

LungStage

**Die Zukunft in der Befundung von
Lungenkrebs**

MAS Human Computer Interaction Design
Masterarbeit Hochschule für Technik Rapperswil und Universität Basel
2017/18

Angela Dannhorn, Immanuel Häussermann, Stefan Pfister
Betreuer: Christian Heusser

Abstract

Ziel

Diese Arbeit untersuchte, wie die automatisierte Erkennung von Lungenkrebs durch künstliche Intelligenz (nachfolgend «KI» genannt) in einem User Interface (nachfolgend «UI» genannt) bereitgestellt werden kann. Die Diagnose von Lungenkrebs sollte beschleunigt und die Radiologen entlastet werden.

Hintergrund

Durch die Computertomographie entstehen in der Schweiz jährlich Petabytes an radiologischen Bilddaten, welche von Fachärzten für Diagnostik und Behandlungsplanung analysiert werden. Diese Analyse ist kostspielig und zeitaufwändig. Eine PET/CT Tomographie eines Lungenkrebspatienten beinhaltet rund 6'000 Einzelbilder. Eine KI kann darauf trainiert werden, Tumore und Metastasen in den Bildern zu finden und so den Prozess zu beschleunigen. Da die Ergebnisse der KI aber nicht deterministisch sind, musste ein Weg gefunden werden, wie das Resultat der Maschine dem Menschen vertrauenswürdig kommuniziert werden konnte.

Methode

Mit qualitativen Methoden sollte untersucht werden, wie Radiologen Lungenkrebs befunden. In einem theoretischen Teil wurde untersucht, welche Aspekte wichtig waren, damit der Radiologe dem Output der KI vertrauen konnte. Danach wurde anhand von drei Iterationen mit Prototypen in steigender Detailtreue untersucht, wie das UI für LungStage aussehen sollte. Die Prototypen wurden mit Radiologen und Nuklearmedizinern in drei Spitälern evaluiert.

Resultat

Das Hauptergebnis der Arbeit war ein klickbarer, animierter Hi-Fi Prototyp. Er demonstrierte, wie die Benutzer in Zukunft mit Hilfe von KI Lungenkrebs befunden könnten und ermöglichte dem Auftraggeber bereits während dem Projekt, weitere Partner für die Zusammenarbeit und Weiterentwicklung zu finden.

Anwendung und nächste Schritte

Das erarbeitete Design bildete die Grundlage für die weitere Entwicklung durch den Auftraggeber. Das Projektteam empfahl, das Design mit der KI zu verbinden und mit weiteren Testpersonen zu evaluieren, die bisher noch nicht am Projekt beteiligt waren.

Danksagung

Das Projektteam bedankt sich ganz herzlich bei folgenden Personen und Organisationen:

- Flavio Trolese (Mitgründer und Partner 4Quant Ltd.) für sein hohes Engagement, die Begeisterungsfähigkeit und die hartnäckige Rekrutierung.
- Christian Heusser für die gute Unterstützung, das sympathische Coaching und die erfrischende Perspektive auf unsere Arbeit.
- Kevin Mader (4Quant) und Thomas Weikert (USB) für die fortlaufende Unterstützung bei Fragen zu technologischen, radiologischen und auch organisatorischen Aspekten.
- Die RadiologInnen und NuklearmedizinerInnen des Unispitals Basel, des Unispitals Zürich und des Kantonsspitals Luzern.
- Impact Hub Zürich und Ginetta für die Bereitstellung der Räumlichkeiten für die Coachings, Workshops und langen Arbeitstagen.

Erklärung der Selbstständigkeit

Hiermit bestätigen wir folgende Punkte:

- Die vorliegende Arbeit haben wir selber und ohne fremde Hilfe durchgeführt, ausser derjenigen, welche explizit beschrieben ist.
- Sämtliche verwendeten Quellen haben wir erwähnt und gemäss gängigen wissenschaftlichen Regeln korrekt zitiert.
- Wir haben keine durch Copyright geschützten Materialien (z.B. Bilder) in dieser Arbeit in unerlaubter Weise genutzt.
- Wir haben in dieser Arbeit keine Adressen, Telefonnummern und andere persönliche Daten von Personen publiziert, die nicht zum Kernteam gehören.

Zürich, 29.01.2018

Angela Dannhorn

Immanuel Häussermann

Stefan Pfister

Abstract	5	5. Iteration 1	49
1. Ausgangslage	7	5.1 Papier Prototyp	50
1.1 Auftraggeber	8	5.2 Usability Walkthrough mit Papier Prototyp	55
1.2 Team	8	6. Iteration 2	61
1.3 Ausgangslage	8	6.1 Klickbare Wireframes	62
1.4 Projektumfang	9	6.2 Usability Walkthrough mit Wireframes	71
1.5 Zielsetzung	10	7. Iteration 3	77
1.6 Fragestellung	11	7.1 Hi-Fi Prototyp	78
1.7 Abgrenzung	11	8. Fazit	89
2. Vorgehen	13	8.1 Bewertung der Ergebnisse	91
2.1 Vorgehensmodell	14	8.2 Empfehlung für den Auftraggeber	92
2.2 Projektmanagement	16	8.3 Teamreflexion	93
2.3 Fachliche Erkenntnisse	19	9. Quellenverzeichnis	94
3. Nutzungskontext verstehen	21	10. Abbildungsverzeichnis	96
3.1 Stakeholder Interview	22	11. Themenrecherche	98
3.2 Konkurrenzanalyse	24	12. Anhang	112
3.3 Contextual Inquiry	26		
3.4 Affinity Diagramm	29		
3.5 Umfrage	32		
3.6 Artefaktanalyse	35		
3.7 Themenrecherche	36		
4. Nutzungsanforderungen spezifizieren	39		
4.1 User Stories	40		
4.2 Persona	42		
4.3 Kontext Szenarien	45		
4.4 Concept Model	46		



Abb. 1: Canon Celesteion

PET/CT-Scanner

Ein PET/CT-Scanner wird verwendet, um mit Hilfe von Strahlung in das Innere eines Patienten schauen zu können. Eine Untersuchung dauert etwa eineinhalb Stunden und generiert bis zu 6000 Einzelbilder. Die Bilder werden anschliessend in einem Picture Archiving and Communication System (PACS) für die weitere Analyse gespeichert.

1. Ausgangslage

Die vorliegende Masterarbeit wurde im Rahmen des berufsbegleitenden Masterstudiums in Human Computer Interaction Design (MAS HCID) der Hochschule für Technik Rapperswil (HSR) erstellt. Die in den beiden Kursen «CAS Requirements Engineering» und «CAS Interaction Design» erlernten Inhalte wurden in einem realen Projekt mit einem Praxispartner angewendet.

Im folgenden Bericht werden der Projektrahmen und die Vorgehensweise mit ihren Methoden erläutert, die Lösung vorgestellt und eine Diskussion der Ergebnisse geführt. Im Anhang finden sich die erarbeiteten Artefakte, die im Laufe dieser Masterarbeit erarbeitet wurden. Der Bericht als Ganzes folgt chronologisch dem Projektverlauf, welcher sich am gewählten Vorgehensmodell orientiert.

1.1 Auftraggeber

Computer-aided diagnosis, (kurz CAD) beschreibt ein Verfahren zur Unterstützung des Arztes bei der Interpretation von Untersuchungsergebnissen.

Der Auftraggeber und Praxispartner dieser Masterarbeit war das Unternehmen 4Quant Ltd. (nachfolgend «4Quant» genannt), ein Spin-Off der ETH Zürich und des Paul Scherrer Instituts. Das Startup entwickelt Lösungen für die (Teil-)Automatisierung von Bildanalysen. 4Quant entwickelt in Zusammenarbeit mit der Abteilung Radiologieforschung des Unispitals Basel ein CAD-System für Desktop – genannt «LungStage» – für die semi-automatisierte Diagnose von Lungenkrebs.

1.2 Team

Das Projektteam bestand aus zwei Interaction DesignerInnen – Angela und Stefan – und einem Softwareentwickler – Immanuel. Alle drei Teammitglieder besaßen mehrjährige Erfahrung in der Gestaltung und Entwicklung von Software-Applikationen, hatten jedoch kein Vorwissen auf dem Gebiet der Radiologie oder der künstlichen Intelligenz.

1.3 Ausgangslage

Aus Gründen der leichteren Lesbarkeit wird auf «Nuklearmediziner» verzichtet und die männliche Sprachform «Radiologen» verwendet.

Non-small cell lung cancer ist die häufigste Art von Lungenkrebs.

Für die Erkennung von Lungenkrebs werden CT- und PET-Tomographien von Lungenkrebspatienten gemacht, welche danach von Radiologen und Nuklearmedizinern auf Tumore untersucht werden. Dieser Prozess ist zeitaufwändig und fehleranfällig und soll daher durch LungStage semi-automatisiert werden.

Das System LungStage befindet sich im frühen Entwicklungsstadium. Es erkennt und klassifiziert mit Hilfe von Bildanalyse-Technologie und künstlicher Intelligenz (nachfolgend «KI» genannt) die Tumorerkrankung semi-automatisch und präsentiert dem Radiologen die Erkenntnisse. Die KI wurde von 4Quant auf die Erkennung von Nicht-Kleinzelligem Lungenkrebs (NSCLC) trainiert. Langfristig soll die Verwendung von LungStage die Befundung des Patienten beschleunigen, die Qualität der Diagnose verbessern und die Radiologen entlasten.

LungStage verfügte zum Projektstart über kein ausgereiftes User Interface (nachfolgend «UI» genannt). Es existierten verschiedene Technologie-Prototypen, die die Resultate in Form von rudimentären Texten oder Grafiken ausgaben. Diese waren jedoch nicht mit Benutzern validiert.

1.4 Projektumfang

Diese Masterarbeit befasste sich mit einer Teilaufgabe, welche in einen langen Prozess zwischen Ärzten und Patienten eingebettet war:

Schritte	
1	Ein Patient kommt mit Beschwerden in der Lunge zum Hausarzt. Er wird bei Verdacht auf Lungenkrebs an ein Spital überwiesen.
2	Im Spital wird ein Oberkörper-CT gemacht, bei dem ca. 1000 Einzelbilder entstehen.
3	Ein Radiologe analysiert die CT-Bilder. Der Verdacht auf Lungenkrebs erhärtet sich.
4	Ein PET/CT-Scan des Patienten wird gemacht, um Ableger des Tumors zu finden. Es entstehen ca. 6000 Einzelbilder.
5	Ein Nuklearmediziner analysiert die Bilder und erstellt einen Befund der Krankheit, in dem er den Tumor beschreibt. Dies beinhaltet folgende Aufgaben:
5 a	Einen optionalen Vergleich der Bilddaten mit früheren Untersuchungen.
5 b	Eine Untersuchung und Interpretation der aktuellen Bilddaten hinsichtlich allgemeinen Auffälligkeiten.
5 c	Eine Untersuchung und Interpretation der aktuellen Bilddaten im Kontext von Lungenkrebs.
5 d	Einen Befundungsbericht erstellen.
6	Der Befundungsbericht wird von Radiologen und den weiteren Ärzten – Onkologen, Pathologen, Pneumologen, Chirurgen – diskutiert. Es wird eine Entscheidung für die Therapie des Patienten gefällt.

Dieser Prozess der erstmaligen Befundung des Tumors wird als «Staging» bezeichnet. Nach der ersten Behandlung wiederholt sich der Prozess ab Schritt 4 und man kontrolliert den Verlauf der Erkrankung, was als «Re-Staging» bezeichnet wird. Die Applikation LungStage könnte langfristig zu einer Entlastung bei fast allen Arbeitsschritten führen. In dieser Arbeit wird der Einsatz von LungStage für die **erstmalige Untersuchung und Interpretation der Bilddaten (5c)** im Kontext von Lungenkrebs behandelt.

1.5 Zielsetzung

Design des UI

Im Rahmen dieser Arbeit sollte ein UI für LungStage benutzerzentriert erarbeitet und validiert werden. Dieses würde es den Radiologen ermöglichen ihre Arbeit effizienter und effektiver zu erledigen. Basierend auf der konzipierten Lösung sollte der Auftraggeber in der Lage sein, ein lauffähiges System zu entwickeln, das mit realen Daten und weiteren Benutzern getestet werden könnte. Dieses sollte für die Prüfung der medizinischen Zulassung und als Verkaufsinstrument an mögliche Investoren oder Käufer dienen.

Die Effizienz des zu gestaltenden UI sollte mit Zeitmessungen belegt werden. Dies stellte sich im Verlauf der Arbeit als eine Herausforderung dar.

Akzeptanz des Systems

Das System arbeitet mit einer künstlicher Intelligenz im Hintergrund, welches wichtige Arbeitsschritte der Radiologen automatisiert. Problematisch ist, dass nicht transparent ist, wie das Ergebnis zustande kommt («Blackbox»-Charakter von KI). Dies stellte ein Risiko für die Akzeptanz des Systems dar, sollte der Radiologe kein Vertrauen in die präsentierten Ergebnisse haben. Somit muss die Interaktion zwischen Mensch und Maschine so gestaltet werden, dass das System vom Menschen akzeptiert wird. Bei der Erarbeitung des UI wurde versucht, dieses Risiko durch benutzerzentrierte Gestaltung zu entschärfen.

Für die HCID Community ist das Thema interessant, weil Systeme mit KI vermehrt zum Einsatz kommen (Brandt, M., 2017) und sich dadurch Chancen ergeben, das Benutzererlebnis massgeblich zu verbessern.

Lernziele

Die erlernten Methoden der CAS-Zertifikatskurse «Requirements Engineering» und «Interaction Design» sollten eingesetzt und geübt werden. Dabei sollten mögliche Stärken und Schwächen der Methoden beleuchtet werden, um entsprechende Erfahrungen auf die berufliche Praxis übertragen zu können.

Das Projektteam legte den Fokus auf die Einarbeitung in ein neues Feld: die künstliche Intelligenz. Es wollte begreifen, wie man bei Benutzern Akzeptanz in die Resultate von KI Systemen herstellt. Hierfür sollte das Thema basierend auf bestehenden Studien theoretisch untersucht und dokumentiert werden. Basierend auf dieser Analyse sollten die erlernten Massnahmen zur Stärkung der Akzeptanz ausgewählt, im UI umgesetzt und mit Benutzern validiert werden.

1.6 Fragestellung

Ausgehend von der beschriebenen Ausgangslage und Zielsetzung ergaben sich folgende Fragen für das Projekt:

Benutzer

- Wer sind die Benutzer von LungStage und welche Eigenschaften haben diese?
- Welche konkreten Ziele verfolgen die Benutzer bei der Befundung mit Hilfe von LungStage?

System

- Welche Anforderungen ergeben sich aus den Benutzerzielen für das Verhalten und die Funktionalität bei der Befundung?
- Wie müssen Form, Verhalten und Funktionalität der Befundung mit Hilfe von LungStage konkret ausgestaltet sein, um diese Anforderungen zu erfüllen?
- Wie kann der «Blackbox»-Charakter des automatischen Befundes entschärft und somit das Vertrauen in das System gefördert werden?

1.7 Abgrenzungen

Der Umfang dieser Arbeit beschränkt sich auf die Untersuchung und Interpretation der Bilddaten im Kontext von Lungenkrebs. Die übrigen Arbeitsschritte werden im Prototyp nicht berücksichtigt. In diesem Projekt wird die Darstellung der Resultate im UI betrachtet und optimiert, jedoch nicht die Qualität der KI Technologie, welche einen entscheidenden Einfluss auf das Vertrauen hat.

Weiterhin beschäftigt sich die Arbeit nicht mit rechtlichen Aspekten wie z.B. Patente auf Interaktionen oder mit den Marktrisiken der Produkteinführung an sich (Marketing, Produkteinführung bei Spitälern, Produktintegration in Umsysteme).

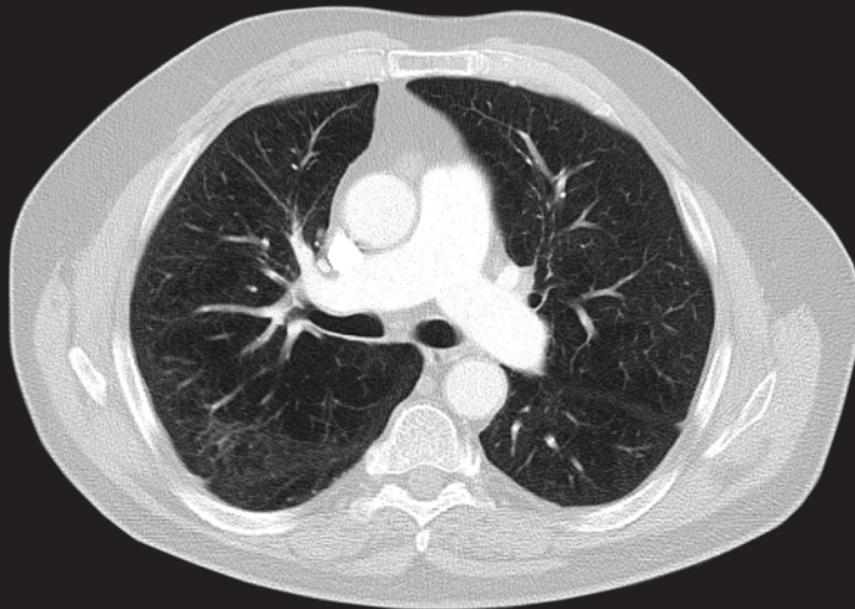


Abb. 2: CT-Bild eines Lungengewebes

CT-Bild

Mithilfe von Röntgenstrahlung werden Querschnittsbilder des Körpers erstellt. Diese Strahlen werden je nach Gewebe unterschiedlich stark abgeschwächt und zeigen dadurch die Anatomie des Patienten wie z.B. Knochen oder die Lunge.

2. Vorgehen

Für die Durchführung des Projektes wurde ein Vorgehensmodell gewählt, welches sich flexibel auf die besonderen Umstände im Projekt anpassen liess. Basierend darauf wurde ein Projektplan erstellt, welcher laufend aktualisiert wurde, aber grob nach den Phasen des Vorgehensmodells eingeteilt war. Er diente als zeitliches Kontrollinstrument und half bei Entscheidungen über den Projektverlauf. Die Erstellung einer Risikoliste ermöglichte es, ein gemeinsames Verständnis zu schaffen, welche Herausforderungen eintreten könnten und wie das Projektteam darauf reagieren würde. Die Lieferobjekte, die als Ergebnis am Ende des Projektes vorliegen sollten, wurden ebenfalls dokumentiert und dienten als Kontrolle, ob das Projekt insgesamt oder teilweise erfolgreich war.

2.1 Vorgehensmodell

Das Projekt startete mit vielen Herausforderungen und schwierig einzuschätzenden Faktoren:

- Eine Person im Projektteam war schwanger und Abwesenheiten waren zu erwarten.
- Die Rekrutierung der Testpersonen wurde durch den Auftraggeber durchgeführt und es war deshalb nicht absehbar, ob und wieviele Benutzer rekrutiert werden konnten.
- Durch das komplexe Themengebiet war schwierig abzuschätzen, wie lange die Erarbeitung des Domänenwissens benötigen wird.
- Der Einfluss einer für das Projektteam unbekanntem Technologie (KI) war schwierig einzuschätzen.

Um mit diesen Herausforderungen optimal umgehen zu können, wählte das Projektteam ein Vorgehen nach dem «EN ISO 9241-210» Standard. Diese DIN-Norm beschreibt in Form von Richtlinien das Vorgehen für die benutzerzentrierte Entwicklung von Mensch-Maschine-Interaktionen.

Der wichtigste Entscheidungsfaktor für das DIN Modell war ein iteratives Vorgehen mit flexibler Methodenwahl. Im Gegensatz dazu stehen die Vorgehensweisen «Goal-Directed Design» (Cooper, A., 2014) oder «Usability-Engineering Lifecycle» (Mayhew, D., 1999), die fixe Methoden, Rollen und Lieferobjekte vorgeben.

Das Hauptziel war die Erarbeitung des Interaktions-Designs für das UI. Das Vorgehen nach «Contextual Design» (Beyer, H., 1998) wurde deshalb als nicht optimal erachtet, da dieses vor allem auf die Research Phase fokussiert. Durch die ungewisse Ressourcenplanung des Teams, war es ebenfalls wichtig keine fixe Rollenverteilung im Team zu etablieren, wie es z.B. im Usability-Engineering Lifecycle der Fall ist.

Durch die Wahl des DIN Modells konnte das Vorgehen auf die spezifischen Bedingungen des Praxisprojektes wie z.B. Zeitbudget, existierender Prototyp und Wissensstand der Gruppe adaptiert werden.

Bei einem Vorgehen nach «ISO 9241-210» wird mit der Planung des Human Centered Design Prozess begonnen. Anschliessend werden die Phasen «Spezifikation des Benutzungskontextes», «Spezifikation der Benutzerbedürfnisse», «Erarbeitung von Designlösungen» und «Evaluation des Designs» iterativ durchlaufen, bis die Lösung den Benutzeranforderungen entspricht.

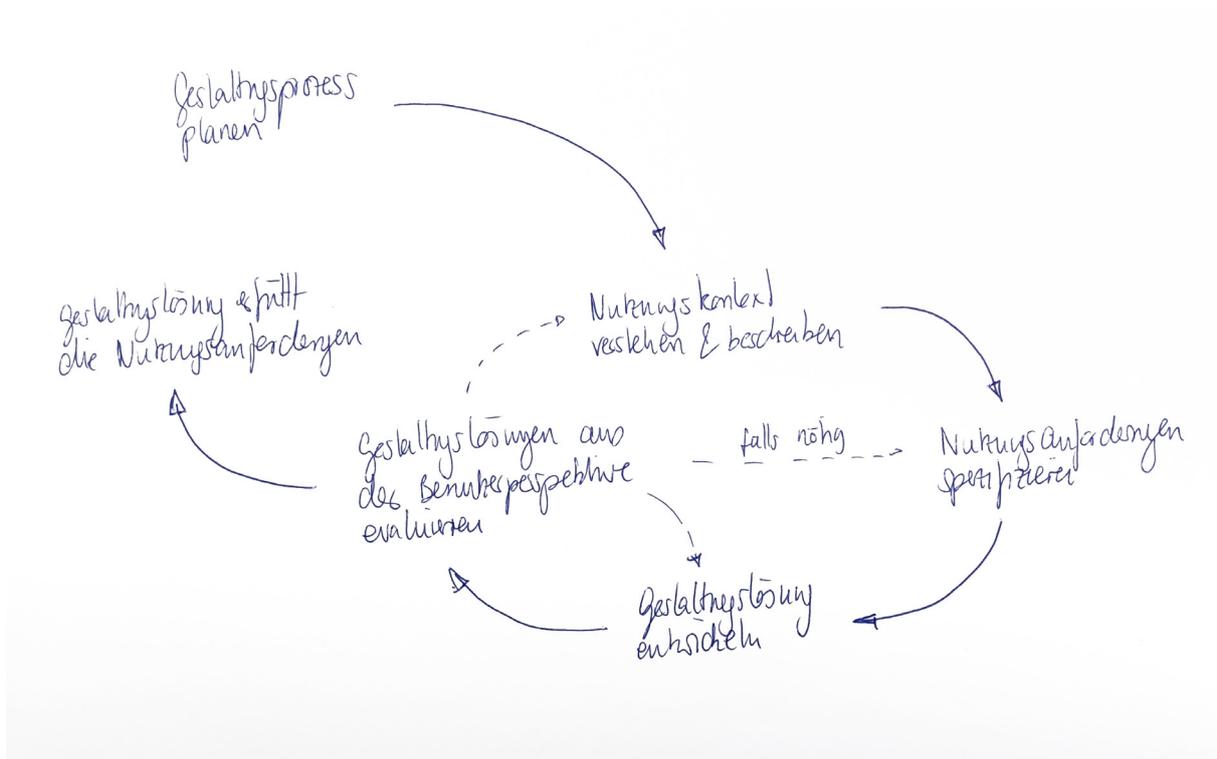
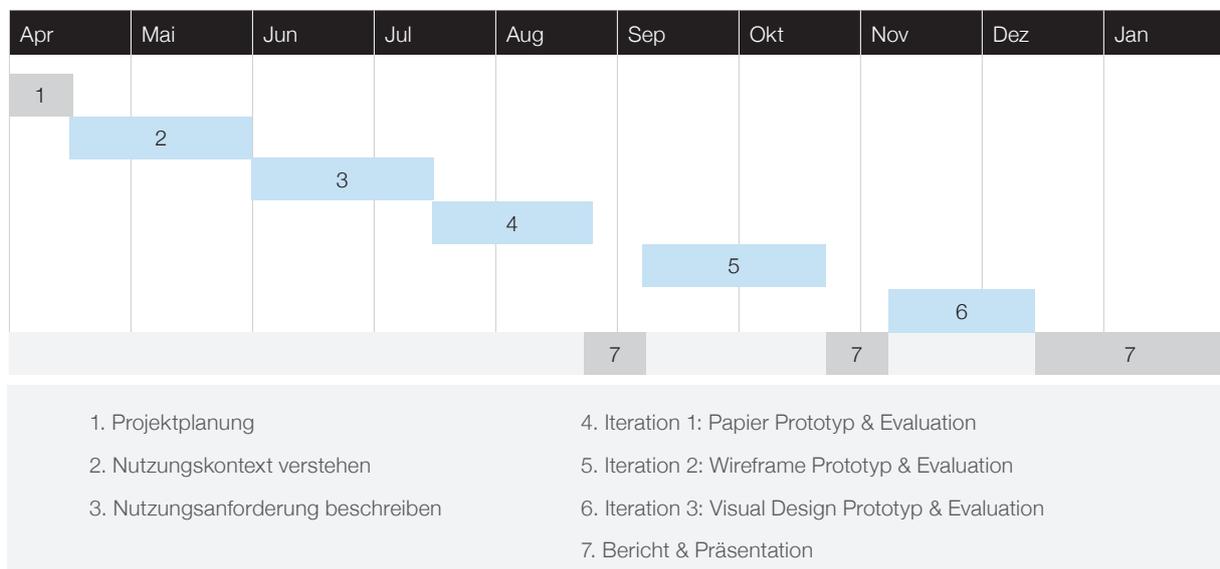


Abb. 3: Darstellung des EN ISO 9241-210 Ablaufs

2.2 Projektmanagement

Projektplan

Zu Beginn des Projektes wurde ein Projektplan erstellt, der es dem Projektteam ermöglichte, Massnahmen und Abwesenheiten zu planen und den Projektfortschritt zu kontrollieren. Der Projektplan wurde grob nach den DIN Phasen aufgeteilt, wobei die Design- und Evaluations-Phasen jeweils in einer Iteration zusammengefasst wurden:



Der vollständige Projektplan ist im Anhang ersichtlich.

Der Schwerpunkt der Design Phasen (Iteration 1 bis 3) lag darin, Lösungen für die vorgängig identifizierten Benutzerbedürfnisse zu entwickeln und mit Benutzern zu testen.

Da die Masterarbeit berufs begleitend durchgeführt wurde, einigte sich das Projektteam auf einen Arbeitstag pro Woche, um gemeinsam zu arbeiten. Das Projektteam passte den Projektplan laufend den aktuellen Bedürfnissen, dem Erkenntnisstand der Arbeit oder der Verfügbarkeit der Teilnehmer an. Die Dokumentation wurde bereits zu Beginn gestartet und laufend ergänzt.

Um allfällige Änderungen bei Lieferobjekten abzustimmen, wurden regelmässige Treffen mit dem Projektpartner vereinbart. Diese fanden alle paar Wochen entweder persönlich oder per Telefon statt.

Risikoliste

In einer zu Beginn erstellten Risikoliste wurden alle Projekt- und Produkt-Risiken dokumentiert. Die Liste half für ein gemeinsames Verständnis, welche Hürden und Herausforderungen eintreten könnten und wie das Projektteam darauf reagieren würde. Die folgenden Risiken wurden als hoch eingestuft:

Beschreibung	Auswirkung	Gegenmassnahme
Nicht genügend Testpersonen, da es sich um hochgradige Experten handelt	Validierung der Prototypen nicht möglich	Frühzeitige Kommunikation mit Auftraggeber und frühzeitige Rekrutierung. In der Vereinbarung festhalten
Lückenhaftes Wissen durch die hohe Komplexität der Materie	Erschwerte Kommunikation mit Stakeholdern und fehlerhafte Gestaltung des Designs	Gute Einarbeitung in das Thema und enge Kollaboration der Wissensträger
Das System bietet nicht die richtigen Parameter, um eine Akzeptanz bei den Radiologen zu ermöglichen	Ablehnung der Lösung durch die Radiologen	Forschungsfokus auf wichtige Parameter setzen

Die komplette Risikoliste ist im Anhang ersichtlich.

Glossar

Die Fachsprache der Ärzte beinhaltet viele Abkürzungen und Fachbegriffe, welche als Laie nicht nachvollziehbar sind. Aufgrund des komplexen Themengebietes wurde ein Glossar erstellt, welches während dem Projekt laufend ergänzt wurde. Dies half dem Projektteam nicht nur, besser mit den Ärzten zu kommunizieren, sondern stellte sich auch als nützlich heraus für die Erstellung der Prototypen.

Das Glossar ist im Anhang ersichtlich.

Lieferobjekte

Die Gesamtheit der Objekte, die am Ende des Projektes vorliegen sollten, wurden bereits zu Beginn festgelegt und mit dem Auftraggeber abgeglichen. Sie dienen als Erfolgskontrolle und als verbindliche Vereinbarung zwischen dem Projektteam und dem Auftraggeber.

Artefakt	Auswirkung
Dokumentation	Projektdokumentation und Beschreibung der Vorgehensweisen, Arbeitsergebnisse sowie der Lerneffekte.
Personas	Segmentierung der Benutzer. Als Grundlage diente eine bestehende Persona des Auftraggebers. Diese wurde, basierend auf Beobachtungen der Benutzer, überarbeitet und validiert.
Szenarien	Beschreibung der einzelnen Situationen. Als Grundlage dienten Beobachtungen der Benutzer.
User Stories	Ausformulierte Software-Anforderung. Als Grundlage dienten Beobachtungen der Benutzer.
Interaction Design	Klickbarer Prototyp der wichtigsten Interaktionen. Als Grundlage dienten die Ergebnisse der Research Phase. Inklusive Validation mit Benutzern.
Visual Design	Look and Feel des Produktes. Dargestellt an ausgewählten Use Cases in einem animierten Prototypen.

2.3 Fachliche Erkenntnisse

In der Phase «Nutzungskontext verstehen» gab es einen zweieinhalbmonatigen ungeplanten Spitalaufenthalt von Angela. Hier bewährte sich die flexible Projektplanung. Als Reaktion auf den Unterbruch an der aktiven Projektarbeit von Angela entschied sich das Team, einen theoretischen Teil über Vertrauen in KI Systeme zu erstellen. Dies hatte zwei Vorteile: Einerseits ermöglichte es Angela weiterhin im Projekt involviert zu bleiben, andererseits stellte der theoretische Teil mitunter eine Grundlage für die spätere Gestaltung des UI dar.

Als weitere Konsequenz der reduzierten Ressourcen zu Beginn des Projektes wurde entschieden, die Phase «Nutzungskontext verstehen» auszudehnen, um weitere Erkenntnisse zu Benutzer und bestehenden Systemen erlangen zu können. Dies ermöglichte es dem Projektteam zudem ein weiteres Contextual Inquiry im Unispital Zürich durchzuführen und dadurch eine breitere Abdeckung von Spitälern zu erreichen.

Die Involvierung des Projektpartners ermöglichte es dem Projektteam viel zusätzliches Wissen zu generieren. Durch die enge Zusammenarbeit bestand die Gefahr, dass das Projekt zu sehr von den Erwartungen des Projektpartners gesteuert wurde (Projektrisiko «ID5»). Das Projektteam vereinbarte deshalb, dass die Projektarbeit unabhängig vom Zeitplan des Auftraggebers durchgeführt werden konnte, um dieses Risiko zu entschärfen. Aus diesem Grund wurde auch das Ziel verworfen, einen Prototypen zu erstellen, der mit dem zu entwickelnden KI System interagierte.

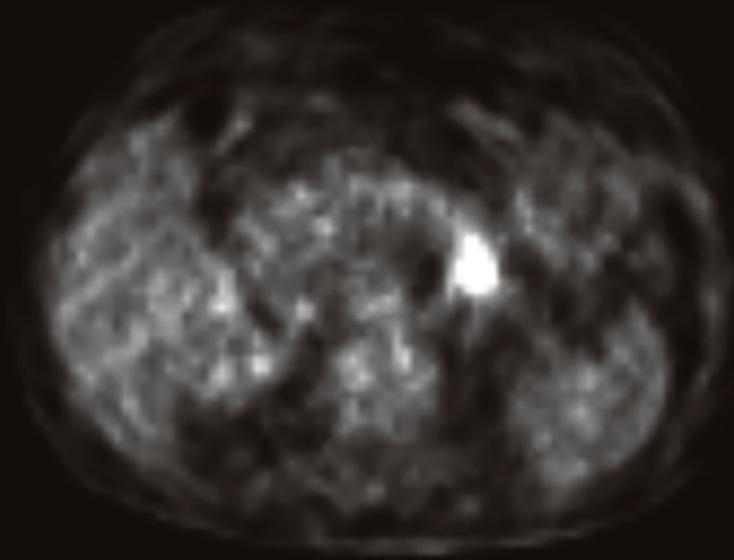


Abb. 4: PET Bild eines Lungengewebes

PET-Bild

Für das PET-Bild wird Energie von einem PET-Scanner aufgefangen und zu einem Bild verarbeitet. Dazu wird dem Patienten radioaktiver Zucker injiziert. Weil Tumorzellen mehr Glukose aufnehmen als normale Zellen, sind sie im PET-Bild deutlich sichtbar. Allerdings können auch Entzündungen zu einer höheren Verstoffwechslung führen. Diese können nur bedingt von Tumorzellen unterschieden werden.

3. Nutzungskontext verstehen

Um die Ziele des Auftraggebers und die Benutzerbedürfnisse zu verstehen, plante das Projektteam eine mehrwöchige Research Phase ein.

In Stakeholder Interviews mit dem Auftraggeber wurden die gemeinsamen Ziele definiert und der Projektrahmen festgelegt. Damit wurde sichergestellt, dass die gegenseitigen Erwartungen geklärt und die Basis für eine erfolgreiche Zusammenarbeit gelegt wurde. Zudem wurde ein früher Einblick in die Domäne der Radiologie gewährt. Anschliessend wurden bestehende Konkurrenzprodukte analysiert, um sich ein erstes Bild der Werkzeuge von Radiologen zu verschaffen.

Das Hauptziel dieser Phase war es, Radiologen bei der Arbeit zu beobachten und ihre Prozesse zu verstehen. Dafür wurden drei Contextual Inquiries in den Unispitälern in Basel und Zürich durchgeführt. Während den Beobachtungen konnten wertvolle Einblicke in die Aufgaben der Radiologen, sowie Erkenntnisse zur Arbeitsumgebung und den verwendeten Werkzeugen gesammelt werden, mit dem Ziel, diese später als Anforderungen zu dokumentieren.

Schliesslich wurde eine Analyse der gesammelten Research-Artefakte durchgeführt, um eine Basis für die ersten Skizzen zu legen. Mit der Durchführung einer Umfrage wurden die Resultate der Contextual Inquiries überprüft.

3.1 Stakeholder Interview

Methode

Das Stakeholder Interview dient dazu, die Rollen der Beteiligten, Risiken und Erfolgskriterien abzuholen und den Kontext der Arbeit zu verstehen. Zudem werden die Ziele und Bedürfnisse und die Art der Zusammenarbeit mit dem Auftraggeber geklärt.

Alternativen

Visioning-Workshop

Warum

Die Vision von 4Quant war in Vorgesprächen schon deutlich geworden und war in einem Businessplan beschrieben. Es bestanden jedoch noch Fragen zum Umfang, zur Domäne, zu den verschiedenen Beteiligten im Projekt und zur Zusammenarbeit, weshalb sich das Projektteam für mehrere Stakeholder Interviews entschied. Ein Workshop mit allen Beteiligten hätte die Koordination aufwändiger gemacht und den Zeitplan bereits nach hinten verschoben.

Durchführung

Das erste Interview fand mit dem Projektleiter in Zürich statt. Er agierte als Vermittler für die weitere Zusammenarbeit mit dem Team und den Testpersonen. Das zweite Interview fand mit den Stakeholdern für die medizinische- und technologische Sicht im Projekt gemeinsam in Basel statt. Dieses Interview war eine Art «Subject Matter Expert» Interview (Goodwin, K., 2009, S.79), in dem das Projektteam bereits vertiefen konnte, was es beim initialen Interview mit dem Projektleiter gelernt hatte.

Die Interviews wurden mit wenigen Fragen vorbereitet, sehr informell durchgeführt und mit Notizen und Fotos dokumentiert.

Ergebnis

Der Auftraggeber offerierte eine enge Zusammenarbeit mit dem Projektteam. Die gegenseitigen Erwartungen konnten geklärt, die Ziele geschärft und organisatorische Entscheide gefällt werden:

- Das Projektteam fokussierte sich auf den Arbeitsschritt des NSCLC Stagings.
- Das Re-Staging und der Befundungsbericht waren ausserhalb des Umfangs.
- Der Auftraggeber war für die Rekrutierung der Testpersonen zuständig.
- Das Projektteam arbeitete ausserhalb des Entwicklungsteams von 4Quant, um einen unabhängigen Zeitplan zu haben. So minimierte sich das Risiko, aufgrund externer Abhängigkeiten blockiert zu sein.
- Regelmässige Treffen wurden vereinbart, um Resultate zu präsentieren, zu diskutieren und zu validieren.
- Ein grober Zeitplan und mögliche Iterationen und Resultate wurden besprochen.
- Das Endresultat sollte als eigenständige Applikation funktionieren und nicht in eine bestehende Radiologie-Software-Suite integriert werden.

Im zweiten Interview fand ein vertiefter Wissenstransfer der medizinischen und technologischen Zusammenhänge statt, was dem Projektteam den Einstieg in das Projekt erleichterte. Folgende Punkte wurden beleuchtet:

- Die generelle Arbeitsweise und Organisation der Radiologen im Unispital Basel.
- Die Annotation und Aufbereitung der Testdaten durch die Radiologen, von denen die KI lernen sollte.
- Die Demonstration eines bestehenden Technologie-Prototypen. Hierbei wurde klar, dass ein Expert Review keinen Sinn machen würde, da praktisch keine Schnittstelle vorhanden war.
- Die Erklärung der Funktionsweise der KI und welche Werte sie ausgeben konnte.

Die Liste der Stakeholder befindet sich im Anhang.

Fachliche Erkenntnisse

Die Stakeholder Interviews waren eine ideale Voraussetzung für den Start des Projektes. Die persönlichen Kontakte zu den Stakeholdern waren im Verlauf des Projektes hilfreich und sorgten für ein angenehmes und kooperatives Klima.

3.2 Konkurrenzanalyse

Methode

Analyse bestehender Konkurrenzprodukte anhand verschiedener Faktoren.

Alternativen

SWOT-Analyse

Warum

Die Konkurrenzanalyse wurde durchgeführt, um ein besseres Bild über bestehende Lösungen von Mitbewerbern zu erhalten. Zweck der Analyse war es, Stärken und Schwächen der Lösungen zu identifizieren, um mögliche Verbesserungspotentiale zu erkennen und sich von der Konkurrenz abzuheben.

Durchführung

Bei der Durchführung war es hilfreich, dass der Auftraggeber bereits eine Konkurrenzanalyse zur Verfügung stellte. Das Projektteam ging die Produkte der Konkurrenten durch und analysierte deren Funktionalitäten, die Interaktionen und das Design, soweit dies anhand der Produktinformationen möglich war.

