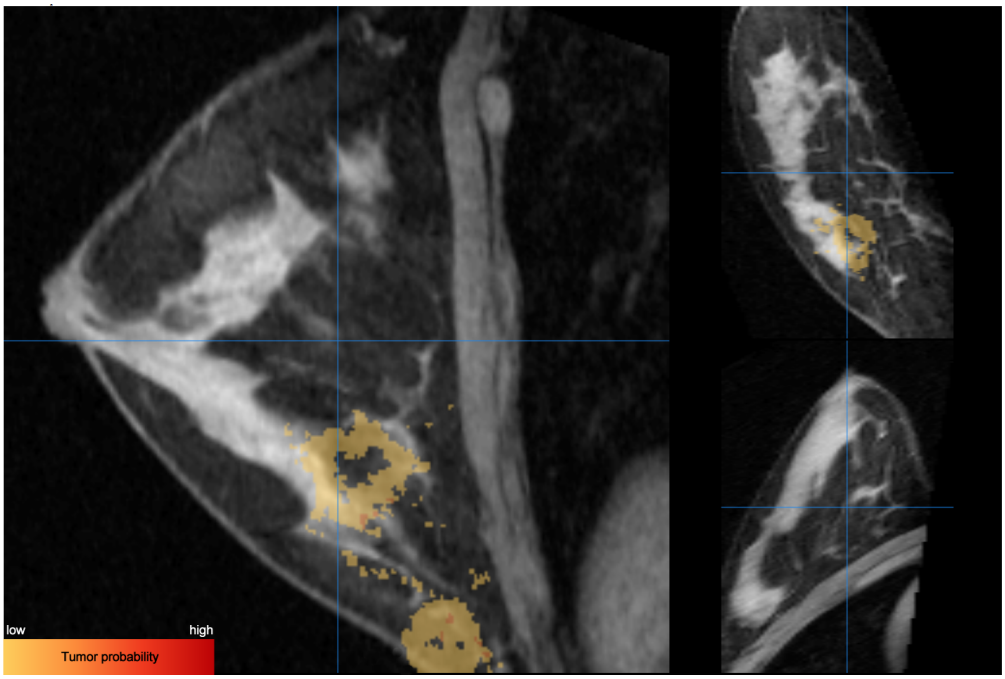
We want to show the result of dividing the cohort by the criteria *age*, *volume change*, and *chemotherapy*. As shown in Figure 6.1, the user brings the desired criteria in the right order and by clicking *Proceed ...* the wheel for selecting a patient will be created, which can be seen in Figure 6.2. As seen in this figure, the center of the wheel contains all patients in the cohort, and each step from the center to the edge represents a division. Just by observing this selection wheel the researcher can already draw a couple of conclusions on the cohort:

- Most of the patients in the cohort are 35 to 55 years old. 30% (19 out of 63) are between 35 and 45 (orange area), and 36.5% (23 out of 63) are aged 45 to 55 (blue area).
- Mostly there is a large change (more than 70%) in the tumor volume. The divisions *Volume Change 70-95%* and *Volume Change >95%* affect the larger part of each age division: 15 out of 23 in the age group 45-55 (65.2%, blue), 14 out of 19 in the age group 35-45 (73.7%, orange), 9 out of 14 in the age group 55-65 (64.3%, green), 4 out of 5 in the age group 25-35 (80%, red) and all patients of the age group >65 (100%, purple).
- Most of the patients received only Anthracycline as chemotherapy (shown by *Chemo: AC*), some also received Taxane (*Chemo: AC/T*) and two patients were undergoing other chemotherapy techniques (*Chemo: FEC100*).

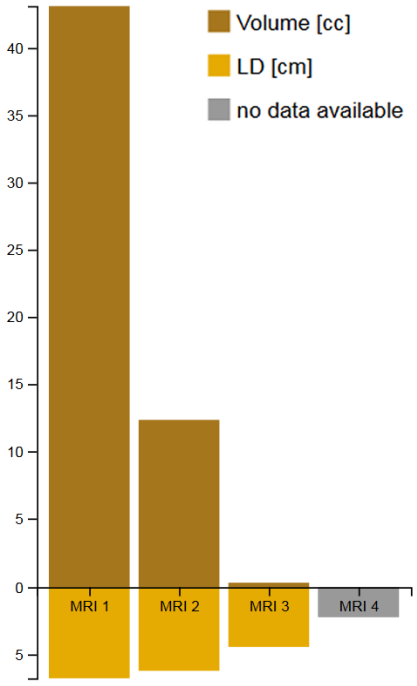
Based on these criteria we decided to do an exploration and analyze some cases we considered to be possibly interesting for a clinical researcher using this framework. Since the visualizations are interactive, not all features can be shown in screenshots. For example, the exact tumor volume and LD values can be preserved by moving the mouse pointer over the concerning bar in the Bar chart.

6.1.2 Young patient

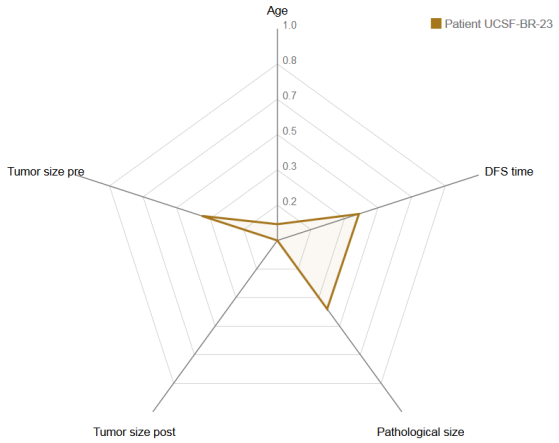
The first patient we want to explore is a very young patient, belonging to the *age 25-35* division and having a large tumor volume change (>95%). We chose a patient not having received Taxane as additional chemotherapy with the purpose to compare this patient later to an old patient (age >65), from a division where no patient received Taxane. Regarding the *Bar chart* (see Figure 6.3b) a user can easily spot the large tumor volume drop going from 43.05 cc in MRI_1, to 12.41 cc in MRI_2 (which already is a change of 71.2%), to finally 0.27 cc in MRI_3, which is a final volume change of 99.4%. Interestingly the longest diameter (LD) did not change to this extent (6.8 cm in MRI_1, 6.2 cm in MRI_2 (8.8%), 4.5 cm in MRI_3 (33.8%). This could possibly mean that the shape of the tumor might be elongated, which can be verified by viewing the tumor visualization in Figure 6.3a. When looking at the *Radar chart* (Figure 6.3c) the large change of the



(a) Tumor segmentation visualization of young patient

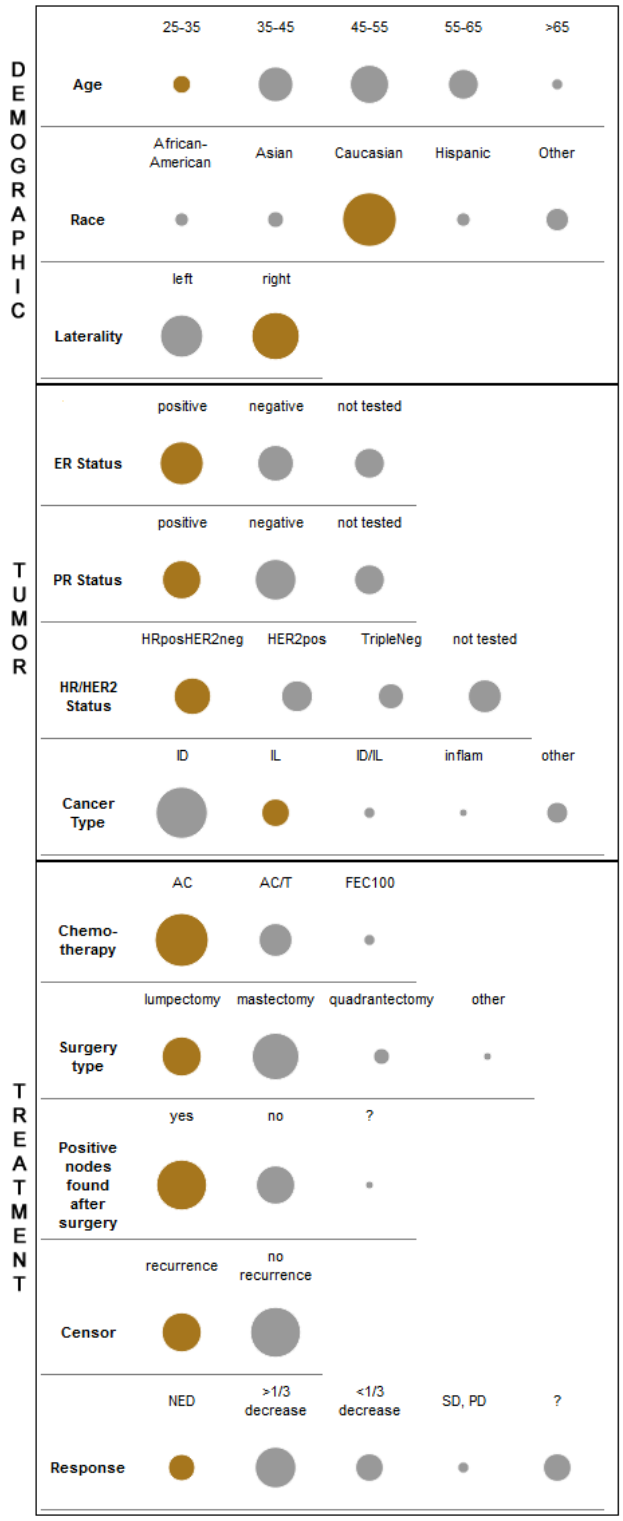


(b) Bar chart of young patient showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of young patient

Figure 6.3: Example charts for the exploration of a young patient (age 25-35).



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.4: Spotmatrix of young patient

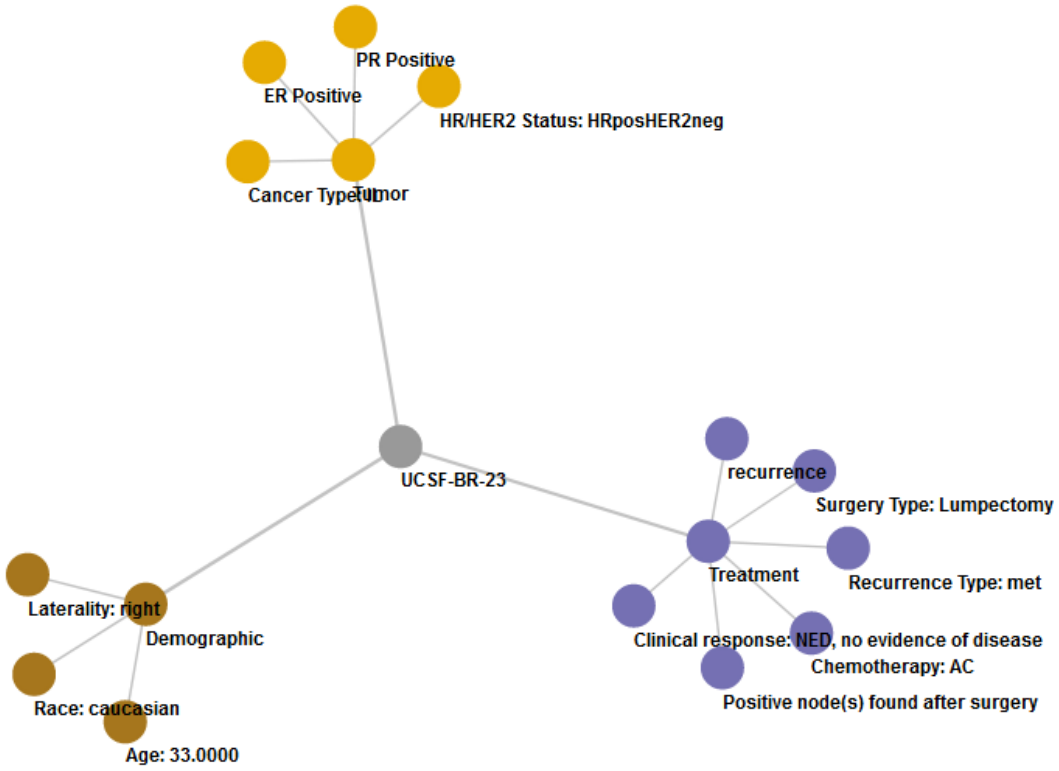
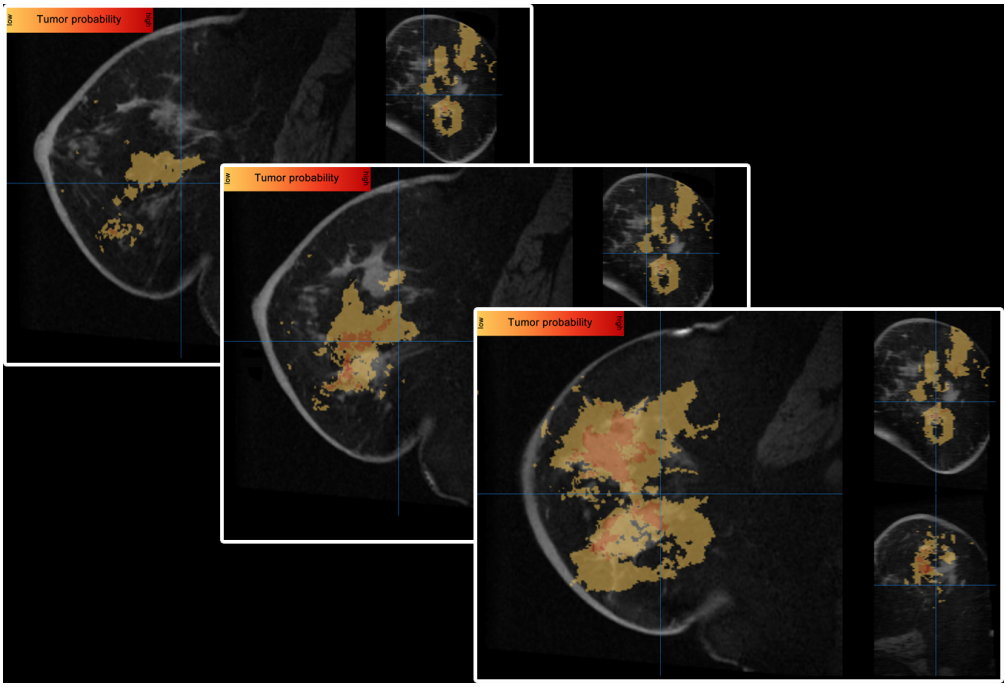


Figure 6.5: Force-directed graph of young patient

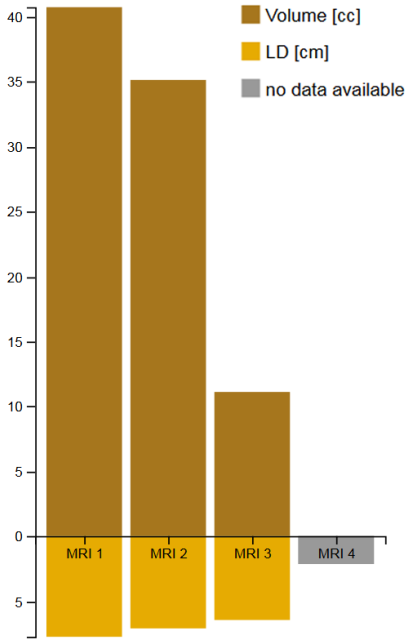
tumor size can also be seen when analyzing the difference between *Tumor size pre* and *Tumor size post* (very large drop). That the patient is one of the youngest patients in the cohort can be seen on the *age* axis in this chart. Although this patient is in one of the smallest age divisions, most of the other divisions (such as race, breast cancer laterality, chemotherapy, etc.) the patient belongs to, are composed of a large number of patients in the cohort, which can be explored in the *Spotmatrix* (Figure 6.4). An overview of all demographic data and their relation to each other can easily be spotted in the *Force-directed graph* (Figure 6.5), which can be used by the researcher as preparation for a possible comparison to another patient.

6.1.3 Old patient

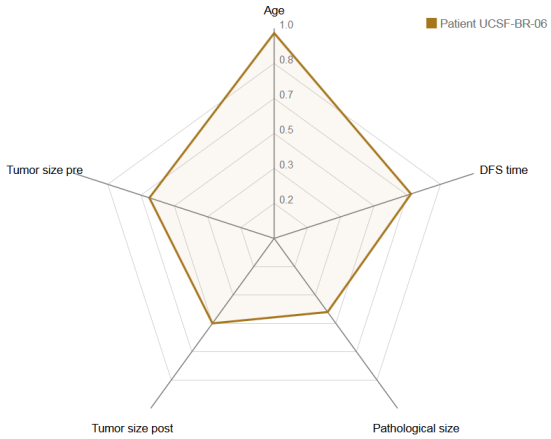
In contrast to the aforementioned young patient, we want to explore a contrasting example, i.e. one of the oldest patients in the cohort (age >65), also having a large tumor volume change (70-95%) and not having received Taxane as additional chemotherapy. The large volume drop can be spotted in the *Bar chart* (see Figure 6.6b). From MRI_1 (40.72 cc) to MRI_2 (35.12 cc) the decrease is only 13.8%, but MRI_3 (11.09 cc) shows already a drop of 72.8%. The LD did not change a lot (7.70 cm in MRI_1 to finally 6.46 cm in MRI_3), which could be an indication of a tumor reduction in only one or two spatial directions. Using the mouse wheel, scrolling through the slices in the tumor segmentation (Figure 6.6a) could help verifying this statement. Regarding the *Radar chart* (see Figure 6.6c), a researcher can easily see that the age of the patient is one of the highest in the cohort (age axis at highest value). When analyzing the *Spotmatrix* (see Figure 6.7), it can be recognized that concerning demographic data this patient is kind of an outsider, but concerning treatment and tumor related data the patient belongs to the majority of the cohort in nearly every property. Finally, as already described for the young patient, the *Force-directed graph* (see Figure 6.8) can be used to get an overview of all non-imaging data to prepare for a possible comparison of this patient to another one.



(a) Various slices of the tumor segmentation visualization of old patient

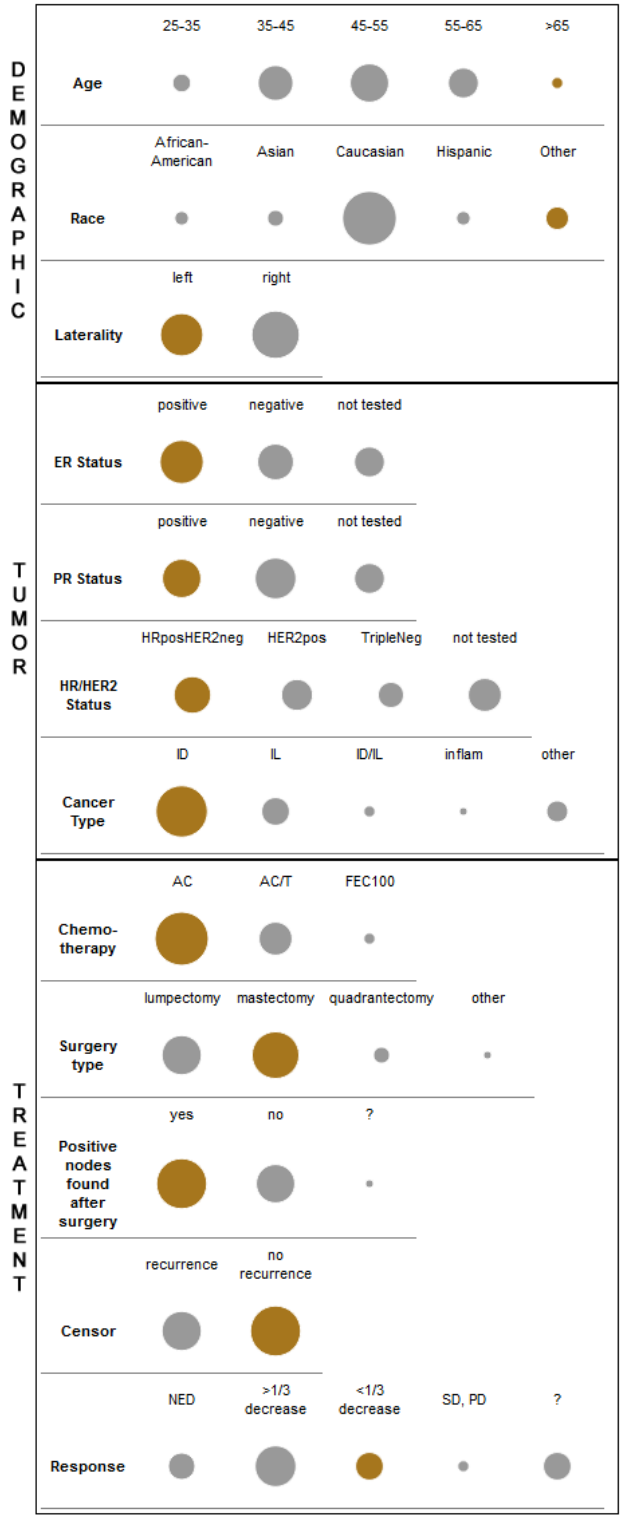


(b) Bar chart of old patient showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of old patient

Figure 6.6: Example charts for the exploration of an old patient (age >65).



