Gefördert durch:

Bundesministerium für Forschung, Technologie und Raumfahrt

DATIpilot Modul 1 Innovationssprints

Sachbericht zum Verwendungsnachweis

Kurzbericht und Eingehende Darstellung öffentlich

HealthDatiPilot: Plattform für eine effiziente und standardisierte Entwicklung, Prüfung und Anwendung von KI-Modellen in der Medizin

HealthDatiPilot

	Institution	Name Projektleitung	Förderkennzeichen
Teilprojekt A „Anwendung und Evaluation“ (Koordination)	Universität Hamburg (UHH), Institute for Computational Systems Biomedicine	Prof. Dr. Jan Baumbach	03DPS1041A
Teilprojekt B „Implementierung und Integration“	Bitspark GmbH (BIT)	Dr. Julian Matschinske	03DPS1041B

LAUFZEIT: 01.10.2024 - 31.03.2026

Hamburg, den **29.09.2026**

Autorenschaft: Prof. Dr. Jan Baumbach; Dr. Julian Matschinske; Andreas Santoro

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Inhalt

1 Kurzbericht 2
1.1 Ausgangslage des Vorhabens 2
1.1.1. Ursprüngliche Aufgabenstellung 2
1.1.2. Ursprüngliche wissenschaftliche und/oder technische Arbeitsziele 2
1.1.3. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde 3
1.2 Ablauf des Vorhabens 3
1.3 Ergebnisse des Vorhabens 4
1.3.1. Wesentliche Ergebnisse 4
1.3.2. Zusammenarbeit mit anderen Stellen 4
2 Eingehende Darstellung 5
2.1 Durchgeführter Arbeitsplan 5
2.2 Erreichte Meilensteine 13
2.3 Erreichte Ergebnisse 15
2.3.1 Erreichung der Gesamtziele 15
2.3.2 Erreichung der wissenschaftlichen und/oder technischen Arbeitsziele 16
2.4 Verwertung 17
2.4.1 Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses 17
2.4.2 Fortschritt Dritter auf dem Gebiet des Vorhabens 18
2.4.3 Erfolgte oder geplante Veröffentlichungen der Ergebnisse 18
2.4.4 Notwendigkeit und Angemessenheit der geleisteten Projektarbeit 19

1 Kurzbericht

1.1 Ausgangslage des Vorhabens

1.1.1. Ursprüngliche Aufgabenstellung

Gesamtziel des Innovationssprints	Ziel des Innovationssprints ist die Entwicklung einer Plattform für eine effiziente und standardisierte Entwicklung, Prüfung und Anwendung von KI-Modellen in der Medizin. Dabei sollen die oft fehlenden Bioinformatikkenntnisse im klinischen Umfeld verfügbar gemacht werden, ohne dass Daten ausgetauscht werden müssen.
Beitrag des Teilprojekts A „Anwendung und Evaluation“ (UHH)	Ziel ist die Entwicklung einer Plattform, die einen effizienten und standardisierten Einsatz von KI-Modellen in medizinischen Umgebungen erlaubt. Dabei sollen die im klinischen Umfeld oft fehlenden Bioinformatik-Kapazitäten über eine Plattform zur Verfügung gestellt werden, ohne dass ein Austausch von Daten erforderlich ist. Das Projekt steht damit unter der Überschrift “datenschutzkonforme KI in der Medizin”. Das Teilprojekt “Anwendung und Evaluation” beinhaltet die Entwicklung der initialen KI-Modelle, Aufbereitung der zu integrierenden Daten und Erstellung der graphischen Benutzeroberflächen sowie die technische und praktische Evaluation der Plattform.
Beitrag des Teilprojekts B „Implementierung und Integration“ (BIT)	Ziel ist die Entwicklung einer Plattform, die einen effizienten und standardisierten Einsatz von KI-Modellen in medizinischen Umgebungen erlaubt. Dabei sollen die im klinischen Umfeld oft fehlenden Bioinformatik-Kapazitäten über eine Plattform zur Verfügung gestellt werden, ohne dass ein Austausch von Daten erforderlich ist. Das Projekt steht damit unter der Überschrift “datenschutzkonforme KI in der Medizin”. Das Teilprojekt „Implementierung und Integration“ umfasst die Anbindung der HealthDatiPlatform an andere Systeme, das gemeinsame Datenmodell mit Datenbank und Schnittstellen sowie die technische Umgebung für das Training und die Anwendung von KI-Modellen.

1.1.2. Ursprüngliche wissenschaftliche und/oder technische Arbeitsziele

Geplante konkrete Arbeitsziele: Gesamtprojekt
Verteiltes System auf klinischen und wissenschaftlichen Rechenressourcen mit zentraler Steuereinheit; klinische Daten bleiben verteilt gespeichert.
Einheitliches, harmonisiertes Datenmodell zur Integration bestehender heterogener Daten; Training und Evaluation von KI-Modellen ohne direkten Datenzugriff („delegiertes Lernen“).
Trainierte Modelle stehen plattformweit zur Verfügung und sind über das gemeinsame Datenmodell unmittelbar auch mit Daten anderer Partner zur Vorhersage nutzbar.
Geplante konkrete Arbeitsziele: Teilprojekt A
Auswahl geeigneter Datensätze; Sichtung der bisher verwendeten Systeme.

Auswahl geeigneter KI-Algorithmen.
Entwicklung erster Modelle auf dem entwickelten Datenmodell.
Nachvollziehbarkeit der Modellvorhersagen sicherstellen.
Evaluation in praktischen Anwendungen.
Geplante konkrete Arbeitsziele: Teilprojekt B
Entwurf und Umsetzung des gemeinsamen Datenmodells.
Entwicklung der verteilten Datenbank auf Grundlage des gemeinsamen Datenmodells.
Entwicklung der erweiterbaren verteilten Trainings- und Ausführungsumgebung.
Entwicklung und Dokumentation der Benutzerschnittstellen und technischen Schnittstellen.
Betrieb und Wartung der zentralen Ressourcen zur Aufrechterhaltung der Funktionsfähigkeit des Gesamtsystems.

1.1.3. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Biomedizinische und klinische Daten werden in unterschiedlichen Formaten erfasst. Für ihre gemeinsame Nutzung bestehen Standards zum Austausch und zur einheitlichen Beschreibung medizinischer Daten. Für Analysen unter Wahrung des Datenschutzes werden verschiedene Verfahren untersucht: Beim föderierten Lernen werden Modelle an mehreren Datenstandorten trainiert und die Ergebnisse zusammengeführt. Andere Ansätze ermöglichen gemeinsame Berechnungen, bei denen die Beteiligten ihre Eingabedaten voreinander schützen, oder Berechnungen mit verschlüsselten Daten. Das Projekt verfolgt den verwandten Ansatz des delegierten Lernens: Eine beauftragte Analyse wird an einem Datenstandort ausgeführt; die einzelnen Patientendaten verbleiben dort. Die Beschränkung auf einen Datensatz vereinfacht die technische Umsetzung. Anwendungen von KI in der Medizin umfassen beispielsweise radiologische Assistenzsysteme zur Erkennung von Tumoren und Frakturen.

1.2 Ablauf des Vorhabens

Geplant war ein 18-monatiger Ablauf vom 01.10.2024 bis zum 31.03.2026 mit fünf Arbeitsbereichen: Daten (DAT), Algorithmen (ALG), System (SYS), Benutzeroberflächen (GUI) und Evaluation (EVA). Die Partner sollten sich wöchentlich abstimmen und die Entwicklung schrittweise planen, umsetzen und prüfen. Die UHH übernahm die Koordination und den regelmäßigen Austausch in Besprechungen und Workshops; BIT gestaltete den Entwicklungsprozess.