Ergebnis

Das Projektteam stützte sich auf die Ergebnisse des Auftraggebers und ging davon aus, dass die getroffene Auswahl umfassend genug war. Gemäss 4Quant gab es kein vergleichbares Produkt, welches das Staging vollautomatisiert durchführt. Jedoch waren einige Produkte auf dem Markt, die bereits teilautomatisierte Unterstützung anboten. In der Analyse zeichnete sich ab, dass immer mehr Software-Produkte automatisierte Features für computergestützte Diagnosen bereitstellten.

Bei der Betrachtung der UI war zudem auffällig, dass alle sehr kleine Schriften und UI Elemente verwendeten, sie eine grosse Menge an Daten zeigten und dass sie sehr dunkel gefärbt waren.

Die Konkurrenzanalyse befindet sich im Anhang.

Fachliche Erkenntnisse

Aufgrund des fehlenden Vorwissens war es schwierig, Kriterien zu definieren, aufgrund deren das Projektteam die verschiedenen Konkurrenzprodukte beurteilen konnte. Ausserdem war es schwierig, den Inhalt der Werkzeuge zu beurteilen, da dem Projektteam die Materie zu dem Zeitpunkt nicht sehr vertraut war. Das Projektteam war der Meinung, dass man Expertentools in der Tiefe anschauen und dafür auch das entsprechende Expertenwissen aufbauen muss. Eine Beurteilung anhand einer laufenden Software war nicht möglich, da diese nur an den Radiologie-Arbeitsplätzen vorhanden war. Folglich musste sich das Projektteam vorwiegend auf Screenshots und Produktbroschüren stützen. Es konnten nur Eindrücke gewonnen und keine Aussagen über Ergebnisse oder sogar Potenziale getroffen werden.

Im Nachhinein stellte das Projektteam fest, dass die durchgeführte Konkurrenzanalyse wohl eher der später durchgeführten Artefaktanalyse entsprach.

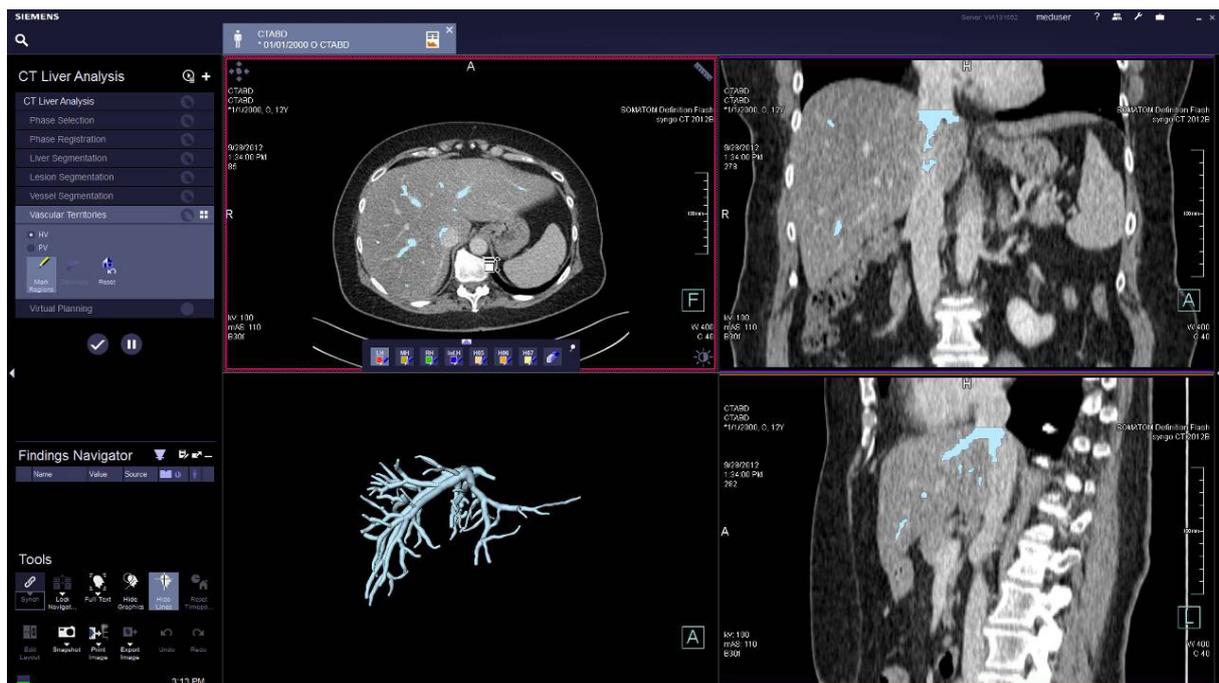


Abb. 5: Beispielscreen Siemens syngo.via

3.3 Contextual Inquiry

Methode

Contextual Inquiry (folgend «CI» genannt) ist eine Methode des Contextual Design Prozesses (Beyer, H., 1998). Durch die Beobachtung in der gewohnten Umgebung können Erkenntnisse zu Aufgabe, Benutzer und Umfeld gemacht werden. Die dabei gesammelten Daten werden interpretiert und die Erkenntnisse u.a. für Personas, Szenarien oder das Design angewendet.

Alternativen

Interviews, Beobachtung, Tagebuch

Warum

CI's eigneten sich, um zu verstehen, in welchem Kontext die Benutzer arbeiteten (Lichtverhältnisse, Unterbrechungen, Gruppendynamik) und welche Werkzeuge eingesetzt wurden. Das Meister-Lehrling-Verhältnis erlaubte es dem Projektteam, während der Arbeit bei Unklarheiten nachzufragen. Dies war aufgrund der schwierigen medizinischen Domäne unverzichtbar.

Durchführung

Bei der Auswahl der Interviewpartner achtete das Projektteam auf eine geeignete Bandbreite der Benutzergruppen, die sich nach Zielen, Verhalten und anderen Aspekten unterschieden. Darum wollte das Projektteam in drei Spitälern jeweils 1 bis 2 CI's machen. Auf Wunsch des Auftraggebers wurde die Rekrutierung durch ihn koordiniert. Die Kontakte der gewünschten Spitälern konnten jedoch nicht zeitnah hergestellt werden, da der Auftraggeber zu Beginn nur mit dem Unispital Basel zusammenarbeitete. Im ausgedehnten Verlauf der Research Phase konnte ein Kontakt zum Unispital Zürich hergestellt werden. Somit konnten zwei CI's in Basel sowie ein CI in Zürich durchgeführt werden.

Um die Benutzer in Ihrer gewohnten Umgebung zu beobachten und den Kontext zu verstehen, wurden die CI's in den jeweiligen Spitälern direkt am Arbeitsplatz geplant. Vom Projektteam waren immer zwei Mitglieder anwesend, wobei eine Person primär mit dem Arzt interagierte, während die andere Notizen machte und die Videoaufnahme steuerte.

Auch der Auftraggeber war vor Ort, um ebenfalls zu beobachten und die Zusammenarbeit mit dem jeweiligen Spital für spätere Iterationen vorzubereiten. Das CI wurde mit einem Leitfaden vorbereitet, welcher die Struktur der Untersuchung festlegte, die Fragestellung und Fokuspunkte definierte und Interviewfragen enthielt. Der detaillierte Leitfaden kann dem Anhang entnommen werden.

Zu Beginn wurde der Radiologe über das Vorhaben aufgeklärt und es gab eine kurze Einleitung mit wenigen Fragen zur Person und seiner Funktion. Danach konnte das Projektteam den Radiologen während der Befundung von Patientenbildern beobachten und immer wieder Rückfragen stellen. Diese Schritte wurden mit Video und Notizen dokumentiert. Abschliessend wurde das Gesehene diskutiert und offene Fragen geklärt.

Ergebnis

Das Projektteam sammelte viele Erkenntnisse zur Arbeit der Radiologen. Im folgenden Abschnitt folgt ein Auszug davon:

- Jeder Radiologe geht bei der Befundung anders vor, jedoch verwenden alle das international standardisierte «TNM» System zur Klassifizierung von Lungenkrebs.
- Ob ein Krebs vorhanden ist, schliesst der Radiologe aus der visuellen Anomalie beim Durchscrollen der Bilder. Diese wird als **Läsion** bezeichnet.
- Es gibt keine 100% Sicherheit für die Richtigkeit der Krankheitsbestimmung, es gibt nur eine Wahrscheinlichkeit. Die endgültige Bestimmung wird anhand einer Gewebeprobe gemacht. Für die Gewebeentnahme durch den Pathologen ist die Lokalisation entscheidend, die vom Radiologen geliefert wird.
- Radiologen verstehen sich als diagnostische Dienstleister für die behandelnden Ärzte.
- Die Befundung wird mit dem Verdacht auf Lungenkrebs durchgeführt. Das Bildmaterial wird aber immer auch auf Nebenbefunde untersucht (z.B. Osteoporose oder vergessenes Operationsmaterial).
- Es gibt viele Faktoren, die den Befund beeinflussen. Z.B. Unterbrüche durch Telefonate, Software die abstürzt, lange Ladezeiten des Bildmaterials, lange Arbeitszeiten, Konzentrationsfähigkeit vor der Mittagspause uvm.

Klassifikation für Tumore. T steht für Primärtumor, N für Metastasen in Lymphknoten und M für Fernmetastasen

Eine Schädigung, Verletzung oder Störung einer anatomischen Struktur oder physiologischen Funktion

Die komplette Liste befindet sich im elektronischen Anhang.



Abb. 6: Arbeitsumgebung der Radiologen im Unispital Basel

Fachliche Erkenntnisse

Während der CI's konnte das Projektteam viel medizinisches Wissen sammeln. In den ersten zwei CI's war das Projektteam ohne Beobachtungsfokus unterwegs und von der Informationsflut überwältigt. Daraufhin wurde der Beobachtungsfokus nochmals geschärft und verkleinert.

Das zweite CI konnte erst einen Monat später gemacht werden. Bei der Auswertung zeigte sich, dass diese nicht zeitnah genug war. Beim dritten CI wurde die Interpretation direkt nach der Beobachtung durchgeführt, was die Auswertung einfacher und schneller machte. Es wäre sinnvoll gewesen, die Subject Matter Interviews vor den CI's durchzuführen, um besser abschätzen zu können, was in den CI's auf das Projektteam zukommen könnte. Dies wurde jedoch erst nach dem ersten CI gemacht.

Das Projektteam hatte selber keinen Zugang zu den Radiologen. Auch der Versuch, selbständig via Email zu rekrutieren, war nicht erfolgreich. Die Anfragen des Projektteams blieben unbeantwortet. Drei CI's schienen auf den ersten Blick zu wenig. Der Auftraggeber war jedoch überzeugt, dass es keine grossen Unterschiede zwischen den Spitälern gäbe. Dies wurde auch von Radiologen in Basel, sowie in Zürich bestätigt. Ein Radiologe hatte zudem vorher in St. Gallen gearbeitet und konnte deshalb aus erster Hand Auskunft geben.

Die Rekrutierung neuer Testpersonen stellte sich weiterhin als schwierig heraus. Aus diesen Gründen entschied das Projektteam, keine weiteren CI's durchzuführen und mit der Gestaltung der Lösung zu beginnen. Aufgrund der kleinen Menge an CI's hatte das Projektteam das Bedürfnis, die Beobachtungen besser abzustützen. Darum beschloss es, die Erkenntnisse auszuwerten und Hypothesen zu formulieren, die im Verlauf der Arbeit mit weiteren Methoden verifiziert werden sollten (siehe Kapitel 3.5 «Umfrage»).

3.4 Affinity Diagramm

Methode

Affinity Diagramme eignen sich zum Aufbereiten von Erkenntnissen, Ideen, Aussagen oder Problemen, indem diese aufgrund ihrer Zusammengehörigkeit gruppiert und mit einem Titel versehen werden. Dadurch entsteht eine Übersicht, welche fassbar wird und weiterverarbeitet werden kann (Goodwin, K., 2009, S.215).

Alternativen

Gruppierung in Excel-Tabelle

Warum

In den CI's wurden viele Erkenntnisse gesammelt und in mehreren Dokumenten festgehalten. Die Projektgruppe benötigte einen systematischen Ansatz um die Erkenntnisse zu strukturieren und eine Übersicht zu schaffen, so dass später daraus Anforderungen an das UI abgeleitet werden konnten. Aufgrund der grossen Anzahl wäre eine Gruppierung innerhalb von Excel sehr mühsam geworden.

Durchführung

Die Erkenntnisse und Aussagen aus den CI's wurden in einer Tabelle zusammengeführt und mit einer ID referenziert. Die Erkenntnisse wurden auf Kärtchen gedruckt und in einem Workshop zusammen mit dem Auftraggeber nach Zusammengehörigkeit gruppiert. Dem Projektteam erschien es sinnvoll, den Auftraggeber zu integrieren, weil dieser ein vollständigeres Domänenwissen mitbrachte.

Nach dem Workshop war nicht klar, wie mit den Gruppen weitergearbeitet werden konnte. Deshalb wurde der Workshop ein zweites Mal ohne den Auftraggeber durchgeführt. Dabei wurden die Erkenntnisse, welche sich zum Teil überschneiden, auf neue Kärtchen zusammengefasst und nach Typ (Hypothese, Idee, Observation, Frage) eingeteilt. Anschliessend wurden neue Gruppen gebildet, welche eine überschaubare Grösse boten.

Ergebnis

Aus den beiden Workshops entstanden 9 Themengebiete, die anschliessend in Anforderungen überführt und priorisiert wurden (siehe Kapitel 4.1 «User Stories»). Die komplette Tabelle befindet sich im Anhang.

Die vielen Hypothesen und Ideen zeigten eine gewisse Unsicherheit auf, welche durch die wenigen CI's zu begründen sind. Da es schwierig war, weitere CI's zu organisieren, beschloss das Projektteam, die folgenden Hypothesen mit einer Umfrage weiter zu validieren.

ID	Hypothese
C2-1	Radiologen wünschen automatische Unterstützung.
C2-13	Die genaue Kontur einer Läsion ist nicht so relevant, die abstrakte Einzeichnung genügt.
C2-14	Anomalien werden einfacher erkannt, wenn man durch mehrere Bilder scrollen kann.
C2-16	Oben-nach-unten Ansicht (axial) ist am wichtigsten, bei Bedarf zieht man andere Ansichten bei.
C2-2	Verortung der Läsion kann automatisch gemacht werden.
C2-21	PET-Signal immer zeigen, wenn es vorhanden ist.
C2-24	Die wichtigsten Parameter sind Wahrscheinlichkeit, Durchmesser für T, SUV MAX und Invasion.

Standard Uptake Value.
 Normierter Wert, der den
 Zucker-Umsatz im Stoffwechsel
 beschreibt.

Fachliche Erkenntnisse

Durch die Breite und Vielfalt der Erkenntnisse war die Auswertung sehr zeitaufwändig. Dies hätte durch fokussiertere CI's eingegrenzt werden können, war jedoch aufgrund des komplexen Themengebietes schwierig umzusetzen.

Obwohl die Auswertung zeitaufwändig war, bewährte es sich, einen zweiten Workshop durchzuführen. Beim ersten Workshop wurden aus Zeitdruck zu schnell Gruppen gebildet und mit Titeln versehen. Ausserdem wurde das Vorgehen und Ziel des Workshops dem Auftraggeber nicht klar kommuniziert, wodurch dieser nur bedingt mithelfen konnte. Es gab auch sehr viele Kärtchen, was eine gute Übersicht erschwerte.

Insgesamt führte die Auswertung mittels Affinity Diagramm zu einem gemeinsamen Verständnis und einer guten Diskussion und bildete zugleich die Basis für die Modellierung der Personas.



Abb. 7: Clustering zusammen mit dem Auftraggeber

3.5 Umfrage

Methode

Umfragen mittels Online-Fragebögen sind eine Methode, um Antworten von einer grossen Anzahl Personen zu erhalten. Die Methode eignet sich, wenn offene Fragen aus der qualitativen Forschung bestehen. Die Qualität der Antworten hängt dabei vor allem von den gestellten Fragen, der Auswahl der befragten Personen, sowie der richtigen Auswertung der Daten ab (Richter, M., 2013, S.87).

Alternativen

Interview

Warum

Das Projektteam entschied sich für einen Online-Fragebogen, um mit wenig Zeitaufwand seitens der Radiologen die bisherigen Research-Ergebnisse zu verifizieren. Die Umfrage ist als qualitative – nicht quantitative – Methode zu verstehen.

Durchführung

Aufgrund der CI's hatte das Projektteam sieben Hypothesen formuliert, welche mit einer Online-Umfrage genauer untersucht werden sollten. Für die Erstellung der Umfrage und Auswertung wurde «Google Forms» verwendet.

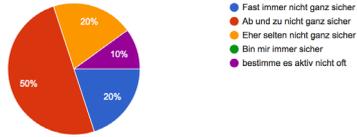
Für die Umfrage wurden sowohl offene, wie auch geschlossene Fragen verwendet. Bei den geschlossenen Fragen wurde eine Skala von 1–7 verwendet, wobei 1 «Stimme nicht zu», 7 «Stimme voll zu» bedeutete.

Die Fragen wurden von einer Person des Projektteams formuliert und von einer weiteren Person geprüft. Um sicherzustellen, dass die Fragen aus medizinischer Sicht verständlich waren, wurden diese zusätzlich mit einem Radiologen des Forschungsteams des Unispitals Basel besprochen.

Der Fragebogen wurde an Radiologen der Unispitäler Basel und Zürich versandt und von insgesamt acht Radiologen und zwei Nuklearmedizinern ausgefüllt.

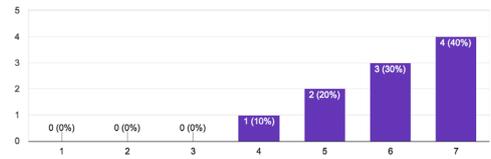
Wie oft sind sie sich nicht ganz sicher beim Bestimmen des TNM?

10 responses



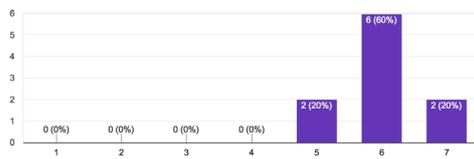
Die axiale Ebene ist die wichtigste PET/CT Bildansicht zur Erkennung von NSCLC

10 responses



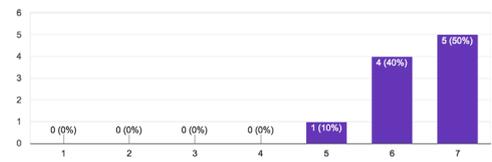
Eine Fehldiagnose ist die schwerwiegendste Belastung in meiner Arbeit

10 responses



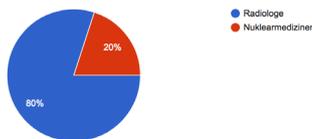
In der axialen Ebene erkenne ich Anomalien einfacher, wenn ich durch mehrere Slices hin- und her wechseln kann (z.B. durch scrollen).

10 responses



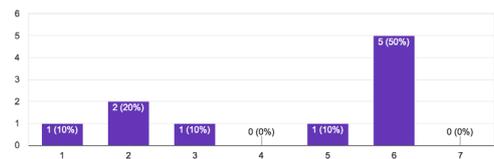
Ich arbeite als...

10 responses



Die genaue Kontur einer Läsion ist nicht so relevant, eine abstrakte Einzeichnung der Region of Interest genügt zur Erkennung im Bild.

10 responses



Ich wünsche mir automatische Unterstützung beim Staging von NSCLC

10 responses



Ein standardisierter Workflow zum Staging von NSCLC würde mir und meinem Team helfen und Objektivität und Transparenz steigern

10 responses

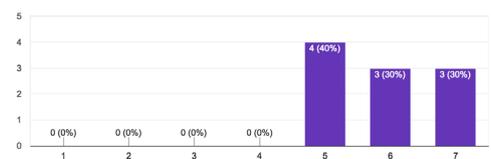


Abb. 8: Umfrageresultate in Google Forms

Ergebnis

Untenstehende Tabelle zeigt einen Auszug der Umfrageresultate.

Hypothese	Zustimmung
Radiologen wünschen automatisierte Unterstützung.	10 von 10
Die genaue Kontur einer Läsion ist nicht so sehr relevant, die abstrakte Einzeichnung genügt.	6 von 10
Anomalien werden einfacher erkannt, wenn man durch mehrere Bilder scrollen kann.	10 von 10
Oben-nach-unten Ansicht ist am wichtigsten, bei Bedarf zieht man andere Ansichten bei.	5 von 10
Falschdiagnosen sind die schwerste Belastung, nicht das Durchführen des Stagings an sich.	10 von 10
Ein standardisierter Workflow ist von den Ärzten gewünscht und hilft der Objektivität und Fehlervermeidung.	10 von 10
Default-Windows (Grauwerte) sind gut genug, manuelle Einstellung ist sekundär oder gar nicht notwendig. Ebenso beim PET-Signal.	8 von 10

Für das UI ergaben sich damit folgende Hauptkenntnisse:

- Die Radiologen wünschten eine automatische Unterstützung und einen standardisierten Workflow für das Staging.
- Die Bilder sollen scrollbar sein, weil durch die Bewegung durch die Bilder die Anomalie besser erkennbar wird.
- Die Ansichten sollen schnell gewechselt werden können, um Serien besser vergleichen zu können.

Die Umfrage und Resultate befinden sich im Anhang.

Fachliche Erkenntnisse

Qualitative Online-Umfragen lassen sich mit wenig Aufwand umsetzen und bieten mit wenig Antworten bereits einen Mehrwert. Sie haben den Vorteil, dass sie – wie bei diesem Projekt – parallel zur Projektarbeit laufen können. Dank der offenen Fragen konnten qualitative Erkenntnisse gewonnen werden, welche die CI's ergänzten.

3.6 Artefaktanalyse

Methode

In der Artefaktanalyse (Courage, C., 2005, S.623) ist der Fokus auf den Objekten, nicht auf den Benutzern. Durch systematisches Evaluieren der Artefakte wird versucht zu erkennen, wozu sie benutzt werden und welche Funktion sie haben.

Alternativen

Cognitive Walkthrough

Warum

Im Gegensatz zum Cognitive Walkthrough (Wharton, C., 1994) fokussiert sich die Artefaktanalyse auf die Objekte, ihre Inhalte und ihre Funktionen und nicht auf den Benutzer. Dies sollte beim Projektteam Wissenslücken bezüglich der in der Radiologie eingesetzten Werkzeuge aufdecken.

Durchführung

Während der initialen Recherche und der ersten CI's sammelte das Projektteam Fotos und Screenshots von Werkzeugen, die in der Radiologie eingesetzt wurden. Zusätzlich wurde der technische Prototyp von LungStage untersucht. Diese Artefakte wurden in einer Präsentation gesammelt und danach Stück für Stück analysiert. Dabei achtete das Projektteam darauf, Funktionen oder Werkzeuge zu identifizieren, von denen nicht klar war, wozu sie dienten und wie sie funktionierten.

Ergebnis

Die durch die Artefaktanalyse offengelegten Wissenslücken wurden dokumentiert, um im dritten CI diesbezüglich fokussiert nachfragen zu können. Durch die gesammelte Übersicht von Screenshots bestehender Werkzeuge erkannte das Projektteam, dass es Konventionen gab, die für die Lösung nützlich sein könnten. Dies betraf vor allem Abkürzungen, die unter Radiologen gängig sind, sowie Muster, wie die Bilddaten von den Radiologen angeschaut und organisiert wurden.

Fachliche Erkenntnisse

Die Artefaktanalyse war wertvoll, um Wissenslücken aufzuzeigen und um das Concept Model (siehe Kapitel 4.4 «Concept Model») zu erarbeiten.

3.7 Themenrecherche

Methode

Recherche zu einem spezifischen Thema mit einer klaren Zielsetzung. Als Quellen dienen verschiedene Kanäle, z.B. Internet, Fachliteratur, Fachpersonen/Experten, Videos. Die wichtigsten Ergebnisse wurden dokumentiert, damit diese in der Praxis angewandt werden konnten, sowie einen Nutzen für den Auftraggeber boten.

Alternativen

Internetrecherche

Warum

Im Gegensatz zu einer kurzen Internetrecherche entschied sich das Projektteam für eine ausgedehnte Recherche, da noch wenig Wissen in der Gestaltung von KI Systemen bestand. Das Projektteam wollte auf Thesen und Erkenntnisse vorhandener Literatur zurückgreifen um bei der Gestaltung auf wissenschaftlich fundierte Grundlagen zurückgreifen zu können.

Durchführung

Zu Beginn stand die Suche nach relevanten Artikeln, wissenschaftlichen Arbeiten und Publikationen von Usability Tests zum Thema «Vertrauen in KI Systeme». Die Quellen wurden in einer Tabelle (siehe Anhang) gesammelt und aufgrund von Nützlichkeit, Zweck und Zielgruppe bewertet. Relevante Informationen wurden in einem Dokument gegliedert und referenziert. In einem letzten Schritt wurde zusammengefasst, was in den Veröffentlichungen vorgefunden wurde und was sie zu dieser Masterarbeit beitragen konnten.

Ergebnis

Das Ergebnis aus der Themenrecherche ist ein mehrseitiger Bericht, welcher im Anhang verfügbar ist. Die Haupteckenkenntnisse um Vertrauen zu fördern waren:

- Besitzt ein Benutzer das Gefühl, Kontrolle über die Situation zu haben, ist es wahrscheinlicher, dass er den Berechnungen vertraut und das UI weiter nutzt.
- Menschen schätzen Transparenz.
- Eine gemeinsame Sprache fördert das Vertrauen.

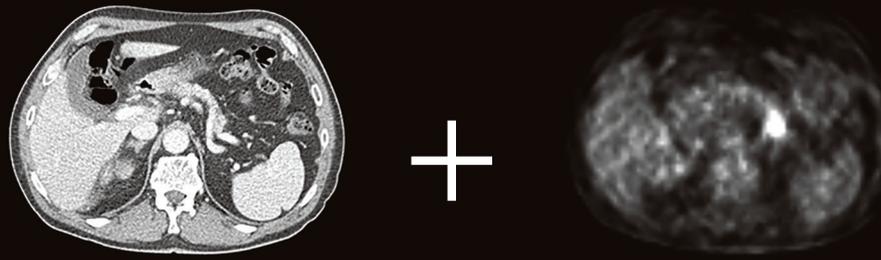
Bei der Recherche hatte sich gezeigt, dass bei der Verwendung von KI die gleichen Usability Regeln gelten wie bei allen Software-Applikationen. Die Technologie hinter einer Gestaltungslösung tritt in den Hintergrund. Was zählt ist der Benutzer und sein Verhalten.

Die Themenrecherche befindet sich im Anhang.

Fachliche Erkenntnisse

Durch die Erarbeitung der Themenrecherche gewann das Projektteam wertvolle Erkenntnisse zu der Interaktion mit KI Systemen. Sie sollten dem Projektteam vor allem zu Beginn des Gestaltungsprozesses helfen, informierte Design Entscheidungen zu treffen.

Das Ziel des theoretischen Teils war aber zu weit und ungenau definiert. Einerseits war der Anspruch, dass der Teil als separate Wissensquelle für andere Designer dienen sollte, andererseits sollte es eine Grundlage für das eigene Projekt sein. Es stellte sich heraus, dass dies eine Gratwanderung war zwischen einer wissenschaftlichen Arbeit und einem Arbeitsdokument. So war es kaum möglich, eine geeignete Flughöhe zu erlangen, die beiden Ansprüchen genügen würde.



CT Bild

PET Bild

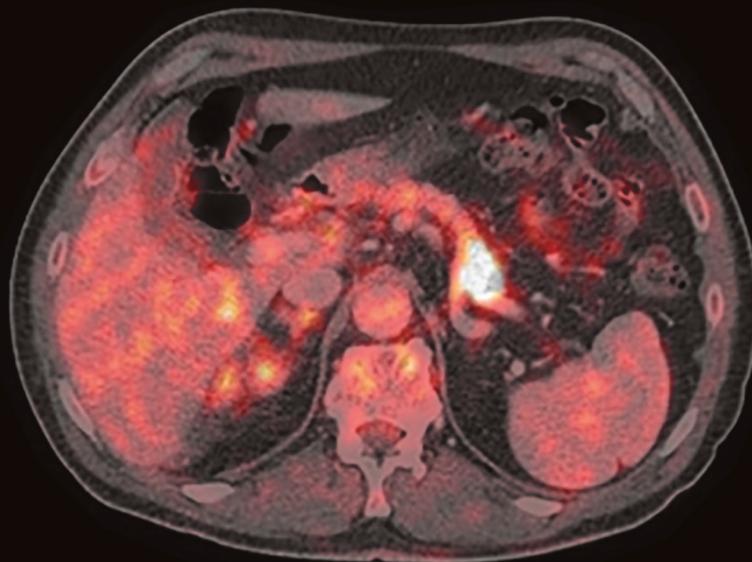


Abb. 9: Fusions-Bild

Fusions-Bild

Das fusionierte PET/CT-Bild zeigt die Überlagerung aus beiden Bildern, wobei das CT-Bild die Anatomie des Patienten und das PET-Bild den Zuckerstoffwechsel zeigt. Die PET-Information ist farbig kodiert (je wärmer die Farbe, um so höher der Stoffwechsel).

4. Nutzungsanforderungen spezifizieren

In diesem Teil des gewählten Vorgehensmodell (ISO 9241-210) beschäftigte sich das Projektteam mit der Spezifizierung der Nutzungsanforderungen mit dem Ziel, die gewonnenen Erkenntnisse in konkrete Anforderungen an das UI zu überführen.

Die gewonnenen Erkenntnisse aus den Untersuchungen wurden in eine Anforderungsliste überführt und priorisiert. Im nächsten Schritt wurden die Anforderungen als User Stories definiert, um eine einheitliche und lösungsneutrale Form der Anforderungen zu erhalten.

Für die Beschreibung der Benutzer und ihrer Tätigkeiten, wurden eine Persona und ein Kontext-Szenario einer typischen Befundung erstellt. Um die Zusammenhänge der verschiedenen Daten besser zu verstehen, wurde ein Concept Model erstellt. Alle Artefakte wurden mit dem Auftraggeber besprochen, um sein Feedback einfließen zu lassen.

4.1 User Stories

Methode

User Stories beschreiben in einer vorgegebenen Satzstruktur das Ziel eines Benutzers bei einer bestimmten Aktion innerhalb eines Systems: Als Rolle ... will ich ..., damit ... erreicht wird.

Alternativen

Use Cases, Jobs To Be Done

Warum

Die Form der User Stories erschien dem Projektteam optimaler als Use Cases, da sich aus der Satzstruktur das Ziel des Benutzers sofort ablesen liess und Use Cases vor allem zur Spezifikation von technischen Systemen verwendet werden (Richter, M., 2013, S.63). Das Projektteam entschied sich gegen Jobs To Be Done (Ulwick, A., 2016), da es schwierig war die angemessene Flughöhe zu finden.

Durchführung

Basierend auf den Erkenntnissen wurde eine Liste erstellt, mit dem Ziel, diese in fassbare Anforderungen zu überführen. In einem nächsten Schritt wurden daraus User Stories formuliert. Gemeinsam wurde für jede User Story die Quelle der Anforderung in einer Tabelle dokumentiert, um die Nachvollziehbarkeit zu gewährleisten. Jede User Story wurde vom Projektteam kategorisiert (funktional, nicht funktional oder out-of-scope) und mit einem Status (Vorschlag) versehen.

Weiter wurde für jede User Story das Projektrisiko (hoch, mittel oder tief) bewertet sowie die Projektpriorität (kritisch, wichtig oder nützlich) eingestuft. Anschliessend wurden diese mit dem Auftraggeber validiert und die Businesspriorität (kritisch, wichtig oder nützlich) bewertet. Anhand einer gewichteten Formel wurde die resultierende Priorität der User Stories bewertet:

- Projektrisiko: Faktor 1.0 * (3=hoch, 2=mittel, 1=tief)
- Businesspriorität: Faktor 1.0 * (3=kritisch, 2=wichtig, 1=nützlich)
- Projektpriorität: Faktor 1.5 * (3=kritisch, 2=wichtig, 1=nützlich)

In der Formel wurde der Faktor Projektpriorität mit einem Faktor 1.5 gewichtet. Damit erlaubte sich das Projektteam, die eigenen Interessen im Projekt etwas höher zu gewichten. Die Summe dieser drei ergab die Gesamtpriorität, anhand welcher das Projektteam den Fokus beim Design setzen wollte.

Ergebnis

Das Ergebnis war eine lösungsneutrale und mit dem Auftraggeber validierte Liste von User Stories. Folgende User Stories wurden mit der höchsten Priorität versehen:

ID	User Story
A10	Als Radiologe möchte ich Einblick in die automatische Bestimmung des TNM haben, um der KI vertrauen zu können und nicht alle Bilddaten selber anschauen zu müssen.
A13	Als Radiologe möchte ich ein objektives Staging, um individuellen Bias und negative Folgen für die Therapie zu vermeiden.
A14	Als Radiologe möchte ich bei der Diagnose unterstützt werden, um Fehldiagnosen zu vermeiden.
A27	Als Radiologe möchte ich mit der KI interagieren können, um besser zu verstehen, welchen Effekt meine Eingaben auf das Resultat haben.

Die komplette Liste der User Stories befindet sich im Anhang.

Fachliche Erkenntnisse

Die User Stories waren lösungsneutral formuliert, was dazu führte, dass sie sehr abstrakt wirkten. Für das Design wären mehr Kontext-Informationen, wie z.B. ausführliche Akzeptanzkriterien oder zusätzliche Beschreibungen, hilfreich gewesen.

In der ersten Design Iteration verwendete das Projektteam die User Stories, um einen Fokus zu setzen. So gesehen waren die User Stories hilfreich, jedoch der Aufwand zu hoch. Die User Stories wurden im Verlauf des Projektes zu wenig referenziert, um wirklich als Quelle für das Design nützlich zu sein. Dafür fehlten zu viele Detailinformationen.

4.2 Persona

Methode

Eine Persona ist ein Modell eines fiktiven Benutzers, in dem sein Verhalten, seine Bedürfnisse und seine Ziele abgebildet sind (Richter, M., 2013, S.39).

Alternativen

User Roles

Warum

Das Projektteam wählte Personas, da es im Gegensatz zu User Roles eine konkrete Person repräsentiert. Dies erleichterte es dem Projektteam die Bedürfnisse, Herausforderungen und Handlungen zu verstehen und auf diese Weise die Lösung besser zu gestalten. Der Auftraggeber hatte bereits eine Persona modelliert, die validiert werden sollte.

Durchführung

Der Auftraggeber konnte nicht belegen, wie seine Persona zustande gekommen war. Da sich die vorhandene Persona mit der Erfahrung aus der Research Phase deckte, entschied sich das Projektteam gegen eine systematische Neumodellierung. Die Persona wurde von einem Mitglied der Projektgruppe mit weiteren Details zu ihren Motivationen, Zielen und Hürden der Benutzer ergänzt, im Team besprochen und anschliessend dem Auftraggeber präsentiert.

Dr. med. Roman Stebler



Alter	36
Lebt in	Pratteln
Wohnverhältnis	Lebt in einer Mietwohnung mit seiner Frau und seiner 1 Jahr alten Tochter
Beruf	Doppelfacharzt für Radiologie und Nuklearmedizin
Technologie & Tools	Er ist technikaffin, sein Arbeitsplatz besteht aus 3 Screens. Nebst Applikationen für administrative Aufgaben ist er es gewohnt mit komplexer Expertensoftware zur Analyse und Auswertung von PET/CT Daten zu arbeiten und Krankheiten zu erkennen. Ein Diktiergerät dient zur Aufzeichnung von Befundungen. Das Festnetztelefon dient zur Kommunikation innerhalb des Spitals. Sein persönliches Smartphone verwendet er für die Kommunikation mit Freunden und Familie.
Interessen & Hobbys	Er fährt jeden Tag mit dem Velo 11 km zur Arbeit um fit zu bleiben. Er möchte gerne Zeit mit seiner Familie verbringen und deshalb nicht zu spät nach Hause kommen. Durch einen langen Arbeitstag ist dies jedoch nicht immer möglich. Generell ist die Work-Life-Balance ein schwieriges Thema. Die Arbeitstage sind lang und die Wochenenden würde er auch gerne für die Forschung verwenden, da dafür im Arbeitsalltag kaum Zeit bleibt.

Ergebnis

Aufgrund der CI's, Interviews und Gesprächen mit dem Stakeholder zeigte sich, dass es verschiedene Rollen (Radiologen, Nuklearmediziner) und Hierarchien (Assistenzarzt, Facharzt, Oberarzt) gab. Radiologen analysierten lediglich CT-Bilder, während Nuklearmediziner zusätzlich PET-Bilder sowie deren Fusion mit CT-Bildern analysierten. Sie hatten jedoch ein gemeinsames Ziel: die Befundung des Patienten hinsichtlich NSCLC. Zum Teil wurde dies auch organisatorisch in den Spitälern optimiert, d.h. die Ärzte hatten eine Doppelfacharzt-Ausbildung und konnten somit beide Rollen – Radiologe und Nuklearmediziner – wahrnehmen.

Innerhalb der Rollen und Hierarchien waren keine weiteren Ziele und Aufgaben auszumachen, welche eine sekundäre Persona gerechtfertigt hätten. Auf die Modellierung einer Non-Persona wurde verzichtet, da dies für die Arbeit keinen Mehrwert gehabt hätte, weil die Benutzergruppe bereits eng definiert war.

Die ausgearbeitete Persona befindet sich im Anhang.

Fachliche Erkenntnisse

Das Projektteam ist sich bewusst, dass die Persona aufgrund der geringen Anzahl beobachteter Benutzer kein starkes Fundament besitzt. Mehr Interviews hätten die Persona breiter abgestützt und ermöglicht, neue Verhaltensvariablen zu finden.

Das Risiko bestand, dass Annahmen enthalten waren und somit falsche Personas kreiert wurden. Das Risiko war aber aus Sicht des Projektteams gering. Da alle Benutzer die gleiche radiologische Ausbildung hatten und die Aufgabenstellung stark eingegrenzt war, beschränkte sich das Projektteam auf die Ausarbeitung der Persona mit den vorhandenen Daten.

4.3 Kontext-Szenarien

Methode

Kontext-Szenarien (Goodwin, K., 2009, S.309) erklären den Kontext, in dem sich die Benutzer befinden. Sie erzählen die Geschichte, wie die Benutzer zum Punkt kommen, an dem sie mit dem zu gestaltenden System interagieren und definieren, wie eine erfolgreiche Interaktion damit aussehen könnte.

Alternativen

Key-Path Szenarien, Testszenarien

Warum

Das Projektteam wählte die Methode, um die komplexen Sachverhalte im Alltag eines Radiologen in Form einer einfachen Geschichte greifbar zu machen. Key-Path Szenarien beschreiben dagegen bereits Schritte im UI, was zu Beginn des Projekts zu früh war. Die Testszenarien wurden später bei der Evaluation des Designs eingesetzt.

Durchführung

Nachdem die CI's durchgeführt worden waren, sammelte das Projektteam prägende Momente aus den Beobachtungen. Diese wurden dann in den typischen Alltag eines Radiologen integriert und als Geschichte modelliert.

Ergebnis

Das ausformulierte Kontext-Szenario befindet sich im Anhang.

Fachliche Erkenntnisse

Die Erstellung des IST-Szenarios half dem Projektteam, die einzelnen Schritte der Benutzer zu reflektieren und in der Diskussion ein gemeinsames Verständnis zu erlangen, wie die Befundung ablief. Das erstellte Szenario wurde sehr objektiv geschrieben. Um die Geschichte eindrücklicher zu schreiben, hätte das Projektteam mehr Stolpersteine und emotionale Momente einbauen können, wie z.B. die häufigen Unterbrechungen, die Müdigkeit am Abend und daraus resultierende Fehler bei der Befundung sowie weitere Informationen, die beim Gestalten der Lösung mehr Empathie für den Benutzer aufgebaut hätten.

4.4 Concept Model

Methode

Das Concept Model (Brown, D., 2009) zeigt Strukturen und Relationen von Objekten eines Systems oder einer Domäne.

