

**Integration von klinischen Daten für die UNet-basierte
Prädiktion der Infarktausbreitung nach ischämischem
Schlaganfall**

Dissertation

**zur Erlangung des akademischen Grades eines
Doktors der Medizin (Dr. med.)
an der
Medizinischen Fakultät der Universität Hamburg**

**vorgelegt von
Valentin Honus
aus
Düsseldorf**

2024

Betreuer/Gutachter der Dissertation: Prof. Dr. rer. nat. René Werner

Vorsitz der Prüfungskommission: PD Dr. habil. Thomas Lindner

Mitglied der Prüfungskommission: PD Dr. med. Bastian Cheng

Mitglied der Prüfungskommission: PD Dr. med. Susanne Gellißen

Datum der mündlichen Prüfung: 28.02.2025

Inhaltsverzeichnis

1 Arbeitshypothese und Fragestellung	1
1.1 Aufbau der Arbeit	1
2 Einleitung.....	2
2.1 Medizinischer Hintergrund.....	3
2.1.1 Grundlagen	3
2.1.1.1 Pathogenese.....	4
2.1.1.2 Pathophysiologie	6
2.1.1.3 Infarktläsion	7
2.1.1.4 Zeitliche Einteilung.....	8
2.1.2 Bildgebende Diagnostik	9
2.1.2.1 Darstellung des Infarktkerns	10
2.1.2.2 Darstellung der Penumbra	12
2.1.2.3 Darstellung der Gefäße	15
2.1.3 Therapiemöglichkeiten	15
2.1.3.1 Unklares Zeitfenster.....	17
2.1.4 Therapieevaluation	18
2.1.4.1 Klinischer Outcome.....	18
2.1.4.2 Bildbasierte Scores	19
2.1.4.3 Infarktausbreitung	20
2.1.4.3.1 Einfluss auf klinischen Outcome.....	22
2.1.4.3.2 Zeitlicher Verlauf.....	23
2.1.4.3.3 Klinische Einflussfaktoren.....	24
2.2 Methodischer Hintergrund	25
2.2.1 Deep-Learning	26
2.2.1.1 Netzwerktypen	27
2.2.1.1.1 Perzeptron	27
2.2.1.1.2 Multilayer-Perzeptron	29
2.2.1.1.3 Convolutional-Neural-Networks	31
2.2.1.1.3.1 nnUNet-Methode	34
2.2.1.2 Lernprozess	35
2.2.1.2.1 Backpropagation.....	36
2.2.1.2.2 Verlustfunktionen.....	36
2.2.1.2.3 Optimierung der Hyperparameter und Evaluation	38

2.2.2 Deep-Learning-basierte Prädiktion der Infarktausbreitung.....	39
2.2.2.1 Bisherige Studien.....	39
2.2.2.2 Integration klinischer Informationen	45
3 Material und Methoden	52
3.1 Verwendete Daten	52
3.1.1 WAKE-UP-Studie.....	52
3.1.2 Vorverarbeitung der Bilddaten	53
3.1.2.1 Nachsegmentierung der Infarktareale.....	55
3.1.3 Finaler Datensatz	58
3.2 Verwendete Methoden und Modelle.....	60
3.2.1 Prädiktion der räumlichen Infarktausbreitung.....	60
3.2.2 Training und Evaluation	61
3.2.3 Prädiktionsmodelle.....	62
3.2.3.1 nnUNet-Methode	62
3.2.3.2 Baseline-UNet.....	62
3.2.3.3 Multitask-UNet: Integration klinischer Informationen.....	63
4 Ergebnisse	66
4.1 Identifikation von Prädiktoren anhand linearer Regression	66
4.2 Prädiktionsergebnisse.....	66
4.2.1 nnUNet-Modelle	67
4.2.2 Baseline-UNet.....	67
4.2.3 Multitask-UNet: Integration klinischer Informationen.....	68
4.3 Vergleich der Modelle	70
4.3.1 Dice-Scores	70
4.3.2 Mittlerer absoluter Fehler	74
5 Diskussion	77
6 Zusammenfassung.....	84
7 Anhang.....	86
8 Literaturverzeichnis	88
9 Abkürzungsverzeichnis	105
10 Abbildungsverzeichnis.....	106
11 Tabellenverzeichnis.....	107
12 Erklärung des Eigenanteils.....	108
13 Eidesstattliche Versicherung	109

14 Danksagung	110
---------------------	-----

1 Arbeitshypothese und Fragestellung

In der vorliegenden Arbeit soll untersucht werden, ob sich die Deep-Learning-basierte Prädiktion der Infarktausbreitung mittels UNet-Modellen durch die Integration klinischer Informationen verbessern lässt und auf welche Weise sich die klinischen Informationen am besten in UNet-Modelle integrieren lassen. In einem ersten Schritt soll dafür die Infarktausbreitung im Verlauf ausschließlich mit Hilfe von initialen Bilddaten prädiziert werden. Anschließend soll untersucht werden, ob sich die Ergebnisse der Prädiktionen durch die zusätzliche Integration klinischer Informationen verbessern lassen. Dabei werden die Prädiktionen der selbstgestellten Modelle mit den Prädiktionsergebnissen der nnUNet-Methode, einer selbstkonfigurierenden, Deep-Learning-basierten State-of-the-Art-Segmentierungsmethode für medizinische Bilddaten, in einen Kontext gesetzt (Isensee et al. 2021). Die verwendeten Daten entstammen dabei der WAKE-UP-Studie (Thomalla et al. 2018).

1.1 Aufbau der Arbeit

In Kapitel 2.1 werden zunächst die medizinischen Grundlagen des ischämischen Schlaganfalls erarbeitet. Ein Schwerpunkt besteht darin das Infarktwachstum als Outcome zu definieren und Zusammenhänge zu anderen klinischen Informationen zu beschreiben (Kapitel 2.1.4.3). Der methodische Hintergrund wird anschließend in Kapitel 2.2 erläutert. Dazu werden die Grundzüge des Deep-Learnings beschrieben (Kapitel 2.2.1). Anschließend werden das Konzept der Deep-Learning-basierten Prädiktion der Infarktausbreitung beschrieben und bisherige Ansätze vorgestellt (Kapitel 2.2.2). Im Zentrum steht hierbei die Beschreibung, welche klinischen Informationen bisher in Deep-Learning-basierte Modelle integriert wurden und auf welche Weise diese Integration gelang (Kapitel 2.2.2.2).

In Kapitel 3 werden der verwendete Datensatz sowie die Methoden und Modelle beschrieben, anhand derer die Infarktausbreitung in dieser Studie prädiziert wurde. In Kapitel 4 werden anschließend die Prädiktionsergebnisse vorgestellt. Dabei liegt ein besonderes Augenmerk auf der Integration der klinischen Informationen (Kapitel 4.2.3). Abschließend werden die Erkenntnisse dieser Arbeit in Kapitel 5 diskutiert und in den wissenschaftlichen Kontext gesetzt.

2 Einleitung

Das Krankheitsbild des Schlaganfalls zählt zu den zerebrovaskulären Erkrankungen. Klinisch zeichnet sich ein Schlaganfall durch eine akut auftretende Schädigung des Hirnparenchyms aus, die mit einer Vielzahl an neurologischen Defiziten einhergehen kann. Ursächlich lassen sich zwei verschiedene Formen des Schlaganfalls unterscheiden. Während der ischämische Schlaganfall durch einen Gefäßverschluss bedingt wird, liegt beim hämorrhagischen Schlaganfall kausal eine Hirnblutung vor. Dabei sind ca. 80% aller Schlaganfälle ischämisch bedingt (Robert Koch-Institut 2015).

Der Schlaganfall ist in Deutschland und weltweit die zweithäufigste Todesursache (Robert Koch-Institut 2015; GBD-2015-Mortality-and-Causes-of-Death-Collaborators 2016). Obwohl die Inzidenz und die Mortalität in den letzten Jahrzehnten, bedingt durch präventive Maßnahmen und verbesserte Diagnostik- und Therapiekonzepte, in westlichen, hochentwickelten Ländern stetig gesunken ist (Feigin et al. 2010; Robert Koch-Institut 2015), steigen die absoluten Krankheitsfälle aufgrund des demographischen Wandels in Deutschland weiter an (Robert Koch-Institut 2015). Innerhalb des ersten Jahres nach stattgehabtem Schlaganfall versterben ein Drittel bis ein Viertel aller Erkrankten (Kolominsky-Rabas et al. 1998; Palm et al. 2010; Robert Koch-Institut 2015). Bei Überlebenden verbleiben häufig persistente Einschränkungen im alltäglichen Leben, die oftmals zu einer Pflegebedürftigkeit führen (Feigin et al. 2010; Luengo-Fernandez et al. 2013; Robert Koch-Institut 2015; Crichton et al. 2016). Somit stellt das Krankheitsbild des Schlaganfalls weiterhin eine große Belastung für die betroffene Person und damit einhergehend für das deutsche Gesundheitssystem dar (van den Bussche et al. 2010; Günster 2011).

Mit den neuen Errungenschaften im Bereich des Deep-Learning einhergehend, entwickelten sich verschiedene Konzepte, diese Technologie im Kontext des Krankheitsbildes des ischämischen Schlaganfalls zu nutzen. So können Deep-Learning-basierte Modelle zur Diagnose (Abedi et al. 2017), zur automatisierten Bestimmung des Alberta-Stroke-Program-Early-CT-Scores (ASPECTS) (Cheng et al. 2021), zur Detektion von Gefäßokklusionen und -blutungen (McLouth et al. 2021), zur Bestimmung der Infarktdauer (Ho et al. 2018), zur Segmentierung des Infarktkerns (Chen et al. 2017a) und zur Prädiktion des klinischen Outcomes (Heo

et al. 2019) sowie der Infarktausbreitung verwendet werden. Im Hinblick auf die Prädiktion der Ausbreitung eines Infarktes konnte eine Überlegenheit Deep-Learning-basierter Verfahren gegenüber anderen Machine-Learning-Verfahren, wie zum Beispiel Random-Forests, aufgezeigt werden (Winzeck et al. 2018). Die Prädiktion der Infarktausbreitung könnte in Zukunft dafür genutzt werden, den Nutzen eines Therapieverfahrens zu simulieren und Zusammenhänge zwischen der Infarktausbreitung und weiteren klinischen Informationen zu untersuchen. Dennoch bleiben viele Fragen bisher unbeantwortet. Deep-Learning-basierte Modelle werden aufgrund des Mangels an frei zugänglichen und großen Bilddatensammlungen anhand der Daten von einzelnen Studien trainiert und evaluiert. Dies führt dazu, dass sich die verschiedenen Ansätze schwer miteinander vergleichen lassen. Eine über die Versuchsdaten hinausgehende Generalisation wird zudem schon durch die Ein- und Ausschlusskriterien (klinische Symptomatik, Untersuchungszeitpunkte, verfügbare Bilddaten, klinische Schwere des Schlaganfalls) der einzelnen Studien erschwert.

Obwohl sich der ischämische und der hämorrhagische Schlaganfall in ihrer Klinik ähnlich präsentieren können, unterscheiden sie sich grundlegend in ihrer Pathophysiologie sowie dem therapeutischen Vorgehen. Es ist daher anzumerken, dass sich die vorliegende Arbeit vorwiegend auf das Krankheitsbild des ischämisch bedingten Schlaganfalls fokussiert.

2.1 Medizinischer Hintergrund

2.1.1 Grundlagen

Das Gehirn (Cerebrum) bezeichnet die Anteile des Zentralen Nervensystems (ZNS), die sich von den Hirnhäuten (Meningen) umhüllt im Inneren der Schädelhöhle befinden. Dabei lässt sich das Gehirn in vier Bereiche unterteilen: Großhirn (Telencephalon), Zwischenhirn (Diencephalon), Kleinhirn (Cerebellum) und Hirnstamm (Truncus cerebri). Wie das gesamte ZNS besteht auch das Gehirn aus der grauen (Substantia grisea) und weißen Substanz (Substantia alba). Dabei besteht die graue Substanz vorwiegend aus Nervenzellkörpern (Perikaryen) und die weiße Substanz vorwiegend aus Nervenfortsätzen (Axonen). Im Rückenmark (Medulla spinalis) wird die graue Substanz von der weißen Substanz umgeben. Die Rinde (Kortex) von Groß- und Kleinhirn besteht aus grauer Substanz und

umgibt somit das Marklager, welches aus weißer Substanz besteht. Auf der Großhirnrinde lassen sich Rindfelder abgrenzen, denen verschiedenste Funktionen zugeordnet werden können. Jede Hemisphäre des Großhirns lässt sich dabei in 5 Lappen aufteilen: Frontallappen, Parietallappen, Okzipitallappen, Temporallappen, Insellappen (Insula). Das Gehirn reguliert lebensnotwendige vegetative Prozesse und steuert höhere kognitive Funktionen. Eine Schädigung des Gehirns kann in Abhängigkeit von Ursache und Lokalisation in zahlreichen fokalen und diffusen Funktionsstörungen resultieren.

Das Gehirn wird über den Circulus arteriosus cerebri mit Blut versorgt. Dieser Kollateralkreislauf bildet sich aus Verbindungsarterien und den paarig angelegten Zuflüssen der Arteria carotis interna (vorderer Kreislauf) und der Arteria vertebralis (Hinterer Kreislauf). Die Arteria carotis interna teilt sich auf jeder Hemisphäre in die Arteria cerebri anterior und media auf. Durch diese werden der Frontal-, Parietal- und der größte Teil des Temporallappens sowie Anteile des Zwischenhirns versorgt. Die paarige Arteria vertebralis vereinigt sich zu der unpaarigen Arteria basilaris. Das Kleinhirn und der Hirnstamm werden aus direkten Abgängen der paarigen Arteria vertebralis und Arteria basilaris versorgt. Die Arteria basilaris teilt sich wiederum in die paarige Arteria cerebri posterior auf, die den Okzipital- und Teile des Temporallappens sowie Anteile des Zwischenhirns versorgt. Die Arteria cerebri anterior, media und posterior stellen die drei großen zerebralen Hauptarterien dar, die zusammen die gesamte Großhirnrinde versorgen. Die Arteria cerebri media versorgt das größte Gebiet und kann dabei in 4 Abschnitte (M1-4) von medial nach lateral unterteilt werden.

2.1.1.1 Pathogenese

Das Krankheitsbild des ischämischen Schlaganfalls stellt sich in vielen Teilaspekten sehr heterogen dar. Ursächlich lassen sich Makro- und Mikroangiopathien unterscheiden, die eine typische Infarktmorphologie aufweisen (siehe Abbildung 1).

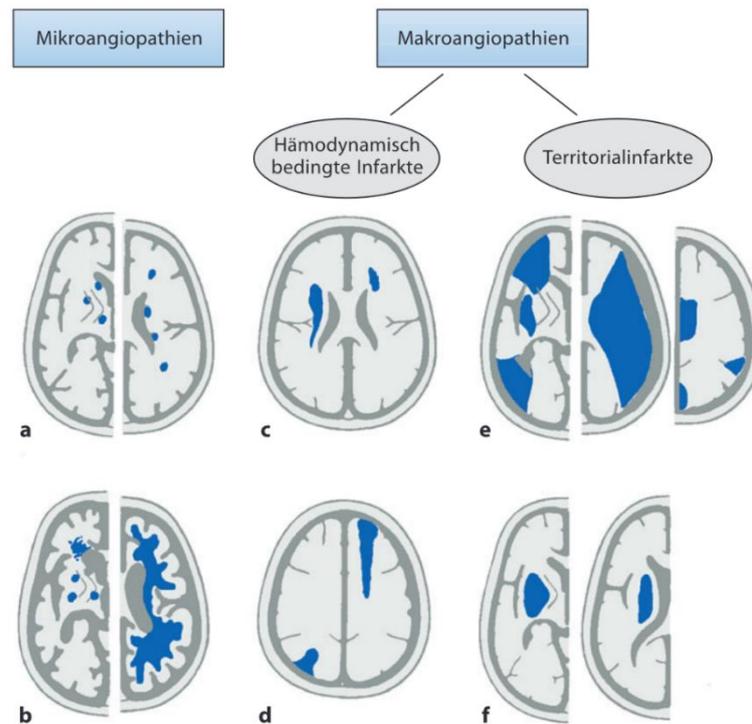


Abbildung 1: Infarkte unterschiedlicher Pathogenese (Klötzsch & Röther 2020)

a: Multiple lakunäre Infarkte, **b:** subkortikale arteriosklerotische Enzephalopathie, **c:** Endstrominfarkt, **d:** vorderer und hinterer Grenzzoneninfarkt, **e:** Territorialinfarkte des vorderen, mittleren und hinteren Mediagebietes (links), des gesamten Mediagebietes (mittig) und des Anterior- und Posteriorgebietes (rechts), **f:** Linsenkerninfarkt (Klötzsch & Röther 2020).

Von einer Makroangiopathie spricht man, falls sich die Gefäßpathologie im Bereich einer der großen zerebralen Hauptarterien oder im Bereich einer mittelgroßen Arterie befindet. Ätiologisch werden hierbei Territorialinfarkte und hämodynamische Infarkte unterschieden. Territorialinfarkte werden dabei durch eine embolische Okklusion, zum Beispiel in Folge eines bestehenden Vorhofflimmerns, ausgelöst. Ein Territorialinfarkt stellt sich oftmals als großes, keilförmiges Infarktareal mit kortikomedullärer Ausdehnung dar. Häufig kann ein embolischer Verschluss durch die klinische Symptomatik einem Stromgebiet zugeordnet werden. Hämodynamische Infarkte entstehen im Rahmen eines Perfusionsabfalls, zum Beispiel ausgelöst durch einen Blutdruckabfall oder durch eine Stenose eines zuführenden Gefäßes. Dabei unterscheidet man morphologisch Endstrominfarkte und Grenzzoneninfarkte. Endstrominfarkte bezeichnen hierbei einen subkortikal gelegenen Infarkt im Bereich des Endstromgebietes einer Arterie. Grenzzoneninfarkte andererseits entstehen an der Grenze von zwei arteriellen Stromgebieten und sind häufig frontoparietal und parietookzipital lokalisiert.

Eine Mikroangiopathie hingegen zeichnet sich durch arteriosklerotische Veränderungen im Bereich von Arteriolen oder Kapillaren aus, die oftmals in Folge eines langjährigen Hypertonus entstehen. Typisch hierbei ist die Morphologie des lakunären Infarktes. Dieser beschreibt einzelne oder multiple kleinere subkortikale Infarkte, die typischerweise im Bereich des Marklagers, der Stammganglien, dem Thalamus oder dem Hirnstamm lokalisiert sind.

2.1.1.2 Pathophysiologie

Infolge der Ischämie unterliegen die Bereiche der Infarktläsion einem pathophysiologischen Wandel. In einer ischämischen Zelle kommt es in Folge der Minderperfusion zu einem Mangel an Adenosintriphosphat (ATP) und daraus resultierend zu einem Ausfall ATP-abhängiger, primär aktiver Ionenporter. Infolgedessen überwiegen sekundär aktive und passive Ionenporter und es kommt zu einem Anstieg der intrazellulären Natriumkonzentration, der, bedingt durch einen reflektiven Chlorid- und Wassereinstrom, zu einem Anschwellen der Zelle führt. Vor allem im Bereich des irreversibel geschädigten Infarktkerns kommt es innerhalb der ersten 24 Stunden so zur Ausbildung eines sogenannten zytotoxischen Ödems. Die Bezeichnung Ödem ist dabei als historisch zu betrachten, da es zwar zu einer Verschiebung des Volumens aus dem Extrazellularraum in den Intrazellularraum kommt, das Hirnvolumen jedoch nicht ansteigt (Kahle et al. 2009).

Infolge der Ausbildung eines zytotoxischen Ödems kommt es innerhalb des Extrazellularraums zu einem Verlust von Wasser und osmotisch aktiven Substanzen, welcher wiederum gesteigerte transendotheliale Gradienten bedingt. Kommt es nun während der postischämischen Reperfusion zu einem Anstieg der Permeabilität der endothelialen Blut-Hirn-Schranke, führt der Einstrom von Wasser und Ionen in den Extrazellularraum zu einem Anstieg des Hirnvolumens und zur Ausbildung eines Hirnödems. Es lassen sich zwei verschiedene Mechanismen unterscheiden, die beide zu einer gesteigerten transendothelialen Permeabilität führen können (Kahle et al. 2009).

Kommt es infolge einer innerhalb des Endothels erhöhten Expression von Ionenkanälen, Ionenportern und Aquaporinen zu einem Anstieg der transendothelialen Permeabilität, bezeichnet man das resultierende Ödem als ionisches Ödem. Resultiert der Anstieg der Permeabilität hingegen aus einer

Auflockerung der Tight-Junctions zwischen den Endothelzellen, spricht man von der Ausbildung eines vasogenen Ödems, welches zusätzlich mit einem Ausstrom von Plasmaproteinen einhergeht (Kahle et al. 2009).

Während das ionische Ödem unmittelbar mit der Entstehung des zytotoxischen Ödems einhergeht, bildet sich das vasogene Ödem erst nach einigen Stunden aus (Stokum et al. 2016). Im Rahmen des Schlaganfalls kommt es so oftmals zu der Ausbildung eines biphasischen Ödems. Während in der Frühphase das zytotoxische und ionische Ödem dominieren, erreicht die Ausbreitung des vasogenen Ödems ihren Höhepunkt nach zwei bis vier Tagen (Simard et al. 2007; Thrane et al. 2014). Im weiteren Verlauf kommt es zu einer Abnahme der ausgebildeten Ödeme und infolge inflammatorischer Prozesse zu einer Vernarbung des irreversibel ischämisch geschädigten Gewebes (Burda & Sofroniew 2014).

2.1.1.3 Infarktläsion

Innerhalb des betroffenen Hirnparenchyms lassen sich zu einem definierten Zeitpunkt verschiedene Bereiche innerhalb der Infarktläsion voneinander unterscheiden. Der Infarktkern definiert die bereits irreversibel geschädigten Bereiche des ischämischen Hirnparenchyms. In Abgrenzung dazu beschreibt die Penumbra die ischämischen und dysfunktionalen Bereiche, die den Infarktkern umgeben und deren Funktionalität im Rahmen einer zeitigen Reperfusion reversibel wiederhergestellt werden kann. Bleibt eine solche Reperfusion jedoch aus, wird sich der irreversibel geschädigte Infarktkern mit voranschreitender Zeit immer weiter auf die Bereiche der Penumbra ausweiten. Die Penumbra ist ihrerseits von einem benignen Ödem umgeben. Das benigne Ödem definiert die Bereiche, die zwar hypoperfundiert sind, deren Funktionalität jedoch erhalten bleibt. Auch ohne Reperfusion wird sich der Infarktkern nicht auf die Bereiche des benignen Ödems ausweiten, da die Versorgung dieses Bereiches durch angrenzende Kollateralgefäße ausreichend gesichert wird (Goyal et al. 2013). In Abbildung 2 sind die unterschiedlichen Bereiche der Infarktläsion schematisch dargestellt.

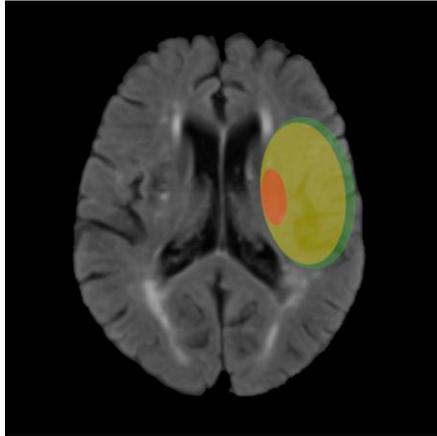


Abbildung 2: Bereiche der Infarktläsion

Die Farben markieren die verschiedenen Bereiche einer Infarktläsion. **rot**: irreversibel geschädigter Infarktkern, **gelb**: potentiell reversibel geschädigte Penumbra, **grün**: benignes Ödem.

2.1.1.4 Zeitliche Einteilung

Im klinischen Verlauf lassen sich verschiedene Phasen unterscheiden, die teils unterschiedlich definiert werden. In dieser Arbeit werden die Definitionen von Bernhardt et al. verwendet, die in Abbildung 3 visualisiert sind (Bernhardt et al. 2017). In der Hyperakutphase, die die ersten 24 Stunden nach stattgehabtem Schlaganfall umfasst, kommt es zu einem Absterben von Nerven- und Gliazellen. Folgend kommt es in der Akutphase, die sich bis zur Vollendung der ersten Woche erstreckt, zu inflammatorischen Prozessen und das geschädigte Hirnparenchym beginnt zu vernarben. An die Akutphase schließt sich die frühe Subakutphase an, die bis zum Ende des dritten Monats andauert. In dieser kommt es infolge der endogenen Plastizität des Hirnparenchyms zu einer klinischen Verbesserung der neurologischen Symptomatik. Dieser Prozess setzt sich mit abnehmender Geschwindigkeit in der späten Subakutphase, die sich bis zur Vollendung des sechsten Monats erstreckt, und in der anschließenden chronischen Phase fort (Bernhardt et al. 2017).

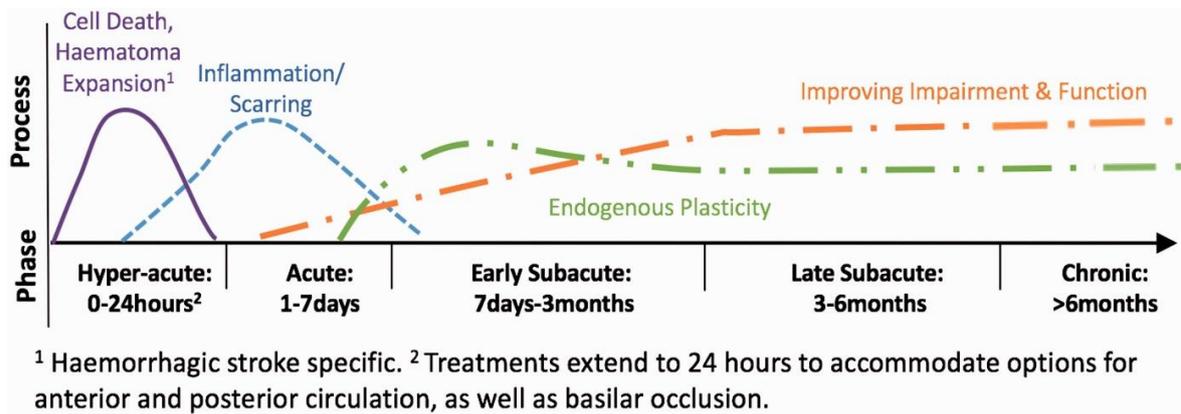


Abbildung 3: Zeitliche Einteilung des Schlaganfalls (Bernhardt et al. 2017)

2.1.2 Bildgebende Diagnostik

Die Computertomographie (CT) und die Magnetresonanztomographie (MRT) sind in der Lage, die Korrelate der beschriebenen pathophysiologischen Prozesse, die mit einem ischämischen Schlaganfall einhergehen, abzubilden und sind so elementarer Bestandteil der modernen Diagnostik. Die MRT stellt mit einer Vielzahl an variablen Parametern und unterschiedlichen Sequenzen das weitaus komplexere Verfahren dar und wird häufig im Rahmen von Studien verwendet. Die CT bietet den großen Vorteil einer kürzeren Aufnahmezeit, geht gleichzeitig jedoch auch mit einer Strahlenbelastung für die zu behandelnde Person und einem niedrigeren Weichteilkontrast einher. Es konnte gezeigt werden, dass die Verwendung der MRT Vorteile hinsichtlich der Darstellung des Infarktkerns in der Hyperakutphase eines ischämischen Schlaganfalls bietet (Fiebach et al. 2002). Der Nutzen beider Aufnahmetechniken im klinischen Kontext ist jedoch als nahezu gleichwertig anzusehen (Vymazal et al. 2012). Im akuten Setting hat sich aufgrund der Verfügbarkeit sowie der schnelleren Aquisitionszeit die CT als Methode der Wahl etabliert.

Da sich der ischämische und der hämorrhagische Schlaganfall anhand der klinischen Symptomatik nicht unterscheiden lassen, ist zunächst eine schnelle Unterscheidung unabdingbar. Ein hämorrhagischer Schlaganfall kann durch den Nachweis einer intrazerebralen Blutung in der Bildgebung diagnostiziert werden. Eine Blutung lässt sich in der nativen Computertomographie und innerhalb der MRT anhand der T2*-gewichteten Gradientenecho-Sequenz (T2*w-MRT) mit hoher Sensitivität abbilden (Kidwell et al. 2004).

Nach erfolgtem Blutungsausschluss kann die Lokalisation der Infarktläsion und die Abgrenzung des irreversibel geschädigten Infarktkerns von der potentiell reversibel geschädigten Penumbra erfolgen. Für diese Abgrenzung spielen die pathophysiologischen Prozesse, die innerhalb der Infarktläsion stattfinden, eine fundamentale Rolle.

2.1.2.1 Darstellung des Infarktkerns

Wie bereits in Kapitel 2.1.1.2 beschrieben kommt es vor allem in dem Bereich des ischämisch geschädigten Infarktkerns zur Ausbildung eines zytotoxischen Ödems, welches wiederum mit der zeitversetzten Ausbildung eines vasogenen Ödems einhergeht. Innerhalb der diffusionsgewichteten MRT-Bildgebung (Diffusion-Weighted-Imaging, DWI) können bereits die Prozesse, die mit der Ausbildung des zytotoxischen Ödems einhergehen, als hyperintens dargestellt werden (Schaefer et al. 1997). So konnte gezeigt werden, dass sich der Infarktkern bereits nach 11 Minuten im DWI darstellen lässt (Hjort et al. 2005). Zudem wird dieser im Vergleich zur nativen CT mit einer größeren Sensitivität und Genauigkeit abgebildet (Fiebach et al. 2002). Jedoch muss erwähnt werden, dass in der diffusionsgewichteten MRT-Bildgebung nicht ausschließlich der Infarktkern dargestellt wird, sondern auch Anteile der Penumbra miterfasst werden. So können initial hyperintense Bereiche im Verlauf eine Reversibilität infolge einer Reperfusion aufweisen (DWI Reversibility, DWIR) (Kidwell et al. 2000). Zudem kann sich auch ein durchscheinender T2-Effekt (T2-Shine-Through-Effekt) innerhalb der DWI als hyperintens darstellen. Durch die Berechnung des Apparent-Diffusion-Coefficient (ADC) kann dieser Effekt negiert und können selektiv Veränderungen der Diffusion als hypointens dargestellt werden (Birenbaum et al. 2011). Es konnte gezeigt werden, dass ein ADC-Schwellenwert von kleiner als $620 \times 10^{-6} \text{mm}^2/\text{s}$ den Infarktkern identifiziert (Purushotham et al. 2015). In der Akutphase nach ein bis zwei Wochen kommt es durch die zunehmende Zersetzung der Zellmembranen der ischämischen Zellen zu einer Abnahme des zytotoxischen Ödems. Diese Prozesse führen zu einer Normalisierung des ADC, während sich der Infarktkern innerhalb der DWI-Maps durch T2-Effekte, die mit der Ausbildung des vasogenen Ödems einhergehen, weiterhin als hyperintens darstellt. Mit voranschreitender Gliose stellt sich der

Infarktkern in ADC-Maps zunehmend hyperintens und in DWI-Maps hyper- bis isointens dar (Lansberg et al. 2001b).

Mit voranschreitender Zeit kann die Infarktläsion durch die Flüssigkeitszunahme im Extrazellularraum, die vor allem mit der Entstehung des vasogenen Ödems einhergeht, auch innerhalb der T2-gewichteten Sequenz (T2w-MRT) und der Fluid-Attenuated-Inversion-Recovery-Sequenz (FLAIR-MRT) als hyperintens dargestellt werden (Lansberg et al. 2001b). Es konnte gezeigt werden, dass die FLAIR-Sequenz den Infarktkern in einem Zeitfenster von 3 bis 6 Stunden nach Symptombeginn mit großer Sensitivität erfasst (Thomalla et al. 2009). Zwischen der ersten bis zur sechsten Woche kann es, bedingt durch die Einwanderung von inflammatorischen Zellen, zu einer Pseudonormalisierung der T2- und FLAIR-Sequenzen (Fogging) kommen (O'Brien et al. 2004). Auch nach abgeschlossener Vernarbung stellt sich der Infarktkern in der chronischen Phase weiterhin als hyperintens dar.

In Abbildung 4 wird die Darstellung des Infarktkerns anhand der zuvor beschriebenen MR-Sequenzen im zeitlichen Verlauf veranschaulicht.

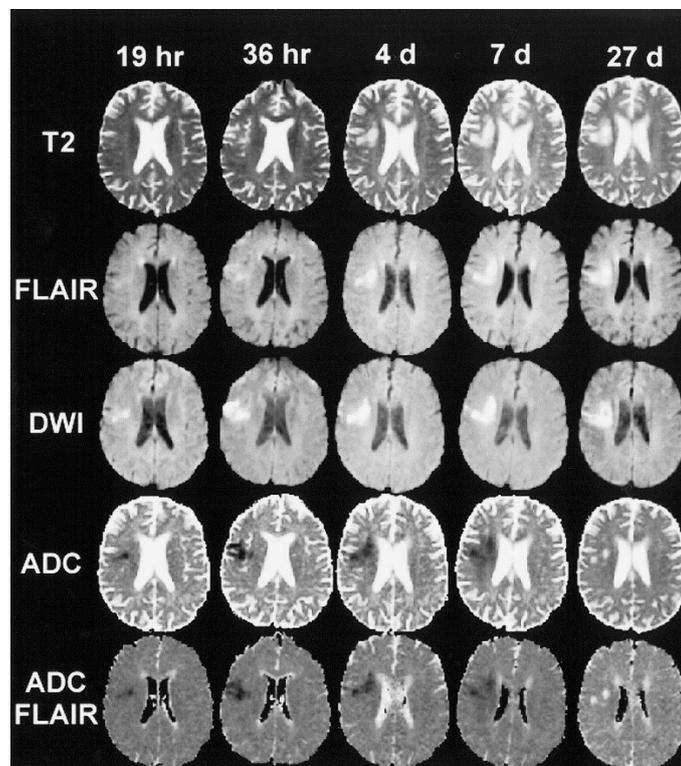


Abbildung 4: Darstellung des Infarktkerns im zeitlichen Verlauf (Lansberg et al. 2001b)

In der ADC-Map stellt sich der Infarktkern bis zum Tag 7 als hypointens dar und im Verlauf als hyperintens. In der DWI-Map stellt sich der Infarktkern über den gesamten Zeitraum als hyperintens dar. In der T2w- und FLAIR-Map wird der Infarktkern zunehmend besser abgebildet und stellt sich als hyperintens dar.

Die Flüssigkeitszunahme im Extrazellularraum, die mit der verzögerten Entstehung des vasogenen Ödems einhergeht, lässt sich mit fortschreitender Zeit auch innerhalb der nativen CT direkt abbilden und stellt sich innerhalb dieser als hypodens dar. In einigen Fällen lässt sich ein ischämischer Infarkt auch früher durch entsprechende indirekte Korrelate darstellen. Zu diesen Korrelaten gehören eine verminderte Abgrenzbarkeit von weißer und grauer Substanz, insbesondere in den Bereichen der Gyri, der Basalganglien, der Inselrinde sowie der Basalganglien, verstrichene Sulci, und das hyperdense Arterienzeichen (Muir et al. 2006). Dennoch besitzt die native CT eine sehr niedrige Sensitivität innerhalb der ersten drei Stunden (Chalela et al. 2007). Zwischen der zweiten und dritten Woche kann es in der CT zum Fogging des Infarktkerns kommen (Skriver & Olsen 1981).

2.1.2.2 Darstellung der Penumbra

Die Darstellung der Penumbra erfolgt mithilfe der CT-Perfusionsbildgebung (CTP) oder der MR-Perfusionsbildgebung (Perfusion-Weighted-Imaging, PWI). In beiden Verfahren erfolgt die Messung der Perfusion durch die venöse Kontrastmittel-Applikation und die anschließende Detektion der An- und Abflutung dieses Kontrastmittels im Hirnparenchym. So kann die Kontrastmittelkonzentration eines Voxels (Tissue-Concentration) über die Zeit gemessen werden. Sie ist das Ergebnis der Faltung von der Arterial-Input-Function (AIF) mit der Residuumsfunktion (Residue-Function). Die Arterial-Input-Function beschreibt dabei den Kontrastmitteleinstrom des zuführenden Gefäßes. Auch dieser kann über die Zeit gemessen werden und ist abhängig von der Bolus-Injektion, dem Herzzeitvolumen sowie der Gefäßanatomie der zu behandelnden Person. Die Residuumsfunktion hingegen beschreibt das gewebsspezifische Ausschwemmen eines idealisierten Kontrastmitteleinstroms und kann durch die Entfaltung der über die Zeit detektierten Arterial-Input-Function und Gewebekonzentration berechnet werden. Anhand dieser Funktionen lassen sich verschiedene Perfusionsparameter bestimmen (Copen et al. 2011).

Die Time-To-Peak (TTP) beschreibt den Zeitpunkt, an dem die größte Signalintensität des Kontrastmittels registriert wurde. Das Cerebral-Blood-Volume (CBV) errechnet sich durch das Integral der über die Zeit registrierten Gewebekonzentration. Der Cerebral-Blood-Flow (CBF) entspricht der maximalen Amplitude der entfalteten Residuumsfunktion und die Time-To-Maximum (Tmax)

entspricht dem Zeitpunkt, zu dem dieses Maximum erreicht wird. Die Mean-Transit-Time (MTT) beschreibt den Quotienten von CBV und CBF (Copen et al. 2011) (siehe Abbildung 5).

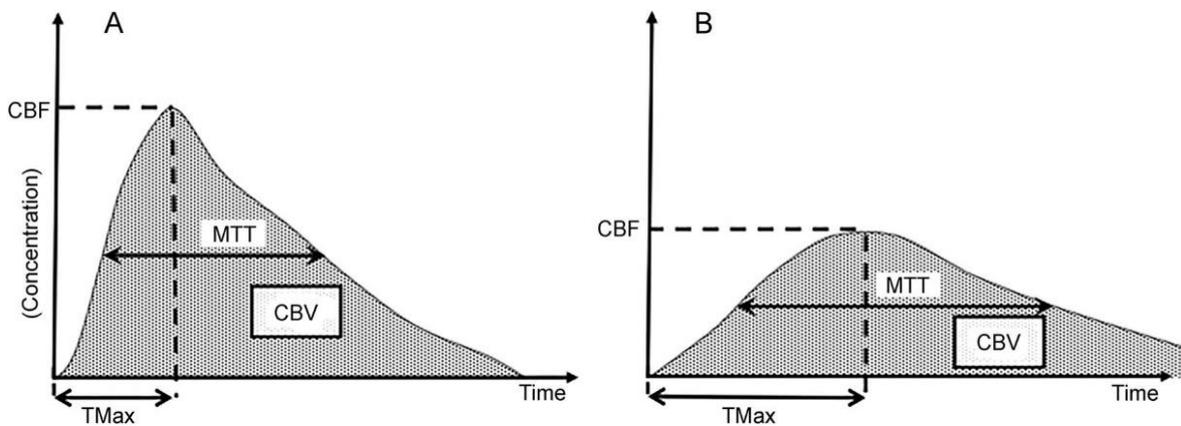


Abbildung 5: Diagramm der Perfusionsparameter (Carrera & Wintermark 2017)

Schematische Darstellung der Perfusionsparameter im gesunden (A) und im hypoperfundierten (B) Hirngewebe.

Infolge des Verschlusses eines zerebralen Gefäßes sinkt der Cerebral-Blood-Flow innerhalb des ischämischen Infarktkerns. Im Bereich der Penumbra kann ein starkes Absinken des Cerebral-Blood-Flows zunächst durch die Vasodilatation von Gefäßen und eine Rekrutierung alternativer Gefäßkollaterale verhindert werden. Diese autoregulatorischen Mechanismen gehen mit einer Erhöhung der Time-To-Maximum, der Time-To-Peak und der Mean-Transit-Time einher (Copen et al. 2011). Dabei konnte gezeigt werden, dass sich eine relevante Perfusionseinschränkung durch einen Tmax-Threshold von größer als 6 Sekunden beschreiben lässt (Olivot et al. 2009). Der Infarktkern wird in der CT-Perfusionsbildgebung gängiger Weise durch einen relativen Abfall des CBFs im Vergleich zu einem festgelegten Baseline-Niveau von mehr als 30 Prozent definiert (Campbell et al. 2011). Jedoch unterscheiden sich die optimalen Schwellenwerte für die verschiedenen Perfusionsparameter je nach Studie, was durch Unterschiede in der Bildakquirierung und der Datenverarbeitung zu erklären ist (Demeestere et al. 2020). In der MR-Bildgebung wird der Infarktkern, wie zuvor beschrieben, anhand von DWI- oder ADC-Maps bestimmt (siehe Abbildung 6).