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.7: Spotmatrix of old patient

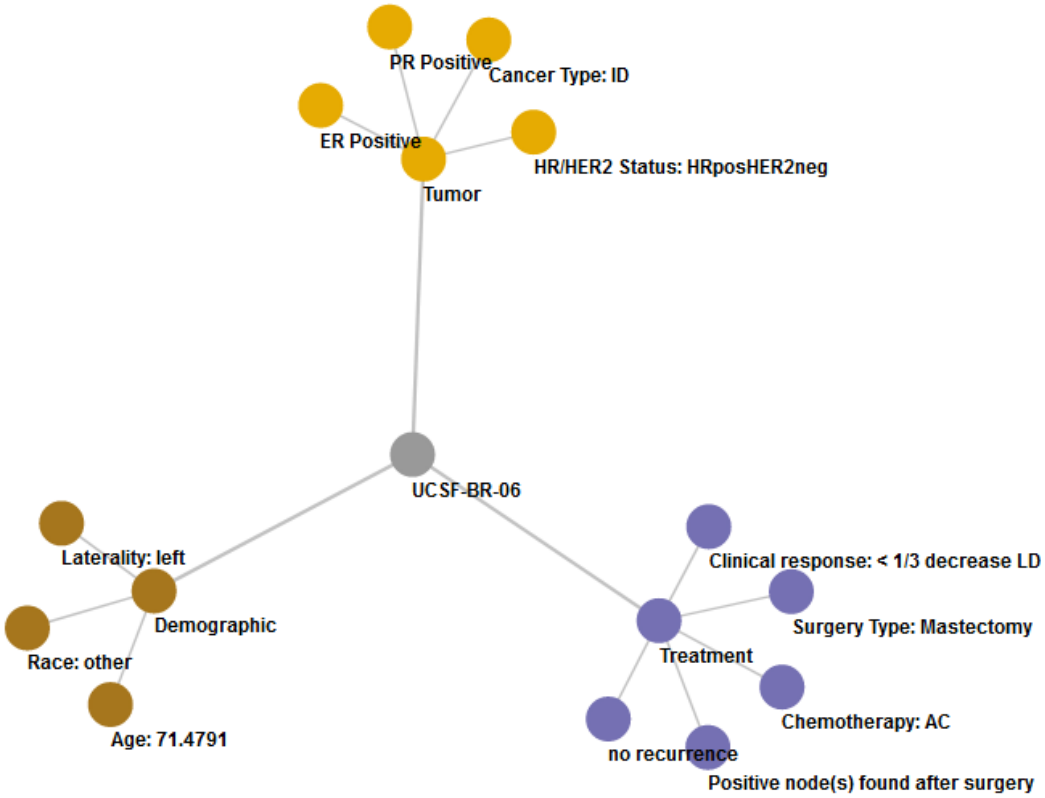
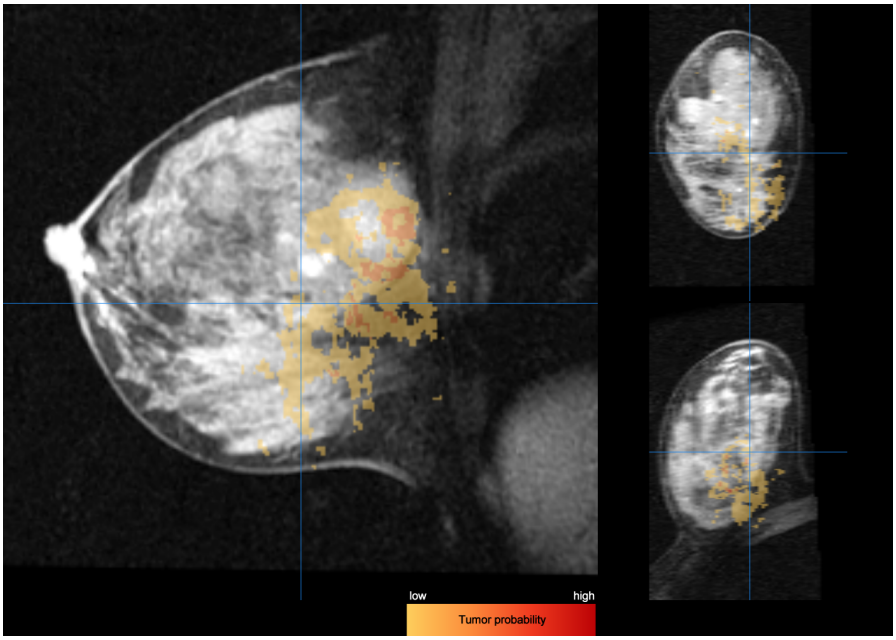


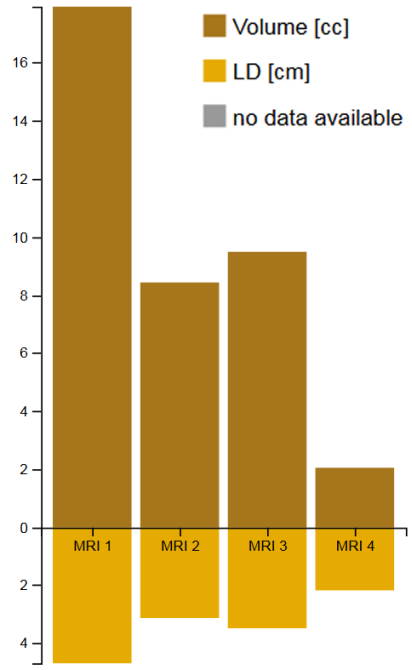
Figure 6.8: Force-directed graph of old patient

6.1.4 Patient having received Taxane

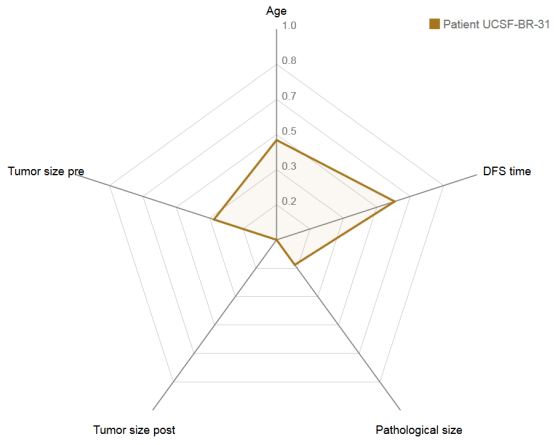
In this section we want to explore one of the patients who received Taxane as additional chemotherapy after completing the Anthracycline therapy. These patients were scheduled for a fourth MRI scan, thus we can see four bars in the *Bar chart*. Another criterion to choose this patient was a large tumor volume decrease. The tumor segmentation visualization (Figure 6.9a) shows that the tumor in this patient is situated proximal (close to the body). In the *Bar chart* (Figure 6.9b) this big volume change from MRI_1 (17.93 cc) to the last MRI (2.04 cc) can easily be recognized. Interesting in this case is a small tumor volume and LD increase in the third MRI scan. The volume increased from 8.44 cc to 9.51 cc (+12.7%) and the LD from 3.14 cm to 3.5 cm (+12.5%). Nevertheless, obviously the additional treatment with Taxane seems to have been successful in this patient, so the final tumor volume in MRI_4 was only 2.04 cc, which means a decrease of 88.6% since the first MRI scan and of 78.5% since MRI_3. Also the LD decreased from 4.69 cm (MRI_1) to finally 2.19 cm (-53.3%), when regarding only the Taxane cycle (time period between MRI_3 and MRI_4) the LD decreased from 3.5 cm to 2.19 cm (-37.4%).



(a) Tumor segmentation visualization of patient with Taxane treatment

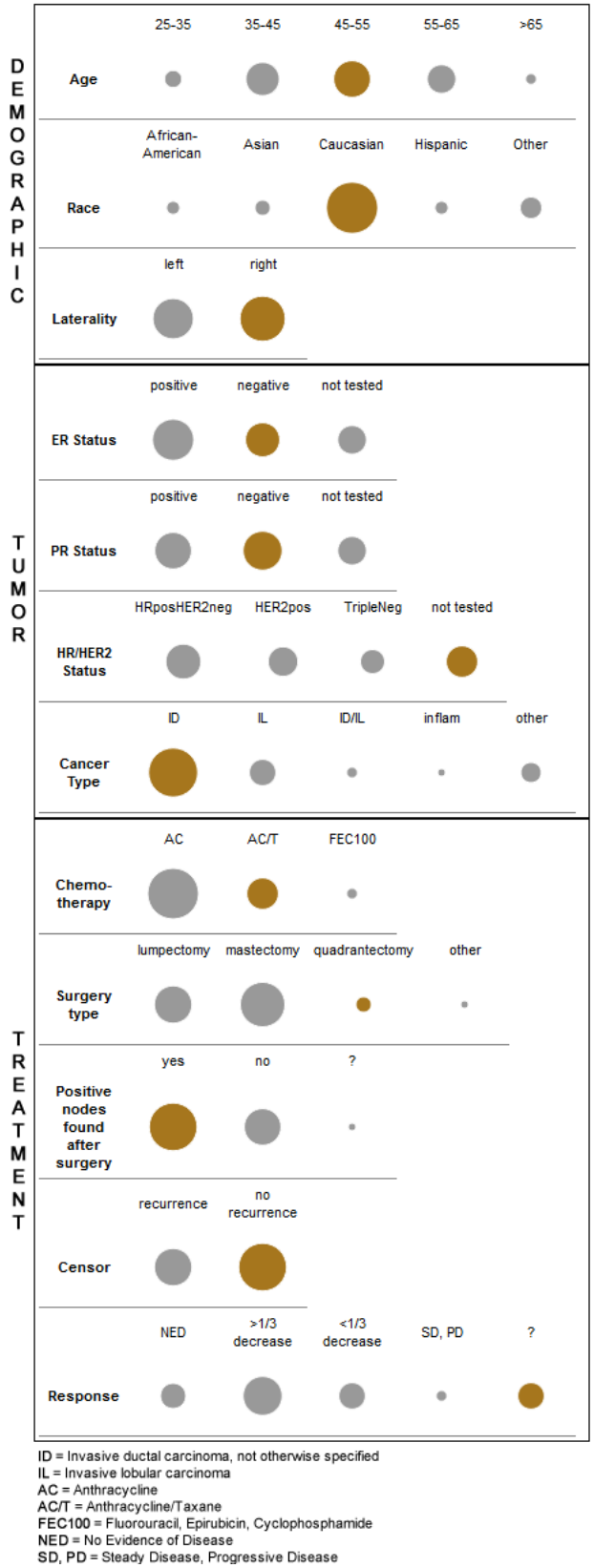


(b) Bar chart of patient with Taxane treatment showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of patient with Taxane treatment

Figure 6.9: Example charts for the exploration of patient with Taxane treatment.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.10: Spotmatrix of patient with Taxane treatment

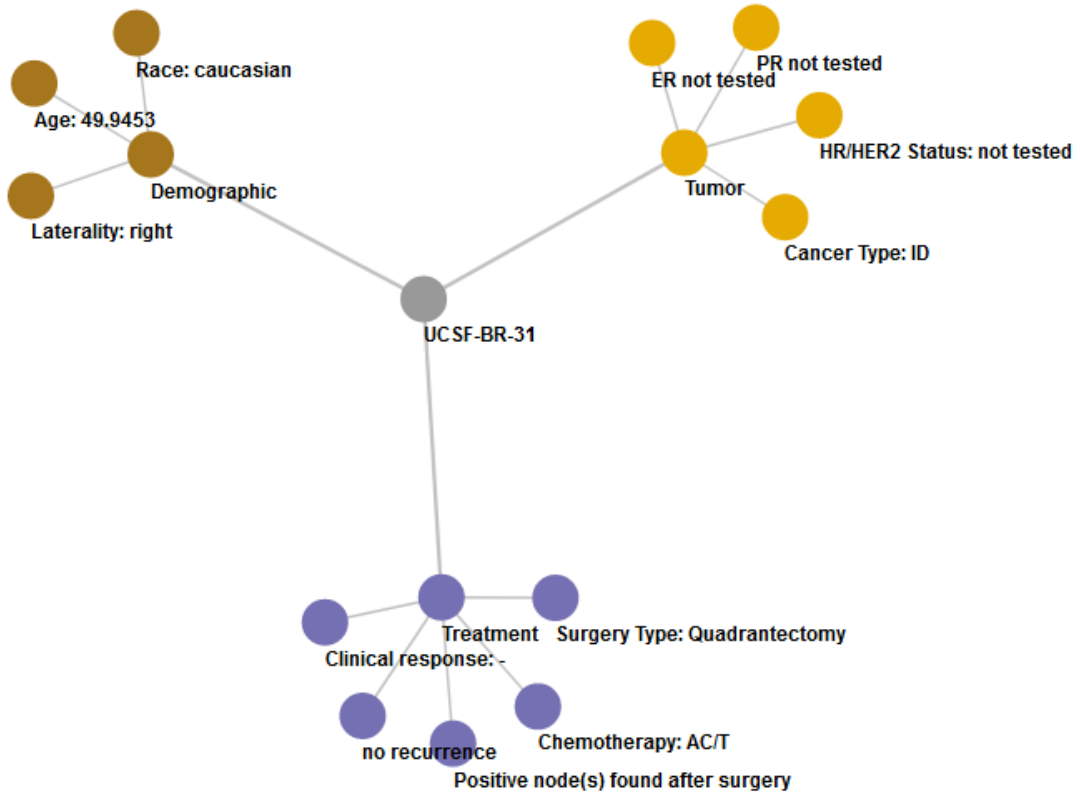
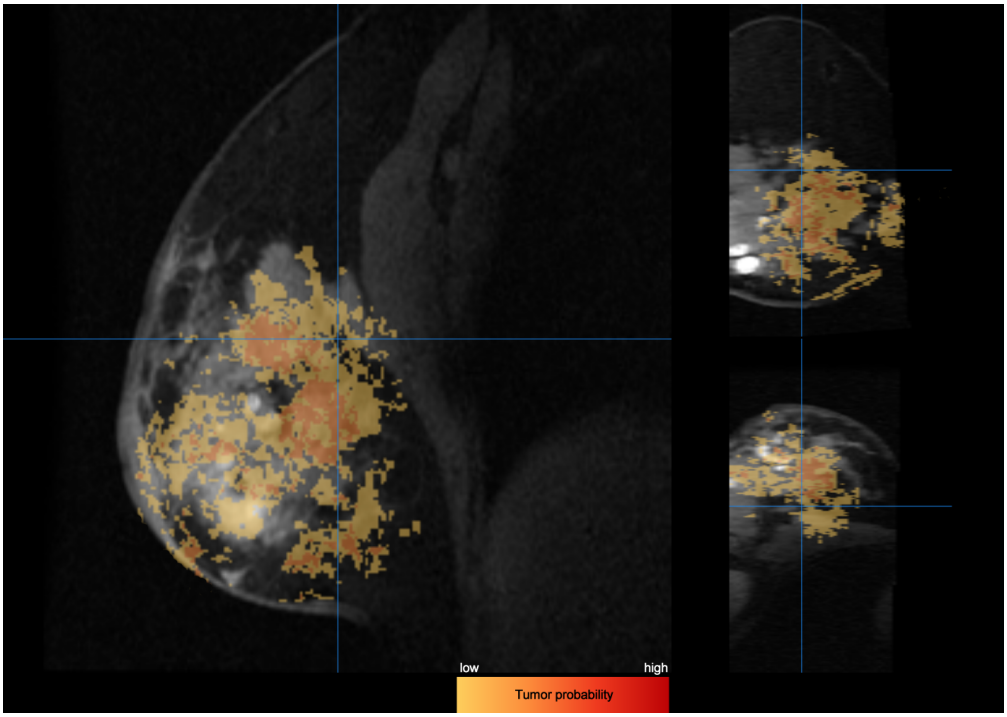


Figure 6.11: Force-directed graph of patient with Taxane treatment

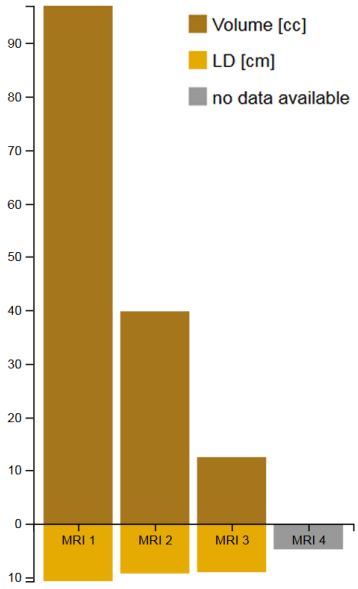
When looking at the *Radar chart* (Figure 6.9c), it also can be seen that the tumor size changed a lot from pre-chemo to post-chemo measuring. The additional Taxane chemotherapy can be seen in the *Force-directed graph* (Figure 6.11) - *Chemotherapy: AC/T* is one of the nodes in the treatment group. That the group of patients having received Taxane is smaller than the one having only received Anthracycline or another chemotherapy can be seen in the *Spotmatrix* (Figure 6.10) - the highlighted circle *AC/T* in the chemotherapy section is smaller than the *AC*-circle.

6.1.5 Patient having only received Anthracycline

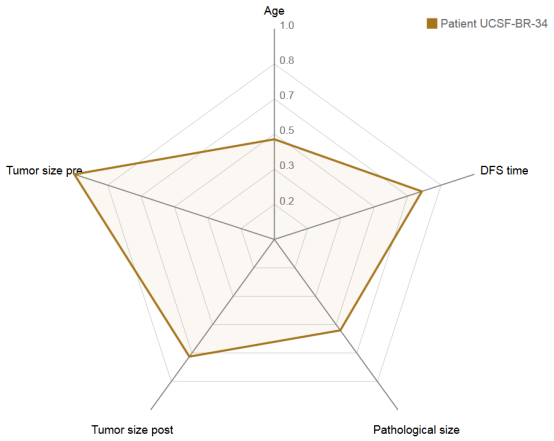
Next a patient having only received Anthracycline as chemotherapy will be analyzed. We selected a patient having a large tumor volume drop in order to compare this patient later on to the aforementioned patient, who received Taxane after MRI_3. The tumor segmentation visualization (Figure 6.12a) shows that the tumor is distributed over a large area of the breast. The large tumor volume decrease can easily be perceived in the *Bar chart* (Figure 6.12b). The tumor volume develops from 97.00 cc in MRI_1, over



(a) Tumor segmentation visualization of patient with only Anthracycline treatment

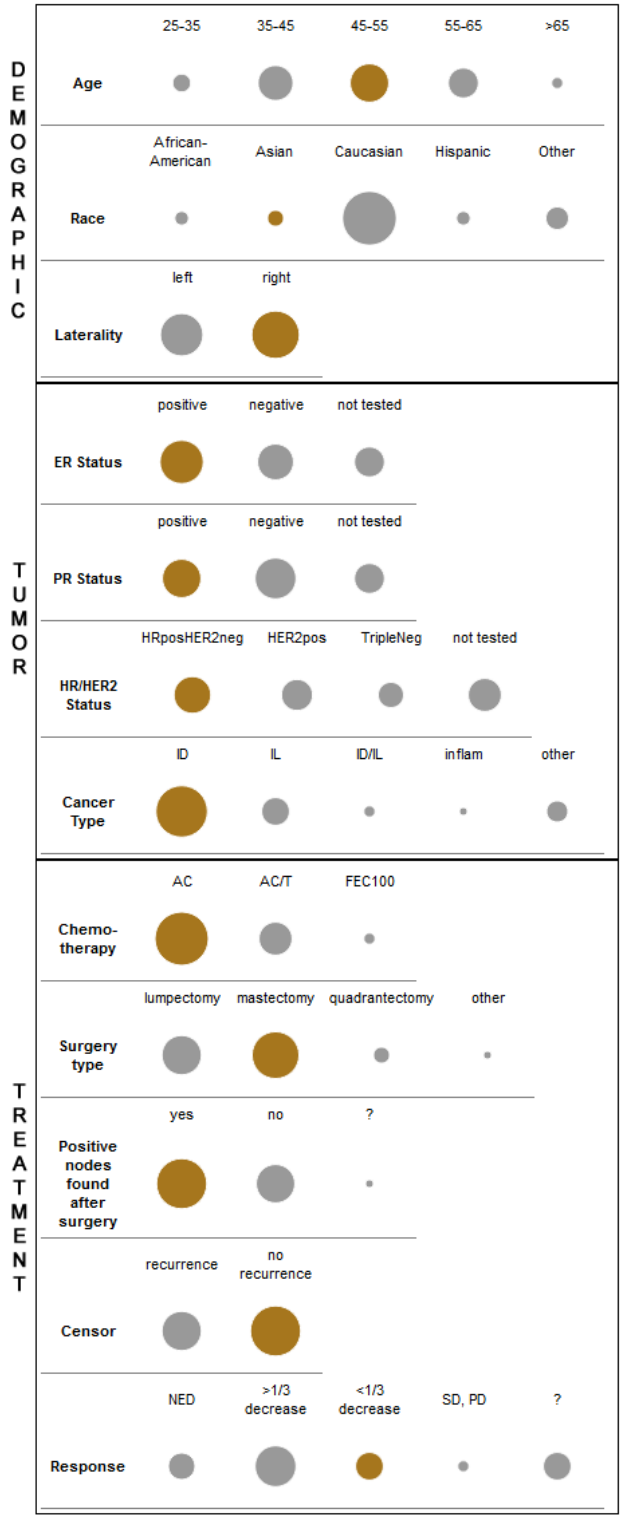


(b) Bar chart of patient with only Anthracycline treatment showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of patient with only Anthracycline treatment

Figure 6.12: Example charts for the exploration of patient with only Anthracycline treatment.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.13: Spotmatrix of patient with only Anthracycline treatment

