UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Department of Diagnostic and Interventional Neuroradiology

Prof. Dr. med. Jens Fiehler

Radiomics Based Machine Learning Outcome Prediction of Acute Intracranial Hemorrhage on Computed Tomography

Dissertation

zur Erlangung des Grades eines Doktors der Zahnmedizin
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Constanze Veronika Friedrich
aus Tegernsee

Hamburg 2022

Table of Contents

List of Abbreviations

WHO	–	World Health Organisation
ICH	–	Intracerebral hemorrhage
SAH	–	Subarachnoid hemorrhage
NECT	–	Non-enhanced computed tomography
CAA	–	Cerebral amyloid angiopathy
CBF	–	Cerebral blood flow
PHE	–	Perihematomal edema
mRS	–	Modified Rankin Scale
IVH	–	Intraventricular hemorrhage
EVD	–	External ventricular drain
CSF	–	Cerebrospinal fluid
CT	–	Computed tomography
MRI	–	Magnetic resonance imaging
CTA	–	Computed tomographic angiography
HU	–	Hounsfield Unit
GCS	–	Glasgow Coma Scale
HE	–	Hematoma expansion
ML	–	Machine learning
RIS	–	Radiology Information System
DICOM	–	Digital Imaging and Communications in Medicine
ROI	–	Region of Interest
AI	–	Artificial intelligence
ROC	–	Receiver operating characteristics
AUC	–	Area under the curve
CI	–	Confidence interval
MCC	–	Matthews correlation coefficient
TP	–	True positive
TN	–	True negative
FP	–	False positive
FN	–	False negative
CSF	–	Cerebrospinal fluid
BHS	–	Black hole sign
IS	–	Island sign

1 Introduction

1.1 Definition

The World Health Organisation (WHO) defines stroke as "rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin" (Truelsen et al., 2006).

Strokes are subdivided according to their pathological distinction and the affected vascular distribution. Subdivision is important for fast choice of therapeutic procedure. The next steps depend strongly on the type of stroke. Although the course of action differs for the different types of a stroke it is always important to make a diagnosis as soon as possible to initiate the right therapy and to plan the next steps (Yew and Cheng, 2015, Knöß et al., 2012). The stroke subtypes are further distinguished between hemorrhagic and ischemic disturbance of blood flow in the brain (Truelsen et al., 2006).

The WHO has a clear definition of stroke. Symptoms caused by subdural hematomas, tumors, poisoning or trauma do not fall into the category of stroke, nor do transient ischemic attacks, because by definition they last less than 24 hours.(Truelsen et al., 2006).

Most acute strokes fall into the ischemic category. These account for 87% of all stroke types and are often the result of large-scale arteriosclerosis, cardio embolism, small vascular occlusion or indeterminate cause (Yew and Cheng, 2015). A circulatory disorder of the brain occurs, resulting in an acute focal neurological deficit. The interruption of the blood and thus oxygen supply leads to a loss of function and finally to the death of brain tissue (Diener and Weimar, 2012).

Of the 15 million strokes worldwide annually (Grunwald et al., 2017), hemorrhagic strokes make up the smaller but more fatal group of strokes (Yew and Cheng, 2015). Hemorrhagic strokes describe intracerebral and subarachnoid hemorrhages (Yew and Cheng, 2015).

Spontaneous intracerebral hemorrhages (ICH) of the hemorrhagic group of strokes are often due to arteriolar hypertension, vascular malformations in the brain or nutrition, such as alcohol consumption or low cholesterol levels in the blood, and less frequently due to clotting disorders.

Cerebral amyloid angiopathy (CAA) is the result of high blood pressure. It occurs primarily in older people and is an increasingly frequent cause of ICH among our aging society in Germany (Truelsen et al., 2006).