Alternativen

User Environment Design

Warum

Das Concept Model sollte dem Projektteam eine Übersicht bieten, womit Radiologen täglich zu tun haben. Es interessierte sich primär für die Zusammenhänge zwischen den Objekten, um die Komplexität zu verstehen, und weniger um die Funktionen der Objekte, so wie sie im «User Environment Design» (Beyer, H., 1998) beschrieben werden.

Durchführung

Auf Papier notierte das Projektteam sein Verständnis der Informationen, so wie sie während den CI's und der Artefaktanalyse beobachtet wurden. Durch die gemeinsame Diskussion des Modells wurde es in mehreren Iterationen weiterentwickelt, bis ein gemeinsames Verständnis entstand. Im Anschluss wurde das Modell in eine Excel Liste übertragen und mit Eigenschaften der Objekte erweitert, welches dem Modell des «User Environment Design» nahe kam.

Ergebnis

Bei der Erstellung des Modells offenbarte sich, dass Unklarheit bestand, wie sich eine Bildserie definierte und welche Inhalte und Eigenschaften sie hatte. Ebenfalls war unklar, ob die Bestimmung des TNM pro Serie passierte oder ob sie serienübergreifend im Befund gemacht wurden. Diese Fragen konnten telefonisch mit einem Radiologen beantwortet und in einem nachgelagerten CI am Unispital Zürich überprüft und ergänzt werden.

Das erweiterte Modell befindet sich im Anhang.

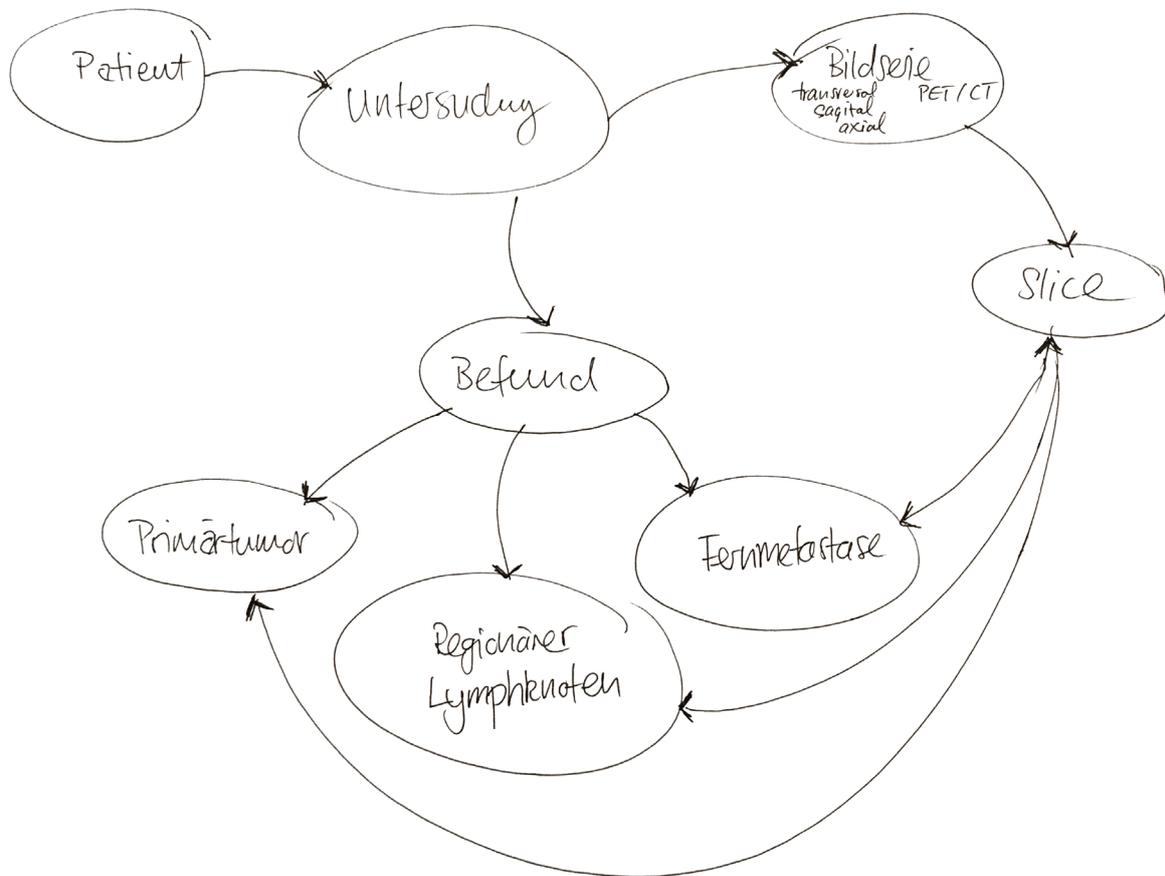


Abb. 11: Übersicht der Objekte

Fachliche Erkenntnisse

Das Concept Model gab dem Projektteam ein gemeinsames Verständnis des Systems und deckte ungeklärte Fragen auf. Es half in der Design Phase, das Layout zu strukturieren und dokumentierte, welche Detailinformationen und Eigenschaften dargestellt werden sollten.

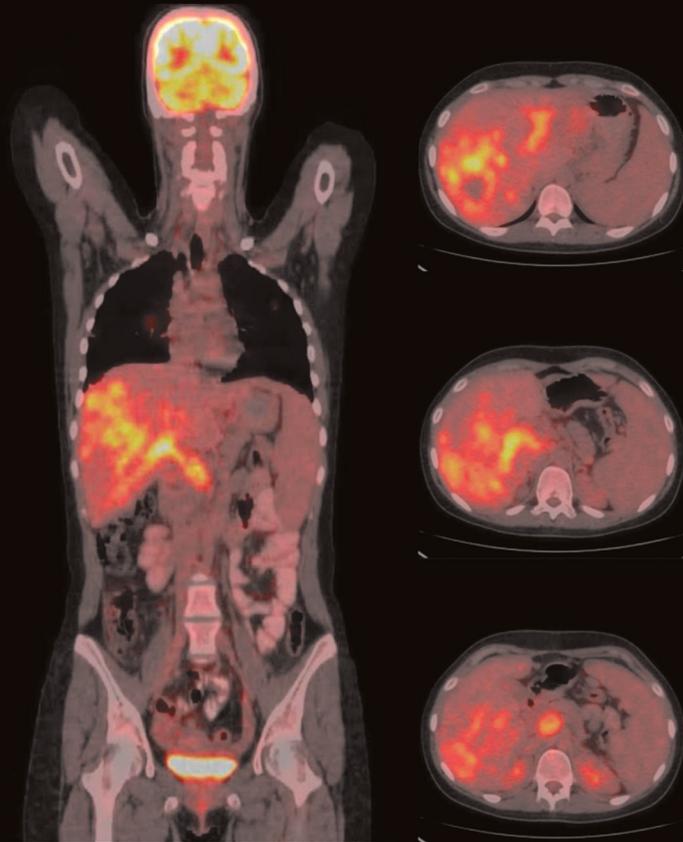


Abb. 12: Läsionen im rechten Lungenflügel

Läsion

In der Radiologie werden Bereiche mit veränderter Anatomie als Läsionen zusammengefasst. Dabei sind Entzündungen und Tumore nicht immer voneinander zu unterscheiden: Auch entzündetes Gewebe hat einen erhöhten Zuckerumsatz. Ein weiteres Problem: Manchmal haben Tumore keinen auffallenden Energieumsatz. Dort, wo weniger oder kein Stoffwechsel stattfindet, ist die Glukoseaufnahme entsprechend verringert. Dann sind weitere Untersuchungen notwendig, wie etwa eine Gewebeentnahme oder andere bildgebende Verfahren (dkfz, 2015).

5. Iteration 1

Ziel dieser Iteration war es, ausgehend von den Erkenntnissen der Research Phase, einen ersten Entwurf des Layouts und der Struktur von LungStage in Form eines Papier Prototypen zu gestalten und mit Benutzern der Zielgruppe zu validieren. Die Ergebnisse dienten als Basis für die nächsten Iterationen.

5.1 Papier Prototyp

Methode

Beim Papier Prototyping werden Ideen für das UI mit Papier und Stift visualisiert. Sie dienen in der frühen Projektphase dazu, möglichst schnell viele Lösungsansätze zu diskutieren und mit Benutzern zu validieren.

Alternativen

Wireframes

Warum

Die Methode eignete sich in der frühen Phase des Projektes, da der Detailgrad noch gering war und die Skizzen unverbindlicher wirkten als ausgearbeitete Wireframes. Sie vermittelten, dass sich das Projekt in einem frühen Stadium befand und Änderungen ohne grossen Aufwand möglich waren.

Durchführung

In einem ersten Schritt skizzierte jedes Projektmitglied während 30 Minuten mit Papier und Stift selbständig erste Lösungen. Die Papier Prototypen zeigten jeweils nur einen Screen, ohne Interaktionsmöglichkeiten. Die Varianten wurden im Team besprochen und in einer zweiten Runde verfeinert. Es liess sich jedoch kein Favorit bestimmen. Deshalb entschied das Projektteam alle Skizzen mithilfe eines Walkthroughs zu validieren.

Ergebnis

Das Ergebnis dieser Phase waren drei Design Varianten. Diese unterschieden sich in den unten genannten Merkmalen und bildeten die Grundlage für die Walkthroughs.

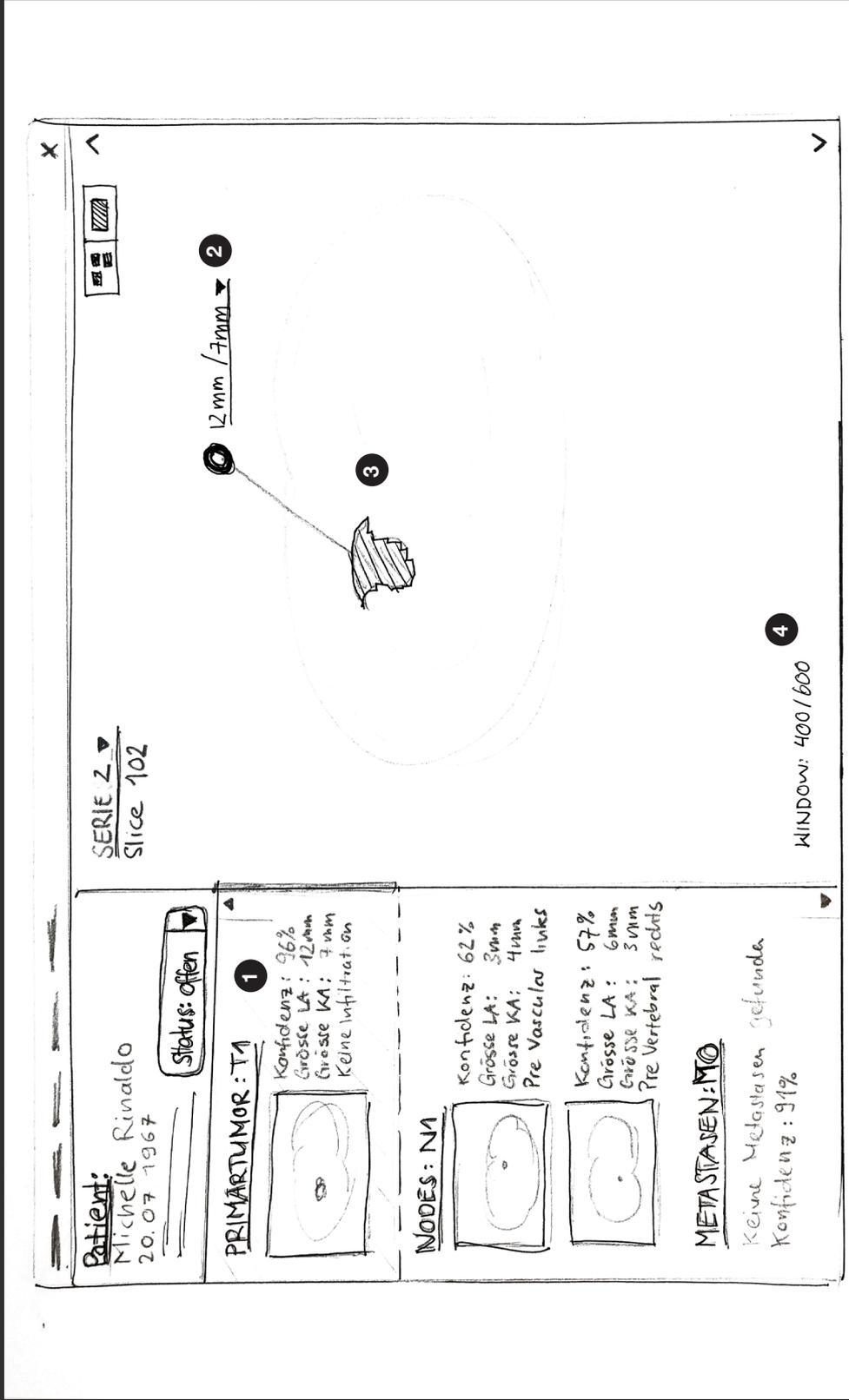


Abb. 13: Skizze Variante 1

- 1 Kompakte Darstellung des automatischen Befundes mit Vorschaubildern in der linken Spalte, Fokus auf wenige wichtige Attribute der Läsion.
- 2 Zusatzinfos können optional angezeigt werden.
- 3 Fokus auf axiale Ansicht.
- 4 Technische Angaben wie Windowing Level und der Serie innerhalb des Hauptfensters.

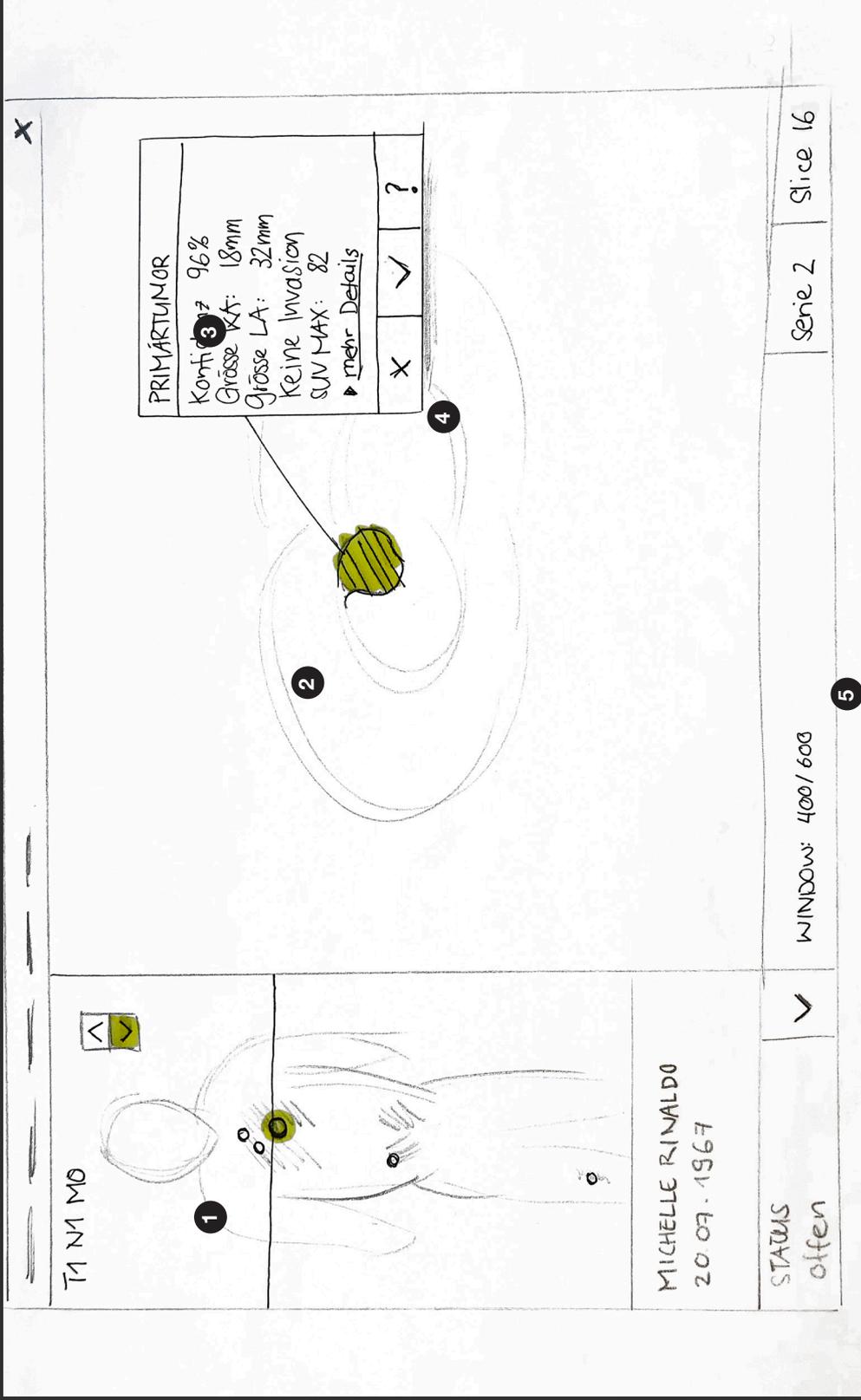


Abb. 14: Skizze Variante 2

- 1 Ganzkörper-Ansicht des Patienten als navigierbare Übersicht aller gefundenen Läsionen.
- 2 Fokus auf axiale Ansicht.
- 3 Detailangaben zur Läsion im Pop-up.
- 4 Der Radiologe hat die Möglichkeit, die gefundene Läsion als korrekt, inkorrekt oder unklar zu klassifizieren.
- 5 Technische Angaben wie Window und der Serie ausserhalb des Hauptfensters.

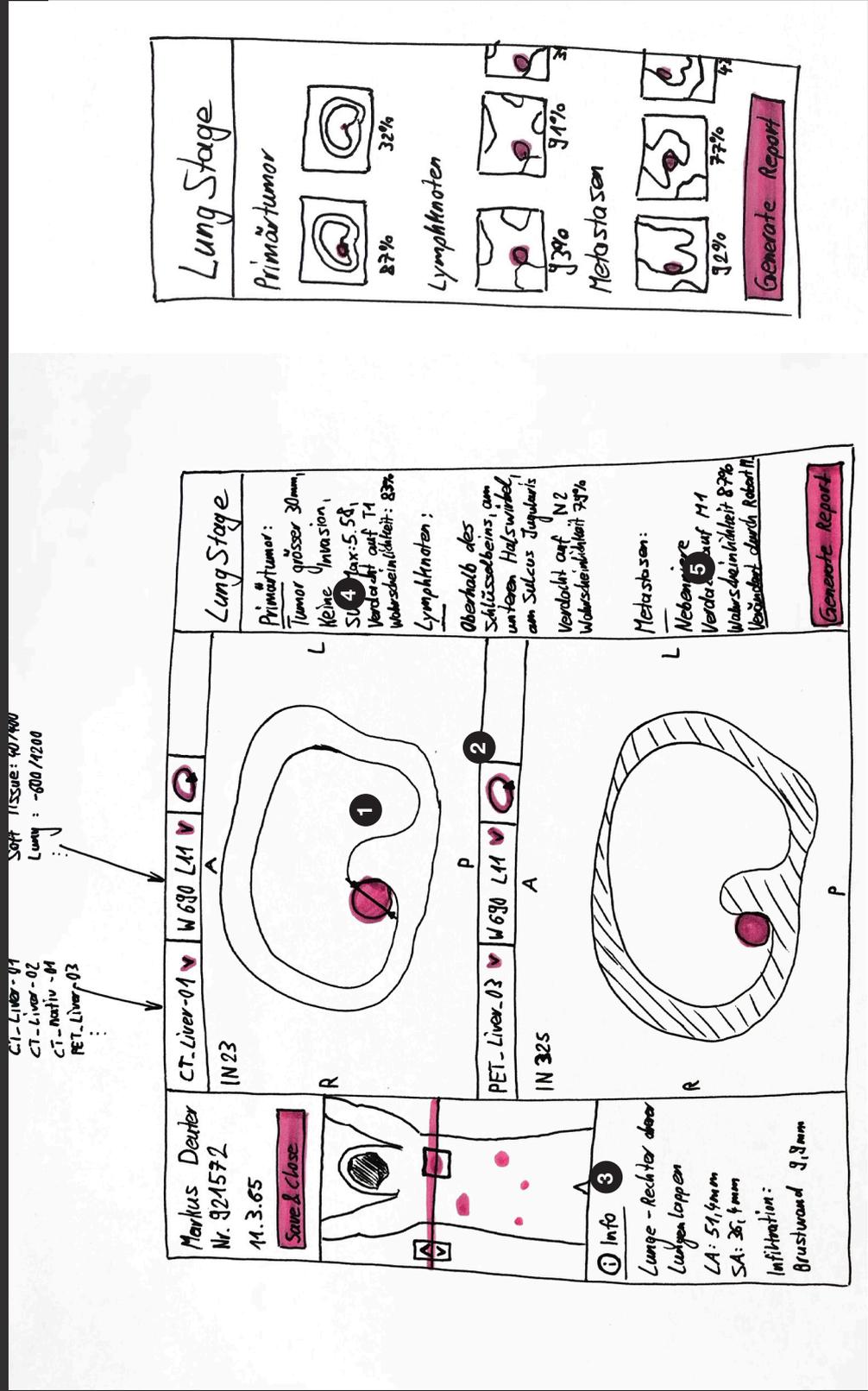


Abb. 15: Skizze Variante 3

- 1 Fokus auf Vergleich von Serien durch parallele Ansichten.
- 2 Einzeichnen von weiteren Läsionen mit «Lasso»-Werkzeug.
- 3 Detailinformation zur ausgewählten Läsion im «Info» Bereich in der linken Spalte.
- 4 Automatische Befundung in der rechten Spalte in Textform oder als Bilder mit Konfidenzangabe.
- 5 Automatische Generierung eines Befundes in Textform.

Fachliche Erkenntnisse

Durch das gegenseitige Vorstellen der Varianten konnten die Projektmitglieder neue Ideen und Elemente aufgreifen und weiterentwickeln. Dieser kollaborative Designprozess ermöglichte es innerhalb kurzer Zeit viele Versionen zu generieren.

Nebst dem effizienten Gestalten entstand der Mehrwert, dass die Skizzen als Diskussionsgrundlage dienten, um Fragen der einzelnen Projektmitglieder zu klären. Es half einen gemeinsamen Wissensstand zu erlangen und Fragen mit dem Stakeholder zu klären.

5.2 Usability Walkthrough mit Papier Prototyp

Methode

Beim Usability Walkthrough (nachfolgend «Walkthrough» genannt) handelt es sich um eine Evaluationsmethode, bei der die Benutzbarkeit eines Produktes ermittelt wird. Bei dieser Methode werden typische Handlungsabläufe mit einem Benutzer durchgespielt mit dem Ziel, Hürden und Probleme in der Interaktion und bei der Bedienung aufzudecken (Richter, M., 2013, S.85).

Alternativen

Cognitive Walkthrough

Warum

Im Gegensatz zum Cognitive Walkthrough (Wharton, C., 1994) wurde die Methode nicht mit Experten durchgeführt, sondern mit Benutzern. Trotz aufwändiger Rekrutierung entschied sich das Projektteam für diese Variante, da dadurch tatsächliche Anwender beobachtet und befragt werden konnten und das Projektteam nicht über ausreichend Expertenwissen verfügte. So konnte die Iteration mit fundierten, benutzerzentrierten Erkenntnissen abgeschlossen werden.

Durchführung

Das Projektteam wollte herausfinden, welche der drei Varianten am ehesten den Bedürfnissen der Benutzer entsprach. Zur Strukturierung des Interviews bereitete es einen Leitfaden vor. Die drei Papier Prototypen wurden mit zwei Personen aus dem Unispital Basel und einer Person aus dem Unispital Zürich vor Ort validiert. Vom Projektteam waren immer zwei Mitglieder anwesend, wobei eine Person primär mit dem Arzt interagierte, während die andere Notizen und Audioaufnahmen machte. Die Varianten wurden in abwechselnder Reihenfolge gezeigt, um keinen Reihenfolge-Effekt (Wikipedia, Fragekontexteffekt, 2017) zu generieren. Dabei wurden die Testpersonen gebeten, während der Durchführung zu kommentieren, was sie auf den Skizzen sahen und welches Verhalten sie erwartet hätten.

Bei der Evaluation wollte das Projektteam herausfinden, ob die Funktionalitäten und das Layout von den Benutzern verstanden wurden. Ein zusätzlicher Walkthrough mit einem Oberarzt des Unispitals Basel war vorbereitet, jedoch konnte dieser nicht durchgeführt werden, da er in der Klinik aufgeboden wurde.

Beschreibt den Effekt, bei der die Reihenfolge der gestellten Fragen Einfluss auf die Interpretation und die Bewertung der Fragen seitens des Befragten hat.

Ergebnis

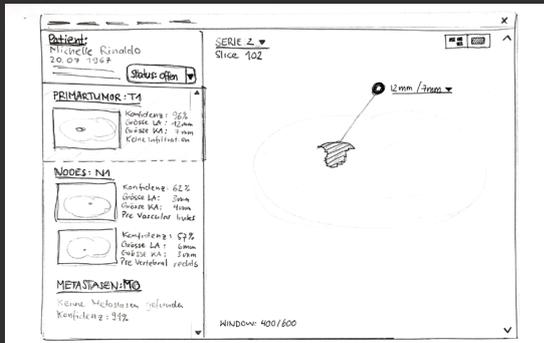


Abb. 17: Skizze Variante 1

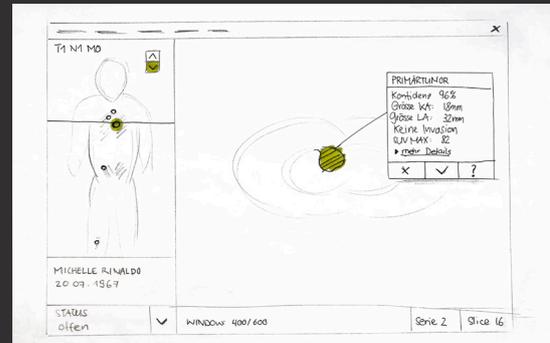


Abb. 18: Skizze Variante 2

Positiv

- Die Daten waren kompakt und strukturiert dargestellt.
- Die Angabe der Konfidenz war hilfreich.
- Der Fokus auf das PET/CT-Bild war positiv.
- Da nur die wichtigsten Infos auf dem Bild waren, wurden Benutzer nicht abgelenkt.
- Die Möglichkeit bestand zwischen Einzel- auf Mehrbild-Ansicht umzuschalten.
- Die korrekten und wichtigen Infos zur Läsion waren vorhanden.

Negativ

- Die Markierung sollte die Läsion nicht überdecken.
- Der Wechsel zwischen Einzel- und Mehrbild-Ansicht mit Button war ungewohnt (Umschalten mit Doppelklick erwartet).
- Die Lokalisierung der Läsion war unklar d.h. in welchem Lungensegment sie sich befand.
- Die Maximumintensitätsprojektion (nachfolgend MIP genannt) war für Radiologen unverständlich, sie hätten eine schematische Ansicht bevorzugt.
- Das TNM war nicht an einem zentralen Ort sichtbar.

Positiv

- Der Status im Popup neben der Läsion zum bestätigen/verwerfen war positiv.
- Es befanden sich keine unnötigen Informationen im Bild.
- Die navigierbare Körperübersicht und die Synchronisation zwischen der gewählten Läsion und dem Hauptbild wurde verstanden.
- Das TNM war klar ersichtlich.

Negativ

- Das Popup neben der Läsion verdeckte wichtige Bildbereiche.
- Es war nicht klar, wie das Popup angezeigt/versteckt werden konnte.
- Es war unklar, wie die Klassifizierung der erkannten Läsion geändert werden konnte.

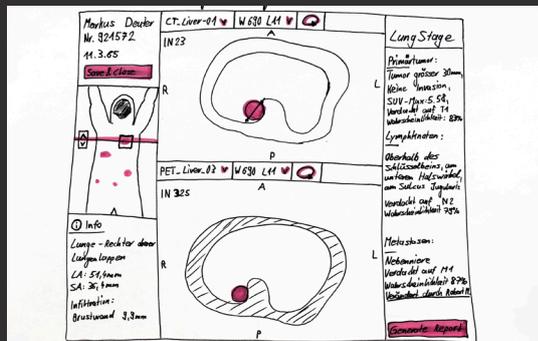


Abb. 19: Skizze Variante 3

Positiv

- Der Start mit zwei Serien gleichzeitig erlaubte den direkten Vergleich. Das Bild gross anzuzeigen mit einem Doppelklick war ein bekanntes Verhalten.
- Es befanden sich keine unnötigen Informationen im Bild.
- Die navigierbare Körperübersicht wurde verstanden.
- Die korrekten und wichtigen Infos zur Läsion waren vorhanden.

Negativ

- Die Symbolik war nicht ganz klar. Das «Lasso» Icon wurde als Ein- und Ausschalten der Synchronisierung zwischen den Serien verstanden.
- Die Informationen zu den Läsionen waren von der eingezeichneten Läsion örtlich getrennt.
- Befund in Textform wurde nicht als effizient oder effektiv beurteilt.

Weitere Ergebnisse

- Die Radiologen waren sich Interaktionsmöglichkeiten mit Maus und Tastatur von anderen Applikationen gewohnt.
- Die Radiologen würden die Applikation entweder als initiales Werkzeug, oder als Zweitmeinung verwenden. Während der Benutzung von LungStage sei ein Vertrauensaufbau zu erwarten (in ca. 6 Monaten). Wenn die Qualität der Befundung gut wäre, würde der Applikation evt. in Zukunft sogar ganz vertraut werden.
- Der Radiologe erkennt anhand des axialen Bildes sofort, wo er sich im Körper befindet.

Der Leitfaden und die vollständige Liste der Erkenntnisse befindet sich im Anhang.

Fachliche Erkenntnisse

Die drei Varianten zeigten unterschiedliche Ansätze, welche halfen, eine Diskussion mit den Radiologen zu führen. Jedoch zeichnete sich kein klarer Favorit ab. Ein möglicher Grund für diese Tatsache könnte die geringe Menge an Testpersonen sein. Trotzdem konnten aus allen Skizzen Erkenntnisse gewonnen werden, welche dem Projektteam halfen, weitere Entscheide zu treffen, bzw. eine verfeinerte und digitale Version zu gestalten. Bei den Walkthroughs wurde wieder deutlich, wie schwierig die Rekrutierung von Experten war.

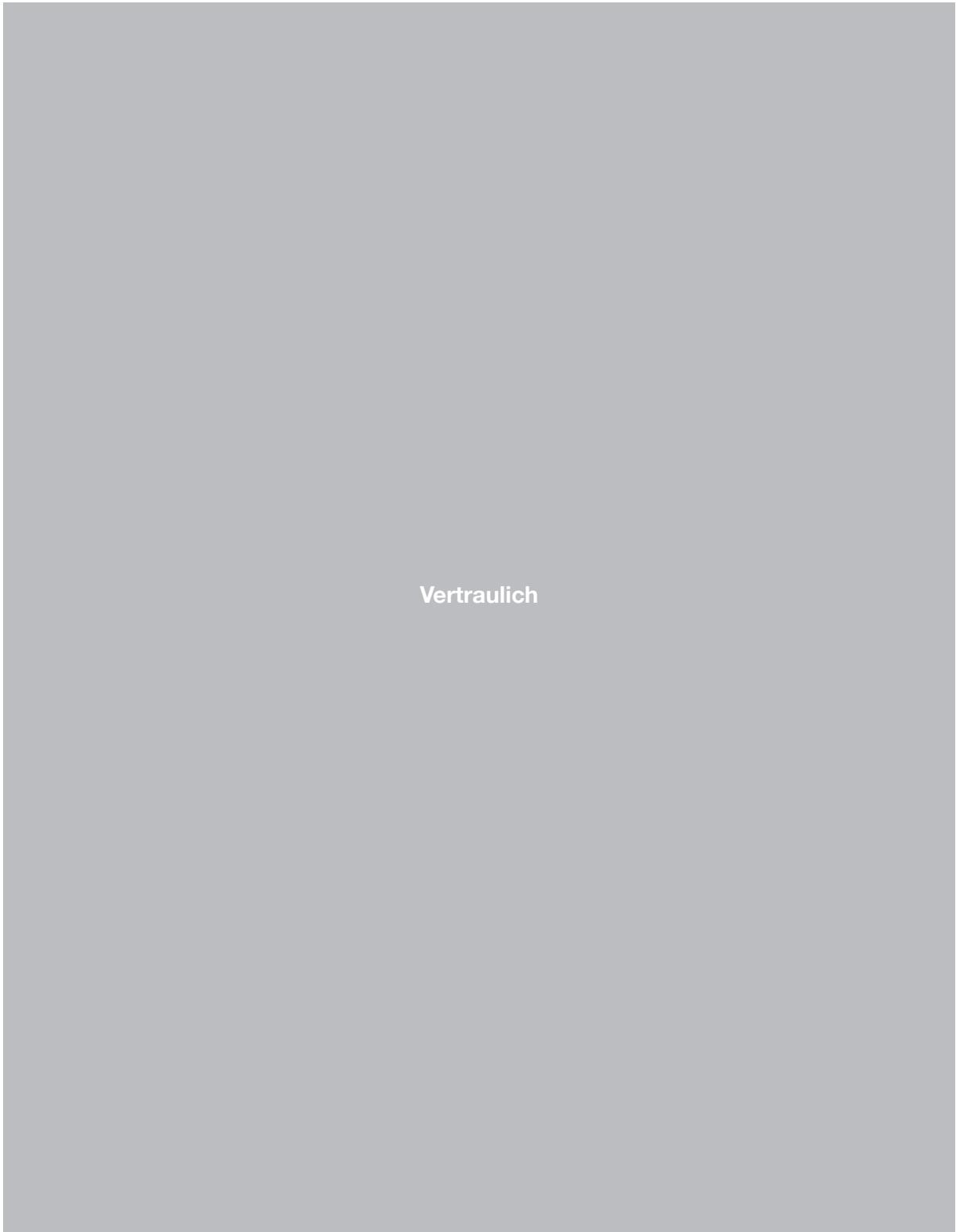


Abb. 20: Walkthrough am Unispital Basel

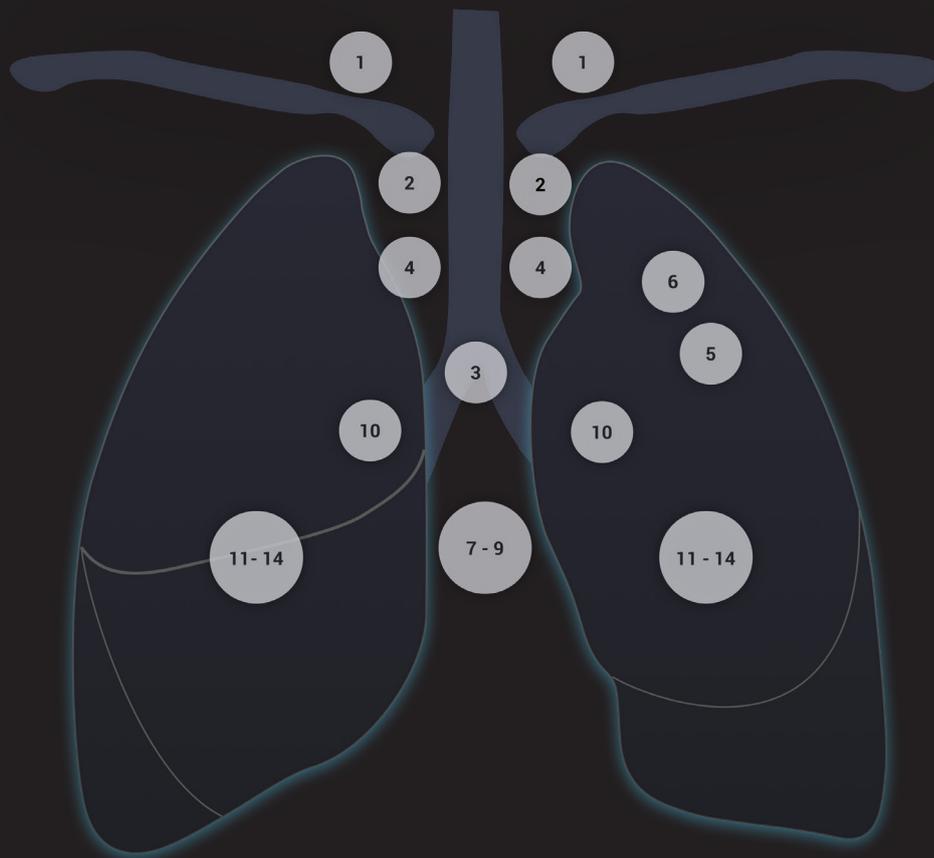


Abb. 21: Illustration aus LungStage Prototyp

Lymphstationen

Ein Bereich, in dem mehrere Lymphknoten zusammen liegen, wird als Lymphknotenstation bezeichnet. In der Lunge werden insgesamt 14 verschiedene Lymphknotenstationen definiert

6. Iteration 2

In der ersten Iteration wurden mehrere Lösungsvarianten in Form von Skizzen ausgearbeitet, um genauer zu verstehen, welche Ansätze sich aus der Nutzerforschung umsetzen bzw. kombinieren lassen.

Das Ziel der zweiten Iteration war, die gewonnenen Erkenntnisse zu konsolidieren, mit realen Daten anzureichern und in klickbare Wireframes zu überführen. Für die Validierung des Prototypen wurde wieder ein Usability Walkthrough gewählt.

Die folgenden Fragen aus der ersten Iteration wollte das Projektteam primär klären:

- Wie und in welcher Form sollten die Zusatzinformationen zu einer Läsion angezeigt werden?
- Wie soll die navigierbare Körperübersicht dargestellt werden?
- Wie funktioniert die Interaktion, um eine Läsion zu bestätigen oder zu verwerfen?
- Braucht es unstrukturierte Informationen, die der Radiologe einer Läsion hinzufügen kann (z.B. über ein Kommentarfeld)?
- Wie kann die Nachvollziehbarkeit der Änderungen von Radiologen gewährleistet werden?

6.1 Klickbare Wireframes

Methode

Wireframes sind reduzierte visuelle Darstellungen eines UI. Sie fokussieren auf die Struktur und Hierarchie der Informationen. Aspekte des visuellen Designs (Farben, Typographie, Branding, etc.) werden bewusst weglassen. Wireframes eignen sich sehr gut, um zu verstehen, ob alle Informationen und Bedienelemente vorhanden und die Abläufe und Interaktionen benutzbar sind.

Alternativen

Papier Prototyp, Mockups

Warum

Der Wechsel auf ein digitales Medium ermöglichte es, reale Daten und Bildmaterial in den Prototypen zu integrieren und die Oberfläche detaillierter und interaktiver zu gestalten. Dies hatte den Vorteil, dass sich der Prototyp realer anfühlte und neue Erkenntnisse aus den Tests gewonnen werden konnten. Mit einem Papier Prototyp wäre dies nicht möglich gewesen.

Mockups zeigen nur einzelne Aspekte des UI (z.B. einen Teil-Ablauf, eine Funktion oder eine Stilrichtung für das visuelle Design) und eignen sich deshalb vor allem, um Potenzial zu erkennen.

Durchführung

Die priorisierten User Stories halfen zu entscheiden, welche Ansätze für die Wireframes übernommen werden sollten. Als Basis für die Wireframes wurde die Variante 1 gewählt, da sie viele positive Eigenschaften der anderen Entwürfe bereits enthielt.

