

Vorhaben	KIAFlex Interaktive KI-Assistenz zur prädiktiven und flexiblen Steuerung im Entlass- und Überleitungsmanagement
Titel	Abschlussbericht Teil II: Eingehende Darstellung
Förderkennzeichen	16SV9005
Zuwendungsempfänger	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
Bewilligungszeitraum	01.08.2022 - 30.11.2025
Autoren	Christian Zeyen, Joscha Grüger und Ralph Bergmann
Erstellungsdatum:	15.01.2026

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt unter dem Förderkennzeichen 16SV9005 gefördert. Dem Muster zu Nr. 4.1 NKBF 2017 (BMFTR-Vordr. 0701/07.25) folgend, handelt es sich bei diesem Dokument um Teil II, eine eingehende Darstellung des Vorhabens. Die Teile I und III liegen als gesonderte Dokumente vor. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Inhaltsverzeichnis

1	Verwendung der Zuwendung	1
1.1	Bedarfsanalyse und nutzerorientierte/partizipative Gestaltung (AP 2)	1
1.2	(Semantische) Datenintegration und Wissensmodelle (AP 3)	4
1.3	KI-Modelle zur kontinuierlichen Prädiktion des Nachsorgebedarfs (AP 4)	5
1.3.1	Aufbau der Fallbasis (AP 4.1)	7
1.3.2	Entwicklung von Ähnlichkeitsmaßen (AP 4.2)	8
1.3.3	Umsetzung der Prognosefunktionalität (AP 4.3)	9
1.3.4	Implementierung der Erklärungskomponente (AP 4.4)	17
1.4	Proaktive Kommunikation und Dokumentation im Prozess (AP 5)	17
1.5	Systemintegration und Entwicklung des Demonstratorsystems (AP 6)	18
2	Wichtigste Positionen des zahlenmäßigen Verwendungsnachweises	20
3	Notwendigkeit und Angemessenheit der geleisteten Arbeit	21
4	Verwertbarkeit	22
4.1	Wirtschaftliche Erfolgsaussichten	22
4.2	Wissenschaftliche Erfolgsaussichten	22
4.3	Wissenschaftliche und wirtschaftliche Anschlussfähigkeit	23
5	Bekannt gewordener Fortschritt	24
6	Erfolgte und geplante Veröffentlichungen der Ergebnisse	25

1 Verwendung der Zuwendung

Im Folgenden werden die Ziele des Vorhabens und die erzielten Ergebnisse im Einzelnen anhand der Arbeitspakete mit Beteiligung des DFKI dargestellt.

Der Fokus des DFKI lag im Projekt KIAFlex auf der automatisierten Prädiktion des Nachsorgebedarfs von Patient:innen, sodass erste Schritte der Entlassung bereits bei Aufnahme vorbereitet werden können. Durch die frühzeitige und systematische Erfassung des Nachsorgebedarfs wird eine optimale individuelle Versorgung der Patient:innen ermöglicht und Versorgungskontinuität sichergestellt.

1.1 Bedarfsanalyse und nutzerorientierte/partizipative Gestaltung (AP 2)

Seitens des DFKI wurde zur Bedarfsanalyse und Szenarioentwicklung (AP 2.1) ein Fragenkatalog zur strukturierten Wissensakquisition erstellt und den medizinischen Partnern zur Beantwortung vorgelegt. Die Beantwortung der Fragen sowie die Ergebnisauswertung waren ein iterativer Prozess. Ergebnisse aus den ersten Iterationen ermöglichten ein gemeinsames Verständnis des Forschungsgegenstandes und die Definition von konkreteren Use Cases. Hierzu wurde zusätzlich, im Sinne der Durchführung von Co-Creation-Aktivitäten (AP 2.2), auch eine dreiteilige Workshop-Reihe mit den Sozialdiensten der am Projekt beteiligten Universitätskliniken durchgeführt. Ziel war es, die Anforderungen aus der Praxis systematisch zu erfassen und eine fundierte Grundlage für die Ausgestaltung der Lösung zu schaffen. In den Workshops wurde strukturiert erhoben, wie die Festlegung von Nachsorgemaßnahmen aktuell erfolgt, welche Parameter dabei berücksichtigt werden und welche Indikatoren in der Entscheidungsfindung eine zentrale Rolle spielen. Ebenso wurden Fallstricke im Prozess sowie typische Schwierigkeiten identifiziert. Ein weiterer Schwerpunkt lag auf den Problemen und Herausforderungen in der täglichen Arbeit der Sozialdienste. Dazu zählen unter anderem aufwendige Kontaktschleifen, hohe Abstimmungsaufwände sowie Rahmenbedingungen wie Datenschutzerfordernungen, die telefonische Rückfragen häufig erschweren. Zudem wurde herausgearbeitet, welche Arbeitsschritte besonders zeitintensiv sind und an welchen Stellen Entlastungspotenziale bestehen. Abschließend wurden gemeinsam Bedarfe und Ideen für KI-basierte

Hilfsmittel diskutiert, die Mitarbeitende im Sozialdienst gezielt unterstützen und wiederkehrende Aufgaben effizienter gestalten könnten.

Zusammenfassend wurden die folgenden für die Nachsorgeplanung und insbesondere die Konkretisierung von Bedarfsmaßnahmen relevanten Patientendaten identifiziert:

- Patientenalter
- Diagnosen
- Lebenssituation
- Psychische Verfassung
- Patienteneigene Ressourcen für die Versorgung
- Barthel-Index
- Pflegeanamnese
- OP-Bericht
- Ärztlicher Verlauf
- Beatmungsprotokoll
- Mikrobiologie/Besiedelung
- Entlassdatum im Anforderungskonsil für Sozialdienst
- Medikamentenplan
- Demenz
- ISAR-Score
- Prozeduren
- Pflegegrad
- ICF-Klassifikation

Auf Basis der gewonnenen Erkenntnisse wurde ein Konzept zur Darstellung des aktuellen Nachsorgeprozesses und dessen Zustand sowie der aktuellen Prognose des Nachsorgebedarfs erarbeitet (AP 2.2). Der Fokus bei der Entwicklung einer Darstellung lag hierbei auf Verständlichkeit und Nachvollziehbarkeit, um den Nutzen für das anwendende Personal zu gewährleisten. Es wurden Mockups zur Erweiterung von Entlassmanagementsoftware um eine KI-basierte Nachsorgeplanung erstellt, welche als Diskussionsgrundlage mit den klinischen Partnern dienen, um einen anwendungsorientierten Konsens zu erzielen. Abbildungen 1.1 und 1.2 zeigen exemplarisch zwei Mockups von zentralen Benutzeroberflächen. Abbildung 1.1 zeigt eine Übersicht von Patient:innen, für die bezüglich der Nachsorgeplanung prognostizierte Änderungen proaktiv durch die Software gemeldet werden. Im vorliegenden Beispiel wird eine voraussichtlich notwendige Verlängerung des Krankenhausaufenthaltes für einen Patienten gemeldet. Per Mouseover wird eine entsprechende kurze textuelle Begründung angezeigt. Abbildung 1.2 zeigt eine Detailansicht eines Patienten, in der die konkreten Nachsorgemaßnahmen geplant werden können. In dieser Ansicht werden durch den KI-Assistenten noch zu

Abbildung 1.1 Mockup der Entlassmanagementsoftware - Übersicht der Nachsorgeplanung

Name	Vorname	Entlassdatum (geplant)	Entlassdatum (prognostiziert)	Planungsstatus	Nachrichten	Detailansicht
Müller	Max	20.05.2023	23.05.2023	60%	[Icon]	[Icon]
Frey	Andrea			95%	[Icon]	[Icon]
Kruger	Lukas			80%	[Icon]	[Icon]
Wirth	Anne	31.05.2023	31.05.2023	75%	[Icon]	[Icon]

Patient: Max Müller
Begründung: Entzündungsparameter
In den letzten Tagen haben sich diverse Entzündungs-parameter des Patienten verschlechtert. Aus diesem Grund ist eine Entlassung zu einem späteren Zeitpunkt wahrscheinlich.

planende Nachsorgemaßnahmen vorgeschlagen. Für jeden Vorschlag wird eine Signifikanz zusammen mit einer kurzen Begründung angegeben. Ein Vorschlag kann hier beispielsweise durch den Verweis auf ähnliche Patient:innen begründet werden. Bei Bedarf können die Eigenschaften der Patient:innen im Detail miteinander verglichen werden, sodass ähnliche Nachsorgemaßnahmen geplant oder begründete Abweichungen in der Nachsorge getroffen werden können. Die durch den Sozialdienst tatsächlich geplanten Nachsorgemaßnahmen werden abschließend dem KI-Assistenten gemeldet, sodass dieser aus übernommenen Vorschlägen und getroffenen Anpassungen von Vorschlägen lernt und dadurch die Prädiktion kontinuierlich verbessert.

Abbildung 1.2 Mockup der Entlassmanagementsoftware – Detailansicht der Nachsorgeplanung

Geplante Nachsorgemaßnahmen			
Maßnahme	Termin	Anmerkung	
Reha	21.05.2023 – 21.07.2023	Rehabesuch vollständig geplant.	
Nachsorgemaßnahme 1	21.05.2023	Nachsorgemaßnahme 1 vollständig geplant.	

Zu planende Nachsorgemaßnahmen			
Maßnahme	Termin	Anmerkung	
Nachsorgemaßnahme 2	unbekannt	Es wird auf Antwort von Dr. XYZ gewartet.	
Nachsorgemaßnahme 3	21.07.2023	Termin noch nicht 100% sicher.	

Vorgeschlagene Nachsorgemaßnahmen			
Maßnahme	Signifikanz	Anmerkung	Hinzufügen
Nachsorgemaßnahme 4	95%	Hohe Übereinstimmung mit Patient 34764. Klicken Sie hier , um Einsicht in diesen Patienten zu erhalten.	[+]
Nachsorgemaßnahme 5	90%	Hohe Übereinstimmung mit Patient 58932. Klicken Sie hier , um Einsicht in diesen Patienten zu erhalten.	[+]
Nachsorgemaßnahme 6	45%	Ähnliche Patienten haben diese Nachsorgemaßnahme auch erhalten.	[+]

Nachsorgemaßnahme anpassen und zur Planung hinzufügen.