UHH und BIT bearbeiteten das Vorhaben gemeinsam: Die UHH brachte die wissenschaftlichen Methoden und deren Bewertung ein; BIT setzte die technischen Komponenten und ihren Betrieb um. Die entwickelte Plattform führt Analysen dort aus, wo die Daten gespeichert sind. Forschende können eine Berechnung beauftragen und zusammengefasste Ergebnisse erhalten, ohne die einzelnen Patientendaten zu übertragen. Für die technische Erprobung wurde der Forschungsdatensatz MIMIC-IV lokal verarbeitet. Der gewählte Ansatz des delegierten Lernens und die einzelnen Entwicklungsschritte werden in Teil II erläutert. BIT beschäftigte während des gesamten Förderzeitraums von Oktober 2024 bis März 2026 einen Mitarbeiter in Vollzeit ausschließlich für HealthDati. Die im Antrag vorgesehene Unterstützung durch einen Werkstudenten wurde nicht in Anspruch genommen.

1.3 Ergebnisse des Vorhabens

1.3.1. Wesentliche Ergebnisse

Entwickelt wurden ein gemeinsames Datenmodell, die technische Umgebung für das Training und die Anwendung von KI-Modellen sowie Benutzeroberflächen zur Steuerung der Analysen und Darstellung der Ergebnisse. Die Erprobung mit bereits erhobenen Krankenhausdaten zeigte, dass sich der Analyseablauf lokal ausführen und die Vorhersagequalität des untersuchten Modells systematisch bewerten lässt. Dabei wurden auch die Verlässlichkeit der angezeigten Risiken und Unterschiede zwischen Patientengruppen untersucht. Die Funktionsfähigkeit der Software wurde auf Systemen geprüft, die mit der vorgesehenen Klinikhardware vergleichbar waren. Die Benutzeroberflächen wurden ergänzend aus der Perspektive einer klinischen Anwendung besprochen. Die Rückmeldungen flossen in die bis Projektende umgesetzte Vereinfachung der Oberfläche und die Ausarbeitung des Benutzerhandbuchs ein. Weitere Rückmeldungen werden insbesondere nach Veröffentlichung des Quellcodes erwartet. Die Untersuchungen und ihre Ergebnisse sind in Teil II beschrieben.

1.3.2. Zusammenarbeit mit anderen Stellen

Für die technische Erprobung wurde MIMIC-IV v3.1 verwendet, ein Forschungsdatensatz mit bereits erhobenen Krankenhausdaten, der nach Freigabe über PhysioNet zugänglich ist. Er ermöglichte die Untersuchung des Analyseablaufs und der Vorhersagequalität auf realen klinischen Daten. Die Verarbeitung erfolgte lokal; ausgegeben wurden zusammengefasste Auswertungsergebnisse.

Die Erprobung erfolgte unter Berücksichtigung der Datenschutzanforderungen und klinikinternen Rahmenbedingungen auf Systemen, die mit der vorgesehenen Klinikhardware vergleichbar waren. Dabei wurde die Funktionsfähigkeit der Software auf entsprechender Hardware geprüft.

Ergänzend zur technischen Evaluation wurden die Benutzeroberflächen aus der Perspektive einer klinischen Anwendung besprochen. Die Rückmeldungen zeigten, wie die Bedienung verständlicher gestaltet und stärker an den Anforderungen der vorgesehenen Nutzer ausgerichtet werden konnte. Sie flossen in die Vereinfachung der Oberfläche und die Ausarbeitung des Benutzerhandbuchs ein. Die Vereinfachung der Oberfläche wurde bis Projektende umgesetzt. Weitere Rückmeldungen werden insbesondere nach Veröffentlichung des Quellcodes erwartet.

2 Eingehende Darstellung

2.1 Durchgeführter Arbeitsplan

Arbeitspaket	Partner	abgeschlossen	nicht abgeschlossen	nicht durchgeführt
DAT.1 Sichtung der Datenlandschaft	UHH	☒	☐	☐
DAT.2 Harmonisiertes Datenmodell	BIT (verantw.), UHH	☒	☐	☐
DAT.3 Einbindung des Datenmodells und Datenimport	BIT (verantw.), UHH	☒	☐	☐
ALG.1 Sichtung der bestehenden KI-Ansätze	UHH	☒	☐	☐
ALG.2 Training erster KI-Modelle	UHH	☒	☐	☐
ALG.3 Integration von KI-Algorithmen	UHH	☒	☐	☐
ALG.4 Umgebung für die Anwendung von KI-Modellen	BIT	☒	☐	☐
SYS.1 Entwurf der Systemarchitektur	BIT	☒	☐	☐
SYS.2 Technische Grundlage und automatisierte Entwicklungsabläufe	BIT	☒	☐	☐
SYS.3 Umgebung und Schnittstelle für das Training am Datenstandort	BIT	☒	☐	☐
SYS.4 Integration mit bestehenden Systemen (technische Umsetzung und Erprobung abgeschlossen)	UHH (verantw.), BIT	☒	☐	☐
SYS.5 Auslieferungsfähiges System	BIT	☒	☐	☐
GUI.1 Benutzeroberfläche für das KI-Training	UHH (verantw.), BIT	☒	☐	☐
GUI.2 Benutzeroberfläche für die Anwendung der KI-Modelle	UHH (verantw.), BIT	☒	☐	☐
EVA.1 Evaluation der KI-Modelle	UHH	☒	☐	☐
EVA.2 Evaluation des Gesamtsystems (technische Bewertung und Rückmeldungen abgeschlossen; Erprobung im klinischen Alltag als Folgeanwendung)	UHH	☐	☒	☐

Arbeitspaket DAT.1 - Sichtung der Datenlandschaft (UHH, Monat 1-3)

Geplantes Ziel: Eine systematische und möglichst vollständige Übersicht über relevante Datentypen, verfügbare Datensätze und Standards für ihre Vereinheitlichung. Dazu gehören Datenmodelle und Schnittstellen von Krankenhausinformationssystemen (KIS) und Laborinformationssystemen (LIS).

Das geplante Ergebnis wurde erreicht. Der Sichtungsbericht erfasst relevante Datentypen, öffentliche Datensätze, medizinische Klassifikationen und Schnittstellen klinischer Informationssysteme. Untersucht wurden die Datensätze MIMIC-IV v3.1, eICU und Synthea sowie unter anderem FHIR R4, HL7 v2 und OMOP-CDM als Grundlagen für Datenaustausch und gemeinsame Datenstrukturen. Der entstandene Katalog bildete die Grundlage für DAT.2. Als erste Fragestellung wurde die Vorhersage der Sterblichkeit während eines Krankenhausaufenthalts anhand von MIMIC-IV ausgewählt. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in medizinischen Datenstrukturen; Fachliteratur, Standards und Dokumentation öffentlicher Datensätze. Personalansatz laut Antrag: UHH 1,5 Personenmonate.

Arbeitspaket DAT.2 - Harmonisiertes Datenmodell (BIT (verantw.), UHH, Monat 4-6)

Geplantes Ziel: Ein gemeinsames Datenmodell auf Grundlage von DAT.1, das die identifizierten Daten zusammenführt und ihre Nutzung durch die ausgewählten KI-Modelle ermöglicht.

Das geplante Ergebnis und Meilenstein M1 wurden im März 2025 erreicht. Das gemeinsame Datenmodell beschreibt Personen, Krankenhausaufenthalte, Messwerte und Diagnosen. Es bildet die Grundlage für den Datenaustausch und die Datenbank. Für das erste KI-Modell wurde die Sterblichkeit während des Krankenhausaufenthalts als Zielgröße festgelegt. Prüfungen erkennen widersprüchliche Datendefinitionen und die unzulässige Nutzung von Informationen, die bei einer Vorhersage noch nicht vorliegen. Laborwerte werden mit einheitlichen Maßeinheiten erfasst. Für die Erprobung lassen sich künstliche Beispieldaten mit vorgegebenen Zusammenhängen reproduzierbar erzeugen. Das Datenmodell wurde automatisiert geprüft. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in Datenmodellierung und Softwareentwicklung; gemeinsames Datenmodell, Prüfwerkzeuge und künstlich erzeugte Beispieldaten. Personalansatz laut Antrag: UHH 1; BIT 2 Personenmonate.