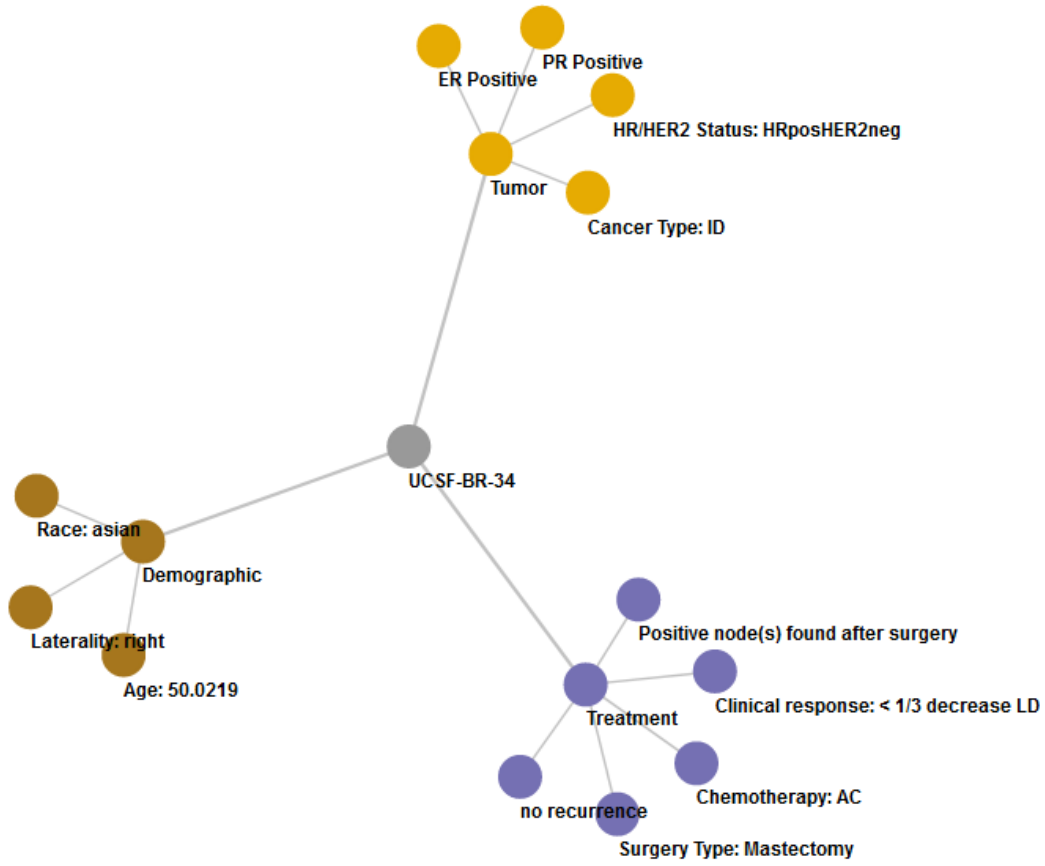
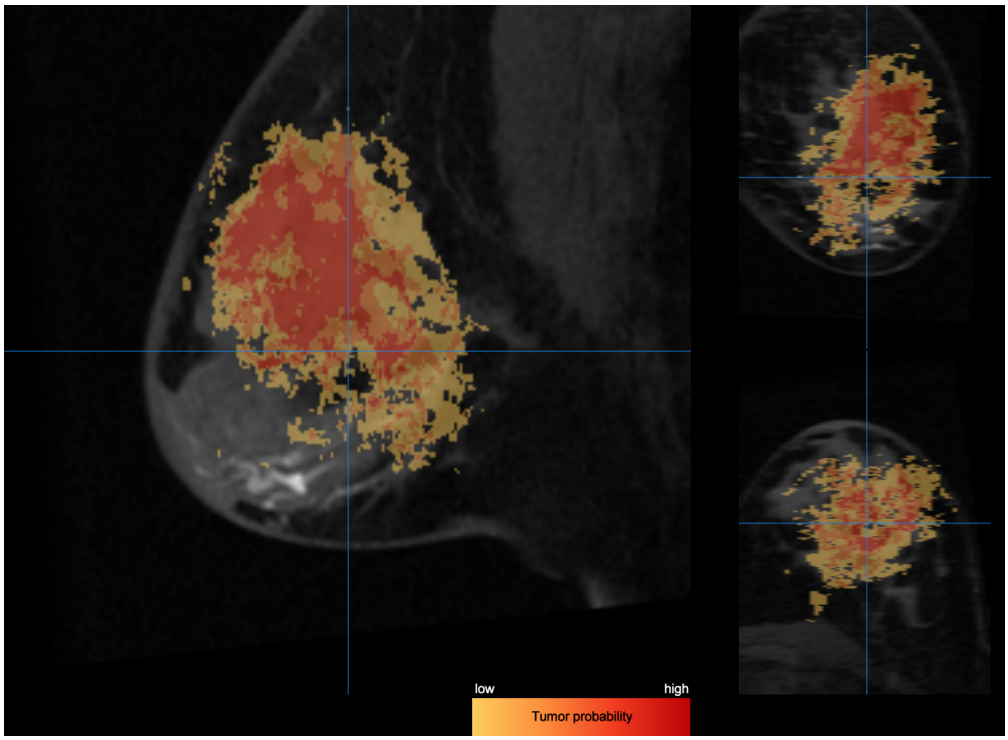


Figure 6.14: Force-directed graph of patient with only Anthracycline treatment

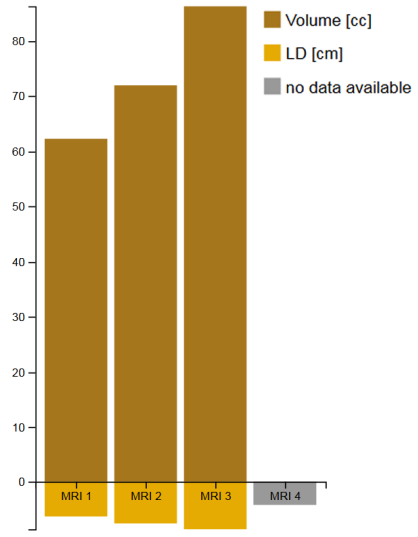
39.71cc in MRI_2 (-59.1%), to 12.39 cc in MRI_3 (-87.2%). Nevertheless, the LD did not change very much (10.76 cm in the first MRI scan to 8.98 cm in the third MRI scan, which means a drop of only 16.5%) meaning that the tumor possibly decreased only in one or two spatial directions. When regarding the *Radar chart* (Figure 6.12c) the decrease of the tumor size can be seen when comparing the *Tumor size pre* and *Tumor size post* values. Another conclusion from this chart is that the patient is about average age in comparison to the cohort. The *Spotmatrix* (Figure 6.13) shows among other things that the patient is Asian, which is rare in this cohort (illustrated by the small circle). Again, all non-imaging demographic data can be observed in the *Force-directed graph* (Figure 6.14) in organized groups.

6.1.6 Patient with tumor volume increase

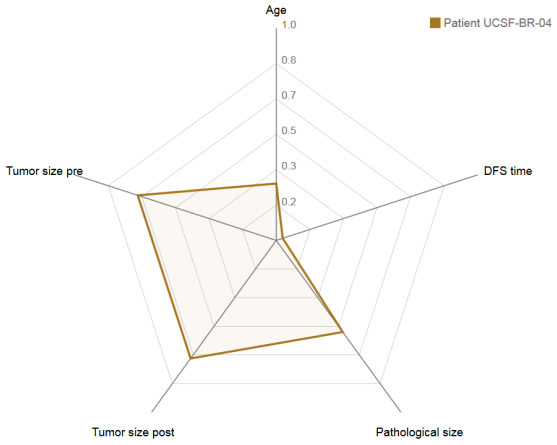
In this section we will explore a patient who had a tumor with an increasing size during the study. This growth can be figured out by analyzing the *Bar chart* (Figure 6.15b).



(a) Tumor segmentation visualization of patient with tumor volume increase

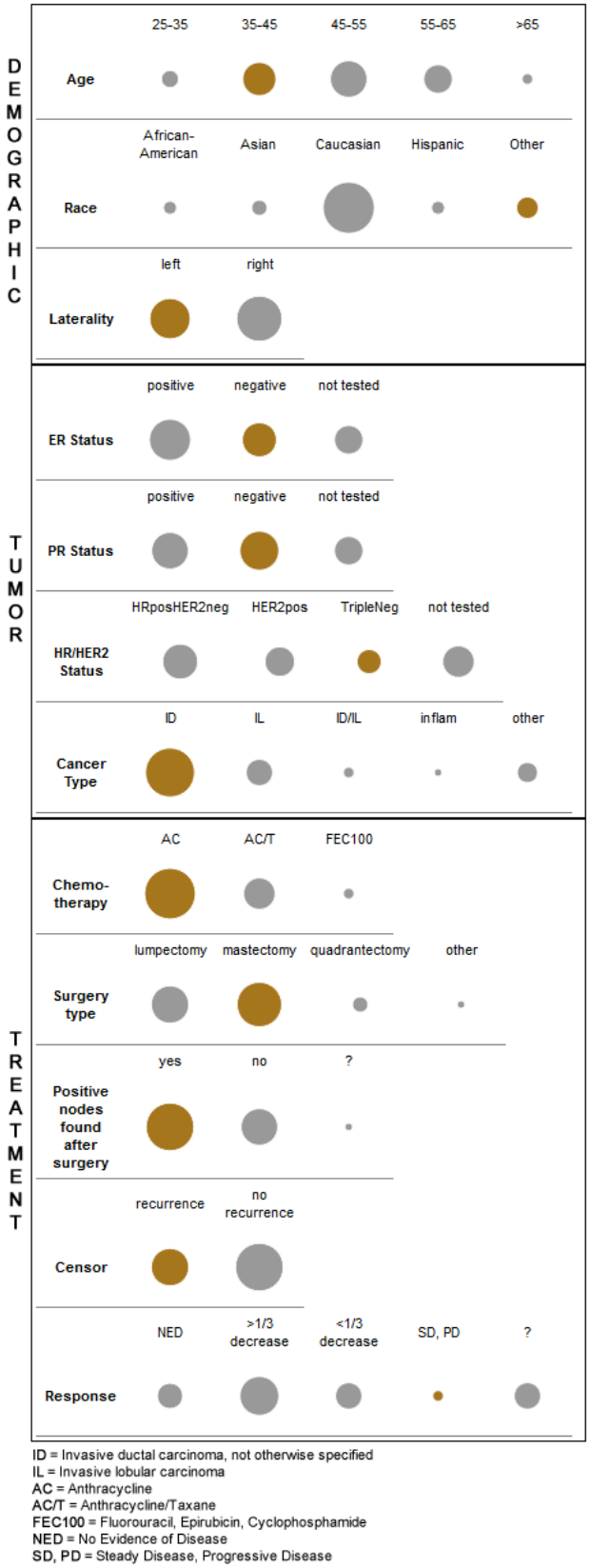


(b) Bar chart of patient with tumor volume increase showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of patient with tumor volume increase

Figure 6.15: Example charts for the exploration of patient with tumor volume increase.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.16: Spotmatrix of patient with tumor volume increase

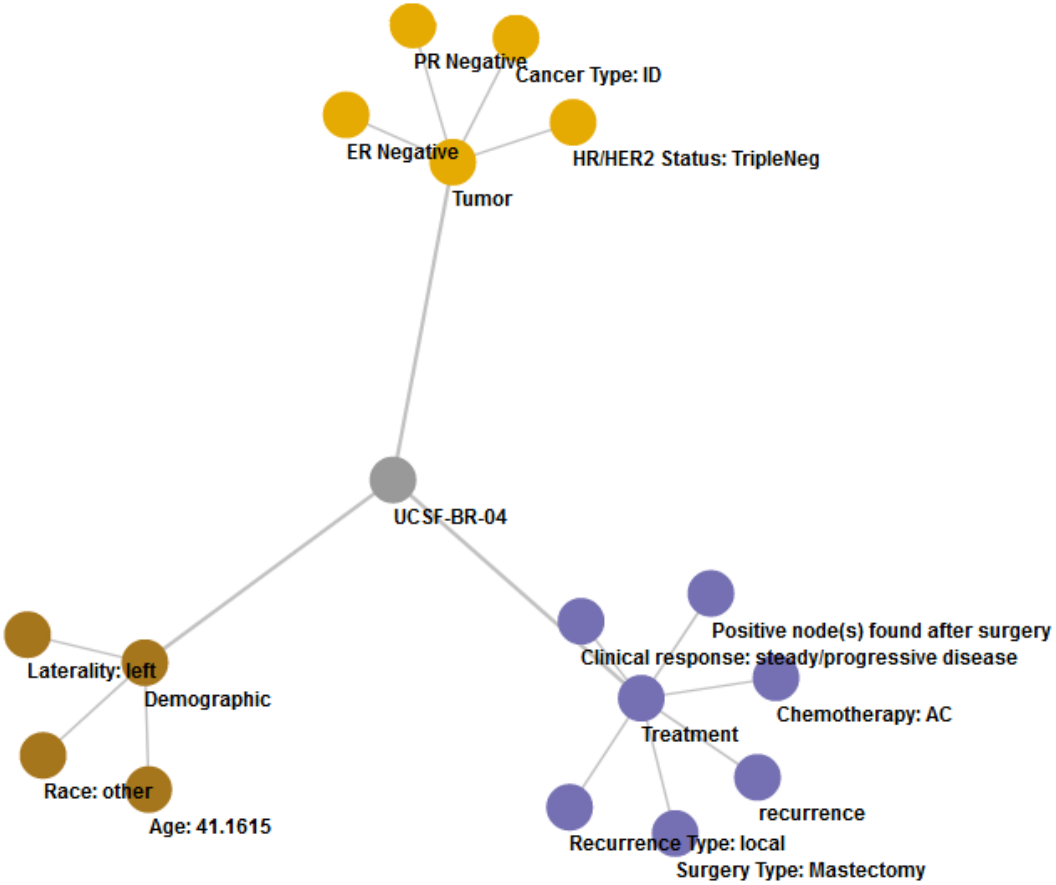
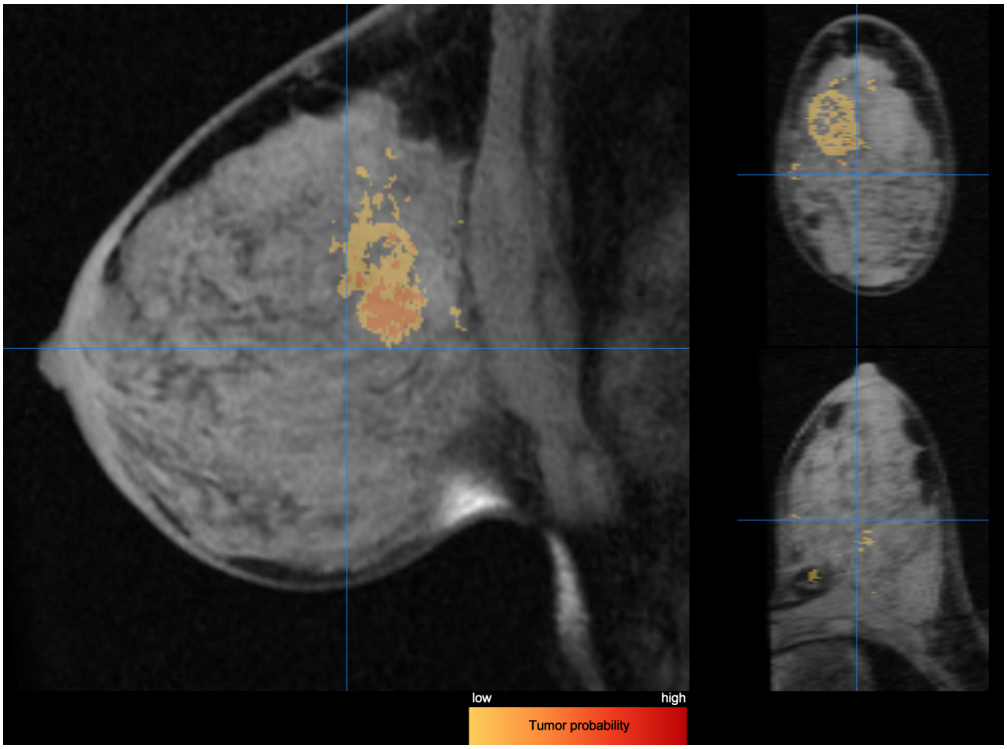
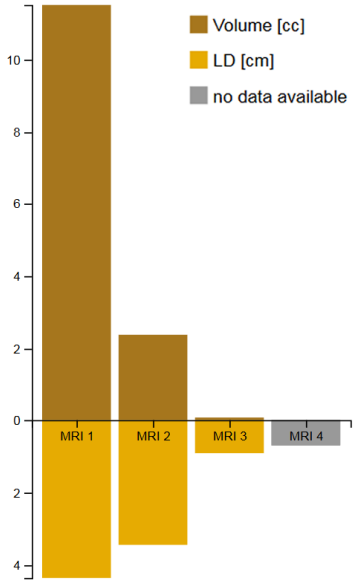


Figure 6.17: Force-directed graph of patient with tumor volume increase

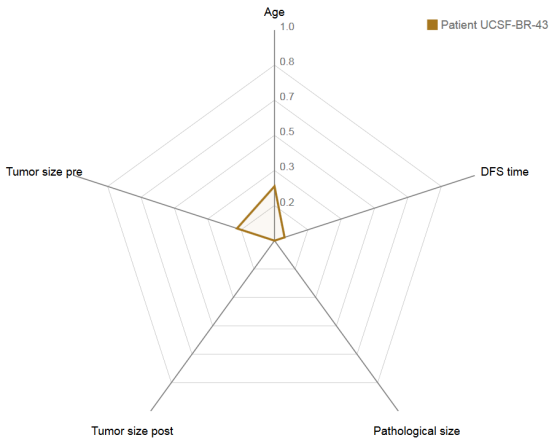
Since this patient did not receive Taxane as additional chemotherapy, there is no data for MRI_4, which was only scheduled for patients receiving Taxane. This can be spotted in the *Spotmatrix* in Figure 6.16 or in the *Force-directed graph* in Figure 6.17. In the tumor segmentation visualization (Figure 6.15a) it can be seen that the tumor in this patient is very large. The tumor volume was increasing from initially 62.29 cc in MRI_1, over 71.86 cc in MRI_2 (+15.4%) to 86.29 cc in MRI_3 (+38.5% compared to MRI_1). The LD had about the same percentage growth like the tumor volume, increasing from 6.23 cm in the first MRI scan to finally 8.55 cm in MRI_3 (+37.2%). In the *Radar chart* (Figure 6.15c) it can be seen that this patient had one of the shortest disease-free survival (DFS) times in the cohort, which most probably implies that the patient died shortly after the end of the treatment. Other indications of this could be the clinical response (steady disease or progressive disease), the fact that positive nodes were found after surgery (both shown in the *Spotmatrix* in Figure 6.16) and that local recurrence was diagnosed (see *Force-directed graph* in Figure 6.17).



(a) Tumor segmentation visualization of patient with large tumor volume decrease



(b) Bar chart of patient with large tumor volume decrease showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart of patient with large tumor volume decrease

Figure 6.18: Example charts for the exploration of patient with large tumor volume decrease.