The other subgroup of hemorrhagic strokes consists of the subarachnoid hemorrhages (SAH). They are often caused by the rupture of aneurysms at the branching of large arteries at the lower brain surface. Although some of the affected patients develop symptoms that correspond to the definition of a stroke, they are excluded from many stroke studies, since they often do not cause direct damage to the brain (Truelsen et al., 2006).
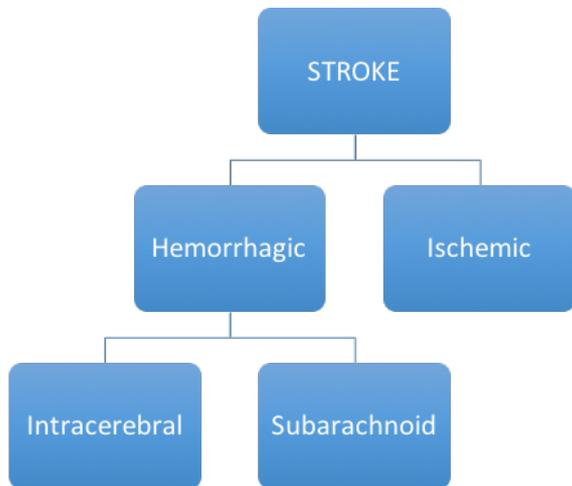

Fig. 1:
Overview of the division of subcategories under the stroke definition

1.2 Incidence

ICH is the second most common subtype of strokes, making up 10-20% of all strokes (An et al., 2017). In countries such as the UK, Australia or the US, they are responsible for 8-15%, in Europe 15% (Weber and Nordmeyer, 2015) and in Japan and Korea for 18-24% of all strokes (An et al., 2017).

The one-month fatality rate is 40% and the one-year fatality rate 54% (An et al., 2017). Only 12-39% of the affected patients reach a long-term functional independent outcome. The occurrence in low- and middle-income countries is almost twice as high as in industrial nations (22 vs. 10 per 100,000 person/years). It is more common in Asians than in Blacks, Whites or Hispanics, increases with age and occurs more frequently in men than in women (An et al., 2017).

1.3 Objectives

Acute primary ICH is particularly time-sensitive and associated with high morbidity and mortality rates. This high mortality rate is highly dependent on the extent of the ICH volume and prognostication in this context is of great importance, to guide goals of care discussions, clinical decision-making, and risk stratification. Neuroimaging is a central criterion for the assessment of prognostication tools as exemplarily shown in the clinically most established

tool, the ICH Score (incorporating different clinical and imaging variables). Failures of recent ICH trials (Sprigg N, 2018, Mayer et al., 2008, Hanley et al., 2019) have given the opportunity to rethink the paradigms of prognostication tools for ICH patients.

In addition to clinical factors and clinically established imaging markers, more recently novel and promising imaging markers have been described (Morotti et al.). There is growing evidence that these novel imaging markers in non-enhanced computed tomography (NECT) have great value for outcome prediction of patients with ICH (Sporns et al., 2018a), yet lack prospective validation and therefore generalizability. One possible novel technology that has been in discussion allows the extraction of high throughput imaging features (radiomics) which in combination with Machine Learning (ML) algorithms have the potential to enhance the accuracy in the risk stratification and clinical decision making.

There are already algorithms that are applied in neuroradiology. However, these mainly deal with other neurological conditions, such as ischemic stroke (Heo et al., 2019). Wang et al were the first to develop an ML approach for outcome prediction of ICHs (Wang et al., 2019). Limitations of this study resulted from the fact that many patients with large hematomas or very critical conditions were excluded from the study after consultation with their relatives. In addition, this was a relatively small sample of patients. Information on early hematoma growth and extent of edema were not included in the model. These reasons and the circumstance that no external validation was performed may limit the generalizability of the results and minimize the predictive power of the model (Wang et al). Integrating high-quality features into such a model could contribute an improved predictive power of such models.

Other authors have also investigated radiomics and ML models for predicting outcomes in ICH. But some focused on hypertensive ICHs, for example, whereas others focused on the predictive power of hematoma expansion after ICH by radiomics. In these studies it was noted that the power of the results might be limited by very different treatment procedures that were performed on the patients as well as by the studies´ retrospectivity. Furthermore, there was no standardized image processing procedure and different parameters were applied for CT image acquisition and reconstruction. A larger patient collective would be useful for better generalization and robustness of the conclusions (Chen, 2021, Xu, 2021).

Based on this understanding the objective of this study is to investigate how well new quantitative imaging markers (radiomics) allow the prediction of clinical outcome in order to improve early triage of high-risk patients for poor outcome using a ML approach.

