

Aus der Klinik für Neurochirurgie  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

**Vergleich von Outcomeprädiktionsmodellen bei der aneurysmatischen  
Subarachnoidalblutung:  
klinische und radiografische Scores versus Machine Learning Algorithmen**

Outcome prediction in aneurysmal subarachnoid haemorrhage: a comparison of machine  
learning methods and established clinico-radiological scores

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Sophie-Charlotte Brune

aus Berlin

Datum der Promotion: 03. März 2023



## **Vorwort**

Teilergebnisse der vorliegenden Arbeit wurden veröffentlicht in:

Dengler NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, Livne M, Wolf S, Vajkoczy P, Frey D. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. Neurosurg Rev. 20. Januar 2021

## **Inhaltsverzeichnis**

<b>Vorwort</b> .....	<b>3</b>
<b>Inhaltsverzeichnis</b> .....	<b>4</b>
<b>Tabellen und Abbildungen</b> .....	<b>6</b>
<b>Abkürzungen</b> .....	<b>7</b>
<b>Zusammenfassung</b> .....	<b>8</b>
<b>Abstract</b> .....	<b>9</b>
<b>I. Einleitung</b> .....	<b>10</b>
<b>I.1. Aneurysmatische Subarachnoidalblutungen</b> .....	<b>13</b>
1.1. Definition, Ätiologie und Pathogenese.....	13
1.2. Therapie: Coiling oder Clipping? .....	14
1.3. Komplikationen der aSAB und Morbidität .....	16
<b>I.2 Outcome einer aSAB: Messung und Prädiktion</b> .....	<b>19</b>
2.1. Die modified Rankin Scale (mRS).....	19
2.2. Klinische, radiologische und gemischte Scores zur Outcome-Vorhersage .....	20
2.3. Machine Learning Modelle: Erweiterung der Prädiktionskapazität? .....	24
<b>I.3 Fragestellungen und Ziele der Arbeit</b> .....	<b>30</b>
3.1. Fragestellungen.....	30
3.2. Ziele der Arbeit.....	30
<b>II. Material, Methodik und statistische Auswertung</b> .....	<b>31</b>
<b>II.1. Material</b> .....	<b>31</b>
1.1. Patient*innenpool und Outcomes .....	31
1.2. Behandlungsalgorithmen für Patient*innen mit aneurysmatischer SAB .....	31
<b>II.2. Methodik</b> .....	<b>32</b>
2.1. Scores, Selektion einzelner Variablen und Modellierung .....	32
2.2. Modellierung und angewandte Algorithmen .....	34
2.3. Das Machine-Learning Framework.....	34
2.4. Modell-Training und Validierung .....	35
<b>II.3. Statistische Auswertung</b> .....	<b>35</b>

3.1.	Statistische Auswertung der Outcomeprädiktion per AUROC-Analyse.....	35
3.2.	Multikollinearität .....	36
<b>III. Ergebnisse.....</b>		<b>37</b>
<b>III.1. Outcome-Prädiktionskraft von klinischen, radiologischen und gemischten Scores im Vergleich – Ergebnisse der ersten AUROC Analyse .....</b>		<b>40</b>
1.1.	Outcome-Vorhersage klinischer und radiologischer Scores.....	41
1.2.	Outcome-Vorhersage gemischter Scores.....	45
<b>III.2. Outcome-Vorhersage von Machine Learning-Algorithmen im Vergleich – Ergebnisse der zweiten AUROC-Analyse.....</b>		<b>46</b>
2.1.	Kann durch die Anwendung von ML-Algorithmen die Outcome-Vorhersage bei der aSAB gegenüber den existierenden Score-Modellen verbessert werden? .....	46
2.2.	Erreichen DL-Algorithmen eine gleiche oder bessere Performance als ML-Algorithmen in der Outcome-Vorhersage? .....	50
2.3.	Besteht eine unterschiedliche Relevanz einzelner Variablen vom ‚Merkmal 8‘? ...	50
<b>IV. Diskussion .....</b>		<b>53</b>
1.	Machine Learning vs. etablierte Scores.....	54
2.	Relevanz radiografischer Parameter und von Kombinationsscores .....	55
3.	Relevanz klinischer Parameter .....	56
<b>Ausblick.....</b>		<b>59</b>
1.	Hinzuziehen weiterer Parameter.....	59
2.	Einbringen später Parameter und Vorhersage von Komplikationen .....	60
3.	Limitationen.....	62
<b>Schlussfolgerung.....</b>		<b>62</b>
<b>Literaturverzeichnis.....</b>		<b>63</b>
<b>Eidesstattliche Versicherung und Anteilserklärung.....</b>		<b>73</b>
<b>Anteilserklärung an etwaigen erfolgten Publikationen.....</b>		<b>75</b>
<b>Lebenslauf.....</b>		<b>76</b>
<b>Publikationsliste .....</b>		<b>78</b>
<b>Danksagung .....</b>		<b>79</b>
<b>Bescheinigung Statistik.....</b>		<b>80</b>

## **Tabellen und Abbildungen**

### **Tabellen**

Tabelle 2.1	Modified Rankin-Scale (mRS) - Übersicht	19
Tabelle 2.2	Hunt und Hess Score	20
Tabelle 2.3	WFNS-Score	21
Tabelle 2.4	Fisher Skala	21
Tabelle 2.5	modifizierte Fisher Skala	22
Tabelle 2.6	BNI-Skala	22
Tabelle 2.7	VASOGRADE Score	23
Tabelle 2.8	Überblick der Scores für die Variablen-Auswahl	33
Tabelle 3.1	Klinische und radiologische Daten der Patient*innen mit aSAB	39
Tabelle 3.2	Prädiktive Performance der klinischen, radiologischen und kombinierten Scores, sowie kombinierter Variablenset für ‚ungünstiges Patient*innen Outcome‘ (mRS 3-6) gemessen per AUC	40
Tabelle 3.3	Prädiktive Performance der klinischen, radiologischen und kombinierten Scores, kombinierter Variablenset in den Machine-Learning Modellen für ‚ungünstiges Patient*innen Outcome‘ gemessen per AUC	48

### **Abbildungen**

Abb. 3.1	ROC-Kurve für die Glasgow Coma Scale	42
Abb. 3.2	ROC-Kurve für den Hunt und Hess Score	42
Abb. 3.3	ROC-Kurve für den WFNS-Score	43
Abb. 3.4	ROC-Kurve für den originalen Fisher Score	43
Abb. 3.5	ROC-Kurve für den modifizierten Fisher Score	44
Abb. 3.6	ROC-Kurve für den BNI-Score	44
Abb. 3.7	ROC-Kurve für den VASOGRADE Score	45
Abb. 3.8.A	Grafische Darstellung der Area Under the Curve	49
Abb. 3.8.B	Grafische Darstellung der Leistung und Variablen Rating für das klinisch-radiologische Modell	52

## **Abkürzungen**

*ANN* – Artificial Neural Networks

*aSAB* – aneurysmatische Subarachnoidalblutung

*AUC* – area under the curve

*BNI* – Barrow Neurological Institute scale

*CT* – computed tomography

*CT-A* – computed tomography + angiography

*DCI* – Delayed Cerebral Ischaemia

*DIND* – Delayed Ischemic Neurological Deficit

*DL* – deep learning

*DSA* – Digitale Subtraktionsangiographie

*GCS* – Glasgow Coma Scale

*ICB* – intracerebrale Blutung

*ITS* – Intensivstation

*IVB* – intraventrikuläre Blutung

*KNN* – künstliche neuronale Netzwerke

*mFish* – modified Fisher Scale

*MAP* – Mean Arterial Pressure

*ML* – machine learning

*MLP* – Multilayer Perceptron artificial neural net

*mRS* – modified Rankin scale

*MRT* – Magnetic Resonance Tomography

*ROC* – receiver operating characteristics

*SDH* - Subduralhämatom

*VIF*- Variance Inflation Factor

*WFNS* – World Federation of Neurosurgical Societies

## Zusammenfassung

**Einleitung:** Verlässliche Vorhersagen über das zu erwartende Therapieergebnis bei aneurysmatischen Subarachnoidalblutungen (aSAB) auf der Grundlage von Merkmalen, die bei der Patient\*innenaufnahme zur Verfügung stehen, könnten positiven Einfluss ausüben auf Therapieentscheidungen und die Verteilung von Ressourcen. Radiographische und klinische Scores helfen Kliniker\*innen bereits, den Schweregrad einer SAB einzuschätzen. Ihre prädiktiven Fähigkeiten sind jedoch begrenzt, insbesondere als Grundlage für Therapieentscheidungen. In dieser Studie wurde untersucht, ob Machine Learning (ML) Algorithmen basierend auf Merkmalen/Variablen, die bei Patient\*innenaufnahme vorhanden sind, die Outcome-Vorhersage bei Patient\*innen mit aSAB im Vergleich zu bereits bestehenden klinischen Scores verbessern.

**Methoden:** Kombinierte klinische und radiographische Merkmale, sowie klassische Scores (Hunt and Hess Score, World Federation of Neurosurgical Societies Score, Fisher Score, modified Fisher Score, Barrow National Index, VASOGRADE Score) wurden für 388 Patient\*innen mit aSAB (n=388) an der Charité – Universitätsmedizin Berlin bei der Patient\*innenaufnahme erhoben und ausgewertet. Verschiedene ML-Modelle (insgesamt sieben Algorithmen, davon drei Typen des traditionellen generalisierten Linearen Modells, ein Tree-Boosting Algorithmus, ein Naive Bayes Classifier, ein Support Vector Machine Classifier und ein Multilayer Perceptron) wurden in drei Modellen antrainiert für (i) Scores, (ii) kombinierte Merkmalssets („*combined features*“) sowie (iii) einzelne Merkmale („*single features*“) mit einer zufallsbestimmten Teilung von Trainings- und Testset (4:1), zehnfacher Kreuzvalidierung und 50 Durchgängen. Für kombinierte Merkmalssets wurde die Merkmalsrelevanz errechnet.

**Resultate:** Es zeigte sich kein Unterschied zwischen ML-Algorithmen und etablierten Scores in der Vorhersagegenauigkeit bezüglich des Outcome einer aSAB. Es war ebenfalls kein Unterschied festzustellen zwischen einem Set kombinierter klinisch-radiologischer Merkmale bei Aufnahme (höchste AUC 0,78, Tree Boosting) und der Glasgow Coma Scale (GCS) bei Aufnahme - dem klinischen Score mit der besten Performance (höchste AUC 0,76, Tree Boosting). Der GCS-Wert und das Alter waren die Variablen mit der besten Outcomeprädiktion in der Merkmalskombination.

**Schlussfolgerungen:** In dieser Patient\*innenkohorte mit aSAB war die Vorhersage des Therapieergebnisses durch ML-Algorithmen vergleichbar mit den traditionellen, etablierten klinischen Scores. Weitere Studien sind nötig, um zusätzliche Inputvariablen zu untersuchen, die nicht bereits Teil der traditionellen klinisch-radiologischen Merkmale sind, um zu prüfen, ob eine bessere Performance der Outcome-Vorhersage für aSAB-Patient\*innen erreicht werden kann.



## Abstract

**Introduction:** Up to now, clinical outcome prediction of aneurysmal subarachnoid haemorrhage (aSAH) has mainly been based on clinical and radiographic scoring systems. They may help as a rough indicator of outcome, but their precise predictive abilities are limited. Their limitations are particularly salient with regard to advocating for one treatment strategy over another, which – in an age of scarcity of medical resources in the midst of a global pandemic—is of critical importance. In the present study, we aimed to examine whether an approach based on Machine Learning (ML) algorithms, using variables readily available on patient admission, may improve outcome prediction compared to the status quo based on current clinical and radiological scoring systems.

**Methods:** Using a consecutive single-centre database of 388 patients suffering from aSAH, we implemented an analysis of combined clinical and radiographic features as well as classic scoring systems, such as Hunt and Hess, WFNS, BNI, Fisher, modified Fischer and Vasograde. We trained a total of seven different ML-algorithms for scores, single features and combined features. The algorithms included a tree boosting algorithm, a Naïve Bayes classifier, a support vector machine classifier, a multilayer perceptron artificial neural net (MLP) as well as three different types of generalized linear models. A random split into training and test-sets was implemented according to a four to one ratio. We proceeded with ten-fold cross-validations and fifty shuffles. We calculated feature importance for combined features.

**Results:** Our findings show that there is no performative difference between traditional and more modern ML models using clinico-radiographic features. No significant difference was found when comparing a set of clinico-radiological features available on admission and the Glasgow Coma Scale (GCS) – the best performing clinical score (highest AUC 0,76, running a Tree Boosting model).

**Discussion:** The GCS and patient age turned out to be the most relevant variables within the feature combination. Functional outcome prediction by advanced ML-techniques did not outperform established scores in our cohort of 388 aSAH Patients. In order to reap the benefits and the full power of data mining models such as ML-algorithms to improve functional outcome prediction in aSAH, future research efforts will need to examine input variables that have not yet been taken into account.

## **I. Einleitung**

Klinische, radiologische und gemischt klinisch-radiologische Scores erlauben Kliniker\*innen, den Schweregrad einer Erkrankung und ihren wahrscheinlichen natürlichen Verlauf einzuschätzen (1, 2). Als solche stellen sie eine Hilfestellung dar, um adäquate Therapieentscheidungen und eine Vorhersage bezüglich des Behandlungsergebnisses einer Erkrankung zu treffen. Nicht nur zu Zeiten einer globalen Pandemie wie COVID-19 ist die Schnelligkeit der Prognoseeinschätzung entscheidend, um die begrenzten Kapazitäten des Gesundheitssystems – wie z.B. Betten auf Intensivstationen und notfallmedizinische Versorgung – verantwortlich zu verteilen. Eine bessere Outcome-Vorhersage kann ebenso dabei helfen, die therapeutische Entscheidungsfindung zu verbessern. Durch eine frühe Identifizierung von Hochrisikopatient\*innen, bei denen ein schlechter Outcome sehr wahrscheinlich ist, können das Monitoring und die Therapie frühzeitig angepasst werden (3). Nicht zuletzt können effiziente Scores es ebenfalls ermöglichen, Patient\*innen und ihre Angehörigen besser bezüglich der Therapieoptionen und Prognose zu beraten.

Seit den späten 1960er Jahren begannen erste Bestrebungen, klinische Scores für das Krankheitsbild der aneurysmatischen Subarachnoidalblutung (aSAB) zu entwickeln, die besagte Ansprüche zumindest im Ansatz erfüllten. Die Hunt und Hess Skala und der Score der World Federation of Neurosurgical Societies (WFNS-Score) sind für die aSAB mittlerweile seit Jahrzehnten Teil des klinischen Alltags. Beide Scores erheben klinische Patient\*innencharakteristika bezüglich der Vigilanz und spezifischer neurologischer Defizite. Der WFNS-Score wurde entwickelt, um die starke Variabilität in der Erhebung des Hunt und Hess Scores zu kompensieren („interobserver Variability“) – und auch, weil die Hunt und Hess Skala klinische Parameter unzureichend definierte. Der WFNS-Score beruht auf der Glasgow Coma Skala (GCS) sowie wesentlichen neurologischen Defiziten. Die Einführung der Computertomographie (CT), sowie ihre schrittweise Verbreitung in der Routinediagnostik führte zur Fortentwicklung und Veränderung verschiedener radiologischer Scores, u.a. durch den Einschluss von qualitativen Bildgebungskriterien wie z.B. die Verteilung von subarachnoidalen Blutansammlungen, intraventrikulären Blutungen (IVB) oder intrazerebralen Blutungen (ICB). Die erste teilquantitative radiologische Skala wurde vom Barrow Neurological Institute (BNI) im Jahr 2012 entwickelt. Gemischte Scores, die auf einer Kombination von klinischen und

radiographischen Scores beruhen, um eine verbesserte Vorhersage z.B. von Vasospasmen und dem klinischen Outcome von Patient\*innen zu treffen, wurden erst später entwickelt (z.B. der VASOGRADE Score). Jeder neu- oder weiterentwickelte Score ging einher mit der Hoffnung auf eine verbesserte Vorhersagekraft. Die Vergleichbarkeit der Prädiktionskraft verschiedener Scores ist in der Vergangenheit jedoch stets durch eine fehlende Standardisierung der Erhebung von Daten (z.B. bezüglich des Outcomes), sowie durch unterschiedliche Patient\*innenkohorten begrenzt gewesen. Weder die bestehenden klinischen noch die radiographischen Scores bieten bislang die prognostische Genauigkeit und Sicherheit, um darauf basierend Entscheidungen über Leben oder Tod zu treffen. Zweifelsohne besteht der klinische Bedarf, über Instrumente zu verfügen, die eine individuelle Patient\*innen-zentrierte und frühzeitige Risikoeinschätzung ermöglichen.

In den letzten zehn Jahren hat die Anzahl der Forschungsvorhaben exponentiell zugenommen, in denen die Anwendbarkeit diverser computer-gestützter Algorithmen zur Deckung dieses Bedarfs geprüft wird. Insbesondere werden sogenannte *Machine Learning (ML) Algorithmen* in verschiedenen Teildisziplinen der Medizin untersucht, u.a. um die Prädiktionskapazität existierender klinischer Methoden zu optimieren. ML-basierte Vorhersagemodelle, die auf sogenannten Support Vector Machines, Entscheidungsbäumen, Bayesische Methoden oder Künstlichen Neuronalen Netzwerken (KNN, *eng: Artificial Neural Networks, ANN*) basieren, haben in verschiedenen medizinischen Fachgebieten bereits eine bessere Vorhersagekraft als traditionelle Methoden gezeigt. Insbesondere KNN und tree-boosting Methoden (basierend auf Entscheidungsbäumen) haben im Vergleich zu klassischen ML Methoden (wie z.B. Lineare oder Logistische Regressionen) in vielfachen medizinischen Teilgebieten erfolgsversprechende Resultate geliefert (4–6).

Inwiefern ML-Algorithmen die bekannten Schwächen klinischer Scoringsysteme kompensieren und ggf. die klinische Outcome-Vorhersage und Prognose bezüglich der aSAB verbessern können, ist Gegenstand dieser Arbeit. Ihr Ziel ist die Prüfung der Vorhersagegenauigkeit einzelner ML-Algorithmen, insb. diejenige von künstlichen neuronalen Netzwerken und von Tree-Boosting Ansätzen, im Vergleich zu einer Auswahl etablierter klinischer, radiologischer und kombinierter Scores. Hierzu wurde innerhalb der Forschungsgruppe, deren Arbeit diese Monographie wiedergibt, erstens geprüft, inwiefern existierende klinische Scores in ihrer Vorhersagegenauigkeit durch die Anwendung von ML-Algorithmen verbessert werden können.

Zweitens wurde ein Set von Variablen aus klinischen und radiologischen Scores erstellt, die für das Behandlungsergebnis relevant und bei Patient\*innenaufnahme vorhanden sind. Die Vorhersagefähigkeit dieses Sets wurde mit der Vorhersagefähigkeit von traditionellen klinischen Scores verglichen, um Nachvollziehbarkeit und Vergleichbarkeit mit anderen Arbeiten in diesem Themenbereich zu gewährleisten.

## **I.1. Aneurysmatische Subarachnoidalblutungen**

### *1.1. Definition, Ätiologie und Pathogenese*

Eine Subarachnoidalblutung (SAB) ist eine Einblutung in den liquorgefüllten Subarachnoidalraum. Letzterer wird durch den Raum zwischen Arachnoidea und Pia mater gebildet. Hier verlaufen die großen arteriellen Gefäße des Gehirns, so zum Beispiel der Circulus arteriosus Willisii. Die Ätiologie der SABs lässt sich in traumatische und nicht-traumatische Ursachen einteilen. Die überwiegende Mehrheit der SABs entstehen im Zuge eines Traumas. Bis zu 40% aller schweren Schädel-Hirn Traumata werden durch eine SAB begleitet und gehen in der Regel mit einem schwereren Verlauf und einer schlechteren Prognose einher.

Nicht-traumatische SABs sind für ca. 5% aller Schlaganfälle ursächlich. Sie haben eine Inzidenz von ca. 9/100.000 Einwohnern pro Jahr (7). Das mittlere Lebensalter der Betroffenen liegt bei 50 Jahren, wobei Frauen häufiger betroffen sind als Männer (im Verhältnis 3:2). Nicht-traumatische, spontane SABs sind zu ca. 80-85% die Folge der Ruptur eines Aneurysmas der basalen Hirnarterien (7). Weitere Ursachen einer nicht-traumatischen SAB (ca. 15%) sind unter anderem perimesenzephalale Blutungen (= präpontine SAB) (7). Hierbei handelt es sich am wahrscheinlichsten um venöse Blutungen deren Ätiologie unbekannt bleibt. Selten können nicht aneurysmatische, nicht traumatische SABs Tumor-bedingt sein, sowie im Rahmen einer arterio-venösen Malformation stattfinden (7). Diese Monographie konzentriert sich jedoch spezifisch auf das Krankheitsbild der aneurysmatischen SAB (aSAB), dessen Pathogenese sich der nächste Abschnitt eingehender widmet.

Bei einem Aneurysma handelt es sich um die Gefäßwandschwäche einer Arterie, die zu einer Gefäßaussackung führt. Unterschiedliche Schichten der Gefäßwand können betroffen sein. Die Anlage der zerebralen Aneurysmen erfolgt wahrscheinlich kongenital. Im Verlauf - a.e. im frühen Erwachsenenalter - nehmen sie durch hämodynamische Faktoren an Größe zu (9). Prädisponierend für die Ausbildung von Aneurysmata sind u.a. eine kongenitale Wandschwäche der *Tunica media*. In diesem Fall kann es bevorzugt an Gefäßgabelungen zu Aussackung der Gefäßwand kommen, da dort eine erhöhte Druckbeanspruchung der Gefäßwand herrscht (9). Weitere prädisponierende Faktoren sind erworbene Gefäßveränderungen durch Arteriosklerose oder Entzündungen, Bluthochdruck, sowie genetische Erkrankungen wie die autosomal-dominante polyzystische

Nierenerkrankung (ADPKD), das Marfan- oder das Ehler-Danlos Syndrom (7). Allgemeine Risikofaktoren für eine aSAB sind überwiegend verhaltensabhängig (Rauchen, Alkoholkonsum, Konsum sympathomimetischer Drogen (z.B. Kokain)), zum Teil jedoch auch abhängig von Geschlecht und ethnischer Herkunft (7).