1.2 (Semantische) Datenintegration und Wissensmodelle (AP 3)

Die Planung und Beantragung der Datennutzung stellte sich als langwieriger Prozess dar. Insgesamt haben die Einreichung und Genehmigung der Ethikanträge, sowie die Klärung der rechtlichen Aspekte für die Datennutzung und -weiterverarbeitung deutlich mehr Zeit erfordert als geplant und somit den Start von Arbeiten, für die das Vorhandensein von Daten erforderlich war, verzögert. Zur Kompensation der Verzögerung der technischen Umsetzung wurden konzeptionelle Arbeiten vorgezogen. Nach erfolgreichem Abschluss dieses Prozesses konnte der projektinterne Datenaustausch stattfinden und die Analyse der Datenqualität sowie die Datenaufbereitung (AP 3.2) durchgeführt werden. Anschließend wurden die Daten mithilfe von Expertenwissen semantisch angereichert (AP 3.3).

In Bezug auf die Eignung der Daten für die nachfolgenden Arbeitspakete war eine umfassende Aufbereitung der Daten zur Verbesserung der Datenqualität notwendig. Von den am Projekt beteiligten klinischen Partnern Universitätsklinikum Bonn (UKB) und Universitätsmedizin Mannheim (UMM) wurden alle stationären Behandlungsfälle im DRG-Entgeltbereich (somatische Medizin) mit Entlassung von 01.01.2019 bis 31.12.2023 von volljährigen Patient:innen in einem Datensatz je Klinik bereitgestellt. Ausgeschlossen wurden alle stationären Behandlungsfälle im PEPP-Entgeltbereich (Psychiatrie, Psychosomatik, Gerontopsychiatrie), da hier für die Mitarbeitenden des Sozialdienstes andere Rahmenbedingungen gelten. Die Anzahl von Fällen der erhaltenen Datensätze und die resultierende Anzahl nach Bereinigung der Daten können Abbildung 1.3 entnommen werden. Der Datensatz

Abbildung 1.3 Datensätze der UKB und UMM aus den Jahren 2019 – 2023

	UKB	UMM
MAIN	219.511 Fälle 180.339 Fälle bereinigt*	179.381 Fälle 179.372 Fälle bereinigt*
SOZ	52.116 Fälle 39.319 Fälle mit Status = vidiert 36.689 Fälle bereinigt* 32.105 Fälle (ca. 18%) mit Nachsorge	Keine Daten vorliegen
ICD	1.831.044 Einträge	1.273.577 Einträge (keine Zeitstempel)
OPS	1.401.165 Einträge 1.333.945 Einträge bereinigt*	1.026.424 Einträge 1.024.108 Einträge bereinigt*
FAB	367.599 Einträge 356.136 Einträge bereinigt*	232.533 Einträge 232.404 Einträge bereinigt*

* Nach dem Entfernen von Einträgen mit inkonsistenten Werten nach dem Abgleich der Datensätze (z.B. IDs, Zeitstempel)

MAIN umfasst hierbei die Patientendaten, die auf Basis des § 21 des Krankenhausentgeltgesetzes durch die Kliniken erfasst werden. SOZ enthält die Falldaten des Sozialdienstes mit den tatsächlich durchgeführten Nachsorgemaßnahmen. Diese Informationen lagen nur für die Fälle der UKB vor. Die

weiteren Daten enthielten ergänzende Informationen zu den Diagnosen (ICD), Prozeduren (OPS) und Verlegungen auf Fachabteilungen (FAB).

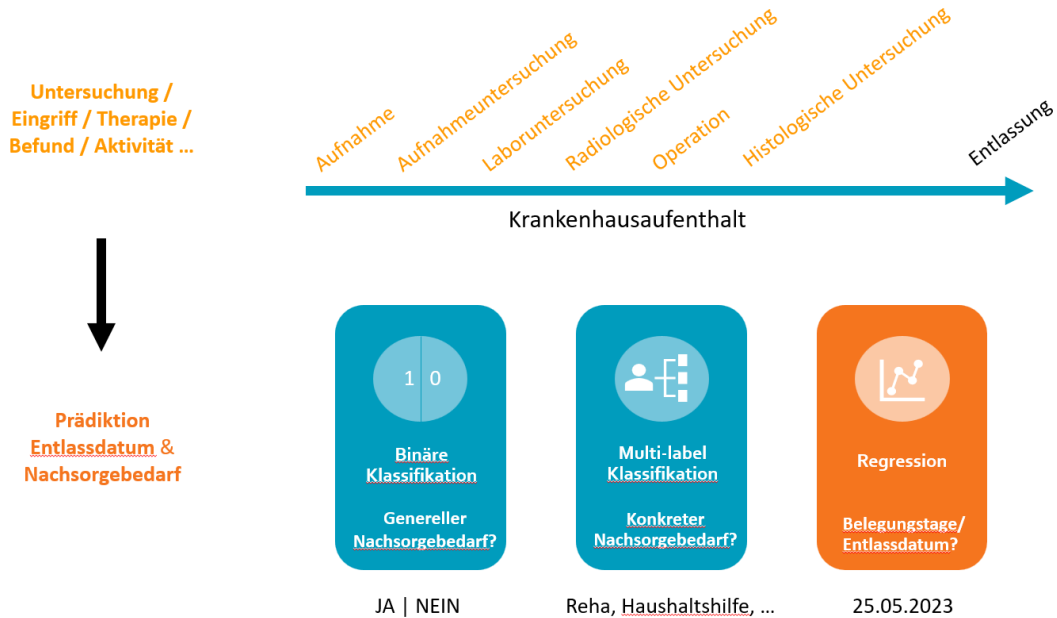
In der Datenaufbereitung und -validierung wurden zunächst die verschiedenen Informationen eines Behandlungsfalls zusammengeführt. Fälle mit fehlenden Werten bei erforderlichen Attributen oder inkonsistenten Werten wie Zeitstempeln wurden hierbei herausgefiltert. Attribute mit Listen von Werten, welche mit einem Zeitstempel versehen werden konnten, wie Fachabteilungen (Verlegungskette), Diagnosen oder Prozeduren, wurden entsprechend im Datensatz abgebildet, sodass die Prognose des Nachsorgebedarfs zu unterschiedlichen Zeitpunkten mit einer unterschiedlichen Informationsmenge mit den historischen Patientendaten simuliert werden konnte. Ergänzend zu den bereitgestellten Daten wurden mithilfe von zusätzlichem Domänenwissen weitere Informationen aus den vorhandenen Daten abgeleitet. So wurde unter anderem die Intensivpflichtigkeit aus dem Stationsaufenthalt abgeleitet oder der Pflegegrad aus den OPS-Codes extrahiert.

1.3 KI-Modelle zur kontinuierlichen Prädiktion des Nachsorgebedarfs (AP 4)

In diesem Arbeitspaket wurden für die Prädiktion des Nachsorgebedarfs verschiedene Ansätze erforscht. Für die Prädiktion mit fallbasiertem Schließen (engl. Case-Based Reasoning, CBR) wurde zunächst eine Wissensbasis modelliert (AP 4.1) und spezifische Ähnlichkeitsmaße entwickelt (AP 4.2). Anschließend erfolgte die Umsetzung der Prognosefunktionalität (AP 4.3) sowohl für verschiedene CBR-Modelle als auch zu Vergleichszwecken für verschiedene Modelle des Maschinellen Lernens (ML). Für die CBR-Modelle wurde zudem eine Erklärungskomponente entwickelt (AP 4.4) und im Rahmen der prototypischen Implementierung validiert.

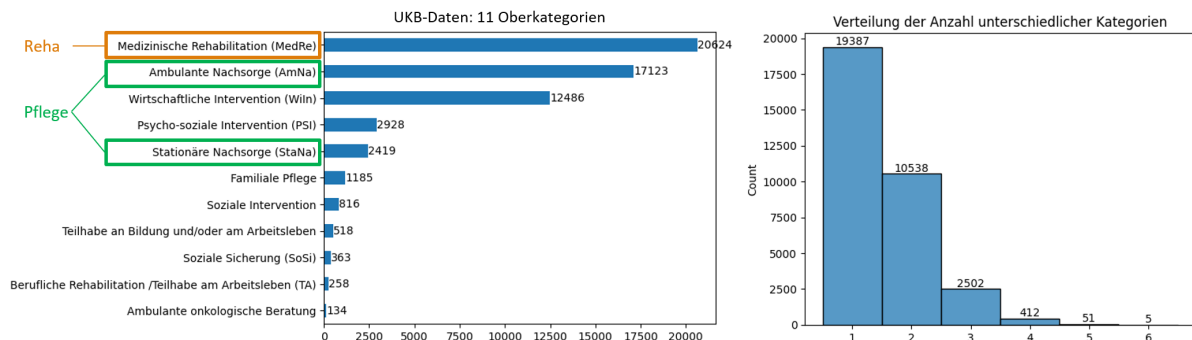
Im Allgemeinen stützen sich die KI-basierten Prädiktionen auf die Annahme, dass ähnliche Fälle einen ähnlichen Nachsorgebedarf haben. Demzufolge wurde eine retrospektive Betrachtung von Behandlungsfällen verfolgt und durch die KI-Prädiktionsmodelle nach Ähnlichkeiten und Mustern untersucht. Insgesamt wurden drei verschiedene Arten von Prädiktionen untersucht. In Abbildung 1.4 werden die Prädiktionsarten anhand von Beispielen veranschaulicht. Basierend auf den Patientendaten, die im Verlauf des Krankenhausaufenthaltes kontinuierlich erhoben werden, können ab dem Zeitpunkt der Aufnahme kontinuierlich Prädiktionen des Entlassdatums und des Nachsorgebedarfs durchgeführt werden. Es wird angenommen, dass die Prädiktionsgüte mit fortschreitendem Krankenhausaufenthalt zunimmt, sofern die relevanten Informationen zeitnah durch das klinische Personal digital erfasst werden. Voraussetzung für die Nachsorgeplanung ist die Festlegung eines voraussichtlichen Entlassdatums, welches mithilfe von Regressionsmodellen vorhergesagt werden kann. Ergänzend dazu kann

Abbildung 1.4 KI-basierte kontinuierliche Prädiktion des Nachsorgebedarfs



Mithilfe einer binären Klassifikation der generelle Nachsorgebedarf vorhergesagt und im Bedarfsfall mithilfe einer Multi-Label-Klassifikation letztlich der konkrete Nachsorgebedarf auf verschiedenen Granularitätsebenen prognostiziert werden.