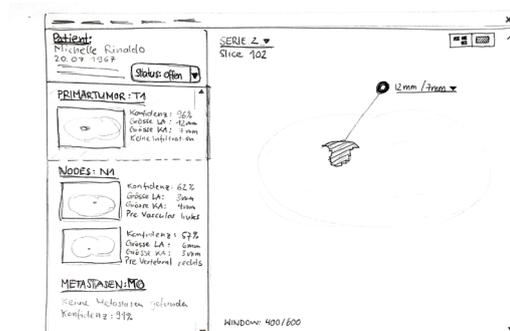


Abb. 22: Skizze Variante 1

Als Leitfaden für den Prototypen definierte das Projektteam drei Testszenarien:

- Überprüfen des automatischen Befundes (Überblick verschaffen)
- Korrektur einer Läsion
- Einzeichnen einer neuen Läsion

Vorgängig wurden im Team Grundsatzfragen bezüglich Funktionsumfang, Darstellung, Interaktivität und Informationsgehalt definiert (Richter, M., 2013, S.53). Ein technischer Prototyp wäre ideal gewesen, um das Verhalten der KI auf die Interaktion der Benutzer im Hinblick auf das Vertrauen zu testen. Das Entwicklungsteam des Auftraggebers hatte bereits einen technischen Prototypen entwickelt, jedoch war es aus zeitlichen Gründen nicht realistisch, diesen so anzupassen, dass damit getestet werden konnte. Um trotzdem medizinisch korrekte Daten zu integrieren, wurde der Auftraggeber für Text- und Bildmaterial eines realen Falls angefragt.

Die Wireframes wurden mit einer hohen Auflösung erstellt, um auf den Monitoren der Radiologen vor Ort möglichst realitätsnah gezeigt werden zu können. Der visuelle Stil wurde simplistisch, mit wenigen Farben und Comic Sans als Schrift, umgesetzt. Da die bestehenden Tools eine dunkle Bedienoberfläche boten und ein helles UI in den dunklen Befundungsräumen zu stark leuchten würde, entschied das Projektteam, dies in die Wireframes einzubauen. Es beschloss ausserdem, die Detail-Interaktionen noch nicht zu gestalten und stattdessen auf die Abläufe zu fokussieren. Für die Umsetzung wurde das Prototypen-Tool «Invision» gewählt.

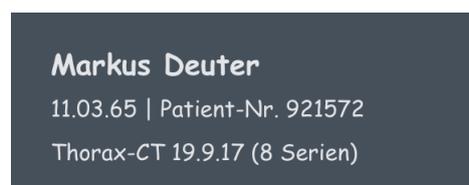


Abb. 23: Die Verwendung von Comic Sans für die Wireframes

Für die Umsetzung gestaltete jedes Projektmitglied eine eigene Variante im Grafikprogramm «Sketch». Nach zwei Stunden wurden diese besprochen und Vor- und Nachteile abgewogen. Das Projektteam einigte sich auf die grundlegenden Elemente und die Schlüsselseiten. Diese wurden von einer Person soweit ausgearbeitet, dass die Testszenarien durchgespielt werden konnten. Nach dem ersten Walkthrough hatte sich jedoch gezeigt, dass Details aus medizinischer Sicht unklar oder sogar falsch dargestellt waren. Da es sich um offensichtliche Fehler handelte, wären dieselben Probleme bei den weiteren Walkthroughs aufgetreten. Deshalb nahm das Projektteam die folgenden Änderungen nach dem ersten Test vor:

- Die medizinischen Daten des Patienten wurden korrigiert und es wurde ein Editiermodus für die Zusatzdaten eingefügt.
- Die genaue Kontur der Läsion wurde ersetzt mit Lang- und Kurzachse, welche die Ausmasse der Läsion zeigten, sowie einem Rahmen um die Läsion für die schnelle Erkennung.
- Die schematische Darstellung wurde überarbeitet, indem Lymphregionen geclustert und die Lymphstationen mit Zahlen referenziert wurden, welche den Radiologen von der Ausbildung her bekannt waren. Ausserdem wurden weitere Körperteile (z.B. Niere) ergänzt, welche ebenfalls von Metastasen betroffen sein können. Die schematische Darstellung wurde wie folgt angepasst:

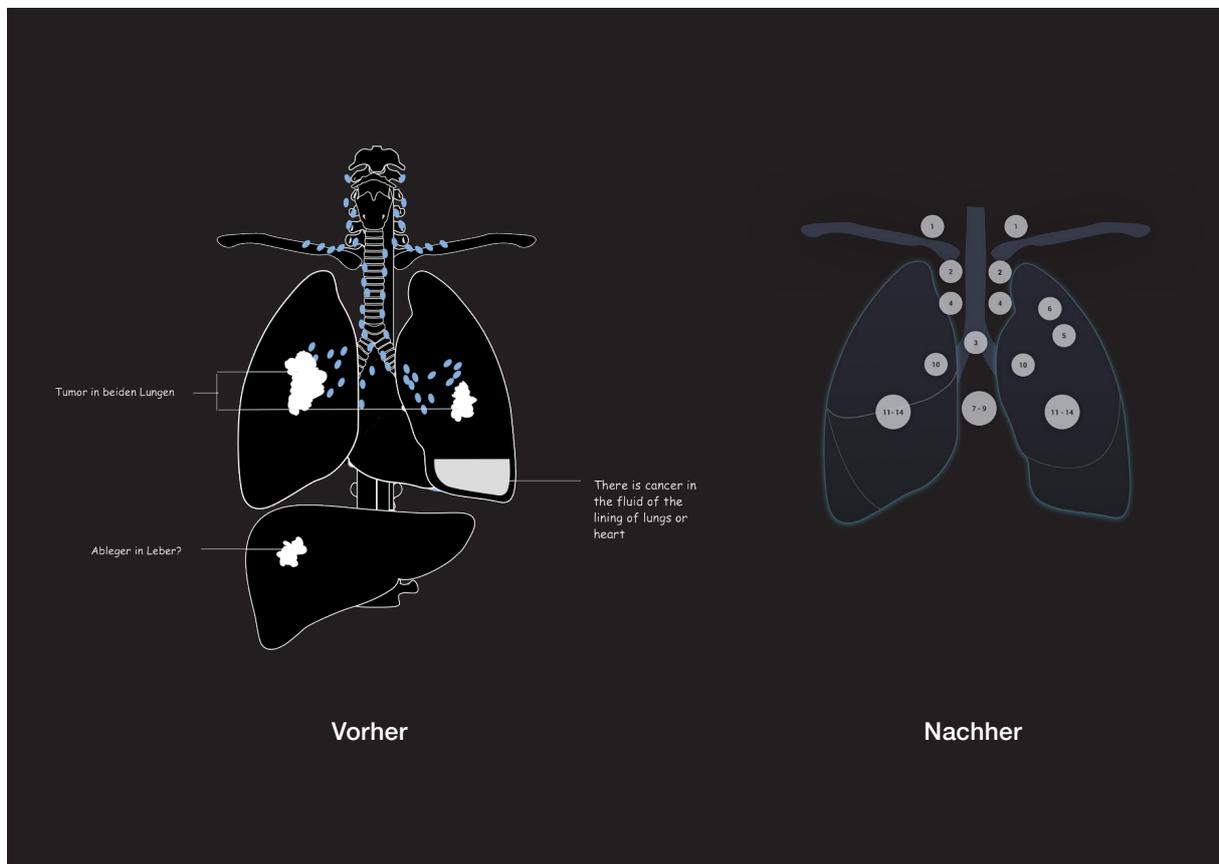


Abb. 24: Anpassungen der schematischen Darstellung

Ergebnis

Das Projektteam entschied sich für ein flexibles Interaktionsmodell, welches dem Benutzer zwei Möglichkeiten gab, die gefundenen Läsionen zu sichten. Er konnte in der linken Spalte strukturiert die Läsionen durchgehen, oder unabhängig davon durch die Bilder scrollen und wie bisher anatomisch in den Bilddaten navigieren.

Läsionen liessen sich optional verifizieren und bearbeiten. Das Bearbeiten sowie das Einzeichnen von neuen Läsionen wurde via Kontextmenü implementiert, da sich die Benutzer diese Art der Interaktion von anderen Applikationen gewohnt waren. Zusatzinformationen zur Läsion wurden in die rechte Seitenleiste integriert. Die Seitenleiste wurde initial nicht angezeigt, liess sich jedoch mit einem Button öffnen. Editierte Läsionen wurden in der linken Spalte markiert, um Änderungen durch den Radiologen sichtbar zu machen.

Im Theorieteil wurde herausgefunden, dass das Vertrauen in das Resultat von KI Systemen dadurch erhöht werden kann, indem die Benutzer mit dem Resultat interagieren können (siehe Themenrecherche). Aus diesem Grund führte das Projektteam ein neues Element ein, um die Sensitivität der Algorithmik zu beeinflussen. Der Algorithmus gewichtete die gefundenen Läsionen und zeigte nur diejenigen an, welche einen gewissen Schwellenwert überschritten. Durch die Sensitivität konnte der Benutzer diesen Schwellenwert manipulieren und dadurch erkennen, wie sich das Resultat der KI veränderte.

Das Ergebnis war ein klickbarer Prototyp, mit welchem sich die definierten Test-szenarien durchspielen liessen. Der Prototyp bestand primär aus den folgenden drei Ansichten:

Datei Bearbeiten Ansicht Hilfe

Markus Deuter
11.03.65 | Patient-Nr. 921572
Thorax-CT 19.9.17 (8 Serien)

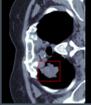
Sensitivität Tief Mittel Hoch

Übersicht



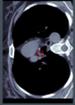
Stadium IIA
T2 N2 M0
Konfidenz 96%

Primärtumor (1)



Kategorie T2
Lokalisation Visc. Pleura
LA 46.87mm KA 42.97mm
Infiltration in viscerele Pleura

Lymphknoten (2)

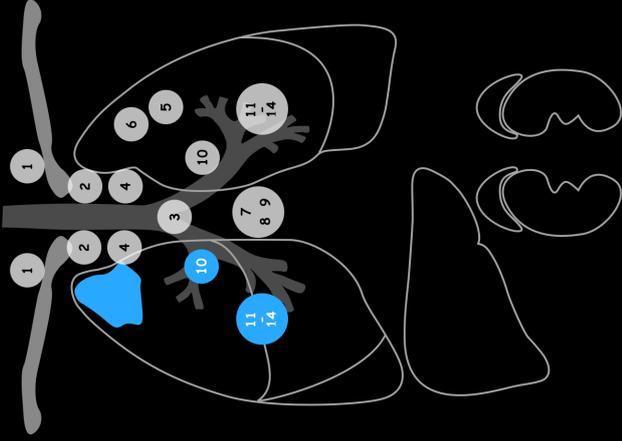


Kategorie N2
Lokalisation Ipsilateral superior mediastinal node S4
LA 16.35mm KA 11.72mm

Kategorie N1
Lokalisation Ipsilateral hilar node
LA 14.65mm KA 9.81mm

Metastasen (0)

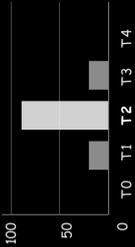
1



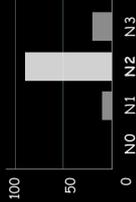
2

Konfidenz

T2



N2



M0

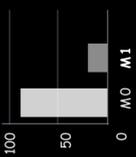


Abb. 25: Wireframe von der Übersicht der Befundung

- 1** Schematische Darstellung der Befundung
- 2** Konfidenz von LungStage für das TNM
- 3** Menüleiste um z.B. einen neuen Patienten zu laden
- 4** Patienteninformationen inklusive Angabe zum Scan
- 5** Sensitivitäts-Einstellung
- 6** Liste der Läsionen inklusive Informationen zu Kategorie, Ort und Grösse

Datei
Bearbeiten
Ansicht
Hilfe

Markus Deuter
 11.03.65 | Patient-Nr. 921572
 Thorax-CT 19.9.17 (8 Serien)

Sensitivität

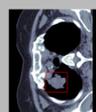
Tief
Mittel
Hoch

Übersicht



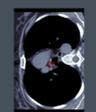
Stadium IIA
 T2 N2 M0
 Konfidenz 96%

Primärtumor (1)



Kategorie T2
 Lokalisation Visc Pleura
 LA 46.87mm KA 42.97mm

Lymphknoten (2)



Kategorie N2
 Lokalisation Ipsilateral superior mediastinal node 54
 LA 16.35mm KA 11.72mm

Metastasen (0)



Kategorie N1
 Lokalisation Ipsilateral hilar node
 LA 14.65mm KA 9.81mm

Details

Verifiziert

Alle korrekt bewertet

Ihr Kommentar...

5

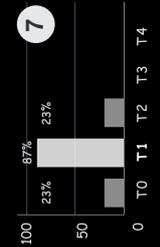
Zusatzinfos

Kategorie T2
 Lokalisation Infiltration in viscerale Pleura
 LA 46.87mm KA 42.97mm
 SUV-Max 2.5
 HU-Mean 38.4
 HU-Abweichung 14

6

Bearbeiten

Konfidenz von LungStage

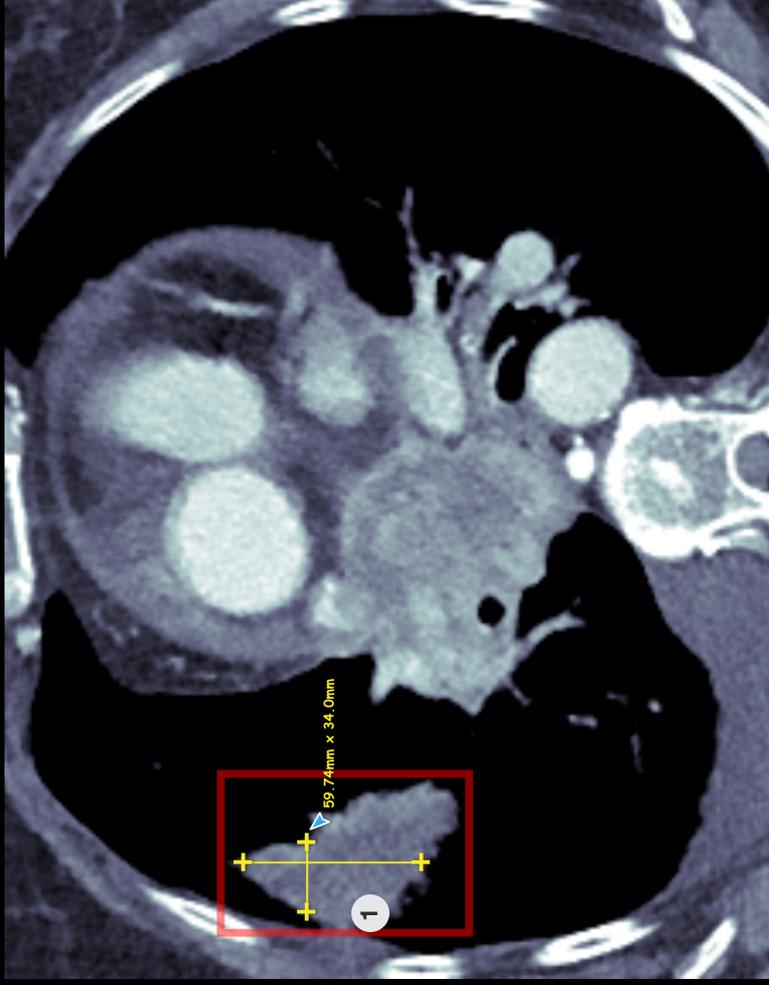


Kategorie	Konfidenz
T0	23%
T1	87%
T2	23%
T3	0%
T4	0%

7

Abb. 26: Wireframe von der Detailsansicht einer Läsion

- 1** Axiale Ansicht der CT-Serie
- 2** Koronale Ansicht der CT-Serie
- 3** Sagittale Ansicht der CT-Serie
- 4** PET/CT-Fusion (Übereinanderlegung der beiden Serien)
- 5** Option zur Verifikation der Läsion inklusive Kommentarfeld
- 6** Zusatzinfos wie z.B. SUV- und HU-Wert
- 7** Konfidenz von LungStage für diese Läsion



0003_ct_soft_ax1a1

Abb. 27: WWireframe des Editiermodus

- 1** Direktes Bearbeiten, bzw. Einzeichnen der Läsion inklusive Angabe zu Längs- und Querachse
- 2** Bearbeiten der Zusatzinformationen wie z.B. Art, Kategorie und Eigenschaften inklusive Kommentarfeld

Informationen

Art der Läsion

Primärtumor

2

Kategorie

T1

Lokalisation

Rechten Lappen/Mitte

Eigenschaften

- Hauptbronchus
- Invasion Viszerale Pleura
- Invasion der Brustwand
- Invasion Mediastinum
- Lappenobstruktion
- Nodulus im gleichen Lappen
- Tumor im anderen Lappen

Ihre Bewertung

Beschreiben Sie Ihr Staging...

Sichern

Verwerfen

Durch die Diskussion mit einer Testperson des Unispitals Basel entstand ein alternatives Layout, welches die PET/CT-Bilder im Hauptfenster zeigte und die Übersicht als Navigation in der linken Seitenspalte. Diese Ansicht hatte den Vorteil, dass die relevanten Bilder für das Staging direkt in der Übersicht angeschaut und bei Bedarf zu der Detailansicht des TNM gewechselt werden konnte. Es hatte jedoch den Nachteil, dass auf der Detailansicht die Liste der gefundenen «Regions of Interest» nicht mehr sichtbar war und dadurch die Navigation erschwert wurde.

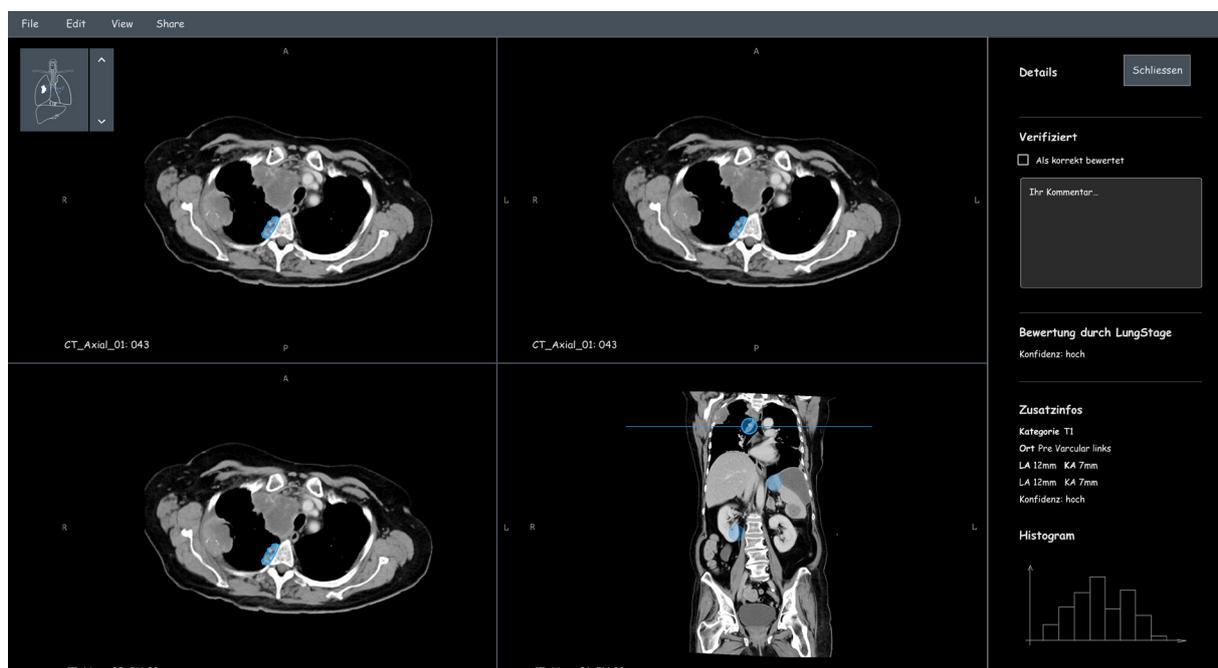


Abb. 28: Alternatives Layout - Detailansicht

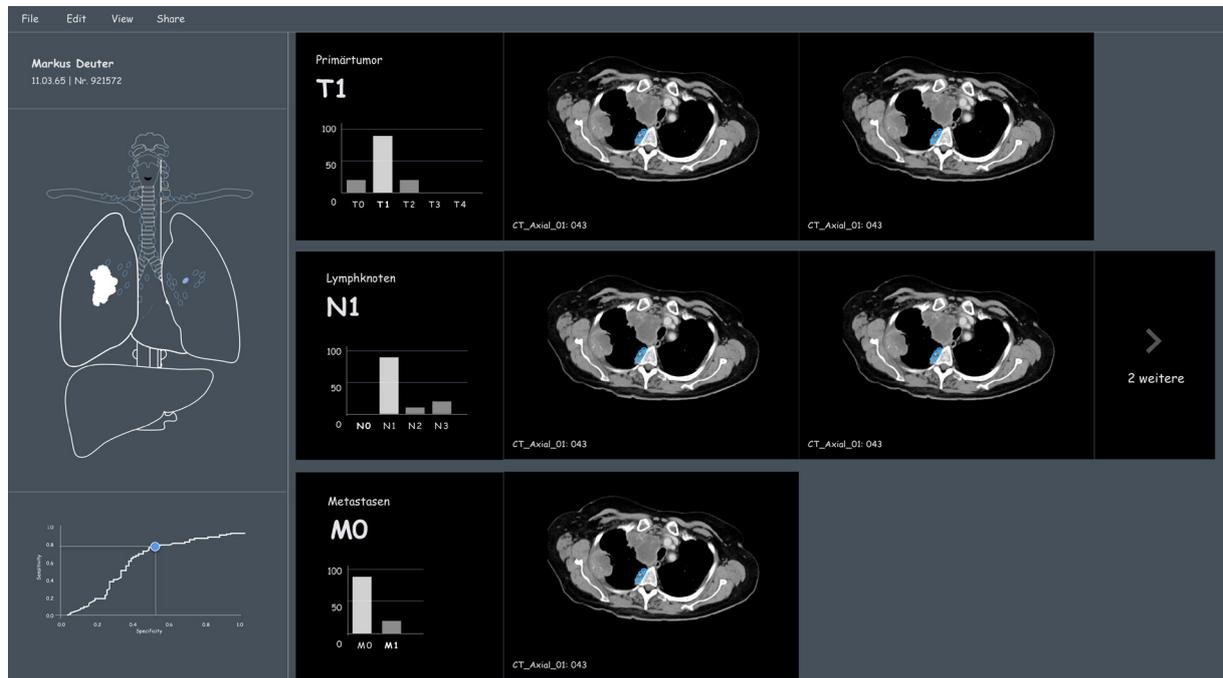


Abb. 29: Alternatives Layout - Übersicht

Fachliche Erkenntnisse

Die Diskussionen innerhalb des Projektteams halfen, schnell eine gemeinsame Richtung zu finden. Dadurch, dass jeder zuerst eine eigene Variante gezeichnet hatte, konnten Unklarheiten geklärt, ein gemeinsames Verständnis aufgebaut und der Prototyp effizient erstellt werden.

Die Änderungen, welche beim ersten Walkthrough offensichtlich wurden, zeigten, dass es besser gewesen wäre, den Prototypen vor den Walkthroughs mit einem Experten zu besprechen und auf fachliche Fehler zu prüfen.

6.2 Usability Walkthrough mit Wireframes

Methode

Siehe Kapitel 5.2 «Usability Walkthrough mit Papier Prototyp»

Alternativen

Usability-Test

Warum

Das Projektteam entschied sich für einen Walkthrough, da es offene Fragen zum Konzept gab und der Prototyp nur bedingt interaktiv war. Bei einem klassischen Usability-Test bekommen die Testpersonen Aufgaben gestellt, welche Sie ohne die Hilfe des Moderators durchführen sollen. Dies wäre mit dem aktuellen Stand des Prototypen schwierig gewesen. Walkthroughs bieten den Vorteil, dass während dem Test Fragen gestellt und dadurch Erkenntnisse zu spezifischen Themen gewonnen werden können. Zudem konnten die Testpersonen bei den Walkthroughs in ihrer üblichen Arbeitsumgebung beobachtet werden.

Durchführung

Es wurden Walkthroughs in Basel, Luzern und Zürich mit jeweils zwei Testpersonen durchgeführt. Der Testaufbau war analog zum ersten Usability Walkthrough. Für die Durchführung der Walkthroughs erstellte das Projektteam einen Testleitfaden (ersichtlich im Anhang). Der klickbare Prototyp wurde auf den Computern der Radiologen geöffnet, um eine möglichst realitätsnahe Testumgebung zu erreichen. Zu Beginn wurden Fragen zur Person gestellt, gefolgt von drei Aufgaben:

- Eine Übersicht verschaffen und die sichtbaren Informationen beurteilen
- Eine Läsion editieren
- Eine neue Läsion hinzufügen

Am Ende jedes Tests füllten die Kandidaten einen System Usability Scale (nachfolgend «SUS» genannt) Fragebogen aus, um die subjektiv wahrgenommene Gebrauchstauglichkeit des Systems zu bewerten. Die Auswertung des SUS befindet sich im Anhang.

Es waren jeweils zwei bis drei Projektmitglieder bei den Tests anwesend, welche sich in den Rollen des Moderators und des Notizen-Schreibers abwechselten. Die Tests wurden auf Video aufgezeichnet, um später auftretende Unklarheiten nachvollziehen zu können. Die erkannten Probleme aus Walkthroughs wurden direkt nach den Tests in eine Tabelle überführt und nach Kategorie (Strukturell, Funktional, Visuell) eingeteilt. Danach definierte das Projektteam die Massnahmen, welche in der nächsten Design Iteration umgesetzt werden sollten.

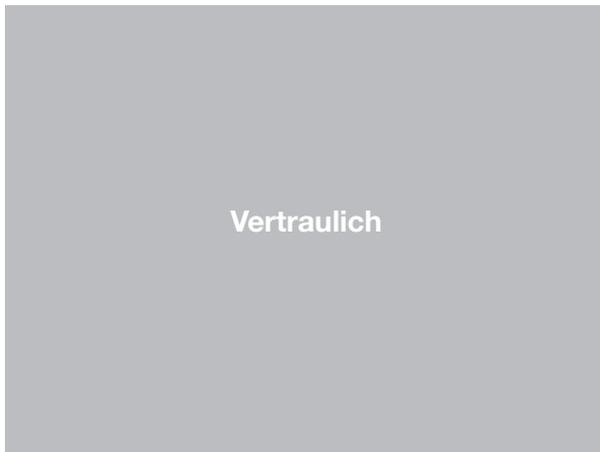


Abb. 30: Walkthrough mit einer Testperson



Abb. 31: Besprechung mit dem Entwicklungsteam

Ergebnis

Sensitivitätsregler

Die Testpersonen sagten aus, dass es interessant sei, mit dem Sensitivitätsregler zu spielen, um den Einfluss auf das Resultat zu sehen. Bei der weiteren Befragung stellte sich jedoch heraus, dass die Bedeutung der Sensitivität für die Benutzer nicht klar und im klinischen Alltag nicht erwünscht war. Aus diesen Gründen wurde der Sensitivitätsregler in der nächsten Iteration weggelassen.

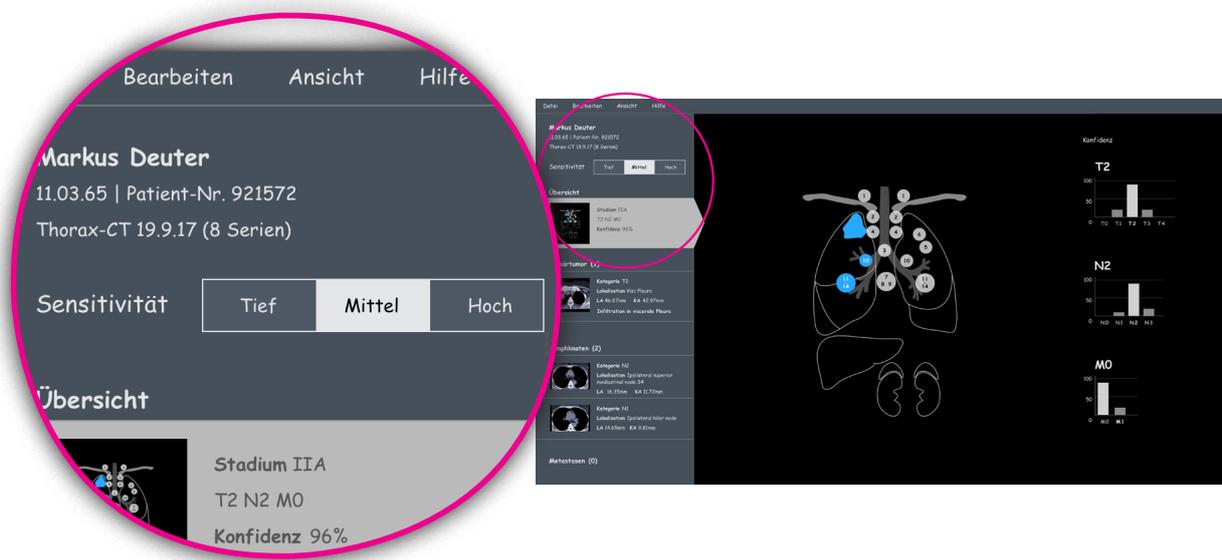


Abb. 32: Sensitivitätsregler

Verifikations-Funktion

Bei den Walkthroughs zeigte sich, dass es effizienter war, wenn man die erkannten Läsionen als korrekt betrachtete, solange diese nicht editiert wurden. Die Testpersonen hatten ausgesagt, dass sie es vorziehen würden, die gesamte Befundung beim Abschliessen des Befundes zu verifizieren.

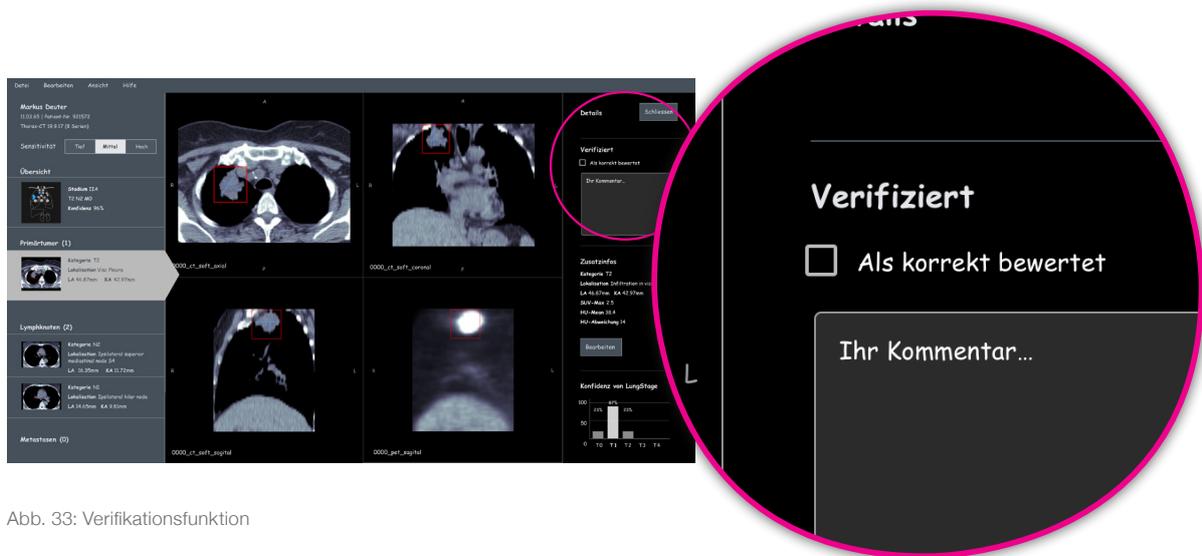


Abb. 33: Verifikationsfunktion

Weitere Ergebnisse

Das Projektteam beobachtete, dass das Bearbeiten, Erstellen und Löschen von Läsionen über das Kontextmenü nicht offensichtlich war. Der Bearbeitenmodus in der Seitenleiste funktionierte zwar gut, aber beim Beobachten der Testpersonen zeigte sich, dass dieser direkter gestaltet werden könnte.

Das alternative Layout wurde von den Testpersonen schlechter bewertet, als das ursprüngliche Layout. Damit wurde die Meinung innerhalb des Projektteams bekräftigt, das ursprüngliche Layout beizubehalten. In der Evaluation ergaben die Messungen mit dem SUS Fragebogen eine durchschnittliche Gebrauchstauglichkeit von 72%, was das Projektteam als eine weitere Indikation für die Akzeptanz des Systems interpretierte.

Die umfassenden Ergebnisse, die Resultate des SUS-Fragebogens, sowie der Testleitfaden befinden sich im Anhang.

Fachliche Erkenntnisse

Es bewährte sich, die Erkenntnisse aus den Walkthroughs am gleichen Tag niederzuschreiben, da sie so noch präsent waren. Somit konnte das Projektteam eine Erkenntnis aus der Research Phase bereits umsetzen.

Beim Einsatz des SUS-Fragebogens zeigte sich, dass die Testpersonen klar instruiert werden mussten, was sie bewerten sollten. So bewertete eine Testperson den effektiven Prototypen, welcher in dieser Form untauglich für den klinischen Einsatz war. Die anderen Testpersonen bewerteten die Vorstellung, wie der umgesetzte Prototyp funktionieren müsste.

T0

keine Anzeichen eines Primärtumors



T1

Tumorausdehnung höchstens 2 cm



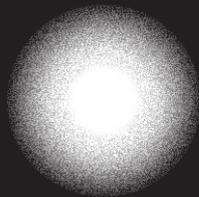
T2

Tumorausdehnung mehr als 2 cm,
aber höchstens 5 cm



T3

Tumorausdehnung mehr als 5 cm



T4

Tumor jeder Grösse mit direkter Ausdehnung auf die Brustwand oder Haut

Abb. 34: TNM-Klassifikation

TNM-Klassifikation

Die TNM-Klassifikation dient in der Medizin zur Bestimmung von bösartigen Tumoren in Stadien. Die drei wichtigsten Kategorien des TNM-Systems entsprechen den drei Buchstaben: T (Ausdehnung und Verhalten des Primärtumors), N (Fehlen bzw. Vorhandensein von Lymphknotenmetastasen), M (Fehlen bzw. Vorhandensein von Fernmetastasen) (Wikipedia, TNM-Klassifikation, 2017).

7. Iteration 3

In der dritten Iteration fokussierte das Projektteam auf das visuelle Erscheinungsbild und die Detail-Interaktionen von LungStage. Das Ziel war, einen Prototypen zu gestalten, der als Basis für die Implementierung genutzt werden konnte.

Das Projektteam entschied zusammen mit dem Auftraggeber, einen Hi-Fi Prototypen zu erstellen, mit welchem sowohl die visuelle Design-Richtung aufgezeigt, wie auch die wichtigsten Erkenntnisse aus der zweiten Iteration verfeinert werden sollten.

Dadurch sollte ein Prototyp entstehen, der auch als Kommunikationsmittel eingesetzt werden konnte, um neue Spitäler für Projektpartnerschaften anzusprechen und somit neue Testpersonen für die weitere Evaluation zu finden.

Aus der zweiten Iteration ergaben sich folgende Überlegungen:

- Das Gesamtergebnis der KI sollte bestätigt werden, das Validieren und Kommentieren von einzelnen Läsionen war für den klinischen Alltag zu aufwändig.
- Das Bearbeiten von Läsionen und ihren Daten sollte direkter gemacht werden.
- Es sollten reale Patientendaten und Ergebnisse der KI gezeigt werden.
- Der Sensitivitätsregler sollte entfernt werden.

7.1 Hi-Fi Prototyp

Methode

Hi-Fi Prototypen zeichnen sich dadurch aus, dass sie mit einem hohen visuellen Detailgrad gestaltet werden, welcher dem finalen Design am ehesten entspricht. Das visuelle Design definiert dabei die Anwendung und das Zusammenspiel von Farben, Typografie und Formen für das UI.

Alternativen

Mockups auf Basis von bestehenden Design-Systemen, wie z.B. Google Material Design (Google LLC, Material Design, 2017).

Warum

Der Auftraggeber verfügte über keine Richtlinien hinsichtlich der Marke und der visuellen Gestaltung und zeigte deshalb Interesse an der Entwicklung einer individuellen Design-Sprache durch das Projektteam. Er wollte sein Produkt gegenüber bestehenden Applikationen, die in der Radiologie im Einsatz waren, abheben. Somit entschied sich das Projektteam auch gegen bestehende Design Systeme.

Ein wichtiger Aspekt für diese Arbeit ist das Vertrauen in die automatischen Resultate von Lung-Stage. Das UI soll dieses Ziel unterstützen. Eine Software, deren Oberfläche in den Augen der Benutzer als professionell und ausgereift wahrgenommen wird, hat eine höhere Chance, von den Benutzern als vertrauenswürdiger beurteilt zu werden (Lindgaard, G., 2011).

Durchführung

Auf Basis der klickbaren Wireframes aus der zweiten Iteration definierte das Projektteam, welche Bereiche verfeinert werden und in welche Richtung sich das visuelle Design bewegen sollte. Schnell wurde klar, dass das Design unter den Bedingungen der Benutzer – Radiologen in abgedunkelten Räumen – bewertet werden musste. Grauwerte mit schwachem Kontrast wirkten bei Tageslicht sehr schlicht, wurden jedoch in dunklen Räumen beinahe als leuchtend wahrgenommen. Aus diesem Grund fand die visuelle Gestaltung schliesslich in einem abgedunkelten Raum statt. Das visuelle Design wurde iterativ erarbeitet, wobei ein Projektmitglied den Lead hatte. Die verschiedenen Entwürfe wurden gemeinsam besprochen und weiterentwickelt.

Nachdem die grundsätzliche Richtung festgelegt worden war, teilte sich das Projektteam auf, um die folgenden Aspekte als interaktiven Prototypen auszugestalten:

- Hover-Effekte und Bedienelemente
- Direkte Bearbeitung innerhalb eines Fensters (ohne Bearbeitungsmodus)
- Funktionsweise der linken Seitenleiste

Um die Detailinteraktionen zu demonstrieren, erstellte das Projektteam einen Prototypen im Animationsprogramm «Principle». Dieser zeigte wie die Navigation in der linken Seitenspalte, das Übersichtsdiagramm, sowie das Einzeichnen einer neuen Läsion funktionierte. Die Animationen waren ein wichtiger Bestandteil des visuellen Designs, um die räumlichen Beziehungen und Zusammenhänge des Layouts zu zeigen.

Ergebnisse

Typografie

Das Projektteam entschied sich für die Schriftart «Roboto». Sie kombinierte eine hohe Informationsdichte mit guter Lesbarkeit (Google, Material Design: Typography, 2017) und war mit ihren 12 Schriftschnitten flexibel einsetzbar.