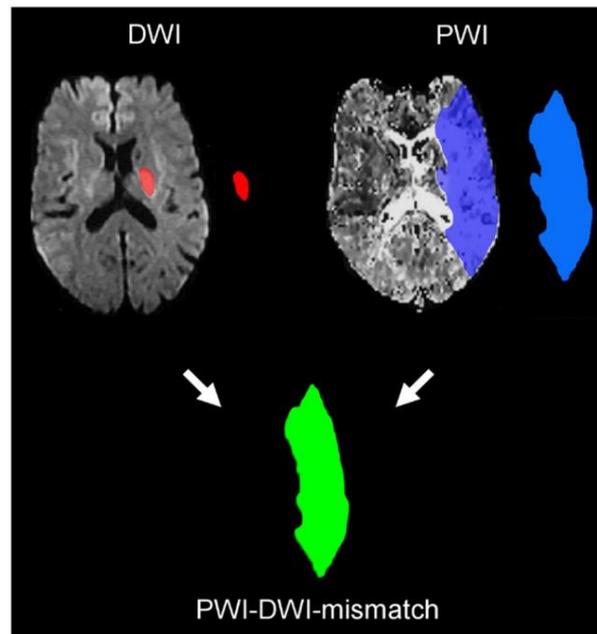


Abbildung 6: PWI-DWI-Mismatch (Rimmele & Thomalla 2014)

Die **rote Fläche** entspricht dem irreversibel geschädigtem Infarktkern, der bereits eine Diffusionsminderung aufweist. Das Areal, in dem eine Perfusionseinschränkung besteht, ist **blau** markiert. Die **grüne Fläche** repräsentiert die Penumbra. In dieser ist die Perfusion eingeschränkt, es liegt jedoch noch keine irreversible Schädigung des Gewebes vor.

Folgend lassen sich der Infarktkern und die Penumbra je nach Bildgebung durch verschiedene Mismatch-Konzepte voneinander abgrenzen. Im Rahmen von MRT-basierten Verfahren wird das sogenannte PWI-DWI-Mismatch zur Abgrenzung von Infarktkern und Penumbra verwendet. Im Rahmen der CTP werden zur Abgrenzung ausschließlich Perfusionsparameter verwendet (Demeestere et al. 2020). Es konnte gezeigt werden, dass mit beiden Ansätzen der Infarktkern und die Penumbra ähnlich gut umschrieben werden können (Campbell et al. 2012a). Jedoch besteht bis zum jetzigen Zeitpunkt noch kein wissenschaftlicher Konsens darüber, welche Parameter und Schwellenwerte sich am besten dazu eignen, um Infarktkern und Penumbra voneinander abzugrenzen, und wie ein günstiges Mismatch zu definieren ist. Zudem definiert das Mismatch-Konzept vielmehr die Wahrscheinlichkeit einer Infarzierung zum Zeitpunkt der Bildakquirierung, die abhängig ist von weiteren Faktoren, wie der Hypoperfusionsschwere und -dauer, und die sich mit fortschreitender Zeit ändert (Demeestere et al. 2020). In Tabelle 1 sind die verschiedenen bildgebenden Verfahren, mit denen die Bereiche der Infarktläsion dargestellt werden können, zusammenfassend dargestellt.

Tabelle 1: Darstellung der Infarktläsion anhand der verschiedenen bildgebenden Verfahren

Modalität		Infarktkern	Perfusionsstörung
natives CT		Niedrige Sensitivität in den ersten drei Stunden, ggf. indirekte Infarktfrühzeichen, ggf. Fogging zwischen 2.-3. Woche, im Verlauf hypodens	
MRT	DWI	Nach wenigen Minuten hyperdens, ggf. T2-Shine-Through, im Verlauf hyper- bis isodens	anteilig erfasst
	ADC	Nach wenigen Minuten hypodens, Schwellenwert: $<620 \times 10^{-6} \text{mm}^2/\text{s}$, nach 1-2 Wochen iso- bis hyperdens	anteilig erfasst
	FLAIR	Nach 3-6 Stunden hyperdens, ggf. Fogging zwischen 1.-6. Woche, im Verlauf hyperdens	
Perfusionsparameter		CBF↓↓, CBV↓↓, Tmax/MTT/TTP↑, Schwellenwert: rCBF<30%	CBF↓, CBV↔, Tmax/MTT/TTP↑, Schwellenwert: Tmax>6s

DWI: Diffusion-Weighted-Imaging, **ADC:** Apparent-Diffusion-Coefficient, **FLAIR:** Fluid-Attenuated-Inversion-Recovery, **CBF:** Cerebral-Blood-Flow, **CBV:** Cerebral-Blood-Volume, **Tmax:** Time-To-Maximum, **MTT:** Mean-Transit-Time, **TTP:** Time-To-Peak

2.1.2.3 Darstellung der Gefäße

Zur genaueren Darstellung der Gefäße werden die CT-Angiographie (CTA) und die MR-Angiographie (MRA) eingesetzt. Die MR-Angiographie kann Kontrastmittel-gestützt (CE-MRA) oder mit der Time-Of-Flight-Sequenz (TOF-Sequenz) erfolgen. Anhand einer Angiographie können therapierrelevante Informationen erhoben werden. So kann zum einen die genaue Lokalisation der Okklusion dargestellt werden und zum anderen der Grad der Versorgung durch Gefäßkollaterale quantifiziert werden. Im Vorfeld einer endovaskulären Thrombektomie müssen zudem anatomische Besonderheiten sowie mögliche Stenosen der supraaortalen Gefäße eruiert werden (Muir et al. 2006).

2.1.3 Therapiemöglichkeiten

Da bei einer unbehandelten Person mit ischämischem Schlaganfall jede Minute durchschnittlich 1.9 Millionen Neuronen irreversibel geschädigt werden, bedarf es neben einer schnell durchzuführenden Diagnostik auch der raschen Initiierung einer adäquaten Therapie (Saver 2006). Neben den allgemeinen Maßnahmen des Schlaganfallmanagements nehmen verschiedene Therapieansätze eine zentrale Rolle bei der Therapie eines ischämischen Schlaganfalls ein. Zu den am häufigsten angewendeten Therapiemöglichkeiten zählen die intravenöse Thrombolyse sowie die mechanische Thrombektomie.

Im Rahmen der intravenösen Thrombolyse (IVT) wird der zu behandelnden Person Alteplase (Recombinant-Tissue-Type-Plasminogen-Activator, rtPA) verabreicht. Alteplase wirkt durch die Umwandlung von an Fibrin gebundenem Plasminogen zu Plasmin fibrinolytisch auf die Blutgerinnung ein. Es konnte gezeigt werden, dass eine solche Therapie den klinischen Outcome signifikant verbessert, sofern sie innerhalb eines Zeitfensters von 4,5 Stunden nach Einsetzen der neurologischen Symptomatik (Time-Since-Stroke, TSS) eingeleitet wird (NINDS-Study-Group 1995; Hacke et al. 2008; Wahlgren et al. 2008). Außerdem kann eine Therapie in einem erweiterten Zeitfenster bei einer TSS zwischen 4,5 und 9 Stunden in Abhängigkeit von einer Mismatch-Bildgebung erwogen werden (Campbell et al. 2019). Zudem korreliert der erzielte Outcome mit der Zeit, die zwischen dem Einsetzen der Symptome und der Therapie (Time-To-Treatment, TTT) verstreicht, und es zeigte sich, dass die intrazerebrale Blutung die schwerwiegendste Komplikation der intravenösen Thrombolyse hinsichtlich der Mortalität darstellt (Emberson et al. 2014). In Kombination mit einer endovaskulären Therapie kann alternativ eine Therapie mit Tenecteplase innerhalb eines 4,5-Stunden-Zeitfensters erwogen werden (Warach et al. 2020; Ringleb et al. 2022).

Die mechanische Thrombektomie (MT) zählt neben der intraarteriellen Thrombolyse zu den endovaskulären Therapieverfahren (EVT). Sie beschreibt ein Verfahren, bei dem ein Thrombus mit Hilfe eines Stent-Retrievers oder Aspirationskatheters mechanisch entfernt wird. Der Nutzen dieses Verfahrens konnte innerhalb eines Zeitfensters von 6 Stunden nach Einsetzen der neurologischen Symptomatik für ein umschriebenes Patient:innenkollektiv in verschiedenen Studien bestätigt werden (Berkhemer et al. 2015; Campbell et al. 2015; Goyal et al. 2015; Jovin et al. 2015; Saver et al. 2015; Bracard et al. 2016; Mocco et al. 2016; Khoury et al. 2017; Muir et al. 2017). Darüber hinaus kann eine Therapie für definierte Patient:innengruppen auch in einem erweiterten Zeitfenster von bis zu 24 Stunden nach Beginn der Symptome in Abhängigkeit von einer erweiterten Bildgebung erwogen werden (Albers et al. 2018; Nogueira et al. 2018).

Ringleb et al. beschreiben in der S2e-Leitlinie „Akuttherapie des ischämischen Schlaganfalls“ das therapeutische Vorgehen im Rahmen des ischämischen Schlaganfalls. Patient:innen, die mit einer TSS von 4,5 Stunden oder weniger behandelt werden können und bei denen keine Kontraindikationen vorhanden

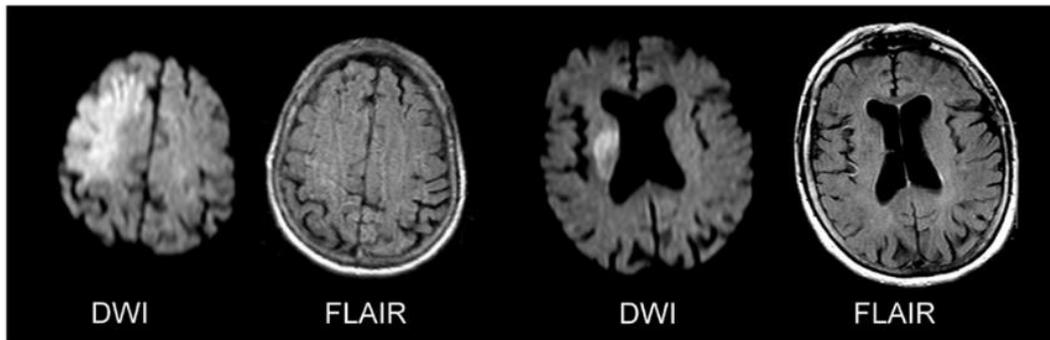
sind, sollen mit einer intravenösen Thrombolyse behandelt werden. Patient:innen mit neurologisch relevantem Defizit sollen mit einer TTT von bis zu 6 Stunden zusätzlich eine mechanische Thrombektomie erhalten. Zudem können die Therapieoptionen auch innerhalb der erweiterten Zeitfenster unter Berücksichtigung der klinischen Symptomatik und der Darstellung von Infarktkern, Penumbra und Kollateralen erwogen werden (Ringleb et al. 2022).

2.1.3.1 Unklares Zeitfenster

In 14 bis 27% aller ischämischen Schlaganfälle sind die therapieentscheidenden Zeitfenster jedoch unklar (Fink et al. 2002, Mackey et al. 2011). Oftmals erleiden Patient:innen Schlaganfälle im Schlaf (Wake-Up-Stroke) oder der Symptombeginn kann aufgrund von Aphasie oder Konfusion der betroffenen Patient:innen bei fehlender Fremdanamnese nicht ermittelt werden. Es konnte gezeigt werden, dass der Schlaganfall bei Wake-Up-Stroke-Patient:innen häufig während der letzten Stunden vor dem Erwachen auftritt und diese deshalb oftmals von einer Therapie profitieren können (Rimmele & Thomalla 2014).

Thomalla et al. konnten zeigen, dass man mit Hilfe eines DWI-FLAIR-Mismatches einen Teil dieser Patient:innen selektieren kann, bei denen das Einsetzen des Schlaganfalls nicht länger als 4.5 Stunden zurückliegt (Thomalla et al. 2011). Ein DWI-FLAIR-Mismatch besteht, falls sich eine ischämische Läsion innerhalb der DWI-Map als hyperintens oder innerhalb der ADC-Map als hypointens darstellt und sich zeitgleich keine entsprechende parenchymale Hyperintensität innerhalb der FLAIR-Map zeigt (siehe Abbildung 7). Thomalla et al. analysierten die Bilddaten von 643 Patient:innen, die von einem akuten ischämischen Schlaganfall betroffen waren und bei denen innerhalb von 12 Stunden nach Einsetzen der Symptomatik eine entsprechende MRT-Bildgebung durchgeführt wurde. Ein solches Mismatch bestand bei 254 Patient:innen und konnte innerhalb dieser Gruppe die Patient:innen, bei denen der Beginn des Schlaganfalls weniger als 4.5 Stunden zurücklag, mit einer Sensitivität von 78% und einem positiv prädiktiven Wert von 83% erfassen (Thomalla et al. 2011).

DWI-FLAIR-mismatch



No DWI-FLAIR-mismatch

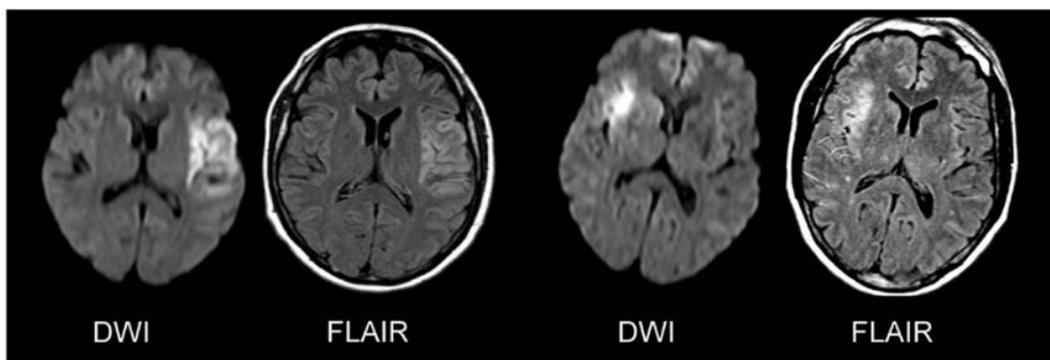


Abbildung 7: DWI-FLAIR-Mismatch (Rimmele & Thomalla 2014)

In einer auf diesen Ergebnissen aufbauenden Studie konnten Thomalla et al. den klinischen Nutzen des DWI-FLAIR-Mismatches nachweisen. Dazu nutzten sie das Mismatch, um Patient:innen mit unklarem Zeitfenster für eine intravenöse Thrombolyse zu selektieren (Thomalla et al. 2018). Die entsprechende Studie wird in Kapitel 3.1.1 näher erläutert.

2.1.4 Therapieevaluation

Durch die Therapie sollen die Ausbreitung des irreversibel geschädigten Infarktkerns und die daraus resultierenden neurologischen Symptome auf ein mögliches Minimum begrenzt werden. Das Therapieergebnis kann anhand des klinischen Outcomes, bildbasierten Scores und der Infarktausbreitung im Verlauf evaluiert werden.

2.1.4.1 Klinischer Outcome

Der klinische Outcome umfasst die Gesamtheit aller erworbenen neurologischen Defizite und die damit verbundenen Einschränkungen im alltäglichen Leben. Es existieren verschiedenste Scores und Skalen, um den klinischen Outcome zu quantifizieren. Im Folgenden werden zwei oft verwendete Skalen beschrieben.

Die modifizierte Rankin-Skala (mRS) erfasst das Ausmaß der durch den Schlaganfall entstandenen Behinderung im Alltag in einer Ordinalskala (siehe Tabelle 2). Diese reicht vom Skalenwert 0, gleichbedeutend mit keiner Symptomatik, bis zum maximalen Skalenwert von 6, gleichbedeutend mit dem Tod infolge des Schlaganfalls (Rankin 1957, van Swieten et al. 1988). In klinischen Studien wird der mRS-Score häufig als primärer Endpunkt verwendet.

Tabelle 2: Modifizierte Rankin-Skala

mRS-Score	Definition
0	Keine Symptomatik und Einschränkung
1	Keine relevante Einschränkung trotz Symptomatik
2	Leichte Einschränkung
3	Moderate Einschränkung
4	Mittelschwere Einschränkung, auf Hilfe angewiesen
5	Schwere Einschränkung, Pflegebedürftigkeit
6	Tod

Die National-Institutes-of-Health-Stroke-Skala (NIHSS) ermöglicht die Quantifizierung der klinischen Symptomatik eines akuten ischämischen Schlaganfalls (siehe Tabelle 3). Dazu werden in der erweiterten Version elf verschiedene klinische Parameter, die innerhalb der neurologischen Befunderhebung erfasst werden, quantifiziert. Ist ein Defizit stark ausgeprägt, wird eine hohe Punktzahl vergeben. Durch die Aufsummierung der einzelnen Punktzahlen kann so maximal ein Score von 42 Punkten erreicht werden (Brott et al. 1989, Meyer & Lyden 2009). In der Hyperakutphase kann die Erhebung des NIHSS-Scores den Schweregrad eines Schlaganfalls quantifizieren. Im Verlauf erhoben, kann der NIHSS-Score aber auch für die Bestimmung des klinischen Outcomes verwendet werden.

Tabelle 3: National-Institutes-of-Health-Stroke-Skala

NIHSS-Score	Interpretation
0	Keine Symptomatik
1-4	Leichter Schlaganfall
5-15	Moderater Schlaganfall
16-20	Mittelschwerer Schlaganfall
21-42	Schwerer Schlaganfall

2.1.4.2 Bildbasierte Scores

Eine Quantifizierung des Infarktareals kann mittels Volumetrie oder für Infarkte des Mediastrom-Gebietes anhand des Alberta-Stroke-Program-Early-CT-Scores (ASPECTS) durchgeführt werden (Barber et al. 2000). Zur Bestimmung des ASPECTS wird das Mediastrom-Gebiet in zehn vordefinierte Bereiche unterteilt

und folgend lässt sich der ASPECTS durch die Aufsummierung aller läSIONsfreien Bereiche bestimmen (Barber et al. 2000). Eine Variation des ASPECTS wurde für Infarkte des Hinteren Kreislaufes entwickelt (Puetz et al. 2011). Zudem wurde die Anwendbarkeit des ASPECTS anhand von DWI-Maps bestätigt (Barber et al. 2005). Es konnte gezeigt werden, dass der initiale ASPECT-Score mit dem klinischen Outcome korreliert und dass eine Infarzierung bestimmter ASPECTS-Areale mit einem schlechten klinischen Outcome assoziiert ist (Barber et al. 2000; Rangaraju et al. 2015; Khan et al. 2017).

Mithilfe der modifizierten Treatment-In-Cerebral-Ischemia-Skala (mTICI-Skala) lässt sich der Grad der Rekanalisierung nach erfolgter Therapie anhand von angiographischen Bilddaten bestimmen (siehe Tabelle 4). Diese reicht vom Skalenwert 0, gleichbedeutend mit fehlender Reperfusion, bis zum maximalen Skalenwert von 3, gleichbedeutend mit einer kompletten antegraden Reperfusion der zuvor okkludierten Arterie ohne Anzeichen für Okklusionen in distal gelegenen Abschnitten (Higashida et al. 2003; Zaidat et al. 2013). Es konnte ein starker Zusammenhang zwischen dem Grad der Rekanalisierung, der infolge einer mechanischen Thrombektomie erreicht wurde, und dem klinischen Outcome nachgewiesen werden (Liebeskind et al. 2019).

Tabelle 4: Modifizierte Treatment-In-Cerebral-Ischemia-Skala

mTICI-Score	Definition
0	Keine Reperfusion
1	Minimale Reperfusion
2a	Teilweise Reperfusion < 50% des Gefäßterritoriums
2b	Teilweise Reperfusion ≥ 50% des Gefäßterritoriums
2c	Nahezu komplette Reperfusion mit langsamem Fluss oder distaler kortikaler Embolie
3	Komplette Reperfusion

2.1.4.3 Infarktausbreitung

Während läSIONsbezogene Endpunkte in Tierversuchen gängiger Weise verwendet werden, spielen diese oftmals in klinischen Studien am Menschen eine untergeordnete Rolle. Aus den insgesamt elf wesentlichen Studien zur mechanischen Thrombektomie (MR CLEAN: (Berkhemer et al. 2015), EXTEND 1A: (Campbell et al. 2015), ESCAPE: (Goyal et al. 2015), REVASCAT: (Jovin et al. 2015), SWIFT PRIME: (Saver et al. 2015), THRACE: (Bracard et al. 2016), THERAPY: (Mocco et al. 2016), EASI: (Khoury et al. 2017), PISTE: (Muir et al. 2017), DEFUSE-3: (Albers et al. 2018), DAWN: (Nogueira et al. 2018)) definierten

lediglich 5 Studien sekundäre oder tertiäre Outcomes, die die Ausbreitung des Infarktes abbildeten (siehe Tabelle 5). Vergleicht man die Ansätze der Studien, so wird deutlich, dass sich diese in verschiedenen Aspekten unterscheiden. In drei Studien werteten die Autor:innen das mittlere Infarktwachstum aus (Campbell et al. 2015; Albers et al. 2018; Nogueira et al. 2018). Im Gegensatz dazu betrachteten zwei Studien das Infarktvolumen zu einem festgelegten Zeitpunkt (Berkhemer et al. 2015; Jovin et al. 2015). Die Bildgebung im Verlauf wurde in vier Studien 24 Stunden nach der initialen klinischen Untersuchung, der initialen Bildgebung oder der erfolgten Therapie erhoben (Campbell et al. 2015; Jovin et al. 2015; Albers et al. 2018; Nogueira et al. 2018) und in einer Studie zwischen fünf und sieben Tagen nach der initialen klinischen Untersuchung (Berkhemer et al. 2015). Die Ausbreitung des Infarktes wurde anhand von nativen CT-Maps (Berkhemer et al. 2015; Jovin et al. 2015; Albers et al. 2018), CTP-Maps (Campbell et al. 2015; Nogueira et al. 2018) und/oder DWI-Maps bestimmt (Jovin et al. 2015; Albers et al. 2018; Nogueira et al. 2018).

Tabelle 5: Studien zur endovaskulären Therapie mit läsionsbezogenem Endpunkt

Studie	Inkludierte Patient:innen	Initiale Bilddaten	Bildgebung im Verlauf	Ausbreitung des Infarktes
MR CLEAN (Berkhemer et al. 2015)	Kontrolle: 267 EVT: 233	natives CT (nCT) + CTP oder CTA oder DWI + CTA oder MRA oder Digitale Subtraktionsangiographie (DSA)	CTA oder MRA oder DSA nach 24 Stunden, natives CT zwischen 5 und 7 Tagen	Infarktvolumen: Kontrolle(160): 79 ml MT(138): 49 ml
EXTEND-1A (Campbell et al. 2015)	Kontrolle: 35 MT: 35	CTP + CTA	CTP + Angiographie nach 24 Stunden	Mittleres Infarktwachstum: Kontrolle: 35.3 ml MT: 10.9 ml, p=0.007
REVASCAT (Jovin et al. 2015)	Kontrolle: 103 MT: 103	natives CT + CTP oder CTA + DWI	natives CT + CTA oder DWI + MRA nach 24 Stunden	Infarktvolumen: Kontrolle(nCT:91, DWI:10): 16.3 ml MT(nCT:98, DWI:5): 38.6 ml, p=0.02
DEFUSE-3 (Albers et al. 2018)	Kontrolle: 92 MT: 90	CTP oder DWI + PWI + CTA oder MRA	DWI oder natives CT nach 24 Stunden	Mittleres Infarktwachstum: Kontrolle(90): 33 ml MT(89): 23 ml, p=0.08
DAWN (Nogueira et al. 2018)	Kontrolle: 99 MT: 107	CTP + CTA oder DWI + MRA	CTP, CTA oder DWI, MRA nach 24 Stunden	Mittleres Infarktwachstum: Kontrolle: 13 ml MT: 1 ml, p<0.001

Zusammenfassend lässt sich feststellen, dass kein einheitlicher Ansatz besteht, die Ausbreitung des Infarktes als Outcome zu erfassen. Vielmehr wird ein solcher Outcome in Abhängigkeit von den in der Studie verwendeten Aufnahmeverfahren, Aufnahmezeitpunkten und Bildmodalitäten definiert. In den beschriebenen Studien wurde der bildbasierte Outcome somit meistens nach 24 Stunden ausgewertet, da zu diesem Zeitpunkt auch eine Angiographie zur Bestimmung der Rekanalisation durchgeführt wurde.

2.1.4.3.1 Einfluss auf klinischen Outcome

Die Ausbreitung des Infarktes besitzt klinische Relevanz. So konnten Zusammenhänge zwischen der Infarktausbreitung zu verschiedenen Zeitpunkten und verschiedenen definierten klinischen Outcomes nachgewiesen werden. Es konnte gezeigt werden, dass das Wachstum des Infarktes innerhalb der ersten 24 Stunden mit einer Zunahme des NIHSS-Scores in den ersten 24 Stunden korreliert, jedoch nicht das Volumen des Infarktes nach 24 Stunden (Simpkins et al. 2017). Eine andere Studie zeigte, dass das Infarktvolumen nach 24 Stunden mit dem mRS-Score nach drei Monaten korreliert (Drewer-Gutland et al. 2015). Dabei konnte nachgewiesen werden, dass das Volumen des Infarktes nach 24 Stunden den klinischen Outcome nach drei Monaten stärker beeinflusst als der Grad der Rekanalisierung (Drewer-Gutland et al. 2015). Zudem konnte gezeigt werden, dass das Infarktvolumen, welches zwischen 24 und 48 Stunden nach endovaskulärer Therapie erhoben wurde, zu den stärksten Prädiktoren des mRS-Scores nach 3 Monaten gehört (Zaidi et al. 2012). Das Wachstum des Infarktes innerhalb der ersten Woche korreliert mit der Änderung des NIHSS-Scores in der ersten Woche (Kimura et al. 2011) und dem mRS-Score nach 3 Monaten (Barrett et al. 2009). Weiterhin konnten Korrelationen zwischen dem Infarktvolumen innerhalb der ersten 3 Wochen und dem NIHSS-Score nach mehr als einem Monat (Schwamm et al. 1998), dem NIHSS-Score nach 3 Monaten (Saver et al. 1999) und dem mRS-Score nach 3 Monaten (Yoo et al. 2012; Compagne et al. 2019) aufgezeigt werden. Der Zusammenhang zwischen dem Infarktvolumen und dem NIHSS-Score konnte auch für spätere Zeitpunkte und den initialen Zeitpunkt nachgewiesen werden (Lövlad et al. 1997; Beaulieu et al. 1999).

2.1.4.3.2 Zeitlicher Verlauf

Die Ausbreitung des Infarktkerns unterliegt mit voranschreitender Zeit einem räumlichen Wandel, der schon in mehreren Studien anhand verschiedener Aufnahmeverfahren für verschiedene Zeitabschnitte untersucht wurde. Für die Ausbreitung des Infarktkerns im zeitlichen Verlauf sind die pathophysiologischen Prozesse ursächlich, die bereits in Kapitel 2.1.1.2 beschrieben wurden. Das anhand von DWI-Maps erhobene Infarktvolumen steigt zu Beginn und erreicht die maximale Ausdehnung im Verlauf der ersten Woche (Schwamm et al. 1998; Beaulieu et al. 1999; Lansberg et al. 2001a; Kimura et al. 2011). In diesem Zeitraum sind zwei Prozesse ursächlich für das Wachstum des Infarktkerns. Zum einen infarzieren mit fortschreitender Zeit zunehmend Anteile der Penumbra. Zum anderen bilden sich infolgedessen zunächst das zytotoxische und später das vasogene Ödem aus, welche wiederum mit inflammatorischen Prozessen einhergehen (Kahle et al. 2009). Auch das anhand von T2w- und FLAIR-Maps erhobene Volumen des Infarktes erreicht die maximale Ausdehnung in der ersten Woche und verkleinert sich kontinuierlich im Verlauf (siehe Abbildung 8) (Schwamm et al. 1998; Beaulieu et al. 1999; Lansberg et al. 2001a; Ritzl et al. 2004; Gaudinski et al. 2008; Krongold et al. 2015). Die Abnahme des Infarktvolumens im Verlauf ist dabei zunächst durch den Rückgang des vasogenen Ödems und der anschließenden Vernarbung des infarzierten Gewebes zu erklären (Burda & Sofroniew 2014).

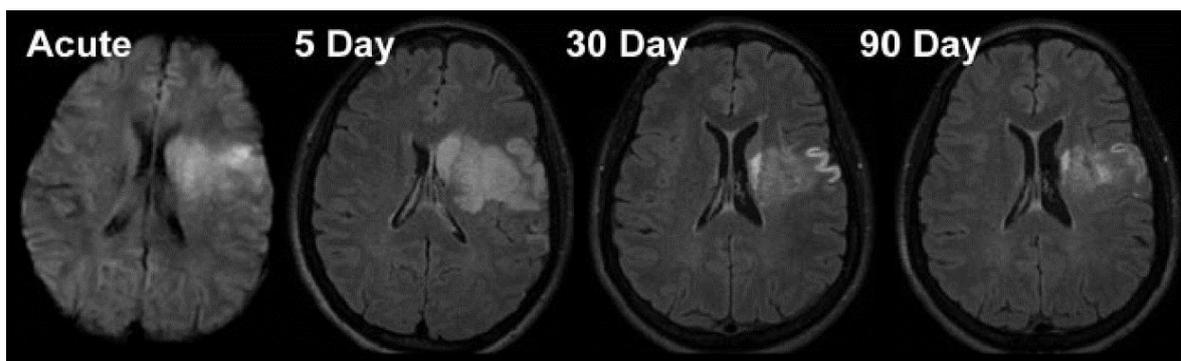


Abbildung 8: Darstellung eines Infarktes im zeitlichen Verlauf (Gaudinski et al. 2008)

Der Infarkt wird anhand einer während der Hyperakutphase aufgenommenen DWI-Map (hier gekennzeichnet mit „Acute“) und drei FLAIR- Maps, die nach 5, 30 und 90 Tagen im Verlauf aufgenommen wurden, dargestellt.

Das Volumen der initialen DWI-Bildgebung korreliert mit dem Volumen der im Verlauf erhobenen FLAIR-Bildgebung (Krongold et al. 2015). Dabei kann sich das Volumen der initialen DWI-Bildgebung in Abhängigkeit vom Patient:innenkollektiv

und dem Zeitpunkt der initialen Bildgebung größer (Schwamm et al. 1998; Ritzl et al. 2004; Gaudinski et al. 2008; Kate et al. 2015) oder kleiner (Beaulieu et al. 1999; Pialat et al. 2005; Kate et al. 2015; Krongold et al. 2015) darstellen als das Volumen einer nach mehr als einem Monat im Verlauf erhobenen T2w- oder FLAIR-Bildgebung.

2.1.4.3.3 Klinische Einflussfaktoren

Es konnte gezeigt werden, dass eine mechanische Thrombektomie in definierten Patient:innenkollektiven das Infarktvolumen, welches 24 Stunden nach erfolgter Therapie erfasst wurde (Campbell et al. 2015; Nogueira et al. 2018), und das mittlere Läsionswachstum über 24 Stunden (Jovin et al. 2015) signifikant verringert. Ein direkter Zusammenhang zwischen einer intravenösen Thrombolyse und dem Läsionswachstum wurde in keiner der drei wesentlichen Studien (NINDS: (NINDS-Study-Group 1995), ECASS-III: (Hacke et al. 2008), SITS-ISTR: (Wahlgren et al. 2008)) untersucht, konnte jedoch in einer kleineren Studie aufgezeigt werden (Parsons et al. 2002).

Die Auswirkung einer erfolgreichen Rekanalisierung auf die Entwicklung eines Infarktes konnte durch verschiedene Studien aufgezeigt werden: Eine Rekanalisierung verkleinert das Läsionswachstum über 24 Stunden und 90 Tage (Campbell et al. 2012b) und über sieben und mehr als 30 Tage (Krongold et al. 2015) signifikant. Auch für Patient:innen, die mit einer intravenösen Thrombolyse behandelt wurden und initial ein PWI-DWI-Mismatch aufwiesen, konnte der Effekt einer Rekanalisierung auf das Läsionswachstum über 30 Tage nachgewiesen werden (Olivot et al. 2008). In einer weiteren Studie zeigte sich, dass sich das Infarktvolumen in den ersten 24 Stunden unabhängig von der Rekanalisierung vergrößerte und sich dann über 60 Tage verkleinerte, wobei eine Rekanalisierung mit einem geringeren Volumen nach 60 Tagen assoziiert war (Pialat et al. 2005). Der durch eine mechanische Thrombektomie erzielte Grad der Rekanalisierung ist dabei abhängig von der Zeit, die zwischen der Klinikaufnahme und der Therapieinitiierung vergeht (Bourcier et al. 2019).

Zudem konnte auch ein Einfluss der Reperfusion, die neben der Rekanalisierung beispielsweise auch durch Gefäßkollateralen beeinflusst wird, auf das Läsionswachstum (Davis et al. 2008) und das Infarktvolumen nach 3 Stunden (Cho et al. 2015) nachgewiesen werden. Eine weitere Studie betrachtete den

Zusammenhang der Reperfusion und der Kollateralen auf das Läsionswachstum nach endovaskulärer Therapie in differenzierten Zeitfenstern. Dabei wurde das Wachstum des Infarktes innerhalb der ersten 24 Stunden als frühes Wachstum und das Wachstum zwischen dem ersten und fünften Tag als spätes Wachstum definiert. In dieser Studie war eine Reperfusion ausschließlich mit einem verminderten spätem Wachstum und schlechte Kollateralen waren ausschließlich mit einem größeren frühem Wachstum assoziiert (Tate et al. 2021). Der Grad der Kollateralen konnte zudem als von der Rekanalisierung unabhängiger Prädiktor des Infarktvolumens beschrieben werden (Angermaier et al. 2011) und kann mittels Perfusionsbildgebung über die Hypoperfusion-Intensity-Ratio (HIR) erfasst werden (Olivot et al. 2014). Hyperglykämie, Hypertonus, Rauchen und stattgehabte Infarkte sind Faktoren, die mit schlechten Kollateralen assoziiert sind (Liebeskind et al. 2014).

In einer weiteren Studie konnte gezeigt werden, dass ein großes initiales Läsionsvolumen, ein hoher initialer NIHSS-Score, das Fehlen einer intravenösen Thrombolyse oder endovaskulären Therapie, ein hoher Blutglukosewert und schlechte Kollateralen mit einem großen Infarktwachstum assoziiert waren. Den einzigen unabhängigen Prädiktor stellte das Clinical-Diffusion-Mismatch (CDM) dar, welches sich durch einen initialen NIHSS-Score größer gleich 8 und ein initiales DWI-Volumen kleiner gleich 25 ml definierte (Nogueira et al. 2017). Der Zusammenhang einer Hyperglykämie auf das Läsionswachstum konnte auch für Patient:innen gezeigt werden, die mit einer intravenösen Thrombolyse behandelt wurden (Ribo et al. 2007). Eine andere Studie konnte den Einfluss von Leukoaraiose auf die Infarktausbreitung von nicht therapierten Patient:innen nachweisen. So konnte gezeigt werden, dass das Ausmaß der Leukoaraiose mit dem im Verlauf infarzierten Anteil der Penumbra korrelierte (Ay et al. 2008).

Eine Abnahme des Therapieeffektes der intravenösen Thrombolyse und der mechanischen Thrombektomie mit zunehmender TTT wurde für klinische Outcomes beschrieben (Emberson et al. 2014; Saver et al. 2016). Demnach lässt sich auch ein Zusammenhang auf das Infarktwachstum vermuten.

2.2 Methodischer Hintergrund

Der Begriff Künstliche Intelligenz beschreibt das Fachgebiet der Informatik, in dem Maschinen oder Computer ein intelligentes Problemlösungsverhalten erlernen und

anwenden. Beim Machine-Learning, einem Teilgebiet der Künstlichen Intelligenz, wird dieses intelligente Problemlösungsverhalten anhand von Datenmerkmalen durch statistische Modelle und Algorithmen erlernt, ohne dass dabei explizite Instruktionen vorgegeben werden müssen. Innerhalb des Machine-Learnings lassen sich zwei unterschiedliche Lernverfahren voneinander unterscheiden. Beim Supervised-Learning müssen die zu erlernenden Ergebnisse vordefiniert sein und werden als Ground Truth bezeichnet. Innerhalb der Trainingsphase werden Datenmerkmale erlernt, mit denen sich die Daten der Ground Truth entsprechend präzisieren lassen. Beim Unsupervised-Learning werden Zusammenhänge anhand von Datenmerkmalen selbständig und unüberwacht erlernt. Hierbei wird keine vorannotierte Ground Truth verwendet. Bei der Prädiktionsaufgabe im Rahmen des Supervised-Learnings lassen sich wiederum zwei Anwendungsbereiche unterscheiden. Bei Klassifikationsaufgaben sollen die Daten in vordefinierte Klassen kategorisiert werden. Bei Regressionsaufgaben sollen die Werte einer kontinuierlichen Variablen vorhergesagt werden.

2.2.1 Deep-Learning

Deep-Learning beschreibt wiederum ein Teilgebiet des Machine-Learnings (siehe Abbildung 9). Mittels Deep-Learning können Artificial-Neural-Networks, die sich durch ihre spezifische Architektur auszeichnen, lernen verschiedenste Datenmerkmale zu abstrahieren. In Abgrenzung zu anderen Machine-Learning-Verfahren müssen dabei relevante Datenmerkmale nicht im Vorhinein definiert werden (Feature-Engineering). Nachteilig ist, dass für diesen Lernprozess große Datenmengen benötigt werden und die erlernten Datenmerkmale für die anwendende Person schwer nachzuvollziehen sind.

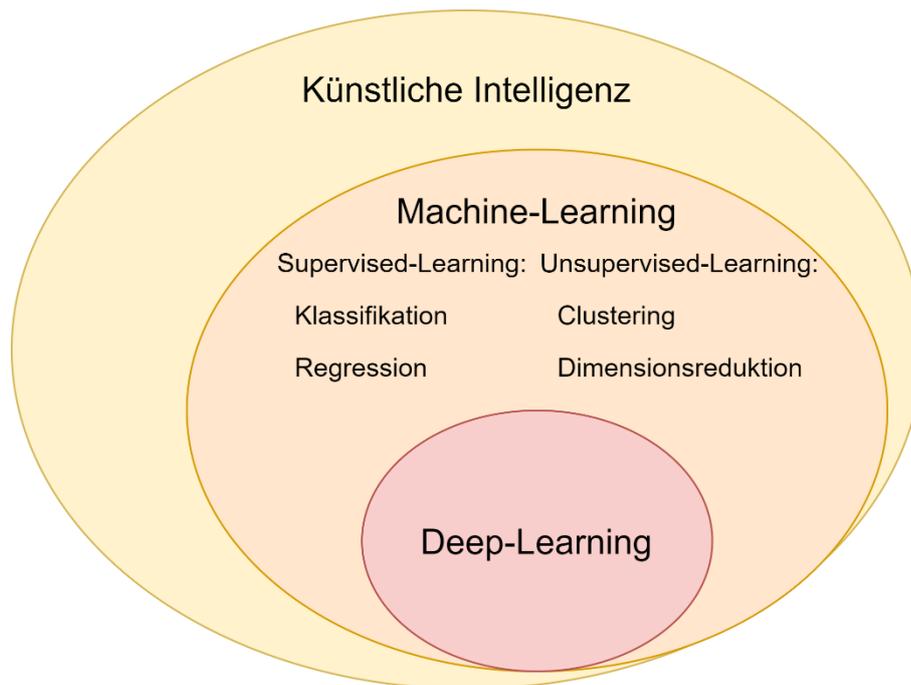


Abbildung 9: Bereiche der Künstlichen Intelligenz

Artificial-Neural-Networks finden schon heute in verschiedenen medizinischen Fachbereichen Anwendung (Litjens et al. 2017). Zu den zahlreichen Anwendungsbereichen gehören unter anderem die Segmentierung relevanter Bildinhalte (Isensee et al. 2021), die Klassifikation von Tumorentitäten (Esteva et al. 2017) sowie die Diagnose (Rajpurkar et al. 2017) als auch die Prognose verschiedenster Erkrankungen (Chaudhary et al. 2018).