Abbildung 1.5 Vorkommen von Nachsorgemaßnahmen nach Oberkategorie



Für die Prädiktion des Nachsorgebedarfs wurde der Fokus auf den Datensatz der UKB gelegt, da dort Informationen zum Nachsorgebedarf unmittelbar zur Verfügung standen und somit eine automatische Evaluation der Prädiktionsgüte ermöglicht wurde. Der Datensatz der UMM konnte zusätzlich für die Prädiktion des Entlassdatums genutzt werden. Abbildung 1.5 zeigt die Anzahl an Vorkommen von Nachsorgemaßnahmen nach Oberkategorie in den Falldaten der UKB. Es wurden insgesamt 11

Oberkategorien in den Daten unterschieden. Von den ca. 32.000 Fällen mit erhaltener Nachsorge macht medizinische Rehabilitation rund 63% aller Maßnahmen aus. In ca. zwei Dritteln aller Fälle mit Nachsorge wurde genau eine Maßnahme dokumentiert. In den Fällen wurden jeweils bis zu sechs unterschiedliche Maßnahmen aufgezeichnet. Bei der Verteilung der Oberkategorien wurden mehrfache Vorkommen unterschiedlicher Maßnahmen derselben Oberkategorie je Fall mehrfach gezählt.

1.3.1 Aufbau der Fallbasis (AP 4.1)

Die Grundlage für eine fallbasierte Prädiktion bildet die Fallbasis. Abbildung 1.6 gibt einen Überblick über die genutzten Attribute aus den Behandlungsfällen, eine kurze Beschreibung der Attributausprägungen und je einen exemplarischen Attributwert in einem fiktiven Beispielfall. Zu beachten ist, dass aus den Daten kein expliziter Nachweis über nicht erkannten oder zu spät erkannten Nachsorgebedarf abgeleitet werden kann. Ferner waren für die Nachsorgeplanung wesentliche Informationen, welche im Rahmen der Bedarfsanalyse identifiziert wurden, wie die Lebenssituation, Pflegeanamnese, Barthel-Index oder patienteneigene Ressourcen für die Versorgung, nicht im initialen Datensatz enthalten. Durch die UKB konnten im späteren Projektverlauf ergänzende Daten aus einer manuellen Auswertung der Pflegeanamnese für eine ausgewählte Patientengruppe (geriatrische Patient:innen) bereitgestellt werden.

Abbildung 1.6 Struktur der Fälle in der Fallbasis

Attribut	Ausprägungen	Beispielfall
key_FallNr	ID des Falls	bssdh-f1b5-kgj07
Geschlecht	m,w	w
AlterJ	Alter in Jahren >= 18	85
TageVorstat	Anzahl an Tagen vorstationär	0
AufnAnlass	A, E, G, N, V	N
AufnGrund	Vierstelliger Zahlencode: 0101 - 0807	0107
AufnAbt	Zahlencode: 100 - 3752	1600
AD_Stunde	Zeitpunkt der Aufnahme auf Stunde gerundet	7
Diagnosen	Liste der ICD-Codes	R26, I10, ...
Prozeduren	Liste der durchgeführten Prozeduren (OPS-Codes)	3-200, 8-550, ...
Pflegegrad	Aktueller Pflegegrad: 0-5	2
Verlegungskette	Liste der besuchten Fachabteilung	HA1600, HA3600, HA1600
Intensiv	Intensivpflichtigkeit während Aufenthalt: Ja,Nein	Nein
Belegungstage	Anzahl an Tagen	5
LeistungErhalten	Nachsorgebedarf: Ja,Nein	Ja
DurchgefLeistung	Liste der durchgeführten Nachsorgemaßnahmen (Auswahl aus 107 Maßnahmen)	Ambulante Nachsorge (AmNa): Haushaltshilfe / Familienpflege, Ambulante Nachsorge (AmNa): Hausnotruf, Ambulante Nachsorge (AmNa): Pflegedienst, Ambulante Nachsorge (AmNa): Pflegehilfsmittel

Eine Auswertung der vorliegenden Falldaten zeigte, dass sich der dokumentierte Nachsorgebedarf mit den im Datensatz enthaltenen Attributen nicht verlässlich erklären ließ. In einer Vielzahl von Fällen fanden sich große Patientenkohorten mit identischen Attributausprägungen, die sowohl Behandlungsfälle ohne als auch mit dokumentiertem Nachsorgebedarf umfassten. Nach eingehender Analyse im Konsortium wurde dieses Ergebnis auf drei wesentliche Einflussfaktoren zurückgeführt: (1) nicht dokumentierte Nachsorge bzw. Nachsorge, die nicht oder fehlerhaft angefordert wurde, (2) der individuelle, in den verfügbaren Daten nicht abgebildete Wille der Patientin bzw. des Patienten, sowie (3) kontextuelle Rahmenbedingungen wie die häusliche Situation, die im Datensatz nicht enthalten sind. Abbildung 1.7 illustriert dies anhand eines Beispiels mit sehr ähnlichen Behandlungsfällen, die sich hinsichtlich der Nachsorge unterscheiden und zugleich signifikante Abweichungen in der Anzahl der Belegungstage aufweisen.

Abbildung 1.7 Unterschiedlicher Nachsorgebedarf bei sehr ähnlichen Fällen (Epilepsie)

Geschlecht	Alter	TageVorstat	AufnAnlass	AufnGrund	AufnAbt	AD_Stunde	AufnDiagn	Diagnosen	AnzDiagnGesamt	Prozeduren	AnzProzGesamt	Pflegegrad	Verlegungskette	Intensiv	Belegungstage	LeistungErhalten	DurchgefLeistung
m	80	0	E	0101	3700	8	[G40.2]	[G40.2, Z11]	2	[1-210]	1	<NA>	[HA3700]	Nein	12	Nein	
m	51	0	E	0101	3700	8	[G40.2]	[G40.2, Z11]	2	[1-210]	1	<NA>	[HA3700]	Nein	5	Nein	
m	46	0	E	0101	3700	8	[G40.2, Z11]	[G40.2, Z11]	2	[1-210]	1	<NA>	[HA3700]	Nein	15	Ja	[Psycho-soziale Intervention (PSI): Bewältigun...
m	38	0	E	0101	3700	9	[G40.2, Z11]	[G40.2, Z11]	2	[1-210]	1	<NA>	[HA3700]	Nein	2	Nein	
m	33	0	E	0101	3700	8	[G40.2, Z11]	[G40.2, Z11]	2	[1-210]	1	<NA>	[HA3700]	Nein	6	Ja	[Psycho-soziale Intervention (PSI): Krankheits...

Die Fälle in der Fallbasis wurden mit einem Zeitbezug in Form der Anzahl an Tagen nach dem Zeitpunkt der Aufnahme abgespeichert, sodass die Falldaten für einen spezifischen Tag X abgeleitet werden können, die ausschließlich die Informationen enthalten, die bis zum X-ten Tag nach stationärer Aufnahme eines Patienten dokumentiert wurden. Zu beachten ist, dass keine Nachsorgeinformationen mit Zeitstempel vorlagen, sondern nur die tatsächlichen Nachsorgemaßnahmen nach Entlassung. Als Zeitpunkte wurden insbesondere Tag 1 (Tag der stationären Aufnahme) und Tag n (Tag der Entlassung) untersucht.

1.3.2 Entwicklung von Ähnlichkeitsmaßen (AP 4.2)

Für die Diagnosen (ICD-Codes), Prozeduren (OPS-Codes) und Nachsorgemaßnahmen wurden Taxonomien aufgebaut, um einen semantischen Vergleich der Codes zu ermöglichen. Für die unterschiedlichen Attributtypen wurden ergänzend zu üblichen Standardmaßen wie *binärer Ähnlichkeit* mit Test auf identische Werte (z.B. bei Geschlecht), *lineare Distanzfunktionen* bei numerischen Werten (z.B. bei Alter) und *Mengen-Ähnlichkeitsmaße* zum paarweisen Vergleich von Listeneinträgen (z.B. bei Diagnosen) spezifische Ähnlichkeitsmaße entwickelt. Die spezifischen Ähnlichkeitsmaße umfassen einerseits die Taxonomie-Ähnlichkeitsmaße, die auf Basis der modellierten Taxonomien die semantische Ähnlichkeit von ICD- und OPS-Codes bestimmen und andererseits Maße, für die

Expertenwissen einbezogen wurde, um paarweise Ähnlichkeiten von Attributausprägungen explizit abzubilden oder um spezifische Codierungen der Datenerfassung nach § 21 zu berücksichtigen (z.B. Aufnahmegrund).

1.3.3 Umsetzung der Prognosefunktionalität (AP 4.3)

Die Umsetzung der Prognosefunktionalität umfasste die iterative Entwicklung und Evaluation von Prädiktionsmodellen auf Basis von CBR- und ML-Modellen und wird im Folgenden näher erläutert. Gegenüber dem ursprünglichen Arbeitsplan wurden die Arbeiten in diesem Arbeitspaket deutlich intensiviert, um die Herausforderungen der Datenqualität und -verfügbarkeit und der Prädiktionsgüte zu adressieren. So wurden umfangreichere Erprobungen von Prädiktionsansätzen sowie Datenanalysen durchgeführt.

Umsetzung der CBR-basierten Prädiktionsmodelle

Zur fallbasierten Prädiktion mit CBR wurden verschiedene Modelle entwickelt. Die Implementierung wurde mit dem gemeinsam vom DFKI und der Universität Trier entwickelten CBR-Framework CBRkit¹ umgesetzt.