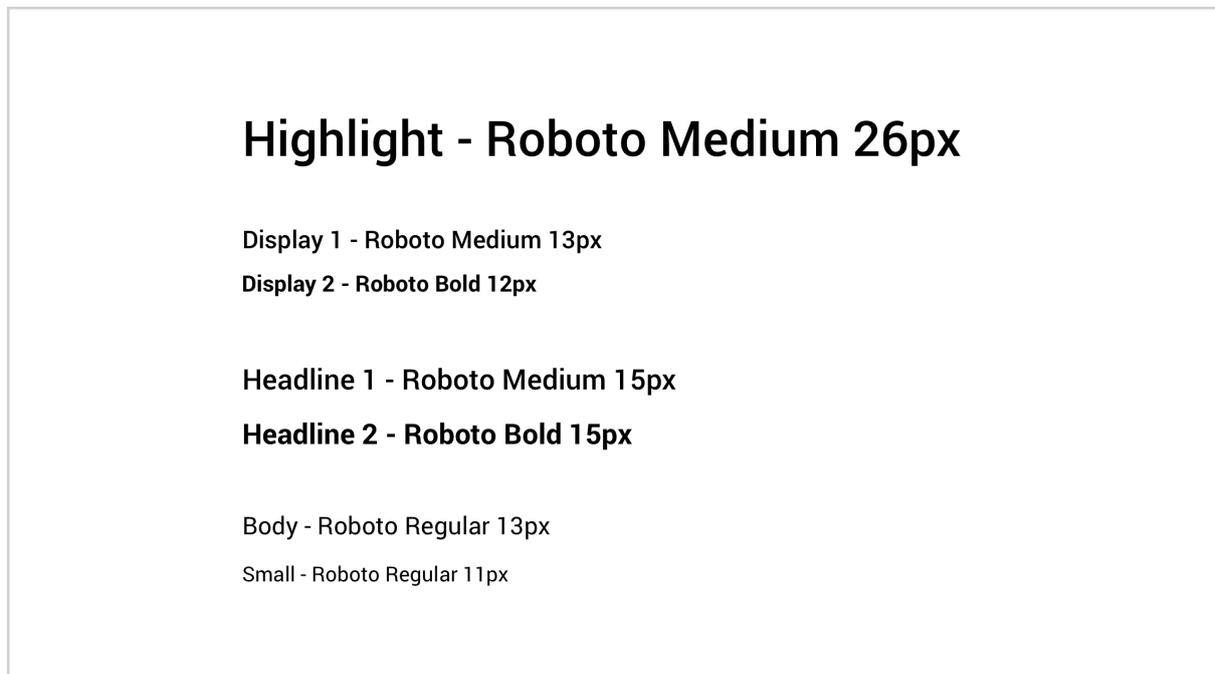


Abb. 35: Verwendet Schriftschnitte für LungStage

Farben

Die Farben für das Visual Design wurden so gewählt, dass sie für eine Anwendung in dunklen Räumen optimiert waren. Unter diesen Gegebenheiten sind dunkle Grautöne auf die Dauer angenehmer und lassen die PET/CT-Bilder besser zur Geltung kommen. Als Inspiration hierfür dienten bestehende Applikationen der Radiologie (z.B. Mint Lesion™).

Damit die Oberfläche nicht «grau-in-grau» wirkte, wurde Blau beigemischt. Diese Farbe wird als vertrauenswürdig und stabil wahrgenommen und deshalb oft für Medizinal-Software eingesetzt (Goodwin, K., 2009, S.487). Blau hat aber noch eine weitere Eigenschaft: es blockiert die Ausschüttung von Melatonin – dem Hormon, das uns schläfrig macht (Universität Basel, 2017). Obwohl der Blauanteil für LungStage insgesamt sehr klein war, könnte er trotzdem unterstützend wirken, um bei langen Arbeitszeiten wach und konzentriert zu bleiben. Abbildung 36 zeigt die feine Abstufung der Layout-Farben. Unter normalen Bedingungen sind diese Farbunterschiede nur schwer zu erkennen, in dunklen Räumen hingegen sind die Unterschiede gut ersichtlich.



Abb. 36: Feine Abstufungen von grau-blau bis schwarz

Als Interaktionsfarbe für klickbare und aktive Elemente wurde ein helles Blau gewählt, welches sich genügend vom Hintergrund abhob. Für die Auszeichnung von Läsionen wurde pink gewählt, um auf den CT-Bildern einen klaren Kontrast zu erreichen.



Abb. 37: Zusätzliche Farben für Interaktion und Auszeichnungen

Interaktion

Ein Ziel aus der Auswertung der zweiten Iteration war es, die Läsionen direkter zu manipulieren. Zu diesem Zweck wurde der Bearbeitungsmodus entfernt und die Läsionen direkt editierbar gestaltet, sowie die Eigenschaften der Läsionen als Formularfelder in die rechte Seitenleiste integriert.

Um die Bedienelemente intuitiver zu gestalten, wurde eine Liste aller Interaktionen erstellt. Die Bedienung wurde für Maus und Tastatur optimiert, da sich die Radiologen dies von bestehenden Applikationen gewohnt waren. Die am häufigsten verwendeten Aktionen wurden zudem in die Werkzeugleiste integriert. Damit das UI nicht überladen wirkte und der Benutzer besser geführt werden konnte, wurde die Werkzeugleiste nur auf dem aktiven Fenster angezeigt.



Abb. 38: Werkzeugleiste mit den wichtigsten Funktionen

Aktion	Bedienung
Bild wechseln	Scrollen mit der Maus
Bild in der Ebene verschieden	Verschiebewerkzeug auswählen und Bild mit der Maus ziehen
Läsion bearbeiten	Die Enden der Markierung der Läsion ziehen und verschieben
Läsion entfernen	Via Kontextmenü
Neue Läsion einzeichnen	Mess-Werkzeug auswählen oder Mouse-Over auf Läsion mit tiefer Konfidenz
Window ändern	Klick auf Window-Angabe (z.B. W500 L50)
Serie wechseln	Klick auf Serienname
Bildgrösse ändern (zoom)	Lupen-Werkzeug auswählen und mit der Maus scrollen

Die Maximumintensitätsprojektion ist eine 2D-Darstellung mehrerer Ansichten von PET-Daten. Dabei werden aktive Stoffwechselprozesse deutlich sichtbar gemacht (Wikipedia, Maximumintensitätsprojektion, 2017).

Grafiken und Diagramme

Die Übersichtsgrafik konnte weiter vereinfacht werden durch die Einführung des MIP-Bildes auf der Patientenübersicht. Somit konnten auch Fernmetastasen erkannt werden, während sich die schematische Ansicht auf die Lunge konzentrierte. Die Diagramme für die Konfidenz des automatischen Befundes wurden mit den fehlenden Zuständen von T, N und M ergänzt und ebenfalls pink eingefärbt, um die Assoziation mit den befallenen Lymphstationen zu erleichtern.

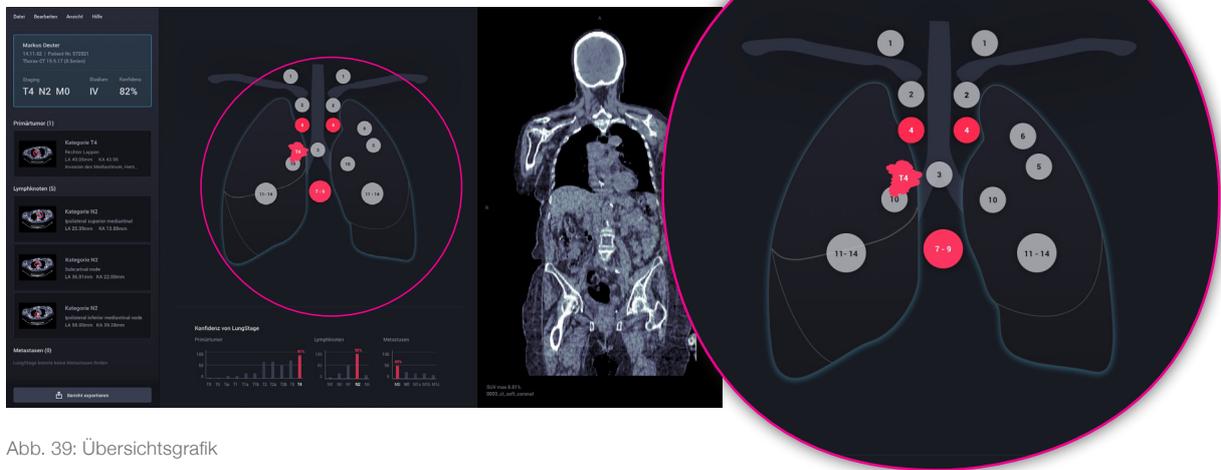


Abb. 39: Übersichtsgrafik

Sensitivität und Konfidenz

In der zweiten Iteration wurde entschieden, den Sensitivitäts-Regler zu entfernen, weil die Bedienung dessen unklar war. Das Projektteam versuchte jedoch einen anderen Weg zu finden, die Informationen zu erkannten Läsionen mit tiefer Konfidenz zu integrieren. Daraus entstand die Idee, die Läsionen mit tiefer Konfidenz als Vorschläge zu integrieren, welche bei «Mouse-Over» angezeigt werden. Dadurch kann LungStage dem Radiologen bei der Einzeichnung von neuen Läsionen helfen und die Zusammenarbeit von Mensch und Maschine erleichtern.

Layout

Um die Detailinformationen direkter zugänglich zu machen, wurde die Seitenleiste fix eingeblendet. Die Patientenübersicht wurde mit den Patienteninformationen kombiniert, da diese zusammengehörten und sich das Element dadurch besser von den erkannten Läsionen abheben konnte.

Während dem Research-Teil und den Walkthroughs zeigte sich, dass die Radiologie-Abteilungen der untersuchten Spitäler über Desktop-Monitore mit hoher Auflösung und Farbwiedergabetreue verfügten.

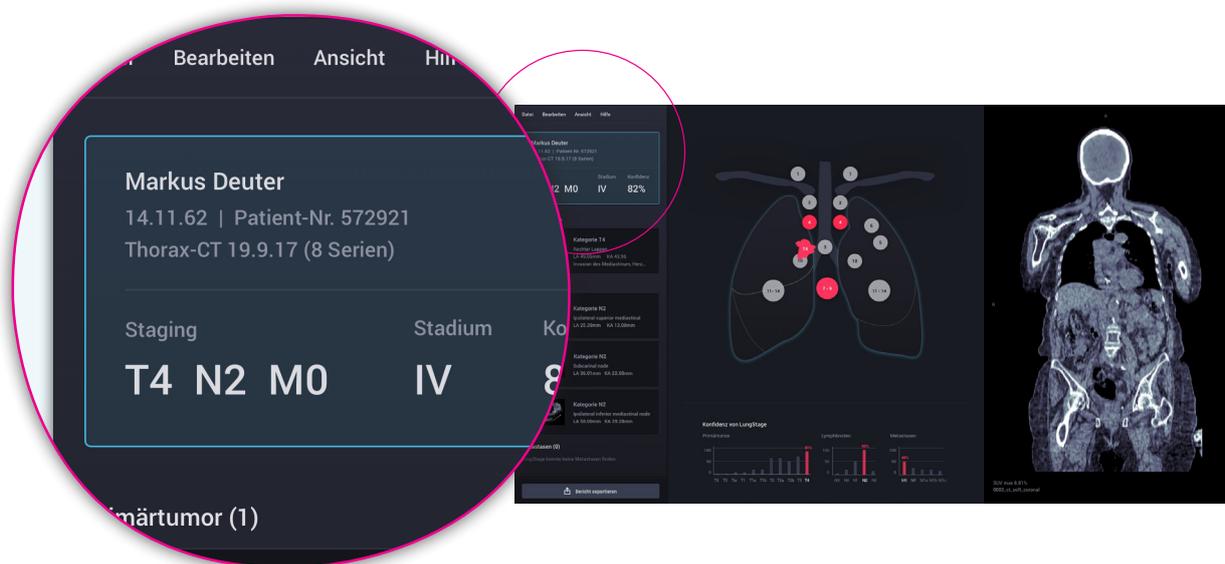


Abb. 40: Patienteninformationen

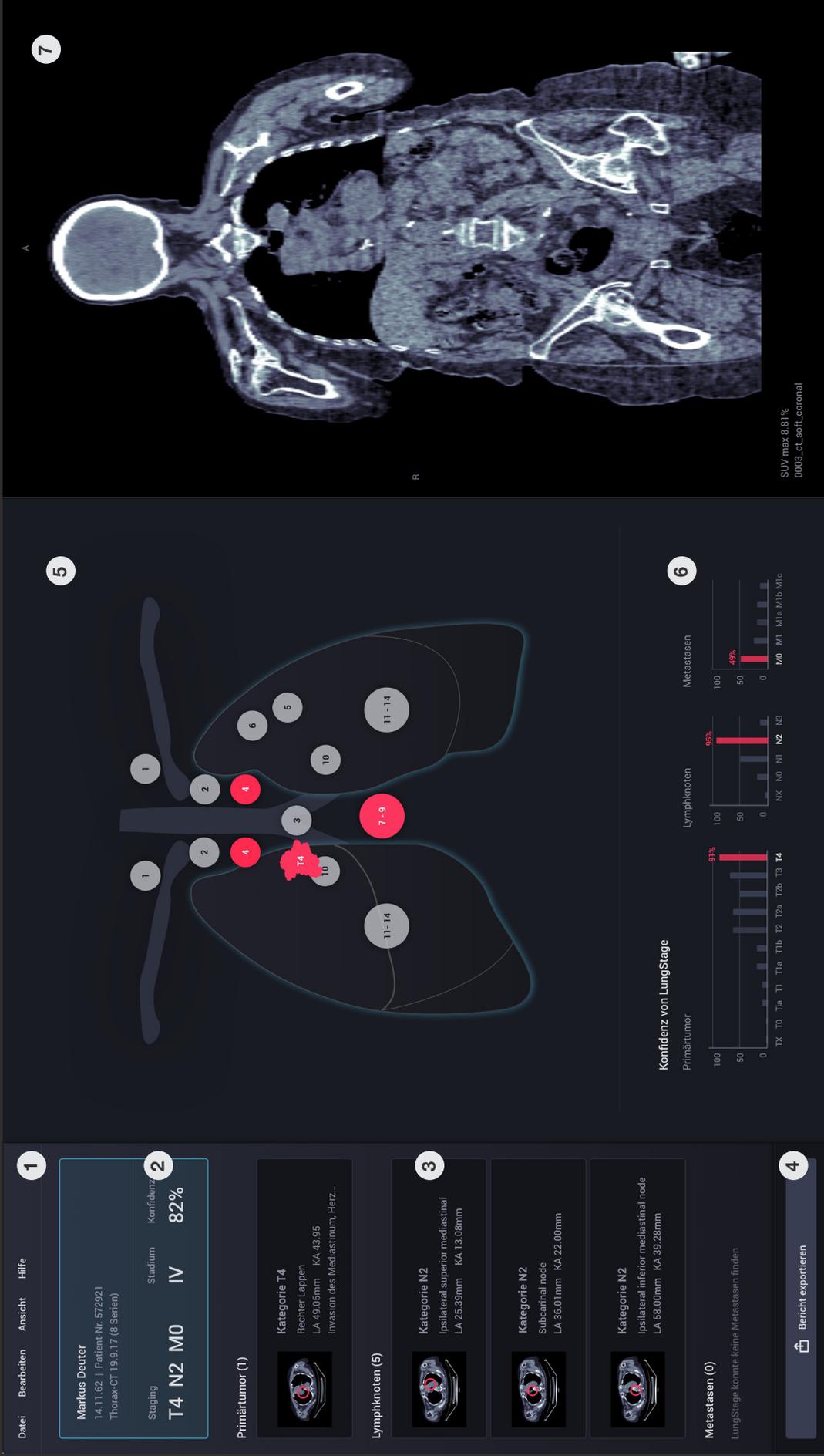


Abb. 41: Visual Design von der Übersicht der Befundung

- 1** Menüleiste (fixiert)
- 2** Patientenübersicht (scrollbar)
- 3** Liste der erkannten, bzw. manuell hinzugefügten Läsionen (scrollbar)

- 4** Exportfunktion (fixiert)
- 5** Übersicht von Tumor und befallenen Lymphstationen

- 6** Konfidenzangabe von LungStage bezüglich Primärtumor, Lymphknoten und Metastasen
- 7** MIP (inklusive SUV-Max Wert)

Datei Bearbeiten Ansicht Hilfe

Markus Deuter
14.11.62 | Patient-Nr. 572921
Thorax-CT 19.9.17 (8 Serien)

Stadium **IV** Konfidenz **82%**

Staging **T4 N2 M0**

Primärtumor (1)

Kategorie T4
Rechter Lappen
LA 49.05mm KA 43.95
Invasion des Mediastinum, Herz...

Lymphknoten (5)

Kategorie N2
Ipsilateral superior mediastinal
LA 25.39mm KA 13.08mm

Kategorie N2
Subcarinal node
LA 36.01mm KA 22.00mm

Kategorie N2
Ipsilateral inferior mediastinal node
LA 58.00mm KA 39.28mm

Metastasen (0)
LungStage konnte keine Metastasen finden

Läsion

Art der Läsion Kategorie **3**

Primärtumor T4

Lokalisation Rechter Mittellappen

Eigenschaften

- Hauptbronchus
- Invasion Viscerale Pleura
- Invasion der Brustwand
- Invasion Mediastinum
- Lappenobstruktion
- Nodulus im gleichen Lappen
- Tumor im anderen Lappen

Messwerte

LA	49.05mm
KA	43.95mm
Volumen	210 mm3
SUV-Max	8.81
SUV-Abweich.	3.2
HU-Mean	38.4
HU-Abweich.	14

Konfidenz von LungStage

Kategorie	Konfidenz (%)
TX	~10
T0	~10
T1	~10
T1a	~10
T1b	~10
T2	~10
T2a	~10
T2b	~10
T3	~10
T4	91

IN: 097
0003_ct_soft_axial

IN: 262
0003_ct_soft_coronal

IN: 097
0003_ct_soft_axial

IN: 244
0003_ct_soft_sagittal

1 4 Serien von Patientenbildern (inkl. Bildnummer und Serie-Titel)

2 Toolbar für die Manipulation der aktiven Serie

3 Einstellungen für die aktuelle Läsion

4 Messwerte (z.B. Länge und Breite der Läsion) und Konfidenzangabe von LungStage

Abb. 42: Visual Design von der Detailsicht einer Läsion

Fachliche Erkenntnisse

Die Entwicklung einer dunklen Oberfläche war eine neue Herausforderung für das Projektteam. Bei Tageslicht wären die feinen Farbunterschiede und die kleine Schriftgröße nicht geeignet, da diese schwierig wahrzunehmen wären. In dunklen Räumen hingegen gelten andere Regeln.

Die Beschaffung von qualitativem Bildmaterial stellte sich als schwieriger heraus als erwartet und bedingte einige Abgleiche mit dem Auftraggeber. Der Aufwand lohnte sich jedoch, um den Prototypen hochwertig wirken zu lassen.

Die regelmässigen Abgleiche im Team ermöglichten eine effiziente Arbeitsweise und beeinflussten das Resultat positiv. Dadurch konnten sich alle Projektmitglieder einbringen und parallel arbeiten.

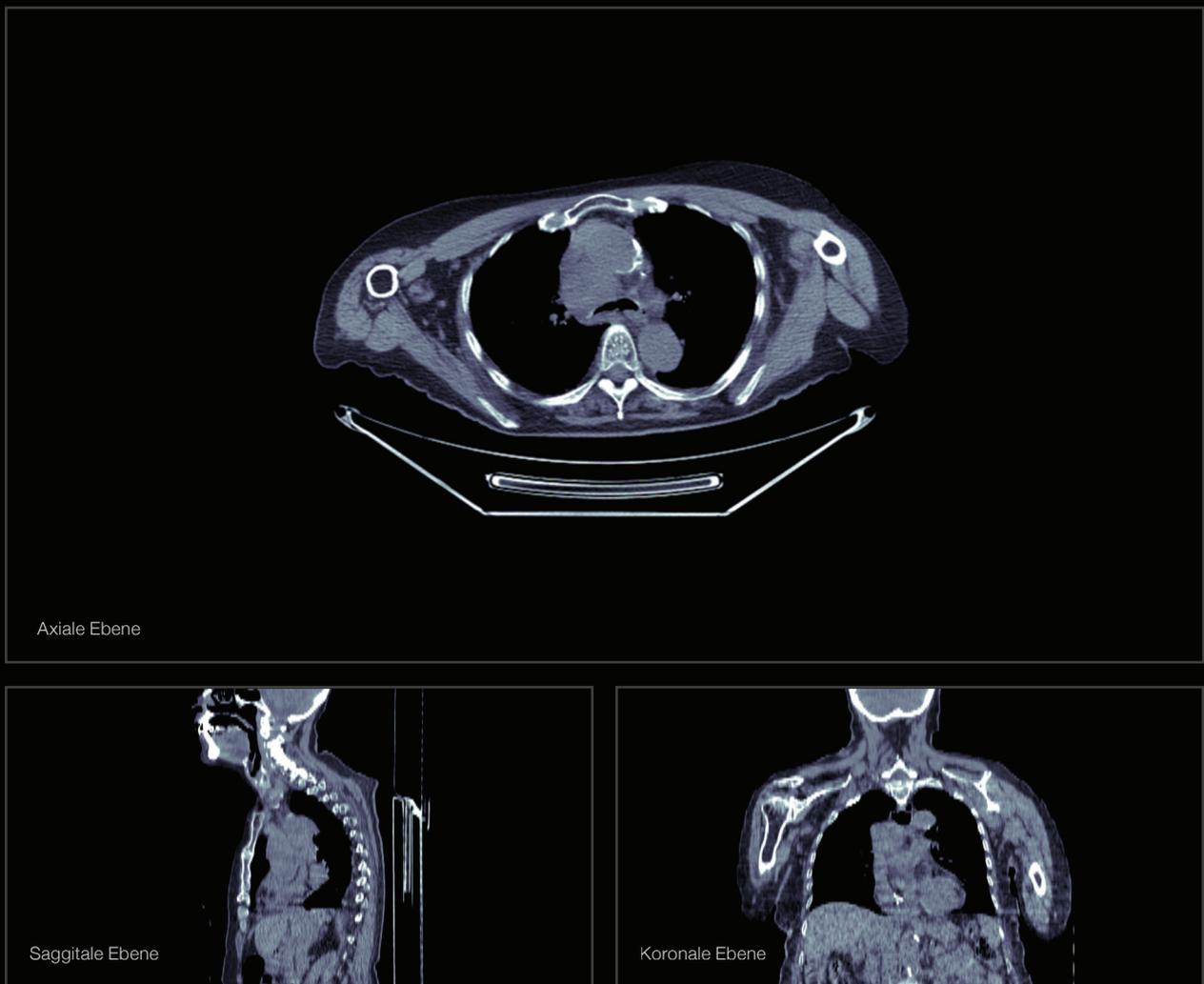


Abb. 43: CT Bilder aus technischem Prototyp von LungStage

Ebenen

Bei der tomographischen Untersuchung werden verschiedene Ansichten des Patienten erstellt, um diesen «scheibchenweise» in sogenannten Slices betrachten zu können. Anomalien lassen sich einfacher erkennen, indem verschiedene Ansichten verglichen werden und spielen deshalb in der Radiologie eine wichtige Rolle. Man unterscheidet «axial» (Ansicht von oben), «sagital» (Ansicht von der Seite) und «koronar» (Ansicht von vorne).

8. Fazit

Das Hauptergebnis der Arbeit war ein klickbarer, animierter Hi-Fi Prototyp für LungStage. Er demonstrierte, wie die Benutzer in Zukunft mit Hilfe von KI Lungenkrebs befunden könnten und ermöglichte dem Auftraggeber bereits während dem Projekt, weitere Partner für die Zusammenarbeit und Weiterentwicklung zu finden.

Die Benutzer von LungStage waren Radiologen und Nuklearmediziner, welche trotz langen Arbeitstagen und vielen Unterbrechungen schnell und exakt arbeiten mussten. Für diese Benutzergruppe wurde der Prototyp im benutzerzentrierten Verfahren in drei Iteration ausgearbeitet.

Das Ergebnis der drei Iterationen wurde dem Auftraggeber präsentiert und sowohl das Ergebnis als auch das Vorgehen wurden von ihm als überzeugende Grundlage für die weitere Entwicklung von LungStage eingeschätzt.

Picture Archiving and Communication System ist im Spital das System, in dem die PET/CT-Bilder abgelegt werden.

User Interface

Das UI bot Radiologen zwei Möglichkeiten, um sich effektiv und effizient einen Überblick über Tumore, Metastasen und Fernmetastasen zu verschaffen, die von der KI in den PET/CT-Bildern entdeckt wurden. Einerseits die Liste der erkannten Läsionen, gruppiert nach dem bekannten TNM System. Andererseits die schematische Darstellung der Lunge, die nach den Lymphknotenregionen der Lunge organisiert war, sowie die Darstellung der Maximum Intensity Projektion.

Die erkannten Läsionen konnten bearbeitet, gelöscht und hinzugefügt werden, um den Radiologen die vollständige Kontrolle über das Resultat zu geben. Das UI zeigte für jede gefundene Läsion die Konfidenz der KI an. Läsionen mit zu tiefer Konfidenz wurden anders dargestellt als Läsionen mit hoher Konfidenz und mussten explizit durch den Radiologen hinzugefügt werden.

LungStage wurde als eigenständige Applikation gestaltet, obwohl die Integration in ein «PACS» System aus Sicht der Benutzer vorzuziehen gewesen wäre. Der Auftraggeber benötigte aber einen Prototypen, der eigenständig zertifiziert und mit dem das Potential der Software gegenüber neuen Projektpartnern und Investoren demonstriert werden konnte.

Akzeptanz des Systems

Das Grundgerüst für die Akzeptanz des Systems – das sekundäre Ziel der Projektarbeit – war das iterative, benutzerzentrierte Vorgehen. Daraus wurden Massnahmen erarbeitet, um eine erfolgreiche Zusammenarbeit von Radiologen und KI zu ermöglichen:

- Das Einhalten der Usability Grundsätze war wichtig, die Technologie war nicht entscheidend.
- Ein UI, das die medizinische Terminologie verwendete. Lediglich die «Konfidenz» beschrieb eine abstrakte Eigenschaft, die sich aus dem technischen Modell der KI ergab, jedoch einfach zu interpretieren war.
- Das mentale Modell der Befundung blieb unverändert. Nach wie vor standen die PET/CT-Bilder des Patienten im Zentrum, und der Arzt konnte sich im Zweifelsfall am Ergebnis der KI selber Gewissheit verschaffen.
- Läsionen mit tiefer Konfidenz wurden dem Radiologen vorenthalten, um keine Unsicherheit zu erzeugen. Diese Läsionen konnten jedoch als Vorschläge angezeigt und vom Radiologen zum Befund hinzugefügt werden.

In der Evaluation ergaben die Messungen mit dem SUS Fragebogen eine durchschnittliche Gebrauchstauglichkeit von 72%, was das Projektteam als eine weitere Indikation für die Akzeptanz des Systems sah.

8.1 Bewertung der Ergebnisse

Bei beiden gesetzten Zielen war die Messbarkeit eine Herausforderung. Zwar fokussierte das Projektteam bei vielen Design Entscheidungen darauf, eine bessere Effizienz und Effektivität zu erreichen. Jedoch war problematisch, dass es dafür eine vergleichbare Einheit brauchte, um bei der abschliessenden Evaluation eine Steigerung zu belegen.

Der Interaktivitätsgrad des erarbeiteten Prototypen war nicht vergleichbar mit einem System, das im klinischen Einsatz zur Befundung von Lungenkrebs verwendet wurde. Zudem wäre die Effizienz stark von der Erfahrung der Radiologen abhängig, was die Messbarkeit der Effizienzsteigerung durch das UI verhindern würde. Dieses Risiko war bereits zu Beginn des Projektes bekannt, konnte aber nicht entschärft werden.

Die Akzeptanz des Systems war ebenfalls schwierig zu belegen. Für eine quantitativ gestützte Aussage anhand des SUS war die Messgrösse mit nur sechs Personen zu gering und wurde deshalb vom Projektteam nur als Indikation einbezogen. In den Gesprächen mit den Radiologen konnten keine Bedenken gegenüber der Einführung von KI im klinischen Alltag festgestellt werden. Allerdings war es möglich, dass die Testpersonen wohlwollend der vorgeschlagenen Lösung gegenüberstanden, da sie sich in der Forschungsgruppe in Basel engagierten oder ohnehin schon Interesse demonstrierten, weil sie sich als Testpersonen zur Verfügung stellten.

Das Risiko einer geringen Akzeptanz der Lösung könnte in Zukunft weiter ausgeräumt werden, indem zusätzliche Spitäler in die Zusammenarbeit einbezogen werden. Die Menge der involvierten Spitäler (3) und Testpersonen (7), welche während dieser Arbeit zur Verfügung standen, war zu gering, um klare Schlüsse zu ziehen.

Für die abschliessende Lösung ist zu beachten, dass die Qualität der KI massgeblich die Effizienz, aber auch die Akzeptanz des Systems beeinflussen wird. Liefert sie zu viele falsche Vorschläge und Klassifizierungen, verliert sie sofort an Vertrauen.

Für den Auftraggeber waren die Ergebnisse insgesamt sehr zufriedenstellend. Dies konnte in einer abschliessenden Präsentation abgeholt werden. Durch laufende Kommunikation während allen Projektphasen wurde sichergestellt, dass die Erwartungen erfüllt und die richtigen Lieferobjekte erstellt werden konnten.

8.2 Empfehlung für den Auftraggeber

Dem Auftraggeber empfahl das Projektteam, weiterhin benutzerzentriert vorzugehen und eng mit den medizinischen Stakeholdern zusammenzuarbeiten. Das Vorgehen bewährte sich, wie zahlreiche Rückmeldungen während dem Projekt zeigten.

Die Walkthroughs, welche in dieser Arbeit durchgeführt wurden, basierten auf Mock-Daten. Dies reichte aus, damit sich die Testpersonen vorstellen konnten, wie das UI später aussehen sollte und konnten dadurch qualitatives Feedback geben. Die Qualität der KI ist aber ein entscheidender Aspekt für das Vertrauen in das Resultat von LungStage.

Darum empfahl das Projektteam, den erarbeiteten Prototypen mit der KI zu verbinden und das Vertrauen in die automatischen Resultate mit Benutzern zu testen. Dieses Vorgehen war auch im Sinne des Auftraggebers und wurde bereits so geplant.

Während dem Projekt wurden verschiedene Themen festgehalten, welche Potenzial zeigten, jedoch nicht Teil dieser Arbeit waren:

- Es wurde mehrmals erwähnt, dass ein entscheidender Mehrwert geschaffen werden könnte, indem weitere Datenquellen aggregiert würden: z.B. die medizinische Vorgeschichte des Patienten, oder Datenbanken von ähnlichen Fällen.
- Ebenfalls spannend war die Aussage, dass oft andere Faktoren als das TNM relevant sind für die Behandlung des Patienten. Wenn z.B. eine Entzündung vorhanden ist, welche es nicht erlaubt, den Patienten zu operieren. Das TNM alleine ist nicht entscheidend für die Behandlung des Patienten.
- Ein kollaborativer Ansatz mehrerer Radiologen innerhalb der Klinik sei spannend, jedoch sei das Umfeld auch sehr kompetitiv geprägt. Klinik-übergreifend wurde es nicht als sinnvoll erachtet, weil die Abläufe in den Spitälern zu unterschiedlich organisiert sind.

8.3 Teamreflexion

Das Projektteam empfand die Ausgangslage des Projektes als extrem spannend. KI ist ein hochaktuelles Thema, welches immer mehr an Bedeutung gewinnt. Der Auftraggeber war hochmotiviert, womit er auch das Projektteam ansteckte.

Die Einblicke in die Radiologie waren herausfordernd, da die Komplexität der Materie sehr hoch war und es war zuweilen bewegend, ernste Patientenfälle zu sehen. Gleichzeitig war die Lockerheit und das Engagement der Radiologen überraschend. Sie nahmen sich viel Zeit, was bei ihrem Arbeitsalltag nicht selbstverständlich war, und interessierten sich für das Vorgehen und die HCID Methodik. War die Rekrutierung von Radiologen erst einmal geschafft, so war die Zusammenarbeit sehr unkompliziert. Die schwierige Rekrutierung bremste jedoch immer wieder den Projektfluss. Gerne hätte das Projektteam öfter und schneller iteriert, was sich jedoch nicht umsetzen liess in diesem Expertenfeld.

Immer wieder musste das Projektteam während der Arbeit neu fokussieren, da ständig weitere interessante Themengebiete auftauchten. Auch auf das Ausarbeiten eines komplett interaktiven Prototypen, oder die Einbindung der KI in den Prototypen musste leider verzichtet werden, da dies zu aufwändig gewesen wäre. Zudem musste viel Zeit für die Dokumentation eingeplant werden.

Herausfordernd war auch der Arbeitsmodus. Das Projektteam reservierte sich jeweils einen Tag pro Woche, jedoch waren die Unterbrüche dazwischen nicht förderlich für ein effizientes Arbeiten. Das Projektteam überlegte sich während der Projektplanung, Fokuswochen einzubauen, was aber nicht kompatibel mit privaten und beruflichen Verpflichtungen war.

Die Zusammenarbeit im Projektteam machte Spass, denn alle konnten ihre Stärken einbringen und gleich viel zum Ergebnis beitragen. Somit kann das Projektteam insgesamt auf ein erfolgreiches Projekt zurückblicken.

9. Quellenverzeichnis

Literatur

Goodwin, K. (2009). **Designing for the Digital Age: How to Create Human Centered Products and Services**. Wiley Publishing, Indianapolis.

Courage, C., Baxter K. (2005). **Understanding Your Users: A Practical Guide to User Requirements Methods, Tools and Techniques**. Morgan Kaufmann, San Francisco.

Cooper, A., Reimann, R., Cronin, D., Noessel C. (2014). **About Face: The Essentials of Interaction Design**. Wiley Publishing, Indianapolis.

Richter, M., Flückiger M. (2013). **Usability Engineering kompakt: Benutzbare Produkte gezielt entwickeln**, 3. Auflage. Springer, Berlin Heidelberg.

Beyer, H., Holtzblatt, K. (1998). **Contextual Design: Defining Customer-Centered Systems**. Morgan Kaufmann Publishers, San Francisco.

Mayhew, D. (1999). **The Usability Engineering Lifecycle: a practitioner's handbook for user interface design**. Morgan Kaufmann, San Francisco.

Lidwell, W., Holden, K., Butler, J. (2003). **Design: Die 100 Prinzipien für erfolgreiche Gestaltung**. Rockport Publishers, Inc., Gloucester.

Deining, M., Lichter, H., Ludewig, J., Schneider, K. (2005). **Studien-Arbeiten: ein Leitfaden zur Vorbereitung, Durchführung und Betreuung von Studien-, Diplom-, Abschluss- und Doktorarbeiten am Beispiel Informatik**. vdf Hochschulverlag, Zürich.

Wharton, C., Rieman, J., Lewis, C., Polson, P. (1994). **The Cognitive Walkthrough: A practitioner's guide**. In Nielsen J., Mack, R., Usability inspections methods. Wiley, New York.

Ulwick, W. A. (2016). **Jobs to be Done: Theory to Practice**. Idea Bite Press, USA.

Websites

Google LLC (2017). **Material Design**. <https://material.io/> (20.01.2018)

Google LLC (2017). **Material Design Typography**. <https://material.io/guidelines/style/typography.html#typography-typeface> (20.01.2018)

dkfz (2015). **PET: Technik, Forschung, Fachinformationen**. <https://www.krebsinformationsdienst.de/untersuchung/pet-technik.php> (20.01.2018)

Brandt, M. (2017). **Künstliche Intelligenz im Aufwärtstrend**. <https://de.statista.com/infografik/8496/investitionen-und-unternehmenszukauefe-im-bereich-kuenstliche-intelligenz/> (20.01.2018)

Wikipedia (2017). **Fragekontexteffekt**. <https://de.wikipedia.org/wiki/Fragekontexteffekt> (20.01.2018)

Wikipedia (2017). **Maximumintensitätsprojektion**. <https://de.wikipedia.org/wiki/Maximumintensit%C3%A4tsprojektion> (20.01.2018)

Wikipedia (2017). **TNM-Klassifikation**. <https://de.wikipedia.org/wiki/TNM-Klassifikation> (20.01.2018)

Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P. (2011). **An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages**. <https://dl.acm.org/citation.cfm?id=1959023> (20.01.2018)

Universität Basel (2017). **Licht am Abend verbessert Leistung im Endspurt**. <https://www.unibas.ch/de/Aktuell/News/Uni-Research/Licht-am-Abend-verbessert-Leistung-im-Endspurt.html> (20.01.2018)

Brown, D. (2009). **In Which a Concept Model Makes Me Giddy**. https://articles.uei.com/concept_models/ (20.01.2018)

10. Abbildungsverzeichnis

Abb. 1: Canon Celesteion, <https://us.medical.canon/products/computed-tomography/> (20.01.2018)

Abb. 2: CT-Bild eines Lungengewebes, <http://www.radiologieherne.de/mrt-ct-untersuchungsregion-lunge> (20.01.2018)

Abb. 3: Darstellung des EN ISO 9241-210 Ablaufs, <http://www.procontext.com/aktuelles/2010/03/iso-9241210-prozess-zur-entwicklung-gebrauchstauglicher-interaktiver-systeme-veroeffentlicht.html> (20.01.2018)

Abb. 4: PET Bild eines Lungengewebes, <https://www.radiologie-nuklearmedizin-bonn.de/PET-und-PET-CT.14.0.html?&l=1we> (20.01.2018)

Abb. 5: Beispielscreen Siemens syngo.via <https://i.ytimg.com/vi/cst59bJJ960/maxresdefault.jpg> (20.01.2018)

Abb. 6: Arbeitsumgebung der Radiologen im Unispital Basel

Abb. 7: Clustering zusammen mit dem Auftraggeber

Abb. 8: Umfrageresultate in Google Forms

Abb. 9: Fusions-Bild, <https://www.radiologie-nuklearmedizin-bonn.de/PET-und-PET-CT.14.0.html?&l=1we> (20.01.2018)

Abb. 10: Auszug aus der entwickelten Persona

Abb. 11: Übersicht der Objekte

Abb. 12: Läsionen im rechten Lungenflügel, <http://www.elsevier.es/es-revista-revista-espanola-medicina-nuclear-e-125-articulo-incidental-diagnosis-tumor-thrombosis-on-S2253654X15000542> (20.01.2018)

Abb. 13: Skizze Variante 1

Abb. 14: Skizze Variante 2

Abb. 15: Skizze Variante 3

Abb. 16: Walkthrough am Unispital Basel

Abb. 17: Skizze Variante 1

Abb. 18: Skizze Variante 2

Abb. 19: Skizze Variante 3

Abb. 20: Walkthrough am Unispital Basel

Abb. 21: Illustration aus LungStage Prototyp

Abb. 22: Skizze Variante 1

Abb. 23: Die Verwendung von Comic Sans für die Wireframes

Abb. 24: Anpassungen der schematischen Darstellung

Abb. 25: Wireframe von der Übersicht der Befundung

Abb. 26: Wireframe von der Detailansicht einer Läsion

Abb. 27: Wireframe des Editiermodus

Abb. 28: Alternatives Layout - Detailansicht

Abb. 29: Alternatives Layout - Übersicht

Abb. 30: Walkthrough mit einer Testperson

- Abb. 31: Besprechung mit dem Entwicklungsteam
- Abb. 32: Sensitivitätsregler
- Abb. 33: Verifikationsfunktion
- Abb. 34: TNM-Klassifikation
- Abb. 35: Verwendet Schriftschnitte für LungStage
- Abb. 36: Feine Abstufungen von grau-blau bis schwarz
- Abb. 37: Zusätzliche Farben für Interaktion und Auszeichnungen
- Abb. 38: Werkzeuggestreife mit den wichtigsten Funktionen
- Abb. 39: Übersichtsgrafik
- Abb. 40: Patienteninformationen
- Abb. 41: Visual Design von der Übersicht der Befundung
- Abb. 42: Visual Design von der Detailansicht einer Läsion
- Abb. 43: CT Bilder aus technischem Prototyp von LungStage

11. Themenrecherche

Vertrauen in Künstliche Intelligenz

Die Erkenntnisse aus dieser theoretischen Abhandlung zum Thema «Vertrauen in Künstliche Intelligenz» (nachfolgend «KI» genannt) waren Teil der Grundlage für das Vorgehen und die Gestaltung von LungStage.

Die Verbreitung von Softwaresystemen, welche auf künstlicher Intelligenz basieren, nimmt laufend zu (Brandt, M., 2017). So werden wir uns in Zukunft mit selbstfahrenden Autos fortbewegen, finanzielle Entscheidungen aufgrund von automatisierten Vorhersagen treffen oder medizinische Diagnosen erhalten. Die heutigen KI-Systeme haben aber meist «Blackbox» Charakter und Benutzer können automatisierte Entscheidungen, Vorhersagen oder Handlungen nicht nachvollziehen. Dies führt häufig zu Misstrauen, Verwirrung und Frustration. Vertrauen ist jedoch ein Schlüssel-Faktor bei der Akzeptanz dieser KI-Systeme (Vetrov, Y., 2017). Leider ist die Rolle des Menschen ein oft übersehener Aspekt auf dem Gebiet. Im Mittelpunkt steht häufig die Technologie und Modelle solcher KI-Systeme. Es wird kaum ein Blick auf den Benutzer und seine Bedürfnisse geworfen. So entstehen Systeme, die aufwendig und teuer sind, ohne sicher gehen zu können, dass sie den Bedürfnissen der Benutzer entsprechen und akzeptiert und genutzt werden.

