

Aus der Klinik für Psychiatrie und Psychotherapie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Prädiktion klinischer Outcomes nach ischämischem
Schlaganfall mit maschinellem Lernen

Prediction of clinical outcomes
after ischemic stroke using machine learning

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Lea Fast

Datum der Promotion: 30.06.2024

Inhaltsverzeichnis

Tabellenverzeichnis	iii
Abbildungsverzeichnis	iv
Abkürzungsverzeichnis	v
Zusammenfassung.....	1
Abstract	2
1 Einleitung	3
1.1 Schlaganfall	3
1.2 Maschinelles Lernen.....	6
1.3 Fragestellung	8
2 Methodik	10
2.1 Datensatz.....	10
2.1.1 Datenselektion	10
2.1.2 Input-Variablen	11
2.1.3 Klinische Outcomes	14
2.2 Maschinelles Lernen.....	15
2.2.1 Vorverarbeitung	17
2.2.2 Algorithmen	17
2.2.3 Kreuzvalidierungsverfahren.....	19
2.2.4 Leistungskennzahlen.....	20
2.2.5 Permutationstest.....	20
2.2.6 Shapley-Werte	20
3 Ergebnisse	22
3.1 Modifizierte Rankin Skala	25
3.2 Barthel Index	25
3.3 Mini-Mental-Status-Test	25
3.4 Modifiziertes Telefoninterview für den kognitiven Status.....	25

3.5 Allgemeine Depressionsskala	26
3.6 Mortalität	26
4 Diskussion	28
4.1 Modifizierte Rankin Skala	29
4.2 Barthel Index.....	30
4.3 Mini-Mental-Status-Test.....	31
4.4 Modifiziertes Telefoninterview für den kognitiven Status	31
4.5 Allgemeine Depressionsskala.....	32
4.6 Mortalität	33
4.7 Klinische Implikationen	34
4.8 Hürden der Methodenentwicklung	34
4.9 Limitationen	35
5 Schlussfolgerungen	36
Literaturverzeichnis	37
Eidesstattliche Versicherung	50
Anteilerklärung an der erfolgten Publikation.....	51
Auszug aus der Journal Summary List	52
Druckexemplar der ausgewählten Publikation: Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M, Ritter K. Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. Front. Neurol. 2023. 14:1114360.....	57
Lebenslauf.....	72
Komplette Publikationsliste	73
Danksagung	74

Tabellenverzeichnis

Tabelle 1: Beschreibung der Patient*innenpopulation	12
Tabelle 2: Grenzwerte und Verteilung der Outcomes in der Patient*innenpopulation ...	15
Tabelle 3: Hyperparameter der maschinellen Lernmodelle	18
Tabelle 4: Die besten Prädiktionsergebnisse und wichtigsten Prädiktoren laut Shapley- Werten für alle Ergebnisse	26

Abbildungsverzeichnis

Abbildung 1: Selektionsprozess der Patient*innendaten	11
Abbildung 2: Überblick der Analysen	16
Abbildung 3: Ablauf des Kreuzvalidierungsverfahrens	19
Abbildung 4: Prädiktionsergebnisse angegeben in ausgewogener Genauigkeit (BA, englisch <i>Balanced accuracy</i>) für alle Outcomes, Zeitpunkte und Input-Subdomänen ..	24
Abbildung 5: Darstellung der Entscheidungsfindung des Entscheidungsbaums mit Gradientenverstärkung (GB, englisch <i>Gradient Boosting</i>) für die modifizierte Rankin Skala (mRS) bei Entlassung	28

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AUC	Fläche unter der Grenzwertoptimierungskurve (englisch <i>Area under the receiver operating characteristic curve</i>)
ASPECTS	Alberta Stroke Program Early CT Scores
BA	Ausgewogene Genauigkeit (englisch <i>Balanced accuracy</i>)
bE	Bei Entlassung
BI	Barthel Index
BMI	Körpermasseindex (englisch <i>Body-Mass-Index</i>)
CES-D	Allgemeine Depressionsskala (englisch <i>Center for Epidemiologic Studies Depression Scale</i>)
CDSS	Klinisches Entscheidungsunterstützungssystem (englisch <i>Clinical Decision Support System</i>)
CSB	Centrum für Schlaganfallforschung Berlin
CT	Computertomografie
DNN	Tiefes neuronales Netzwerk (englisch <i>Deep Neural Network</i>)
DWI	Diffusions-gewichtete Sequenz (englisch <i>Diffusion-weighted imaging</i>)
eGFR	Geschätzte glomeruläre Filtrationsrate (englisch <i>Estimated Glomerular Filtration Rate</i>)
EKG	Elektrokardiografie
FIM	Funktionaler Selbständigkeitsindex (englisch <i>Functional Independence Measure</i>)
FLAIR	Fluid-attenuated inversion recovery
GB	Entscheidungsbaum mit Gradientenverstärkung (englisch <i>Gradient Boosting</i>)
HbA1c	Glykiertes Hämoglobin
HDL	High-density Lipoprotein
HRS	Health and Retirement Study
hsCRP	Hochsensitives C-reaktives Protein
IDI	Integrated Discrimination Improvement
KI	Konfidenzintervall
KM	Kontrastmittel
LDL	Low-density Lipoprotein
LR	Plausibilitätsquotiententest (englisch <i>Likelihood ratio test</i>)
MICE	Multiple Imputation durch verkettete Gleichungen (englisch <i>Multiple Imputation using Chained Equations</i>)
ML	Maschinelles Lernen
MMSE	Mini-Mental-Status-Test (englisch <i>Mini-Mental State Examination</i>)
MoCA	Montreal-Cognitive-Assessment-Test
mRS	Modifizierte Rankin Skala
MRT	Magnetresonanztomographie
MW	Mittelwert
NIHSS	National Institute of Health Stroke Scale
PROSCIS-B	PROSpective Cohort with Incident Stroke Berlin
RF	Zufallswald (englisch <i>Random Forest</i>)
SD	Standardabweichung (englisch <i>Standard deviation</i>)
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
SVM	Stützvektormaschine (englisch <i>Support Vector Machine</i>)
SVM-lin	Stützvektormaschine mit linearem Kernel

SVM-rbf	Stützvektormaschine mit Kernel der radialen Basisfunktion
TICS-M	Modifiziertes Telefoninterview für den kognitiven Status (englisch <i>Modified Telephone Interview for Cognitive Status</i>)
TOAST	Trial of ORG 10172 in Acute Stroke Treatment
WHO	Weltgesundheitsorganisation (englisch <i>World Health Organization</i>)

Zusammenfassung

Ziele: Der Schlaganfall zählt weltweit zu den Hauptursachen von Tod und Behinderung im Erwachsenenalter. Daher ist es von entscheidender Bedeutung die wichtigsten Einflussfaktoren für den Verlauf dieses Krankheitsbildes zu identifizieren. In dieser Arbeit nutzen wir maschinelles Lernen (ML), um klinische Profile nach erstmaligem ischämischem Schlaganfall anhand funktioneller und kognitiver Beeinträchtigung, Depression und Mortalität zu charakterisieren.

Methoden: Wir nutzten drei ML-Modelle, einschließlich einer Stützvektormaschine mit linearem Kernel und Kernel der radialen Basisfunktion (SVM-lin und SVM-rfb) sowie eines Entscheidungsbaums mit Gradientenverstärkung (GB, englisch *Gradient Boosting*), um gute versus schlechte klinische Outcomes bei 307 Patient*innen mit jeweils 43 Input-Variablen aus der PROSpective Cohort with Incident Stroke Berlin (PROSCIS-B) zu klassifizieren. Die Patient*innenpopulation bestand aus 151 Frauen und 156 Männern mit einem Durchschnittsalter von 68 ± 14 Jahren. Unter Anwendung einer wiederholten fünffach verschachtelten Kreuzvalidierung wurden funktionelle und kognitive Beeinträchtigung, Depression und Mortalität zu verschiedenen Zeitpunkten prädiziert sowie die wichtigsten Einflussfaktoren mittels Shapley-Werten identifiziert. Die modifizierte Rankin Skala (mRS) und Barthel Index (BI) dienten zur Messung des funktionellen Outcomes, der Mini-Mental-Status-Test (MMSE, englisch *Mini-Mental State Examination*) und das modifizierte Telefoninterview für den kognitiven Status (TICS-M, englisch *Modified Telephone Interview for Cognitive Status*) als Indikatoren für kognitiven Beeinträchtigung und die allgemeine Depressionsskala (CES-D, englisch *Center for Epidemiologic Studies Depression Scale*) als Maßstab für Depression.

Ergebnisse: Unsere Analysen ergaben signifikante Ergebnisse in Hinblick auf die Prädiktion von mRS, BI und MMSE bei Entlassung, TICS-M und CES-D nach einem Jahr sowie TICS-M nach drei Jahren. Bildung war der wichtigste Prädiktor für alle MMSE-, TICS-M- und CES-D-Outcomes, die National Institutes of Health Stroke Scale für mRS und BI bei Entlassung sowie Bauchumfang für mRS nach einem Jahr.

Schlussfolgerung: Mittels eines systematischen Vergleichs von ML-Methoden konnten wir die klinischen Outcomes für Patient*innen nach erstmaligem ischämischem Schlaganfall prädizieren und die wichtigsten Vorhersagefaktoren identifizieren.

Abstract

Background: The worldwide impact of stroke as a leading cause of death and disability highlights the importance of identifying the main factors contributing to poor long-term outcomes. In this study, we predict and systematically compare clinical profiles of patients experiencing their first-ever ischemic stroke using machine learning (ML) techniques. Our primary objective is to predict functional recovery, cognitive impairment, depression and mortality, while also uncovering the leading predictors.

Methods: We used three ML models, including a Support Vector Machine with linear kernel (SVM-lin), a Support Vector Machine with radial basis function kernel (SVM-rfb), and a Gradient Boosting Classifier (GB), to predict good versus poor clinical outcomes in 307 patients from the PROSpective Cohort with Incident Stroke Berlin (PROSCIS-B) dataset. The dataset consisted of 151 females and 156 males, with an average age of 68 ± 14 years and 43 input variables per patient. We aimed to predict functional outcome, cognitive impairment, depression and mortality at various follow-up intervals based on repeated 5-fold nested cross-validation and to identify key predictors via Shapley additive explanations. The Modified Rankin Scale (mRS) and Barthel Index (BI) were used as indicators of functional outcome, the Mini-Mental State Examination (MMSE) and Modified Telephone Interview for Cognitive Status (TICS-M) were utilized to assess cognitive impairment and the Center for Epidemiologic Studies Depression Scale (CES-D) served as a measure of depression.

Results: Our analysis showed significant prediction results for mRS, BI and MMSE at patient discharge, TICS-M and CES-D after one year as well as for TICS-M after three years. Education level was the leading predictor for all MMSE, TICS-M and CES-D outcomes while National Institutes of Health Stroke Scale was the most important predictor for mRS and BI at patient discharge and waist circumference for mRS at one-year follow-up.

Conclusion: Through a systematic comparison of ML methods, we were able to successfully predicted clinical profiles for patients experiencing their first-ever ischemic stroke while also pinpointing the associated leading predictors.

1 Einleitung

1.1 Schlaganfall

Laut der Weltgesundheitsorganisation (WHO, englisch *World Health Organization*) ist der Schlaganfall definiert als ein plötzlich auftretendes klinisches Zeichen einer fokalen oder globalen Störung der zerebralen Funktion, das länger als 24 Stunden anhält oder zum Tod führt und für das es keine andere offensichtliche Ursache als eine vaskuläre gibt [1]. Diese akut eintretende Schädigung des Gehirns geht mit einem irreversiblen Ausfall von Neuronen durch einen Gefäßverschluss (ischämisch) oder eine Hirnblutung (hämorrhagisch) einher [2]. Mit einem Anteil von rund 85% stellt der ischämische Schlaganfall den häufigsten Subtypen dar. Dabei wird die vaskuläre Versorgung der Neuronen durch Embolien, Thrombosen oder systemische Hypoperfusion unterbrochen. Der hämorrhagische Schlaganfall hingegen äußert sich als intrazerebrale Blutung oder Subarachnoidalblutung. Im Gegensatz zum ischämischen Schlaganfall wird das zerebrale Parenchym hierbei durch den zunehmenden intrakraniellen Druck beeinträchtigt [3]. Zu den modifizierbaren Risikofaktoren des Schlaganfalls zählen arterielle Hypertonie, koronare Herzkrankheit, Herzrhythmusstörungen, strukturelle Herzerkrankungen, Diabetes Mellitus, Hypercholesterinämie, Hypertriglyceridämie, Adipositas, Bewegungsmangel, Einnahme oraler Kontrazeption sowie Nikotin-, Alkohol- und Drogenabusus [4]. Alter, biologisches Geschlecht, zerebrovaskuläre Vorerkrankungen, genetische Disposition, Ethnizität und sozioökonomischer Status sowie die Umweltfaktoren Temperatur, Jahreszeit und Klima gehören hingegen zu den nicht-modifizierbaren Risikofaktoren [5,6].

Die typischen klinischen Symptome des Schlaganfalls sind Sprach- und Sprechstörungen, einseitige Lähmungserscheinungen, plötzlich eintretender starker Kopfschmerz, Gesichtsfeldausfälle, Krampfanfälle, Bewusstseinsverlust und Schwindel [7,8]. Neben der Anamnese und körperlichen Untersuchung spielen die bildgebenden Verfahren der Computertomografie (CT) und Magnetresonanztomografie (MRT) eine zentrale Rolle in der Diagnostik. In erster Linie ist dabei eine Unterscheidung in ischämischen und hämorrhagischen Schlaganfall essenziell. Die CT hat sich durch ihre breite Verfügbarkeit als diagnostischer Goldstandard etabliert. Durch Gabe von Kontrastmittel (KM) kann eine Darstellung der extra- und intrakraniellen Gefäße erfolgen sowie mittels multimodaler Bild-

gebung die Gewebperfusion errechnet werden. Diese erlaubt es abgestorbenes Gewebe (Infarkt) von minderperfundierten Arealen (Penumbra) zu unterscheiden [9]. Obwohl die Verfügbarkeit der MRT begrenzt ist und teilweise Kontraindikationen bestehen, kann durch einen Vergleich von Diffusions-gewichteten Sequenz (DWI, englisch *Diffusion-weighted imaging*) und der Fluid-attenuated inversion recovery (FLAIR) Sequenz das Infarktalter abgeschätzt werden [10]. Zudem ist im Vergleich zur CT die Sensitivität der MRT für den Nachweis infarzierter Areale in der Frühphase höher [11]. Ebenso ist die Darstellung von Gefäßpathologie und Perfusionsuntersuchung mittels KM-Gabe möglich. Neben CT und MRT kommen im Behandlungsverlauf auch die Doppler-Sonografie, Elektrokardiografie (EKG) und Kardioechografie zur weiteren Abklärung kardiovaskulärer Ursachen des Schlaganfalls zum Einsatz [9].

Für die Prognose ist vor allem eine zeitnahe Überführung und Behandlung in einer spezialisierten Einrichtung entscheidend, einer sogenannten Stroke-Unit [12]. In einer Stroke-Unit erfolgt unter anderem die kontinuierliche Erfassung von Blutdruck, Herzfrequenz, EKG, Atemfrequenz und Sauerstoffsättigung. Das ärztliche Personal dokumentiert zudem regelmäßig den neurologischen Status der Patient*innen, um ein Fortschreiten des Schlaganfalls oder Rezidive frühzeitig zu erkennen [9,13]. Die systemische Thrombolyse-Therapie und die mechanischen Thrombektomie sind die zentralen Bestandteile der Akuttherapie des Schlaganfalls. Es besteht die Möglichkeit einer intravenösen Gabe fibrinolytischer Medikamente (beispielsweise Alteplase) innerhalb eines Zeitfensters von viereinhalb Stunden nach Symptombeginn oder ab dem Zeitpunkt zu dem Patient*innen zuletzt gesund gesehen wurden [9]. Gemäß der Leitlinie wird innerhalb von sechs Stunden, sowie unter speziellen Bedingungen, eine endovaskuläre Therapie in Betracht gezogen, um das funktionelle Ergebnis mittels Rekanalisation zu verbessern [9,13]. Neben der akuten Therapie zählt auch die frühzeitige Rehabilitation zum Stroke-Unit-Konzept in Deutschland. Spezialisiertes Pflegepersonal sowie ein interdisziplinäres Team aus den Bereichen der Physiotherapie, Ergotherapie und Logopädie kommen dabei zum Einsatz [9].

Etwa die Hälfte aller Schlaganfallpatient*innen erleidet motorische Ausfälle, wobei Arm- und Beinpareesen zu den häufigsten Symptomen zählen [4]. Fast 70% der Betroffenen weisen dabei zumindest geringgradige Funktionseinschränkungen auf [1]. Dies ist besonders relevant hinsichtlich der Mobilität und Selbstversorgung sowie für die berufliche

und soziale Teilhabe der Patient*innen. Ausschlaggebend für eine Rückerlangung der bestmöglichen Funktionalität sind intensive physio- und ergotherapeutische Maßnahmen innerhalb der ersten sechs Monate [14,15]. In bis zu 80% aller Fällen treten zudem kognitive Störungen auf, welche sich durch Einschränkungen von Gedächtnis, Sprache und Aufmerksamkeit äußern [16]. Diese sind mit dem funktionellen Outcome der Patient*innen assoziiert, da die Teilnahme an Rehabilitationsmaßnahmen dadurch maßgeblich beeinflusst wird [13]. Neben einer genauen Diagnostik der einzelnen Funktionen zählen eine individuell angepasste Ergotherapie mit Hirnleistungstraining, das Erlernen von Kompensationsstrategien für den Alltag und die Schulung des sozialen Umfelds zu den wichtigsten Maßnahmen [17]. Hinzu kommt, dass innerhalb der ersten zwei Jahre nach Schlaganfall rund 30% aller Patient*innen eine klinisch relevante Depression entwickeln [18]. Diese wird als Post-Stroke Depression bezeichnet und ist vor allem mit rechtshemipariellen Schlaganfällen assoziiert [19]. Besonders begünstigt wird die Entwicklung einer Depression durch eine aus dem Schlaganfall resultierende Aphasie sowie durch eine bestehende depressive Vorerkrankung [20]. Zu den therapeutischen Maßnahmen zählen kognitive Verhaltenstherapie, Ausdauersport sowie eine pharmakologische Therapie bei persistierenden Beschwerden [21]. Die schlaganfallbedingte Mortalitätsrate innerhalb des ersten Monats beträgt etwa 30% und steigt danach sukzessive auf rund 40% innerhalb eines Jahres sowie bis auf 60% innerhalb von fünf Jahren an [22]. In der Akutphase trägt allen voran die Dysphagie zu einer erhöhten Rate an Aspirationspneumonien bei. Langfristig wird der Anstieg der Mortalität hingegen anderen Faktoren zugeschrieben, wie beispielsweise kardiovaskulären und onkologische Erkrankungen [23,24].

Im Jahr 2016 ereigneten sich weltweit rund 13,7 Millionen Schlaganfälle mit 5,5 Millionen Todesfällen [25]. Damit ist der Schlaganfall die weltweit zweithäufigste Todesursache und eine der bedeutensten Ursachen für Behinderung im Erwachsenenalter [26–28]. In Deutschland kam es 2008 zu etwa 262.000 Schlaganfällen, von denen 196.000 Erstmanifestationen waren und 66.000 Rezidive. Mit rund acht Prozent aller Todesfälle war der Schlaganfall damit die dritthäufigste Todesursache sowie die häufigste Ursache für Behinderung im Erwachsenenalter in Deutschland [29]. Frauen waren insgesamt mit rund 145.000 Fällen häufiger betroffen als Männer, wovon 108.000 erstmalige Ereignisse waren und 36.000 wiederholte. Die Gesamtzahl der Schlaganfälle bei Männern wurde auf 117.000 geschätzt, mit 88.000 Erstmanifestationen und 29.000 Rezidiven [29]. Die damit

assoziierten Kosten beliefen sich 2017 in Europa auf ca. 60 Milliarden Euro. Davon entfielen 27 Milliarden Euro auf das Gesundheitssystem, fünf Milliarden auf professionelle Pflege und soziale Versorgung, 13 Milliarden auf den Verlust von Produktivität und 16 Milliarden auf die Pflege durch Angehörige [30]. Die Gesamtkosten für Schlaganfall in Deutschland beliefen sich auf 17,6 Milliarden, was 0,54% des Bruttoinlandsproduktes entsprach. Die Kosten für die medizinische Versorgung betragen dabei etwa 9 Milliarden und machten 2,6% der gesamten Kosten im Gesundheitssystem aus. Die Aufwendungen für professionelle Pflege erreichten 800 Millionen und die Kosten für Pflege durch Angehörige beinahe fünf Milliarden. Einbußen in der wirtschaftlichen Produktivität durch Arbeitsausfälle oder vorzeitigen Tod wurden auf rund 2,5 Milliarden geschätzt. Deutschland war damit nach Estland das am zweitstärksten von Kosten betroffene Land in Europa [30]. Obwohl die Mortalitätsrate und Inzidenz in den westlichen Industrienationen seit 1990 rückläufig sind, steigt die absolute Anzahl an Schlaganfällen aufgrund der weltweiten demografischen Entwicklung [31–33]. Daher ist es ein dringliches wissenschaftliches und gesellschaftliches Anliegen jene Faktoren zu identifizieren, welche den klinischen Verlauf dieser Erkrankung mitbestimmen.

1.2 Maschinelles Lernen

Der Begriff maschinelles Lernen (ML) wurde erstmals im Jahr 1959 von Arthur Lee Samuel geprägt [34]. Die Kernidee von ML besteht darin einen Computer zu befähigen autonom Aufgaben zu lösen. Im Vergleich zu statischen Programmen nutzen ML-Algorithmen keine expliziten Anweisungen, was eine Analyse großer und komplexer Datenmengen ermöglicht [35]. Dies geschieht unter Anwendung spezifischer mathematischer und statistischer Konzepte, welche zur Entwicklung der ML-Modelle genutzt werden. Das ML lässt sich grob in überwachtes und unüberwachtes Lernen einteilen. Beim überwachten Lernen wird der ML-Algorithmus auf bereits annotierten Daten trainiert und danach auf neue, unbekannte Daten angewendet. Zu den wichtigsten Vertretern in dieser Kategorie zählen neuronale Netzwerke, Stützvektormaschinen (SVM) und Entscheidungsbäume. Beim unüberwachten Lernen versucht der ML-Algorithmus hingegen in einem Datensatz selbstständig relevante Muster zu erkennen. Beispiele für Algorithmen aus dieser Kategorie sind selbstorganisierende Karten und Clusteranalysen [36].

In der medizinischen Forschung haben ML-Methoden in den letzten Jahrzehnten zahlreiche Anwendungen gefunden. Beispielsweise wurden sie bei der Klassifizierung medizinischer Bilddaten, Planung von Radiotherapien oder der Dosierung von Medikamenten genutzt [37–39]. Des Weiteren haben diese Technologien ihre Vielseitigkeit bezüglich personalisierter Medizin und klinischer Entscheidungsunterstützungssysteme (CDSS, englisch *Clinical Decision Support System*) erwiesen [40]. Trotzdem ist das volle Potential von ML-Methoden und ihrer Anwendung in der Medizin noch nicht vollständig ausgeschöpft [41,42]. In der Neurologie fand ML bisher insbesondere Anwendung in der Analyse von Bilddaten und Interpretation von Elektroenzephalografie-Daten [43]. Beispielsweise ermöglicht ML die automatische Segmentierung von hämorrhagischen Infarkten, die Erkennung früher Anzeichen einer Ischämie auf CT-Bildern sowie eine automatische Bestimmung des Alberta Stroke Program Early CT Scores (ASPECTS) [44] in der Frühphase des Schlaganfalls [45–47]. Mouridsen et al. betonen jedoch, dass der Vorhersage klinischer Langzeitergebnisse eine größere Bedeutung beigemessen werden sollte [48]. Es existieren ebenfalls bereits zahlreiche Studien, die sich mit der Prädiktion solcher Outcomes nach akutem, ischämischen Schlaganfall und deren Einflussfaktoren beschäftigen [49,50]. Die modifizierte Rankin Skala (mRS) gilt als der wichtigste Score zur Beurteilung funktionellen Outcomes nach Schlaganfall [51]. Aufgrund dieser hohen Relevanz fokussieren sich die meisten Studien auf die Prädiktion der mRS. Als weiterer Vertreter der funktionellen Outcome Scores wird auch der Barthel Index (BI) [52] in der wissenschaftlichen Literatur genannt. Goldstandard für die Diagnostik kognitiver Funktion sind der Montreal-Cognitive-Assessment-Test (MoCA) [53] und der Mini-Mental-Status-Test (MMSE, englisch *Mini-Mental State Examination*) [54]. In diesem Bereich ist die Studienlage wesentlich spärlicher als bei funktionellen Outcomes. Die ML-basierte Forschung bezüglich Depression nach Schlaganfall ist ebenfalls begrenzt. In der PROSpective Cohort with Incident Stroke Berlin (PROSCIS-B) Studie [55] wurde neben dem MMSE das modifizierte Telefoninterview für den kognitiven Status (TICS-M, englisch *Modified Telephone Interview for Cognitive Status*) [56] für Nachfolgeuntersuchungen herangezogen sowie die allgemeine Depressionsskala (CES-D, englisch *Center for Epidemiologic Studies Depression Scale*) [57] für Depression. Obwohl es somit bereits eine etablierte Forschung im Bereich der ML-basierten Outcome-Prognose nach Schlaganfall gibt, bestehen noch einige grundlegende Herausforderungen.

Die Anwendung von ML-Methoden in der Medizin birgt allgemein einige problematische Aspekte. Eine große Herausforderung sind die begrenzten medizinischen Datenmengen, welche zudem oft heterogen, unvollständig und fehlerhaft sind [58]. Erschwerend kommen die dezentrale Speicherung der Datensätze sowie die hohen Ansprüche an Datenschutz hinzu [59]. Ein weiteres Problem ist die oft mangelnde Generalisierbarkeit der ML-Modelle auf neue Daten wegen unzureichender interner oder externer Validierung [60]. Ebenfalls wird die Interpretierbarkeit von ML-Analysen als besonders wichtig erachtet, da diese einen Einfluss auf potenziell lebenswichtige Entscheidungen haben können [61]. Nicht zuletzt existieren auch ethische und rechtliche Bedenken im Zusammenhang mit der Anwendung dieser Methoden in der Medizin. Einige Studien haben beispielsweise demonstriert, dass diskriminierende Ergebnisse in der Anwendung von ML-Methoden vorkommen können [62]. Die Zuweisung von Verantwortung für die Entscheidungen der ML-Modelle stellt damit in der Medizin eine zentrale Frage dar. Hinsichtlich der ML-basierten Schlaganfallforschung im Speziellen erschwert die Heterogenität der methodologischen Herangehensweisen einen Vergleich der bestehenden Literatur. Besonders der Einsatz einer Vielzahl unterschiedlicher ML-Modelle und Leistungskennzahlen stellt hier eine Hürde dar. Zudem trägt die Diversität der Datensätze und der Einsatz sehr spezifischer Input-Variablen weiter zur Unübersichtlichkeit bei [50]. Im Gesamten betrachtet mangelt es somit an systematischer Forschung bezüglich dieser Einflüsse [63,64].