2.2.1.1 Netzwerktypen

Im Folgenden werden einige ausgewählte Netzwerktypen näher erläutert. Dabei wird deutlich, dass viele Netzwerktypen architektonisch Strukturen aus der Biologie nachempfunden wurden. Ihre tatsächliche Funktionsweise unterscheidet sich jedoch von den biologischen Korrelaten.

2.2.1.1.1 Perzeptron

Das Perzeptron bildet den grundlegenden Baustein vieler Netzwerktypen. Ein einzelnes Perzeptron stellt im Grunde genommen ein sehr rudimentäres Netzwerk dar (siehe Abbildung 10). Der Funktionsweise eines Neurons nachempfunden empfängt das Perzeptron verschieden gewichtete numerische Eingabedaten und prozessiert diese intern. Nach erfolgter Prozessierung der Daten leitet es eine numerische Ausgabe weiter.

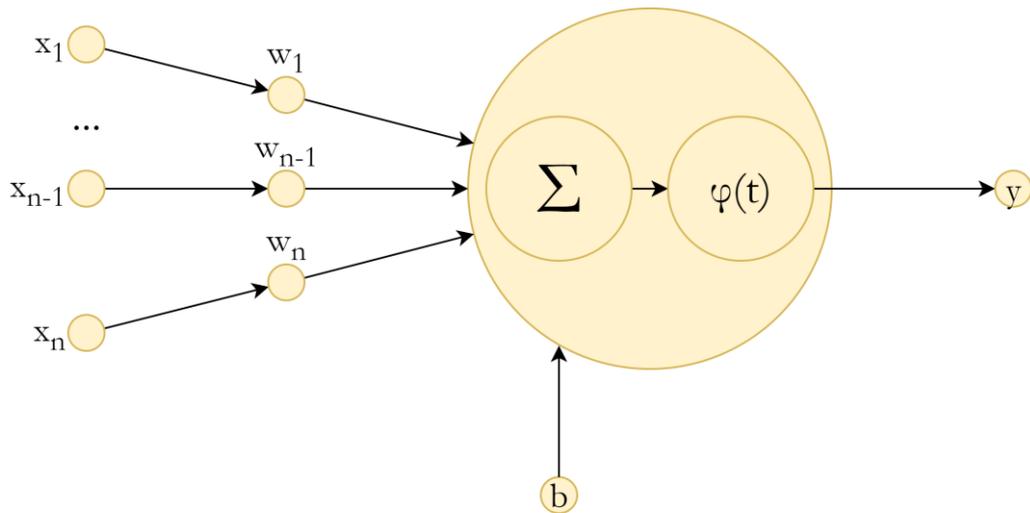


Abbildung 10: Aufbau eines Perzeptrons

Die Ausgabe y eines solchen Perzeptrons kann beschrieben werden durch

$$y = \varphi \left(\sum_{i=1}^n w_i x_i + b \right),$$

wobei $\varphi()$ die Aktivierungsfunktion und $\sum_{i=1}^n w_i x_i + b$ die aufsummierten Eingabedaten beschreiben. Dabei beschreibt x_i den numerischen Wert der i ten Eingabe mit dem zugehörigen Gewicht w_i , b das Bias und n die Menge der Eingabedaten x_i (Huang et al. 2010).

Die interne Prozessierung der Daten erfolgt demensprechend durch die Anwendung einer Aktivierungsfunktion φ auf die aufsummierten Eingabedaten. Im Rahmen des Deep-Learning finden verschiedene Aktivierungsfunktionen Anwendung. Folgend sind einige häufig verwendete Aktivierungsfunktionen aufgeführt:

Lineare Funktion: $\varphi(t) = a \cdot t$

Sigmoidfunktion: $\varphi(t) = \frac{1}{1+e^{-t}}$

Tangens hyperbolicus: $\varphi(t) = \frac{2}{e^{2t}+1}$

Rectified-Linear-Unit (ReLU)-Funktion: $\varphi(t) = \begin{cases} t & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$

Leaky-ReLU-Funktion: $\varphi(t) = \begin{cases} t & \text{if } t \geq 0 \\ 0.01t & \text{if } t < 0 \end{cases}$

Exponential-Linear-Unit (ELU)-Funktion: $\varphi(t) = \begin{cases} t & \text{if } t \geq 0 \\ a(e^{t-1}) & \text{if } t < 0 \end{cases}$

Die unterschiedlichen Eigenschaften der verschiedenen Aktivierungsfunktionen gehen mit verschiedenen Vor- und Nachteilen einher. Beispielsweise konnte gezeigt werden, dass die Verwendung der ReLU-Funktion zu einer Verminderung der Rechenzeit führt, da diese ohne Brüche oder Exponenten berechnet wird (Zeiler et al. 2013). Verglichen mit der Sigmoidfunktion ist die ReLU-Funktion jedoch sensibler für Overfitting (Glorot et al. 2011). Ein in Relation zur Gewichtung großer negativer Bias kann dazu führen, dass ein Perzeptron, welches die reLU-Aktivierungsfunktion verwendet, die Gewichtung nicht mehr trainieren kann, da die ReLU-Funktion für negative Werte keine Steigung aufweist (Goodfellow et al. 2016). Die Verwendung der Leaky-ReLU-Funktion umgeht wiederum dieses Problem. Die optimale Wahl der innerhalb der internen Prozessierung eines Netzwerkes verwendeten Aktivierungsfunktion folgt somit keinen festen Regeln, sondern muss je nach verwendetem Modell und Aufgabenbereich explorativ ermittelt werden.

2.2.1.1.2 Multilayer-Perzeptron

Sogenannte Layer bilden sich durch die vertikale Aufreihung mehrerer einzelner Perzeptren. Einige Netzwerktypen bestehen aus mehreren, horizontal angeordneten Layern. Zwischen diesen Layern werden Daten über die Verbindungen der einzelnen Perzeptren übertragen. Dabei definieren diese einzelnen Verbindungen neben der Datenflussrichtung auch ihre Einflussgröße auf das nachfolgende Perzeptron durch eine sogenannte Gewichtung. Der erste Layer wird als Input-Layer bezeichnet. Dieser Input-Layer empfängt die zu verarbeitenden Eingabedaten und leitet diese ins Netzwerk weiter. Nach erfolgter Transformation der Daten innerhalb des Netzwerkes werden die Ausgabedaten durch den Output-Layer ausgegeben. Im Rahmen von Klassifikationsaufgaben entspricht die Anzahl der Perzeptren des Output-Layers meist der Anzahl der zu differenzierenden Klassen. Bei Regressionsaufgaben wird meist nur ein einzelnes Perzeptron im Output-Layer verwendet. Layer zwischen dem Input- und dem Output-Layer werden als Hidden-Layer bezeichnet. Besitzt ein Netzwerk zwei oder mehr Hidden-Layer, wird es gängiger Weise als Deep-Network bezeichnet.

In Abbildung 11 ist ein exemplarisches Multilayer-Perzeptron (MLP) dargestellt. Das Multilayer-Perzeptron ist ein einfaches Netzwerk, welches sich durch zwei elementare Eigenschaften definiert. Jedes Perzeptron eines Layers ist hier mit

allen Perzeptronen des nachfolgenden Layers verbunden (Fully-Connected-Layer) und die Transformation der Daten läuft stringent vom Input-Layer in Richtung des Output-Layers (Feed-Foward-Neural-Network). Die damit einhergehende Architektur des Multilayer-Perzeptrons erinnert an die neuronale Struktur des zentralen Nervensystems.

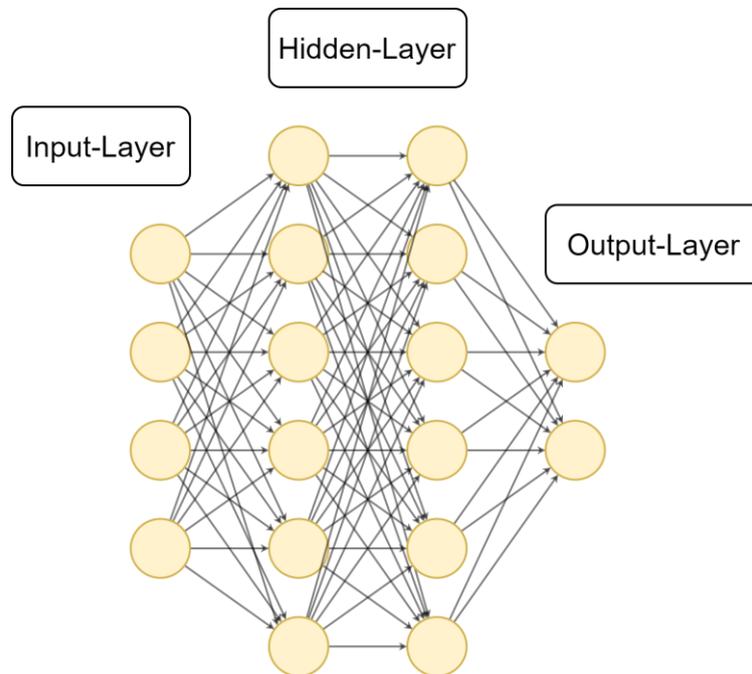


Abbildung 11: Architektur eines Multilayer-Perzeptrons

Folgend kann die Transformation eines Multilayer-Perzeptrons mit zwei Hidden-Layern durch die Formel

$$y_l(x; \theta) = \gamma \left(\sum_{k=1}^K w_{kl}^{(3)} \varphi \left(\sum_{j=1}^J w_{jk}^{(2)} \varphi \left(\sum_{i=1}^I w_{ij}^{(1)} x_i + b_j^{(1)} \right) + b_k^{(2)} \right) + b_l^{(3)} \right)$$

beschrieben werden. Dabei definiert y_l die Ausgabe des l ten Perzeptrons des Output-Layers mit $y = [y_l] \in \mathbb{R}^L$, $x = [x_i] \in \mathbb{R}^I$ den Eingabevektor des Input-Layers, θ die Parameter, entsprechend den Gewichten W und die Bias b , mit $\theta = \{W^{(1)} = [W_{ji}^{(1)}] \in \mathbb{R}^{J \times I}, W^{(2)} = [W_{kj}^{(2)}] \in \mathbb{R}^{K \times J}, W^{(3)} = [W_{lk}^{(3)}] \in \mathbb{R}^{L \times K}\}$, $\varphi(\cdot)$ die Aktivierungsfunktion der Hidden-Layer und $\gamma(\cdot)$ die Aktivierungsfunktion des Output-Layers (Shen et al. 2017). I beschreibt dabei die Anzahl der Perzeptronen des Input-Layers, J die Anzahl der Perzeptronen des ersten Hidden-Layers, K die Anzahl der Perzeptronen des zweiten Hidden-Layers und L die Anzahl der Perzeptronen des Output-Layers (Shen et al. 2017).

Während ein einzelnes Perzeptron unabhängig von der gewählten Aktivierungsfunktion keine nichtlinearen Klassifikations- oder Regressionsaufgaben lösen kann, können Multilayer-Perzeptronen nichtlineare Zusammenhänge der Eingabedaten erfassen. So konnte gezeigt werden, dass ein Multilayer-Perzeptron jede stetige Funktion approximieren kann, sofern es mindestens einen Hidden-Layer besitzt (Hornik et al. 1989).

2.2.1.1.3 Convolutional-Neural-Networks

Die Architektur und die interne Struktur eines Convolutional-Neural-Networks (CNNs) erinnern an den Aufbau und die Funktionsweise des primär visuellen Cortex des Menschen. Im Gegensatz zu einem Multilayer-Perzeptron, welches ausschließlich vektorbasierte Eingabedaten prozessiert, ist ein CNN fähig zwei- oder dreidimensionale Eingabedaten zu transformieren und so auch räumliche Datenmerkmale zu abstrahieren. Dies wird durch die sogenannten Convolutional-Layer und Pooling-Layer ermöglicht.

In den Convolutional-Layern werden die Eingabebilddaten durch die Convolutional-Kernel gefiltert. Ein Kernel beschreibt dabei eine meist quadratische oder quaderförmige Matrix, die analog zu einem rezeptiven Feld die Informationen aus einem Bildausschnitt verarbeitet. Die einzelnen Matrixelemente entsprechen dabei erlernbaren Gewichtungen. Da dieser Bildausschnitt zumeist kleiner ist als die Eingabebilddaten, wird der Kernel in einem festgelegten Schema, dem sogenannten Stride, verschoben, bis die gesamten Eingabebilddaten erfasst wurden. Die Ergebnisdaten dieser Transformation werden auf einer so neugenerierten Feature-Map zusammengefasst. Die Transformation $A_j^{(l)}$ zwischen der Feature-Map i des Layers $l - 1$ und der Feature-Map j des Layers l kann beschrieben werden als

$$A_j^{(l)} = \gamma \left(\sum_{i=1}^{M^{(l-1)}} A_i^{(l-1)} * k_{ij}^{(l)} + b_j^{(l)} \right),$$

wobei $M^{(l-1)}$ die Anzahl der Feature-Maps in Layer $l - 1$, $A_i^{(l-1)}$ die i te Feature-Map des Layers $l - 1$, $k_{ij}^{(l)}$ die Convolutional-Kernels zwischen den Layern $l - 1$ und l mit den dazugehörigen Bias $b_j^{(l)}$, $*$ einen Faltungsoperator und γ die Aktivierungsfunktion definieren (Shen et al. 2017).

In den Pooling-Layern wird die Dimensionalität der Feature-Maps reduziert (Downsampling). Dies führt zu einer dazu, dass in dem darauffolgenden Convolutional-Layer größere Anteile der ursprünglichen Eingabebilddaten durch einen Kernel erfasst werden, und zum anderen die Komplexität des Netzwerkes und die damit einhergehende Rechenintensität verringert wird. Auch die Pooling-Layer erfassen die Eingabebilddaten durch Kernel. Jedoch werden hier die Daten nicht anhand von erlernbaren Gewichtungen transformiert, sondern durch eine festgelegte Funktion, die auf den erfassten Bildabschnitt angewandt wird. Beim Maximum-Pooling wird dabei der Maximalwert der erfassten Voxel verwendet, beim Average-Pooling wird ein Mittelwert aus allen erfassten Voxeln gebildet. Alternativ dazu kann die Dimensionalität einer Feature-Map auch mittels eines Convolutional-Layers reduziert werden. Hierbei muss der Stride, mit dem der Convolutional-Kernel die Eingabebilddaten erfasst, entsprechend gewählt werden (Strided-Convolution). In Abbildung 12 ist die Funktionsweise von Convolutional- und Pooling-Layern verbildlicht dargestellt.

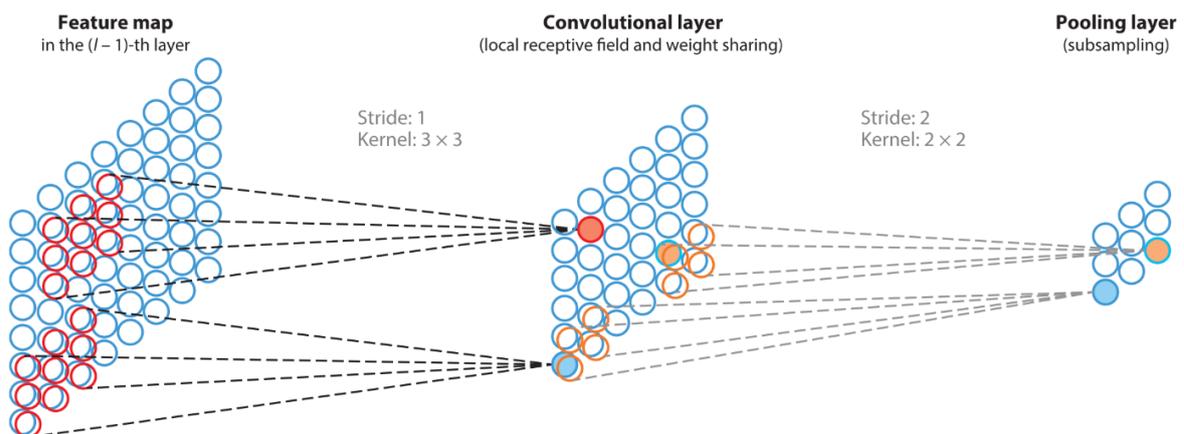


Abbildung 12: Funktion eines Convolutional- und eines Pooling-Layers (Shen et al. 2017)

Ein klassisches CNN, welches für Klassifikationsaufgaben verwendet wird, besteht aus zwei Abschnitten. Der erste Abschnitt setzt sich aus aneinandergereihten Convolutional- und Pooling-Layern zusammen und wird als Encoder bezeichnet. Innerhalb des Encoders werden für eine Aufgabe relevante Bildmerkmale extrahiert. Dabei werden räumlich nahe Bildmerkmale in den ersten Layern erfasst, wohingegen globale Zusammenhänge in den tiefergelegenen Layern erkannt werden. Die durch den Encoder transformierten zwei- oder dreidimensionalen Feature-Maps werden anschließend in einen Vektor umgewandelt. Den zweiten Abschnitt bildet ein Multilayer-Perzeptron, welches

letztendlich das ursprüngliche Eingabebild anhand der durch den Encoder extrahierten Bildmerkmale klassifiziert. Bekannte, auf diesem grundlegenden Aufbau basierende Netzwerke sind beispielsweise das LeNet (Lecun et al. 1998), das in Abbildung 13 dargestellte VGGNet (Simonyan & Zisserman 2015), das GoogLeNet (Szegedy et al. 2015), das ResNet (He et al. 2016) oder das AlexNet (Krizhevsky et al. 2017).

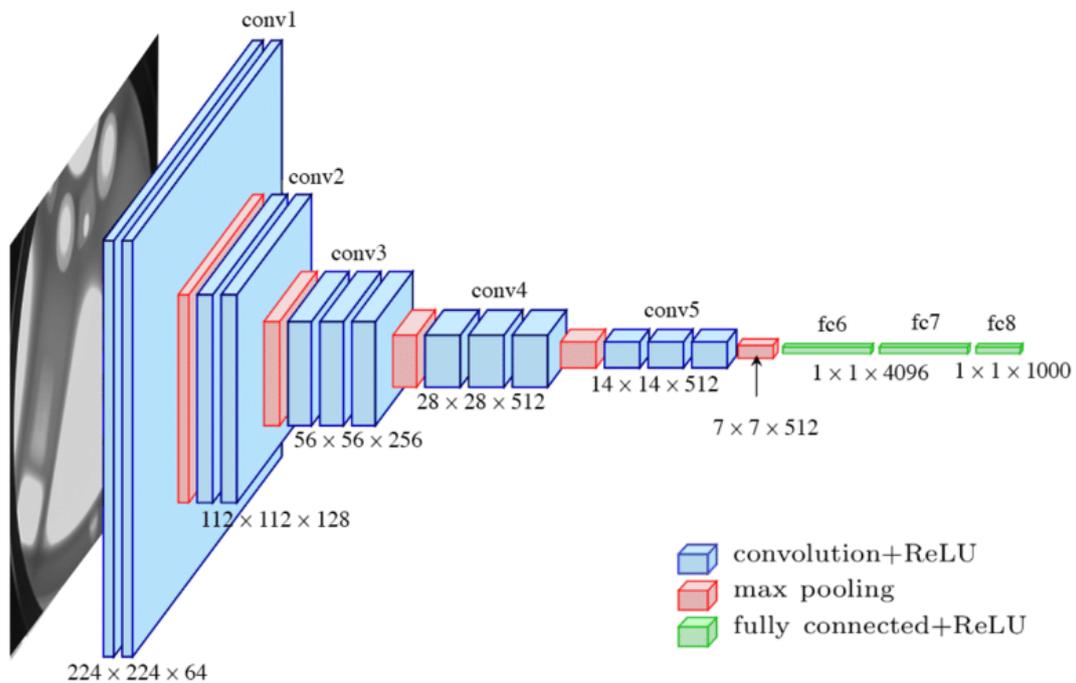


Abbildung 13: VGGNet (Ferguson et al. 2017)

Für die Segmentierung relevanter Bildbereiche werden Fully-Convolutional-Networks (FCN) verwendet, die ausschließlich aus Convolutional-Layern und Pooling-Layern bestehen, sich aber ebenso aus zwei Abschnitten zusammensetzen. Analog zu den klassischen CNNs entspricht der erste Abschnitt einem Encoder, in dem relevante Bildmerkmale extrahiert werden. In dem zweiten als Decoder bezeichneten Abschnitt werden die durch den Encoder extrahierten Feature-Maps zurück in die Größe der Eingabebilddaten transformiert (Upsampling), um so den ursprünglichen räumlichen Kontext wiederherzustellen. Der Bereich zwischen Encoder und Decoder wird als Bottleneck bezeichnet. Das Upsampling wird durch Upsampling-Layer oder Transposed-Convolutional-Layer erreicht. Zwischen dem Encoder- und Decoder-Abschnitt bestehen auf mehreren Ebenen sogenannte Skip-Connections, die die extrahierten Bildmerkmale aus dem ersten Abschnitt mit dem räumlichen Kontext des zweiten Abschnitts kombinieren.

Bekannte und auf diesem grundlegenden Aufbau basierende Netzwerke sind beispielsweise das U-Net (siehe Abbildung 14) (Ronneberger et al. 2015), das V-Net (Milletari et al. 2016) oder das SegNet (Badrinarayanan et al. 2017).

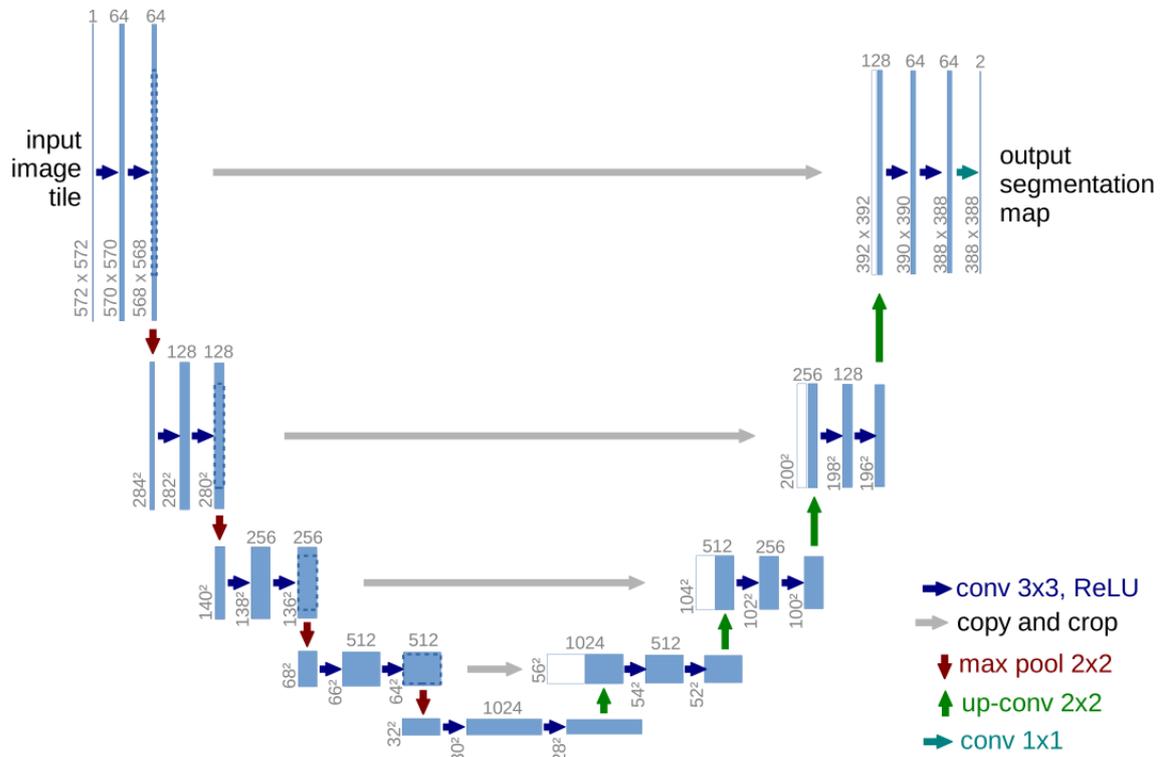


Abbildung 14: U-Net (Ronneberger et al. 2015)

2.2.1.1.3.1 nnUNet-Methode

Die nnUNet-Methode beschreibt eine Deep-Learning-basierte, sich selbst konfigurierende Segmentierungsmethode für Bilddaten. Das Besondere hierbei ist die Methodik der Selbstkonfiguration, bei der verschiedene Parameter der Bilddatenvorverarbeitung, der Netzwerkarchitektur, des Trainings und des Postprocessings in Abhängigkeit vom verwendeten Datensatz ermittelt werden (Isensee et al. 2021). Während einige Parameter stets fest vorgegeben sind, werden andere je nach Datensatz und verfügbarer Rechenleistung regelbasiert oder empirisch ermittelt. Die Autor:innen konnten zeigen, dass die Methodik der Konfiguration dieser Parameter mehr Einfluss auf die Performance eines Modells hat als architektonische Variationen. Zudem ist es fraglich, ob sich die Performanceverbesserungen vieler dieser architektonischen Variationen überhaupt generalisieren lassen, da diese häufig ausschließlich anhand eines einzelnen Datensatzes erstellt und evaluiert werden (Isensee et al. 2021).

Architektonisch orientieren sich die nnUNet-Modelle an dem bereits vorgestellten UNet (Ronneberger et al. 2015). Je nach Datensatz und verfügbarer Rechenleistung werden verschiedene Modell-Variationen konfiguriert. Zu den möglichen Konfigurationen gehören ein 2D-nnUNet-Modell, ein 3D-nnUNet-Modell, welches dreidimensionale Bilddaten in voller Auflösung verarbeitet, und ein 3D-nnUNet-Kaskade-Modell, bei dem das erste UNet zunächst die Bilddaten in reduzierter Auflösung segmentiert und anschließend ein zweites UNet diese Segmentierungen in die volle Auflösung überträgt. In allen Modellen werden ausschließlich Convolutional-Layer verwendet (Isensee et al. 2021). Ein Computational-Block bildet sich dabei aus einer Convolution, gefolgt von einer Instance-Normalisierung (Ulyanov et al. 2016) und einer Leaky-ReLU-Aktivierung (Maas et al. 2013). Jede Ebene des Modells besteht aus zwei Computational-Blocks. Das Downsampling geschieht mittels Strided-Convolutions, das Upsampling mittels Transposed-Convolutions. Die Feature-Maps werden bis auf eine minimale Größe von $[4 \times 4 \times 4]$ gepoolt. Anisotrope Bilddaten werden solange in einer Ebene gepoolt, bis sich die kleinste und größte Achse maximal um den Faktor 2 unterscheiden (Isensee et al. 2021). Ein Deep-Supervision-Loss wird aus allen Ebenen des Decoders, abzüglich der 2 am tiefsten gelegenen Ebenen, berechnet (Lee et al. 2015). Das Training erfolgt mit Hilfe des Stochastic-Gradient-Decent-Verfahrens unter Verwendung einer polynomialen Reduktion der Lernrate (Chen et al. 2017b) und einer Batchgröße von 2. Die Verlustfunktion bildet sich aus der Binary-Cross-Entropy und dem Dice-Score (Raman 1998).

Von 53 internationalen Segmentierungs-Challenges konnten selbstkonfigurierte nnUNet-Modelle in 33 Challenges den ersten Rang belegen und performten in den restlichen Challenges auf Augenhöhe mit den jeweiligen erstplatzierten Modellen, obwohl diese speziell auf die entsprechende Challenge zugeschnitten waren (Isensee et al. 2021).

2.2.1.2 Lernprozess

In den folgenden Kapiteln wird die Methodik beschrieben, anhand derer Artificial-Neural-Networks verschiedenste Aufgaben erlernen können. Dafür wird zunächst das Backpropagation-Verfahren umrissen. Anschließend werden die gängigsten Verlustfunktionen, die im Rahmen des Supervised-Learnings verwendet werden, und der Prozess der Modell-Optimierung beschrieben.

2.2.1.2.1 Backpropagation

Im Rahmen des Supervised-Learnings lernen Artificial-Neural-Networks in der Trainingsphase mittels Backpropagation. Zunächst müssen dafür die Eingabedaten durch das Netz propagiert werden. Nach erfolgter Propagation wird die Ausgabe des Netzwerkes mit der vordefinierten Ground Truth verglichen. Dazu wird ein Fehler durch eine vordefinierte Verlustfunktion bestimmt. Dieser Fehler wird anschließend mit dem Output-Layer beginnend zurück durch das Netzwerk propagiert, in dem die erlernbaren Parameter, abhängig von ihrem Einfluss auf den Fehler, optimiert werden.

Zu den erlernbaren Parametern gehören die Gewichtungen und Bias, die je nach Netzwerktyp unterschiedliche Funktionen erfüllen. Während bei einem Multilayer-Perzeptron die Gewichtungen und Bias die Verbindungen zwischen den einzelnen Perzeptronen definieren, bilden sie bei einem FCN die einzelnen Matrixelemente der Convolutional-Kernel. Die erlernbaren Parameter werden mittels Gradientenabstiegsverfahren optimiert. Zu den am häufigsten verwendeten Optimizer-Algorithmen zählen dabei der Stochastic-Gradient-Decent-Optimizer und der Adam-Optimizer (Kingma & Ba 2015).

2.2.1.2.2 Verlustfunktionen

Die Aktivierungsfunktion des Output-Layers und die verwendete Verlustfunktion werden dabei in Abhängigkeit von der zu lösenden Aufgabe gewählt. Für Regressionsaufgaben wird als Aktivierungsfunktion des Output-Layers häufig eine lineare Aktivierung gewählt, da diese anders als die ReLU-Aktivierung auch negative Ausgaben abbilden kann. Als Verlustfunktion wird häufig der Mean-Squared-Error (MSE), der Mean-Absolute-Error (MAE) oder der Huber-Loss verwendet. Diese sind folgend definiert, wobei y_i die i te Ausgabe mit der dazugehörigen Ground Truth \hat{y}_i und δ einen festgelegten Schwellenwert repräsentieren:

$$\text{Mean-Squared-Error: } L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\text{Mean-Absolute-Error: } L = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$\text{Huber-Loss: } L = \begin{cases} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (y_i - \hat{y}_i)^2 & \text{if } |y_i - \hat{y}_i| \leq \delta \\ \frac{1}{n} \sum_{i=1}^n \delta \left(|y_i - \hat{y}_i| - \frac{1}{2} \delta \right) & \text{if } |y_i - \hat{y}_i| > \delta \end{cases}$$

Für binäre Klassifikationsaufgaben wird hingegen die Sigmoidfunktion als Aktivierungsfunktion des Output-Layers verwendet. Als Verlustfunktion wird komplementär die Binary-Cross-Entropy (BCE) gewählt. Diese wird folgend definiert, wobei y_1 die Ausgabe und \hat{y}_1 die dazugehörige binäre Ground Truth repräsentieren:

$$\text{Binary-Cross-Entropy-Loss: } L = -y_1 \cdot \log \hat{y}_1 - (1 - y_1) \cdot \log(1 - \hat{y}_1)$$

Für mehrklassige Klassifikationsaufgaben wird die Softmax-Funktion als Aktivierungsfunktion des Output-Layers zusammen mit der Cross-Entropy als Verlustfunktion verwendet. Diese sind wie folgt definiert, wobei x_i die Eingabe des i ten Perzeptrons des Output-Layers, x_j die Eingabe des j ten Perzeptrons des Output-Layers, n die Anzahl der Perzeptronen des Output-Layers, y_i die i te Ausgabe und \hat{y}_i die dazugehörige Ground Truth repräsentieren:

$$\text{Softmax-Funktion: } \varphi(y_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$$

$$\text{Cross-Entropy-Loss: } L = -\sum_{i=1}^n \hat{y}_i \cdot \log y_i$$

Segmentierungsaufgaben können als voxelweise Klassifikationsaufgabe betrachtet werden. Es existieren viele verschiedene Metriken, mit denen sich die Qualität einer Segmentierung quantifizieren lässt. Zu den gängigsten Evaluationsmetriken gehören unter anderem der Dice-Score, die Hausdorff-Distance oder die Average-Symmetric-Surface-Distance. Der Dice-Score wird dabei auch als F1-Score bezeichnet. Anstatt diese Metriken während des Trainings eines Modells indirekt über den Cross-Entropy-Loss zu minimieren, können der Dice-Score und die Hausdorff-Distance auch direkt als Verlustfunktion verwendet werden (Milletari et al. 2016; Karimi & Salcudean 2020). Der Dice-Score-Loss zweier binarisierter Volumina definiert sich dabei wie folgt:

$$\text{Dice-Score-Loss: } L = 1 - \frac{2 TP}{2 TP + FP + FN} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N \hat{y}_i^2},$$

wobei TP die richtig-positiv klassifizierten Voxel, FP die falsch-positiv klassifizierten Voxel, FN die falsch-negativ klassifizierten Voxel und N die Anzahl der Voxel des gesamten Prädiktionsvolumens definieren. y_i ist hierbei der

prädizierte Wert des Voxels und \hat{y}_i der entsprechende Wert der Ground Truth (Milletari et al. 2016).

Der F-Beta-Score bezeichnet eine allgemeinere Formulierung des F1-Scores, bei dem die Gewichtung falsch positiver und falsch negativer Klassifikationen in Abhängigkeit des Beta-Wertes gewählt werden kann. Der F-Beta-Loss definiert sich dabei wie folgt:

$$\text{F-Beta-Loss: } L = 1 - \frac{(1+\beta^2)TP}{(1+\beta^2)TP+FP+\beta^2FN} = 1 - \frac{(1+\beta^2) \cdot \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N \beta^2 y_i^2 + \sum_{i=1}^N \hat{y}_i^2}$$

2.2.1.2.3 Optimierung der Hyperparameter und Evaluation

Für das Training und die anschließende Evaluation eines Deep-Learning-Modells müssen die verfügbaren Daten in Testfälle sowie Trainings- und Validierungsfälle aufgeteilt werden. Während der Trainingsphase trainiert und optimiert man das Modell anhand der Trainings- und Validierungsfälle. In dieser Phase können verschiedene Hyperparameter optimiert werden. Zu diesen optimierbaren Hyperparametern gehören zum einen Parameter, die sich aus dem Aufbau des Netzwerkes ergeben, wie zum Beispiel die Layer-Abfolge und die Netzwerktiefe, die Anzahl der verwendeten Filter, die Verwendung von Regularisierungsmechanismen, die Wahl der netzwerkinternen Normalisierung oder die Wahl der verwendeten Aktivierungsfunktionen. Zum anderen gibt es Parameter, die die Backpropagation direkt beeinflussen, wie zum Beispiel die Wahl des Optimizers und der initialen Lernrate, die Anzahl an Epochen, die Größe der Batchsize und die verwendete Verlustfunktion.

All diese optimierbaren Hyperparameter können dabei nicht nach festdefinierten Regeln bestimmt werden, sondern müssen unabhängig von den Testfällen explorativ ermittelt werden. Um die Hyperparameter unabhängig von den Testfällen zu optimieren, wird das Kreuzvalidierungsverfahren angewandt. Die für das Training und die Validierung verfügbaren Fälle werden dabei mehrfach in Trainings- und Validierungsfälle unterteilt, sodass jeder dieser Fälle einmal den Validierungsfällen zugeordnet wurde. Die verschiedenen Aufteilungen werden dabei als Folds bezeichnet. Für jede Modellvariation wird für jeden Fold ein separates Modell trainiert. Die Validierung der Modellvariationen erfolgt anschließend anhand der gemittelten Ergebnisse der Validierungsfälle. Anhand der so ausgewählten Modellvariation können anschließend die Testfälle

ausgewertet werden. Dazu werden die Testfälle mittels der separat trainierten Fold-Modelle prädiziert und diese Prädiktionen anschließend gemittelt. Alternativ kann ein finales Modell mit Hilfe aller für das Training und die Validierung verfügbaren Fälle mit den optimierten Hyperparametern trainiert und zur Prädiktion der Testfälle verwendet werden. Zudem ist es möglich die Aufteilung des Datensatzes in Testfälle und Trainings- und Validierungsfälle zu iterieren, um so jeden Fall des Datensatzes einmal in den Testfällen enthalten zu haben (verschachtelte Kreuzvalidierung). Die Aufteilung in Trainings-, Validierungs- und Testfälle folgt dabei jedoch auch keinen festdefinierten Regeln und die optimale Aufteilung kann sich je nach Aufgabe, verwendetem Modell und vorhandenen Daten unterscheiden. Im optimalen Fall erfolgt die Evaluation anhand eines unabhängigen Datensatzes.

2.2.2 Deep-Learning-basierte Prädiktion der Infarktausbreitung

Die Prädiktion der Infarktausbreitung eines ischämischen Schlaganfalls kann als Variation einer klassischen Segmentierungsaufgabe betrachtet werden. Anhand der Bilddaten einer initialen Bildgebung soll dabei nicht das Läsionsvolumen des initialen Zeitpunktes segmentiert, sondern die Ausbreitung des Infarktes, die zu einem späteren Zeitpunkt im Verlauf erfasst wurde, prädiziert werden. Im Rahmen Deep-Learning-basierter Modelle werden dazu vorwiegend FCNs verwendet. Als Eingabedaten dienen meist Bilddaten, die während der Hyperakutphase aufgenommen wurden. Als Ground Truth werden Segmentierungen des Infarktgebiets verwendet, die auf im Verlauf erhobenen Bilddaten basierend erstellt werden. Innerhalb der Trainingsphase erlernt das Netzwerk so Bildmerkmale der initialen Bildgebung zu extrahieren, mit denen sich die entsprechende Infarktausbreitung prädizieren lässt.

2.2.2.1 Bisherige Studien

Die Infarktausbreitung nach einem ischämischen Schlaganfall wurde schon in verschiedensten Studien mittels Deep-Learning-basierter Modelle prädiziert. Die bisherigen Kontributionen sind in Tabelle 6 zusammenfassend dargestellt. Es ist anzumerken, dass ausschließlich die Studien beachtet wurden, die Convolutional-Neural-Networks verwendeten. Studien, deren Prädiktion auf Multilayer-

Perzeptronen basierte, wurden nicht berücksichtigt (Huang et al. 2010; Bagher-Ebadian et al. 2011; Kasasbeh et al. 2019).

