

Schlussbericht

zum Teilvorhaben

Interaktives Maschinelles Lernen

Förderkennzeichen: 16SV8638

**Fördermaßnahme: Adaptive Technologien für die
Gesellschaft – Intelligentes Zusammenwirken von
Mensch und KI (MeKI)**

**Deutsches Forschungszentrum für Künstliche Intelli-
genz GmbH
Trippstadter Straße 122
D-67663 Kaiserslautern**

Laufzeit: 15.03.2021 - 14.03.2024

des Verbundprojektes

**Ophthalmology-AI – Intelligente, kooperative Diagnose- und
Therapieunterstützung in der Augenheilkunde**



Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 16SV8638 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Inhalt

I.	Kurze Darstellung zu Zielen, Voraussetzungen, Planung und Ablauf des Vorhabens und Stand der Technik	1
1	Aufgabenstellung	1
2	Voraussetzungen, unter denen das Vorhaben durchgeführt wurde	4
3	Planung und Ablauf des Vorhabens	5
4	Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	8
II.	Eingehende Darstellung	10
1	Verwendung der Zuwendung und erzielte Ergebnisse	10
2	Wichtigste Positionen des zahlenmäßigen Nachweises.....	21
3	Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	22
4	Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit der Ergebnisse und der Erfahrungen.....	22
5	Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens	23
6	Erfolgte oder geplante Veröffentlichung der Ergebnisse	23

I. Kurze Darstellung zu Zielen, Voraussetzungen, Planung und Ablauf des Vorhabens und Stand der Technik

1 Aufgabenstellung

Zielsetzung von Ophthalmo-AI war es, durch eine effektive Zusammenarbeit von maschineller und menschlicher Expertise (Interaktives Maschinelles Lernen IML) bessere Diagnose- und Therapieentscheidungen in der Augenheilkunde zu treffen; besser heißt hier vergleichbare Genauigkeit wie der (kostenintensivere) Goldstandard und eine vergleichbare Akzeptanz seitens der Anwender. Dabei wurden Teile des Entscheidungsprozesses durch Machine Learning unterstützt, was Reliabilität und Objektivität erhöht bei entsprechender Transparenz zur Fehlerabschätzung und zur Reduzierung des Haftungsrisikos. Zusätzlich wurde der Diagnose- und Therapieprozess effizienter gestaltet, indem aus weniger Diagnoseschritten ein differenziertes Bild erstellt wird. Dieses konnte dann über bestehende Strukturen zurückverfolgt werden, um die entsprechende Erklärbarkeit im Mensch-Maschinen-Dialog sicherzustellen.

Konkret war es daher Ziel des Gesamtvorhabens (siehe auch Abbildung 1):

1. diagnose-/prognoserelevante Merkmale durch neue ML-Modelle aus Rohdaten zu extrahieren und diese beispielsweise mit Hilfe von automatisch generierten Visualisierungen erklärbar zu machen;
2. den Stand der Kunst an hoch performanten maschinellen Input-Output Lernverfahren anzuwenden, um klinisch bekannte Merkmale verlässlicher, objektiver und letztendlich effizienter aus Rohdaten zu extrahieren (effizienter = mit geringerem manuellen Aufwand). Hierzu sollen über Konzepte des interaktiven Maschinellen Lernens menschliche Agenten ihr Expertenwissen in den Lernprozess rückkoppeln;
3. große Mengen an heterogenen sensiblen Patientendaten mit hoher Qualität aufzubereiten und insbesondere Imaging-Daten und Texte semantisch anzureichern und dann unter Berücksichtigung von Datenschutz und Datensicherheit für ML zur Verfügung zu stellen;
4. eine lernende Schnittstelle für eine effektive Mensch-Maschine Diagnose- und Therapieinitiative bereitzustellen, um Ziele (2) und (3) im Dialog mit dem medizinischem Fachpersonal zu erreichen. Die Schnittstelle wird auf den menschlichen Experten abgestimmt, um ML reibungslos in den klinischen Alltag zu integrieren, um Diagnosen zu treffen, die optimale Therapie auszuwählen und den Therapieverlauf zu prognostizieren;
5. die einzelnen Komponenten eines klinischen Augmented-Intelligence-Systems (Interaktive Diagnose- und Therapieschnittstelle, ML Werkzeuge, Visualisierungs- und Erklärmodule) aus den Anwendungsfällen zu einem Funktionsdemonstrator in zwei unterschiedlichen Ausprägungsformen für die Augenheilkunde zu generalisieren,

6. zuvor genannte Ziele in dem konkreten medizinischen Anwendungsfeld der Ophthalmologie zu demonstrieren und zu evaluieren. Dazu wird der Demonstrator aus (5) in einer Pilotstudie zur Diagnose und Therapieunterstützung bei Patienten mit Makuladegeneration und Diabetischer Retinopathie auf Praxistauglichkeit überprüft.

Das Teilvorhaben des Deutschen Forschungszentrums für Künstliche Intelligenz hat zum Ziel, klinische Leitlinien und Wissen des medizinische Fachpersonals (Expertenwissen/ bzw. Humane Intelligenz) interaktiv mit maschinellen Lern-verfahren (Künstliche Intelligenz) in den Diagnoseprozess einzubeziehen in einem sogenannten Augmented-Intelligence System. Um dies zu erreichen, werden auf technischer Seite nicht nur diagnostische ML Modelle erarbeitet, sondern auch Erklärbarkeitskomponenten (z.B. Visualisierungstools), die den Medizinern Feedback über die Entscheidungen des ML-Modells liefern und damit einen entscheidenden Beitrag zur Transparenz im Diagnose- und Therapieprozess leisten. Dieses System findet perspektivisch nicht nur in dem Diagnose-Szenario Anwendung, sondern auch in der niederschweligen automatischen Unterstützung bei der Aufbereitung/Annotation von OCT Bilddaten.

Spezielle wissenschaftlich-technische Ziele dieses Teilvorhabens

Zur Umsetzung des dargestellten Ansatzes eines Augmented-Intelligence-Systems in der Augenheilkunde trägt dieses Teilvorhaben mit den folgenden wissenschaftlich-technischen Arbeitszielen bei.

- Erweiterung und Training eines state-of-the-art ML-Modells zur direkten Vorhersage von Diagnosen und Therapien unter Einbezug von Expertenwissen und klinischen Leitlinien.
- Erweiterung der ML-Modelle um Erklärbarkeit durch die Rückführung der ML-Vorhersagen auf medizinische Konzepte und Leitlinien.
- Entwicklung einer IML-Schnittstelle, die Experten aktiv in den Lernprozess integriert (System-Initiative).
- Entwicklung einer IML-Schnittstelle, die es Experten ermöglicht, ihr Wissen auf Anwendungsebene in den Lern-prozess einzubringen, z.B. in einem konkreten Behandlungsfall (Nutzer-Initiative).

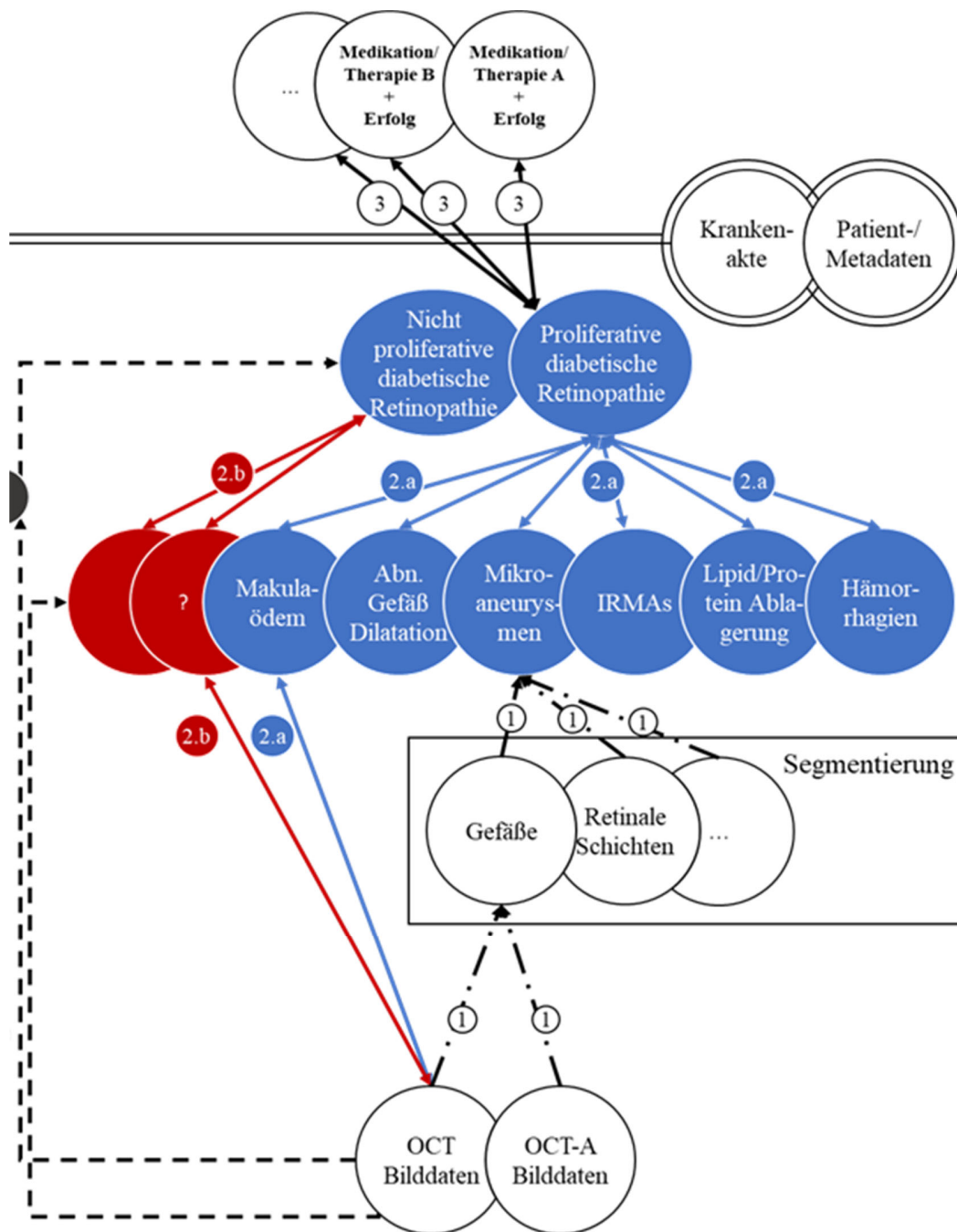


Abbildung 1: Mehrstufiges ML-Konzept: (1) Segmentierung biologischer Einheiten in OCT-Bilddaten (OCT, OCT-A), (2) Ableitung von Pathologien in Bildern, (3) Ableitung von Diagnosen und Therapien durch Hinzunahme weiterer Parameter aus Krankenakte.