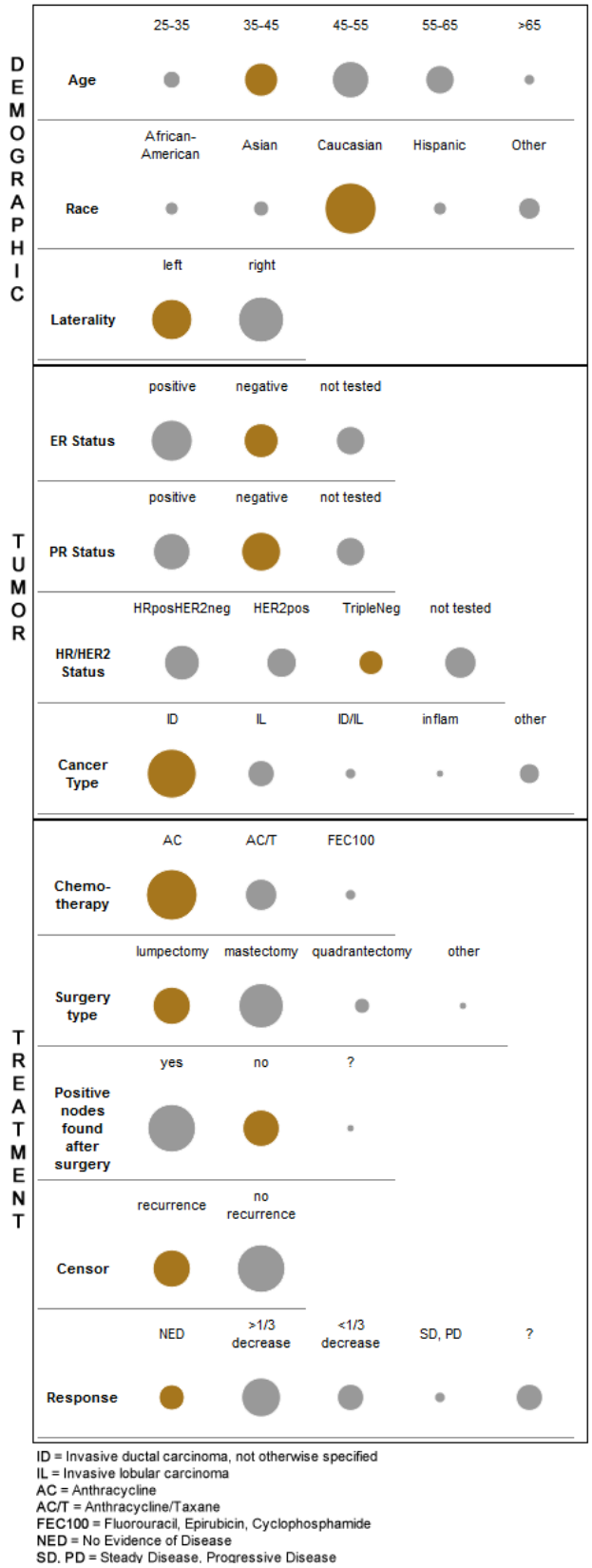


Figure 6.19: Spotmatrix of patient with large tumor volume decrease

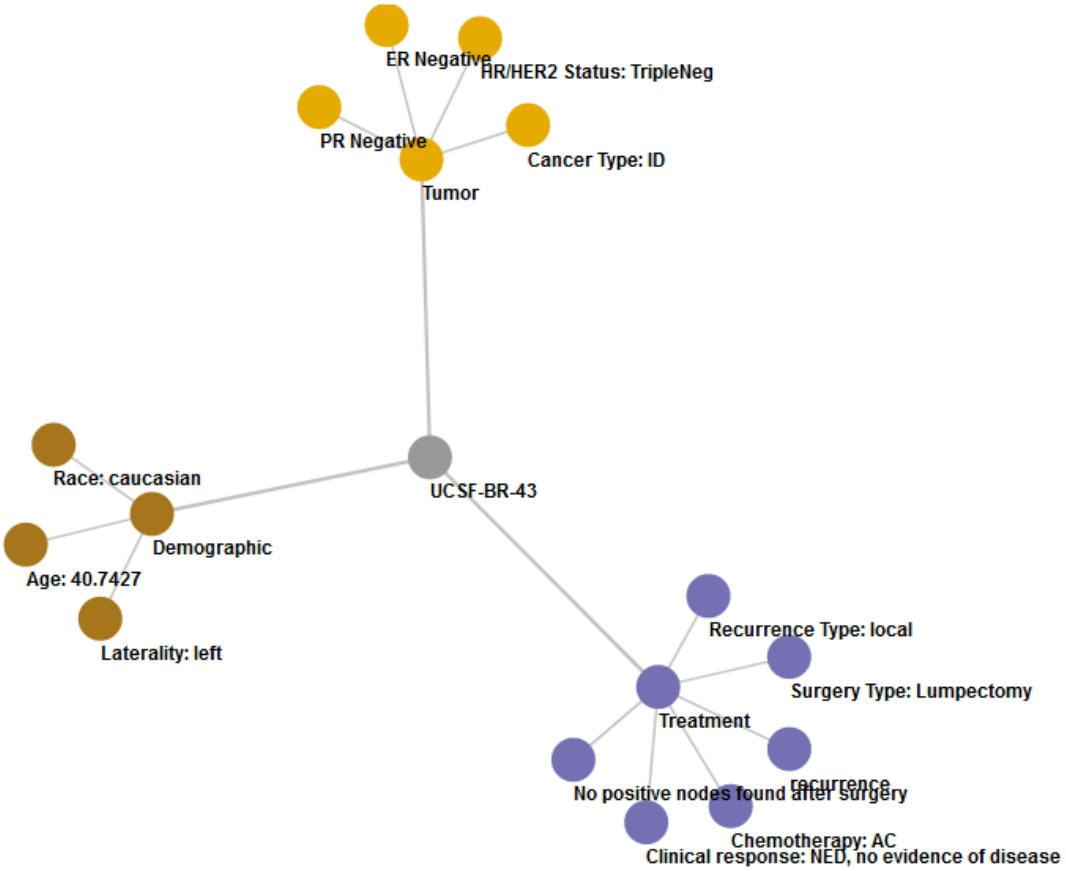


Figure 6.20: Force-directed graph of patient with large tumor volume decrease

6.1.7 Patient with large tumor volume decrease

The progress of the treatment of the patient we want to analyze in the following was different. This patient had a very large decrease of tumor volume and LD, which is shown in the *Bar chart* in Figure 6.18b. The tumor volume decreased from 11.51 cc in MRI_1 first to 2.39 cc in MRI_2 (-79.2%) and finally to 0.07 cc in MRI_3 (-99.4%). Also the LD dropped from 4.37 cm in the first MRI scan, over 3.44 cm in MRI_2 (-21.3%) to 0.90 cm in MRI_3 (-79.4%). Regarding the *Radar chart* (Figure 6.18c), the decrease of the tumor size can also be seen concerning this patient, as well as a low age. The tumor segmentation visualization (Figure 6.18a) shows that the tumor only affected a small area of the breast in this patient. In the *Spotmatrix* (Figure 6.19), especially when looking at the features concerning treatment, it can be seen that in this patient no positive nodes were found after surgery and the clinical response was *no evidence of disease (NED)*. An overview of all non-imaging demographic data can be seen in the *Force-directed graph* (Figure 6.20).

6.2 Pairwise inter-patient study

In the following we want to use the framework we introduced in the second task (Section 4.5), to compare each pair of patients we explored in Section 6.1. Therefore we created the selection wheel we introduced in the first task again, but with the intention to compare patients. Therefore we selected *Pairwise inter-patient* before arranging the desired criteria (as shown for *Intra-patient* in Figure 6.1).

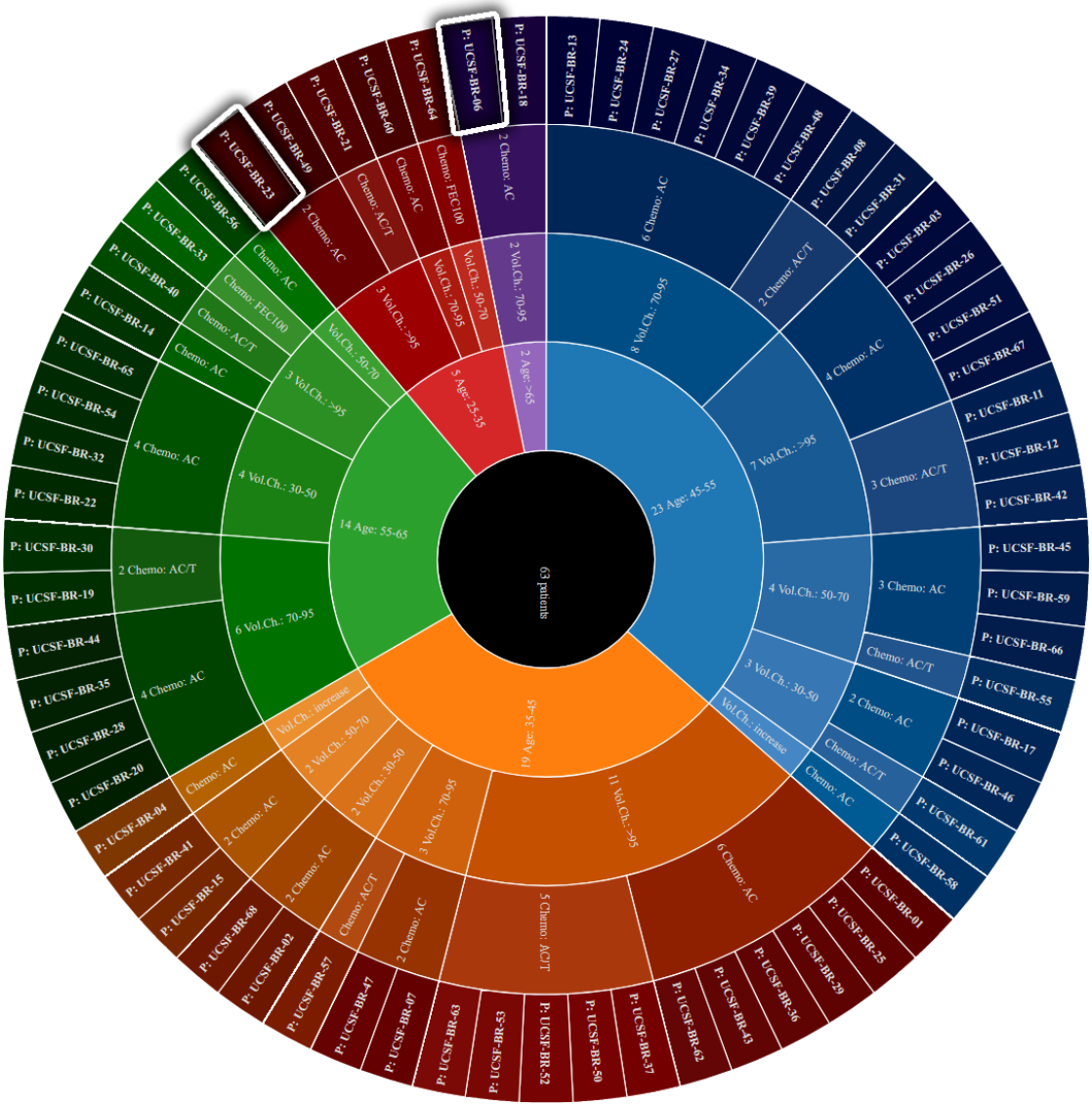
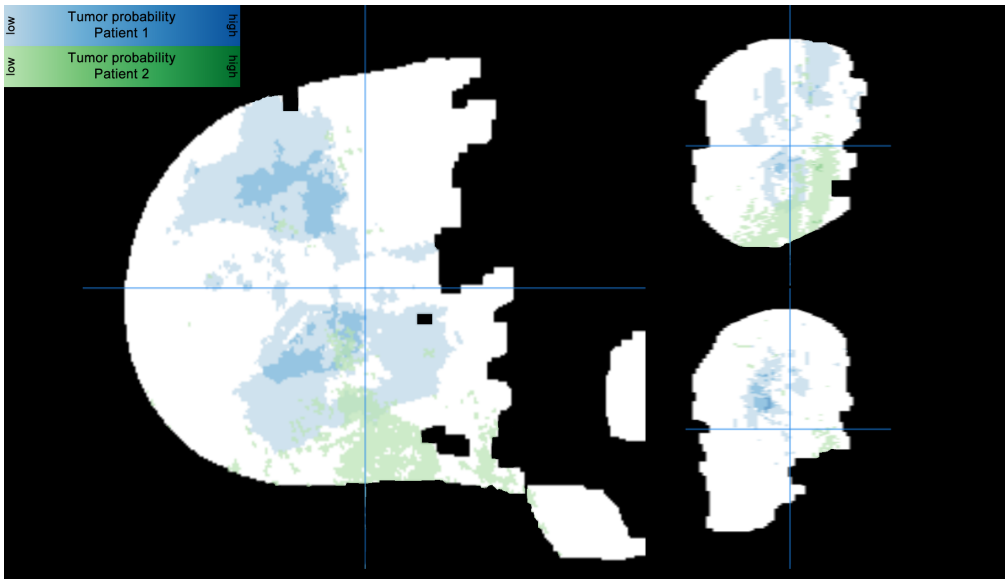


Figure 6.21: Selection wheel for pairwise inter-patient study comparing young and old patient. The selected patients are outlined in white.

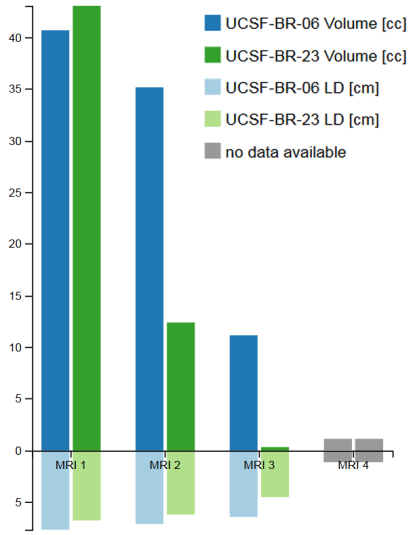
6.2.1 Young patient compared to old patient

The selection of these patients was done by successively clicking both patients in the selection wheel (Figure 6.21). We selected patients differing in their age - one being in the youngest age division (25-35), the other one in the oldest age division (>65). We selected two patients not having received Taxane and having a big tumor volume drop (more than 70%).

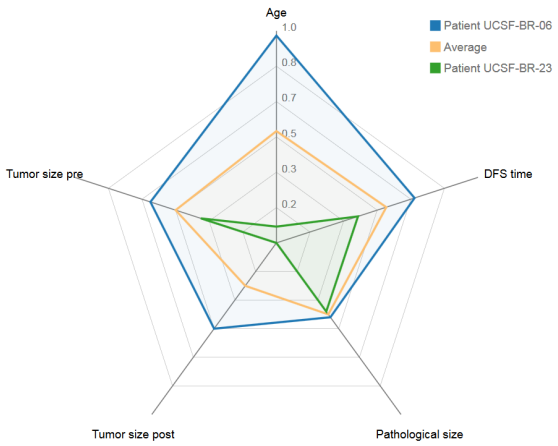
The overlapping probability map of the tumors can be seen in Figure 6.22a. When looking at the *Bar chart* (Figure 6.22b), it can be seen that both patients had about the same tumor volume in MRI_1 (40.72 cc for the older patient, colored in blue, and 43.05 cc for the younger patient, colored in green). The treatment of the younger patient can be characterized as faster and more effective than of the older patient, since a large volume drop can already be distinguished in MRI_2. Regarding the *Radar chart* (Figure 6.22c), the huge difference in the age of the patients can be recognized, as well as the difference in the post-chemo tumor size. Commonalities and differences can best be seen in the *Spotmatrix* (Figure 6.23). While these patients do not have anything in common concerning demographic data such as age, race, or breast laterality, the tumor-concerning data is almost the same when comparing the hormone receptor status of each tumor (both patients have positive values for all three hormone receptors). Also in both patients positive nodes were found after surgery. These commonalities and differences can also be recognized in the *Force-directed graph* (Figure 6.24). The nodes labeled *Recurrence* and *Clinical response* in the treatment-concerning part are split, showing that the young patient (green) had recurrence, but the old patient (blue) did not. For the young patient no evidence of disease (NED) was diagnosed, the older patient shows a decrease of less than $<1/3$. Commonalities can be found in, e.g. chemotherapy (both Anthracycline).



(a) Tumor segmentation of young (green) and old (blue) patient projected onto a standardized breast.

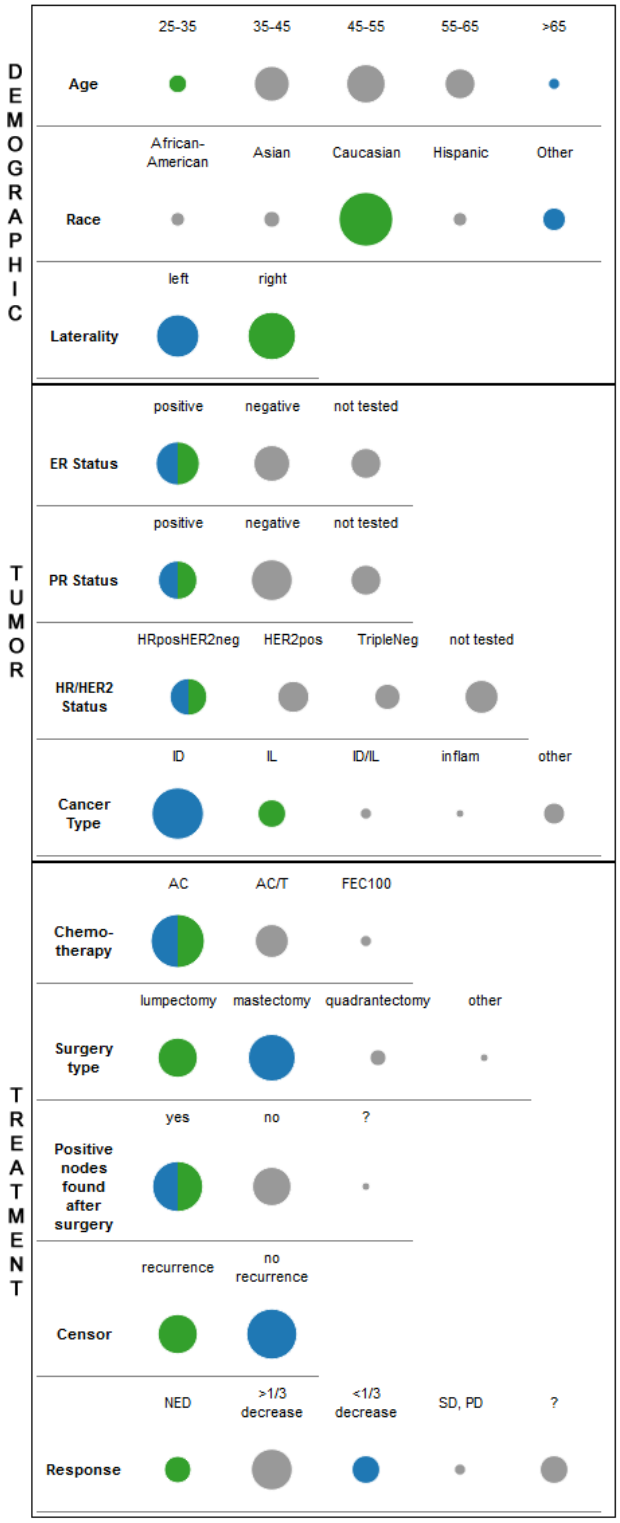


(b) Bar chart for comparing young (green) and old (blue) patient showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart for comparing a young patient (green), old patient (blue) and the average of both (orange).

Figure 6.22: Example charts for the comparison of a young and an old patient.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.23: Spotmatrix for comparing young (green) and old (blue) patient. When a value shows up in both patients, the circle in split.

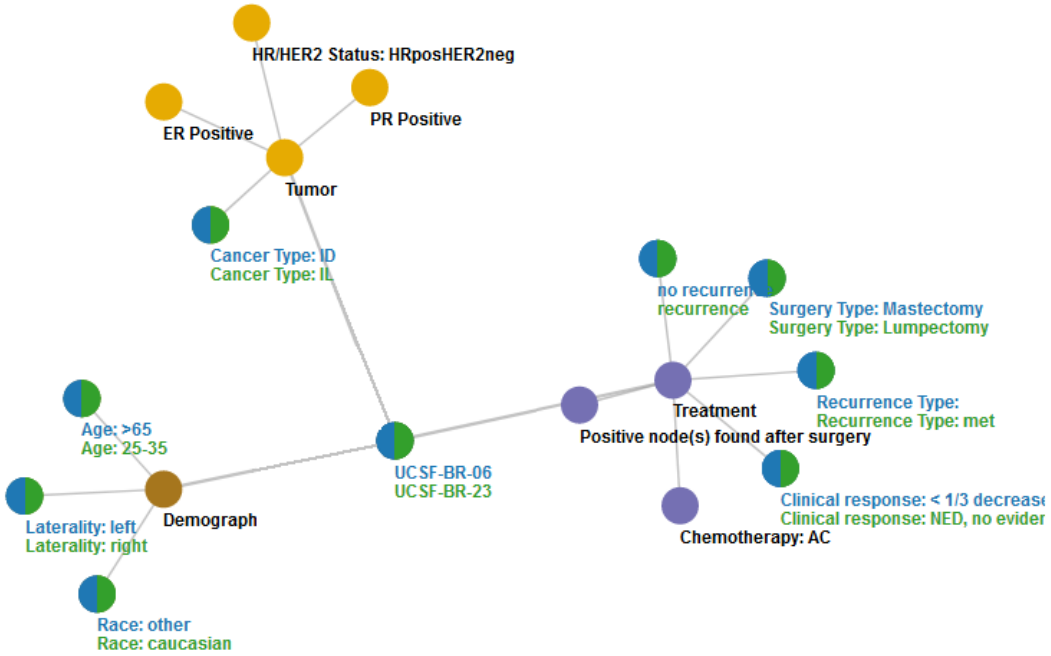


Figure 6.24: Force-directed graph for comparing young (green) and old (blue) patient. If a property has the same value in both patients the node is in standard color, otherwise the node is split.

6.2.2 Comparison of patient having received Taxane to patient only having received Anthracycline

To compare a patient, who only received Anthracycline as chemotherapy, to a patient, who additionally received Taxane after MRI_3, we selected two suitable patients in the selection wheel (see Figure 6.25). We chose two patients at about the same age (both in age division 45-55) and with a large tumor volume drop (more than 70%).

