

Beschluss

des Innovationsausschusses beim Gemeinsamen Bundesausschuss gemäß § 92b Absatz 3 SGB V zum abgeschlossenen Projekt CASSANDRA (01VSF20015)

Vom 21. August 2026

Der Innovationsausschuss beim Gemeinsamen Bundesausschuss hat in seiner Sitzung am 21. August 2026 zum Projekt CASSANDRA - *Clinical ASSist AND aleRt Algorithms - Frühe Detektion von postoperativen Komplikationen mit machine learning Algorithmen* (01VSF20015) folgenden Beschluss gefasst:

- I. Die im Projekt erzielten Ergebnisse werden an die Deutsche Gesellschaft für Allgemein- und Viszeralchirurgie e. V. (DGAV), die Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e. V. (DGAI), die Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V. (FhG), den Bundesverband Medizintechnologie e. V. (BVMed) sowie den Bundesverband Gesundheits-IT e. V. (bvitg) zur Information weitergeleitet.

Begründung

Das Projekt hat erfolgreich die Anwendung von Machine Learning (ML)-Modellen zur frühzeitigen Detektion von schwerwiegenden postoperativen Komplikationen bei Patientinnen und Patienten mit viszeralchirurgischen Major-Resektionen erprobt und evaluiert. Hierfür wurden Patientinnen und Patienten nach Eingriffen u. a. am Ösophagus, Magen oder Kolorektum, über den prä-, intra- und postoperativen Verlauf umfangreicher überwacht, als durch herkömmliche klinische Beobachtung, um Risiken für intraabdominelle schwere Infektionen („Deep Surgical Site Infection“, SSI-III) früher zu identifizieren. Ziel war die Entwicklung und Validierung eines Modells zur Prognose von Infektionen auf Basis longitudinaler, multimodaler Patienten- und Behandlungsdaten. Hierzu wurden während des stationären Aufenthalts ärztlich und pflegerisch erhobene Verlaufs- und Vitalparameter (u. a. Herzfrequenz, Herzratenvariabilität, Körpertemperatur sowie Aktivitätsmuster in Echtzeit) mittels sogenannter Wearables bei Patientinnen und Patienten erhoben. Anhand der erfassten Daten wurden prädiktive Modellierungen vorgenommen, um postoperative Komplikationen in Echtzeit besser abschätzen und therapeutische Maßnahmen zeitnah einzuleiten zu können. Für die Evaluation von ML-gestützten Detektions- und Vorhersagemodellen wurde eine prospektive, monozentrische, nicht-interventionelle Studie zur prädikativen Genauigkeit durchgeführt. Zu Projektbeginn erfolgten der Aufbau einer leistungsfähigen Studieninfrastruktur, die Aggregation klinischer Routinedaten aus verschiedenen Quellsystemen (u. a. dem klinischen Informationssystem), die Messung und Integration kontinuierlicher Wearable-Daten sowie deren Zusammenführung für die Modellentwicklung und Validierung. Als primärer Endpunkt wurde das Auftreten einer SSI-III im Operationsgebiet mit konsekutiver therapeutischer Maßnahme festgelegt. Nach der Bewertung der Modellpräzision anhand von area under the receiver operator curve (AUROC) und precision-recall curve (AUPRC) erfolgten abschließend die explorative Entwicklung, Simulation und Evaluation eines klinischen Alarmsystems zur Vorhersage der Zielkomplikation.

Nach Etablierung der Infrastruktur konnten für die prospektive Kohortenstudie Daten von 1.194 Patientinnen und Patienten mit elektivem, abdominellem, chirurgischem Eingriff verwendet werden. Das mediane Alter bei Aufnahme betrug 63 Jahre. 60 % der Patientinnen und Patienten waren männlich. Am häufigsten wurden hepatische Resektionen (41 %), gefolgt von kolorektalen Eingriffen (29 %) und pankreatischen Resektionen (25 %) durchgeführt. Die Prävalenz relevanter Begleiterkrankungen in der Studienpopulation war ausgeprägt: arterielle Hypertonie (51 %), metastasierte solide Tumore (42 %), chronische Lungenerkrankungen (41 %) und, maligne Grunderkrankungen (40 %) waren am häufigsten vertreten. Von den 1.194 analysierten Patientinnen und Patienten zeigten knapp zwei Drittel einen komplikationslosen postoperativen Verlauf. Bei 10 % traten ausschließlich kleinere Komplikationen auf. Mindestens eine schwerwiegende Komplikation wurde bei 22 % der Patientinnen und Patienten dokumentiert und 4 % der Fälle entsprach Komplikationen gemäß SSI-III.