Die häufigste Lokalisation der cerebralen Aneurysmata sind die Arteria cerebri anterior /Arteria communicans anterior (30-40%), die Arteria carotis interna (ca. 30%), die Arteria cerebri media (20-25%) und die Arteria basilaris/Arteria vertebrales (ca. 10%) (9). Die Morphologie der Aneurysmata kann sehr divers sein und erlaubt Rückschlüsse über die Blutungswahrscheinlichkeit: Höchstes Blutungsrisiko haben die sog. sakkulären Aneurysmata. Diese sackförmigen, runden oder bläschenförmigen Ausstülpungen befinden sich meist an Gefäßabgabelungen und haben eher eine multifaktorielle Genese (z.B. familiäre Veranlagung oder ADPKD). Fusiforme Aneurysmen haben ein geringeres Blutungsrisiko, sind spindelförmig und eher langstreckig. Sie treten meist an nicht-verzweigten Gefäßabschnitten auf und sind i.d.R. arteriosklerotisch bedingt.

### *1.2. Therapie: Coiling oder Clipping?*

Die primäre Behandlungsstrategie einer aSAB besteht in der intensivmedizinischen Überwachung und Therapie sowie dem Verschluss des Aneurysmas zur Vermeidung einer Rezidivblutung (10). In Abhängigkeit von Lage, Größe und Konfiguration des Aneurysmas werden endovaskuläre oder chirurgische Therapiekonzepte verfolgt. Endovaskuläre Lösungen beinhalten die Versorgung des Aneurysmas mittels Stent- oder nicht Stent-gestütztem Coiling, Flussdiversion oder anderweitiger „Embolisationssysteme“. Das Coiling stellt das häufigste endovaskuläre Versorgungsmittel dar, und erzielt eine Thrombosierung des Aneurysmas durch elektrolytisch ablösbare Platinspiralen (GDC-Coils), die durch arterielle Mikrokatheter im Aneurysma-Sack platziert werden (11). Die chirurgischen Behandlungsstrategien beinhalten das sog. Clipping, Wrapping, oder den Bypass. Die primäre neurochirurgische Versorgung des rupturierten Aneurysmas erfolgt meist per sog. Clipping, wenn das Aneurysma im vorderen Kreislauf liegt (Arteria cerebri media, Arteria cerebri anterior, Arteria communicans anterior). Zumeist erfolgt zuerst eine Liquorentlastung, z.B. durch die Anlage einer externen Ventrikeldrainage (EVD). Diese Entlastung kann jedoch auch lumbal

oder zisternal erfolgen (11). Bei der Operation von cerebralen Aneurysmen ist zuerst die Darstellung der zum Aneurysma ziehenden Gefäße vor der Präparation des eigentlichen Aneurysmas nötig. Zumeist erfolgt dann die Anlage eines temporären Clips, und nach Kontrolle einer suffizienten Durchblutung ( $\text{ptiO}_2 > 8\text{mmHg}$ ) die Anlage eines endgültigen Clips (11). Intraoperativ kann die Indocyaningrün Videoangiografie (ICG) mit Hilfe spezieller Mikroskopfilter zur Bestätigung der Durchgängigkeit der Gefäße und der suffizienten Ausschaltung des Aneurysmas genutzt werden (12, 13). Alternativ wird teilweise noch ein intraoperativer Doppler-Ultraschall zur Bestätigung der Durchgängigkeit von Gefäßen genutzt, sowie zum Aufsuchen von begleitenden intrazerebralen Blutungen oder einer plötzlich zunehmenden Schwellung (13).

Die chirurgische und interventionelle cerebrale Aneurysmaversorgung nach akuter aSAB wurde in den vergangenen Jahren durch vier relevante prospektive, randomisierte und kontrollierte Studien miteinander verglichen. Der „International Subarachnoid Aneurysm Trial“ (ISAT) und der „Barrow Ruptured Aneurysm Trial“ (BRAT) (14–16) suggerieren in der einjährigen Nachbeobachtung ein signifikant besseres klinisches Ergebnis durch die interventionelle, endvaskuläre Behandlung. Dieser Vorteil scheint sich im Langzeitverlauf jedoch zu relativieren, da die Reperforationsrate und die Rate an Blutungsrezidiven bei gecoilten Aneurysmen höher als bei chirurgisch geclippten Aneurysmen zu sein scheint (17).

ISAT zeigte anhand von 1644 eingeschlossenen Patient\*innen in der ein- und fünf-Jahresanalyse ein signifikant niedrigeres Mortalitätsrisiko nach Coiling, bei jedoch erhöhter Blutungsrezidivrate (14, 15). Die Sterblichkeitsraten glichen sich zwischen Coiling und Clipping allerdings nach 18 Jahren an (15). Hauptkritikpunkt am ISAT-Studiendesign ist ein Selektionsbias zu Gunsten einer endovaskulären Versorgung. Die Versorgungszentren mit dem höchsten Rekrutierungsanteil der aSAB-Fälle waren demnach ebenfalls diejenigen Zentren, die über besonders viel Erfahrung bezüglich der endovaskulären Versorgung von aSABs verfügten, so dass diese Therapiemodalität überproportional oft angewandt worden sei (20). Ein weiterer Kritikpunkt betrifft die höhere Blutungsrezidivrate in der Clipping-Gruppe, bevor das Aneurysma ausgeschaltet wurde. Die Ausschaltung des Aneurysmas erfolgte in der Coiling-Gruppe tatsächlich signifikant früher, als in der Clipping-Gruppe (1.1 Tage fürs Coiling vs. 1.7 Tage fürs Clipping) (21). In der BRAT-Studie (N=365) wurden dagegen alle einer Therapie zugänglichen aSAB-Patient\*innen eingeschlossen

und ein Crossover zwischen den Gruppen ‚Coiling‘ und ‚Clipping‘ ermöglicht (15). Dazu bot BRAT eine bessere Vergleichbarkeit beider Gruppen durch höhere Expertise der involvierten Therapeuten und durch eine Zeitspanne bis zur Therapie von unter 24 Stunden.

Im 1-Jahresverlauf nach den Interventionen zeigte sich im BRAT eine leichte Überlegenheit des Coilings gegenüber dem Clipping (22). Im 3-Jahres- und 6-Jahresverlauf nach Aneurysma-Verschluss bei stattgehabter aSAB zeigte sich für Aneurysmen des anterioren Stromgebiets eine Gleichwertigkeit von Clipping und Coiling bezüglich des Patient\*innen-Outcomes. Im hinteren Stromgebiet war das Coiling jedoch dem Clipping überlegen. Im Rahmen der BRAT Studie war die Häufigkeit einer notwendigen Nachbehandlung bei insuffizientem Aneurysma-Verschluss in der Coiling-Gruppe gegenüber dem Clipping deutlich erhöht. Die Obliterationsrate durch Clipping lag bei 93% gegenüber nur 22% durch Coiling. Die Vergleichbarkeit der beiden Behandlungsformen im Rahmen der Studie wurde jedoch aufgrund einer sehr hohen Wechselrate der Patient\*innen (38 – 42 %) nach Randomisierung von der Coiling- in die Clipping-Gruppe in Frage gestellt (23, 24).

### *1.3. Komplikationen der aSAB und Morbidität*

Der klinische Verlauf einer aSAB ist sehr häufig durch Komplikationen geprägt - eine Tatsache, die das Outcome maßgeblich prägt und zur hohen Letalität des Krankheitsbildes beiträgt (7). Die Einblutung in den Subarachnoidalraum führt in der Regel zum Anstieg des intrakraniellen Drucks (Eng., *Intracranial Pressure*, ICP) und zum charakteristischen Vernichtungskopfschmerz und Meningismus, Übelkeit und Erbrechen. Die häufigste frühe Komplikation einer aSAB ist die Rezidivblutung. Durch die körpereigene Fibrinolyse kommt es in bis zu 5% der Fälle innerhalb der ersten 24 Stunden zu einer Rezidivblutung. Ist die Blutungsquelle nicht primär versorgt worden, so ergibt sich eine Rezidivrate von ca. 50% innerhalb der ersten sechs Monate. Eine frühzeitige Behandlung des Aneurysmas beugt dieser Komplikation vor (25, 26).

Die häufigsten sekundären Komplikationen einer aSAB sind Vasospasmen und sogenannte sekundäre Ischämien, sowie Ischämien ohne Vasospasmen (z.B. aufgrund von Spreading Depolarisation) (27). Durch die Freisetzung vasokonstriktiver Faktoren unter dem Einfluss von Blutabbauprodukten kommt es in 30-70% der aSABs zu diffusen Vasospasmen. Diese führen oft



zu einer lokalen Hypoperfusion des betroffenen Gebiets und entsprechend zu sekundären ischämischen Hirninfarkten (sog. *Delayed Cerebral Ischaemia*, DCI) (28–31). DCIs treten bei 12–57% der Patient\*innen mit aSAB auf und sind für ca. 28% der post-aSAB Mortalität verantwortlich, sowie für bis zu 39% der permanenten Hirnschäden in Folge einer aSAB (10, 32). Die Pathogenese der Vasospasmen bleibt teilweise noch unklar. Lokale Entzündungsvorgänge und durch die Blutung veränderte Ladungsverhältnisse entlang der Endothelzellen werden u.a. als pathogene Faktoren vermutet. Ursächlich ist die Störung der vaskulären Autoregulation durch Blutabbauprodukte im Subarachnoidalraum und die Freisetzung von vasoaktiven Substanzen und Blutabbauprodukte im Subarachnoidalraum (33).

Während proximale Vasospasmen größere Gefäße an der Hirnbasis betreffen, haben periphere Vasospasmen eher Auswirkungen auf intraparenchymale Gefäße (34). Der Häufigkeitsgipfel des Auftretens von Vasospasmen findet zwischen dem dritten und dem 14. postiktalen Tag statt (selten bis zu 20 Tage postiktal). In etwa 20% der Fälle erleiden betroffene Patient\*innen bleibende Defizite in Folge sekundärer Ischämien.

Durch die mechanische Schädigung, die durch eine Massenblutung hervorgerufen werden kann, kommt es unter Umständen zu einer Schädigung der Blut-Hirn-Schranke und somit zu einem vasogenen Hirnödem, das durch Ausschüttung proinflammatorischer Zytokine exazerbiert und aufrechterhalten wird. Infolge einer stressinduzierten Sympathikusaktivierung können systemische, v.a. kardiopulmonale Komplikationen auftreten (Hypo-/Hypertension, Arrhythmie, Hypoxämie, respiratorische Alkalose/ Azidose, Lungenembolie) (35). Ein Großteil der aSAB-Patient\*innen entwickelt in der Frühphase ein Systemisches Inflammatorisches Response Syndrom (SIRS) (36).

Weitere Komplikationen sind die Meningitis sowie die Takotsubo Kardiomyopathie (neurogene Stress-Kardiomyopathie) (37). Zwei maßgebliche Ursachen für ein schlechtes Outcome von aSAB-Patient\*innen sind kardiale Komplikationen, die u.a. durch den Cushing Reflex induziert werden (Blutdrucksteigerung mit gelegentlichem Herzfrequenzabfall bei Anstieg des Hirndrucks). Innerhalb der ersten Stunden und Tage entwickelt sich bei aSAB-Patient\*innen oft eine arterielle Hypertonie durch den kompensatorischen Cushing-Reflex. Der zentrale Sympathikotonus ist dauerhaft erhöht, vermutlich u.a. durch die Reizung oder Verletzung des Hypothalamus durch die

Subarachnoidalblutung (7). Durch den erhöhten Sympathikotonus steigt der Blutdruck reaktiv an. Durch die Stimulation des *Glomus caroticum* (Barorezeptor) kann es zu einer vagal verursachten Bradykardie kommen. Der sog. Cushing Reflex stellt einen einfachen Kompensationsmechanismus dar, welcher der Aufrechterhaltung der zerebralen Durchblutung dienen soll. Er ist jedoch meist aufgrund der gestörten Autoregulation der zerebralen Durchblutung im Rahmen einer aSAB nicht erfolgreich. Vielmehr stellen die teilweise extremen Blutdruckanstiege eine kardiale Belastung dar, die zu Komplikationen führen können und die Prognose des Patient\*innen i.d.R. weiter verschlechtern. Selbst wenn sich meist nur selten lebensbedrohliche kardiale Komplikationen entwickeln, wie zum Beispiel EKG-Veränderungen, Troponin-T- sowie CK-MB-Anstiege, so sind jedoch ebenfalls ventrikuläre und supraventrikuläre Herzrhythmusstörungen bis hin zum Herzinfarkt möglich, oder ein kardiogenes Lungenödem. Selten tritt eine neurogene Stress-Kardiomyopathie auf (Takotsubo Kardiomyopathie) (37).

Ungefähr 10-15% der Patient\*innen mit einer aSAB sterben, bevor eine medizinische Versorgung möglich ist (7). Die Sterblichkeitsrate der medizinisch anbehandelten Patient\*innen beträgt ca. 10% innerhalb der ersten Tage, insbesondere aufgrund von Komplikationen wie z.B. einem neurogenen Lungenödem oder einem neurogenen myocardialen ‘Stunning’ (‘*stunned myocardium*‘) (7). Für die überlebenden Patient\*innen ohne chirurgische Versorgung ist ein Blutungsrezidiv die Hauptursache für Morbidität und Mortalität. Das Risiko eines Rezidivs liegt innerhalb der ersten 14 Tage bei ca. 15-20% (7, 8). Innerhalb der ersten sechs postiktalen Monate beträgt die Gefahr einer Rezidivblutung ca. 50%. (38). Die Mortalität einer aSAB liegt bei 20%, wenn bei Aufnahme ein Hunt und Hess score von Grad I° oder II° bestand. Die Mortalität von Patient\*innen mit einem Hunt und Hess Score von Grad I° oder II°, bei denen ein chirurgischer Eingriff (i.d.R. Clipping) durchgeführt wurde, sinkt auf 14% (7). Die wichtigste Todesursache für Patient\*innen mit einem Hunt und Hess Score von I° oder II° sind Rezidivblutungen (7).

## I.2 Outcome einer aSAB: Messung und Prädiktion

### 2.1. Die modified Rankin Scale (mRS)

Das mittel- bis langfristige Outcome wird anhand der sog. modified Rankin-Scale (mRS) erfasst. Die mRS - von J. Rankin in 1957 erstmalig vorgestellt (39) und 1988 von van Swieten u.a. modifiziert (40) - dient der groben Einschätzung des Behinderungsgrades nach einem Schlaganfall (sog. funktionales Outcome). SABs stellen 1-7% aller Schlaganfälle dar (41). Die mRS errechnet sich anhand eines Punktwertes von 0 bis 6, je nach Grad der Behinderung (s. Tabelle 2.1.). Je höher der Punktwert, desto schwerer die Beeinträchtigung.

**Tabelle 2.1.: Modified Rankin Scale (mRS)**

Punkte	Behinderungsgrad	Klinik
0	Keine Behinderung, keine Symptomatik	<ul style="list-style-type: none"><li>• Überhaupt keine Symptome</li></ul>
1	Keine signifikante Behinderung trotz Symptomatik	<ul style="list-style-type: none"><li>• Kann sich eigenständig versorgen</li></ul>
2	Leichte Behinderung	<ul style="list-style-type: none"><li>• Braucht in einzelnen Bereichen wenig Hilfe</li><li>• Fast selbständige Versorgung</li></ul>
3	Mäßige Behinderung	<ul style="list-style-type: none"><li>• Braucht Hilfe in mehreren Bereichen</li><li>• Weitgehend selbständige Versorgung</li><li>• Gehen noch ohne Hilfe</li></ul>
4	Schwere Behinderung	<ul style="list-style-type: none"><li>• Braucht Hilfe bei Aktivitäten des tägl. Lebens</li><li>• Keine selbstständige Versorgung möglich</li><li>• Gehen nur mit Hilfe</li></ul>
5	Sehr schwere Behinderung	<ul style="list-style-type: none"><li>• Ständige Hilfe und Überwachung notwendig</li><li>• Inkontinenz</li><li>• Bettlägerigkeit</li></ul>
6	Tod	-

Quelle: Amboss, modified Rankin-Scale, 12.06.2020

## 2.2. Klinische, radiologische und gemischte Scores zur Outcome-Vorhersage

Im Folgenden werden Aufbau, Parameter und Vorhersagegenauigkeit einzelner klinischer, radiologischer und gemischter Scores zur Beurteilung einer aSAB vorgestellt und ihre jeweilige prognostischen Reichweite und Leistungsfähigkeit erläutert.

### 2.2.1. Klinische Scores: Hunt und Hess, WFNS

Die weltweit verbreitetsten klinischen Scores zur Klassifikation von aSABs sind die Einteilung der klinischen Merkmale einer SAB nach der Hunt und Hess Skala (s. Tabelle 2.2), sowie nach der WFNS-Skala (s. Tabelle 2.3). Sie erlauben z.T. eine prognostische Einschätzung. Dr. William E. Hunt und Dr. Robert McDonald Hess veröffentlichten im Jahr 1968 die Hunt und Hess

**Tabelle 2.2 – Hunt und Hess Skala**

Grad	Klinik
1	Asymptomatisch oder leichter Kopfschmerz und/oder leichter Meningismus
2	Mäßige bis starke Kopfschmerzen, Meningismus, möglich: Hirnnervenausfälle
3	Somnolenz und/oder Verwirrtheit und/oder leichte fokal-neurologische Ausfälle
4	Sopor, mäßige bis schwere fokal-neurologische Ausfälle, vegetative Störungen
5	Koma, Zeichen der Einklemmung

Klassifikation, die heutzutage vielleicht den am weitesten verbreiteten Score zur klinischen Einordnung einer aSAB bezüglich ihrer Letalität darstellt. Er unterliegt einer vergleichsweise hohen subjektiven Variabilität zwischen Erhebern (die sog. ‚Interobserver-Variabilität‘). Die Hunt und Hess Klassifikation besteht aus fünf Graden. Grad 1 bezeichnet eine aSAB, die asymptomatisch oder lediglich durch leichte Kopfschmerzen und/oder einen leichten Meningismus gekennzeichnet ist;

Grad 2 wird durch mäßige bis starke Kopfschmerzen, Meningismus und möglicherweise durch Hirnnervenausfälle charakterisiert; Grad 3 wird durch Somnolenz und/oder Verwirrtheit und/oder leichte fokal-neurologische Ausfälle charakterisiert; unter Grad 4 fallen Sopor, mäßige bis schwere fokal-neurologische Ausfälle und vegetative Störungen; Grad 5 geht mit Koma sowie Zeichen der Einklemmung einher. Ein hoher Grad im Hunt und Hess Score ist mit einem schlechteren Outcome und einer höheren Letalität verbunden (41, 42). Im Jahr 1988 wurde von der WFNS eine Skala veröffentlicht, die auf dem GCS und dem Vorhandensein eines fokalen neurologischen Defizits fußt (Outcome = GCS) (44).

**Tabelle 2.3 – WFNS-Skala**

Grad	GCS	Fokales Defizit
1	15	Nein
2	13-14	Nein
3	13-14	Ja
4	7-12	Für Einteilung irrelevant
5	3-6	

Ein hoher Grad auf der WFNS-Skala ist mit einem schlechteren Outcome und einer höheren Letalität verbunden (44).

*2.2.2. Radiologische Scores: Fisher Skala, modifizierte Fisher Skala, BNI*

Eine Einteilung der aSAB kann ebenfalls anhand der Blutmenge im CT und somit durch primär radiologische Scores erfolgen. Vornehmlich sind hier die Fisher-Skala (s. Tabelle 2.4) und die modifizierte Fisher-Skala (s. Tabelle 2.5) zu erwähnen, sowie die BNI-Skala (s. Tabelle 2.6). In Deutschland kommen den radiologischen Skalen eine vergleichbar untergeordnete Bedeutung hinsichtlich des klinischen Managements von aSAB-Fällen zu (45).

Die Fisher-Skala wurde im Jahr 1980 veröffentlicht (46), mit dem Ziel das Risiko für angiographische Vasospasmen vorherzusagen (= Outcome).

**Tabelle 2.4 – Fisher Skala**

Grad	CT-Befund
1	Kein Blut im Subarachnoidalraum oder im Ventrikel
2	Diffuse Subarachnoidalblutung, Dicke < 1mm, keine Clots
3	SAB, Dicke >1mm und/oder umschriebene Clots
4	Diffuse SAB jeglicher Dicke mit intrazerebraler oder intraventrikulärer Blutung

Der Fisher Score basiert auf der Einteilung der Dicke einer Blutung in der CT-Bildgebung (in mm) sowie der Lokalisierung

- der aSAB,
- einer möglichen intraventrikulären Ein-

blutung (*intraventrikular haemorrhage*, IVB),

- einer möglichen intrazerebralen Blutung.

Der Fisher Score war anfänglich kaum statistisch validiert und seine prädiktive Kraft bleibt weiterhin unklar (40, 46–50). Seit dem Jahr 2006 wird die sog. modifizierte Fisher-Skala (mFS) genutzt. Grundlage ist die Auswertung der Ausdehnung einer IVB und der Ausdehnung möglicher

**Tabelle 2.5 – modifizierte Fisher Skala**

Grad	CT-Befund
0	Keine subarachnoidale oder intraventrikuläre Blutung
1	Fokale oder diffuse SAB, Dicke <1mm, ohne intraventrikuläre Blutung
2	Fokale oder diffuse SAB, Dicke <1mm, mit intraventrikulärer Blutung
3	Fokale oder diffuse SAB, Dicke >1mm, ohne intraventrikuläre Blutung
4	Fokale oder diffuse SAB, Dicke >1mm, mit intraventrikulärer Blutung

Blutansammlungen in den Zisternen. Prädiktions-Outcome bleiben angiographische Vasospasmen. Die mFS ist mittlerweile klinisch etabliert, da sie besser mit dem Auftreten von Vasospasmen korre-

liert (47), und sekundäre Hirnischämien und das Outcome der Patient\*innen besser vorhersagt (51).