2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Die Realisierung der ehrgeizigen Projektziele erfolgte durch eine enge Bündelung der Kompetenzen zwischen den beteiligten Forschungspartnern aus den Bereichen Medizin, Künstliche Intelligenz, Computerlinguistik sowie Medizin- und Bioinformatik:

- Das im März 2021 gestartete und auf 3 Jahre ausgelegte Verbundvorhaben »Ophthalmo-AI« wurde vom **Fraunhofer-Institut für Biomedizinische Technik IBMT** koordiniert, das die Federführung bei der Entwicklung der Datenplattform und Daten-Layer innehatte und Kernkomponenten zur semantischen Informationsextraktion, -integration und -analyse, sowie Sicherstellung der Datenqualität beitrug.
- Die Firma **LangTec** trug Werkzeuge zur Pseudonymisierung und Informationsextraktion aus klinischen Textdokumenten bei.
- Die Modellentwicklung erfolgte durch die **DFKI GmbH** und die **Heidelberg Engineering GmbH**.
- Die Realisierung der Demonstratoren erfolgte durch die **DFKI GmbH**
- Klinische Expertise und Daten wurden durch das **Augenkl. Sulzbach** sowie durch das **Augenzentrum am St. Franziskus-Hospital Münster** bereitgestellt.
- Die **Universität des Saarlandes, Lehrstuhl Arbeits- und Organisationspsychologie** bearbeitete hauptverantwortlich die ethischen, rechtlichen und sozialen Aspekte des Vorhabens, plante die Evaluierung der Prototypen in Zusammenarbeit mit den klinischen Partnern und wertete sie wissenschaftlich aus

Das Deutsche Forschungszentrum für künstliche Intelligenz ist auf dem Gebiet innovativer Mensch-Technik-Interaktion die führende wirtschaftsnahe Forschungseinrichtung in Deutschland. Im Kontext der Implementierung kognitiver Technologien für das medizinische Interaktionssystem mit IML-Komponente bringt das DFKI entsprechende Kompetenz mit ein (<https://iml.dfki.de/>). Zu den Forschungsschwerpunkten gehören multimodale Multisensor-Schnittstellen für Medizin- und Gesundheitssysteme und insbesondere Industry 4.0, Common-Sense- und (interaktive) maschinelle Lernmethoden für Human-Computer-Schnittstellen, Wissensermittlung, Informationsextraktion und kognitive Modellierung mit Ontologien. Das DFKI bringt Erfahrung bisheriger Projekte aus dem medizinischen Kontext mit ein, dazu gehören das BMWi-Projekt Klinische Daten Intelligenz (<https://www.klinische-datenintelligenz.de/>), in dem unter anderem Bildmaterial der Radiologie mit Hilfe von ML-Methoden analysiert wurde, das BMBF-Projekt Intera-KT - Interaktives Kognitives Assessment Tool (<https://intera-kt.de/>), welches eine IML-Diagnosekomponente enthält, sowie das Projekt Skincare, welches Deep Learning zur Analyse von Bilddaten im Bereich Hautkrebsdetektion einsetzt (Nunnari & Sonntag, 2019; Nunnari et al. 2020). 2011 wurde

Radspeech (Sonntag, 2012), einem digitalen Interaktionssystem für Radiologen, der German High Tech Award verliehen.

DFKI trägt die Hauptlast der KI-Komponentenentwicklung. Dazu zählen die ML-Werkzeuge zur Diagnose- und Therapieunterstützung sowie die Umsetzung des interaktiven ML-Konzeptes in der Benutzerschnittstelle mit den Erklärungs- und Visualisierungstools.

3 Planung und Ablauf des Vorhabens

Fraunhofer IBMT koordinierte das Vorhaben und stellt die IT-Plattform für die Aufnahme und das Management der benötigten klinischen Datenbestände zur ML-Modellerstellung bereit. Hierfür erweiterte es seine XpIOit-Datenintegrationsplattform und schafft mit ihr auch das Data-Layer des Ophthamo-AI-Augmented-Intelligence-Systems (Teilvorhaben Datenplattform und Data-Layer). Als Spezialist für Sprachverarbeitung hat die Firma LangTec Werkzeuge zur Extraktion von Informationen aus Befundtexten bereitgestellt, die für die Modellentwicklung und ihre Ausführung benötigt wurden. Diese wurden als NLP-Pipeline der Datenplattform vorgeschaltet (Teilvorhaben Textuelle Informationsextraktion). Die beiden augenärztlichen Versorger AKS und AZM stellten retrospektiv und prospektiv umfangreiche klinische Datenbestände für das Training von ML-Modellen in der XpIOit-Plattform zur Verfügung, annotierten OCT-Bilder, legten die Anwendungsfälle und ihre Anforderungen fest, begleiteten die Entwicklung und erproben die geschaffenen Demonstratoren im Klinikeinsatz (Teilvorhaben Datenbereitstellung und Piloterprobung). Durch die Einbindung von zwei Expertenzentren wurde zum einen ein Bias in den Daten und damit in den ML-Modellen abgefangen und zum anderen ein Systemtest durch eine signifikante Zahl von Spezialisten sichergestellt. Das DFKI trug die Hauptlast der KI-Komponentenentwicklung. Dazu zählten die ML-Werkzeuge zur Annotation (Segmentierung) der OCT-Bilder, zur Diagnose- und Therapieunterstützung sowie die Umsetzung des interaktiven ML-Konzeptes in der Benutzerschnittstelle mit den Erklärungs- und Visualisierungstools (Teilvorhaben Interaktives Maschinelernen). Das Medizintechnikunternehmen HE entwickelte ML-Verfahren zur Merkmalsextraktion und Segmentierung in OCT-Bildern (Teilvorhaben OCT-Feature-Extraktion und Segmentierung). Die UdS bearbeitete hauptverantwortlich die ethischen, rechtlichen und sozialen Aspekte des Vorhabens, plante die Piloterprobung in Zusammenarbeit mit den klinischen Partnern und wertete sie wissenschaftlich aus (Teilvorhaben ELSA und Piloterprobung). Es wurden keine weiteren Akteure als assoziierte Partner in das Vorhaben eingebunden.

Der Arbeitsplan (siehe Abbildung 2) sah eine Vielzahl von Aufgaben gruppiert in 11 Arbeitspakete vor, die sich über eine dreijährige System- und Modellentwicklungsphase erstreckten. Schwerpunkte der Aufgaben des DFKI bildeten das AP1 ‚Anforderungen und Systemdesign‘, (separates Dokument zur Gesamtarchitektur, Meilenstein 1), AP 5 ‚OCT Feature Extraktion und Segmentierung‘ (Entwicklung eines Annotationstools mit ML Un-

terstützung), AP 6 'ML Werkzeuge zur Diagnose- und Therapieunterstützung (ML Modelle), AP 7 'IML Benutzerschnittstelle mit Erklärungs- und Visualisierungstools' (Dashboard, Meilenstein 3), sowie AP 8 'Systemintegration und Demonstratorbereitstellung'. Darüber hinaus unterstützte das DFKI intensiv die Evaluation der Demonstratoren in AP 9 'Piloterprobung' (Meilenstein 4).

Arbeitspaket und Aufgaben	1. Jahr				2. Jahr				3. Jahr			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
AP1 Anforderungen und Systemdesign		MS1										
1.1 Anforderungsanalyse und Anwendungsszenariospezifikation												
1.2 Ermittlung technischer Anforderungen												
1.3 Systemspezifikation												
1.4 Bedienkonzept und Benutzerschnittstelle												
AP2 Datenbereitstellung			MS2									
2.1 Festlegung und Modellierung relevanter Parameter												
2.2 Bereitstellung retrospektiver Bild (OCT)- und Patientendaten												
2.3 Annotation und Aufbereitung retrospektiver Daten												
2.4 Erhebung + Bereitstellung prospektiver Daten												
AP3 Semantische Anreicherung & Datenplattform												
3.1 Wissensbasis Ophthalmologie zur Datenintegration												
3.2 Umsetzung Datenpipelines für OCT-Daten und strukturierte KIS-Daten												
3.3 Umsetzung Datenschutzkonzept												
3.4 Statistische ML-Methoden zur Optimierung der Datenqualität												
3.5 Data-Layer-Anbindung an ML-Modelle												
3.6 Systembereitstellung und Betrieb für klinische Partner												
AP4 Textuelle Informationsextraktion												
4.1 NLP-Datenpipeline für Befunde und Diagnosen												
4.2 NLP-Datenpipeline für Medikamente, Therapien u. Therapiefortschritt												
4.3 Extraktionsdienst für Pseudonymisierungsinformationen aus ...												
4.4 Anbindung an die Datenplattform												
AP5 OCT-Feature-Extraktion und Segmentierung												
5.1 ML Vorverarbeitung von Rohdaten												
5.2 Extraktion klinisch bekannter Marker aus Längsschnittrohdaten												
5.3 Interaktives ML zur Segmentierung und semantischen Anreicherung von OCT-Bildern												
AP6 ML-Werkzeuge zur Diagnose- und Therapieunterstützung												
6.1 Analyse von Diagnose- und Therapieverfahren bzgl. maschineller Umsetzung												
6.2 Transfer von klinischen ML-Modellen aus Forschungsprojekten												
6.3 ML Modell zur Vorhersage von Diagnosen und Therapien unter Einbezug von Expertenwissen und klinischen Leitlinien												
AP7 IML Benutzerschnittstelle mit Erklärungs- & Visualisierungstools												
7.1 Visualisierung relevanter Daten und Entscheidungsprozesse der ML Modelle												
7.2 Erweiterung der ML Modelle (aus AP 6) um Erklärbarkeit												
7.3 Implementierung IML-Schnittstelle für Experten (Systeminitiative)												
7.4 Implementierung IML-Schnittstelle für Experten (Nutzerinitiative)												
AP8 Systemintegration und Demonstratorbereitstellung										MS3		
8.1 KIS-Schnittstellen zur Dateneinspeisung												
8.2 Schnittstelle OCT-Bilddatenbank												
8.3 Komponentenintegration und Systemtest												
8.4 Demonstratorbereitstellung und Betrieb												
AP9 Piloterprobung												MS4
9.1 Pilotplanung inkl. Datenmanagement												
9.1 Demonstratorevaluierung I												
9.2 Demonstratorevaluierung II												
9.4 Auswertung												
9.5 Systemoptimierung												
AP10 Ethische, rechtliche und soziale Aspekte												
10.1 Ethikanträge für Pilotstudien und zum Datenzugang												
10.2 Datenschutzkonzept												
10.3a Akzeptanz-, und Vertrauenswürdigkeitsstudie Patienten												
10.3b Akzeptanz-, und Vertrauenswürdigkeitsstudie Ärzte												
10.4 Medizinprodukterecht und Patientensicherheit												
AP11 Projektkoordination und Querschnittsthemen												
11.1 Wissenschaftliche und entwicklungstechnische Projektkoordination												

Abbildung 2: Arbeitspakete, Aufgaben und Meilensteine des Verbundvorhabens und ihr Timing gemäß der Projektskizze.

Der Arbeitsplan des Projektes wurde mit Ausnahme der OCT-Annotierungen ohne nennenswerte Abweichungen umgesetzt. Projektergebnisse aus den einzelnen Aufgaben wurden in den Jahres-Zwischenberichten beschrieben. Alle Meilensteine des Verbund-Vorhabens wurden erreicht, Meilenstein 2 ‚Annotierter klinischer Datenbestand verfügbar in Datenplattform aus AP3 für ML-Aufgaben‘ verzögerte sich zunächst um drei Monate, weil der Antrag bei der Ethikkommission des Saarlandes mit Verweis auf das Medizinproduktrecht zunächst zurückgewiesen wurde und neu eingereicht werden musste. Dies verzögerte auch den Beginn der OCT-Annotierungen, die danach bis zum Projektende fortgeführt wurden, um eine möglichst große Zahl von annotierten OCT-Aufnahmen für das Training der ML Modelle zu haben. Die Annotierung musste durch besonders geschulte klinische Fachkräfte und entlang eines speziellen Schemas und mit einem eigens vom DFKI dafür entwickelten Tool erfolgen. Vor der Durchführung von Annotierungen mussten die Fachkräfte außerdem eine eigens hierzu definierte Zertifizierung durchlaufen, um die Qualität und Vergleichbarkeit der Annotierungen gewährleisten zu können. Das erwies sich deutlich zeitaufwendiger als erwartet, was u.a. auch in der hohen Arbeitsbelastung der Fachkräfte begründet lag.