Für die HCI Community stellen solche Systeme neue Chancen dar. KI-Systeme können helfen das Benutzererlebnis massgeblich zu verbessern indem sie Vorgänge automatisieren (z.B. selbstfahrende Autos), empfehlen (z.B. Spotify oder Netflix), erkennen (z.B. Tagging von Menschen in Fotos) oder auf neue Art und Weise interagieren (z.B. Chatbots, Siri). Um diese Schnittstellen zu gestalten muss der Benutzer und seine Bedürfnisse in den Mittelpunkt gerückt werden. Eine wissenschaftliche Betrachtung der Akzeptanz und Nachvollziehbarkeit von Analyseresultaten solcher Systeme ist von fundamentaler Bedeutung. Diese theoretische Abhandlung verfolgt die Frage wie KI-Systeme gestaltet werden müssen, damit diese den Benutzerbedürfnissen gerecht werden.

Grundverständnis von KI

KI ist ein Teilgebiet der Informatik, welches sich mit der Automatisierung intelligenten Verhaltens befasst. Innerhalb der KI unterscheidet man die sogenannte «starke» von der «schwachen» KI.

Die «starke» KI

Die «starke» KI beschreibt eine Maschine, die zu allem fähig ist, wozu auch ein Mensch in der Lage wäre (Petereit, D., 2016). Ein System, das intelligent auf seine Umwelt reagiert und sich wie ein Mensch verhält. Bei starker KI geht es um die Schaffung von Bewusstsein. Dieses Konzept wird in den letzten Jahren immer häufiger von Filmemachern thematisiert. Tatsächlich ist starke KI lediglich immer noch eine Idee und noch nicht über die konzeptionelle Ebene hinausgekommen.

So zum Beispiel die Filme «Her» (2013), «Ex Machina» (2015).

Die «schwache» KI

Die «schwache» KI hingegen übernimmt lediglich einzelne Fähigkeiten des Menschen. Zum Beispiel das Erkennen von Texten, Bildern oder der Sprache. Ein Hauptkriterium ist die Fähigkeit mit Unsicherheit und Wahrscheinlichkeiten umzugehen. Letztlich geht es um die Simulation intelligenten Verhaltens mit Mitteln der Mathematik und der Informatik. Hier werden seit Jahren schnelle Fortschritte gemacht. Einsatzgebiete sind allgegenwärtig in Medizin, im Finanzwesen, in der Mobilität, etc (Frey, C. und Osborne M., 2013). Am schwierigsten zu automatisieren sind Tätigkeiten, die hohe soziale Intelligenz (z.B. PR-Beratung), Kreativität (z.B. Mode-Design) oder Feingefühl und Flexibilität bei den Bewegungen (z.B. Chirurgie) erfordern (Mannino, A., 2015). In diesen Bereichen ist der Stand der KI-Forschung noch weit vom Niveau menschlicher Experten/innen entfernt.

Diese Arbeit fokussiert sich vor allem auf die Analyse «schwacher» KI-Systeme.

Funktionsweise von KI-Systemen

Die Fähigkeit zu Lernen ist eine Hauptanforderung an KI-Systeme und ist ein integraler Bestandteil.

Maschinelles Lernen

Maschinelles Lernen hilft, neues Wissen durch ein künstliches System zu erwerben (Manhart, K., 2017). Der Computer erstellt selbstständig Wissen aus Erfahrung und findet eigenständig Lösungen für neue und unbekanntere Probleme. Dazu analysiert ein Programm Beispiele und erkennt mit Hilfe selbstlernender Algorithmen, in den Daten bestimmte Muster und Gesetzmässigkeiten. Zu Beginn wird der lernenden Software Daten präsentiert. Beispielsweise vermerken Programmierer, dass ein bestimmtes Bild «ein Hund» und ein anderes «kein Hund» ist. So erhält die Lernsoftware ständig Feedback, die es nutzt, um das Modell anzupassen und zu optimieren: Mit jedem präsentierten Bild wird das Modell genauer und kann schlussendlich Hunde von Nicht-Hunden unterscheiden. Das Ziel von Machine Learning ist es, Daten miteinander zu verknüpfen, Zusammenhänge zu erkennen, Rückschlüsse zu ziehen und Vorhersagen zu treffen.

Deep Learning

Das am weitesten verbreitete maschinelle Lernverfahren ist mittlerweile Deep Learning. Ein wichtiges Einsatzgebiet ist neben der Verarbeitung von Sprache das Erkennen von Objekten in Bildern.

Deep Learning funktioniert ähnlich wie ein Kind den Begriff «Hund» lernt: Zum Training wird dem Programm zunächst Daten zur Verfügung gestellt, z.B. Bilder, die mit den Meta-Tags «Hund» oder «nicht Hund» markiert wurden. Das Programm verwendet diese, um ein Feature-Set für Hunde und ein Vorhersagemodell zu bauen.

Zuerst registriert das Programm lediglich Helligkeitswerte der Pixel. Danach erkennt es, dass Pixel zu Linien verbunden sind und unterscheidet zwischen horizontalen und vertikalen Linien. Dies geht so weiter bis eine Ebene erreicht wird, in der Beine unterschieden werden können. In einem weiteren Schritt könnte das Programm vorhersagen, dass alles mit vier Beinen ein Hund ist. Und schlussendlich ist es soweit Hunde von Nicht-Hunden unterscheiden zu können. Bei jeder Iteration wird das Vorhersagemodell immer komplexer und genauer.

Vor- und Nachteile von KI-Systemen

Vorteile

KI hilft effizienter zu arbeiten

KI kann helfen, Bilder in kürzerer Zeit und genauer zu organisieren und zu editieren, als sein menschliches Gegenstück. Sie übernimmt sehr komplexe Aufgaben, wie etwas die Erkennung von Mustern und Gesetzmässigkeiten und kann so z.B. Fehlermuster entdecken. Die Verarbeitung von grossen Datenmengen, ist einer der grössten Vorteile.

KI ist häufig präziser

Durch die enorme Rechenleistung können KI-Systeme genauere Aktionen oder Resultate erzielen. So z.B. bei der Erkennung von Krebstumoren und bei der Empfehlung von Therapien. Diese sind laut Dr. Klaus Manhart häufig präziser als die menschlichen Experten.

Beide Aspekte führen dazu, dass Benutzer solcher Systeme repetitive oder aufwändige Arbeiten an diese Systeme delegieren und sich auf Aspekte der Arbeit konzentrieren können, die komplexere oder kreativere Denkvorgänge benötigen.

Nachteile von KI-Systemen

Fehlschlüsse sind nicht auszuschliessen

Je grösser der Einsatz von KI ist, desto weitreichender sind auch mögliche Fehlschlüsse oder Ungenauigkeiten. Wenn es beispielsweise um medizinische Prognosen, autonomes Fahren oder um die Steuerung automatischer Waffensysteme geht (Wolfangel, E., 2016). Experten bezweifeln, dass man dieses Problem lösen kann, denn die Schwierigkeit liegt im System und in der Anwendung: Es gibt keine Regeln bei der Programmierung, welches Berechnungsverfahren für welchen Einsatz geeignet ist. Bei der Programmierung und beim Training des Systems werden häufig Annahmen getroffen. So kann man bei vielen Algorithmen die Richtigkeit des Ergebnisses nicht überprüfen. Ein möglicher Grund ist die «Long Chain of Responsibility» (Lange Verantwortungskette).

Zuerst entwickelt jemand einen Algorithmus, eine andere Person implementiert ihn, wiederum ein anderer wählt die Daten aus und jemand interpretiert das Ergebnis.

KI kann (noch) nicht auf veränderte Situationen reagieren

Maschinen können ihre Reaktionen nicht an veränderte Situationen anpassen. Die Systeme sind zwar in der Lage, enorme Mengen an Daten zu speichern und wiederholte Aufgaben lange durchzuführen, aber sie sind (noch) nicht in der Lage, anders zu handeln, als sie programmiert sind. Ganz anders als das menschliche Gehirn. Dies wird meist als ein Vorteil gesehen, ist jedoch ein Nachteil in heiklen Situationen.

Fehlendes Vertrauen führt zu Ablehnung von KI-Systemen

Eine weitere Hürde in der Nutzung von KI-Systemen kann in der Interaktion von Mensch und Maschine und das fehlende Vertrauen (Mankins, P., 2015) der Benutzer in diese Systeme liegen. Nur wenn Benutzer den Systemen vertrauen, werden sie die Berechnungen, Entscheidungen und Empfehlungen, die zukünftig den Alltag beeinflussen, akzeptieren und nutzen. Die Wichtigkeit dieses Aspekts ist einer der Hauptherausforderungen für die Konzeption und die Gestaltung von Mensch-Maschinen-Schnittstellen. Um richtige Schlussfolgerungen für die Gestaltung zu ziehen, müssen wir erst die Hintergründe und Ursachen verstehen:

Gründe für fehlendes Vertrauen in KI

Fehlende Transparenz

Einer der Schlüsselfaktoren ist die fehlende Transparenz dieser Systeme (Ismail, N., 2016). Intelligente Systeme kommen bereits heute in vielen Bereichen zum Einsatz. Die Datenverarbeitung und -sammlung passiert meist für den Nutzer unsichtbar in einer Art Blackbox. Er kennt oft nur das Ergebnis. Dabei spielen zwei Aspekte eine Rolle. Das Vertrauen in eine Vorhersage. Das heisst, ob ein Benutzer eine individuelle Prognose ausreichend vertraut um eine Handlung darauf zu nehmen. Und das Vertrauen in die zugrundeliegende Berechnung der Vorhersage, also dem Algorithmus.

Menschen schätzen Transparenz und Interaktion in KI-Systemen. In einer Studie (Kulesza, T., 2012) wurden Benutzer eines Musik-Empfehlungssystems analysiert, dass die Empfehlungen durch Rückkopplungs-Steuerungen generierte. Z.B. durch Song-Ratings, oder durch das Auswählen von bestimmten Musikmerkmalen. Dabei wurde den Benutzern ein 15-minütiges Tutorial präsentiert, wie die Musikempfehlungen zustande kamen. Die Teilnehmer schätzten es, die Details des zugrundeliegenden Systems zu verstehen. Darüber hinaus fanden die Forscher folgendes heraus: Je mehr die Teilnehmer mit dem System interagierten und je mehr sie über das System lernten, desto zufriedener waren sie mit den Empfehlungen. Diese Fallstudie zeigt ausserdem, dass Benutzer von «Blackbox» Systemen nuanciertere Rückmeldungen liefern wollen.

In einer weiteren Studie wurde die Auswirkung von Transparenz auf einer sozialen Ebene anhand des MovieLens Empfehlungssystems untersucht (Rashid, A., 2006). Dafür zeigte er den Nutzern den Wert ihrer potenziellen Filmbewertungen für die Community. Benutzer, die Informationen über den Wert ihres Beitrags für die MovieLens-Community erhielten, lieferten mehr Bewertungen. Personen, die Informationen über den Wert für eine Gruppe von Benutzern mit ähnlichem Geschmack erhielten gaben sogar noch mehr Bewertungen.

Fehlende Kontrollmöglichkeiten

Ein weiterer Punkt, sind fehlende Kontrollmöglichkeiten (Reeps, I., 2004). So ist es meist nicht möglich auf die Berechnungen, Entscheidungen und Empfehlungen der KI-Resultate Einfluss zu nehmen. Aktionen werden häufig einfach vom Computer ausgeführt ohne zu erklären wieso. Ergebnisse und Aktionen ohne das Gefühl der Kontrolle führen zu Stress und Frustration. Besitzt ein Benutzer das Gefühl, keine Kontrolle über die Situation zu haben, ist es wahrscheinlicher, dass er den Berechnungen nicht vertraut und nicht mehr nutzt.

Menschen vertrauen Menschen

Wir akzeptieren, dass Irren menschlich ist. Nicht so bei Maschinen. Wenn diese versagen, meiden wir zukünftig ihre Hilfe (Coren, M., 2016). Dies haben Forscher der University of Wisconsin herausgefunden indem sie folgendes testeten. Die Forscher beauftragten 160 Hochschulabsolventen, sie sollen die Belegung von Krankenhaus-Räumlichkeiten planen. Dabei sollten sie sich auf die Ratschläge von einem Computersystem oder einem menschlichen Spezialisten im OP-Management stützen. Nach sieben von 14 Planungsversuchen erhielten die Teilnehmer von ihren jeweiligen Beratern fehlerhafte Ratschläge.

Zu Beginn des Experiments berichteten die Testpersonen, dass sie gleiches Vertrauen in beide Quellen hatten. Dies änderte sich jedoch schnell nach dem fehlerhaften Ratschlag. Nachdem der «Computerberater» einen Fehler gemacht hatte, ignorierten viele Teestpersonen seinen Rat in späteren Versuchen. Computer-Konsultationen fielen von 70% auf weniger als 45%. Bei den menschlichen Beratern sahen die Forscher nur etwa einen 5%igen Abfall.

Prahl und sein Co-Autor Lyn Van Swol analysierten welche psychologischen Kräfte diesem Phänomen zugrunde liegen könnten. Die Teilnehmer in seiner Studie berichteten, dass sie fühlten «mehr gemeinsam zu haben» mit dem menschlichen als mit dem automatisierten Berater. Dies könnte ein Grund für dieses Phänomen sein: Das Gemeinsamkeitsgefühl der Testpersonen mit dem Berater. So empfanden sich die Personen dem menschlichen Berater näher: Das Gefühl nicht perfekt zu sein, den Willen die eigenen Fehlern korrigieren zu wollen und das Streben etwas richtig machen zu wollen. Um mehr Vertrauen in KI-Systeme zu zu schaffen glaubt Prahl, dass künstliche Intelligenz so programmiert werden sollten, einige dieser Merkmale zu zeigen.

Andrew Prahl, einer der Autoren der Studie stellt ausserdem fest, dass die Ergebnisse darauf hindeuten, dass potenzielle Effizienzgewinne aus der Automatisierung am Arbeitsplatz verloren gehen könnten, wenn Menschen das Vertrauen in diese Systeme

verlieren und diese nicht mehr nutzen. Während wir heute fehlerfreie Leistung von alltäglichen automatisierten Systemen erwarten, die unsere Laptops laufen lassen und Kreditkarten verarbeiten, treten wir in eine Ära ein, wo Computer weitreichende Prognosen und Urteile für uns machen werden. So fällt KI bereits rechtliche Urteile oder empfiehlt finanzielle Investitionen. Die Prognosen beschäftigen sich immer mehr mit einer ungewissen Welt und werden viel häufiger falsch sein als wir es von derzeitigen automatisierten Systemen kennen. «Das ist ein psychologisches Problem für die Benutzer solcher Systeme», schreibt Prahl. Menschen übertragen ihre Erwartungen von trivialen automatisierten Systemen auf komplexere KI-Systeme. Wir müssen aber akzeptieren, dass diese Systeme nicht immer richtig liegen.

Sorge um die Privatsphäre

Um KI-Systeme zu trainieren, werden viele Daten benötigt. Dadurch entsteht bei den Nutzern auch die Sorge um die Privatsphäre (Beschnitt, M., 2016) und somit auch das fehlende Vertrauen gegenüber den Anbietern. Gegenüber dem Nutzer gibt es keine Transparenz, was mit den Daten passiert und wie diese verarbeitet werden.

Implikationen für das User Interface

Um den obengenannten Gründen entgegenzuwirken, werden im folgenden Massnahmen aufgezeigt, wie das Vertrauen gefördert werden kann.

Massnahmen um Transparenz zu fördern

- Der Benutzer versteht die Berechnungen, Entscheidungen und Empfehlungen, visuelle oder textliche Erklärungen unterstützen dabei.
- Der Benutzer kann bei Bedarf weiterführende Informationen zum zugrundeliegenden Modell abrufen.
- Berechnungen bei denen Unsicherheit besteht, werden als solche ausgezeichnet.
- Informationen sind vollständig und aussagekräftig.
- Visuelle oder textliche Erklärungen der Berechnungen sind den Fähigkeiten und Vorkenntnissen des Benutzers angepasst.
- Die Menge oder die Gestaltung der Informationen überfordern den Benutzer nicht.

Massnahmen um Kontrollmöglichkeiten zu fördern

Die Wahrscheinlichkeit ist höher, dass der Benutzer den Berechnungen vertraut und weiter nutzt, wenn er das Gefühl der Kontrolle besitzt. Je mehr der Benutzer seine Fähigkeiten einsetzen kann, desto positiver ist auch das Benutzer-Erlebnis. Die Verbindung zwischen der Aktion und dem Ergebnis sollte deutlich gemacht werden, damit die Benutzer wissen, was und warum es passiert ist. Ein einfaches Mittel ist hier Benutzer-Feedback, um dem Benutzer das Gefühl der Kontrolle zu geben.

- Der Benutzer erhält Feedback zu seinen Aktionen und den Aktionen des Systems
- Der Benutzer hat Kontrolle über die Ergebnisse und kann diese nach seinen Bedürfnissen adaptieren, korrigieren oder ignorieren
- Die Kontrollmöglichkeiten sind verständlich und den Fähigkeiten angemessen
- Es gibt Variationen und Entscheidungsfreiraum
- Es besteht die Möglichkeit der sozialen Interaktion (z.B. Ratings)

Implikationen für die Gestaltung

Obwohl sich alle hier vorgestellten Massnahmen auf die Usability einer neuen Technologie beziehen, spiegeln diese doch grundlegende und universelle menschliche Bedürfnisse, Fähigkeiten und Anforderungen wieder. Die Technologie hinter einer Gestaltungslösung tritt somit in den Hintergrund. Was zählt ist der Benutzer und sein Verhalten. Um Mensch-Maschinen-Interaktionen (egal welcher Technologie) systematisch benutzerfreundlich zu gestalten wurden verschiedenste Richtlinien oder Heuristiken entwickelt.

Für medizinische Produkte werden von der U.S. Food and Drug Administration die 10 Usability-Prinzipien von Nielsen empfohlen:

- **Sichtbarkeit des Systemstatus** – Das System sollte den Benutzer immer auf dem Laufenden halten, indem es angemessenes Feedback in einer angemessenen Zeit liefert.
- **Übereinstimmung zwischen dem System und der realen Welt**– Das System sollte die Sprache der Benutzer sprechen und systemorientierte Terminologien vermeiden.
- **Benutzerkontrolle und -freiheit** – Ein System sollte Benutzer nie in Situationen geraten lassen, aus denen sie nicht wieder zurückfinden.
- **Konsistenz und Standards** – Benutzer sollten sich nicht über unterschiedliche Wortwahl für gleiche Situationen oder Aktionen wundern müssen.
- **Fehlerverhütung** – Besser als gute Fehlermeldungen ist ein gutes Design, welches das Eintreten von Fehlern erst gar nicht zulässt.
- **Wiedererkennen statt sich erinnern** – Das Kurzzeitgedächtnis eines Benutzers ist begrenzt. Deshalb sollten sie sich nicht an Informationen erinnern müssen, die in einem ganz anderen Bereich des Dialogs von Bedeutung waren.
- **Flexibilität und Effizienz der Benutzung** – Features, die es einfach machen ein System zu erlernen, sind für den erfahrenen Benutzer häufig lästig. Gut gewählte Abkürzungen, die der Anfänger nicht sieht, können in ein System aufgenommen werden, so dass beide Nutzergruppen zufrieden sind.
- **Ästhetik und minimalistisches Design** – Jegliche Information sollte in einer natürlichen und logischen Ordnung erscheinen.
- **Hilfe beim Erkennen, Diagnostizieren und Beheben von Fehlern** – Gute Fehlermeldungen sind defensiv, präzise und konstruktiv. Präzise Fehlermeldungen geben dem User genaue Informationen über die Ursache des Problems.
- **Hilfe und Dokumentation** – Jede kleine Website sollte ohne Hilfe auskommen.

Weitere Überlegungen

Gemeinsame Sprache

Um erfolgreiche Schnittstellen für KI-Systeme zu gestalten benötigt es die Entwicklung einer gemeinsamen Designsprache, die auf die Bedürfnisse der Benutzer zugeschnitten ist. Diese Sprache ermöglicht es interdisziplinär mit Psychologen, Designern und Ingenieuren zusammenarbeiten zu können.

Amershi, S. (2012) untersuchte KI-Systeme über mehrere Felder hinweg und identifizierte spezifische Designfaktoren, die das Zusammenspiel von Mensch und Maschine beeinflussen (z.B. der Fokus der Aufmerksamkeit einer Person während einer Interaktion) und Design Dimensionen, um diese Faktoren zu adressieren (z.B. Benutzer Feedback). Solche Untersuchungen helfen eine gemeinsame Sprache für KI-Systeme zu entwickeln.

Neben der Entwicklung einer gemeinsamen Sprache braucht es Prinzipien und Richtlinien für die Art und Weise, wie wir künftig Interaktion mit KI-Systeme gestalten. Ähnlich wie bei der Gestaltung von traditionellen Schnittstellen. Einige Prinzipien für die Gestaltung traditioneller Schnittstellen können direkt in die Gestaltung von KI-Systeme eingesetzt werden.

Fazit

Wenn die Technologie sich den Bedürfnissen und den Fähigkeiten der Benutzer unterordnet, steigt das Vertrauen und die Akzeptanz in KI-Systeme. Das heisst, es gilt Systeme zu entwickeln, die transparent in ihren Funktionen und Resultaten sind und gleichzeitig dem Nutzer das Gefühl der Kontrolle vermitteln. Hersteller müssen das Thema der Privatsphäre adressieren und ihren Verbrauchern Sicherheit vermitteln. Die Entwicklung von KI-Systemen ist rasant. In Zukunft werden sie einen grossen Teil unseres Alltages bestimmen. Es reicht nicht aus, die grundlegenden Usability-Prinzipien zu berücksichtigen. Die Gestaltung dieser Systeme muss, durch stetige Evaluation mit dem Benutzer optimiert und verbessert werden.

Nielsen & Norman (2016) haben dies in folgendem Zitat treffend zusammengefasst:

«No matter how different the technology, the people who are using it haven't changed. And most usability principles have more to do with human capabilities and limitations than with technology.»

Literaturverzeichnis

Brandt, M. (2017). **Künstliche Intelligenz im Aufwärtstrend.** <https://de.statista.com/infografik/8496/investitionen-und-unternehmenszukauefe-im-bereich-kuenstliche-in-telligenz/> (22.01.2018)

Vetrov, Y. (2017). **Algorithm-Driven Design: How Artificial Intelligence Is Changing Design.** www.smashingmagazine.com/2017/01/algorithm-driven-design-how-artificial-intelligence-changing-design/ (22.01.2018)

Petereit, D. (2016). **Was ist eigentlich der Unterschied zwischen AI, Machine Learning, Deep Learning und Natural Language Processing?** <https://t3n.de/news/ai-machine-learning-nlp-deep-learning-776907/> (22.01.2018)

Frey, C. und Osborne, M. (2013). **The future of employment.** https://web.archive.org/web/20150109185039/http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf (22.01.2018)

Mannino, A. (2015). **Künstliche Intelligenz: Chancen und Risiken.** <https://ea-stiftung.org/files/Kuenstliche-Intelligenz-Chancen-und-Risiken.pdf> (22.01.2018)

Manhart, K. (2017). **Was Sie über Maschinelles Lernen wissen müssen.** www.computer-woche.de/a/was-sie-ueber-maschinelles-lernen-wissen-muessen,3329560 (22.01.2018)

Wolfangel, E. (2016). **Die Grenzen der künstlichen Intelligenz.** www.spektrum.de/news/die-grenzen-der-kuenstlichen-intelligenz/1409149 (22.01.2018)

Mankins, P. (2015). **Can We Design Trust Between Humans and Artificial Intelligence?** www.fastcodesign.com/3047500/can-we-design-trust-between-humans-and-artificial-intelligence. (22.01.2018)

Ismail, N. (2016). **Artificial intelligence: increasing transparency while keeping humans in the loop.** www.information-age.com/keeping-humans-loop-ai-world-123463556/ (22.01.2018)

Kulesza, T. (2012). **Power to the People: The Role of Humans in Interactive Machine Learning.** www.aaai.org/ojs/index.php/aimagazine/article/view/2513 (22.01.2018)

Reeps, I. (2004). **State-of-the-Art Analyse. Usability, Design und Joy of Use.** http://hci.uni-konstanz.de/downloads/STAR_Reeps.pdf (22.01.2018)

Coren M. (2016). **People dump AI advisors that give bad advice, while they forgive humans for doing the same.** qz.com/693194/people-dump-ai-advisors-that-give-bad-advice-while-they-forgive-humans-for-doing-the-same/ (22.01.2018)

Beschnitt, M. (2016) **Die Angst vor intelligenten Systemen** <https://www.usability-blog.de/die-angst-vor-intelligenten-systemen/> (22.01.2018)

U.S. Department of Health and Human Services. **Applying Human Factors and Usability Engineering to Medical Devices** www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm259760.pdf (22.01.2018)

Nielson, N., (1995). **10 Usability Heuristics for User Interface Design** <https://www.nngroup.com/articles/ten-usability-heuristics/> (22.01.2018)

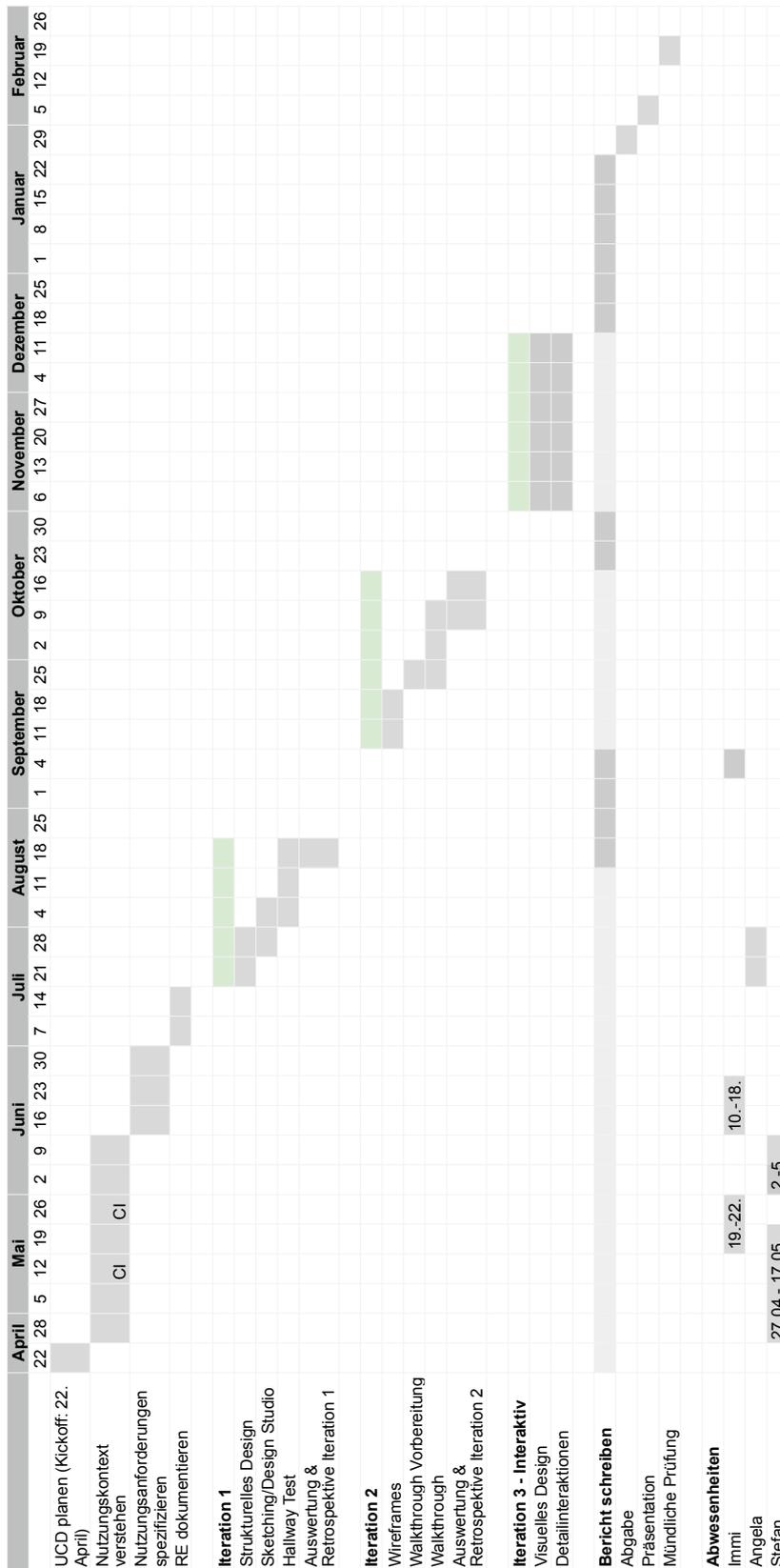
Amershi, S. (2012). **Power to the People: The Role of Humans in Interactive Machine Learning** <http://bradknox.net/public/papers/AIMagazine2015-IML.pdf> (22.01.2018)

Nielson, N. (2016) **Voice Interaction UX: Brave New World...Same Old Story.** <https://www.nngroup.com/articles/voice-interaction-ux/> (22.01.2018)

12. Anhang

12. Inhaltsverzeichnis	112
12.1 Projektplan	113
12.2 Risikoliste	114
12.3 Glossar	115
12.4 Stakeholder Liste	116
12.5 Leitfaden für CI	117
12.6 Affinity Diagramm Clustering	120
12.7 Umfrage Ergebnisse	122
12.8 Themenrecherche Quellen	132
12.9 User Stories	133
12.10 Persona	136
12.11 Kontext Szenario	138
12.12 Concept Model	139
12.13 Leitfaden Usability Walkthroughs mit Papier Prototyp	140
12.14 Ergebnisliste Usability Walkthroughs mit Papier Prototyp	141
12.15 Testszenarien Usability Walkthroughs mit Wireframes	144
12.16 Ergebnisliste Usability Walkthroughs mit Wireframes	146
12.17 SUS Ergebnis	151
12.18 Clustering Ergebnis Affinity Diagramm	151

12.1 Projektplan



12.2 Risikoliste

ID	Beschreibung	Art	Wahrscheinlichkeit	Auswirkung	Gegenmassnahme	Verantwortlich
1	Nicht genügend Testpersonen, da es sich um hochgradige Experten handelt	Projekt	Hoch	Validierung der Prototypen nicht möglich	Frühzeitige Kommunikation mit Flavio T. und frühzeitige Rekrutierung. In der Vereinbarung festhalten.	FT (4Quant)
2	Lückenhaftes Wissen durch die Hohe Komplexität der Materie	Projekt	Hoch	Fehlerhafte und unbrauchbare Designs	Gute Einarbeitung in das Thema und enge Kollaboration der Wissensträger	Projektteam
9	Das System bietet nicht die richtigen Parameter, um eine Akzeptanz bei den Radiologen zu ermöglichen	Produkt	Hoch	Keine Akzeptanz bei den Benutzern	Researchfokus auf wichtige Parameter setzen	Projektteam
3	Nicht genug Ressourcen, durch Krankheit, unvorhersehbare Ereignisse, Mutterschaft etc.	Projekt	Mittel	Geplante Aktivitäten können nicht durchgeführt werden	Flexible und langfristige Planung im Projektplan. Priorisieren der Massnahmen und Aktivitäten. Benötigte Ressourcen/Tasks sichtbar machen: Trello-Board mit wöchentlicher Taskplanung	Projektteam
4	Lange Reaktionszeit des Stakeholders	Projekt	Mittel	Zu wenig Feedback	Regelmässige Check-ins vereinbaren, aktive Kommunikation, Erwartungsmanagement	Projektteam
5	Stakeholderzentriertes Vorgehen, durch zu starke Beeinflussung des Stakeholders.	Projekt	Mittel	Zu wenig Fokus auf die Benutzer des Produkts, Marketingsicht	Benutzerzentriertes Vorgehen kommunizieren (regelmässig) und in der Vereinbarung festhalten	Projektteam
6	Detailgrad und Interaktivität des Prototypen reichen nicht aus um die Aufgaben zufriedenstellend zu testen	Projekt	Mittel	Evaluation wird zu wenig aussagekräftig sein	Abwägen zwischen Aufwand und Output um geeigneten Prototypen zu erreichen. Ressourcen beim Stakeholder für HiFi Prototyp abklären.	Projektteam
7	Erwartungen des Stakeholder an die Deliverables können nicht erreicht werden	Projekt	Mittel	Stakeholder ist nicht zufrieden	Transparenz an was gearbeitet wird, Gemeinsames Verständnis der Möglichkeiten	Projektteam
10	Keine Akzeptanz des Produkts, da Ärzte Angst haben, dadurch ersetzt zu werden	Produkt	Mittel	Keine Testpersonen oder weitere Zusammenarbeit möglich	Involvierung der Ärzte in die Produktentwicklung, Demonstrieren des Mehrwerts	FT (4Quant), Projektteam
11	Keine branchenweit akzeptierte Lösung, durch sehr unterschiedliche Arbeitsweisen der versch. Radiologen	Produkt	Mittel	Nischenlösung, die nur einen sehr geringen Mehrwert bietet	Einbezug versch. Ärzte an mehreren Spitälern.	FT (4Quant)
12	Nichterreichen der Produktziele (5min für das Staging)	Produkt	Mittel	Kein Anreiz, die vorgeschlagene Lösung weiter zu verfolgen	Benutzerzentriertes und iteratives Vorgehen	FT (4Quant)
8	Zu wenig Zeit für Berichtschreiben	Projekt	Tief	Stress am Ende des Projektes, bzw. niedrige Qualität des Berichts	Regelmässige Updates des Berichts, Taskverteilung nach jedem Treffen	Projektteam

12.3 Glossar

CAD	Computer-aided detection / diagnosis
CADe	Computer-aided detection
CADx	Computer-aided diagnosis
CIS	Clinical Information System
CT	Computer-Tomographie, ein Verfahren zum Erzeugen von morphologischen Schnittbildern.
FDG	18F-Fluordesoxyglucose, mit einem radioaktiven Tracer angereicherter Zucker. Wird dem dem Patienten für den PET Scan injiziert. Damit kann der erhöhte Stoffwechsel in Krebszellen im PET Scan sichtbar gemacht werden.
HU	Hounsfield Units, beschreibt die Gewebedichte
Läsion	Eine Schädigung, Verletzung oder Störung einer anatomischen Struktur oder physiologischen Funktion
NSCLC	Non-small cell lung cancer Lungenkarzinom, nicht-kleinzellig
Nuklearmediziner	Nuklearmediziner sind Radiologen mit einer Zusatzausbildung.
PACS	Picture Archiving And Communication System, z.B. GE Centricity
PET	Positronen-Emissions-Tomographie, ein Verfahren zum Erzeugen von Schnittbildern mit Stoffwechsel Informationen.
Radiologen	Radiologen sind Ärzte mit Fachausbildung Radiologie.
RIS	Radiology Information System, Scheduling der Patienten und Untersuchungen, Befundung
Sensitivität	Die Proportion von Positiven, die korrekt als Positive identifiziert wurden, also Tumore korrekt als Tumore erkannt wurden.
Serie	Ein oder mehrere zusammengehörige Bilder eines Scans eines Patienten
Spezifität	Die Proportion von Negativen, die korrekt als Negative identifiziert wurden, z.B. Entzündungen oder Wunden, die korrekt nicht als Tumor erkannt wurden.
SUV	Standard Uptake Value. Normierter Wert, der den Zucker-Umsatz im Stoffwechsel beschreibt.
TNM	Klassifikation für Tumore. T steht ein für den Primärtumor, N für Metastasen in Lymphknoten nahe des Primärtumors und M für Fernmetastasen, entfernt vom Primärtumor
USB	Universitätsspital Basel
USZ	Universitätsspital Zürich

12.4 Stakeholder Liste

Rolle	Kontakt	Kennt sich aus mit...	Interesse
Produkt Manager	Flavio Trolese	Produkt Vision, Kontakte und Rekrutierung	Muss die Vision kommunizieren können
CTO	Kevin Mader	System (Machine Learning, AI), Prototyp	Möglichkeiten der Technologie demonstrieren
Radiologie Experte	Thomas Weikert	Radiologieforschung	Forschungsergebnisse zum Wohlfühl der Menschen, Arbeitsplatz der Zukunft mitgestalten
Kunden	Unispital Basel, Unispital Zürich		Showcase, Entwicklungspartner und Early Adopter

12.5 Leitfaden für CI

Theorie zu Contextual Inquiry

Kontext: das Beobachtungsinterview findet vor Ort, am Arbeitsplatz des Benutzers statt, damit der Kontext der Handlungen verstanden wird.

Partnerschaft: in einer Meister-Lehrling-Beziehung versucht der Forscher in der Rolle des Lehrlings, die Arbeit des Benutzers zu verstehen.

Interpretation: Beobachtungen und Interpretationen des Forschers werden gemeinsam diskutiert und gegebenenfalls vom Benutzer korrigiert. (Neutral, nicht suggestiv)

Mitnehmen

- Beobachtungsleitfaden
- Schreibunterlage + Stift
- Foto- oder Filmkamera
- Tonaufnahme
- Screen recorder (for one or multiple screens; MS Windows)
- Voice recorder
- Incentive
- Einverständniserklärung

Vorgehen

- Intro & Warm up & NDA (15min)
- Observation/Interview eines Falles (1h 30min)
- Zeitmessung aller 3 Fälle (30min)
- Abschluss (15min)

Interview Leitfaden

Warm-up (15min)

Intro & NDA unterschreiben lassen

Fragen

Wie lange sind sie schon Radiologe?

Seit wann sind sie in diesem Spital?

An welchen Spitalern sind sie tätig?

Wie sieht ihr Tagesablauf aus (kurz)?

Wieviele Fälle bearbeiten sie an einem Tag?

Fragen zur Persona

- Was motiviert sie?
- Was stresst oder nervt sie?
- Was ist die schwerwiegendste Belastung in Ihrer Arbeit?
- Was ist für sie das Hauptziel beim Staging? (eher schnell oder eher gute Qualität)

Contextual Inquiry (1h)

Arbeitsablauf

Bitte stellen Sie sich vor wir sind Ihre Assistenten und Sie sind morgen abwesend. Wir müssten morgen die Analyse selber durchführen. Können Sie uns an einem konkreten Beispiel zeigen, wie Sie Bilddaten untersuchen und interpretieren.

Nicht relevant für uns ist die Akquisition von medizinischen Bilddaten, der Vergleich der Bilddaten mit früheren Untersuchungen oder die Untersuchung hinsichtlich allgemeinen Auffälligkeiten die nicht im Kontext von NSCLC liegen. Wichtig: Bitte erklären Sie uns immer den gesamten Prozess von der Initiierung bis zum Ende. □ Screenshots erstellen

Fokusfragen

Tools

- Welche Tools verwenden Sie beim Staging?
- Wie gerne arbeiten sie mit den bestehenden Tools?
- Was gefällt ihnen?
- Was gefällt ihnen nicht?
- Welche Ansichten verwenden Sie? (axial, etc) Ist Axial die Wichtigste oder verwenden Sie alle gleich oft?
- Wann und weshalb ändern Sie das Window? Sind Standard-Windows genügend?
- Wann und weshalb ändern Sie die Serie?
- Wie gut ist die (automatische) Unterstützung des Tools beim Staging?