**Tabelle 2.6 – BNI Skala**

Punktezahl	Dicke des Hämatoms in mm	Inzidenz symptomatischer Vasospasmen (%)
1	Keine bildgebend sichtbare SAB	0
2	≤5mm	12,5
3	>5-10mm	21,6
4	>10-15mm	29,6
5	>15-20mm	50

Der BNI-Score wurde im Jahr 2012 entwickelt, um angiographische und symptomatische Vasospasmen (= Outcome) nach einer aSAB vorherzusagen (52). Der BNI-Score wird anhand einer computer-tomographischen Bildgebung erhoben und

quantifiziert die Dicke möglicher subarachnoidaler Hämatome. Seine Vorhersagekraft bezüglich angiographischer und symptomatischer Vasospasmen, sekundärer Hirnischämien und ungünstiger Outcomes sowie der Vorhersage eines Shunt-abhängigen Hydrocephalus ist im Vergleich zum mFish genauer und verlässlicher (52, 53). Nichtsdestotrotz bleiben klinische Scores wie der Hunt

and Hess Score oder der WFNS Score bis heute dem BNI in ihrer prädiktiven Kapazität bezüglich des Patient\*innen-Outcome überlegen (52).

### 2.2.3. Kombinierter klinisch-radiologischer Score: VASOGRADE

Um den Schwächen der klinischen und der radiologischen Scores entgegenzuwirken, wurde eine neue Generation kombinierter Scores entwickelt, die sowohl klinische wie auch radiologische Daten erfassen. Ein Hauptbeispiel ist der VASOGRADE Score (s. Tabelle 2.7) (54–58).

Der VASOGRADE Score wurde entwickelt, um das Auftreten von Sekundärischämien vorherzusagen (= Outcome). Er wurde aber auch entwickelt, um zwischen guten WFNS Scores mit und ohne signifikanter aSAB zu differenzieren, sowie zwischen schlechten WFNS Scores bezüglich der Wahrscheinlichkeit einer sekundären Hirnischämie (45). Der VASOGRADE Score

**Tabelle 2.7 –VASOGRADE Score**

VASOGRADE Kategorie	mFish-Entsprechung	WFNS-Entsprechung
GRÜN	mFish 1 oder 2	WFNS 1 oder 2
GELB	mFish 3 oder 4	WFNS 1-3
ROT	-	WFNS 4 oder 5

wird anhand einer Kombination aus dem mFish und dem WFNS Score erhoben (54). Er besteht aus den Kategorien Grün (mFish 1 oder 2, WFNS 1 oder 2), Gelb (mFish 3 oder 4, WFNS 1-3), oder Rot (WFNS 4 oder 5, egal welcher mFS).

### 2.2.4. Scores und Outcome-Vorhersage bei aSAB im Vergleich

Die Vielfalt existierender Scores, die den Schweregrad einer aSAB beschreiben und die Outcomevorhersage, sowie die Heterogenität der erhobenen Parameter und die recht starke Interobserver-Variabilität haben eine retrospektive vergleichende statistische Auswertung der Outcome-Vorhersagekapazitäten der wesentlichen radiologischen, klinischen und kombinierten Scores motiviert (45). Das Ergebnis zeigt, dass die Vorhersagekraft etablierter klinischer, radiologischer und kombinierter Scores für die Outcome-Vorhersage nach einer aSAB variieren von sehr niedrig (Fisher Score: AUC 0,55) bis moderat gut (Hunt und Hess sowie GCS: AUC 0,76). Die Leistungsfähigkeit der anderen Scores befindet sich zwischen diesen zwei Polen.

### 2.3. *Machine Learning Modelle: Erweiterung der Prädiktionskapazität?*

Die sogenannte Künstliche Intelligenz (KI, engl.: *Artificial Intelligence, AI*) ist ein Teilgebiet der Informatik, das sich u.a. mit dem maschinellen Lernen befasst. Das geführte sowie das selbstständige Anlernen von Algorithmen zur Durchführung diverser Aufgaben in wachstumsträchtigen Sektoren wie dem Finanzwesen, der Robotik und Automatisierung von Verfahren, sowie im Gesundheitswesen sind zentrale Bestreben der KI. Im Rahmen der vorliegenden Arbeit stellt sich die Frage nach dem Potenzial der KI –insbesondere von ML-Algorithmen – um die Outcomeprädiktion für aSAB zu verbessern. Einfache Prädiktionsmodelle beruhen auf der Grundlage von Regressionsmodellen. Diese können jedoch nur eine begrenzte Anzahl von Variablen sinnvoll testen. Durch ihre Kapazität komplexere, multidimensionale Variablen zu verarbeiten, versprechen ML-Algorithmen hingegen die Performance prädiktiver Modelle zu verbessern.

#### 2.3.1. *Was ist Künstliche Intelligenz?*

Als Teilgebiet der Informatik umfasst die KI u.a. die Entwicklung maschinellen Lernens sowie die Automatisierung intelligenten Verhaltens. Meist bezeichnet KI den Versuch, bestimmte Entscheidungsstrukturen des Menschen nachzubilden, indem z.B. ein Computer so gebaut und programmiert wird, dass er eigenständig Probleme bearbeiten kann (47). Ein Vergleich zum konventionellen Programmieren kann hilfreich sein, um den Anspruch der KI besser darzustellen. Konventionelles Programmieren kann nur bereits stringent organisierte und klassifizierte Informationen verarbeiten wie z.B. solche, die in einer Datenbank gespeichert sind. KI hingegen ist ein Zweig der Informatik, der angeleitetes oder sogar selbständiges, intelligentes Verhalten zu simulieren versucht. Typisch für KI-Systeme sind Algorithmen, die versuchen, Verhaltensweisen nachzuahmen, die der menschlichen Intelligenz zugeschrieben werden, wie z.B. Planen, Lernen, logisches Denken, Problemlösung, Wissensdarstellung, Wahrnehmung, Bewegung und Manipulation von Objekten, und in geringerem Ausmaß soziale Intelligenz und Kreativität (47). Das Themenfeld KI umfasst eine rasch wachsende Anzahl von Anwendungen in der heutigen Informationsgesellschaft (Musteranalyse und Mustererkennung in Suchmaschinen, Gesichtserkennung, der Erkennung von Fingerabdrücken, industrielle Qualitätskontrolle und Fertigungsautomatisierung in Kombination mit Robotik, Sprachsynthese, Spracherkennung; Mustervorhersage als Erweiterung der Mustererkennung, Robotik). Das Feld der KI lässt sich in einige Teilgebiete gliedern (47). Allen voran befasst sich das sogenannte Machine Learning damit,



große Datenmengen zu durchforsten, um induktiv Rückschlüsse über bislang unbekannte kausale Zusammenhänge oder Korrelationen zu ziehen.

### 2.3.2. Maschinelles Lernen und die drei Lernmodalitäten

Machine Learning (ML) ist ein Sammelbegriff für Algorithmen, die Daten durchforsten, von Datensätzen empirisch lernen, und dann die induktiv erlernten kausalen Zusammenhänge nutzen, um selbstständig fundierte Entscheidungen zu treffen (47). So ist es ML-Algorithmen nach einer Anlernphase auf der Grundlage von Datensätzen möglich, Probleme selbstständig zu lösen und adäquate Vorhersagen zu treffen. ML-Algorithmen erlauben auch eine Selbstoptimierung über die Zeit, brauchen dafür allerdings noch externe Orientierungshilfe (47).

ML-Algorithmen lassen sich entsprechend entlang ihrer Lernmodalität einteilen. So gibt es das supervidierte Lernen (Eng.: *supervised*), das nicht-supervidierte Lernen (Eng.: *unsupervised*), sowie das sog. bestärkende Lernen (Eng.: *reinforcement learning*). Supervidiertes Lernen beinhaltet die Klassifizierung von Daten innerhalb eines Datensets anhand von Klassifizierungsmarkern (Eng.: *label*). Diese Marker werden dann genutzt, um ein Klassifizierungsmodell zu erstellen. Dank dieses Klassifizierungsmodells können neu eingehende Daten benannt und ihren jeweiligen Klassen zugeordnet werden. Nicht-supervidiertes Lernen hingegen nutzt keine Klassifizierungsmarker deduktiv, sondern identifiziert anhand geteilter Charakteristika die Klassen innerhalb unstrukturierter Datensätze. Das sog. bestärkende Lernen, arbeitet wiederum mit einer Belohnungsfunktion, die erfolglose Aktionen bestraft und erfolgreiche Aktionen belohnt (59).

Klassifikationsmodelle nutzen Dateneingaben zu verschiedenen Variablen, um eine Klasse vorherzusagen. Formen der Klassifikation beinhalten Entscheidungsbäume, Support Vektor Maschinen (SVM), Logistische Regressionsanalyse oder Random Forest Modelle. Klassifikationsmodelle arbeiten auf der Basis von Lernvorgängen und von Trainingsdaten, um zu verstehen, welchen Bezug Inputdaten zu gesuchten Klassen oder Kategorien haben (59).

Ein Lernvorgang gestaltet sich typischerweise so, dass ein Lernalgorithmus genutzt wird, um die Parameter des gewünschten Modells zu entwickeln und zu definieren. Soll ein Modell z.B. eine Vorhersage treffen bezüglich der Wahrscheinlichkeit eines Herzinfarktes anhand von Richtig/Falsch-Aussagen, wird es mit Daten konfrontiert, die mit einem Herzinfarkt korrelieren,

dann mit Daten, die keine Korrelation aufweisen. Im Zuge zahlreicher Wiederholungen dieses Vorgangs adaptiert der Algorithmus interne Werte bis es erlernt hat, Datensätze, die einen Herzinfarkt vorhersagen, von denen zu unterscheiden, die dieses nicht tun.

Machine Learning nutzt in der Regel im Rahmen eines Lernprozesses einen Datensatz, der in drei Sets geteilt wird: ein Trainingsset, ein Validierungsset und ein Test-Set. Das Trainingsset beinhaltet die Daten, die zum Trainieren des Algorithmus genutzt werden. Das Validierungsset wird genutzt, um die Resultate der Trainingsphase zu validieren, und den Algorithmus präziser zu justieren. Das Testset beinhaltet Daten, die der Algorithmus noch nie verarbeitet hat und wird zu allerletzt genutzt, um die Generalisierungsfähigkeit des Lernverfahrens zu testen (59). Die Qualität des Modells kann abschließend entlang der Kategorien Richtigkeit (Eng.: *accuracy*), Präzision (Eng.: *precision*), Erinnerungsfähigkeit (Eng.: *recall*) beurteilt werden.

Beim ML unterscheidet man zwischen sog. symbolischen Systemen, in denen Wissen (sowohl als Beispiel als auch als induzierte Regel) explizit repräsentiert ist; und sub-symbolischen Systemen, wie z.B. neuronalen Netzwerken, denen zwar ein berechenbares Verhalten “antrainiert” wird, die allerdings keinen Einblick in die erlernten Lösungswege erlauben. Neuronale Netzwerke werden meist als Paradebeispiel des Deep Learnings aufgezählt.

### 2.3.3. Deep Learning und Künstliche Neuronale Netzwerke

Deep Learning ist ein spezialisiertes Teilgebiet des ML und besteht aus supervidierten Lernalgorithmen. Die wesentlichen Vertreter dieser Algorithmen sind u.a. sogenannte Support Vector Machines (SVM), Entscheidungsbäume, Bayes'sche Ansätze und künstliche neuronale Netzwerke (KNN). Die aktuellen state-of-the-art DL-Algorithmen beruhen meist auf der Grundlage von sog. geschichteten, künstlichen neuronalen Netzwerken, die menschliche Entscheidungsfindung simulieren. Ein KNN – auch oft einfach als neuronales Netzwerk bezeichnet - ist eine Ansammlung kleiner Recheneinheiten (Eng.: *computing units*), die als ‘Neurone’ bezeichnet werden und die Struktur und Funktionsweise von neuronalen Verbänden im Gehirn nachahmen. Jedes KNN besitzt eine Input-Schicht und eine Output-Schicht mit einer oder mehreren versteckten Schichten. Versteckte Schichten nehmen gewichtete Input-Variablen zur Verarbeitung an und produzieren ein Output durch eine Aktivierungsfunktion. Ein KNN, das aus

mehr als einer Schicht besteht, wird als tiefe neuronale Netze (eng.: *Deep Neural Network*) bezeichnet.

Deep Learning-Algorithmen verbinden also nicht direkt Input und Output-Variablen, sondern lassen diese durch mehrere Schichten neuronaler Recheneinheiten laufen. Neuronale Netzwerke bestehen also meist aus vielen überlagerten Schichten (*'deep'*), was die Nachverfolgung von Ergebnissen teilweise erschwert bis unmöglich macht. Die Entwickler des Algorithmus definieren Anzahl der Schichten und die Aufgaben, welche das Output einer Schicht mit der Input-Ebene der nächsten Schicht verbinden. In der Folge wird der Algorithmus mit großen Datenbeständen von gelabelten Daten konfrontiert (z.B. Bild + Name/ Bezeichnung). Der Algorithmus wird dann diese Beispiel-Datensätze durch die Schichten seines KNN verarbeiten und die Gewichtung der Variablen in jeder Schicht anpassen, um Verbindungs-/ Korrelationsmuster zu erkennen, die diejenigen Bilder mit ähnlichen Labels definieren.

Diese Einheiten scannen eingehende Daten und lernen über die Zeit, auf dieser Grundlage Entscheidungen zu treffen. Die Mehrschichtigkeit bewirkt u.a. auch, dass DL-Algorithmen mit der Zeit und zunehmendem Datenvolumen im Vergleich zu herkömmlichen ML-Algorithmen effizienter werden. Letztere tendieren eher dazu, mit zunehmender Datenlage in ihrem Lernverhalten zu stagnieren.

Deep Learning-Algorithmen können Informationen markieren und kategorisieren, und vermögen es auch, Muster zu erkennen. Dieser Algorithmustyp ermöglicht KI-Systemen, kontinuierlich zu lernen und die Qualität und Genauigkeit der Resultate durch eine Einschätzung zu verbessern, ob Entscheidungen korrekt waren oder nicht.

KNN lernen durch den sog. *Backpropagation*-Prozess. Dieser nutzt einen Trainingsdatensatz, der bekannte Inputdaten mit erwünschten Output-Daten verbindet. Als erstes werden Inputvariablen und erwünschte Output-Variablen definiert. Als zweites wird eine sog. Fehler-Funktion definiert, die festlegt, um wieviel die Output-Werte variieren dürfen, um noch als akzeptabel zu gelten. Sog. Perzeptrons – vom Prinzip her sehr einfache, lineare Klassifizierer – sind die ältesten und einfachsten Formen eines KNN. Es handelt sich hierbei um einschichtige neuronale Netzwerke, die aus Inputknoten mit sofortigem Bezug zum Output-Knoten bestehen (59).

Vereinfacht kann das lineare Perzeptron vorerst als lernender Agent konzipiert werden, d.h. als eine mathematische Funktion, die einen Merkmalsvektor auf einen Funktionswert abbildet (59). Das Perzeptron ist das einfachste neuronale Netzwerkmodell äquivalent zu Naive-Bayes, dem einfachsten Typ vom Bayes-Netz (59). Als eines der einfachsten neuronalen Lernverfahren kann das zweischichtige Perzeptron nur linear separable Klassen trennen (59). Mehrschichtige Netze sind hingegen wesentlich potenter. Der sog. Backpropagation Algorithmus, z.B., ist eine Verallgemeinerung des Perzeptrons, das mehrere Schichten verwendet und so linear trennbare Mengen teilen kann und somit auf der Grundlage einer verbesserten Lernregel operiert (59).

#### 2.3.4. KI-Anwendungen in der Medizin

Aktuell wird das Forschungsfeld v.a. von einem Wildwuchs an heterogenen klinischen Fragestellungen, abhängigen und unabhängigen Variablen und ausgewählten Algorithmen charakterisiert. Im Fachbereich Neurochirurgie wurden ML-Outcomeprädiktionsmodelle für einige relevante Krankheitsbilder getestet – bislang mit variabler prädiktiver Leistungsfähigkeit (AUC 0,71 - 0,96) (60). ML-basierte Outcome-Vorhersagen wurden u.a. bei sog. Flow-Diverter Implantationen (61) für sehr kleine Patient\*innenkohorten durchgeführt (61). Sie wurden ebenfalls für die Vorhersage von angiographischen Vasospasmen in einer Datenserie genutzt, die neurochirurgisch sowie interventionell versorgte Patient\*innen beinhaltete (62). Auch die radiologische Diagnostizierung intrakranieller Aneurysmata wurde u.a. durch Künstliche Neuronale Netzwerke und DL-Algorithmen in der Auswertung von CT-A und MR TOF verbessert (63,64). Daten zur Anwendung von Machine Learning-Algorithmen in der Outcome-Prädiktion von aSABs sind jedoch bis dato selten (10).

Zahlreiche Publikationen im Bereich der interventionellen Radiologie konzentrieren sich auf die Vorhersage von postiktalen angiographischen Vasospasmen und Sekundärischämien (*eng: delayed cerebral ischaemia*, DCI) nach kombinierten chirurgischen und endovaskulären Eingriffen (62). Einzelne Beiträge untersuchen die Anwendbarkeit von ML-Algorithmen zur Outcome-Vorhersage von ICBs (66) und traumatischen Hirnverletzungen (67).

Einige Arbeiten erstreben die Outcome-Vorhersage durch ML- (68) und DL- (69) Algorithmen bei Gliomen, sowie bei Patient\*innen mit Neuroblastomen (69) oder fokussieren sich auf die Optimierung der bildgebenden Diagnostik von Glioblastomen (70). Einzelne Beiträge

beschäftigen sich mit ML- und DL-Anwendungen im Bereich der Wirbelsäulenchirurgie und der Frage, welche Patient\*innendaten sinnvollerweise für die präoperative Planung sowie für die Outcome-Vorhersage genutzt werden können (71). Unter anderem wird hier die Frage gestellt, ob ein DL-Algorithmus zur Auswertung der präoperativen Bildgebung genutzt werden sollte, um Aussagen zur Outcome-Vorhersage zu treffen (72). Bisherige Arbeiten untersuchten die Fähigkeit von ML-Algorithmen, die Outcome-Vorhersage einer aSAB zu verbessern (73).

Vor dem Hintergrund lässt sich das hiesige Vorhaben der Testung diverser ML-Algorithmen für die Optimierung der Outcome Prädiktion von aSABs verstehen. Kann eine prädiktive Modellierung auf der Grundlage von ML- und DL-Algorithmen die Outcome-Vorhersage von Patient\*innen mit aSABs im Vergleich zu den gängigen klinischen, radiologischen und gemischten Scores verbessern?

## **I.3 Fragestellungen und Ziele der Arbeit**

### **3.1. Fragestellungen**

Kann durch die Anwendung von ML- und DL-Algorithmen die Outcome-Vorhersage bei der aSAB gegenüber den existierenden klinischen, radiologischen und gemischten Score-Modellen verbessert werden? Erreichen DL-Algorithmen eine gleiche oder bessere Performance als ML-Algorithmen in der Outcome-Vorhersage? Besteht eine unterschiedliche Relevanz einzelner Variablen innerhalb traditioneller Scores? Lässt sich der Einsatz von ML- und DL-Algorithmen zumindest anhand eines Vergleichs mit der Outcome-Vorhersage etablierter Scores in derselben Kohorte nachvollziehen?

### **3.2. Ziele der Arbeit**

Ziel der Arbeit ist es, die Vorhersagekraft von supervidierten ML- und DL-Algorithmen bezüglich des Outcomes nach einer aSAB auf Grundlage der bei Aufnahme routine-mäßig erhobenen Parameter zu testen (65). Verglichen wird die Prädiktionskraft von ML-Algorithmen vom Treeboosting Typ (*CatBoost treeboosting Algorithmus*), von traditionellen ML-Algorithmen (*L1 regularized Generalized Linear Model*, einer *Lasso logistic regression* gleichzustellen) und state-of-the-art DL-Algorithmen (künstlichen neuronalen Netzwerken wie z.B. dem *Multilayer Perceptron artificial neural net*), sowie etablierten klinischen Scores (klinische, radiologische und kombinierte klinisch-radiologische Scores).

## II. Material, Methodik und statistische Auswertung

### II.1. Material

#### 1.1. Patient\*innenpool und Outcomes

Die Studie wurde vom Ethikkomitee der Charité - Universitätsmedizin Berlin genehmigt (EA 1/291/14). Es wurden Patient\*innen eingeschlossen, die entweder CT-gestützt oder nach Lumbalpunktion eine nachgewiesene aSAB erlitten hatten. Klinische Scores wurden bei Aufnahme angewandt und radiographische Scores auf der Grundlage des Aufnahme-cCTs errechnet. Als primärer Endpunkt der Arbeit wurde das funktionale Outcome genutzt, das anhand der mRS dichotom definiert wurde.

Ein gutes Outcome wurde mit einem mRS-Wert von null bis zwei ( $\leq 2$ ), ein schlechtes Outcome mit einem mRS-Wert größer als zwei (mRS 3-6) bemessen. Das klinische Outcome wurde anhand der mRS-Erhebung im Rahmen von Kontrollvisiten 6-12 Monate nach der initialen aSAB definiert. Bei unzureichender Information zur Bestimmung des mRS erfolgten systematische Telefoninterviews. Beide Formen der mRS-Erhebung wurden für den initialen Schweregrad der aSAB verblindet. Der funktionale Outcome wurde nach 10 Monaten (Median) ausgewertet (Interquartile Range: 6-17). In der Summe befanden sich 46,7% der Patient\*innen (n=181) bei einem mRS 0-2 (*„günstig“*) und 53,3% (n=207) bei einem mRS 3-6 (*„ungünstig“*) (65).