4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

ML in der medizinischen Diagnostik

Die Forschung zeigt, dass ML vor allem bei bildgebenden Medizinanwendungen die komplette Spanne von Rohdaten bis Diagnoseeinschätzung abdecken kann. Ende 2016 konnte eine Gruppe von Google Research zeigen, dass ein mit Deep Learning Verfahren trainierter Algorithmus eine diabetische Retinopathie (eine Netzhauterkrankung hervorgerufen durch Diabetes mellitus) basierend auf Fundusbildern des Augenhintergrunds klassifizieren und dabei eine Sensitivität und Spezifität von über 90% erreichen kann (Gulshan et. al, 2016). Die Stärke - und gleichzeitig Schwäche für die praxisnahe Medizinanwendung - ist hierbei, dass keinerlei Expertenwissen oder explizite Modellierung für die Definition von aussagekräftigen Merkmalen notwendig ist. Das maschinelle Lernverfahren ist im Gegenteil in der Lage, auf großen Sets annotierter Fundusbilder (100k+ Patientenbilder von Ophthalmologen klinisch bewertet) selbstständig relevante mehrdimensionale Merkmalkombinationen zu finden. Die darauf basierenden diagnostischen Entscheidungen können dann für das Fachpersonal aber nicht darstellbar gemacht werden und sind auch nicht nachvollziehbar.

Kombination von Input-Output ML und strukturellen ML-Verfahren

Besonders im Bereich der medizinischen Bildgebung finden sich Beispiele, wo ML effektiv in der Kombination mit medizinischen Experten angewandt wird. Besonders hervorzuhe-

ben ist hierbei das Potenzial von ML-Verfahren zur computerassistierten Detektion/Diagnose. Unabhängig von der konkreten Domäne kann ML effizient eingesetzt werden, um die Verteilung von latenten Merkmalen in medizinischen Bilddaten zu lernen, die dann effektiv lokale Merkmale in Bildern beschreiben. Zudem wird ML genutzt, um Bildanalyseverfahren für zusätzliche Bildgebungsmodalitäten basierend auf einem bereits erlernten Gerüst von Merkmalsverteilungen zu entwickeln. Klassische Anwendungen sind Segmentierungsalgorithmen (z.B. voll-/teilautomatisierte Lokalisierung und Segmentierung von Gewebeklassen in MRT-Scans, siehe beispielsweise Kleesiek, 2016; Trebeschi et al., 2017). In einem finalen Schritt können zuvor genannte Ansätze auch zur computerassistierten Detektion oder sogar Diagnose zusammengesetzt werden. Hierbei werden zunächst die potentiellen Regionen in MRT-Bildern detektiert, anschließend die Regionen durch zuvor trainierte latente Merkmale dargestellt (siehe beispielsweise Shin et al., 2016). Nur wenn sich diese Merkmale an medizinisch interpretierbaren morphologischen Strukturen orientieren, kann hier eine Schnittstelle für den medizinischen Experten implementiert werden. Abschließend kann ein vortrainierter Klassifikationsalgorithmus basierend auf den extrahierten Merkmalen auch eine Einschätzung bezüglich wahrscheinlicher Diagnosen abgeben.

Methoden zur Erhöhung von Transparenz und Fehlerabschätzung

Visualisierung von Entscheidungsmechanismen in Neuronalen Netzen: Sogenannte Perturbationstechniken stören/maskieren Teile des Eingangssignals und evaluieren deren Änderungen in der Vorhersagewahrscheinlichkeit des ML-Modells. Sobald entsprechende Abschnitte, die maximal zur Vorhersage beitragen, maskiert/gestört werden, sinkt die Klassifikationswahrscheinlichkeit (Zeiler & Fergus, 2014). Dieser Ansatz zu entfaltenden Netzen (De-Convolutional Neural Networks) kann Heatmaps für jede Schicht eines faltenden Netzes (Convolutional Neural Network) visualisieren und damit die Verknüpfung von Merkmal und Rohdatum auf der jeweiligen hierarchischen Ebene zeigen; Merkmale können so über Rohdatenausschnitte, für die sie sensitiv sind, erklärt werden.

Unüberwachtes Strukturlernen: Hierbei können selbst ohne oder mit lückenhaften Trainingslabels latente Merkmale in Rohdaten identifiziert werden -unüberwachtes Lernen. Autoencoder oder Generative Adversarial Networks können automatisch Output abstrahieren und dann wieder rekonstruieren oder Rohdaten erzeugen, die ein Klassifikator als „echt“ einstufen würde. Die komprimierte Repräsentation sowie der Rekonstruktions-/Produktions-Workflow stellen dabei eine latente generalisierbare Repräsentation der Daten dar, ein Modell der darunterliegenden abstrakten Struktur. Durch das Hinzugeben bestehender medizinischer Diagnosekonzepte als Bedingungen/Optimierungskriterien, können weitere inkrementelle Strukturen über die bekannten prädiktiven Konzepte hinaus automatisch aus Rohdaten extrahiert werden. Zur Repräsentation verschiedener hierarchischer diagnostischer Ebenen (siehe auch Abbildung 1) können beispielsweise mehrere Autoencoder vortrainiert und gestapelt werden -Stacked Autoencoder-, um somit eine im Abstraktionsniveau zunehmende latente Repräsentation der Daten zu erhalten.

Bedingte Vorhersagewahrscheinlichkeiten: Bayes'sche Netzwerke (BN) propagieren (bedingte) Wahrscheinlichkeiten durch das Netzwerk, die mit konkreten Diagnosen verbunden werden können. Diese Netze können vorzugsweise dazu verwendet werden, um basierend auf bereits extrahierten Basis-Features, Expertenwissen anzuwenden. Als Ergebnis liefern die Netze dann eine Aussage (z.B. eine Diagnose oder die Beurteilung des Trends) mit Konfidenzabschätzung. Die Aussage kann dann analog zurückverfolgt und begründet werden.

Interaktives maschinelles Lernen (IML): Die vollständige Automatisierung von Aufgaben kann extrem schwierig oder sogar unerwünscht sein: Mit IML kann die Reichweite von ML-Lösungen für Endanwender im medizinischen Bereich erhöht werden, indem intelligente Benutzerschnittstellen zum interaktiven Lernen und Anwenden von ML-Modellen eingesetzt werden. So kann durch Visualisierung und interaktive Exploration der Patientendaten eine Therapievorhersage gemeinsam mit dem behandelnden Arzt getroffen werden (Prange et al., 2017). Dabei ist es essenziell, dass die vom System getroffenen Entscheidungen transparent an den Endanwender weitergegeben werden, um so beispielsweise eine Zeitersparnis im Diagnoseprozess erzielen zu können (Prange & Sonntag, 2019).

II. Eingehende Darstellung

1 Verwendung der Zuwendung und erzielte Ergebnisse

Der wesentliche Teil der Zuwendung wurde für Personalkosten verwendet. Diese Ausgaben spiegeln sich in den erzielten Ergebnissen der Arbeitspakete wider.

1.1 Zusammenfassung der Ergebnisse des Teilvorhabens des DFKI

Der Arbeitsplan des Gesamtvorhabens sah nach einer Spezifikationsphase und Klärung der ethischen und datenschutzrechtlichen Voraussetzung die Bereitstellung der für das Trainieren von ML-Modellen nötigen klinischen Datenbestände durch die Kliniken in der XpIOit-Datenmanagementplattform des IBMT vor. Aufbauend darauf entwickelte das DFKI anhand der zusammengeführten Daten die nötigen ML-Modelle und Werkzeuge zur OCT-Bildverarbeitung und zur Diagnose- und Therapieunterstützung inklusive der IML-Benutzerschnittstelle des Gesamtsystems zur Erklärbarkeit der ML-Ergebnisse, in das die Einzelkomponenten integriert wurden. Mehrere Demonstratoren mit unterschiedlichen Schwerpunkten wurden daraus abgeleitet und anschließend auf Praxistauglichkeit im Klinikbetrieb untersucht. Die folgende Abbildung 3 zeigt das Gesamtsystem und seine Hauptkomponenten im Zusammenspiel der Entwicklungsarbeit.

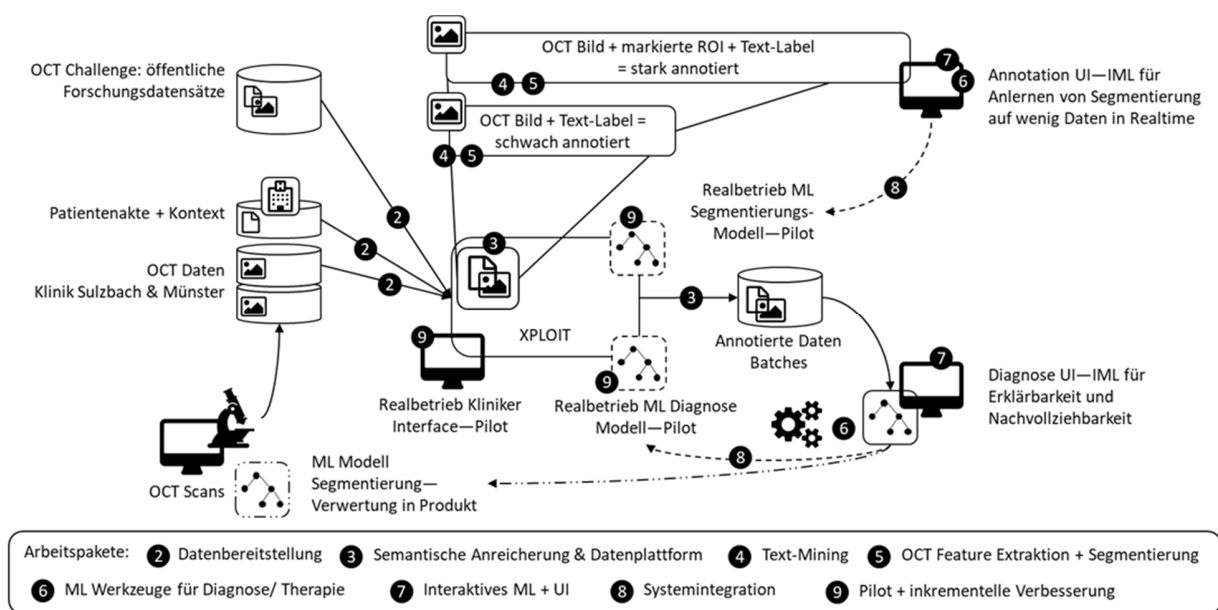


Abbildung 3: Zusammenwirken der Arbeitspakete.

Dieses Teilvorhaben fokussierte sich auf die Arbeitspakete 1 und 5 bis 9, an denen das DFKI den größten Anteil bzw. die Leitung hatte.

1.2 Details zu den Ergebnissen des Teilvorhabens des DFKI

Nachfolgend werden die Ergebnisse dieses Teilvorhabens zu den betreffenden Arbeitspaketen des Gesamtvorhabens im Detail dargestellt.

AP 1: Anforderungen und Systemdesign

Ziel dieses Arbeitspakets des Gesamtvorhabens war die Erstellung der Systemspezifikation. Als Beitrag des Teilvorhabens erarbeiteten die Partner schwerpunktmäßig Anforderungen an das DSGVO konforme Datenmanagement und spezifizierte Eigenschaften und Dienste des Datenservers und des Daten-Layer zu anderen Systemkomponenten. Das Ergebnis war ein vom DFKI erstelltes Dokument zur Gesamtarchitektur (siehe auch Diagramm zur Gesamtarchitektur, Abbildung 4) und zum Bedienkonzept als Meilenstein 1.

T1.1: Anforderungsanalyse und Spezifikation der medizinischen Anwendungsszenarios

In diesem Teilarbeitspaket wurde zusammen mit den klinischen Partnern das erforderliche klinische Hintergrundwissen zu den Zielerkrankungen Diabetische Retinopathie (DRP) und Altersbedingter Makuladegeneration (AMD) sowie deren Diagnostik mittels OCT-Bildgebung und Therapie erarbeitet und zusammengetragen. Die Behandlungsprozesse und Datenerhebung in der Augenklinik Sulzbach (AKS) und im Augenzentrum Münster (AZM) wurden untersucht. Basierend auf den verfügbaren Daten zu AMD und

Diabetischem Makula-Ödem (DMÖ) wurden medizinische Anwendungsszenarien für interaktive KI-Lösungen diskutiert und konkretisiert.