2 State of Science and Research

2.1 Subdivision

ICH can be further divided into lobar and deep bleeding. This distinction is important for the underlying cause of the bleeding and its prognosis. Deep bleeding is usually due to age-related arteriosclerosis, hypertension or other conventional vascular risk factors (Smith and Romero, 2018).

Lobar bleeding is localised in the cortex or white matter below. It is more likely associated with CAA that results from vascular amyloid deposition (Smith and Romero, 2018). While deep bleeding with a closely adjusted and controlled blood pressure has only relatively low recurrence rates (1-2% per year), lobar bleeding is associated with a significantly higher risk of recurrence.

There are also patients who show a mixed pattern of bleeding. Of the two possible scenarios, it is either a mixed occurrence of CAA and non-CAA related reasons or it is only non-CAA related arteriosclerosis. It seems unlikely, however, that a strongly developed CAA is responsible for deep bleeding since the basal ganglia and the brainstem are relatively well protected from vascular amyloid accumulations (Smith and Romero, 2018).

For this reason and because the pattern of risk factors indicates it, patients with mixed bleeding are more likely to be associated with hypertensive-related ICH and less with a combination of hypertension and CAA (Smith and Romero, 2018).

The classification of the ICH according to the bleeding site can be further categorised into typical and atypical ICHs. The "loco typico" ICH is limited to the region of the stem ganglia, sometimes with involvement of the thalamus, and is caused by hypertension. The atypical ICH includes the lobar hemorrhages and is characterised by the way it affects cortical and subcortical regions. It is often associated with CAA and often shows microbleedings and later recurring macrobleedings (Kuramatsu et al., 2017).

2.2 Risk Factors

There are factors that increase the likelihood of a stroke. Some of them are congenital and cannot be controlled by humans. These include increasing age, the male sex, CAA, the Asian ethnic group, sleep apnea, chronic kidney disease, diabetes mellitus or family history (Yew and Cheng, 2015, An et al., 2017).

However, there are also modifiable risk factors. According to the Interstroke Study, 88.1% of the risk attributable to the population is attributable to those modifiable factors (Zeng et

al., 2017). They include hypertension, smoking, excessive alcohol consumption, decreased low-density lipoprotein cholesterol, low triglycerides, drugs and medications such as anticoagulants, antithrombotic agents or sympathomimetics (An et al., 2017).

Nutrition and physical activity also affect the risk of a bleeding event. People who exercise infrequently, regularly drink sweetened drinks and generally eat little fish, fruit and vegetables have an increased risk. In women oral contraceptives, the immediate postpartum period or preeclampsia can have a negative effect on risk behaviour (Yew and Cheng, 2015).

The most important risk factor for hemorrhagic stroke remains hypertension (An et al., 2017). Pathophysiological changes due to hypertension are characterised by hypertrophy of the smooth muscles along the vessels. This can lead to vessel remodeling, resulting in vascular stenosis and reduced venous stretching. Finally, these endothelial dysfunctions can lead to disruption of the blood-brain barrier and an impairment of the vascular tone modulation (Xiao et al., 2017).

Another well-described and important risk factor is diabetes mellitus. The clinical picture shows a reduced response of the ß-cells of the pancreatic islets and the adipose tissue to persistent energy excess, resulting in the disturbance of glucose homeostasis, which is usually further supported by nutrition excess and lack of exercise. Initially insulin resistance is increased and metabolic stress is triggered, eventually leading to endothelial dysfunction, including in the cerebral venous system. As a result, vascular changes lead to increased platelet adhesion, coagulation cascades, vasoconstriction and inflammation.

The overall consequences may then be cerebral venous thrombi, hyperemia and cerebral perihematomal edema (PHE) (Xiao et al., 2017).

2.3 Symptoms

The clinical manifestation of patients with an acute stroke can be seen with various symptoms. These range from sudden neurological deficiencies, speech disorders, unilateral body weakness to seizures, severe headaches or conversion disorder. Also laboratory findings such as low blood sugar level can be a symptom of an acute stroke (Yew and Cheng, 2015).

