UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QUERSCHNITTSBEREICH 1: MEDIZINISCHE INFORMATIK

CDSS kritisch bewerten – Chancen, Risiken und ärztliche Verantwortung

Prof. Toralf Kirsten
Dr. Franziska Jahn
Dr. Birgit Schneider

imise.

MDS
Medical Data Science

INHALT

1. Evidenz zum Nutzen
2. Risiken und Fallstricke
3. Regulatorik und Recht
4. Fallbeispiel: System sagt X, Klinik sagt Y
5. Zusammenfassung und Prüfungsbezug

EINSTIEG

WÜRDEN SIE DIESEM SYSTEM VERTRAUEN?

WERBEPROSPEKT EINES HERSTELLERS

„Unsere KI erkennt Pneumonien im Röntgenbild mit 95 % Genauigkeit.“

fiktives Beispiel

Was müssten Sie wissen, bevor Sie es bei Frau M. nutzen?

An welchen Patient:innen getestet?

In welchem Krankenhaus?

Verbessert es auch die Behandlung?

Evidenz zum Nutzen

Was CDSS nachweislich verbessern – und was nicht

STUDIENLAGE

PROZESSE VERBESSERN SICH, OUTCOMES SELTEN BELEGT

METAANALYSE, 108 STUDIEN

+5,8 %

mehr Patient:innen erhalten die empfohlene Maßnahme (Median)

Kwan et al., BMJ 2020

Gut belegt

Leitlinientreue, Medikationsfehler, Dosisanpassung, Impf- und Vorsorgeerinnerungen

Unsicher

Harte Endpunkte wie Mortalität, Komplikationen, Liegedauer – selten untersucht, Effekte klein und uneinheitlich

Große Streuung

Dasselbe Prinzip wirkt in einer Klinik, in der anderen nicht

BEWERTUNG VON KI-SYSTEMEN

VIER STUFEN BIS ZUM NUTZENNACHWEIS

1

Interne Validierung

Testdaten aus derselben Klinik

2

Externe Validierung

Andere Kliniken, andere Geräte, andere Patient:innen

3

Prospektiver Einsatz

Im echten Workflow, mit Ärzt:innen

4

Randomisierte Studie

Geht es Patient:innen mit System besser?

Die meisten veröffentlichten KI-Modelle kommen nicht über Stufe 1 oder 2 hinaus.

Risiken und Fallstricke

Vier Risiken, wie ein gutes System schaden kann

RISIKO 1 · MENSCH

ALERT FATIGUE: ZU VIELE WARNUNGEN, KEINE WIRD GELESEN

ÜBERGANGENE WARNUNGEN

49–96 %

der Arzneimittelwarnungen werden übergangen

van der Sijs et al., JAMIA 2006

Ursachen

Niedriger prädiktiver Wert, klinisch irrelevante Hinweise, Unterbrechung im falschen Moment

Gegenmittel

Nach Schweregrad stufen, nur Schwerwiegendes unterbrechen, Warnregeln regelmäßig ausdünnen, Override-Raten messen

RISIKO 2 · MENSCH

AUTOMATION BIAS: DEM SYSTEM ZU SEHR VERTRAUEN

FEHLER DURCH UNTERLASSEN

Das System schweigt – also schaut niemand genauer hin.

Keine Warnung heißt nicht „kein Problem“.

FEHLER DURCH BEFOLGEN

Das System empfiehlt falsch – und man folgt.

Auch gegen den eigenen klinischen Eindruck.

DESKILLING

Wer nie selbst befundet, verlernt es.

Kritisch, wenn das System ausfällt oder irrt.

Je zuverlässiger ein System meistens ist,
desto weniger fallen seine seltenen Fehler auf.

RISIKO 3 · SYSTEM

BIAS: DAS MODELL LERNT DIE VERZERRUNG MIT

Ziel

Patient:innen mit hohem Versorgungsbedarf finden

→

Ersatzgröße im Training

Bisherige Gesundheitskosten

→

Folge

Gruppen mit schlechterem Zugang wirken gesünder, als sie sind

Quellen von Bias

Wer fehlt in den Daten? Was wurde dokumentiert? Was misst das Label wirklich?

Prüfen

Leistung getrennt nach Alter, Geschlecht, Herkunft und Versorgungsstufe berichten

RISIKO 3 · SYSTEM

ÜBERTRAGBARKEIT: DAS SEPSIS-MODELL IM PRAXISTEST

UNTERSCHEIDUNGSFÄHIGKEIT

AUC 0,63

deutlich unter den Angaben des Herstellers

ÜBERSEHEN

67 %

der Sepsisfälle ohne rechtzeitigen Alarm

ALARMIERT

18 %

aller stationären Aufenthalte

Andere Klinik, andere Patient:innen, andere Dokumentation – das Modell muss vor Ort neu geprüft werden.