1.3 Fragestellung

Das primäre Ziel dieser Studie war es einen systematischen Vergleich ML-basierter Prädiktion von Outcomes nach erstmaligem ischämischem Schlaganfall zu erstellen. Im Vordergrund stand der Einsatz standardisierter ML-Algorithmen, ein einheitliches Patient*innenkollektiv mit identischen klinischen Parametern und die Messung der Ergebnisse mittels uniformer Leistungskennzahlen. Für die ML-Analyse wurde als lineares Modell eine SVM mit linearem Kernel (SVM-lin) sowie eine SVM mit Kernel der radialen Basisfunktion (SVM-rbf) und ein Entscheidungsbaum mit Gradientenverstärkung (GB, englisch *Gradient Boosting*) als nicht-lineare Modelle ausgewählt. Zu den genutzten Input-Daten zählten demografische, klinische, serologische und MRT-Variablen. Die untersuchten Outcomes beinhalteten funktionelle (mRS und BI) und kognitive Indikatoren (MMSE und TICS-M) sowie Maßstäbe zur Erfassung von Depression (CES-D) und Mortalität zu verschiedenen

Beobachtungszeitpunkten. Zusätzlich wurden als sekundärer Endpunkt die wichtigsten Input-Variablen für jedes individuelle Ergebnis mithilfe von Shapley-Werten erhoben.

2 Methodik

2.1 Datensatz

Der zugrundeliegende Datensatz sowie die im Folgenden beschriebenen ML-Analysen entsprechen der von Fast et al. geschilderten Methodik [50]. Als Datengrundlage diente die prospektive Kohortenstudie PROSCIS-B, welche im März 2010 vom Centrum für Schlaganfallforschung Berlin (CSB) und der Charité - Universitätsmedizin Berlin initiiert wurde. Sie rekrutierte über einen Zeitraum von drei Jahren Proband*innen mit einer dreijährigen Periode für Nachfolgeuntersuchungen. Zu den Einschlusskriterien zählten ein Alter über 18 Jahre, Eintritt des ersten akuten Schlaganfalles gemäß der WHO Kriterien innerhalb der letzten sieben Tage und eine schriftliche Einwilligung zur Studienteilnahme durch die Patient*innen oder durch deren rechtliche*n Vertreter*in [55]. Als Pathologien wurden der ischämische Schlaganfall, intrakranielle Hämorrhagien und Sinusvenenthrombosen laut den Kriterien der WHO berücksichtigt [1]. Exkludiert wurden Patient*innen mit bereits vorangegangenen Schlaganfall laut WHO-Kriterien, bei Vorhandensein von zerebralen Tumorerkrankung oder Metastase, sowie bei zeitgleicher Teilnahme an einer Interventionsstudie oder einer Studie nach dem Arzneimittelgesetz. Eine Auflistung der Ein- und Ausschlusskriterien für PROSCIS-B befindet sich auf <https://www.clinicaltrials.gov> (NCT01363856) [55].

2.1.1 Datenselektion

Im Rahmen dieser Arbeit wurden nur Patient*innen mit ischämischem Schlaganfall aus dem PROSCIS-B Datensatz eingeschlossen. Darüber hinaus wurden lediglich Input-Variablen mit einem Fehlwertanteil von weniger als 15% in den Originaldaten für die Analysen herangezogen [65]. Aus diesem Grund wurde der serologische Marker Procalcitonin nicht berücksichtigt, welcher laut Yan et al. [66] für die 30-Tages Mortalität nach Schlaganfall prognostisch relevant ist.

Um die Charakteristika der Schlaganfälle zu quantifizieren, wurden alle akuten und chronischen Infarktläsionen in der DWI TRACE und FLAIR Sequenz von zwei Studierenden manuell mit MRlcron [67] vom Center for Advanced Brain Imaging (University of South Carolina, Chris Rordan, USA) abgegrenzt. Die Supervision und Definition aller weiteren MRT-Variablen erfolgte durch zwei unabhängige, erfahrene Expertinnen - einer Neurora-

diologin und einer Neurologin mit über neun Jahren Erfahrung in der Schlaganfalldiagnostik. Aufgrund größerer Unterschiede in der Alters- und Geschlechterverteilung der Patient*innen wurde der Datensatz stratifiziert. Hierzu wurden alters- und geschlechtsspezifische Untergruppen erstellt (z.B.: Frauen zwischen 70-80 Jahren) und innerhalb dieser Untergruppen zufällig Patient*innen selektiert, bis keine signifikanten Unterschiede mehr innerhalb des Datensatzes vorhanden waren (bis $p \leq 0,1$). Dies sollte sogenanntes shortcut learning [68] unterbinden und sicherstellen, dass die Prädiktion der ML-Modelle nicht auf einer verzerrten Stichprobe in den Trainingsdaten basieren (z.B.: generell schlechtere klinische Outcomes bei Frauen aufgrund eines höheren Durchschnittsalters). Dies resultierte in einem balancierten Datenset mit 307 Patient*innen (siehe Abbildung 1 und Tabelle 1).

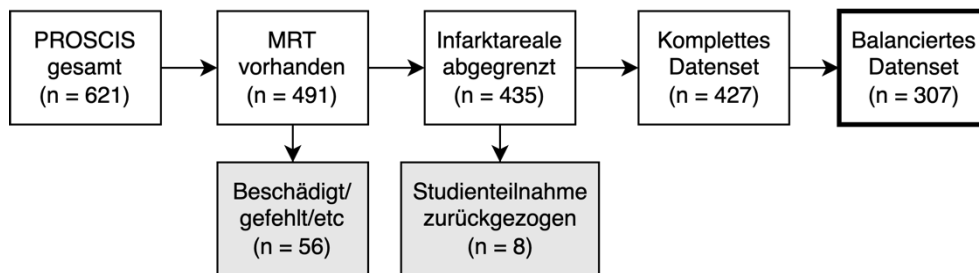


Abbildung 1: Selektionsprozess der Patient*innendaten.

Abkürzungen: PROSCIS, PROSpective Cohort with Incident Stroke; MRT, Magnetresonanztomografie

Quellenangabe: Abbildung modifiziert aus Fast et al. [50]

2.1.2 Input-Variablen

Es wurden insgesamt 43 Variablen unterteilt in vier Input-Subdomänen (demografisch, klinisch, serologisch und MRT) als Input für die ML-Analysen genutzt. Die demografische Input-Subdomäne einhielt die sechs Variablen biologisches Geschlecht, Alter, Körpermasseindex (BMI, englisch *Body-Mass-Index*), Bauchumfang, Hüftumfang und Bildung in Jahren. Die klinische Input-Subdomäne enthielt 16 Variablen mit systolischem und diastolischem Blutdruck, Tabak- und Alkoholkonsum, Pflegebedürftigkeit, körperliche Aktivität, Thrombolyse, Revaskularisierung, National Institutes of Health Stroke Scale (NIHSS) [69] bei Aufnahme sowie die Vorerkrankungen Diabetes Mellitus, Hypertonie, Vorhofflimmern, Angina Pectoris, Myokardinfarkt, periphere arterielle Verschlusskrankheit und Hypercholesterinämie. Die serologische Input-Subdomäne bestanden aus den

folgenden zehn Parametern: Glucose, glykiertes Hämoglobin (HbA1c), Gesamtcholesterol, high-density Lipoprotein (HDL), low-density Lipoprotein (LDL), Triglyceride, Kreatinin, geschätzte glomeruläre Filtrationsrate (eGFR, englisch *estimated glomerular filtration rate*), hochsensitives C-reaktives Protein (hsCRP, englisch *high-sensitivity C-reactive protein*) und Leukozytenzahl. In der MRT Input-Subdomäne wurden insgesamt elf Parameter zusammengefasst, darunter das Volumen des akuten Infarkts in DWI TRACE und FLAIR, die Seite des Infarkts, die supra-/infratentorieller Lokalisation, die uni-/bilaterale Verteilung, der vaskuläre Ursprung und das Muster des Infarkts, das Vorhandensein und Volumen von chronischen Infarkten in der FLAIR Sequenz sowie der Wahlund Score [70] und die trial of ORG 10172 in acute stroke treatment (TOAST) Klassifikation [71]. Diese vier Input-Subdomänen und der vollständige Datensatz wurden jeweils einzeln herangezogen, um die Outcomes vorherzusagen.

Tabelle 1: Beschreibung der Patient*innenpopulation.

Parameter	Gesamt	Frauen	Männer	p-Wert	
Demografisch					
Biologisches Geschlecht, n (%)	307	151 (49,2)	156 (50,8)		
Alter [Jahre], MW±SD	68±14	69±15	67±13	0,1	
Bildung [Jahre], MW±SD	14,1±4,6	12,9±4,3	15,3±4,6	0,0	
BMI, MW±SD	27,3±4,8	27,5±5,4	27,1±4,0	0,16	
Bauchumfang [mm], MW±SD	997,4±131,6	964,6±142,1	1028,1±113,0	0,0	
Hüftumfang [mm], MW±SD	1035,8±112,5	1034,7±128,5	1036,8±95,8	0,58	
Klinisch					
Blutdruck, MW±SD	Systolisch	140,1±22,3	139,8±22,8	140,3±21,7	0,57
	Diastolisch	76,1±13,7	74,7±13,0	77,4±14,3	0,03
Alkoholkonsum, n (%)	108 (35,2)	38 (25,2)	70 (44,9)	0,0	
Raucher*in, n (%)	Aktiv	102 (33,2)	39 (25,8)	63 (40,4)	0,1
	Nie	130 (42,3)	71 (47,0)	59 (37,8)	
	Früher	70 (22,8)	39 (25,8)	31 (19,9)	
Pflegebedürftig, n (%)	39 (12,7)	22 (14,6)	17 (10,9)	0,39	
Körperlich aktiv, n (%)	101 (32,9)	44 (29,1)	57 (36,5)	0,09	
Thrombolyse, n (%)	60 (19,5)	33 (21,9)	27 (17,3)	0,48	
Revaskularisierung, n (%)	5 (1,6)	0 (0,0)	5 (3,2)	0,04	
NIHSS, Median [IQR]	2 [1-4]	2 [1-4]	2 [1-4]	0,41	
Vorerkrankungen					
Diabetes Mellitus, n (%)	71 (23,1)	35 (23,2)	36 (23,1)	0,61	
Hypercholesterinämie, n (%)	66 (21,5)	32 (21,2)	34 (21,8)	0,95	
Hypertonie, n (%)	198 (64,5)	98 (64,9)	100 (64,1)	0,98	
Vorhofflimmern, n (%)	64 (20,8)	31 (20,5)	33 (21,2)	1,00	
Angina pectoris, n (%)	37 (12,1)	14 (9,3)	23 (14,7)	0,18	
Myokardinfarkt, n (%)	10 (3,3)	2 (1,3)	8 (5,1)	0,33	
pAVK, n (%)	15 (4,9)	7 (4,6)	8 (5,1)	0,95	

Serologisch					
Glucose [mmol/L], MW±SD		7,3±3,2	7,6±3,9	7,1±2,5	0,41
HbA1c [mmol/mol], MW±SD		10,9±15,5	12,2±18,9	9,5±10,7	0,21
Cholesterol [mmol/L], MW±SD		11,1±2,7	11,3±2,7	10,9±2,7	0,10
HDL [mmol/L], MW±SD		2,9±0,9	3,2±0,9	2,6±0,8	0,00
LDL [mmol/L], MW±SD		6,8±2,2	6,8±2,3	6,7±2,1	0,46
Triglyceride [mmol/L], MW±SD		7,6±4,8	7,0±4,5	8,1±5,1	0,00
Kreatinin [µmol/L], MW±SD		82,4±25,0	75,9±23,1	88,6±25,3	0,00
eGFR [ml/min], MW±SD		76,8±20,6	73,5±21,2	79,9±19,6	0,02
hsCRP [mg/L], MW±SD		1,2±1,9	1,3±2,1	1,0±1,7	0,15
Leukozyten [Zellen/mm ³], MW±SD		8,0±2,8	8,2±2,8	7,8±2,8	0,47
MRT					
Akuter Infarkt DWI TRACE [ml], MW±SD		6,1±14,5	5,6±11,4	6,6±17,1	0,17
Akuter Infarkt FLAIR [ml], MW±SD		5,0±12,9	4,6±10,4	5,4±15,0	0,11
Supra-/ Infratentoriell, n (%)	Supratentoriell	225 (73,3)	116 (76,8)	109 (69,9)	0,18
	Infratentoriell	52 (16,9)	21 (13,9)	31 (19,9)	
	Beides	30 (9,8)	14 (9,3)	16 (10,3)	
Seite des Infarkts, n (%)	Links	138 (45,0)	67 (44,4)	71 (45,5)	0,26
	Rechts	132 (43,0)	69 (45,7)	63 (40,4)	
	Beides	37 (12,1)	15 (9,9)	22 (14,1)	
Unilateraler Infarkt, n (%)		270 (87,9)	136 (90,1)	134 (85,9)	0,46
Chronischer Infarkt, n (%)		79 (25,7)	37 (24,5)	42 (26,9)	0,90
Chronischer Infarkt [ml], MW±SD		1,5±2,5	1,3±2,4	1,6±2,6	0,34
Wahlund Score [70], Median [IQR]		6 [3-10]	6 [3,5-11]	5 [2-8]	0,02
Ursprung des Infarkts, n (%)	MCA	143 (46,6)	75 (49,7)	68 (43,6)	0,58
	ACA	1 (0,3)	1 (0,7)	0 (0,0)	
	PCA	17 (5,5)	10 (6,6)	7 (4,5)	
	AchA	18 (5,9)	8 (5,3)	10 (6,4)	
	Infratentoriell	52 (16,9)	21 (13,9)	31 (19,9)	
	Thalamus	19 (6,2)	8 (5,3)	11 (7,1)	
	Multiple	57 (18,6)	28 (18,5)	29 (18,6)	
Muster des Infarkts, n (%)	Territorial	96 (31,3)	52 (34,4)	44 (28,2)	0,61
	Subkortikal	74 (24,1)	37 (24,5)	37 (23,7)	
	Scattered	72 (23,5)	35 (23,2)	37 (23,7)	
	Lakunar	1 (0,3)	0 (0,0)	1 (0,6)	
	Infratentoriell	52 (16,9)	21 (13,9)	31 (19,9)	
TOAST, n (%)	Makroangiopathie	88 (28,7)	46 (30,5)	42 (26,9)	0,4
	Kardiale Embolie	89 (29,0)	50 (33,1)	39 (25,0)	
	Mikroangiopathie	14 (4,6)	8 (5,3)	6 (3,8)	
	Andere Ursachen	40 (13,0)	17 (11,3)	23 (14,7)	
	Unklare Ätiologie	76 (24,8)	30 (19,9)	46 (29,5)	

Die Darstellung der Daten erfolgt durch Mittelwert mit Standardabweichung (MW±SD, englisch *standard deviation*) für kontinuierliche Variablen, Median mit Interquartilsabstand [25.-75. Perzentile] für ordinale Variablen und absolute (n) sowie relative (%) Häufigkeit für kategoriale Variablen. Um signifikante Unterschiede zwischen Frauen und Männern zu untersuchen, wurde ein t-Test für kontinuierliche Variablen sowie ein Chi-Quadrat-Test für kategoriale Variablen durchgeführt und die resultierenden p-Werte ermittelt.

Abkürzungen: ACA, Arteria cerebri anterior; AchA, Arteria choroidea anterior; BMI, Körpermasseindex (englisch *Body-Mass-Index*); DWI, Diffusions-gewichtete Sequenz (englisch *Diffusion-weighted imaging*); eGFR, Geschätzte glomeruläre Filtrationsrate (englisch *estimated glomerular filtration rate*); FLAIR, Fluid attenuated inversion recovery; HDL, High-density Lipoprotein; HbA1c, Glykiertes Hämoglobin; hsCRP, Hochsensitives C-reaktives Protein; IQR, Interquartilabstand; LDL, Low-density Lipoprotein; MCA, Arteria cerebri media; MW±SD, Mittelwert±Standardabweichung (englisch *standard deviation*); NIHSS, National Institute of Health Stroke Scale; pAVK, Periphere arterielle Verschlusskrankheit; PCA, Arteria cerebri posterior; TOAST, Trial of ORG 10172 in Acute Stroke Treatment

Quellenangabe: Tabelle modifiziert aus Fast et al. [50]

2.1.3 Klinische Outcomes

Das funktionelle Outcome wurde mit mRS [72] und BI [52] bei Aufnahme und nach einem Jahr gemessen. Kognitive Beeinträchtigung wurde mittels MMSE [54] bei Aufnahme sowie TICS-M [56] nach einem und drei Jahren quantifiziert. Depression wurde mittels CES-D [57] nach einem und drei Jahren gemessen. Die Mortalität wurde innerhalb von einem und drei Jahren betrachtet. Die Verteilung dieser Outcomes innerhalb der Patient*innenpopulation sowie die Grenzwerte für gutes versus schlechtes klinisches Outcome laut Goldstandards können der Tabelle 2 entnommen werden.

Nach einem Schlaganfall ist die mRS das wichtigste Instrument zur Messung des funktionellen Outcomes. Die Punkteskala reicht von null mit "Keine Einschränkung" über fünf mit "Schwere Beeinträchtigung, bettlägerig, inkontinent, auf Pflege angewiesen" bis hin zu sechs, dem Tod [72]. Der BI ermöglicht eine systematische Erfassung grundlegender Alltagstätigkeiten, der sogenannten Aktivitäten des täglichen Lebens. Er umfasst Kategorien wie Nahrungsmittelaufnahme, Körperpflege, Kontinenz und Mobilität und wird von 0-100 Punkten bewertet [52]. Der MMSE erfasst mit insgesamt 30 Punkten die kognitiven Fähigkeiten von Patient*innen. Zu den abgefragten Bereichen zählen die Orientierung, Merkfähigkeit, Aufmerksamkeit, Rechenfähigkeit, Erinnerungsfähigkeit und Sprache [54]. Das TICS-M ist ein telefonischer Fragebogen zur Erfassung kognitiver Fähigkeiten mit 13 Fragen und 0-50 Punkten. Es werden unter anderem Orientierung, Merkfähigkeit und Rechenfähigkeit betrachtet [56]. Die CES-D ist ein Fragebogen zur Beurteilung von Depression mit 0-60 Punkten. Er besteht aus 20 Fragen zur affektiven Lage von Patient*innen und beinhaltet beispielsweise Fragen zu Appetit, Antrieb und Schlaf [57].

Tabelle 2: Grenzwerte und Verteilung der Outcomes in der Patient*innenpopulation.

Verteilung der Outcomes innerhalb der Patient*innenpopulation			
Outcome	Zeitpunkt	Gutes Outcome, n (g/f/m)	Schlechtes Outcome, n (g/f/m)
mRS	bE	221/110/111	86/41/45
	Jahr 1	193/89/104	40/27/13
BI	bE	263/125/138	44/26/18
	Jahr 1	195/90/10	7/6/1
MMSE	bE	271/126/145	29/21/8
TICS-M	Jahr 1	147/69/78	48/32/16
	Jahr 3	125/60/65	19/8/11
CES-D	Jahr 1	163/79/93	48/35/13
	Jahr 3	132/53/79	30/19/11
Mortalität	Jahr 1	271/132/139	36/19/17
	Jahr 3	142/78/64	165/73/92
Grenzwerte für gutes versus schlechtes Outcome			
Outcome	Gesamt	Gutes Outcome	Schlechtes Outcome
mRS	0-6	0-2	3-6
BI	0-100	61-100	0-60
MMSE	0-30	24-30	0-23
TICS-M	0-50	30-50	0-29
CES-D	0-60	0-15	16-60

Die Verteilung der Outcome ist angegeben in absoluten Zahlen für die Gesamtzahl der Patient*innen sowie getrennt nach Frauen und Männern (g/f/m).

Abkürzungen: BI, Barthel Index; bE, bei Entlassung; CES-D, Allgemeine Depressionsskala (englisch *Center for Epidemiologic Studies Depression Scale*); mRS, Modifizierte Rankin Skala; MMSE, Mini-Mental-Status-Test (englisch *Mini-Mental State Examination*); TICS-M, Modifiziertes Telefoninterview für den kognitiven Status (englisch *Modified Telephone Interview for Cognitive Status*)

Quellenangabe: Tabelle modifiziert aus Fast et al. [50]

2.2 Maschinelles Lernen

Die ML-Analyse bestand im Kern aus den drei Algorithmen SVM-lin, SVM-rbf und GB. Die Input-Daten wurden vorverarbeitet und danach sowohl im Gesamten als auch in den einzelnen Input-Subdomänen (demografisch, klinisch, serologisch, MRT) für die Berechnung herangezogen (siehe Abbildung 2). Um robuste Ergebnisse zu generieren wurden die ML-Modelle mittels Rastersuche und einer wiederholten fünffach verschachtelten Kreuzvalidierung optimiert [73]. Die Auswertung der Ergebnisse erfolgte anhand des Mittelwertes (MW) mit Standardabweichung (SD, englisch *Standard deviation*) der Leistungskennzahlen ausgewogene Genauigkeit (BA, englisch *Balanced Accuracy*), Fläche

unter der Grenzwertoptimierungskurve (AUC, englisch *Area under the receiver operating characteristic curve*), Sensitivität, Spezifität, Plausibilitätsquotiententest (LR, englisch *Likelihood ratio test*) [74] und Integrated Discrimination Improvement (IDI) [75] sowie mittels eines Permutations-Tests mit Bonferroni-Korrektur. Zur Bestimmung der wichtigsten Prädiktoren für jedes Outcome wurden Shapley-Werte ermittelt. Shapley-Werte sind ein statistischer Erklärungsansatz, um den Einfluss einzelner Variablen auf das Prädiktionsergebnis zu quantifizieren [76]. Für alle Analysen und Visualisierungen wurde die Python 3.6 Programmiersprache mit den Paketen scikit-learn, pandas, statsmodel, matplotlib und seaborn verwendet.

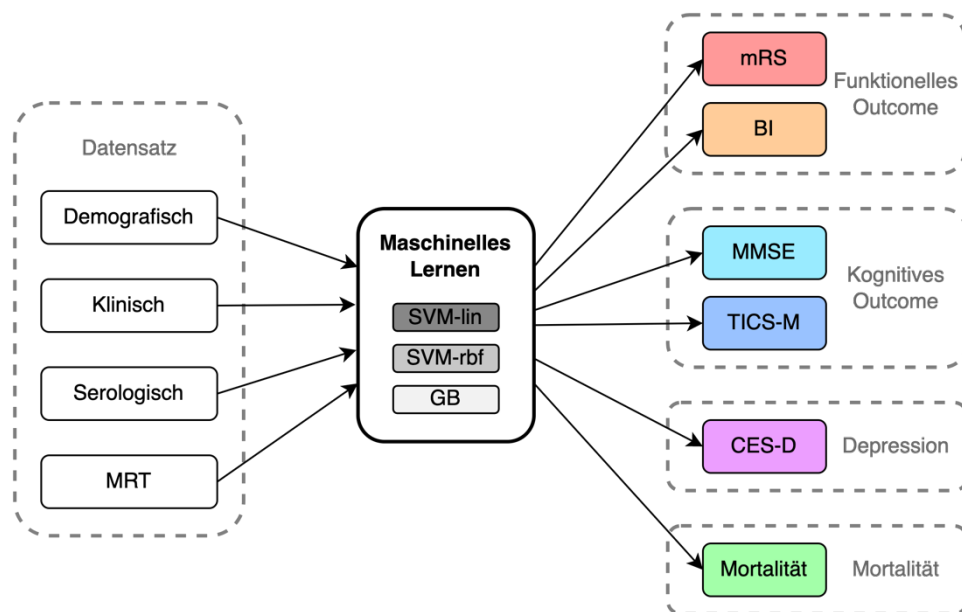


Abbildung 2: Überblick der Analysen.

Die Input-Daten werden sowohl im Gesamten als auch in vier Input-Subdomänen zur Analyse mittels drei maschinellen Lernalgorithmen herangezogen. Es erfolgt eine Prädiktion von gutem versus schlechtem Outcome für insgesamt sechs klinische Maßstäbe in vier Kategorien.

Abkürzungen: CES-D, Allgemeine Depressionsskala (englisch *Center for Epidemiologic Studies Depression Scale*); GB, Entscheidungsbaum mit Gradientenverstärkung (englisch *Gradient Boosting*); mRS, Modifizierte Rankin Skala; MMSE, Mini-Mental-Status-Test (englisch *Mini-Mental State Examination*); MRT, Magnetresonanztomografie; SVM-lin, Stützvektormaschine mit linearem Kernel; SVM-rbf, Stützvektormaschine mit Kernel der radialen Basisfunktion; TICS-M, Modifiziertes Telefoninterview für den kognitiven Status (englisch *Modified Telephone Interview for Cognitive Status*)

Quellenangaben: Abbildung modifiziert aus Fast et al. [50]

2.2.1 Vorverarbeitung

Im ersten Schritt wurde der Datensatz manuell kontrolliert und Eingabefehler korrigiert. Danach wurden fehlende Datenpunkten mittels multipler Imputation durch verkettete Gleichungen (MICE, englisch *Multiple Imputation using Chained Equations*) [65] kompensiert. Die MICE-Methode läuft in mehreren Schritten ab, angefangen bei der Erzeugung eines Platzhalters (z.B. MW) für die fehlenden Datenpunkte. Im nächsten Schritt wird einer dieser Platzhalterwerte wiederum entfernt und mittels Regression anhand des restlichen Datensatzes erneut ermittelt. Dieser Prozess wird nun so lange wiederholt, bis alle ursprünglich fehlenden Werte imputiert wurden. Dieser Vorgang kann mehrfach durchlaufen werden, wobei die imputierten Werte jedes Mal erneut aktualisiert werden [65]. Entsprechend der Erkenntnisse von Sterne et al. [77] wurde MICE im Rahmen unserer Studie lediglich auf die Daten des Trainings- und Validierungs-Sets angewendet. Fehlende Daten im Test-Set wurden durch den entsprechenden MW aus dem Trainings- und Validierungs-Set ersetzt.

In Tabelle 2 wird die Verteilung der Outcomes innerhalb der Patient*innenpopulation dargestellt, wobei ein Missverhältnis der meisten Outcome-Klassen (gutes versus schlechtes Outcome) deutlich wird. Da eine solche Verzerrung ein effektives Training der ML-Algorithmen erschwert, wurden die Synthetic Minority Oversampling Technique (SMOTE) [78] und randomisiertes Oversampling [79] im Trainings-Set eingesetzt. SMOTE generiert statistisch ähnliche Beispiele einer unterrepräsentierten Klasse während randomisiertes Oversampling bereits vorhandene Beispiele dupliziert. Aus technischen Gründen wurden zudem alle kategorischen Input-Variablen mittels one-hot encoding [80] transformiert.