Tabelle 6: Übersicht über bisherige Studien

Referenz	Netzwerk-Architektur	Datensatz	Eingabedaten:	Ground Truth	Ergebnis
(Stier et al. 2015)	2D-CNN	19 Fälle	MR: Tmax 2D-Patches	MR: FLAIR nach 4 Tagen	Mean-Accuracy: 85.3%
(Choi et al. 2016)	3D-Multiscale-UNet (UNet)	ISLES-2016-Datensatz: 30 Training 19 Test	UNet + FCN: MR: ADC, CBF, CBV, MTT, TTP, Tmax 3D-Patches	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: UNet: 0.31 Ensemble 1: 0.3 Ensemble 2 : 0.31
	3D-Multiscale-UNet + 3D-FCN (Ensemble)		Nur FCN: Klinische Informationen: Time-Since-Stroke, Time-To-Treatment, TICI-Score		
(Lucas et al. 2018b)	2D-UNet mit Short-Skip-Connections und Multi-Loss	ISLES-2017-Datensatz: 43 Training 32 Test	MR: ADC, CBF, CBV, MTT, TTP, Tmax 2D-Slices	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: 0.29
(Lucas et al. 2018a)	3D-UNet (UNet-2in)	TRAVE-STROKE-Datensatz: 29 Fälle	CT: CBV,TTD 3D-Patches	CT nach Thrombektomie	Dice-Score: Cascade-Network: 0.43 UNet-2in: 0.34 UNet-4in: 0.22 Individuelles Läsionswachstum in Abhängigkeit der Time-Onset-To-Imaging und der Time-Imaging-To-Treatment
	3D-UNet (UNet-4in)		CT: CBV,TTD Klinische Informationen: TSS, Time-Imaging-To-Treatment 3D-Patches	CT nach Thrombektomie	
	3D-UNet und Convolutional-Auto-Encoder (Cascade-Network)		UNet: CT: CBV,TTD 3D-Patches	Initiale CT: Segmentierung Infarktkern und Penumbra	
			CAE: 3D-UNet Segmentierungen Klinische Informationen: Time-Onset-To-Imaging, Time-Imaging-To-Treatment	CT nach Thrombektomie	
(Pinto et al. 2018)	2D-UNet mit Gated-Recurrent-Unit, Custom-Verlustfunktion und Integration einer TICI-Feature-Map	ISLES-2017-Datensatz: 43 Trainingsfälle 32 Testfälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax 3D-Patches Klinische Informationen: TICI-Score	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: UNet+TICI: 0.29 UNet ohne TICI: 0.24 Individuelles Läsionswachstum in Abhängigkeit des TICI-Scores

Referenz	Netzwerk-Architektur	Datensatz	Eingabedaten:	Ground Truth	Ergebnis
(Robben et al. 2020)	3D-FCN	MR-CLEAN-Datensatz: 188 Fälle	CT: 4D-Concentration-Curve 3D-Patches in zwei räumlichen Ausdehnungen Klinische Informationen: Time-To-CTP, Time-CTP-To-End-Thrombectomy, mTICI-Score, Reperfusion nach 24 Stunden in CTA Arterial-Input-Function	CT: Nativ nach 24 Stunden oder 5 Tagen	Dice-Score: 0.48 Ablationsstudie: MAE: Läsionsvolumen: Ohne Time-CTP-To-End-Thrombectomy: +12% mit p<0.005
(Nielsen et al. 2018)	2D-CNN (CNN deep und CNN shallow)	I-KNOW- und Remote-Ischemic-Preconditioning-Datensatz: 187: Lyse-Fälle 158: Training 29 Test 35: ohne Therapie	MR: DWI, ADC, FLAIR, CBF, CBV, MTT, CMRO2, rTTH, Delay 2D-Patches Training durch Lyse-Fälle	MR: FLAIR nach 30 Tagen	AUC: CNN deep: 0.88 CNN deep -rtPA: 0.85 CNN shallow: 0.85 CNN Tmax: 0.72 ADC threshold: 0.66 Gemittelttes Infarktvolumen: 29 Lyse-Testfälle: CNN deep: 16.44 ml CNN deep -rtpa: 29.40 ml
	2D-CNN (CNN deep-rtpa)		Nachtraining durch Fälle ohne Therapie		
	2D-CNN (CNN Tmax) (Stier et al. 2015)		Training durch Lyse-Fälle (Tmax: 2D-Patches)		
(Ho et al. 2019)	2D- und 3D-CNN	48 Fälle	MR: CBF,CBV, Tmax, MTT, TTP (global AIF) 2D- oder 3D-Patches	MR: FLAIR nach 3-7 Tagen	AUC: 2.5D-CNN: 0.871 3D-CNN: 0.799 2D-CNN: 0.783
	2.5D-CNN mit initialer Convolution über die Zeitachse		MR: 4D-Concentration-Curve 2D-Patches mit zusätzlicher Zeit-Dimension und kontralateralem Patch		
(Hu et al. 2020)	3D-UNet mit Refinement-Modul	ISLES-2017-Datensatz: 43 Trainingsfälle 32 Testfälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax 3D-Patches	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: 0.3
(Kumar et al. 2020)	3D-UNet mit Dense-Layern zwischen Encoder und Decoder	ISLES-2017-Datensatz: 43 Trainingsfälle 32 Testfälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax 3D-Patches selektiert durch CNN	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: 0.28

Referenz	Netzwerk-Architektur	Datensatz	Eingabedaten:	Ground Truth	Ergebnis
(Yu et al. 2020)	3D-UNet mit Attention-Gating	iCAS- und DEFUSE-2-Datensatz: 182 Fälle: 35: Geringe Reperfusion 67: Große Reperfusion	MR: DWI, ADC, CBF, CBV, MTT, Tmax, ADC Maske, Tmax Maske 3D-Patches	MR: FLAIR nach 3-7 Tagen	Dice-Score: 0.53 Subgruppen-Analyse: Große Reperfusion: UNet: 0.48 ADC-Threshold: 0.45 Geringe Reperfusion: UNet: 0.58 Tmax-ADC-Threshold: 0.55
(Debs et al. 2021)	2D-Multi-Modal-UNet (UNet General)	HIBISCUS-STROKE-Datensatz und I-KNOW-Datensatz: 109 Fälle: 74: Reperfusion 35: ohne Reperfusion	MR: DWI, ADC, CBF, CBV, Tmax 2D-Slices Training anhand aller 109 Fälle :	MR: FLAIR nach 6 oder 30 Tagen	Subgruppen-Analyse: Dice-Score: Fälle mit Reperfusion: UNet Reperused: 0.44 UNet General: 0.43 UNet Non-Reperused: 0.35 DWI-Threshold: 0.41 Fälle ohne Reperfusion: UNet Non-Reperused: 0.47 UNet General: 0.44 UNet Reperused: 0.44 DWI-PWI-Threshold: 0.26 Ablationsstudie: Fälle mit Reperfusion: Alle Modalitäten: 0.44 DWI: 0.42 Fälle ohne Reperfusion: Alle Modalitäten: 0.47 DWI: 0.45
	2D-Multi-Modal-UNet (UNet Reperused)		Training anhand der Fälle mit Reperfusion		
	2D-Multi-Modal-UNet (UNet Non-Reperused)		Training anhand der Fälle ohne Reperfusion		
(Hokkinen et al. 2021b)	3D-CNN	83 Fälle	CT: CTA 3D-Patches mit kontralateralem Patch	CT: CTP um 12-36 Stunden	Gemittelttes Läsionsvolumen: Ground Truth: 36 ml CNN: 23 ml CBF-Threshold: 56 ml Korrelation: CNN: r=0.83 CBF-Threshold: r=0.91
(Hokkinen et al. 2021a)	Vortrainiertes 3D-CNN (Hokkinen et al. 2021b)	89 Fälle	CT: CTA 3D-Patches mit kontralateralem Patch	CT: Nativ oder MR: DWI innerhalb von 5 Tagen	Gemittelttes Läsionsvolumen: Ground Truth: 36 ml CNN: 54 ml CBF-Threshold: 28 ml Korrelation: Läsionsvolumen CNN: r=0.67 CBF-Threshold: r=0.82
(Pinto et al. 2021a)	2D-UNet (OPW-SM)	ISLES-2017-Datensatz: 43 Trainingsfälle 32 Testfälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax 2D-Patches	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: TPW: 0.31 OPW-SM: 0.3
	2D-Multipath-UNet mit gesondertem Pfad für 4D-Perfusions-MRT-Daten (TPW)		MR: ADC, CBF, CBV, MTT, TTP, Tmax + 4D-Concentration-Curve (vorverarbeitet) 2D-Patches		

Referenz	Netzwerk-Architektur	Datensatz	Eingabedaten:	Ground Truth	Ergebnis
(Pinto et al. 2021b)	zwei 3D-Restricted-Boltzmann-Machines + 2D-UNet mit Gated-Recurrent-Unit	ISLES-2017-Datensatz: 43 Trainingsfälle 32 Testfälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax UNet: 2D-Patches RBMs: 3D-Patches	MR: T2w oder FLAIR nach 90 Tagen	Dice-Score: 3D-RBMs+2D-UNet+GRU: 0.38 Ablationsstudie: 3D-RBMs+2D-UNet: 0.32 2D-UNet+GRU: 0.3 2D-UNet: 0.3
(Vupputuri et al. 2021)	2D-UNet mit Multi-Path-Attention-Convolution	ISLES-2017-Datensatz: 55 Fälle	MR: ADC, CBF, CBV, MTT, TTP, Tmax UNet: 2D-Slices	MR: T2w oder FLAIR nach 90 Tagen	Nicht offiziell gelistet Nur 55 von 75 Fällen verwendet Dice-Score: 0.47
(Yu et al. 2021)	3D-UNet mit Attention-Gating	iCAS-, DEFUSE- und DEFUSE-2-Datensatz: 237 Fälle	MR: DWI, ADC, CBF, CBV, MTT, Tmax, ADC Maske, Tmax Maske 3D-Patches	MR: DWI oder FLAIR nach 3-7 Tagen oder 30 Tagen	Dice-Score: Vortrainiertes UNet: 0.57 Korrelation: Läsionsvolumen Vortrainiertes UNet: 0.756 UNet: 0.657 ADC-Threshold: 0.625
	Vortrainiertes 3D-UNet mit Attention-Gating		Zunächst Training anhand von Fällen mit mittlere Reperfusion oder unbekannter Reperfusion, anschließend Training anhand von Fällen mit großer Reperfusion		
	3D-UNet mit Attention-Gating		Training anhand von Fällen mit großer Reperfusion		
(Amador et al. 2021)	2D Multipath-UNet mit Temporal-Convolutional-Network zwischen Encoder und Decoder	ProvelIT- und ERASER-Datensatz sowie weitere Fälle	CT: 4D-Concentration-Curve 2D-Slices mit 8, 16, oder 32 Zeitpunkten	CT: Nativ oder MR	Dice-Score: UNet mit TCN (32 Zp.): 0.3361 UNet mit TCN (16 Zp.): 0.2836 UNet mit TCN (8 Zp.): 0.2564 UNet: 0.2387
	UNet	187 Fälle	CT: CBF, CBV, MTT, Tmax und baseline CTP 2D-Slices		
(Nazari-Farsani et al. 2022)	3D-UNet mit Attention-Gating	iCAS-, DEFUSE- und DEFUSE-2- und UCLA-Datensatz: 445 Fälle 180: geringe Reperfusion 81: große Reperfusion	MR: ADC, DWI, DWI Maske	MR: DWI oder FLAIR nach 3-7 Tagen oder 30 Tagen	Dice-Score: UNet: 0.5 ADC Threshold: 0.18 Subgruppen-Analyse: Große Reperfusion: Dice-Score: 0.46 Geringe Reperfusion: Dice-Score: 0.56

Referenz	Netzwerk-Architektur	Datensatz	Eingabedaten:	Ground Truth	Ergebnis
(Winder et al. 2022)	2D-UNet (Param-UNet)	ERASER-Datensatz sowie weitere Fälle: 145 Fälle	CT: CBF, CBV, MTT, Tmax 2D-Slices	CT: Nativ innerhalb von 3 Tagen	Dice-Score: CTC-Causal: 0.296 Param-UNet: 0.287 RC-Causal: 0.286 RC-Simple: 0.276 Tmax-Threshold: 0.249
	2D-UNet mit Simple-Temporal-Processing (RC-Simple)		CT: 4D-Residual-Funktion		
	2D-UNet mit Causal-Temporal-Processing (RC-Causal)		2D-Slices mit 32 Zeitpunkten CT: 4D-Residual-Funktion		
	2D-UNet mit Causal-Temporal-Processing (CTC-Causal)		2D-Slices mit 32 Zeitpunkten CT: 4D-Concentration-Curve		
(Wouters et al. 2022)	3D-FCN (Robben et al. 2020)	MR-CLEAN-Datensatz: 127 Fälle	MR: 4D-Concentration-Curve 3D-Patches in zwei räumlichen Ausdehnungen	CT: Nativ nach 24 Stunden oder 5 Tagen	MAE: Läsionsvolumen Fälle: mTICI 0 und 2b +3 MR CLEAN: 3D-FCN: 34.5 ml PI-Threshold: 52.4 ml CRISP: 3D-FCN: 41.2 ml PI-Threshold: 52.4 ml Individuelles Läsionswachstum in Abhängigkeit von Time-CTP-To-End-Thrombectomy und mTICI-Score
		External-Testing: CRISP-Datensatz: 104 Fälle	Klinische Informationen: Time-To-CTP, Time-CTP-To-End-Thrombectomy, mTICI-Score, Reperfusion nach 24 Stunden in CTA Arterial-Input-Function	MR: FLAIR nach 5 Tagen	

Neben den allgemeinen methodischen Modell-Variationen lassen sich weitere Erkenntnisse aus den bisherigen Studien ableiten. Wie bereits in Kapitel 2.1.4.3 beschrieben, zeigt sich, dass sich die Aufnahmezeitpunkte und die verwendeten Aufnahmeverfahren der verwendeten Ground Truth erheblich voneinander unterscheiden. In einigen Studien wurden die Prädiktionsergebnisse mit Threshold-basierten Prädiktionen verglichen (Nielsen et al. 2018; Yu et al. 2020; Debs et al. 2021; Hokkinen et al. 2021a, 2021b; Yu et al. 2021; Nazari-Farsani et al. 2022; Wouters et al. 2022; Winder et al. 2022). Als Eingabedaten wurden zumeist Bilddaten verwendet, die sowohl den Infarktkern als auch die Penumbra abbilden können. Eine Studie verwendete ausschließlich DWI-basierte Eingabedaten (Nazari-Farsani et al. 2022). In zwei anderen Studien wurden ausschließlich CTA-Maps zur Prädiktion verwendet (Hokkinen et al. 2021a, 2021b). Eine Anzahl an Studien verwendete native oder präprozessierte 4D-Perfusionsdaten, um die Ausbreitung des Infarktes zu präzisieren (Robben et al.

2020; Ho et al. 2019; Pinto et al. 2021a; Amador et al. 2021; Winder et al. 2022). Einige Studien untersuchten einen möglichen Therapie-Effekt innerhalb der dazugehörigen Kohorten. In zwei Studien wurden dafür Modelle anhand unterschiedlicher Trainingsdaten trainiert (Nielsen et al. 2018; Debs et al. 2021). In drei Studien erfolgte die Evaluation anhand von Subgruppen, die sich durch den Grad der festgestellten Reperfusion voneinander unterschieden (Debs et al. 2021; Yu et al. 2020; Nazari-Farsani et al. 2022). Nur in fünf Studien prozessierten die Modelle neben Bilddaten auch additive klinische Informationen (Choi et al. 2016; Lucas et al. 2018a; Pinto et al. 2018; Robben et al. 2020; Wouters et al. 2022).

2.2.2.2 Integration klinischer Informationen

In den ISLES-Challenges 2016 und 2017 sollte die Infarktausbreitung anhand von MR-Bilddaten aus der Hyperakutphase und klinischen Informationen prädiziert werden (Winzeck et al. 2018).

Der ISLES-2016-Datensatz besteht aus 30 Trainingsfällen und 19 Testfällen. Als Ground Truth dienen Segmentierungen, die anhand nach 90 Tagen erhobener T2w- oder FLAIR-Maps erstellt wurden. Als Bilddaten sind ADC-, CBF-, CBV-, MTT-, TTP- sowie Tmax-Maps vorhanden. Als klinische Informationen sind die Time-Since-Stroke, die Time-To-Treatment und der nach Thrombektomie erhobene TICI-Score verfügbar (Winzeck et al. 2018). Die ISLES-Challenge 2016 konnte von Choi et al. mit einem Rank-Score von 3.42 und einem Dice-Score von 0.31 gewonnen werden. Dazu verwendeten sie ein 3D-Multiscale-UNet, um die Ausbreitung des Infarktes anhand von 3D-Patches zu prädizieren (Choi et al. 2016). Der Multiscale-Ansatz wurde verwendet, um verschiedene räumliche Zusammenhänge abzubilden (Kamnitsas et al. 2017). Als Eingabedaten wurden dafür lokale Patches und auf dieselbe Größe downgesampelte globale Patches, die weitere räumliche Zusammenhänge erfassten, verwendet. Jede Patchart wurde durch ein eigenes UNet prozessiert und die Ausgabe dieser Netzwerke wurden in einem anschließenden Fully-Convolutional-Abschnitt zusammengeführt. Die verfügbaren klinischen Informationen wurden nicht verwendet. Neben diesem Ansatz reichten die Autor:innen auch zwei Ensemble-Modelle ein, die jedoch schlechter performten als das alleinige 3D-Multiscale-UNet. Für die Ensemble-Modelle wurden die Prädiktionen des 3D-Multiscale-UNets mit den Prädiktionen eines FCN in Abhängigkeit von einer Gewichtung verrechnet. Das FCN

klassifizierte voxelweise, ob die Eingabepatches Läsionsanteile abbildeten. Auch das FCN bestand dabei aus zwei separaten Pfaden. Während der eine Pfad klassifizierte, ob ein Patch eine Läsion enthielt, klassifizierte der andere Pfad, ob der zentrale Voxel des Patches zur Läsion zugehörig war. Der Output beider Pfade wurde mit den klinischen Eingabedaten zusammengeführt und zunächst durch ein Multilayer-Perzeptron transformiert, welches nach dem Training in Convolutional-Layer umgewandelt wurde (siehe Abbildung 15). Es wurde nicht untersucht, inwiefern die Prädiktionsergebnisse durch die Integration der klinischen Daten beeinflusst wurde (Choi et al. 2016).

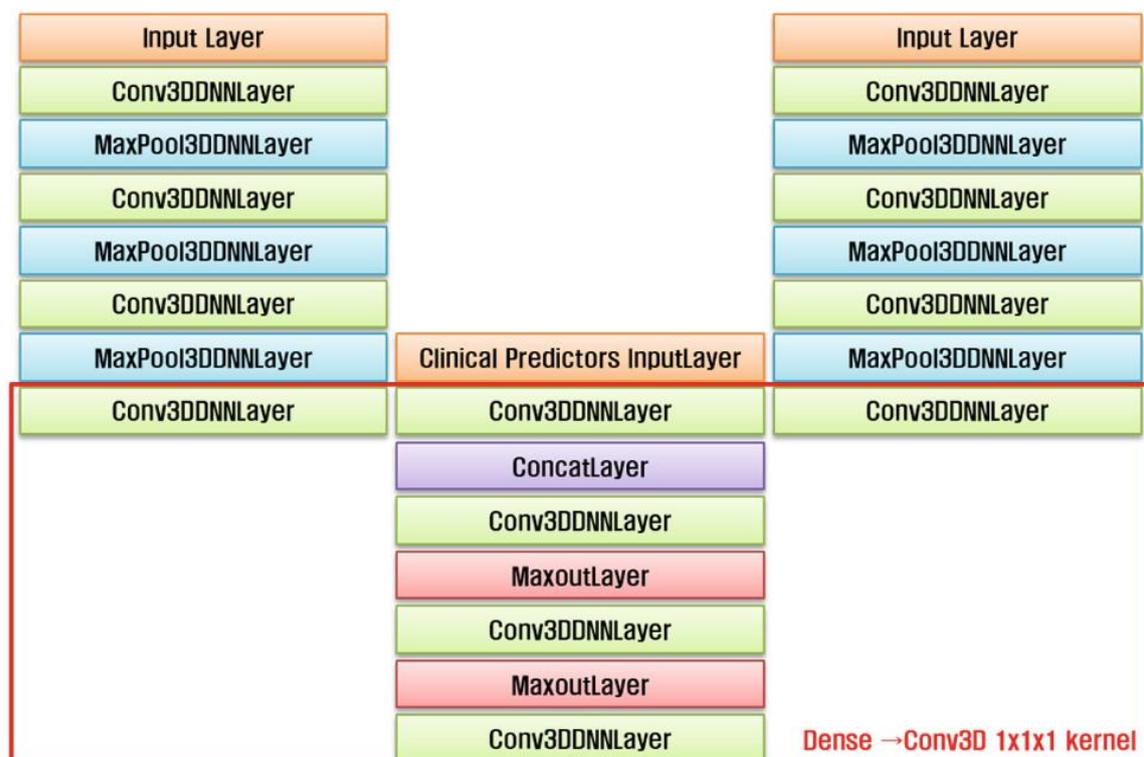


Abbildung 15: Fully-Convolutional-Network (Choi et al. 2016)

Der ISLES-2017-Datensatz besteht aus 43 Trainingsfällen und 32 Testfällen. Als Ground Truth dienen Segmentierungen, die anhand nach 90 Tagen erhobener T2w- oder FLAIR-Maps erstellt wurden. Als Bilddaten sind ADC-, CBF-, CBV-, MTT-, TTP- sowie Tmax-Maps vorhanden. Als klinische Informationen sind die Time-Since-Stroke, die Time-To-Treatment und der nach Thrombektomie erhobene TICI-Score verfügbar (Winzeck et al. 2018). Pinto et al. entwickelten anhand des ISLES-2017-Datensatzes ein 2D-UNet mit Gated-Recurrent-Unit, welches den TICI-Score als klinischen Parameter zusätzlich zu den Bilddaten verarbeitete (Pinto et al. 2018). Die Gated-Recurrent-Unit wurde verwendet, um

lokale und globale Bildzusammenhänge zu erfassen (Cho et al. 2014). Die Integration der TICI-Scores erfolgte über zwei Mechanismen. Zum einen wurde der TICI-Score während der Trainingsphase zur Berechnung einer Custom-Verlustfunktion verwendet, so dass bei hohen TICI-Scores falsch-negativ klassifizierte Voxel und bei niedrigen TICI-Scores falsch-positiv klassifizierte Voxel bevorzugt wurden. Die verwendete Verlustfunktion entsprach dabei dem in Kapitel 2.2.1.2.2 beschriebenen F-Beta-Loss, für die der Beta-Wert für jedes Sample in Abhängigkeit des TICI-Scores gewählt wurde. Dabei wurde der Beta-Wert wie folgt bestimmt:

$$\beta = \begin{cases} 2 & \text{if } TICI \in \{0,1\} \\ 1 & \text{if } TICI \in \{2,2a,2b\} \\ 0.5 & \text{if } TICI = 3 \end{cases}$$

Zusätzlich wurde der TICI-Score als Feature-Map mit der Ausgabe der Gated-Recurrent-Units kurz vor dem Output-Layer verbunden (siehe Abbildung 16). Der Dice-Score konnte durch die Integration des TICI-Scores auf diese Weise von 0.24 auf 0.29 verbessert werden. Die Infarktausbreitung eines Falles konnte so zudem in Abhängigkeit verschieden gewählter TICI-Scores prädiziert werden (Pinto et al. 2018).

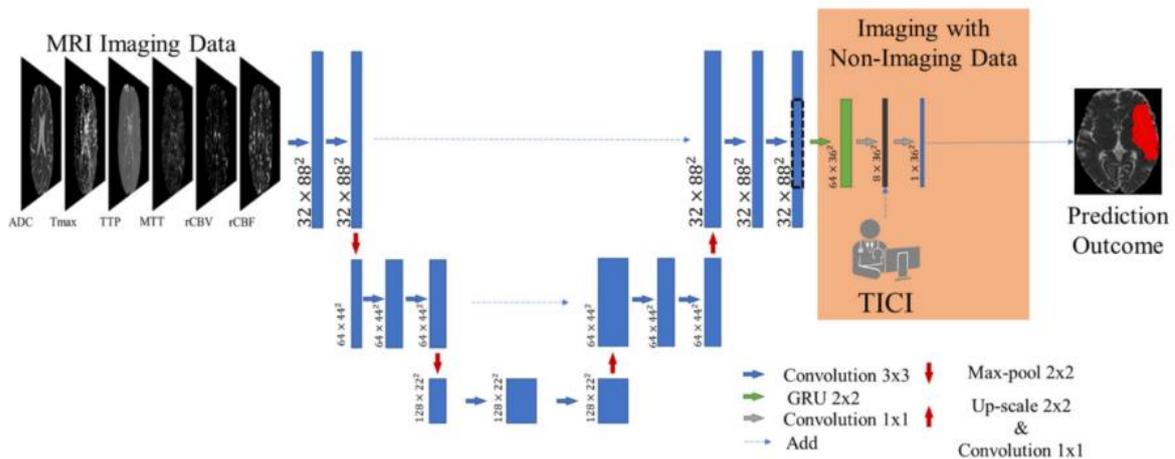


Abbildung 16: UNet mit Integration des TICI-Scores (Pinto et al. 2018)

Lucas et al. verwendeten ein Cascade-Network, bestehend aus einem 3D-UNet mit anschließendem Convolutional-Auto-Encoder (CAE), zur Prädiktion der Infarktausbreitung. Dazu verwendeten sie 29 Fälle aus dem TRAVESTROKE-Datensatz. Der Infarktkern und die Penumbra der initialen Bildgebung wurden anhand von CBV- und Time-To-Drain-Maps segmentiert. Die Aufgabe des UNets war es den Infarktkern und die Penumbra zum initialen Zeitpunkt zu segmentieren.

Als Eingabedaten dienten 3D-Patches der CBV- und Time-To-Drain-Maps. Die Infarktausbreitung wurde anhand von CT-Maps segmentiert, die nach erfolgter Therapie erhoben wurden. Die Aufgabe des Convolutional-Auto-Encoders war es die tatsächliche Infarktausbreitung zu prädizieren. Als Eingabedaten dienten dabei die UNet-Segmentierungen des Infarktkerns und der Penumbra zum initialen Zeitpunkt sowie die klinischen Informationen Time-Onset-To-Imaging und Time-Imaging-To-Treatment. Die Segmentierungen von Infarktkern und Penumbra wurden dafür in den Latent-Space transformiert. Beruhend auf der Annahme, dass sich der Infarktkern mit fortschreitender Zeit auf das Volumen der Penumbra ausweitet, wurde die Ausbreitung des Infarktes durch eine lineare Interpolation, abhängig von den Zeitparametern, im Latent-Space prädiziert und anschließend in den tatsächlichen Raum zurücktransformiert (siehe Abbildung 17). Dabei ist anzumerken, dass die lineare Interpolation im Latent-Space einer nicht linearen Interpolation im tatsächlichen Raum entspricht und dass die Zeitspanne, die vergeht, bis der Infarkt sich vollständig auf die Penumbra ausgebreitet hat, in Abhängigkeit von den verfügbaren Daten auf 10 Stunden festgelegt wurde. Die Formel der linearen Interpolation definierte sich dabei als

$$\hat{y}_i = \hat{y}_c + \eta(\hat{y}_{cp} - \hat{y}_c) \text{ mit } \eta = \frac{t_{Imaging \rightarrow Treatment}}{10 - t_{Onset \rightarrow Imaging}},$$

wobei \hat{y}_i der linear transformierten Infarktausbreitung, \hat{y}_c dem initialen Infarktkern, \hat{y}_{cp} der initialen Penumbra im Latent-Space und $t_{Imaging \rightarrow Treatment}$ der Time-Imaging-To-Treatment sowie $t_{Onset \rightarrow Imaging}$ der Time-Onset-To-Imaging entsprechen (Lucas et al. 2018a).

Das Cascade-Network erreichte einen Dice-Score von 0.43. Zum Vergleich trainierten die Autor:innen zwei UNet-Modelle, die die Infarktausbreitung direkt prädizierten. Das UNet-2in nutzte dabei 3D-Patches der CBV- und Time-To-Drain-Maps als Eingabedaten und erzielte einen Dice-Score von 0.34. Das UNet-4in nutzte zusätzlich 3D-Patches aus den klinischen Informationen, der Time-Onset-To-Imaging und der Time-Imaging-To-Treatment als Eingabedaten und erzielte einen weitaus geringeren Dice-Score von 0.22.

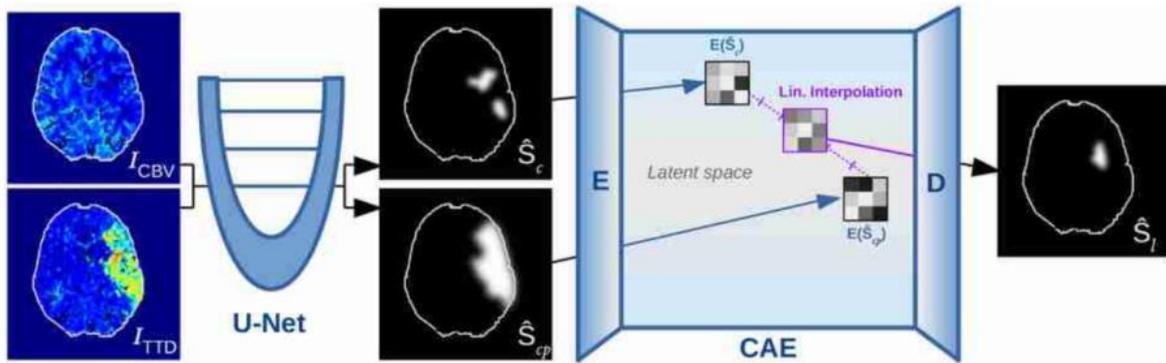


Abbildung 17: Cascade-Network (Lucas et al. 2018a)

Robben et al. trainierten ein 3D-FCN, um die Infarktausbreitung zu präzisieren. Dabei verwendeten sie 188 Fälle aus dem MR-CLEAN-Datensatz. Als Ground Truth dienten Segmentierungen, die anhand von nativen CT-Daten erstellt wurden, welche nach 24 Stunden oder 5 Tagen erhoben wurden. Als Eingabedaten dienten 3D-Patches mit zeitlicher Dimension aus nativen 4D-CT-Perfusion-Daten. Dabei wurden lokale Patches und auf dieselbe Größe downgesampelte globale Patches verwendet. Zudem wurden die Arterial-Input-Funktion sowie weitere klinische Parameter in das Modell integriert. Zu den klinischen Informationen zählten die Time-To-CTP, die Time-CTP-To-End-Thrombectomy, der mTICI-Score sowie der Grad der Reperfusion, welcher nach 24 Stunden durch eine CT-Angiographie quantifiziert wurde. Alle Eingabedaten wurden durch Convolutional-Layer, Upsampling-Layer und Repeat-Operationen in dieselbe Größe transformiert und anschließend vereinigt und durch weitere Convolutional-Layer transformiert (siehe Abbildung 18). Das Modell erzielte einen Dice-Score von 0.48. In einer Ablationsstudie konnte gezeigt werden, dass sich der Dice-Score und der mittlere absolute Fehler ohne die Integration der Time-CTP-To-End-Thrombektomie signifikant verschlechterten (Robben et al. 2020).

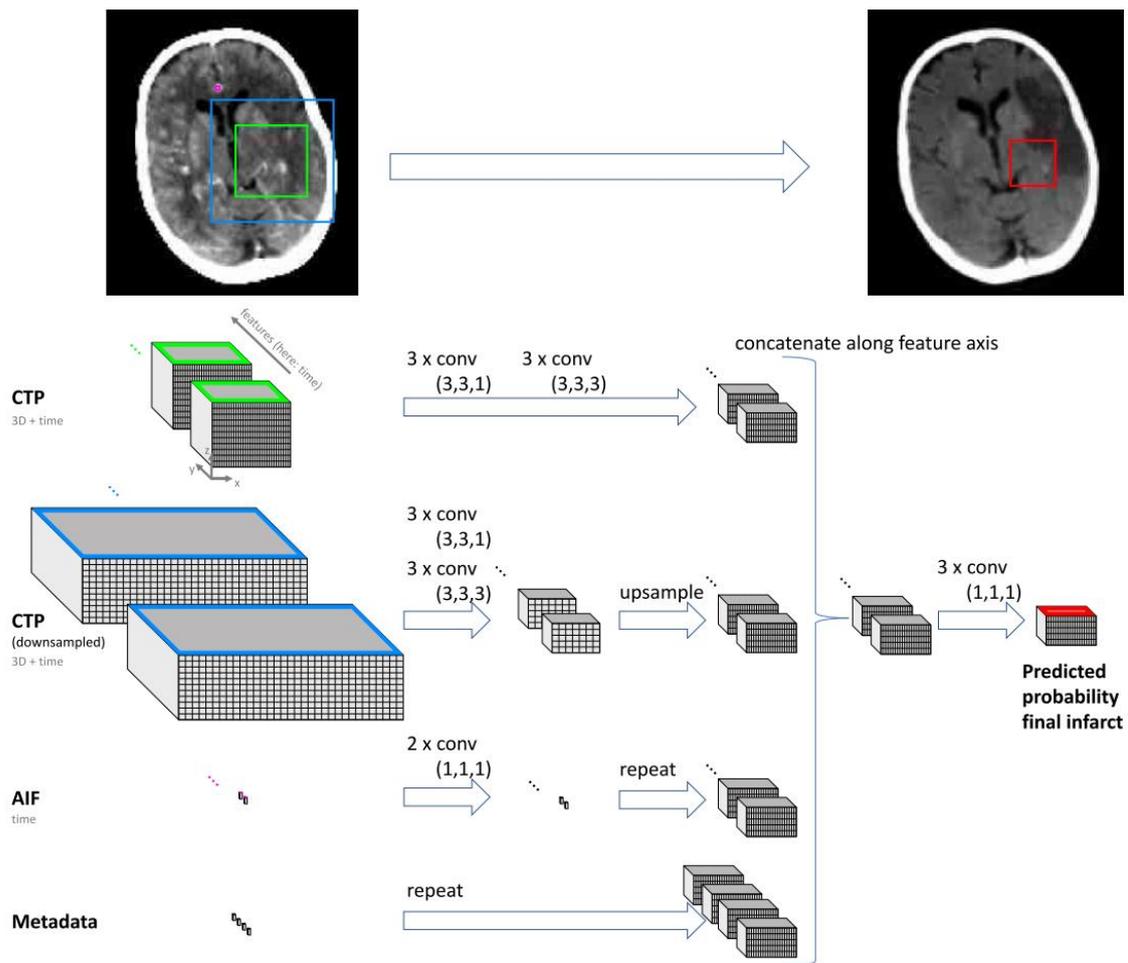


Abbildung 18: Fully-Convolutional-Network (Robben et al. 2020)

In einer weiteren Studie von Wouters et al. wurde die gleich Methodik erneut zur Prädiktion der Infarktausbreitung angewandt. Das Modell wurde anhand von 127 Fällen des MR-CLEAN-Datensatzes trainiert und evaluiert. Zudem wurde das Modell zusätzlich anhand von 101 Fällen des CRISP-Datensatzes evaluiert. Für die Fälle des MR-CLEAN-Datensatzes wurden die Segmentierungen der zu prädizierenden Infarktausbreitung anhand von nativen CT-Daten erstellt, welche nach 24 Stunden oder 5 Tagen erhoben wurden. Für die Fälle des CRISP-Datensatzes wurden die Segmentierungen anhand von FLAIR-Maps erstellt, welche nach 5 Tagen erhoben wurden. Als Eingabedaten dienten, wie zuvor beschrieben, 3D-Patches aus nativen 4D-CT-Perfusion-Daten, die Arterial-Input-Funktion, die Time-To-CTP, die Time-CTP-To-End-Thrombectomy, der mTICI-Score und der Grad der Reperfusion, welcher nach 24 Stunden durch eine CT-Angiographie quantifiziert wurde. Auf diese Weise konnte das Infarktvolumen beider Kohorten besser prädiziert werden als durch Threshold-basierte Verfahren. Zudem konnte die Infarktausbreitung eines Falles in Abhängigkeit der Time-CTP-

To-End-Thrombectomy und verschiedener Reperfusionen prädiziert werden. In einer Subanalyse konnte zudem gezeigt werden, dass der Grad der Kollateralisierung, der durch die Hypoperfusion-Intensity-Ratio erfasst wurde (Guenego et al. 2020), mit dem prädizierten Läsionsvolumen korrelierte. Der Effekt der klinischen Informationen auf die Prädiktionsqualität wurde nicht untersucht (Wouters et al. 2022).

3 Material und Methoden

3.1 Verwendete Daten

Die in dieser Arbeit verwendeten Daten entstammen der WAKE-UP-Studie (Thomalla et al. 2018). Diese Studie wird in Kapitel 3.1.1 beschrieben. Anschließend wird in Kapitel 3.1.2 die Vorverarbeitung der Daten näher erläutert.

3.1.1 WAKE-UP-Studie

In der WAKE-UP-Studie, einer multizentrischen, randomisierten, doppelt verblindeten, Placebo-kontrollierten klinischen Studie konnten Thomalla et al. erstmals zeigen, dass Patient:innen, die einen Schlaganfall während der Nacht erlitten hatten, von einer intravenösen Thrombolyse profitieren konnten. 503 Patient:innen, bei denen der Beginn der neurologischen Symptomatik nicht zu ermitteln war, stattdessen aber ein bestehendes DWI-FLAIR-Mismatch (siehe Kapitel 2.1.3.1) nachgewiesen werden konnte, wurden der Intention-To-Treat-Population zugeordnet. Bei einem Großteil der Patient:innen war der Beginn der neurologischen Symptomatik nicht zu ermitteln, weil sich der Schlaganfall während des Schlafes ereignete. Patient:innen, bei denen eine intrakranielle Hämorrhagie nachgewiesen werden konnte, deren Läsion größer als ein Drittel des Mediagebietes war, die einen NIHSS-Score von mehr als 25 aufwiesen, bei denen eine Thrombektomie geplant war oder die eine Kontraindikation gegen die intravenöse Thrombolyse aufwiesen, wurden nicht in die Studie mit eingeschlossen. Von den 503 Patient:innen der Intention-To-Treat-Population wurden 254 der Thrombolyse-Gruppe und 249 Patient:innen der Placebo-Gruppe zugeordnet. Insgesamt 424 Patient:innen wurden der Per-Protocol-Population zugeordnet, von denen 217 Patient:innen mit Alteplase und 207 Patient:innen mit einem Placebo behandelt wurden (Thomalla et al. 2018).

Die Patient:innen wurden initial zum Zeitpunkt der klinischen Vorstellung und im Verlauf nach 22 bis 36 Stunden, nach 5 bis 9 Tagen und nach 90 Tagen untersucht. Dabei wurde unter anderem der klinische Outcome anhand dem mRS-Score und dem NIHSS-Score erfasst. Zur Beurteilung einer intrakraniellen Blutung und zur Quantifizierung des Infarktolumens wurde zum Zeitpunkt der klinischen Vorstellung und nach 22-36 Stunden eine MRT-Bildgebung durchgeführt. Als primärer Endpunkt wurde der mRS-Score definiert, der 90 Tage nach

Randomisierung erhoben wurde. Dabei wurde ein günstiger Outcome durch einen mRS-Score von 0 oder 1 definiert. Neben einigen anderen klinischen Outcomes wurde zudem das Infarktvolumen im Verlauf, welches anhand der MR-Bildgebung nach 22 bis 36 Stunden nach Randomisierung erhoben wurde, als sekundärer Endpunkt betrachtet (Thomalla et al. 2018).

Es konnte gezeigt werden, dass die Patient:innen der Intention-To-Treat-Population, die einer Therapie mit Alteplase zugeordnet wurden, mit einem p-Wert von 0.02 signifikant häufiger einen günstigen Outcome erreichten. Hinsichtlich des Infarktvolumens im Verlauf konnte jedoch kein Unterschied innerhalb der beiden Subgruppen festgestellt werden. In der Thrombolyse-Gruppe betrug die gemittelte Infarktgröße im Verlauf 3.0 ml [0.8-17.7] und in der Placebo-Gruppe 3.3 ml [1.1-16.6], was einem p-Wert von 0.32 entsprach (Thomalla et al. 2018).

3.1.2 Vorverarbeitung der Bilddaten

Zur Durchführung des Promotionsvorhabens standen insgesamt 483 Fallordner zur Verfügung. Nach der heuristischen Bestimmung der Zielordner für die einzelnen Bildmodalitäten wurden die Bilddaten mit dem dcm2niix-Tool in das Nifti-Format konvertiert (Li et al. 2016). Fehlende ADC-Maps wurden bei vorhandenen DWI-Maps mit Hilfe der FSL-Software generiert (Jenkinson et al. 2012). Die Bilddaten wurden dann mit den Advanced-Normalization-Tools (ANTs) N4-Bias-Field-korrigiert (Avants et al. 2011). Für die Registrierung mittels ANTs wurde der T2w-MNI152-ICBM-2009b-Atlas verwendet (Fonov et al. 2009; Avants et al. 2011). Die Auflösung der hirnbildenden Bereiche des unbearbeiteten T2w-Atlas beträgt $(290 \times 362 \times 310)$ Voxel mit einem Voxelspacing von $(0.5 \text{ mm} \times 0.5 \text{ mm} \times 0.5 \text{ mm})$. Vor der Registrierung wurde der Atlas gesampelt. In den registrierten Bilddaten besitzen die hirnbildenden Bereiche dementsprechend eine Auflösung von $(189 \times 236 \times 35)$ mit einem Voxelspacing von $(0.77 \text{ mm} \times 0.77 \text{ mm} \times 4.43 \text{ mm})$. In einem letzten Schritt wurde in der x- und y-Achse gepaddet und der oberste Slice sowie die untersten beiden Slices in der z-Achse entfernt, damit die finale Auflösung von $(256 \times 256 \times 32)$ in allen Dimensionen Zweierpotenzen entsprach, was bei der Verwendung von Pooling- sowie Convolutional-Layern vorteilhaft ist. Dieses Vorgehen führte zwar zu anisotropen Bilddaten entlang der z-Achse, ermöglichte jedoch unter der Verwendung einer GeForce RTX 3080-Grafikkarte mit 10 GB Grafikspeicher das Training von

Modellen, die die vorverarbeiteten Bilddaten in ihrer ganzen Auflösung verarbeiten konnten.