#### 1.2. Behandlungsalgorithmen für Patient\*innen mit aneurysmatischer SAB

Der Behandlungsalgorithmus einer aSAB erfolgte in Anlehnung an die Empfehlungen der Leitlinien der American Heart Association/ American Stroke Association (2012) und der Europäischen Stroke Organization (2013) unter Berücksichtigung folgender Aspekte: Bei Aufnahme erfolgt initial ein cCT und eine CT-Angiographie (CTA) des Kopfes. Im Falle einer signifikanten Mittellinienverlagerung und/oder lichtstarrten, weiten Pupillen sowie einer suffizienten Lokalisation des Aneurysmas in der CTA erfolgt eine sofortige chirurgische Therapie. Bei allen anderen Patient\*innen erfolgt eine diagnostische digitale Subtraktionsangiografie (DSA) im Verlauf mit Diskussion der Therapieoptionen durch Chirurgen und interventionelle Radiologen in einem neurovaskulären Board. Eine bildgebende Verlaufskontrolle per cCT erfolgt in der Regel

nach einer Intervention. Patient\*innen werden dann intensivmedizinisch weiterbetreut. Im Falle der Notwendigkeit einer Fortführung von Sedierung, Intubation und Beatmung nach Intervention/OP, d.h. bei frustranen Extubationsversuchen oder schlechtem neurologischen Status, erfolgt in der Regel eine invasive Messung des intrakraniellen Drucks (ICP) mittels einer externen Ventrikeldrainage, sowie auch die Messung der intraparenchymatösen Sauerstoffsättigung (PtiO<sub>2</sub>). Tägliche transkranielle dopplersonographische (TCD) Messungen sind empfohlen, um frühzeitige Anzeichen von Vasospasmen zu identifizieren. Eine zweite DSA kann als Teil der klinischen Routine zwischen den postiktalen Tagen sechs und 12 durchgeführt werden, oder bei Anzeichen einer Vasospasmusphase im TCD, sowie bei einer neurologischen Verschlechterung oder einem signifikanten Abfall im PtiO<sub>2</sub>. Beim Auftreten von DCI oder Vasospasmen ist eine Euvolämie empfohlen, sowie die Erhöhung des mittleren arteriellen Drucks (*eng.*: Mean Arterial Pressure, MAP) auf zwischen 100 und 120mmHg anhand von Vasopressoren wie Noradrenalin unter strenger Beobachtung des kardialen Outputs und der cerebralen parenchymatösen Sauerstoffsättigung als Ersatzmarker für den zerebralen Blutfluss. Bei Vasospasmen wird eine Nimodipin-Therapie empfohlen (60mg alle 4h). Bei starkem Blutdruckabfall in Folge dieser initialen Dosierung kann eine reduzierte Dosierung von 30mg alle 4h weiter verabreicht werden. (34, 74–76)

## **II.2. Methodik**

### *2.1. Scores, Selektion einzelner Variablen und Modellierung*

Als Grundlage des Vergleichs von Scores in der Outcome-Vorhersage mit ML-Algorithmen erfolgte zunächst die Selektion relevanter Merkmale. Klinische, radiologische und kombinierte Scores wurden entsprechend den Referenzpublikationen auf die Primärdaten angewandt. Nach Aufnahme wurden der Hunt und Hess Score (s. Tabelle 2.2) und der GCS erhoben, sowie etwaige neurologische Defizite ermittelt. Der WFNS-Score (s. Tabelle 2.3) wurde auf der Grundlage des bereits erhobenen GCS errechnet. Die radiographischen Scores wurden *post hoc* von einem erfahrenen Neurochirurgen erhoben, der für das Outcome verblindet war. Der VASOGRADE Score (s. Tabelle 2.7) wurde errechnet auf der Basis der bereits erfolgten retrospektiven und prospektiven Datenauswertung und entsprechend des in der Literatur empfohlenen Vorgehens (54). Die Ergebnisse der verschiedenen klinischen, radiologischen und gemischten Scores sind als



Merkmale 1 bis 7 dargestellt (s. Tabelle 2.8) und bilden numerische Werte auf der Grundlage binärer Werteinträge.

Klinische Parameter, die als Teil des Variablen-Clusters ‚Merkmal 8‘ erhoben wurden (s. Tabelle 2.8), waren zudem Alter, Geschlecht, Pupillenstatus, sowie weitere radiographische Eigenschaften, wie das Vorhandensein einer intrazerebralen Blutung (ICB), einer intraventrikulären Blutung (IVB), eines Subduralthämatoms (SDH), sowie einer Mittellinienverlagerung (MLV) um mehr als 5mm. Die Größe des Aneurysmas und seine Lokalisation (dichotomisiert in hinteres und vorderes Stromgebiet) wurden anhand von CTA und/oder DSA kurz nach Aufnahme erhoben.

Merkmal 8 greift somit zentrale Variablen der gängigen Scores separat auf, und erlaubt sowohl die Testung dieses Variablenbündels als auch die individuelle Testung ihrer individuellen Outcome-Prädiktionskapazität um die Nachvollziehbarkeit der Testung der ML-Algorithmen zu gewährleisten.

**Tabelle 2.8 -- Überblick der Scores für die Variablen-Auswahl**

<b>Input-Feature</b>	<b>Ausprägung</b>
1	Wert des Hunt und Hess Score
2	Wert des WFNS-Score
3	Wert des Original Fisher Score
4	Wert des Modified Fisher Score
5	Wert des VASOGRADE Score
6	Wert des BNI-Score
7	Wert der GCS
8	Alter, GCS-Score, Geschlecht, Pupillenstatus, Intraventrikuläre Blutung, Intrazerebrale Blutung, Mittellinienverlagerung um >5mm, Subduralthämatom, Lokalisierung des Aneurysma im anterioren Stromgebiet, BNI-Score

Kategorische Variablen mit zwei oder mehr Kategorien wurden in binäre Variablen übersetzt, wenn nur wenige Kategorien zur Verfügung standen. Der Pupillenstatus wurde zum Beispiel dichotomisiert (‚beide Pupillen sind lichtreagibel‘ vs. ‚pathologische Lichtreaktion‘). Das

Vorhandensein einer ICB in der Bildgebung wurde in ‚ja/nein‘ dichotomisiert. Eine radiologisch bestimmte MLV wurde in ‚Verlagerung um mehr als 5mm: ja/nein‘ übersetzt. Die Lokalisation des Aneurysmas wurde in ‚anteriores Stromgebiet: ja/nein‘ dichotomisiert. So waren zuletzt alle Variablen entweder binär oder kontinuierlich (65).

## 2.2. Modellierung und angewandte Algorithmen

Insgesamt drei Modelle wurden für die sieben ML-Algorithmen antrainiert:

- eines für die einzelnen Scores (GCS, WFNS, Hunt und Hess, BNI, Fisher, modified Fisher, VASOGRADE);
- eines für die kombinierten Variablen des Merkmals 8 (‚combined features‘) (65), so dass wichtige Parameter der Patient\*innenaufnahme integriert werden konnten und
- eines zur Identifizierung der jeweiligen Outcome-Prädiktionskapazität der einzelnen Variablen aus Merkmal 8 (‚single features‘), um die Nachvollziehbarkeit zu gewährleisten (s. Tabelle 2.8 für das Set der Input Variablen für jedes getestete Modell).

## 2.3. Das Machine-Learning Framework

Die ML-Programme wurden mit Python geschrieben unter Verwendung von standardisierten ML-Bibliotheken. Der Code ist auf Github frei zugänglich.<sup>1</sup> Als Teil eines supervidierten ML-Algorithmus wurden bereits erwähnte klinische Parameter und Scores genutzt, um das endgültige Outcome von Patient\*innen mit aSAB anhand einer dichotomisierten mRS (mRS 0-2 ‚positives‘ Outcome, mRS 3-6 ‚negatives‘ Outcome) vorherzusagen. Diese Dichotomisierung erbrachte 181 positive (‚gutes‘ Outcome) und 207 negative (‚schlechtes‘ Outcome) Fälle. Trotz dieses leichten Gefälles ergab sich keinen Bedarf für eine ‚sub-sampling‘ Methode, die die Datenmenge für das Trainieren der Modelle reduziert hätte (65).

Sieben unterschiedliche Algorithmen wurden antrainiert:

1. Ein einfacher Generalized Linear Model (GLM) Algorithmus
2. Ein L1 Regularized Generalized Linear Model (Lasso) Algorithmus (entspricht einer Lasso Logistischen Regressionsanalyse)

---

<sup>1</sup> Siehe: <https://github.com/prediction2020/explainable-predictive-models>

3. Ein GLM Elastic Net (Elastic Net) Algorithmus plus einer zusätzlichen L2 Regularization
4. Ein Tree Boosting Algorithmus (Tree Boosting)
5. Ein Support Vector Machine Classifier (SVMC) Algorithmus
6. Ein 'Naive Bayes' Algorithmus (NB)
7. Ein Multilayer Perceptron (MLP), (künstliches Neuronales Netz)

Ein vollständiger technischer Überblick der Algorithmen und der Wichtung der einzelnen Variablen ist in der Open Source Veröffentlichung des angewandten Modells und auf GitHub einsehbar.

#### 2.4. Modell-Training und Validierung

Die vorhandenen Daten wurden nach dem Zufallsprinzip in Trainings- und Testsets in einem 4:1 Verhältnis eingeteilt. Für beide Sets wurden Mittelwert/Modalwert berechnet, sowie das Skalieren der Variablen (*feature scaling*) anhand der Z-Transformation/ der statistischen Standardisierung (*zero mean unit variance normalization*) basierend auf dem Trainingsset. Die Modelle wurden danach anhand einer zehnfachen Kreuzvalidierung optimiert. Das Vorgehen wurde insgesamt in 50 Runden wiederholt.

### **II.3. Statistische Auswertung**

#### 3.1. Statistische Auswertung der Outcomeprädiktion per AUROC-Analyse

Die Leistungsfähigkeit der Modelle wurde an den Testdaten anhand einer Receiver-Operating-Characteristic (ROC)-Analyse geprüft sowie durch die Messung der Fläche unter der Grenzwertoptimierungskurve (sog. *Area Under the Curve (AUC)*). Eine AUROC-Analyse kann als Optimierungsmethode genutzt werden, oder als Qualitätsmaß. In letzterem Sinne wird die AUROC-Analyse oft in Themenbereichen genutzt, die sich mit dem Abrufen von Information, z.B. aus Datenbanken befassen (eng.: *Information retrieval*). Dieses Fachgebiet der Informatik befasst sich mit der Optimierung computergestützter Suchvorgängen nach komplexen Inhalten. Bei einer AUROC-gestützten Qualitätsanalyse wird die ROC-Kurve für eine Stichprobe von Testwerten oder allen verfügbaren Testwerten berechnet, um die Qualität eines Sets von Testwerten unabhängig zu beurteilen. Die Grenzwertoptimierungskurve stellt die Sensitivitätswerte eines

Datensets auf der Y-Achse (richtig positiv, Trefferrate) gegen die Spezifität auf der X-Achse (falsch positiv, Fehlerrate) auf. Es erfolgt die Berechnung der Fläche unter der Grenzwertoptimierungskurve, der sog. AUC. Dieser Wert liegt zwischen null und eins, wobei 0,5 der schlechteste Wert ist. Eine ROC-Kurve nahe der Diagonalen ist a.e. das Ergebnis eines Zufallsprozesses, der eine Fläche von 0,5 hat. Eine optimale Kurve hat eine Fläche zwischen 0,5 und 1. Je mehr der AUC-Wert in Richtung 1,0 tendiert, umso größer ist die Performance des untersuchten Phänomens.

Bei den ML-Modellen, wurde als Maß für die Leistungsfähigkeit der Medianwert über die 50 Wiederholungsrunden genommen. Die Absolutwerte der errechneten Wichtungen der Variablen (*calculated feature importance score*) wurden skaliert, um eine vergleichbare Einstufung der Variablen für alle Modelle zu gewährleisten: Für jede der 50 Runden wurde die errechnete Wichtung der Scores neu auf ein [0,1] Spektrum skaliert, wobei die Summe stets gleich eins war. Daraufhin wurden für jede Variable die mittlere Abweichung und die Standardabweichung für jede Runde errechnet und als endgültiges Rating abgespeichert.

### 3.2. Multikollinearität

Die ausgewählten Variablen der klinischen, radiologischen und gemischten Scores, die in der Selektion 8 zusammengefasst wurden, erwiesen eine vernachlässigbare Multikollinearität mit Variance Inflation Faktoren (VIF)  $< 3.48$ . (Alter 1.08, GCS 1.81, Geschlecht 2.77, Pupillen Status 1.44, IVB 2.41, ICB 2.17, MLV 1.85, SDH 1.17, Lokalisation des Aneurysmas 3.47, BNI Score 1.24) (65). In der Regressionsanalyse benennt der Begriff der Multikollinearität eine Situation in der zwei oder mehrere unabhängige Variablen sehr stark miteinander korrelieren. Mit steigender Multikollinearität wird die Schätzung der Regressionskoeffizienten instabil und Aussagen zur Schätzung der Regressionskoeffizienten werden ungenau. Auch ist die Modellinterpretation nicht mehr eindeutig. Ein niedrige bis vernachlässigbare Multikollinearität bedeutet also, dass eine Modellinterpretation ohne Verzerrung möglich ist. Der sog. *Variance Inflation Factor* (VIF) bietet ein Maß dafür, ob - und wenn ja, wie stark – eine Multikollinearität zwischen den unabhängigen Variablen besteht. Wenn ein VIF-Wert gleich 1 ist, liegt keine Multikollinearität vor. Wenn ein VIF- Wert größer 1 ist besteht eine Korrelation der unabhängigen Variablen.

### III. Ergebnisse

Die Auswertung bezieht sich auf eine Kohorte von 388 Patient\*innen, die aufgrund einer aSAB in der Charité Berlin zwischen 2009 und 2015 behandelt wurden. Das mediane Alter betrug 54 Jahre (IQR 46; 63) und das Verhältnis von Frau zu Mann lag bei 2,3. Klinische und radiologische Daten der Patient\*innen mit aSAB sind in Tabelle 3.1 zusammengefasst.

Die Mehrheit der Aneurysmata (35,8%, n=136) im Patient\*innenpool lagen im vorderen Stromgebiet (ACA), gefolgt von Aneurysmata der Arteria cerebri media (MCA) (26,1%, n=99). Die Lokalisation der Aneurysmata war für 380 von 388 Patient\*innen verfügbar. Insgesamt 52 Patient\*innen (13,4%) erwiesen eine pathologische Pupillenreaktion bei Aufnahme. Eine pathologische Pupillenreaktion beschreibt eine Anisokorie und eine fehlende Lichtreaktion. Eine begleitende IVB erlitten insgesamt 44,3% (n=172) der Patient\*innen, während die aSAB in 32% der Fälle (n= 124) ebenfalls von einer ICB begleitet wurde, und in 6,5% (n=25) von einem SDH. Eine MLV um >5mm lag bei 23,1% (n=89) vor. Die häufigste Dicke der aSAB (relevant für den BNI) lag in 32% der Fälle (n=124) bei 15-20mm (BNI 4°), die zweithäufigste mit 29,9% (n= 116) bei 11-15mm (BNI 3°).

Eine frühzeitige Versorgung des Aneurysmas erfuhren insgesamt 57,9% (n=220) der Patient\*innen durch Coiling und 27,9% der Patient\*innen (n=106) durch Clipping. Ausschließlich klinisch/symptomatisch behandelt oder anhand von invasivem Neuromonitoring auf Intensivstation versorgt wurden insgesamt 2,6% (n=10) der Patient\*innen. Insgesamt 10,9% (n=25) erfuhren keinerlei Behandlung. Von den selektierten Patient\*innen hatten 50% einen GCS von 13, 14 oder 15 Punkten (n=194). Insgesamt 32,3% (n=125) der Patient\*innen hatten den niedrigsten GCS von 3 Punkten. Fast ein Drittel der Patient\*innen (29,4%; n=114) hatte den höchsten Wert auf der Hunt und Hess Skala (Kategorie V). Die zweitgrößte Anzahl von Patient\*innen fiel in die Kategorie I (24,2%, n=94). Der Großteil der Patient\*innen fiel in die WFNS Kategorie I (36,6%, n=142) und V (n=146). In der Kategorie IV waren es lediglich 12,6% (n=49), in Kategorie III waren es 3,4% (n=13) und in Kategorie II 9,8% (n=38).

Die Kategorie des mFS, die am häufigsten vertreten war, war die Kategorie 4 mit 50,5% (n=196), gefolgt von Kategorie 3 (26,8%, n= 104) und Kategorie 1 (12,1%, n=47). Die am meisten vertretene Kategorie der VASOGRADE Skala war die Kategorie ‚Rot‘ mit 50% (n=194), gefolgt von der Kategorie ‚Gelb‘ (34,1%, n=132).

Insgesamt 53,9% (n= 209) der Patient\*innen konnten entsprechend ihrer Platzierung auf der mRS mit einem ‚guten Outcome‘ (mRS 0-2) klassifiziert werden, 46,1% der Patient\*innen (n= 179) hingegen als ‚schlechtes Outcome‘ (mRS 3-6).

**Tabelle 3.1 - Klinische und radiografische Daten der Patient\*innen mit aSAB**

			<b>% (n)</b>
<b>Klinische Merkmale</b>	Pathol. Pupillenreaktion GCS bei Aufnahme	3	13,4% (52)
		4-8	32,3% (125)
		9-12	8,7% (34)
		13-15	9,0% (35)
			50,0% (194)
<b>Klinische Scores</b>	WFNS	I	36,6% (142)
		II	9,8% (38)
		III	3,4% (13)
		IV	12,6% (49)
		V	37,6% (146)
	Hunt und Hess	I	24,2% (94)
		II	17,5% (68)
		III	14,7% (57)
		IV	14,25 (55)
		V	29,4% (114)
<b>Radiograf. Merkmale</b>	IVB		44,3% (172)
	ICB		32,0% (124)
	SDH		6,5% (25)
	MLV (>= 5mm)		23,1% (89)
	SAB-Dicke (BNI)	<5mm (1°)	6,4% (25)
		6-10mm (2°)	16,0% (62)
		11-15 (3°)	29,9% (116)
		15-20mm (4°)	32,0% (124)
		>25mm (5°)	15,7% (61)
	Lokalisation Aneurysma	ACA	35,8% (136)
		ACI	19,2% (73)
ACM		26,1% (99)	
Hintere Strombahn		18,9% (72)	
<b>Radiograf.Scores</b>	Mod. Fisher	0	4,6% (18)
		1	12,1% (47)
		2	5,9% (23)
		3	26,8% (104)
		4	50,5% (196)
<b>Combined Scores</b>	VASOGRADE	Green	15,9% (62)
		Yellow	34,1% (132)
		Red	50,0% (194)
<b>Outcome</b>	Positiv	mRS 0	22,2% (86)
		mRS 1	18,3% (71)
		mRS 2	6,2% (24)
	Negativ	mRS 3	7,2% (28)
		mRS 4	7,7% (30)
		mRS 5	4,9% (19)
		mRS 6	33,5% (130)
<b>Therapie</b>	Coiling		57,9% (220)
	Clipping		27,9% (106)
	andere		2,6% (10)
	keine		10,9% (25)

*IVB – Intraventriculäre Blutung, ICB – Intracerebrale Blutung; SDH – Subduralhämatom; ACA – Arteria Cerebri Anterior; ACI – Arteria Carotis Interna; ACM – Arteria Cerebri Media*

### III.1. Outcome-Prädiktionskraft von klinischen, radiologischen und gemischten Scores im Vergleich – Ergebnisse der ersten AUROC Analyse

Um die Nachvollziehbarkeit der Auswertung von ML-Algorithmen zu ermöglichen, erfolgte zunächst eine statistische AUROC-Analyse und Auswertung der klinischen, radiologischen und gemischten Scores anhand ihrer jeweiligen AUC-Werte in Bezug auf ihre Vorhersagekraft eines negativen oder positiven Outcomes im Sinne einer dichotomisierten mRS. Die Ergebnisse sind in Tabelle 3.2. zusammengefasst.

**Tabelle 3.2 – Prädiktive Performance aller Scores und kombinierter Variablen für ‚schlechtes Patient\*innen-Outcome‘ (modified Rankin Scale 3-6) gemessen per Area Under the Curve**

	Merkmale	> mRS/ pos. Q	< mRS/ neg. Q	fehlend	eAUC	r (stn. Fehler)	Konfidenz Intervall
<b>Klinische Scores</b>	GCS	181	207	20	<b>0,752</b>	0,025	[0,703 - 0,802]
	Hunt and Hess	181	207	20	<b>0,750</b>	0,025	[0,701 - 0,799]
	WFNS	181	207	20	<b>0,740</b>	0,026	[0,689 - 0,790]
<b>Radiografische Scores</b>	Original Fisher	181	207	20	<b>0,553</b>	0,029	[0,495 - 0,610]
	Modified Fisher	181	207	20	<b>0,645</b>	0,028	[0,590 - 0,700]
	BNI	181	207	20	<b>0,620</b>	0,028	[0,565 - 0,675]
<b>Gemischter Score</b>	VASOGRADE	180	204	24	<b>0,722</b>	0,026	[0,671 - 0,774]

Bildlegende: **Q:** Outcome; **fehlend:** Benutzerdefinierte fehlende Werte gelten als fehlend. Die Statistik basiert auf allen Fällen mit gültigen Daten für alle Variablen in der Analyse; **mRS** – modified Rankin Scale; **eAUC** = 'einfache AUC'; **GCS** – Glasgow Coma Scale; **WFNS** – World Federation of Neurosurgical Societies Scale; **BNI** – Barrow Neurological Institute Index;



### *1.1. Outcome-Vorhersage klinischer und radiologischer Scores*

Der Vergleich der Prädiktionskraft klinischer und radiologischer Scores gemessen an dem Verhältnis ihrer Sensitivität und Spezifität in der AUROC-Analyse ergab folgende Ergebnisse: Die klinischen Scores (GCS, Hunt und Hess [HH], WFNS) zeigten eine mäßig gute und deutlich größere Prädiktionskraft als die radiografischen Scores (original Fisher [Fish], modified Fisher [mFish], BNI). So umschreiben die AUC-Kurven der klinischen Scores (AUC-HH: 0,750; AUC-GCS: 0,752; AUC-WFNS: 0,740) eine deutlich größere Fläche als diejenigen der radiografischen Scores (Fisher score (AUC-Fish 0,553); modified Fisher score (AUC-mFish: 0,645); Barrow Neurological Index (AUC-BNI: 0,620)) – ein Hinweis auf eine größere Performance in der Prädiktion eines negativen Outcomes im Sinne der dichotomisierten modifizierten Rankin-Scale ( $mRS \leq 3-6$ ) (s. Abb. 3.1 - 3.6, Vgl. Tab. 3.2).