Arbeitspaket DAT.3 - Integration des Datenmodells und Datenimport (BIT (verantw.), UHH, Monat 7-9)

Geplantes Ziel: Umsetzung des Datenmodells in einer Datenbank mit Schnittstellen, über die Daten für das Training und die Anwendung von KI-Modellen bereitstehen.

Das geplante Ergebnis wurde erreicht. Daten im FHIR-R4- oder CSV-Format werden eingelesen, in das gemeinsame Datenmodell überführt und vor dem Speichern geprüft. Bei Fehlern wird die Übernahme abgebrochen; wiederholte Importe derselben Daten erzeugen keine zusätzlichen Einträge. Das Datenmodell wurde um weitere Wertearten, Angaben zur Krankenhausaufnahme und die Kennzeichnung bereits bei Aufnahme vorliegender Erkrankungen ergänzt. Bisherige Datenstrukturen bleiben nutzbar. Der gesamte Ablauf wurde mit einer laufenden Datenbank automatisiert geprüft. Künstlich erzeugte Beispieldaten stehen zur Verfügung. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in Datenmodellierung und Softwareentwicklung; Datenbank, Programme zur Datenübernahme und Testdaten in FHIR- und CSV-Formaten. Personalansatz laut Antrag: UHH 2; BIT 2 Personenmonate.

Arbeitspaket ALG.1 - Sichtung der bestehenden KI-Ansätze (UHH, Monat 1-3)

Geplantes Ziel: Eine systematische Übersicht über bestehende biomedizinische KI-Verfahren. Verglichen werden Daten- und Ressourcenbedarf, erforderliche Schnittstellen, Zuverlässigkeit, Erklärbarkeit und klinische Relevanz.

Das geplante Ergebnis wurde erreicht. Die Übersicht konzentrierte sich auf standortübergreifendes Lernen, Erklärbarkeit und Ressourcenbedarf. Untersucht wurden föderiertes Lernen, die geschützte Zusammenführung von Ergebnissen, Verfahren zur Begrenzung von Rückschlüssen auf einzelne Personen und Berechnungen auf verschlüsselten Daten. Als Referenzen dienten FeatureCloud, FAMHE, PriCell und FeTS. Ein Versuch mit dem Verschlüsselungsverfahren CKKS ergab einen etwa um den Faktor 1.000 höheren Aufwand; dieser Ansatz wurde deshalb nicht weiterverfolgt. Ein Prototyp für föderiertes Lernen mit FedAvg erreichte eine AUROC von 0,68 gegenüber 0,71 bei zentralem Training. Die AUROC beschreibt, wie gut ein Modell die beiden Ergebnisgruppen unterscheiden kann. Bei der Modellwahl wurden die Eignung für tabellarische klinische Daten, der Rechenbedarf und die Erklärbarkeit der Vorhersagen berücksichtigt. Für die weitere Entwicklung wurden logistische Regression als Vergleichsmodell, verstärkte Entscheidungsbäume und ein mehrschichtiges neuronales Netz ausgewählt. Ergänzende Vergleiche dieser Modellfamilien sind unter EVA.2 beschrieben. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in maschinellem Lernen und Datenschutz; Fachliteratur, Modellvergleiche und Prototypen zu föderiertem Lernen und Verschlüsselung. Personalansatz laut Antrag: UHH 1,5 Personenmonate.

Arbeitspaket ALG.2 - Training erster KI-Modelle (UHH, Monat 4-6)

Geplantes Ziel: Entwicklung erster KI-Modelle und Prüfung ihrer Eignung für das gemeinsame Datenmodell.

Das geplante Ergebnis wurde erreicht. Das Modell K1 zur Vorhersage der Sterblichkeit während eines Krankenhausaufenthalts wurde für das Training am Datenstandort entwickelt. Es nutzt ausgewählte Laborwerte aus den ersten 24 Stunden, Begleiterkrankungen nach Elixhauser und demografische Angaben. Prüfungen verhindern die Nutzung von Angaben, die zum Vorhersagezeitpunkt noch nicht verfügbar sind. Daten derselben Person werden nicht zugleich für Training und Prüfung verwendet. Umgesetzt wurden logistische Regression als Standardmodell und verstärkte Entscheidungsbäume. Multimodale Ansätze wurden im Rahmen der Sichtung veröffentlichter Arbeiten zu medizinischen Bilddaten, insbesondere FeTS, betrachtet. Die praktische Modellentwicklung konzentrierte sich auf die gemeinsame Auswertung tabellarischer Labor-, Diagnose- und demografischer Daten. Angaben zum Einfluss der Eingangsmerkmale werden ausschließlich zusammengefasst weitergegeben. Mit künstlich erzeugten Daten wurde eine AUROC von 0,715 erreicht; die erreichbare Modellgüte war durch den darin vorgegebenen Zusammenhang begrenzt. Für reale Daten wurde eine AUROC von etwa 0,85 oder höher angestrebt. Die Umsetzung wurde automatisiert geprüft. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in maschinellem Lernen und Datenaufbereitung; Programme zur Datenaufbereitung und zum Modelltraining, künstlich erzeugte Daten und eine lokale Trainingsumgebung. Personalansatz laut Antrag: UHH 2 Personenmonate.

Arbeitspaket ALG.3 - Integration von KI-Algorithmen (UHH, Monat 10-12)

Geplantes Ziel: Anbindung der vorbereiteten KI-Verfahren an die Schnittstellen aus SYS.3 und Umsetzung erster funktionsfähiger Beispiele.

Das geplante Ergebnis wurde als Beitrag zu Meilenstein M3 erreicht. K1 wurde in den Analyseablauf mit MIMIC-IV eingebunden. Datenimport und Modelltraining verwenden dieselben Datendefinitionen. Die vorhergesagten Wahrscheinlichkeiten können am Datenstandort nach Platt oder mit isotonischer Regression an beobachtete Häufigkeiten angepasst werden. Ein optionales Verfahren begrenzt redundante Eingangsmerkmale. Die Untersuchung mit MIMIC-IV ergab eine AUROC von 0,860 (95-%-Konfidenzintervall 0,824–0,893). Durch die Anpassung nach Platt sank der mittlere Kalibrierungsfehler ECE von 0,160 auf 0,007; die vorhergesagten Wahrscheinlichkeiten stimmten damit besser mit den beobachteten Häufigkeiten überein. Die automatisierten Tests wurden erfolgreich durchlaufen; auch die automatische Prüfung auf dem Server war in Betrieb. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in Modelltraining und statistischer Auswertung; Abläufe für das Modell K1, Software zur Ausführung am Datenstandort und lokale MIMIC-IV-Daten. Personalansatz laut Antrag: UHH 1,5 Personenmonate.

Arbeitspaket ALG.4 - Entwicklung der Umgebung zur Modellanwendung (BIT, Monat 13-15)

Geplantes Ziel: Trainierte Modelle sollen am Datenstandort die benötigten Daten abrufen, passend aufbereiten und Vorhersagen berechnen können. Die Ergebnisse werden über die Benutzeroberfläche aus GUI.2 zugänglich.

Das geplante Ergebnis wurde erreicht. Trainierte Modelle können direkt am Datenstandort angewendet werden. Das Training und die anschließende Anwendung sind technisch voneinander getrennt. Gegenüber einer Referenzberechnung lag die Abweichung bei etwa $5 \cdot 10^{-7}$. Für den in EVA.2/M4 gewählten Entscheidungsschwellenwert ist die Anpassung der Wahrscheinlichkeiten nach Platt vorgeschrieben. Bei der Wahl dieses Schwellenwerts wurden übersehene positive Fälle zehnmal so stark gewichtet wie fälschlich als positiv eingestufte Fälle (10:1). Einzelne Risikowerte werden am Datenstandort angezeigt; übertragen werden ausschließlich zusammengefasste Ergebnisse. Eine zusätzliche Prüfung kontrolliert diese Beschränkung. Die Software zur Modellanwendung lässt sich gesondert installieren; Modelle können optional für die Nutzung mit anderen Programmen exportiert werden. Die automatisierten Tests wurden erfolgreich durchlaufen. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in Softwareentwicklung und Modellanwendung; gesonderte Umgebung für die Modellanwendung, gespeicherte Modelle und Referenzberechnungen. Personalansatz laut Antrag: BIT 2 Personenmonate.