2.2.2 Algorithmen

Die Analysen wurden auf Basis der drei ML-Algorithmen SVM-lin, SVM-rbf und GB durchgeführt. Die entsprechenden Hyperparameter der ML-Modelle wurden im Rahmen einer Rastersuche optimiert (siehe Tabelle 3). Eine SVM zeichnet sich als mathematisches Verfahren dadurch aus, dass Input-Daten in einen höhendimensionalen Raum transformiert werden, um diese linear separieren zu können. Der Algorithmus ermittelt dafür eine Trennebene zwischen den Datenpunkten der verschiedenen Outcome-Klassen. Die SVM strebt danach einen möglichst großen Abstand zu den Datenpunkten in der Nähe der Trennebene einzuhalten, was mithilfe sogenannter Stützvektoren (englisch *support vec-*

tors) erreicht wird [81]. Die SVM-lin wurde mit den zwei Hyperparametern k und C optimiert. Dabei steht k für die Anzahl der verwendeten Input-Features und C als ein Maß für die Regularisierung. Regularisierung beschränkt die Komplexität eines ML-Modells, damit es adäquat auf neue Daten verallgemeinern kann. Bei k wurden stets alle Variablen verwendet, da wir in dieser Arbeit bereits eine Vorauswahl anhand von Input-Subdomänen getroffen haben. Bei der SVM-rbf waren k und der Kernel ebenfalls bereits definiert, während C und Gamma optimiert wurden. Der Parameter Gamma bestimmte dabei den Einfluss eines einzelnen Trainingsbeispiels auf das finale Ergebnis. Der GB basiert auf einem mehrstufigen Entscheidungsprozess, in dem alle möglichen Ergebnisse einer Entscheidung exploriert werden. Diese Methode zeichnet sich durch iteratives Lernen aus, wobei zusätzlich auch falsch prädizierte Beispiele berücksichtigt werden [81]. Der Hyperparameter `max_depth` beschreibt die maximale Anzahl der Verzweigungen pro Entscheidungsbaum und `max_features` hingegen die Menge an Input-Variablen. Der Wert `n_estimators` bezeichnet die Anzahl der Berechnung für das finale Ergebnis, `min_samples_split` die Anzahl der notwendigen Beispiele innerhalb des Rechenganges und `min_samples_leaf` die Mindestzahl der Beispiele für die jeweils letzte Berechnung [82].

Tabelle 3: Hyperparameter der maschinellen Lernmodelle.

Modell	Hyperparameter	Werte
SVM-lin	<code>k</code>	all
	<code>C</code>	1; 10; 100; 1000
SVM-rbf	<code>k</code>	all
	<code>kernel</code>	rbf
	<code>C</code>	1; 10; 100; 1000
	<code>Gamma</code>	0,001; 0,0001
GB	<code>max_depth</code>	8; 16; 32; 40; 50
	<code>max_features</code>	sqrt; auto; log2
	<code>n_estimators</code>	25; 50; 75; 100
	<code>min_samples_split</code>	0,1; 0,3
	<code>min_samples_leaf</code>	0,3; 0,5

Abkürzungen: GB, Entscheidungsbaum mit Gradientenverstärkung (englisch *Gradient Boosting*); SVM-lin, Stützvektormaschine mit linearem Kernel; SVM-rbf, Stützvektormaschine mit Kernel der radialen Basisfunktion

Quellenangabe: Tabelle modifiziert aus Fast et al. [50]

2.2.3 Kreuzvalidierungsverfahren

Grundsätzlich werden ML-Modelle optimiert, um das bestmögliche Prädiktionsergebnis anhand der Trainingsdaten zu erzielen. Sobald das ML-Modell auf neue Daten (z.B. aus dem Test-Set) angewendet wird, sinkt daher in der Regel die Leistung. Der Unterschied ist typischerweise vor allem dann besonders groß, wenn die Anzahl der Beispiele im Trainings-Set gering ist. Um diese Problematik zu umgehen können sogenannte Kreuzvalidierungsverfahren eingesetzt werden [83]. In dieser Studie wurde eine fünffach verschachtelte Kreuzvalidierung mit zehn Wiederholungen genutzt (siehe Abbildung 3). Diese bestand aus einer äußeren und einer inneren Schleife für Training, Validierung und Testung. In der äußeren Schleife wurde der gesamte Datensatz fünfmal in zufällig gewählte Training-Sets (80%) und Test-Sets (20%) separiert. In der inneren Schleife wurde danach jedes Trainings-Set in fünf weitere verschachtelte Training-Sets (80%) und Validierungs-Sets (20%) gesplittet. Die innere Schleife wurde genutzt, um die ML-Modelle zu trainieren und deren Hyperparameter zu optimieren, bevor diese an den Test-Sets der äußeren Schleife getestet wurden. Die Berechnungen dieser Analysen wurden als Gesamtergebnis der wiederholten fünffach verschachtelten Kreuzvalidierung gewertet. Dieser gesamte Prozess wurde zehnfach wiederholt und die Ergebnisse als MW mit Konfidenzintervall (KI) angegeben.

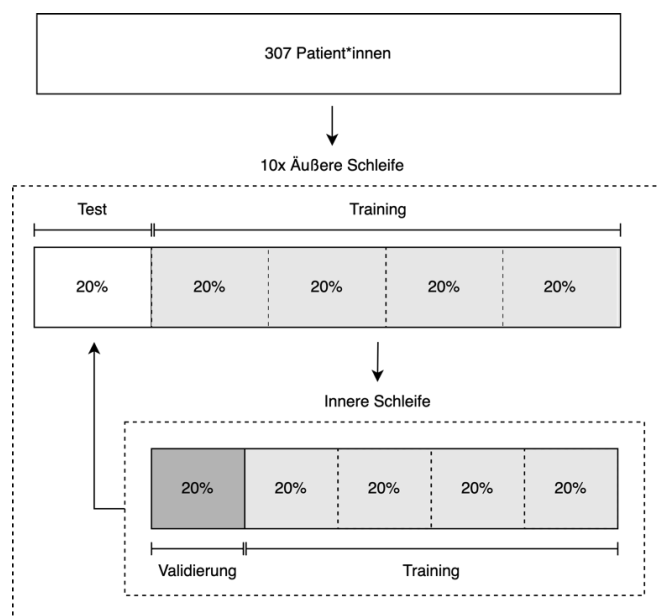


Abbildung 3: Ablauf des Kreuzvalidierungsverfahrens.

Quellenangabe: Eigene Abbildung.

2.2.4 Leistungskennzahlen

Die Leistung der ML-Modelle wurde mit BA, AUC, Sensitivität, Spezifität, LR und IDI sowie mittels eines Permutationstest mit Bonferroni-Korrektur verglichen. Die BA ist der MW aus Sensitivität und Spezifität. Die Sensitivität oder Richtig-positiv-Rate gibt die Wahrscheinlichkeit an mit der ein richtig positives Objekt auch korrekt positiv klassifiziert wird. Die Spezifität oder Richtig-negativ-Rate hingegen beschreibt die Wahrscheinlichkeit mit der ein richtig negatives Objekt auch negativ klassifiziert wird [84]. Die AUC gibt die Fläche unter der Grenzwertoptimierungskurve an. Diese ergibt sich aus der Gegenüberstellung von Sensitivität und Falsch-positiv-Rate. Das LR beschreibt die relative Übereinstimmung eines ML-Modells mit den beobachteten Daten. Dabei werden das getestete Modell und ein Vergleichsmodell geprüft und bewertet [85]. Das IDI ist ein quantitatives Maß für die Verbesserung der Diskriminationsleistung eines Modells im Vergleich zu einer Referenzstrategie. Die Berechnung des IDI erfolgt durch Integration der Änderungen in der Steigung der Diskriminierungskurve [75].

2.2.5 Permutationstest

Die statistische Signifikanz der Ergebnisse wurde mittels Permutationstest ermittelt. Hierbei handelt es sich um einen exakten Test aus der nichtparametrischen Statistik. Es werden zufällige Stichprobenwiederholungen unter Annahme der Nullhypothese gezogen und die resultierende Verteilung der Teststatistik bestimmt. Konkret bedeutete diese im Rahmen unserer Studie, dass die Labels des Datensatzes zufällig permutiert und die gleichen ML-Analysen wie an den Originaldaten durchgeführt wurden. Dieser Vorgang wurde hundertfach wiederholt und jeweils mit den ursprünglichen Ergebnissen verglichen. Um der Problematik von multiplen Vergleichen entgegenzuwirken, wurde zusätzlich eine Bonferroni-Korrektur (drei ML-Algorithmen x fünf Input-Subdomänen) angewendet. Die Ergebnisse galten nur dann als statistisch signifikant, wenn der p-Wert $\leq 0,05$ beziehungsweise $\leq 0,01$ lag.

2.2.6 Shapley-Werte

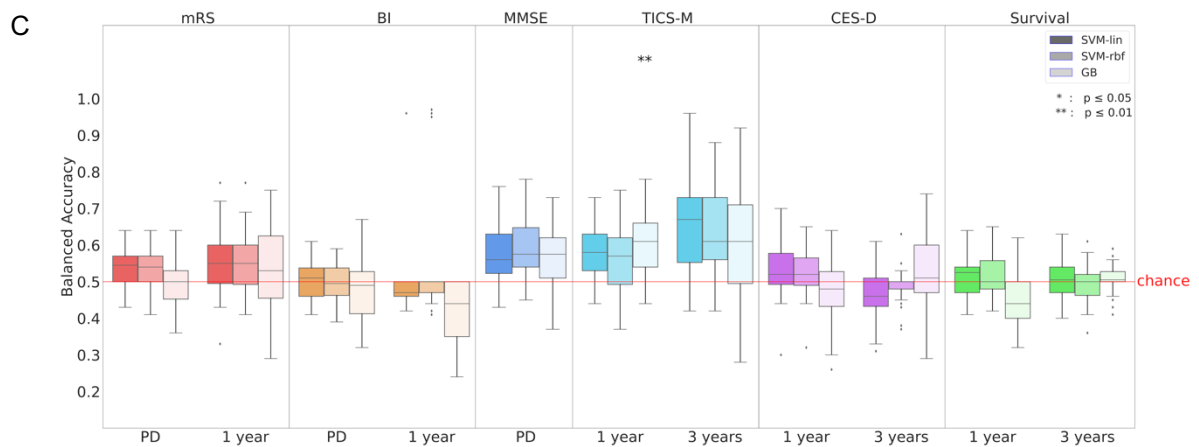
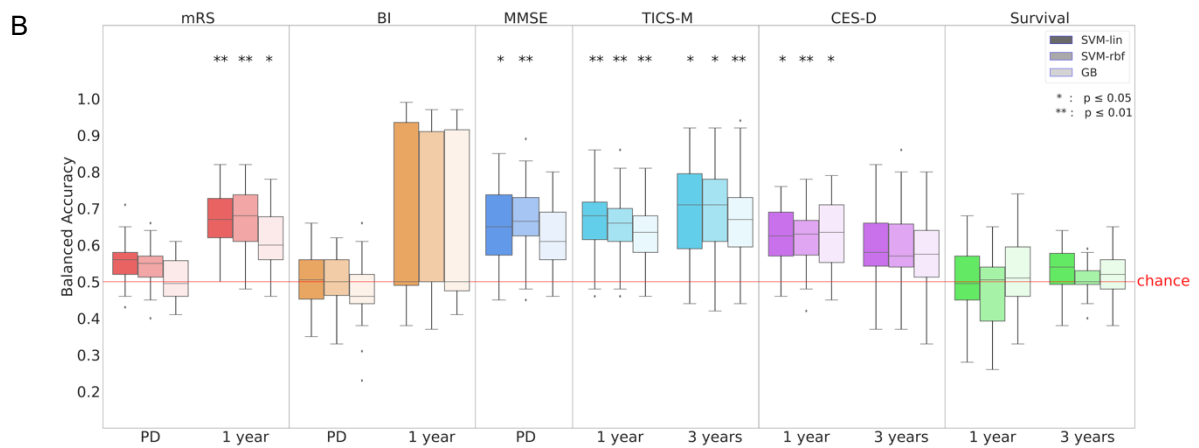
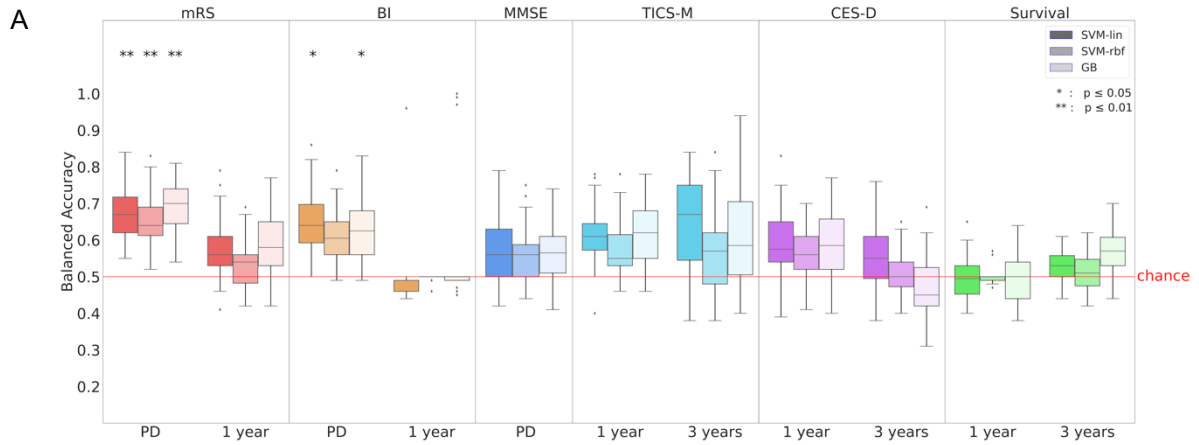
Shapley-Werte sind ein statistisches Konzept aus der kooperativen Spieltheorie. Sie wurden 1951 von Lloyd Shapley entwickelt, um den Beitrag individueller Spieler*innen zum finalen Ergebnis zu berechnen [86]. In Hinsicht auf ML kann dieses Konzept angewendet werden, um den Beitrag einzelnen Variablen zum Prädiktionsergebnis zu ermitteln. Die-

ser Erklärungsansatz wurde bereits mehrfach erfolgreich in biomedizinischer und klinischer Forschung angewandt [76]. Die Berechnung erfolgt anhand der Auswirkung auf das Ergebnis, wenn eine Variable in der Analyse enthalten ist, beziehungsweise wenn diese ausgeschlossen wird. Anschließend werden die Ergebnisse aller Variablen gemittelt und die Shapley-Werte berechnet. In unserer Studie wurden die Shapley-Werte als absoluter MW aller Wiederholungen der fünffach verschachtelten Kreuzvalidierung angegeben. Die Implementierung erfolgte mithilfe des SHAP (SHapley Additive exPlanations) Frameworks [87]. Der Kernel SHAPexplainer nutzt gewichtete lineare Regression, um die Relevanz der einzelnen Variablen zu bestimmen.

3 Ergebnisse

Von den 621 Patient*innen in PROSCIS-B erhielten 496 eine MRT-Bildgebung mit DWI TRACE und FLAIR Sequenzen auf Geräten der Firma Siemens, wobei in 5 Fällen die Bilddaten nicht gefunden werden konnten. Von den verbleibenden 491 Patient*innen wurden 255 mit einem 3T Scanner am CSB untersucht und 236 an der Charité - Universitätsmedizin Berlin mit Magnetfeldstärken zwischen 1-1,5T. In 56 Fälle konnten aufgrund fehlender Sequenzen oder Bewegungsartefakten die Infarktareale nicht abgegrenzt werden. Weitere 8 Patient*innen hatten zudem ihre Einwilligung in die Studie widerrufen. Daraus resultierten 427 Patient*innen und schlussendlich ein balancierter Datensatz von 307 Patient*innen. Die Verlustraten der Nachfolgeuntersuchungen betragen 24,1% für mRS, 34,2% für BI, 26,2% für TICS-M und 23,2% für CES-D, während bei der Mortalität keine Verluste auftraten.

Die Leistung der ML-Modelle wurden anhand von BA und AUC gemessen. Als weiterer Maßstab wurden IDI und LR berechnet, wobei sich bei LR keine signifikanten Unterschiede zwischen den ML-Modellen zeigten. Hier ist es allerdings wichtig zu beachten, dass diese Ergebnisse unabhängig voneinander betrachtet werden sollten, da diese auf verschiedenen Methoden zur Bewertung der Prädiktionsleistung basieren. Abbildung 4 zeigt eine Visualisierung der Ergebnisse, angegeben in BA für alle Outcomes, Zeitpunkte und ML-Modelle. In Tabelle 4 sind die wichtigsten Prädiktoren für alle Outcomes gemäß der Shapley-Werte angegeben. Das beste Ergebnis in dieser Arbeit wurde durch SVM-rbf für TICS-M nach drei Jahren mit einer $BA \pm SD = 0,7 \pm 0,13$ und einer $AUC \pm SD = 0,76 \pm 0,13$ in der demografischen Input-Subdomäne erzielt. Im folgenden Abschnitt werden die jeweils besten Prädiktionsergebnisse für jedes Outcome und jeden Zeitpunkt beschrieben.



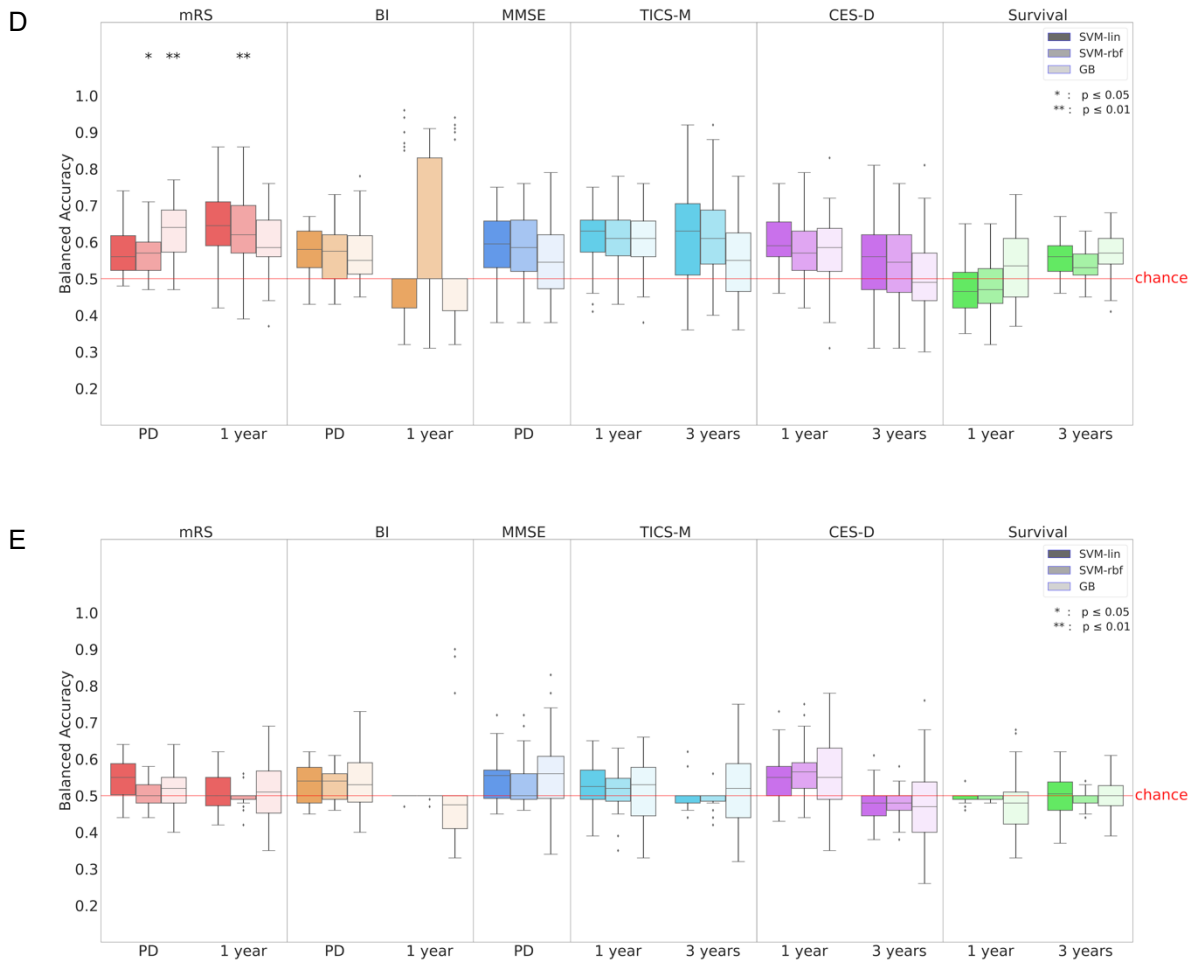


Abbildung 4: Prädiktionsergebnisse angegeben in ausgewogener Genauigkeit (BA, englisch *Balanced accuracy*) für alle Outcomes, Zeitpunkte und Input-Subdomänen.

In A wurden alle Input-Variablen in Betracht gezogen, während in B – E die Ergebnisse der B demografischen, C klinischen, D serologischen und E MRT Input-Subdomäne gezeigt werden. Aufgrund der ungleichen Verteilung innerhalb des Datensatzes (siehe Tabelle 2) waren die Ergebnisse für BI nach einem Jahr nicht eindeutig.

Abkürzungen: BI, Barthel Index; CES-D, Allgemeine Depressionsskala (englisch *Center for Epidemiologic Studies Depression Scale*); GB, Entscheidungsbaum mit Gradientenverstärkung (englisch *Gradient Boosting*); MMSE, Mini-Mental-Status-Test (englisch *Mini-Mental State Examination*); MRT, Magnetresonanztomografie; mRS, Modifizierte Rankin Skala; PD, bei Entlassung (englisch *at patient discharge*); SVM-lin, Stützvektormaschine mit linearem Kernel; SVM-rbf, Stützvektormaschine mit Kernel der radialen Basisfunktion; TICS-M, Modifiziertes Telefoninterview für den kognitiven Status (englisch *Modified Telephone Interview for Cognitive Status*)

Quellenangaben: Abbildungen aus Fast et al. [50]

3.1 Modifizierte Rankin Skala

Das beste Ergebnis für mRS bei Entlassung wurden mittels GB mit $BA \pm SD = 0,69 \pm 0,07$ und einer $AUC \pm SD = 0,77 \pm 0,06$ unter Verwendung aller Input-Variablen erzielt. Die fünf wichtigsten Prädiktoren waren NIHSS, hsCRP, Glucose, Cholesterol und supra-/infratentorieller Infarkt. Für mRS nach einem Jahr erzielte SVM-rbf mit $BA \pm SD = 0,68 \pm 0,09$ und $AUC \pm SD = 0,73 \pm 0,01$ in der demografischen Input-Subdomäne das beste Ergebnis. Die führenden Variablen hierbei waren Bauchumfang, biologisches Geschlecht, Alter, Bildung und BMI.

3.2 Barthel Index

Für BI bei Entlassung zeigte SVM-lin mit $BA \pm SD = 0,65 \pm 0,08$ und $AUC \pm SD = 0,73 \pm 0,11$ die beste Leistung unter Verwendung aller Input-Variablen. Die stärksten Prädiktoren waren NIHSS, Rauchen, Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Klassifikation, Muster und Ursprung des Infarkts. Für BI nach einem Jahr konnten keine signifikanten Resultate erreicht werden.

3.3 Mini-Mental-Status-Test

Das beste Ergebnis für MMSE bei Entlassung erzielte SVM-rbf mit $BA \pm SD = 0,67 \pm 0,09$ und $AUC \pm SD = 0,71 \pm 0,11$ in der demografischen Input-Subdomäne, mit Bildung, biologischem Geschlecht, Alter, Bauchumfang und BMI als wichtigste Prädiktoren.

3.4 Modifiziertes Telefoninterview für den kognitiven Status

Für TICS-M nach einem Jahr demonstrierte SVM-lin mit $BA \pm SD = 0,67 \pm 0,09$ und $AUC \pm SD = 0,73 \pm 0,09$ die beste Leistung in der demografischen Input-Subdomäne. Die führenden Variablen waren Bildung, Alter, BMI, biologisches Geschlecht und Hüftumfang. TICS-M nach drei Jahren wurde am besten durch SVM-rbf in der demografischen Input-Subdomäne modelliert, mit $BA \pm SD = 0,7 \pm 0,13$ und $AUC \pm SD = 0,76 \pm 0,13$. Bildung, Alter, biologisches Geschlecht, Bauchumfang und Hüftumfang zählten hier zu den wichtigsten Variablen.

3.5 Allgemeine Depressionsskala

Bei der Prädiktion der CES-D nach einem Jahr war GB mit $BA \pm SD = 0,63 \pm 0,09$ und $AUC \pm SD = 0,7 \pm 0,1$ in der demografischen Input-Subdomäne am erfolgreichsten. Die stärksten Prädiktoren waren Bildung, biologisches Geschlecht, BMI sowie Bauch- und Hüftumfang. Es konnte jedoch kein signifikantes Ergebnis für CES-D nach drei Jahren erzielt werden.

3.6 Mortalität

Es wurden keine signifikanten Ergebnisse bezüglich der Mortalität innerhalb von einem und drei Jahren festgestellt.

Tabelle 4: Die besten Prädiktionsergebnisse und wichtigsten Prädiktoren laut Shapley-Werten für alle Ergebnisse.

Outcome	Zeitpunkt	Modell	Input	Absoluter SHAP-Mittelwert	Variable
mRS	bE	GB	Alle Variablen	0,68 [0,64-0,72]	NIHSS
				0,44 [0,41-0,47]	hsCRP
				0,21 [0,19-0,24]	Glucose
				0,18 [0,15-0,21]	Cholesterol
				0,18 [0,15-0,20]	Supra-/Infratentoriell
	Jahr 1	SVM-rbf	Demografisch	0,52 [0,47-0,57]	Bauchumfang [cm]
				0,50 [0,46-0,54]	Biologisches Geschlecht
				0,47 [0,43-0,51]	Alter
				0,37 [0,33-0,41]	Bildung [Jahre]
				0,19 [0,15-0,22]	BMI
BI	bE	SVM-lin	Alle Variablen	1,11 [1,05-1,18]	NIHSS
				0,61 [0,57-0,65]	Raucher
				0,46 [0,42-0,49]	TOAST Klassifikation
				0,41 [0,36-0,45]	Muster des Infarkts
				0,37 [0,34-0,41]	Ursprung des Infarkts
TICS-M	Jahr 1	SVM-lin	Demografisch	0,68 [0,62-0,73]	Bildung [Jahre]
				0,51 [0,46-0,56]	Alter
				0,40 [0,34-0,46]	BMI
				0,20 [0,17-0,23]	Biologisches Geschlecht
				0,19 [0,16-0,21]	Hüftumfang [cm]
	Jahr 3	SVM-rbf	Demografisch	1,32 [1,16-1,49]	Bildung [Jahre]
				0,54 [0,48-0,60]	Alter
				0,42 [0,36-0,47]	Biologisches Geschlecht
				0,38 [0,31-0,44]	Bauchumfang [cm]
				0,36 [0,31-0,42]	Hüftumfang [cm]

MMSE	bE	SVM-rbf	Demografisch	0,48 [0,43-0,53]	Bildung [Jahre]
				0,36 [0,33-0,38]	Biologisches Geschlecht
				0,35 [0,30-0,41]	Alter
				0,14 [0,11-0,17]	Bauchumfang [cm]
				0,13 [0,10-0,15]	BMI
CES-D	Jahr 1	GB	Demografisch	0,52 [0,49-0,55]	Bildung [Jahre]
				0,42 [0,38-0,46]	Biologisches Geschlecht
				0,36 [0,31-0,41]	BMI
				0,29 [0,25-0,32]	Hüftumfang [cm]
				0,21 [0,17-0,26]	Bauchumfang [cm]

Der absolute SHAP-Mittelwert wurde mit dem 95%-Konfidenzintervall angegeben.