Weiterhin ist ebenfalls zu beobachten, dass die insgesamt abgeflachten ROC-Kurven der radiografischen Scores im Vergleich zu den ROC-Kurven der klinischen Scores auf eine Prädiktionskapazität hinweisen, die einem Zufallsprozess nahekommt (s. Abb. 3.4, 3.5, 3.6). Eine ROC-Kurve, die sich der Diagonalen zwischen Sensitivität und Spezifität nähert, deutet in der Tat auf einen Zufallsprozess (eine ähnlich bis gleiche Treffer- und Fehlerrate entsprechen in etwa der zu erwartenden Trefferhäufigkeit eines Zufallsprozesses).

1.1.1. ROC-Kurven der klinischen Scores

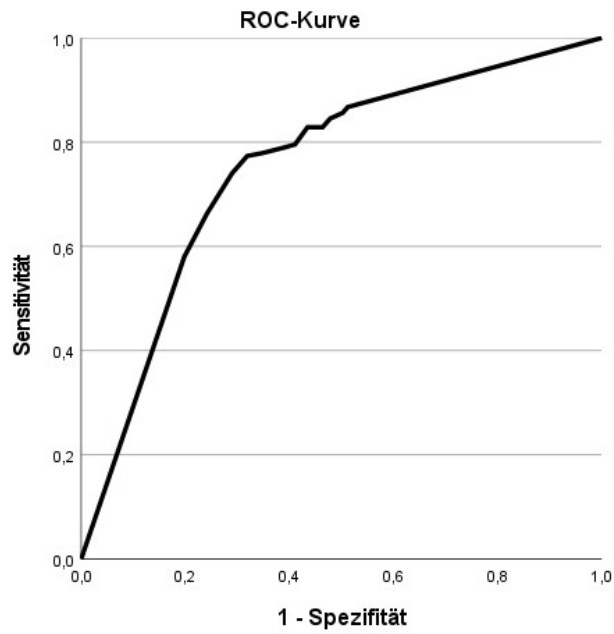


Abb. 3.1: ROC-Kurve für die Glasgow Coma Scale

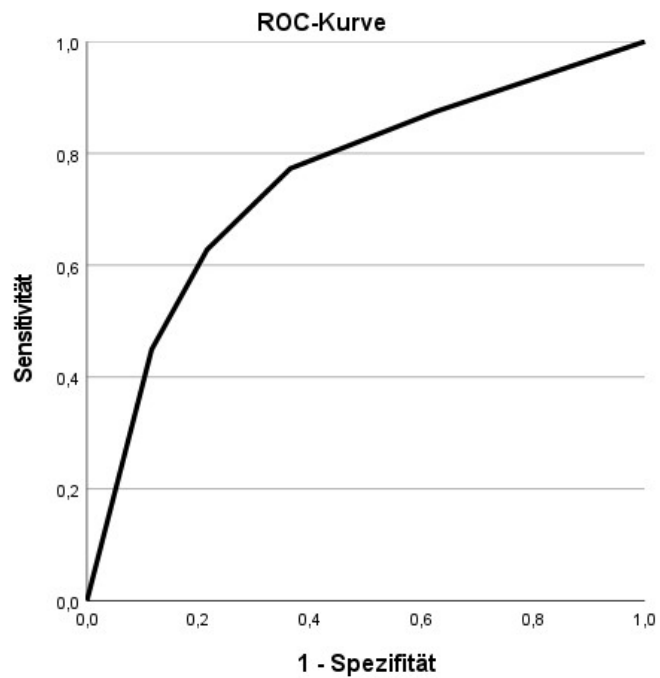
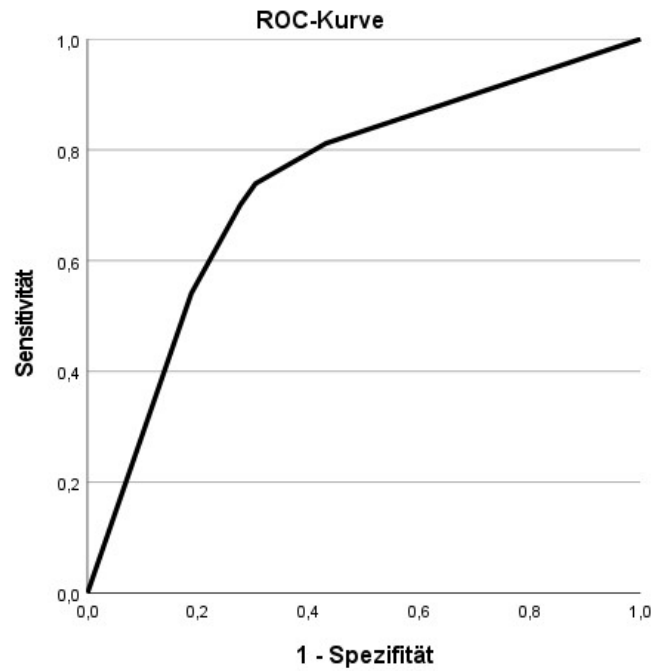


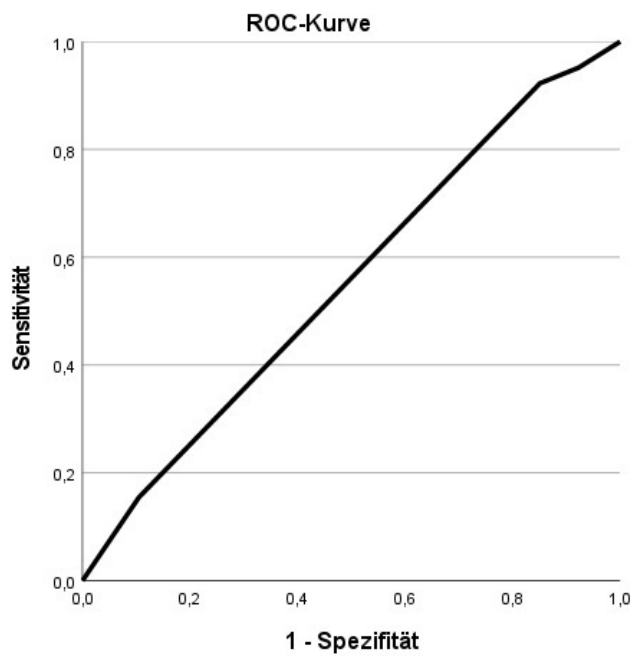
Abb. 3.2: ROC-Kurve für den Hunt und Hess Score



Diagonale Segmente ergeben sich aus Bindungen.

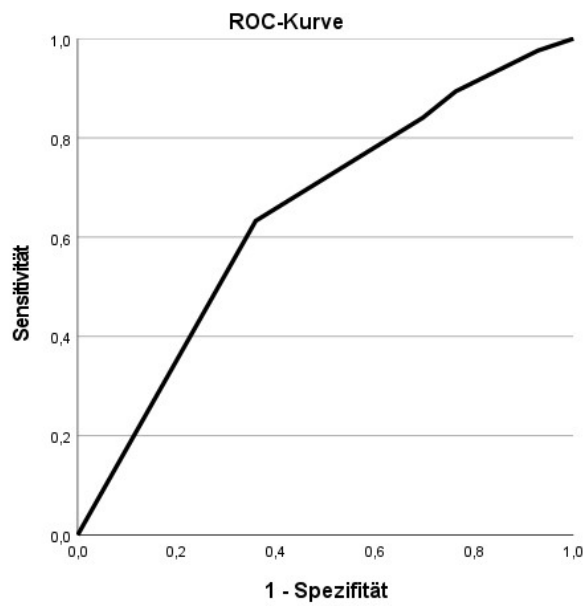
Abb. 3.3: ROC-Kurve für den WFNS-Score

1.1.2. ROC-Kurven der radiografischen Scores



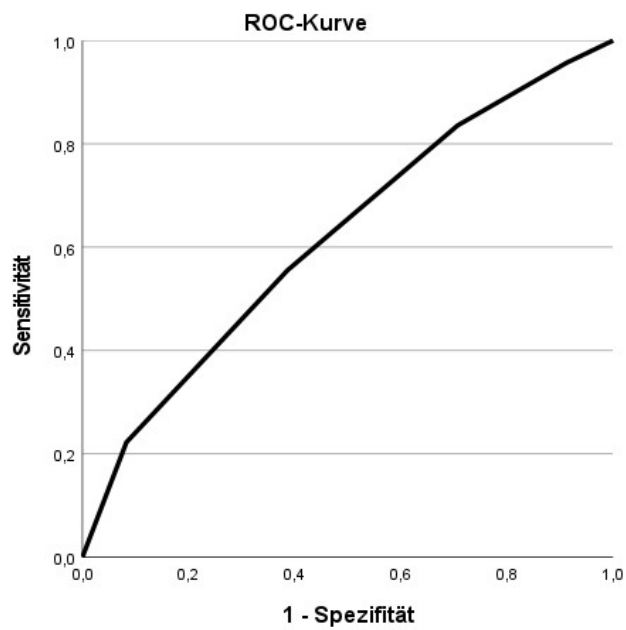
Diagonale Segmente ergeben sich aus Bindungen.

Abb. 3.4: ROC-Kurve für den originalen Fisher Score



Diagonale Segmente ergeben sich aus Bindungen.

**Abb. 3.5: ROC-Kurve für den modifizierten Fisher Score**



Diagonale Segmente ergeben sich aus Bindungen.

**Abb. 3.6: ROC-Kurve für den BNI-Score**

### 1.2. Outcome-Vorhersage gemischter Scores

Gemischte Scores waren den klinischen Scores im klinischen Alltag *nicht* überlegen; alleine eine Kombination aus klinischen, radiologischen *und* gemischten Scores bot ein verlässliches prädiktives Bild (45, 77).

Der VASOGRADE Score zeigte eine gute Vorhersagekraft bezüglich des Endpunkts ‚*negatives Patient\*innen-Outcome*‘ (AUC-VG: 0,722, s. Abb.3.7, Vgl. Tab. 3.2), war allerdings weder dem GCS (AUC-GCS: 0,752), dem Hunt und Hess Score (AUC-HH: 0,75) noch dem WFNS (AUC-WFNS: 0,740) überlegen.

Klinische Scores waren somit einem kombinierten Score wie dem VASOGRADE (VG) durchweg leicht überlegen.

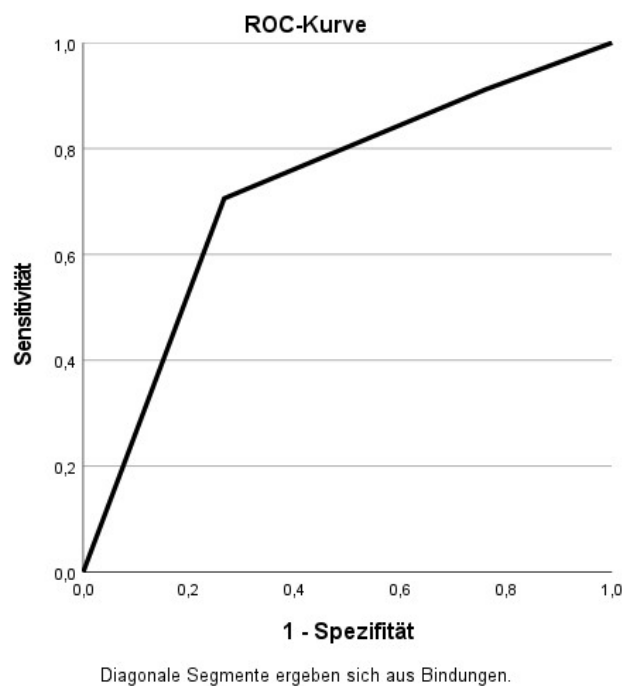


Abb. 3.7: ROC-Kurve für den VASOGRADE Score

### **III.2. Outcome-Vorhersage von Machine Learning-Algorithmen im Vergleich – Ergebnisse der zweiten AUROC-Analyse**

#### **2.1. Kann durch die Anwendung von ML-Algorithmen die Outcome-Vorhersage bei der aSAB gegenüber den existierenden Score-Modellen verbessert werden?**

Nach der ersten erfolgte nun eine zweite AUROC-Analyse der Performance der selektierten ML-Algorithmen basierend auf den etablierten Scores in derselben Patient\*innenkohorte. Die in der ersten statistischen Auswertung erhobenen Ergebnisse ließen sich ebenfalls in den ML-Analysen der Prädiktionskraft verschiedener klinischer und radiologischer Scores nachweisen (s. Tabelle 3.3, Abb. 3.8. A): Gemessen an ihrer AUC in der AUROC-Analyse boten klinische Scores (HH, GCS, WFNS) weiterhin im Schnitt eine stärkere prädiktive Performance bezüglich der Outcome-Vorhersage als radiologische Scores (Fisher score, modified Fisher score, Barrow Neurological Index). Klinische Scores waren den kombinierten Scores weiterhin im Durchschnitt in etwa ebenbürtig bis überlegen.

Eindeutig beste prädiktive Performance für ein ‚ungünstiges‘ Patient\*innen-Outcome (mRS 3-6) bietet erneut laut allen ML-Algorithmen das ‚Merkmal 8‘ – das Merkmalbündel, das Alter, GCS, Geschlecht, Pupillenstatus, die Anwesenheit einer IVB, ICB, SDH oder einer MLV um >5mm erhob, sowie den BNI-Score und ob eine Blutung im anterioren Stromgebiet stattfand. ‚Merkmal 8‘ erweist eine AUC von 0,78 für das Tree Boosting Modell und eine AUC von 0,77 für das GLM, das Lasso, das Elastic Net, das MLP und das SVMC-Modell. Das Naive Bayes Modell erteilt ‚Merkmal 8‘ die schlechteste AUC von 0,75.

Zweitbeste Performance in Puncto Prädiktionskapazität kommen dem GCS zu Gute mit einer AUC von 0,76 für sechs von sieben der Algorithmen (GLM, Lasso, Tree Boosting, MLP, SVMC, NB). Die GCS wird dicht gefolgt von der Hunt und Hess Skala, der vier von sieben Algorithmen ebenfalls eine AUC von 0,76 errechnen (GLM, Tree Boosting, MLP, NB). Dem WFNS-Score wird einheitlich mit einem AUC von 0,74 die vierte Stelle in Puncto prädiktive Performance eines schlechten Outcomes von allen sieben Algorithmen erteilt. Sechs von sieben Algorithmen errechnen dem VASOGRADE einen AUC von 0,72, der auch der letzte Score mit einem AUC > 0,70 ist (GLM, Lasso, Elastic Net, MLP, SVMC, NB). An sechster Stelle befindet sich der modified Fisher Score, dem fünf der sieben Modelle einen AUC von 0,65 erteilten (GLM, Lasso,

Tree Boosting, MLP, NB). An siebter Stelle befindet sich der BNI mit einem AUC von 0,63 für das GLM, dem Lasso Modell sowie dem Naive Bayes Modell und einem AUC von 0,62 für das Tree Boosting Modell, das MLP sowie das SVMC-Modell. An letzter Stelle befindet sich der original Fisher Score, der von vier Modellen eine AUC von 0,55 erteilt bekommt, vom SVMC sogar einen AUC von nur 0,47 (GLM, Lasso, Tree Boosting, MLP).

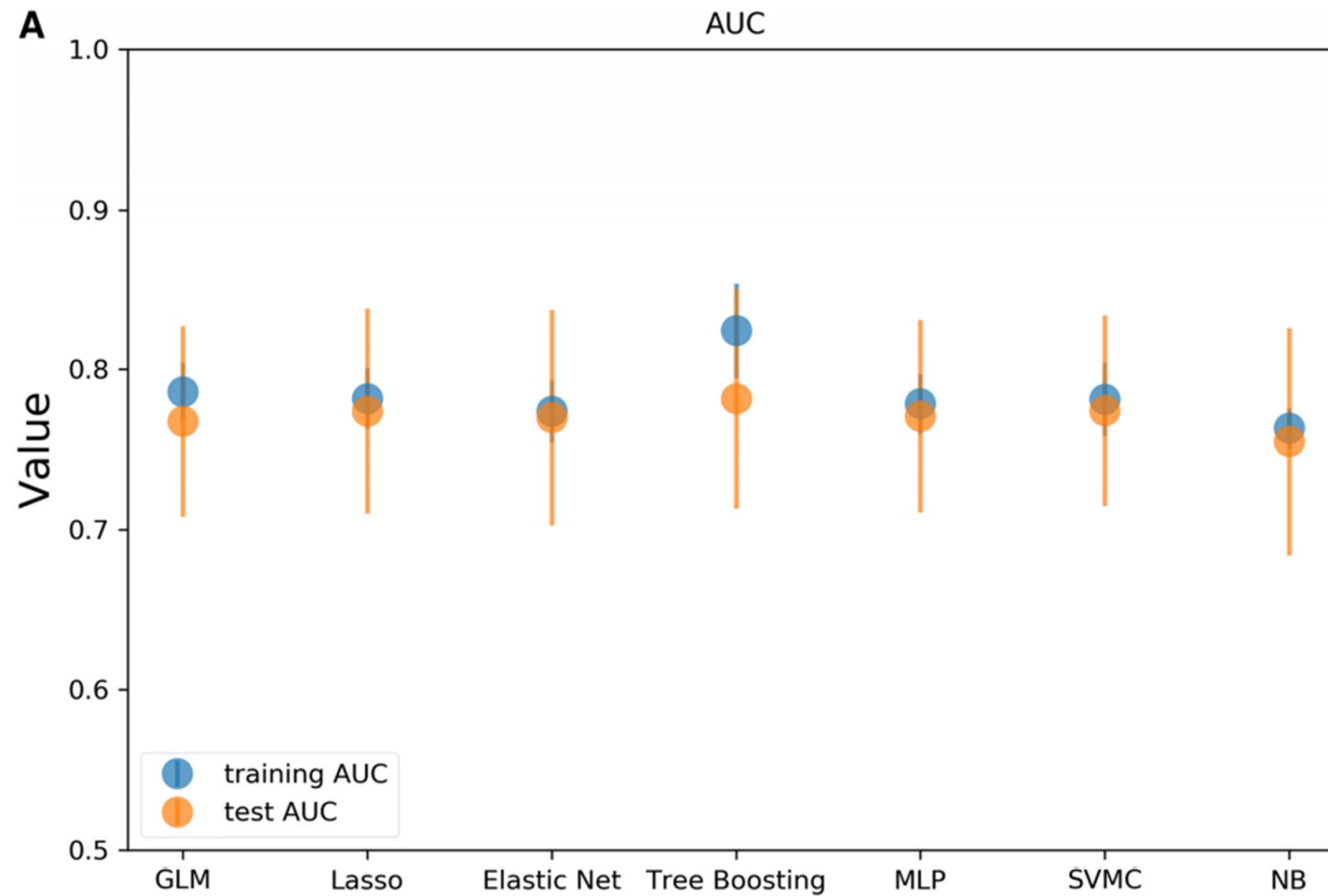
**Tabelle 3.3 – Prädiktive Performance der klinischen, radiologischen und kombinierten Scores, sowie kombinierter Variablenset in den Machine Learning Modellen für ‚ungünstiges Patient\*inn-Outcome‘ (mRS 3-6) gemessen per AUC**

<b>Merkmale</b>	<b>GLM</b>	<b>Lasso</b>	<b>Elastic Net</b>	<b>Tree Boosting</b>	<b>MLP</b>	<b>SVMC</b>	<b>NB</b>
<b>Hunt and Hess</b>	0,75/0,76 (0,07)	0,75/0,75 (0,08)	0,75/0,75 (0,08)	0,75/0,76 (0,07)	0,75/0,76 (0,07)	0,75/0,75 (0,07)	0,75/0,76 (0,07)
<b>WFNS</b>	0,74/0,74 (0,04)	0,74/0,74 (0,04)	0,73/0,74 (0,09)	0,74/0,74 (0,04)	0,74/0,74 (0,04)	0,74/0,74 (0,05)	0,74/0,74 (0,04)
<b>Modified Fisher</b>	0,65/0,65 (0,07)	0,65/0,65 (0,07)	0,64/0,62 (0,15)	0,65/0,65 (0,07)	0,64/0,65 (0,07)	0,65/0,64 (0,07)	0,65/0,65 (0,07)
<b>Original Fisher</b>	0,55/0,55 (0,04)	0,55/0,55 (0,05)	0,55/0,52 (0,11)	0,55/0,55 (0,06)	0,55/0,55 (0,04)	0,49/0,47 (0,11)	0,55/0,54 (0,08)
<b>VASOGRADE</b>	0,72/0,72 (0,07)	0,72/0,72 (0,07)	0,72/0,72 (0,09)	0,72/0,71 (0,06)	0,72/0,72 (0,06)	0,72/0,72 (0,07)	0,72/0,72 (0,07)
<b>BNI</b>	0,62/0,63 (0,06)	0,62/0,63 (0,06)	0,61/0,60 (0,15)	0,62/0,62 (0,07)	0,62/0,62 (0,07)	0,62/0,62 (0,08)	0,62/0,63 (0,06)
<b>GCS</b>	0,75/0,76 (0,05)	0,75/0,76 (0,05)	0,75/0,75 (0,07)	0,76/0,76 (0,06)	0,75/0,76 (0,06)	0,75/0,76 (0,05)	0,75/0,76 (0,05)
<b>Alter, GCS score, Geschlecht, Pupillenstatus, IVB, ICB, MLV um &gt;5mm, SDH, Lokalisierung der Blutung im anterioren Stromgebiet, BNI-Score</b>	0,79/0,77 (0,06)	0,78/0,77 (0,06)	0,77/0,77 (0,07)	0,82/0,78 (0,07)	0,78/0,77 (0,06)	0,78/0,77 (0,06)	0,76/0,75 (0,07)

*AUC für das training-/ und test set, sowie die Interquartile Range (IQR) für das Test Set (in Klammern) über 50 Iterationen*



Abbildung 3.8.A –Area Under the Curve der ML-Algorithmen: Darstellung der Test und Training-Sets



Der höchste AUC-Test betrug 0.78 für das Tree-Boosting Modell, mit Ausnahme des Naive Bayes Modell (0.75). Die restlichen Modelle hatten einen test-AUC von 0.77. Einen größeren Unterschied zwischen Training und Test-Set konnte für das Tree Boosting Modell beobachtet werden, was als Hinweis auf ein sog. ‚overfitting‘ gedeutet werden kann. **Quelle:** Dengler, N.F., Madai, V.I., Unteroberdörster, M., Zihni, E., Brune, SC., Hilbert, A., Livne, M., Wolf, S., Vajkoczky, P., Frey, D., Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinic-radiological scores, Neurosurgical Review, 1.December 2020.