Arbeitspaket SYS.1 - Entwurf der Systemarchitektur (BIT, Monat 1-2)

Geplantes Ziel: Entwurf des technischen Gesamtaufbaus für ein System, das in unterschiedlichen Umgebungen einsetzbar und um weitere Funktionen erweiterbar ist.

Das geplante Ergebnis wurde erreicht. Der technische Gesamtaufbau verbindet eine zentrale Auftragssteuerung mit der Datenverarbeitung an den beteiligten Standorten. Die Konzeption umfasst die lokale Aufbereitung von FHIR- und HL7-Daten, ihre Speicherung im gemeinsamen Datenmodell und die getrennte Bereitstellung der Softwarekomponenten. Die wesentlichen Entscheidungen sind begründet und mit Diagrammen dokumentiert. Ergänzend zum im Antrag vorgesehenen delegierten Lernen wurden föderiertes Lernen und die geschützte Zusammenführung von Ergebnissen untersucht. Ein funktionsfähiger Prototyp zeigte die zentrale Übersicht sowie grundlegende Abläufe der Nachrichtenübermittlung und Ergebniszusammenführung. Die lokale

Datenaufbereitung wurde konzeptionell skizziert. Der Entwurf wurde zum Stand des Zwischenberichts überprüft. Die Installation unter verschiedenen Betriebssystemen ist unter SYS.5 beschrieben. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse im Entwurf verteilter Softwaresysteme und im Datenschutz; Dokumentation des Systemaufbaus und technische Prototypen. Personalansatz laut Antrag: BIT 2 Personenmonate.

Arbeitspaket SYS.2 - Technische Grundlage und automatisierte Entwicklungsabläufe (BIT, Monat 3-6)

Geplantes Ziel: Aufbau des Grundsystems und eines Entwicklungsprozesses mit kurzen Arbeitsabschnitten. Änderungen am Programmcode werden automatisch geprüft, getestet und für die Bereitstellung vorbereitet.

Meilenstein M2 wurde im März 2025 erreicht. Für die Softwarekomponenten wurden automatische Prüf- und Bereitstellungsabläufe eingerichtet. Vor dem Modelltraining werden das Datenmodell, ein kurzer Trainingslauf und der Ausschluss unzulässiger Eingangsmerkmale geprüft. Die Entwicklung erfolgte in zweiwöchigen Arbeitsabschnitten mit regelmäßigen Besprechungen. Anfangs war der vorgesehene gemeinsame Prüfserver nicht verfügbar; vorübergehend galten deshalb erfolgreiche Tests auf den lokalen Entwicklungsrechnern als Abnahmekriterium. Ab Phase 3 führte ein eigener Server die automatisierten Prüfungen durch.
Ressourcen: Kenntnisse in Softwareentwicklung und Softwareprüfung; gemeinsame Entwicklungsumgebung, lokale Entwicklungsrechner, automatische Prüf- und Bereitstellungsabläufe sowie später ein eigener Prüfserver. Personalansatz laut Antrag: BIT 2 Personenmonate.

Arbeitspaket SYS.3 - Umgebung und Schnittstelle für das Training am Datenstandort (BIT, Monat 8-12)

Geplantes Ziel: Dokumentierte Schnittstellen und geeignete Softwareumgebungen sollen das Hinzufügen weiterer Algorithmen und die zentrale Steuerung von Trainingsaufträgen ermöglichen.

Meilenstein M3 wurde im September 2025 erreicht. Über eine dokumentierte Schnittstelle lassen sich Trainingsaufträge am Datenstandort auslösen, ihr Bearbeitungsstand verfolgen und Ergebnisse abrufen. Die lokale Software holt Aufträge selbst ab; dafür sind ausschließlich ausgehende Verbindungen erforderlich. Eine zusätzliche Prüfung kontrolliert die weitergegebenen Ergebnisse. Der vollständige Ablauf wurde mit MIMIC-IV erprobt. Die Einbindung des Referenzmodells K1 ist unter ALG.3 beschrieben. Die Datengrundlage umfasste 6.066 Personen und 14.525 Krankenhausaufenthalte. Bei 2,33 % der Aufenthalte verstarb die behandelte Person im Krankenhaus. Das lokal trainierte Modell K1 erreichte eine AUROC von 0,860. Entsprechend der Datennutzungsvereinbarung verließen ausschließlich zusammengefasste Ergebnisse den Datenstandort. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in der Entwicklung verteilter Softwaresysteme; zentrale Steuerungssoftware, dokumentierte Schnittstellen und Programme zur Auftragsbearbeitung am Datenstandort. Personalansatz laut Antrag: BIT 2 Personenmonate.

Arbeitspaket SYS.4 - Integration mit bestehenden Systemen (UHH (verantw.), BIT, Monat 13-14)

Geplantes Ziel: Bestehende klinische Systeme soweit möglich anbinden, um aktuelle Daten in die eigene Datenbank übernehmen zu können.

Die technischen Schnittstellen wurden umgesetzt. Daten aus FHIR-, HL7-v2- und CSV-Quellen werden an den Import aus DAT.3 übergeben. Vorgesehen sind verschlüsselte Verbindungen, das Einlesen größerer Datenbestände in Teilmengen und die Ersetzung direkter Personenkennungen durch Pseudonyme am Datenstandort. Der Ablauf wurde mit künstlich erzeugten FHIR-Daten geprüft. Damit liegt die technische Grundlage für die Anbindung klinischer Systeme vor. Auf dieser Grundlage kann die Plattform künftig mit aktuellen Daten aus den laufenden Systemen einer externen Klinik erprobt werden.
Ressourcen: Kenntnisse in klinischen Datenformaten und der Entwicklung von Schnittstellen; Schnittstellen für FHIR-, HL7-v2- und CSV-Quellen sowie künstlich erzeugte Testdaten. Personalansatz laut Antrag: UHH 1; BIT 1,5 Personenmonate.

Arbeitspaket SYS.5 - Auslieferungsfähiges System (BIT, Monat 15-18)

Geplantes Ziel: Fertigstellung des Systems für die eigenständige Installation in unterschiedlichen technischen Umgebungen, einschließlich dokumentierter Schnittstellen, Installationshilfen und weiterer Tests.

Meilenstein M6 wurde im März 2026 erreicht. Das Gesamtsystem kann einschließlich seiner Datenbank installiert werden. Die Modellverwaltung prüft die Identität und Unversehrtheit gespeicherter Modelle; bei Abweichungen wird ihre Verwendung gesperrt. Die Installation auf Windows, macOS und Linux wird unterstützt und kann wiederholt ausgeführt werden. Installation und Bereitstellung sind dokumentiert. Das Zusammenspiel der Komponenten wurde im laufenden System geprüft; dabei blieb der Zustand von Aufträgen auch nach einem Neustart erhalten. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in Softwareinstallation und im Zusammenspiel der Komponenten; Installationshilfen, Datenbank und Modellverwaltung. Personalansatz laut Antrag: BIT 2 Personenmonate.

Arbeitspaket GUI.1 - Benutzeroberfläche für das KI-Training (UHH (verantw.), BIT, Monat 10-13)

Geplantes Ziel: Eine Benutzeroberfläche zur Datenauswahl, Einrichtung des Modelltrainings und Anzeige des Fortschritts. Das Training soll am Datenstandort gesteuert werden können, ohne direkten Zugriff auf die zugrunde liegenden Datensätze zu gewähren.