T1.2: Ermittlung technischer Anforderungen

Das IBMT hat basierend auf den Behandlungsprozessen technische Anforderungen an das Datenmanagement abgeleitet. Dafür wurde die Datenintegrationsplattform XploIt verwendet und erforderliche Systemanpassungen und Erweiterungen identifiziert.

T1.3: Systemspezifikation: Architektur, Funktionalität und Schnittstellen

Nach Vorstellung der Datenintegrationsplattform XploIt wurden Architekturvorschläge für deren Nutzung im Datenmanagement für die Modellentwicklung und den Demonstrator des Assistenzsystems unter den Partnern erarbeitet. Es wurden Designvorschläge zur Datenintegrationsarchitektur diskutiert und in dem Dokument zur Gesamtarchitektur festgehalten.

Die Spezifikation der REST-Schnittstellen von XploIt wurde entsprechend den Projektanforderungen für das Datenmanagement in der Modellentwicklung angepasst und optimiert. Dies umfasst insbesondere die datenschutz-konforme Synchronisation von Bild-Annotationen mit dem OCT-Annotations-Tool des DFKI und die Abfrage von EHR-Daten über die OCT-Befundungen. Diese REST-Schnittstelle bildet den zentralen Ausgangspunkt zum funktionalen Zugriff auf alle Daten für die darauf aufbauenden Funktionalitäten der Partner (siehe Abbildung 4). Diese Anpassungen wurden im Dokument zur Gesamtarchitektur beschrieben und fortlaufend aktualisiert.

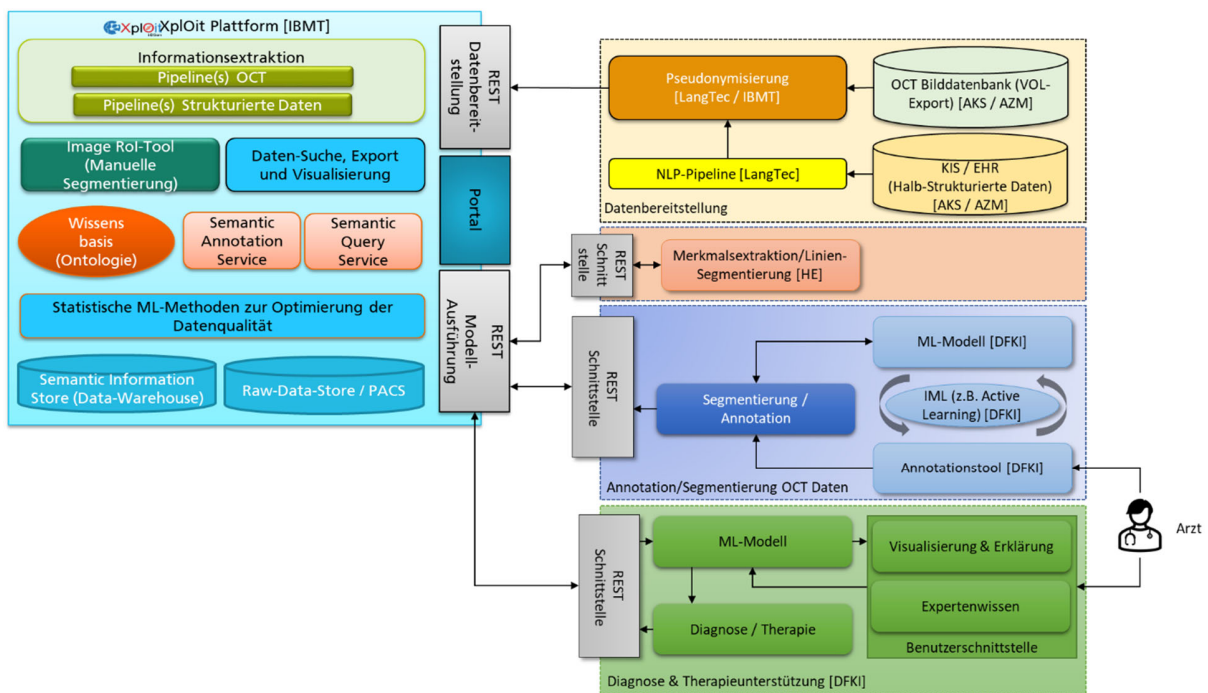


Abbildung 4: Gesamtarchitektur zum Ophthamo-AI-Systems.

T1.4: Design und Spezifikation von Bedienkonzept und Benutzerschnittstellen

Es wurden Erweiterungen der Benutzerschnittstelle (Portal) der Xploit-Plattform identifiziert und skizziert, um den Umgang mit OCT-Bilddaten und deren manuellen Annotation zu verbessern. Das IBMT hat eine Dashboard-Ansicht spezifiziert, um die Übersicht und Visualisierung der bereitgestellten Daten mit ihren Kennzahlen zu verbessern. Diese Dashboard-Ansicht wurde in die Xploit-Benutzeroberfläche integriert. Darüber hinaus wurden Komponenten als Erweiterungen der Xploit-GUI spezifiziert, um ein Daten-Monitoring und die Überprüfung der Datenqualität zu verbessern.

AP 5 'OCT Feature Extraktion und Segmentierung'

Ziel dieses Arbeitspakets war die Entwicklung interaktiver ML-Werkzeuge zur Extraktion und Segmentierung klinischer Marker in OCT-Bildern. Die Hauptkomponenten umfassen: (1) ein spezialisiertes Annotationstool für OCT-Bilder, (2) neuronale Segmentierungsmodelle für klinische Marker und (3) „MedDeepCyleAL“, ein Framework für *aktives Lernen*, das das Annotationstool und die Modelle integriert, um die Modelle kontinuierlich zu verbessern und den Annotationsaufwand zu minimieren und den Prozess für die Nutzer zu vereinfachen und zu beschleunigen.

Das Annotationstool wurde in enger Zusammenarbeit mit den zukünftigen Nutzern, den Fachärzten der Partnerkliniken, entwickelt und speziell für OCT-Annotationen angepasst. Das Tool bietet eine Reihe von Möglichkeiten entweder freihand oder mit Unterstützung durch Bildverarbeitungsverfahren Segmentierungen in OCT-Bilder einzuzeichnen. Außerdem werden die Segmentierungen von bereits trainierten Modellen als Vorschläge genutzt, die vom Nutzer nur noch angepasst werden müssen, wodurch der Zeit- und Arbeitsaufwand der Nutzer minimiert wird. Unter Verwendung dieses Tools wurde ein Teil der gesammelten Daten der Kliniken zum Training von Neuronalen Netzen annotiert. Dazu wurden ausführlich die Anforderungen an die Annotationen sowie das geplante Vorgehen festgelegt. Auf dieser Grundlage wurde eine Annotationsanleitung erstellt, welche genau bestimmt, welche Merkmale auf welche Weise zu annotieren sind. Die insgesamt 20 Merkmale setzen sich zusammen aus acht Membranen und Schichten, neun Flüssigkeiten und drei kategoriellen Merkmalen. Darauf aufbauend wurde zuerst ein kleinerer Zertifizierungsdatensatz von einer Fachärztin mit dem Annotationstool erstellt. Diese Annotationen wurden in mehreren Iterationen überarbeitet, bis sie so genau wie möglich den Anforderungen entsprachen. Diese Annotationen wurden dann sowohl in der Annotationsanleitung als Beispiele verwendet als auch zur Ausbildung und Evaluation von weiteren Fachärzten, welche dann nach erfolgter Ausbildung die Annotation von größeren Datenmengen übernahmen. Insgesamt wurden 1486 Bilder innerhalb von ca 450 OCT Aufnahmen annotiert.

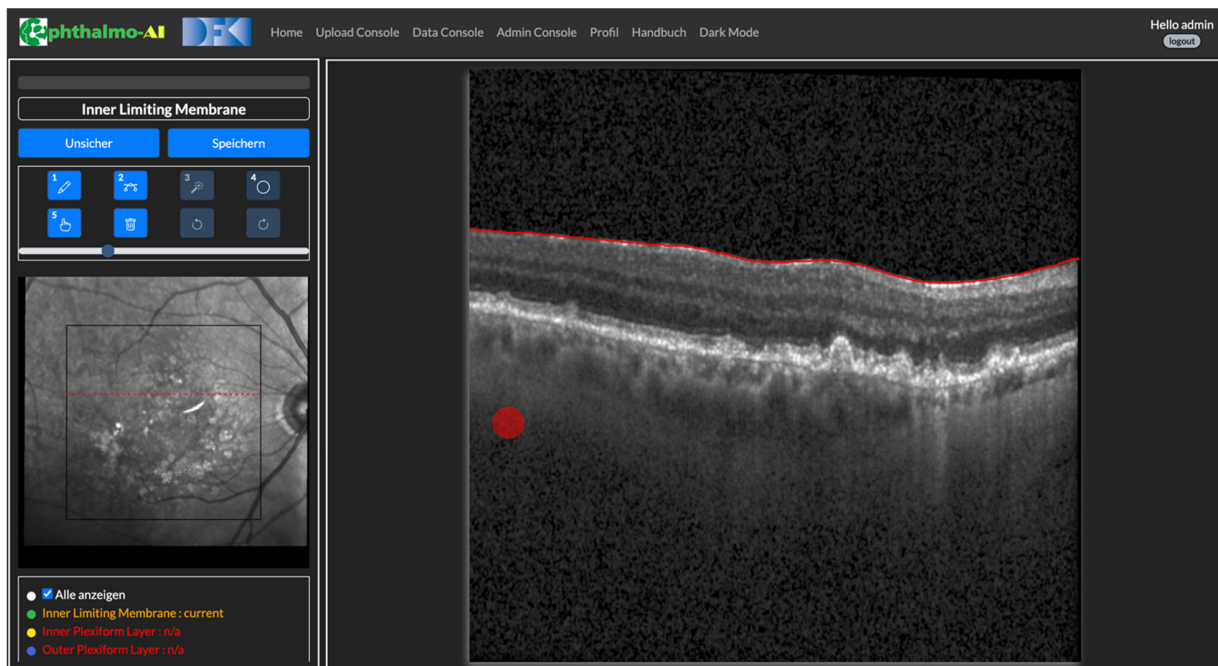


Abbildung 5 Die Oberfläche des Annotationstools

Die so gesammelten Daten wurden über die Xpl0it Plattform an die Projektpartner verteilt. So konnten die Annotationen für andere Projektzwecke, beispielsweise das Trainieren von Segmentierungsmodellen durch Heidelberg Engineering verwendet werden.

Die Human-in-the-Loop-Umgebung „MedDeepCyleAL“ wurde entwickelt, um Segmentierungsmodelle zu trainieren und Experimente für verschiedene Active-Learning-Methoden durchzuführen. Active Learning ist ein Verfahren, bei dem durch eine intelligente Auswahl der zu annotierenden Daten (Daten, die den größten Wissensgewinn versprechen, da das System bei ihrer Beurteilung am unsichersten ist) und ein Modelltraining auf diesen ausgewählten Daten sowohl der Annotations- als auch der Lernprozess beschleunigt werden. Dieses Framework wurde erfolgreich für eine Teilmenge der Annotationen eingesetzt. Veröffentlichungen zu den Ergebnissen des AL-Einsatzes sind für Ende 2024 geplant.

AP 6 ‘ML Werkzeuge zur Diagnose- und Therapieunterstützung

und AP 7 ‘IML Benutzerschnittstelle mit Erklärungs- und Visualisierungstools‘

Ziele der Arbeitspakete 6 und 7 war die Entwicklung von ML Werkzeugen zur Diagnose- und Therapieunterstützung und die Integration dieser Werkzeuge in eine Benutzeroberfläche mit Visualisierungs- und Erklärungskomponenten. ML Werkzeuge stellen hier im Wesentlichen neuronale Modelle dar, die in der Lage sind, Ärzte bei zeitaufwändigen Aufgaben zu unterstützen bzw. ihnen diese abzunehmen. Die beiden zentralen Datenquellen bei Entscheidungen zu Diagnosen und Therapien bilden die Patientenakten (EHR – electronic health record) und die Aufnahmen des Augenhintergrunds in Form von OCT Bildern (siehe hierzu auch Abbildung 10 zum Datenverarbeitungsprozess).

