

Evaluation von Entscheidungsbaummodellen des maschinellen Lernens für das akute Leberversagen nach Reanimation

Evaluation of decision-tree models of machine learning for the prediction of acute liver failure after resuscitation

A. Luckscheiter¹ · W. Zink¹ · M. Thiel^{2,4} · T. Viergutz^{3,4}

► **Zitierweise:** Luckscheiter A, Zink W, Thiel M, Viergutz T: Evaluation von Entscheidungsbaummodellen des maschinellen Lernens für das akute Leberversagen nach Reanimation. *Anästh Intensivmed* 2022;63:350–361. DOI: 10.19224/ai2022.350

Zusammenfassung

Hintergrund

Maschinelles Lernen kann mithilfe digitaler Gesundheitsdaten Prognosemodelle für die frühzeitige Entstehung von Organinsuffizienzen erstellen. Ziel der Studie ist es, Entscheidungsbaummodelle für die Vorhersage des akuten Leberversagens (ALF) nach Reanimation zu evaluieren, da ein solches mit erhöhter Letalität und einer schlechteren Gesamtprognose assoziiert ist.

Patienten und Methodik

Mittels Sekundäranalyse von 347 Patienten einer belgischen Kohorte der Jahre 2007–2015, deren anonymisierter Datensatz unter der Creative Commons Attribution Licence veröffentlicht wurde, wurden ein Random Forest (RF)- sowie ein J48-Entscheidungsbaummodell erstellt und letzteres zusätzlich mittels AdaBoost verstärkt. Die Evaluation erfolgte mittels Sensitivität, Spezifität, positivem und negativem prädiktivem Wert (PPV/NPV) sowie der Fläche unter der Receiver-Operator-Kurve (AUC-ROC).

Ergebnisse

184 (53 %) Patienten wiesen ein ALF auf. Der J48-Baum teilte sich hauptsächlich an den Aufnahmewerten von International Normalized Ratio (INR, Grenzwert 1,48) und Gesamtbilirubin (Grenzwert 0,99 mg/dl) auf. Weitere wichtige Knoten waren die chronische Nieren- und Herzinsuffizienz, Adrenalinmenge während der Reanimation und Betalaktamgabe (Sensitivität 0,8, PPV

0,73, Spezifität 0,72, NPV 0,79, AUC-ROC 0,87). AdaBoost verringerte die Sensitivität auf 0,79 bei Steigerung der PPV auf 0,75 und der Spezifität auf 0,74 (NPV 0,77, AUC-ROC 0,85). RF erzielte eine Sensitivität von 0,86 (PPV 0,74) bei einer Spezifität von 0,74 (NPV 0,84) und hatte verglichen zu den anderen Modellen eine signifikant höhere AUC-ROC (0,88).

Schlussfolgerung

Maschinelles Lernen mittels Entscheidungsbaummodellen scheint für die Vorhersage eines ALF nach Reanimation geeignet. Auf Basis der verwendeten Attribute und errechneten Modelle könnten weiterführende Studien an größeren Kohorten zur Weiterentwicklung von Prognosemodellen führen und so in der frühzeitigen Diagnostik und Entscheidungsfindung unterstützen. Klinisch gesehen könnten nach einer prolongierten Reanimation initial erhöhte INR- und Bilirubinwerte, Herz- und Niereninsuffizienz das Risiko für die Entwicklung eines akuten Leberversagens erhöhen.

Summary

Background

Patients after cardiac arrest developing acute liver failure (ALF) show higher fatality rates and worse outcomes. As machine learning is able to support physicians in their decision-making with the help of big data in health records, the aim of this study is to evaluate decision-tree models for the prediction of ALF after resuscitation.

- 1 Klinik für Anästhesiologie, Operative Intensivmedizin und Notfallmedizin, Klinikum der Stadt Ludwigshafen (Direktor: Prof. Dr. W. Zink)
- 2 Klinik für Anästhesiologie und Intensivmedizin, Universitätsklinikum Mannheim (Direktorin: Prof. Dr. G. Beck)
- 3 Klinik für Anästhesie, Intensivmedizin und Schmerzmedizin, Berufsgenossenschaftliche Unfallklinik Tübingen (Chefarzt: Prof. Dr. T. Viergutz)
- 4 Medizinische Fakultät Mannheim, Universität Heidelberg

Danksagung

Die Autoren möchten Enrica Iesu sowie den Mitgliedern ihrer Arbeitsgruppe für die öffentliche Bereitstellung der Daten danken.

Interessenkonflikt

Die Autoren geben an, dass keine Interessenkonflikte bestehen.

Schlüsselwörter

Maschinelles Lernen – Entscheidungsbaume – Kardiopulmonale Wiederbelebung – Leberversagen, akut – Entscheidungsfindung – Intensivmedizin

Keywords

Maschine Learning – Decision Trees – Cardiopulmonary Resuscitation – Liver Failure, Acute – Decision-making – Critical Care

Patients and methods

The study is a secondary analysis of 347 Belgian patients after successful resuscitation from 2007 to 2015, whose anonymised dataset was published under the Creative Common Attribution Licence. For machine learning, a J48 decision-tree model as well as its AdaBoost variation and a random forest model were created. Evaluation was performed by sensitivity, specificity, positive/negative predictive value (PPV/NPV) and by the area under the receiver operator curve (AUC-ROC).

Results

184 patients (53 %) developed ALF after resuscitation. The main split points in J48 were the attributes international normalised ratio ad admission (INR, threshold 1.48) and total bilirubin ad admission (threshold 0.99 mg/dl). Further splitting was performed in chronic renal and heart failure, administered amount of epinephrine during resuscitation, and the use of beta-lactam antibiotics. J48

yielded a sensitivity of 0.8 and a specificity of 0.72 (0.73 PPV/0.79 NPV, AUC-ROC 0.83). AdaBoost yielded a comparable sensitivity (0.79) and a higher PPV (0.75) and specificity (0.74, NPV 0.77, AUC-ROC 0.85). Random forest reached a sensitivity of 0.86 (PPV 0.74) and a specificity of 0.74 (NPV 0.84, AUC-ROC 0.87). The random forest achieved a significantly higher test quality in AUC-ROC than the other models.

Conclusion

Machine learning with decision-tree models seems to be suitable for the prediction of ALF after resuscitation. The clinical risk of developing ALF could be increased in conditions with elevated INR and total bilirubin, chronic renal and heart failure as well as prolonged resuscitation. Based on the applied attributes and developed models, further studies in large cohorts could lead to an advancement of the prediction models and therefore support physicians in diagnosis and decision-making.

Einleitung

Die zunehmende Verbreitung elektronischer Patientenakten und digitaler Dokumentationssysteme hat dazu geführt, dass große medizinische Datenmengen im Gesundheitswesen entstanden sind. Zu deren Analyse werden vermehrt Methoden des maschinellen Lernens verwendet, um neue Muster und Gesetzmäßigkeiten zur frühzeitigen Diagnose bestimmter Krankheiten und deren Prognoseabschätzungen abzuleiten und so Modelle zu erstellen, welche etablierte Systeme verbessern oder erweitern [1]. Hierzu werden Trainingsdatensätze mit Attributen, welche eine thematische Korrelation bzw. Kausalität aufweisen, selektiert und daraus Modelle errechnet. Anschließend werden diese an zuvor nicht verwendeten Testdatensätzen evaluiert [2]. Neben klassischen Verfahren wie der logistischen Regressionsanalyse sind neben künstlichen neuronalen Netzwerken und Stützvektorverfahren vor allem Entscheidungsbaummodelle

etabliert. Diese bieten den Vorteil, nicht nur anschauliche Entscheidungswege aufzuzeigen, sondern sich auch dank ihrer schnellen Rechenoperationen gut in elektronische Qualitäts- und Diagnostikprozesse integrieren zu lassen [3].