Zur Entwicklung und Validierung eines ML-Modells erfolgte zunächst eine Analyse der Daten unter idealen Laborbedingungen (Modellkohorte), in denen ausschließlich Patientinnen und Patienten mit der Zielkomplikation (SSI-III), ohne anderweitige Komplikationen berücksichtigt wurden, um die ML-Modelle entsprechend zu trainieren. Nach der Entwicklung und Evaluation der Modellkohorte, wurden die ML-Modelle analog auf die Gesamtkohorte (Real-World-Kohorte) trainiert. Ziel war die Überprüfung der Modellleistungsfähigkeit hinsichtlich des primären Endpunkts auch bei potenziellen Signalconfoundern durch andere Komplikationen. Die Ergebnisse zeigen, dass das optimale ML-Modell die Zielkomplikation mit einer AUROC von 0,88 in der Modellkohorte sowie 0,88 in der Gesamtkohorte detektierte. Unmittelbar zum klinisch dokumentierten Komplikationszeitpunkt lag die Detektionsleistung bei AUROCs von 0,91 bzw. 0,88, wobei robuste Vorhersagen bereits 24-Stunden zuvor möglich waren (AUROCs 0,82 bzw. 0,78). Die primäre Hypothese konnte damit im Projekt bestätigt werden. Unter den getesteten Modellarchitekturen wiesen ‚XGBoost‘ und ‚LightGBM‘ über alle Evaluationsszenarien hinweg die höchste Vorhersageleistungen und Robustheit auf, während Deep-Learning-Modelle eine geringere Performance zeigten. Insgesamt konnte das Projekt erfolgreich die grundsätzliche Machbarkeit einer patientennahen, prädiktiven Risikostratifizierung im Versorgungsalltag zeigen und einen technologisch und methodisch neuartigen Ansatz erproben, der Neuerungen gegenüber konventionellen Early-Warning-Scores bietet.

Die Methoden waren insgesamt geeignet zur Beantwortung der Fragestellungen. Limitation ergeben sich aus der begrenzten Fallzahl sowie aus der Beschränkung auf ein Zentrum. Zudem erfolgte keine Validierung des Modells anhand weiterer/externer Daten. In der Praxis müsste sich auch erst zeigen, ob die KI-gestützten Alarmer als klinisch handlungsweisend (actionable) erweisen und die vorgesehenen Gefahren auch abgewendet werden können.

Vor dem Hintergrund der wachsenden Bedeutung von KI-Anwendungen im klinischen Setting müssten die im Projekt erprobten und evaluierten ML-Modelle insbesondere in Bezug auf externe Validität und Wirksamkeit weiterverfolgt werden. Daher werden die Ergebnisse zur Information an die o. g. Adressatinnen und Adressaten weitergeleitet.

- II. Dieser Beschluss sowie der Ergebnisbericht des Projekts *CASSANDRA* werden auf der Internetseite des Innovationsausschusses beim Gemeinsamen Bundesausschuss unter www.innovationsfonds.g-ba.de veröffentlicht.
- III. Der Innovationsausschuss beauftragt seine Geschäftsstelle mit der Weiterleitung der gewonnenen Erkenntnisse des Projekts *CASSANDRA* an die unter I. genannten Institutionen.

Berlin, den 21. August 2026

Innovationsausschuss beim Gemeinsamen Bundesausschuss
gemäß § 92b SGB V
Die Vorsitzende

Dr. Optendrenk