The overlapping probability map of the tumor segmentation can be seen in Figure 6.26a. It can be perceived that the tumor is situated in about the same area in the breast in both patients. In the *Bar chart* (Figure 6.26b), it can be seen that the patient only having received Anthracycline had a much bigger tumor volume at the beginning than the other patient (97.00 cc compared to 17.93 cc in MRI_1), but the volume decreased in both patients. Since MRI_4 was only scheduled for patients having received Taxane, data for this MRI scan is only available for one of the patients. When looking at the *Radar chart* (Figure 6.26c), it can be seen that the two patients are at about the same age. Concerning the progress of the tumor size, it can be recognized that although the tumor size decreased for both patients, the patient only having received Anthracycline still had

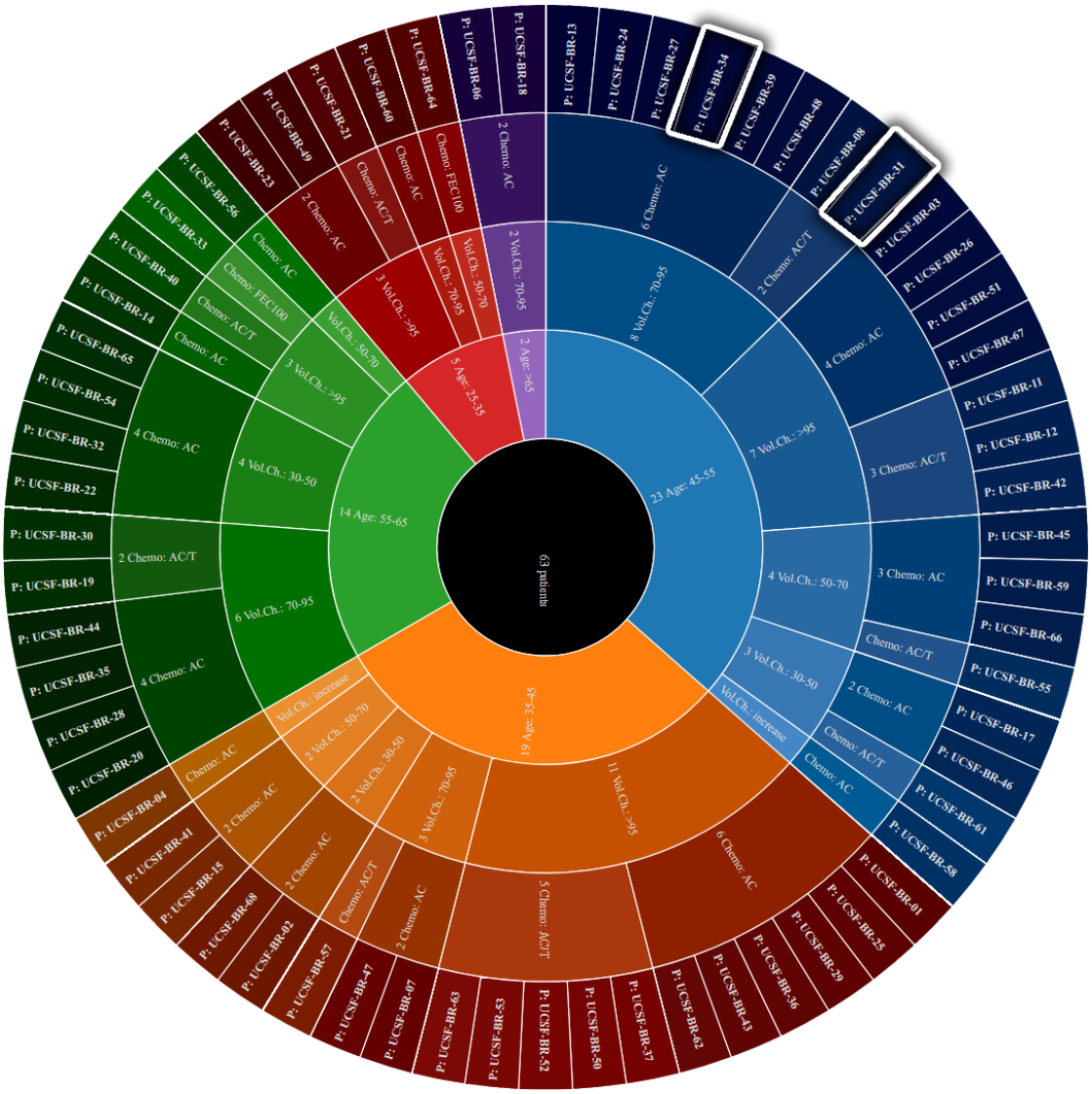
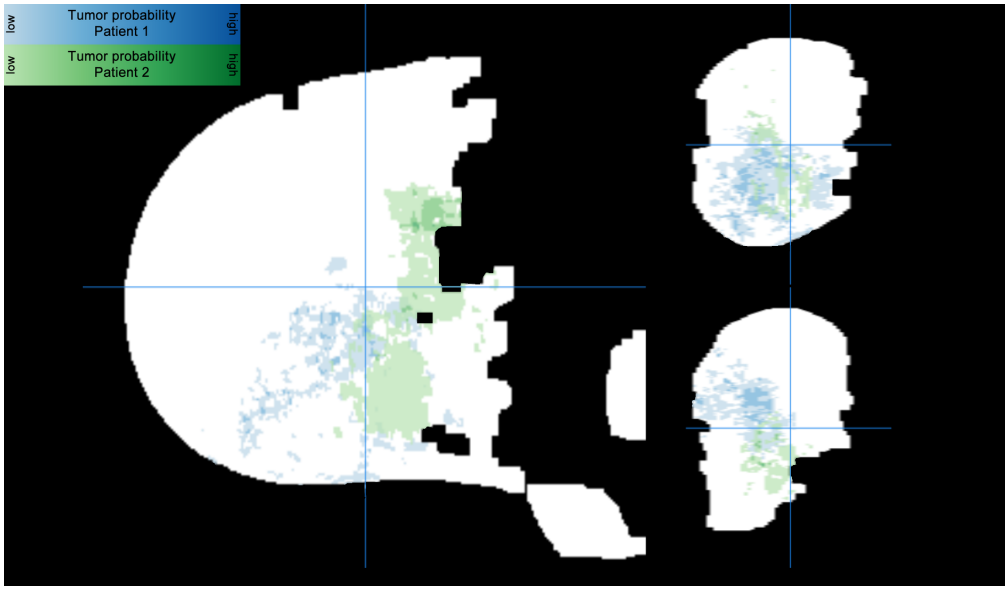
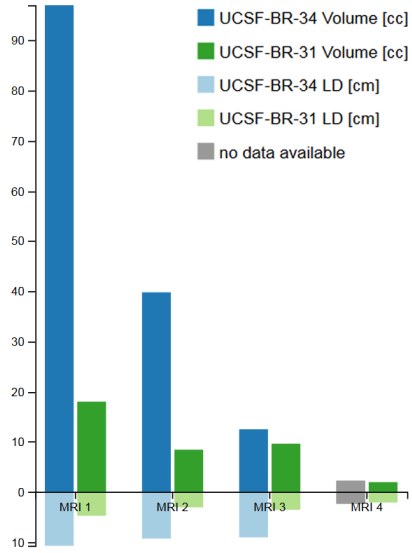


Figure 6.25: Selection wheel for pairwise inter-patient study, comparing a patient, who received only Anthracycline as chemotherapy and a patient, who additionally received Taxane. The selected patients are outlined in white.

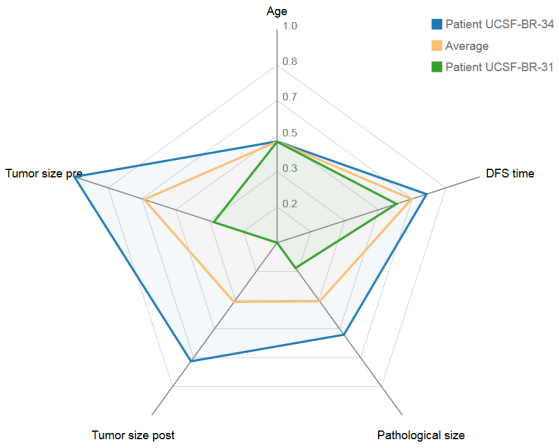
one of the largest tumor sizes after the chemotherapy in the cohort, in contrast to the patient who received Taxane, whose tumor size post-chemo was one of the smallest in the cohort. From the *Spotmatrix* (Figure 6.27) and the *Force-directed graph* (Figure 6.28) it can be figured out that these two patients had commonalities in their age, breast laterality, and cancer type. In both patients positive nodes were found after surgery and both had no recurrence of the tumor. Differences can be found in their race, the hormone receptor status of ER and PR, and the surgery type (the patient only having



(a) Tumor segmentation of a patient having only received Anthracycline as chemotherapy (blue) and a patient having additionally received Taxane (green), projected onto a standardized breast.

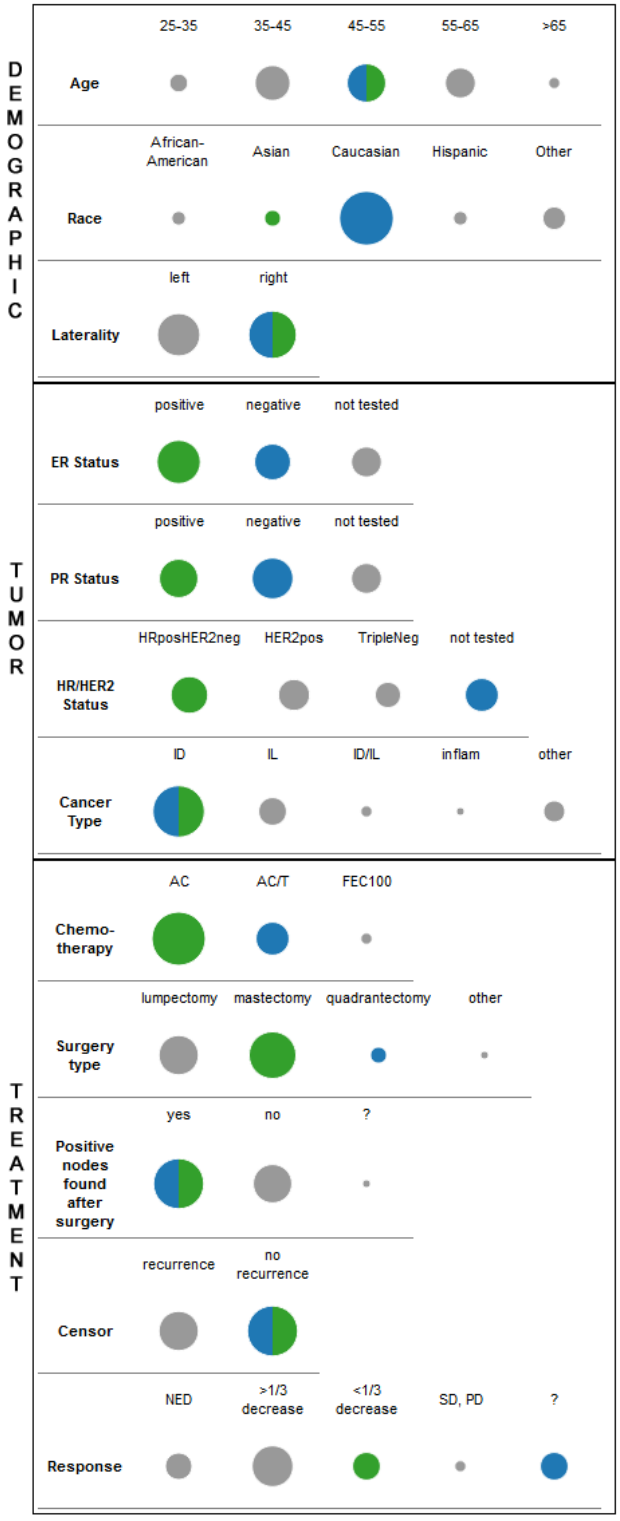


(b) Bar chart for comparing a patient only having received Anthracycline as chemotherapy (blue) and a patient additionally having received Taxane (green) showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart for comparing a patient only having received Anthracycline as chemotherapy (blue), a patient additionally having received Taxane (green), and the average of both (orange).

Figure 6.26: Example charts for the comparison of a patient only having received Anthracycline as chemotherapy and a patient additionally having received Taxane.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.27: Spotmatrix for comparing a patient only having received Anthracycline as chemotherapy (blue) and a patient additionally having received Taxane (green). If a value shows up in both patients, the circle is split.

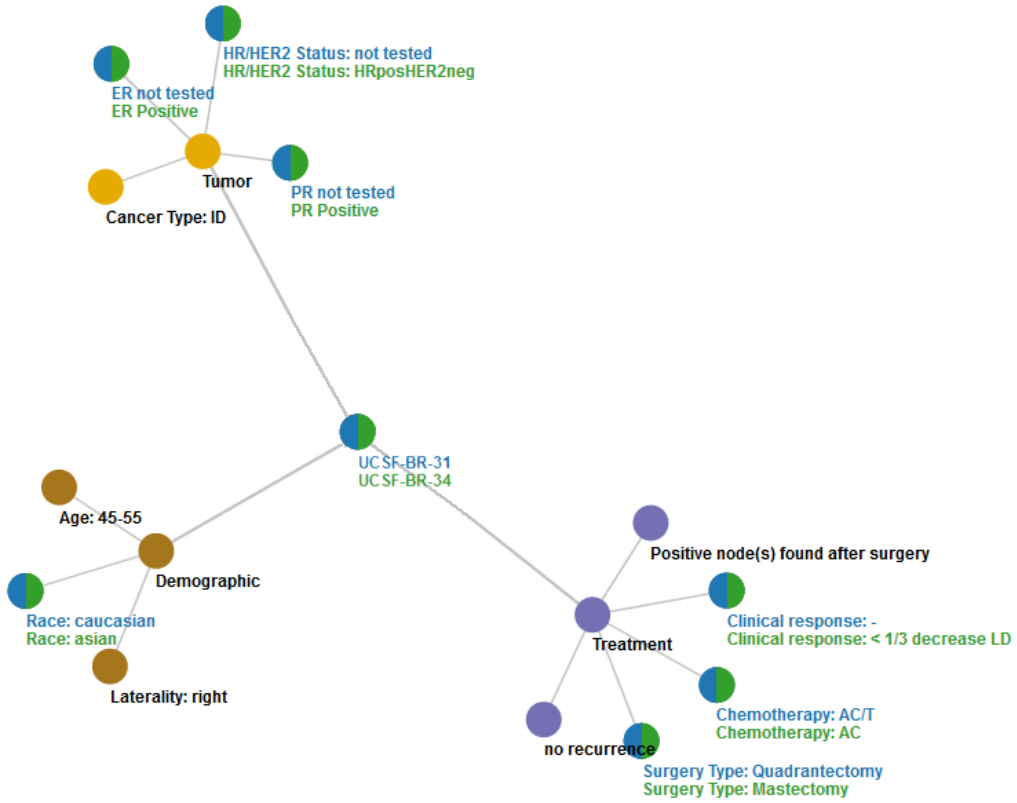


Figure 6.28: Force-directed graph for comparing a patient only having received Anthracycline as chemotherapy (blue) and a patient additionally having received Taxane (green). If a property has the same value in both patients, the node is in standard color, otherwise the node is split.

received Anthracycline underwent a mastectomy while the other patient only underwent a quadrantectomy). Clinical response is only known for the patient who did not receive Taxane as an additional chemotherapy treatment.

6.2.3 Comparison of patient with large tumor volume drop to patient with tumor volume increase

As final example of the pairwise inter-patient study framework to compare two patients, we chose a comparison of a patient with a large tumor volume decrease (>95%) and a patient with a tumor volume increase (see Figure 6.30). Again we selected two patients in the same age division (35-45). Both patients did not additionally receive Taxane as chemotherapy.

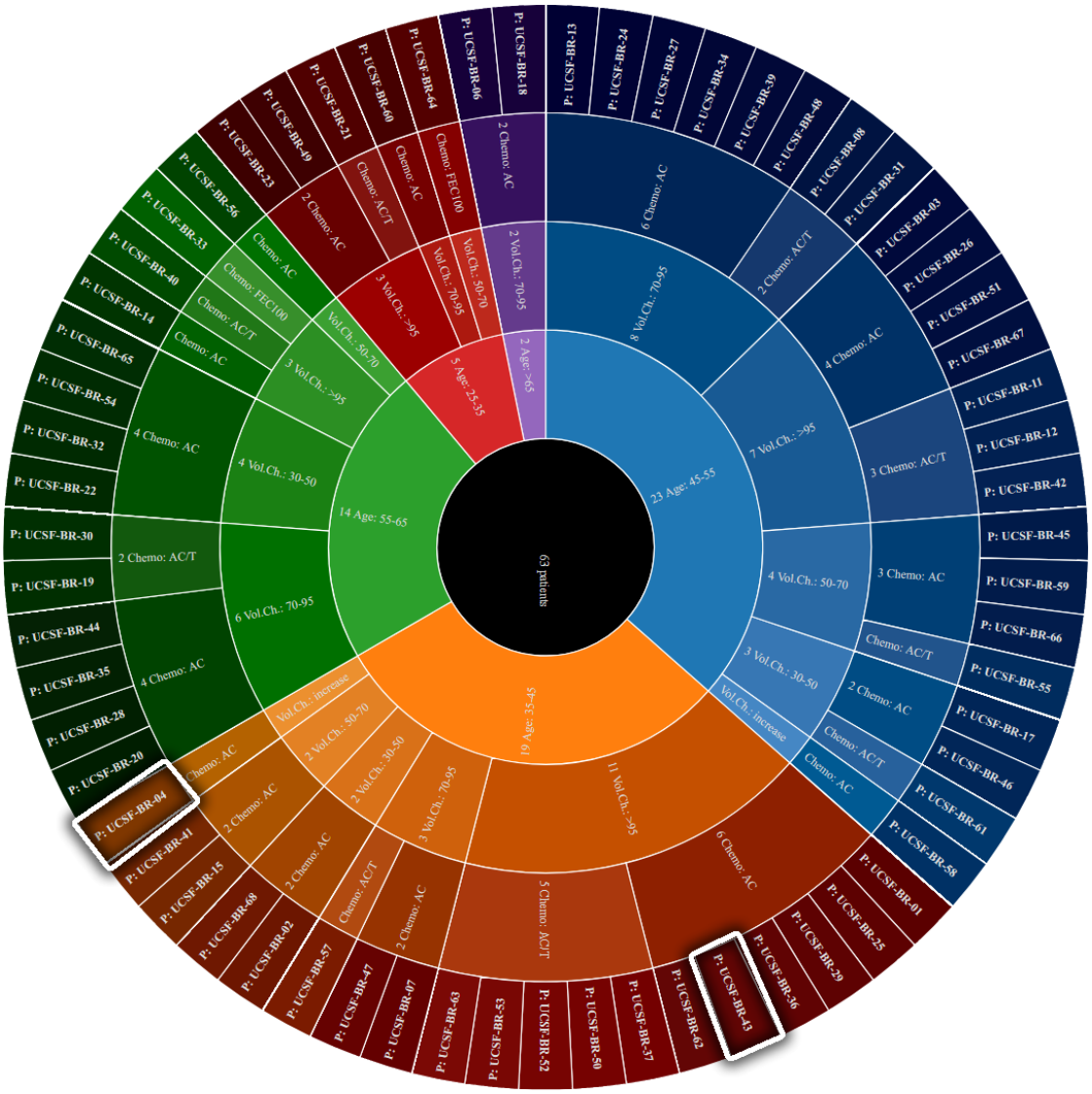
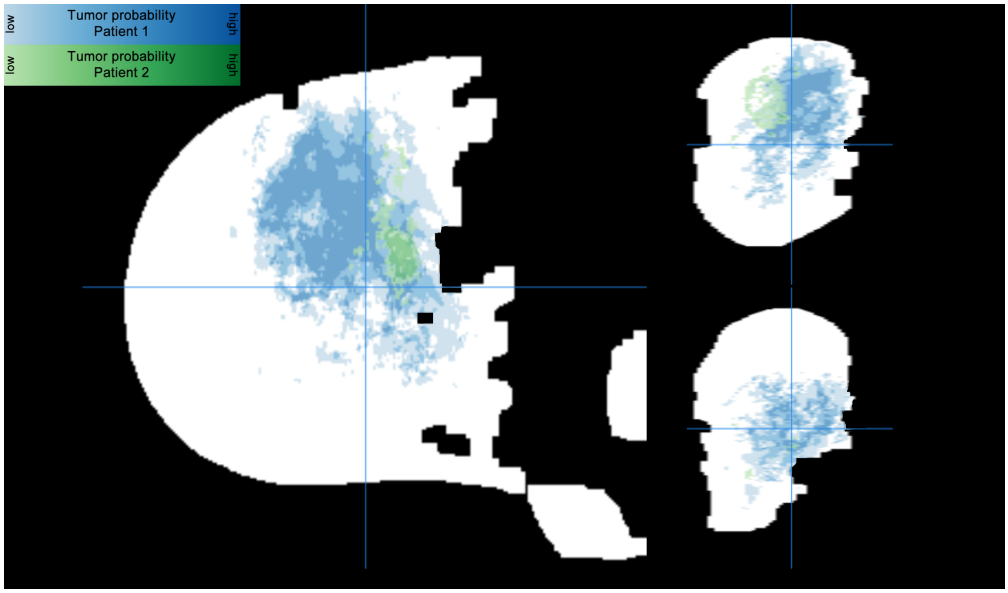
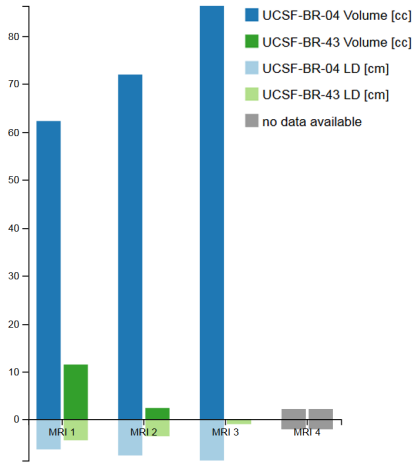


Figure 6.29: Selection wheel for pairwise inter-patient study comparing a patient with large tumor volume drop and a patient with tumor volume increase. The selected patients are outlined in white.

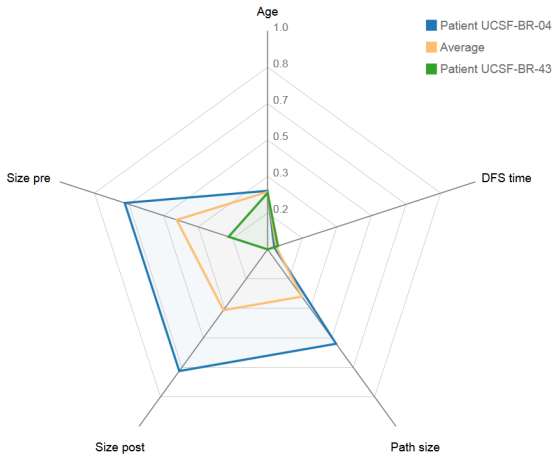
The probability map of the tumor of these two patients is shown in Figure 6.30a. Regarding the *Bar chart* (Figure 6.30b) the tumor volume and LD increase of the first patient (shown in blue colors) can easily be spotted, in contrast to the tumor volume and LD decrease of the other patient (shown in green colors). Since both patients did not receive Taxane, no data is available for MRI_4. In the *Radar chart* (Figure 6.30c) the difference in the progress of the tumor size can also be seen when comparing *Tumor size pre* to *Tumor size post* of both patients. The disease-free survival (DFS) time is very short for



(a) Tumor segmentation of a patient with large tumor volume drop and a patient with tumor volume increase, projected onto a standardized breast.