Wong et al., JAMA Intern Med 2021

RISIKO 4 · SYSTEM

BLACK BOX: WARUM SAGT DAS SYSTEM DAS?

	Regelbasiert	Neuronales Netz
Begründung	Regel lesbar	Millionen Gewichte
Fehlersuche	Regel korrigieren	schwierig
Erklärhilfe	nicht nötig	Heatmap, Einflussfaktoren
Leistung bei Bildern	begrenzt	oft hoch

VORSICHT ABKÜRZUNG

Eine Heatmap zeigt, wohin das Modell schaut – nicht, ob es medizinisch richtig denkt.

Schaut es auf das Infiltrat oder auf die Beschriftung des Röntgengeräts?

Regulatorik und Recht

Medizinprodukt, KI-Verordnung und wer haftet

MEDIZINPRODUKTERECHT · MDR

DIE ZWECKBESTIMMUNG MACHT SOFTWARE ZUM MEDIZINPRODUKT

Medizinprodukt

Software, die für einzelne Patient:innen Diagnose oder Therapie unterstützen soll – etwa der Pneumonie-Assistent

Kein Medizinprodukt

Reine Verwaltung, Abrechnung, Archiv oder Nachschlagewerk ohne patientenbezogene Auswertung

REGEL 11 · RISIKOKLASSEN

IIa	Grundfall für Diagnose- und Therapiesoftware
IIb	Fehler kann schwere Verschlechterung verursachen
III	Fehler kann Tod oder irreversiblen Schaden verursachen

Ab IIa prüft eine Benannte Stelle vor der CE-Kennzeichnung.

Verordnung (EU) 2017/745, Anhang VIII

KI-VERORDNUNG · EU AI ACT

KI IM MEDIZINPRODUKT GILT ALS HOCHRISIKO-KI

PFLICHTEN HERSTELLER

Risikomanagement, Datenqualität, technische Dokumentation, Protokollierung, Transparenz, menschliche Aufsicht ermöglichen

PFLICHTEN KLINIK

Nach Gebrauchsanweisung nutzen, Aufsicht durch geschultes Personal, Betrieb überwachen, Vorfälle melden

GELTUNG FÜR MEDIZINPRODUKTE

2. Aug. 2028

nach Verschiebung durch den Digital-Omnibus

MDR und KI-VO gelten nebeneinander – geprüft wird möglichst in einem Verfahren.

HAFTUNG

DREI BETEILIGTE, EINE LETZTVERANTWORTUNG

ÄRZTIN, ARZT

Behandelt nach fachärztlichem Standard und entscheidet selbst

§ 630a BGB, Behandlungsfehler

KLINIK

Wählt geeignete Systeme, schult, organisiert Betrieb und Wartung

Organisationspflichten, MPBetreibV

HERSTELLER

Haftet für fehlerhafte Produkte – künftig ausdrücklich auch für Software und KI

Produkthaftungsrichtlinie (EU) 2024/2853

Wer der Empfehlung folgt oder sie verwirft, muss die Entscheidung medizinisch begründen und dokumentieren können.

System sagt X, die Klinik sagt Y

Ein letzter Blick auf Frau M.

FALLBEISPIEL · GERIATRIE, TAG 3, 22 UHR

FRÜHWARNSYSTEM: „NIEDRIGES RISIKO“

DAS SYSTEM

0,12

Verschlechterungsrisiko in 24 h, berechnet aus den Werten von 14 Uhr

DIE PFLEGE BERICHTET

Neu verwirrt, zupft an der Infusion

Atemfrequenz 28/min

Blutdruck 95/55 mmHg

nur als Freitext im Pflegebericht

1 · Was tun Sie? **2 · Was dokumentieren Sie?** **3 · Warum lag das System daneben?**

AUFLÖSUNG

DIE PATIENTIN GEHT VOR, DER SCORE NICHT

1	**Handeln**	Patientin sofort sehen, Vitalwerte und Labor neu erheben, eskalieren – unabhängig vom Score
2	**Dokumentieren**	Befund, Uhrzeit, Abweichung vom System und Begründung strukturiert festhalten
3	**Ursache verstehen**	Das System sah nur alte, strukturierte Werte – Verwirrtheit stand im Freitext. Fehlfunktion an die Verantwortlichen melden

Schlechte Daten rein, falsche Sicherheit raus

– zurück zu Block 1

FÜR DIE PRAXIS

SECHS FRAGEN AN JEDES CDSS

1. Wofür ist es zugelassen – und wofür nicht?
2. An welchen Patient:innen wurde es validiert?
3. Wie hoch ist der prädiktive Wert bei uns?
4. Welche Daten nutzt es – und sind sie aktuell?
5. Kann ich die Empfehlung nachvollziehen?
6. Wer überwacht die Leistung im Betrieb?