Abkürzungen: BI, Barthel Index; BMI, Körpermasseindex (englisch *Body-Mass-Index*); bE, bei Entlassung; CES-D, Allgemeine Depressionsskala (englisch *Center for Epidemiologic Studies Depression Scale*); GB, Entscheidungsbaum mit Gradientenverstärkung (englisch *Gradient Boosting*); hsCRP, Hochsensitives C-reaktives Protein; mRS, Modifizierte Rankin Skala; MMSE, Mini-Mental-Status-Test (englisch *Mini-Mental State Examination*); NIHSS, National Institute of Health Stroke Scale; SHAP, SHapley Additive exPlanations; SVM-lin, Stützvektormaschine mit linearem Kernel; SVM-rbf, Stützvektormaschine mit Kernel der radialen Basisfunktion; TICS-M, Modifiziertes Telefoninterview für den kognitiven Status (englisch *Modified Telephone Interview for Cognitive Status*); TOAST, Trial of ORG 10172 in Acute Stroke Treatment.

Quellenangabe: Tabelle modifiziert aus Fast et al. [50]

4 Diskussion

Im Rahmen dieser Arbeit wurden zum ersten Mal standardisierte ML-Modelle zur Prädiktion von funktionellem Outcome, kognitiver Beeinträchtigung, Depression und Mortalität nach erstmaligem ischämischem Schlaganfall eingesetzt. Sie basierte auf einer homogenen Gruppe von Patient*innen und prädizierte Langzeitergebnisse in einem Zeitraum von bis zu drei Jahren. Während in der bisherigen Literatur vor allem funktionelle Outcomes wie mRS oder BI als primäre Endpunkte herangezogen wurden, haben wir in dieser Studie auch kognitive Beeinträchtigungen und Depressionen berücksichtigt, da diese eine zentrale Rolle bei der Rehabilitation nach Schlaganfall spielen. Des Weiteren wurden die wichtigsten Prädiktoren der einzelnen Outcomes mithilfe von Shapley-Werten ermittelt. Diese und weitere Aspekte werden im Folgenden näher erläutert.

A



B

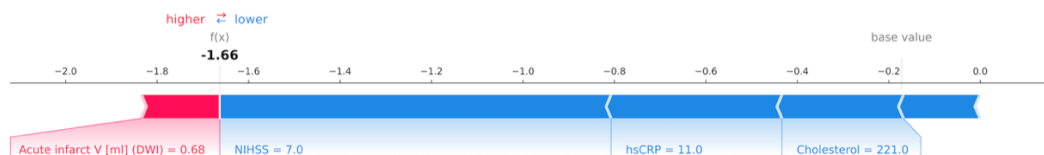


Abbildung 5: Darstellung der Entscheidungsfindung des Entscheidungsbaums mit Gradientenverstärkung (GB, englisch *Gradient Boosting*) für die modifizierte Rankin-Skala (mRS) bei Entlassung.

Abbildungen A und B zeigen die Relevanz einzelner Variablen basierend auf Shapley-Werten für die Prädiktion bei zwei verschiedenen Patient*innen. Die Länge der Balken repräsentiert dabei den Einfluss der jeweiligen Variable auf das Ergebnis. Die Farbe rot steht dabei für einen positiven Effekt und blau für einen negativen Effekt. Abbildung A zeigt eine Person mit einem mRS von einem Punkt, welche von GB korrekt als gutes Outcome klassifiziert wurde. Die Entscheidung wurde aufgrund von niedrigen Werten bei der National Institutes of Health Stroke Scale (NIHSS), hochsensitivem C-reaktiven Protein (hsCRP, englisch *high-sensitivity C-reactive protein*), Cholesterol und akutem Infarktvolume in der Diffusionsgewichteten Sequenz (DWI, englisch *Diffusion-weighted imaging*) getroffen, im Gegensatz zu einem hohen Wahlund Score. Die Person in Abbildung B hingegen hatte ein schlechtes Outcome mit einem mRS von

vier Punkten. Basierend auf hohen Werten bei NIHSS, hsCRP und Cholesteroll klassifizierte der GB auch dieses Outcome korrekt, trotz niedrigem akuten Infarktvolume in DWI. Die Entscheidung des GB wurde dabei stets auf Grundlage des Gesamteffekts aller betrachteten Variablen getroffen.

Quellenangabe: Abbildungen aus Fast et al. [50]

4.1 Modifizierte Rankin Skala

Die mRS zählt zu den wichtigsten Maßstäben in der Messung funktioneller Outcomes nach Schlaganfall. Die Ergebnisse unserer Analysen decken sich daher größtenteils mit dem bereits etablierten Forschungsstand. Schon Wouters et al. [88] und Kazi et al. [89] stellten NIHSS als wichtigsten Prädiktor für mRS nach akutem Schlaganfall fest. Erhöhtes hsCRP und erniedrigtes Cholesteroll hingegen wurden mit schlechteren mRS Outcomes assoziiert [90,91]. Bei untergewichtigen Patient*innen ($\text{BMI} < 18,5 \text{ kg/m}^2$) wurden zudem ungünstige funktionelle Outcomes und erhöhte Mortalitätsraten festgestellt [92]. In Abbildung 5 wird der Entscheidungsfindungsprozess von GB für mRS bei Entlassung anhand zweier Beispielpatient*innen illustriert.

Die erste bekannte Applikation von ML-Methoden in der Prädiktion von mRS stammt von Monteiro et al. [93]. Es wurden 425 Patient*innen mit 152 Input-Variablen einbezogen, um das Outcome nach drei Monaten zu modellieren. Die Modelle setzten sich aus einer regularisierten logistischen Regression, einem Entscheidungsbaum, einer SVM, einem Zufallswald (RF, englisch *Random Forest*) und Xgboost mit einer zehnfachen Kreuzvalidierung zusammen. An dieser Studie war besonders interessant, dass NIHSS zu verschiedenen Zeitpunkten einbezogen wurde - bei Aufnahme, zwei Stunden nach Beginn der Behandlung, nach 24 Stunden, nach sieben Tagen und bei Entlassung. Der RF Klassifikator erreichte damit eine $\text{AUC} = 0,936 \pm 0,034$. Die wichtigsten Prädiktoren waren NIHSS, eine NIHSS Komponente bezüglich Okulomotorik, Glucose, systolischer Blutdruck und Alter. In einer Studie mit 3522 Patient*innen von Heo et al. [94] wurde ein Deep Neural Network (DNN) mit drei Ebenen und 15 Knotenpunkten verwendet. Dies erzielte eine $\text{AUC} = 0,888$ in der Prädiktion von mRS. Die Analyse inkludierte 38 Parameter mit Variablen wie NIHSS, demografische Variablen, Zeit zwischen Erstmanifestation der Symptome und Aufnahme, Schlaganfall-Subtyp, Vorerkrankungen, Medikamenteneinnahme und serologische Werte. Es wird jedoch nicht erwähnt, ob die Analysen wiederholt wurden oder eine Resampling-Methode zur Anwendung kam. Die Anwendung dieser Techniken ist notwendig, um robuste ML oder Deep Learning (DL) Modelle zu entwickeln

und sogenanntes Overfitting zu vermeiden. Trotz dieser potenziellen Einschränkungen suggerieren die Ergebnisse, dass DL-Anwendungen möglicherweise geeigneter sind für die Prädiktion von mRS. Chiu et al. [95] fokussierten sich auf die Prädiktion multipler Klassen innerhalb der mRS (jeweils 0-2, 3-4 und 5-6 Punkte) bei 590 Patient*innen unter Reperfusionstherapie. Ihr Modell erzielte bessere Ergebnisse als der DRAGON score [96] in der Klassifikation der genannten Kategorien. Wie bei allen genannten Studien wurde NIHSS abermals als die wichtigste Input-Variable erkannt. Vergleichbare Ergebnisse fanden sich auch in den Studien von Alaka et al. [97], Ramos et al. [98] und Li et al. [99].

4.2 Barthel Index

Obwohl signifikante Ergebnisse für BI bei Entlassung erzielt wurden, konnte der BI nach einem Jahr von keinem der ML-Modelle zuverlässig prädiziert werden. Ursächlich dafür könnte die extreme Ungleichverteilung der Outcome-Klassen im Datensatz sein (siehe Tabelle 2). Obwohl die SMOTE Technik dies bis zu einem gewissen Maß kompensieren kann, stößt auch dieses Verfahren bei kleineren Datenmengen an Grenzen. In Analogie zur mRS galt auch hier NIHSS als der führende Prädiktor. Diese Ähnlichkeit wird durch inhaltliche Überschneidungen und Korrelationen zwischen mRS und BI erklärt [100]. Darüber hinaus konnte, in Übereinstimmung zu Musa et al., der vaskuläre Ursprung des Infarkts als wichtiger Prädiktor erfasst werden [101].

Sale et al. [102] entwickelten ein ML-Modell basierend auf einer SVM-lin für die Prädiktion prognostischer Faktoren hinsichtlich motorischer und kognitiver Entwicklung durch rehabilitatives Training nach Schlaganfall. Das Modell erfasste Alter, biologisches Geschlecht, Schlaganfalltyp und eine Reihe serologischer Parameter bei 55 Patient*innen. Die Ergebnisse des Test-Sets für ischämischen Schlaganfall ergaben eine Korrelation von 0,75, eine mittlere absolute Abweichung von 87,03% und einen mittleren quadratischen Fehler von 21,74 für BI. Als wichtigste Parameter ergaben sich der funktionale Selbständigkeitsindex (FIM, englisch *Functional Independence Measure*) und BI bei Aufnahme sowie das Thrombozyten-Lymphozyten-Ratio, die Erythrozytenverteilungsbreite und die Lymphozytenzahl. In einer Studie von De Wit et al. [103] identifizierten die Autor*innen eine Reihe von BI Variablen bei Entlassung (Fähigkeiten bezüglich Körperpflege und eigenständiges Ankleiden) als wichtigste Prädiktoren hinsichtlich des funktionellen Outcomes (BI > 95) fünf Jahre nach Schlaganfall bei 153 Patient*innen. In einem Review von Meyer et al.

[104] wurden die Parameter Funktionslevel bei Aufnahme (BI oder FIM), NIHSS, Dysphasie, Impulsivität, Neglect, vorangegangener Schlaganfall und Alter als die wichtigsten Input-Variablen bezüglich BI Outcome nach Rehabilitation identifiziert.

4.3 Mini-Mental-Status-Test

Der MMSE wurde als Messinstrumente für kognitive Beeinträchtigung herangezogen. Dabei erbrachte die demografische Input-Subdomäne die besten Ergebnisse für den MMSE bei Entlassung. Bereits durch Ojala-Oksala et al. [105] wurde die protektive Wirkung eines höheren Bildungsstatus gegenüber kognitivem Abbau etabliert. Geschlechtsspezifische Aspekte wurden zudem durch eine Studie von Dong et al. [106] belegt. In deren Arbeit hatten Frauen 90 Tage nach einem Schlaganfall generell schlechtere Ergebnisse im MMSE im Vergleich zu Männern. Daneben ist auch der BMI in Hinsicht auf den MMSE relevant, da einerseits Untergewicht ($\text{BMI} < 20 \text{ kg/m}^2$) mit schlechteren kognitiven Outcomes sowie Übergewicht mit einer spezifischen Dysfunktion des Frontallappens assoziiert wurde [107].

Youn et al. [108] entwickelten ein ML-Modell basierend auf 10185 Patient*innen mit 24 Input-Variablen und erzielten damit eine $\text{AUC} = 0,77$ für den MMSE. Es wurden jedoch weder das ML-Modell genauer spezifiziert noch die wichtigsten Input-Variablen eruiert. Das sogenannte COMPASS-Modell wurde von Zhu et al. [109] beschrieben und schätzt die MMSE-Testergebnisse von Patient*innen der Alzheimer's Disease Neuroimaging Initiative Datenbank innerhalb von 24 Monaten ein. Eine Abnahme der MMSE-Testergebnisse von mehr als drei Punkten konnte mit einer $\text{AUC} = 0,814$ berechnet werden.

4.4 Modifiziertes Telefoninterview für den kognitiven Status

Rund ein Drittel aller Patient*innen entwickeln eine demenzielle Symptomatik innerhalb eines Jahres nach Schlaganfall [110]. Das Alter ist dabei einer der stärksten Prädiktoren für kognitive Defizite nach einem Schlaganfall [111,112]. Ebenso ist der langfristig protektive Effekt eines höheren Bildungsniveaus [113] sowie wieder die Relevanz des biologischen Geschlechts ein bekannter Einflussfaktor [111]. Ähnlich wie bei dem MMSE bestätigen auch hier die Ergebnisse von Lee et al. [107] den Einfluss von BMI, Hüft- und Bauchumfang.

Casanova et al. [113] haben sich in einer Studie mit 7142 Patient*innen aus der Health and Retirement Study (HRS) Kohorte mit der Prädiktion von kognitivem Abbau beschäftigt. Die Ergebnisse wurden anhand einer abgekürzten Version des TICS-M gemessen. Die Studie inkludiert genetische Faktoren und epidemiologische Information sowie eine Reihe klinischer Parameter. Bildung, Alter, sozioökonomischer Status, Geschlecht, Schlaganfall und Diabetes Mellitus waren die wichtigsten Variablen bei der Nutzung eines RF Klassifikator, welcher damit eine Genauigkeit von 0,78, Sensitivität von 0,75 und Spezifität von 0,81 erreichte. In einer weiteren Arbeit mit 9979 Patient*innen aus der HRS und basierend auf einem RF untersuchten Aschwanden et al. [114] den Einfluss von 52 Parametern in der Prädiktion von kognitivem Abbau und Demenz. Den größten Effekt zeigten hier afroamerikanische Ethnizität, ökonomischer Wohlstand, Bildung, BMI, subjektives Gesundheitsempfinden, emotionale Belastung und hispanische Ethnizität mit einer Vorhersage-Fehlerquote = 0,34. Die Ergebnisse dieser beiden Studien implizieren, dass sozioökonomischer Status und Ethnizität eine wichtige Rolle in der Entwicklung kognitiver Verlaufsparemeter spielen.

4.5 Allgemeine Depressionsskala

Etwa ein Drittel aller Patient*innen nach ischämischem Schlaganfall entwickelt eine klinisch relevante Depression innerhalb eines Jahres [115]. In unserer Studie erzielten alle ML-Modelle mit der demografischen Input-Subdomäne die besten Ergebnisse für CES-D. Als wichtigster Prädiktor für Depression nach Schlaganfall erwies sich die Bildung [116,117]. Darüber hinaus wurden in einer Studie von Shi et al. [118] weibliches Geschlecht und ein Alter unter 70 Jahren mit einem höheren Risiko für die Entstehung einer Depression assoziiert. Laut Xu et al. [119] gilt das gleiche für einen erhöhten Bauchumfang.

Eine gekürzte Variante der CES-D wurde in einer weiteren Studie von Xu et al. [120] genutzt, um Patient*innen aus dem HRS-Datensatz zu analysieren. Dafür wurden elf Input-Variablen in einer Population älterer Menschen herangezogen, um Prädiktoren für Depression zu eruieren. Ein rückgekoppeltes neuronales Netzwerk wurde genutzt, um eine AUC = 0,873 in der Prädiktion von Depression zu erreichen. Eine Studie von Hama et al. [121] erreichte eine besonders hohe AUC von über 0,90 in der Prädiktion für Depression nach Schlaganfall, gemessen anhand der Hospital Anxiety and Depression

Scale, unter Anwendung eines logarithmisch-linearen Netzwerks mit gaußscher Mischverteilung bei 274 Patient*innen des Hibino Hospital. Als wichtigste Prädiktoren in dieser Studie zeigten sich die Japanese Perceived Stress Scale, der Symbol Digit Modalities Test, Tapping span backwards, Visual cancellation Kana time und der Continuous Performance Test. Hier zeigte sich, dass die Nutzung sehr spezifischer Testmarker die Genauigkeit drastisch erhöhen kann. Li et al. [122] entwickelten ein Modell zur Klassifizierung anhand der Hamilton Rating Scale for Depression-17 mit mehr als 17 Punkten drei Monate nach initialem Schlaganfall. Basierend auf einer Logistischen Regression erreichte das Modell eine Kontingenzkoeffizient $C = 0,81$ (95%-KI: 0,791-0,829) mit 71,1% Sensitivität und 78,6% Spezifität. Zu den Variablen im entwickelten Nomogramm zählten Alter, NIHSS, mRS, Bildungsniveau, Calcium-Phosphatprodukt sowie die Vorerkrankungen Hypertonie und Vorhofflimmern. Damit zeigt sich eine Überschneidung hinsichtlich Alter und Bildung mit den Ergebnissen unserer Studie.

4.6 Mortalität

In dieser Studie konnten keine signifikanten Ergebnisse für Mortalität erzielt werden. In der wissenschaftlichen Literatur existieren jedoch einige erfolgreiche Arbeiten für diese Fragestellung. Cheon et al. [123] nutzen ein DNN, um Mortalität nach Schlaganfall bei 15009 Patient*innen zu modellieren. Mit elf Input-Variablen erreichten sie eine $AUC = 0,835$ wobei Typ des Schlaganfalls der wichtigste Prädiktor war. In einer Studie von Scrutinio et al. [124] mit 1207 Patient*innen nach Rehabilitation wurde eine $AUC > 0,90$ für drei ML-Modelle (RF, ADA-Boost und GB) mit einer zehnfachen Kreuzvalidierung erzielt. Allerdings ist dieses bemerkenswerte Ergebnis unter der Anwendung der SMOTE Technik auf dem Test-Set entstanden, was methodisch kritisch zu betrachten ist [78]. Trotzdem erreichte der RF Klassifikator eine $AUC = 0,844$ ohne Anwendung von SMOTE. In dieser Analyse waren Alter, die Dauer der Nachfolgeuntersuchung, Case-Mix Gruppen, Dauer zwischen Erstmanifestation des Schlaganfalls und Einweisung zur Rehabilitationsbehandlung sowie der FIM die wichtigsten Prädiktoren von den insgesamt 35 Input-Variablen. Die methodischen Differenzen und Unterschiede der verwendeten Datensätzen könnten die Abweichung unserer Ergebnisse erklären.

4.7 Klinische Implikationen

Für die Abschätzung von funktionellen Outcomes nach Schlaganfall existieren mit dem DRAGON [96] und ASTRAL Score [125] bereits etablierte Methoden zur Prädiktion der mRS nach Schlaganfall. Zum Zeitpunkt dieser Studie galt dies allerdings nicht für den BI. Der SIGNAL2 [126] und CHANGE Score [127] sind validierte prognostische Scores hinsichtlich der Entwicklung kognitiver Defizite nach Schlaganfall, gemessen an MMSE und MoCA. Beide Scores beziehen in ihre Berechnungen neben Alter und Bildungsniveau der Patient*innen eine Reihe MRT-spezifischer Parameter mit ein. Für die Prognose hinsichtlich Depression nach Schlaganfall entwickelten Li et al. [122] mit ihrem Nomogramm den ersten Score zur klinischen Anwendung. Die Überschneidung mit den Ergebnissen aus unseren Analysen ist dabei vielversprechend. Der iScore [128] und PLAN Score [129] bieten eine Einschätzung des Mortalitätsrisikos innerhalb von 30 Tagen und eines Jahres nach Schlaganfall.

Das Ziel zukünftiger Studien sollte sein ein universelles Prädiktionsmodell zu entwickeln, das die Gesamtheit von funktionellem Outcome, kognitiver Beeinträchtigung, Depression und Mortalität modellieren kann, basierend auf einer kleinen Gruppe von Variablen. Dazu könnten beispielsweise NIHSS, Bildung oder BMI zählen. Ein solches Modell könnte als ML-basiertes CDSS einen kosteneffektiven Beitrag zur Effizienz im klinischen Alltag liefern. CDSS können auf klinisch wichtige Informationen hinweisen und damit Behandlungsfehler reduzieren, was letztendlich die Sicherheit der Patient*innen im Gesundheitswesen erhöht [130]. Beispielsweise hängt die funktionelle Rehabilitation nach einem Schlaganfall maßgeblich von einer ausreichenden kognitiven Funktion und einer adäquaten Behandlung von Depression ab [14,17]. CDSS könnten auf gefährdete Patient*innen hinweisen, um personalisierte Rehabilitationsmaßnahmen und spezifische Sekundärpräventionsstrategien zu entwickeln.

4.8 Hürden der Methodenentwicklung

Ein problematischer Aspekt in der Anwendung von ML in der medizinischen Forschung sind die unterschiedlichen technischen Herangehensweisen. Der mangelnde Einsatz von Resampling-Methoden oder fehlende Informationen zur Anzahl der Wiederholung von Analysen lässt die Reliabilität einiger Ergebnisse anzweifeln. Darüber hinaus wird eine Vielzahl von nicht standardisierten ML-Algorithmen verwendet, oft in Kombination mit

sehr spezifischen Input-Variablen, was zu teilweise erheblichen Unterschieden in den Ergebnissen führt. Es bleibt somit unklar, worauf diese Diskrepanz zurückzuführen ist. Um diesen Ansprüchen gerecht zu werden haben wir in dieser Studie eine Reihe von Methoden zur Entwicklung robuster ML-Modelle genutzt - angefangen bei einem balancierten Patient*innenkollektiv, einem standardisiertes Set von drei ML-Algorithmen, wiederholter fünffach verschachtelter Kreuzvalidierung sowie einem Permutationstest mit Bonferroni-Korrektur.

4.9 Limitationen

Limitierend für diese Studie waren eine Reihe von methodologischen und inhaltlichen Aspekten. Allen voran stand eine limitierte Anzahl von Patien*innendaten aus dem PROSCIS-B Datensatz zur Verfügung sowie keine externe Kohorte zur Validierung der Ergebnisse. Da Patient*innen nur in Berlin rekrutiert wurden, könnte zudem ein regionales Bias vorliegen. Die Bilddaten wurden an unterschiedlichen Scannern innerhalb des CSB und der Charité - Universitätsmedizin Berlin erhoben, wodurch möglicherweise die MRT-Parameter einer Verzerrung unterliegen. Vor allem bei der manuellen Abgrenzung der Infarktareale könnte dies zu einer ungenaueren Erhebung der Volumina bei Patient*innen mit qualitativ schlechten MRT-Bilddaten geführt haben. Ein weiterer Kritikpunkt ist, dass der MoCA nur unvollständig erhoben und somit nicht in diese Studie eingeschlossen werden konnte. Ähnlich gilt, dass keine detaillierteren Informationen zu den verabreichten Medikationen oder Unterschiede in den Rehabilitationsmaßnahmen erfasst wurden. Darüber hinaus unterlag der gesamte Datensatz einer starken Dysbalance hinsichtlich der Outcome-Klassen (siehe Tabelle 2). Um diese Problematik zu umgehen, wurden SMOTE und die wiederholte fünffach verschachtelte Kreuzvalidierung eingesetzt. Trotzdem ist es nicht auszuschließen, dass diese Einschränkung die Ergebnisse beeinflusst und damit die Anwendbarkeit auf neue Daten vermindert. Aus technischer Perspektive wäre zudem ein Vergleich der Shapley-Werte aus den unterschiedlichen ML-Modellen interessant, um mögliche Gemeinsamkeiten und Unterschiede zu untersuchen. Zudem wurden Patient*innen nur dann in die Studie eingeschlossen, wenn diese selbst oder durch eine*n gesetzlichen Vertreter*in schriftlich in die Studie eingewilligt hatten. Dies hatte zur Folge, dass vom Schlaganfall sehr schwer betroffene Patient*innen aus der Studie ausgeschlossen wurden, was wiederum die allgemeine Aussagekraft der Ergebnisse einschränkt [131].

5 Schlussfolgerungen

Mittels eines systematischen Vergleichs von drei ML-Modellen (SVM-lin, SVM-rbf, GB) wurden in dieser Studie funktionelles Outcome, kognitive Beeinträchtigung, Depression und Mortalität nach der Erstmanifestation eines ischämischen Schlaganfalls prädiziert. Dafür wurden eine Reihe von Input-Subdomänen (demografisch, klinisch, serologisch und MRT) mit der Gesamtheit aller Input-Variablen verglichen und die jeweils stärksten Prädiktoren mittels Shapley-Werten ermittelt. Die Ergebnisse unserer Studie belegten somit bereits etablierte als auch potenziell neue Prädiktoren. Es sind jedoch weitere Studien notwendig, um diese Erkenntnisse zu bestätigen und ihre klinische Anwendbarkeit zu überprüfen.

Literaturverzeichnis

1. Hatano S. Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ.* 1976;54(5):541–53.
2. Deb P, Sharma S, Hassan KM. Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. *Pathophysiology.* 2010 Jun;17(3):197–218.
3. Caplan LR. Chapter 2 - Basic Pathology, Anatomy, and Pathophysiology of Stroke. In: Caplan LR, editor. *Caplan's Stroke (Fourth Edition)* [Internet]. Philadelphia: W.B. Saunders; 2009. p. 22–63. Available from: <https://www.sciencedirect.com/science/article/pii/B9781416047216500043>
4. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin e.V. (DEGAM). S3-Leitlinie Schlaganfall. 2020.
5. Gardener H, Sacco RL, Rundek T, Battistella V, Cheung YK, Elkind MSV. Race and Ethnic Disparities in Stroke Incidence in the Northern Manhattan Study. *Stroke.* 2020 Apr 1;51(4):1064–9.
6. Lavados PM, Olavarría VV, Hoffmeister L. Ambient Temperature and Stroke Risk. *Stroke.* 2018 Jan 1;49(1):255–61.
7. Nimptsch U, Mansky T. Trends in acute inpatient stroke care in Germany--an observational study using administrative hospital data from 2005-2010. *Dtsch Arztebl Int.* 2012 Dec;109(51–52):885–92.
8. Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. *Lancet.* 2008 May;371(9624):1612–23.
9. Deutsche Gesellschaft für Neurologie e.V. (DGN). S2e-Leitlinie Akuttherapie des ischämischen Hirninfarktes. 2021.
10. Thomalla G, Cheng B, Ebinger M, Hao Q, Tourdias T, Wu O, Kim JS, Breuer L, Singer OC, Warach S, Christensen S, Treszl A, Forkert ND, Galinovic I, Rosenkranz M, Engelhorn T, Köhrmann M, Endres M, Kang DW, Dousset V, Sorensen AG, Liebeskind DS, Fiebach JB, Fiehler J, Gerloff C. DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4·5 h of symptom onset (PRE-FLAIR): a multicentre observational study. *Lancet Neurol.* 2011 Nov 1;10(11):978–86.
11. Fiebach JB, Schellinger PD, Jansen O, Meyer M, Wilde P, Bender J, Schramm P, Jüttler E, Oehler J, Hartmann M, Hähnel S, Knauth M, Hacke W, Sartor K. CT and Diffusion-Weighted MR Imaging in Randomized Order. *Stroke.* 2002 Sep 1;33(9):2206–10.