Basierend auf einer initialen Analyse der Attributausprägungen in der Fallbasis wurden unterschiedliche *Attributfilter* (Feature Sets) definiert, auf Basis derer die Ähnlichkeitsberechnung umgesetzt wurde.

Mithilfe von unterschiedlichen *Ergebnisfiltern* wurde die Auswahl von ähnlichsten Fällen zur Ermittlung der Prädiktionswerte variiert. Es wurde dabei nach dem Rang der ähnlichsten Fälle im Intervall [1,50] sowie den ermittelten Ähnlichkeitswerten im Intervall [0,1] in verschiedenen Konstellationen gefiltert.

Mithilfe unterschiedlicher *Aggregationen* wurden aus einer Menge von ähnlichen Fällen die Prädiktionswerte bestimmt. Hierfür wurden je nach Art der Prädiktion verschiedene Aggregationen erprobt. Zur Regression der Belegungstage wurden der *Mittelwert*, der *Median* und ein *mit Ähnlichkeitswerten gewichteter Mittelwert* berechnet. Zur Binären und Multi-Label-Klassifikation kamen der *Multi-Modus* und ein *mit Ähnlichkeiten gewichteter Multi-Modus* zum Einsatz.

¹Lenz, M., Malburg, L., Bergmann, R. (2024). CBRkit: An Intuitive Case-Based Reasoning Toolkit for Python. In: Recio-Garcia, J.A., Orozco-del-Castillo, M.G., Bridge, D. (eds) Case-Based Reasoning Research and Development. ICCBR 2024. Lecture Notes in Computer Science, vol 14775. Springer, Cham. https://doi.org/10.1007/978-3-031-63646-2_19

Umsetzung der ML-basierten Prädiktionsmodelle

In Abbildung 1.8 werden die ML-Algorithmen aus der Python-Bibliothek scikit-learn² für die jeweilige Prädiktionsaufgabe aufgelistet, die zu Vergleichszwecken implementiert und evaluiert wurden. Zu beachten ist, dass die ML-Modelle auf denselben vorverarbeiteten Daten trainiert wurden,

Abbildung 1.8 Erprobte ML-Algorithmen der scikit-learn-Bibliothek

Belegungstage (Regression):	Binäre Klassifikation (Nachsorgebedarf):	Multi-Label-Klassifikation (Nachsorgekategorien/-maßnahmen):
1. LinearRegression	1. LogisticRegression	1. LogisticRegression (MultiOutputClassifier)
2. Lasso	2. DecisionTreeClassifier	2. DecisionTreeClassifier
3. Ridge	3. RandomForestClassifier	3. RandomForestClassifier
4. ElasticNet	4. GradientBoostingClassifier	4. GradientBoostingClassifier (MultiOutputClassifier)
5. DecisionTreeRegressor	5. XGBClassifier	5. XGBClassifier (MultiOutputClassifier)
6. RandomForestRegressor	6. LGBMClassifier	6. LGBMClassifier (MultiOutputClassifier)
7. GradientBoostingRegressor	7. CatBoostClassifier	7. CatBoostClassifier (MultiOutputClassifier)
8. XGBRegressor	8. KNeighborsClassifier	8. KNeighborsClassifier
9. LGBMRegressor	9. Dense	9. Dense
10. CatBoostRegressor	10. LSTM	10. LSTM
11. Dense	11. GRU	11. GRU
12. LSTM		
13. GRU		

wie die CBR-Modelle. Einige ML-Algorithmen erforderten eine weitergehende Vorverarbeitung wie Label-Encoding, One-Hot-Encoding oder Hashing von Listen-Attributen. Die Umsetzung der Prädiktionsmodelle mit Maschinellem Lernen sollte eine Einschätzung der Leistungsfähigkeit der verschiedenen Modelle in der Breite ermöglichen und diene primär der Einschätzung der Performance im Vergleich zu den CBR-Modellen. Aus diesen Gründen wurde auf ein tiefergehendes Fine-Tuning der Hyperparameter der ML-Modelle verzichtet.

Training und Testen der Prädiktionsmodelle

Die Evaluation der Prädiktionsmodelle erforderte das Aufteilen der Datensätze in Trainings- und Testdaten. Für die UKB-Daten mit Nachsorgeinformationen wurde eine fünffache zufällige Auswahl von 1400 Fällen mit gleichem Anteil (ca. 18%) von Fällen mit Nachsorgebedarf (Ja/Nein) (Stratified Shuffle Split) bestimmt. Zusätzlich wurden Trainings- und Testdaten ausschließlich aus den ca. 32.000 Fällen mit Nachsorgebedarf mit fünffacher zufälliger Auswahl von je 1400 Fällen (Shuffle Split) bestimmt. Da für die UMM-Daten keine Nachsorgeinformationen vorlagen, wurde eine fünffache zufällige Auswahl von je 1400 Fällen (Shuffle Split) bestimmt. Es wurden somit in fünf Iterationen je 1400 Fälle zum Testen verwendet und die übrigen Daten zum Training genutzt.

²<https://scikit-learn.org>

Ergebnisse der Prädiktion von Belegungstagen

Die Zielgröße der Belegungstage wies in beiden Datensätzen der klinischen Partner eine breite Streuung auf. Die Abbildungen 1.9 und 1.10 zeigen die deskriptive Statistik für beide Datensätze.

Abbildung 1.9 Verteilung der Belegungstage im Datensatz der UKB

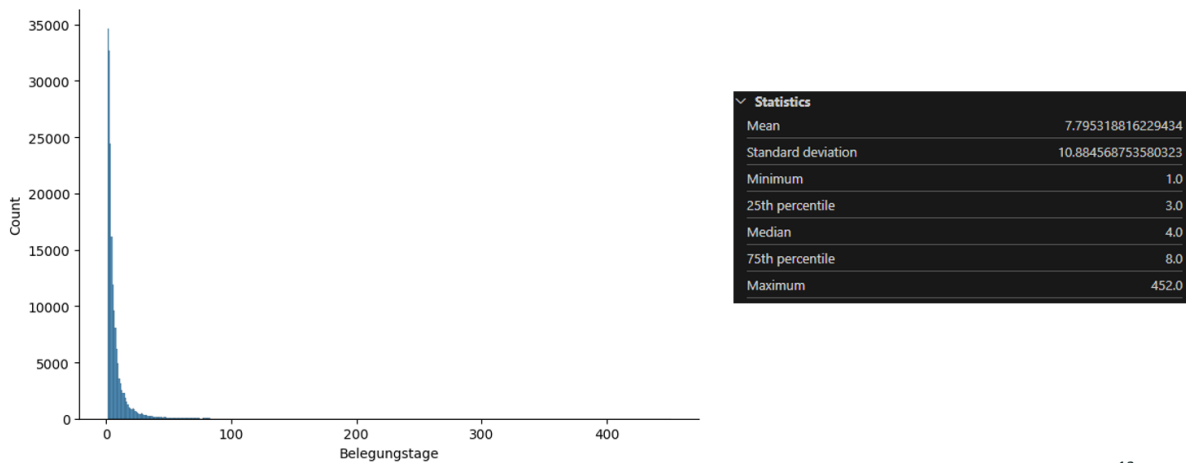
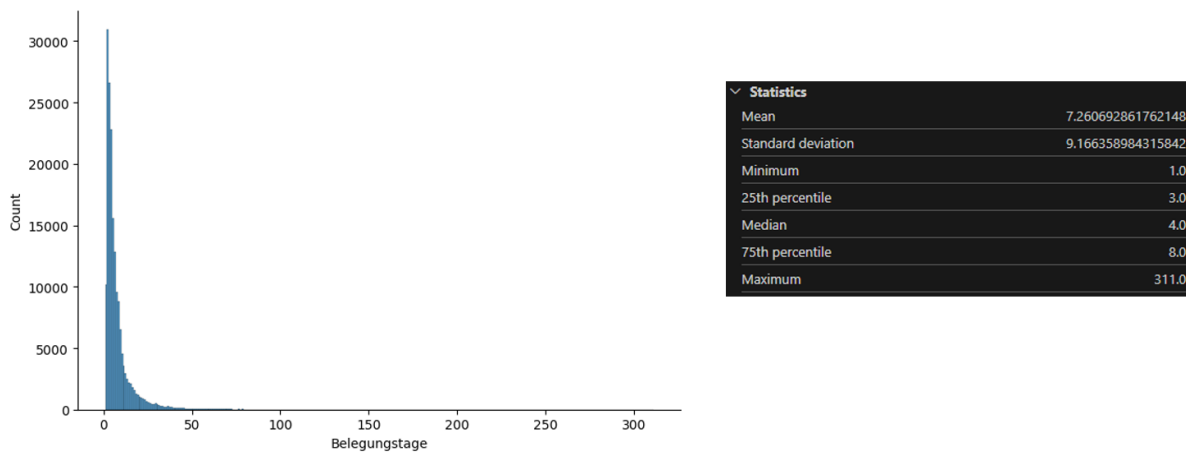


Abbildung 1.10 Verteilung der Belegungstage im Datensatz der UMM



Zur Bewertung der Prädiktionsgüte der Regression wurden der mittlere absolute Fehler (MAE), die Quadratwurzel des mittleren quadratischen Fehlers (RMSE) und das Bestimmtheitsmaß zur Beurteilung der Anpassungsgüte einer Regression (R^2) herangezogen. Die Ergebnisse der jeweils fünf besten ML- und CBR-basierten Prädiktionsmodelle für die Prädiktion der Belegungstage werden in Abbildung 1.11 für die UKB-Daten und in Abbildung 1.12 für die UMM-Daten präsentiert. Die

Ergebnisse sind absteigend nach der RMSE-Metrik sortiert, welche große Abweichungen höher gewichtet als die MAE-Metrik. Für die UKB-Daten erzielten die besten Prädiktionsmodelle eine durchschnittliche Abweichung nach RMSE von ± 9 Tagen und eine Anpassungsgüte von ca. 28% für den Tag 1 und eine etwas geringere durchschnittliche Abweichung von ± 5 Tagen und eine Anpassungsgüte von ca. 74% für den Tag n. Es ließen sich keine signifikanten Unterschiede zwischen den ML- und CBR-Modellen feststellen. Für die UMM-Daten erzielten die Prädiktionsmodelle