Untersuchung/Interpretation

- Gibt es einen expliziten oder impliziten Workflow, dem sie folgen?
- Welche Aspekte des NSCLC Stagings empfinden sie als eher mühsam? Könnte etwas davon automatisiert werden? Wenn ja, was wünschst du dir, bzw. wie könnte dies aussehen?
- Wenn sie ein Staging eines Kollegen anschauen, wie gehen sie vor um es zu verifizieren? (selber Workflow wie alles selber machen, oder checkt er nur bestimmte Parameter)

- Welche Bilder sind wichtig oder aussagekräftig?
- Welche Probleme können beim Staging auftreten?
- Wie vergewissern Sie sich, wenn Sie unsicher sind? Welche Mittel setzen Sie ein?
- Wie sicher ist er sich jeweils, dass das TNM stimmt? (Oft, selten unsicher?)
- Wie und wann/warum holt er sich eine Zweitmeinung?
- Wenn sie sich unsicher sind, geben sie dies an im Befund bzw. Wie transparent ist dies für die behandelnden Ärzte?
- Zeichnen sie die Läsionen immer genau ein?
- Welche Personen interagieren mit ihnen und dem Befund? (Interessant für Interaktion mit AI)
- Wenn die schwerwiegendste Läsion einer Kategorie gefunden wurde, wie wichtig sind die Weiteren noch?

Fragen zum Kontext

- Was sind gängige Dinge, die beim Staging und am Tumorboard zu Diskussionen führen?
- Wie geht man bei verschiedenen Meinungen vor?
- Welche Fälle werfen unterschiedliche Beurteilungen auf?
- Welche Argumente sind es, die dann überzeugen?
- Wie präsentierst du deine Befundung den anderen Ärzten? Kannst du uns ein Beispiel senden?
- Wie sieht der Raum aus?
- Wird er gestört/abgelenkt durch Telefonate, andere Personen, Geräusche etc?
- Ist er alleine im Raum?
- Wie lange ist er im Raum?

12.6 Affinity Diagramm Clustering

ID	Finding	Typ	Cluster	Quellen
C2-1	Radiologen wünschen automatische Unterstützung	Beobachtung/ Quote	Automatisierung	34, 36, 63, 212, 64
C2-10	Wird Staging in LungStage zurückgespielt in PACS und kann via PACS Viewer angeschaut werden?	Beobachtung/ Quote	Integration	23, 140, 320
C2-11	Nebenbefunde werden im PACS Viewer gemacht	Beobachtung/ Quote	Integration	143, 211, 216, 52, 217
C2-12	Gibt bereits viele Tools, mit verschiedenen Spezialisierungen, Switching zwischen den Tools findet statt	Beobachtung/ Quote	Integration	1, 6, 215, 217, 333, 321
C2-13	Die genaue Kontur einer Läsion ist nicht so sehr relevant, die abstrakte Einzeichnung genügt	Hypothese	Interaktion mit Läsionen	221, 211, 205, 62, 28, Survey
C2-14	Anomalien werden einfacher erkannt, wenn man durch mehrere Bilder scrollen kann	Beobachtung/ Quote	Interaktion mit Läsionen	45, 121, 126, 28, 127, Survey
C2-15	Bilder mit guten Default-Settings im Vordergrund, Statistik und Parameter im Hintergrund	Idee	Interaktion mit Läsionen	131, 106, 69, 121, 118, 133, 26
C2-16	Oben-nach-unten Ansicht ist am wichtigsten, bei Bedarf zieht man andere Ansichten bei	Beobachtung/ Quote	Interaktion mit Läsionen	44, 56, 46, 201, Survey
C2-17	Hauptinteraktion mit Maus, Scrollrad und Keyboard Shortcuts	Beobachtung/ Quote	Interaktion mit Läsionen	142, 45, 51, 24, 47, 321
C2-18	User Interfaces sind dunkel gefärbt	Beobachtung/ Quote	Interaktion mit Läsionen	129, 350
C2-19	Welche Patientendaten muss man im Bild zeigen?	Beobachtung/ Quote	Interaktion mit Läsionen	132, 33, 325
C2-2	Verortung der Läsion kann automatisch gemacht werden	Hypothese	Automatisierung	207, 335
C2-20	Warum hat sich eine 3D Ansicht nicht durchgesetzt?	Frage	Interaktion mit Läsionen	Survey
C2-21	PET Signal immer zeigen wenn es vorhanden ist	Beobachtung/ Quote	Nuk-Med	100, 50, 136, 19, 124, 51, 328
C2-22	Nuk-Med arbeitet mit zusätzlichem Bildmaterial und Parametern	Beobachtung/ Quote	Nuk-Med	135, 57, 49, 68, 101-2, 101, 128, 50, 100, 56, 48, 328
C2-23	LungStage als Tool für die objektive Zweitmeinung mit bester UX positionieren	Idee	Positionierung	60, 153, 154, 335
C2-24	Die wichtigsten Parameter sind Wahrscheinlichkeit, Durchmesser für T, SUV MAX und Invasion	Beobachtung/ Quote	TNM	131, 307
C2-25	Pro T, N, M muss jeweils nur das aussagekräftigste Bild dargestellt werden bzw. definierend ist für das TNM	Beobachtung/ Quote	TNM	104, 151, 114, 134, 115, 116, 66, 210, Survey

ID	Finding	Typ	Cluster	Quellen
C2-26	TNM ist die wichtigste Information	Beobachtung/ Quote	TNM	103, 209, 113
C2-27	Mint Lesion Report ist gut aufbereitet	Beobachtung/ Quote	TNM	200, 332
C2-28	Geben Ärzte die Wahrscheinlichkeit an?	Frage	TNM	58, 152, 335, Survey
C2-29	Ausweisen, ob Bewertung der AI übernommen wurde oder vom Arzt manipuliert wurde. Ziel: Objektivität und Transparenz	Idee	TNM	145
C2-3	Falschdiagnosen sind die schwerste Belastung, nicht das Durchführen des Stagings an sich	Beobachtung/ Quote	Belastung	61, Survey
C2-30	Vertrauen: Sensitivity Slider um TNM Regions of Interest zu bestücken	Idee	TNM	220
C2-31	Ein standardisierter Workflow ist von den Ärzten gewünscht und hilft der Objektivität und Fehlervermeidung	Beobachtung/ Quote	Zusammenarbeit	37, 204, 55, 13, 48, 144, Survey
C2-32	Arbeiten in Fachgruppen (bis. ~10 Personen pro Raum) in verdunkelten Räumen an (zum Teil) Stehtischen	Beobachtung/ Quote	Zusammenarbeit	8, 9, 130, 10, 7
C2-4	Gibt bereits viele Tools, mit verschiedenen Spezialisierungen, Switching zwischen den Tools findet statt	Beobachtung/ Quote	Belastung	1, 6, 214, 217, 333, 321
C2-5	Zeiterlust ist das Schlimmste	Beobachtung/ Quote	Belastung	110, 11, 102, 5, 41, 22, 218, 32, 35, 147, 333, 343, 344
C2-6	Slice einer Läsion kann direkt geändert werden, wenn der automatisch definierte Slice dem Radiologen nicht passt	Idee	Customizations	219, 39, 203
C2-7	Default-Windows (Grauwerte) sind gut genug, manuelle Einstellung ist sekundär oder gar nicht notwendig. Ebenso beim PET Signal.	Beobachtung/ Quote	Customizations	122, 100, 27, 31, Survey
C2-8	Voreinstellungen müssen veränderbar sein	Beobachtung/ Quote	Customizations	123, 125, 126, 109, 122, 31, 327
C2-9	Fliesen Tumorboard Entscheide zurück in das TNM Rating in LungStage?	Frage	Integration	

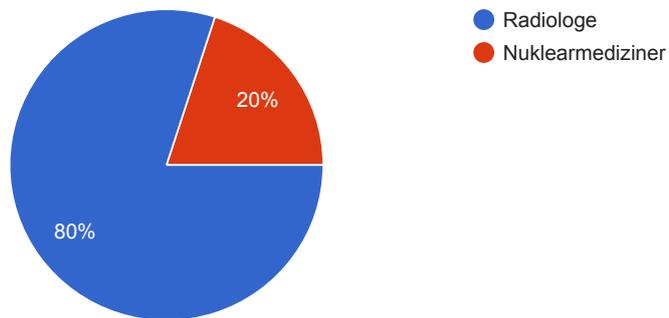
12.7 Umfrage Ergebnisse

Umfrage zum Staging von NSCLC

10 responses

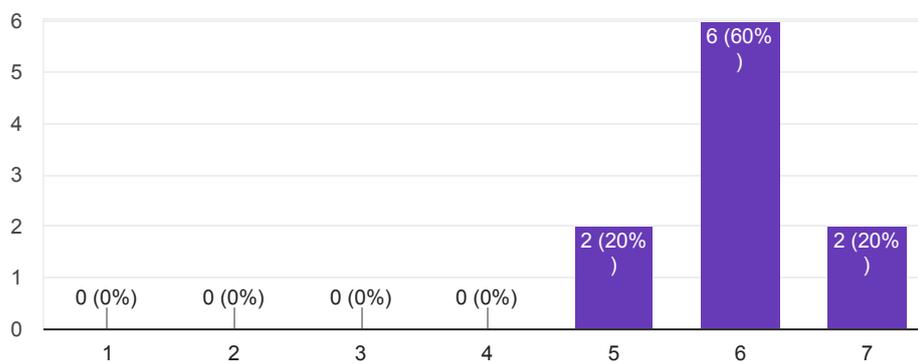
Ich arbeite als...

10 responses



Eine Fehldiagnose ist die schwerwiegendste Belastung in meiner Arbeit

10 responses



Welche Möglichkeiten nutzen Sie zur Verhinderung von Fehldiagnosen?

10 responses

Bei Unsicherheiten ziehe ich einen Kollegen bei oder dokumentiere meine Unsicherheit.

Genügend Schlaf, Konzentration

Strukturierte Befundung nach gleichem Schema. 4-Augen-Prinzip.

OA, Abteilungsleiter, Kollegen, Literatur (Bücher, statdx, Radiopedia)

Systematisches Vorgehen, zweimalige Kontrolle

Wissen, Internet, Bücher, Kollegen

supervision

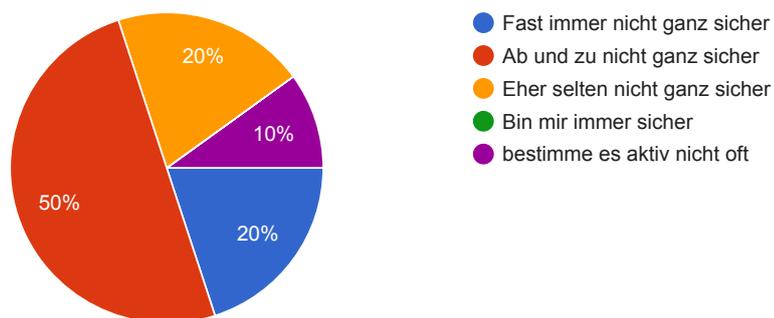
Zeit nehmen (besprechen, recherchieren und einfach mal kurz nachdenken)

Kollegen/Oberarzt fragen, Radiopaedia

Zweitmeinungen bei Unsicherheiten

Wie oft sind sie sich nicht ganz sicher beim Bestimmen des TNM?

10 responses



Stellen sie sich vor, ein Kollege hat als Resultat eines Befundes T1 bestimmt. Wie stellen Sie sicher, dass kein T3 übersehen wurde?

10 responses

Ich gehe die Slices des Scans genau durch.

- Ich überprüfe, ob im Scan ein Tumor > 7cm zu finden ist
- Ich überprüfe, ob der Tumor in die Brustwand, das Diaphragma, den Herzbeutel oder die Pleura hineingewachsen ist.
- Ich überprüfe, ob ein Lungenkollaps oder eine Pneumonie droht.
- Ich überprüfe, ob zwei oder mehr Tumore im selben Lungenflügel vorhanden sind.

Vollständiges ansehen des kompletten Datensatzes.

sorgfältige Begutachtung, Diskussion mit Kollegen

Erneute systematische Durchsicht des Befundes

Selbst nochmals nachmessen

Den ganzen Datensatz anschauen

(T1 vs. T3 ist meist eine Grössenmessung, die wahrscheinlichkeit ist relativ gering, dass dies passiert. Es geht eher um eine kleine, neue Läsion, die im Verlauf gesehen wird und dann deutlich grösser ist. Der Patient hätte vorher schon eine andere Behandlung bekommen können).

überlegen welche Organe primär/häufig befallen sind

alles anschauen

klinische Informationen ausnutzen (und danach suchen)

Das TNM Stadium wird eigentlich nicht direkt angegeben sondern eher ob es suspekte Lymphknoten (N) oder Fernmetastasen (M) gibt

gucken ob angrenzende Strukturen infiltriert wurden, davor nachlesen wie T3 in der Tumorentität genau definiert ist

Läsion in 3 Ebenen genau Anschauen / Messen.

Wie gehen sie vor, wenn sie sich beim Bestimmen des TNM nicht ganz sicher sind? Wie transparent ist dies für behandelnde Ärzte?

9 responses

Ich kommuniziere die Unsicherheit im Bericht und am Tumorboard.

Ich erstelle einen ausführlichen Befundungsbericht.

Fragen eines Kollegen. Nicht transparent.

s.o. und Literatur

Rücksprache mit Oberarzt, telefonische Mitteilung an Zuweiser und entsprechende Formulierung in der Beurteilung.

dies wird nicht dokumentiert

Bei zB Rektumkarzinom werden softwaregestützt alle relevanten, anatomischen Strukturen/für die TNM Klassifikation wichtigen Informationen Schritt für Schritt als Pflichtfelder abgefragt

Kollegen, Oberarzt fragen

TNM wird bei uns nicht im Bericht festgehalten, sondern nur die entsprechenden Angaben zu den LK / die Grösse... welche eine Einteilung nach TNM erlauben.

Stellen sie sich vor, T3 ist das Resultat einer Bestimmung. Wie interessant sind die weiteren Läsionen, welche z.B. nur zu einem T1 geführt hätten?

8 responses

Irrelevant.

Verstehe die Frage nicht.

Für das Staging irrelevant. Es ist aber zu beachten: ein niedriger eingestufte Tumor kann trotzdem Behandlungsrelevant sein- Beispiel: T4 Tumor; Zusätzlicher "nur" T3-Tumor, der das Zwerchfell infiltriert.

Müssen ebenfalls erwähnt werden, jedoch untergeordnet

recht interressant

ich glaube nach ein paar Fragen Missverständnisse bei der TNM Klassifikation zu finden
T steht für die grösse des Tumors bzw. sein lokales Wachstumsverhalten (Invasion von Gefäss, anderen Organen, etc). Die Frage macht so keinen Sinn, da andere Läsionen nicht in T sondern in N oder M einfließen

interessant da man sie zum Beispiel als Findings in MINT Lesion markieren kann

Zentral - können 2. Tumore oder Metastasen sein - das ändert alles.

Welche Aspekte des NSCLC Stagings empfinden sie als eher mühsam?

9 responses

Zeitdruck.

Das Ausmessen der Läsionen.

Patienten mit multiplen Läsionen und v.a. deren Beschreibung.

Ausmessen, Vergleich mit Voruntersuchungen

Metastasensuche

Beurteilung von LK und Infiltration

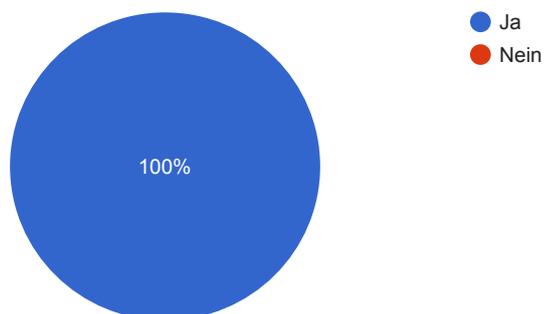
wann ein N Stadium

Alle fraglich suspekten Noduli beschreiben, vermessen, im Follow Up wieder suchen, etc

Screening der Lunge nach weiteren Noduli

Ich wünsche mir automatische Unterstützung beim Staging von NSCLC

10 responses



Welche Form von automatischer Unterstützung wäre wünschenswert beim Staging von NSCLC?

10 responses

System findet Läsionen (schnell!).

Software-Tool, dass sehr rasch (real-time!) alle Läsionen findet und segmentiert.

Automatisches Reporting. Sinnvolle Darstellung aller Läsionen

Anzeige von Hinweisen auf Abnormalitäten, Irregularitäten

Messung, Vergleich mit Voruntersuchungen

Ankreuzsystem mit Auswertung

Detection, classifizierung

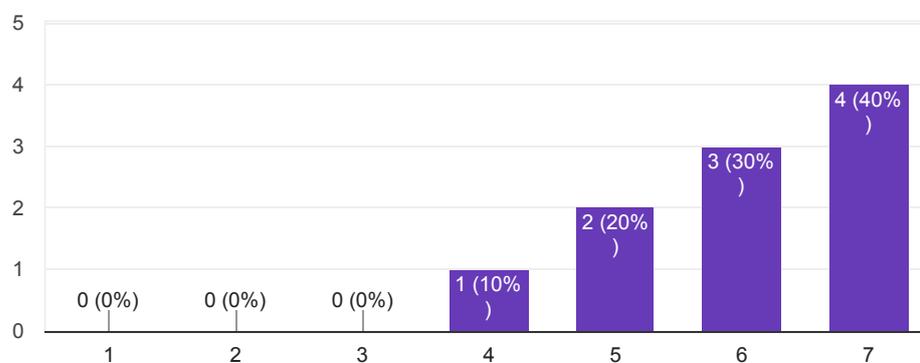
eine Schritt für Schritt Durchführung, angepasst an die Guidelines

Automatische Segmentierung/Volumetrie von Target Lesions ohne dass man sie manuell "einzeichnen"/ die Kontur umfahren muss

System das auf gewisse Dinge Aufmerksam macht und diese zur weiteren Beurteilung anzeigt. Es sollte ersichtlich sein, aufgrund welcher Läsion welches Stadium gegeben wurde.

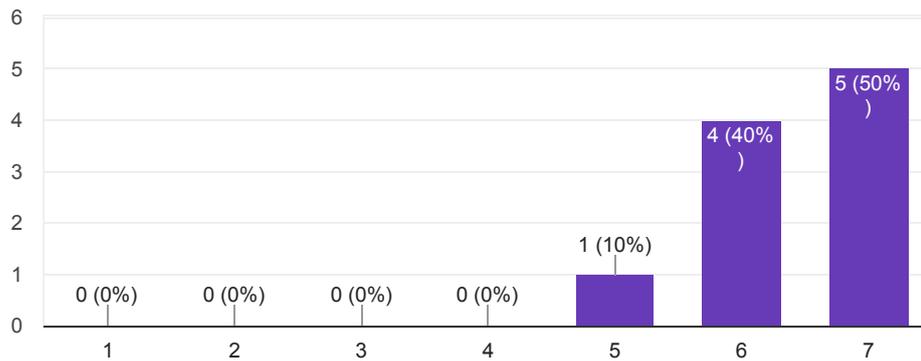
Die axiale Ebene ist die wichtigste PET/CT Bildansicht zur Erkennung von NSCLC

10 responses



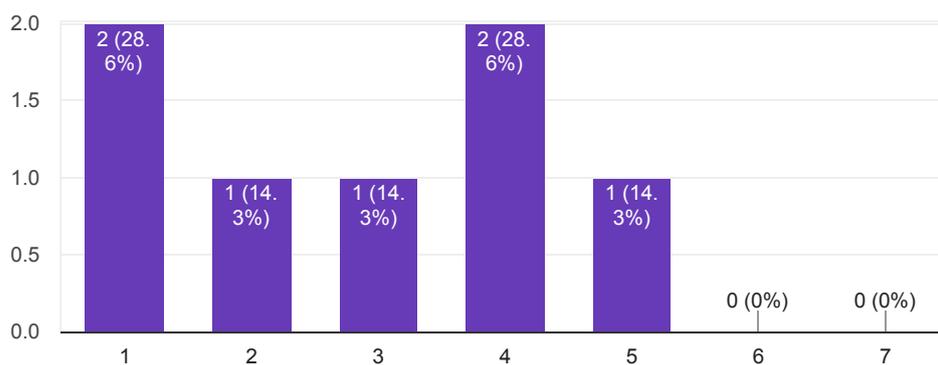
In der axialen Ebene erkenne ich Anomalien einfacher, wenn ich durch mehrere Slices hin- und her wechseln kann (z.B durch scrollen).

10 responses



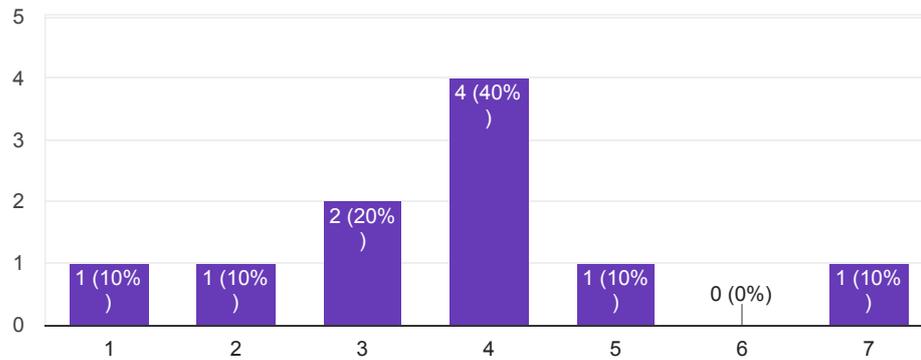
Weitere Ebenen (sagittal, frontal) benutze ich eher selten, für die Befundung.

7 responses



Eine 3D Ansicht wäre hilfreich, um Anomalien besser zu erkennen

10 responses



Warum könnte eine 3D Ansicht der PET/CT Daten hilfreich bzw. nicht hilfreich sein?

5 responses

Läsionen weniger oft übersehen.

Der überlap verhindert die eindeutige Beurteilung

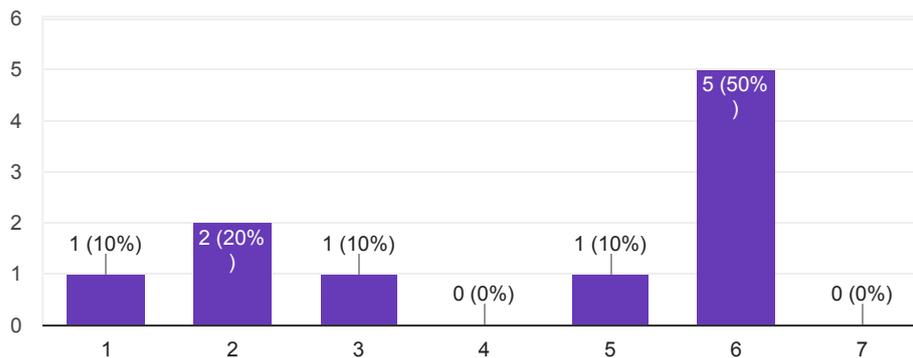
für die Befundung eher weniger wichtig, für die Befunddemonstration an Kollegen z.B. in Rapporten sehr hilfreich

Um in anatomisch unübersichtlichen Regionen eine anschaulichere Darstellung zu haben, Lagebeziehung zu Bronchien, Gefässen etc

Sag und Cor ist für die Beurteilung der Lungenlappen zentral wichtig

Die genaue Kontur einer Läsion ist nicht so relevant, eine abstrakte Einzeichnung der Region of Interest genügt zur Erkennung im Bild.

10 responses



Präzisieren sie, warum die genaue Kontur relevant bzw. irrelevant sein könnte (optional)

7 responses

Ich kann mir nicht vorstellen, dass eine Software in der Lage ist, eine genaue Kontur zu bestimmen.

Relevant nur zur genauen Bestimmung des max. T-Stadiums.

Je nachdem was in der Umgebung infiltriert wird

Verlauf.

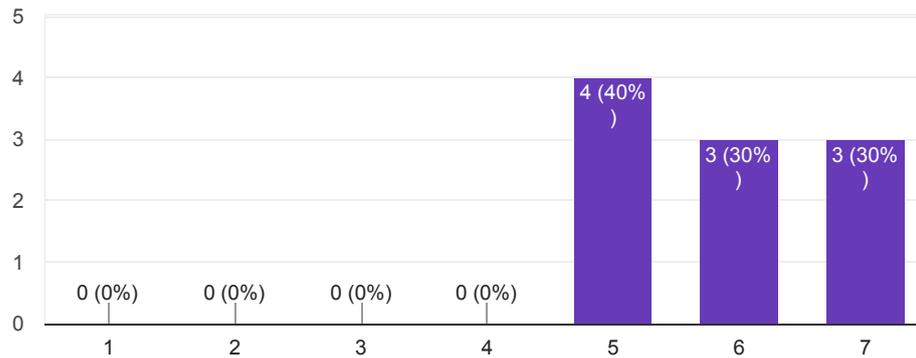
ist relevantes Kriterium für das staging

Kontur weist oft auf Dignität hin, irregulär ist zum Beispiel ein Hinweis auf maligne Läsion

Irrelevant zur Erkennung - jedoch sehr relevant zur Beurteilung der Infiltration.

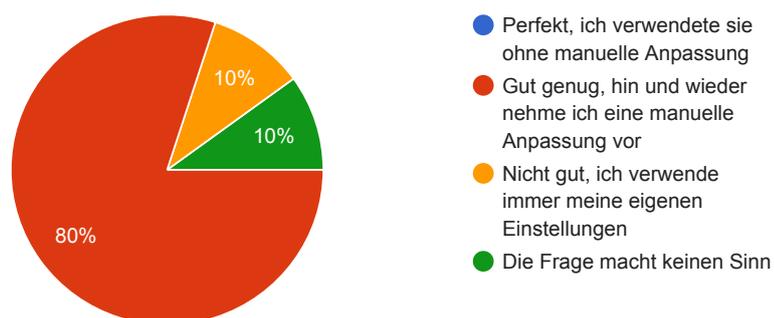
Ein standardisierter Workflow zum Staging von NSCLC würde mir und meinem Team helfen und Objektivität und Transparenz steigern

10 responses



Wie nützlich sind die standard PET/CT Windows in ihrem PACS Viewer? (Soft Tissue, Lung, LungNodes, Bone, etc)

10 responses



12.8 Themenrecherche Quellen

ID	Titel/Link	Autor(en)	Veröffentlichung	Zweck & Zielgruppe	Format	Relevanz	Nützlichkeit für LungStage
1	Can We Design Trust Between Humans and Artificial Intelligence?	Patrick Mankins	06.19.15	Designer und Entwickler	Onlineartikel	Vertrauen in Technik (AI)	Nicht bewertet
2	Artikel zum Thema 'Akzeptanz von AI-Systemen': Power to the People: The Role of Humans in Interactive Machine Learning	Saleema Amershi, Maya Cakmak, W. Bradley Knox, Todd Kulesza	01.12.2014	Designer, Researcher, Entwickler	Onlineartikel	Case studies zu Benutzerverhalten mit Machine Learning. Mit dem Ziel eine bessere user experience zu gestalten. Nicht viel zu Trust	Tiefe Relevanz
3	People dump AI advisors that give bad advice, while they forgive humans for doing the same	Michael J. Coren	27.05.2016	Designer, Researcher	Onlineartikel	Studie an der University of Wisconsin. Vergleich zu Vertrauen zu Menschen und zu AI	Hohe Relevanz
4	"Why Should I Trust You?" Explaining the Predictions of Any Classifier	Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin	09.08.2016	Designer, Researcher, Entwickler	Onlineartikel	Thema ist, wie man Vertrauen in AI Vorhersagen schaffen kann. Es wird das der Algorithmus LIME vorgeschlagen, der Vorhersagen erklären kann.	Hohe Relevanz
5	Keeping humans in the loop in an AI-driven world	Nick Ismail	12.12.2016	Designer, Researcher, Entwickler	Onlineartikel	Behandelt EU Datenschutz-Grundverordnung und das "Recht auf Erläuterung". Und Akzeptanz AI zu verwenden. Eher oberflächlich	Hohe Relevanz
6	AI can excel at medical diagnosis, but the harder task is to win hearts and minds first	Owen A. Johnson	12.08.2016	Entwickler	Onlineartikel	Betsehende Lösungen in der Medizin Technik	Tiefe Relevanz
7	Machine Learning with Humans in the Loop—Lessons from StitchFix	Jared Polivka	13.02.2017	Entwickler, Researcher	Onlineartikel	Erklärt wie sie bei StitchFix den Benutzer in den Prozess involvieren. Keine Infos zu vertrauen	Tiefe Relevanz
8	Advanced Visualization Critical to Health IT Data Analytics	Elizabeth O'Dowd	27.03.2017	Researcher	Onlineartikel	Wichtigkeit von Visualisierungen in der Medizinbranche. Kein Bezug zu Vertrauen und AI	Tiefe Relevanz
UI/UX für AI-Systeme:							
10	Towards Better Analysis of Machine Learning Models: A Visual Analytics Perspective	Shixia Liu, Xiting Wang, Mengchen Liu, Jun Zhu	04.02.2017	Entwickler	Onlineartikel	Für unseren Usecase Zu technisch	Tiefe Relevanz
11	Using Visual Analytics to Interpret Predictive Machine Learning Models	Josua Krause, Adam Perer, Enrico Bertini	17.08.2016	Entwickler	Onlineartikel	Für unseren Usecase Zu technisch	Tiefe Relevanz
12	A visual introduction to machine learning	Stephanie Yee, Tony Chu	2015	Entwickler	Onlineartikel	Datengrafische aufbereitung von Decision Trees, zu technisch	Tiefe Relevanz
13	An Explanation-Centric Approach for Personalizing Intelligent Agents	Todd Kulesza	14.02.2017	Designer, Entwickler, Researcher	Onlineartikel	Analyse und Vorschlag für UI, wie Prognosen in AI und Machine Learning beschrieben werden können.	Hohe Relevanz
15	Explainable Artificial Intelligence (XAI)	David Gunning					Nicht bewertet
16	Supporting Trust in Autonomous Driving	Renate Haeuselichmid, Max von Buelow, Bastian Pfleging, Andreas Butz	7.3.2017	Researcher			Nicht bewertet

12.9 User Stories

ID	Status	Typ	Projekt-Risiko/Unsicherheit	Business-Priorität	Projekt-Priorität	Priorität Total	User Story	Beschreibung	Quelle
A10	Akzeptiert	Nicht funktional	Hoch	Kritisch	Kritisch	10.5	Als Radiologe möchte ich Einblick in die automatische Bestimmung des TNM haben, um der AI vertrauen zu können und nicht alle Bilddaten selber anschauen zu müssen	Die Ärzte sollen Vertrauen haben in das automatisch bestimmte TNM	Business, C2-23
A13	Akzeptiert	Nicht funktional	Hoch	Kritisch	Kritisch	10.5	Als Radiologe möchte ich ein objektiveres Staging, um individuellen Bias und negative Folgen für die Therapie zu vermeiden	Das System soll individuellen Bias vermeiden und Objektivität und Transparenz fördern beim Staging	Business, C2-31
A14	Akzeptiert	Nicht funktional	Hoch	Kritisch	Kritisch	10.5	Als Radiologe möchte ich bei der Diagnose unterstützt werden, um Fehldiagnosen zu vermeiden	Das System soll helfen, Fehldiagnosen zu vermeiden	Business, C2-3, C2-31
A27	Akzeptiert	Nicht funktional	Hoch	Kritisch	Kritisch	10.5	Als Radiologe möchte ich mit der AI interagieren können, um besser zu verstehen welchen Effekt meine Eingaben auf das Resultat haben	Der Benutzer kann durch Interaktion mit der AI lernen wie sie funktioniert	Business, C2-30, Business
A11	Akzeptiert	Funktional	Mittel	Kritisch	Kritisch	9.5	Als Radiologe möchte ich einen schnellen Überblick über die Krebs Regionen im Körper haben, um die Krankheit effektiv zu diagnostizieren	Das User Interface soll einen Überblick über die betroffenen Krebs Regionen bieten	Business, C2-25, C2-27, C2-2
A12	Akzeptiert	Funktional	Mittel	Kritisch	Kritisch	9.5	Als Radiologe möchte ich Detailinformationen zu den Krebs Regionen haben, um bei Unsicherheiten ein klareres Bild zu bekommen	Das User Interface soll detaillierte Informationen zu den betroffenen Krebs Regionen anbieten	Business, C2-24, C2-26, C2-15, C2-27
A18	Akzeptiert	Funktional	Mittel	Kritisch	Kritisch	9.5	Als Radiologe möchte ich mit den Slices interagieren, um durch die Bewegung zwischen den Bildern das Resultat der AI besser verifizieren zu können	Der Benutzer soll mit der Serie von Slices interagieren können	C2-14
A25	Akzeptiert	Funktional	Mittel	Kritisch	Kritisch	9.5	Als Radiologe möchte ich die Konfidenz der AI in die eigene Diagnose sehen, um das Resultat kritisch hinterfragen zu können	Das UI weist die Konfidenz in das Resultat als Wahrscheinlichkeit aus	C2-28
A26	Akzeptiert	Funktional	Mittel	Kritisch	Kritisch	9.5	Als Radiologe möchte ich Korrekturen am Resultat der AI erkennen und ausweisen, um transparent zu machen, wie die Diagnose zustande kommt	Korrekturen des automatisierten Resultats durch den Arzt sind klar ersichtlich	C2-29
A8	Akzeptiert	Nicht funktional	Hoch	Kritisch	Wichtig	9	Als Radiologe will ich ein schnelles Staging, um den hohen Workload bewältigen zu können	Das Staging von NSCLC soll beschleunigt werden (messbares Ziel?)	Business, C2-5
A1	Akzeptiert	Nicht funktional	Tief	Kritisch	Kritisch	8.5	Als Hersteller möchte ich ein eigenständiges Tool, um die Zulassung dessen vor der Integration in bestehende Systeme sicherzustellen	LungStage soll als eigenständiges Tool laufen und nicht im PACS UI integriert sein	Business
A6	Akzeptiert	Funktional	Tief	Kritisch	Kritisch	8.5	Als Radiologe möchte ich von der KI gesetzte Parameter manuell überschreiben können, um die AI zu korrigieren und weil ich letztendlich die Verantwortung für das Resultat trage	Vom System bestimmte Parameter müssen manuell überschrieben werden können	C2-6
A17	Akzeptiert	Funktional	Tief	Kritisch	Kritisch	8.5	Als Radiologe möchte ich die erkannten Krebs Regionen im Bild klar erkennen, um das Resultat der AI schnell verifizieren zu können	Eine Krebs Region muss im Bild klar erkennbar sein	C2-13

ID	Status	Typ	Projekt-Risiko/Unsicherheit	Business-Priorität	Projekt-Priorität	Priorität Total	User Story	Beschreibung	Quelle
A2	Akzeptiert	Funktional	Mittel	Kritisch	Wichtig	8	Als Radiologe möchte ich immer sehen, für welchen Patienten ich das Staging gerade durchführe, um Verwechslungen zu vermeiden	Es muss klar sein, welcher Patient gerade gestaged wird	C2-19
A21	Akzeptiert	Nicht funktional	Mittel	Kritisch	Wichtig	8	Als Radiologe möchte ich ein User Interface, das gut in dunklen Umgebungen funktioniert, um im Reading Room ergonomisch arbeiten zu können	Das UI muss in den verdunkelten Reading-Rooms gut funktionieren	C2-18, C2-32
A23	Akzeptiert	Funktional	Mittel	Kritisch	Wichtig	8	Als Nuklearmediziner möchte ich die Fusion von PET und CT Bildern, um durch die stoffwechselaktiven Regionen die Krankheit besser einschätzen zu können	Das UI muss die Fusion von PET und CT Bildern ermöglichen	C2-21
A9	Akzeptiert	Funktional	Tief	Kritisch	Wichtig	7	Als Radiologe möchte ich eine automatische Bestimmung des TNM in den Bilddaten, um ein Staging effizient durchzuführen und mehr Objektivität über alle Stagings insgesamt zu haben	Das TNM soll automatisch bestimmt werden	Business, C2-1, C2-2
A28	Akzeptiert	Nicht funktional	Mittel	Wichtig	Wichtig	7	Als Radiologe möchte ich bei Unterbrechungen unterstützt werden, um meine Arbeit fehlerfrei fortführen zu können	Das UI soll unterstützen, um nach Unterbrechungen die Arbeit wieder aufzunehmen	C2-32
A19	Akzeptiert	Funktional	Mittel	Nützlich	Wichtig	6	Als Radiologe möchte ich mit den gewohnten Ansichten arbeiten, um mein damit geschultes Auge nutzen zu können	Das UI soll die Oben-nach-unten Ansicht vorziehen, jedoch müssen zusätzliche Ansichten möglich sein	C2-16
A3	Akzeptiert	Nicht funktional	Tief	Kritisch	Nützlich	5.5	Als Radiologe möchte ich ein performantes und stabiles Tool, um keine Zeit zu verlieren	LungStage soll performant und stabil laufen	C2-5
A4	Akzeptiert	Nicht funktional	Tief	Kritisch	Nützlich	5.5	Als Radiologe möchte ich LungStage schnell erkennen, um während der Arbeit mit verschiedenen Tools schnell zum richtigen Tool wechseln zu können	Das UI soll klar als LungStage erkennbar sein	C2-4, C2-12
A5	Akzeptiert	Funktional	Tief	Kritisch	Nützlich	5.5	Als Radiologe möchte ich mit den Daten aus bestehenden Systemen arbeiten können, um doppelte Eingaben zu vermeiden und Zeit zu sparen	Patientendaten müssen aus bestehenden Systemen geladen werden können	Business
A20	Akzeptiert	Funktional	Tief	Nützlich	Wichtig	5	Als Radiologe möchte ich mit Maus, Scrollrad und Keyboard Shortcuts arbeiten, um wie in den bestehenden Tools arbeiten zu können	Das UI soll mit Maus, Scrollrad und Keyboard Shortcuts bedient werden können	C2-17
A24	Akzeptiert	Funktional	Mittel	Nützlich	Nützlich	4.5	Als Nuklearmediziner möchte ich PET-spezifische Parameter in den Details sehen, um die PET Bilder besser verstehen zu können	Das UI muss die PET spezifischen Parameter zusätzlich anzeigen können	C2-22, C2-24
A7	Akzeptiert	Funktional	Tief	Nützlich	Nützlich	3.5	Als Radiologe möchte ich Voreinstellungen im User Interface bearbeiten können, um sie meiner persönlichen Arbeitsweise anpassen zu können	Voreinstellungen im User Interface müssen bearbeitbar sein	C2-8, C2-7

ID	Status	Typ	Projekt-Risiko/Unsicherheit	Business-Priorität	Projekt-Priorität	Priorität Total	User Story	Beschreibung	Quelle
A16	Akzeptiert	Funktional	Tief	Nützlich	Nützlich	3.5	Als Radiologe möchte ich die Diagnose speichern können, um sie später zu verwenden und anderen Radiologen zugänglich zu machen	Das Befundungsergebnis in LungStage kann gespeichert werden	C2-10
A15	Vorschlag	Out of scope						Nebenbefundungen werden nicht in LungStage durchgeführt	C2-11
A22	Vorschlag	Out of scope						Eine 3D-Ansicht wird nicht von LungStage bereitgestellt	C2-20
A29	Vorschlag	Out of scope						Entscheidungen aus dem Tumorboard werden nicht in LungStage zurückgeführt	C2-9
A30	Vorschlag	Out of scope					Als Radiologe möchte ich ein im PACS integriertes Tool, um nicht noch zwischen einem weiteren zusätzlichen Tool switchen zu müssen	LungStage soll im PACS integriert werden, möchten nicht noch ein zusätzliches Tool öffnen	C2-4

12.10 Persona

Dr. med. Roman Stebler



Alter	36
Lebt in	Pratteln
Wohnverhältnis	Lebt in einer Mietwohnung mit seiner Frau und seiner 1 Jahr alten Tochter
Beruf	Doppelfacharzt für Radiologie und Nuklearmedizin
Technologie & Tools	Er ist technikaffin, sein Arbeitsplatz besteht aus 3 Screens. Nebst Applikationen für administrative Aufgaben ist er es gewohnt mit komplexer Expertensoftware zur Analyse und Auswertung von PET/CT Daten zu arbeiten und Krankheiten zu erkennen. Ein Diktiergerät dient zur Aufzeichnung von Befundungen. Das Festnetztelefon dient zur Kommunikation innerhalb des Spitals. Sein persönliches Smartphone verwendet er für die Kommunikation mit Freunden und Familie.
Interessen & Hobbys	Er fährt jeden Tag mit dem Velo 11 km zur Arbeit um fit zu bleiben. Er möchte gerne Zeit mit seiner Familie verbringen und deshalb nicht zu spät nach Hause kommen. Durch einen langen Arbeitstag ist dies jedoch nicht immer möglich. Generell ist die Work-Life-Balance ein schwieriges Thema. Die Arbeitstage sind lang und die Wochenenden würde er auch gerne für die Forschung verwenden, da dafür im Arbeitsalltag kaum Zeit bleibt.