Datenquellen

Die Patientenakten wurden von Projektpartner LangTec annotiert und die Ergebnisse in das zentrale Data Warehouse XplOit hochgeladen. Dieser Prozess erstreckte sich bis zum Ende des Projekts, da die Annotationsalgorithmen immer wieder angepasst werden mussten und sich damit auch die verfügbaren Parameter, die aus XplOit abgerufen werden konnten, geändert haben. Daher mussten auch die lokalen Prozesse, die diese Parameter auswerten und auf ihnen Modelle trainieren sollten, immer wieder angepasst werden.

Die zweite zentrale Datenquelle bilden die OCT Bilder, die von den Kliniken bereit gestellt wurden und ebenfalls über XplOit abgerufen werden konnten. Zur medizinischen Auswertung der Bilder, die aus einer zentralen Draufsicht und einer Reihe von einer unterschiedlichen Zahl Schnitten (Slices) bestehen, müssen auch diese Schnitte annotiert werden. Dabei werden vor allem die verschiedenen Schichten und einige weitere Features (Läsionen, Flüssigkeiten) segmentiert. Dazu wurde ein separates Annotationstool entwickelt und eingesetzt, das unter AP 5 näher beschrieben ist. Auch dieser Prozess zog sich bis zum Projektende hin.

Im Rahmen des Projekts wurden drei verschiedene Demonstratoren entwickelt:

- „DR Grading“: Diagnoseunterstützung für Diabetische Retinopathie
- Intelligentes Tutorssystem: Nutzung des DR Systems als Lernsoftware
- Dashboard: System zur Entscheidungsunterstützung für Therapien

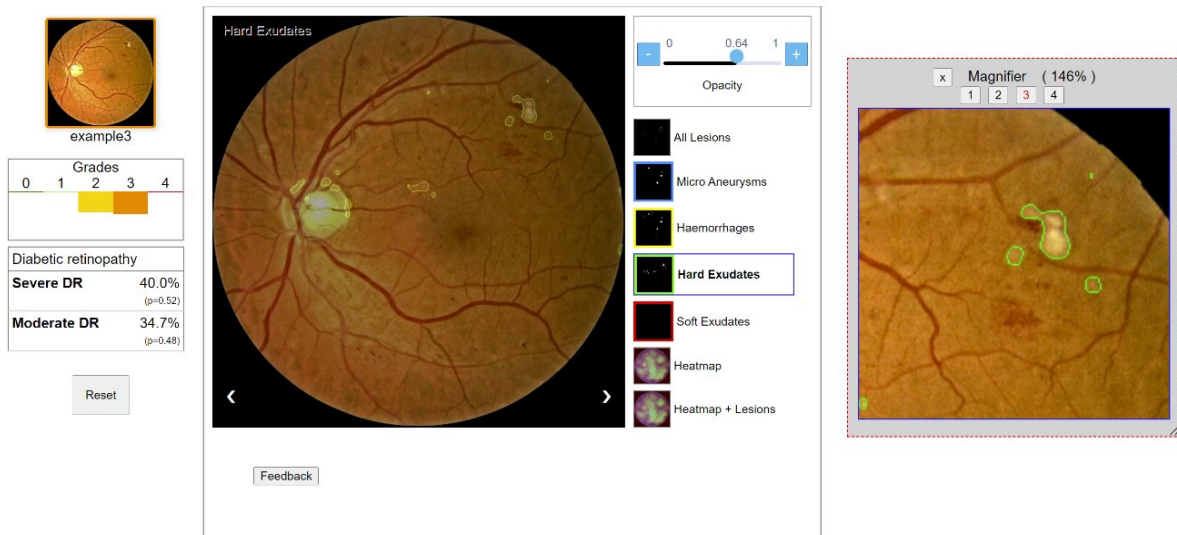
„DR Grading“ Demonstrator

Als erstes Teilziel wurde ein webbasierter Prototyp zur Diagnoseunterstützung für die Diabetische Retinopathie entwickelt (siehe Screenshot in Abbildung 6). Die dazu notwendigen Modelle basieren auf frei verfügbaren Fundusbildern, da lange keine annotierten OCT-Bilder in ausreichend großer Zahl zum Trainieren von Modellen zur Verfügung standen. Die Software wurde so entworfen, dass ein Wechsel zu dem geplanten Bildformat (OCT) ohne größeren Aufwand möglich ist, sobald diese Annotationen vorliegen. Der Demonstrator wurde auf dem Projektmeeting im Herbst 2022 präsentiert und vor allem seine Erklärungskomponenten wurden von den Fachärzten sehr positiv bewertet. Wesentliche Merkmale dieses Demonstrators sind:

- linke Spalte: Unterscheidung des Schweregrads der Erkrankung in fünf Klassen, deren Prognosen durch das Modell auch visuell mit der jeweiligen Wahrscheinlichkeit und Konfidenz angezeigt werden (Balkendarstellung plus die Liste unter dem Originalbild)
- mittlere Spalte: fünf verschiedene Merkmalstypen, die für die Diagnose relevant sind, können von dem Modell prognostiziert werden; diese können als Overlays über das Originalbild gelegt werden und geben dem Benutzer damit eine Erklärung für die Klassifikationsergebnisse. Darüber hinaus können auch Heatmaps (Markierung der für das Ergebnis relevantesten Bildbereiche) und die Kombination aller genannten Masken über dem Original gezeigt werden.
- rechte Spalte: über eine Lupe kann sich der Benutzer Teile des Bildes inklusive der ausgewählten Merkmalsmasken zur genaueren Analyse vergrößert anzeigen lassen

Ophthamo-AI Classification Demo

Classification of Fundus images for DR (diabetic retinopathy)



Diabetic retinopathy	
Severe DR	40.0% ($p=0.52$)
Moderate DR	34.7% ($p=0.48$)

Abbildung 6 "DR Grading" Demonstrator

Der Demonstrator wurde für eine Studie durch die Projektpartner der Universität des Saarlandes zur Verfügung gestellt und von ihnen umfangreich evaluiert. Dabei standen vor allem die Erklärungskomponenten im Mittelpunkt. Eine zentrale Erkenntnis daraus ist, dass Erklärungen nicht in jedem Fall und für jeden Benutzer nur positive Effekte haben, da Benutzer sie auch als überflüssig (bei einfachen Fällen oder wenn sie sich in ihrem Urteil sehr sicher sind) und zeitraubend empfinden können. Die Endergebnisse der Studie wurden auf dem 53. GDP Congress in Wien unter dem Titel „Caught my (A)eye: Effects of explanations of AI-based diagnostic support in ophthalmology“ eingereicht (und angenommen). Weitere Publikationen sind in Arbeit.

Tutorsystem

Bei der Vorstellung des „DR Grading“ Systems auf einem Konsortialmeeting stellte sich heraus, dass die Funktionalitäten des Systems eine Schwachstelle in der Ausbildung der Assistenzärzte offenbarten: das systematische Erlernen des Erkennens verschiedener Krankheitsmerkmale fällt in der Ausbildung meist dem Zeitdruck zu Opfer und es gibt nur wenige Hilfsmittel, die die angehenden Ärzte nutzen können, um ihre Fähigkeiten darin weiter zu entwickeln oder zu überprüfen.

Im Rahmen einer Masterarbeit wurde daher das Backend des Systems (d.h. vor allem die Klassifikationsmodelle und Segmentierungsfunktionalitäten) genutzt, um ein Web-basier-

tes Lernsystem zu realisieren, mit dem die Assistenzärzte die Erkennung und Segmentierung der Merkmale üben und testen können. Nach Auswahl eines Beispielbildes (es können auch eigene Bilder hochgeladen werden) können die verschiedenen Merkmale von Hand auf der Weboberfläche markiert und am Ende mit den Ergebnissen des Systems verglichen werden. Abbildung 7 zeigt den Zustand des Systems bei der Präsentation der einzelnen Ergebnisse, bei der die vom Benutzer eingegebenen Markierungen der „Ground truth“ des Systems gegenübergestellt werden. Eine umfangreiche Evaluationsstudie bestätigte das Potential eines ML basierten Lernsystems in der medizinischen Ausbildung, vor allem im Hinblick auf die Zeitrestriktionen und die Möglichkeit des unabhängigen Lernens von Assistenzärzten. Die Veröffentlichung der Ergebnisse ist unter dem Titel „Human-Centered Development of an ML-based Learning Tool for Medical Education in Ophthalmology“ angestrebt.

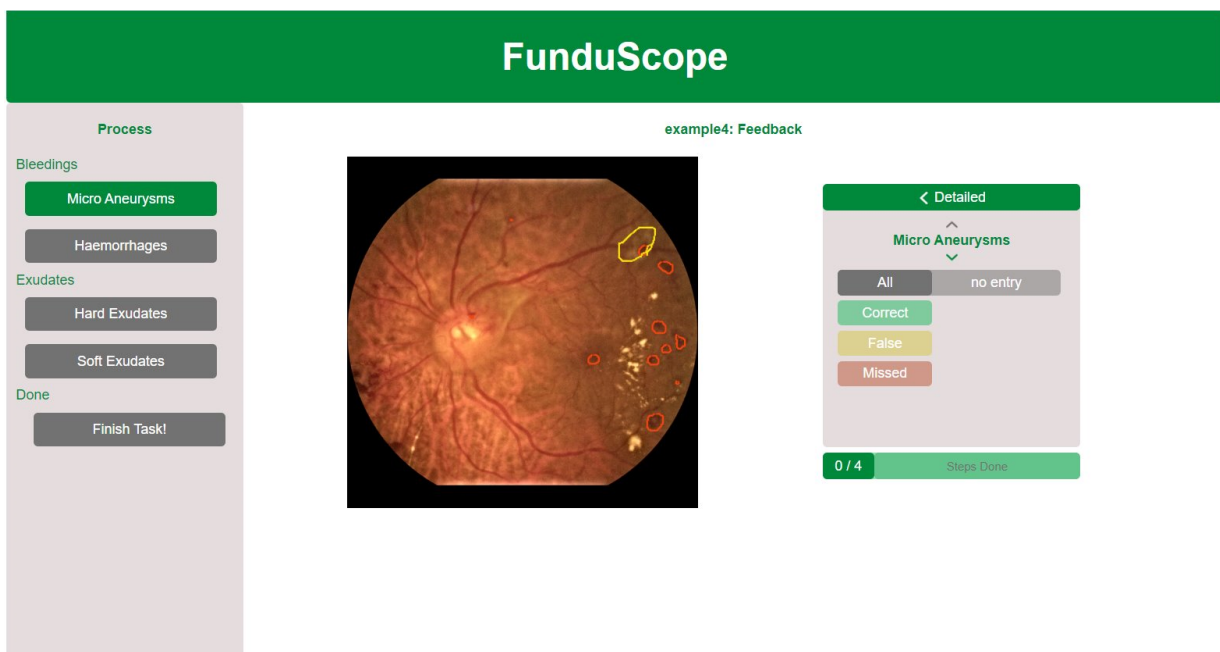


Abbildung 7 Oberfläche des intelligenten Tutorsystems

Dashboard

Nachdem im Projektverlauf zunehmend mehr annotierte Daten zur Verfügung standen wurde begonnen, verschiedene neuronale Modelle zu trainieren. Segmentierte OCT Bilder wurden genutzt, um Segmentierungsmodelle zu trainieren, die nicht nur wieder im Annotationstool weitere Annotationen erleichtern (siehe Arbeitspaket AP 5), sondern auch für weitere Unterstützungsfunktionalitäten eingesetzt werden konnten.