Ein Themengebiet, das aufgrund seiner Verbindung zwischen standardisierter, elektronischer Datenerfassung einerseits und klinischer Relevanz andererseits sehr im Fokus des maschinellen Lernens liegt, ist die Reanimationsforschung. So existieren bereits in der Reanimatologie Modelle für die Wahrscheinlichkeit eines wiedereinsetzenden Spontankreislaufs (Return Of Spontaneous Circulation, ROSC), die Vorhersage einer Reanimationspflichtigkeit in der Sepsis oder das neurologische Outcome [4–7]. Für letztgenanntes spielen bei der Reanimation nicht nur frühzeitige, qualitativ hochwertige Reanimations- und Oxygenierungsmaßnahmen und die anschließende Hypothermievermeidung, sondern auch die Insuffizienz anderer Organsysteme eine Rolle [8]. Neben pulmonalen und kardialen Komplikationen (z. B. Lungenversagen, Rhythmusstörungen, Herzinsuffizienz oder Nierenversagen) rücken vermehrt die Auswirkungen des Leberversagens auf andere Organsysteme in den wissenschaftlichen Fokus. Roedl et al. zeigten, dass Patienten mit Leberzirrhose eine 3-fach erhöhte Einjahressterblichkeit nach erfolgreicher Reanimation respektive ein häufig deutlich schlechteres Outcome aufweisen [9]. Jedoch können auch zuvor vermeintlich lebergesunde Patienten nach erfolgreicher Reanimation von einem Leberversagen betroffen sein. Es wird hier zwischen einem häufigeren akuten Leberversagen (Acute Liver Failure, ALF, definiert als Bilirubin $>1,2$ mg/dl, einer International Normalized Ratio (INR) $>1,5$ ohne chronischen Leberschaden sowie Zeichen der hepatischen Enzephalopathie) und einer selteneren hypoxischen Hepatitis (definiert als ein 20-facher Anstieg der Aminotransferasen) unterschieden [10–12]. Für beide Formen konnten Iesu et al. eine signifikant erhöhte Letalität sowie für die hypoxische Hepatitis zudem ein schlechteres neurologisches Outcome nach Reanimation zeigen [11]. Für das

akute Leberversagen, welches im Median drei Tage nach Reanimation auftrat, ermittelten sie mittels logistischer Regression die Parameter Schock, vorbestehende Herzinsuffizienz, erhöhtes Laktat bei Aufnahme, niedrige zentral-/gemischtvenöse Sättigung während des Intensivaufenthalts sowie Dialysenotwendigkeit als Risikofaktoren.

Die frühzeitige Vorhersage einer Organinsuffizienz nach Reanimation würde für Intensivmediziner die Chance eröffnen, entweder sich besser auf diese vorzubereiten oder sogar präventiv zu handeln. Die vorliegende Studie soll daher evaluieren, ob das maschinelle Lernen am Beispiel der Vorhersage des akuten Leberversagens nach Reanimation dazu geeignet ist, hierfür klinisch plausible und aussagekräftige Entscheidungsbaummodelle zu entwickeln.

Methodik

Die vorliegende Studie ist eine Sekundäranalyse einer Studie von Iesu et al, die seinerzeit von der lokalen Ethikkommission genehmigt wurde (Comité d’Ethique Hospitalo-Facultaire Erasme-ULB, P2017/264, Brüssel, Belgien) [11]. Als Grundlage dienen hierzu die Daten von über 340 reanimierten Patienten, die unter der Creative Common Attribution Licence für Forschungszwecke frei zugänglich veröffentlicht wurden [11].

Datengrundlage

Iesu et al. führten eine monozentrische, retrospektive Studie im Erasmus Hospital Brüssel durch, in die inner- und präklinisch reanimierte Patienten der Jahre 2007–2015 eingeschlossen wurden. Ziel der Studie war die erstmalige Abschätzung der Inzidenz des akuten Leberversagens bzw. der hypoxischen Hepatitis nach Reanimation, deren Auswirkung auf die Gesamtprognose und auf das neurologische Outcome sowie die Ermittlung von Risikofaktoren. In der Originaluntersuchung wurden 61 von 435 Patienten wegen fehlender Leberwertbestimmungen bei Aufnahme bzw. Tod in weniger als 24 Stunden nach Übernahme auf die Intensivstation ausgeschlossen, sodass 374 in die Studie auf-

genommen wurden. Ursprünglich wurde ein ALF aufgrund der Beatmungstherapie respektive der möglichen späteren hypoxischen Enzephalopathie lediglich definiert als Bilirubin $>1,2$ mg/dl sowie einem INR $>1,5$ ohne chronischen Leberschaden. Die hepatische Enzephalopathie wurde somit nicht berücksichtigt. Die in der Studie bereits durchgeführte Klassifizierung in eine Gruppe mit bzw. ohne ALF wurde übernommen. Erfasst wurden damals demografische Attribute (Alter, Geschlecht), Charakteristika der Reanimation (u. a. beobachtet, By-stander-CPR, Zeit bis ROSC, Adrenalinmenge, Defibrillation, Hypothermietherapie), Komorbiditäten, Charakteristika des Intensivaufenthalts (u. a. Dialyse, Vasopressortherapie, Beatmung, hepatotoxische Medikamente wie Paracetamol oder Betalaktamantibiotika ohne weitere Unterdifferenzierung) sowie initiale Blutgas- und Laborwerte (u. a. Leberwerte, Gerinnung, Kreatinin sowie Thrombozyten), SOFA (Sequential Organ Failure Assessment)- und APACHE II-Score (Acute Physiology and Chronic Health Evaluation) bei Aufnahme. Iesu et al. ermittelten hieraus mittels logistischer Regression zuvor genannte Risikofaktoren für ein ALF. Für weitere Einzelheiten zur Intensivtherapie und Datenerfassung sei auf die Studie von Iesu et al. verwiesen [11].

In vorliegender Studie wurden die öffentlich zugänglichen Daten zunächst entsprechend den Formatvorgaben für die Analysesoftware WEKA 3.8.4 (Waikato Environment for Knowledge Analysis, University of Waikato, Neu Zeeland) aufbereitet. Wegen eines fehlenden Kausalzusammenhangs mit einem akuten Leberversagen wurden dann die Attribute Intensiv- bzw. Krankenhaussterblichkeit sowie das neurologische Outcome für die Modellberechnung ausgeschlossen. Der niedrigste zentralvenöse Sättigungswert und die niedrigste Thrombozytenzahl wurden ebenfalls für die Berechnung nicht verwendet, da das Modell nur aus Aufnahmelaborwerten bestehen sollte. Zusätzlich wurde der DeRitis-Quotient sowie der MELD-Score (Model of End Stage Liver Disease) aus den Aufnahmelaborwerten berechnet

(Tab. 1 und 2). Der MELD-Score wurde zwar ursprünglich für die Abschätzung der Letalität von Patienten im Endstadium einer Leberzirrhose validiert. Wegen seiner guten Korrelation zu klinischen Endpunkten und zur Sterblichkeit beispielsweise beim fulminanten Leberversagen wurde er ebenso zur Modellierung herangezogen [13].

Entscheidungsbaummodelle

Es wurden drei verschiedene Entscheidungsbaummodelle erstellt, welche mit allen Skalenniveaus sowie mit einzelnen fehlenden Daten umgehen können. Als erstes Entscheidungsbaummodell wurde der C4-Algorithmus in der Version J48 gewählt. Der C4-Algorithmus erstellt einen Baum, indem er eine Aufteilung vom Attribut mit dem höchstmöglichen Informationsgewinn (d. h. hier mit der größten Unterschiedlichkeit zwischen zwei Klassen) vornimmt [14]. Die anschließenden Aufteilungen an den folgenden Knotenpunkten werden unter demselben Prinzip unter Verwendung der restlichen Attribute vorgenommen, bis der gesamte Datensatz aufgeteilt ist und eine zuvor definierte minimale Fallzahl pro Blatt repräsentiert wird. Dies führt dazu, dass Attribute in verschiedenen Zweigen wiederholt vorkommen können, allerdings auch nicht alle Attribute verwendet werden müssen. Damit solch ein Baum nicht durch maximale Aufteilung auf eine unüberschaubare Größe anwächst und damit eine Überanpassung (overfitting) einhergeht, muss er getrimmt (pruning) werden. Dies wurde in vorliegender Studie durch eine minimale Fallanzahl pro Blatt von 8 sowie durch Unterbaumzucht (subtree raising) sichergestellt. Dabei werden innere Knoten getrimmt und der Unterbaum darunter um eine Ebene zur Wurzel hin angehoben (weitere Parameter: Konfidenzfaktor 0,25, MDL-Korrektur für numerische Attribute, LaPlace-Korrektur). Ziel war es, so einen Baum mit möglichst hoher klinischer Nachvollziehbarkeit zu entwickeln.