Abbildung 1.11 Ergebnisse der Regression der Belegungstage (je Top 5 ML- & CBR-Modelle)

UKB - Tag 1:

Run Name	Metrics				filter
	Cr	Dt	mae	r2	
● CatBoostRegressor	4.312381127...	0.281239510...	9.006589140...	-	
● LGBMRegressor	4.387486443...	0.278062025...	9.026297830...	-	
● XGBRegressor	4.310269521...	0.278112873...	9.026448126...	-	
● RandomForestRegressor	4.174200358...	0.268803051...	9.070519390...	-	
● cbr_feature_set_3_filter_set_13_weighted_mean	4.083714285...	0.266610260...	9.103829196...	rank <= 50	
● cbr_feature_set_3_filter_set_1_weighted_mean	4.083714285...	0.266540846...	9.104180249...	similarity >= 0.7	
● cbr_feature_set_3_filter_set_13_mean	4.087285714...	0.265996590...	9.107595726...	rank <= 50	
● cbr_feature_set_3_filter_set_1_mean	4.087571428...	0.265879620...	9.108225817...	similarity >= 0.7	
● cbr_feature_set_3_filter_set_12_weighted_mean	4.043857142...	0.262177071...	9.131639387...	rank <= 20	
● GradientBoostingRegressor	4.601756394...	0.236308656...	9.289064988...	-	

UKB - Tag n:

Run Name	Metrics				filter
	Cr	Dt	mae	r2	
● CatBoostRegressor	2.811500472...	0.742929761...	5.298874520...	-	
● LGBMRegressor	2.983009393...	0.728943720...	5.457208847...	-	
● XGBRegressor	2.864833700...	0.726878731...	5.504176808...	-	
● RandomForestRegressor	2.812820831...	0.718727111...	5.568962185...	-	
● GradientBoostingRegressor	3.220397805...	0.684576515...	5.88567139...	-	
● cbr_feature_set_3_filter_set_10_weighted_mean	2.813714285...	0.674382589...	6.036041309...	rank <= 10	
● cbr_feature_set_3_filter_set_10_mean	2.812428571...	0.673427103...	6.045405961...	rank <= 10	
● cbr_feature_set_3_filter_set_11_weighted_mean	2.828142857...	0.668260584...	6.094649050...	rank <= 15	
● cbr_feature_set_3_filter_set_11_mean	2.829714285...	0.667310714...	6.103880996...	rank <= 15	
● cbr_feature_set_3_filter_set_1_weighted_mean	2.869	0.665712498...	6.125838501...	similarity >= 0.7	

geringfügig bessere Ergebnisse mit einem RMSE von ± 8 Tagen für Tag 1 und einem RMSE von ± 5 Tagen für Tag n. Die Anpassungsgüte ist auf einem vergleichbaren Niveau.

Abbildung 1.12 Ergebnisse der Regression der Belegungstage (je Top 5 ML- & CBR-Modelle)

UMM - Tag 1:

Run Name	Metrics				filter
	Cr	Dt	mae	r2	
● CatBoostRegressor	3.790564746...	0.276215593...	8.000546924...	-	
● XGBRegressor	3.814759033...	0.268274474...	8.042043060...	-	
● LGBMRegressor	3.930672102...	0.258418733...	8.096974959...	-	
● cbr_feature_set_4_filter_set_11_weighted_mean	3.826714285...	0.239564470...	8.202089409...	rank <= 15	
● cbr_feature_set_1_filter_set_11_weighted_mean	3.826857142...	0.239532668...	8.202292732...	rank <= 15	
● cbr_feature_set_4_filter_set_11_mean	3.828142857...	0.239137587...	8.204306778...	rank <= 15	
● cbr_feature_set_1_filter_set_11_mean	3.828142857...	0.239137587...	8.204306778...	rank <= 15	
● cbr_feature_set_1_filter_set_12_weighted_mean	3.810714285...	0.237982865...	8.206832571...	rank <= 20	
● RandomForestRegressor	3.934172284...	0.218808033...	8.291885611...	-	
● GradientBoostingRegressor	4.192037323...	0.220721748...	8.300295673...	-	

UMM - Tag n:

Run Name	Metrics				filter
	Cr	Dt	mae	r2	
● CatBoostRegressor	2.524532546...	0.750738353...	4.654966448...	-	
● XGBRegressor	2.584677426...	0.730896389...	4.840776816...	-	
● RandomForestRegressor	2.542884451...	0.721435298...	4.922426683...	-	
● LGBMRegressor	2.716092748...	0.710836892...	5.037172693...	-	
● GradientBoostingRegressor	2.972232888...	0.688673806...	5.207634258...	-	
● cbr_feature_set_3_filter_set_12_weighted_mean	2.635428571...	0.646430133...	5.590954131...	rank <= 20	
● cbr_feature_set_3_filter_set_12_mean	2.637428571...	0.646264390...	5.592497671...	rank <= 20	
● cbr_feature_set_3_filter_set_11_mean	2.637	0.642881933...	5.620090279...	rank <= 15	
● cbr_feature_set_3_filter_set_11_weighted_mean	2.636428571...	0.642793559...	5.620431360...	rank <= 15	
● cbr_feature_set_3_filter_set_10_weighted_mean	2.653571428...	0.639765183...	5.633211924...	rank <= 10	

Ergebnisse der Prädiktion von generellem Nachsorgebedarf

Zur Bewertung der Prädiktionsgüte der binären Klassifikation des generellen Nachsorgebedarfs wurden die Metriken *Precision*, *Recall* und *F1-Score* betrachtet. Die größte Relevanz wurde dem Recall beigemessen, sodass möglichst alle Fälle mit Nachsorgebedarf gefunden werden. In Abbildung 1.13 werden die Ergebnisse dargestellt. Es ist ersichtlich, dass die generalistischen Prädiktionsmodelle

Abbildung 1.13 Ergebnisse der binären Klassifikation des generellen Nachsorgebedarfs

UKB - Tag 1:

Run Name	Metrics				Parameters
	Cr	Dt	f1_binary	precision_binary	
cbr_feature_set_3_filter_set_7_weighted_multimode			0.538123349...	0.536140574...	0.540392156... rank <= 1
cbr_feature_set_3_filter_set_7_multimode			0.538123349...	0.536140574...	0.540392156... rank <= 1
cbr_feature_set_3_filter_set_8_weighted_multimode			0.564039328...	0.607781357...	0.526274509... rank <= 3
cbr_feature_set_3_filter_set_8_multimode			0.564039328...	0.607781357...	0.526274509... rank <= 3
cbr_feature_set_8_filter_set_7_weighted_multimode			0.523126382...	0.529236373...	0.517647058... rank <= 1
DecisionTreeClassifier			0.495529664...	0.484522802...	0.507450980... -
CatBoostClassifier			0.578698911...	0.702659803...	0.492549019... -
XGBClassifier			0.562446759...	0.688318984...	0.476862745... -
RandomForestClassifier			0.545614206...	0.720756534...	0.439215686... -
LGBMClassifier			0.535422497...	0.713211066...	0.429803921... -

UKB - Tag n:

Run Name	Metrics				Parameters
	Cr	Dt	f1_binary	precision_binary	
LSTM			0.628805294...	0.648131386...	0.614901960... -
cbr_feature_set_3_filter_set_8_weighted_multimode			0.640929344...	0.674448065...	0.610980392... rank <= 3
cbr_feature_set_3_filter_set_8_multimode			0.640929344...	0.674448065...	0.610980392... rank <= 3
cbr_feature_set_3_filter_set_7_weighted_multimode			0.594993787...	0.584446821...	0.607058823... rank <= 1
cbr_feature_set_3_filter_set_7_multimode			0.594993787...	0.584446821...	0.607058823... rank <= 1
cbr_feature_set_3_filter_set_10_weighted_multimode			0.654453674...	0.714488774...	0.603921568... rank <= 10
XGBClassifier			0.639135491...	0.705575177...	0.584313725... -
CatBoostClassifier			0.644778758...	0.723399709...	0.580392156... -
DecisionTreeClassifier			0.560926738...	0.556725915...	0.565490196... -
LGBMClassifier			0.619396490...	0.709519377...	0.549803921... -

in der Gesamtheit der Falldaten keine hinreichend genauen Muster finden konnten. Dies wird durch die Recall-Werte von 54% für Tag 1 und 61% für Tag n deutlich. Es liegt die Vermutung nahe, dass dies durch das Fehlen von relevanten Informationen wie die Lebens- und Versorgungssituation von Patient:innen bedingt ist.