Basierend auf Interviews mit Klinikärzten wurde begonnen, einen weiteren Demonstrator „Dashboard“ (ebenfalls im Rahmen einer Masterarbeit) zu realisieren. Dieses Websystem soll den Ärzten einen schnellen Überblick über den bisherigen Krankheitsverlauf und speziell die Veränderungen zur letzten Visite ermöglichen. Zusätzlich nutzt es die angelernten Segmentierungsmodelle für eine 3D Ansicht der OCT Schichten und zur Berechnung der

Flüssigkeiten in den einzelnen Netzhautschichten und weiterer Biomarker. Ein Prognosemodell berechnet dann die besten Behandlungsoptionen und erklärt die zu erwartenden Verbesserungen.

Abbildung 8 zeigt die verschiedenen visuellen Komponenten des Systems:

- Über die Sidebar kann ein Patient inklusive Augenseite ausgewählt werden; die Auswahl kann auf Patienten eines Klinikums bzw. eines Krankheitsbildes beschränkt werden. Außerdem kann das Datum einer Visite gewählt werden, deren Daten angezeigt werden sollen.
- In VC1 werden neben vorhandenen Metadaten wie Geschlecht und Alter auch die bisherigen Visiten in einer Timeline angezeigt (inkl. des verabreichten Medikaments); außerdem findet sich hier eine Anzeige der „Position“ innerhalb einer Reihe von IVOM Uploads (z.B. vor zehn Tagen wurde die zweite von drei Spritzen verabreicht)
- VC2 enthält verschiedene Optionen zur Visualisierung des OCTs (siehe auch Abbildung 9): neben der Draufsicht („IRSLO“) können die verschiedenen Slices angezeigt und automatisch mit Hilfe des trainierten ML Modells segmentiert werden. Außerdem ist ein Vergleich der OCTs zweier Visiten möglich, wobei die Ansichten automatisch passend zueinander ausgerichtet werden und die Überlappung über einen Slider verändert werden kann. Eine dritte Möglichkeit stellt die 3D Rekonstruktion dar, die alle Schichten inklusive der Segmentierungen in einer beliebig drehbaren Darstellung zusammenfasst. Da die Dicke einiger Netzhautschichten medizinisch relevant ist, wurde eine weitere Sichtoption integriert, bei der zu jeder Schicht die jeweilige Dicke über dem IRSLO Bild farblich kodiert gezeigt wird; optional kann auch die Veränderung der Dicken einer Schicht zwischen zwei Visiten angezeigt werden.
- In VC 3 wird der bisherige Verlauf verschiedener relevanter Größen dargestellt: Visus, Flüssigkeiten und Augendruck. Die Darstellung ist ähnlich wie in der Timeline darüber. Zusätzlich kann hier auch eine Prognose für verschiedene Therapieoptionen eingeblendet werden. Das dahinterliegende Forecast-Modell wurde mit den wichtigsten Annotationen aus der Patientenakte plus allen berechneten Daten trainiert. Dazu gehören vor allem die Anzahl und Gesamtvolumen von Flüssigkeiten, die für die Therapieentscheidung entscheidend sind, und aufgrund aller vorhandenen OCT Annotationen rekonstruiert und berechnet wurden.
- VC 4 zeigt die Veränderung der Merkmale Visus, Flüssigkeiten und Augendruck zwischen den ausgewählten Visiten
- VC 5 enthält eine Behandlungsempfehlung, die sich auf die Ergebnisse des Prognosetools in Kombination mit den existierenden Leitlinien für die IVOM Behandlung stützt.
- VC 6 zeigt als Voreinstellung eine Begründung („Reasoning“) für die vorgeschlagene Entscheidung (es können auch andere Optionen ausgewählt werden) in Form der zu erwartenden Änderungen der medizinisch relevanten Merkmale. In

diesem Bereich können alternativ auch die einzelnen Annotationen aus der Patientenakte („Visit Diff“) oder die Zahlenwerte in den Schichtendicken („Mean Thickness“) gegenübergestellt werden

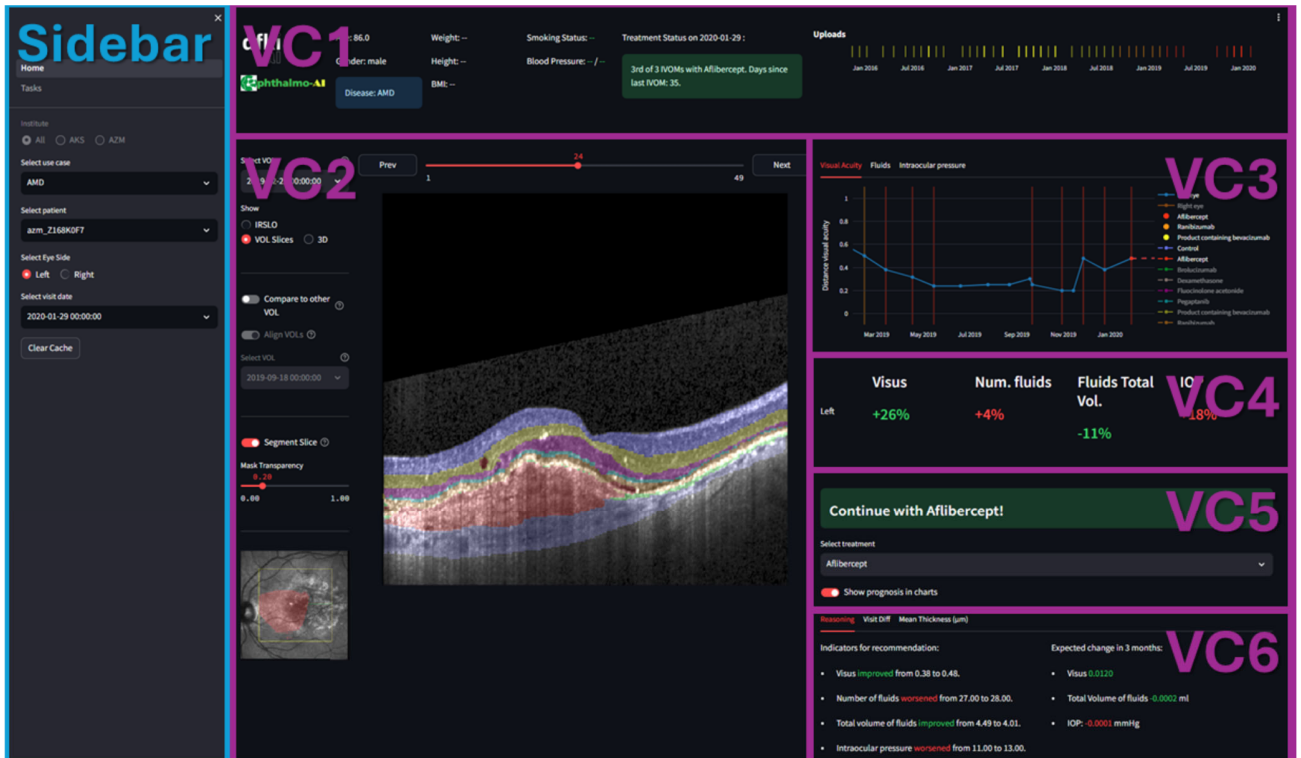


Abbildung 8 Visuelle Komponenten des Dashboards

Das Dashboard wurde ebenfalls an der Augenklinik Sulzbach evaluiert. Die Ergebnisse zeigen, dass das System überwiegend eine deutliche Verbesserung bei Effektivität, Informativität und User Experience gegenüber dem bisher in Kliniken benutzten System aufweist. Die innovativen visuellen Elemente (Timeline, Segmentierung, 3D Darstellung) wurden überwiegend positiv bewertet. Mangelndes Vertrauen in das System, v.a. bezüglich der ML-Ergebnisse zu Segmentierungen und Prognosen, ist ein kritischer Punkt, der allen neuen Systemen, und dabei insbesondere ML basierten Tools entgegengebracht wird. Die Studie zeigte, dass das Vertrauen durch längere Nutzung des Systems und die Möglichkeit von Feedback durch den Benutzer gesteigert werden kann. Feedback speziell zu den Segmentierungen des OCT waren vorgesehen, konnten aber aufgrund des Projektendes nicht mehr realisiert werden. Dieses Feedback könnte genutzt werden, um betroffene OCT Bilder wieder dem Annotationstool zuzuspielen, so dass die Annotation durch einen menschlichen Benutzer korrigiert werden kann. Teilergebnisse des Dashboards werden auf dem DOG Kongress 2024 in Berlin in zwei Vorträgen vorgestellt. Eine weitere Publikation zu den Ergebnissen des Dashboards ist in Vorbereitung.

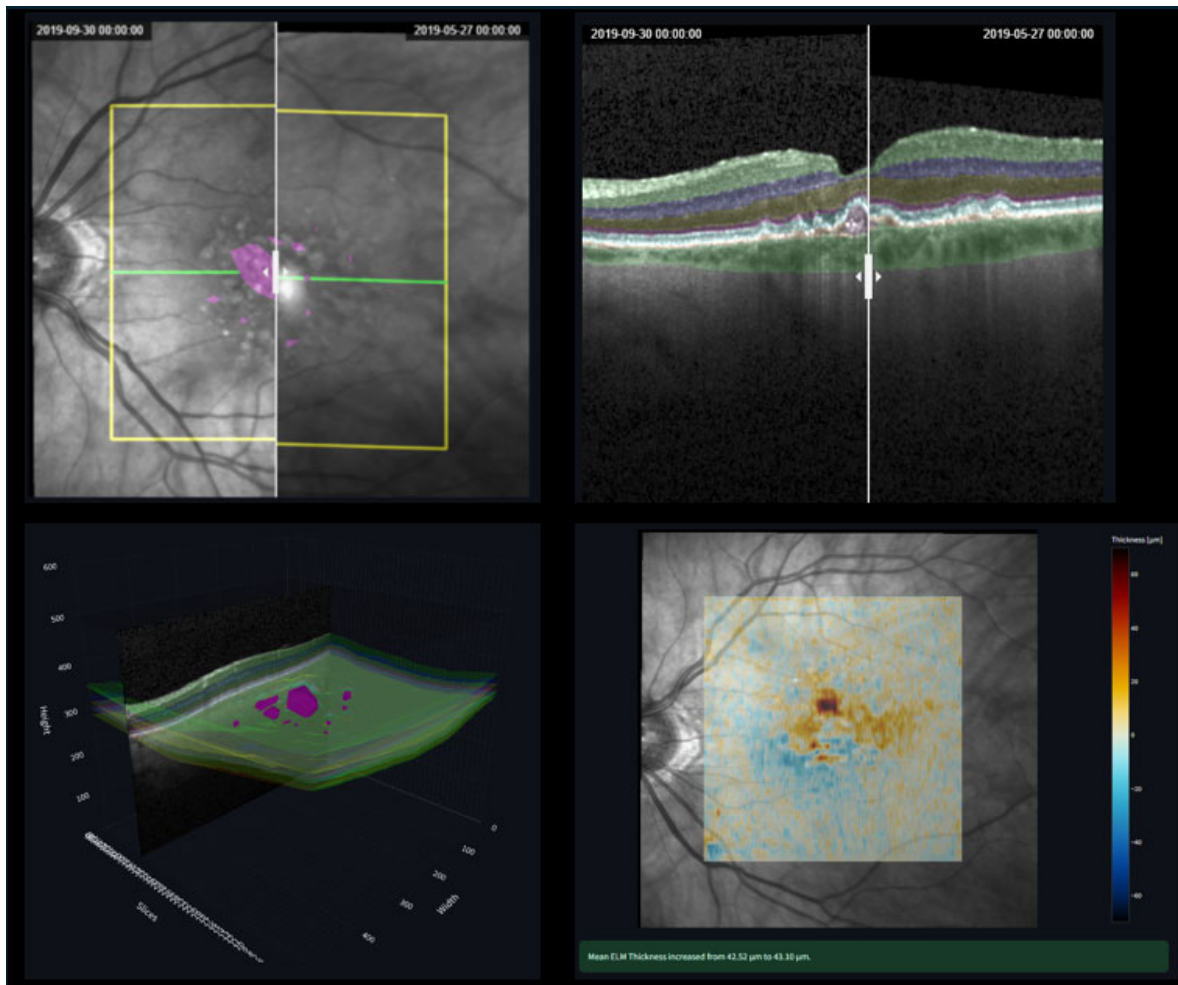


Abbildung 9 Visualisierungsoptionen (v.l.n.r.): Vergleich zweier IRSLO bzw Slices, 3D Darstellung und Darstellung der Schichtdicken

AP 8 Systemintegration und Demonstratorbereitstellung

Gesamtziel des Arbeitspaketes ist die Integration der technischen Komponenten zu zwei Demonstrator-Systemen und deren Bereitstellung für die klinischen Partner. Das IBMT hat mit dem DFKI und den klinischen Partnern die Möglichkeiten für eine direkte Anbindung der Demonstratoren an die KIS im Realbetrieb geprüft, die sich als sehr aufwändig herausstellte und den hohen Performanzansprüchen eines Lifesystems nicht genügt. Daher wurde eine Offline-Lösung spezifiziert, bei der die bereits verwendete Infrastruktur und Schnittstellen der Tools und Modelle genutzt werden (siehe Abbildung 10). Die Anwendungssysteme laden die Daten aus der zentralen Datenintegrationsplattform in lokale Datenbanken, die den verschiedenen lokalen Systemen für Training, Tests und Validierung zur Verfügung stehen. Das Ergebnis des Arbeitspakets waren die einsatzbereiten Demonstratoren am AZM und AKS als Meilenstein 3, die für die Piloterprobung genutzt wurden.