J48 zählt zu den sogenannten schwachen Lernern, d. h. seine Baumstruktur ändert sich erheblich bei nur geringfügigen Veränderungen in der Daten-

struktur. Seine Leistung und Robustheit können daher oft durch das sogenannte Boosting-Verfahren gesteigert werden. Als zweiter Ansatz wurde daher das J48-Baummodell mit den äquivalenten Parametern mittels dem AdaBoost-Algorithmus geboostet [15]. Hierbei wird pro Durchlauf ausgehend von einem zufälligen Merkmal ein J48-Baum und dessen Testgüte berechnet. Anschließend gewichtet AdaBoost entsprechend der Testgüte des zuvor berechneten Baums die korrekt und inkorrekt klassifizierten Attribute für den nächsten Durchlauf, um diese erneut zu klassifizieren (hier: 10 Durchläufe, keine Stichprobenwiederholung, sondern Wiedergewichtung). Die Modelle der jeweiligen Durchläufe werden schlussendlich zu dem geboosteten Modell zusammengeführt (ensemble learning) [15].

Als drittes Verfahren soll ein Random Forest-Modell (RF) evaluiert werden, welches zunächst eine bestimmte Anzahl an Zufallsbäumen erstellt (hier: 200) [3]. Die mehrheitliche Entscheidung der Bäume wird dann als Vorhersage ausgegeben (bagging, weitere Parameter Batch-Größe 100, keine zufälligen Brüche bei gleichgewichteten Attributen, unlimitierte Baumgröße). Zudem können alle Attribute zur Baumerstellung verwendet werden. Dafür sind die Entscheidungswege des RF wie auch des AdaBoost-Modells schlechter interpretierbar. So können beide Modelle im Gegensatz zum einzelnen J48-Modell keinen einzelnen finalen Entscheidungsbaum visualisieren. Durch das Bagging- bzw. Boosting-Verfahren wird jedoch im Allgemeinen eine erhöhte Robustheit gegenüber Veränderungen in der Datenstruktur erreicht und der Aussagekraft so mehr Stabilität verliehen [15].

Modellevaluation

Alle Modelle durchliefen zehnmal eine zehnfache Kreuzvalidierung. Bei der zehnfachen Kreuzvalidierung wird der Datensatz in zehn zufällige und gleichgroße Abschnitte unterteilt. Neun werden dann zum Training und der zehnte zur Testung herangezogen. Das Verfahren wird so oft wiederholt, bis jeder Abschnitt genau einmal als Testabschnitt

herangezogen wurde. Hieraus werden dann die Gütekriterien Sensitivität und Spezifität, positiver und negativer prädiktiver Wert (PPV, NPV), die Fläche unter der Grenzwertoptimierungskurve (Receiver Operating Characteristic, AUC-ROC) sowie die Fläche unter der Precision-Recall-Kurve (AUC-PRC, Sensitivität-PPV-Kurve) gemittelt. Ebenso wurden die Durchschnittswerte von Kappa als Maß für die Reliabilität und Matthews Korrelationskoeffizient (MCC) als Maß für die Güte der binären Klassifikation ermittelt [2,16]. Die Testung auf Signifikanz erfolgte mittels Chi-Quadrat- oder t-Test ($p < 0,05$ als signifikant) in Microsoft Excel 2016 (Microsoft Corporation, Redmond, Seattle, USA).

Ergebnisse

Datensatz, Reanimations- und Intensivverlauf

Von 374 in der Originalstudie aufgeführten Patienten waren Datensätze verfügbar (92,8 %). 184 Patienten (53 %) erlitten seinerzeit ein ALF, während dieser Verlauf in der Kontrollgruppe ($n = 163$, 47 %) nicht beobachtet wurde. Bei 21 von 184 Patienten mit ALF (11 %) lag bereits initial eine pathologische Laborkonstellation gemäß der oben aufgeführten ALF-Definition vor. Die restlichen Patienten entwickelten das ALF erst im Verlauf (Kontrollgruppe 0 % bei Aufnahme). Von den in der Originalstudie aufgelisteten Attributen waren keine Daten zu HIV-Infektionen sowie zur hypoxischen Hepatitis im öffentlichen Datensatz verfügbar. Patienten mit ALF hatten verglichen zur Kontrollgruppe eine signifikant längere Zeit bis ROSC (19,7 vs. 15,39 min), bekamen mehr Adrenalin appliziert (4,73 vs. 3,13 mg) und hatten häufiger einen präklinischen Kreislaufstillstand (47 % vs. 64 %), eine vorbestehende Herzinsuffizienz (28 % vs. 15 %), vorbestehende Niereninsuffizienz (22 % vs. 9 %) und Leberzirrhose (7 % vs. 2 %). Auf der Intensivstation musste die ALF-Gruppe zudem häufiger im Schockgeschehen (61 % vs. 39 %) mit Vasopressoren (79 % vs. 68 %, Dobutamin 60 % vs. 40 %), extrakorpora-

len Membranoxygenierungsverfahren (ECMO, 15 % vs. 6 %) sowie einer Dialyse (18 % vs. 9 %, alle $p < 0,05$) therapiert werden. An hepatotoxischen Medikamenten wurde ihnen Paracetamol (41 % vs. 74 %) zwar signifikant seltener, dafür aber Betalaktamantibiotika (54 % vs. 34 %) signifikant vermehrt verabreicht. Bezüglich der Laborwerte bei

Aufnahme fanden sich in der Gruppe mit akutem Leberversagen zudem signifikant niedrigere Thrombozytenwerte (125 vs. 174 /nl), erhöhte CRP-Spiegel (68,99 vs. 52,33 mg/dl) sowie Laktatwerte (6,73 vs. 5,13 mmol/l). Bezüglich der initialen Leberwerte waren sowohl GOT (147 vs. 117 IU/l), INR (1,89 vs. 1,18) als auch die Cholestasewerte (AP 96 vs. 86 IU/l,

GGT 109 vs. 88 IU/l) signifikant erhöht in der ALF-Gruppe. Zudem wies diese Gruppe bei Aufnahme auch einen signifikant niedrigeren pH (7,28 vs. 7,31) und mittleren arteriellen Druck (MAP, 87 vs. 95 mmHg) und einen höheren MELD-Score (16 vs. 10) sowie DeRitis-Quotienten (1,6 vs. 1,4) auf (Tab. 1 und 2).