12. W. Sturm, Munte TF. Klinische Neurologie, in Lehrbuch der klinischen Neuropsychologie. Grundlagen, Methoden, Diagnostik, Therapie. Springer Spektrum. 2009;740–50.
13. Nabavi DG, Koennecke HC, Ossenbrink M, Grau A, Busse O, die Stroke Unit Kommission, den Zertifizierungsausschuss der DSG, den Vorstand der DSG. Zertifizierungskriterien für Stroke-Units in Deutschland. *Nervenarzt*. 2019 Apr 1;90(4):335–42.
14. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, Kwakkel G. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. *PLoS One*. 2014;9(2):e87987.
15. Schneider EJ, Lannin NA, Ada L, Schmidt J. Increasing the amount of usual rehabilitation improves activity after stroke: a systematic review. *J Physiother*. 2016 Oct;62(4):182–7.
16. Leśniak M, Bak T, Czepiel W, Seniów J, Członkowska A. Frequency and prognostic value of cognitive disorders in stroke patients. *Dement Geriatr Cogn Disord*. 2008;26(4):356–63.
17. das Nair R, Cogger H, Worthington E, Lincoln NB. Cognitive rehabilitation for memory deficits after stroke. *Cochrane Database Syst Rev*. 2016 Sep;9:CD002293.
18. Hackett ML, Pickles K. Part I: frequency of depression after stroke: an updated systematic review and meta-analysis of observational studies. *Int J Stroke*. 2014 Dec;9(8):1017–25.
19. Wei N, Yong W, Li X, Zhou Y, Deng M, Zhu H, Jin H. Post-stroke depression and lesion location: a systematic review. *J Neurol*. 2015 Jan;262(1):81–90.
20. Worrall L, Ryan B, Hudson K, Kneebone I, Simmons-Mackie N, Khan A, Hoffmann T, Power E, Togher L, Rose M. Reducing the psychosocial impact of aphasia on mood and quality of life in people with aphasia and the impact of caregiving in family members through the Aphasia Action Success Knowledge (Aphasia ASK) program: study protocol for a randomized controlled trial. *Trials*. 2016 Mar;17:153.
21. Hackett ML, Anderson CS, House A, Xia J. Interventions for treating depression after stroke. *Cochrane Database Syst Rev*. 2008 Oct;(4):CD003437.
22. Brønnum-Hansen H, Davidsen M, Thorvaldsen P. Long-Term Survival and Causes of Death After Stroke. *Stroke J Cereb Circ*. 2001;32:2131–6.
23. Feng MC, Lin YC, Chang YH, Chen CH, Chiang HC, Huang LC, Yang YH, Hung CH. The Mortality and the Risk of Aspiration Pneumonia Related with Dysphagia in Stroke Patients. *J Stroke Cerebrovasc Dis*. 2019 May;28(5):1381–7.

24. S3-Leitlinie Klinische Ernährung in der Neurologie. Dtsch Ges Für Ernährungsmedizin DGEM. 2013;
25. Johnson CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, Abd-Allah F, Abdelalim A, Abraha HN, Abu-Rmeileh NM, Adebayo OM, Adeoye AM, Agarwal G, Agrawal S, Aichour AN, Aichour I, Aichour MTE, Alahdab F, Ali R, Alvis-Guzman N, Anber NH, Anjomshoa M, Arabloo J, Arauz A, Ärnlöv J, Arora A, Awasthi A, Banach M, Barboza MA, Barker-Collo SL, Bärnighausen TW, Basu S, Belachew AB, Belayneh YM, Bennett DA, Bensenor IM, Bhattacharyya K, Biadgo B, Bijani A, Bikbov B, Bin Sayeed MS, Butt ZA, Cahuana-Hurtado L, Carrero JJ, Carvalho F, Castañeda-Orjuela CA, Castro F, Catalá-López F, Chaiah Y, Chiang PPC, Choi JYJ, Christensen H, Chu DT, Cortinovis M, Damasceno AAM, Dandona L, Dandona R, Daryani A, Davletov K, de Courten B, De la Cruz-Góngora V, Degefa MG, Dharmaratne SD, Diaz D, Dubey M, Duken EE, Edessa D, Endres M, FARAON EJA, Farzadfar F, Fernandes E, Fischer F, Flor LS, Ganji M, Gebre AK, Gebremichael TG, Geta B, Gezae KE, Gill PS, Gnedovskaya EV, Gómez-Dantés H, Goulart AC, Grosso G, Guo Y, Gupta R, Haj-Mirzaian A, Haj-Mirzaian A, Hamidi S, Hankey GJ, Hassen HY, Hay SI, Hegazy MI, Heidari B, Herial NA, Hosseini MA, Hostiuc S, Irvani SSN, Islam SMS, Jahanmehr N, Javanbakht M, Jha RP, Jonas JB, Jozwiak JJ, Jürisson M, Kahsay A, Kalani R, Kalkonde Y, Kamil TA, Kanchan T, Karch A, Karimi N, Karimi-Sari H, Kasaeian A, Kassa TD, Kazemini H, Kefale AT, Khader YS, Khalil IA, Khan EA, Khang YH, Khubchandani J, Kim D, Kim YJ, Kisa A, Kivimäki M, Koyanagi A, Krishnamurthi RK, Kumar GA, Lafranconi A, Lewington S, Li S, Lo WD, Lopez AD, Lorkowski S, Lotufo PA, Mackay MT, Majdan M, Majdzadeh R, Majeed A, Malekzadeh R, Manafi N, Mansournia MA, Mehndiratta MM, Mehta V, Mengistu G, Meretoja A, Meretoja TJ, Miazgowski B, Miazgowski T, Miller TR, Mirrakhimov EM, Mohajer B, Mohammad Y, Mohammadoo-khorasani M, Mohammed S, Mohebi F, Mokdad AH, Mokhayeri Y, Moradi G, Morawska L, Moreno Velásquez I, Mousavi SM, Muhammed OSS, Muruet W, Naderi M, Naghavi M, Naik G, Nascimento BR, Negoï RI, Nguyen CT, Nguyen LH, Nirayo YL, Norrving B, Noubiap JJ, Ofori-Asenso R, Ogbo FA, Olagunju AT, Olagunju TO, Owolabi MO, Pandian JD, Patel S, Perico N, Piradov MA, Polinder S, Postma MJ, Poustchi H, Prakash V, Qorbani M, Rafiei A, Rahim F, Rahimi K, Rahimi-Movaghar V, Rahman M, Rahman MA, Reis C, Remuzzi G, Renzaho AMN, Ricci S, Roberts NLS, Robinson SR, Roeber L, Roshandel G, Sabbagh P, Safari H, Safari S, Safiri S, Sahebkar A, Salehi Zahabi S, Samy AM, Santalucia P, Santos IS, Santos JV, Santric Milicevic MM, Sartorius

- B, Sawant AR, Schutte AE, Sepanlou SG, Shafieesabet A, Shaikh MA, Shams-Beyranvand M, Sheikh A, Sheth KN, Shibuya K, Shigematsu M, Shin MJ, Shiue I, Siabani S, Sobaih BH, Sposato LA, Sutradhar I, Sylaja P, Szoeki CEI, Te Ao BJ, Temsah MH, Temsah O, Thrift AG, Tonelli M, Topor-Madry R, Tran BX, Tran KB, Truelsen TC, Tsadik AG, Ullah I, Uthman OA, Vaduganathan M, Valdez PR, Vasankari TJ, Vasanathan R, Venketasubramanian N, Vosoughi K, Vu GT, Waheed Y, Weiderpass E, Weldegewergs KG, Westerman R, Wolfe CDA, Wondafrash DZ, Xu G, Yadollahpour A, Yamada T, Yatsuya H, Yimer EM, Yonemoto N, Yousefifard M, Yu C, Zaidi Z, Zamani M, Zarghi A, Zhang Y, Zodpey S, Feigin VL, Vos T, Murray CJL. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019 May 1;18(5):439–58.
26. Busch MA, Kuhnert R. 12-Monats-Prävalenz von Schlaganfall oder chronischen Beschwerden infolge eines Schlaganfalls in Deutschland. 2017;2(1).
27. Wang H. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* 2016 Oct;388(10053):1459–544.
28. Gorelick PB. The global burden of stroke: persistent and disabling. *Lancet Neurol.* 2019;18(5):417–8.
29. Heuschmann PU, Busse O, Wagner M, Endres M, Villringer A, Röther J, Kolomin-sky-Rabas PL, Berger K, für das Kompetenznetz Schlaganfall die DSG sowie die SDSH. Schlaganfallhäufigkeit und Versorgung von Schlaganfallpatienten in Deutschland. *Akt Neurol.* 2010 Oct 5;37(07):333–40.
30. Luengo-Fernandez R, Violato M, Candio P, Leal J. Economic burden of stroke across Europe: A population-based cost analysis. *Eur Stroke J.* 2020 Mar 1;5(1):17–25.
31. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol.* 2009 Apr;8(4):355–69.
32. Kolomin-sky-Rabas PL, Wiedmann S, Weingärtner M, Liman TG, Endres M, Schwab S, Buchfelder M, Heuschmann PU. Time trends in incidence of pathological and etiological stroke subtypes during 16 years: the Erlangen Stroke Project. *Neuroepidemiology.* 2015;44(1):24–9.
33. Busch M, Schienkiewitz A, Nowossadeck E, Gößwald A. [Prevalence of stroke in adults aged 40 to 79 years in Germany : Results of the German Health Interview and

Examination Survey for Adults (DEGS1)]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2013 May 1;56:656–60.

34. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev*. 1959;3(3):210–29.

35. Mitchell TM. *Machine learning*. Vol. 1. McGraw-hill New York; 1997.

36. Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning*. 2012 Jan 1;

37. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb 1;542(7639):115–8.

38. Tseng HH, Luo Y, Cui S, Chien JT, Ten Haken R, El Naqa I. Deep Reinforcement Learning for Automated Radiation Adaptation in Lung Cancer. *Med Phys*. 2017 Oct 16;44.

39. Nemati S, Ghassemi MM, Clifford GD. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. 2016 38th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2016;2978–81.

40. Valliani AA, Ranti D, Oermann EK. Deep Learning and Neurology: A Systematic Review. *Neurol Ther*. 2019 Dec;8(2):351–65.

41. Wang DZ, Schwamm LH, Qian T, Dai Q. Decoding the brain through research—the future of brain health. *The BMJ*. 2020;371.

42. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. *Biomed. Inform. Insights* 8. BII-S31559. 2016;

43. Auger SD, Jacobs BM, Dobson R, Marshall CR, Noyce AJ. Big data, machine learning and artificial intelligence: a neurologist's guide. *Pract Neurol*. 2021 Feb 1;21(1):4.

44. Pop N Ovidiu, Tit D Mirela, Diaconu C Cristina, Munteanu M Alexandru, Babes E Elena, Stoicescu M, Popescu M Ioachim, Bungau S. The Alberta Stroke Program Early CT score (ASPECTS): A predictor of mortality in acute ischemic stroke. *Exp Ther Med*. 2021 Dec 1;22(6):1371.

45. Scherer M, Cordes J, Younsi A, Sahin YA, Götz M, Möhlenbruch M, Stock C, Bösel J, Unterberg A, Maier-Hein K, Orakcioglu B. Development and Validation of an Automatic Segmentation Algorithm for Quantification of Intracerebral Hemorrhage. *Stroke*. 2016 Nov 1;47(11):2776–82.

46. Nagamine M, Kohanteb P, Yu W, Chang P, Chow D. Accuracy of Artificial Intelligence in Measuring Intracerebral Hemorrhage Volumes and Expansion Compared to Human Estimates (5018). *Neurology*. 2020 Apr 14;94(15 Supplement):5018.

47. Takahashi N, Kondo Y, Tsai DY, Matsuyama E, Kinoshita T, Ishii K. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiol Phys Technol*. 2013 Sep 1;7.
48. Mouridsen K, Thurner P, Zaharchuk G. Artificial Intelligence Applications in Stroke. *Stroke*. 2020;51(8):2573–9.
49. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, Wang Y, Douiri A, Wolfe CD, Bray B. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS One*. 2020;15(6):e0234722.
50. Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M, Ritter K. Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. *Front Neurol* [Internet]. 2023;14. Available from: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1114360>
51. Quinn T, Dawson J, Walters M. Dr John Rankin; His Life, Legacy and the 50th Anniversary of the Rankin Stroke Scale. *Scott Med J*. 2008;53(1):44–7.
52. Dewing J. A critique of the Barthel Index. *Br J Nurs Mark Allen Publ*. 1992;1(7):325–9.
53. Nasreddine Z, Phillips N, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings J, Chertkow H. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *J Am Geriatr Soc*. 2005 May 1;53:695–9.
54. Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the Mini-Mental State Examination and the Standardized Mini-Mental State Examination: Implications for clinical practice. *Appl Nurs Res*. 2000 Nov 1;13(4):209–13.
55. Liman T, Zietemann V, Wiedmann S, Jungehülsing G, Endres M, Wollenweber F, Wellwood I, Dichgans M, Heuschmann P. Prediction of vascular risk after stroke - protocol and pilot data of the Prospective Cohort with Incident Stroke (PROSCIS). *Int J Stroke Off J Int Stroke Soc*. 2012 Aug 29;8.
56. Cook SE, Marsiske M, McCoy KJM. The use of the Modified Telephone Interview for Cognitive Status (TICS-M) in the detection of amnesic mild cognitive impairment. *J Geriatr Psychiatry Neurol*. 2009 Jun;22(2):103–9.
57. Carleton RN, Thibodeau MA, Teale MJN, Welch PG, Abrams MP, Robinson T, Asmundson GJG. The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. *PLoS One*. 2013/03/01 ed. 2013;8(3):e58067–e58067.
58. Zhu X, Vondrick C, Fowlkes CC, Ramanan D. Do We Need More Training Data?

Int J Comput Vis. 2016 Aug;119(1):76–92.

59. Ranti D, Valliani AAA, Costa A, Oermann EK. Chapter 20 - Artificial intelligence as applied to clinical neurological conditions. In: Xing L, Giger ML, Min JK, editors. *Artificial Intelligence in Medicine* [Internet]. Academic Press; 2021. p. 395–413. Available from: <https://www.sciencedirect.com/science/article/pii/B978012821259200020X>
60. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Med*. 2018 Nov 6;15(11):e1002683.
61. Clark T, Nyberg E. Creating the Black Box: A Primer on Convolutional Neural Network Use in Image Interpretation. *Curr Probl Diagn Radiol*. 2020 Nov 1;49(6):365–7.
62. Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In: *NIPS*. 2016.
63. Bratic B, Kurbalija V, Ivanović M, Oder I, Bosnic Z. Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors. *J Med Syst*. 2018;42:1–15.
64. Graham SA, Lee EE, Jeste DV, Van Patten R, Twamley EW, Nebeker C, Yamada Y, Kim HC, Depp CA. Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry Res*. 2020;284:112732.
65. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011 Mar;20(1):40–9.
66. Yan L, Wang S, Xu L, Zhang Z, Liao P. Procalcitonin as a prognostic marker of patients with acute ischemic stroke. *J Clin Lab Anal*. 2020 Jul;34(7):e23301.
67. Rorden C, Brett M. Stereotaxic display of brain lesions. *Behav Neurol*. 2000;12(4):191–200.
68. Geirhos R, Jacobsen JH, Michaelis C, Zemel R, Brendel W, Bethge M, Wichmann FA. Shortcut learning in deep neural networks. *Nat Mach Intell*. 2020 Nov 1;2(11):665–73.
69. Kwah LK, Diong J. National Institutes of Health Stroke Scale (NIHSS). *J Physiother*. 2014 Mar;60(1):61.
70. Wahlund LO, Barkhof F, Fazekas F, Bronge L, Augustin M, Sjögren M, Wallin A, Ader H, Leys D, Pantoni L, Pasquier F, Erkinjuntti T, Scheltens P. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke*. 2001 Jun;32(6):1318–22.

71. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993 Jan;24(1):35–41.
72. Saver JL, Filip B, Hamilton S, Yanes A, Craig S, Cho M, Conwit R, Starkman S, FAST-MAG Investigators and Coordinators. Improving the reliability of stroke disability grading in clinical trials and clinical practice: the Rankin Focused Assessment (RFA). *Stroke*. 2010/04/01 ed. 2010 May;41(5):992–5.
73. Claesen M, Moor BD. Hyperparameter Search in Machine Learning. ArXiv CsLG [Internet]. 2015; Available from: <http://arxiv.org/abs/1502.02127>
74. Birkes D. Likelihood Ratio. In: Wiley StatsRef: Statistics Reference Online [Internet]. 2014 [cited 2023 Feb 12]. Available from: <https://doi.org/10.1002/9781118445112.stat05882>
75. Pickering J, Endre Z. New Metrics for Assessing Diagnostic Potential of Candidate Biomarkers. *Clin J Am Soc Nephrol CJASN*. 2012 Jun;7:1355–64.
76. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, Liston DE, Low DKW, Newman SF, Kim J, Others. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. 2018;2(10):749.
77. Sterne J, White IR, Carlin JB, Spratt MP, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009 Jun;338:157–60.
78. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority over-Sampling Technique. *J Artif Int Res*. 2002 Jun;16(1):321–57.
79. Mohammed R, Rawashdeh J, Abdullah M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. In 2020. p. 243–8.
80. Potdar K, Pardawala T, Pai C. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *Int J Comput Appl*. 2017;175:7–9.
81. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Springer New York Inc.; 2001.
82. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-Learn: Machine Learning in Python. *J Mach Learn Res*. 2011 Nov;12(null):2825–30.

83. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc Ser B Methodol.* 1974;36(2):111–47.
84. Binney N, Hyde C, Bossuyt PM. On the Origin of Sensitivity and Specificity. *Ann Intern Med.* 2021 Mar 16;174(3):401–7.
85. King G. Unifying Political Methodology: The Likelihood Theory of Statistical Inference [Internet]. Ann Arbor: University of Michigan Press; 1998. Available from: <http://www.press.umich.edu/titleDetailDesc.do?id=23784>
86. Shapley LS. Notes on the n-Person Game — II: The Value of an n-Person Game. Santa Monica, CA: RAND Corporation; 1951.
87. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. *Advances in Neural Information Processing Systems 30* [Internet]. Curran Associates, Inc.; 2017. p. 4765–74. Available from: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
88. Wouters A, Nysten C, Thijs VN, Lemmens R. Prediction of Outcome in Patients With Acute Ischemic Stroke Based on Initial Severity and Improvement in the First 24 h. *Front Neurol.* 2018;9.
89. Kazi SA, Siddiqui M, Majid S. Stroke Outcome Prediction Using Admission Nihss In Anterior And Posterior Circulation Stroke. *J Ayub Med Coll Abbottabad JAMC.* 2021;33 2:274–8.
90. Koton S, Molshatzki N, Bornstein NM, Tanne D. Low cholesterol, statins and outcomes in patients with first-ever acute ischemic stroke. *Cerebrovasc Dis.* 2012;34(3):213–20.
91. den Hertog HM, van Rossum JA, van der Worp HB, van Gemert HM, de Jonge R, Koudstaal PJ, Dippel DW, van Gemert HM, Dippel DW, van der Worp HB, Vermeij FH, Kwa VI, Bienfait HP, Jansen BP, de Kort PL, van der Ree TC, Saxena R, Oomes PG, Huisman UW, Wouda EJ, Bouwsma C, Keizer K, Kleyweg RP, Kerkhoff H, Verbiest HB, Bakker SL, De Schryver EL, Donders RC, van der Kruijk RA, Heerema J, Bernsen RA, Herderschêe D, Keuter EJ, Baart WC, Meijer RJ, Alting van Geusau RB, de Ruitter JP, Wessel JN, Bollen AE. C-reactive protein in the very early phase of acute ischemic stroke: association with poor outcome and death. *J Neurol.* 2009 Dec;256(12):2003–8.
92. Sun W, Huang Y, Xian Y, Zhu S, Jia Z, Liu R, Li F, Wei J, Wang JG, Liu M, Anderson C. Association of body mass index with mortality and functional outcome after acute ischemic stroke. *Sci Rep.* 2017 May 31;7.

93. Monteiro M, Fonseca AC, Freitas AT, Pinho E Melo T, Francisco AP, Ferro JM, Oliveira AL. Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients. *IEEEACM Trans Comput Biol Bioinform.* 2018;15(6):1953–9.
94. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke.* 2019;50(5):1263–5.
95. Chiu IM, Zeng WH, Cheng CY, Chen SH, Lin CR. Using a Multiclass Machine Learning Model to Predict the Outcome of Acute Ischemic Stroke Requiring Reperfusion Therapy. *Diagn Basel.* 2021 Jan;11(1).
96. Wang A, Pednekar N, Lehrer R, Todo A, Sahni R, Marks S, Stiefel MF. DRAGON score predicts functional outcomes in acute ischemic stroke patients receiving both intravenous tissue plasminogen activator and endovascular therapy. *Surg Neurol Int.* 2017;8:149.
97. Alaka SA, Menon BK, Brobbey A, Williamson T, Goyal M, Demchuk AM, Hill MD, Sajobi TT. Functional Outcome Prediction in Ischemic Stroke: A Comparison of Machine Learning Algorithms and Regression Models. *Front Neurol.* 2020;11:889.
98. Ramos LA, Kappelhof M, van Os HJA, Chalos V, Van Kranendonk K, Kruyt ND, Roos YBWEM, van der Lugt A, van Zwam WH, van der Schaaf IC, Zwinderman AH, Strijkers GJ, van Walderveen MAA, Wermer MJH, Olabariaga SD, Majoie CBLM, Marquering HA. Predicting Poor Outcome Before Endovascular Treatment in Patients With Acute Ischemic Stroke. *Front Neurol.* 2020;11:580957.
99. Li X, Pan X, Jiang C, Wu M, Liu Y, Wang F, Zheng X, Yang J, Sun C, Zhu Y, Zhou J, Wang S, Zhao Z, Zou J. Predicting 6-Month Unfavorable Outcome of Acute Ischemic Stroke Using Machine Learning. *Front Neurol.* 2020;11:539509.
100. Cioncoloni D, Piu P, Tassi R, Acampa M, Guideri F, Taddei S, Bielli S, Martini G, Mazzocchio R. Relationship between the modified Rankin Scale and the Barthel Index in the process of functional recovery after stroke. *NeuroRehabilitation.* 2012;30(4):315–22.
101. Musa KI, Keegan TJ. The change of Barthel Index scores from the time of discharge until 3-month post-discharge among acute stroke patients in Malaysia: A random intercept model. *PLoS One.* 2018;13(12):e0208594.
102. Sale P, Ferriero G, Ciabattini L, Cortese AM, Ferracuti F, Romeo L, Piccione F, Masiero S. Predicting Motor and Cognitive Improvement Through Machine Learning Algorithm in Human Subject that Underwent a Rehabilitation Treatment in the Early Stage of Stroke. *J Stroke Cerebrovasc Dis.* 2018 Nov;27(11):2962–72.
103. De Wit L, Putman K, Devos H, Brinkmann N, Dejaeger E, De Weerdts W, Jenni W,

- Lincoln N, Schuback B, Schupp W. Long-term prediction of functional outcome after stroke using single items of the Barthel Index at discharge from rehabilitation centre. *Disabil Rehabil.* 2014;36(5):353–8.
104. Meyer MJ, Pereira S, McClure A, Teasell R, Thind A, Koval J, Richardson M, Speechley M. A systematic review of studies reporting multivariable models to predict functional outcomes after post-stroke inpatient rehabilitation. *Disabil Rehabil.* 2015;37(15):1316–23.
105. Ojala-Oksala J, Jokinen H, Kopsi V, Lehtonen K, Luukkonen L, Paukkunen A, Seck L, Melkas S, Pohjasvaara T, Karhunen P, Hietanen M, Erkinjuntti T, Oksala N. Educational history is an independent predictor of cognitive deficits and long-term survival in postacute patients with mild to moderate ischemic stroke. *Stroke.* 2012 Nov;43(11):2931–5.
106. Dong L, Briceno E, Morgenstern LB, Lisabeth LD. Poststroke Cognitive Outcomes: Sex Differences and Contributing Factors. *J Am Heart Assoc.* 2020;9(14):e016683.
107. Lee M, Oh MS, Jung S, Lee JH, Kim CH, Jang MU, Kim YE, Bae HJ, Park J, Kang Y, Lee BC, Lim JS, Yu KH. Differential effects of body mass index on domain-specific cognitive outcomes after stroke. *Sci Rep.* 2021;11(1):14168.
108. Youn YC, Choi SH, Shin HW, Kim KW, Jang JW, Jung JJ, Hsiung GR, Kim S. Detection of cognitive impairment using a machine-learning algorithm. *Neuropsychiatr Treat.* 2018;14:2939–45.
109. Zhu F, Panwar B, Dodge HH, Li H, Hampstead BM, Albin RL, Paulson HL, Guan Y. COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer’s disease. *Sci Rep.* 2016;6:34567.
110. Al-Qazzaz NK, Ali SH, Ahmad SA, Islam S, Mohamad K. Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment. *Neuropsychiatr Treat.* 2014;10:1677–91.
111. Srithumsuk W, Kabayama M, Gondo Y, Masui Y, Akagi Y, Klinpuatan N, Kiyoshige E, Godai K, Sugimoto K, Akasaka H, Takami Y, Takeya Y, Yamamoto K, Ikebe K, Ogawa M, Inagaki H, Ishizaki T, Arai Y, Rakugi H, Kamide K. The importance of stroke as a risk factor of cognitive decline in community dwelling older and oldest peoples: the SONIC study. *BMC Geriatr.* 2020;20(1):24.
112. Ariyanto Haroso A. Risk Factors for Cognitive Impairment after Ischemic Stroke. *KnE Life Sci.* 2018 Dec;4(9):152–160.
113. Casanova R, Saldana S, Lutz MW, Plassman BL, Kuchibhatla M, Hayden KM.

Investigating Predictors of Cognitive Decline Using Machine Learning. *J Gerontol B Psychol Sci Soc Sci.* 2020;75(4):733–42.