## 2.2. Erreichen DL-Algorithmen eine gleiche oder bessere Performance als ML-Algorithmen in der Outcome-Vorhersage?

Die Performance der ML-Modelle (Tree Boosting und MLP) war vergleichbar mit traditionellen GLM-Methoden (Lasso). Die Auswertung ergab homogene Ergebnisse und somit keinen beobachtbaren Unterschied in der Performance zwischen state-of-the-art DL-Algorithmus-Anwendungen (Tree Boosting und MLP) und ML-Algorithmus-Anwendungen (Lasso) auf der Grundlage traditioneller klinisch-radiologischer Variablen, die bei Patient\*innenaufnahme standardisiert erhoben wurden (s. Tabelle 3.3).

## 2.3. Besteht eine unterschiedliche Relevanz einzelner Variablen vom ‚Merkmal 8‘?

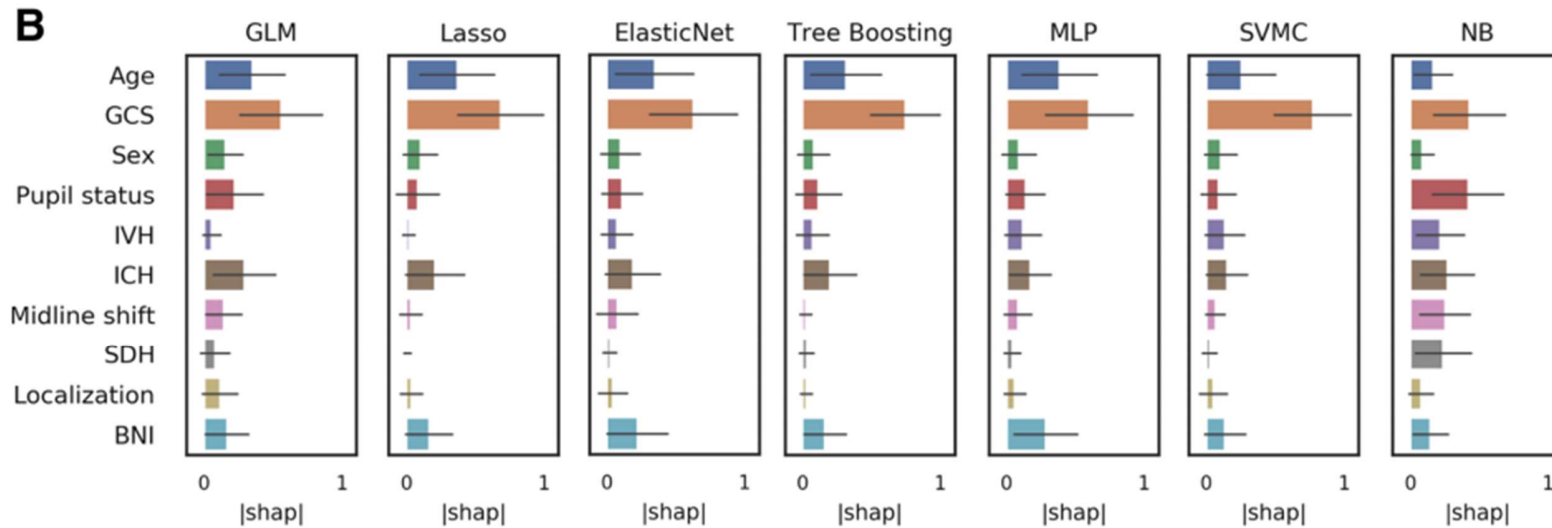
Das kombinierte Set klinisch-radiologischer Variablen (Merkmal 8) erhob Werte für einzelne Schlüsselvariablen (Alter, GCS, Geschlecht, Pupillenstatus, IVB, ICB, MLV>5mm, SDH, vorderes Stromgebiet, BNI), um jenseits der existierenden Scores ihre Wertigkeit als Variablenbündel sowie ihre individuelle, relative Wertigkeit zu testen. Nach den bereits dargelegten Auswertungen des sog. kombinierten Modells (AUROC Analyse der Outcome-Vorhersagekapazität der Scores, sowie der ML-Algorithmen) erfolgte eine Auswertung der Outcome-Vorhersagekapazität der einzelnen Variablen, die in ‚Merkmal 8‘ beinhaltet sind (sog. ‚Single Score‘ Modell).

Eine eindeutige Überlegenheit des kombinierten Modells über das Single Score Modell zeigte sich nicht. Die GCS wurde durch das Variablen-Rating (*feature importance rating*) als wichtigste Variable identifiziert (Abb.3.8.B). Die zweitwichtigste Variable war über alle Scores hinweg das Alter. Die Modelle schrieben auch dem BNI-Score und dem Auftreten einer intrazerebralen Blutung Relevanz zu.

Das Naive Bayes Modell war das einzige, das dem Pupillenstatus sehr hohe Relevanz zuschrieb. Im Lasso Modell waren die zweit- und drittwichtigsten Variablen die Anwesenheit einer ICB und das Alter. Für das Tree Boosting Modell waren dies jeweils das Vorliegen einer ICB und der BNI-Score. Für das MLP-Modell waren zweit- und drittwichtigste Variable das Alter und die Lokalisierung des Aneurysmas.

Der GCS und das Alter wurden als wichtigste Variablen innerhalb der Variablenkombinationen identifiziert. Schlussfolgernd ergab sich aus diesem ersten Forschungsschritt, dass ML- und DL-Anwendungen den klassischen klinischen Scores in Punkto Prädiktion von funktionellem Outcome nach aSAB nicht überlegen waren. Es zeigte sich kein relevanter Unterschied in der Performance von einer Kombination aus klinisch-radiologischen Faktoren (AUC 0.78) und der GCS (AUC 0.77).

Abbildung 3.8.B - Grafische Darstellung der Leistung und Variablen Rating für das klinisch-radiologische Modell



Das Ranking der Wichtung einzelner Variablen identifizierte durchgehend die GCS als den wichtigsten Faktor. Die GCS erreicht ein AUC von 0.76.  
**Quelle:** Dengler, N.F., Madai, V.I., Unteroberdörster, M., Zihni, E., Brune, SC., Hilbert, A., Livne, M., Wolf, S., Vajkoczky, P., Frey, D., Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinic-radiological scores, Neurosurgical Review, 1.December 2020.

#### **IV. Diskussion**

Das Hauptergebnis unserer Studie ist, dass sich bei der Genauigkeit der Vorhersage des Behandlungsergebnisses nach einer aSAB kein relevanter Unterschied zwischen ML-Algorithmen (die auf klassische klinische und radiografische Daten zugreifen) und etablierten klinischen und radiografischen Scores zeigt. Selbst ein Kombinationsdatenset aus klinischen und radiologischen Merkmalen bei Aufnahme zeigte keinen Vorteil gegenüber dem traditionell verwendeten GCS. Die besten Outcome-Prädiktoren waren der GCS-Wert selbst und das Patient\*innenalter.

Mit einer Fallzahl von 388 Patient\*innen rangiert unsere Studie im Mittelfeld ähnlicher Arbeiten (78). Kleinere Studien erfolgten bereits mit einer Fallzahl von 147 Patient\*innen (3), während die aktuell größte aSAB-Outcome-bezogene Studie mit einer Fallzahl von 1866 Patient\*innen auf dem landesweiten Schweizer aSAB-Patient\*innenregister beruht (79).

Die meisten Forschungsvorhaben, die sich auf die funktionale Outcome-Vorhersage einer aSAB fokussieren, basierten ihre Arbeit auf ähnliche Variablen wie wir. Diese sind das Alter der Patient\*innen, der WFNS Score, der Fisher Score, das Geschlecht, die Lokalisation des Aneurysmas, sowie dessen Größe, die Anwesenheit einer ICB oder IVB (3, 79–81). Einige Studien testeten ebenfalls die Relevanz einzelner radiologischer Variablen, wie z.B. die Mean Transit Time (MTT) und Time-to-Maximum (Tmax) einer CT Perfusionsbildgebung (3). Bei der MTT handelt es sich um die durchschnittliche Zeit in Sekunden, in der Blut - oder ein Bestandteil dessen, wie z.B. ein Erythrozyt, innerhalb eines Gefäßes in einem spezifischen Hirnareal verbringt. Time-to-Maximum ist ein Parameter der im Falle einer Perfusions-CT die Zeit beschreibt zwischen Bolusweg des Kontrastmittels von einem großen proximalen arteriellen Hirngefäß zum Hirnparenchym. Beide Werte werden z.B. in der Auswertung von Perfusions-CTs als Marker eines pathologischen Vorgangs herangezogen, so z.B., wenn sie erhöht sind.

Unsere Arbeit bietet die bislang umfassendste Testung aller gebräuchlichen klinischen (GCS, Hunt und Hess Skala, WFNS-Skala), radiologischen (Fischer, mFisher, BNI-Skala) und kombiniert klinisch-radiologischen Scores (VASOGRADE) sowie die separate, einzelne Testung der wichtigsten klinischen Parameter (Alter, Geschlecht, Pupillenstatus, IVB, ICB, MLV um >5mm, SDH, Lokalisierung der Blutung im anterioren Stromgebiet).

### 1. Machine Learning vs. etablierte Scores

Um eine Baseline zur Vergleichbarkeit vorbestehender Scores mit ML-Algorithmen herzustellen, nutzte die hier vorliegende Arbeit etablierte Scoring Systeme, sowie die Primärdaten, die für deren Erhebung notwendig ist. In einem ersten Schritt erfolgte eine AUROC-Analyse gängiger klinischer, radiologischer und gemischter Scores. In einem zweiten Schritt wandten wir nun eine Vielzahl von ML-Methoden am selben Datenset an und erreichten vergleichsweise ähnliche Ergebnisse in Punkto prädiktive Leistung, Sensitivität und Spezifität, aber auch einige überraschende Ergebnisse bezüglich der Relevanz einzelner Variablen.

Die prädiktive Leistung des Hunt und Hess Scores war in der AUROC-Analyse der ML-Algorithmen die stärkste (AUC-Range 0,76 - 0,75) und spiegelte die prädiktive Leistung in der einfachen statistischen Auswertung wider (eAUC= 0,75). Der AUC-Wert für den Hunt und Hess Score betrug 0,76 für drei Algorithmen (GLM, CatBoost, MLP) und 0,75 für drei weitere (GLM-Lasso, GLM\_elastic net, SVMC). Der zweitbeste Score in der ML-Auswertung war der WFNS-Score. Der AUC-Wert des WFNS-Score betrug für alle sechs Algorithmen 0,74 und war dem AUC-Wert in der einfachen statistischen Auswertung identisch. Der drittbeste Score in der ML-Auswertung war der VASOGRADE Score (AUC-Range 0,72-0,71, fünf von sechs Algorithmen errechnen eine AUC von 0,72). Auch hier war die ML-Auswertung quasi identisch zur einfachen statistischen Auswertung (eAUC: 0,722). Alle weiteren Scores befanden sich sowohl in der einfachen statistischen Auswertung wie in der ML-Auswertung unterhalb einer AUC von 0,70 (*Fisher-Score* - AUC-Range für ML: 0,47-0,55, eAUC: 0,553; *modified Fisher Score* - AUC-Range für ML: 0,62-0,65, eAUC: 0,645, *BNI-Score* - AUC-Range für ML: 0,60-0,63, eAUC: 0,620).

Bemerkenswert war das Ergebnis, dass der GCS als isoliertes Merkmal genauso gute *bis bessere* AUC-Werte als der Hunt und Hess Score in der ML-Testung erwies: Fünf von sechs getesteten Algorithmen gaben dem GCS einen AUC-Wert von 0,76 (der GLM-Elastic net Algorithmus lediglich einen AUC-Wert von 0,75). Auch in der einfachen statistischen AUROC Auswertung erwies der GCS-Score die beste prädiktive Kraft mit einem eAUC-Wert von 0,752, gefolgt vom Hunt und Hess Score (eAUC 0,750).

Ein zweites, vielleicht noch bemerkenswerteres Ergebnis, ist die Tatsache, dass die gebündelten einzelnen Variablen unter dem Namen ‚Merkmal 8‘ (Alter, GCS score, Geschlecht, Pupillenstatus, IVB, ICB, MLV um  $>5\text{mm}$ , SDH, Lokalisierung der Blutung im anterioren Stromgebiet, BNI-Score) durchweg die besten AUC-Werte in der ML-Auswertung erhielten. Fünf von sechs Algorithmen gaben dem Variablenbündel einen AUC-Wert von 0,77 (CatBoost sogar 0,78).

## 2. Relevanz radiografischer Parameter und von Kombinationsscores

Zur Wertigkeit von radiologischen Scores sowie Kombinationsscores, die klinische und radiologische Daten zusammenführen, hatten wir in einer kürzlich veröffentlichten Studie unserer Arbeitsgruppe bereits die Beobachtung gemacht, dass diese keinen signifikanten Vorteil in der Vorhersagegenauigkeit bezüglich des Outcomes von Vasospasmen, Schlaganfällen oder schlechtem Therapieergebnis nach aSAB gegenüber weitaus einfacheren Scoring-Systemen boten (45). Die Unterlegenheit radiografischer Scores und Kombinationsscore gegenüber klinischen Scores in der Outcomevorhersage einer aSAB gemessen an der modified Rankin Scale konnten wir nun auch in der vorliegenden ML-gestützten Arbeit anhand der AUROC-Analysen bestätigen. Dies wirft die Frage der Relevanz der Variablen für die Outcome-Prädiktion einer aSAB auf, die in den radiologischen Scores verwendet werden, wie z.B. die Dicke einer Blutung in der CT-Bildgebung, oder die Lokalisation einer IVB oder ICB (Fisher Score).

Der BNI schnitt in unserer Modelltestung der radiologischen Scores am besten ab und somit sind ihre Variablen wie z.B. die Dicke des subarachnoidalen Blutclots auch für die Outcome-Vorhersage im Rahmen unserer Arbeit von größerer Relevanz. Die Dicke der subarachnoidalen Blutung als Variable wurde in unserer Studie im Rahmen der BNI-Skala semiquantitativ berücksichtigt. Es erfolgte also keine genauere Quantifizierung der Menge oder der Lokalisation des intraventrikulären oder intraparenchymatösen Blutes. Jüngste Quantifizierungsbestrebungen weisen allerdings darauf hin, dass Patient\*innen mit einer hohen und diffusen SAB-Hämatomlast (definiert als ein aSAB-Hämatom mit  $\geq 4\text{mm}$  Dicke, das sich in mindestens drei Zisternen ausgebreitet hat) im Vergleich zu geringer ausgeprägten aSAB Blutungen eine deutlich höhere Mortalität und Komorbidität im Verlauf erweisen (82). Patient\*innen mit einer hohen und diffusen SAB-Hämatomlast zeigen im Vergleich zu Patienten mit einer geringeren Hämatomlast eine erhöhte Sterblichkeitsrate und Vasospasmen-Inzidenz sechs Wochen postiktal (36,1% vs 14,7%).

Sie unterliegen auch einer erhöhten Wahrscheinlichkeit, einen Sekundärfarkt zu erleiden (19,4% vs. 6,8%), sowie verzögerte Ischämie-bedingte neurologische Ausfälle (28,2% vs. 9,4%) und in der Summe eine größere Wahrscheinlichkeit, einen schlechten klinischen Outcome zu erleiden (32,7% vs. 6,2%) (82).

### 3. Relevanz klinischer Parameter

In unserer Arbeit zeigte sich letztlich kein Vorteil von state-of-the-art ML gegenüber routinemäßig bei Patient\*innenaufnahme zur Verfügung stehender klinischer Diagnostik. Lediglich ein kombiniertes Set relevanter radiologischer und klinischer Variablen zeigte eine leichte Überlegenheit im Vergleich zu einfachen und etablierten klinischen Scores (z.B. Tree Boosting ‚Merkmal 8‘ mit kombinierten Variablen im Vergleich zum Hunt und Hess Score oder der GCS alleine). Vergleichbare Resultate erzielte ein auf Entscheidungsbäume fundiertes Modell, das ähnliche Vorhersagekraft wie eine logistische Regression in einer anderen Studie zeigte (78).

Anhand der Feature-Bewertung in unserer Arbeit ist hervorzuheben, dass klinische Parameter eine erhebliche Relevanz für die Outcomevorhersage zeigen. Wie bereits hervorgehoben, stellte sich der einfache, klinisch erhobene GCS-Wert in unseren Auswertungen als Score mit der stärksten Outcome-Prädiktionskraft dar. Das Variablenbündel ‚Merkmal 8‘ birgt zudem großes Potential bezüglich der Outcomevorhersage einer aSAB.

Das Einbeziehen traditioneller klinischer und radiologischer Diagnostik bleibt also weiterhin nicht wegzudenken bei der Beratung von Patient\*innen und ihren Angehörigen zu dem Krankheitsbild der aSAB. Dazu bleiben sie unabdingbar für die Abschätzung der Prognose, die Kliniker\*innen direkt vor Ort bei Aufnahme treffen, um basierend darauf die therapeutische Strategie festzulegen oder Krankenhausträger bei der Zuteilung von medizinischen Ressourcen zu unterstützen (83–85).

Die Mehrheit bislang etablierter Modelle zur Outcome-Vorhersage bezüglich aSABs basieren auf den bereits aufgezählten klinischen Kriterien wie z.B. der Erhebung neurologischer Ausfälle bei Patient\*innenaufnahme oder radiologischer Kriterien wie der Dicke der aSAB in der initialen Bildgebung und die Anwesenheit einer IVB oder einer ICB. Aktuellere Studien verweisen jedoch darauf, dass andere Faktoren eine Rolle in der genauen Outcome-Vorhersage einer aSAB spielen,



wie z.B. das Alter, der Pupillenstatus (70–72), oder Größe und Lokalisation des Aneurysmas (81, 86).

Auch in unserer Arbeit stellte sich das Alter der Patient\*innen durchweg für all unsere Modelle als die zweitrelevanteste Variable (nach dem GCS-Wert) in der Outcome-Vorhersage einer aSAB dar. Dieses Ergebnis deckt sich mit zahlreichen Arbeiten über die Outcomeprädiktion bei aSAB-Patient\*innen, die ebenfalls in ihren Auswertungen das Alter als eine der drei Variablen mit größter Relevanz für die Outcome-Vorhersage von aSABs identifizieren (3, 78, 79, 86).

Die klinische Relevanz des Pupillenstatus in der Outcome-Vorhersage einer aSAB wird ebenfalls seit geraumer Zeit diskutiert (85, 87, 88). Eine weite oder lichtstarre Pupille kann ein Zeichen für einen erhöhten intrakraniellen Druck darstellen, der zu einer Uncusherniation führt und somit zu einer Kompression des ipsilateralen Nervus oculomotorius, an der Stelle, wo er das Sphenoid kreuzt. Eine weitere Kompression kann zu einer veränderten Reagibilität und Größe der zweiten Pupille führen. Die prädiktive Relevanz des Pupillenstatus wurde jüngst von einer Arbeitsgruppe hervorgehoben (86), die zwei neue Scoring-Methoden für die Outcome-Vorhersage einer aSAB entwickelte, eine davon basiert auf der Hunt und Hess Skala und dem Pupillenstatus, die andere auf der Hunt und Hess Skala, dem Pupillenstatus sowie dem Alter. Beide Scoring Systeme zeigten sich in der AUROC-Analyse sowohl dem einfachen Hunt und Hess Score (AUC 0,794) sowie dem WFNS-Score (AUC 0,775) in der Outcome-Vorhersage für Krankenhaus-Mortalität überlegen. Das System, das sowohl Hunt und Hess Skala, wie auch den Pupillenstatus in der Outcomevorhersage mit einbezog, war diesen zwei klassischen Scoring Systemen überlegen (AUC 0,821). Das System, das zusätzlich zur Hunt und Hess Skala sowohl Pupillenstatus als auch Alter miteinbezog zeigte sich allen anderen überlegen (AUC 0,841).

Der Pupillenstatus wurde in unserer Arbeit aufgrund seiner klinischen Relevanz mit in die Variablenauswahl eingeschlossen, obwohl etwas mehr als 10% der Daten fehlten (13,4% fehlender Daten für Patient\*innen mit pathologischem Pupillenstatus, d.h. ein abweichender Status von seitengleichen Pupillengrößen und einer beidseitigen Lichtreagibilität der Pupillen). In unserer Modelltestung war das Naive Bayes Modell jedoch das einzige, das dem Pupillenstatus einen sehr hohen Stellenwert zuschrieb.