Tabelle 1

Demografische Daten der Kohorte zu Vorerkrankungen sowie zur Reanimation und Intensivtherapie.

	akutes Leberversagen (n = 184)	kein Leberversagen (n = 163)	p-Wert	Attributgewichtung RF
Alter (a)	61,93 (62,5) ± 15,7	62,31 (62) ± 15,17	0,82	0,48 ± 0,02
Gewicht (kg)	75,92 (75) ± 14,01	78,45 (78) ± 15,35	0,11	0,47 ± 0,02
Anteil Männer	75%	71%	0,35	0,53 ± 0,01
Dauer Intensivaufenthalt	8,8 (5) ± 10,98	7,12 (4) ± 8,12	0,10	0,46 ± 0,02
beobachteter Kreislaufstillstand	87 %	85 %	0,54	0,47 ± 0,03
Bystander CPR	72 %	64 %	0,11	0,48 ± 0,03
Zeit bis ROSC (min)	19,7 (15) ± 16,33	15,39 (14) ± 10,81	<0,01*	0,43 ± 0,01
Adrenalinosis (mg)	4,73 (3) ± 4,21	3,13 (3) ± 2,68	<0,01*	0,41 ± 0,01
präklinischer Kreislaufstillstand	47 %	64 %	<0,01*	0,49 ± 0,04
Hypothermietherapie	89 %	88 %	0,82	0,45 ± 0,04
nicht-kardiologische Ätiologie	55 %	64 %	0,09	0,46 ± 0,03
nicht defibrillierbarer Rhythmus	37 %	48 %	0,03*	0,5 ± 0,03
Tod auf Intensivstation	52 %	46 %	0,30	n. b.
Tod im Krankenhausaufenthalt	58 %	50 %	0,14	n. b.
CPC 1–2 nach 3 Monaten	40 %	45 %	0,34	n. b.
Kortikoidtherapie	24 %	18 %	0,16	0,47 ± 0,03
dauerhafte Antikoagulation	18 %	18 %	0,97	0,44 ± 0,04
Herzinsuffizienz	28 %	15 %	<0,01*	0,41 ± 0,05
arterielle Hypertonie	41 %	45 %	0,38	0,5 ± 0,03
KHK	41 %	39 %	0,61	0,5 ± 0,05
Diabetes mellitus	22 %	27 %	0,31	0,47 ± 0,04
COPD/Asthma	16 %	17 %	0,95	0,45 ± 0,04
neurologische Vorerkrankung	13 %	17 %	0,28	0,46 ± 0,05
chronische Niereninsuffizienz	22 %	9 %	<0,01*	0,32 ± 0,04
Leberzirrhose	7 %	2 %	0,02*	0,3 ± 0,08
IABP	8 %	4 %	0,14	0,3 ± 0,04
ECMO	15 %	6 %	<0,01*	0,34 ± 0,06
Schock	61 %	39 %	<0,01*	0,42 ± 0,03
Vasopressortherapie	79 %	68 %	0,02*	0,43 ± 0,06
Dobutamin	60 %	40 %	<0,01*	0,43 ± 0,04
invasive Beatmungstherapie	99 %	98 %	0,56	0,32 ± 0,07
Dialyse	18 %	9 %	0,01*	0,34 ± 0,06

CPR: kardiopulmonale Reanimation; **ROSC:** wiederhergestellter Spontankreislauf; **CPC 1–2:** sehr gutes bis gutes neurologisches Outcome; **COPD:** chronisch obstruktive Lungenerkrankung; **KHK:** koronare Herzerkrankung; **IABP:** intraaortale Ballonpumpe; **ECMO:** extrakorporale Membranoxygenierung; **RF:** Random Forest; * signifikant $p < 0,05$; **n. b.:** nicht bestimmbar; Daten gemäß der Creative Commons Attribution Licence frei abdruckbar.

Tabelle 2

Hepatotoxische Medikamente, Laborbefunde bei Aufnahme und Scores.

	akutes Leberversagen (n = 184)	kein Leberversagen (n = 163)	p-Wert	Attributgewichtung RF
mind. 1 hepatotox. Medikament	90 %	89 %	0,83	0,38 ± 0,05
Anzahl hepatotox. Med.	1,49 (1) ± 0,84	1,67 (2) ± 0,95	0,06	0,37 ± 0,03
Paracetamol	41 %	74 %	<0,01*	0,27 ± 0,02
Amiodaron	52 %	53 %	0,91	0,45 ± 0,04
Betalaktame	54 %	34 %	<0,01*	0,31 ± 0,02
Fluorchinolone	1 %	2 %	0,33	0,31 ± 0,06
Azole	0 %	4 %	0,01*	0,26 ± 0,1
INH	0 %	0 %	n. b.	0
Sulfonamide	0 %	0 %	n. b.	0
Metronidazol	1 %	1 %	0,93	0,26 ± 0,17
Chemotherapie	0 %	0 %	n. b.	0
akutes Nierenversagen	61 %	52 %	0,06	0,42 ± 0,04
Laktat (mmol/l)	6,73 (5,6) ± 3,42	5,13 (4,6) ± 2	<0,01*	0,34 ± 0,01
CRP (mg/dl)	68,99 (44) ± 77,13	52,33 (32) ± 62,43	0,03*	0,37 ± 0,01
Kreatinin (mg/dl)	1,55 (1,2) ± 1,07	1,39 (1,1) ± 1,25	0,19	0,35 ± 0,03
ScvO ₂ bzw. SvO ₂ (%)	69,17 (68,7) ± 8,91	69,69 (69,1) ± 8,95	0,59	0,36 ± 0,02
GOT (IU/l)	147,13 (95) ± 157,5	117,08 (78) ± 114,33	0,04*	0,33 ± 0,02
GPT (IU/l)	112,21 (59,5) ± 140,58	93,04 (63) ± 92,69	0,13	0,35 ± 0,01
LDH (IU/l)	415,15 (328,5) ± 301,92	353,41 (322) ± 166,16	0,02*	0,34 ± 0,01
AP (IU/l)	96,82 (76) ± 79,83	86,89 (75) ± 46,65	0,15	0,33 ± 0,02
GGT (IU/l)	109,69 (66) ± 162,36	88,56 (71) ± 91,09	0,13	0,34 ± 0,02
Gesamtbilirubin (mg/dl)	1,34 (0,7) ± 2,42	0,47 (0,4) ± 0,23	<0,01*	0,31 ± 0,01
aPTT (sec)	50,75 (37,1) ± 33,57	32 (29,2) ± 12,94	<0,01*	0,3 ± 0,01
Prothrombinzeit (%)	54,56 (52,5) ± 23,2	75,48 (74) ± 15,72	<0,01*	0,31 ± 0,01
INR	1,89 (1,5) ± 1,52	1,18 (1,2) ± 0,14	<0,01*	0,31 ± 0,01
Thrombozyten (/nl)	185,69 (170,5) ± 104,45	242,84 (220) ± 105,39	<0,01*	0,28 ± 0,02
Gesamtprotein (mg/dl)	6,43 (5,6) ± 6,49	8,25 (5,8) ± 11,13	0,07	0,3 ± 0,02
Glukose (mg/dl)	232,55 (197) ± 122,26	234,75 (209) ± 104,7	0,86	0,29 ± 0,01
pH	7,28 (7,3) ± 0,15	7,31 (7,3) ± 0,11	0,02*	0,29 ± 0,01
pO ₂ (mmHg)	161,41 (119,5) ± 113,29	140,24 (106) ± 96,08	0,06	0,3 ± 0,02
pCO ₂ (mmHg)	38,89 (37) ± 11,44	40,25 (38) ± 10,33	0,25	0,29 ± 0,02
MAP (mmHg)	87,93 (84) ± 21,47	95,15 (94,5) ± 20,64	<0,01*	0,28 ± 0,02
MELD	16,07 (14,6) ± 7,41	10,41 (9) ± 4,7	<0,01*	0,25 ± 0,01
DeRitis	1,64 (1,4) ± 1,04	1,43 (1,3) ± 0,73	0,03*	0,28 ± 0,02
APACHE II	24,38 (25) ± 7,32	23,93 (24) ± 6,66	0,55	0,28 ± 0,01
SOFA	11,26 (11) ± 3,62	10,91 (11) ± 3,33	0,35	0,27 ± 0,02
nicht verwendete Laborwerte aus dem Krankenhausaufenthalt				
höchster INR	2,25 (1,2) ± 3,25	0,64 (0,61) ± 0,23	<0,01*	n. b.
höchstes Gesamtbilirubin (mg/dl)	2,65 (1,72) ± 3,02	1,26 (1,25) ± 0,13	<0,01*	n. b.
niedrigste ScvO ₂	60,13 (62) ± 8,63	62,38 (63) ± 8,19	0,01*	n. b.
niedrigste Thrombozyten (/nl)	125,52 (106,5) ± 85,5	174,33 (161) ± 86,58	<0,01*	n. b.