114. Aschwanden D, Aichele S, Ghisletta P, Terracciano A, Kliegel M, Sutin AR, Brown J, Allemand M. Predicting Cognitive Impairment and Dementia: A Machine Learning Approach. *J Alzheimers Dis.* 2020;75(3):717–28.

115. Shi Y, Xiang Y, Yang Y, Zhang N, Wang S, Ungvari GS, Chiu HF, Tang WK, Wang Y, Zhao X, Wang Y, Wang C. Depression after minor stroke: Prevalence and predictors. *J Psychosom Res.* 2015 Aug;79(2):143–7.

116. Khedr EM, Abdelrahman AA, Desoky TE, Zaki AF, Gamea A. Post-stroke depression: frequency, risk factors, and impact on quality of life among 103 stroke patients—hospital-based study. *Egypt J Neurol Psychiatry Neurosurg.* 2020;56:1–8.

117. Ayasrah SM, Ahmad MM, Basheti IA. Post-Stroke Depression in Jordan: Prevalence Correlates and Predictors. *J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc.* 2018;27 5:1134–42.

118. Shi Y, Yang D, Zeng Y, Wu W. Risk Factors for Post-stroke Depression: A Meta-analysis. *Front Aging Neurosci.* 2017;9:218.

119. Xu Q, Anderson D, Lurie-Beck J. The relationship between abdominal obesity and depression in the general population: A systematic review and meta-analysis. *Obes Res Clin Pr.* 2011;5(4):267–360.

120. Xu Z, Zhang Q, Li W, Li M, Yip PSF. Individualized prediction of depressive disorder in the elderly: A multitask deep learning approach. *Int J Med Inf.* 2019;132:103973.

121. Hama S, Yoshimura K, Yanagawa A, Shimonaga K, Furui A, Soh Z, Nishino S, Hirano H, Yamawaki S, Tsuji T. Relationships between motor and cognitive functions and subsequent post-stroke mood disorders revealed by machine learning analysis. *Sci Rep.* 2020;10(1):19571.

122. Li G, Jing P, Chen G, Mei J, Miao J, Sun W, Lan Y, Zhao X, Qiu X, Zhu Z, Zhu S. Development and Validation of 3-Month Major Post-Stroke Depression Prediction Nomogram After Acute Ischemic Stroke Onset. *Clin Interv Aging.* 2021;16:1439–47.

123. Cheon S, Kim J, Lim J. The Use of Deep Learning to Predict Stroke Patient Mortality. *Int J Env Res Public Health.* 2019;16(11).

124. Scrutinio D, Ricciardi C, Donisi L, Losavio E, Battista P, Guida P, Cesarelli M, Paganò G, D’Addio G. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Sci Rep.* 2020;10(1):20127.

125. Saposnik G. An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score. *Neurology*. 2012 Dec;79(23):2293–4.
126. Kandiah N, Chander RJ, Lin X, Ng A, Poh YY, Cheong CY, Cenina AR, Assam PN. Cognitive Impairment after Mild Stroke: Development and Validation of the SIGNAL2 Risk Score. *J Alzheimers Dis*. 2016;49(4):1169–77.
127. Chander RJ, Lam BYK, Lin X, Ng AYT, Wong APL, Mok VCT, Kandiah N. Development and validation of a risk score (CHANGE) for cognitive impairment after ischemic stroke. *Sci Rep*. 2017 Sep;7(1):12441.
128. Saposnik G, Kapral MK, Liu Y, Hall R, O'Donnell M, Raptis S, Tu JV, Mamdani M, Austin PC. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation*. 2011 Feb;123(7):739–49.
129. O'Donnell MJ, Fang J, D'Uva C, Saposnik G, Gould L, McGrath E, Kapral MK, Silver F, Kapral M, Tu J, Hall R, Richards J, Robertson A, Bayer N, Black S, Brunet D, Hachinski V, Howse D, Maher J, Oczkowski W, Phillips S, Selchen D, Sharma M, Silver F, Zalan P. The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke. *Arch Intern Med*. 2012 Nov;172(20):1548–56.
130. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *Npj Digit Med*. 2020 Feb 6;3(1):17.
131. Malsch C, Liman T, Wiedmann S, Siegerink B, Georgakis MK, Tiedt S, Endres M, Heuschmann PU. Outcome after stroke attributable to baseline factors-The PROSpective Cohort with Incident Stroke (PROSCIS). *PLoS One*. 2018;13(9):e0204285.

Eidesstattliche Versicherung

„Ich, Lea Fast, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Prädiktion klinischer Outcomes nach ischämischem Schlaganfall mit maschinellem Lernen (englisch *Prediction of clinical outcomes after ischemic stroke using machine learning*) 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, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; 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.“

Datum

Unterschrift

Anteilserklärung an der erfolgten Publikation

Publikation:

Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M, Ritter K. Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. *Front. Neurol.* 2023. 14:1114360. doi: 10.3389/fneur.2023.1114360

Beitrag im Einzelnen – Lea Fast:

- Verfassung und Einreichung des Datennutzungsantrags
- Kontrolle und manuelle Bereinigung des Datensatzes
- Manuelle Abgrenzung der Infarktareale mit MRICron in den DWI TRACE und FLAIR Sequenzen von ca. 50% der MRT-Bilddaten
- Programmierung der deskriptiven Statistik und Anteilen der Daten-Vorverarbeitung (Bereinigung, Binarisierung, Grenzwerte)
- Entwicklung und Durchführung der Analysen in Zusammenarbeit mit Uchralt Temuulen auf Basis der vorgegebenen Idee und Datensatzes
- Auswertung und Interpretation der Ergebnisse unter Supervision von Prof. Dr. rer. nat. Kerstin Ritter und Dr. med. Kersten Villringer
- Verfassung, Einreichung und Korrektur des Manuskripts unter Supervision von Prof. Dr. rer. nat. Kerstin Ritter, Dr. med. Kersten Villringer und Anna Kufner, MD PhD
- Erstellung von Tabellen (1-4, S1) und Abbildungen (1-3) sowie Tabellen (S2-S11) und Abbildungen (4-5) in Zusammenarbeit mit Uchralt Temuulen
- Literaturrecherche zu der vorliegenden Arbeit

Unterschrift der Doktorandin

Auszug aus der Journal Summary List

Journal Data Filtered By: **Selected JCR Year: 2021** Selected Editions: SCIE,SSCI
 Selected Categories: **"CLINICAL NEUROLOGY"** Selected Category
 Scheme: WoS

Gesamtanzahl: 212 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
1	LANCET NEUROLOGY	49,221	59.935	0.06262
2	Nature Reviews Neurology	18,852	44.711	0.02950
3	JAMA Neurology	21,007	29.907	0.05043
4	Alzheimers & Dementia	26,516	16.655	0.03970
5	ACTA NEUROPATHOLOGICA	30,046	15.887	0.03384
6	BRAIN	69,241	15.255	0.05988
7	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	38,495	13.654	0.02508
8	NEURO-ONCOLOGY	20,825	13.029	0.02439
9	PSYCHIATRY AND CLINICAL NEUROSCIENCES	6,445	12.145	0.00577
10	NEUROLOGY	115,450	11.800	0.09703
11	SLEEP MEDICINE REVIEWS	12,620	11.401	0.01356
12	Neurology-Neuroimmunology & Neuroinflammation	5,161	11.360	0.01049
13	ANNALS OF NEUROLOGY	45,647	11.274	0.03862
14	STROKE	86,009	10.170	0.07454
15	Stroke and Vascular Neurology	1,741	9.893	0.00369
16	MOVEMENT DISORDERS	35,923	9.698	0.02884
17	Brain Stimulation	10,760	9.184	0.01497
18	Alzheimers Research & Therapy	7,513	8.823	0.01287
19	Journal of Stroke	2,036	8.632	0.00391
20	JOURNAL OF HEADACHE AND PAIN	6,592	8.588	0.01107

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
21	PAIN	46,662	7.926	0.02957
22	Epilepsy Currents	1,428	7.872	0.00170
23	BRAIN PATHOLOGY	7,083	7.611	0.00689
24	NEUROSCIENTIST	6,603	7.235	0.00538
25	International Journal of Stroke	7,698	6.948	0.01415
26	Translational Stroke Research	4,125	6.800	0.00500
27	EPILEPSIA	36,294	6.740	0.02447
28	JOURNAL OF NEUROLOGY	24,906	6.682	0.02746
29	JOURNAL OF AFFECTIVE DISORDERS	59,622	6.533	0.06912
30	CNS DRUGS	6,627	6.497	0.00642
31	Therapeutic Advances in Neurological Disorders	2,816	6.430	0.00398
32	SLEEP	31,283	6.313	0.01904
33	EUROPEAN JOURNAL OF NEUROLOGY	17,087	6.288	0.01965
34	CURRENT OPINION IN NEUROLOGY	7,258	6.283	0.00814
35	NEUROPATHOLOGY AND APPLIED NEUROBIOLOGY	5,197	6.250	0.00447
36	Neurotherapeutics	7,998	6.088	0.00899
37	CEPHALALGIA	13,467	6.075	0.01470
38	Current Neurology and Neuroscience Reports	5,335	6.030	0.00686
39	European Stroke Journal	1,210	5.894	0.00411
40	Multiple Sclerosis Journal	15,617	5.855	0.01778
41	EUROPEAN ARCHIVES OF PSYCHIATRY AND CLINICAL NEUROSCIENCE	6,005	5.760	0.00577
42	INTERNATIONAL JOURNAL OF NEUROPSYCHOPHARMACOLOGY	8,630	5.678	0.00655

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
43	CLINICAL AUTONOMIC RESEARCH	2,484	5.625	0.00248
44	JOURNAL OF PAIN AND SYMPTOM MANAGEMENT	16,737	5.576	0.01740
45	NEUROLOGIA	2,146	5.486	0.00235
46	Annals of Clinical and Translational Neurology	5,916	5.430	0.01380
47	EUROPEAN NEUROPSYCHOPHARMACOLOGY	9,792	5.415	0.00876
48	JOURNAL OF NEUROSURGERY	44,049	5.408	0.02790
49	NEUROEPIDEMIOLOGY	4,955	5.393	0.00350
50	JOURNAL OF PAIN	13,915	5.383	0.01264
51	BIPOLAR DISORDERS	6,472	5.345	0.00541
52	NEUROSURGERY	34,433	5.315	0.02469
53	HEADACHE	11,263	5.311	0.01220
54	JOURNAL OF SLEEP RESEARCH	9,708	5.296	0.00912
55	PROGRESS IN NEURO-PSYCHOPHARMACOLOGY & BIOLOGICAL PSYCHIATRY	15,970	5.201	0.01295
56	JOURNAL OF THE PERIPHERAL NERVOUS SYSTEM	2,488	5.188	0.00249
57	JPAD-Journal of Prevention of Alzheimers Disease	873	5.020	0.00220
58	AMERICAN JOURNAL OF NEURORADIOLOGY	29,851	4.966	0.02313
59	NEUROREHABILITATION AND NEURAL REPAIR	7,555	4.895	0.00615
60	JOURNAL OF NEUROTRAUMA	19,581	4.869	0.01717
61	DEVELOPMENTAL MEDICINE AND CHILD NEUROLOGY	17,461	4.864	0.01202
62	CLINICAL NEUROPHYSIOLOGY	25,162	4.861	0.01585
63	SLEEP MEDICINE	17,340	4.842	0.01628
64	Multiple Sclerosis and Related Disorders	7,487	4.808	0.01284

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
65	Journal of Neurogastroenterology and Motility	2,606	4.725	0.00349
66	Journal of Neuromuscular Diseases	1,350	4.693	0.00275
67	Journal of Neurologic Physical Therapy	1,571	4.655	0.00153
68	CNS SPECTRUMS	3,558	4.604	0.00290
69	JOURNAL OF NEURORADIOLOGY	1,789	4.600	0.00198
70	JOURNAL OF PSYCHOPHARMACOLOGY	9,324	4.562	0.00949
71	JOURNAL OF THE NEUROLOGICAL SCIENCES	23,403	4.553	0.01804
72	Movement Disorders Clinical Practice	2,088	4.514	0.00399
73	JOURNAL OF NEURO-ONCOLOGY	16,733	4.506	0.01490
74	Neurology and Therapy	917	4.446	0.00196
75	JOURNAL OF NEURO-OPHTHALMOLOGY	2,897	4.415	0.00286
76	PARKINSONISM & RELATED DISORDERS	13,677	4.402	0.01510
77	Pain Physician	6,200	4.396	0.00630
78	CLINICAL NEUROPSYCHOLOGIST	6,059	4.373	0.00451
79	Neurosurgical Focus	10,856	4.332	0.01139
80	Journal of Clinical Sleep Medicine	11,389	4.324	0.01245
81	REVUE NEUROLOGIQUE	3,295	4.313	0.00325
82	Spine Journal	13,705	4.297	0.01567
83	Expert Review of Neurotherapeutics	5,795	4.287	0.00476
83	BRAIN TOPOGRAPHY	3,529	4.275	0.00351
85	Journal of Movement Disorders	709	4.229	0.00126
86	PEDIATRIC NEUROLOGY	7,875	4.210	0.00711
87	Sleep Health	2,865	4.207	0.00513



Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
87	Frontiers in Neurology	27,128	4.086	0.04871
89	Journal of Neurodevelopmental Disorders	2,054	4.074	0.00278
90	Epilepsia Open	1,125	4.026	0.00263
91	Current Treatment Options in Neurology	1,872	3.972	0.00239
92	JOURNAL OF NEUROSURGICAL ANESTHESIOLOGY	1,988	3.969	0.00148
92	Pain and Therapy	941	3.960	0.00171
94	NEUROGASTROENTEROLOGY AND MOTILITY	11,600	3.960	0.01224
94	ACTA NEUROLOGICA SCANDINAVICA	8,854	3.915	0.00637
96	Current Pain and Headache Reports	3,600	3.904	0.00341
97	MUSCLE & NERVE	16,166	3.852	0.01052
98	JOURNAL OF NEURAL TRANSMISSION	9,768	3.850	0.00683
99	NEUROLOGICAL SCIENCES	10,952	3.830	0.01262
100	NEUROLOGIC CLINICS	3,546	3.787	0.00283
101	EUROPEAN JOURNAL OF PAEDIATRIC NEUROLOGY	4,458	3.692	0.00541
102	Neurology-Genetics	1,641	3.663	0.00455
103	EUROPEAN JOURNAL OF PAIN	9,804	3.651	0.00876
104	NEUROMUSCULAR DISORDERS	7,104	3.538	0.00684
105	Neurocritical Care	6,230	3.532	0.00762
106	Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration	4,833	3.528	0.00437
107	Behavioral Sleep Medicine	2,419	3.492	0.00270
108	JOURNAL OF NEUROSURGERY-SPINE	10,739	3.467	0.01197
109	CLINICAL JOURNAL OF PAIN	8,913	3.423	0.00665
110	Neurodegenerative Diseases	1,871	3.417	0.00115

Druckexemplar der ausgewählten Publikation: Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M, Ritter K. Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. Front. Neurol. 2023. 14:1114360.

Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S,
Piper SK, Sperber PS, Liman T, Endres M, Ritter K.

**Machine learning-based prediction of clinical outcomes
after first-ever ischemic stroke.**

Front. Neurol. 2023. 14:1114360. doi: 10.3389/fneur.2023.1114360.

Impact Factor: 4,086



OPEN ACCESS

EDITED BY

Nishant K. Mishra,
Yale University, United States

REVIEWED BY

Amit Mehndiratta,
Indian Institute of Technology Delhi, India
Shubham Misra,
Yale University, United States

*CORRESPONDENCE

Kerstin Ritter
✉ kerstin.ritter@charite.de

SPECIALTY SECTION

This article was submitted to
Stroke,
a section of the journal
Frontiers in Neurology

RECEIVED 02 December 2022

ACCEPTED 31 January 2023

PUBLISHED 21 February 2023

CITATION

Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M and Ritter K (2023) Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. *Front. Neurol.* 14:1114360. doi: 10.3389/fneur.2023.1114360

COPYRIGHT

© 2023 Fast, Temuulen, Villringer, Kufner, Ali, Siebert, Huo, Piper, Sperber, Liman, Endres and Ritter. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke

Lea Fast¹, Uchralt Temuulen², Kersten Villringer², Anna Kufner^{2,3,4}, Huma Fatima Ali⁵, Eberhard Siebert⁶, Shufan Huo^{2,4,7}, Sophie K. Piper^{3,8,9}, Pia Sophie Sperber^{2,10,11,12}, Thomas Liman^{2,7,13,14}, Matthias Endres^{2,3,4,7,10,13} and Kerstin Ritter^{1,15*}

¹Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Psychiatry and Psychotherapy, Berlin, Germany, ²Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Center for Stroke Research Berlin (CSB), Berlin, Germany, ³Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Berlin, Germany, ⁴Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Neurology with Experimental Neurology, Berlin, Germany, ⁵Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany, ⁶Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Neuroradiology, Berlin, Germany, ⁷German Center for Cardiovascular Research (Deutsches Zentrum für Herz-Kreislaufkrankungen, DZHK), Partner Site Berlin, Berlin, Germany, ⁸Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Biometry and Clinical Epidemiology, Berlin, Germany, ⁹Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Informatics, Berlin, Germany, ¹⁰Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, NeuroCure Cluster of Excellence, NeuroCure Clinical Research Center (NCRC), Berlin, Germany, ¹¹Experimental and Clinical Research Center, A Cooperation Between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin, Berlin, Germany, ¹²Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany, ¹³German Center for Neurodegenerative Diseases (Deutsches Zentrum für Neurodegenerative Erkrankungen, DZNE), Partner Site Berlin, Berlin, Germany, ¹⁴Department of Neurology, Evangelical Hospital Oldenburg, Carl von Ossietzky-University, Oldenburg, Germany, ¹⁵Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Bernstein Center for Computational Neuroscience (BCCN), Berlin, Germany

Background: Accurate prediction of clinical outcomes in individual patients following acute stroke is vital for healthcare providers to optimize treatment strategies and plan further patient care. Here, we use advanced machine learning (ML) techniques to systematically compare the prediction of functional recovery, cognitive function, depression, and mortality of first-ever ischemic stroke patients and to identify the leading prognostic factors.

Methods: We predicted clinical outcomes for 307 patients (151 females, 156 males; 68 ± 14 years) from the PROSpective Cohort with Incident Stroke Berlin study using 43 baseline features. Outcomes included modified Rankin Scale (mRS), Barthel Index (BI), Mini-Mental State Examination (MMSE), Modified Telephone Interview for Cognitive Status (TICS-M), Center for Epidemiologic Studies Depression Scale (CES-D) and survival. The ML models included a Support Vector Machine with a linear kernel and a radial basis function kernel as well as a Gradient Boosting Classifier based on repeated 5-fold nested cross-validation. The leading prognostic features were identified using Shapley additive explanations.

Results: The ML models achieved significant prediction performance for mRS at patient discharge and after 1 year, BI and MMSE at patient discharge, TICS-M after 1 and 3 years and CES-D after 1 year. Additionally, we showed that National Institutes of Health Stroke Scale (NIHSS) was the top predictor for most functional recovery outcomes as well as education for cognitive function and depression.

Conclusion: Our machine learning analysis successfully demonstrated the ability to predict clinical outcomes after first-ever ischemic stroke and identified the leading prognostic factors that contribute to this prediction.

KEYWORDS

stroke, machine learning, outcome prediction, post-stroke depression, mortality, functional outcome, cognitive impairment

1. Introduction

Stroke is the second most common cause of death and a major cause of disability on a worldwide scale (1). It occurs when the blood supply to brain tissue is interrupted by either blockage (ischemic stroke) or bleeding caused by rupture of cerebral blood vessels (haemorrhagic stroke) ultimately resulting in irreversible neuronal death (2). The incidence of stroke is set to rise due to the demographic shift affecting populations across the globe (3). Thus, it is paramount to identify parameters that can aid in accurate prediction of long-term clinical outcome post-stroke.

In recent years the move toward electronic health records and the application of machine learning (ML) techniques in the medical research field have opened new frontiers of personalized medicine and decision support. The key advantage is that—in contrast to traditional statistical analyses—not only can predictors and biomarkers be identified on a group level, but ML techniques also enable prediction on an individual patient level. In other words, the outcome for a single patient can be predicted by considering a vast array of variables (4). Numerous studies have successfully demonstrated the ability of ML models to predict specific clinical outcomes after stroke with remarkable accuracy and identified leading baseline factors that carry high prognostic value (5–8). Most studies so far have focused on the prediction of the modified Rankin Scale (mRS) (9) as it is the gold standard for determining functional recovery after stroke. While there are some studies investigating the ML-based prediction of the Barthel Index (BI) (10) and Modified Telephone Interview for Cognitive Status (TICS-M) (11), research regarding the Center for Epidemiologic Studies Depression Scale (CES-D) (12) and Mini-Mental State Examination (MMSE) (13) is sparse. In addition, the heterogeneity of ML techniques, clinical outcomes and datasets used in these studies makes it difficult to assess the broader implications of their findings (4).

The primary aim of the present study was therefore to conduct a systematic comparison of ML-based outcome prediction after first-ever ischemic stroke featuring measures of functional recovery (mRS, BI), cognitive function (MMSE, TICS-M), depression (CES-D), and mortality. The analysis was based on three powerful ML models and an array of baseline features including demographic, clinical, serological and MRI variables. As a secondary aim, we set out to identify to the key prognostic markers for each outcome using state-of-the-art visualization techniques.

2. Methods

2.1. Dataset and feature selection

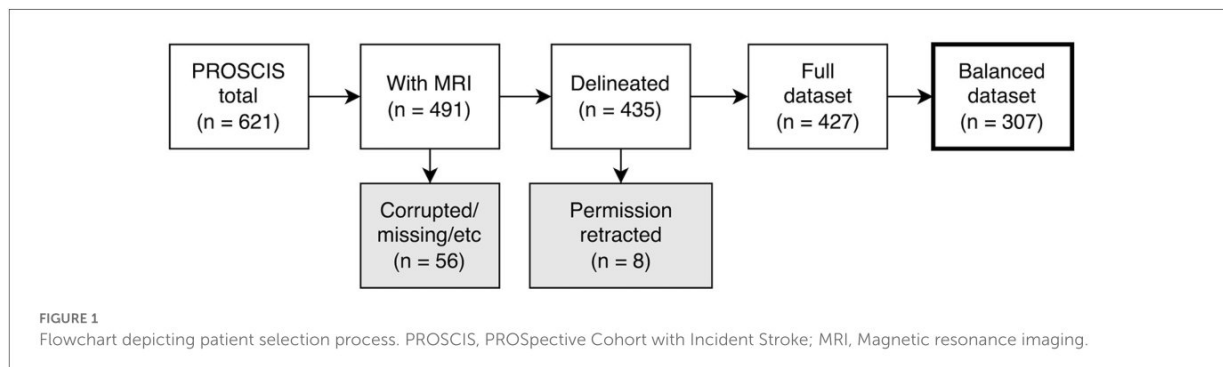
The patients included in these analyses were selected from the PROSpective Cohort with Incident Stroke Berlin (PROSCIS-B) study. Recruitment for this prospective cohort study was conducted over a three-year period starting in March 2010 at the Center for Stroke Research Berlin and Charité University Hospital with a consecutive three-year follow-up period. The study population consists of patients aged 18 years and over with acute first-ever stroke according to the WHO stroke criteria (14). The complete inclusion and exclusion criteria are described in detail on <https://clinicaltrials.gov> (NCT01363856). The study was approved by the ethics committee of the Charité - Universitätsmedizin Berlin (EA1/218/09) and was conducted in accordance with the Declaration of Helsinki. For the purposes of this exploratory analysis only patients with ischemic stroke and input features with no more than 15% missing values were included.

MRI data was collected after study completion from clinical routine data. In order to quantify the characteristics of the imaging data all acute and chronic stroke lesions were delineated on Diffusion-weighted imaging (DWI) and Fluid-attenuated inversion recovery (FLAIR) sequences, respectively, using MRIcron (15) from the Center for Advanced Brain Imaging (University of South Carolina, Chris Rordan, USA). The delineation and volume extraction for acute and chronic stroke lesions were performed by medical students supervised by two independent expert neuroradiologists while all further MRI parameters were obtained by expert neuroradiologists.

Due to significant differences in the number and mean age of female and male patients, we balanced the dataset by separating all patients into groups according to sex and age and then randomly selecting patients within these groups until there were no more significant differences (up to $p \leq 0.1$). This was necessary to ensure the predictions of our models were not based on an inherent bias in the training data (e.g., women being older on average and thus having worse outcomes) (16). The patient selection process is shown in Figure 1 and the characteristics of the dataset are described in Table 1.

2.2. Input data and outcomes

This study includes a total of 43 stroke-related baseline variables in four input subdomains. They consisted of 6



demographic and 16 clinical variables, 10 serological markers and 11 MRI parameters as listed in Table 1. Procalcitonin serum levels, which have previously been identified as a prognostic marker for 30-day mortality after stroke (18), had to be excluded since this variable had more than 15% missing values. The outcomes included measures of functional recovery (mRS and BI), cognitive function (MMSE and TICS-M), depression (CES-D) and survival. The mRS and BI were assessed at patient discharge, and 1 year post-stroke. Cognitive impairment was evaluated using the MMSE at discharge and later with the TICS-M at 1 and 3 years. CES-D and survival were also assessed 1 and 3 years after the index event. The follow-up process included an initial telephone assessment of cognitive function, followed by a structured interview conducted either by phone or mail. Table 2 shows the distribution of outcomes in the dataset, their respective follow-up time points, and the cut-off points for good vs. poor clinical outcome as defined by clinical scoring gold standards.

2.3. Machine learning analysis

The aim of this study was to conduct a systematic comparison of ML-based outcome prediction models after first-ever ischemic stroke. To accomplish this, a linear model, a non-linear model, and a tree-based model were selected for comparison (see Figure 2). To reduce complexity and potential problems brought on by multiple comparisons, a small set of three ML algorithms were selected. A Support Vector Machine (SVM) with linear kernel (SVM-lin) (19) and a SVM with radial basis function kernel (SVM-rbf) (20) were chosen as linear and non-linear models due to their strong performance in previous studies and the ability to directly compare them (6, 16, 21). Similarly, Gradient Boosting (GB) (22) was chosen as the tree-based classifier due to its superior performance and when compared to other tree-based models (23, 24). We compensated for missing data in the training and validation set with Multiple Imputation using Chained Equations (MICE) (25). The outcome class imbalances in the training set were counteracted with the Synthetic Minority Over-sampling Technique (SMOTE) (26) and random oversampling (27). Categorical input features were transformed using one-hot encoding. Then, models were carefully evaluated using ten times repeated 5-fold nested cross-validation with fixed seed to increase robustness (28). Here the data is split into five training (80%) and test sets (20%). Each of these

training sets is then subdivided into further five training (80%) and validation sets (20%). The hyperparameters of the ML models (listed in Supplementary Table S1) have been optimized on these training and validation sets *via* grid search before finally being evaluated on the unseen data of the test sets.