Das geplante Ergebnis wurde erreicht. Die Benutzeroberfläche führt schrittweise durch die Auswahl von Datenstandort, Datenmodell und KI-Modell sowie die Einrichtung des Trainings. Sie zeigt Fortschritt, bisherige Aufträge und zusammengefasste Ergebnisse an. Dazu gehören Kennzahlen zur Vorhersagegüte, ein Vergleich vorhergesagter Wahrscheinlichkeiten mit beobachteten Häufigkeiten vor und nach der Kalibrierung sowie Angaben zum Einfluss einzelner Eingangsmerkmale. Neben einem Demonstrationsmodus mit künstlich erzeugten Daten besteht die Anbindung an die Trainingsschnittstelle aus SYS.3. Die Oberfläche wurde im Betrieb überprüft und in die automatisierten Tests aufgenommen. Entsprechend der Datennutzungsvereinbarung werden ausschließlich zusammengefasste Ergebnisse angezeigt. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in der Gestaltung und Entwicklung von Benutzeroberflächen; Benutzeroberfläche, Schnittstellen zur Steuerungssoftware und zusammengefasste Kennzahlen aus dem Modelltraining. Personalansatz laut Antrag: UHH 1,5; BIT 2 Personenmonate.

Arbeitspaket GUI.2 - Benutzeroberfläche für die Anwendung der KI-Modelle (UHH (verantw.), BIT, Monat 16-18)

Geplantes Ziel: Eine Benutzeroberfläche, mit der medizinisches Personal die Vorhersagen der ausgewählten Modelle nutzen und einordnen kann. Die Ergebnisse sollen in medizinisch geeigneter Form dargestellt werden. Als Beispiel für eine modellspezifische Darstellung nennt der Antrag die Markierung relevanter Bereiche in MRT-Bildern.

Das geplante Ergebnis wurde als Beitrag zu Meilenstein M5 erreicht. Die Benutzeroberfläche zeigt die vom Pilotmodell K1 berechneten Risikowerte und erläutert den Entscheidungsschwellenwert. Bei dessen Wahl wurden übersehene positive Fälle zehnmal so stark gewichtet wie fälschlich als positiv eingestufte Fälle (10:1). Angezeigt werden der Anteil tatsächlich erkannter positiver Fälle und der Anteil zutreffender positiver Vorhersagen. Für einzelne Patientengruppen wird die Übereinstimmung vorhergesagter Wahrscheinlichkeiten mit beobachteten Häufigkeiten dargestellt; bei Personen ab 75 Jahren ist besondere Vorsicht bei der Interpretation erforderlich. Die Oberfläche nutzt die in ALG.4 geprüfte Modellanwendung. Am Datenstandort lässt sich der Einfluss einzelner Eingangsmerkmale auf eine Vorhersage erläutern. Weitergegeben werden ausschließlich zusammengefasste Ergebnisse. Automatisierte Tests prüfen, ob bereits geprüfte Funktionen auch nach Änderungen weiter funktionieren. Die Rückmeldungen zu den Benutzeroberflächen aus der Perspektive einer klinischen Anwendung und die daraus abgeleitete Vereinfachung sind unter EVA.2 beschrieben.
Ressourcen: Kenntnisse in der Gestaltung von Benutzeroberflächen und der Darstellung von Modellvorhersagen; Benutzeroberfläche, Schnittstelle zur Modellanwendung und Darstellungen der Modellvorhersagen. Personalansatz laut Antrag: UHH 3; BIT 0,5 Personenmonate.

Arbeitspaket EVA.1 - Evaluation der KI-Modelle (UHH, Monat 7-9)

Geplantes Ziel: Prüfung der biomedizinischen Nachvollziehbarkeit und der Vorhersagegüte unter unterschiedlichen Bedingungen. Dafür werden unter anderem wiederholt getrennte Trainings- und Testdaten verwendet.

Das geplante Ergebnis wurde erreicht. Der Bewertungsablauf für K1 gibt ausschließlich zusammengefasste Ergebnisse weiter. Untersucht wurden der Umgang mit veränderten Datenbedingungen und fehlenden Werten, die Kalibrierung, die Erklärbarkeit, Unterschiede zwischen Patientengruppen und der Ausschluss unzulässiger Informationen. Drei voneinander unabhängige Prüfungen des letzten Bereichs wurden erfolgreich durchlaufen. Der wichtigste Befund betraf zu sichere Wahrscheinlichkeitsangaben: Der maximale Kalibrierungsfehler MCE betrug 0,281. Deshalb wurde eine nachträgliche Anpassung empfohlen. Die Untersuchung nutzte künstlich erzeugte Daten zur Prüfung der Bewertungsmethoden; die damit erreichbare AUROC lag aufgrund des vorgegebenen Datenzusammenhangs bei etwa 0,72. Die Bewertung anhand realer MIMIC-IV-Daten ist unter EVA.2 beschrieben. Die automatisierten Tests wurden erfolgreich durchlaufen. Wesentliche Abweichungen gab es nicht.
Ressourcen: Kenntnisse in statistischer Modellbewertung; künstlich erzeugte Evaluationsdaten und Werkzeuge zur Bewertung der Modellgüte und Erklärbarkeit. Personalansatz laut Antrag: UHH 1 Personenmonat.

Arbeitspaket EVA.2 - Evaluation des Gesamtsystems (UHH, Monat 13-18)

Geplantes Ziel: Bewertung der Bedienbarkeit und Praxistauglichkeit bei einem klinischen Partner sowie Prüfung der technischen Erweiterbarkeit.

Der technische Teil der Gesamtsystembewertung wurde mit MIMIC-IV durchgeführt. Das Referenzmodell K1 verwendet eine logistische Regression und berücksichtigt Merkmale aus den ersten 24 Stunden sowie Begleiterkrankungen nach Elixhauser. Dokumentiert sind eine AUROC von 0,860 (95-%-Konfidenzintervall 0,824–0,893) und eine AUPRC von 0,213. Die AUROC beschreibt die Trennfähigkeit des Modells. Die AUPRC fasst über verschiedene Entscheidungsschwellen hinweg zusammen, welcher Anteil der positiven Fälle erkannt wird und welcher Anteil der positiven Vorhersagen zutrifft. Durch eine Anpassung nach Platt sank der mittlere Kalibrierungsfehler ECE von 0,160 auf 0,007; die vorhergesagten Wahrscheinlichkeiten stimmten danach besser mit den beobachteten Häufigkeiten überein. Die Auswertung umfasste außerdem einen Entscheidungsschwellenwert, der unterschiedliche Folgen von Fehlvorhersagen berücksichtigt, eine Entscheidungskurvenanalyse zum möglichen Nutzen verschiedener Schwellenwerte, eine Prüfung anhand zeitlich getrennter Daten und Auswertungen nach Patientengruppen. Die Datengrundlage umfasst 6.066 Personen und 14.525 Krankenhausaufenthalte; Daten für Training, Kalibrierung und abschließende Prüfung wurden getrennt. Die Untersuchung vorhandener Daten prüft den Analyseablauf und die Modellgüte.

Die Erprobung erfolgte unter Berücksichtigung der Datenschutzanforderungen und klinikinternen Rahmenbedingungen auf Systemen, die mit der vorgesehenen Klinikhardware vergleichbar waren. Dabei wurde die Funktionsfähigkeit der Software auf entsprechender Hardware geprüft.

Die Prüfung des Speicher- und Rechenbedarfs wurde im Förderzeitraum begonnen. Bei künstlich erzeugten Daten im Umfang von 14.525 Krankenhausaufenthalten gelang das Training mit einem zugewiesenen Prozessorkern und 1 GB Arbeitsspeicher ohne Grafikkarte. Für den fünffachen Datenumfang reichten zwei Prozessorkerne und 4 GB Arbeitsspeicher aus.