Ergebnisse der Prädiktion von Nachsorgekategorien

Abbildung 1.14 Ergebnisse der Multi-Label-Klassifikation der Nachsorgekategorien für Tag 1

UKB - Tag 1 (Ausschließlich Fälle mit Nachsorgebedarf):

Run Name	Metrics											Parameters		
	Cr	Dt	all_hit	f1_macro	f1_samples	f1_weighted	precision_macro	precision_sampli	precision_weight	recall_macro	recall_sampli		recall_weighted	single_hit
cbr_feature_set_3_filter_set_8_all			0.270571428...	0.327821824...	0.635080385...	0.639011268...	0.257278750...	0.564637414...	0.533026824...	0.471378626...	0.846223809...	0.810502001...	0.925428571...	rank <= 3
cbr_feature_set_8_filter_set_8_all			0.263285714...	0.319172278...	0.627662471...	0.631270099...	0.249211947...	0.556041836...	0.523054174...	0.462104504...	0.842073809...	0.807963219...	0.919285714...	rank <= 3
cbr_feature_set_6_filter_set_8_all			0.267999999...	0.318906599...	0.629162358...	0.635178745...	0.251658163...	0.560115646...	0.530315842...	0.453873112...	0.838607142...	0.804111519...	0.919142857...	rank <= 3
cbr_feature_set_1_filter_set_9_all			0.178714285...	0.318554335...	0.610054885...	0.631601303...	0.234319216...	0.509054251...	0.503074198...	0.546411810...	0.899554761...	0.876071803...	0.952285714...	rank <= 5
cbr_feature_set_1_filter_set_8_all			0.267285714...	0.317986458...	0.629831220...	0.636002728...	0.250998990...	0.560649319...	0.531082181...	0.451133029...	0.839145238...	0.804502295...	0.919571428...	rank <= 3
MultiOutputClassifier_LGBMClassifier			0.656285714...	0.279161968...	0.660157142...	0.627628573...	0.449638846...	0.725452380...	0.701658070...	0.238182997...	0.659585714...	0.594982460...	0.794142857...	-
DecisionTreeClassifier			0.521285714...	0.276628242...	0.597937755...	0.584143716...	0.283111023...	0.636566666...	0.586293982...	0.275143563...	0.630833333...	0.583019428...	0.760714285...	-
MultiOutputClassifier_CatBoostClassifier			0.664714285...	0.262236188...	0.665270748...	0.631187014...	0.450253542...	0.731333333...	0.706923397...	0.225138473...	0.664104761...	0.599027566...	0.798285714...	-
MultiOutputClassifier_XGBClassifier			0.643571428...	0.260803450...	0.657873469...	0.629064729...	0.411726400...	0.719285714...	0.691397223...	0.226943674...	0.661071428...	0.598837256...	0.794999999...	-
RandomForestClassifier			0.672857142...	0.254992514...	0.654357142...	0.611885232...	0.446495736...	0.727845238...	0.704554060...	0.218851123...	0.64525	0.577316456...	0.783142857...	-

Die Prädiktion der Nachsorgekategorien wurde für verschiedene Ausschnitte der UKB-Daten untersucht. Im Folgenden werden die Prädiktionsergebnisse für die Gesamtheit an Behandlungsfällen (ca. 32.000 Fälle) mit dokumentierter Nachsorge für alle Kategorien von Nachsorgemaßnahmen (11 Oberkategorien) präsentiert. Zur Bewertung der Prädiktionsgüte der Multi-Label-Klassifikation von Nachsorgekategorien wurden die Metriken *Precision*, *Recall* und *F1-Score* für jede Kategorie berechnet und auf verschiedene Weisen (*Macro*, *Weighted* und *Samples*) aggregiert. Die Ergebnisse werden in Abbildung 1.14 für Tag 1 und in Abbildung 1.15 für Tag n dargestellt. Die Ergebnisse sind nach dem *F1-Score* mit Aggregation *Macro* sortiert, sodass auch seltene Kategorien mit gleicher

Abbildung 1.15 Ergebnisse der Multi-Label-Klassifikation der Nachsorgekategorien für Tag n
UKB - Tag n (Ausschließlich Fälle mit Nachsorgebedarf):

Run Name	Metrics													Parameters
	Cr	Di	all_hit	f1_macro	f1_weighted	precision_macro	precision_sample	precision_weight	recall_macro	recall_samples	recall_weighted	single_hit	filter	
cbr_feature_set_3_filter_set_8_all	0.299714285...			0.350575486...	0.654162213...	0.656346580...	0.276699056...	0.586446598...	0.551702905...	0.499554241...	0.856071428...	0.823734896...	0.930285714...	rank <= 3
cbr_feature_set_3_filter_set_9_all	0.208571428...			0.347227233...	0.628624819...	0.645293957...	0.255903112...	0.530777040...	0.518182806...	0.592477518...	0.907185714...	0.883667240...	0.957857142...	rank <= 5
cbr_feature_set_1_filter_set_8_all	0.293428571...			0.346326886...	0.648947917...	0.652315480...	0.273675043...	0.581599829...	0.548696607...	0.492007046...	0.848426190...	0.816850564...	0.925428571...	rank <= 3
cbr_feature_set_6_filter_set_8_all	0.285714285...			0.341510347...	0.644734498...	0.648927923...	0.269268656...	0.576282482...	0.545153366...	0.487724622...	0.847230952...	0.814494952...	0.925428571...	rank <= 3
cbr_feature_set_8_filter_set_8_all	0.294428571...			0.339555832...	0.646849546...	0.648013168...	0.268698848...	0.580244557...	0.543489278...	0.478883432...	0.845688095...	0.814631408...	0.921428571...	rank <= 3
MultiOutputClassifier_LGBMClassifier	0.669857142...			0.315072166...	0.679640476...	0.650145272...	0.487038767...	0.742142857...	0.719482686...	0.269988708...	0.680888095...	0.617180364...	-	-
MultiOutputClassifier_XGBCClassifier	0.641142857...			0.307668638...	0.673141156...	0.650168211...	0.457583801...	0.727261904...	0.702532791...	0.266589018...	0.683326190...	0.623832837...	0.813428571...	-
MultiOutputClassifier_CatBoostClassifier	0.671142857...			0.302464183...	0.685882993...	0.654392495...	0.505302085...	0.747214285...	0.724157208...	0.259121391...	0.688550000...	0.625360282...	0.823000000...	-
MultiOutputClassifier_GradientBoostingClassifier	0.648285714...			0.301990708...	0.657804081...	0.626345713...	0.485407683...	0.720809523...	0.707460295...	0.254449336...	0.659083333...	0.590857678...	0.793571428...	-
DecisionTreeClassifier	0.538857142...			0.295122485...	0.619962244...	0.603042374...	0.304659192...	0.657976190...	0.605069083...	0.290013166...	0.652514285...	0.601764850...	0.784285714...	-

Relevanz berücksichtigt werden. Es ist ersichtlich, dass die Prädiktionsmodelle nur eine geringe Prädiktionsgüte erzielten.

Ergebnisse der Prädiktion für ausgewählte Patientengruppen und Nachsorgekategorien

Im weiteren Verlauf wurden Experimente zur Prädiktion der Nachsorge mit einem Fokus auf die als besonders relevant identifizierten Nachsorgemaßnahmen/-kategorien und Patientengruppen durchgeführt. Hierbei wurden die Prädiktionsmodelle explizit mit den gewählten Nachsorgemaßnahmen

Abbildung 1.16 Besonders relevante Nachsorgemaßnahmen oder -kategorien

- N1: Medizinische Rehabilitation (MedRe): Geriatrische Rehabilitation
- N2: Medizinische Rehabilitation (MedRe): Neurologische/neurochirurgische Frührehabilitation
- N3: Ambulante Nachsorge (AmNa)
- N4: Stationäre Nachsorge (StaNa)
- N5: Ambulante Nachsorge (AmNa) und Stationäre Nachsorge (StaNa)
- N6: Stationäre Nachsorge (StaNa): Kurzzeitpflege und Stationäre Nachsorge (StaNa): Langzeitpflege
- N7: Stationäre Nachsorge (StaNa): Hospiz
- N8: Ambulante Nachsorge (AmNa): Pflegedienst

(siehe Abbildung 1.16) in Kombination mit verschiedenen Patientenkohorten (siehe Abbildung 1.17) trainiert und evaluiert. Die Patientengruppen basieren auf strukturellen, medizinischen und diagnostischen Merkmalen und wurden durch die klinischen Partner definiert. Abbildung 1.18 zeigt die besten Ergebnisse eines CBR-Modells für die binäre Klassifikation des Nachsorgebedarfs und Tag n des UKB-Datensatzes. Zu beachten sind die deutlich reduzierten Fallzahlen der Kombinationen. Die Anzahlen werden durch die Betrachtung früherer Tage des Krankenhausaufenthaltes weiter reduziert, da dann durch die geringere Menge an Patientendaten weniger Personen gemäß der Definition der Patientengruppen (siehe Abbildung 1.17) identifiziert werden können. In der Tabelle werden in der ersten Zeile die Gesamtzahlen an Fällen ohne Nachsorgebedarf für die jeweiligen Patientengruppen aufgeführt. In jeder Zeile für eine Patienten-Nachsorge-Kombination werden die

Abbildung 1.17 Besonders relevante Patientengruppen

- P1: Patienten mit Alter 80+
- P2: Patienten mit diagnostizierter Demenz (F00-F03)
- P3: Patienten mit diagnostiziertem Delir (F05)
- P4: Patienten mit motorischer Funktionseinschränkung (U50)
- P5: Patienten mit Senilität (R54)
- P6: Geriatrische Fälle
 - Patienten mit Alter 70+
 - Patienten haben mindestens eine oder mehrere der folgenden Komorbiditäten
 - o R29.6: Sturzneigung
 - o R42: Schwindel und Taumel
 - o R52.2: Sonstiger chronischer Schmerz
 - o F05: Delir
 - o R32: Nicht näher bezeichnete Harninkontinenz
 - o R15: Stuhlinkontinenz
 - o R26: Störungen des Ganges und der Mobilität
 - o F00: Demenz bei Alzheimer-Krankheit
 - o F01: Vaskuläre Demenz
 - o F02: Demenz bei anderenorts klassifizierten Krankheiten
 - o F03: Nicht näher bezeichnete Demenz
 - o M62.5: Muskelschwund und -atrophie, anderenorts nicht klassifiziert
 - o R54: Senilität
 - o I69.4: Folgen eines Schlaganfalls, nicht als Blutung oder Infarkt bezeichnet
 - o M15: Polyarthrose
 - o G62: Sonstige Polyneuropathien
 - o G45: Zerebrale transitorische Ischämie und verwandte Syndrome
 - o G20: Primäres Parkinson-Syndrom
- P7: Geriatrische Fälle nach P6 mit Verlegung auf Intensivstation (Fachabteilung 3600)
- P8: Geriatrische Fälle nach P6 mit einer Operation (OPS-Code beginnend mit 5-)
- P9: Geriatrische Fälle nach P6 mit Verlegung auf Orthopädie (Fachabteilung 2300)
- P10: Geriatrische Fälle nach P6 mit Verlegung auf Unfallchirurgie (Fachabteilung 1600)
- P11: Patienten mit somatischer Erkrankung
 - alle Diagnosen beginnend mit F
 - ohne F10.0 bis F19.0
 - ohne F10.1 bis F19.1