Performance of each model was evaluated using balanced accuracy (BA), area under the receiver operating characteristic curve, sensitivity, specificity, likelihood ratio (LR) and Integrated Discrimination Improvement index (IDI). BA is the arithmetic mean of sensitivity and specificity while the receiver operating characteristics curve (ROC) plots the true positive rate in relation to the false positive rate of the ML models. The area under the curve (AUC) of the ROC is routinely used as a measure of performance in ML. For each outcome, we reported the mean BA and AUC along with their standard deviation (SD) for ten iterations of 5-fold nested cross-validation. The LR compares the fit of two models by taking the ratio of their likelihoods (29) while the IDI ranks the model according to the change of the discrimination slopes (30). To test for statistical significance, we performed non-parametric permutation testing (31). Here, the exact same ML analysis and nested cross-validation procedure was performed a hundred times on randomly permuted ground truth labels before being compared to the original results. Results were considered statistically significant below $p \leq 0.05$ and $p \leq 0.01$ after Bonferroni correction for multiple comparisons (3 ML algorithms \times 5 feature subsets). We used the Python 3.6 programming language with the scikit-learn, pandas, statsmodel, matplotlib and seaborn packages for all analyses and visualizations.

2.4. Feature importance and Shapley values

In order to discern feature importance we implemented Shapley values using the SHAP (SHapley Additive exPlanations) framework (32). This statistic is a solution concept originating from cooperative game theory which calculates the relative importance of an input feature for the final prediction result and has already demonstrated convincing results in biomedical and clinical research applications (33, 34). Shapley values are calculated by determining the average marginal contribution of each feature over all possible combinations of input features. This is done by analyzing the effect of each feature on the prediction when it is included or excluded, while also taking into account the

TABLE 1 Baseline characteristics of patient population.

Parameter		Total <i>n</i> = 307	Female	Male	<i>P</i> -value
Demographic					
Sex, <i>n</i> (%)		307	151 (49.2)	156 (50.8)	
Age in years, Mean ± SD		68 ± 14	69 ± 15	67 ± 13	0.1
Education in years, Mean ± SD		14.1 ± 4.6	12.9 ± 4.3	15.3 ± 4.6	0.0
BMI, Mean ± SD		27.3 ± 4.8	27.5 ± 5.4	27.1 ± 4.0	0.16
Waist circumference [mm], Mean ± SD		997.4 ± 131.6	964.6 ± 142.1	1,028.1 ± 113.0	0.0
Hip circumference [mm], Mean ± SD		1,035.8 ± 112.5	1,034.7 ± 128.5	1,036.8 ± 95.8	0.58
Clinical					
Blood pressure, Mean ± SD	Systolic	140.1 ± 22.3	139.8 ± 22.8	140.3 ± 21.7	0.57
	Diastolic	76.1 ± 13.7	74.7 ± 13.0	77.4 ± 14.3	0.03
Alcohol consumption, <i>n</i> (%)		108 (35.2)	38 (25.2)	70 (44.9)	0.0
Smoking, <i>n</i> (%)	Active	102 (33.2)	39 (25.8)	63 (40.4)	0.1
	Never	130 (42.3)	71 (47.0)	59 (37.8)	
	Former	70 (22.8)	39 (25.8)	31 (19.9)	
Dependent before stroke, <i>n</i> (%)		39 (12.7)	22 (14.6)	17 (10.9)	0.39
Physically active, <i>n</i> (%)		101 (32.9)	44 (29.1)	57 (36.5)	0.09
Thrombolysis, <i>n</i> (%)		60 (19.5)	33 (21.9)	27 (17.3)	0.48
Revascularization, <i>n</i> (%)		5 (1.6)	0 (0.0)	5 (3.2)	0.04
NIHSS, Median [IQR]		2 [1–4]	2 [1–4]	2 [1–4]	0.41
Pre-existing conditions					
Diabetes Mellitus, <i>n</i> (%)		71 (23.1)	35 (23.2)	36 (23.1)	0.61
Hypercholesterinemia, <i>n</i> (%)		66 (21.5)	32 (21.2)	34 (21.8)	0.95
Hypertension, <i>n</i> (%)		198 (64.5)	98 (64.9)	100 (64.1)	0.98
Atrial fibrillations, <i>n</i> (%)		64 (20.8)	31 (20.5)	33 (21.2)	1.00
Angina pectoris, <i>n</i> (%)		37 (12.1)	14 (9.3)	23 (14.7)	0.18
Myocardial infarction, <i>n</i> (%)		10 (3.3)	2 (1.3)	8 (5.1)	0.33
Peripheral artery disease, <i>n</i> (%)		15 (4.9)	7 (4.6)	8 (5.1)	0.95
Serological markers					
Glucose [mmol/L], Mean ± SD		7.3 ± 3.2	7.6 ± 3.9	7.1 ± 2.5	0.41
HbA1c [mmol/mol], Mean ± SD		10.9 ± 15.5	12.2 ± 18.9	9.5 ± 10.7	0.21
Cholesterol [mmol/L], Mean ± SD		11.1 ± 2.7	11.3 ± 2.7	10.9 ± 2.7	0.10
HDL [mmol/L], Mean ± SD		2.9 ± 0.9	3.2 ± 0.9	2.6 ± 0.8	0.00
LDL [mmol/L], Mean ± SD		6.8 ± 2.2	6.8 ± 2.3	6.7 ± 2.1	0.46
Triglycerides [mmol/L], Mean ± SD		7.6 ± 4.8	7.0 ± 4.5	8.1 ± 5.1	0.00
Creatinine [μmol/L], Mean ± SD		82.4 ± 25.0	75.9 ± 23.1	88.6 ± 25.3	0.00
eGFR [ml/min], Mean ± SD		76.8 ± 20.6	73.5 ± 21.2	79.9 ± 19.6	0.02
hsCRP [mg/L], Mean ± SD		1.2 ± 1.9	1.3 ± 2.1	1.0 ± 1.7	0.15
White blood cells [cells/mm ³], Mean ± SD		8.0 ± 2.8	8.2 ± 2.8	7.8 ± 2.8	0.47
MRI					
Acute infarct DWI [ml], Mean ± SD		6.1 ± 14.5	5.6 ± 11.4	6.6 ± 17.1	0.17
Acute infarct FLAIR [ml], Mean ± SD		5.0 ± 12.9	4.6 ± 10.4	5.4 ± 15.0	0.11

(Continued)

TABLE 1 (Continued)

Parameter		Total <i>n</i> = 307	Female	Male	<i>P</i> -value
Infarct location, <i>n</i> (%)	Supratentorial	225 (73.3)	116 (76.8)	109 (69.9)	0.18
	Infratentorial	52 (16.9)	21 (13.9)	31 (19.9)	
	Both	30 (9.8)	14 (9.3)	16 (10.3)	
Infarct side, <i>n</i> (%)	Left	138 (45.0)	67 (44.4)	71 (45.5)	0.26
	Right	132 (43.0)	69 (45.7)	63 (40.4)	
	Both	37 (12.1)	15 (9.9)	22 (14.1)	
Unilateral infarct, <i>n</i> (%)		270 (87.9)	136 (90.1)	134 (85.9)	0.46
Chronic infarct, <i>n</i> (%)		79 (25.7)	37 (24.5)	42 (26.9)	0.90
Chronic infarct [ml], Mean ± SD		1.5 ± 2.5	1.3 ± 2.4	1.6 ± 2.6	0.34
Wahlund Score, (17) Median [IQR]		6 [3–10]	6 [3.5–11]	5 [2–8]	0.02
Infarct origin, <i>n</i> (%)	MCA	143 (46.6)	75 (49.7)	68 (43.6)	0.58
	ACA	1 (0.3)	1 (0.7)	0 (0.0)	
	PCA	17 (5.5)	10 (6.6)	7 (4.5)	
	AchA	18 (5.9)	8 (5.3)	10 (6.4)	
	Infratentorial	52 (16.9)	21 (13.9)	31 (19.9)	
	Thalamus	19 (6.2)	8 (5.3)	11 (7.1)	
	Multiple	57 (18.6)	28 (18.5)	29 (18.6)	
Infarct pattern, <i>n</i> (%)	Territorial	96 (31.3)	52 (34.4)	44 (28.2)	0.61
	Subcortical	74 (24.1)	37 (24.5)	37 (23.7)	
	Scattered	72 (23.5)	35 (23.2)	37 (23.7)	
	Lacunar	1 (0.3)	0 (0.0)	1 (0.6)	
	Infratentorial	52 (16.9)	21 (13.9)	31 (19.9)	
TOAST, <i>n</i> (%)	Large-artery	88 (28.7)	46 (30.5)	42 (26.9)	0.4
	Cardioembolism	89 (29.0)	50 (33.1)	39 (25.0)	
	Small-vessel	14 (4.6)	8 (5.3)	6 (3.8)	
	Other	40 (13.0)	17 (11.3)	23 (14.7)	
	Undefined	76 (24.8)	30 (19.9)	46 (29.5)	

BMI, body mass index; NIHSS, National Institute of Health Stroke Scale; TOAST, Trial of ORG 10172 in Acute Stroke Treatment; MCA, Middle cerebral artery; ACA, Anterior cerebral artery; PCA, Posterior cerebral artery; AchA, Anterior choroidal artery.

Data are given as mean ± standard deviation (SD) for continuous variables, median with limits of the interquartile range [25th–75th percentile] for ordinal variables and absolute (*n*) as well as relative (%) frequency for categorical variables. To determine significant differences between female and male patients we performed a t-test for continuous variables and a chi-squared test for categorical variables and reported the resulting *p*-values.

dependencies between features. For the purposes of this study, we implemented the Kernel SHAPexplainer which acts as a specially-weighted local linear regression (32).

3. Results

Out of the 621 PROSCIS-B patients 125 had no MRI associated with their study ID and in 5 further cases we were unable to locate the MRI data. This resulted in 491 patients with imaging data out of which 255 had received a 3T scan at the Center of Stroke Research Berlin (CSB) and 236 had been processed on scanners at Charité - Universitätsmedizin Berlin ranging from 1 to 1.5T, all of which were Siemens MRI units. In 56 cases the imaging data could not

be delineated due to missing sequences or motion artifacts and in 8 cases participants had retracted their consent for the study which resulted in a total of 427 fully delineated cases. The final balanced dataset consisted of 307 patients. There was a loss to follow-up of 74 patients (24.1%) in mRS, 105 patients (34.2%) in BI, 51 patients (26.2%) in TICS-M, and 49 patients (23.2%) in CES-D from the initial sample size. No loss was observed for mortality.

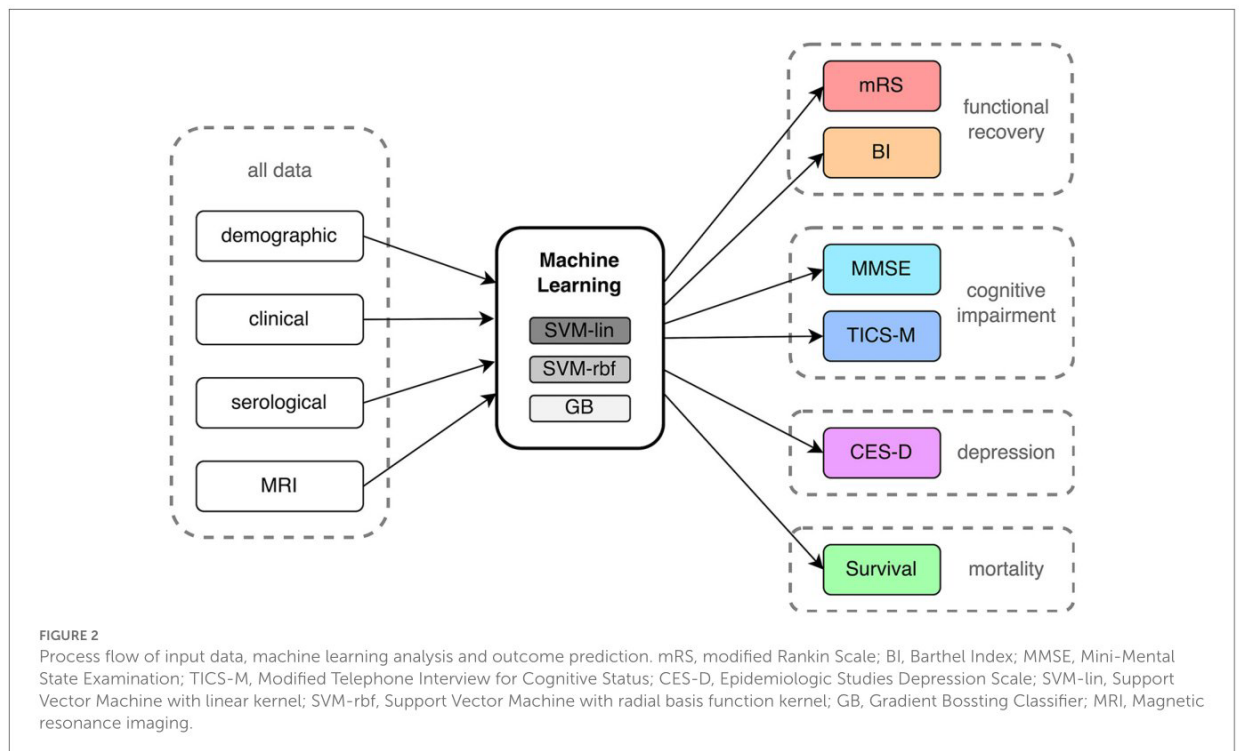
We evaluated and ranked the performance of the ML models using the metrics of BA and AUC. The results of these analyses can be found in [Supplementary Tables S2–S6](#). In [Figure 3](#), we show the performance in BA for all outcomes (mRS, BI, MMSE, TICS-M, CES-D, and survival), time points, and ML models (SVM-lin, SVM-rbf and GB). Additionally, we calculated the Integrated IDI and LR to provide further insight into the models' performance.

TABLE 2 Cut-offs and distribution of outcomes, listed as frequency for patient numbers in total, males, and females.

Distribution of outcomes in patient population			
Outcome	Time points	Good outcome, <i>n</i> (total/female/male)	Poor outcome, <i>n</i> (total/female/male)
mRS	PD	221/110/111	86/41/45
	Year 1	193/89/104	40/27/13
BI	PD	263/125/138	44/26/18
	Year 1	195/90/10	7/6/1
MMSE	PD	271/126/145	29/21/8
TICS-M	Year 1	147/69/78	48/32/16
	Year 3	125/60/65	19/8/11
CES-D	Year 1	163/79/93	48/35/13
	Year 3	132/53/79	30/19/11
Mortality	Year 1	271/132/139	36/19/17
	Year 3	142/78/64	165/73/92

Cut-off points for good vs. poor outcome			
Outcome	Total points	Good outcome	Poor outcome
mRS	0–6	0–2	3–6
BI	0–100	61–100	0–60
MMSE	0–30	24–30	0–23
TICS-M	0–50	30–50	0–29
CES-D	0–60	0–15	16–60

mRS, modified Rankin Scale; BI, Barthel Index; MMSE, Mini-Mental State Examination; TICS-M, Modified Telephone Interview for Cognitive Status; CES-D, Epidemiologic Studies Depression Scale; PD, patient discharge.



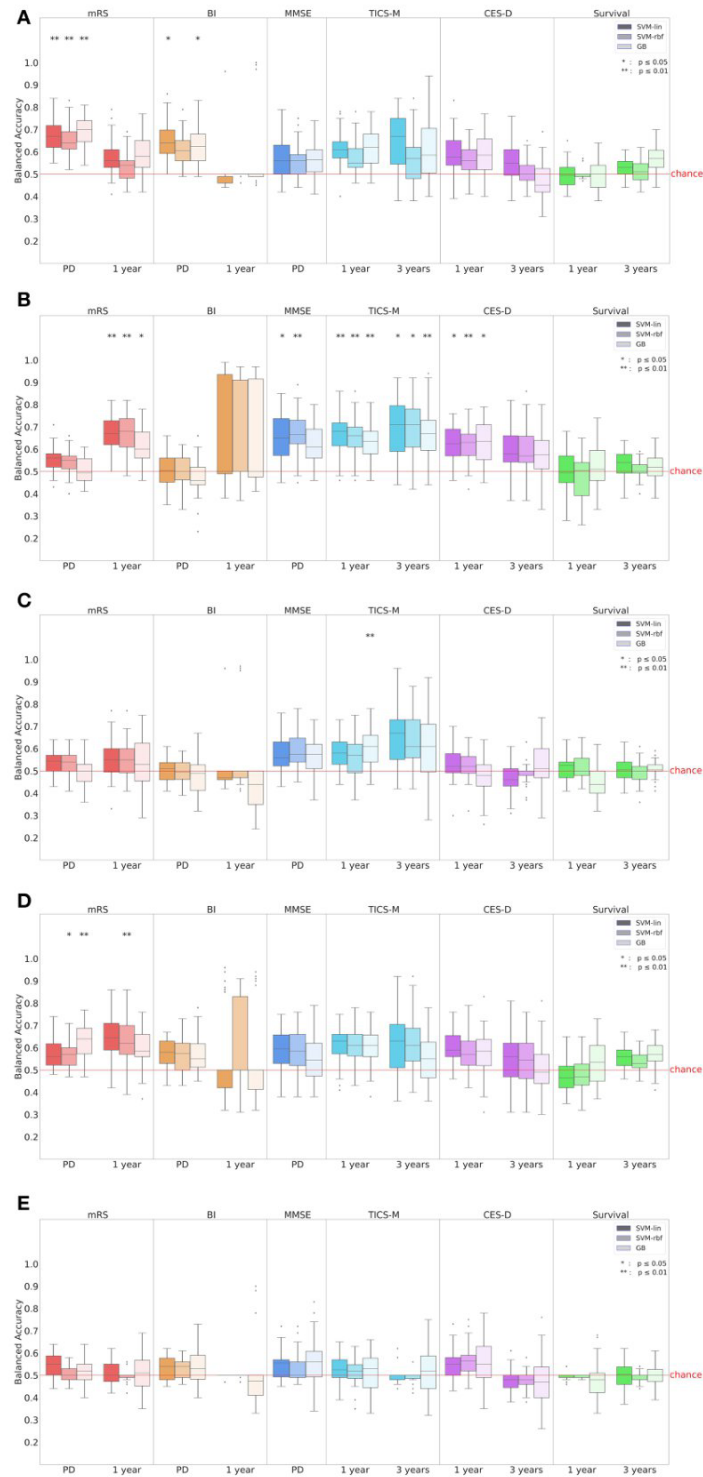


FIGURE 3 Prediction performance in balanced accuracy (BA) for all outcomes, time points and input subdomains. In (A) all input parameters were considered while (B–E) show the results of the (B) demographic, (C) clinical, (D) serological and (E) MRI input subdomain. Results for BI after 1 year were unreliable due to the extreme class imbalance in the dataset (see Table 2). mRS, modified Rankin Scale; BI, Barthel Index; MMSE, Mini-Mental State Examination; TICS-M, Modified Telephone Interview for Cognitive Status; CES-D, Epidemiologic Studies Depression Scale; SVM-lin, Support Vector Machine with linear kernel; SVM-rbf, Support Vector Machine with radial basis function kernel; GB, Gradient Boosting Classifier; MRI, Magnetic resonance imaging.

TABLE 3 Best prediction results and most important predictors for all outcomes as determined via Shapley values.

Outcome	Time	Model	Input	Mean absolute SHAP value	Variables
mRS	PD	GB	All	0.68 [0.64, 0.72]	NIHSS
				0.44 [0.41, 0.47]	hsCRP
				0.21 [0.19, 0.24]	Glucose
				0.18 [0.15, 0.21]	Cholesterol
				0.18 [0.15, 0.20]	Supra-/Infratentorial
	Year 1	SVM-rbf	Demographic	0.52 [0.47, 0.57]	Waist circumference [cm]
				0.50 [0.46, 0.54]	Sex
				0.47 [0.43, 0.51]	Age
				0.37 [0.33, 0.41]	Education [years]
				0.19 [0.15, 0.22]	BMI
BI	PD	SVM-lin	All	1.11 [1.05, 1.18]	NIHSS
				0.61 [0.57, 0.65]	Smoking
				0.46 [0.42, 0.49]	TOAST classification
				0.41 [0.36, 0.45]	Infarct pattern
				0.37 [0.34, 0.41]	Infarct origin
TICS-M	Year 1	SVM-lin	Demographic	0.68 [0.62, 0.73]	Education
				0.51 [0.46, 0.56]	Age
				0.40 [0.34, 0.46]	BMI
				0.20 [0.17, 0.23]	Sex
				0.19 [0.16, 0.21]	Hip circumference [cm]
	Year 3	SVM-rbf	Demographic	1.32 [1.16, 1.49]	Education [years]
				0.54 [0.48, 0.60]	Age
				0.42 [0.36, 0.47]	Sex
				0.38 [0.31, 0.44]	Waist circumference [cm]
				0.36 [0.31, 0.42]	Hip circumference [cm]
MMSE	PD	SVM-rbf	Demographic	0.48 [0.43, 0.53]	Education [years]
				0.36 [0.33, 0.38]	Sex
				0.35 [0.30, 0.41]	Age
				0.14 [0.11, 0.17]	Waist circumference [cm]
				0.13 [0.10, 0.15]	BMI
CES-D	Year 1	GB	Demographic	0.52 [0.49, 0.55]	Education [years]
				0.42 [0.38, 0.46]	Sex
				0.36 [0.31, 0.41]	BMI
				0.29 [0.25, 0.32]	Hip circumference [cm]
				0.21 [0.17, 0.26]	Waist circumference [cm]

mRS, modified Rankin Scale; BI, Barthel Index; MMSE, Mini-Mental State Examination; TICS-M, Modified Telephone Interview for Cognitive Status; CES-D, Epidemiologic Studies Depression Scale; PD, patient discharge; SVM-lin, Support Vector Machine with linear kernel; SVM-rbf, Support Vector Machine with radial basis function kernel; GB, Gradient Boosting Classifier; BMI, body mass index; NIHSS, National Institute of Health Stroke Scale; TOAST, Trial of ORG 10172 in Acute Stroke Treatment. The mean absolute SHAP value is reported with 95% confidence interval.

The detailed results are reported in [Supplementary Tables S7–S11](#). While the LR revealed no significant differences between the ML models it is important to note that the results obtained from the BA, AUC and the LR should be viewed independently, as they are based on different methods of evaluating the models' performance. Although in many cases the performance of the three ML models was at a comparable level the strongest predictive performance overall was achieved by SVM-rbf for TICS-M after 3 years (BA \pm SD = 0.7 ± 0.13 ; AUC \pm SD = 0.76 ± 0.13 ; $p \leq 0.05$) using the demographic input subdomain. [Table 3](#) states the most important predictors according to the Shapley values. The following paragraphs will list significant results ($p \leq 0.05$ or $p \leq 0.01$ Bonferroni corrected) according to the permutation test for each outcome per input subdomain.

3.1. Modified Rankin Scale

The highest prediction score for mRS at patient discharge was achieved by GB (BA \pm SD = 0.69 ± 0.07 ; AUC \pm SD = 0.77 ± 0.06 ; $p \leq 0.01$) followed by SVM-lin (BA \pm SD = 0.67 ± 0.07 ; AUC \pm SD = 0.74 ± 0.07 ; $p \leq 0.01$) and SVM-rbf (BA \pm SD = 0.65 ± 0.06 ; AUC \pm SD = 0.77 ± 0.06 ; $p \leq 0.01$) using all input parameters. In the serological input subdomain GB (BA \pm SD = 0.63 ± 0.07 ; AUC \pm SD = 0.68 ± 0.08 ; $p \leq 0.01$) and SVM-rbf (BA \pm SD = 0.57 ± 0.06 ; AUC \pm SD = 0.63 ± 0.07 ; $p \leq 0.05$) attained significant prediction results. The top five predictors using all input parameters were National Institutes of Health Stroke Scale (NIHSS), hsCRP, glucose, cholesterol and supra-/infratentorial infarct location.

The mRS after 1 year could best be predicted using the demographic input subdomain by SVM-rbf (BA \pm SD = 0.68 ± 0.09 ; AUC \pm SD = 0.73 ± 0.01 ; $p \leq 0.01$) followed by SVM-lin (BA \pm SD = 0.67 ± 0.08 ; AUC \pm SD = 0.73 ± 0.01 ; $p \leq 0.01$) and GB (BA \pm SD = 0.61 ± 0.08 ; AUC \pm SD = 0.66 ± 0.09 ; $p \leq 0.05$). In the serological input subdomain, SVM-rbf (BA \pm SD = 0.63 ± 0.1 ; AUC \pm SD = 0.64 ± 0.12 ; $p \leq 0.01$) led in prediction results. Waist circumference, sex, age, education, and BMI were the leading predictors in the demographic input subdomain.

3.2. Barthel Index

For BI at patient discharge, SVM-lin (BA \pm SD = 0.65 ± 0.08 ; AUC \pm SD = 0.73 ± 0.11 ; $p \leq 0.05$) and GB (BA \pm SD = 0.63 ± 0.08 ; AUC \pm SD = 0.74 ± 0.07 ; $p \leq 0.05$) achieved significant prediction results using all input parameters. The strongest predictors were NIHSS, smoking, the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification, infarct pattern and infarct origin. However, BI after 1 year could not be predicted by any model.

3.3. Mini-Mental State Examination

The leading ML models for predicting MMSE at patient discharge were SVM-rbf (BA \pm SD = 0.67 ± 0.09 ; AUC \pm SD =

0.71 ± 0.11 ; $p \leq 0.01$) and SVM-lin (BA \pm SD = 0.65 ± 0.1 ; AUC \pm SD = 0.7 ± 0.1 ; $p \leq 0.05$) using the demographic input subdomain with education, sex, age, waist circumference and BMI being the most important predictors.

3.4. Modified Telephone Interview for Cognitive Status

The best predictions for TICS-M after 1 year were by SVM-lin (BA \pm SD = 0.67 ± 0.09 ; AUC \pm SD = 0.73 ± 0.09 ; $p \leq 0.01$), SVM-rbf (BA \pm SD = 0.65 ± 0.09 ; AUC \pm SD = 0.72 ± 0.09 ; $p \leq 0.01$) and GB (BA \pm SD = 0.63 ± 0.08 ; AUC \pm SD = 0.69 ± 0.11 ; $p \leq 0.01$) using the demographic input subdomain. Further significant prediction results were achieved by GB (BA \pm SD = 0.6 ± 0.08 ; AUC \pm SD = 0.66 ± 0.1 ; $p \leq 0.01$) using the clinical input subdomain. The top five predictors in the demographic input subdomain were education, age, BMI, sex, and hip circumference. TICS-M after 3 years was most successfully predicted by SVM-rbf (BA \pm SD = 0.7 ± 0.13 ; AUC \pm SD = 0.76 ± 0.13 ; $p \leq 0.05$), SVM-lin (BA \pm SD = 0.69 ± 0.14 ; AUC \pm SD = 0.77 ± 0.13 ; $p \leq 0.05$) and GB (BA \pm SD = 0.68 ± 0.12 ; AUC \pm SD = 0.74 ± 0.13 ; $p \leq 0.01$) using the demographic input subdomain. Education, age, sex, waist circumference, and hip circumference were the leading variables.