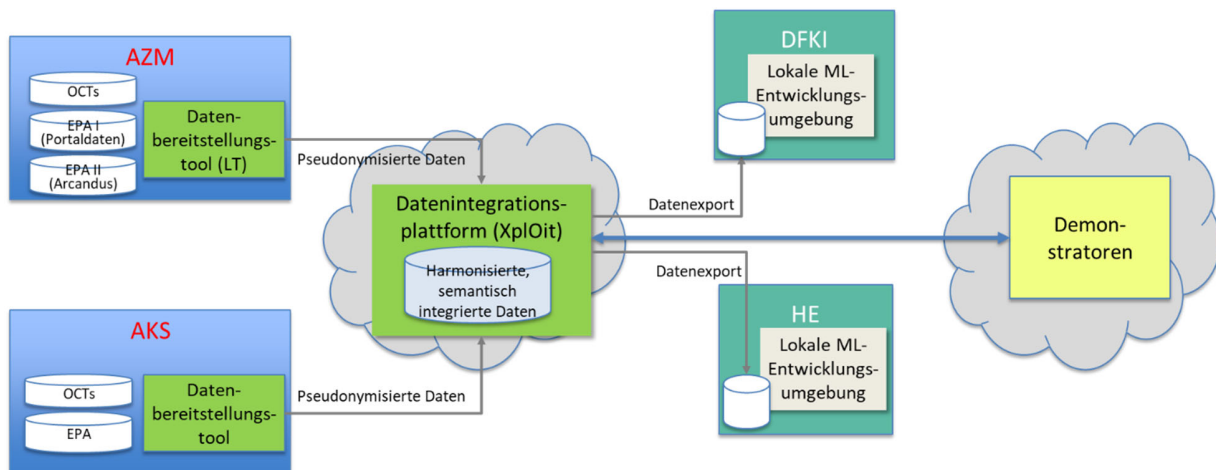


Abbildung 10: Datenverarbeitungsprozess in Ophthamo-AI.

AP 9 Piloterprobung

Ziel dieses Arbeitspakets war die Evaluierung der Prototypen, die in den Arbeitspaketen AP 6 und AP 7 vom DFKI entwickelt wurden. Dazu wurden drei Studien durchgeführt, bei denen das DFKI die Verfügbarkeit und Funktionalität der Webtools sichergestellt hat:

- Der Demonstrator „DR Grading“ wurde unter Leitung des Partners UdS¹ am Klinikum in Sulzbach evaluiert. Die Ergebnisse wurden in einer Publikation zusammengefasst, weitere sind in Arbeit.
- Das Lerntool wurde im Rahmen einer Masterarbeit am DFKI an den beiden Kliniken und in Zusammenarbeit mit weiteren Augenärzten von verschiedenen Kliniken Deutschlands evaluiert. Die Ergebnisse wurden in einer Publikation beschrieben.
- Der Prototyp „Dashboard“ wurde in Zusammenarbeit mit der UdS am Klinikum in Sulzbach evaluiert, auch diese Ergebnisse wurden (und werden) sowohl in der Masterarbeit am DFKI als auch in weiteren in Arbeit befindlichen Veröffentlichungen beschrieben.

Die Publikationen zu allen Studien sind in Abschnitt 6 „Veröffentlichungen“ aufgeführt.

2 Wichtigste Positionen des zahlenmäßigen Nachweises

Im Teilvorhaben des IBMT war der wesentliche Teil der Ausgaben durch Personalkosten entstanden. An zweiter Stelle folgten Ausgaben für vorhabensspezifische Abschreibungen, die insbesondere für die Anschaffung einer Server-Maschine entstanden. Die Kosten für Dienstreisen, Material und sonstige Sachkosten nehmen eine untergeordnete Rolle im

¹ Lehrstuhl für Arbeits- und Organisationspsychologie der Universität des Saarlandes

Vergleich zu den Personalkosten ein. Die folgende Tabelle zeigt die Ausgaben im Überblick.

Tabelle 1: Entstandene Kosten im Teilvorhaben.

Kostenart	Entstandene Kosten
Personal	560.747,03
Unteraufträge	0,00
Material	0,00
Reisen	8.740,19
Abschreibungen	1.046,21
Sonstige	2.254,90
Gemeinkosten	274.633,46
Summe	847.421,79

3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die durchgeführten F&E-Arbeiten sowie die dafür aufgewandten Ressourcen waren notwendig und angemessen, da sie der im Projektantrag beschriebenen Arbeitsplanung entsprachen. Alle wesentlichen in den Arbeitspaketen beschriebenen Aufgaben wurden erfolgreich bearbeitet. Bei den geplanten Kosten für Reisen konnten Mittel eingespart und für eine Erweiterung der Personalkosten genutzt werden.

Die in diesem Teilvorhaben des Ophthalmo-AI-Projektes realisierten informationstechnischen Funktionalitäten sind in dieser Form bisher so noch nicht realisiert worden, obwohl ein konkreter Bedarf vorhanden ist. Nur durch die Zusammenarbeit der Partner innerhalb dieses Verbundprojektes kam es zu einer Bündelung der notwendigen Kompetenzen, um die bestehenden wissenschaftlich-technischen Risiken einer solch höchst interdisziplinären Systementwicklung zur Unterstützung der Vorhersageforschung in der Medizin zu meistern.

4 Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit der Ergebnisse und der Erfahrungen

Die in Ophthalmo AI gewonnenen technischen und wissenschaftlichen Ergebnisse bilden die Grundlage für weitere Projekte im Rahmen der „KI in der Medizin“. Die Ergebnisse der Evaluationen und die Rückmeldung durch die medizinischen Projektpartner haben gezeigt, dass die entwickelten KI-Tools für die Unterstützung der medizinischen Lehre (Lern-tool) und bei Diagnose- und Therapieentscheidungen (Dashboard) wertvolle Hilfestellungen bieten können. Die wissenschaftlichen Arbeiten im Rahmen des Active Learning konnten bereits in weiteren Projekten u.a. zusammen mit dem DFKI Gesellschafter

Google genutzt und erweitert werden. Die Erfahrungen im Rahmen des Projekts im Hinblick auf automatische Segmentierungen und Auswertungen von ophthalmologischen Bildern, Design und Umsetzung von XAI Ansätzen und Design und Umsetzung von komplexen Dashboards zur Unterstützung des medizinischen Personals wurden in zahlreichen Publikationen veröffentlicht und auf Veranstaltungen präsentiert und bilden eine wertvolle Grundlage für weitere Forschungen, kommende Projekte und in der Lehre. Erste Anfragen zur Nutzung der Modelle liegen von der Firma TeleMedC vor, die sich weltweit auf Telemedizin spezialisiert hat und den Einsatz dieser Modelle für ein frühes Augen-Screening in medizinisch unterversorgten Gebieten (Indien, Australien) weltweit prüfen will. Auch andere Formen der Zusammenarbeit wie gemeinsame Projekte auf Basis der gewonnenen Erfahrungen werden aktuell mit ihnen diskutiert.

5 Während der Durchführung des Vorhabens dem Anwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens

Während der Durchführung des Vorhabens sind von anderer Seite keine Ergebnisse oder Fortschritte bekannt geworden, die in unmittelbarem Zusammenhang mit den Zielen unseres Teilvorhabens standen und es obsolet gemacht hätten.

6 Erfolgte oder geplante Veröffentlichung der Ergebnisse

Die wissenschaftlich-technischen Ergebnisse des Forschungsverbundes wurden und werden in Form von gemeinsamen Publikationen in Fachzeitschriften und Konferenzen verwertet. Diese sind in wissenschaftlichen Literaturdatenbanken gelistet und sichern langfristig die Sichtbarkeit der Projektergebnisse.

Publikationen im Zusammenhang mit dem Ophthalmo-AI Projekt:

- Bittner, S.-J., Barz, M., Profitlich, H.-J., Sonntag, D.: "Human-Centered Development of an ML-based Learning Tool for Medical Education in Ophthalmology" (2024, eingereicht)
- Leist, R., "Development of an AI-Driven Clinical Decision Support System for Treating Age-Related Macular Degeneration and Diabetic Retinopathy", Master Thesis an Universität des Saarlandes, Juni 2024 (Konferenzversion in Arbeit)
- Kadir, A., Alam, H., Maul, P., Profitlich, H.-J., Wolf, M., Sonntag, D.: "Modular Deep Active Learning Framework for Image Annotation: A Technical Report for the Ophthalmo-AI Project" CoRR <https://arxiv/abs/2403.15143> (2024)
- Nguyen, H.M.D., Alam, H., Nguyen, T., Srivastav, D., Profitlich, H.-J., Le, N., Sonntag, D.: "AI in Ophthalmology - Current Trends and Future Works", eingereicht bei Medical Image Analysis Journal, 2024