Zusammenfassend zeigt die Mehrheit der publizierten Evidenz die Relevanz folgender Variablen für die Outcome-Vorhersage einer aSAB – die meisten davon klinische Parameter:

- 1) das Alter (als wichtigste (81), zweitwichtigste (3, 45), drittwichtigste Variable (79));
- 2) der WFNS Score (als zweitwichtigste (79, 81), oder viertwichtigste Variable (3))
- 3) die Lokalisation des Aneurysmas (als zweitwichtigste (79) oder fünfwichtigste Variable (3));
- 4) die Größe des Aneurysmas (als drittwichtigste Variable (81)).

## **Ausblick**

Eine bekannte Stärke des Machine Learnings ist die Fähigkeit, eine Vielzahl von Variablen zu analysieren und induktiv Korrelationen und Kausalketten aufzudecken. In zahlreichen medizinischen Disziplinen haben sich ML-basierte Prädiktionsmodelle bereits bewährt. Auch im Bereich der Neurochirurgie wurden ML-basierte Outcome-Vorhersagemodelle für verschiedene Pathologien bereits erfolgreich getestet und prädiktive Leistungen zwischen AUC 0.71 bis 0.96 gezeigt (89). Bezüglich der Vorhersage eines Shunt-abhängigen Hydrozephalus nach einer aSAB waren ML-Methoden klassischen Scores überlegen (90). Sie schlossen dynamische Variablen wie z.B. Infektionen mit ein oder den zeitlichen Beginn einer Therapie in Abhängigkeit zum Beginn eines Fiebers, sowie die Abfolge von Therapieansätzen. Auch in Bezug auf Frühkomplikationen nach zerebralen Tumorsektionen zeigten sich ML-basierte Outcomeprädiktionsmodelle den klassischen Methoden geringfügig überlegen (91). Im aktuellen Forschungsvorhaben wurden u.a. Tree-Boosting und KNN-Methoden angewandt - zwei der vielversprechendsten state-of-the-art ML-Methoden - um das funktionale Outcome nach einer aSAB vorherzusagen. Beide Methoden haben nennenswerte Fortschritte im Vergleich zu linearen oder logistischen Regressionstechniken in der jüngeren Vergangenheit erwiesen.

### *1. Hinzuziehen weiterer Parameter*

In der Outcome-Vorhersage der aSAB sind unseren Resultaten entsprechend klinische Parameter von oberster Relevanz. Ein bislang noch gering erforschter, vielversprechender Weg, um die ML-basierte mittel- bis langfristige Outcome-Vorhersage zu verbessern, ist die Testung weiterer Variablen mit komplexeren Datenstrukturen wie z.B. Labordaten.

Erste Studien belegen die Relevanz von Laborparametern in der Outcome-Vorhersage einer aSAB. Eine Forschungsgruppe zeigte jüngst, dass laborchemische Parameter, die am Aufnahmetag erhalten wurden, die wichtigsten differenzierenden Faktoren für eine erfolgreiche Outcome-Vorhersage bei aSAB in der AUROC-Analyse waren (92). Variablen wie das Procalcitonin (PCT), das C-reaktive Protein oder die Leukozytenzahl waren hier besonders erwähnenswert. Der PCT-Wert war zu jedem Zeitpunkt der Outcome-Prädiktion der wichtigste Faktor zur Ausdifferenzierung eines positiven Outcomes (92). Das PCT ist ein Akute-Phase-Protein, das unter vielfältigen Stressbedingungen ansteigen kann, wie z.B. ein Polytrauma, ein Herzinfarkt oder

höhergradige Verbrennungen. Im Falle einer aSAB ist zu vermuten, dass ein Anstieg des PCT-Wertes eine akute systemische Stressreaktion auf die Blutung darstellt. In besagter Arbeit waren frühe, stark erhöhte PCT-Werte innerhalb von 24 Stunden nach Aufnahme ein starker Prädiktor für ein ungünstiges neurologisches Outcome nach einer transienten globalen Ischämie, wie sie im Rahmen einer schwerwiegenden aSAB erfolgen kann (92).

In der Prädiktionsfähigkeit bezüglich der Überlebensrate stellte sich in einer Arbeit von Hostettler et al. (2018) (92) wiederum die Leukozytenzahl am Tag 1 der Aufnahme als wichtigster früher Parameter dar. Eine Leukozytose innerhalb der ersten 24h nach Aufnahme war verbunden mit einer erhöhten Wahrscheinlichkeit des Auftretens von Vasospasmen, einem Sekundärinfarkt, sowie insgesamt einem schlechten neurologischen Outcome auf der mRS. Variablen wie das Procalcitonin, das C-reaktive Protein oder die Leukozytenzahl weisen auf den Zusammenhang zwischen initialer systemischer Stressreaktion bei Iktus und dessen Auswirkung auf das mittel- bis langfristige Outcome einer aSAB (78, 92). Diese Marker einer systemischen Stressreaktion waren den konventionellen Scoring Systemen wie z.B. die Hunt und Hess und die WFNS-Skala in der Outcome-Vorhersage überlegen. Im Übrigen wurde auch in der Outcome-Vorhersage für nicht-aneurysmatische SABs jüngst die Relevanz laborchemischer Parameter wie z.B. eine Erhöhung des C-Reaktiven Proteins sowie einer Leukozytose bei Aufnahme gezeigt (93).

## 2. Einbringen später Parameter und Vorhersage von Komplikationen

Was aktuelle Forschungsaktivitäten bisher jedoch nicht zu genüge geleistet haben, ist Komplikationen, Komorbiditäten und Behandlungsverläufe in die Outcomevorhersage einer aSAB mit einzuberechnen (78, 94). Meist fokussierte sich bislang die Entwicklung neuer Outcomeprädiktionsmodelle auf große retrospektive Kohorten von Patient\*innen, die am Krankheitsbild der aneurysmatischen SAB erkrankten, und konzentrieren sich auf ‚frühe‘ klinische und radiologische Variablen, die bei Patient\*innenaufnahme zu erhalten sind, um eine mittel bis langfristige Outcome-Vorhersage zu treffen. Es ist selbsterklärend, dass Komorbiditäten, Komplikationen und Behandlungsverläufe eine ausschlaggebende Rolle in dem Outcome einer aSAB spielen – wenn auch als ‚späte‘ Faktoren. Auch ignorieren die meisten Outcome-Prädiktionsmodelle in der Regel Vorkehrungen, die Patient\*innen ggf. bereits getroffen hatten bezüglich der Gestaltung der Versorgung im Falle einer kritischen Erkrankung (z.B.

Patient\*innenverfügungen) oder ebensolche Fälle, wo im Lichte der Erkrankung des Patient\*innen sein Wille gemeinsam mit den Angehörigen, Vorsorgebevollmächtigten, Betreuern dahingehend ausgelegt wird, dass lebenserhaltende Maßnahmen nicht weiter erwünscht sind und es somit zu verfrühtem Todeseintritt kommt (94).

Ein Outcome-Prädiktionsmodell, das erste Ansätze bietet, sowohl eine erweiterte Datenlage der frühen Parameter zu integrieren (Labordaten), wie auch Komorbiditäten und Komplikationen miteinzubeziehen, wird durch Maldaner et al. (2021) geboten (79). Das Timing der Erhebung Outcome-relevanter Daten ist im Großteil der Studiendesigns auf Parameter begrenzt, die man bei Aufnahme erheben kann. Maldaner et al. plädieren dafür, sowohl ‚frühe‘ wie auch ‚späte‘ Parameter in der Outcome-Vorhersage zu berücksichtigen und entsprechend in der Modellbildung miteinzubeziehen (79). ‚Frühe‘ Parameter, die sich als Outcome-relevant herausstellten, sind z.B. der Blutdruck zum Aufnahmezeitpunkt, die Dicke der Blutung (*clot thickness*), das Vorkommen einer intrazerebralen Blutung, der klinische Status zum Aufnahmezeitpunkt, das Vorhandensein einer Hyperglykämie zum Aufnahmezeitpunkt, das Verhältnis von Plättchenvolumen zu Plättchenanzahl, ein erhöhtes C-Reaktives Protein, sowie die Leukozytenzahl zum Aufnahmezeitpunkt. Sog. ‚späte‘ Parameter sind z.B. das Auftreten einer Meningitis, von Vasospasmen oder einer Sekundärischämie, oder die Entwicklung eines chronischen Hydrocephalus (55, 68, 79, 81, 82). In der Modellierung und der AUROC-Analyse der Autoren war das ‚späte‘ Modell mit einer AUC von 0,85 dem ‚frühen‘ Modell (AUC 0,79) in der Outcome-Vorhersage deutlich überlegen (79).

Weitere Fragestellungen, bei denen KI-Anwendungen für die aSAB relevant werden könnten, sind die Vorhersage von Ereignissen, die zu einer Verschlechterung des Outcomes beitragen können, wie z.B. einer Sekundärischämie. Die ML-Modellierung zur Vorhersage des Auftretens von *Delayed Cerebral Ischaemia* auf der Grundlage von extrahierter Datensätze von cCTs bei Aufnahme kombiniert mit klinischen Daten hat bereits einige erfolgsversprechende Resultate erzielt (97–99).

### 3. Limitationen

Die hier präsentierte Studie weist gewisse Limitationen auf. Erstens ist zu erwähnen, dass es sich um eine monozentrische Studie handelt und daher die Generalisierbarkeit unserer Ergebnisse eingeschränkt sein könnte. Zweitens wurden die Daten retrospektiv erhoben, so dass ein Selektions-Bias nicht ausgeschlossen werden kann. Drittens ist unsere Patient\*innenkohorte kleiner als die einiger anderer vergleichbarer Studien (78, 81). Viertens ist die von uns verwendete mean/mode-Imputation keine state-of-the-art Imputationsmethode. Solche state-of-the-art Imputationsmethoden sind momentan nicht zugeschnitten auf Vorhersagemodellierungen. Da jedoch nur sehr wenige fehlende Werte in unseren Datensätzen zu verzeichnen waren, erscheinen uns die diskutierbaren Schwächen der mean/mode-Imputationsmethode in unserem Kontext vernachlässigbar. Fünftens kann nicht ausgeschlossen werden, dass die geringgradige Unausgeglichenheit der dichotomisierten Outcome-Variablen eine gewisse Ungenauigkeit der Analyse bedingt.

### **Schlussfolgerung**

Die in dieser Studie angewandten ML- und DL-Methoden zur Vorhersage des Therapieergebnisses nach aSAB basierten ausschließlich auf klinischen und radiologischen Variablen, die Kliniker\*innen bei Patient\*innenaufnahme zur Verfügung stehen. Zwar zeigte sich, dass ML gut geeignet war zur Vorhersage des Therapieergebnisses nach aSAB, jedoch war diese Outcome-Vorhersage nicht besser geeignet als die auf einfachen traditionellen klinischen Scores basierende. Dennoch sprechen unsere Ergebnisse dafür, dass die weitere Erforschung von ML als Outcome-Prädiktionsmethode lohnenswert sein könnte. ML-basierte Prädiktionsmodelle können mit großer Wahrscheinlichkeit davon profitieren, dass ein breiteres Variablenspektrum miteinbezogen wird als diejenigen der alleinigen, existierenden klinischen Scores. Diese Parameter könnten z.B. folgende sein: diverse laborchemische Parameter, bildgebende Daten oder Nebenerkrankungen. Des Weiteren könnten auch ITS-Scores wie APACHE oder SOFA zur Auswertung herangezogen werden. Eine solche Ausweitung der zu analysierenden Variablen könnte dabei helfen, die klinische Relevanz von ML-Algorithmen zu steigern. Inwiefern zusätzliche Patient\*innendaten die Outcome-Vorhersagekapazität von ML- und DL Algorithmen verändern, bleibt im Rahmen eines Anschlussforschungsvorhaben zu klären (100, 101).

## Literaturverzeichnis

1. Brimblecombe FS, Stoneman ME. Score for respiratory-distress syndrome. *Lancet*. 1969;1:7601.
2. Mullie A, Verstringe P, Buylaert W, Houbrechts H, Michem N, Deloos H, Verbruggen H, Van den Broeck L, Corne L, Lauwaert D. Predictive value of Glasgow coma score for awakening after out-of-hospital cardiac arrest. Cerebral Resuscitation Study Group of the Belgian Society for Intensive Care. *Lancet Lond Engl*. 23. Januar 1988;1(8578):137–40.
3. Rubbert C, Patil KR, Beseoglu K, Mathys C, May R, Kaschner MG, Sigl B, Teichert NA, Boos J, Turowski B, Caspers J. Prediction of outcome after aneurysmal subarachnoid haemorrhage using data from patient admission. *Eur Radiol*. Dezember 2018;28(12):4949–58.
4. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. *Nat Med*. Januar 2019;25(1):24–9.
5. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevskaya O. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med*. 2019;(7).
6. Livne M, Boldsen JK, Mikkelsen IK, Fiebach JB, Sobesky J, Mouridsen K. Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke. *Stroke*. 2018;(49):912–8.
7. Greenberg MS, Herausgeber. *Handbook of Neurosurgery*. 8th Edition. Thieme; 2016.
8. Hölper B, Eichler M. *Kompandium Neuro- und Wirbelsäulen Chirurgie*. 3. Auflage. Pro-Inn Verlag; 2012.
9. Texakalidis P, Sweid A, Mouchtouris N, Peterson EC, Sioka C, Rangel-Castilla L, Reavy-Canwell J, Jabbour P. Aneurysm Formation, Growth, and Rupture: The Biology and Physics of Cerebral Aneurysms. *World Neurosurg*. Oktober 2019;130:277–84.
10. Kassell NF, Torner JC, Haley ECJ, Jane JA, Adams HP, Kongable GL. The International Cooperative Study on the Timing of Aneurysm Surgery. Part 1: Overall management results. *J Neurosurg*. Juli 1990;73(1):18–36.
11. Brisman J. Cerebral Aneurysm Clipping/Coiling 101. *World Neurosurg*. August 2017;104:990–2.
12. Gekka M, Nakayama N, Uchino H, Houkin K. Factors influencing cerebral aneurysm obliteration and reliability of indocyanine green video-angiography. *Acta Neurochir (Wien)*. Februar 2018;160(2):269–76.

13. Mielke D, Malinova V, Rohde V. Comparison of intraoperative microscopic and endoscopic ICG angiography in aneurysm surgery. *Neurosurgery*. September 2014;10 Suppl 3:418–25; discussion 425.
14. Molyneux AJ, Kerr RSC, Yu LM, Clarke M, Sneade M, Yarnold JA, Sandercock P. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. *Lancet Lond Engl*. 3. September 2005;366(9488):809–17.
15. Molyneux AJ, Birks J, Clarke A, Sneade M, Kerr RSC. The durability of endovascular coiling versus neurosurgical clipping of ruptured cerebral aneurysms: 18 year follow-up of the UK cohort of the International Subarachnoid Aneurysm Trial (ISAT). *Lancet Lond Engl*. 21. Februar 2015;385(9969):691–7.
16. Ecker A, Riemenschneider PA. Arteriographic demonstration of spasm of the intracranial arteries, with special reference to saccular arterial aneurysms. *J Neurosurg*. November 1951;8(6):660–7.
17. Petridis AK, Kamp MA, Cornelius JF, Beez T, Beseoglu K, Turowski B, Steiger HJ. Aneurysmatische Subarachnoidalblutung: Diagnostik und Therapie. *Dtsch Arztebl Int*. 2017;114:226–36.
18. Molyneux A, Kerr R, Stratton I, Sandercock P, Clarke M, Shrimpton J, Holman R. International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. *Lancet Lond Engl*. 26. Oktober 2002;360(9342):1267–74.
19. Molyneux AJ, Kerr RSC, Birks J, Ramzi N, Yarnold J, Sneade M, Rischmiller J. Risk of recurrent subarachnoid haemorrhage, death, or dependence and standardised mortality ratios after clipping or coiling of an intracranial aneurysm in the International Subarachnoid Aneurysm Trial (ISAT): long-term follow-up. *Lancet Neurol*. Mai 2009;8(5):427–33.
20. Sade B, Mohr G. Critical appraisal of the International Subarachnoid Aneurysm Trial (ISAT). *Neurol India*. März 2004;52(1):32–5.
21. Nichols DA, Brown RDJ, Meyer FB. Coils or clips in subarachnoid haemorrhage? *Lancet Lond Engl*. 26. Oktober 2002;360(9342):1262–3.
22. McDougall CG, Spetzler RF, Zabramski JM, Partovi S, Hills NK, Nakaji P, Albuquerque FC. The Barrow Ruptured Aneurysm Trial. *J Neurosurg*. Januar 2012;116(1):135–44.
23. Spetzler RF, McDougall CG, Albuquerque FC, Zabramski JM, Hills NK, Partovi S, Nakaji P, Wallace RC. The Barrow Ruptured Aneurysm Trial: 3-year results. *J Neurosurg*. Juli 2013;119(1):146–57.



24. Spetzler RF, McDougall CG, Zabramski JM, Albuquerque FC, Hills NK, Russin JJ, Partovi S, Nakaji P, Wallace RC. The Barrow Ruptured Aneurysm Trial: 6-year results. *J Neurosurg.* September 2015;123(3):609–17.
25. Abdalkader M, Raymond J, Mian A, Naragum V, Cronk K, Roy D, Weill A, Nguyen TN. Early major recurrence of cerebral aneurysms after satisfactory initial coiling. *Interv Neuroradiol J Peritherapeutic Neuroradiol Surg Proced Relat Neurosci.* April 2021;27(2):172–80.
26. Han Y, Liu J, Tian Z, Lv M, Yang X, Wu Z, u. a. Factors affecting recurrence and management of recurrent cerebral aneurysms after initial coiling. *Interv Neuroradiol J Peritherapeutic Neuroradiol Surg Proced Relat Neurosci.* Juni 2020;26(3):300–8.
27. Woitzik J, Dreier JP, Hecht N, Fiss I, Sandow N, Major S, Winkler M, Dahlem YA, Manville J, Diepers M, Muench E, Kasuya H, Schmiedeck P, Vajkoczy P. Delayed cerebral ischemia and spreading depolarization in absence of angiographic vasospasm after subarachnoid hemorrhage. *J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab.* Februar 2012;32(2):203–12.
28. Foreman B. The Pathophysiology of Delayed Cerebral Ischemia. *J Clin Neurophysiol Off Publ Am Electroencephalogr Soc.* Juni 2016;33(3):174–82.
29. Dumont AS, Dumont RJ, Chow MM, Lin CL, Calisaneller T, Ley KF, Kassell NF, Lee KS. Cerebral vasospasm after subarachnoid hemorrhage: putative role of inflammation. *Neurosurgery.* Juli 2003;53(1):123–33; discussion 133-135.
30. Crowley RW, Medel R, Dumont AS, Ilodigwe D, Kassell NF, Mayer SA, Ruefenacht D, Schmiedek P, Weidauer S, Pasqualin A, Macdonald RL. Angiographic vasospasm is strongly correlated with cerebral infarction after subarachnoid hemorrhage. *Stroke.* April 2011;42(4):919–23.
31. Vergouwen MDI, Vermeulen M, van Gijn J, Rinkel GJE, Wijdicks EF, Muizelaar JP, Mendelow AD, Juvela S, Yonas H, Terbrugge KG, Macdonald RL, Diinger MN, Broderick JP, Dreier JP, Roos YBWEM. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: proposal of a multidisciplinary research group. *Stroke.* Oktober 2010;41(10):2391–5.
32. Kassell NF, Sasaki T, Colohan AR, Nazar G. Cerebral vasospasm following aneurysmal subarachnoid hemorrhage. *Stroke.* August 1985;16(4):562–72.
33. Sarrafzadeh AS, Kaisers U, Boemke W. Aneurysmal subarachnoid hemorrhage. Significance and complications. *Anaesthesist.* September 2007;56(9):957–66; quiz 967.
34. Connolly ESJ, Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, Hoh BL, Kirkness CJ, Naidech AM, Ogilvy CS, Patel AB, Thompson BG, Vespa P. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* Juni 2012;43(6):1711–37.

35. Leinonen V, Vanninen R, Rauramaa T. Raised intracranial pressure and brain edema. *Handb Clin Neurol.* 2017;145:25–37.
36. Hokari M, Uchida K, Shimbo D, Gekka M, Asaoka K, Itamoto K. Acute systematic inflammatory response syndrome and serum biomarkers predict outcomes after subarachnoid hemorrhage. *J Clin Neurosci Off J Neurosurg Soc Australas.* August 2020;78:108–13.
37. Nasr DM, Tomasini S, Prasad A, Rabinstein AA. Acute Brain Diseases as Triggers for Stress Cardiomyopathy: Clinical Characteristics and Outcomes. *Neurocrit Care.* Dezember 2017;27(3):356–61.
38. Alfotih GTA, Li F, Xu X, Zhang S. Risk factors for re-bleeding of aneurysmal subarachnoid hemorrhage: meta-analysis of observational studies. *Neurol Neurochir Pol.* 2014;48(5):346–55.
39. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J.* Mai 1957;2(5):200-15).
40. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* Mai 1988;(19(5):604-7).
41. Feigin VL, Lawes CMM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol.* Januar 2003;2(1):43–53.
42. Cedzich C, Roth A. Neurological and psychosocial outcome after subarachnoid haemorrhage, and the hunt and hess scale as a predictor of clinical outcome. *Zentralbl Neurochir.* August 2005;66(3):112–8.
43. Konczalla J, Seifert V, Beck J, Güresir E, Vatter H, Raabe A, Marquardt G. Outcome after Hunt and Hess Grade V subarachnoid hemorrhage: a comparison of pre-coiling era (1980-1995) versus post-ISAT era (2005-2014). *J Neurosurg.* Januar 2018;128(1):100–10.
44. Teasdale GM, Drake CG, Hunt W, Kassell N, Sano K, Pertuiset B, De Villiers JC. A universal subarachnoid hemorrhage scale: report of a committee of the World Federation of Neurosurgical Societies. *J Neurol Neurosurg Psychiatry.* November 1988;51(11):1457.
45. Dengler NF, Sommerfeld J, Diesing D, Vajkoczy P, Wolf S. Prediction of cerebral infarction and patient outcome in aneurysmal subarachnoid hemorrhage: comparison of new and established radiographic, clinical and combined scores. *Eur J Neurol.* 2018;(25):S. 111-119.
46. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. *Neurosurgery.* Januar 1980;6(1):1–9.