Abbildung 1.18 Ergebnisse der binären Klassifikation des Nachsorgebedarfs für ausgewählte Patientengruppen und Nachsorgekategorien

Tag n	P1: Alter 80+	P2: Demenz (F00-F03)	P3: Delir (F05)	P4: motorische Funktionseinschränkung (U50)	P5: Senilität (R54)	P6: Geriatrische Fälle	P7: Geriatrische Fälle + Intensivstation	P8: Geriatrische Fälle + Operation	P9: Geriatrische Fälle + Orthopädie	P10: Geriatrische Fälle + Unfallchirurgie	P11: Somatische Erkrankung
Fälle mit und ohne Nachsorgebedarf an Tag n	28983	2590	4782	1928	39	17491	2530	6588	537	872	19370
N0: Genereller Nachsorgebedarf	6851	505	2567	1723	14	5845	1616	3297	348	429	5877
			p=0,65 r=0,73	p=0,91 r=0,99		p=0,68 r=0,50	p=0,71 r=0,85	p=0,67 r=0,79	p=0,66 r=0,95	p=0,59 r=0,45	p=0,69 r=0,48
N1: Medizinische Rehabilitation (MedRe): Geriatrische Rehabilitation	368	22	113	30	2	309	96	209	32	25	135
N2: Medizinische Rehabilitation (MedRe): Neurologische/neurochirurgische Frührehabilitation	438	34	403	451	0	686	89	238	7	8	659
N3: Ambulante Nachsorge (AmNa)	3046	243	734	153	8	2068	457	955	209	263	1886
N4: Stationäre Nachsorge (StaNa)	1044	223	380	46	5	896	241	434	105	208	717
N5: Ambulante Nachsorge (AmNa) und Stationäre Nachsorge (StaNa)	3667	396	963	176	11	2621	603	1211	262	392	2343
								p=0,57 r=0,69	p=0,59 r=0,44		
N6: Stationäre Nachsorge (StaNa): Kurzzeitpflege und Stationäre Nachsorge (StaNa): Langzeitpflege	822	174	275	38	3	669	159	305	77	144	535
N7: Stationäre Nachsorge (StaNa): Hospiz	48	8	33	4	2	73	6	27	0	0	74
N8: Ambulante Nachsorge (AmNa): Pflegedienst	441	38	123	35	1	347	77	169	34	40	295

Anzahl an Fällen
p=Precision r= Recall

Anzahl an Fällen angegeben, die Nachsorge dieser Kategorie erhielten, sowie die erzielte Precision und der Recall der CBR-basierten Prädiktion. Es wurde eine Leave-One-Out-Evaluation genutzt, sodass jeder Fall einzeln zum Testen des Modells herangezogen wurde. Die zehn besten Ergebnisse

des CBR-Modells sind in den markierten Zellen angegeben. Das beste Ergebnis im Hinblick auf Kombinationen, bei denen die Anzahl an Fällen mit und ohne Nachsorge annähernd gleich verteilt ist, wurde für die Patientengruppe der Geriatrischen Fälle mit Verlegung auf die Intensivstation (P7) und der Geriatrischen Fälle mit Verlegung auf die Orthopädie (P9) der Prädiktion des generellen Nachsorgebedarfs erzielt. Für P7 wurden ein Recall von 85% und eine Precision von 71% gemessen. Für P9 wurden ein Recall von 95% und eine Precision von 66% gemessen. Die Experimente mit ML-Modellen haben durchweg zu schlechteren Ergebnissen geführt, was auf die sehr geringen Fallzahlen zurückzuführen ist.

Ergebnisse der Prädiktion für die Patientengruppe P9 mit Informationen aus der Pflegeanamnese

Für die Patientengruppe P9 wurde durch die UKB für eine Stichprobe an Fällen mit und ohne Nachsorge die Pflegeanamnese ausgewertet und in ausgewählten Attributen codiert. In 243 Fällen lag eine Pflegeanamnese vor und konnte ausgewertet werden. Die Attribute können Abbildung 1.19 entnommen werden. Die besten Ergebnisse der CBR- und ML-basierten Prädiktionen werden in

Abbildung 1.19 Datensatz für die Patientengruppe P9 mit Pflegeanamnese (n=537)



Abbildung 1.20 dargestellt. Mithilfe der Pflegeanamnesedaten konnte der Recall auf 97% erhöht werden. Mit den ML-Modellen konnten die Ergebnisse der CBR-Modelle nicht erreicht werden. Durch die Betrachtung von Korrelationen zwischen den Attributausprägungen nach Baak et al.³ konnte eine Auswahl von 5 Attributen gefunden werden, für die ein CBR-Modell einen Recall von 90% und

³Baak, M., Koopman, R., Snoek, H., Klous, Sander. (2020). A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Computational Statistics & Data Analysis, vol 152. 107043. <https://doi.org/10.1016/j.cstda.2020.107043>

Abbildung 1.20 Ergebnisse der binären Klassifikation des Nachsorgebedarfs für die Patientengruppe P9 mit Pflegeanamnese

Experimente mit 537 Fällen (348 mit Nachsorge/ 189 ohne Nachsorge)	CBR			ML		
	F1	Precision	Recall	F1	Precision	Recall
Datensatz P9 + Attribute Pflegeanamnese	0,78	0,66	0,97	0,65	0,65	0,65
Ausgewählte Attribute mit hoher Korrelation (Infos zu Nachsorgebedarf, Gefahren, Lebenssituation, Pflegegrad, AufnAnlass)	0,77	0,67	0,90	0,63	0,63	0,63

eine Precision von 67% erreichte. Dies zeigt das Potenzial, dass durch eine gezielte Erfassung von für die Nachsorgeplanung relevanten Attributen eine hohe Prädiktionsgüte in der Zukunft erreicht werden kann.

1.3.4 Implementierung der Erklärungskomponente (AP 4.4)

Ein Vorteil der CBR-Modelle gegenüber ML-Modellen und insbesondere Deep Learning-Modellen kann in der Transparenz und damit der Interpretierbarkeit der Schlussfolgerungen gesehen werden. Die fallbasierte Prädiktionskomponente wurde aus diesem Grund um eine Erklärungskomponente erweitert, welche die der Prädiktion zugrundeliegenden Fälle dem Anwender anzeigen und bei Bedarf das resultierende Ranking von Fällen erläutern kann. Abbildung 1.21 veranschaulicht dies anhand einer einfachen textuellen Beschreibung der der Prädiktion zugrunde liegenden Fälle. In der Erklärung werden neben der Gesamtähnlichkeit eines zur Anfrage ähnlichen Falls auch die Attribute mit ihren Ähnlichkeiten nach Schwellwerten gefiltert aufgeführt.

1.4 Proaktive Kommunikation und Dokumentation im Prozess (AP 5)

Die Beteiligung des DFKI in AP 5 fand primär in beratender Funktion statt. Die Arbeiten in AP 4 konnten dadurch gegenüber dem ursprünglichen Arbeitsplan deutlich intensiviert werden, um die Herausforderungen der Datenqualität und -verfügbarkeit und der Prädiktionsgüte zu adressieren. Während der parallelen Bearbeitung von AP 4 durch DFKI und AP 5 durch Empolis wurden die Anknüpfungspunkte in regelmäßigen Projekttreffen besprochen, konzeptionell im Rahmen von AP 6 ausgearbeitet und bei der prototypischen Implementierung berücksichtigt. So ermöglicht die implementierte REST-Schnittstelle der Prädiktionskomponente dem Sozialberatungsassistenten, Prädiktionen anzufragen, sodass im Sinne einer hybriden KI eine Prognose des Nachsorgebedarfs

Abbildung 1.21 Beispiel einer fallbasierten Prädiktion



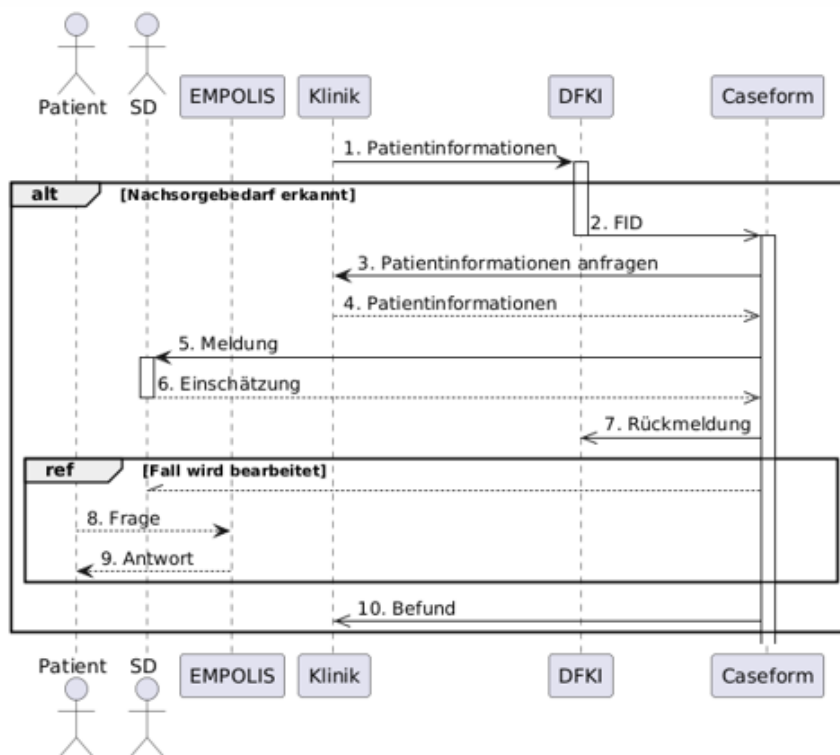
auf Basis beider KI-Komponenten erfolgen kann. Das Potenzial einer vollständigen Integration und praktischen Erprobung der Prädiktionskomponente und des Sozialberatungsassistenten wird weiterhin als groß eingeschätzt und könnte zukünftig durch das Zusammenspiel von menschlichen Experten mit den KI-Systemen im Sinne einer interaktiven Prädiktion die Prädiktionsgüte hinreichend verbessern.