3.5. Center for epidemiologic studies depression scale

For the prediction of CES-D after 1 year the use of the demographic input subdomain led to a significant prediction performance by GB (BA \pm SD = 0.63 ± 0.09 ; AUC \pm SD = 0.7 ± 0.1 ; $p \leq 0.05$), SVM-lin (BA \pm SD = 0.63 ± 0.08 ; AUC \pm SD = 0.68 ± 0.1 ; $p \leq 0.05$) and SVM-rbf (BA \pm SD = 0.62 ± 0.07 ; AUC \pm SD = 0.7 ± 0.09 ; $p \leq 0.01$). The strongest predictors were education, sex, BMI as well as hip and waist circumference. No ML model achieved significant prediction results for CES-D after 3 years.

3.6. Survival

Survival within 1 or 3 years could not be predicted reliably by any model.

4. Discussion

To the best of our knowledge, this is the first study to apply highly comparable standardized ML models to predict a wide range of long-term patient outcomes including functional recovery, cognitive impairment, depression, and mortality from a single, homogenous patient collective. While functional recovery scores like mRS and BI are often used as primary outcome endpoints in most major stroke cohorts, cognitive impairment and depression play a vital role in terms of long-term patient outcome. Up to 80% of patients are affected by cognitive impairment post-stroke and up



to 30% will develop a clinically relevant depression within 2 years after the index event (35, 36). These factors not only negatively affect functional recovery by decreasing a patient's capability for actively participating in rehabilitation measures but also disrupt their social integration. Although numerous previous studies have used similar ML models to predict functional recovery after stroke (5), here we demonstrate the accuracy of ML models to predict post-stroke cognitive status and depression up to 3 years post-stroke, as well as functional recovery.

Our results are in line with previous studies in identifying NIHSS as the leading predictor for mRS at patient discharge amongst all input variables (37, 38). Increased levels of hsCRP were correlated with poor clinical outcome which supports findings reported by den Hertog et al. (39) in acute stroke. Interestingly, waist circumference was the leading predictor for mRS after 1 year. Being underweight (BMI < 18.5 kg/m²) has previously been associated with unfavorable outcomes in terms of mortality and functional recovery in previous studies (40). Figure 4 illustrates the decision-making process of GB for mRS at patient discharge on a single-subject level.

In a study by Monteiro et al. (6) various ML models were applied to predict mRS after 3 months from 425 patients using 152 input variables. The best performance using baseline variables was achieved using a Random Forest (RF) classifier with an AUC of 0.808 ± 0.085 . In a separate study by Heo et al. (7) a DNN was used on 3,522 patients and achieved a classification accuracy of AUC = 0.888 with no reported SD. However, the authors did not mention whether cross-validation or repetition were used, which are important for developing a robust ML model and avoiding overfitting. In a study by Li et al. (21) predicting mRS after 6 months a SVM (AUC = 0.865; 95% CI 0.823–0.907) performed comparably well with six other models, including a RF classifier (AUC = 0.874;

95% CI 0.835–0.912) and a DNN (AUC 0.867; 95% CI 0.827–0.908). In contrast, in our study, for mRS at patient discharge the SVM-lin (AUC \pm SD = 0.74 ± 0.07) was outperformed by GB (AUC \pm SD = 0.77 ± 0.06). However, comparing the results of these studies is challenging due to variations in follow-up time points, input variables, methodology, and performance measures. Nevertheless, it appears that SVMs tend to perform similarly to, or worse than, tree-based classifiers or DNNs for predicting mRS outcomes.

Considerable overlap exists between mRS and BI in the development of functional recovery post stroke (41). This is reflected in NIHSS being the leading predictor for BI at patient discharge. Our results also confirm the relative importance of stroke origin for this outcome (42). The BI after one year could not be predicted—this may be due to the extreme class imbalance of this outcome (see Table 2). In contrast, in a study by den Hertog et al. (39) a ML model for identifying prognostic factors for motor and cognitive improvement after post-stroke rehabilitative training was developed based on a SVM-lin. The model included 55 patients and the results of the ischemic test set reported performance scores of correlation = 0.75, MADP = 87,03% and RMSE = 21,74 for BI. The most important parameters for the prediction were identified as the Functional Independence Measure and BI at patient discharge as well as serological markers such as Platelet-to-lymphocyte ratio, Red Cell Distribution Width and Lymphocytes.

Amongst the leading predictors for cognitive function post-stroke were demographic factors such as education, age and BMI which confirms previously published results (43, 44). While our findings are in line with the results by Casanova et al. (45) and Aschwanden et al. (46) their studies additionally identified the importance of socioeconomic status and ethnicity in terms of cognitive function post-stroke. Unfortunately, in the current study, these variables could not be accounted for.

Education being the top predictor for levels of depression after 1 year is in accordance with several studies linking low education level to an increased risk of post-stroke depression (47). Previous studies have found a significant association between higher waist circumference with an elevated rate of depression (48). In the current analysis, female sex was also identified as an important predictor of depression (49). A study by Hama et al. (50) achieved an impressive AUC above 0.90 for the prediction of post-stroke depression using a probabilistic artificial neural network on 274 stroke inpatients at the Hibino Hospital. The predicted clinical score was the Hospital Anxiety and Depression Scale and its lead predictors were the Japanese Perceived Stress Scale, the Symbol Digit Modalities Test, tapping span backward, visual cancellation Kana time and the Continuous Performance Test. This jump in prediction accuracy may be explained in part by the inclusion of these very specific test scores.

4.1. Methodological considerations

While many previous ML-based studies achieved noteworthy results, there are some potentially problematic methodological factors to consider: ideally, a ML model is trained and tested on numerous different samples in order to create a robust predictor for new, unseen data (51). In face of limited clinical data, it is crucial to include a re-sampling procedure to ensure effective training (52). Additionally, few studies performed more than one iteration of their analyses which negatively impacts robustness (28). In our study, we accounted for these factors by using a repeated 5-fold nested cross-validation. Furthermore, many studies use datasets and ML methods specific to the purpose of predicting an individual outcome. This impedes comparability as it remains unclear whether differences in performance are based on variations in input data or technical aspects of the ML analysis (5). Neglecting to balance these datasets regarding age and sex may also lead to biased results (53). We therefore balanced the dataset according to age and sex and predicted a range of clinical outcomes from the same dataset using three classical ML models while ensuring independence between training and test data. In addition, and in contrast to previous ML studies, we estimated the relative importance of features using Shapley values allowing to assess the impact of different input features for clinical outcome prediction in individual patients (see Figure 4).

4.2. Clinical implications

In the coming years, the advancement of big data analytics based on collaboration networks and electronic health records is set to drive a paradigm shift in clinical research (54). Novel automated and computer-based methods will play a key role in making use of increasing datasets and processing power. Therefore, we take a crucial step forward in the application of ML-based research methods to one of the most common and severe diseases around the globe and show that established as well as less traditional risk predictors can be identified and reproduced with ML techniques even in a limited sample size.

There is currently no established prediction score for depression outcomes following ischemic stroke. However, there are already a variety of scores available in the scientific literature for predicting functional outcomes (such as the Wang et al. (55) and ASTRAL (56) scores), cognitive outcomes (such as the CHANGE (57) and SIGNAL2 (58) scores), and mortality outcomes (such as the iScore (59) and PLAN (60) scores). In future studies, the aim should be to develop a universal model that can predict multiple outcomes-including functional recovery, cognitive impairment, depression, and mortality outcomes-using a basic set of variables such as NIHSS, education, sex, age, or BMI. This model would ideally be an easy-to-use tool for clinicians in real-world medical practice and act as an AI-based clinical decision support system (CDSS). The implementation of CDSS has been shown to be a cost-effective and efficient method for enhancing clinical workflow and decision-making (61). CDSSs have the potential to enhance patient safety by mitigating the occurrence of oversights and treatment errors. In the case of stroke, functional recovery is heavily dependent on rehabilitation measures which in turn requires adequate cognitive function and management of post-stroke depression (62, 63). The ability of CDSSs to alert providers to potential challenges in the management process can provide valuable guidance for more personalized rehabilitation programs and patient-tailored secondary prevention strategies, ultimately improving post-stroke outcomes.

4.3. Limitations

This study has several limitations that warrant discussion. First and foremost, this study had a limited sample size, the outcome classes were imbalanced, and an external control dataset was lacking. The application of 5-fold nested cross-validation, SMOTE and random oversampling partially counteract these limitations. To avoid shortcut learning and develop a model representative of the general population, we balanced our dataset by age and sex. Shortcut learning occurs when the model relies heavily on easily observable features like age rather than underlying causes, leading to potential biases and inaccuracies when applied to individuals outside the trained age range. However, this approach does not account for the natural incidence variation within the population, which may impact the ML model's predictions. Additionally, most of the patients included in this study had relatively mild to moderate strokes (NIHSS median of 2 (1–4)); this may have negatively affected prediction performance and limits generalizability to more severely affected stroke cohorts. There was also no data available on whether patients entered a rehabilitation program post-stroke, or which secondary prevention strategies were initiated. Therefore, these factors could not be accounted for in terms of post-stroke outcome endpoints in this analysis.

5. Conclusion

Based on a systematic comparison, the results of this study demonstrated the viability of ML-based outcome prediction after first-ever ischemic stroke for functional recovery, cognitive function, depression, and mortality. Compared to group-based

statistical analyses, the advantage of ML-techniques is their ability to make predictions on a single-subject level by considering a multitude of variables which is key for future application in clinical routine. Furthermore, we extracted the most important prognostic variables for each outcome. On the one hand, the results confirmed several already established prognostic markers and on the other identified novel candidates such as education, hsCRP and waist circumference as relevant predictors of important clinical endpoints. However, further studies are needed to confirm these findings and to establish their clinical viability.

Data availability statement

The PROCIS-B data is available upon request from TL. The code and results data are available upon request from KR.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the Charité - Universitätsmedizin Berlin (EA1/218/09). The patients/participants or their legal representative provided their written informed consent to participate in this study.

Author contributions

LF, KV, AK, and KR: conceptualization. LF, UT, KV, AK, HA, SP, and KR: data curation. LF, UT, and KR: formal analysis, methodology, visualization, and software. LF, TL, and KR: project administration. LF: writing-original draft. KV, AK, ES, SH, SP, PS, TL, ME, and KR: writing-review and editing. KV, TL, and KR: resources. KV and KR: supervision. All authors contributed to the article and approved the submitted version.

Funding

We acknowledge support from the German Research Foundation (DFG, 389563835; 402170461-TRR 265; 414984028-CRC 1404; 42075332-RU 5187) and the Manfred and

Ursula-Müller Stiftung. ME received funding from DFG under Germany's Excellence Strategy-EXC-2049-390688087, Collaborative Research Center ReTune TRR 295-424778381, Bundesministerium für Bildung und Forschung (BMBF), Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Deutsches Zentrum für Herz-Kreislaufkrankungen (DZHK), EU, Corona Foundation, and Fondation Leducq.

Acknowledgments

We thank Evert de Man for supplying the ML toolbox, Ralf Mекle for access to the server infrastructure and Jane Thümmel for data management.

Conflict of interest

ME reports grants from Bayer and fees paid to the Charité from Abbot, Amgen, AstraZeneca, Bayer, 296 Boehringer Ingelheim, BMS, Daiishi Sankyo, Sanofi, Novartis, Pfizer, all outside the submitted work.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1114360/full#supplementary-material>

References

- Gorelick PB. The global burden of stroke: persistent and disabling. *Lancet Neurol.* (2019) 18:417–8. doi: 10.1016/S1474-4422(19)30030-4
- Deb P, Sharma S, Hassan KM. Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. *Pathophysiology.* (2010) 17:197–218. doi: 10.1016/j.pathophys.2009.12.001
- Donkor ES. Stroke in the 21st century: a snapshot of the burden, epidemiology, and quality of life. *Stroke Res Treat.* (2018) 2018:3238165. doi: 10.1155/2018/3238165
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7
- Wang W, Kiik M, Peek N, Curcin V, Marshall JJ, Rudd AG, et al. systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE.* (2020) 15:e0234722. doi: 10.1371/journal.pone.0234722
- Monteiro M, Fonseca AC, Freitas AT, Pinho E, Melo T, Francisco AP, Ferro JM, Oliveira AL. Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEEACM Trans Comput Biol Bioinform.* (2018) 15:1953–9. doi: 10.1109/TCBB.2018.2811471
- Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke.* (2019) 50:1263–5. doi: 10.1161/STROKEAHA.118.024293
- Chiu IM, Zeng WH, Cheng CY, Chen SH, Lin CR. Using a Multiclass machine learning model to predict the outcome of acute ischemic stroke requiring reperfusion therapy. *Diagn Basel.* (2021) 11:80. doi: 10.3390/diagnostics11010080
- Saver JL, Filip B, Hamilton S, Yanes A, Craig S, Cho M, et al. FAST-MAG investigators and coordinators. Improving the reliability of stroke disability grading

- in clinical trials and clinical practice: the Rankin Focused Assessment (RFA). *Stroke*. (2010) 41:992–5. doi: 10.1161/STROKEAHA.109.571364
10. Dewing J. A critique of the Barthel Index. *Br J Nurs Mark Allen Publ.* (1992) 1:325–9. doi: 10.12968/bjon.1992.1.7.325
 11. Cook SE, Marsiske M, McCoy KJM. The use of the modified Telephone Interview for Cognitive Status (TICS-M) in the detection of amnesic mild cognitive impairment. *J Geriatr Psychiatry Neurol.* (2009) 22:103–9. doi: 10.1177/0891988708328214
 12. Carleton RN, Thibodeau MA, Teale MJN, Welch PG, Abrams MP, Robinson T, et al. The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. *PLoS ONE.* (2013) 8:e58067–e58067. doi: 10.1371/journal.pone.0058067
 13. Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Appl Nurs Res.* (2000) 13:209–13. doi: 10.1053/apnr.2000.9231
 14. Liman T, Zietemann V, Wiedmann S, Jungehülsing G, Endres M, Wollenweber F, et al. Prediction of vascular risk after stroke - protocol and pilot data of the Prospective Cohort with Incident Stroke (PROSCIS). *Int J Stroke Off J Int Stroke Soc.* (2012) 8:484–90. doi: 10.1111/j.1747-4949.2012.00871.x
 15. Rorden C, Brett M. Stereotaxic display of brain lesions. *Behav Neurol.* (2000) 12:191–200. doi: 10.1155/2000/421719
 16. Rane RP, de Man EF, Lim JH, Gorgen K, Tschorn M, Rapp MA, et al. Structural differences in adolescent brains can predict alcohol misuse. *eLife.* (2022) 11:e77545. doi: 10.7554/eLife.77545
 17. Wahlund LO, Barkhof F, Fazekas F, Bronge L, Augustin M, Sjögren M, et al. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke.* (2001) 32:1318–22. doi: 10.1161/01.STR.32.6.1318
 18. Yan L, Wang S, Xu L, Zhang Z, Liao P. Procalcitonin as a prognostic marker of patients with acute ischemic stroke. *J Clin Lab Anal.* (2020) 34:e23301. doi: 10.1002/jcla.23301
 19. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory (COLT '92)*. New York, NY: Association for Computing Machinery (1992). p. 144–52. doi: 10.1145/130385.130401
 20. Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. *Mach Learn.* (2002) 46:131–59. doi: 10.1023/A:1012450327387
 21. Li X, Pan X, Jiang C, Wu M, Liu Y, Wang F, et al. Predicting 6-month unfavorable outcome of acute ischemic stroke using machine learning. *Front Neurol.* (2020) 11:539509. doi: 10.3389/fneur.2020.539509
 22. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* (2001) 29:1189–232. doi: 10.1214/aos/1013203450
 23. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevskaya O. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med.* (2019) 7:152. doi: 10.21037/atm.2019.03.29
 24. Ogutu JO, Piepho HP, Schulz-Streeck T. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* (2011) 5:S11. doi: 10.1186/1753-6561-5-S3-S11
 25. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* (2011) 20:40–9. doi: 10.1002/mpr.329
 26. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE synthetic minority over-sampling technique. *J Artif Int Res.* (2002) 16:321–57. doi: 10.1613/ja.ir.953
 27. Mohammed R, Rawashdeh J, Abdullah M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. In: *11th International Conference on Information and Communication Systems (ICICS)*. Irbid (2020). p. 243–8. doi: 10.1109/ICICS49469.2020.239556
 28. Qayyum A, Qadir J, Bilal M, Al-Fuqaha A. Secure and robust machine learning for healthcare: a survey. *IEEE Rev Biomed Eng.* (2021) 14:156–80. doi: 10.1109/RBME.2020.3013489
 29. Birknes D. Likelihood Ratio. In: Armitage P, and Colton T, editors. *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons. (2005). doi: 10.1002/0470011815.b2a15073
 30. Pickering J, Endre Z. New metrics for assessing diagnostic potential of candidate biomarkers. *Clin J Am Soc Nephrol CJASN.* (2012) 7:1355–64. doi: 10.2215/CJN.09590911
 31. Nichols T, Holmes A. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp.* (2002) 15:1–25. doi: 10.1002/hbm.1058
 32. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*. Long Beach, CA: Curran Associates, Inc. (2017). p. 4765–4774.
 33. Shapley LS. *Notes on the n-Person Game — II: The Value of an n-Person Game*. Santa Monica, CA: RAND Corporation. (1951).
 34. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* (2018) 2:749. doi: 10.1038/s41551-018-0304-0
 35. Leśniak M, Bak T, Czepiel W, Seniów J, Członkowska A. Frequency and prognostic value of cognitive disorders in stroke patients. *Dement Geriatr Cogn Disord.* (2008) 26:356–63. doi: 10.1159/000162262
 36. Hackett ML, Pickles K. Part I: frequency of depression after stroke: an updated systematic review and meta-analysis of observational studies. *Int J Stroke.* (2014) 9:1017–25. doi: 10.1111/ijvs.12357
 37. Wouters A, Nysten C, Thijs VN, Lemmens R. Prediction of Outcome in Patients With Acute Ischemic Stroke Based on Initial Severity and Improvement in the First 24 h. *Front Neurol.* (2018) 9:308. doi: 10.3389/fneur.2018.00308
 38. Kazi SA, Siddiqui M, Majid S. Stroke outcome prediction using admission nihns in anterior and posterior circulation stroke. *J Ayub Med Coll Abbottabad JAMC.* (2021) 2:274–8.
 39. den Hertog HM, van Rossum JA, van der Worp HB, van Gemert HM, de Jonge R, Koudstaal PJ, et al. C-reactive protein in the very early phase of acute ischemic stroke: association with poor outcome and death. *J Neurol.* (2009) 256:2003–8. doi: 10.1007/s00415-009-5228-x
 40. Sun W, Huang Y, Xian Y, Zhu S, Jia Z, Liu R, et al. Association of body mass index with mortality and functional outcome after acute ischemic stroke. *Sci Rep.* (2017) 7:2507. doi: 10.1038/s41598-017-02551-0
 41. Cioncoloni D, Piu P, Tassi R, Acampa M, Guideri F, Taddei S, et al. Relationship between the modified Rankin Scale and the Barthel Index in the process of functional recovery after stroke. *NeuroRehabilitation.* (2012) 30:315–22. doi: 10.3233/NRE-2012-0761
 42. Musa KI, Keegan TJ. The change of Barthel Index scores from the time of discharge until 3-month post-discharge among acute stroke patients in Malaysia: a random intercept model. *PLoS ONE.* (2018) 13:e0208594. doi: 10.1371/journal.pone.0208594
 43. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA.* (1993) 269:2386–91. doi: 10.1001/jama.1993.03500180078038
 44. Lee M, Oh MS, Jung S, Lee JH, Kim CH, Jang MU, et al. Differential effects of body mass index on domain-specific cognitive outcomes after stroke. *Sci Rep.* (2021) 11:14168. doi: 10.1038/s41598-021-93714-7
 45. Casanova R, Saldana S, Lutz MW, Plassman BL, Kuchibhatla M, Hayden KM. Investigating predictors of cognitive decline using machine learning. *J Gerontol B Psychol Sci Soc Sci.* (2020) 75:733–42. doi: 10.1093/geronb/gby054
 46. Aschwanden D, Aichele S, Ghisletta P, Terracciano A, Kliegel M, Sutin AR, et al. Predicting cognitive impairment and dementia: a machine learning approach. *J Alzheimers Dis.* (2020) 75:717–28. doi: 10.3233/JAD-190967
 47. Shi Y, Yang D, Zeng Y, Wu W. Risk factors for post-stroke depression: a meta-analysis. *Front Aging Neurosci.* (2017) 9:218. doi: 10.3389/fnagi.2017.00218
 48. Xu Q, Anderson D, Lurie-Beck J. The relationship between abdominal obesity and depression in the general population: a systematic review and meta-analysis. *Obes Res Clin Pr.* (2011) 5:267–360. doi: 10.1016/j.orcp.2011.04.007
 49. Poynter B, Shuman M, Diaz-Granados N, Kapral M, Grace SL, Stewart DE. Sex differences in the prevalence of post-stroke depression: a systematic review. *Psychosomatics.* (2009) 50:563–9. doi: 10.1016/S0033-3182(09)70857-6
 50. Hama S, Yoshimura K, Yanagawa A, Shimonaga K, Furui A, Soh Z, et al. Relationships between motor and cognitive functions and subsequent post-stroke mood disorders revealed by machine learning analysis. *Sci Rep.* (2020) 10:19571. doi: 10.1038/s41598-020-76429-z
 51. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res.* (2010) 11:2079–107. Available online at: <https://ueaeprints.uea.ac.uk/id/eprint/3640>
 52. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform.* (2014) 6:10. doi: 10.1186/1758-2946-6-10
 53. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv.* (2021) 54:1–35. doi: 10.1145/3457607
 54. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *Npj Digit Med.* (2018) 1:18. doi: 10.1038/s41746-018-0029-1
 55. Wang A, Pednekar N, Lehrner R, Todo A, Sahni R, Marks S, et al. DRAGON score predicts functional outcomes in acute ischemic stroke patients receiving both intravenous tissue plasminogen activator and endovascular therapy. *Surg Neurol Int.* (2017) 8:149. doi: 10.4103/2152-7806.210993

56. Saposnik G. An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score. *Neurology*. (2012) 79:2293–4. doi: 10.1212/WNL.0b013e31827a3c0a
57. Chander RJ, Lam BYK, Lin X, Ng AYT, Wong APL, Mok VCT, et al. Development and validation of a risk score (CHANGE) for cognitive impairment after ischemic stroke. *Sci Rep*. (2017) 7:12441. doi: 10.1038/s41598-017-12755-z
58. Kandiah N, Chander RJ, Lin X, Ng A, Poh YY, Cheong CY, et al. Cognitive Impairment after Mild Stroke: Development and Validation of the SIGNAL2 Risk Score. *J Alzheimers Dis*. (2016) 49:1169–77. doi: 10.3233/JAD-150736
59. Saposnik G, Kapral MK, Liu Y, Hall R, O'Donnell M, Raptis S, et al. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation*. (2011) 123:739–49. doi: 10.1161/CIRCULATIONAHA.110.983353
60. O'Donnell MJ, Fang J, D'Uva C, Saposnik G, Gould L, McGrath E, et al. The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke. *Arch Intern Med*. (2012) 172:1548–56. doi: 10.1001/2013.jamainternmed.30
61. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *Npj Digit Med*. (2020) 3:17. doi: 10.1038/s41746-020-0221-y
62. das Nair R, Cogger H, Worthington E, Lincoln NB. Cognitive rehabilitation for memory deficits after stroke. *Cochrane Database Syst Rev*. (2016) 9:CD002293. doi: 10.1002/14651858.CD002293.pub3
63. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. *PLoS ONE*. (2014) 9:e87987. doi: 10.1371/journal.pone.0087987

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Publikationen:

- 1) Fast L, Temuulen U, Villringer K, Kufner A, Ali HF, Siebert E, Huo S, Piper SK, Sperber PS, Liman T, Endres M, Ritter K. Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke. *Front. Neurol.* 2023. 14:1114360. doi: 10.3389/fneur.2023.1114360. Impact Factor: 4,086

- 2) Ali HF, Fast L, Khalil A, Siebert E, Liman T, Endres M, Villringer K, Kufner A. White matter hyperintensities are an independent predictor of cognitive decline 3 years following first-ever stroke-results from the PROSCIS-B study. *J Neurol.* 2022 Dec 6. doi: 10.1007/s00415-022-11481-5. Epub ahead of print. PMID: 36471099. Impact Factor: 6,682

- 3) Pucher E, Klemm F, Gadjanski M, Fast L. LNG Masterplan for Rhine-Main-Danube - Supply chain analysis & assessment of options for Danube region. Bericht für European Commission / DG MOVE / TEN-T; Berichts-Nr. D 1.2.2 V 1.3, 2015; 208 S.

Abschlussarbeiten:

- 1) Fast L. Urania: Kommunikationskampagne für die Urania Wien. Universität für angewandte Kunst Wien. 2015.

Danksagung

An erster Stelle danke ich meiner Doktormutter, Frau Prof. Dr. rer. nat. Kerstin Ritter, für die Möglichkeit diese Doktorarbeit durchzuführen. Ihre freundliche Unterstützung, wissenschaftliche Anleitung und vielfältigen Ideen haben dieses Projekt überhaupt erst ermöglicht. Ebenso danke ich meiner Zweitbetreuerin, Frau Dr. med. Kersten Villringer, für ihre wertvollen wissenschaftlichen Beiträge und die organisatorische Unterstützung.

Herrn Uchralt Temuulen danke ich für die unermüdliche Zusammenarbeit an diesem Forschungsprojekt und die daraus entstandene Freundschaft. Außerdem möchte mich bei Frau Anna Kufner, MD PhD für ihren Rückhalt und Optimismus bedanken. Herrn Dr. med. Thomas Liman und dem gesamten Team der PROSCIS Forschungsgruppe danke ich für die Bereitstellung der Daten und ihr wissenschaftliches Feedback. Ebenso danke ich dem Team des Centrums für Schlaganfallforschung Berlin für die Bereitstellung der Server-Infrastruktur und die Geduld mit diesem Projekt. Ein herzliches Dankeschön geht auch an alle Mitglieder der AG Ritter für unser kollegiales Miteinander.

In persönlicher Hinsicht bin ich meinem Partner, Konrad Froböse, aus tiefstem Herzen dankbar für seine liebevolle und selbstlose Unterstützung während des Medizinstudiums und der Doktorarbeit. Meiner Freundin Samipa Pudasaini danke ich für die vielen guten Ratschläge und den emotionalen Rückhalt auch in schwierigen Zeiten. Meinen Freundinnen Gina Varnavides und Laura Bragagna bin ich dankbar für ihre jahrelange treue Freundschaft und die viele Gespräche aus der Ferne. Darüber hinaus danke ich all den anderen Freund*innen und Bekannten, die mir zur Seite standen.

Zum Abschluss danke ich meiner Familie, insbesondere meinen Eltern, ohne deren Unterstützung diese Doktorarbeit nicht möglich gewesen wäre.