47. Frontera JA, Claassen J, Schmidt JM, Wartenberg KE, Temes R, Connolly ESJ, Macdonald RL, Mayer SA. Prediction of symptomatic vasospasm after subarachnoid hemorrhage: the modified fisher scale. *Neurosurgery*. Juli 2006;59(1):21–7; discussion 21-27.
48. Claassen J, Bernardini GL, Kreiter K, Bates J, Du YE, Copeland D, Connolly ES, Mayer SA. Effect of cisternal and ventricular blood on risk of delayed cerebral ischemia after subarachnoid hemorrhage: the Fisher scale revisited. *Stroke*. September 2001;32(9):2012–20.
49. Klimo PJ, Schmidt RH. Computed tomography grading schemes used to predict cerebral vasospasm after aneurysmal subarachnoid hemorrhage: a historical review. *Neurosurg Focus*. 15. September 2006;21(3):E5.
50. Etminan N, Vergouwen MDI, Macdonald RL. Angiographic vasospasm versus cerebral infarction as outcome measures after aneurysmal subarachnoid hemorrhage. *Acta Neurochir Suppl*. 2013;115:33–40.
51. Kramer AH, Hehir M, Nathan B, Gress D, Dumont AS, Kassell NF, Bleck TP. A comparison of 3 radiographic scales for the prediction of delayed ischemia and prognosis following subarachnoid hemorrhage. *J Neurosurg*. August 2008;109(2):199–207.
52. Dengler NF, Diesing D, Sarrafzadeh A, Wolf S, Vajkoczy P. The Barrow Neurological Institute Scale Revisited: Predictive Capabilities for Cerebral Infarction and Clinical Outcome in Patients With Aneurysmal Subarachnoid Hemorrhage. *Neurosurgery*. 1. August 2017;81(2):341–9.
53. Wilson DA, Nakaji P, Abila AA, Uschold TD, Fusco DJ, Oppenlander ME, Albuquerque FC, McDougall CG, Zabramski JM, Spetzler RF. A simple and quantitative method to predict symptomatic vasospasm after subarachnoid hemorrhage based on computed tomography: beyond the Fisher scale. *Neurosurgery*. Oktober 2012;71(4):869–75.
54. de Oliveira Manoel AL, Jaja BN, Germans MR, Yan H, Qian W, Kouzmina E, Marotta TR, Turkel-Parrella D, Schweizer TA, Macdonald RL. The VASOGRADE: A Simple Grading Scale for Prediction of Delayed Cerebral Ischemia After Subarachnoid Hemorrhage. *Stroke*. Juli 2015;46(7):1826–31.
55. Witsch J, Frey HP, Patel S, Park S, Lahiri S, Schmidt JM, Agarwal S, Falo MC, Velazquez A, Jaja B, Macdonald RL, Connolly ES, Claassen J. Prognostication of long-term outcomes after subarachnoid hemorrhage: The FRESH score. *Ann Neurol*. Juli 2016;80(1):46–58.
56. Witsch J, Kuohn L, Hebert R, Cord B, Sansing L, Gilmore EJ, Hwang DY, Petersen N, Falcone GJ, Matouk C, Sheth K. Early Prognostication of 1-Year Outcome After Subarachnoid Hemorrhage: The FRESH Score Validation. *J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc*. Oktober 2019;28(10):104280.
57. Naval NS, Kowalski RG, Chang TR, Caserta F, Carhuapoma JR, Tamargo RJ. The SAH Score: a comprehensive communication tool. *J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc*. Juni 2014;23(5):902–9.

58. Lee VH, Ouyang B, John S, Conners JJ, Garg R, Bleck TP, Temes RE, Cutting S, Prabhakaran S. Risk stratification for the in-hospital mortality in subarachnoid hemorrhage: the HAIR score. *Neurocrit Care*. August 2014;21(1):14–9.
59. Ertel, W. *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*. Wiesbaden: Springer Fachmedien; 2008.
60. Senders JT, Arnaout O, Karhade AV, Dasenbrock HH, Gormley WB, Broekman ML, Smith T. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. *Neurosurgery*. 1. August 2018;83(2):181–92.
61. Paliwal N, Jaiswal P, Tutino VM, Shallwani H, Davies JM, Siddiqui AH, Rai R, Meng H. Outcome prediction of intracranial aneurysm treatment by flow diverters using machine learning. *Neurosurg Focus*. 1. November 2018;45(5):E7.
62. Dumont TM, Rughani AI, Tranmer BI. Prediction of Symptomatic Cerebral Vasospasm after Aneurysmal Subarachnoid Hemorrhage with an Artificial Neural Network: Feasibility and Comparison with Logistic Regression Models. *World Neurosurg*. 2011;75:57–63.
63. Faron A, Sichtermann T, Teichert N, Luetkens JA, Keulers A, Nikoubashman O, Freiherr J, Mpotsaris A, Wiesmann M. Performance of a Deep-Learning Neural Network to Detect Intracranial Aneurysms from 3D TOF-MRA Compared to Human Readers. *Clin Neuroradiol*. 2019;
64. Park A, Chute C, Rajpurkar P, Lou J, Ball RL, Shpanskaya K, Jabarkheel R, Kim LH, McKenna E, Tseng J, Ni J, Wishah F, Wittber F, Hong DS, Wilson TJ, Halabi S, Basu S, Patel BN, Lungren MP, Ng AY, Yeom KW. Deep-Learning-Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. *JAMA Netw Open*. 2019;2.
65. Dengler NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, Livne M, Wolf S, Vajkoczy P, Frey D. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. *Neurosurg Rev*. 20. Januar 2021;
66. Wang HL, Hsu WY, Lee MH, Weng HH, Chang SW, Yang JT, Tsai YH. Automatic Machine-Learning-Based Outcome Prediction in Patients With Primary Intracerebral Hemorrhage. *Front Neurol*. 2019;10:910.
67. Yao H, Williamson C, Gryak J, Najarian K. Automated hematoma segmentation and outcome prediction for patients with traumatic brain injury. *Artif Intell Med*. Juli 2020;107:101910.
68. Liu L, Zhang H, Rekik I, Chen X, Wang Q, Shen D. Outcome Prediction for Patient with High-Grade Gliomas from Brain Functional and Structural Networks. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv*. Oktober 2016;9901:26–34.

69. Tranchevent LC, Azuaje F, Rajapakse JC. A deep neural network approach to predicting clinical outcomes of neuroblastoma patients. *BMC Med Genomics*. 20. Dezember 2019;12(Suppl 8):178.
70. Shaver MM, Kohanteb PA, Chiou C, Bardis MD, Chantaduly C, Bota D, Filippi CG, Weinberg B, Grinband J, Chow DS, Chang PD. Optimizing Neuro-Oncology Imaging: A Review of Deep Learning Approaches for Glioma Imaging. *Cancers*. 14. Juni 2019;11(6).
71. Schwartz JT, Gao M, Geng EA, Mody KS, Mikhail CM, Cho SK. Applications of Machine Learning Using Electronic Medical Records in Spine Surgery. *Neurospine*. Dezember 2019;16(4):643–53.
72. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine*. März 2019;2(1):e1044.
73. Lo BWY, Macdonald RL, Baker A, Levine MAH. Clinical Outcome Prediction in Aneurysmal Subarachnoid Hemorrhage Using Bayesian Neural Networks with Fuzzy Logic Inferences. *Comput Math Methods Med*. 2013;
74. Sandow NF, Diesing D, Sarrafzadeh A, Vajkoczy P, Wolf S. Nimodipine Dose Reductions in the Treatment of Patients with Aneurysmal Subarachnoid Hemorrhage. *Neurocrit Care*. August 2016;25(1):29–39.
75. Steiner T, Juvela S, Unterberg A, Jung C, Forsting M, Rinkel G. European Stroke Organization guidelines for the management of intracranial aneurysms and subarachnoid haemorrhage. *Cerebrovasc Dis Basel Switz*. 2013;35(2):93–112.
76. Diringer MN, Bleck TP, Claude Hemphill J 3rd, Menon D, Shutter L, Vespa P, Bruder N, Conolly ESJ, Citerio G, Gress D, Hänggi D, Hoh BL, Lanzino G, Le Roux P, Rabinstein A, Schmutzhard E, Stocchetti N, Suarez JJ, Treggiari M, Tseng MY, Vergouwen MDI, Wolf S, Zipfel G. Critical care management of patients following aneurysmal subarachnoid hemorrhage: recommendations from the Neurocritical Care Society's Multidisciplinary Consensus Conference. *Neurocrit Care*. September 2011;15(2):211–40.
77. Dengler NF, Sommerfeld J, Diesing D, Vajkoczy P, Wolf S. Prediction of cerebral infarction and patient outcome in aneurysmal subarachnoid hemorrhage: comparison of new and established radiographic, clinical and combined scores. *Eur J Neurol*. Januar 2018;25(1):111–9.
78. de Toledo P, Rios PM, Ledezma A, Sanchis A AJF, Lagares A. Predicting the outcome of patients with subarachnoidal hemorrhage using machine learning techniques. *IEEE Trans Inf Technol Biomed*. 2009;13:794–801.
79. Maldaner N, Zeitlberger AM, Sosnova M, Goldberg J, Fung C, Bervini D, May A, Bijlenga P, Schaller K, Roethlisberger M, Rychen J, Zumofen DW, D'Alonzo D, Marbacher S, Fandino J, Daniel RT, Burkhardt JK, Chiappini A, Robert T, Schatlo B, Schmid J, Maduri R, Staartjes VE, Seule MA, Weyerbrock A, Serra C, Stienen MN, Bozinov O, Regli L. Development of a Complication- and Treatment-Aware Prediction Model for Favorable

Functional Outcome in Aneurysmal Subarachnoid Hemorrhage Based on Machine Learning. *Neurosurgery*. 13. Januar 2021;88(2):E150–7.

80. Jaja BNR, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, Lantigua H, Le Roux P, Lo B, Louffat-Olivares A, Mayer S, Molyneux A, Quinn A, Schweizer TA, Schenk T, Spears J, Todd M, Torner J, Vergouwen MDI, Wong GKC, Sing J, Macdonald RL. Clinical prediction models for aneurysmal subarachnoid hemorrhage: a systematic review. *Neurocrit Care*. Februar 2013;18(1):143–53.
81. van Donkelaar CE, Bakker NA, Birks J, Veeger NJGM, Metzemaekers JDM, Molyneux AJ, Groen RJM, van Dijk JMC. Prediction of Outcome After Aneurysmal Subarachnoid Hemorrhage. *Stroke*. April 2019;50(4):837–44.
82. Aldrich EF, Higashida R, Hmissi A, Le EJ, Macdonald RL, Marr A, Mayer SA, Roux S, Bruder N. Thick and diffuse cisternal clot independently predicts vasospasm-related morbidity and poor outcome after aneurysmal subarachnoid hemorrhage. *J Neurosurg*. 22. Mai 2020;134(5):1553–61.
83. Grewal P, Hall JP, Jhaveri M, Dafer RM. Cerebral Vasculopathy and Spinal Arachnoiditis: Two Rare Complications of Ventriculitis Post Subarachnoid Hemorrhage. *Cureus*. 23. Dezember 2020;12(12):e12241.
84. Sun PY, Poblete RA, Nguyen PL, Bulic SF, Kim-Tenser MA, Marehbian J, Cen SY, Emanuel BA. Association Between Glycemic Gap and In-hospital Outcomes in Aneurysmal Subarachnoid Hemorrhage. *Front Neurol*. 2021;12:714341.
85. Ndieugnou Djangang N, Ramunno P, Izzi A, Garufi A, Menozzi M, Diaferia D, Peluso L, Prezioso C, Talamonti M, Njimi H, Schuind S, Vincent JL, Creteur J, Taccone FS, Gouvea Bogossian E. The Prognostic Role of Lactate Concentrations after Aneurysmal Subarachnoid Hemorrhage. *Brain Sci*. 17. Dezember 2020;10(12).
86. Mader MM, Piffko A, Dengler NF, Ricklefs FL, Dührsen L, Schmidt NO, Regelsberger J, Westphal M, Wolf S, Czorlich P. Initial pupil status is a strong predictor for in-hospital mortality after aneurysmal subarachnoid hemorrhage. *Sci Rep*. 16. März 2020;10(1):4764.
87. Brennan PM, Murray GD, Teasdale GM. Simplifying the use of prognostic information in traumatic brain injury. Part 1: The GCS-Pupils score: an extended index of clinical severity. *J Neurosurg*. Juni 2018;128(6):1612–20.
88. Murray GD, Brennan PM, Teasdale GM. Simplifying the use of prognostic information in traumatic brain injury. Part 2: Graphical presentation of probabilities. *J Neurosurg*. Juni 2018;128(6):1621–34.
89. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, Smith TR, Arnaout O. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. *World Neurosurg*. 2018;109:476–486.e1.

90. Muscas G, Matteuzzi T, Becattini E, Orlandini S, Battista F, Laiso A, Nappini S, Limbucci N, Renieri L, Carangelo BR, Mangiafico S, Della Puppa A. Development of Machine Learning Models to prognosticate chronic shunt-dependent hydrocephalus after aneurysmal subarachnoidal hemorrhage. *Acta Neurochir.* 2020;162:3093–105.
91. van Niftrik CHB, van der Wouden F, Staartjes VE, Fierstra J, Stienen MN, Akeret K, Sebök M, Fedele T, Sarnthein J, Bozinov O, Krayenbühl N, Regli L, Serra C. Machine Learning Algorithm identifies patients at high risk for early complications after intracranial tumor surgery: registry-based cohort study. *Neurosurgery.* 85:756–64.
92. Hostettler IC, Muroi C, Richter JK, Schmid J, Neidert MC, Seule M, Boss O, Panglu A, Germans MR, Keller E. Decision tree analysis in subarachnoid hemorrhage: prediction of outcome parameters during the course of aneurysmal subarachnoid hemorrhage using decision tree analysis. *J Neurosurg.* 1. Dezember 2018;129(6):1499–510.
93. Schuss P, Hadjiathanasiou A, Brandecker S, Güresir Á, Vatter H, Güresir E. Elevated C-reactive protein and white blood cell count at admission predict functional outcome after non-aneurysmal subarachnoid hemorrhage. *J Neurol.* Dezember 2018;265(12):2944–8.
94. Wartenberg KE, Hwang DY, Haeusler KG, Muehlschlegel S, Sakowitz OW, Madžar D, Hamer HM, Rabinstein AA, Greer DM, Hemphill JC, Meixensberger J, Varelas PN. Gap Analysis Regarding Prognostication in Neurocritical Care: A Joint Statement from the German Neurocritical Care Society and the Neurocritical Care Society. *Neurocrit Care.* Oktober 2019;31(2):231–44.
95. Eagles ME, Tso MK, Ayling OGS, Wong JH, MacDonald RL. Unfavorable Outcome After Good Grade Aneurysmal Subarachnoid Hemorrhage: Exploratory Analysis. *World Neurosurg.* Dezember 2020;144:e842–8.
96. Schupper AJ, Eagles ME, Neifert SN, Mocco J, Macdonald RL. Lessons from the CONSCIOUS-1 Study. *J Clin Med.* 14. September 2020;9(9).
97. Ramos LA, van der Steen WE, Sales Barros R, Majoie CBLM, van den Berg R, Verbaan D, Vandertop WP, Zijlstra IJAJ, Zwinderman AH, Strijkers GJ, Olabarriaga SD, Marquering HA. Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. *J Neurointerventional Surg.* Mai 2019;11(5):497–502.
98. Savarraj JPI, Hergenroeder GW, Zhu L, Chang T, Park S, Megjhani M, Vahidy FS, Zhao Z, Kitagawa RS, Choi HA. Machine Learning to Predict Delayed Cerebral Ischemia and Outcomes in Subarachnoid Hemorrhage. *Neurology.* 26. Januar 2021;96(4):e553–62.
99. Katsuki M, Kawamura S, Koh A. Easily Created Prediction Model Using Automated Artificial Intelligence Framework (Prediction One, Sony Network Communications Inc., Tokyo, Japan) for Subarachnoid Hemorrhage Outcomes Treated by Coiling and Delayed Cerebral Ischemia. *Cureus.* Juni 2021;13(6):e15695.

100. Dasenbrok HH, Rudy RF, Smith TR, Guttieres D, Frerichs KU, Gormley WB, Aziz-Sultan MA, Du R. Hospital-Acquired Infections after Aneurysmal Subarachnoid Hemorrhage: A Nationwide Analysis. *World Neurosurg.* 2016;88:459–74.
101. Kahn JM, Caldwell EC, Deem S, Newell DW, Heckbert SR, Rubenfeld GD. Acute lung injury in patients with subarachnoid hemorrhage: incidence, risk factors and outcome. *Crit Care Med.* 2006;34:196–202.



## Eidesstattliche Versicherung und Anteilserklärung

### Eidesstattliche Versicherung

„Ich, Sophie-Charlotte BRUNE, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: *„Vergleich von Outcomeprädiktionsmodellen bei der aneurysmatischen Subarachnoidalblutung: klinische und radiografische Scores versus Machine Learning Algorithmen/ Outcome prediction in aneurysmal subarachnoid haemorrhage: a comparison of machine learning methods and established clinico-radiological scores“* selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren:innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit der Erstbetreuerin, **PD Dr. med. Nora Dengler**, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; [www.icmje.org](http://www.icmje.org)) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

## **Anteilerklärung an etwaigen erfolgten Publikationen**

**Frau Sophie-Charlotte Brune** hatte folgenden Anteil an den folgenden Publikationen:

**Dengler** NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, Livne M, Wolf S, Vajkoczy P, Frey D. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. *Neurosurg Rev.* 2021 Jan 20. doi: 10.1007/s10143-020-01453-6. Epub ahead of print. PMID: 33474607.

Beitrag im Einzelnen: Sophie-Charlotte Brune hat die Erhebung der Daten, die in der o.g. Publikation analysiert wurden unterstützt. Es erfolgte die Erhebung der radiologischen Scores anhand der CTs der Patient\*innen bei Aufnahme sowie die Zusammenstellung der klinischen Daten aus den Medizinischen Dokumentationssystemen der Klinik (SAP, COPRA). Weiterhin hat Sie bei der Erstellung der Figures und Tabellen sowie bei der Literaturrecherche und Manuskripterstellung wichtige Funktionen übernommen.

## **Lebenslauf**

Der Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version dieser Arbeit nicht veröffentlicht.



## **Publikationsliste**

- Dengler NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, Livne M, Wolf S, Vajkoczy P, Frey D. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. Neurosurg Rev. 20. Januar 2021;

## **Danksagung**

Ich danke Frau PD Dr. Nora Franziska Dengler für die Überlassung des Themas.

PD Dr. Nora Franziska Dengler und Dr. Dietmar Frey möchte ich für ihre Ausdauer, Motivation, vielfache konstruktive Kritik und wissenschaftliche Expertise danken.

Dr. Vince Istvan Madai, Esra Zihni, Adam Hilbert, Michelle Livne und Dr. Stefan Wolf danke ich für die Aufnahme in die Arbeitsgruppe KI und aSAB, sowie für informative inhaltliche Diskussionen.

Danke an Dr. Meike Unteroberdörster für die anfängliche Beratung und den Support.

Meinem Partner und meiner Familie danke ich für ihre Geduld und bedingungslose Unterstützung.

# Bescheinigung Statistik



CharitéCentrum für Human- und Gesundheitswissenschaften

Charité | Campus Charité Mitte | 10117 Berlin

Institut für Biometrie und Klinische Epidemiologie (iBikE)

Direktor: Prof. Dr. Geraldine Rauch

Name, Vorname: Brune, Sophie-Charlotte

E-mailadresse

Matrikelnummer

PromotionsbetreuerIn: PD Dr. Nora Dengler

Promotionsinstitution / Klinik: CC15 Kliniken für Neurochirurgie CBF/CVK

Postanschrift:

Charitéplatz 1 | 10117 Berlin

Besucheranschrift:

Reinhardtstr. 58 | 10117 Berlin

Tel. +49 (0)30 450 562171

geraldine.rauch@charite.de

<https://biometrie.charite.de/>



## Bescheinigung

Hiermit bescheinige ich, dass Frau Sophie-Charlotte Brune innerhalb der Service Unit Biometrie des Instituts für Biometrie und Klinische Epidemiologie (iBikE) bei mir eine statistische Beratung zu einem Promotionsvorhaben wahrgenommen hat. Folgende Beratungstermine wurden wahrgenommen:

- Termin 1: 13.07.2020
- (Termin am 17.08.2021 aus technischen Gründen abgebrochen)
- Termin 2: 08.09.2021

Folgende wesentliche Ratschläge hinsichtlich einer sinnvollen Auswertung und Interpretation der Daten wurden während der Beratung erteilt:

- Einführung in die Datenauswertung und grafische Datenaufbereitung mit SPSS
- Diskussion der bereits veröffentlichten Publikation:
  - Interpretation von Parametern diagnostischer Testverfahren (Sensitivität, Spezifität, positiv/negativ prädiktiver Wert, ROC-Kurve, AUC)
  - Erwartungshaltung an Test-Performance in Trainings- und Validierungsdatensatz (Overfitting)
  - Überprüfung der tabellarischen Ergebnisdarstellung, nötige Korrekturen

Diese Bescheinigung garantiert nicht die richtige Umsetzung der in der Beratung gemachten Vorschläge, die korrekte Durchführung der empfohlenen statistischen Verfahren und die richtige Darstellung und Interpretation der Ergebnisse. Die Verantwortung hierfür obliegt allein der Promovierenden. Das Institut für Biometrie und klinische Epidemiologie übernimmt hierfür keine Haftung.

Datum: 13.09.2021

[Redacted signature area]

Unterschrift BeraterIn, Institutsstempel