Ergänzende Ressourcenmessungen und Modellvergleiche zu den im Förderzeitraum begonnenen Untersuchungen werden mit dem bis zur Berichtserstellung vorliegenden Auswertungsstand zusammengefasst. Beim Training mit 14.525 realen MIMIC-IV-Krankenhausaufenthalten betrug der maximale Arbeitsspeicherbedarf 650 bis 704 MB und die Prozessorzeit 36 bis 71 Sekunden. Alle drei Trainingsabläufe gelangen mit einem zugewiesenen Prozessorkern und 1 GB Arbeitsspeicher ohne Grafikkarte; die vorgelagerte Datenaufbereitung wurde gesondert durchgeführt. Im Modellvergleich anhand von fünf Aufteilungen derselben Kohorte erreichten verstärkte Entscheidungsbäume im Mittel eine AUROC von 0,894 gegenüber 0,870 bei logistischer Regression. Die untersuchten mehrschichtigen neuronalen Netze zeigten über die Datenaufteilungen hinweg keinen durchgängigen Vorteil gegenüber der logistischen Regression.

Die technische Erweiterbarkeit zeigt sich in den dokumentierten Trainingsschnittstellen, den untersuchten Modellvarianten und der Möglichkeit, weitere Versionen des Referenzmodells einzubinden. Logistische Regression und verstärkte Entscheidungsbäume wurden im Training untersucht; für die Anwendung trainierter Modelle wurde die Einbindung von Varianten der logistischen Regression umgesetzt.

Ergänzend zur technischen Evaluation wurden die Benutzeroberflächen aus der Perspektive einer klinischen Anwendung besprochen. Die Rückmeldungen zeigten, wie die Bedienung verständlicher gestaltet und stärker an den Anforderungen der vorgesehenen Nutzer ausgerichtet werden konnte. Sie flossen in die Vereinfachung der Oberfläche und die Ausarbeitung des Benutzerhandbuchs ein. Die Vereinfachung der Oberfläche und das Benutzerhandbuch wurden bis Projektende fertiggestellt.

EVA.2 umfasst damit die technische Auswertung, die Prüfung auf vergleichbarer Hardware und die Besprechung der Benutzeroberflächen aus der Perspektive einer klinischen Anwendung. Weitere Rückmeldungen zur überarbeiteten Oberfläche werden insbesondere nach Veröffentlichung des Quellcodes erwartet.

Ressourcen: Kenntnisse in statistischer Auswertung, Softwareprüfung und der Bewertung von Benutzeroberflächen; Systeme mit vergleichbarer Hardware, lokale MIMIC-IV-Daten, Werkzeuge zur Modell- und Gesamtsystembewertung sowie die Benutzeroberflächen zur Besprechung ihrer Verständlichkeit und Bedienbarkeit. Personalansatz laut Antrag: UHH 2 Personenmonate.

2.2 Erreichte Meilensteine

Meilensteine Gesamtprojekt

Beschreibung: messbare Zielparameter, Steuerungsfunktion	Nr.	geplanter Monat	realer Monat	Erläuterung (planabweichend/nicht erreicht)
Gemeinsames Datenmodell mit festgelegter Datenstruktur - Festlegung und Dokumentation der Datenstruktur in einem gängigen elektronischen Format. - Kriterium: Das Datenmodell soll alle relevanten biomedizinischen Daten aufnehmen können, die in DAT.1 identifiziert wurden. - Entscheidungsfunktion: Falls sich nicht alle Daten in ein gemeinsames Modell aufnehmen lassen, wird dies bei der Auswahl der Algorithmen im Arbeitsbereich ALG berücksichtigt. Gründe können technische oder konzeptionelle Unterschiede sein.	1	6	6	planmäßig erreicht
Bedienbares Gesamtsystem mit Benutzeroberfläche - Anwendung, die unmittelbar von klinischem Fachpersonal sowie Bioinformatikern verwendet werden kann. - Kriterium: Fertige Benutzeroberfläche mit Dokumentation und Bedienungsanleitung.	5	18	18	Planmäßig erreicht. Die Benutzeroberfläche wurde anhand der Rückmeldungen aus der Perspektive einer klinischen Anwendung bis Projektende vereinfacht; das umfassende Benutzerhandbuch wurde bis Projektende fertiggestellt. Weitere Rückmeldungen werden insbesondere nach Veröffentlichung des Quellcodes erwartet.

Meilensteine Teilprojekt A

Beschreibung: messbare Zielparameter, Steuerungsfunktion	Nr.	geplanter Monat	realer Monat	Erläuterung (planabweichend/nicht erreicht)
Schnittstelle für das Training von KI-Modellen am Datenstandort - Umsetzung der Schnittstelle und Dokumentation ihrer Nutzung. - Kriterium: Externe Entwicklerinnen und Entwickler müssen anhand der Schnittstelle und ihrer Dokumentation eigene Algorithmen für das System erstellen können.	3	12	12	planmäßig erreicht
Bericht zur Prüfung des KI-Trainings - Bericht über den Bedarf an Speicher und Rechenleistung sowie die Leistungsfähigkeit des KI-Trainings anhand geeigneter	4	15	15	planmäßig erreicht

Kennzahlen. Die Ergebnisse zum Speicher- und Rechenbedarf sind unter EVA.2 in Abschnitt 2.1 beschrieben. - Kriterium: Die in Kliniken eingesetzten Teile des Systems müssen mit der dort typischerweise verfügbaren Rechentechnik auskommen. - Entscheidungsfunktion: Geeignete Algorithmen werden entsprechend gekennzeichnet oder ausgewählt, damit sie unter den vorgesehenen Bedingungen wie erwartet funktionieren.				

Meilensteine Teilprojekt B

Beschreibung: messbare Zielparameter, Steuerungsfunktion	Nr.	geplanter Monat	realer Monat	Erläuterung (planabweichend/nicht erreicht)
Automatisierte Softwareprüfung und Bereitstellung - Einrichtung eines dauerhaft nutzbaren Ablaufs für automatisierte Tests, die Prüfung des Programmcodes und die Bereitstellung der Software. Die Maßnahmen zur Qualitätssicherung werden dokumentiert. - Kriterium: Vollständig eingerichteter Ablauf und Bericht über die Maßnahmen zur Fehlervermeidung und zur zuverlässigen Funktion des Systems. Die Testabdeckung soll in jeder Teilkomponente mindestens 80 % betragen. Sie beschreibt, welche Teile des Programmcodes bei den Tests ausgeführt werden. - Steuerungsfunktion: Falls die fortlaufenden Prüfungen aus SYS.2 kritische Fehler nicht ausreichend verhindern, werden zusätzliche Maßnahmen entwickelt. Dazu zählen Tests einzelner Funktionen und vollständiger Abläufe.	2	6	6	Planmäßig erreicht. Die Tests sind in den automatisierten Entwicklungsablauf eingebunden.
Integriertes Gesamtsystem mit einsatzfähiger Datenbank, Installationspaketen und Dokumentation - Bereitstellung von Installationspaketen und Skripten für unterschiedliche Einsatzumgebungen mit zugehöriger Dokumentation. - Kriterium: Das System ist unter Windows, macOS, gängigen Linux-Distributionen und mit Docker installierbar.	6	18	18	planmäßig erreicht