INH: Isoniazid; ScvO₂/SvO₂: zentralvenöse bzw. gemischtvenöse Sättigung; CRP: C-reaktives Protein; GOT: Glutamat-Oxalacetat-Transaminase; GPT: Glutamat-Pyruvat-Transaminase; LDH: Laktatdehydrogenase; AP: alkalische Phosphatase; GGT: Gamma-Glutamin-Transferase; aPTT: aktivierte partielle Thromboplastinzeit; INR: International Normalized Ratio; MAP: mittlerer arterieller Druck; MELD: Model of End Stage Liver Disease; SOFA: Sequential Organ Failure Assessment; APACHE: Acute Physiology and Chronic Health Evaluation; RF: Random Forest; * signifikant p < 0,05; n. b.: nicht bestimmbar; Daten gemäß der Creative Common Attribution Licence frei abdruckbar.

Entscheidungsmodelle

Der J48-Entscheidungsbaum sieht den höchsten Informationsgewinn für ein akutes Leberversagen nach Reanimation bei einem INR nach Aufnahme von größer 1,48. Kleiner-gleich einem INR von 1,48 votiert der Baum für ein ALF bei einem initialen Gesamtbilirubin $>0,99$ mg/dl. Neben der Gabe von Paracetamol bei chronischer Niereninsuffizienz mit erniedrigtem Gesamtprotein $\leq 5,7$ g/dl scheint die Gabe von Betalaktamantibiotika bei verlängerter aPTT ($>25,8$ s) sowie erhöhtem Kreatinin ($>1,1$ mg/dl) mit einem erhöhten Risiko für das Entstehen eines akuten Leberversagens assoziiert zu sein. Umgekehrt scheint Dobutamin bei normaler Nierenfunktion und Betalaktamgabe keinen negativen Einfluss auf die Leberfunktion zu haben. Bei Verzicht auf Paracetamol führen hohe Dosen von Adrenalin (>7 mg) während der Reanimation vermehrt zu einem akuten Leberversagen. Werden weniger

als 7 mg verabreicht, scheint jedoch eine vorbestehende Herzinsuffizienz sich negativ auszuwirken. Bei herzgesunden Patienten ist die Betalaktamgabe bei Laktatwerten über 4,2 mmol/l mit einem akuten Leberversagen assoziiert. Werden keine Betalaktame verabreicht, führen erhöhte Cholestaseparameter ($AP > 73$ IU/l) und ein Gesamtprotein $\leq 5,5$ g/dl zur Klassifizierung als ein akutes Leberversagen (Abb. 1). Der J48-Baum erzielte eine Sensitivität von 0,8 sowie einen PPV von 0,73 (AUC-PRC 0,87). Die AUC-ROC lag bei 0,83 (Tab. 3 sowie Abb. 2 und 3). Mittels AdaBoost verminderte sich die Sensitivität auf 0,79 ($p > 0,05$) bei einer geringgradigen Steigerung der PPV auf 0,75 ($p = 0,01$, AUC-PRC 0,88, $p > 0,05$). Die AUC-ROC war mit 0,85 nicht signifikant höher ($p > 0,05$).

Das RF-Modell erzielte eine Sensitivität von 0,86 bei einem PPV von 0,74 (AUC-ROC 0,87, AUC-PRC 0,91, alle $p < 0,05$ zu allen anderen Modellen mit

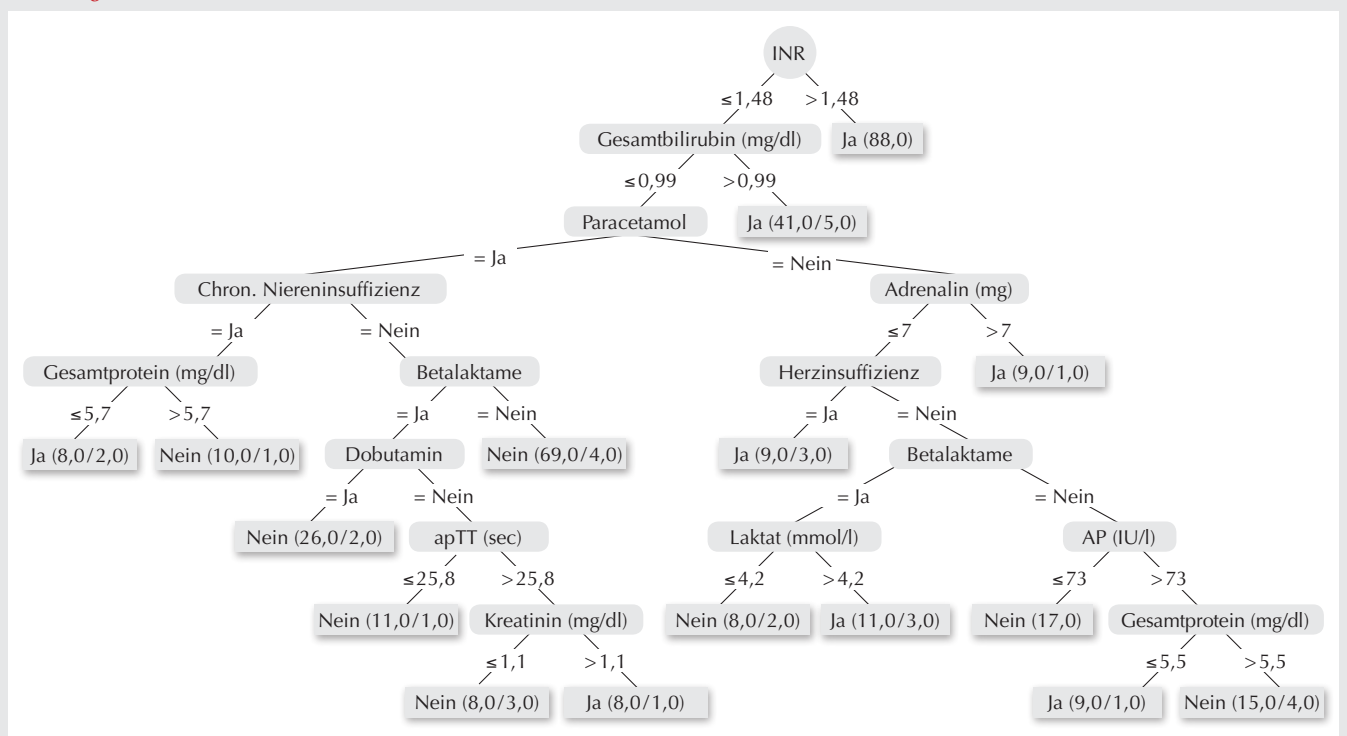
Ausnahme der höheren PPV bei AdaBoost 0,75, $p = 0,02$). Die Spezifität von 0,74 war verglichen zu J48 signifikant höher (0,72, $p < 0,05$). Die Unterschiede im NPV waren mit 0,84 (RF) nicht signifikant verglichen zu J48 (0,79) und AdaBoost (0,77, alle $p > 0,05$). Verglichen zu AdaBoost war die Spezifität bei J48 leicht verringert (0,72 bzw. 0,74, $p > 0,05$), der NPV aber größer (0,77 vs. 0,79, $p > 0,05$) (Tab. 3).

Die größten Gewichtungen im Random Forest ergaben sich bei den Attributen männliches Geschlecht (0,53), arterielle Hypertonie, koronare Herzkrankheit sowie nicht defibrillierbarer Rhythmus (alle 0,5; Tab. 1 und 2).

Diskussion

In der vorliegenden Studie sollte untersucht werden, inwieweit die Anwendbarkeit von Entscheidungsbaummodellen aus dem Gebiet des maschinellen

Abbildung 1



J48-Entscheidungsbaum für ein akutes Leberversagen (= Ja) oder dagegen (= Nein). Alle Laborwerte wurden bei Aufnahme erfasst. Die erste Zahl am Blatt ist die Anzahl der Fälle, die das Blatt erreichen, die zweite Zahl die davon falsch berechneten.