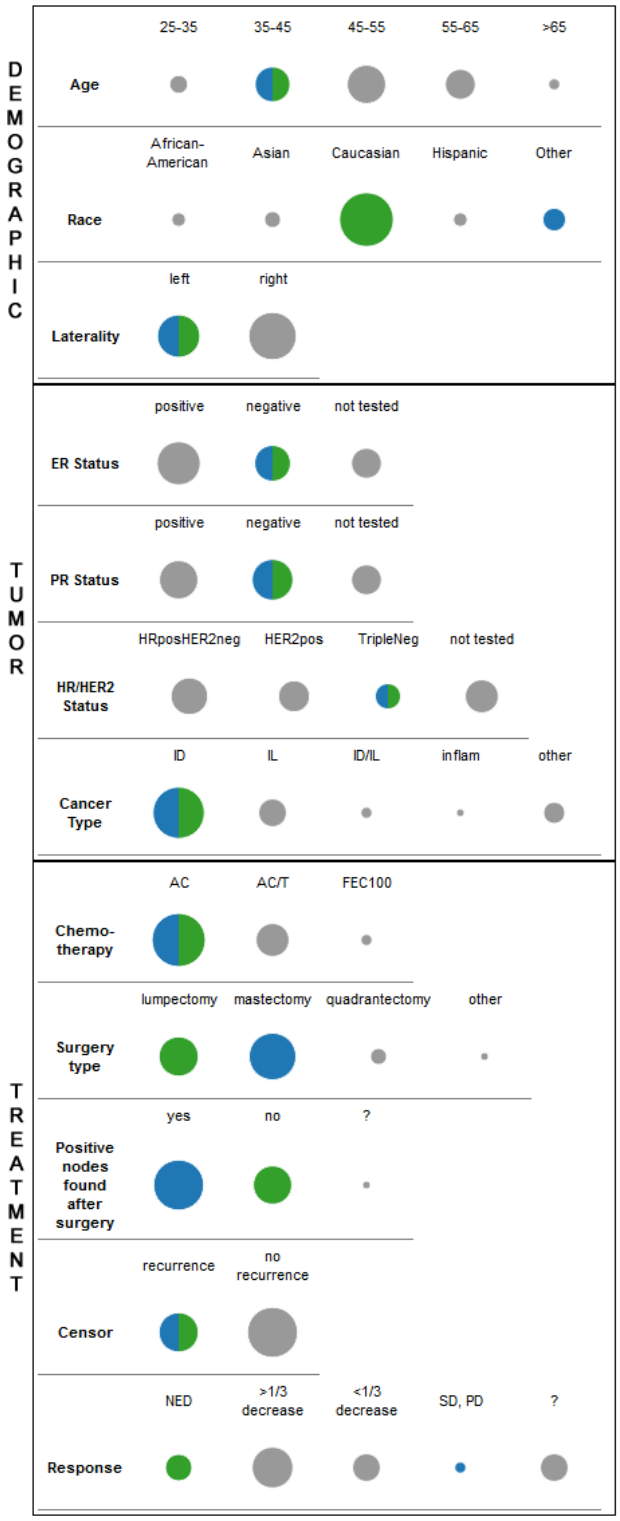


(b) Bar chart for comparing a patient with large tumor volume drop (green) and a patient with tumor volume increase (blue) showing the volume and longest diameter (LD) for each MRI.



(c) Radar chart for comparing a patient with large tumor volume drop (green), a patient with tumor volume increase (blue), and the average of both (orange).

Figure 6.30: Example charts for the comparison of a patient with large tumor volume drop and a patient with tumor volume increase.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.31: Spotmatrix for comparing a patient with large tumor volume drop (green) and a patient with tumor volume increase (blue). If a value shows up in both patients, the circle in split.

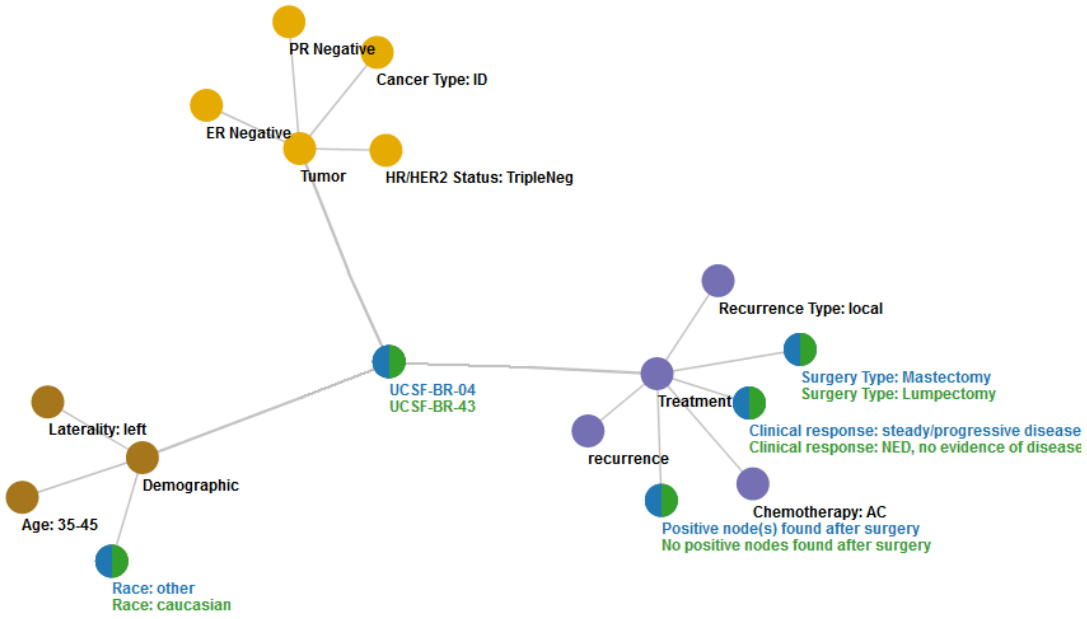


Figure 6.32: Force-directed graph for comparing a patient with large tumor volume drop (green) and a patient with tumor volume increase (blue). If a property has the same value in both patients, the node is in standard color, otherwise the node is split.

both patients. A possible explanation for this can be found when additionally looking at the *Spotmatrix* (Figure 6.31) and the *Force-directed graph* (Figure 6.32): The patient with tumor volume and LD increase had a steady or progressive disease as response and positive nodes were found after surgery, which most probably indicates that the patient died not long after surgery. The other patient shows no evidence of disease (NED) as response and no positive nodes were found after surgery. Nevertheless tumor recurrence seem to be recognized after only a short DFS time.

6.3 Groupwise inter-patient study

Finally we want to show an example for comparing two different groups of patients by the use of the framework we introduced in Section 4.6. To do the selection of the groups, we first had to create the selection wheel by choosing *Groupwise inter-patient* and arranging the desired criteria (as shown in Figure 6.33). In the generated selection wheel (see Figure 6.34) no patients are shown on the edge of the circle, the outermost sections represent the chosen divisions. By successively clicking on the two desired groups, the various visualizations are generated and displayed.

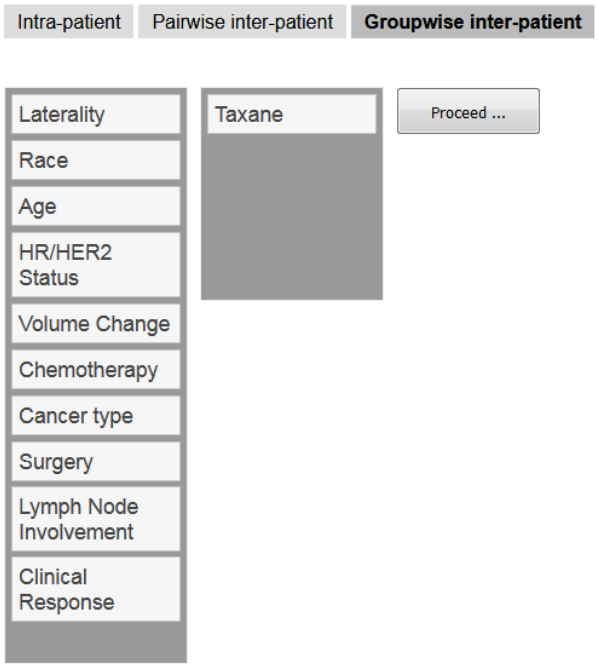


Figure 6.33: Selection of criteria to create the selection wheel for the comparison of two groups. *Taxane* is chosen as criterion for dividing the cohort.

6.3.1 Comparison of patient group having received Taxane to patient group not having received Taxane

In this section we want to compare all patients that additionally received Taxane as chemotherapy after completing their Anthracycline therapy, to all other patients having their chemotherapy treatment stopped after MRI_3. Therefore we selected *Taxane* as only selection criterion (see Figure 6.33) to generate the selection wheel, which is shown in Figure 6.34. After having successively clicked both groups the visualizations shown in Figure 6.35 (tumor probability maps, Boxplots and Radar chart), Figure 6.36 (Spotmatrix) and Figure 6.37 (Force-directed graph) are displayed.

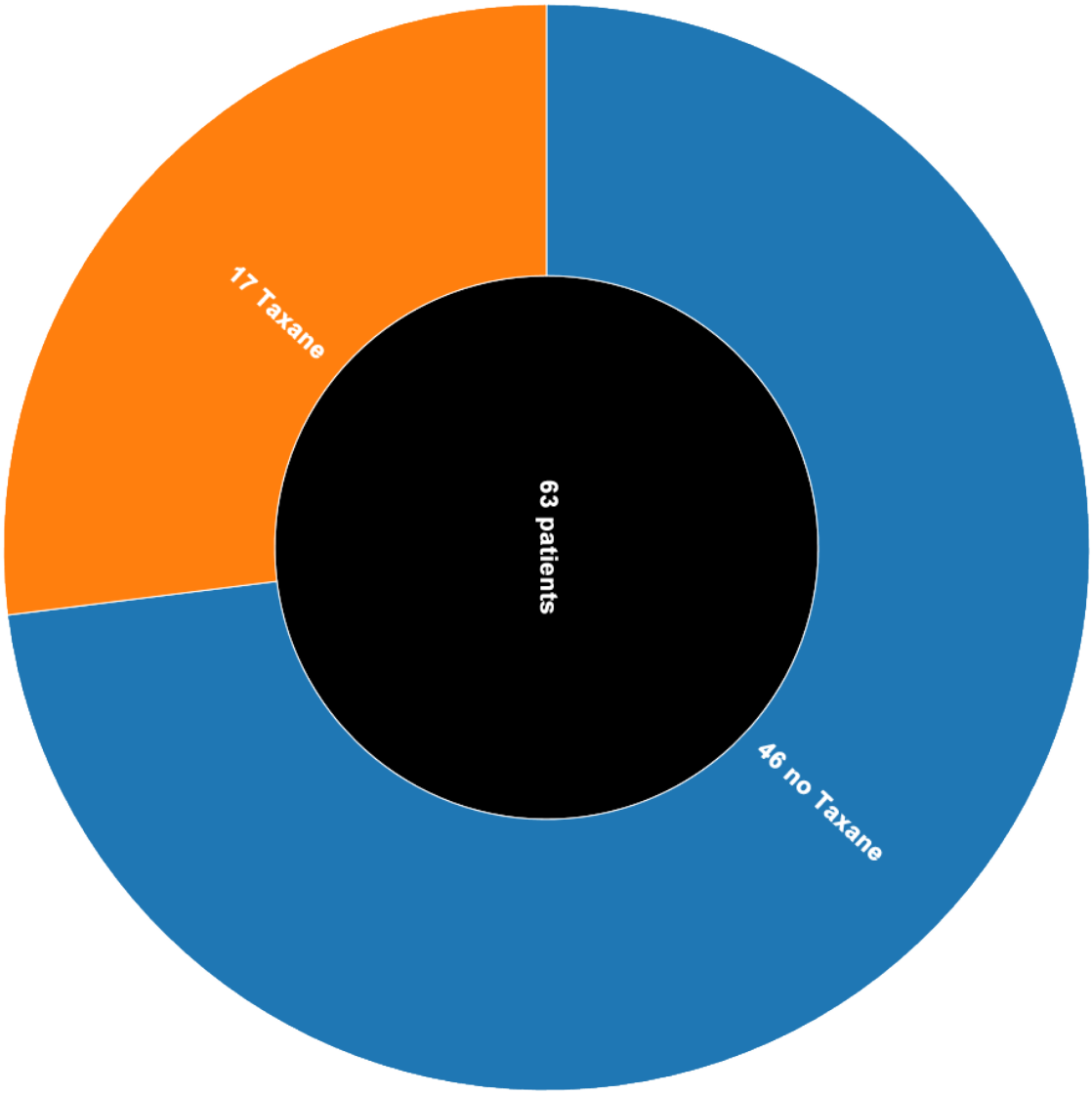
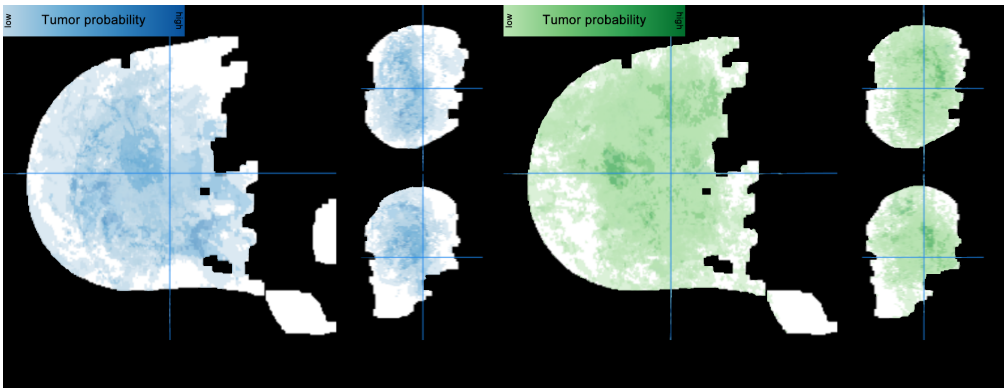
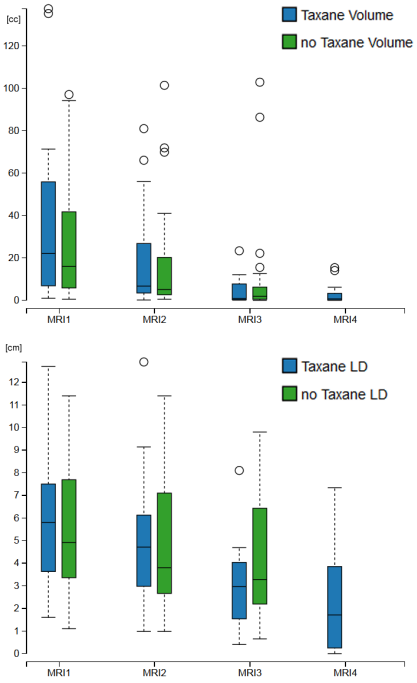


Figure 6.34: Selection wheel for a groupwise inter-patient study comparing all patients, who received Taxane as additional chemotherapy, and all patients, who did not received Taxane.

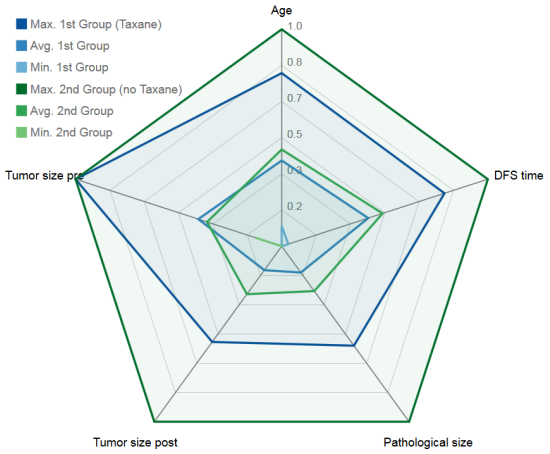
When looking at the comparison of the tumor visualization maps (Figure 6.35a), it can be recognized that the distribution of the tumor probability in the Taxane-group (left) is more concentrated than in the Anthracycline-group (right). The tumor probability map on the right side is more diffuse, which could be because of the size of the group - the Antracycline-group is larger resulting in higher and more complex variability. In the *Boxplots* (Figure 6.35b), it can easily be seen that in both groups the average of the tumor volume and LD decreased. It seems that in the first group (the group having received



(a) Comparative tumor segmentation of patients having received Taxane as additional chemotherapy after MRI_3 (left) to patients not having received Taxane (right), each projected onto a standardized breast.

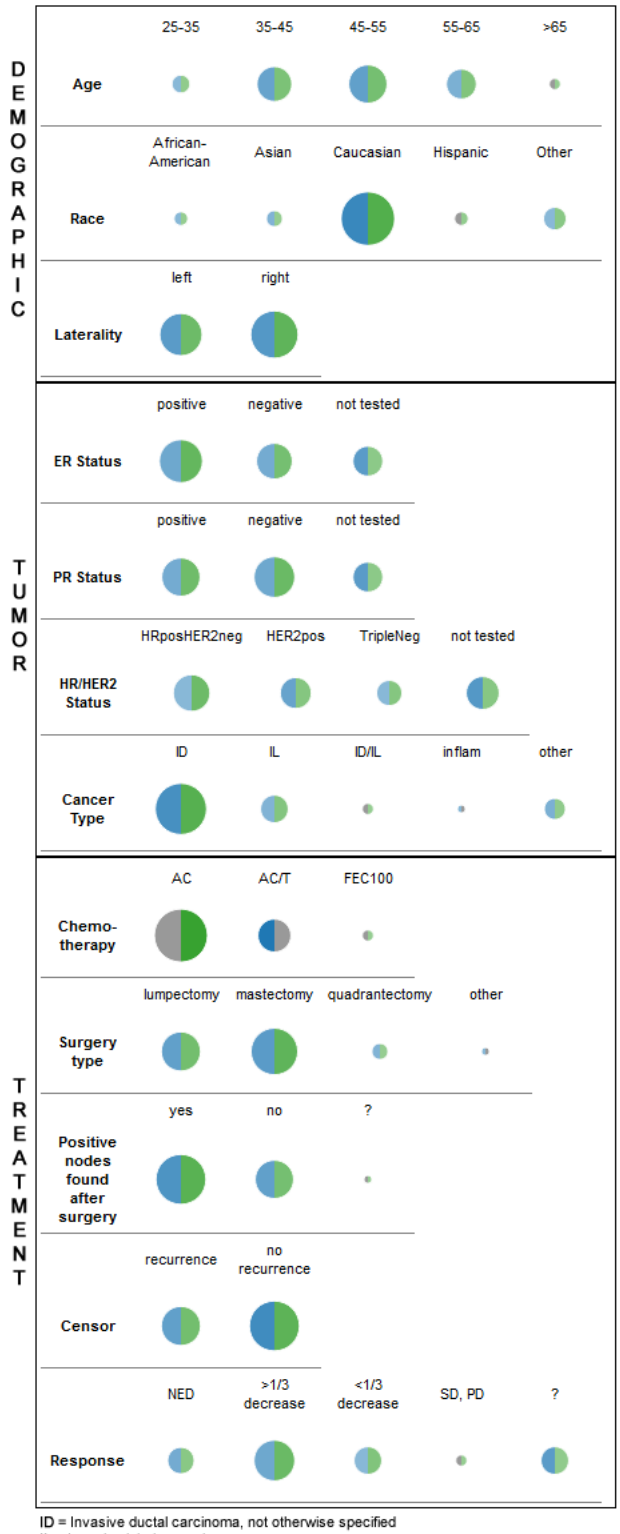


(b) Boxplots for comparing tumor volume and longest diameter (LD) distribution of patients having received Taxane as additional chemotherapy after MRI_3 (blue) to patients not having received Taxane (green).



(c) Radar chart for comparing maximum (dark), average (medium), and minimum (light) of multivariate data of patients having received Taxane as additional chemotherapy after MRI_3 (blue) to patients not having received Taxane (green).

Figure 6.35: Example charts for the comparison of patients having received Taxane as additional chemotherapy after MRI_3 to patients not having received Taxane.



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.36: Spotmatrix for comparing patients having received Taxane as additional chemotherapy after MRI_3 (blue) to patients not having received Taxane (green). The luminance of a circle shows the quantity of the particular property within the patient group (the more luminous, the more patients are affected).

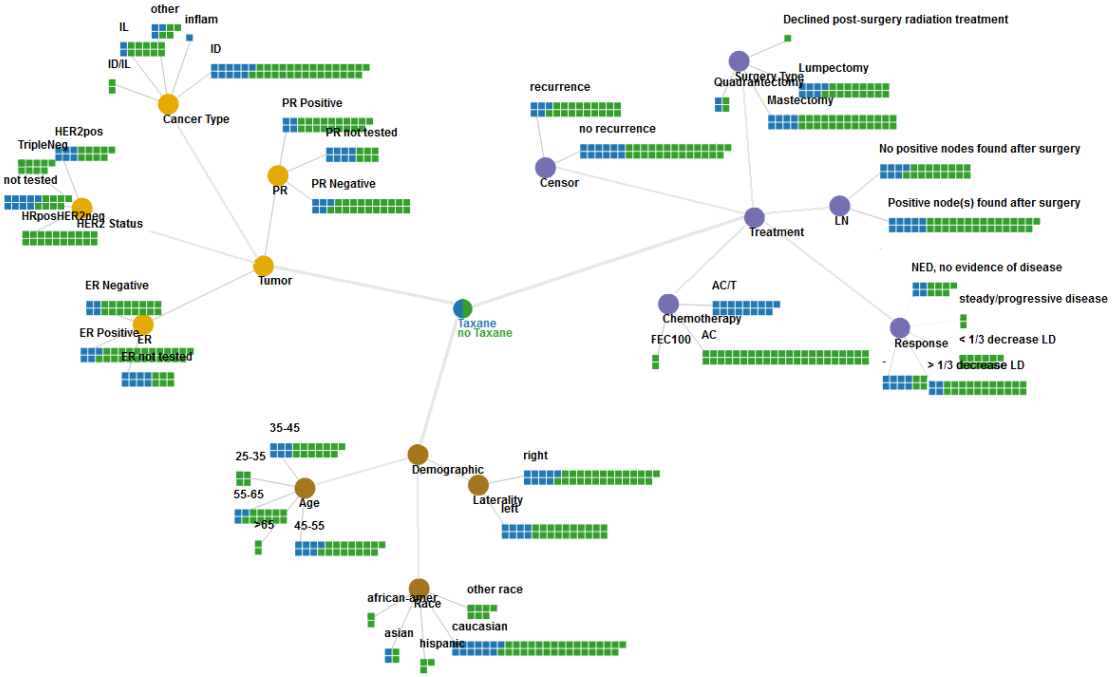


Figure 6.37: Force-directed graph for comparing patients having received Taxane as additional chemotherapy after MRI_3 (blue) to patients not having received Taxane (green). Each property is shown as a *Waffle chart*, showing the number of patients affected by the property. Properties that do not show up in both groups are not displayed.