2.3 Erreichte Ergebnisse

2.3.1 Erreichung der Gesamtziele

Gesamtziel des Innovationssprints	Ziel des Innovationssprints ist die Entwicklung einer Plattform für eine effiziente und standardisierte Entwicklung, Prüfung und Anwendung von KI-Modellen in der Medizin. Dabei sollen die oft fehlenden Bioinformatikkenntnisse im klinischen Umfeld verfügbar gemacht werden, ohne dass Daten ausgetauscht werden müssen.	Erreicht. Die Plattform lässt sich installieren und für Analysen am jeweiligen Datenstandort nutzen. Ein erstes Modell zur Vorhersage der Krankenhaussterblichkeit (K1) wurde anhand realer klinischer Daten trainiert, ohne die einzelnen Patientendaten zu übertragen. Die Bewertung der Vorhersagequalität ergab eine AUROC von 0,860. Auch die Verlässlichkeit der berechneten Risiken wurde untersucht und verbessert. Das trainierte Modell wurde zur Anwendung bereitgestellt.
Beitrag des Teilprojekts A (UHH)	Ziel ist die Entwicklung einer Plattform, die einen effizienten und standardisierten Einsatz von KI-Modellen in medizinischen Umgebungen erlaubt. Dabei sollen die im klinischen Umfeld oft fehlenden Bioinformatik-Kapazitäten über eine Plattform zur Verfügung gestellt werden, ohne dass ein Austausch von Daten erforderlich ist. Das Projekt steht damit unter der Überschrift “datenschutzkonforme KI in der Medizin”. Das Teilprojekt “Anwendung und Evaluation” beinhaltet die Entwicklung der initialen KI-Modelle, Aufbereitung der zu integrierenden Daten und Erstellung der graphischen Benutzeroberflächen sowie die technische und praktische Evaluation der Plattform.	Erarbeitet wurden das Modell K1, die Datenaufbereitung, Benutzeroberflächen für das Training und die Anwendung der Modelle sowie die technische Prüfung der Modelle und des Gesamtsystems. Die Benutzeroberflächen wurden ergänzend aus der Perspektive einer klinischen Anwendung besprochen. Die Rückmeldungen flossen in die bis Projektende umgesetzte Vereinfachung der Oberfläche und die Ausarbeitung des Benutzerhandbuchs ein. Weitere Rückmeldungen werden insbesondere nach Veröffentlichung des Quellcodes erwartet.
Beitrag des Teilprojekts B (BIT)	Ziel ist die Entwicklung einer Plattform, die einen effizienten und standardisierten Einsatz von KI-Modellen in medizinischen Umgebungen erlaubt. Dabei sollen die im klinischen Umfeld oft fehlenden Bioinformatik-Kapazitäten über eine Plattform zur Verfügung gestellt werden, ohne dass ein Austausch von Daten erforderlich ist. Das Projekt steht damit unter der Überschrift “datenschutzkonforme KI in der Medizin”. Das Teilprojekt „Implementierung und Integration“ umfasst die Anbindung der HealthDatiPlatform an andere Systeme, das gemeinsame Datenmodell mit Datenbank und Schnittstellen sowie die technische Umgebung für das Training und die Anwendung von KI-Modellen.	Erreicht. Umgesetzt wurden das gemeinsame Datenmodell und eine zugehörige Datenbank, die zentrale Steuerung der Plattform und die Ausführung am jeweiligen Datenstandort. Dokumentierte Schnittstellen ermöglichen das Training und die Anwendung der Modelle. Hinzu kommen automatisierte Softwareprüfungen, die zentrale Verwaltung der verfügbaren Modelle und Installationspakete für das Gesamtsystem (M6).

2.3.2 Erreichung der wissenschaftlichen und/oder technischen Arbeitsziele

Geplante konkrete Arbeitsziele: Gesamtprojekt	**Erfüllung**
Verteiltes System auf klinischen und wissenschaftlichen Rechenressourcen mit zentraler Steuereinheit; klinische Daten bleiben verteilt gespeichert.	Technisch umgesetzt: Eine zentrale Steuerung koordiniert die Analyseaufträge. Die Software am Datenstandort ruft diese Aufträge ab und führt sie lokal aus; die einzelnen Patientendaten verbleiben dort. Die Funktionsfähigkeit der Software wurde auf Systemen geprüft, die mit der vorgesehenen Klinikhardware vergleichbar waren. Ergänzend wurden die Benutzeroberflächen aus der Perspektive einer klinischen Anwendung besprochen. Die Ergebnisse bilden zugleich eine Grundlage für die gemeinsame Erprobung an mehreren Klinikstandorten.
Einheitliches, harmonisiertes Datenmodell zur Integration bestehender heterogener Daten; Training und Evaluation von KI-Modellen ohne direkten Datenzugriff („delegiertes Lernen“).	Erfüllt. Das gemeinsame Datenmodell ist umgesetzt. Das Training am Datenstandort wurde mit MIMIC-IV erprobt, ohne die einzelnen Patientendaten zu übertragen. Die Bewertung der Vorhersagequalität ergab eine AUROC von 0,860.
Trainierte Modelle stehen plattformweit zur Verfügung und sind über das gemeinsame Datenmodell unmittelbar auch mit Daten anderer Partner zur Vorhersage nutzbar.	Die Verwaltung der verfügbaren Modelle ist umgesetzt. Damit liegt die technische Grundlage vor, trainierte Modelle an weiteren Datenstandorten anzuwenden. Ihre Übertragbarkeit auf weitere klinische Daten ist gesondert zu prüfen.
Geplante konkrete Arbeitsziel: Teilprojekt A	**Erfüllung**
Auswahl geeigneter Datensätze; Sichtung der bisher verwendeten Systeme.	Erfüllt. MIMIC-IV wurde für die Untersuchung der Krankenhaussterblichkeit ausgewählt (DAT.1).
Auswahl geeigneter KI-Algorithmen.	Erfüllt. Geeignete Verfahren zum Schutz der Daten und geeignete Modelle wurden ausgewählt (ALG.1). Erarbeitet wurden ein Prototyp für föderiertes Lernen und das Modell K1 auf Grundlage der logistischen Regression.
Entwicklung erster Modelle auf dem entwickelten Datenmodell.	Erfüllt. Das Modell K1 wurde auf Grundlage des gemeinsamen Datenmodells am Datenstandort trainiert (ALG.2/ALG.3). Begleiterkrankungen wurden anhand der Elixhauser-Klassifikation berücksichtigt.
Nachvollziehbarkeit der Modellvorhersagen sicherstellen.	Erfüllt. Die Beiträge einzelner Eingangsgrößen zu den Vorhersagen wurden mit dem SHAP-Verfahren untersucht. Die wichtigsten Einflussgrößen waren klinisch plausibel. Zusätzlich wurde geprüft, ob unzulässig einbezogene Informationen die Entwicklung oder Bewertung der Modelle beeinflussten.
Evaluation in praktischen Anwendungen.	Die Modelle und der technische Ablauf wurden geprüft (EVA.1/EVA.2). Die Funktionsfähigkeit der Software wurde auf Systemen geprüft, die mit der vorgesehenen Klinikhardware vergleichbar waren. Ergänzend wurden die Benutzeroberflächen aus der Perspektive einer klinischen Anwendung besprochen. Die Rückmeldungen flossen in die bis Projektende umgesetzte Vereinfachung der Oberfläche und die

	Ausarbeitung des Benutzerhandbuchs ein. Weitere Rückmeldungen werden insbesondere nach Veröffentlichung des Quellcodes erwartet.
Geplante konkrete Arbeitsziele: Teilprojekt B	**Erfüllung**
Entwurf und Umsetzung des gemeinsamen Datenmodells.	Erfüllt (DAT.2/DAT.3, M1).
Entwicklung der verteilten Datenbank auf Grundlage des gemeinsamen Datenmodells.	Erfüllt. Der Datenimport wurde in einer laufenden Datenbank erprobt. Dabei wurde geprüft, ob die importierten Daten der festgelegten Datenstruktur entsprechen.
Entwicklung der erweiterbaren verteilten Trainings- und Ausführungsumgebung.	Erfüllt. Eine Schnittstelle ermöglicht das Training am Datenstandort (SYS.3/M3). Bei der Anwendung der Modelle stimmen die Ergebnisse bis auf eine Abweichung von etwa $5 \cdot 10^{-7}$ mit den Referenzberechnungen überein (ALG.4). Schnittstellen für FHIR, HL7 v2 und CSV sind umgesetzt (SYS.4).
Entwicklung und Dokumentation der Benutzerschnittstellen und technischen Schnittstellen.	Erfüllt. Die technischen Schnittstellen und ihre Nutzung sind für Entwickler dokumentiert. Die Benutzeroberflächen sind mit den zugehörigen Funktionen der Plattform verbunden (GUI.1/GUI.2).
Betrieb und Wartung der zentralen Ressourcen zur Aufrechterhaltung der Funktionsfähigkeit des Gesamtsystems.	Erfüllt. Automatisierte Prüfungen und die Bereitstellung der Software sind eingerichtet. Dazu gehören eine eigene Umgebung für automatisierte Tests, die Verwaltung der Modelle sowie Installationspakete (SYS.5).