INR: International Normalized Ratio; **aPTT:** aktivierte partielle Thromboplastinzeit; **AP:** alkalische Phosphatase.

Lernens zur Vorhersage des akuten Leberversagens nach Reanimation gegeben sein könnte. Auf Grundlage einer Sekundäranalyse des Reanimationsablaufs, der Intensivtherapie, initialen Laborwerten und Komorbiditäten einer belgischen Kohorte wurden drei verschiedene Entscheidungsbaummodelle erstellt. Diese hoben als klinische Risikofaktoren INR, Gesamtbilirubin, chronische Nieren- und Herzinsuffizienz, Adrenalinmenge während der Reanimation, nicht defibrillierbare Rhythmen sowie männliches Geschlecht hervor. Es zeigte sich im Rahmen der Kreuzvalidierung innerhalb der kleinen Kohorte eine robuste Vorhersagegenauigkeit über alle Modelle hinweg, wobei Random Forest die höchste Testdiskrimination und -genauigkeit aufwies. Die Ergebnisse vorliegender Studie legen nahe, dass unter Verwendung von Entscheidungsbaummodellen die Vorhersagbarkeit eines ALF nach Reanimation gegeben sein könnte. Soweit den Autoren bekannt, stellt diese Studie die erste Untersuchung des akuten Leberversagens nach Reanimation mit Verfahren des maschinellen Lernens dar.

Die Prognose von Patienten mit hepatischer Insuffizienz bei intensivpflichtigen Erkrankungen ist schlecht. So ermittelten Jentzer et al. für das Leberversagen auf Intensivstation eine bis zu 6,7-fach erhöhte Krankenhaussterblichkeit [17]. Møller und Bernardi beschrieben in ihrem Review zur Herz-Leber-Interaktion einen Circulus vitiosus zwischen unzureichender hepatischer Durchblutung einerseits und einem passiven Blutstau in der Leber andererseits. Letzteres wird durch ein rechtsventrikuläres Versagen mitverursacht und kann sich u. a. im Anstieg der Cholestaseparameter bemerkbar machen. In der Folge kommt es zu einem verminderten systemisch-vaskulären Widerstand und einer gestörten systolischen und diastolischen Herzfunktion mit vermindertem adrenergen Ansprechen sowie zu QT-Verlängerungen durch eine Hyperbilirubinämie [18]. Berücksichtigt man nun noch die gestörte toxische Clearance, erklärt sich so die erhöhte Letalität und das schlechtere neurologische Outcome nach Reanimation [11].

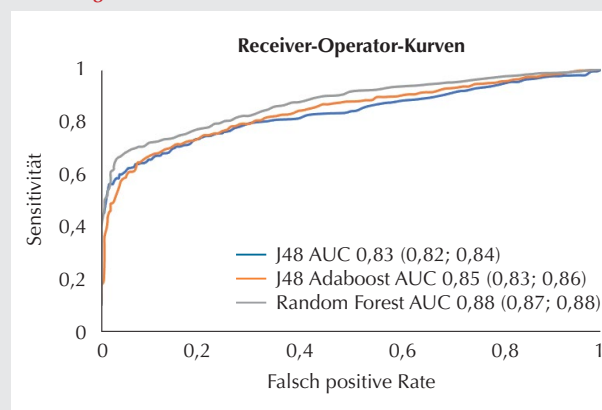
Tabelle 3

Testcharakteristika der drei Modelle (Angabe in Mittelwert, Standardabweichung sowie 95 %-Konfidenzintervall in Klammern).

Testkriterium	J48	Vs. p-Wert	J48 AdaBoost	Vs. p-Wert	Random Forest	Vs. J48 P-Wert
% korrekt	75,99 ± 1,22 (75,24; 76,75)	0,61	76,4 ± 1,98 (75,17; 77,62)	<0,01*	79,54 ± 0,91 (78,97; 80,1)	<0,01*
Kappa	0,52 ± 0,03 (0,51; 0,54)	0,68	0,53 ± 0,04 (0,5; 0,55)	<0,01*	0,59 ± 0,02 (0,58; 0,6)	<0,01*
AUC-ROC	0,83 ± 0,012 (0,82; 0,84)	0,06	0,85 ± 0,022 (0,83; 0,86)	<0,01*	0,87 ± 0,002 (0,87; 0,88)	<0,01*
MCC	0,52 ± 0,03 (0,51; 0,54)	0,80	0,53 ± 0,04 (0,5; 0,55)	<0,01*	0,6 ± 0,02 (0,59; 0,61)	<0,01*
Schwelle	0,58 ± 0,027 (0,56; 0,6)	<0,01*	0,51 ± 0,004 (0,5; 0,51)	<0,01*	0,5 ± 0 (0,5; 0,5)	<0,01*
Sensitivität	0,8 ± 0,03 (0,79; 0,82)	0,27	0,79 ± 0,02 (0,78; 0,8)	<0,01*	0,86 ± 0,01 (0,85; 0,87)	<0,01*
PPV	0,73 ± 0,02 (0,72; 0,74)	0,01*	0,75 ± 0,02 (0,74; 0,77)	0,26	0,74 ± 0,01 (0,73; 0,74)	0,26
AUC-PRC	0,87 ± 0,012 (0,87; 0,88)	0,08	0,88 ± 0,013 (0,88; 0,89)	<0,01*	0,91 ± 0,001 (0,91; 0,91)	<0,01*
Spezifität	0,72 ± 0,01 (0,72; 0,73)	0,05	0,74 ± 0,02 (0,73; 0,75)	0,5	0,74 ± 0,01 (0,74; 0,75)	<0,01
NPV	0,79 ± 0,04 (0,77; 0,82)	0,13	0,77 ± 0,03 (0,75; 0,79)	0,05	0,84 ± 0,09 (0,78; 0,89)	0,21
AUC-PRC	0,74 ± 0,03 (0,72; 0,76)	0,03*	0,77 ± 0,02 (0,76; 0,79)	<0,01*	0,81 ± 0,01 (0,8; 0,81)	<0,01*

AUC-ROC: Fläche unter der Grenzwertoptimierungskurve (Receiver Operating Characteristic); **PPV bzw. NPV:** positiver und negativer prädiktiver Wert; **AUC-PRC:** Fläche unter der Precision-Recall Kurve (Sensitivität-PPV-Kurve); **MCC:** Matthews Korrelationskoeffizient; * signifikant $p < 0,05$.

Abbildung 2



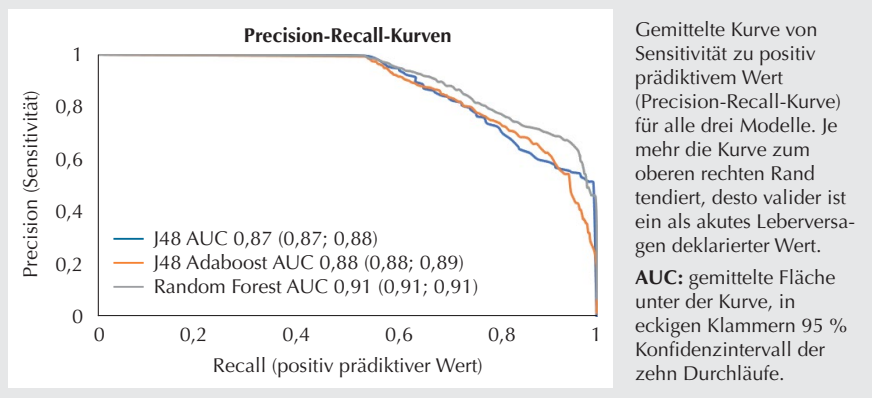
Gemittelte Grenzwertoptimierungskurve (Receiver-Operator-Kurve) für alle drei Modelle. Je näher die AUC-Kurve zum oberen linken Rand tendiert, desto besser ist die Testdiskrimination zwischen Sensitivität und falsch positivem Wert sowohl für als auch gegen ein akutes Leberversagen.