Taxane) the tumor volume and LD, measured after MRI_4, decreased more than in the other group. However, when applying an *unpaired t test* this difference is considered to be *not statistically relevant* (p-value 0.6760). Since only the patients undergoing Taxane chemotherapy were scheduled to have a fourth MRI scan, no data is available for MRI_4 in the second group. Regarding the first group, we can see that treating the patients with Taxane did decrease the tumor volume and the LD. When looking at the *Radar chart* (Figure 6.35c), it can be recognized that there are large differences in every multivariate value in both groups, which can be seen in the large difference between maximum (dark) and minimum (light) value. When comparing the two groups, the best would be to compare only the average (medium) values. We can see that the drop of tumor size in the first group (Taxane, blue) is bigger than in the second group (no Taxane, green). In all other multivariate values no significant differences can be seen. Also when looking at the *Spotmatrix* (Figure 6.36), no big difference in the non-imaging demographic data can be spotted. Every property is affected by approximately the same proportion of patients within the group (each circle shows about the same luminance), except for the chemotherapy, which is because of the initial selection of the groups. The *Force-directed graph* (Figure 6.37), which was extended by a *Waffle chart*, shows in detail how many patients are affected by the various properties. For example, we can see that most of

the patients underwent a *mastectomy* as surgery, the second largest group underwent a *lumpectomy*, but the distribution of the two groups is about the same.

6.3.2 Comparison of patient group with clinical response of *no evidence of disease* to patient group with clinical response of *steady or progressive disease*

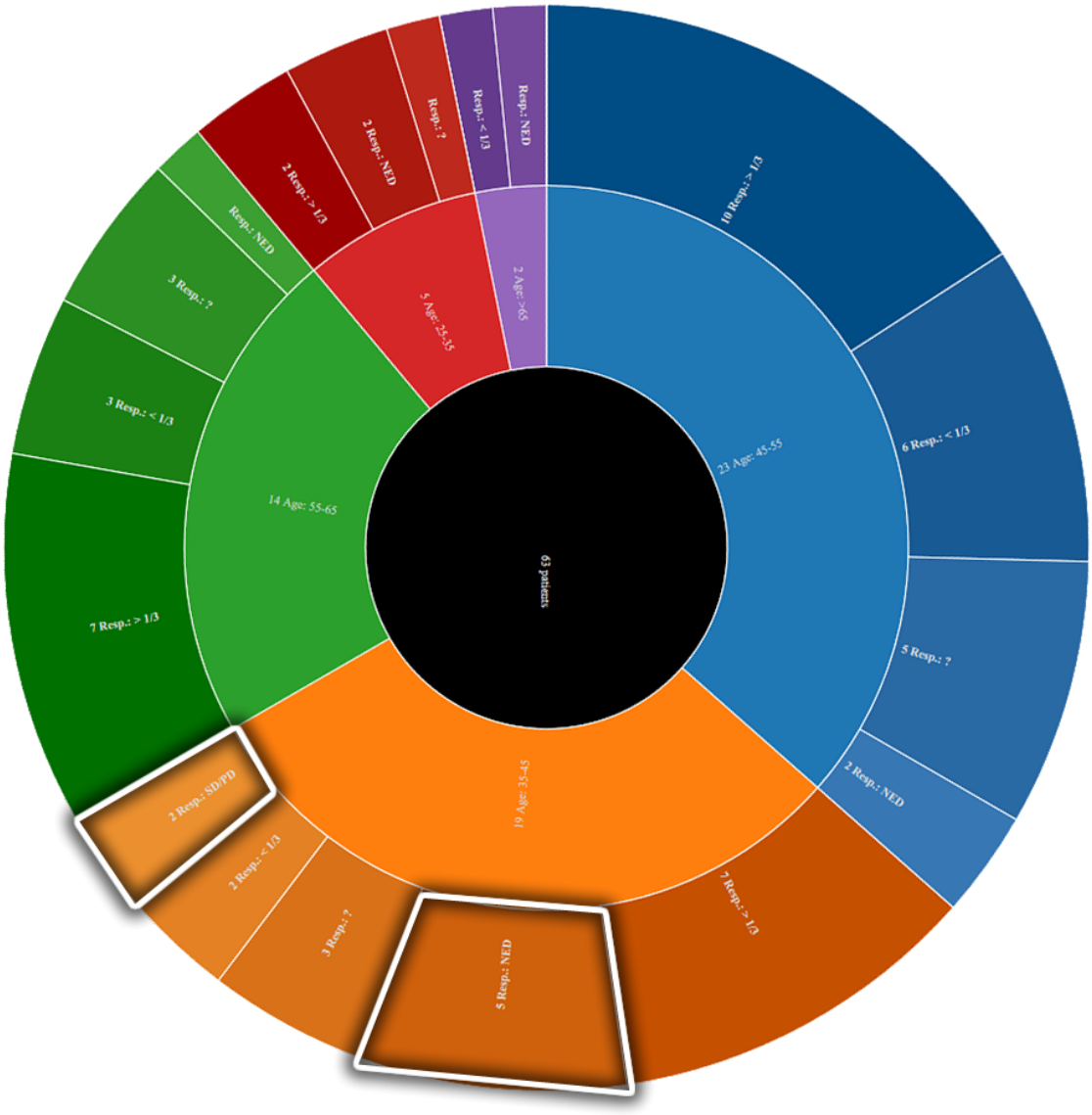
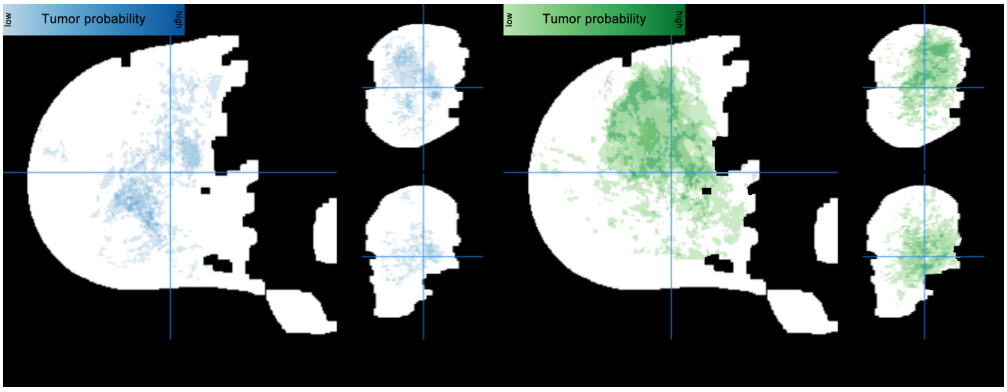
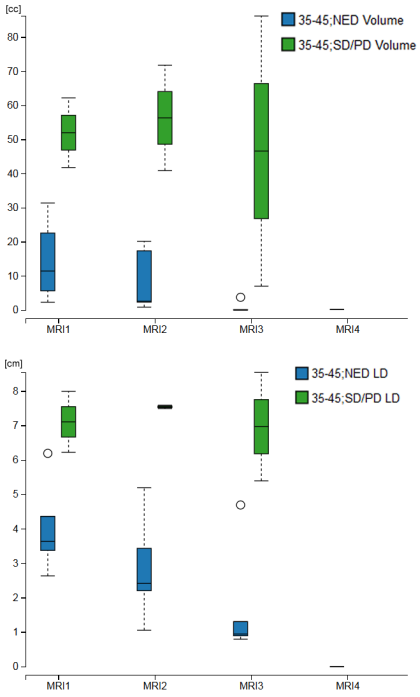


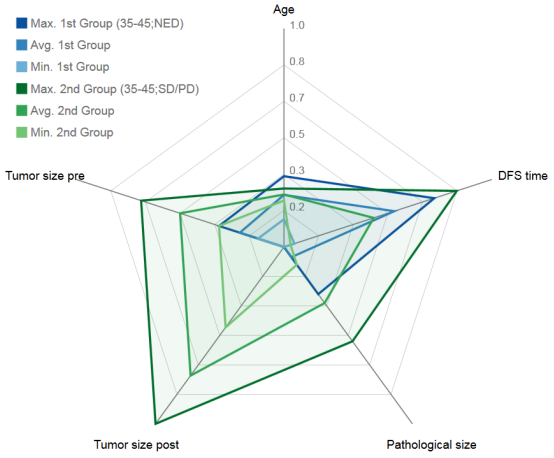
Figure 6.38: Selection wheel for groupwise inter-patient study comparing patients having a clinical response of *no evidence of disease* (NED) to patients having a clinical response of *steady or progressive disease* (SD/PD). Both groups are in the same age division, i.e. 35-45.



(a) Comparative tumor segmentation of patients having a clinical response of *no evidence of disease* (left) to patients having a clinical response of *steady or progressive disease* (right), each projected onto a standardized breast.

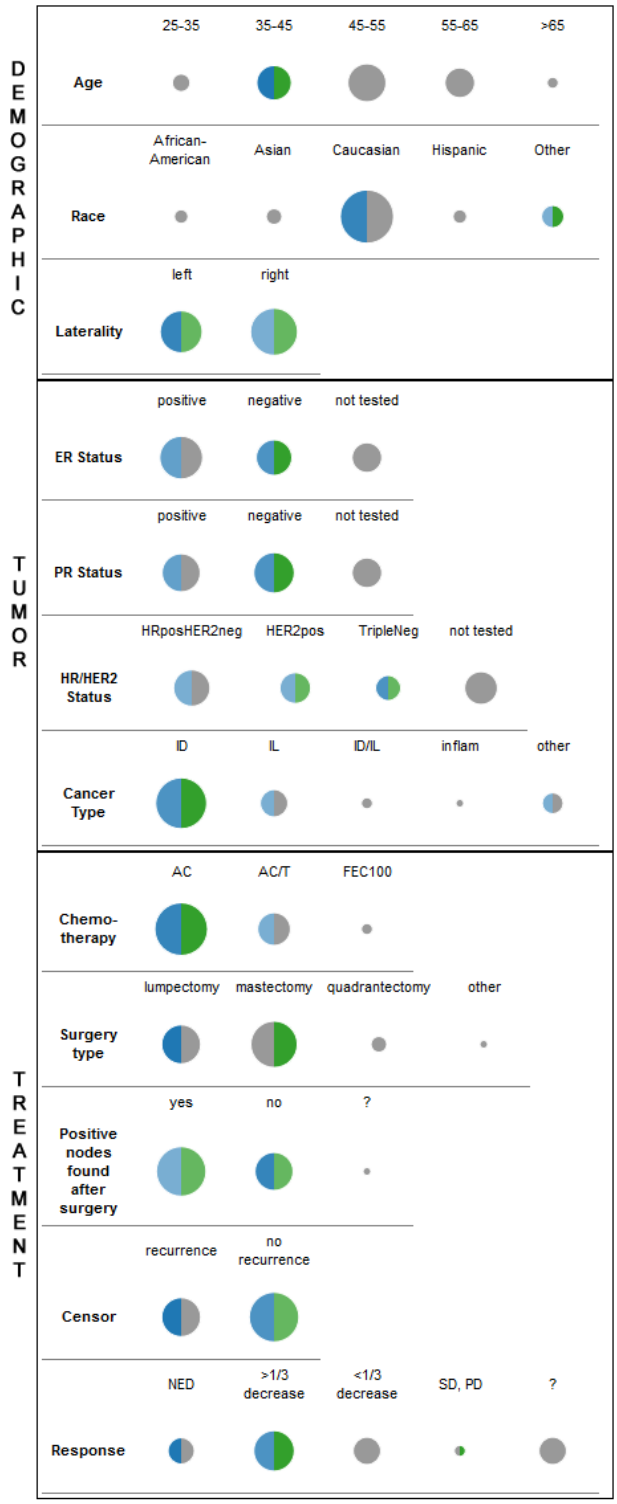


(b) Boxplots for comparing tumor volume and longest diameter (LD) distribution of patients having a clinical response of *no evidence of disease* (blue) to patients having a clinical response of *steady or progressive disease* (green).



(c) Radar chart for comparing maximum (dark), average (medium), and minimum (light) of multivariate data of patients having a clinical response of *no evidence of disease* (blue) to patients having a clinical response of *steady or progressive disease* (green)

Figure 6.39: Example charts for the comparison of patients having a clinical response of *no evidence of disease* (NED) to patients having a clinical response of *steady or progressive disease* (SD/PD)



ID = Invasive ductal carcinoma, not otherwise specified
IL = Invasive lobular carcinoma
AC = Anthracycline
AC/T = Anthracycline/Taxane
FEC100 = Fluorouracil, Epirubicin, Cyclophosphamide
NED = No Evidence of Disease
SD, PD = Steady Disease, Progressive Disease

Figure 6.40: Spotmatrix for comparing patients having a clinical response of *no evidence of disease* (blue) to patients having a clinical response of *steady or progressive disease* (green). The luminance of a circle shows the quantity of the particular property within the patient group (the more luminous, the more patients are affected).

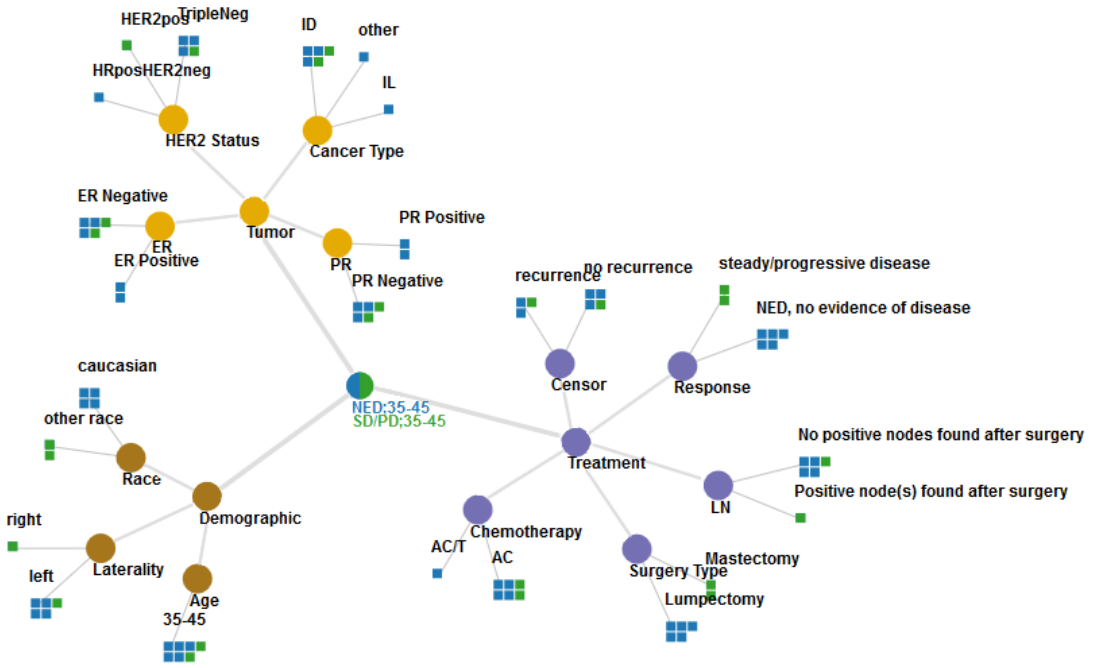


Figure 6.41: Force-directed graph for comparing patients having a clinical response of *no evidence of disease* (blue) to patients having a clinical response of *steady or progressive disease* (green). Each property is shown as a *Waffle chart*, showing the number of patients affected by the property. Properties that do not show up in both groups are not displayed.

Finally we want to show an example of comparing a group of patients having a clinical response of *no evidence of disease* (NED) to a group of patients having a clinical response of *steady or progressive disease* (SD/PD). We also chose *age 25-35* as preselection to compare groups of patients of about the same age. The corresponding selection wheel is shown in Figure 6.38.

Regarding the tumor probability maps projected onto a standardized breast (Figure 6.39a), it can be seen that the patients having a clinical response of *steady or progressive disease* (right) have a more expanded probability map and therefore a larger volume where tumor appearance is probable. In the *Boxplots* (Figure 6.39b) it can be recognized that the average tumor volume and LD is much higher in the SD/PD-group than in the NED-group. Also the progress of the tumor volume and LD seems to be a lot better in the NED-group. In contrast the average tumor volume of the SD/PD-group did not decrease very much, whereby the average tumor volume of MRI_2 even showed an increase. Comparing these two groups, a t-test shows that the difference is *statistically extremely significant* (p-value 0.0009). This difference can also be seen in the *Radar chart* (Figure 6.39c). Again we only want to analyze the average of both groups (medium-bright colored) and we can see that the only commonalities can be found in the age (which was preselected) and the disease-free survival (DFS) time. Regarding the *Spotmatrix* (Figure 6.40), it can be seen

that the patients in the NED-group (blue) are mostly Caucasian, while all the patients in the SD/PD-group are *other race*. This could be understood that (considering this small sample of patients) the applied treatments work better in patients of Caucasian race than in other races. When exploring the *Force-directed graph* (Figure 6.41), we see that this comparison affects only a small sample of patients. Some of the properties are not displayed, because no patient in any of the two groups is affected by these properties. This can also be seen in the *Spotmatrix* (see Figure 6.40), where every circle concerning one of these properties is not split and colored in gray.

6.4 Critical reflection

The framework was designed and implemented based on a cohort of 64 breast cancer patients. During the preprocessing procedure the data of one patient turned out to be unusable for registration, most likely caused by deficient MRI data, therefore we excluded this patient and continued working with the other 63 patients. Because of this limitation, some of the visualizations will only work fine in other cohorts of approximately the same size. Also the amount of ascertained non-imaging demographic data is a limitation for some of the charts, which will be discussed in detail in Chapter 7.

Since the aim of this work was to make comparative visualizations of the imaging and non-imaging data possible, we did not focus on the imaging registration process, therefore this step could be improved in later works. The calculation of tumor probability maps was limited to four MRI scans and would not be suitable for follow-up studies with more scheduled scans. An analytical approach for the calculation of the probabilities would be needed in this case, which was considered out of scope for our work.

Nevertheless, clinical researchers working on the selection of chemotherapy treatment are able to explore and analyze the progress of the treatment in different patients and can compare the results to other patients or groups of patients within the cohort. Several visualizations show the multitude of the imaging and non-imaging data and make them comparable, using the concept of multiple linked views [Rob05].

CHAPTER 7

Conclusion and future work

7.1 Summary

This thesis introduced a web-based framework to help clinical researchers working on the selection of chemotherapy treatment in breast cancer to explore and analyze the multitude of both imaging and non-imaging data. Therefore three main tasks for the exploration and analysis of the available data were designed and implemented. A functionality for single patient follow-up studies (intra-patient study), a functionality to compare two different patients (pairwise inter-patient study), and a functionality to compare groups of patients (groupwise inter-patient study) have been realized to answer the research questions: "How can differences between patients that were treated in differing ways be explored?" and "How can conclusions be drawn while comparing imaging and non-imaging data of patients or groups?". The intra-patient study allows a researcher to explore and analyze follow-up data of one patient, to spot differences in tumor data at diverse time points. The pairwise inter-patient study assists a researcher in comparing data of two patients to understand the efficiency of the treatment. Finally the groupwise inter-patient study helps a researcher to understand how different groups of patients respond to selected treatments. The cohort in the provided data consists of two major groups, one having their treatment stopped after the Anthracycline therapy, the other one having received Taxane as additional treatment after MRI_3. Therefore the design and specific visualizations were based on the assumption that a user wants to compare either two patients - each out of one of these groups - or two groups.

To achieve these goals the available data had to undergo some preprocessing steps. For imaging data registration, segmentation, and calculation of tumor probability maps had to be accomplished. Also, the non-imaging data had to be structured and prepared for easier querying by clinical researchers. Various interactive visualizations have been realized to let a clinical researcher visually recognize characteristics in patients, and

differences and similarities between two patients or two groups of patients using the concept of multiple linked views.

Finally, representative results that have been achieved by the use of the developed framework have been presented, showing that the mentioned research questions can be answered and differences as well as commonalities can easily be spotted in the various visualizations. The results show that the user can browse through MRI images of different patients and compare them to each other in a facilitated way, which was not possible with previous means of exploration.

To give our reader the possibility to test this framework, we made a web page available, that provides all the functionalities that are described in this thesis. The web page can be found at the following link:

`http://nkarall.f-s.at/masterthesis/`

7.2 Limitations

As already stated in Section 6.4, this framework was designed and implemented based on a cohort of 64 breast cancer patients. This would be a limitation when assigning the framework to another cohort. Some of the visualizations will scale in other cohorts of approximately the same size. The overlapping presentation of the tumor probability maps in the groupwise inter-patient study would not be suitable for large groups and a more analytical approach would be required [PW13]. If more ascertained non-imaging demographic data would be available in other cohorts, the Force-directed graph would increase proportionally to the number of properties and the Spotmatrix would also get bigger, which could overwhelm a user with a high number of different properties.