Motivation & Ziele	Sein Ziel ist es schnell und exakt das Stadium von Lungentumoren anhand der Bilddaten zu bestimmen. Er möchte wiederkehrende Probleme in der Zukunft mit einer höheren Sicherheit bestimmen können. Dabei will er aber immer auch den Patienten als Ganzes sehen. Auch eine reibungslose Kommunikation mit behandelnden Ärzten, welche er als seine Kunden sieht, ist wichtig für ihn. Er beteiligt sich aktiv an der Radiologieforschung und möchte neue Erkenntnisse finden und publizieren um seine Arbeit qualitativ zu verbessern und zu erleichtern.
Needs	Er braucht eine leistungsstarke Software, die ihn dabei unterstützt, die Befundung lückenlos, möglichst objektiv und schnell zu erstellen und einen nachvollziehbaren Befund für den Patienten und die behandelnden Ärzte zu erstellen. Er wünscht sich, dass die Software genug Flexibilität bietet, um schnell einen Überblick zu erhalten, aber auch detaillierte Probleme zu können.
Pain points	Das Wissen, dass eine fehlerhafte Befundung oder eine übersehene Läsion lebensbedrohende Konsequenzen haben kann, setzt ihn unter Druck. Zeitdruck bei der Befundung und lange Arbeitstage von 10-12 Stunden können Fehler bei der Befundung provozieren. Software, die langsam ist oder häufiger abstürzt, nervt ihn. Häufige Unterbrechungen durch Telefonate oder andere Personen im Raum sind eine weitere Quelle für mögliche Fehler und Unkonzentriertheiten.

12.11 Kontext Szenario

LungStage Szenario

Das nachfolgende Szenario beschreibt den Ist-Zustand der Befundung von Lungenkrebs. Stand Juni 2017.

Roman kommt soeben vom Mittagessen aus der Kantine zurück und macht sich nun an die Befundung seines nächsten Patienten. Als erstes startet er den **Ladevorgang der CT-Bilder** in MintLesion, weil er weiss dass dies einige Minuten dauern kann. In der Zwischenzeit liest er sich die **Diagnoseliste und die Fragestellung aus Ismed** durch. Dies gibt ihm einen groben Überblick über den Patienten sowie erste Hinweise worauf er bei der Befundung achten muss.

Für die Befundung öffnet er vorgängig ein **Word-Template**, in welcher er seine Erkenntnisse während der Analyse festhält. Nachdem die Bilder geladen sind, wählt er in MintLesion den **Workflow** für Lungenkrebs aus und beginnt durch die CT-Bilder zu scrollen. Sein Ziel ist die Bestimmung des TNM, mit welchem das Stadium des Tumors definiert wird. Dabei folgt er einem bestimmten Muster, um sicherzustellen, dass er nichts übersieht.

Zuerst identifiziert er den **Primärtumor** und markiert diesen auf dem Bild, auf welchem dieser am besten erkennbar ist. Die Software errechnet daraus automatisch den Durchmesser und bestimmt den Wert für "**T**" (Primary Tumor), welchen er kurzerhand akzeptiert. Anschliessend schaut er sich die **Lymphknoten** an, um herauszufinden, ob Metastasen vorhanden sind. Daraus bestimmt er das "**N**" (Regional Lymph Nodes). Um die Erkennung von Metastasen (ROI - Region of interest) zu erleichtern, ändert er zwischendurch das **Windowing** (Grauwerte der Bilder), wodurch die Strukturen besser erkennbar werden und er diese einfacher **einzeichnen** kann. Nach der Analyse der Metastasen ist er bereits in der Lage zu bestimmen, ob Fernmetastasen vorhanden sind. Dadurch wird das "**M**" (Distant Metastasis) definiert.

Während der Analyse der Bilder setzt er **Bookmarks** für wichtige Bilder und ergänzt laufend den Bericht, um nichts zu vergessen. Neben der Analyse des Lungenkrebs schaut er sich zusätzliche Regionen des Körpers an, um **weitere Angaben** zum medizinischen Zustand des Patienten (z.B. vergessene Schere einer früheren OP) machen zu können. Diese Zusatzinformationen können für die weitere Behandlung von Bedeutung sein.

Die ganze Befundung (inklusive Bericht) dauert ungefähr **40 Minuten**. Bevor er den Bericht abspeichert, bespricht er diesen mit dem leitenden **Oberarzt** und geht dazu nochmals durch die relevanten Bilder durch und erklärt seine Überlegungen.

12.13 Leitfaden Usability Walkthroughs mit Papier Prototyp

Fragen zur Diskussion:

- Welche Daten soll die KI anzeigen?
- Macht Unterteilung in “Analyse” und “Resultat Sinn?
- Vergleich von Slices oder Serien nützlich?
- Kannst du dir vorstellen, die Analyse des gesamten Körpers in “unserem” Tool zu machen oder nur das LungStaging?
- Möchtest du zuerst eine Übersicht (des KI-Resultats) oder lieber mit den Details (Analyse) beginnen?
- Sind Text oder Bilder nützlicher für die Beschreibung der Analyse der KI?
- Konfidenz der KI anzeigen?
- Ähnliche Fälle anzeigen? Was könnten Kriterien für ähnliche Fälle sein?
- Kollaborativer Ansatz interessant? (z.B. Fach und Oberarzt arbeiten zusammen am gleichen Tool oder arbeiten sogar remote via Internet)
- Ist es sinnvoll die Befunde der KI zu bewerten (z.B. mit “Akzeptieren”, “Unsicher”, “Falsch”)? Oder lieber direkt die Befunde einzeichnen (implizite Bewertung)?
- Würdest du dem UI (bzw. den Skizzen) vertrauen? Was müsste geändert werden, dass du dem Resultat vertraust, bzw. wie könnte dein Vertrauen schneller aufgebaut werden?

12.14 Ergebnisliste Usability Walkthroughs mit Papier Prototyp

ID	TP	Finding	Kontext
1	1	Er findet, dass ein eigenständiges Tool schwierig ist, versteht aber, dass dies der einzige Weg ist für 4Quant	Strukturell
2	1	Er möchte die Ansicht selber einstellen können (das UI soll voll konfigurierbar sein). z.B. viel Info/wenig Info/Keine Infos, Anzahl Fenster ändern	Funktionell
3	1	Möchte möglichst wenig Infos sehen auf den Bildern, das lenkt ihn eher ab, bzw. die Möglichkeit, Annotationen schnell aus- und einzublenden	Strukturell
4	1	Bei Syngo.via findet er gut, dass die Tools erst on Hover erscheinen und dadurch das UI nicht überladen wird	Funktionell
5	1	Kennt die folgenden Interaktionsmöglichkeiten von anderen Applikationen: - Linksklick zum scrollen - Rechtsklick für Pan - [Space] + Scrollrad für Zoom - Kontextmenü für weitere Funktionen, z.B. Window einstellen	Funktionell
6	1	Nur TNM bringt wenig ("Niemand bezahlt Millionen für so was"). TNM ist nur ein Puzzleteil von vielen. Alter, Blutwerte, etc. sind weitere Puzzle-Teile, welche beachtet werden müssen um einen klaren Mehrwert zu bieten	Strukturell
7		Von 12 Fällen an einem Tag ist das TNM nur bei ca. 2 Fällen relevant	Strukturell
8	1	Ein erfahrener Arzt benötigt 5-10 Minuten für eine Befundung (schwierige Fälle benötigen 15min). Auch wenn das automatische Staging in 2 Minuten vorliegt, bringt dies keinen grossen Mehrwert.	Strukturell
9	1	Den Tumor finden ist nicht schwierig und ob es nun 7 oder 12 Läsionen sind ist für den weiteren Verlauf der Behandlung nicht so relevant	Strukturell
10	1	Oft spielen andere Faktoren eine grössere Rolle. Z.B. kann wegen einer Entzündung nicht operiert werden oder die Frau des Patienten will eine Zweitmeinung, bevor er operiert wird	Strukturell
11	1	Integration von Wissensdatenbanken oder Datenbanken mit früheren Fällen wäre ein klarer Mehrwert, weil dies der Arzt nicht so gut machen kann (ist von seiner Erfahrung abhängig)	Funktionell
12	1	Als Second-Reader ist LungStage eine super Idee und dient definitiv der Verbesserung der Diagnosequalität	Strukturell
13	1	Ein grosser Mehrwert könnte sein, wenn LungStage auch Entzündungen oder andere Erkrankungen erkennen könnte.	Strukturell
14	1	Workflow: Er würde LungStage zuerst als Hintergrundtool/Zweitmeinung verwenden	Strukturell
15	1	Workflow: Verschafft sich zuerst einen Überblick in bestehendem Tool (den ganzen Körper anschauen)	Strukturell
16	1	Workflow: Main Findings von LungStage anschauen und die erkannten Werte (z.B. Durchmesser des Tumors) für den Bericht diktieren	Strukturell
17	1	Workflow: Je mehr er LungStage vertraut, um so mehr würde er sich nur noch die Resultate anschauen und direkt den Bericht diktieren. Er meint, nach ungefähr 6 Monaten hätte er genügend Vertrauen in LungStage.	Strukturell
18	1	Für die Analyse der unterschiedlichen Organe sind unterschiedliche Ansichten sinnvoll (z.B. ist eine grosse Darstellung der Lunge am effizientesten, aber für die Niere benötigt es mehrere Ansichten/Fenster mit CT/PET etc.)	Strukturell
19	1	Er möchte mit Keyboard-Shortcuts zwischen den Ansichten wechseln können (z.B. Taste "L" für Lunge) oder sogar den ganzen Workflow für sich selbst konfigurieren können	Funktionell
20	1	Die meisten Metastasen sind Stammnah, d.h. im Oberkörper angesiedelt. Somit benötigt es keine Ganz-Körperdarstellung für den In-Body-Browser	Strukturell
21	1	Kollaborativer Ansatz innerhalb der Klinik findet er spannend (z.B. um einem Kollegen, der in einem anderen Raum oder Stockwerk sitzt, eine Frage stellen zu können). Eine Kollaboration zwischen Kliniken (z.B. Basel und Zürich) funktioniert aus "politischen Gründen" nicht.	Funktionell
22	1	Hot-Corners oder Regionen-basierte Interaktion, um z.B. Menus dazu zu holen	Funktionell
23	1	Einstellungsmöglichkeit um Menge der Annotationen zu steuern: lo/mid/hi, da er nicht immer alle Infos braucht. Ein schneller Weg, dies einzustellen fände er super	Funktionell
24	1	Michael scheint weniger voreingenommen zu sein vom Setup im USB, findet TNM an sich nicht so wichtig	Strukturell
25	1	Findet Know-How aus Kollektiv der Fälle interessant	Strukturell
26	1	Möchte mit der Software für das Outcome optimieren, nicht für das Stadium	Strukturell
27	1	Möchte Dinge sehen, die er von Auge nicht erkennen kann, z.B. Entzündungshinweise für die Kliniker	Funktionell
28	1	Gärtchendenken im USZ: bekommt auf den Deckel, wenn er gewisse Leute um Hilfe fragt	
29	1	Bestätigte / Verworfenen Läsionen müssen im UI veränderbar sein im Nachhinein	Funktionell
30	1	Könnte sich vorstellen, aus dem UI die Main-Findings anhand der Bilder nur noch in den Befund zu diktieren	Strukturell
31	1	Mit der Zeit und genügend Vertrauen würde vielleicht sogar er nur noch abdiktieren	Strukturell
32	1	Je nach Serie braucht es ein anderes Window	Funktionell
33	1	Könnte sich eine Art Team-Viewer vorstellen, um Findings schnell einem Kollegen zu zeigen	Funktionell
34	1	"Es braucht den Menschen, um den Menschen zu verstehen"	

ID	TP	Finding	Kontext
1	2	Findet es gut, da die wichtigste Infos ist in der Mitte und Basic Infos links oben sind	Strukturell
2	2	Versteht beide PET/CT Bilder so, dass die 2 Bilder gekoppelt sind: zeigen immer das gleiche Bild, man kann beides entkoppeln. Wenn die Kopplung gut ist, macht ein Entkoppeln nicht sinn	Strukturell
3	2	Es reichen 2 axiale Bilder, andere Ansichten sind nur Zusatzinfos	Strukturell
4	2	Sollten 2 Bilder angezeigt werden oder 1? Man könnte sie auch "mergen" aber er meint es ist besser je ein CET und eine PET Ansicht zu haben	Strukturell
5	2	Bei Farben und Überlagerungen der PET aufpassen, dass man die Läsionen nicht übermalt, eventuell mit Kästchen lösen	Visuell
6	2	Koronale Ansicht: Serie macht mehr Sinn als Grafik, wär gut wenn man dort navigieren kann	Strukturell
7	2	Es braucht keine sagitale Ansicht	Strukturell
8	2	Window: möchte lieber die Namen verwenden (z.B. "Softtissue Lung") um schnell zu wechseln	Strukturell
9	2	Bei drüberhovern Einstellungen anzeigen	Funktionell
10	2	Man sollte es auch manuell "schärfer" stellen können	Funktionell
11	2	Versteht das Icon des "Einzeichnen"-Tool nicht	Visuell
12	2	Versteht "Save/Close" Button nicht. Versteht nicht ob es ein Zwischenspeicher ist oder direkt ins Pax speichert. Möchte 3 Buttons (evt. mit Icons): "Exit and discard", "Save and Send", "Pause" (Zwischenspeichern)	Funktionell
13	2	TNM ist für ihn die Hauptinfo und sollte prominenter dargestellt werden	Visuell
14	2	Infos zu erkannter Läsion direkt im Hauptbild zeigen und nicht an der Seite. Info-Kästchen lösen, da man das so in anderen Tools gewöhnt ist. Man muss sonst immer hin- und herschauen.	Strukturell
15	2	Infos: Was passiert mit dem Info-Kästchen, wenn man die Läsion nicht mehr sieht?	Strukturell
16	2	Infos: Art der Läsion (TNM), Ort (z.B. rechter Lungenflügel), Durchmesser, Feature (z.B. Infiltration Mediastinum, etc (Brustwand), Kollaps von...) und SUV Max sollten angezeigt werden	Strukturell
17	2	Infos: Ist SUV Wert relevant? Jeder Nuklearmediziner möchte den SUV Wert sehen, (Stoffwechsel), Info sagt nicht aus ob es Entzündung oder was anderes ist. Kein klarer Cut-off kann aber ein Indikator sein.	Strukturell
18	2	Infos: Kombination aus Text und Bild besser, aber platztechnisch schwierig	Visuell
19	2	Infos: Möchte keinen Fliesstext, braucht z.B. kein Zusatztext "Primärtumor" bei "T"	Visuell
20	2	Möchte Entscheidungsmöglichkeiten ob das Resultat richtig/falsch ist	Funktionell
21	2	Workflow: Würde zuerst das Resultat von LungStage anschauen, dann die Untersuchung im PACS machen und anschliessend Raport schreiben	Strukturell
22	2	Workflow: Würde zuerst die Zusammenfassung anschauen, zuerst durch T, dann N, dann M durchklicken, Eventuell würde es Sinn machen einen Worklow, zuerst T, N, M draus zu machen	Strukturell
23	2	Workflow: Würde zuerst den entschieden Tumor ansehen	Strukturell
24	2	Vertrauen: würde dem Tool vertrauen.	Strukturell
25	2	Vertrauen: Macht keinen Sinn Trainingsdaten zu sehen, macht den Prozess nur langsamer	Funktionell
26	2	Vertrauen: Man fängt an zu Vertrauen, wenn das Tool gut ist	Strukturell
27	2	Vertrauen: man muss so oder so alle Bilder anschauen, weil es rechtlich Probleme gibt wenn man blind einem Tool vertraut	Strukturell
28	2	Tool hilft die Analyse zu bestärken, Interaktionen sollten kurz sein, und helfen einen guten Report (Textbaustein) zu erstellen	Strukturell
29	2	Vertrauen: Vertrauen könnte bestärkt werden, indem man sagt wieso TNM so entstanden ist und Konfidenz angeben	Strukturell
30	2	Resultat: Es gibt unwahrscheinlichere Kombinationen, "Wahrscheinlicher, da in 90% der Fälle mit der gleichen Konstellation..."	Strukturell
31	2	Resultat: Vernetzung mit Kontextinfos (Alter, Raucher, andere Krankheiten, etc.) wäre gut	Visuell
32	2	Technische Schwierigkeit: Deeplearning funktioniert gut bei 2D Bildern, aber nicht bei 3D. Ist häufig auch weniger klar als die Erkennung einer Katze auf einem Bild	Strukturell
34	2	Befundung: Was passiert wenn ich ein Resultat ablehne ("x" drücken)? Wie kann ich zum Beispiel sagen, dass es ein T2 ist anstelle T4? Eventuell in einem zweiten Schritt	Strukturell
35	2	Befundung: Commentfeld könnte hilfreich sein	Funktionell
36	2	Befundung: Teilen mit Kollegen könnte gut sein	Funktionell
37	2	Befundung: Man müsste sehen wer das Resultat der KI verändert, bzw. abgelehnt hat	Strukturell
38	2	Möchte das Tool nur kurz verwenden müssen um sich einen Überblick zu verschaffen (2 Minuten)	Strukturell

ID	TP	Finding	Kontext
200	3	Infos: möchte wissen ob Pet positiv/negativ, Lymphknotenlevel (römische Zahlen), Ort (z.B. Lungenflügel links), metastasierte Organe	Strukturell
201	3	Möchte Info zu Kernel (z.B. mit Serie verbinden): Knochen/Lunge, Weichteil-Kernel	Strukturell
202	3	Möchte zusätzliche Läsionen einzeichnen können, bzw. erkannte Läsionen anpassen	Funktionell
203	3	Die meisten Tools die er kennt verwenden Doppelklick um ein Bild (in 4er Grid) vollflächig darzustellen	Funktionell
204	3	Er kann sich vorstellen dass Zusatzinfo angezeigt wird, wenn er die Läsion anklickt	Funktionell
205	3	Er möchte Läsion bestätigen oder ablehnen	Funktionell
206	3	Möchte Ansichten wechseln können, so dass er entweder 1 oder 4 Bilder sieht. Initial möchte er die 4er Ansicht sehen und z.B. mit Doppelklick Vollbild anzeigen	Funktionell
207	3	Anpassungen und Änderungen des Resultats der KI möchte er sehen, ihm ist aber egal wer etwas geändert hat	Strukturell
208	3	Er möchte alle Informationen an einem Ort finden	Strukturell
209	3	TNM sollte prominent auf einen Blick sichtbar sein	Visuell
210	3	Workflow: würde LungStage als Second Reader (Zweitmeinung) verwenden. Er geht prinzipiell davon aus, dass Geräte und auch Kollegen Fehler machen können.	Strukturell
211	3	Möchte zuerst eine klare Aussage von LungStage sehen (glaubwürdiger) und alternative Läsionen erst im Nachgang sehen	Strukturell
212	3	Möchte ein schlichtes Interface	Visuell
213	3	Übersicht des Körpers ist nicht wichtig für ihn, aber es würde ihn auch nicht stören. Nukis wollen diese Ansicht bestimmt sehen meint er.	Strukturell
214	3	Syngo.via bietet sehr schlechte Usability (viele Infos werden reingeschoben und das UI verändert sich sehr oft)	Visuell
215	3	Das Design von MintLesion ist sehr angenehm. LungStage sollte auf jeden Fall ein dunkles Interface bieten, weil dies in dunklen Räumen besser funktioniert (z.B. dunkelgrau plus akzentfarbe)	Visuell

12.15 Testszzenarien Usability Walkthroughs mit Wireframes

Intro (3min)

-

Fragen zur Person (2min)

- Name/Alter
- Was ist Ihre Funktion?
- Mit welchem Tool arbeiten Sie normalerweise bei Lungenkrebs-Befundung

Einführung

An ihrem Arbeitsplatz wurde eine neue Software eingeführt, die mittels künstlicher Intelligenz Läsionen erkennt und klassifiziert. Dies soll sie bei der Befundung von Lungenkrebs (NSCLC) unterstützen.

Die Daten im Prototyp entsprechen nicht einem realen Case und er ist noch nicht voll interaktiv. Versuchen Sie den Prototyp so zu verwenden, wie sie es sich von bestehenden Tools gewohnt sind.

Komentieren sie laut, was sie dabei denken und welche Fragen aufkommen während der Benutzung.

Wir möchten mit diesem Walkthrough herausfinden, ob die Abläufe so stimmen und ob die Interaktion nützlich sind. Wichtig: Nicht Sie werden getestet, sondern der Prototyp. Sie können also nichts falsch machen.

Test 1: Überprüfen eines Befundes in LungStage

Sie öffnen den Patient "Markus Deuter". Was sehen sie? Überprüfen sie den Befund von LungStage. Wie gehen sie vor? Verstehen Sie, wie das (TNM)-Ergebnis zustande gekommen ist?

Beobachtungsfokus

- Was soll in der Liste, was im Detail angezeigt werden?
- Welche Farbe ist am sinnvollsten für die Läsionen? Gelb scheint besser als blau.
- Slider vs. hoch/mittel/tief-Buttons? Sensitivität und Spezifität zeigen?
- Spezifität pro T/N/M zeigen oder Angabe über alles?
- Braucht er das MIP bei der Übersicht?
- Wie detailliert sollte der Körper in der Übersicht dargestellt werden?
- Versteht der Benutzer wie sich die Befundung zusammensetzt?
- Sind alle relevanten Informationen vorhanden?
- Vertraut der Benutzer der Befundung von LungStage?
- Interaktion mit UI (erwarten Sie z.B. Vollansicht mit Doppelklick?)
- Frage: Braucht es eine Ganzkörper-Ansicht oder reicht nur der Oberkörper?

Test 2: Korrektur einer Läsion in LungStage

Sie machen das NSCLC Staging für den Patient "Peter Müller". Sie entscheiden sich dies zuerst mit LungStage durchzuführen und danach in PACS zu importieren. Überprüfen sie das Ergebnis und nehmen sie Korrekturen vor. Wie gehen sie vor?

Beobachtungsfokus

- Versteht der Benutzer wie er das automatisierte Ergebnis editieren kann?
- Sind alle wichtigen Informationen vorhanden
- Benötigt er ein Kommentarfeld?
- Kann er den Prozess abschliessen und in's PACS exportieren?
- Features korrekt auf Deutsch übersetzen:
- "Tumor im anderen Lappen"
- "Nodus im gleichen Lappen"

Test 3: Einzeichnen einer neuen Läsion

Sie haben eine neue Läsion entdeckt, die nicht eingezeichnet ist. Sie möchten diese ergänzen. Wie gehen dabei vor?

Beobachtungsfokus

- Versteht der Benutzer wie er eine neue Läsion einzeichnen kann?

Abschlussfragen

- Wie interessiert sind sie an weiteren Läsionen, die gefunden wurden, aber wir aus wegen zu geringer Konfidenz nicht anzeigen?
- Wie kannst du durch den Körper navigieren
- Vertrauen Sie den automatisierten Ergebnissen?
- Fehlen Informationen auf den Screens?
- Möchten Sie dem Entwicklungs-Team sonst noch etwas sagen?

12.16 Ergebnisliste Usability Walkthroughs mit Wireframes

ID	Beobachtung/Kommentar	Kategorie	Quelle	Massnahme
	TNM-Liste			
B36	Sieht kein M, weil M0. Fände es gut, wenn M0 trotzdem gezeigt wird in der Liste.	Visuell	TP1	"M(0)" wird in der Liste angezeigt -> M immer anzeigen, mit empty state
B37	Gibt immer definierende Läsion für T, N oder M, davon ist jeweils die Schlimmste ausschlaggebend für das TNM. Diese als erste sicher zeigen.	Strukturell	TP1	Die Läsion mit der höchsten Konfidenz als erstes anzeigen
B38	Er fände Boxen als Shortcuts für die jeweilige Läsion gut	Visuell	TP1	Pfeil für aktiven Tab entfernen, Bookmark-Funktionalität machen, Liste beibehalten, da gewichtet keine Änderung
B39	Würde in der Liste nur die anatomische Lokalisation und das Stadium reintun, Infiltration, Grösse usw in die Details tun, schaut er sich nicht an.	Strukturell	TP2	keine Änderung
D2	Würde zusätzlich Infos in Liste zeigen, die er auch in den Rapport tut, SUV Max wenn vorhanden und Volumne	Strukturell	TP5	keine Änderung
I15	Findet den Hinweis "editiert" gut in der Übersicht, nachdem er die Läsion editiert hat.	Strukturell	TP3	Aussage
I27	Erkennt auf nachfragen, dass in der Übersicht die bearbeitete Läsion als "editiert" angezeigt wird.	Strukturell	TP4	"Editiert" prominenter anzeigen.
	Findet das Wording nicht optimal, wichtig für den Arzt ist, dass es "kontrolliert" bzw gesehen wurde.			Keine Änderung, da eine Kontrollfunktion einen Workflow mit sich ziehen und den Prozess verlangsamen würde
B40	Thumbnail hilft schon, aber fraglich ob es verhältnismäßig ist zum Platzverbrauch.	Visuell	TP2	Keine Änderung
	Allgemein			
B20	Kein "Denglisch", alles Deutsch machen	Strukturell	TP1	Einfügen der deutschen Begriffe
B66	Möchte 3 Ebenen sehen + PET/CT Fusion, oder für das Einblenden der Übersicht mit den Infos der Konfidenz (bei Primärtumor nur das T)	Strukturell	TP2, TP5	Einfügen der axialen, koronalen, sagitalen Ansicht und der Fusion
B3	"Lokalisation" anstelle von "Ort" verwenden	Strukturell	TP1	Einfügen von "Lokalisation" anstelle von "Ort"
	Visualisierung			
I5	Wünscht sich eine MIP Darstellung pro T, N und M	Strukturell	TP3	Variante überlegen mit MIP und Schematischer Darstellung
I9	Man sieht die Metastasen nicht in der Übersicht, was für die Übersicht aber enorm wichtig ist. Hirn, Leber, Niere und Nebenniere oft betroffen, in Knochen müssen sie auch sichtbar sein im ganzen Körper	Strukturell	TP3, TP2	Ganzkörper zeigen. Metastasen evt. als Text darstellen, da sonst der Fokus auf die Lunge verloren geht (Vorschlag von Radiologen)
B1	Rechter/Linker Lungenflügel verdreht	Strukturell	TP1	Korrekte Beschriftung der Lungenflügel
B2	Ausgefüllter Tumor stellt Haupttumor dar.	Visuell	TP1	Aussage

ID	Beobachtung/Kommentar	Kategorie	Quelle	Massnahme
B4	Leerer Tumor im linken Flügel irritiert, unklar was es bedeutet	Visuell	TP1	Nicht vorhandene Tumore ("leere Tumore") werden nicht dargestellt
B7	Die Zahlen der Stationen sind standardisiert (jeder Radiologe versteht das), Bei einigen fehlt die zweite Dimension (Lage hinten/vorne). Für's Stadium nicht immer wichtig, aber für die Visualisierung im Tumorboard wichtig.	Strukturell	TP1, TP2	Einfügen der Zahlen in den Lymphstationen
B11	Lungen sind in Lappen und Segmente unterteilt. Lage in den Segmenten ist wichtig für's Stadium. Lappenübergreifende sind schwerer im Stadium.	Visuell	TP2	Einzeichnen der Lungenlappen
B12	Gibt M1a, M1b, M1c zusätzlich zu M0, M1	Strukturell	TP2	Zusätzliche Kategorien (M1a, Mab etc, T1a, usw) anzeigen
I7	Versucht über das Visual zu navigieren, was jedoch nicht interaktiv ist	Funktional	TP3	(Mockbug) Grafik verlinken
I19	Fragt sich, warum die betroffenen Regionen blau sind in der Übersicht. Blau steht in der Radiologie für tiefe Aktivität, würde eher gelb verwenden.	Visuell	TP4, TP5	Evt. Gelbe oder rote Farbe für Läsionen verwenden
I8	Erkennt sofort die Lymphknotenstationen, verschafft sich einen schnellen Überblick, erkennt die betroffenen Regionen	Strukturell	TP3	Aussage
D1	Nur Teilkörper wird gescannt, Beine bringen selten mehr Infos	Strukturell	TP5	Teilkörper MIP zeigen wenn vorhanden
B13	3D Ansicht 3d Visualisierung wäre höchstens für die Lymphknoten spannend, denkt die 2D Ansicht passt aber schon.	Funktional	TP2	Keine Änderung
B5	Nicht klar, wo die Lymphknoten liegen im 3D Raum. Grafik prinzipiell gut, hilf für den ersten Überblick. Besser nur die Stationen, anstatt alle Lymphknoten zeigen.	Visuell	TP1	Einfügen der Lymphstationen
I2	Added Value von 3D Ansicht ist gering, höchstens wenn das Volumen mal beurteilt werden muss	Funktional	TP3	Keine Änderung
I4	Detailansicht und Editieren Window-Leveling ist schon wichtig, aber nicht soooo wichtig wenn es per Default schon gut eingestellt ist.	Funktional	TP3	Aussage
B48	Einheit fehlt bei der Detaildarstellung der Ausmasse	Strukturell	TP1	Einfügen der Einheit
B43	Würde reinzoomen und direkt im Bild editieren (ohne Editiermodus). Möchte Kontextmenü nur zum Entfernen verwenden	Funktional	TP1, TP5	Direkt im Fenster editierbar machen, ohne Editiermodus
D5	Volumen auf Detail wäre gut, Voluming Doubling Time ist wichtig um beurteilen zu können ob bös oder gut, ebenso TLG (Total Lesion Glycolysis)		TP5	Volumen und TLG ins Detail nehmen
I14	Wenn er im Editiermodus editiert, sollte das T, N oder M automatisch angepasst werden in der Seitenleiste (aufgrund der neuen Masse)	Funktional	TP3	Aussage
I11	Seitenleiste von Detailansicht Finde die Detailansicht auf Anhebung	Strukturell	TP3, TP5	Aussage

ID	Beobachtung/Kommentar	Kategorie	Quelle	Massnahme
B52	Hat Detail entdeckt nach Hinweis	Strukturell	TP2	Bei Klick auf Läsion Details und Hover-Effekt oder Overlay/Tooltip anzeigen
B44	Wenn man den Vorschlag ändert, dann sollte sich ein neuer Vorschlag anzeigen.	Funktional	TP2	TNM muss neu errechnet werden, Feedback dazu (Status) muss irgendwie gezeigt werden
B53	Genauer beschreiben warum es T3 ist, wo er reinwächst usw. Konfidenzdiagramm anzeigen, SUV Wert anzeigen bei PET (mean,max), HU Wert für CT (mean, std. deviation), Organ angeben für Metastasen (Leber, Knochen)	Strukturell	TP1, TP2	Features strukturiert eingeben
B45	Findet Editiermodus gut. Würde kein Freitext machen, sondern die Features anzeigen. Sollte nur noch die KI-Meinung modifizieren müssen. Die Kliniker arbeiten allerdings gerne mit Text, da sie den einfach Copy-Pasten können.	Funktional	TP2, TP4, T3, TP1	Einfügen von strukturierten Daten (Dropdowns) im Editiermodus. -> Freitext löschen
I12	Rechter Lappen Mitte gibt es nicht, nur "rechter Mittellappen"	Visuell	TP3	korrekte Bezeichnung der Lappen einfügen
I24	Wünscht sich beim Hauptbronchus eine Angabe des Abstands zur Karina, da dies die Chirurgen jeweils wissen möchten. Dies wird normalerweise auch mit dem Tool ausgemessen und im Befund angegeben	Strukturell	TP4	Angabe des Abstands zur Karina einfügen
I25	Option "Tumor im anderen Lappen" verwirrt ihn (wahrscheinlich schlecht übersetzt von uns)	Strukturell	TP4	Deutsche Übersetzung abklären
I26	Option "Nodus im gleichen Lappen" findet er zu undefiniert, handelt es sich um Malignant Nodus oder einfach einen Nodus? (wahrscheinlich auch schlecht übersetzt)	Strukturell	TP4	Deutsche Übersetzung abklären
B58	Würde über Histogramm/Konfidenz umklassifizieren in der einzelnen Läsion.	Funktional	TP2	Keine Änderung
D3	Findet Bearbeiten-Modus passend.	Strukturell	TP5, TP6	Ausprobieren: Bei Klick auf Läsion und beim Bearbeiten der Läsion erscheinen Detailinfos in der linken Spalte als Expand-Animation
Patientendaten				
B16	Patientennummer unklar	Strukturell	TP1, TP2	Patientennummer einfügen
B18	Datum und Typ der Untersuchung fehlt (wichtig für die Identifikation der Untersuchung, z.B. PET CT vom 20.4.XXXX)	Strukturell	TP1, TP2	Datum und Typ der Untersuchung einfügen
Sensitivität				
B24	Hoch/Mittel/Tief zuwenig genau. Slider wäre besser	Strukturell/ Funktional	TP1, TP2	Slider ausprobieren, 0-100%, Spezifität ausgeben
B25	Je höher Sensitivität, desto mehr müsste gezeigt werden	Strukturell	TP1	Ja
B26	Sensitivität müsste die Konfidenz und das TNM beeinflussen	Strukturell	TP1	Ja
B27	Will immer hohe Sensitivität und gleichzeitig möglichst hohe Spezifität.	Funktional	TP2	Richtig

ID	Beobachtung/Kommentar	Kategorie	Quelle	Massnahme
B28	Möchte Sensitivität zusammen mit der Sensitivität ändern können, bzw. beide gleichzeitig sehen (Rock Kurve), um optimale Genauigkeit zu finden.	Funktional	TP2	Genau
B30	Default-Wert von Sensitivität und Spezifität müsste auf das Optimum eingestellt sein	Funktional	TP2	Yup
I21	Interpretiert den "Sensitivitätsslider" richtig	Strukturell	TP4	Yeah
I30	Findet den Sensitivitäts-Slider im UI besser als die 5 TNM Vorschläge im technischen Prototyp. Die Vorschläge in der Liste sind zudem nicht intuitiv für den Radiologen, findet es komisch, dass die Algorithmik zuerst ein T2, T2, T1, T3 und T4 vorschlägt, dieser Übergang wäre sehr unwahrscheinlich und würde zeigen, dass die Algorithmik irgend etwas nicht versteht (was ja auch zutrifft, da sie die TNM-Klassifikation nicht kennt und eine ähnliche Systematik selbst gelernt hat)	Strukturell	TP4	Yesss
	Für den klinischen Alltag nicht interessant, da man einfach die beste Lösung erwartet	Funktional	TP5	Sensitivität löschen, für den Forschungsaspekt spannend aber nicht für den klinischen Alltag. Eventuell als Simulationsmodus gestalten
	Konfidenz			
B31	Was bedeutet die Konfidenz Angabe? Muss man aber eigentlich erklären oder über Zeit durch die Benutzung kennenlernen. Ist mit dem Algorithmus verbunden.	Strukturell	TP2	Evt. Info-Text oder Info Icon
B32	Würde Konfidenz dann auch pro T, N, M machen.	Strukturell	TP2	Keine Änderung
B33	Stage und Konfidenz gehört zu den Charts auf der rechten Seite der Übersicht, ist für das Tumorboard wichtig.	Strukturell	TP2	Prozentwerte bei Charts ergänzen, evt. Stage bei den Charts ergänzen
I10	Findet die Konfidenz sehr hilfreich, würde anhand tiefer Konfidenz nochmal genauer hinschauen (Indikation, dass etwas nicht so passt mit dem Algorithmus)	Strukturell	TP3	Aussage
I20	Fragt sich, warum die Balken mit der besten Konfidenz im jeweiligen T/N/M nicht auch blau sind, würde ihm helfen die Verbindung herzustellen	Visuell	TP4	Visuelles Design
	Verifizieren			
B56	Glaut "verifizieren" ist zu aufwändig, würde alles was unbearbeitet ist, als korrekt annehmen. Am Ende dann eine Freigabe machen von der gesamten Stage.	Funktional	TP1, TP2, TP3, TP5	Die "Approval"-Funktion wird aus den Details genommen und als Abschluss gestaltet.
I22	Will den Tumor als nicht korrekt beurteilen (geht jedoch nicht im Prototyp) und versucht dann das Kommentarfeld zu nutzen, später versucht er dann den "Bearbeiten" Button	Strukturell	TP4	
	Interaktionen			
B61	Erwartet Doppelklick auf Fenster zum gross machen des Fensters.	Funktional	TP2	Umsetzen

ID	Beobachtung/Kommentar	Kategorie	Quelle	Massnahme
B62	Soll auch durch scrollen im Fenster zu den nächsten Läsionen scrollen können, nicht nur durch die Liste	Funktional	TP2	Umsetzen
B63	Muss Möglichkeit haben, etwas hinzuzufügen.	Funktional	TP2	Evt. Button für neue Läsion und/oder Hover-Effekt und/oder Läsion anzeigen on hover
D2	Kein Anreiz neue Läsionen einzuzichnen, ausser wenn Report generiert wird	Funktional	TP5	"Report generieren" einfügen
I17	Aussagen "Je weniger Klicks desto besser"	Strukturell	TP3	Aussage
I18	"Das System unterstützt, es ist ok wenn es in 1-2 Fällen nicht funktioniert"	Funktional	TP3	Aussage
I28	Generell sieht er das UI als "reduce to the max", findet er gut. Die anderen Tools sind alle total aufgebläht, z.b. Mint Lesion, andererseits sind sie auch weniger fokussiert auf eine Aufgabe.	Strukturell	TP4	Aussage
I31	Geben bei CT Staging nicht immer ein Stadium an, im PET/CT (d.h. Wenn es ein Nuki macht) jedoch immer.	Strukturell	TP4	Aussage
I1	Hohe Falsch-Positiv-Rate machen die Akzeptanz der Algorithmik kaputt (wenn Nicht-Tumore als Tumore eingestuft werden). Im Tagesgeschäft lieber 3 Läsionen verpassen, 60 Nicht-Tumore wegklicken	Funktional	TP3	Aussage
I3	Faustregel: der SUV Wert ist nicht relevant, wenn die Läsion auf dem MIP ohnehin nicht ersichtlich ist	Funktional	TP3	Aussage
B69	Evtl. TNM Richtlinien als Hilfe anbieten.	Funktional	TP2	Könnte im Hilfe-Menü integriert werden (nicht Teil dieser Arbeit)
B70	Idee: Navigation umkehren. Läsionen als Kacheln gross zur Übersicht anbieten, Grafik als Navigation anbieten.	Strukturell	TP2	Alternative gestalten

12.17 SUS Ergebnis

	P1	P2	P3	P4	P5	P6	Total
F1	1	5	5	5	4	5	69.4
F2	2	2	1	1	2	1	75
F3	4	4	4	5	4	5	72.2
F4	1	1	2	2	1	3	72.3
F5	4	3	4	5	5	5	72.2
F6	5	4	1	1	1	1	63.9
F7	4	5	5	5	5	5	80.5
F8	3	2	1	1	1	1	75
F9	3	4	4	5	4	4	66.6
F10	2	2	2	1	1	1	75
							72.21

12.18 Clustering Ergebnis Affinity Diagramm