While patient history and physical examination are usually sufficient to make a diagnosis of stroke, neuroimaging helps distinguish between the different types of strokes. The type and combinations of symptoms can often provide initial conclusions. It can be observed that a hemorrhagic stroke often leads to vomiting, headaches, meningitis, coma, symptom

progression over minutes or hours and a diastolic blood pressure level of more than 110mmHg (Yew and Cheng, 2015, Goyal et al., 2015).

Subarachnoid hemorrhage on the other hand is followed by sudden onset of "extermination headache" after 2-8 weeks of headache. In addition, vomiting, photophobia, meningism, focal neurological signs, dullness and loss of consciousness can occur (Yew and Cheng, 2015).

Trigger factors for ICH include great exertion or sudden emotional stress with a time-delayed onset during daily routine activity. Associated neurological symptoms usually occur within the first few minutes or a few hours later.

Headaches occur mainly in larger hematomas as the increased volume leads to traction on the meningeal fibers and increases intracranial pressure and partial blood flow to the ventricles. The increased pressure can also lead to compression of the thalamus and brainstem, which then affects the reticular activating system and results in a lowered attention level, dullness or coma (An et al., 2017).

Vomiting is also associated with an increase in pressure and occurs particularly often in cerebellar hemorrhage but also in 50% of patients with hemispheric hemorrhage. Seizures can be observed in 10% of all patients but in 50% of patients with an ICH in the lobar region. They usually occur with the onset of bleeding or within the first 24 hours (An et al., 2017).

Contralateral motor deficits are most commonly observed in supratentorial ICHs involving the basal ganglia or thalamus. In contrast, infratentorial ICHs often cause brainstem dysfunction such as ocular motor or other cranial nerve abnormalities, which may also cause contralateral motor deficits. In higher hemisphere cognitive disorders such as aphasia, neglect, gaze deviation and hemianopia, lobar ICHs are often causative. Neurological worsening may indicate early hematoma enlargement and secondary brain injury due to PHE expansion (An et al., 2017).

2.4 Pathophysiology

Blood flow to the brain is particularly important. For every cardiac output, almost 20% of the body's oxygen and glucose is distributed to the brain.

To ensure a constant supply of oxygen and limit fluctuations of cerebral blood flow (CBF) well-developed autoregulation mechanisms of the brain help to maintain a normal brain function at all times (Xiao et al., 2017).

These regulatory mechanisms include prevailing arterial pressure, intracranial pressure, arterial blood gases, neural activity and metabolic demand. The different cell types of the

communicating networks regulate each other and the CBF. The cerebral arteries, arterioles and capillaries are all interconnected and provide the necessary oxygen, energy metabolites and nutrients. The cerebral veins are the laxative system and eliminate carbon dioxide and metabolic waste products. They enter them into the big system for them to be eliminated by clearance (Xiao et al., 2017).

In the scenario of an ICH, the mass effect of the hematoma compresses the brain parenchyma and leads to physical disruption of the normal parenchymal architecture. The intracranial pressure is further increased by the expansion of the hematoma (An et al., 2017). The autoregulation mechanisms are sensitive to when the system is out of balance and activate reflex mechanisms to secure and defend the CBF (Xiao et al., 2017).

However, the increased pressure can lead to circulatory disturbances, mechanical deformation, neurotransmitter release, mitochondrial dysfunctions and membrane depolarisations (An et al., 2017). Decompensation of the autoregulatory mechanisms may occur and the CBF consequently may decrease substantially. Due to this dysfunction, CBF is further reduced, metabolic substrate delivery is decreased and generalised cerebral ischemia and hypoxia are induced, resulting in neuronal cell death. If CBF is interrupted or reduced brain function is minimised and neurons will be irreversibly damaged in just a few seconds (Xiao et al., 2017).

Secondary brain damage consists, among other things, of inflammatory and cytotoxic reactions to the hematoma and its degradation products. It is accompanied by intertwined cascades. The time course of the development of a PHE correlates with the manifestation of secondary brain injury and consists of three phases in total (Xiao et al., 2017, Ironside et al., 2019).

In the first phase of PHE evolution (within 1-4 hours of hematoma formation) after ICH, PHE is formed by the pressure gradient between parenchyma and capillaries. This is mediated mainly by retraction of the clot and accumulation of serum proteins.