AUC: gemittelte Fläche unter der Kurve, in eckigen Klammern 95 % Konfidenzintervall der zehn Durchläufe.

Iesu et al. ermittelten in ihrer logistischen Regressionsanalyse die Parameter Schock, Herzinsuffizienz, erhöhtes Laktat, eine niedrige zentral-/gemischtvenöse Sättigung und die Dialysenotwen-

digkeit als Risikofaktoren für ein ALF. Auch die Baummodelle zeigten, dass das akute Leberversagen nach Reanimation multifaktoriell bedingt ist. Der Vorteil des J48-Modells verglichen zur

Abbildung 3



logistischen Regressionanalyse von Iesu et al. ist jedoch, dass das Baummodell diese Parameter bzw. ihre Surrogate in direkte Beziehungen setzt. Gleichzeitig spiegelt es aber auch eine absteigende Gewichtung der einzelnen Attribute beginnend von der Wurzel wider. So zweigt der Entscheidungsbaum zunächst am initialen INR bzw. Bilirubinwert als Surrogatparameter für eine vorbestehende Leberstörung auf. Anschließend zeigen sich an den weiteren Knoten

die von Iesu et al. ebenso ermittelten Risikofaktoren für die Entwicklung eines akuten Leberversagens. Während erhöhte Laktat Spiegel und die Herzinsuffizienz direkt im Baum erscheinen, waren die chronische Niereninsuffizienz und das erhöhte Kreatinin sowie die Dobutamingabe Surrogatwerte für eine potenzielle Dialysepflichtigkeit bzw. für eine niedrige zentralvenöse Sättigung. Jedoch werden anders als bei der logistischen Regressionsanalyse im Baum

die Zusammenhänge zwischen ALF und den Risikofaktoren aufgezeigt. So könnte beispielsweise die Dobutamingabe als protektiv bei erhöhter hepatotoxischer Belastung interpretiert werden. Drolz et al. zeigten in ihrer Übersichtsarbeit zur Schockleber, dass gerade ein erniedrigtes Herzzeitvolumen im Rahmen eines Schockgeschehens zum Verlust der Leberfunktion führt [19]. Gerade eine dobutaminvermittelte Erhöhung des Herzzeitvolumens und ein damit einhergehendes verbessertes Sauerstoffangebot könnte für die Leber unter bestimmten Bedingungen zur Durchbrechung des Circulus vitiosus führen [20]. Häufig werden in der Intensivmedizin bei Infekten nach Reanimation Betalaktamantibiotika verabreicht. Der Entscheidungsbaum hob auch die potenzielle Hepatotoxizität dieser Substanzklasse im Rahmen eines akuten Nierenversagens mit etwaiger Koagulopathie oder einer Laktatazidose hervor. Zwar zeigten Leitner et al., dass Betalaktame ein wesentlicher Risikofaktor für ein Leberversagen neben weiblichem Geschlecht,

erhöhten Bilirubinspiegeln und einem gleichzeitigen akuten Leberzellschaden sein können. Jedoch ermittelten sie als Substanz mit dem höchsten Risiko die Kombination Amoxicillin/Clavulansäure, gefolgt von Flucloxacillin. Der toxische Effekt von Cephalosporinen oder Piperacillin/Tazobactam wurde als deutlich geringer eingestuft, weswegen bei fehlender Substanzdifferenzierung in den Studiendaten dieser Punkt als Limitation aufgeführt werden muss [11,21]. Mit Blick auf den Entscheidungsbaum könnte in Summe mit anderen hepatotoxischen Medikamenten wie Paracetamol bereits in der Postreanimationsphase frühzeitig bei noch okkultem Leberschaden eine toxische Grenze erreicht oder gar überschritten worden sein. Zusammenfassend kann für das J48-Modell festgestellt werden, dass unter Einbeziehung von Risikofaktoren und Surrogatparametern aus routinemäßig erhobenen Patienten- und Behandlungscharakteristika sowie Laborwerten ein medizinisch nachvollziehbarer Entscheidungsbaum erstellt werden konnte.

Limitationen

Bezüglich der Auswahl der Attribute INR und Gesamtbilirubin bei Aufnahme muss festgehalten werden, dass sich ein gewisser Bias dadurch ergeben könnte, dass beide Attribute für die Definition des ALF verwendet werden. Jedoch hatten bei Aufnahme nur 11 % der ALF-Gruppe pathologische Laborparameter gemäß der Definition. Durch die deutlich unterschiedlichen INR- (ALF 1,89 vs. 1,18) und Gesamtbilirubinwerte (ALF 1,34 vs. 0,47 mg/dl) bei Aufnahme lassen sich so auch die Selektierung dieser Parameter als Hauptaufteilungspunkte des J48-Baums erklären. Bei dessen weiterer Betrachtung wird ersichtlich, dass ein ALF nur mittels initialem INR und Gesamtbilirubin in 5 von 129 Fällen (3,9 %) falsch vorhergesagt wird. Zudem erfolgte die Gruppeneinteilung bereits in der Originalstudie und wurde hier unverändert übernommen. Iesu et al. gaben einen medianen Zeitraum von 3 Tagen bis zum Auftreten eines ALF an, sodass ein J48-Modell bereits bei

Aufnahme als Frühwarnsystem dienen könnte. Der Vorteil eines RF-Modells würde in der Verwendung aller weiteren Parameter des Intensivaufenthalts und damit in der weiteren klinischen Überwachung liegen [6]. Klinisch betrachtet lenken zudem initial erhöhte Leberwerte in der frühen Intensivtherapie die Aufmerksamkeit auf mögliche spätere hepatische Komplikationen. Sollten initialer INR und Gesamtbilirubin bei Aufnahme keine Berücksichtigung in den Modellen finden, müssten auch daraus kalkulierte Attribute wie MELD- und SOFA-Score aus der Berechnung entfernt werden. Dies würde zu einer deutlichen Minderung der klinischen Aussagekraft aller Modelle führen. Daher wurden beide Attribute für die Modellierung eingeschlossen. Als weiterer Punkt muss aufgeführt werden, dass auf die Erhebungsqualität wegen des Studiendesigns kein Einfluss genommen werden konnte. So waren beispielsweise keine Daten zu Lipidwerten oder dem Serumnatrium, welches bei Leberzirrhose erniedrigt sein kann, vorhanden. Daten zur hypoxischen Hepatitis waren im Datenset nicht verfügbar und konnten somit nicht in die Berechnungen miteinfließen. Der Untersuchungszeitraum erstreckt sich über acht Jahre, in denen die Leitlinien zur Reanimationstherapie bzw. intensivmedizinischen Therapie einem Wandel unterworfen waren [22,23]. Strukturelle Verbesserungen in Prozess- und Therapieabläufen könnten heute zu einer veränderten Attributgewichtung führen. Allgemein gesprochen tendieren Algorithmen des maschinellen Lernens zu einer verbesserten Vorhersage der Mehrheitsgruppe. In vorliegender Studie lagen etwa zwei gleich große Gruppen und damit eine in etwa gleiche Gewichtung vor, sodass auch dies ein Grund für die gute Diskriminierung zwischen den Gruppen mit und ohne ALF darstellt. Da jedoch multizentrische Untersuchungen zur Häufigkeit des ALF fehlen, könnte sich die Gewichtung hin zu einer der Gruppen verschieben. Dies würde in einer geänderten Baumstruktur sowie verminderten ROC- und PRC-Kurven resultieren. Alle hier berechneten Mo-

delle weisen zufriedenstellende Testgütekriterien auf, wobei das RF-Modell die robusteste Klassendiskrimination erzielte. Jedoch muss festgehalten werden, dass nur eine kleine Kohortengröße vorlag und daher ein Kreuzvalidierungsverfahren gewählt wurde. Letzteres bietet gerade bei kleinen Kohorten die Möglichkeit einer Validierung unter unterschiedlichen Häufigkeitsverteilungen. Eine endgültige Validierung an einer unabhängigen, repräsentativen multizentrischen Testgruppe steht allerdings aus. Daher kann die hier vorgestellte gute Testgüte der Modelle nur einen ersten Orientierungswert für weitere Modelle in größeren Kohorten bieten.