QUIZFRAGE

Ein zugelassenes CDSS empfiehlt eine Dosis, die Ihnen zu hoch erscheint. Was gilt?

A Das System ist zugelassen, also haftet der Hersteller.

B Ich prüfe selbst, entscheide und dokumentiere die Begründung.

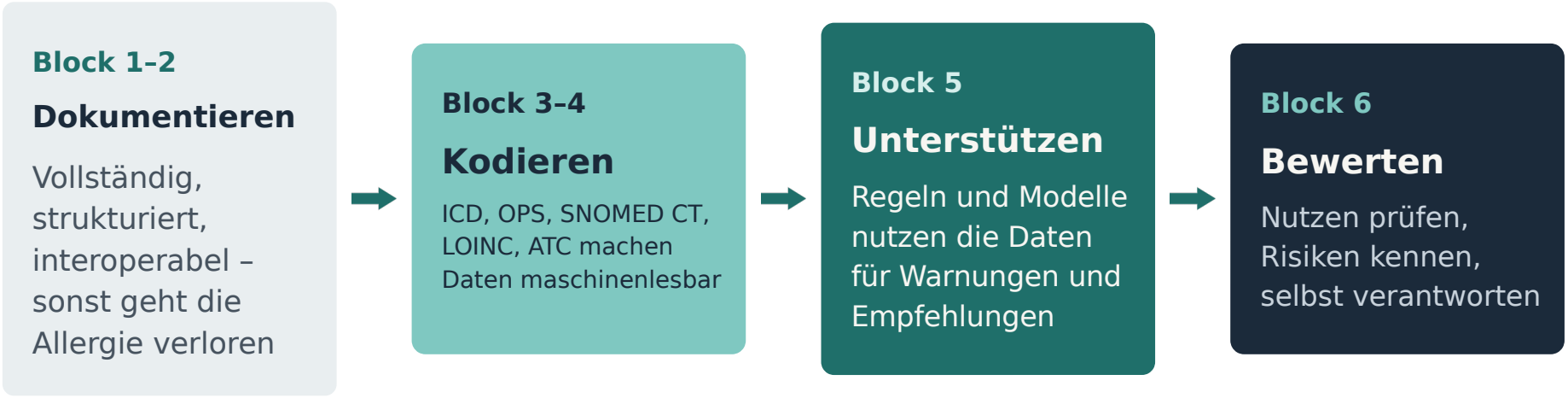
C Abweichen ist nur mit Zustimmung der Oberärztin erlaubt.

Zusammenfassung

Sechs Bereiche, eine Patientin

GESAMTZUSAMMENFASSUNG

VON DER NOTIZ ZUR ENTSCHEIDUNG



Jedes System ist nur so gut wie die Dokumentation, auf der es aufbaut.

PRÜFUNGSBEZUG

DIESE BEGRIFFE SOLLTEN SIE ERKLÄREN KÖNNEN

- Zwecke und Qualität der Dokumentation
- § 630f BGB
- Primär- und Sekundärnutzung
- FAIR
- Interoperabilität, HL7 FHIR
- Klassifikation vs. Nomenklatur
- ICD-10-GM, OPS, DRG
- SNOMED CT, LOINC, ATC
- Wissensbasiert vs. datengetrieben
- Sensitivität, Spezifität, PPV
- KI, KI-Modell, KI-System
- Alert Fatigue
- Automation Bias
- Bias und Dataset Shift
- MDR, KI-VO, ärztliche Letztverantwortung

Typische Prüfungsform: an einem kurzen Fall erklären, warum ein System versagt oder hilft.

ZUSAMMENFASSUNG IN VIER SÄTZEN

- CDSS verbessern nachweislich Prozesse; ein Nutzen für Patient:innen muss eigens belegt werden.
- Alert Fatigue und Automation Bias machen aus guten Systemen gefährliche, wenn der Mensch nicht mitdenkt.
- Modelle übernehmen Verzerrungen ihrer Daten und müssen in jeder Klinik neu geprüft werden.
- MDR und KI-VO regeln das Produkt – die Entscheidung und ihre Dokumentation bleiben ärztlich.

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

VIELEN DANK!

Prof. Toralf Kirsten
Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

Härtelstraße 16-18
tkirsten@uni-leipzig.de
www.imise.uni-leipzig.de