Zusätzliche oder neue Arbeitsergebnisse, die im Projektverlauf entstanden sind:

Zu den nutzbaren technischen Ergebnissen gehören dokumentierte Schnittstellen für das Training und die Anwendung von Modellen am Datenstandort sowie für die Anbindung von Datenquellen in den Formaten FHIR, HL7 v2 und CSV. Die verfügbaren Modelle werden zentral verwaltet. Werkzeuge ermöglichen es, die berechneten Risiken an beobachtete Häufigkeiten anzupassen, die Vorhersagequalität in späteren Zeiträumen zu prüfen und Ergebnisse nach Patientengruppen auszuwerten. Hinzu kommen die Verwaltung gespeicherter Modelle und zugehöriger Dateien, die Anzeige und der Abbruch von Analyseaufträgen sowie Hilfen zur Installation. Die Benutzeroberfläche stellt außerdem dar, welche Eingangsgrößen eine einzelne Vorhersage beeinflussen und wie gut berechnete Risiken mit beobachteten Häufigkeiten übereinstimmen. Die automatisierten Softwaretests sind in den Entwicklungsablauf eingebunden. Die Besprechung der Benutzeroberflächen aus der Perspektive einer klinischen Anwendung und die daraus gewonnenen Rückmeldungen sind unter EVA.2 beschrieben. Die Benutzeroberfläche wurde bis Projektende vereinfacht; ergänzend liegt ein umfassendes Benutzerhandbuch vor. Weitere Rückmeldungen zur Bedienbarkeit werden insbesondere nach Veröffentlichung des Quellcodes erwartet.

2.4 Verwertung

2.4.1 Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses

Ursprünglicher Verwertungsplan (Vollantrag, Kapitel 4): Vorgesehen waren die gemeinsame Verwertung als Open-Source-Projekt, die Einbindung externer Nutzer und einer Entwicklergemein-

schaft sowie der langfristige Betrieb einer Projektwebsite mit Dokumentation und Nutzungsressourcen. BIT plante ein Dienstleistungsangebot für Installation, Betreuung und Anpassung innerhalb eines Jahres nach Projektende sowie die Monetarisierung der zentralen Steuereinheit über ein Gebührenmodell mit den erforderlichen Anpassungen innerhalb von 1,5 Jahren. Die UHH plante wissenschaftliche Kooperationen, Publikationen und Konferenzbeiträge; weitere Datentypen und biomedizinische Folgeprojekte wurden mit einem Horizont von drei bzw. fünf Jahren vorgesehen.

Fortgeschriebener Verwertungsplan: Die Plattform, das gemeinsame Datenmodell und die MIMIC-IV-Auswertungen bilden die Grundlage für die weitere Nutzung. Tutorial und Dokumentation sind zur Einreichung dieses Berichts nach dem Förderzeitraum öffentlich zugänglich. Die Veröffentlichung des Quellcodes ist für 2026 geplant; die Fundstellen sind unter 2.4.3 genannt. BIT strebt ein Dienstleistungsangebot für klinikspezifische Anpassungen, Einführung und laufende Unterstützung auf Grundlage frei nutzbarer Open-Source-Software an. Umfang und Aufbau richten sich nach dem Bedarf konkreter Einrichtungen. Die UHH plant die Nutzung in Forschungskooperationen und Folgeprojekten sowie wissenschaftliche Publikationen und Konferenzbeiträge. Weitere Rückmeldungen zur Nutzung und zur Benutzeroberfläche werden insbesondere nach der Quellcodefreigabe erwartet. Eine mögliche Folgeentwicklung ist das gemeinsame Training an mehreren Klinikstandorten. Die weiteren Schritte richten sich nach konkreten Vorhaben; zusätzliche feste Termine werden dafür nicht angesetzt.

2.4.2 Fortschritt Dritter auf dem Gebiet des Vorhabens

Im Berichtszeitraum erschienen weitere Arbeiten zur Bereitstellung und Auswertung klinischer Daten. Cadman et al. beschreiben mit MOLGENIS Armadillo einen schlanken Server für verteilte Analysen mit Verwaltung von Daten, Zugriffsrechten und Analyseumgebungen (Cadman et al., Bioinformatics 41, btae726, 2025; online Dezember 2024; https://doi.org/10.1093/bioinformatics/btae726). Sarrat-González et al. verbinden mit dsOMOP standardisierte klinische Daten nach dem OMOP-Datenmodell mit der Analyseplattform DataSHIELD (Sarrat-González et al., Bioinformatics 41, btaf286, 2025; https://doi.org/10.1093/bioinformatics/btaf286). Zhu et al. untersuchen mit FedWeight, wie sich Modelle durch die Gewichtung von Patientenfällen an unterschiedliche Patientengruppen anpassen lassen; die Auswertung umfasst eICU- und MIMIC-III-Daten (Zhu et al., npj Digital Medicine 8, 286, 2025; https://doi.org/10.1038/s41746-025-01661-8). Die im Januar 2026 veröffentlichte S-RACE-Plattform verbindet die Aufbereitung klinischer Daten mit Werkzeugen zur Modellentwicklung und Benutzeroberflächen für unterschiedliche Nutzergruppen (Traverso et al., npj Digital Medicine 9, 6, 2026; https://doi.org/10.1038/s41746-025-02132-w). Diese Arbeiten verdeutlichen die Bedeutung einer handhabbaren technischen Bereitstellung, einheitlich strukturierter Daten und der Bewertung von Modellen unter unterschiedlichen Datenbedingungen. HealthDatiPlatform verbindet ein gemeinsames Datenmodell mit der beauftragten Ausführung von Analysen am Datenstandort sowie Benutzer- und Programmierschnittstellen. Für die weitere Verwertung bieten die beschriebenen Entwicklungen Anknüpfungspunkte bei der Anbindung zusätzlicher Datenbestände und der standortbezogenen Bewertung von Modellen.

2.4.3 Erfolgte oder geplante Veröffentlichungen der Ergebnisse

Zum Zeitpunkt der Einreichung dieses Berichts sind Tutorial und Dokumentation unter https://healthdati.net öffentlich zugänglich. Die Veröffentlichung des Quellcodes unter https://github.com/healthdati/healthdati ist für 2026 geplant. Beide Schritte liegen nach dem För-

derzeitraum. Das Tutorial führt mit synthetischen Beispieldaten durch die Vorbereitung einer Analyse, die Anwendung eines Modells und die Einordnung der Ergebnisse. Das Benutzerhandbuch und ergänzende Hinweise unterstützen Bedienung, Installation und Weiterentwicklung. Wissenschaftliche Publikationen und Konferenzbeiträge sind als weitere Verwertung vorgesehen.

2.4.4 Notwendigkeit und Angemessenheit der geleisteten Projektarbeit

Entsprechende Fördermittel der EU waren weder ganz noch teilweise verfügbar. Die Förderung ermöglichte die gemeinsame Entwicklung des Datenmodells, der Plattform und der Methoden zur Bewertung der Modelle sowie deren technische Erprobung auf klinischen Forschungsdaten. Die Plattform setzt den im Antrag vorgesehenen Ansatz des delegierten Lernens um: Berechnungen werden am jeweiligen Datenstandort ausgeführt, ohne die zugrunde liegenden Patientendaten zentral zusammenzuführen. Ergänzend wurden föderierte Verfahren und weitere Verfahren zum Schutz der Daten untersucht. Die Ergebnisse liefern eine technische Grundlage für weitere Modelle, Datenquellen und die klinische Praxiserprobung.