Für die Erstellung der Transformationsmatrizen wurde zunächst der Schädel in allen Bildmodalitäten mit Hilfe des Brain-Extraction-Tools (BET) entfernt (Smith 2002). Danach wurden die ADC-Maps zunächst auf die FLAIR-Maps des dazugehörigen Zeitpunktes registriert. Die initialen FLAIR-Maps wurden auf die im Verlauf erhobenen FLAIR-Maps registriert. Die im Verlauf erhobenen FLAIR-Maps wurden auf den Atlas registriert. Die verarbeiteten Bilddaten wurden letztendlich durch eine einzelne Transformationsmatrix affin auf den Atlas registriert. Diese Transformationsmatrix berechnete sich durch die nötigen einzelnen Transformationsschritte und war somit abhängig von der entsprechenden Bildmodalität und dem dazugehörigen Aufnahmezeitpunkt. Die ursprünglichen Infarktsegmentierungen wurden dabei entsprechend der ADC-Maps transformiert. Nach erfolgter affiner Registrierung wurden alle Bilddaten und die Läsionsmasken durch eine auf die gehirnabbildenden Anteile des Atlas begrenzte Maske maskiert, da dieses Vorgehen bei einer qualitativen Registrierung dem Skull-Stripping anhand des Brain-Extraction-Tools überlegen ist. Zur Verdeutlichung ist eine schematische Darstellung der einzelnen Registrierungsschritte in Abbildung 19 dargestellt.

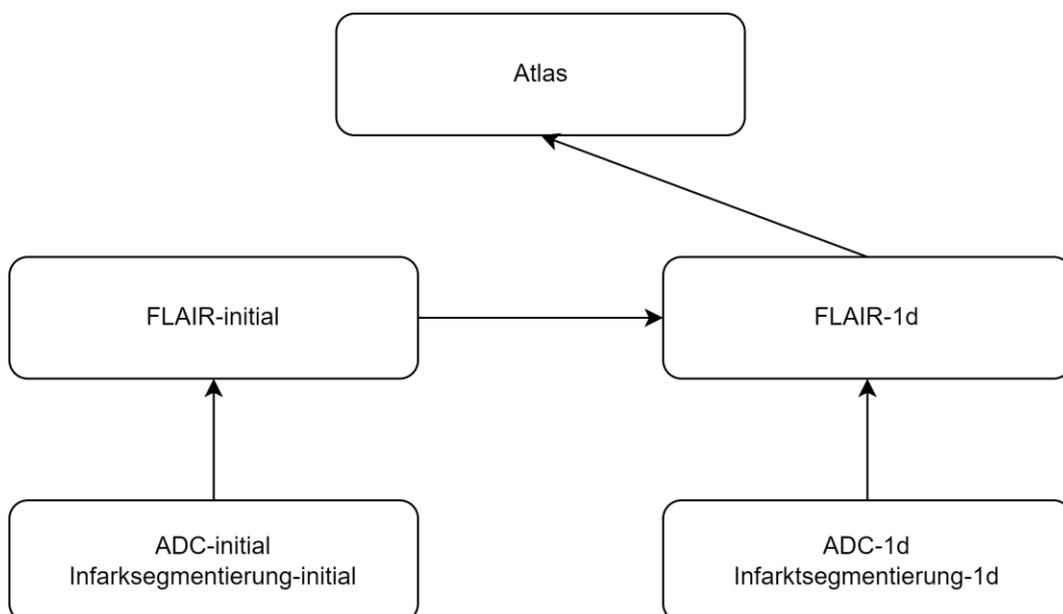


Abbildung 19: Registrierungsschritte

Fälle, die aufgrund von bildtechnischen Ausschlusskriterien nicht zu der Per-Protocol-Population (PPP) gezählt wurden oder deren Bilddaten unvollständig

waren, wurden ausgeschlossen. Erforderlich waren das Vorhandensein von ADC- und FLAIR-Maps sowie die ursprünglichen Segmentierungen des Infarktes der initialen Bildgebung sowie der Bildgebung im Verlauf. Weitere Fälle wurden aufgrund einer fehlerhaften Registrierung oder mangelnder Bildqualität ausgeschlossen. Eine schematische Darstellung der gesamten Vorverarbeitung ist in Abbildung 20 abgebildet.

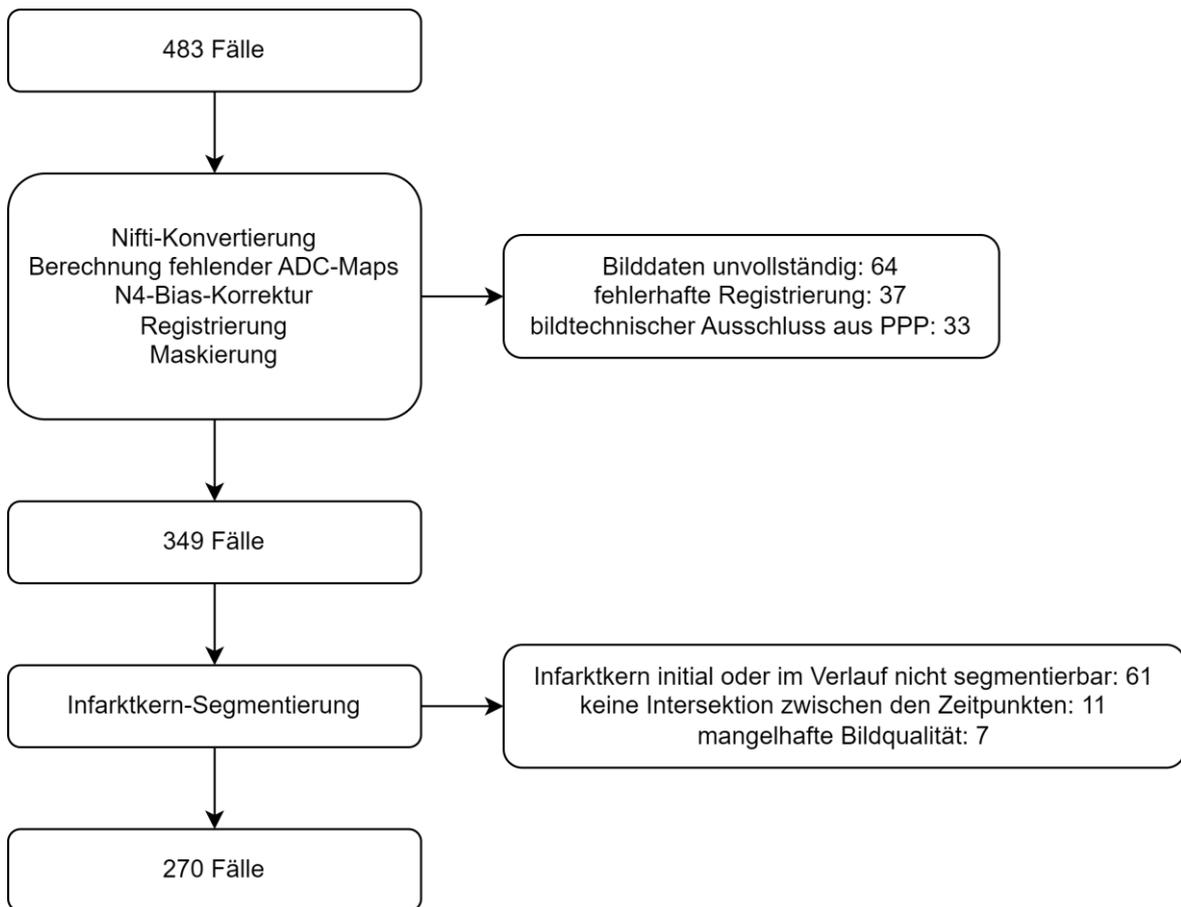


Abbildung 20: Preprocessing der Bilddaten

3.1.2.1 Nachsegmentierung der Infarktareale

Für die verbleibenden 349 Fälle wurden die Infarktgebiete anhand der ursprünglichen Segmentierungen nachsegmentiert, da die ursprünglichen Segmentierungen, bedingt durch das manuelle Schwellenwertverfahren, anhand dessen diese erstellt wurden, Rauschen aufwiesen und die registrierungsbedingte Interpolation die Qualität der Segmentierungen kleiner Infarktareale minderte. Für 50 zufällig ausgewählte Fälle wurden dafür manuell Segmentierungen anhand von ADC-Maps des initialen Zeitpunktes sowie des Verlaufszeitpunktes erstellt. Anhand dieser Ground Truth wurden unter Verwendung der nnUNet-Methode,

welche bereits in Kapitel 2.2.1.1.3.1 beschrieben wurde, Deep-Learning-basierte Modelle trainiert, die Infarktbereiche nachzusegmentieren.

Als Eingabedaten dienten die ADC-Maps und die ursprünglichen Segmentierungen des zu segmentierenden Zeitpunkts sowie die initialen FLAIR-Maps des initialen Zeitpunkts. Die FLAIR-Maps des initialen Zeitpunktes wurden verwendet, da diese zwar aufgrund des bestehenden DWI-FLAIR-Mismatches keine Informationen über den zu segmentierenden Infarktkern enthielten, jedoch alte Läsionen abbildeten. Durch die zwei Zeitpunkte ergaben sich aus den ursprünglichen 50 Fällen somit 100 verfügbare Fälle, die mit verschiedenen nnUNet-Modellen anhand einer 5-Fold-Kreuzvalidierung trainiert und evaluiert wurden. Für jeden Fold wurden die Fälle in je 80 Trainings- und 20 Validierungsfälle aufgeteilt. Dabei wurde darauf geachtet, dass die beiden Zeitpunkte eines ursprünglichen Falles immer zusammen in den Trainings- oder den Validierungsdaten enthalten waren.

2D-Slice-basierte-nnUNet-Modelle und 3D-Modelle, welche die Eingabedaten in voller Größe verarbeiteten, wurden trainiert und anhand der entsprechenden Validierungsfälle evaluiert. Für die Evaluation wurden die Prädiktionen zunächst anhand eines Schwellenwertes von 0.5 binarisiert. Der Dice-Score wurde mit Hilfe der manuell erstellten Ground Truth bestimmt. Die initiale Lernrate und die Anzahl der Epochen wurden anhand der Dice-Score-Ergebnisse optimiert. Das 2D-Slice-basierte-nnUNet-Modell performte am besten, wenn es mit einer initialen Lernrate von 0.01 über 400 Epochen trainiert wurde. Das 3D-basierte-nnUNet-Modell performte am besten, wenn dies mit einer initialen Lernrate von 0.001 über 200 Epochen trainiert wurde. Die Ergebnisse der Validierungsfälle sind in Tabelle 7 zusammengefasst.

Tabelle 7: Segmentierungs-UNet-Modelle: Hyperparameteroptimierung der Epochenanzahl sowie der initialen Lernrate
anhand des über alle Validierungsfälle gemittelten Dice-Scores

Modell	Initiale Lernrate	Epochen	Dice	std	Modell	Initiale Lernrate	Epochen	Dice	std
2D	0.01	200	0.7775	0.1062	3D	0.01	100	0.8145	0.1626
2D	0.01	400	0.8131	0.1474	3D	0.01	200	0.8253	0.1484
2D	0.01	800	0.7996	0.1741	3D	0.01	400	0.8218	0.1686
2D	0.001	100	0.7818	0.1117	3D	0.001	100	0.8211	0.1566
2D	0.001	200	0.7919	0.1687	3D	0.001	200	0.8263	0.1531
2D	0.001	400	0.8114	0.1370	3D	0.001	400	0.8209	0.1671
2D	0.0001	100	0.7705	0.0939	3D	0.0001	100	0.7943	0.1587
2D	0.0001	200	0.7908	0.1463	3D	0.0001	200	0.8101	0.1600
2D	0.0001	400	0.8034	0.1382	3D	0.0001	400	0.8182	0.1492

Zusätzlich wurde ein Ensemble-Modell aus dem performantesten 2D- und 3D-Modell erstellt. Dafür wurden die Prädiktionen der entsprechenden Modelle gemittelt und anschließend anhand eines Schwellenwertes von 0.5 binarisiert. Auch die Evaluation des Ensemble-Modells erfolgte anhand der über alle Validierungsfälle gemittelten Dice-Score-Ergebnisse. Am besten performte das Ensemble-nnUNet-Modell mit einem Dice-Score von 0.8296 bei einer Standardabweichung von 0.1396 (siehe Tabelle 8).

Tabelle 8: Dice-Scores der Segmentierungs-UNet-Modelle
anhand aller Validierungsfälle gemittelt

Modell	Dice	std
Ursprüngliche Segmentierungen	0.6067	0.1527
2D-nnUNet	0.8131	0.1474
3D-nnUNet	0.8263	0.1531
Ensemble-nnUNet	0.8296	0.1396

Die Infarktbereiche aller 349 Fälle wurden dementsprechend anhand des Ensemble-nnUNet-Modells segmentiert. Die Infarktbereiche der 50 Fälle, die zuvor als Trainings- und Validierungsfälle dienten, wurden jeweils durch das Modell segmentiert, in dessen Fold die entsprechenden Fälle als Validierungsfälle verwendet wurden. Die Infarktbereiche der restlichen 299 Fälle wurden durch jedes Modell der Kreuzvalidierung segmentiert. Diese Segmentierungen wurden anschließend über alle Modelle gemittelt. Die Segmentierungen des Ensembles wurden abschließend mit einem Schwellenwert von 0.5 binarisiert. Fälle, bei denen keine Läsion für eine der beiden Zeitpunkte segmentiert werden konnte, Fälle, bei denen keine Überschneidung in den Segmentierungen der beiden Zeitpunkte vorhanden war, und Fälle, die eine mangelhafte Bildqualität aufwiesen, wurden ausgeschlossen (siehe Abbildung 20). In Abbildung 21 sind exemplarisch die Infarkt-Segmentierungen zweier Fälle abgebildet.

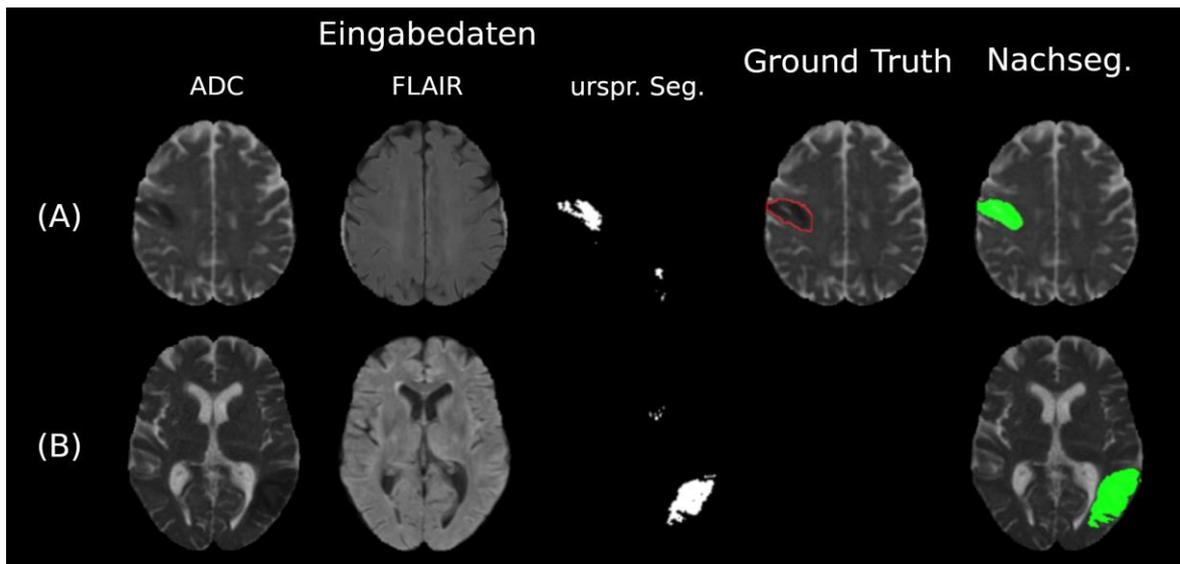


Abbildung 21: Infarkt-Segmentierungen von zwei exemplarischen Fällen

Zu den Eingabedaten zählen jeweils die ADC-Map und die ursprüngliche Segmentierung des entsprechenden Zeitpunkts sowie, unabhängig vom entsprechenden Zeitpunkt, die initiale FLAIR-Map. **Rot umrandet** ist, sofern verfügbar, die manuell erstellte Segmentierung als Ground Truth dargestellt. Die **grün-maskierte Fläche** entspricht der Segmentierung des Ensemble-nnUNet-Modells. Exemplarisch sind zwei Fälle dargestellt. **(A)** zeigt einen Fall, für den eine manuelle Segmentierung erstellt wurde und der somit zum Training und zur Validierung des Segmentierungs-Modells verwendet wurde. **(B)** zeigt einen Fall, der nicht in den Trainings- und Validierungsfällen enthalten war.

3.1.3 Finaler Datensatz

Der finale Datensatz umfasst 270 Fälle. Als Bilddaten sind für jeden Fall für den initialen Zeitpunkt und den Verlaufszeitpunkt jeweils eine ADC- und FLAIR-Map sowie eine Segmentierung des Infarktgebietes vorhanden (siehe Abbildung 22). Als klinische Informationen lagen für jeden Fall die Therapiegruppe, das Alter, das initiale Infarktvolumen (IV-initial) und das Infarktvolumen nach 22 bis 36 Stunden (IV-1d), der NIHSS-Score bei Aufnahme (NIHSS-initial), nach 22 bis 36 Stunden (NIHSS-1d) sowie der mRS-Score nach 90 Tagen (mRS-90d) vor. Da der NIHSS-Score nach 90 Tagen (NIHSS-90d) für 14 Fälle fehlte, wurde diese klinische Information nicht für die Integration der klinischen Daten verwendet. Die Berechnung der Volumenänderung des Infarktes erfolgte dabei über die Formel: $IV\text{-delta} = IV\text{-1d} - IV\text{-initial}$.

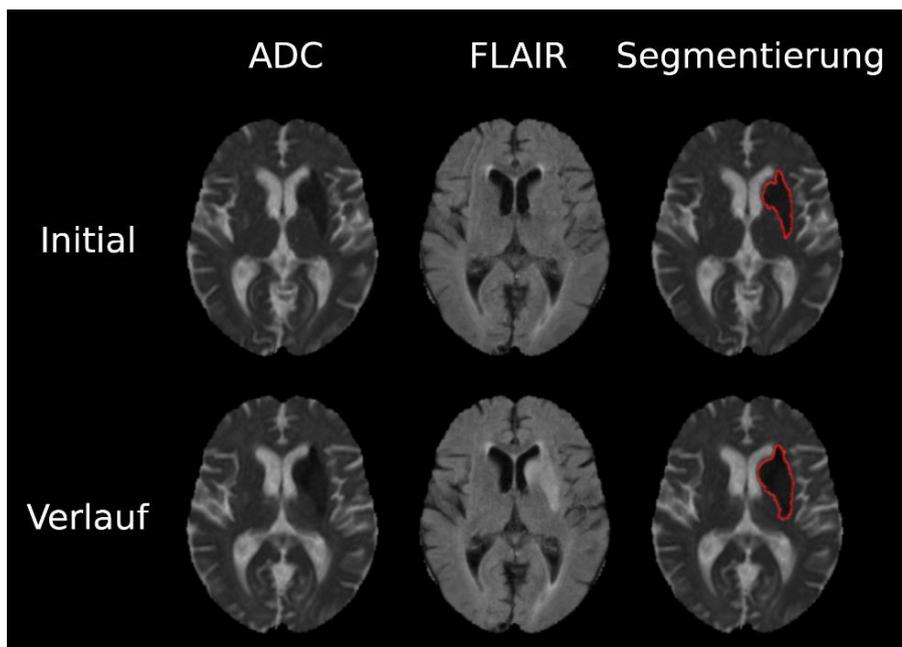


Abbildung 22: Bilddaten eines Falles

Als Bilddaten liegen sowohl für den initialen Zeitpunkt sowie den Verlaufszeitpunkt die ADC-Map, die FLAIR-Map sowie die Segmentierung des Ensemble-nnUNet-Modells vor. Die Segmentierungen des Ensemble-nnUNet-Modells sind **rot umrandet** dargestellt.

Aus den 270 verfügbaren Fällen waren 125 Fälle der Lyse-Gruppe und 145 Fälle der Placebo-Gruppe zugeordnet. Das Durchschnittsalter dieser Fälle lag bei 65 Jahren. Das gemittelte initiale Infarktvolumen betrug 11.20 ml und das gemittelte Infarktvolumen nach 22-36 Stunden betrug 27.58 ml, entsprechend einem mittleren Läsionswachstum von 16.38 ml. Dass die gemittelten Volumina im Vergleich zu den Ergebnissen der originalen Studie höher ausfallen, ist vor allem dadurch zu erklären, dass im Rahmen der Registrierung und der Nachsegmentierung der Infarktbereiche vermehrt Fälle mit kleinem Infarktvolumen verloren gingen. Der gemittelte NIHSS-Score betrug 8 Punkte initial, 6 Punkte nach 22-36 Stunden und 3 Punkte nach 90 Tagen. Dabei wurde der NIHSS-Score nach 90 Tagen nur anhand der Fälle gemittelt, für die dieser vorhanden war. Der gemittelte mRS-Score nach 90 Tagen betrug 2 Punkte (siehe Tabelle 9).

Tabelle 9: Zusammenfassung der klinischen Informationen

Gesamtanzahl der eingeschlossenen Patient:innen	Therapie		Mittelwerte der Gesamtzahl aller eing. Patient:innen							
	Anzahl der Lyse-Gruppe	Anzahl der Placebo-Gruppe	Alter	Infarktvolumen [ml]			NIHSS-Score			mRS-90d
				initial	1d	delta	initial	1d	90d	
270	125	145	65	11.20	27.58	16.38	8	6	3	2

3.2 Verwendete Methoden und Modelle

3.2.1 Prädiktion der räumlichen Infarktausbreitung

Die räumliche Ausbreitung des Infarktes im Verlauf wurde anhand verschiedener Methoden und Modelle prädiziert. Alle Deep-Learning-basierten Modelle wurden mithilfe der Keras-Bibliothek unter der Verwendung des Tensorflow-Backends erstellt, trainiert und evaluiert (Chollet 2015; Abadi et al. 2016).

Die Prädiktion der räumlichen Infarktausbreitung kann als eine voxelweise Regressionsaufgabe betrachtet werden. Obwohl der Source-Code der nnUNet-Methodik öffentlich zugänglich ist (<https://github.com/MIC-DKFZ/nnUNet>) und bestimmte methodische Modifikationen einfach zu implementieren sind, ermöglichte die Verwendung eines selbsterstellten UNet-Modells tiefgreifende architektonische Variationen. Das selbsterstellte Baseline-UNet-Modell wurde in Anlehnung an ein 3D-UNet-Modell entwickelt und die Prädiktionsergebnisse aller Modelle wurden immer im Kontext mit den Ergebnissen der nnUNet-Methode betrachtet. Die architektonischen Variationen des Baseline-UNet-Modells, genannt Multitask-UNet-Modelle, dienen dem Zweck, die Integration von klinischen Informationen über Feature-Maps an verschiedenen Stellen des Modells zu ermöglichen.

Die Bilddaten eines exemplarischen Falles und dessen beispielhafte Prädiktion werden in Abbildung 23 veranschaulicht.

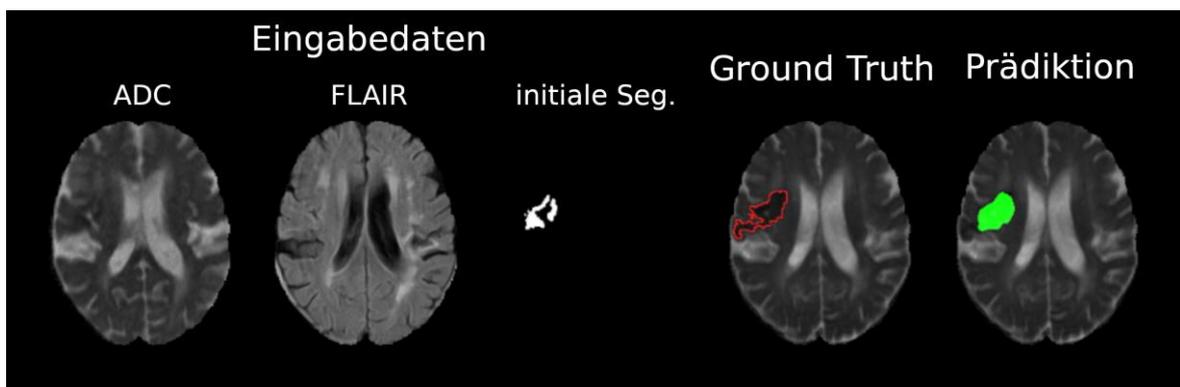


Abbildung 23: Beispielhafte Prädiktion eines exemplarischen Falles

Zu den für die Prädiktion verwendeten Eingabedaten zählen die initiale ADC-, die initiale FLAIR-Map und die Infarkt-Segmentierung des initialen Zeitpunktes. **Rot umrandet** ist die Infarkt-Segmentierung des Verlaufszeitpunktes als Ground Truth über der im Verlauf erhobenen ADC-Map dargestellt. Die **grün maskierte Fläche** entspricht einer beispielhaften Prädiktion der Infarktausbreitung im Verlauf.

3.2.2 Training und Evaluation

Um die Prädiktionen der unterschiedlichen Deep-Learning-basierten Modelle vergleichen zu können, wurden alle Deep-Learning-basierten Modelle anhand derselben verschachtelten Kreuzvalidierung trainiert, optimiert und evaluiert (siehe Abbildung 24). Im Rahmen der verschachtelten Kreuzvalidierung wurden die Prädiktionen aller Modelle stets anhand eines Schwellenwertes von 0.5 binarisiert (binäre Klassifikation) und mit einem Binary-Fill-Holes-Algorithmus nachbearbeitet.

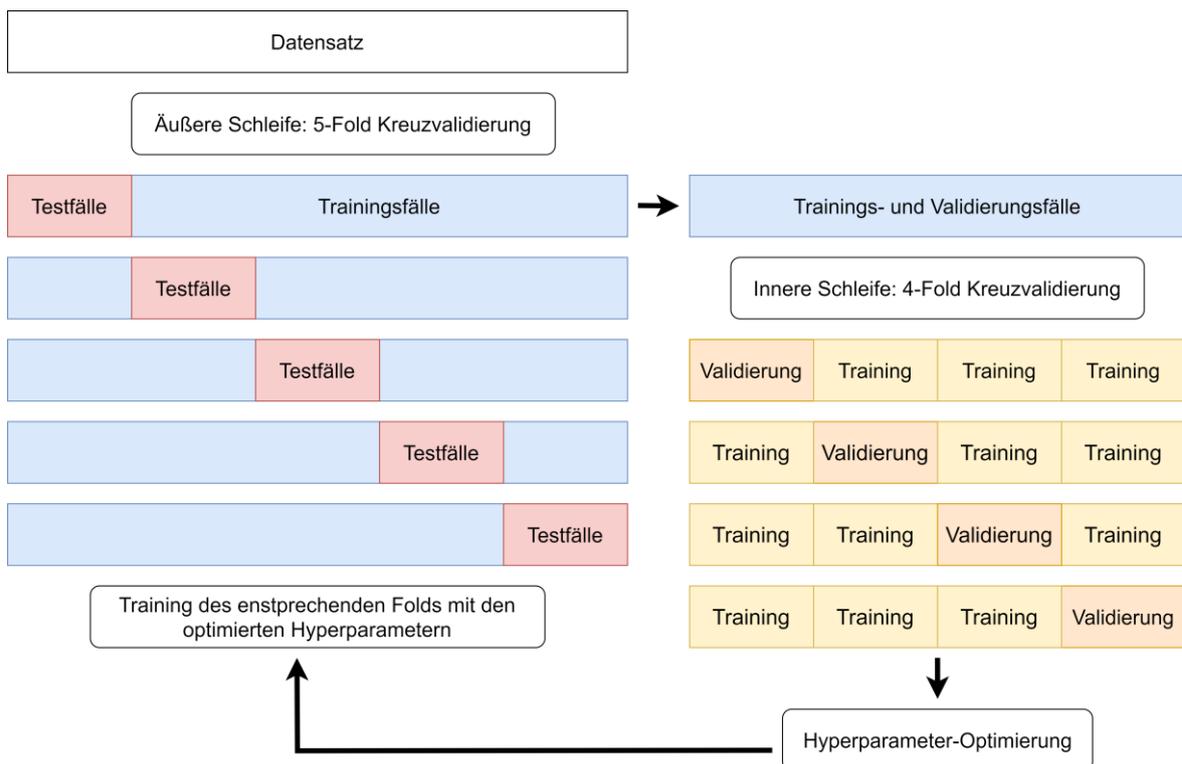


Abbildung 24: Verschachtelte Kreuzvalidierung

Dafür wurden die verbleibenden 270 Fälle des finalen Datensatzes mittels einer 5-Fold-Kreuzvalidierung zufällig in 216 Trainingsfälle und 54 Testfälle pro Fold aufgeteilt (äußere Schleife). Die finale Evaluation der verschiedenen Modelle erfolgte anschließend anhand dieser 5-Fold-Kreuzvalidierung gemittelt über alle 270 Testfälle. Die Optimierung der Hyperparameter der Modelle jedes Folds der äußeren Schleife erfolgte zuvor unabhängig von den Testfällen des jeweiligen Folds. Die Trainingsfälle des entsprechenden Folds der äußeren Schleife wurden dafür mittels einer 4-Fold-Kreuzvalidierung in 162 Trainings- und 54 Validierungsfälle pro Fold aufgeteilt (innere Schleife des entsprechenden Folds der äußeren Schleife). Für das 3D-nnUNet-Modell und das Baseline-UNet-Modell wurde so die Epochenanzahl optimiert. Für die Deep-Learning basierten Modelle,

in die zusätzlich die klinischen Informationen integriert wurden, erfolgte die Bestimmung des Integrationsortes der klinischen Informationen und die Selektion der integrierten klinischen Informationen.

3.2.3 Prädiktionsmodelle

3.2.3.1 nnUNet-Methode

Die nnUNet-Methode ist eine Deep-Learning-basierte selbstkonfigurierende state-of-the-art Segmentierungsmethode für medizinische Bilddaten (Isensee et al. 2021), welche bereits in Kapitel 2.2.1.1.3.1 näher beschrieben wurde. Es wurde ein 3D-nnUNet-Modell, welches die Eingangs-Bilddaten als 3D-Volumina verarbeitete, trainiert und evaluiert. Als Eingabedaten für die nnUNet-Modelle und das Baseline-UNet-Modell dienten die initialen ADC- und FLAIR-Maps sowie die Segmentierungen der initialen Infarktbereiche. Die als Eingabedaten verwendeten Bilddaten wurden dabei bildweise z-standardisiert.

Die Konfigurationen der Methode, die vom Datensatz und der Rechenleistung abhängig sind, definierten sich für die 3D-nnUNet-Modelle wie folgt. Da die Bilddaten bereits zuvor auf einen Atlas registriert wurden, war außer dem Cropping der Bilddaten auf die hirnbildenden Bereiche ($192 \times 256 \times 32$) kein weiteres Resampling nötig. Die Normalisierung erfolgte bildweise über die voxelweise z-Standardisierung, dabei wurden Voxel mit einem Wert von Null in der Berechnung nicht berücksichtigt. Die 3D-nnUnet-Modelle wurden anhand der Bilddaten in voller Auflösung trainiert. Im Rahmen der Datenaugmentierung wurden die Bilddaten skaliert, rotiert, gespiegelt, geblurt, Gamma-, Helligkeits- und Kontrast-korrigiert, eine geringe Auflösung wurde simuliert und Rauschen hinzugefügt. Architektonisch bildete sich ein Modell mit insgesamt 6 Pooling-Operationen. Eine Reduktion der Prädiktionen auf die größte verbundene Komponente wurde während des Postprocessings nicht angewandt (Isensee et al. 2021).

3.2.3.2 Baseline-UNet

Das Baseline-UNet-Modell wurde in Anlehnung an das zuvor beschriebenen 3D-nnUNet-Modell konfiguriert. Auch hier wurden die Eingangs-Bilddaten auf die hirnbildenden Bereiche beschnitten ($192 \times 256 \times 32$). Anschließend erfolgte

jedoch die voxelweise z-Standardisierung jeder Map unter der Berücksichtigung von allen Voxeln innerhalb einer auf die Bereiche des Hirns zugeschnittenen Maske. Auch das Baseline-UNet-Modell wurde anhand der 3D-Volumina trainiert. Im Rahmen der Datenaugmentierung wurden die Bilddaten skaliert, rotiert, gespiegelt und Gamma-korrigiert. Im Unterschied zum 3D-nnUNet-Modell wurde aus Rechenkapazitätsgründen ein Modell mit insgesamt 5 anstatt 6 Pooling-Operationen verwendet, wobei initial 16 anstatt 32 Feature-Maps verwendet wurden, die sich nach jeder Pooling-Operation verdoppelten. Außerdem wurden unabhängig von einer möglichen Anisotropie der räumlichen Auflösung stets 3x3x3-Kernel in den Convolutional-Layern verwendet. Die Berechnung der Verlustfunktion erfolgte ohne die Implementierung einer Deep-Supervision (Lee et al. 2015). Eine schematische Darstellung der Architektur des Baseline-UNet-Modells ist in Abbildung 25 abgebildet.

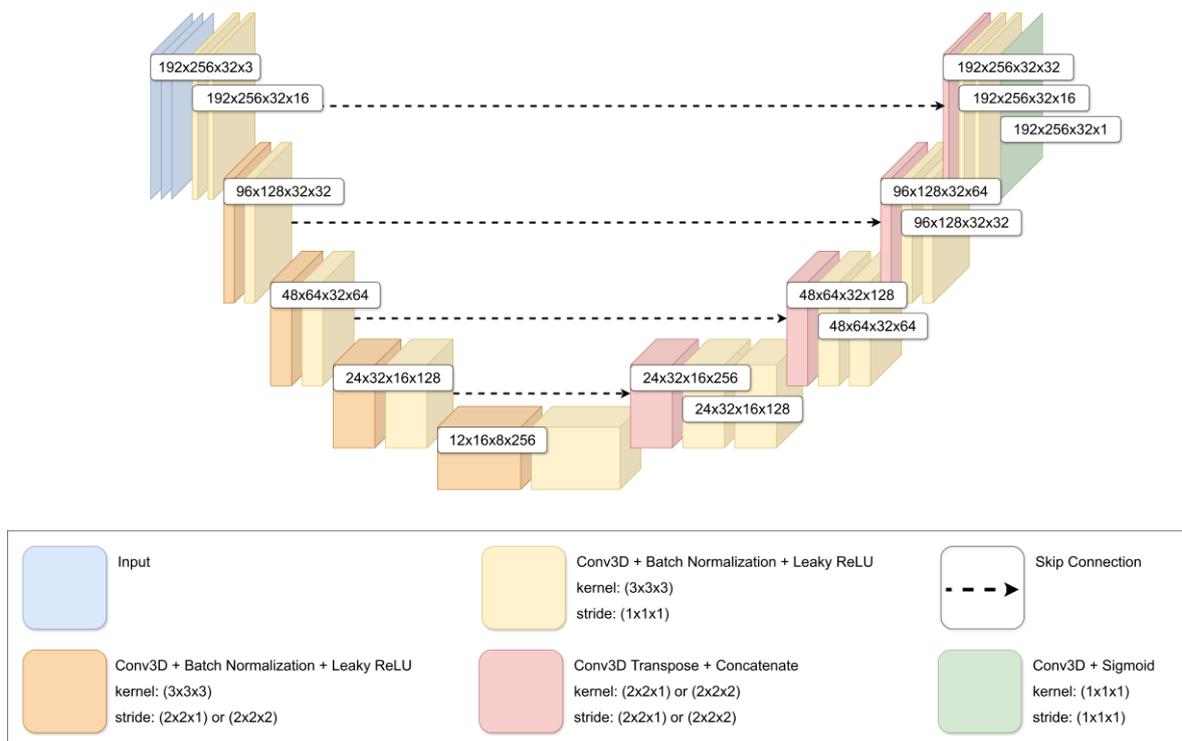


Abbildung 25: Baseline-UNet

3.2.3.3 Multitask-UNet: Integration klinischer Informationen

Zur Klärung der Fragen, an welchem Ort die Feature-Maps aus klinischen Informationen in ein Modell zu integrieren sind und ob die Funktionsweise eines UNets sich dafür eignet, relevante Zusammenhänge aus mehreren klinischen Feature-Maps zu extrahieren, wurde der Aufbau des Baseline-UNet-Modells variiert, damit die Informationen der klinischen Informationen auf verschiedene

Weise durch Feature-Maps an verschiedenen Orten in das Modell integriert werden konnten. Die klinischen Daten wurden dabei über separate Feature-Maps in das Modell integriert. Der Einfluss der klinischen Informationen auf die Infarktausbreitung musste so innerhalb des UNet-Modells abstrahiert werden.

Wie bereits in Kapitel 3.1.3 beschrieben lagen als klinische Informationen für jeden Fall die Therapiegruppe, das Alter, das initiale Infarktvolumen (IV-initial), der NIHSS-Score bei Aufnahme (NIHSS-initial), nach 22 bis 36 Stunden (NIHSS-1d) sowie der mRS-Score nach 90 Tagen (mRS-90d) vor. Zudem lag für jeden Fall das tatsächliche Infarktvolumen nach 22 bis 36 Stunden (IV-1d) vor. Im Rahmen der 4-Fold-Kreuzvalidierung der inneren Schleifen und der 5-Fold-Kreuzvalidierung der äußeren Schleife wurden die klinischen Informationen anhand der Trainingsfälle des entsprechenden Folds z-standardisiert. Die Zugehörigkeit zur Therapiegruppe wurde als binäre Variable betrachtet. Wie die auf den Atlas registrierten Bilddaten wurden die so generierten Feature-Maps stets auf die gehirnabbildenden Anteile des Atlas maskiert (siehe Kapitel 3.1.2). Die jeweiligen Feature-Maps der Modelle wurden am Anfang des Encoders, auf jeder Ebene des Encoders, am Ende des Bottlenecks, auf jeder Ebene des Decoders oder am Ende des Decoders integriert. Bei einer Integration der Feature-Maps hinter dem Encoder wurde der vortrainierte Encoder des Baseline-UNet-Modells verwendet und dessen Gewichte wurden während des Trainings eingefroren.

In Abbildung 26 ist ein Multitask-UNet-Modell, bei dem die klinischen Informationen auf jeder Ebene des Decoders eingefügt werden, beispielhaft dargestellt. Die restlichen Modell-Variationen sind im Anhang dargestellt (siehe Abbildung 32 bis Abbildung 35).