1.5 Systemintegration und Entwicklung des Demonstratorsystems (AP 6)

In diesem Arbeitspaket implementierte das DFKI die Prädiktionskomponente prototypisch (AP 6.3) und validierte sie mittels automatisierter Tests (AP 6.4). Darüber hinaus wurde ein Integrationskonzept erarbeitet. Kern des Konzepts ist der Betrieb der Prädiktionskomponente als Docker-Container lokal im geschützten Netzwerk der Klinik, sodass die erforderlichen Falldaten direkt aus dem Krankenhausinformationssystem (KIS) bezogen werden können. Der Informationsaustausch mit externen Systemen – etwa der Entlassmanagementplattform Caseform des Projektpartners Nubedian – kann dadurch auf Fall-IDs und Prädiktionsergebnisse beschränkt werden; ein Transfer sensibler Patientendaten ist nicht erforderlich. Anfragen an die Prädiktionskomponente können sowohl zeitbasiert (z. B. nächtliche Verarbeitung) als auch ereignisgesteuert (z. B. bei Änderungen an einem Fall) erfolgen,

jeweils für einzelne Fälle oder für Fallmengen. Der konzipierte Ablauf der Nachsorgeplanung sowie das Zusammenspiel der DFKI-Prädiktionskomponente mit den Systemkomponenten der Projektpartner sind in Abbildung 1.22 dargestellt. Ergänzend wurde eine REST-basierte Schnittstelle für die Prädiktionskomponente gemäß den Anforderungen der Projektpartner spezifiziert und implementiert (AP 6.5). Eine Anbindung der Prädiktionskomponente an die KIS der klinischen Partner zum automatisierten Datenaustausch wurde nicht weiterverfolgt, da die Voraussetzung für eine praktische Erprobung durch die Sozialdienste mit aktuellen Behandlungsfällen durch die erzielte Prädiktionsgüte nicht gegeben war. Jedoch ermöglichte die umfassende Auswertung der historischen Behandlungsfälle eine experimentelle Evaluation zu unterschiedlichen Zeitpunkten des stationären Aufenthaltes von Patient:innen.

Abbildung 1.22 Sequenzdiagramm zum Ablauf der Nachsorgeplanung bei Integration der Teilsysteme der Projektpartner



2 Wichtigste Positionen des zahlenmäßigen Verwendungsnachweises

Im Verbundprojekt ist seitens des DFKI der wesentliche Teil der Kosten durch Personalausgaben entstanden. Abgesehen von den Verwaltungskosten stehen an zweiter Stelle der beantragte mobile Entwicklungsarbeitsplatz und an dritter Stelle folgen Reisekosten. Insgesamt wurden die Positionen des Antrages im zahlenmäßigen Nachweis im Rahmen des Verwendungsnachweises in den Selbstkosten des Projektes um 3% überschritten.

3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Notwendigkeit und Angemessenheit der öffentlichen Förderung für die im Vorhaben KIAFlex geleisteten Arbeiten sind einerseits in der Charakteristik des Forschungsprojektes und andererseits in der hohen Allgemeinrelevanz der bearbeiteten Fragestellungen begründet. Die im Projekt geleisteten Arbeiten waren dem Arbeitsplan angemessen. Ebenfalls waren die durchgeführten Forschungsarbeiten im Teilprojekt des Verbundprojektes sowie die dafür aufgewandten Ressourcen notwendig und angemessen, da sie der im Projektantrag formulierten Planung entsprachen und alle wesentlichen im Arbeitsplan formulierten Aufgaben erfolgreich bearbeitet wurden. Darüber hinaus waren keine zusätzlichen Ressourcen für das Projekt notwendig.

4 Verwertbarkeit

In dem Teilvorhaben wurden keine Erfindungen getätigt bzw. Schutzrechtsanmeldungen vorgenommen.

4.1 Wirtschaftliche Erfolgsaussichten

Das Projekt zeigte, dass eine praktische Realisierung der KI-basierten Prädiktionskomponente möglich ist. Das Potenzial der Prädiktionskomponente wird als groß eingeschätzt und könnte zukünftig durch eine vollständige Integration in bestehende Entlassungsmanagementsysteme, die die Interaktion von menschlichen Experten mit den KI-Modellen ermöglicht, voll ausgeschöpft werden.

4.2 Wissenschaftliche Erfolgsaussichten

Die wissenschaftlich-technischen Ergebnisse des DFKI bestehen in der Entwicklung von semantischen Modellen, die integrativ die Domäne des Entlassmanagements umfassen, die Fallbasis, durch die die Repräsentation der Patienten realisiert wird und die Basis zur Bedarfsprognose bildet, sowie die Prognosefunktionalität des Entlassmanagements für die individuellen Patientenbedarfe. Für das DFKI wird KIAFlex ein wesentlicher Baustein sein, um die Anwendung von KI im Bereich des Gesundheitswesens auszubauen und die Sichtbarkeit der Kompetenzen zu steigern. Das DFKI erwartet aus diesen Ergebnissen und den daraus zu erwartenden wissenschaftlichen Publikationen auf Konferenzen und in Journals eine weitere Stärkung seiner wissenschaftlich-technischen Kompetenz und Reputation, die nach Ende von KIAFlex mittel- und langfristig zu thematisch verwandten F&E-Projekten in weiteren Forschungsverbänden und ggf. zu Industriaufträgen führen wird. Außerdem ist Teil des wissenschaftlichen Verwertungskonzepts, dass die Projektergebnisse in die Lehre der beteiligten Professoren einfließen und ein wichtiges Anwendungsfeld in KI-orientierten Studiengängen darstellen. Zusätzlich wird der Aufbau eines strukturierten Promotionsprogrammes angestrebt, um auch Projektmitarbeitenden eine Promotion im Themenfeld zu ermöglichen. Das Projekt hat dazu beigetragen, die Expertise im Bereich der KI-basierten Analyse, Aufbereitung und Nutzung von

medizinischen Daten weiter zu stärken. Daraus folgt, dass zukünftig bei der Bearbeitung thematisch ähnlicher Projekte und möglicher Anschlussprojekte auf dieses Wissen zurückgegriffen werden kann.

4.3 Wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Das DFKI strebt selbst grundsätzlich keine wirtschaftliche Verwertung der Forschungsergebnisse an. Entsprechende Ergebnisse werden stattdessen in Kooperation mit dedizierten Wirtschaftspartnern (in Verbundprojekten oder in direkter Beauftragung) oder im Einzelfall über Ausgründungen in die wirtschaftliche Verwertung überführt. Daher sollen im Sinne der wissenschaftlichen Anschlussfähigkeit, aufbauend auf den Projektergebnissen, weitere Projektvorhaben initiiert werden, die über Kooperationen mit weiteren wissenschaftlichen Partnern, Forschungseinrichtungen und Unternehmen die Ausweitung und Nutzarmachung der Ergebnisse vorantreiben sollen. Ebenso wird der Transfer der Ergebnisse durch wissenschaftliche Publikationen und wiss. Weiterbildungsangebote betrieben.

Auf wissenschaftlicher Ebene hat das KIAFlex-Projekt eine Erweiterung der Forschungsaktivitäten im Themengebiet KI-basierte Prädiktion ermöglicht. Das DFKI nutzt die erzielten Ergebnisse, um in anschließenden weiteren Grundlagen- bzw. Anwendungsprojekten gemeinsam mit weiteren wissenschaftlichen Partnern sowie mit Institutionen und Unternehmen der Gesundheitsbranche daran anknüpfende Fragestellungen zu bearbeiten, sowohl im Kreis des KIAFlex-Partnernetzwerkes aber auch darüber hinaus. Die zentralen F&E-Themen in KIAFlex haben das Potenzial, in zukünftigen Verbundprojekten vertieft zu werden. Das DFKI wird dies aktiv vorantreiben.

5 Bekannt gewordener Fortschritt

Im Projektzeitraum sind uns keine Kenntnisse über Ergebnisse Dritter bekannt geworden, welche für die Durchführung des Vorhabens relevant waren. Hierzu wurden in regelmäßigen Abständen Literaturrecherchen durchgeführt.

6 Erfolgte und geplante Veröffentlichungen der Ergebnisse

Im November 2022 hatte das DFKI auf der MEDICA, der weltweit größten Fachmesse für Medizintechnik, einen Messestand, an dem das Projekt KIAFlex präsentiert wurde. Ziel war die aktive Diskussion der Projektidee und die Sensibilisierung für das Thema KI im Nachsorgemanagement; die Gespräche vor Ort lieferten vielfältige Impulse und konkretes Feedback zur Weiterentwicklung der Lösung.

Im Mai 2023 wurde das Projekt KIAFlex in einem gemeinsamen Vortrag¹ mit den Projektpartnern Nubedian und Empolis auf der KH-IT-Frühjahrstagung vorgestellt; dabei entstanden wertvolle Kontakte sowie Anknüpfungspunkte für weiterführende Kooperationen.

Im Oktober 2024 wurde das Projekt KIAFlex auf der International Conference on Process Mining in Rom im Rahmen des Workshops Process-Oriented Data Science for Healthcare mit einer Posterpräsentation zum Thema „The seamless transition to aftercare: Leveraging analogy-based AI techniques“ vorgestellt; die Diskussionen im internationalen Fachpublikum lieferten wertvolle Rückmeldungen zur methodischen Ausrichtung und eröffneten weitere Vernetzungs- und Kooperationsmöglichkeiten.

Im April 2025 wurde ein Ergebnissteckbrief² in Zusammenarbeit mit den Projektpartnern erstellt und durch das Bundesministerium für Forschung, Technologie und Raumfahrt veröffentlicht; die Veröffentlichung trägt zur breiten Sichtbarkeit bei und unterstützt die weiterführende Verwertung der Ergebnisse.

¹<https://www.linkedin.com/feed/update/urn:li:activity:7066669353743626241/>

²<https://www.interaktive-technologien.de/service/ergebnissteckbriefe/kias/kiaflex>