- Kadir, A., Alam, H., Sonntag, D.: „EdgeAL: An Edge Estimation Based Active Learning Approach for OCT Segmentation”. MICCAI (2) 2023: 79-89, CoRR <https://arxiv.org/abs/2307.10745> (2023), https://link.springer.com/chapter/10.1007/978-3-031-43895-0_8
- Alam, H., Nguyen, Duy M. H., Truong, M.T.N., Nguyen, T.A., Nguyen, B.T., Barz, M., Profitlich, H.-J., Than, N.T.T., Le, N., Xie, P., Sonntag, D.: „DRG-Net: Interactive Joint Learning of Multi-lesion Segmentation and Classification for Diabetic Retinopathy Grading. CoRR”, <https://arxiv.org/abs/2212.14615> (2022)
- Nguyen, H.M.D., Mai, T.T.; Than, N.T.T., Prange, A., Sonntag, D.: “Self-Supervised Domain Adaptation for Diabetic Retinopathy Grading using Vessel Image Reconstruction” Proc. of KI 2021, https://link.springer.com/chapter/10.1007/978-3-030-87626-5_26
- Rothaus K., Heimes-Bussmann B., Wolf M., Weichsel J., Rauch J., von Ehrlich-Treuenstätt G., Luc J., Kemmler-v. Saint Paul R., Hunsicker T., Profitlich H.-J., Boden K., Lommatzsch A., Kiefer S. „High-Level Annotation von OCT-Biomarkern im Rahmen des BMBF-Projekts “Ophthalmology-AI” unter Verwendung von Active-Learning”, DOG 2024, Jahrestagung der Deutschen Ophthalmologischen Gesellschaft DOG, Berlin, 10.-13.10.2024
- Ehrlich-Treuenstätt, G., Hunsicker T., Profitlich H.-J., Leist, R., Julich-Haertel, H., Rothaus, K., Rudolph C., Stanzel B., Lommatzsch A., Boden, K., „Über ein neu entwickeltes Dashboard zur KI-unterstützten Befundung von OCT-Schnittbildern bei altersbedingter Makuladegeneration und diabetischer Retinopathie“, DOG 2024, Jahrestagung der Deutschen Ophthalmologischen Gesellschaft DOG, Berlin, 10.-13.10.2024
- Rauch J., Moghaddam F., Rothaus K., Boden K., Julich-Haertel H., Schilk A., Luc J., McCrae P., Wolf M., Profitlich H.-J., Hunsicker T. and Kiefer S. „Advanced methods for ensuring optimal quality of data extracted from patient records for machine learning applications in ophthalmology”, gesundheit-gemeinsam 2024, GMDS-Jahrestagung, Dresden, 8.-13.9.2024
- Hunsicker, T., Profitlich, H.-J., “Caught my (A)eye: Effects of explanations of AI-based diagnostic support in ophthalmology”, Symposium “Insights into explainable AI: Theory, design, and applications”, 53. DGPs Congress / 15. ÖGP Conference, Wien, 2024

Weitere Publikationen zu den Ergebnissen der Demonstratoren unter der Federführung des DFKI und der UdS sind in Vorbereitung. Darüber hinaus wurde das Projekt in Pressemitteilungen² und als Ergebnissteckbrief³ zur Initiative „Adaptive Technologien für die Gesellschaft – Intelligentes Zusammenwirken von Mensch und Künstlicher Intelligenz“ des BMBF der Öffentlichkeit bekannt gemacht.

²<https://www.ibmt.fraunhofer.de/de/ibmt-presse-uebersicht-2021/presse-ophthalmo-ai-2021-05-04.html>,
<https://www.dfki.de/web/news/ophthalmo-ai>,
https://uk.dfki.de/DFKI_NEWS_ePaper/epaper-DFKI_News_51_d/#8
<https://iml.dfki.de/news/successful-completion-of-the-lighthouse-project-ophthalmo-ai/>

³ <https://www.interaktive-technologien.de/service/ergebnissteckbriefe/meKi>

Ophthalmology-AI – Intelligente, kooperative Diagnose- und Therapieunterstützung in der Augenheilkunde

Kurzbericht zu Teilvorhaben „Interaktives Maschinelles Lernen“

Förderkennzeichen 16SV8638

Motivation und Problemstellung:

Das Verbund-Projekt Ophthalmology-AI hatte das Ziel, in der Augenheilkunde durch eine effektive Zusammenarbeit von maschineller und menschlicher Expertise (interaktives Maschinelles Lernen IML) bessere Diagnose- und Therapieentscheidungen zu treffen. Dabei sollten Teile des Entscheidungsprozesses durch maschinelles Lernen unterstützt werden, um eine erhöhte Reliabilität und Objektivität zu erreichen. Ziel war es, eine vergleichbare Genauigkeit wie der kostenintensive Goldstandard zu erreichen und die Akzeptanz seitens der Anwender zu gewährleisten. Zusätzlich sollte der Diagnose- und Therapieprozess effizienter gestaltet werden, indem aus weniger Diagnoseschritten ein differenziertes Bild erstellt und dadurch die Erklärbarkeit im Mensch-Maschinen-Dialog sichergestellt wird. Das Teilvorhaben des DFKI mit dem Titel „Interaktives Maschinelles Lernen“ hatte das Ziel, intelligente Werkzeuge und Module zur Unterstützung von Diagnose- und Entscheidungsprozessen zu entwickeln, die durch eine effektive Zusammenarbeit von maschineller und menschlicher Expertise und unter Berücksichtigung der medizinischen Leitlinien eine vergleichbare Genauigkeit und Akzeptanz seitens der Anwender erreichen. Um dies zu erreichen, wurden auf technischer Seite nicht nur diagnostische ML Modelle erarbeitet, sondern auch Erklärbarkeitskomponenten (z.B. Visualisierungstools), die den Medizinern Feedback über die Entscheidungen des ML-Modells liefern und damit einen entscheidenden Beitrag zur Transparenz im Diagnose- und Therapieprozess leisten.

Derzeitiger Stand von Wissenschaft und Technik:

ML-Verfahren ermöglichen die automatische Extraktion relevanter Merkmale aus medizinischen Daten, was bei der Detektion und Diagnose von Erkrankungen unterstützt. In der medizinischen Bildgebung können ML-Algorithmen effektiv eingesetzt werden, um latente Merkmale in den Bilddaten zu erlernen und lokale Merkmale präzise zu beschreiben. Das unüberwachte Strukturlernen ermöglicht die Identifizierung latenter Merkmale in den Rohdaten, auch ohne vollständige oder lückenlose Trainingslabels. Bayes'sche Netzwerke bieten eine weitere Möglichkeit zur Verbesserung der Transparenz, indem sie bedingte Vorhersagewahrscheinlichkeiten verwenden und Wahrscheinlichkeiten mit konkreten Diagnosen verknüpfen. Dadurch kann eine präzisere Abschätzung der Wahrscheinlichkeit verschiedener Diagnosen basierend auf den extrahierten Merkmalen erfolgen.

Methodik

Im Projekt OphthalmologyAI dienen die annotierten und kodierte Patientenakten sowie OCT-Bilddaten als zentrale Ausgangsdaten. Die OCT-Bilder, die entscheidend für Therapieentscheidungen sind, werden von Fachpersonal aufwendig annotiert, insbesondere durch die Segmentierung der Netzhautschichten und Markierung von Symptomen wie Läsionen oder Zysten. Diese annotierten Daten werden genutzt, um maschinelle Lernmodelle zu trainieren, die den Annotierungsvorgang unterstützen und zur Therapieunterstützung, Berechnung von Biomarkern sowie zur Vorhersage des Krankheitsverlaufs eingesetzt werden. Intelligente Benutzerschnittstellen gewährleisten, dass das medizinische Fachpersonal diese Informationen nachvollziehbar und verständlich erhält, was die Akzeptanz und das Vertrauen in das KI-System fördert.

Ergebnisse

Das DFKI entwickelte ein intelligentes Tool zur Annotation von OCT Bildern, d.h. zur Segmentierung der Netzhautschichten und weiterer relevanter Merkmale. Bereits vorhandene Annotationen (z.T. auch durch den Projektpartner Heidelberg Engineering zur Verfügung gestellt) wurden genutzt, um ML Modelle

anzulernen, die den Prozess durch Vorschläge unterstützen, so dass diese nur noch überprüft und gegebenenfalls korrigiert werden müssen, was die Zeit für Annotationen drastisch verkürzen kann. Weitere intelligente Werkzeuge wie "Intelligent Scissors" wurden realisiert, um den Segmentierungsprozess zu erleichtern. Um eine möglichst optimale Qualität der Annotationen sicherzustellen, wurde ein aufwändiger Zertifizierungsprozess entworfen und durchgeführt. Zur weiteren Optimierung des gesamten Annotationsprozesses wurde ein Active Learning Framework implementiert, das sicherstellt, dass immer genau die OCT Bilder zur Annotation herangezogen werden, die den größten Wissensgewinn für das ML Modell darstellen.

Da der Annotationsprozess zeitaufwändig war, nutzte das Projekt öffentlich zugängliche Fundusbilder, um die ersten ML-Modelle und Webtools zu entwickeln. Das DFKI trainierte ein Modell, das den Demonstrator „DR Grading“ hervorbrachte, der zur Diagnose der Diabetischen Retinopathie dient. Das System unterscheidet fünf Schweregrade und erklärt seine Ergebnisse durch Segmentierungsmasken, die relevante Läsionen markieren. Dieses System wurde erfolgreich evaluiert und veröffentlicht. Zudem wurde auf Basis der Modelle das Lerntool „FunduScope“ entwickelt, das Augenärzten in der Ausbildung hilft, ihre Diagnosefähigkeiten zu verbessern. Auch dieses Tool wurde evaluiert und publiziert.

Nachdem durch den projektspezifischen Annotationsprozess ausreichend OCT-Annotationen vorlagen, wurden ML-Modelle trainiert, die medizinisch relevante Merkmale wie Schichtdicken und Flüssigkeiten unterstützten, welche entscheidend für Therapieentscheidungen sind. Diese Modelle ermöglichten die Berechnung wichtiger Biomarker und wurden verwendet, um Prognosemodelle zu entwickeln, die die zukünftige Entwicklung von Aspekten wie Flüssigkeiten, Visuswerten und Augendruck in Abhängigkeit von Therapieentscheidungen vorhersagen. Diese Modelle wurden in ein interaktives Dashboard integriert, das alle relevanten Daten visualisiert. Es bietet u.a. Timelines zum Therapieverlauf, 3D-Rekonstruktionen der Netzhaut und Vergleiche von OCT-Bildern, um den bisherigen Therapieerfolg und die Wirkung der verabreichten Medikamente zu überwachen.

Neben diesen Visualisierungsaspekten enthält das Dashboard eine Handlungsempfehlung, d.h. welche Entscheidung (Medikamentengabe, nur Kontrolle, Abbruch der Behandlung) auf Basis des Prognosemodells die beste Entwicklung v.a. des Visuswerts verspricht. Zur Erklärung der Empfehlung wird außerdem die zu erwartende Entwicklung der Biomarker (Visus, Flüssigkeiten, Schichtdicke) sowohl in Text- als auch in grafischer Form gezeigt. Der Benutzer kann sich auch die zu erwartenden Werte für andere mögliche Entscheidungen zum Vergleich anzeigen lassen. Das Dashboard wurde sowohl innerhalb des Konsortiums als auch im Rahmen einer Studie am Fachklinikum Sulzbach von den medizinischen Fachkräften größtenteils sehr positiv bewertet. Die Ergebnisse werden im Rahmen einer Masterarbeit und in weiteren Publikationen veröffentlicht.

Schlussfolgerungen/Verwertung

Die durchgeführten Studien liefern wertvolle Erkenntnisse vor allem in Bezug auf eines der zentralen Themen der KI, XAI, der Erklärbarkeit von Ergebnissen. Sie zeigten beispielsweise, dass Erklärungen nicht per se und in jeder Form positiv aufgenommen werden. Die Benutzer möchten entscheiden können, wann und in welchem Umfang Erklärungen gegeben werden, da sie in (subjektiv für sie) einfachen Fällen von Erklärungen als überflüssig und zusätzlich zeitraubend empfinden. Ein weiterer wesentlicher Aspekt von KI Systemen ist Vertrauen. Die Mediziner zeigten sich grundsätzlich sehr skeptisch in Bezug auf Modellergebnisse, vor allem, da ihnen das System nicht vertraut war (längere Nutzung verspricht hier Verbesserung) und zum anderen, weil sie keinen Einfluss auf die Modelle (v.a. der Segmentierungsmodelle) nehmen konnten. Daher wurde ein Feedbackmechanismus integriert, der es dem Benutzer ermöglicht, Anmerkungen zu einer Schichtsegmentierung zu geben, so dass diese (im Rahmen des Active Learning Prozesses) zur erneuten Annotation herangezogen wird.

Die Arbeiten und Ergebnisse wurden in zahlreichen Publikationen des DFKI veröffentlicht bzw. bilden die Grundlage für Veröffentlichungen der Partner. Des Weiteren wurden einige bereits in neuen Projekten (zum Thema Active Learning) genutzt bzw. bilden wichtige Grundlagen für weitere Forschungen und Anschlussprojekte des DFKI beispielweise zum Thema XAI und HCI.