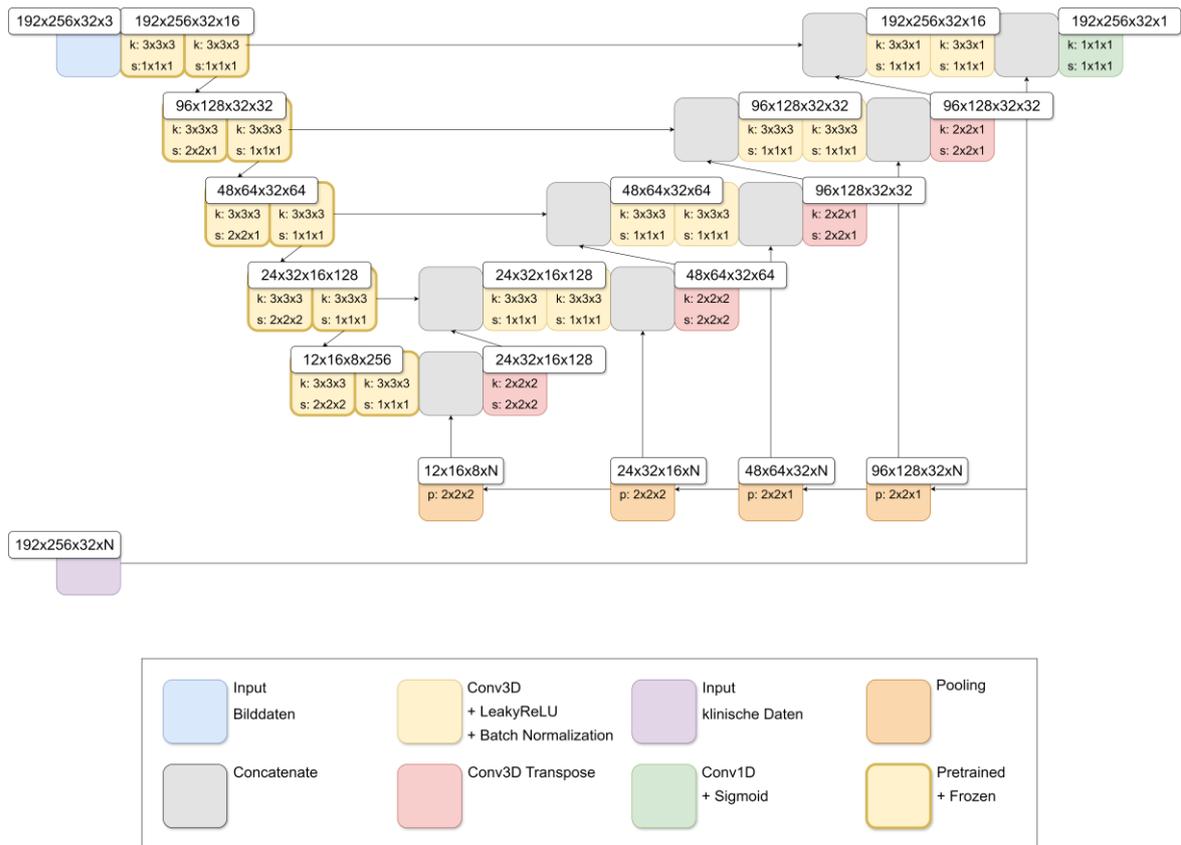


Abbildung 26: Integration klinischer Informationen anhand von separaten Feature-Maps

In diesem Beispiel erfolgt die Integration von N klinischen Informationen über separate Feature-Maps auf jeder Ebene des Decoders. Der Encoder wurde vom Baseline-UNet übernommen und dessen Gewichte wurden während des Trainings eingefroren.

4 Ergebnisse

4.1 Identifikation von Prädiktoren anhand linearer Regression

Anhand univariater, linearer Regressionsanalysen (Seabold & Perktold 2010) zeigten sich das initiale Infarktvolumen, der initiale NIHSS-Score, der NIHSS-Score nach 22-36 Stunden, der NIHSS-Score nach 90 Tagen und der mRS-Score nach 90 Tagen als signifikante Prädiktoren der Volumenänderung des Infarktes (IV-delta) ($p < 0.05$). Für das Patient:innenalter und die Therapiegruppe konnte hingegen keine signifikante lineare Korrelation mit der Volumenänderung des Infarktes (IV-delta) nachgewiesen werden (siehe Tabelle 10).

Tabelle 10: Korrelation der klinischen Informationen mit der Volumenänderung des Infarktes

Prädiktor	Korrelations-Koeffizient	Standardfehler	t-Wert	p-Wert
Alter	-0.0161	0.223	-0.072	0.942
Konstante	17.4361	14.821	1.176	0.240
Therapiegruppe	2.3333	5.058	0.461	0.645
Konstante	15.3002	3.442	4.445	$p < 0.001$
IV-initial	1.0203	0.158	6.472	$p < 0.001$
Konstante	4.9514	2.937	1.686	0.093
NIHSS-initial	3.2053	0.490	6.546	$p < 0.001$
Konstante	-7.8135	4.376	-1.785	0.075
NIHSS-1d	4.2036	0.363	11.571	$p < 0.001$
Konstante	-8.6077	2.985	-2.884	0.004
mRS-90d	13.1781	1.576	8.361	$p < 0.001$
Konstante	-9.0972	3.786	-2.403	0.017

4.2 Prädiktionsergebnisse

Als Ground Truth wurden im Rahmen des Trainings die Segmentierungen der Infarktbereiche verwendet, die anhand der Bilddaten im Verlauf erstellt wurden. Als Eingabedaten für die nnUNet-Modelle und das Baseline-UNet-Modell dienten die initialen ADC- und FLAIR-Maps sowie die Segmentierungen der initialen Infarktbereiche. Für die Multitask-UNet-Modelle lagen als klinische Informationen zusätzlich für jeden Fall die Therapiegruppe, das Alter, das initiale Infarktvolumen (IV-initial), der NIHSS-Score bei Aufnahme (NIHSS-initial), nach 22 bis 36 Stunden (NIHSS-1d), der mRS-Score nach 90 Tagen (mRS-90d) sowie das tatsächliche Infarktvolumen nach 22 bis 36 Stunden (IV-1d) vor.

Die Optimierung der jeweiligen Hyperparameter der Modelle erfolgte unabhängig von den Testfällen der äußeren Schleife anhand der Validierungsfälle der inneren Schleifen. Anschließend wurden die entsprechenden Modelle jedes Folds der

äußeren Schleife anhand der optimierten Hyperparameter trainiert (siehe Kapitel 3.2.2).

4.2.1 nnUNet-Modelle

Die räumliche Infarktausbreitung wurde zunächst anhand eines 3D-nnUNet-Modells prädiziert. Die 3D-UNet-Modelle wurde mit einer initialen Lernrate von 0.001 trainiert. Die Epochenanzahl wurde anhand der verschachtelten Kreuzvalidierung unabhängig von den Testdaten des jeweiligen Folds der äußeren Schleife optimiert (siehe Tabelle 11).

Tabelle 11: nnUNet-Modelle: Hyperparameteroptimierung der Epochenanzahl
anhand des über alle Validierungsfälle des entsprechenden Folds der inneren Schleife gemittelten Dice-Scores

Epochenanzahl	Dice				
	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
10	0.5238	0.5324	0.5337	0.5372	0.5273
25	0.5269	0.5363	0.5356	0.5364	0.5278
50	0.5203	0.5389	0.5312	0.5306	0.5207

Im Rahmen der äußeren Schleife erzielten die auf diese Weise optimierten 3D-nnUNet-Modelle einen Dice-Score von 0.5344 bei einer Standardabweichung von 0.2150. Die entsprechenden Modelle werden folgend mit dem Begriff 3D-nnUNet bezeichnet.

4.2.2 Baseline-UNet

Anschließend erfolgte die Prädiktion der räumlichen Infarktausbreitung anhand des Baseline-UNet-Modells. Analog zu den 3D-nnUNet-Modellen wurden alle Baseline-UNet-Modelle mit einer initialen Lernrate von 0.001 trainiert. Die Epochenanzahl wurde anhand der verschachtelten Kreuzvalidierung unabhängig von den Testdaten des jeweiligen Folds der äußeren Schleife optimiert (siehe Tabelle 12).

Tabelle 12: Baseline-UNet-Modelle: Hyperparameteroptimierung der Epochenanzahl
anhand des über alle Validierungsfälle des entsprechenden Folds der inneren Schleife gemittelten Dice-Scores

Epochenanzahl	Dice				
	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
10	0.5219	0.5321	0.5275	0.5358	0.5264
25	0.5221	0.5347	0.5300	0.5368	0.5295
50	0.5078	0.5268	0.5274	0.5291	0.5096

Im Rahmen der äußeren Schleife erzielten die auf diese Weise optimierten Baseline-UNet-Modelle einen Dice-Score von 0.5334 bei einer

Standardabweichung von 0.2183. Die entsprechenden Modelle werden folgend mit dem Begriff Baseline-UNet bezeichnet.

4.2.3 Multitask-UNet: Integration klinischer Informationen

Alle Modell-Variationen, in die die klinischen Informationen integriert wurden, wurden orientierend an den Baseline-UNet-Modellen unabhängig vom Integrationsort und den verwendeten klinischen Informationen mit einer initialen Lernrate von 0.001 über 25 Epochen trainiert. Der Integrationsort sowie die verwendeten klinischen Informationen wurden anhand der verschachtelten Kreuzvalidierung unabhängig von den Testdaten des jeweiligen Folds der äußeren Schleife optimiert.

Die klinischen Informationen wurden über separate Feature-Maps in die verschiedenen Modelle integriert. Dabei wurden zunächst die klinischen Informationen integriert, die zum initialen Zeitpunkt verfügbar waren (Therapiegruppe, IV-initial, NIHSS-initial). Anschließend wurden zwei weitere Modelle trainiert. Zunächst wurde zusätzlich der NIHSS-Score zum Zeitpunkt der Bildgebung im Verlauf (NIHSS-1d) integriert und anschließend ergänzend der mRS nach 90 Tagen (mRS-90d). Zunächst wurde für jedes Modell der Integrationsort anhand der verschachtelten Kreuzvalidierung unabhängig von den Testdaten des jeweiligen Folds der äußeren Schleife optimiert (siehe Tabelle 13).

Tabelle 13: Multitask-UNet-Modelle: Hyperparameteroptimierung des Integrationsortes der klinischen Informationen
anhand des über alle Validierungsfälle des entsprechenden Folds der inneren Schleife gemittelten Dice-Scores

Klinische Informationen	Integrationsort	Dice				
		Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
Initial	Vor dem Encoder	0.5093	0.5237	0.5263	0.5338	0.5219
	Auf jeder Ebene des Encoders	0.5129	0.5250	0.5262	0.5280	0.5186
	Am Ende des Bottlenecks	0.5194	0.5350	0.5276	0.5329	0.5241
	Auf jeder Ebene des Decoders	0.5218	0.5380	0.5314	0.5336	0.5257
	Am Ende des Decoders	0.5227	0.5384	0.5344	0.5370	0.5280
Initial + NIHSS-1d	Vor dem Encoder	0.5155	0.5290	0.5276	0.5347	0.5243
	Auf jeder Ebene des Encoders	0.5163	0.5293	0.5262	0.5391	0.5291
	Am Ende des Bottlenecks	0.5202	0.5355	0.5288	0.5328	0.5262
	Auf jeder Ebene des Decoders	0.5309	0.5417	0.5392	0.5396	0.5327
	Am Ende des Decoders	0.5265	0.5412	0.5359	0.5386	0.5292
Initial + NIHSS-1d + mRS-90d	Vor dem Encoder	0.5124	0.5134	0.5125	0.5342	0.5214
	Auf jeder Ebene des Encoders	0.5093	0.5239	0.5174	0.5299	0.5171
	Am Ende des Bottlenecks	0.5224	0.5351	0.5364	0.5361	0.5257
	Auf jeder Ebene des Decoders	0.5312	0.5385	0.5358	0.5394	0.5290
	Am Ende des Decoders	0.5269	0.5406	0.5372	0.5382	0.5286

Dabei wurde die Tendenz ersichtlich, dass sich die klinischen Informationen am besten am Ende des Decoders oder auf jeder Ebene des Decoders integrieren

lassen. Der auf diese Weise selektierte Integrationsort für die verschiedenen Modelle ist in Tabelle 14 aufgeführt.

Tabelle 14: Multitask-UNet-Modelle: Optimierter Integrationsort

Äußere Schleife	Initial	Initial + NIHSS-1d	Initial + NIHSS-1d + mRS-90d
Fold 0	Am Ende des Decoders	Auf jeder Ebene des Decoders	Auf jeder Ebene des Decoders
Fold 1	Am Ende des Decoders	Auf jeder Ebene des Decoders	Am Ende des Decoders
Fold 2	Am Ende des Decoders	Auf jeder Ebene des Decoders	Am Ende des Decoders
Fold 3	Am Ende des Decoders	Auf jeder Ebene des Decoders	Auf jeder Ebene des Decoders
Fold 4	Am Ende des Decoders	Auf jeder Ebene des Decoders	Auf jeder Ebene des Decoders

Die Dice-Scores, die durch die entsprechend optimierten Modelle im Rahmen der äußeren Schleife erzielt wurden, sind in Tabelle 15 zusammengefasst. Die entsprechenden Modelle werden folgend als Multitask-UNets bezeichnet, dabei wird in der Klammerung angegeben, bis zu welchem Zeitpunkt die klinischen Informationen integriert wurden.

Tabelle 15: Dice-Scores der Multitask-UNet-Modelle
anhand aller Testfälle der äußeren Schleife gemittelt

Modell	Dice-Score
Multitask-UNet (Initial)	0.5340
Multitask-UNet (1d)	0.5384
Multitask-UNet (90d)	0.5394

In einer weiteren Untersuchung wurde testweise die tatsächliche Volumenänderung des Infarktes integriert und der Integrationsort entsprechend optimiert (siehe Tabelle 16).

Tabelle 16: Multitask-UNet-Modell: Hyperparameteroptimierung des Integrationsortes der tatsächlichen Volumenänderung

anhand des über alle Validierungsfälle des entsprechenden Folds der inneren Schleife gemittelten Dice-Scores

Klinische Informationen	Integrationsort	Dice				
		Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
Volumenänderung	Vor dem Encoder	0.5608	0.5739	0.5757	0.5761	0.5717
Volumenänderung	Auf jeder Ebene des Encoders	0.5653	0.5760	0.5749	0.5822	0.5717
Volumenänderung	Am Ende des Bottlenecks	0.5194	0.5334	0.5347	0.5396	0.5230
Volumenänderung	Auf jeder Ebene des Decoders	0.5508	0.5511	0.5627	0.5676	0.5592
Volumenänderung	Am Ende des Decoders	0.5305	0.5713	0.5442	0.5489	0.5366

Im Rahmen der äußeren Schleife erzielten die auf diese Weise optimierten UNet-Modelle einen Dice-Score von 0.5785 bei einer Standardabweichung von 0.1989 und einen mittleren absoluten Fehler von 5 ml. Im Gegensatz zu den Modellen, in die die klinischen Informationen integriert wurden, erwies sich hierbei der Integrationsort auf jeder Ebene des Encoders oder vor dem Encoder als optimal. Diese Ergebnisse können als Grenzwerte angesehen werden, die man durch die Integration klinischer Informationen erreichen kann.

4.3 Vergleich der Modelle

Die finale Evaluation der verschiedenen Modelle erfolgte anhand des über alle 270 Testfälle gemittelten Dice-Scores der äußeren Schleife.

4.3.1 Dice-Scores

Da es schwierig ist, die Prädiktionsqualität eines Modells anhand des gemittelten Dice-Scores zu bewerten, ist es wichtig, diesen im Kontext mit den Ergebnissen anderer Methoden zu betrachten. Um eine Einordnung der Prädiktionsqualität zu ermöglichen, werden die Ergebnisse der verschiedenen Prädiktions-Modelle mit den Dice-Scores der initialen Infarktsegmentierungen und einem Dilatierungs-Modell in Kontext gesetzt. Für das Dilatierungs-Modell wurden die initialen Segmentierungen mit Hilfe des SciPy-Dilatierungsalgorithmus dilatiert (Virtanen et al. 2020). Dabei wurde die Anzahl der Iterationen des Algorithmus anhand des über alle Testfälle gemittelten Dice-Scores optimiert. Von allen Modellen, die ausschließlich bildbasierte Eingabedaten verarbeiteten, performte das 3D-nnUNet mit einem Dice-Score von 0.5344 bei einer Standardabweichung von 0.2150 am besten (siehe Tabelle 17).

Tabelle 17: Dice-Scores der verschiedenen Prädiktions-Modelle
anhand aller Testfälle der äußeren Schleife gemittelt

Modell	Dice	std
Initiale Infarktsegmentierung	0.4575	0.2014
Dilatierungs-Modell	0.5068	0.2075
3D-nnUNet	0.5344	0.2150
Baseline-UNet	0.5334	0.2183

Die Modelle, die ausschließlich bildbasierte Eingabedaten verwendeten, wurden anschließend mit Hilfe des einseitigen Wilcoxon-Vorzeichen-Rang-Tests mit diesem Modell verglichen (siehe Tabelle 18). Dabei zeigten sich die Dice-Scores der Prädiktionen des 3D-nnUNets signifikant besser als die Dice-Scores der Prädiktionen des Dilatierungs-Modells und als die der initialen Infarktsegmentierungen ($p < 0.05$). Die Prädiktionen des Baseline-UNets hingegen unterschieden sich nicht signifikant von denen des 3D-nnUNets.

Tabelle 18: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem 3D-nnUNet

Modell	Alternativhypothese:
	$\tilde{x}_{3D-nnUNet} > \tilde{x}_{Modell}$
Initiale Infarktsegmentierung	$5.2675 \cdot 10^{-22}$
Dilatierungs-Modell	$2.069 \cdot 10^{-15}$
Baseline-UNet	0.0941

Von allen Modellen performte das Multitask-UNet (90d) mit einem Dice-Score von 0.5394 bei einer Standardabweichung von 0.214 am besten (siehe Tabelle 19).

Tabelle 19: Dice-Scores der verschiedenen Prädiktions-Modelle
anhand aller Testfälle der äußeren Schleife gemittelt

Modell	Dice	std
Initiale Infarktsegmentierung	0.4575	0.2014
Dilatierungs-Modell	0.5068	0.2075
3D-nnUNet	0.5344	0.2150
Baseline-UNet	0.5334	0.2183
Multitask-UNet (initial)	0.5340	0.2171
Multitask-UNet (1d)	0.5384	0.2127
Multitask-UNet (90d)	0.5394	0.2140

Alle Modelle wurden anschließend mit Hilfe des einseitigen Wilcoxon-Vorzeichen-Rang-Tests mit dem Baseline-UNet verglichen (siehe Tabelle 20). Dabei zeigten sich die Prädiktionen des Multitask-UNets (90d) sowie des Multitask-UNets (1d) signifikant besser als die Prädiktionen des Baseline-UNets ($p < 0.05$). Die Integration der klinischen Informationen konnte somit für diese Modelle die Prädiktionen der räumlichen Infarktausbreitung signifikant verbessern.

Tabelle 20: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem Baseline-UNet

Modell	Alternativhypothese:
	$\tilde{x}_{Baseline-UNet} < \tilde{x}_{Modell}$
Initiale Infarktsegmentierung	1
Dilatierungs-Modell	1
3D-nnUNet	0.0941
Multitask-UNet (initial)	0.6533
Multitask-UNet (1d)	0.0002
Multitask-UNet (90d)	$2.6285 \cdot 10^{-5}$

In Abbildung 27 sind die Prädiktionsergebnisse ausgewählter Modelle in einem Geigenplot dargestellt. Dabei fällt auf, dass sich die Verteilung der Prädiktionsergebnisse der verschiedenen Prädiktionsmodelle zum Teil deutlich von den initialen Infarktkernsegmentierungen abhebt. Die Verteilung der Dice-Scores des Multitask-UNets (1d) unterscheidet sich dabei deutlich von der Verteilung des Dilatierungs-Modells, jedoch nur sehr gering von der Verteilung des Baseline-UNet-Modells.

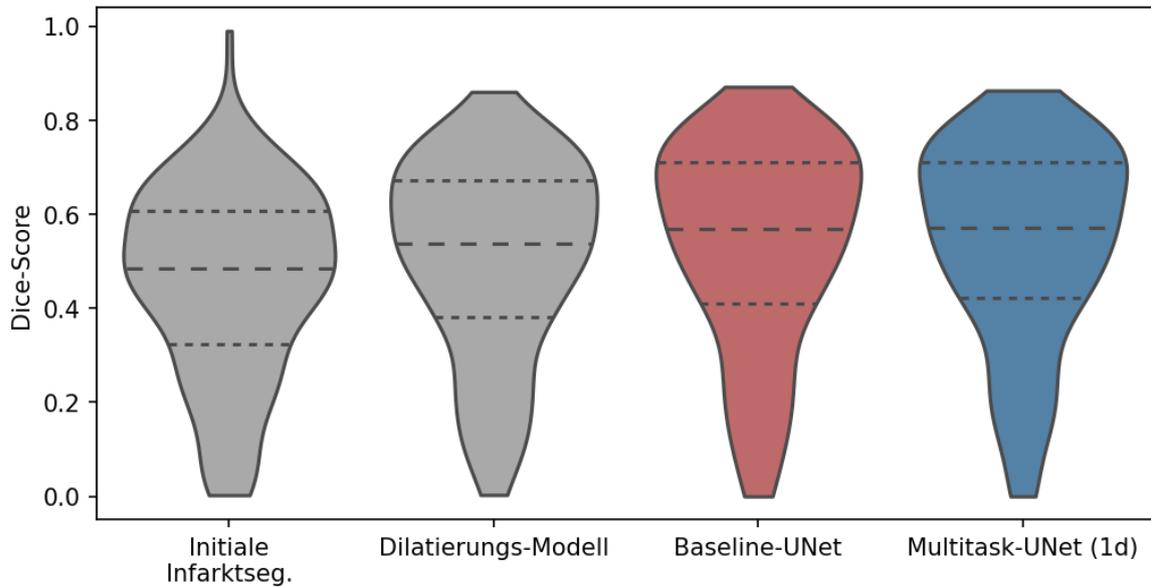


Abbildung 27: Geigenplot verschiedener Prädiktions-Modelle

Abgebildet wird die Verteilung der Dice-Score-Prädiktionsergebnisse. Die kurz-gestrichelten Linien repräsentieren das obere und untere Quartil. Der Median ist grob-gestrichelt dargestellt.

In Abbildung 28 und in Abbildung 29 sind die Prädiktionsergebnisse des Baseline-UNets mit den Prädiktionen des Multitask-UNets (1d) anhand von Box-Plots und Streudiagrammen im Vergleich dargestellt. Dabei sind in Abbildung 28 alle negativen und in Abbildung 29 alle positiven Dice-Score-Veränderungen oberhalb einer Schwellendifferenz von 0.05 gegenüber den Prädiktionsergebnissen des Baseline-UNets verbunden dargestellt. Dabei fällt auf, dass die Fälle, deren Dice-Scores sich durch die Integration der klinischen Informationen verschlechterten, häufig bereits gut durch das Baseline-UNet prädiziert wurden und dass sich einzelne Fälle zum Teil deutlich verschlechterten. Im Gegensatz dazu weisen die Fälle, deren Dice-Scores sich durch die Integration der klinischen Informationen verbesserten, sowohl gute als auch schlechte Prädiktionen durch das Baseline-UNet auf. Dabei verbesserten sich numerisch mehr Fälle, die Verbesserungen fallen jedoch auch weniger gravierend aus.

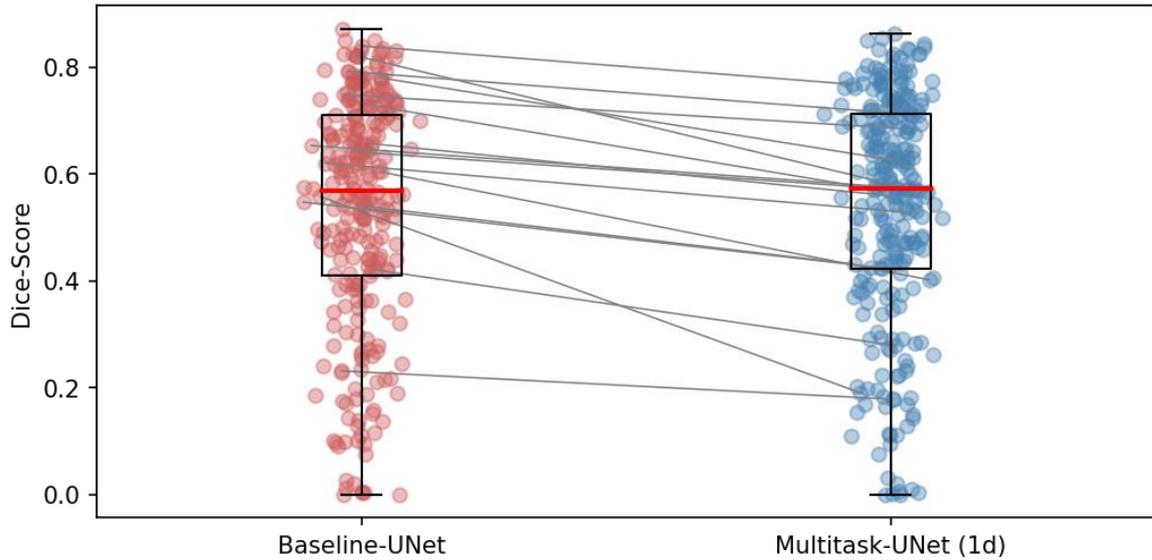


Abbildung 28: Boxplot-Ergebnisse des Baseline-UNets und des Multitask-UNets (1d)

Die Punkte der Streudiagramme repräsentieren die Dice-Score-Ergebnisse der einzelnen Testfälle. Alle Fälle, deren Dice-Scores sich durch die Integration der klinischen Informationen um mehr als 0.05 verschlechterten, sind verbunden dargestellt. Die rote Linie markiert den über alle Testfälle gemittelten Dice-Score.

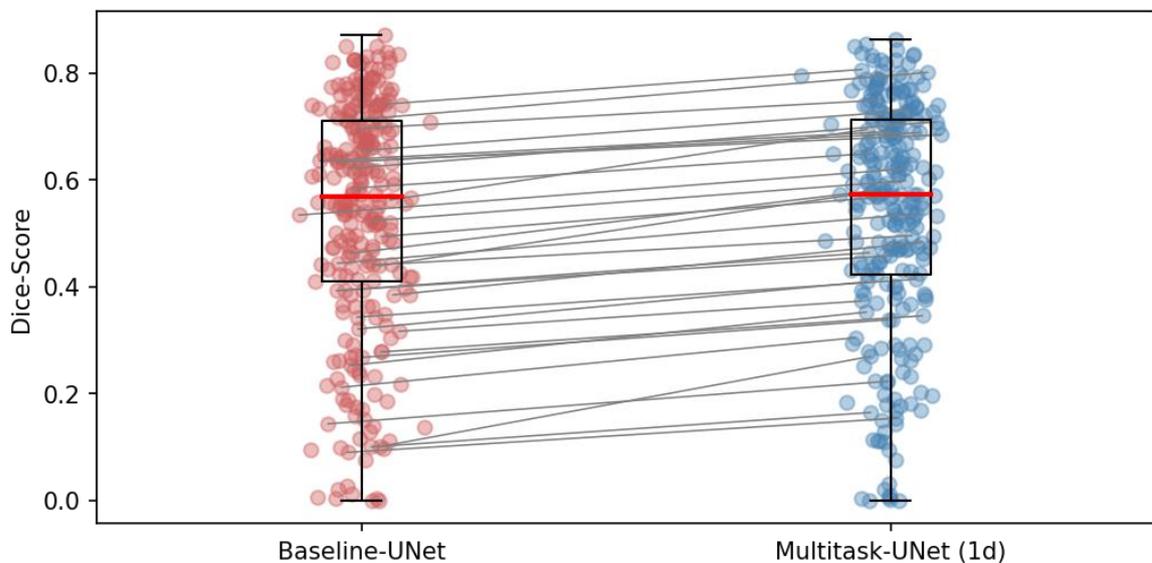


Abbildung 29: Boxplot-Ergebnisse des Baseline-UNets und des Multitask-UNets (1d)

Die Punkte der Streudiagramme repräsentieren die Dice-Score-Ergebnisse der einzelnen Testfälle. Alle Fälle, deren Dice-Scores sich durch die Integration der klinischen Informationen um mehr als 0.05 verbesserte, sind verbunden dargestellt. Die rote Linie markiert den über alle Testfälle gemittelten Dice-Score.

In Abbildung 30 sind die Dice-Score-Veränderungen der Prädiktionen zwischen dem Baseline-UNet und dem Multitask-UNet (1d) dargestellt. Aus der Abbildung wird ersichtlich, dass die Fälle anzahlmäßig überwiegen, deren Dice-Scores sich durch die Integration der klinischen Daten verbesserte. Nur für 13 Fälle änderte

sich der Dice-Score um mehr als 0.1. Bei einem Ausreißer verschlechterte sich der Dice-Score um mehr als 0.5.

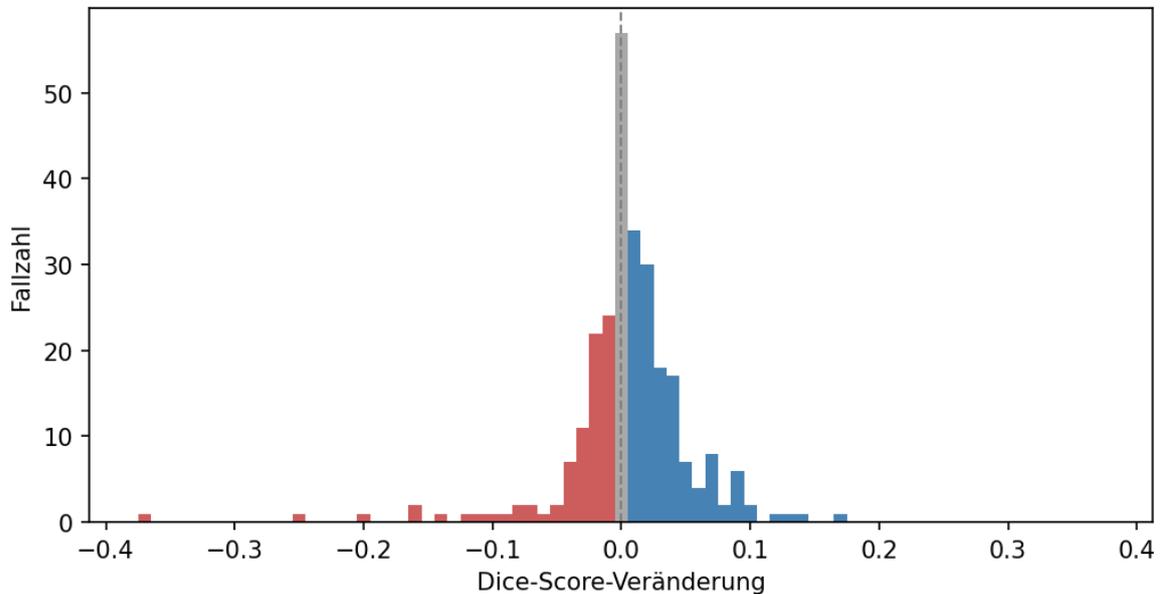


Abbildung 30: Dice-Score-Veränderungen zwischen dem Baseline- und dem Multitask-UNet

In einem letzten Schritt wurden nun auch die Multitask-UNets mit Hilfe des einseitigen Wilcoxon-Vorzeichen-Rang-Tests mit dem 3D-nnUNets verglichen (siehe Tabelle 21). Dabei zeigten sich die Prädiktionen des Multitask-UNets (1d) und des Multitask-UNets (90d) signifikant besser als die Prädiktionen des 3D-nnUNets ($p < 0.05$).

Tabelle 21: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem 3D-nnUnet

Modell	Alternativhypothese: $\tilde{x}_{3D-nnUNet} < \tilde{x}_{Modell}$
Multitask-UNet (initial)	0.8500
Multitask-UNet (1d)	0.0396
Multitask-UNet (90d)	0.0144

4.3.2 Mittlerer absoluter Fehler

Für die Auswertung des mittleren absoluten Fehlers wurden die Prädiktionen der Deep-Learning-basierten Modelle, die die räumliche Infarktausbreitung prädizierten, mit einem Schwellenwert von 0.5 binarisiert und anschließend mit Hilfe der Voxelgröße des Atlas in Infarktvolumina umgerechnet. Für jeden Fall wurde anschließend der absolute Fehler aus der Differenz zwischen dem prädizierten Infarktvolumen zum Verlaufszeitpunkt und dem tatsächlichen Infarktvolumen zum Verlaufszeitpunkt berechnet. Anzumerken bleibt, dass diese Modelle nicht anhand des absoluten Fehlers trainiert wurden, sondern diesen nur

indirekt über den Dice-Score und die Binary-Cross-Entropy minimierten. Am besten performte das Multitask-UNet (90d) mit einem mittleren absoluten Fehler von 14.01 ml (siehe Tabelle 22).

Tabelle 22: Mittlerer absoluter Fehler der verschiedenen Prädiktions-Modelle
anhand aller Testfälle der äußeren Schleife gemittelt

Modell	MAE [ml]
Initiale Infarktsegmentierungen	17.54
Dilatierungs-Modell	14.67
3D-nnUNet	14.78
Baseline-UNet	14.73
Multitask-UNet (initial)	14.65
Multitask-UNet (1d)	14.95
Multitask-UNet (90d)	14.01

Alle Modelle wurden anschließend mit Hilfe des einseitigen Wilcoxon-Vorzeichen-Rang-Tests mit dem Multitask-UNet (90d) verglichen (siehe Tabelle 23). Dabei zeigten sich die Prädiktionen des Multitask-UNets (90d) signifikant besser als die initialen Infarktsegmentierungen sowie die Prädiktionen des Dilatierungs-Modells, des 3D-nnUNets, des Baseline-UNets sowie des Multitask-UNets (initial) ($p < 0.05$).

Tabelle 23: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem Multitask-UNet (90d)

Modell	Alternativhypothese: $\tilde{x}_{\text{Multitask-UNet (90d)}} < \tilde{x}_{\text{Modell}}$
Initiale Infarktsegmentierungen	0.0011
Dilatierungs-Modell	0.0070
3D-nnUNet	0.0008
Baseline-UNet	0.0001
Multitask-UNet (initial)	$9 \cdot 10^{-6}$
Multitask-UNet (1d)	0.5636

In Abbildung 31 sind die prädizierten Volumenänderungen gegenüber den tatsächlichen Volumenänderungen aller Fälle des Baseline-UNets sowie der verschiedenen Multitask-UNets aufgetragen. Hierbei fällt auf, dass die Multitask-UNets in verschieden starker Ausprägung dazu tendieren, Fälle mit großer tatsächlicher Volumenänderung besser zu prädizieren, aber dafür einige Fälle mit kleiner Volumenänderung zu groß zu prädizieren.

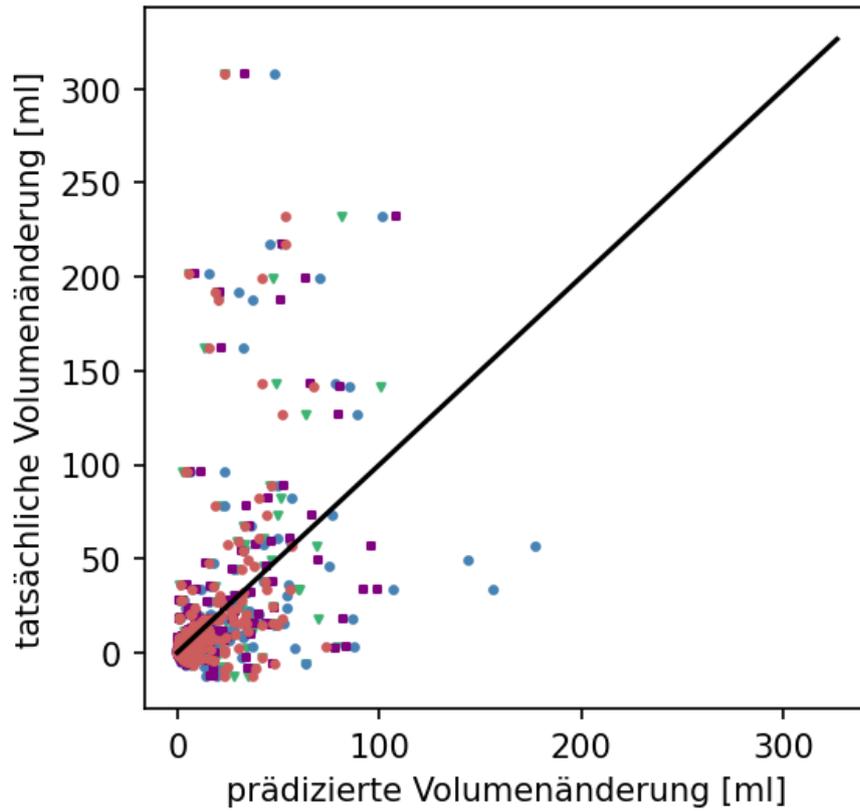


Abbildung 31: Prädizierte und tatsächliche Infarktvolumenänderungen

Dargestellt werden die Volumina aller Fälle. Die Modelle werden durch die unterschiedlichen Farben und Formen repräsentiert. Der **rote Kreis** repräsentiert das Baseline-UNet, das **grüne Dreieck** das Multitask-UNet (initial), der **blaue Kreis** das Multitask-UNet (1d) und das **lila Quadrat** das Multitask-UNet (90d).

5 Diskussion

Die Deep-Learning-basierte Prädiktion der räumlichen Infarktausbreitung nach ischämischem Schlaganfall wurde bereits in mehreren Studien untersucht. Im Rahmen der ISLES-Challenge konnte eine Überlegenheit Deep-Learning-basierter Modelle gegenüber anderen Machine-Learning-Verfahren, wie zum Beispiel Random-Forests, aufgezeigt werden (Winzeck et al. 2018). Die bisher publizierten Studien unterscheiden sich dennoch teils deutlich in mehreren Aspekten. Zum einen wurden die Modelle anhand unterschiedlicher Bilddaten trainiert und evaluiert. Zum anderen unterscheiden sich die verwendeten Modelle in ihrer Architektur. Die meisten Studien verwenden eine UNet-Architektur, die teils durch additive architektonische Features variiert wurde. Die Ergebnisse der einzelnen Studien lassen sich schlecht miteinander vergleichen, da kein frei zugänglicher repräsentativer Datensatz besteht, anhand dessen die Ergebnisse unterschiedlicher Modelle verglichen werden könnten. So bleibt es fraglich, ob sich die Ergebnisse dieser architektonischen Variationen überhaupt generalisieren lassen, da diese häufig ausschließlich anhand eines kleinen Datensatzes erstellt und evaluiert wurden. Ein weiterer Fokus der aktuell publizierten Studien liegt in der Untersuchung, welche Eingabedaten sich für die Prädiktion der Infarktausbreitung eignen. Neuere Studien untersuchen dabei die Verwendung nativer 4D-Perfusiondaten als Eingabedaten. Nur in wenigen Untersuchungen wurden zusätzlich zu den Bilddaten auch weitere klinische Informationen als Eingabedaten genutzt.

Betrachtet man die Studien, die neben bildbasierten Eingabedaten auch klinische Informationen als Eingabedaten verwenden, werden sowohl Gemeinsamkeiten als auch Unterschiede deutlich. Alle Studien verwenden Bildmodalitäten als Eingabedaten, die sowohl den initialen Infarktkern als auch die initiale Penumbra erfassen und abbilden. Die Bilddaten, anhand derer die Ground Truth der Infarktausbreitung im Verlauf segmentiert wurde, unterscheiden sich in der Modalität als auch im Aufnahmezeitpunkt. Auch die klinischen Informationen, welche als Eingabedaten genutzt wurden, unterscheiden sich von Studie zu Studie. Obwohl alle Modelle Convolutional-Layer zur Extrahierung von Bildmerkmalen verwenden, unterscheiden sich die gewählten Modell-Architekturen deutlich. Damit einhergehend unterscheidet sich auch die Art und Weise, mit der die klinischen Eingabeparameter in die Modelle integriert wurden (Choi et al. 2016;

Lucas et al. 2018a; Pinto et al. 2018; Robben et al. 2020; Wouters et al. 2022). Der Ansatz von Lucas et al., die Infarktausbreitung mit Hilfe eines Convolutional-Auto-Encoders durch eine lineare Transformation der initialen Segmentierungen von Infarktkern und Penumbra im Latent-Space zu prädizieren, hebt sich dabei deutlich von den restlichen Ansätzen ab (Lucas et al. 2018a). Choi et al., Robben et al. und Wouters et al. verwendeten Fully-Convolutional-Networks und verbanden die extrahierten Merkmale der klinischen Informationen mit den extrahierten Merkmalen weiterer Eingabedaten, um diese anschließend zusammen weiter zu transformieren (Choi et al. 2016; Robben et al. 2020; Wouters et al. 2022). Pinto et al. und Lucas et al. verwendeten UNet-Modelle und integrierten die klinischen Informationen über Feature-Maps direkt in die UNet-Architektur. Während Pinto et al. den TICl-Score nach dem Decoder-Abschnitt integrierten und so zusammen mit der Verwendung einer Verlustfunktion, die sich während des Trainings in Abhängigkeit von dem TICl-Score des zu prädizierenden Falles definierte, den Dice-Score verbessern konnten, integrierten Lucas et al. die klinischen Informationen zusätzlich zu den Bilddaten als Eingabedaten des Encoders, was zu einer Verschlechterung des Dice-Scores führte (Lucas et al. 2018a; Pinto et al. 2018). Ein Nachteil des Ansatzes von Pinto et al. ist, dass mit dem TICl-Score nur ein einziger klinischer Parameter in das Modell und die Verlustfunktion integriert wurde (Pinto et al. 2018). In keiner dieser Studien wurde untersucht, inwiefern die integrierten klinischen Informationen mit dem Infarktwachstum oder dem Infarktvolumen zum Verlaufszeitpunkt korrelieren (Choi et al. 2016; Lucas et al. 2018a; Pinto et al. 2018; Robben et al. 2020; Wouters et al. 2022). So lässt sich aus diesen Beispielen keine allgemeine Kausalität ableiten, welche klinischen Informationen die Prädiktionsgenauigkeit verbessern können und wie diese additiv in eine UNet-Architektur zu integrieren sind.