Unter Betrachtung zuvor genannter Limitationen konnte in vorliegender Studie jedoch gezeigt werden, dass Methoden des maschinellen Lernens für die Vorhersage des akuten Leberversagens nach Reanimation anwendbar waren und eine prinzipielle Vorhersagbarkeit auf Basis alltäglicher klinischer Parameter möglich erscheint. Zukünftige Modelle könnten von dieser ersten Evaluierung von J48 bzw. seiner geboosteten Varianten sowie von Random Forest profitieren. Diese könnten durch Hinzunahme von weiteren Daten wie Vormedikation (u. a. Vitamin K- und Faktor-Xa-Antagonisten), Echokardiographiebefunden, der zeitlichen Entwicklung der Leberwerterhöhung, der Dosierung von Katecholaminen oder bestimmten hepatotoxischen Medikamenten in der prognostischen Aussagekraft weiter gesteigert werden. Integriert in ein Gesamtprognosemodell zu weiteren Organinsuffizienzen und zur neurologischen Prognose nach Reanimation könnten Intensivmediziner in Zukunft frühzeitig in ihren Entscheidungsprozessen unterstützt werden. Doch auch schon die vorliegenden Ergebnisse können die Aufmerksamkeit von Intensivmedizinern auf hepatische Komplikationen nach Reanimation lenken und so zu Modifizierungen in Behandlungspfaden beispielweise durch eine differenzierte Katecholamintherapie oder den Verzicht auf hepatotoxische Substanzen anregen [10,24].

Literatur

1. Peine A, Lütge C, Poszler F, Celi L, Schöffski O, Marx G et al: Künstliche Intelligenz und maschinelles Lernen in der intensivmedizinischen Forschung und klinischen Anwendung. *Anaesth Intensivmed* 2020;372–384
2. Witten IH, Frank E, Hall MA, Pal CJ: Chapter 5 – Credibility: Evaluating what's been learned. In: Witten IH, Frank E, Hall MA, Pal CJ (Hrsg.): *Data Mining (Fourth Edition)*. Cambridge: Morgan Kaufmann 2017;161–203
3. Witten IH, Frank E, Hall MA, Pal CJ: Chapter 6 – Trees and rules. In: Witten IH, Frank E, Hall MA, Pal CJ (Hrsg.): *Data Mining (Fourth Edition)*. Cambridge: Morgan Kaufmann 2017;209–242
4. Gräsner J-T, Meybohm P, Lefering R, Wnent J, Bahr J, Messelken M, et al: ROSC after cardiac arrest – the RACA score to predict outcome after out-of-hospital cardiac arrest. *European Heart Journal* 2011;32:1649–1656
5. Layeghian JS, Sepehri MM, Layeghian JM, Khatibi T: An intelligent warning model for early prediction of cardiac arrest in sepsis patients. *Computer Methods and Programs in Biomedicine* 2019;178:47–58
6. Shamout F, Zhu T, Clifton DA: Machine Learning for Clinical Outcome Prediction. *IEEE Reviews in Biomedical Engineering* 2021;14:116–126
7. Al-Dury N, Ravn-Fischer A, Hollenberg J, Israelsson J, Nordberg P, Strömsöe A, et al: Identifying the relative importance of predictors of survival in out of hospital cardiac arrest: a machine learning study. *Scand J Trauma Resusc Emerg Med* 2020;28:60
8. Perkins GD, Gräsner J-T, Semeraro F, Olasveengen T, Soar J, Lott C et al: Kurzfassung: Leitlinien des European Resuscitation Council 2021. *Notfall Rettungsmed* 2021;24:274–345
9. Roedl K, Wallmüller C, Drolz A, Horvatits T, Rutter K, Spiel A, et al: Outcome of in- and out-of-hospital cardiac arrest survivors with liver cirrhosis. *Ann Intensive Care* 2017;7:103
10. Nanchal R, Subramanian R, Karvellas CJ, Hollenberg SM, Peppard WJ, Singbartl K, et al: Guidelines for the Management of Adult Acute and Acute-on-Chronic Liver Failure in the ICU: Cardiovascular, Endocrine, Hematologic, Pulmonary, and Renal Considerations. *Critical Care Medicine* 2020;48:e173–e191
11. Iesu E, Franchi F, Zama Cavicchi F, Pozzebon S, Fontana V, Mendoza M, et al: Acute liver dysfunction after cardiac arrest. Rosenberger P, editor. *PLoS ONE* 2018;13:e0206655
12. Taylor RM, Tujios S, Jinjuvadia K, Davern T, Shaikh OS, Han S, et al: Short and Long-Term Outcomes in Patients with Acute Liver Failure Due to Ischemic Hepatitis. *Dig Dis Sci* 2012;57:777–778
13. Kamath PS, Kim WR: The model for end-stage liver disease (MELD). *Hepatology* 2007;45:797–805
14. Witten IH, Frank E, Hall MA, Pal CJ: Chapter 4 – Algorithms: The basic methods. In: Witten IH, Frank E, Hall MA, Pal CJ (Hrsg.): *Data Mining (Fourth Edition)*. Cambridge: Morgan Kaufmann 2017;91–160
15. Witten IH, Frank E, Hall MA, Pal CJ: Chapter 12 – Ensemble learning. In: Witten IH, Frank E, Hall MA, Pal CJ (Hrsg.): *Data Mining (Fourth Edition)*. Cambridge: Morgan Kaufmann 2017;479–501
16. Chicco D, Tötsch N, Jurman G: The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* 2021;14:13. DOI: 10.1186/s13040-021-00244-z
17. Jentzer JC, Wiley B, Bennett C, Murphree DH, Keegan MT, Gajic O, et al: Early noncardiovascular organ failure and mortality in the cardiac intensive care unit. *Clinical Cardiology* 2020;43:516–523
18. Møller S, Bernardi M: Interactions of the heart and the liver. *European Heart Journal* 2013;34:2804–2811
19. Drolz A, Horvatits T, Roedl K, Fuhrmann V: Schockleber und Cholestase beim kritisch Kranken. *Medizinische Klinik – Intensivmedizin und Notfallmedizin* 2014;109:228–234
20. Kasper P, Tacke F, Steffen H-M, Michels G: Hepatische Dysfunktion bei Patienten mit kardiogenem Schock. *Medizinische Klinik – Intensivmedizin und Notfallmedizin* 2019;114:665–676
21. Leitner JM, Graninger W, Thalhammer F: Hepatotoxicity of Antibacterials: Pathomechanisms and Clinical Data. *Infection* 2010;38:3–11
22. Kreimeier U, Holzer M, Toursarkissian M, Arntz HR: ERC-Leitlinien 2005 – Umsetzung im klinischen Bereich. *Notfall + Rettungsmedizin* 2008;11:91–98
23. Nolan JP, Soar J, Zideman DA, Biarent D, Bossaert LL, Deakin C, et al: European Resuscitation Council Guidelines for Resuscitation 2010 Section 1. Executive summary. *Resuscitation* 2010;81:1219–1276
24. Rutter K, Horvatits T, Drolz A, Roedl K, Siedler S, Kluge S, et al: Akutes Leberversagen. *Med Klin Intensivmed Notfmed* 2018;113:174–183.

Korrespondenz-
adresse

Dr. med.
André Luckscheiter

Klinik für Anästhesiologie, Operative
Intensivmedizin und Notfallmedizin
Klinikum der Stadt Ludwigshafen
Bremserstraße 79
67063 Ludwigshafen, Deutschland
Tel.: 0621 503-3000
Fax: 0621 503-3024
E-Mail: luckscha@kllilu.de
ORCID-ID: 0000-0002-5724-7130