The conceptual approach of this thesis was based on the cohort data, which consists of two major groups, one having treatment stopped after the Anthracycline therapy, the other one having received Taxane as additional treatment. Therefore, comparing visualizations is limited to only two patients or two groups. Future works can carry out an approach of comparing more than two patients or groups within one visualization, or even without predetermined limitations, i.e. a comparison of all patients with a certain criterion.

7.3 Directions for future work

Future work based on this thesis could establish various concepts to improve the exploration and analysis of cohort data. One improvement could be the aforementioned extension to make comparisons of more than two patients or groups possible. Therefore, methods for calculating suitable probability maps have to be designed. Also the visualization charts for the non-imaging demographic data have to be adapted to make visualizations for data of more than two patients or groups possible.

For bigger cohorts with a lot of patients some of the presented visualizations might not be suitable. The selection process, which patients or groups to explore or to compare (selection wheel, example see Figure 4.4) should be adapted or exchanged in large groups, as well as the Force-directed graph. Methods to improve the scalability of these visualizations should be researched or designed and implemented. In the groupwise inter-patient study, the Force-directed graph could show details for each node only on demand, i.e. when clicking on it. For larger groups the Waffle chart in this graph could be scaled by a different binning. E.g. one square would show a group of five patients instead of showing only one patient.

Concerning the registration process, an improvement of this work would be to include the imaging preprocessing procedures into the web-based framework and therefore making it possible to dynamically register MRI scans during the use of the system. Also the calculation of the tumor probability maps could be integrated into the framework to make a more flexible generation process of such probability maps possible.

It could also be examined if an aggregation of some or even all presented visualizations would be reasonable and possible. One approach would be to create an interactive overview visualization. The user could choose which visualizations should be displayed or could dynamically enlarge or minimize some visualizations. When displaying a multitude of different imaging and non-imaging data, interactivity should be an important issue to address a lot of interests of different clinical researchers. Also it should be ensured that the visualization data is available in real-time without waiting for calculations to be done.

Another approach for future work could be an examination of including *Machine learning methods* [CHPS05]. Suitable registration procedures, calculations of tumor probability maps, or scalings of visualizations depending on the amount and kind of available imaging and non-imaging data, could automatically be chosen by the framework, if machine learning is included. Therefore, appropriate machine learning methods would have to be researched and implemented, or possibly designed or adapted.

This thesis presents a regressive method of exploring and analyzing data of a cohort. When using different or additional data, this approach could be extended to either be predictive or regressive. Therefore, suitable and reliable prognosis methods would have to be researched and applied, so a researcher could estimate the progress of various possible treatments and their expected effect on the tumor. New patients could be added and the outcome or a risk assessment of a treatment could be predicted by finding similar patients.

To identify possible strong or weak points, an evaluation of our provided framework would be helpful. According to Heidi Lam et al. [LBI+11] the evaluation scenarios *Evaluating Visual Data Analysis and Reasoning (VDAR)* and *Evaluating User Experience (UE)* would fit for this thesis. VDAR facilitates finding out how a tool supports analysis and drawing conclusions of given data and UE supports getting subjective feedback about the framework.

The proposed approach is a first step towards an environment to explore and analyze the progress of cancer and the treatment of patients suffering from cancer, but not limited to breast cancer. It can allow clinical researchers to understand the effect of different therapies on patients with different demographic properties and therefore is a small part of the big area of curing cancer.

List of Figures

List of Tables

Bibliography

[AG] MeVis Medical Solutions AG. MeVisLab. URL: https://www.mevislab.de/. last accessed on February 13, 2018.

[AKM+04] Heiko Alfke, Sven Kohle, Elisabeth Maurer, Ilhan Celik, Richard Rascher-Friesenhausen, Sarah Behrens, Johannes T. Heverhagen, Heinz-Otto Peitgen, and Klaus Klose. Analysis of Mice Tumor Models Using Dynamic MRI Data and a Dedicated Software Platform. 176:1226–31, 10 2004.

[All] Medical Imaging & Technology Alliance. DICOM - Digital Imaging and Communications in Medicine. URL: http://www.dicomstandard.org/. last accessed on February 19, 2018.

[AOH+14] Paolo Angelelli, Steffen Oeltze, Judit Haász, Cagatay Turkay, Erlend Hodneland, Arvid Lundervold, Astri J. Lundervold, Bernhard Preim, and Helwig Hauser. Interactive visual analysis of heterogeneous cohort-study data. *IEEE Computer Graphics and Applications*, 34(5):70–82, 2014.

[BCS12] Tony Blakely, Roy Costilla, and Matthew Soeberg. *Cancer excess mortality rates over 2006-2026 for ABC-CBA*. Department of Public Health, University of Otago, Wellington, 2012.

[Bos] Mike Bostock. D3 - Data-Driven Documents. URL: https://d3js.org/. last accessed on February 1, 2018.

[Bos17a] Mike Bostock. D3 Bubble chart. URL: https://bl.ocks.org/mbostock/4063269, 2017. last accessed on Feburary 20, 2018.

[Bos17b] Mike Bostock. D3 Force-directed graph. URL: https://bl.ocks.org/mbostock/4062045, 2017. last accessed on February 12, 2018.

[Bos17c] Mike Bostock. D3 Treemap. URL: https://bl.ocks.org/mbostock/4063582, 2017. last accessed on February 20, 2018.

[Bos18a] Mike Bostock. D3 Bar chart. URL: https://bl.ocks.org/mbostock/3885304, 2018. last accessed on February 12, 2018.

[Bos18b] Mike Bostock. D3 Circle packing. URL: https://bl.ocks.org/mbostock/4063530, 2018. last accessed on February 20, 2018.

[Bos18c] Mike Bostock. D3 Donut multiples. URL: https://bl.ocks.org/mbostock/3888852, 2018. last accessed on February 20, 2018.

[BSH+97] Gunnar Brix, Laura M. Schreiber, Udo Hoffmann, Friedemann J. Gückel, Hans Hawighorst, and Michael V. Knopp. Methodological approaches to quantitative evaluation of microcirculation in tissues with dynamic magnetic resonance tomography. *Der Radiologe*, 37(6):470–480, 1997.

[BVD+08] Barbara Belletti, Jayant S. Vaidya, Sara D'Andrea, Frank Entschladen, Mario Roncadin, Francesca Lovat, Stefania Berton, Tiziana Perin, Ezio Candiani, Sonia Reccanello, Andrea Veronesi, Vincenzo Canzonieri, Mauro G. Trovò, Kurt S. Zaenker, Alfonso Colombatti, Gustavo Baldassarre, and Samuele Massarut. Targeted intraoperative radiotherapy impairs the stimulation of breast cancer cell proliferation and invasion caused by surgical wounding. *Clinical Cancer Research*, 14(5):1325–1332, 2008.

[CGB+05] Ernesto Coto, Sören Grimm, Stefan Bruckner, Eduard Gröller, Armin Kanitsar, and Omaira Rodriguez. *MammoExplorer: an advanced CAD application for breast DCE-MRI*. Citeseer, 2005.

[CHPS05] Hung-Han Chen, Lawrence Hunter, Harry Towsley Poteat, and Kristin Kendall Snow. Machine learning method, July 12 2005. US Patent 6,917,926.

[Con] World Wide Web Consortium. CSS Cascading Style Sheets. URL: https://www.w3.org/Style/CSS/. last accessed on February 19, 2018.

[Cor] Oracle Corporation. MySQL - My Structured Query Language. URL: https://www.mysql.com/. last accessed on February 19, 2018.

[CVS+13] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

[FAB+02] Bernard Fisher, Stewart Anderson, John Bryant, Richard G. Margolese, Melvin Deutsch, Edwin R. Fisher, Jong-Hyeon Jeong, and Norman Wolmark. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *New England Journal of Medicine*, 347(16):1233–1241, 2002.

[GAW+11] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[GBW17] Mary K. Gospodarowicz, James D. Brierley, and Christian Wittekind. *TNM classification of malignant tumours*. John Wiley & Sons, 2017.

[Gim17] Xavi Gimenez. D3 Waffle chart. URL: `http://bl.ocks.org/XavierGimenez/8070956`, 2017. last accessed on January 24, 2018.

[GPTP10] Sylvia Glaßer, Uta Preim, Klaus Tönnies, and Bernhard Preim. A visual analytics approach to diagnosis of breast DCE-MRI data. *Computers & Graphics*, 34(5):602–611, 2010.

[gro] The PHP group. PHP.net. URL: `http://php.net`. last accessed on February 13, 2018.

[Gru17] Jens Grubert. D3 Boxplot. URL: `http://bl.ocks.org/jensgrubert/7789216`, 2017. last accessed on March 30, 2018.

[Haf17] Chris Hafey. Cornerstone - Browser-based imaging library in JavaScript/HTML5. URL: `https://github.com/cornerstonejs/cornerstone`, 2017. last accessed on February 13, 2018.

[HASH10] Mohammad S. U. Hassan, Jawaher Ansari, David Spooner, and Syed A. Hussain. Chemotherapy for breast cancer. *Oncology reports*, 24(5):1121–1131, 2010.

[HB03] Mark Harrower and Cynthia A. Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[HBG02] Sean Ho, Elizabeth Bullitt, and Guido Gerig. Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 532–535. IEEE, 2002.

[HH01] Joseph V. Hajnal and Derek L. G. Hill. *Medical Image Registration*. Biomedical Engineering. CRC Press, 2001.

[IFP97] Victoria Interrante, Henry Fuchs, and Stephen M. Pizer. Conveying the 3D shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization and Computer Graphics*, 3(2):98–117, 1997.

[ivm16] ivmartel. DWV - Dicom Web Viewer. URL: `https://github.com/ivmartel/dwv`, 2016. last accessed on February 13, 2018.

[JBC+11] Ahmedin Jemal, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[jf] The jQuery fundation. jQuery - A JavaScript library. URL: https://jquery.com/. last accessed on February 19, 2018.

[KAFA14] Vinay Kumar, Abul K. Abbas, Nelson Fausto, and Jon C. Aster. *Robbins and Cotran pathologic basis of disease, professional edition e-book*. Elsevier health sciences, 2014.

[KB88] Jennifer L. Kelsey and Gertrud S. Berkowitz. Breast cancer epidemiology. *Cancer research*, 48(20):5615–5623, 1988.

[KCK17] Kyungyoon Kim, John V. Carlis, and Daniel F. Keefe. Comparison techniques utilized in spatial 3D and 4D data visualizations: A survey and future directions. *Computers & Graphics*, 67:138–147, 2017.

[KERC09] Daniel Keefe, Marcus Ewert, William Ribarsky, and Remco Chang. Interactive coordinated multiple-view visualization of biomechanical motion data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1383–1390, 2009.

[KJK+15] Kyungyoon Kim, Bret Jackson, Ioannis Karamouzas, Moses Adeagbo, Stephen J. Guy, Richard Graff, and Daniel F. Keefe. Bema: A multimodal interface for expert experiential analysis of political assemblies at the pnyx in ancient Greece. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, pages 19–26. IEEE, 2015.

[KSL+05] Christiane K. Kuhl, Simone Schrading, Claudia C. Leutner, Nuschin Morakkabati-Spitz, Eva Wardelmann, Rolf Fimmers, Walther Kuhn, and Hans H. Schild. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of clinical oncology*, 23(33):8469–8476, 2005.

[KWS+99] Michael V. Knopp, Eva C. Weiss, Hans-Peter Sinn, Johannes Mattern, Hans Junkermann, Jannis Radeleff, Achim Magener, Gunnar Brix, Stefan Delorme, Ivan Zuna, and Gerhard van Kaick. Pathophysiologic basis of contrast enhancement in breast tumors. *Journal of Magnetic Resonance Imaging*, 10(3):260–266, 1999.

[Lak12] Sunil R. Lakhani. *WHO Classification of Tumours of the Breast*. International Agency for Research on Cancer, 2012.

[LBI+11] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Seven guiding scenarios for information visualization evaluation. 2011.

[Lin18] David Lin. D3 Radial boxplot. URL: https://bl.ocks.org/davidwclin/ad5d13db260caeffe9b3, 2018. last accessed on February 20, 2018.

[MEV+06] Matej Mlejnek, Pierre Ermes, Anna Vilanova, Rob van der Rijt, Harrie van den Bosch, Frans A. Gerritsen, and Eduard Gröller. Application-oriented extensions of profile flags. In *EuroVis*, pages 339–346, 2006.

[MHG10] Muhammad M. Malik, Christoph Heinzl, and Eduard Gröller. Comparative visualization for parameter studies of dataset series. *IEEE Transactions on Visualization and Computer Graphics*, 16(5):829–840, 2010.

[mrd17] mrdoob. Three.js - JavaScript 3D library. URL: `https://threejs.org`, 2017. last accessed on February 13, 2018.

[MSD00] Klim McPherson, Michael C. Steel, and Michael J. Dixon. ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics. *BMJ: British Medical Journal*, 321(7261):624, 2000.

[Nar] Arpit Narechania. D3 Spotmatrix. URL: `https://arpitnarechania.github.io/d3-spotmatrix/`. last accessed on January 10, 2018.

[NH16] David Newitt and Nola Hylton. Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy, 2016. doi:10.7937/K9/TCIA. 2016.QHsyhJKy.

[NT10] Laurent Najman and Hugues Talbot. Introduction to mathematical morphology. *Mathematical morphology: From theory to applications*, pages 1–33, 2010. Wiley Online Library.

[OEGM09] Adedayo A. Onitilo, Jessica M. Engel, Robert T. Greenlee, and Bickol N. Mukesh. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clinical medicine & research*, 7(1-2):4–13, 2009.

[Org16] World Health Organization. International Classification of Diseases. 2010. 2016. URL: `http://www.who.int/classifications/icd/en`, last accessed on February 16, 2018.

[PB13] Bernhard Preim and Charl P. Botha. *Visual computing for medicine: theory, algorithms, and applications.* Newnes, 2013.

[POM+09] Bernhard Preim, Steffen Oeltze, Matej Mlejnek, Eduard Gröller, Anja Hennemuth, and Sarah Behrens. Survey of the visual exploration and analysis of perfusion data. *IEEE Transactions on Visualization and Computer Graphics*, 15(2):205–220, 2009.

[PPA+12] Madhura N. Phadke, Lifford Pinto, Oluwafemi Alabi, Jonathan Harter, Russell M. Taylor, Xunlei Wu, Hannah Petersen, Steffen A. Bass, and Christopher G. Healey. Exploring ensemble visualization. In *Visualization and Data Analysis 2012*, volume 8294, page 82940B. International Society for Optics and Photonics, 2012.

[PW13] Tobias Pfaffelmoser and Rüdiger Westermann. Visualizing contour distributions in 2D ensemble data. *EuroVis-Short Papers*, pages 55–59, 2013.

[Rai17] Renata G. Raidou. *Visual analytics for digital radiotherapy: towards a comprehensible pipeline*. PhD thesis, Technische Universiteit Eindhoven, 2017.

[RBRH+17] Nicolas Rannou, Jorge Luis Bernal-Rusiel, Daniel Haehn, Patricia Ellen Grant, and Pienaar Rudolph. AMI - Medical Imaging JavaScript ToolKit. URL: `https://github.com/FNNDSC/ami`, 2017. last accessed on February 13, 2018.

[RBV+14] Renata G. Raidou, Marcel Breeuwer, Anna Vilanova, Uulke A. van der Heide, and Petra J. van Houdt. The iCoCooN: Integration of Cobweb Charts with Parallel Coordinates for Visual Analysis of DCE-MRI Modeling Variations. In *VCBM*, pages 11–20, 2014.

[Rhe96] Penny Rheingans. Opacity-modulating triangular textures for irregular surfaces. In *Proceedings of the 7th conference on Visualization'96*, pages 219–ff. IEEE Computer Society Press, 1996.

[Ric17] David Richards. D3 Simple tree. URL: `https://bl.ocks.org/denjn5/4bdda16ab6338772a7882ec2f5f2a9c7`, 2017. last accessed on March 19, 2018.

[Rob05] Jonathan C. Roberts. Exploratory visualization with multiple linked views. In *Exploring geovisualization*, pages 159–180. Elsevier, 2005.

[RU] RII-UTHSCSA. rii-mango/Papaya - a pure JavaScript medical research image viewer. URL: `https://github.com/rii-mango/Papaya`. last accessed on February 21, 2018.

[RWS+10] Christian Rieder, Andreas Weihusen, Christian Schumann, Stephan Zidowitz, and Heinz-Otto Peitgen. Visual support for interactive post-interventional assessment of radiofrequency ablation therapy. In *Computer Graphics Forum*, volume 29, pages 1093–1102. Wiley Online Library, 2010.

[SCGM00] John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International journal of human-computer studies*, 53(5):663–694, 2000.

[SGB13] Johanna Schmidt, Eduard Gröller, and Stefan Bruckner. VAICo: Visual Analysis for Image Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2090–2099, 2013.

[SHO+04] Debbie Saslow, Judy Hannan, Janet Osuch, Marianne H. Alciati, Cornelia Baines, Mary Barton, Janet K. Bobo, Cathy Coleman, Mary Dolan, Ginny Gaumer, Daniel Kopans, Susan Kutner, Dorothy S. Lane, Herschel Lawson, Helen Meissner, Candace Moorman, Henry Pennypacker, Peggy Pierce, Eva Sciandra, Robert Smith, and Ralph Coates. Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA: a cancer journal for clinicians*, 54(6):327–344, 2004.

[Sie17] Andrew Sielen. D3 Violin plot. URL: `http://bl.ocks.org/asielen/1a5e8d77ae8feb464167`, 2017. last accessed on February 20, 2018.

[SK13] Hans-Peter Sinn and Hans Kreipe. A brief overview of the WHO classification of breast tumors. *Breast Care*, 8(2):149–154, 2013.

[SKK+14] David Schroeder, Fedor Korsakov, Carissa Mai-Ping Knipe, Lauren Thorson, Arin M. Ellingson, David Nuckley, John Carlis, and Daniel F. Keefe. Trend-centric motion visualization: Designing and applying a new strategy for analyzing scientific motion collections. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2644–2653, 2014.

[SMvB+10] Martijn D. Steenwijk, Julien Milles, Mark A. van Buchem, Johan H. C. Reiber, and Charl P. Botha. Integrated visual analysis for heterogeneous datasets in cohort studies. In *IEEE VisWeek Workshop on Visual Analytics in Health Care*, volume 3, page 3, 2010.

[SPA+14] Johanna Schmidt, Reinhold Preiner, Thomas Auzinger, Michael Wimmer, Eduard Gröller, and Stefan Bruckner. YMCA—Your mesh comparison application. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pages 153–162. IEEE, 2014.

[STH02] Chris Stolte, Diane Tang, and Pat Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65, 2002.

[Stu17] DecemberCafe Studio. D3 Mind map. URL: `http://www.decembercafe.org/demo/mindmap/`, 2017. last accessed on February 20, 2018.

[SW01] Ben Shneiderman and Martin Wattenberg. Ordered treemap layouts. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on*, pages 73–78. IEEE, 2001.

[TB94] Paul S. Tofts and Bruce A. Berkowitz. Measurement of capillary permeability from the Gd enhancement curve: a comparison of bolus and constant infusion injection methods. *Magnetic resonance imaging*, 12(1):81–91, 1994.

[TFJ12] Christian Tominski, Camilla Forsell, and Jimmy Johansson. Interaction support for visual comparison inspired by natural behavior. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2719–2728, 2012.

[TMA01] Melanie K. Tory, Torsten Möller, and Stella M. Atkins. Visualization of time-varying MRI data for MS lesion analysis. In *Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*, volume 4319, pages 590–599. International Society for Optics and Photonics, 2001.

[Tur09] Lindsay W. Turnbull. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. *NMR in Biomedicine*, 22(1):28–39, 2009.

[Tur18] William Turman. D3 Interactive streamgraph. URL: `http://bl.ocks.org/WillTurman/4631136`, 2018. last accessed on February 20, 2018.

[vBH89] Stef van Buuren and Willem J. Heiser. Clustering n objects into k groups under optimal scaling of variables. *Psychometrika*, 54(4):699–706, 1989.

[vLBK+13] Tatiana von Landesberger, Sebastian Bremm, Matthias Kirschner, Stefan Wesarg, and Arjan Kuijper. Visual analytics for model-based medical image segmentation: Opportunities and challenges. *Expert Systems with Applications*, 40(12):4934–4943, 2013.

[WKB+15] Shandong Wu, Brenda F. Kurland, Wendie A. Berg, Margarita L. Zuley, Rachel C. Jankowitz, Jules Sumkin, and David Gur. Signal enhancement ratio (SER) quantified from breast DCE-MRI and breast cancer risk. *Proc.SPIE*, 9414:9414 – 9414 – 6, 2015.

[WS06] Jonathan Woodring and Han-Wei Shen. Multi-variate, time varying, and comparative visualization with contextual cues. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):909–916, 2006.

[YLC+07] Thomas E. Yankeelov, Martin Lepage, Anuradha Chakravarthy, Elizabeth E. Broome, Kenneth J. Niermann, Mark C. Kelley, Ingrid Meszoely, Ingrid A. Mayer, Cheryl R. Herman, Kevin McManus, Ronald R. Price, and John C. Gore. Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results. *Magnetic resonance imaging*, 25(1):1–13, 2007.

[YLPM+08] Wei T. Yang, Huong T. Le-Petross, Homer Macapinlac, Selin Carkaci, Ana M. Gonzalez-Angulo, Shaheenah Dawood, Erika Resetkova, Gabriel N. Hortobagyi, and Massimo Cristofanilli. Inflammatory breast cancer: PET/CT, MRI, mammography, and sonography findings. *Breast cancer research and treatment*, 109(3):417–426, 2008.

[Zho18] Chris Zhou. D3 Radar chart. URL: `http://bl.ocks.org/chrisrzhou/2421ac6541b68c1680f8`, 2018. last accessed on February 12, 2018.