Infolgedessen ergeben sich verschiedene, bisher unbeantwortete Fragen. Es ist unklar, welche Modell-Architektur sich am besten dafür eignet, neben bildbasierten Eingabedaten auch klinische Informationen zu verarbeiten. Legt man sich auf die Verwendung eines UNet-Modells fest, ist es unklar, auf welche Art und Weise additive klinische Informationen am besten zu integrieren sind. Zum einen haben Pinto et al. gezeigt, dass es möglich ist, den eigentlichen Lernprozess des Netzwerkes über die Verlustfunktion in Abhängigkeit der klinischen Informationen zu beeinflussen (Pinto et al. 2018). Zum anderen können klinische Informationen

aber auch direkt als Eingabedaten über Feature-Maps in ein Modell integriert werden. Es ist jedoch unklar, an welchem Punkt eine Integration dieser Feature-Maps am sinnvollsten ist und in welcher Weise mehrere klinische Informationen integriert werden können. Zudem ist es unklar, ob sich die Funktionsweise eines UNets überhaupt dafür eignet, relevante Zusammenhänge aus mehreren klinischen Feature-Maps zu extrahieren.

In der vorliegenden Studie wurde das räumliche Infarktwachstum anhand von 270 Fällen prädiziert und ausgewertet. Der verwendete Datensatz entstammt der WAKE-UP-Studie (Thomalla et al. 2018). Als Eingabedaten standen den verschiedenen Prädiktions-Modellen für jeden Fall eine initiale ADC- und FLAIR-Map sowie die Segmentierung des initialen Infarktgebietes zu Verfügung. Als Ground Truth wurde die Segmentierung des Infarktgebietes verwendet, die anhand von nach 22-36 Stunden erstellten ADC-Maps erstellt wurde. Das Training und die anschließende Evaluation aller verwendeten Deep-Learning-basierten Methoden und Modelle erfolgte anhand derselben verschachtelten Kreuzvalidierung. Die Hyperparameter der verschiedenen Modelle wurden so unabhängig von den Testfällen für jeden Fold der äußeren Schleife anhand einer 4-Fold-Kreuzvalidierung optimiert. Die Evaluation der verschiedenen Modelle erfolgte anschließend anhand des über alle 270 Testfälle der äußeren Schleife gemittelten Dice-Scores.

Das 3D-nnUNet-Modell performte von allen Modellen, die ausschließlich anhand bildbasierter Eingabedaten trainiert wurden, am besten und erreichte einen Dice-Score von 0.5344 bei einer Standardabweichung von 0.2150. Ein selbst erstelltes Baseline-UNet-Modell, welches in Anlehnung an das 3D-nnUNet-Modell erstellt wurde, erreichte mit einem Dice-Score von 0.5334 bei einer Standardabweichung von 0.2187 ein vergleichbares Ergebnis und unterschied sich damit nicht signifikant von den Prädiktionen des 3D-nnUNet-Modells. Ein Dilatierungs-Modell, welches die Infarktausbreitung mittels Dilatierung der Segmentierung des initialen Infarktgebietes prädizierte, erreichte einen Dice-Score von 0.5068 bei einer Standardabweichung von 0.2075. Die initiale Infarktsegmentierung wies im Vergleich mit der Infarktsegmentierung der Bildgebung im Verlauf einen Dice-Score von 0.4575 bei einer Standardabweichung von 0.2014 auf. Die Prädiktionen des 3D-nnUNet-Modells unterschieden sich dabei signifikant von den Prädiktionen des Dilatierungs-Modells sowie von den initialen Infarktsegmentierungen.

Als klinische Informationen lagen für jeden Fall des Datensatzes die Therapiegruppe, das Alter, das initiale Infarktvolumen, das Infarktvolumen nach 22 bis 36 Stunden, der NIHSS-Score bei Aufnahme, nach 22 bis 36 Stunden sowie der mRS nach 90 Tagen vor. Insgesamt wurden drei verschiedene Modelle trainiert, in die die klinischen Informationen zusätzlich zu den initialen Bilddaten integriert wurden. Dabei wurden in ein Modell ausschließlich die klinischen Informationen integriert, die zum initialen Zeitpunkt verfügbar waren (Multitask-UNet (initial)). In ein weiteres Modell wurde zusätzlich der NIHSS-Score integriert, der zum Zeitpunkt der Verlaufsbildgebung erhoben wurde (Multitask-UNet (1d)). In ein drittes Modell wurde ergänzend auch der mRS-Score integriert, der im Verlauf nach 90 Tagen erhoben wurde (Multitask-UNet (90d)). Im Rahmen der Optimierung des Integrationsortes innerhalb der verschachtelten Kreuzvalidierung zeigte sich dabei für jedes Modell und jeden Fold die Integration am Ende des Decoders oder auf jeder Ebene des Decoders am performantesten. Das Multitask-UNet (initial)-Modell erreichte einen Dice-Score von 0.5340 bei einer Standardabweichung von 0.2171 und unterschied sich somit nicht signifikant von dem Baseline-UNet-Modell, welches ausschließlich die initialen Bilddaten zur Prädiktion verwendete. Das Multitask-UNet (1d)-Modell erreichte einen Dice-Score von 0.5384 bei einer Standardabweichung von 0.2127. Das Multitask-UNet (90d)-Modell erreichte einen Dice-Score von 0.5394 bei einer Standardabweichung von 0.2140. Die Prädiktionen dieser beiden Modelle unterschieden sich dabei signifikant von den Prädiktionen des Baseline-UNet-Modells sowie des 3D-nnUNet-Modells. Dabei bleibt anzumerken, dass ausschließlich das Multitask-UNet (1d)-Modell eine Prädiktion zum initialen Zeitpunkt ermöglicht. Nichtsdestotrotz konnte somit gezeigt werden, dass im Rahmen der verfügbaren Bilddaten und klinischen Informationen sich die klinischen Informationen am besten auf jeder Ebene des Decoders oder am Ende des Decoders in ein UNet-Modell zur Prädiktion der räumlichen Infarktausbreitung integrieren lassen und dass eine Integration mehrerer klinischer Informationen auf diese Weise möglich ist.

Anzumerken bleibt, dass der verwendete Datensatz keine Bilddaten enthielt, die die Penumbra zum initialen Zeitpunkt erfassen konnten. Dementsprechend konnten nur bildbasierte Eingabedaten zur Prädiktion genutzt werden, die den Infarktkern zum initialen Zeitpunkt erfassten. Es ist möglich, dass sich der optimale

Ort der Integration der klinischen Informationen ändert, wenn zusätzlich Eingabedaten verwendet werden, die Informationen über die Penumbra enthalten. Zudem könnte sich der Einfluss der klinischen Informationen auf die Prädiktion durch die Hinzunahme dieser Informationen verstärken. Die Integration von klinischen Informationen sollte daher zukünftig mit Hilfe eines Datensatzes untersucht werden, dessen initiale Bilddaten sowohl den Infarktkern als auch die Penumbra abbilden können. Bei der testweisen Integration der tatsächlichen Infarkt volumenänderung als klinische Information erwies sich der Integrationsort auf jeder Ebene des Encoders oder vor dem Encoder als optimal. Dies könnte damit zusammenhängen, dass das tatsächliche Läsionswachstum die Prädiktion der räumlichen Infarktausbreitung deutlich stärker beeinflusst als die Integration der klinischen Informationen, die im besten Falle mit dem Läsionswachstum korrelieren. So ist es auch möglich, dass die Stärke der Korrelation der klinischen Informationen den optimalen Integrationsort beeinflusst. Des Weiteren sind in diesem Zusammenhang die verfügbaren und verwendeten klinischen Informationen zu nennen. Manche der verwendeten klinischen Informationen wurden nicht zum Zeitpunkt der initialen Bildgebung, sondern erst im klinischen Verlauf erhoben. Demensprechend ermöglicht ausschließlich das Multitask-UNet (initial)-Modell eine echte Prädiktion zum Zeitpunkt der initialen Bildgebung. Zudem erfolgte keine Korrektur für multiples Testen, folglich wird der Fehler 1. Art unterschätzt.

Zu den Stärken dieser Studie zählt zum einen die Größe des verwendeten Datensatzes. Mit 270 prädizierten Fällen ist diese Studie aktuell die zweitgrößte Studie, die die Infarktausbreitung nach einem ischämischen Schlaganfall anhand von Deep-Learning-basierten Modellen untersucht. Zudem ist es die bisher einzige Studie, die die Integration mehrerer klinischer Informationen in unterschiedlicher Zusammensetzung an verschiedenen Integrationsorten in ein UNet-Model strukturiert untersucht. Der Integrationsort wird dabei erstmalig innerhalb einer verschachtelten Kreuzvalidierung als Hyperparameter unabhängig von den Testdaten optimiert. Eine weitere Stärke dieser Studie ist, dass die Ergebnisse mit den Prädiktionsergebnissen der nnUNet-Methode und eines Dilatierungs-Modells sowie mit den initialen Infarktsegmentierungen in einen Kontext gesetzt werden.

Grundlegend bestehen aktuell jedoch weiterhin einige Unklarheiten bezüglich der Deep-Learning-basierten räumlichen Prädiktion von ischämischen Infarkten. Es ist

unklar, anhand welcher Bilddaten eine optimale Ground Truth zu erstellen ist. Zwar konnten bereits verschiedene Studien Zusammenhänge zwischen der Ausbreitung des Infarktgebietes zu verschiedenen Zeitpunkten und verschiedenen definierten klinischen Outcomes nachweisen (Yoo et al. 2012; Zaidi et al. 2012), dennoch gibt es bisher keine allgemein gültige Definition, wie die Ausbreitung eines Infarktes als Outcome zu definieren ist, anhand welcher Bilddaten die Infarktausbreitung segmentiert werden sollte und zu welchem Zeitpunkt das Infarktgebiet am besten mit klinischen Scores korreliert. In den bisherigen Studien wurden die Bilddaten zur Segmentierung der Ground Truth verwendet, die in den verfügbaren Datensätzen vorhanden waren. Die Zeitspanne des Aufnahmezeitpunktes der Bildgebung, anhand derer die Ground Truth segmentiert wurde, reicht so von einem Tag bis zu 3 Monaten nach stattgehabtem Schlaganfall (Robben et al. 2020; Winzeck et al. 2018). Zudem ist es unklar, welche klinischen Informationen die Prädiktion der Infarktausbreitung positiv beeinflussen können. In einigen Studien, die das Nutzen einer endovaskulären Therapie untersuchten, konnte ein signifikanter Zusammenhang auf die Infarktausbreitung nachgewiesen werden (Campbell et al. 2015; Jovin et al. 2015; Nogueira et al. 2018). Des Weiteren korreliert der Grad der Rekanalisierung nach stattgehabter Therapie mit dem Infarktvolumen im Verlauf (Campbell et al. 2012b; Krongold et al. 2015). Der erzielte Grad der Rekanalisierung ist dabei wiederum abhängig von der Time-To-Treatment (Bourcier et al. 2019). Der Grad der Kollateralen konnte zudem als von der Rekanalisierung unabhängiger Prädiktor des Infarktvolumens beschrieben werden (Angermaier et al. 2011). Ein hoher initialer NIHSS-Score und eine Hyperglykämie sind mit einem großen Infarktwachstum assoziiert (Nogueira et al. 2017).

Aus den Ergebnissen dieser Untersuchung lässt sich ableiten, dass ein UNet-Modell in der Lage ist, Informationen aus mehreren numerischen klinischen Informationen zu extrahieren und diese für die räumliche Prädiktion zu nutzen. Entscheidend sind hierbei zum einen die Auswahl geeigneter klinischer Informationen sowie der verwendete Integrationsort. In dieser Studie konnten die Prädiktionsergebnisse durch die Integration von klinischen Informationen im Vergleich zu einem 3D-nnUNet-Modell, welches ausschließlich bildbasierte Eingangsdaten verarbeitete, signifikant verbessert werden. Der Integrationsort der klinischen Informationen wurde dabei im Rahmen einer verschachtelten

Kreuzvalidierung unabhängig von den Testfällen des jeweiligen Folds der äußeren Schleife optimiert. Als optimaler Integrationsort zeigte sich dabei die Integration der klinischen Informationen auf jeder Ebene des Decoders oder am Ende des Decoders.

6 Zusammenfassung

Fragestellung

Kann die Prädiktion der räumlichen Infarktausbreitung anhand eines UNet-Modells durch die Integration mehrerer klinischer Informationen verbessert werden und auf welche Weise sollten diese in das UNet-Modell integriert werden?

Material und Methoden

Anhand von 270 Fällen wurden verschiedene UNet-Modelle unter der Verwendung einer verschachtelten Kreuzvalidierung trainiert und evaluiert. Dabei wurden einige Modelle ausschließlich anhand bildbasierter Eingabedaten trainiert. Zusätzlich verwendeten andere Modelle additiv klinische Informationen. Als bildbasierte Eingabedaten waren für jeden Fall initiale ADC- und FLAIR-Maps sowie eine initiale Infarktsegmentierung vorhanden. Als klinische Informationen lagen für jeden Fall die Therapiegruppe, das Alter, das initiale Infarktvolumen, der NIHSS-Score bei Aufnahme, nach 22 bis 36 Stunden nach Aufnahme sowie der mRS-Score nach 90 Tagen vor. Als Ground Truth dienten Infarktsegmentierungen, die anhand von ADC-Maps erstellt wurden, welche nach 22 bis 36 Stunden aufgenommen wurden. Die Auswertung der verschiedenen Modelle erfolgte anhand des über alle Fälle gemittelten Dice-Scores sowie zwischen den Modellen anhand des einseitigen Wilcoxon-Vorzeichen-Rang-Tests.

Ergebnisse

Durch die Integration der klinischen Informationen, die zum initialen Zeitpunkt verfügbar waren, konnte der Dice-Score im Vergleich zu dem Baseline-UNet-Modell geringfügig verbessert werden. Durch die zusätzliche Integration von klinischen Informationen, die im Verlauf erhoben wurden, konnten die Prädiktionen signifikant im Vergleich zu dem Baseline-UNet-Modell sowie zu einem 3D-nnUNet-Modell verbessert werden. Der verwendete Integrationsort wurde dabei als Hyperparameter unabhängig von den Testdaten im Rahmen einer verschachtelten Kreuzvalidierung optimiert. Als optimal zeigte sich dabei die Integration auf jeder Ebene des Decoders oder am Ende des Decoders.

Schlussfolgerung

Ein UNet-Modell ist in der Lage, Informationen aus multiplen, klinischen Parametern zu extrahieren und diese für die räumliche Prädiktion der Schlaganfallausbreitung zu nutzen. Die Integration von klinischen Informationen sollte zukünftig strukturiert anhand eines Datensatzes untersucht werden, dessen bildbasierte Eingangsdaten sowohl den Infarktkern als auch die Penumbra abbilden und der viele klinische Informationen enthält, die zum initialen Zeitpunkt erhoben wurden.

Abstract

Objective

Can the spatial prediction of the infarct core using a UNet model be improved by integrating multiple clinical parameters and in what way should these be integrated into the UNet model?

Materials and Methods

Various UNet models were trained and evaluated using a nested cross-validation approach based on 270 cases. Some models were exclusively trained on image-based input data, while others additionally incorporated clinical information. Image-based input data for each case included initial ADC and FLAIR maps as well as an initial infarct segmentation. Clinical information for each case included treatment group, age, initial infarct volume, NIHSS score at admission and 22 to 36 hours after admission, as well as mRS score after 90 days. Infarct segmentations created using ADC maps generated after 22 to 36 hours served as the ground truth. The evaluation of different models was based on the mean Dice score across all cases and between the models based on the one-sided Wilcoxon signed-rank test.

Results

By integrating clinical information available at the initial time point, the Dice score was slightly improved compared to the baseline UNet model. Furthermore, the additional integration of clinical information collected over time significantly enhanced predictions compared to both the baseline UNet model and a 3D-nnUNet model. The integration site was treated as a hyperparameter and optimized independently of the test data through nested cross-validation. The optimal approach was found to be integration at each level of the decoder or at the end of the decoder.

Conclusion

A UNet model is capable of extracting information from multiple clinical parameters and utilizing them for spatial prediction. The integration of clinical information should be systematically examined in the future using a dataset that includes image-based input data representing both the infarct core and the penumbra, along with a wealth of clinical information collected at the initial time point.

7 Anhang

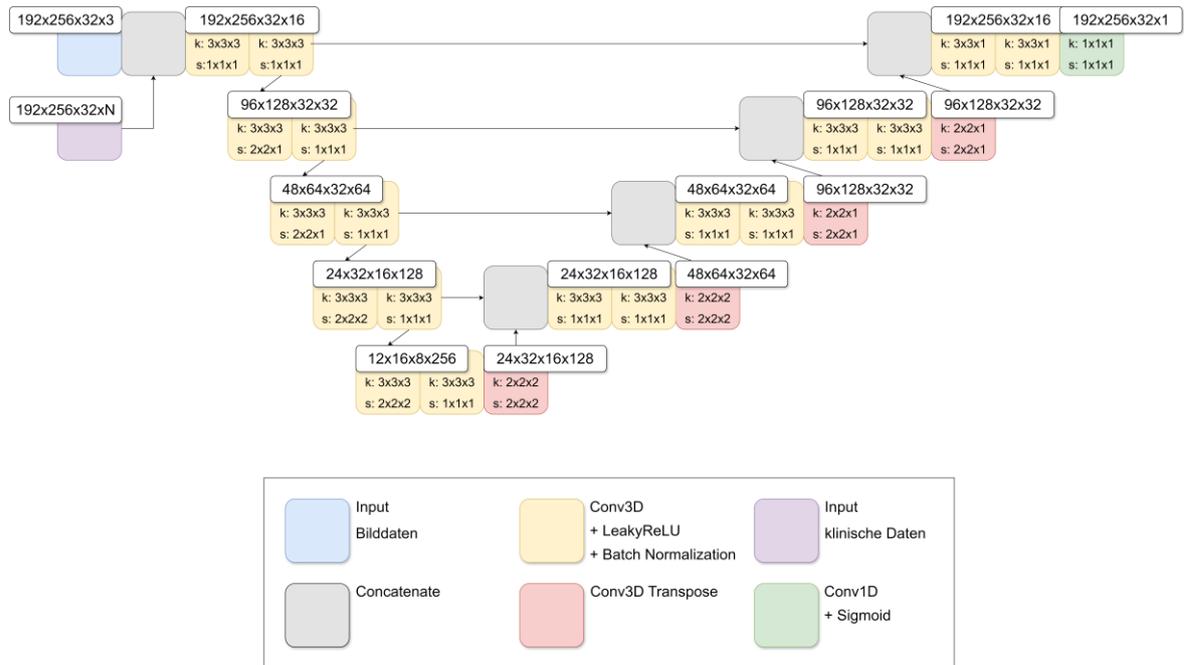


Abbildung 32: Integration klinischer Informationen vor dem Encoder

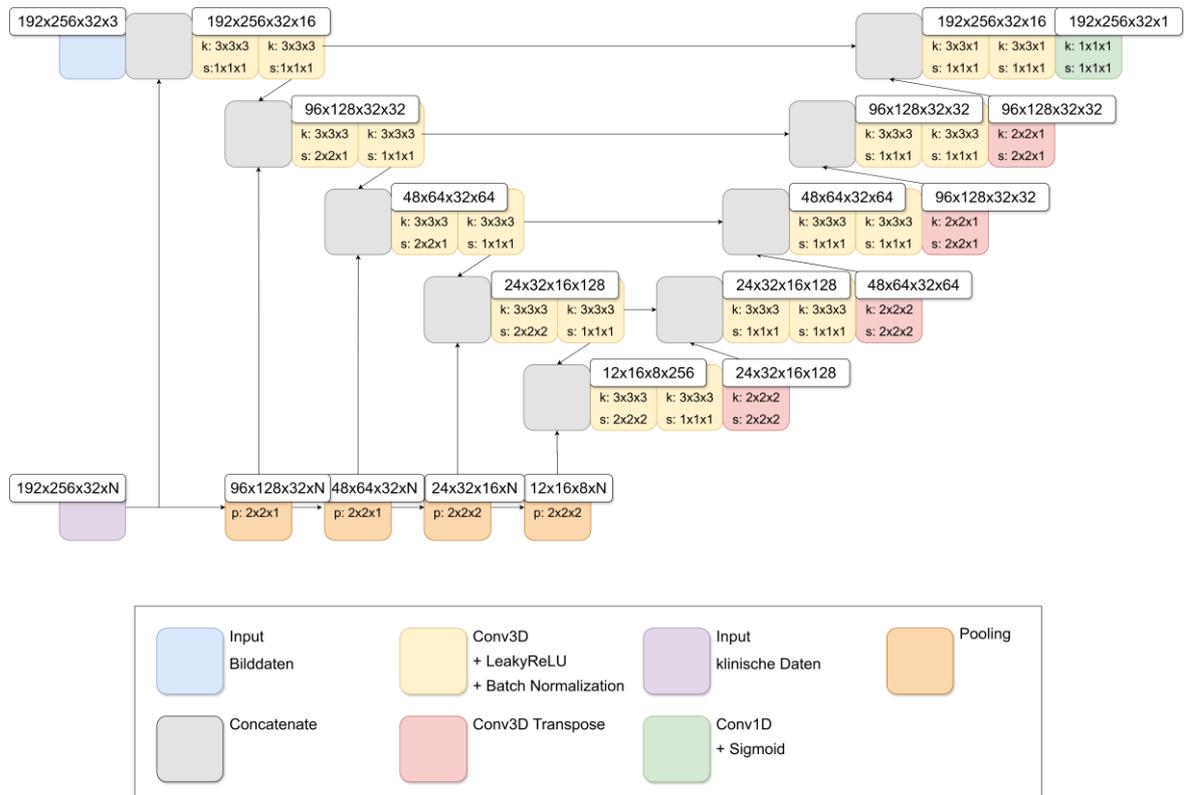


Abbildung 33: Integration klinischer Informationen auf jeder Ebene des Encoders

8 Literaturverzeichnis

- Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M (2016) **Tensorflow: A system for large-scale machine learning**. In: 12th USENIX6 Symposium on Operating Systems Design and Implementation. 5OSDI6 2016, 265–283.
- Abedi V, Goyal N, Tsivgoulis G, Hosseinichimeh N, Hontecillas R, Bassaganya-Riera J, Elijevich L, Metter JE, Alexandrov AW, Liebeskind DS, Alexandrov AV, Zand R (2017) **Novel Screening Tool for Stroke Using Artificial Neural Network**. *Stroke* 48, 1678–1681.
- Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, McTaggart RA, Torbey MT, Kim-Tenser M, Leslie-Mazwi T, Sarraj A, Kasner SE, Ansari SA, Yeatts SD, Hamilton S, Mlynash M, Heit JJ, Zaharchuk G, Kim S, Carrozzella J, Palesch YY, Demchuk AM, Bammer R, Lavori PW, Broderick JP, Lansberg MG (2018) **Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging**. *N Engl J Med* 378, 708–718.
- Amador K, Wilms M, Winder A, Fiehler J, Forkert ND (2021) **Stroke Lesion Outcome Prediction Based on 4D CT Perfusion Data Using Temporal Convolutional Networks**. In: Proceedings of the 4th Conference on Medical Imaging with Deep Learning. PMLR 2021, 22–33.
- Angermaier A, Langner S, Kirsch M, Kessler C, Hosten N, Khaw AV (2011) **CT-angiographic collateralization predicts final infarct volume after intra-arterial thrombolysis for acute anterior circulation ischemic stroke**. *Cerebrovasc Dis* 31, 177–184.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011) **A reproducible evaluation of ANTs similarity metric performance in brain image registration**. *Neuroimage* 54, 2033–2044.
- Ay H, Arsava EM, Rosand J, Furie KL, Singhal AB, Schaefer PW, Wu O, Gonzalez RG, Koroshetz WJ, Sorensen AG (2008) **Severity of leukoaraiosis and susceptibility to infarct growth in acute stroke**. *Stroke* 39, 1409–1413.
- Badrinarayanan V, Kendall A, Cipolla R (2017) **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation**. *IEEE Trans Pattern Anal Mach Intell* 39, 2481–2495.
- Bagher-Ebadian H, Jafari-Khouzani K, Mitsias PD, Lu M, Soltanian-Zadeh H, Chopp M, Ewing JR (2011) **Predicting final extent of ischemic infarction using artificial neural network analysis of multi-parametric MRI in patients with stroke**. *PLOS ONE* 6, e22626.
- Barber PA, Demchuk AM, Zhang J, Buchan AM (2000) **Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy**. *Lancet* 355, 1670–1674.

- Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JHW, Hudon ME, Tomanek A, Frayne R, Buchan AM (2005) **Imaging of the brain in acute ischaemic stroke: comparison of computed tomography and magnetic resonance diffusion-weighted imaging.** *J Neurol Neurosurg Psychiatry* 76, 1528–1533.
- Barrett KM, Ding YH, Wagner DP, Kallmes DF, Johnston KC (2009) **Change in diffusion-weighted imaging infarct volume predicts neurologic outcome at 90 days: results of the Acute Stroke Accurate Prediction (ASAP) trial serial imaging substudy.** *Stroke* 40, 2422–2427.
- Beaulieu C, Crespigny A de, Tong DC, Moseley ME, Albers GW, Marks MP (1999) **Longitudinal magnetic resonance imaging study of perfusion and diffusion in stroke: Evolution of lesion volume and correlation with clinical outcome.** *Ann Neurol* 46, 568–578.
- Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, Schonewille WJ, Vos JA, Nederkoorn PJ, Wermer MJH, van Walderveen MAA, Staals J, Hofmeijer J, van Oostayen JA, Lycklama à Nijeholt GJ, Boiten J, Brouwer PA, Emmer BJ, Bruijn SF de, van Dijk LC, Kappelle LJ, Lo RH, van Dijk EJ, Vries J de, Kort PLM de, van Rooij WJJ, van den Berg JSP, van Hasselt BAAM, Aerden LAM, Dallinga RJ, Visser MC, Bot JCJ, Vroomen PC, Eshghi O, Schreuder THCML, Heijboer RJJ, Keizer K, Tielbeek AV, Hertog HM den, Gerrits DG, van den Berg-Vos RM, Karas GB, Steyerberg EW, Flach HZ, Marquering HA, Sprengers MES, Jenniskens SFM, Beenen LFM, van den Berg R, Koudstaal PJ, van Zwam WH, Roos, Yvo B. W. E. M., van der Lugt A, van Oostenbrugge RJ, Majoie, Charles B. L. M., Dippel DWJ (2015) **A randomized trial of intraarterial treatment for acute ischemic stroke.** *N Engl J Med* 372, 11–20.
- Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, Krakauer JW, Boyd LA, Carmichael ST, Corbett D, Cramer SC (2017) **Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce.** *Int J Stroke* 12, 444–450.
- Birenbaum D, Bancroft LW, Felsberg GJ (2011) **Imaging in acute stroke.** *West J Emerg Med* 12, 67–76.
- Bourcier R, Goyal M, Liebeskind DS, Muir KW, Desal H, Siddiqui AH, Dippel DWJ, Majoie CB, van Zwam WH, Jovin TG, Levy EI, Mitchell PJ, Berkhemer OA, Davis SM, Derraz I, Donnan GA, Demchuk AM, van Oostenbrugge RJ, Kelly M, Roos YB, Jahan R, van der Lugt A, Sprengers M, Velasco S, Lycklama à Nijeholt GJ, Ben Hassen W, Burns P, Brown S, Chabert E, Krings T, Choe H, Weimar C, Campbell BCV, Ford GA, Ribo M, White P, Cloud GC, San Roman L, Davalos A, Naggara O, Hill MD, Bracard S (2019) **Association of Time From Stroke Onset to Groin Puncture With Quality of Reperfusion After Mechanical Thrombectomy: A Meta-analysis of Individual Patient Data From 7 Randomized Clinical Trials.** *JAMA Neurol* 76, 405–411.

- Bracard S, Ducrocq X, Mas JL, Soudant M, Oppenheim C, Moulin T, Guillemin F (2016) **Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial.** *Lancet Neurol* 15, 1138–1147.
- Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V (1989) **Measurements of acute cerebral infarction: a clinical examination scale.** *Stroke* 20, 864–870.
- Burda JE, Sofroniew MV (2014) **Reactive gliosis and the multicellular response to CNS damage and disease.** *Neuron* 81, 229–248.
- Campbell BCV, Christensen S, Levi CR, Desmond PM, Donnan GA, Davis SM, Parsons MW (2011) **Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core.** *Stroke* 42, 3435–3440.
- Campbell BCV, Christensen S, Levi CR, Desmond PM, Donnan GA, Davis SM, Parsons MW (2012a) **Comparison of computed tomography perfusion and magnetic resonance imaging perfusion-diffusion mismatch in ischemic stroke.** *Stroke* 43, 2648–2653.
- Campbell BCV, Ma H, Ringleb PA, Parsons MW, Churilov L, Bendszus M, Levi CR, Hsu C, Kleinig TJ, Fatar M, Leys D, Molina C, Wijeratne T, Curtze S, Dewey HM, Barber PA, Butcher KS, Silva DA de, Bladin CF, Yassi N, Pfaff JAR, Sharma G, Bivard A, Desmond PM, Schwab S, Schellinger PD, Yan B, Mitchell PJ, Serena J, Toni D, Thijs V, Hacke W, Davis SM, Donnan GA (2019) **Extending thrombolysis to 4-5-9 h and wake-up stroke using perfusion imaging: a systematic review and meta-analysis of individual patient data.** *Lancet* 394, 139–147.
- Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, Yan B, Dowling RJ, Parsons MW, Oxley TJ, Wu TY, Brooks M, Simpson MA, Miteff F, Levi CR, Krause M, Harrington TJ, Faulder KC, Steinfort BS, Priglinger M, Ang T, Scroop R, Barber PA, McGuinness B, Wijeratne T, Phan TG, Chong W, Chandra RV, Bladin CF, Badve M, Rice H, Villiers L de, Ma H, Desmond PM, Donnan GA, Davis SM (2015) **Endovascular therapy for ischemic stroke with perfusion-imaging selection.** *N Engl J Med* 372, 1009–1018.
- Campbell BCV, Tu HTH, Christensen S, Desmond PM, Levi CR, Bladin CF, Hjort N, Ashkanian M, Sølling C, Donnan GA, Davis SM, Ostergaard L, Parsons MW (2012b) **Assessing response to stroke thrombolysis: validation of 24-hour multimodal magnetic resonance imaging.** *Arch Neurol* 69, 46–50.
- Carrera E, Wintermark M (2017) **Imaging-based selection of patients for acute stroke treatment: Is it ready for prime time?** *Neurology* 88, 2242–2243.
- Chalela JA, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L, Warach S (2007) **Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison.** *Lancet* 369, 293–298.

- Chaudhary K, Poirion OB, Lu L, Garmire LX (2018) **Deep Learning–Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer.** Clin Cancer Res 24, 1248–1259.
- Chen L, Bentley P, Rueckert D (2017a) **Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks.** Neuroimage Clin 15, 633–643.
- Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017b) **DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.** IEEE Trans Pattern Anal Mach Intell 40, 834–848.
- Cheng X, Su X, Shi J, Liu Q, Zhou C, Dong Z, Xing W, Lu H, Pan C, Li X, Yu Y, Zhang L, Lu G (2021) **Comparison of automated and manual DWI-ASPECTS in acute ischemic stroke: total and region-specific assessment.** Eur Radiol 31, 4130–4137.
- Cho K, van Merriënboer B, Bahdanau D, Bengio Y (2014) **On the Properties of Neural Machine Translation: Encoder-Decoder Approaches.** In: Association for Computational Linguistics (Hg.) Proceedings of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation. SSST 2014.
- Cho T-H, Nighoghossian N, Mikkelsen IK, Derex L, Hermier M, Pedraza S, Fiehler J, Østergaard L, Berthezène Y, Baron J-C (2015) **Reperfusion within 6 hours outperforms recanalization in predicting penumbra salvage, lesion growth, final infarct, and clinical outcome.** Stroke 46, 1582–1589.
- Choi Y, Kwon Y, Lee H, Kim BJ, Paik MC, Won J-H (2016) **Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke.** In: Crimi A (Hg.) Proceedings of the International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2016, 231–243.
- Chollet F (2015) **Keras**, <https://keras.io>.
- Compagne KCJ, Boers AMM, Marquering HA, Berkhemer OA, Yoo AJ, Beenen LFM, van Oostenbrugge RJ, van Zwam WH, Roos YBWEM, Majoie CB, van Es ACGM, van der Lugt A, Dippel DWJ, Lingsma H (2019) **Follow-up infarct volume as a mediator of endovascular treatment effect on functional outcome in ischaemic stroke.** Eur Radiol 29, 736–744.
- Copen WA, Schaefer PW, Wu O (2011) **MR perfusion imaging in acute ischemic stroke.** Neuroimaging Clin N Am 21, 259-283.
- Crichton SL, Bray BD, McKeivitt C, Rudd AG, Wolfe CDA (2016) **Patient outcomes up to 15 years after stroke: survival, disability, quality of life, cognition and mental health.** J Neurol Neurosurg Psychiatry 87, 1091–1098.

- Davis SM, Donnan GA, Parsons MW, Levi C, Butcher KS, Peeters A, Barber PA, Bladin C, Silva DA de, Byrnes G, Chalk JB, Fink JN, Kimber TE, Schultz D, Hand PJ, Frayne J, Hankey G, Muir K, Gerraty R, Tress BM, Desmond PM (2008) **Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial.** *Lancet Neurol* 7, 299–309.
- Debs N, Cho T-H, Rousseau D, Berthezène Y, Buisson M, Eker O, Mechtouff L, Nighoghossian N, Ovize M, Frindel C (2021) **Impact of the reperfusion status for predicting the final stroke infarct using deep learning.** *Neuroimage Clin* 29, 102548.
- Demeestere J, Wouters A, Christensen S, Lemmens R, Lansberg MG (2020) **Review of Perfusion Imaging in Acute Ischemic Stroke: From Time to Tissue.** *Stroke* 51, 1017–1024.
- Drewer-Gutland F, Kemmling A, Ligges S, Ritter M, Dziewas R, Ringelstein EB, Niederstadt TU, Heindel W, Heßelmann V (2015) **CTP-Based Tissue Outcome: Promising Tool to Prove the Beneficial Effect of Mechanical Recanalization in Acute Ischemic Stroke.** *Rofo* 187, 459–466.
- Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, Brott T, Cohen G, Davis S, Donnan G, Grotta J, Howard G, Kaste M, Koga M, Kummer R von, Lansberg M, Lindley RI, Murray G, Olivot JM, Parsons M, Tilley B, Toni D, Toyoda K, Wahlgren N, Wardlaw J, Whiteley W, del Zoppo GJ, Baigent C, Sandercock P, Hacke W (2014) **Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials.** *Lancet* 384, 1929–1935.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) **Dermatologist-level classification of skin cancer with deep neural networks.** *Nature* 542, 115–118.
- Feigin VL, Barker-Collo S, Parag V, Senior H, Lawes CMM, Ratnasabapathy Y, Glen E (2010) **Auckland Stroke Outcomes Study. Part 1: Gender, stroke types, ethnicity, and functional outcomes 5 years poststroke.** *Neurology* 75, 1597–1607.
- Ferguson M, Ak R, Lee Y-TT, Law KH (2017) **Automatic localization of casting defects with convolutional neural networks.** In: IEEE (Hg.) Proceedings of the IEEE International Conference on Big Data. *Big Data* 2017, 1726–1735.
- 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 (2002) **CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke.** *Stroke* 33, 2206–2210.
- Fink JN, Kumar S, Horkan C, Linfante I, Selim MH, Caplan LR, Schlaug G (2002) **The stroke patient who woke up: clinical and radiological features, including diffusion and perfusion MRI.** *Stroke* 33, 988–993.

- Fonov VS, Evans AC, McKinstry RC, Almlí CR, Collins DL (2009) **Unbiased nonlinear average age-appropriate brain templates from birth to adulthood.** *Neuroimage* 47, S102.
- Gaudinski MR, Henning EC, Miracle A, Luby M, Warach S, Latour LL (2008) **Establishing final infarct volume: stroke lesion evolution past 30 days is insignificant.** *Stroke* 39, 2765–2768.
- GBD-2015-Mortality-and-Causes-of-Death-Collaborators (2016) **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* 388, 1459–1544.
- Glorot X, Bordes A, Bengio Y (2011) **Deep Sparse Rectifier Neural Networks.** In: Gordon G, Dunson D (Hg.) *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics.* AISTATS 2011, 315–323.
- Goodfellow I, Courville A, Bengio Y (2016) **Deep learning.** The MIT Press, Cambridge, Massachusetts.
- Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, Roy D, Jovin TG, Willinsky RA, Sapkota BL, Dowlatshahi D, Frei DF, Kamal NR, Montanera WJ, Poppe AY, Ryckborst KJ, Silver FL, Shuaib A, Tampieri D, Williams D, Bang OY, Baxter BW, Burns PA, Choe H, Heo J-H, Holmstedt CA, Jankowitz B, Kelly M, Linares G, Mandzia JL, Shankar J, Sohn S-I, Swartz RH, Barber PA, Coutts SB, Smith EE, Morrish WF, Weill A, Subramaniam S, Mitha AP, Wong JH, Lowerison MW, Sajobi TT, Hill MD (2015) **Randomized assessment of rapid endovascular treatment of ischemic stroke.** *N Engl J Med* 372, 1019–1030.
- Goyal M, Menon BK, Derdeyn CP (2013) **Perfusion imaging in acute ischemic stroke: let us improve the science before changing clinical practice.** *Radiology* 266, 16–21.
- Guenego A, Fahed R, Albers GW, Kuraitis G, Sussman ES, Martin BW, Marcellus DG, Olivot J-M, Marks MP, Lansberg MG, Wintermark M, Heit JJ (2020) **Hypoperfusion intensity ratio correlates with angiographic collaterals in acute ischaemic stroke with M1 occlusion.** *Eur J Neurol* 27, 864–870.
- Günster C (2011) **Schlaganfallversorgung in Deutschland - Inzidenz, Wiederaufnahmen, Mortalität und Pflegerisiko im Spiegel von Routinedaten.** In: Günster C, Klose, Joachim, Klose, Schmacke N (Hg.) *Versorgungs-Report 2011,* Schattauer, Stuttgart, 147–163.
- Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, Larrue V, Lees KR, Medeghri Z, Machnig T, Schneider D, Kummer R von, Wahlgren N, Toni D (2008) **Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke.** *N Engl J Med* 359, 1317–1329.
- He K, Zhang X, Ren S, Sun J (2016) **Deep Residual Learning for Image Recognition.** In: IEEE (Hg.) *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* CVPR 2016, 770–778.

- Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH (2019) **Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke**. *Stroke* 50, 1263–1265.
- Higashida RT, Furlan AJ, Roberts H, Tomsick T, Connors B, Barr J, Dillon W, Warach S, Broderick J, Tilley B, Sacks D (2003) **Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke**. *Stroke* 34, 109-137.
- Hjort N, Christensen S, Sølling C, Ashkanian M, Wu O, Røhl L, Gyldensted C, Andersen G, Østergaard L (2005) **Ischemic injury detected by diffusion imaging 11 minutes after stroke**. *Ann Neurol* 58, 462–465.
- Ho KC, Scalzo F, Sarma KV, Speier W, El-Saden S, Arnold C (2019) **Predicting ischemic stroke tissue fate using a deep convolutional neural network on source magnetic resonance perfusion images**. *J Med Imaging* 6, 26001.
- Ho KC, Speier W, El-Saden S, Arnold CW (2018) **Classifying Acute Ischemic Stroke Onset Time using Deep Imaging Features**. *AMIA Annu Symp Proc* 2017, 892–901.
- Hokkinen L, Mäkelä T, Savolainen S, Kangasniemi M (2021a) **Computed tomography angiography-based deep learning method for treatment selection and infarct volume prediction in anterior cerebral circulation large vessel occlusion**. *Acta Radiol Open* 10, 1-10.
- Hokkinen L, Mäkelä T, Savolainen S, Kangasniemi M (2021b) **Evaluation of a CTA-based convolutional neural network for infarct volume prediction in anterior cerebral circulation ischaemic stroke**. *Eur Radiol Exp* 5, 25.
- Hornik K, Stinchcombe M, White H (1989) **Multilayer feedforward networks are universal approximators**. *Neural Networks* 2, 359–366.
- Hu X, Luo W, Hu J, Guo S, Huang W, Scott MR, Wiest R, Dahlweid M, Reyes M (2020) **Brain SegNet: 3D local refinement network for brain lesion segmentation**. *BMC Med Imaging* 20, 17.
- Huang S, Shen Q, Duong TQ (2010) **Artificial neural network prediction of ischemic tissue fate in acute stroke imaging**. *J Cereb Blood Flow Metab* 30, 1661–1670.
- Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) **nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation**. *Nat Methods* 18, 203–211.
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) **FSL**. *Neuroimage* 62, 782–790.
- Jovin TG, Chamorro A, Cobo E, Miquel MA de, Molina CA, Rovira A, San Román L, Serena J, Abilleira S, Ribó M, Millán M, Urra X, Cardona P, López-Cancio E, Tomasello A, Castaño C, Blasco J, Aja L, Dorado L, Quesada H, Rubiera M, Hernandez-Pérez M, Goyal M, Demchuk AM, Kummer R von, Gallofré M, Dávalos A (2015) **Thrombectomy within 8 hours after symptom onset in ischemic stroke**. *N Engl J Med* 372, 2296–2306.

- Kahle KT, Simard JM, Staley KJ, Nahed BV, Jones PS, Sun D (2009) **Molecular mechanisms of ischemic cerebral edema: role of electroneutral ion transport.** *Physiology (Bethesda)* 24, 257–265.
- Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) **Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.** *Med Image Anal* 36, 61–78.
- Karimi D, Salcudean SE (2020) **Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks.** *IEEE Transactions on Medical Imaging* 39, 499–513.
- Kasasbeh AS, Christensen S, Parsons MW, Campbell B, Albers GW, Lansberg MG (2019) **Artificial Neural Network Computer Tomography Perfusion Prediction of Ischemic Core.** *Stroke* 50, 1578–1581.
- Kate MP, Riaz P, Gioia L, Sivakumar L, Jeerakathil T, Buck B, Beaulieu C, Butcher K (2015) **Dynamic Evolution of Diffusion-Weighted Imaging Lesions in Patients With Minor Ischemic Stroke.** *Stroke* 46, 2318–2321.
- Khan M, Baird GL, Goddeau RP, Silver B, Henninger N (2017) **Alberta Stroke Program Early CT Score Infarct Location Predicts Outcome Following M2 Occlusion.** *Front Neurol* 8, 98.
- Khoury NN, Darsaut TE, Ghostine J, Deschaintre Y, Daneault N, Durocher A, Lanthier S, Poppe AY, Odier C, Lebrun L-H, Guilbert F, Gentic J-C, Batista A, Weill A, Roy D, Bracard S, Raymond J (2017) **Endovascular thrombectomy and medical therapy versus medical therapy alone in acute stroke: A randomized care trial.** *J Neuroradiol* 44, 198–202.
- Kidwell CS, Chalela JA, Saver JL, Starkman S, Hill MD, Demchuk AM, Butman JA, Patronas N, Alger JR, Latour LL, Luby ML, Baird AE, Leary MC, Tremwel M, Ovbiagele B, Fredieu A, Suzuki S, Villablanca JP, Davis S, Dunn B, Todd JW, Ezzeddine MA, Haymore J, Lynch JK, Davis L, Warach S (2004) **Comparison of MRI and CT for detection of acute intracerebral hemorrhage.** *JAMA* 292, 1823–1830.
- Kidwell CS, Saver JL, Mattiello J, Starkman S, Vinuela F, Duckwiler G, Gobin YP, Jahan R, Vespa P, Kalafut M, Alger JR (2000) **Thrombolytic reversal of acute human cerebral ischemic injury shown by diffusion/perfusion magnetic resonance imaging.** *Ann Neurol* 47, 462–469.
- Kimura K, Sakamoto Y, Iguchi Y, Shibasaki K (2011) **Serial changes in ischemic lesion volume and neurological recovery after t-PA therapy.** *J Neurol Sci* 304, 35–39.
- Kingma DP, Ba J (2015) **Adam: A Method for Stochastic Optimization.** In: Bengio Y, Lecun Y (Hg.) *Proceedings of the 3rd International Conference on Learning Representations. ICLR 2015.*

- Klötzsch C, Röther J (2020) **Zerebrale Ischämie: Pathophysiologie, Klassifikation, Epidemiologie und Risikofaktoren**. In: Berlit P (Hg.) *Klinische Neurologie*, 4. Auflage, Springer, Berlin, Heidelberg, 823–829.
- Kolominsky-Rabas PL, Sarti C, Heuschmann PU, Graf C, Siemonsen S, Neundoerfer B, Katalinic A, Lang E, Gassmann KG, Stockert TR von (1998) **A prospective community-based study of stroke in Germany--the Erlangen Stroke Project (ESPro): incidence and case fatality at 1, 3, and 12 months**. *Stroke* 29, 2501–2506.
- Krizhevsky A, Sutskever I, Hinton GE (2017) **ImageNet classification with deep convolutional neural networks**. *Commun ACM* 60, 84–90.
- Krongold M, Almekhlafi MA, Demchuk AM, Coutts SB, Frayne R, Eilaghi A (2015) **Final infarct volume estimation on 1-week follow-up MR imaging is feasible and is dependent on recanalization status**. *Neuroimage Clin* 7, 1–6.
- Kumar A, Upadhyay N, Ghosal P, Chowdhury T, Das D, Mukherjee A, Nandi D (2020) **CSNet: A new DeepNet framework for ischemic stroke lesion segmentation**. *Comput Methods Programs Biomed* 193, 105524.
- Lansberg MG, O'Brien MW, Tong DC, Moseley ME, Albers GW (2001a) **Evolution of cerebral infarct volume assessed by diffusion-weighted magnetic resonance imaging**. *Arch Neurol* 58, 613–617.
- Lansberg MG, Thijs VN, O'Brien MW, Ali JO, Crespigny AJ de, Tong DC, Moseley ME, Albers GW (2001b) **Evolution of apparent diffusion coefficient, diffusion-weighted, and T2-weighted signal intensity of acute stroke**. *AJNR Am J Neuroradiol* 22, 637–644.
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) **Gradient-based learning applied to document recognition**. *Proc IEEE* 86, 2278–2324.
- Lee C-Y, Xie S, Gallagher P, Zhang Z, Tu Z (2015) **Deeply-Supervised Nets**. In: Lebanon, G., Vishwanathan, S. V. N. (Hg.) *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*. AISTATS 2015, 562–570.
- Li X, Morgan PS, Ashburner J, Smith J, Rorden C (2016) **The first step for neuroimaging data analysis: DICOM to NIfTI conversion**. *J Neurosci Methods* 264, 47–56.
- Liebeskind DS, Bracard S, Guillemin F, Jahan R, Jovin TG, Majoie CB, Mitchell PJ, van der Lugt A, Menon BK, San Román L, Campbell BC, Muir KW, Hill MD, Dippel DW, Saver JL, Demchuk AM, Dávalos A, White P, Brown S, Goyal M (2019) **eTICI reperfusion: defining success in endovascular stroke therapy**. *J Neurointerv Surg* 11, 433–438.
- Liebeskind DS, Jahan R, Nogueira RG, Zaidat OO, Saver JL (2014) **Impact of collaterals on successful revascularization in Solitaire FR with the intention for thrombectomy**. *Stroke* 45, 2036–2040.

- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI (2017) **A survey on deep learning in medical image analysis.** *Med Image Anal* 42, 60–88.
- Lövblad KO, Baird AE, Schlaug G, Benfield A, Siewert B, Voetsch B, Connor A, Burzynski C, Edelman RR, Warach S (1997) **Ischemic lesion volumes in acute stroke by diffusion-weighted magnetic resonance imaging correlate with clinical outcome.** *Ann Neurol* 42, 164–170.
- Lucas C, Kemmling A, Bouteldja N, Aulmann LF, Madany Mamlouk A, Heinrich MP (2018a) **Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations.** *Front Neurol* 9, 989.
- Lucas C, Kemmling A, Mamlouk AM, Heinrich MP (2018b) **Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images.** In: IEEE (Hg.) *Proceedings of the IEEE 15th International Symposium on Biomedical Imaging. ISBI 2018*, 1118–1121.
- Luengo-Fernandez R, Paul NLM, Gray AM, Pendlebury ST, Bull LM, Welch SJV, Cuthbertson FC, Rothwell PM (2013) **Population-based study of disability and institutionalization after transient ischemic attack and stroke: 10-year results of the Oxford Vascular Study.** *Stroke* 44, 2854–2861.
- Maas AL, Hannun AY, Ng AY (2013) **Rectifier Nonlinearities Improve Neural Network Acoustic Models.** In: Dasgupta S, McAllester D (Hg.) *Proceedings of the 30th International Conference on Machine Learning. ICML 2013*.
- Mackey J, Kleindorfer D, Sucharew H, Moomaw CJ, Kissela BM, Alwell K, Flaherty ML, Woo D, Khatri P, Adeoye O, Ferioli S, Khoury JC, Hornung R, Broderick JP (2011) **Population-based study of wake-up strokes.** *Neurology* 76, 1662–1667.
- McLouth J, Elstrott S, Chaibi Y, Quenet S, Chang PD, Chow DS, Soun JE (2021) **Validation of a Deep Learning Tool in the Detection of Intracranial Hemorrhage and Large Vessel Occlusion.** *Front Neurol* 12, 656112.
- Meyer BC, Lyden PD (2009) **The modified National Institutes of Health Stroke Scale: its time has come.** *Int J Stroke* 4, 267–273.
- Milletari F, Navab N, Ahmadi S-A (2016) **V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation.** In: IEEE (Hg.) *Proceedings of the IEEE 4th International Conference on 3D Vision. 3DV 2016*, 565–571.
- Mocco J, Zaidat OO, Kummer R von, Yoo AJ, Gupta R, Lopes D, Frei D, Shownkeen H, Budzik R, Ajani ZA, Grossman A, Altschul D, McDougall C, Blake L, Fitzsimmons B-F, Yavagal D, Terry J, Farkas J, Lee SK, Baxter B, Wiesmann M, Knauth M, Heck D, Hussain S, Chiu D, Alexander MJ, Malisch T, Kirmani J, Miskolczi L, Khatri P (2016) **Aspiration Thrombectomy After Intravenous Alteplase Versus Intravenous Alteplase Alone.** *Stroke* 47, 2331–2338.

- Muir KW, Buchan A, Kummer R von, Rother J, Baron J-C (2006) **Imaging of acute stroke**. *Lancet Neurol* 5, 755–768.
- Muir KW, Ford GA, Messow C-M, Ford I, Murray A, Clifton A, Brown MM, Madigan J, Lenthall R, Robertson F, Dixit A, Cloud GC, Wardlaw J, Freeman J, White P (2017) **Endovascular therapy for acute ischaemic stroke: the Pragmatic Ischaemic Stroke Thrombectomy Evaluation (PISTE) randomised, controlled trial**. *J Neurol Neurosurg Psychiatry* 88, 38–44.
- Nazari-Farsani S, Yu Y, Duarte Armindo R, Lansberg M, Liebeskind DS, Albers G, Christensen S, Levin CS, Zaharchuk G (2022) **Predicting final ischemic stroke lesions from initial diffusion-weighted images using a deep neural network**. *Neuroimage Clin* 37, 103278.
- Nielsen A, Hansen MB, Tietze A, Mouridsen K (2018) **Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning**. *Stroke* 49, 1394–1401.
- NINDS-Study-Group (1995) **Tissue plasminogen activator for acute ischemic stroke**. *N Engl J Med* 333, 1581–1587.
- Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, Yavagal DR, Ribo M, Cognard C, Hanel RA, Sila CA, Hassan AE, Millan M, Levy EI, Mitchell P, Chen M, English JD, Shah QA, Silver FL, Pereira VM, Mehta BP, Baxter BW, Abraham MG, Cardona P, Veznedaroglu E, Hellinger FR, Feng L, Kirmani JF, Lopes DK, Jankowitz BT, Frankel MR, Costalat V, Vora NA, Yoo AJ, Malik AM, Furlan AJ, Rubiera M, Aghaebrahim A, Olivot J-M, Tekle WG, Shields R, Graves T, Lewis RJ, Smith WS, Liebeskind DS, Saver JL, Jovin TG (2018) **Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct**. *N Engl J Med* 378, 11–21.
- Nogueira RG, Kemmling A, Souza LM, Payabvash S, Hirsch JA, Yoo AJ, Lev MH (2017) **Clinical diffusion mismatch better discriminates infarct growth than mean transit time-diffusion weighted imaging mismatch in patients with middle cerebral artery-M1 occlusion and limited infarct core**. *J Neurointerv Surg* 9, 127–130.
- O'Brien P, Sellar RJ, Wardlaw JM (2004) **Fogging on T2-weighted MR after acute ischaemic stroke: how often might this occur and what are the implications?** *Neuroradiology* 46, 635–641.
- Olivot JM, Mlynash M, Inoue M, Marks MP, Wheeler HM, Kemp S, Straka M, Zaharchuk G, Bammer R, Lansberg MG, Albers GW (2014) **Hypoperfusion intensity ratio predicts infarct progression and functional outcome in the DEFUSE 2 Cohort**. *Stroke* 45, 1018–1023.
- Olivot J-M, Mlynash M, Thijs VN, Kemp S, Lansberg MG, Wechsler L, Bammer R, Marks MP, Albers GW (2009) **Optimal Tmax threshold for predicting penumbral tissue in acute stroke**. *Stroke* 40, 469–475.

- Olivot J-M, Mlynash M, Thijs VN, Kemp S, Lansberg MG, Wechsler L, Schlaug G, Bammer R, Marks MP, Albers GW (2008) **Relationships between infarct growth, clinical outcome, and early recanalization in diffusion and perfusion imaging for understanding stroke evolution (DEFUSE)**. *Stroke* 39, 2257–2263.
- Palm F, Urbanek C, Rose S, Buggle F, Bode B, Hennerici MG, Schmieder K, Inselmann G, Reiter R, Fleischer R, Piplack K-O, Safer A, Becher H, Grau AJ (2010) **Stroke Incidence and Survival in Ludwigshafen am Rhein, Germany: the Ludwigshafen Stroke Study (LuSSt)**. *Stroke* 41, 1865–1870.
- Parsons MW, Barber PA, Chalk J, Darby DG, Rose S, Desmond PM, Gerraty RP, Tress BM, Wright PM, Donnan GA, Davis SM (2002) **Diffusion- and perfusion-weighted MRI response to thrombolysis in stroke**. *Ann Neurol* 51, 28–37.
- Pialat J-B, Wiart M, Nighoghossian N, Adeleine P, Derex L, Hermier M, Froment J-C, Berthezene Y (2005) **Evolution of lesion volume in acute stroke treated by intravenous t-PA**. *J Magn Reson Imaging* 22, 23–28.
- Pinto A, Amorim J, Hakim A, Alves V, Reyes M, Silva CA (2021a) **Prediction of Stroke Lesion at 90-Day Follow-Up by Fusing Raw DSC-MRI With Parametric Maps Using Deep Learning**. *IEEE Access* 9, 26260–26270.
- Pinto A, Mckinley R, Alves V, Wiest R, Silva CA, Reyes M (2018) **Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information**. *Front Neurol* 9, 1060.
- Pinto A, Pereira S, Meier R, Wiest R, Alves V, Reyes M, Silva CA (2021b) **Combining unsupervised and supervised learning for predicting the final stroke lesion**. *Med Image Anal* 69, 101888.
- Puetz V, Khomenko A, Hill MD, Dzialowski I, Michel P, Weimar C, Wijman CAC, Mattle HP, Engelter ST, Muir KW, Pfefferkorn T, Tanne D, Szabo K, Kappelle LJ, Algra A, Kummer R von, Demchuk AM, Schonewille WJ (2011) **Extent of hypoattenuation on CT angiography source images in basilar artery occlusion: prognostic value in the Basilar Artery International Cooperation Study**. *Stroke* 42, 3454–3459.
- Purushotham A, Campbell BCV, Straka M, Mlynash M, Olivot J-M, Bammer R, Kemp SM, Albers GW, Lansberg MG (2015) **Apparent diffusion coefficient threshold for delineation of ischemic core**. *Int J Stroke* 10, 348–353.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP, Ng AY (2017) **CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning**, arXiv preprint, 1711.05225.
- Raman TV (1998) **Audio system for technical readings**. 1. Auflage, Springer, Berlin.

- Rangaraju S, Streib C, Aghaebrahim A, Jadhav A, Frankel M, Jovin TG (2015) **Relationship Between Lesion Topology and Clinical Outcome in Anterior Circulation Large Vessel Occlusions.** *Stroke* 46, 1787–1792.
- Rankin J (1957) **Cerebral vascular accidents in patients over the age of 60. II. Prognosis.** *Scott Med J* 2, 200–215.
- Ribo M, Molina CA, Delgado P, Rubiera M, Delgado-Mederos R, Rovira A, Munuera J, Alvarez-Sabin J (2007) **Hyperglycemia during ischemia rapidly accelerates brain damage in stroke patients treated with tPA.** *J Cereb Blood Flow Metab* 27, 1616–1622.
- Rimmele DL, Thomalla G (2014) **Wake-up stroke: clinical characteristics, imaging findings, and treatment option - an update.** *Front Neurol* 5, 35.
- Ringleb P, Köhrmann M, Jansen O (2022) **Akuttherapie des ischämischen Schlaganfalls, S2e-Leitlinie, Version 1.1.** In: Deutsche Gesellschaft für Neurologie (Hg.) Leitlinien für Diagnostik und Therapie in der Neurologie. Online: www.dgn.org/leitlinien [abgerufen am 28.12.2022].
- Ritzl A, Meisel S, Wittsack H-J, Fink GR, Siebler M, Mödder U, Seitz RJ (2004) **Development of brain infarct volume as assessed by magnetic resonance imaging (MRI): follow-up of diffusion-weighted MRI lesions.** *J Magn Reson Imaging* 20, 201–207.
- Robben D, Boers AMM, Marquering HA, Langezaal, Lucianne L. C. M., Roos, Yvo B. W. E. M., van Oostenbrugge RJ, van Zwam WH, Dippel DWJ, Majoie, Charles B. L. M., van der Lugt A, Lemmens R, Suetens P (2020) **Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning.** *Med Image Anal* 59, 101589.
- Robert Koch-Institut (2015) **Gesundheit in Deutschland. Gesundheitsberichterstattung des Bundes.** RKI, Berlin.
- Ronneberger O, Fischer P, Brox T (2015) **U-Net: Convolutional Networks for Biomedical Image Segmentation.** In: Navab N, Hornegger J, Wells W, Frangi A (Hg.) Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI 2015, 234–241.
- Saver JL (2006) **Time is brain--quantified.** *Stroke* 37, 263–266.
- Saver JL, Goyal M, Bonafe A, Diener H-C, Levy EI, Pereira VM, Albers GW, Cognard C, Cohen DJ, Hacke W, Jansen O, Jovin TG, Mattle HP, Nogueira RG, Siddiqui AH, Yavagal DR, Baxter BW, Devlin TG, Lopes DK, Reddy VK, Du Mesnil de Rochemont R, Singer OC, Jahan R (2015) **Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke.** *N Engl J Med* 372, 2285–2295.

- Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie, Charles B. L. M., Dippel DW, Campbell BC, Nogueira RG, Demchuk AM, Tomasello A, Cardona P, Devlin TG, Frei DF, Du Mesnil de Rochemont R, Berkhemer OA, Jovin TG, Siddiqui AH, van Zwam WH, Davis SM, Castaño C, Sapkota BL, Fransen PS, Molina C, van Oostenbrugge RJ, Chamorro Á, Lingsma H, Silver FL, Donnan GA, Shuaib A, Brown S, Stouch B, Mitchell PJ, Davalos A, Roos, Yvo B. W. E. M., Hill MD (2016) **Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis**. JAMA 316, 1279–1288.
- Saver JL, Johnston KC, Homer D, Wityk R, Koroshetz W, Truskowski LL, Haley EC (1999) **Infarct volume as a surrogate or auxiliary outcome measure in ischemic stroke clinical trials. The RANTTAS Investigators**. Stroke 30, 293–298.
- Schaefer PW, Buonanno FS, Gonzalez RG, Schwamm LH (1997) **Diffusion-weighted imaging discriminates between cytotoxic and vasogenic edema in a patient with eclampsia**. Stroke 28, 1082–1085.
- Schwamm LH, Koroshetz WJ, Sorensen AG, Wang B, Copen WA, Budzik R, Rordorf G, Buonanno FS, Schaefer PW, Gonzalez RG (1998) **Time course of lesion development in patients with acute stroke: serial diffusion- and hemodynamic-weighted magnetic resonance imaging**. Stroke 29, 2268–2276.
- Seabold S, Perktold J (2010) **Statsmodels: Econometric and Statistical Modeling with Python**. In: van der Walt SJ, Millman J (Hg.) Proceedings of the 9th Python in Science Conference. SciPy 2010, 92–96.
- Shen D, Wu G, Suk H-I (2017) **Deep Learning in Medical Image Analysis**. Annu Rev Biomed Eng 19, 221–248.
- Simard JM, Kent TA, Chen M, Tarasov KV, Gerzanich V (2007) **Brain oedema in focal ischaemia: molecular pathophysiology and theoretical implications**. Lancet Neurol 6, 258–268.
- Simonyan K, Zisserman A (2015) **Very Deep Convolutional Networks for Large-Scale Image Recognition**. In: Bengio Y, Lecun Y (Hg.) Proceedings of the 3rd International Conference on Learning Representations. ICLR 2015.
- Simpkins AN, Dias C, Norato G, Kim E, Leigh R (2017) **Early Change in Stroke Size Performs Best in Predicting Response to Therapy**. Cerebrovasc Dis 44, 141–149.
- Skriver EB, Olsen TS (1981) **Transient disappearance of cerebral infarcts on CT scan, the so-called fogging effect**. Neuroradiology 22, 61–65.
- Smith SM (2002) **Fast robust automated brain extraction**. Hum Brain Mapp 17, 143–155.
- Stier N, Vincent N, Liebeskind D, Scalzo F (2015) **Deep learning of tissue fate features in acute ischemic stroke**. In: IEEE (Hg.) Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. BIBM 2015, 1316–1321.

- Stokum JA, Gerzanich V, Simard JM (2016) **Molecular pathophysiology of cerebral edema.** *J Cereb Blood Flow Metab* 36, 513–538.
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) **Going Deeper with Convolutions.** In: IEEE (Hg.) Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2015, 1–9.
- Tate WJ, Polding LC, Christensen S, Mlynash M, Kemp S, Heit JJ, Marks MP, Albers GW, Lansberg MG (2021) **Predictors of Early and Late Infarct Growth in DEFUSE 3.** *Front Neurol* 12, 699153.
- 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 D-W, Dousset V, Sorensen AG, Liebeskind DS, Fiebach JB, Fiehler J, Gerloff C (2011) **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* 10, 978–986.
- Thomalla G, Rossbach P, Rosenkranz M, Siemonsen S, Krüzelmann A, Fiehler J, Gerloff C (2009) **Negative fluid-attenuated inversion recovery imaging identifies acute ischemic stroke at 3 hours or less.** *Ann Neurol* 65, 724–732.
- Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, Cheripelli B, Cho T-H, Fazekas F, Fiehler J, Ford I, Galinovic I, Gellissen S, Golsari A, Gregori J, Günther M, Guibernau J, Häusler KG, Hennerici M, Kemmling A, Marstrand J, Modrau B, Neeb L, La Perez de Ossa N, Puig J, Ringleb P, Roy P, Scheel E, Schonewille W, Serena J, Sunaert S, Villringer K, Wouters A, Thijs V, Ebinger M, Endres M, Fiebach JB, Lemmens R, Muir KW, Nighoghossian N, Pedraza S, Gerloff C (2018) **MRI-Guided Thrombolysis for Stroke with Unknown Time of Onset.** *N Engl J Med* 379, 611–622.
- Thrane AS, Rangroo Thrane V, Nedergaard M (2014) **Drowning stars: reassessing the role of astrocytes in brain edema.** *Trends Neurosci* 37, 620–628.
- Ulyanov D, Vedaldi A, Lempitsky V (2016) **Instance Normalization: The Missing Ingredient for Fast Stylization,** arXiv preprint, 1607.08022.
- van den Bussche H, Berger K, Kemper C, Barzel A, Glaeske G, Koller D (2010) **Inzidenz, Rezidiv, Pflegebedürftigkeit und Mortalität von Schlaganfall.** *Akt Neurol* 37, 131–135.
- van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J (1988) **Interobserver agreement for the assessment of handicap in stroke patients.** *Stroke* 19, 604–607.
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P (2020) **SciPy 1.0: fundamental algorithms for scientific computing in Python.** *Nat Methods* 17, 261–272.

- Vupputuri A, Gupta A, Ghosh N (2021) **MCA-DN: Multi-path convolution leveraged attention deep network for salvageable tissue detection in ischemic stroke from multi-parametric MRI.** *Comput Biol Med* 136, 104724.
- Vymazal J, Rulseh AM, Keller J, Janouskova L (2012) **Comparison of CT and MR imaging in ischemic stroke.** *Insights Imaging* 3, 619–627.
- Wahlgren N, Ahmed N, Dávalos A, Hacke W, Millán M, Muir K, Roine RO, Toni D, Lees KR (2008) **Thrombolysis with alteplase 3-4.5 h after acute ischaemic stroke (SITS-ISTR): an observational study.** *Lancet* 372, 1303–1309.
- Warach SJ, Dula AN, Milling TJ (2020) **Tenecteplase Thrombolysis for Acute Ischemic Stroke.** *Stroke* 51, 3440–3451.
- Winder AJ, Wilms M, Amador K, Flottmann F, Fiehler J, Forkert ND (2022) **Predicting the tissue outcome of acute ischemic stroke from acute 4D computed tomography perfusion imaging using temporal features and deep learning.** *Front Neurosci* 16, 1009654.
- Winzeck S, Hakim A, Mckinley R, Pinto, José A. A. D. S. R., Alves V, Silva C, Pisov M, Krivov E, Belyaev M, Monteiro M, Oliveira A, Choi Y, Paik MC, Kwon Y, Lee H, Kim BJ, Won J-H, Islam M, Ren H, Robben D, Suetens P, Gong E, Niu Y, Xu J, Pauly JM, Lucas C, Heinrich MP, Rivera LC, Castillo LS, Daza LA, Beers AL, Arbelaezs P, Maier O, Chang K, Brown JM, Kalpathy-Cramer J, Zaharchuk G, Wiest R, Reyes M (2018) **ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI.** *Front Neurol* 9, 679.
- Wouters A, Robben D, Christensen S, Marquering HA, Roos, Yvo B. W. E. M., van Oostenbrugge RJ, van Zwam WH, Dippel DWJ, Majoie, Charles B. L. M., Schonewille WJ, van der Lugt A, Lansberg M, Albers GW, Suetens P, Lemmens R (2022) **Prediction of Stroke Infarct Growth Rates by Baseline Perfusion Imaging.** *Stroke* 53, 569–577.
- Yoo AJ, Chaudhry ZA, Nogueira RG, Lev MH, Schaefer PW, Schwamm LH, Hirsch JA, González RG (2012) **Infarct volume is a pivotal biomarker after intra-arterial stroke therapy.** *Stroke* 43, 1323–1330.
- Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Christensen S, Marks MP, Lansberg MG, Albers GW, Zaharchuk G (2021) **Tissue at Risk and Ischemic Core Estimation Using Deep Learning in Acute Stroke.** *AJNR Am J Neuroradiol* 42, 1030–1037.
- Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, Christensen S, Marks MP, Lansberg MG, Albers GW, Zaharchuk G (2020) **Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging.** *JAMA Netw Open* 3, e200772.

Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, Kummer R von, Saver JL, Marks MP, Prabhakaran S, Kallmes DF, Fitzsimmons B-FM, Mocco J, Wardlaw JM, Barnwell SL, Jovin TG, Linfante I, Siddiqui AH, Alexander MJ, Hirsch JA, Wintermark M, Albers G, Woo HH, Heck DV, Lev M, Aviv R, Hacke W, Warach S, Broderick J, Derdeyn CP, Furlan A, Nogueira RG, Yavagal DR, Goyal M, Demchuk AM, Bendszus M, Liebeskind DS (2013) **Recommendations on angiographic revascularization grading standards for acute ischemic stroke: a consensus statement.** Stroke 44, 2650–2663.

Zaidi SF, Aghaebrahim A, Urra X, Jumaa MA, Jankowitz B, Hammer M, Nogueira R, Horowitz M, Reddy V, Jovin TG (2012) **Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy.** Stroke 43, 3238–3244.

Zeiler MD, Ranzato M, Monga R, Mao M, Yang K, Le QV, Nguyen P, Senior A, Vanhoucke V, Dean J, Hinton GE (2013) **On rectified linear units for speech processing.** In: IEEE (Hg.) Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2013, 3517–3521.

9 Abkürzungsverzeichnis

ADC	Apparent-Diffusion-Coefficient
AIF	Arterial-Input-Function
ANTs	Advanced-Normalization-Tools
ASPECTS	Alberta-Stroke-Program-Early-Computertomographie-Score
ATP	Adenosintriphosphat
BCE	Binary-Cross-Entropy
BET	Brain-Extraction-Tool
CAE	Convolutional-Auto-Encoder
CBF	Cerebral-Blood-Flow
CBV	Cerebral-Blood-Volume
CDM	Clinical-Diffusion-Mismatch
CE-MRA	Contrast-Enhanced-Magnetresonanz-Angiographie
CMRO2	Cerebral-Metabolic-Rate-Of-Oxygen
CNN	Convolutional-Neural-Network
CT	Computertomographie
CTA	Computertomographische Angiographie
CTC	Concentration-Time-Curve
CTP	Computertomographische Perfusionsbildgebung
DSA	Digitale Subtraktionsangiographie
DWI	Diffusion-Weighted-Imaging
DWIR	Diffusion-Weighted-Imaging-Reversibility
ELU	Exponential-Linear-Unit
EVT	endovaskuläre Therapieverfahren
FCN	Fully-Convolutional-Network
FLAIR	Fluid-Attenuated-Inversion-Recovery
GRU	Gated-Recurrent-Unit
HIR	Hypoperfusion-Intensity-Ratio
IV	Infarkt volumen
IVT	intravenöse Thrombolyse
MAE	Mean-Absolut-Error, Mittlerer absoluter Fehler
MLP	Multilayer-Perzeptron
MNI	Montreal-Neurological-Institute
MRA	Magnetresonanzangiographie
MRT	Magnetresonanztomographie
mRS	modifizierte Rankin-Skala
MSE	Mean-Squared-Error, Mittlere quadratische Abweichung
MT	mechanische Thrombektomie
mTICI	Modified-Treatment-In-Cerebral-Infarction
MTT	Mean-Transit-Time
nCT	Native Computertomographie
NIHSS	National-Institutes-of-Health-Stroke-Skala
PPP	Per-Protocol-Population
PWI	Perfusion-Weighted-Imaging
RC	Residual-Curve
rCBV	relative-Cerebral-Blood-Volume
ReLU	Rectified-Linear-Unit
rTTH	relative-Transit-Time-Heterogeneity
rtPA	Recombinant-Tissue-Type-Plasminogen-Activator
std	Standardabweichung
TICI	Treatment-In-Cerebral-Infarction
Tmax	Time-To-Maximum
TOF	Time-Of-Flight
TSS	Time-Since-Stroke
TTD	Time-To-Drain
TTP	Time-To-Peak
TTT	Time-To-Treatment

10 Abbildungsverzeichnis

Abbildung 1: Infarkte unterschiedlicher Pathogenese (Klötzsch & Röther 2020)	5
Abbildung 2: Bereiche der Infarktläsion	8
Abbildung 3: Zeitliche Einteilung des Schlaganfalls (Bernhardt et al. 2017)	9
Abbildung 4: Darstellung des Infarktkerns im zeitlichen Verlauf (Lansberg et al. 2001b).....	11
Abbildung 5: Diagramm der Perfusionsparameter (Carrera & Wintermark 2017)	13
Abbildung 6: PWI-DWI-Mismatch (Rimmele & Thomalla 2014).....	14
Abbildung 7: DWI-FLAIR-Mismatch (Rimmele & Thomalla 2014)	18
Abbildung 8: Darstellung eines Infarktes im zeitlichen Verlauf (Gaudinski et al. 2008).....	23
Abbildung 9: Bereiche der Künstlichen Intelligenz	27
Abbildung 10: Aufbau eines Perzeptrons.....	28
Abbildung 11: Architektur eines Multilayer-Perzeptrons	30
Abbildung 12: Funktion eines Convolutional- und eines Pooling-Layers (Shen et al. 2017)	32
Abbildung 13: VGGNet (Ferguson et al. 2017)	33
Abbildung 14: U-Net (Ronneberger et al. 2015).....	34
Abbildung 15: Fully-Convolutional-Network (Choi et al. 2016)	46
Abbildung 16: UNet mit Integration des TICI-Scores (Pinto et al. 2018).....	47
Abbildung 17: Cascade-Network (Lucas et al. 2018a).....	49
Abbildung 18: Fully-Convolutional-Network (Robben et al. 2020)	50
Abbildung 19: Registrierungsschritte	54
Abbildung 20: Preprocessing der Bilddaten	55
Abbildung 21: Infarkt-Segmentierungen von zwei exemplarischen Fällen	58
Abbildung 22: Bilddaten eines Falles	59
Abbildung 23: Beispielhafte Prädiktion eines exemplarischen Falles	60
Abbildung 24: Verschachtelte Kreuzvalidierung.....	61
Abbildung 25: Baseline-UNet	63
Abbildung 26: Integration klinischer Informationen anhand von separaten Feature-Maps	65
Abbildung 27: Geigenplot verschiedener Prädiktions-Modelle	72
Abbildung 28: Boxplot-Ergebnisse des Baseline-UNets und des Multitask-UNets (1d)	73
Abbildung 29: Boxplot-Ergebnisse des Baseline-UNets und des Multitask-UNets (1d)	73
Abbildung 30: Dice-Score-Veränderungen zwischen dem Baseline- und dem Multitask-UNet	74
Abbildung 31: Prädizierte und tatsächliche Infarktvolumenänderungen	76
Abbildung 32: Integration klinischer Informationen vor dem Encoder.....	86
Abbildung 33: Integration klinischer Informationen auf jeder Ebene des Encoders	86
Abbildung 34: Integration klinischer Informationen am Ende des Bottlenecks	87
Abbildung 35: Integration klinischer Informationen am Ende des Decoders	87

11 Tabellenverzeichnis

Tabelle 1: Darstellung der Infarktläsion anhand der verschiedenen bildgebenden Verfahren	15
Tabelle 2: Modifizierte Rankin-Skala.....	19
Tabelle 3: National-Institutes-of-Health-Stroke-Skala	19
Tabelle 4: Modifizierte Treatment-In-Cerebral-Ischemia-Skala	20
Tabelle 5: Studien zur endovaskulären Therapie mit läsionsbezogenem Endpunkt	21
Tabelle 6: Übersicht über bisherige Studien	40
Tabelle 7: Segmentierungs-UNet-Modelle: Hyperparameteroptimierung der Epochenanzahl sowie der initialen Lernrate.....	56
Tabelle 8: Dice-Scores der Segmentierungs-UNet-Modelle	57
Tabelle 9: Zusammenfassung der klinischen Informationen.....	59
Tabelle 10: Korrelation der klinischen Informationen mit der Volumenänderung des Infarktes	66
Tabelle 11: nnUNet-Modelle: Hyperparameteroptimierung der Epochenanzahl	67
Tabelle 12: Baseline-UNet-Modelle: Hyperparameteroptimierung der Epochenanzahl	67
Tabelle 13: Multitask-UNet-Modelle: Hyperparameteroptimierung des Integrationsortes der klinischen Informationen	68
Tabelle 14: Multitask-UNet-Modelle: Optimierter Integrationsort.....	69
Tabelle 15: Dice-Scores der Multitask-UNet-Modelle	69
Tabelle 16: Multitask-UNet-Modell: Hyperparameteroptimierung des Integrationsortes der tatsächlichen Volumenänderung	69
Tabelle 17: Dice-Scores der verschiedenen Prädiktions-Modelle	70
Tabelle 18: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem 3D-nnUNet.....	71
Tabelle 19: Dice-Scores der verschiedenen Prädiktions-Modelle	71
Tabelle 20: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem Baseline-UNet	71
Tabelle 21: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem 3D-nnUnet	74
Tabelle 22: Mittlerer absoluter Fehler der verschiedenen Prädiktions-Modelle	75
Tabelle 23: Einseitige Wilcoxon-Vorzeichen-Rang-Tests gegenüber dem Multitask-UNet (90d) ...	75

12 Erklärung des Eigenanteils

Die vorliegende Arbeit wurde in der Abteilung „Image Processing and Medical Imaging“ des Instituts für Computational Neuroscience des Universitätsklinikums Hamburg-Eppendorf unter Betreuung von Herrn Prof. Dr. rer. nat. René Werner durchgeführt.

Die Konzeption dieser Arbeit erfolgte in Zusammenarbeit mit Herrn Prof. Dr. rer. nat. René Werner.

Die ursprünglichen DICOM-Bilddaten sowie die dazugehörigen klinischen Informationen entstammen der Wake-Up-Studie und wurden durch Herrn Prof. Dr. med. Götz Thomalla sowie Herrn PD Dr. med. Bastian Cheng zur Verfügung gestellt (Thomalla et al. 2018).

Die Vorverarbeitung der Bilddaten erfolgte in Zusammenarbeit mit meinem Mitdoktoranden Herrn Böhme und beinhaltete die Nifti-Konvertierung, die MNI-Space-Registrierung sowie die Nachsegmentierung der Infarktareale.

Sämtliche experimentelle Versuche wurden von mir eigenständig durchgeführt. Dies umfasst die Generierung, das Training und die Hyperparameter-Optimierung der selbsterstellten UNet-Modelle sowie die Prädiktion anhand dieser Modelle. Die selbsterstellten Modelle wurden mit UNet-Modellen, die anhand der nnUNet-Methode erstellt wurden, verglichen (Isensee et al. 2021).

Die statistische Auswertung der Prädiktionsmodelle sowie der klinischen Informationen erfolgte eigenständig in Beratung durch Herrn Prof. Dr. rer. nat. René Werner.

Datum

Unterschrift

13 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.

Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet. Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum

Unterschrift

14 Danksagung

Ich danke meinem Doktorvater und wissenschaftlichem Betreuer Herrn Prof. Dr. rer. nat. René Werner für seine kompetente fachliche Unterstützung während der experimentellen Durchführung dieser Arbeit sowie für sein kritisches Feedback im Rahmen der wissenschaftlichen Auswertung und Interpretation der experimentellen Ergebnisse.

Für die vielseitige Unterstützung und die angenehme Arbeitsatmosphäre danke ich der Abteilung „Image Processing and Medical Imaging“. Insbesondere danke ich hierbei meinem Mitdoktoranden Herrn Böhme für die gemeinsame Vorverarbeitung der Bilddaten.

Ich danke Herrn Prof. Dr. med. Götz Thomalla sowie Herrn PD Dr. med. Bastian Cheng für die Bereitstellung der verwendeten Bilddaten.

Zu guter Letzt gebührt mein Dank meiner Familie und meinen Freunden für die moralische Unterstützung bei kleineren und größeren Rückschlägen während der experimentellen Durchführung und im Rahmen des Schreibprozesses dieser Arbeit. Ein besonderer Dank gilt dabei meiner Partnerin Sarah sowie meinen Eltern Verena und Daniel.