It is followed by the second phase, the intermediate phase (between 4-72 hours after hematoma formation). The leaked thrombin causes the conversion of fibrinogen to fibrin as part of the coagulation cascade, and pronounced inflammatory responses, such as mitotic induction, leukocyte chemotaxis, platelet aggregation and cytokine release occur (Ironside et al., 2019)

Endothelial cell-cell and cell-matrix interactions altered by thrombin result in further opening of the blood-brain barrier, which is followed by activation of the complement cascade and further amplification of the inflammatory response.

In the late and third phase of PHE evolution (>72 hours of hematoma formation), resolution of the hematoma by erythrocyte lysis and erythrophagocytosis occurs, which in turn involves secondary cell lysis. The resulting hemoglobin deposition in the surrounding parenchyma and its consequences lead and contribute to apoptosis and necrosis (Ironside et al., 2019, An et al., 2017).

The resulting destruction causes further inflammation, red blood cell lysis, iron deposition and further thrombin production, which in turn may directly or indirectly contribute to cerebral venule endothelial dysfunction, microthrombi and outflow reduction.

Combined with other pathophysiological procedures such as oxidative stress and apoptosis, the situation may worsen and lead to destruction of the blood-brain barrier, more brain edema, hydrocephalus and ever-increasing intracranial pressure. The body is caught in a so-called vicious cycle (Xiao et al., 2017).

In summary, CBF is reduced which leads to generalised cerebral ischemia, hypoxia and neuronal cell death. So far, clinical phenomena and strategies of autoregulation after a hemorrhagic stroke are still subject of intensive research (Ironside et al., 2019, Xiao et al., 2017).

2.5 Outcome

In developed countries, stroke is the third leading cause of death, just after heart disease and cancer, the second leading cause of dementia, and the leading cause of disability (Starostka-Tatar et al., 2017, Haacke et al., 2006). Half of all patients with primary ICH die within the first month after the event (Paolucci et al., 2003). But as death rates decline overall in relation to strokes, it is increasingly common for individuals to have to live with their impairments and disabilities after a stroke, which affects overall well-being and quality of life (Haacke et al., 2006).

Hemorrhage in the brain or head can lead to very severe neurological conditions, with accompanying effects ranging from emotional to financial and social costs and lifelong consequences, encompassing a wide range of symptoms and multidimensional limitations. Physical, functional, psychological and social health can be affected (Hakimi and Garg, 2016, Haacke et al., 2006). For motor impairments, patient outcomes range from those with no permanent disabilities to those who are severely disabled, bed-bound, frequently incontinent, and require constant care (Sulter et al., 1999). The well-being of the survivor and the quality of life depend, among other things, on mobility and physical functioning and the ability to perform everyday tasks (Clarke et al., 2002). Relationships and interaction with fellow human beings are often burdened by neuropsychiatric disorders. In addition to

dementia, anxiety and epilepsy, poststroke depression is a serious complication. It has been shown that even mild or moderate difficulties with cognitive function limit the patient's ability to set goals that are part of life satisfaction (Haacke et al., 2006, Starostka-Tatar et al., 2017). Another strong influence on the well-being of patients can be the necessity to change their domicile, as adequate care in their own home is often not possible (Haacke et al., 2006). The size and reliability of the social network around one turns out to be an important factor. Correlations between the support system and well-being after a stroke also highlight the importance of an appropriate rehabilitation program (Clarke et al., 2002).

To monitor and evaluate the patient outcome several scores have been introduced. As for clinical functional outcome after ICH, the modified Ranking Scale (mRS) can be used. It is one of the most frequently used scales to indicate the degree of disability, limitations in activities and changes in everyday life of stroke patients (Sulter et al., 1999). The scale contains points from one through six and describes states between complete health and death. It ranges from 0 = no symptoms, 1 = no significant disability, a few symptoms but able to do all normal activities, 2 = slight disability, does not need assistance but has restrictions, over 3 = moderate disability, needs assistance but can walk alone, 4 = moderately severe disability, unable to walk unassisted, up to 5 = severe disability, bedridden and incontinent or 6 = death (Bruno et al., 2010).

It is a widely accepted measure of functional outcome (Duncan et al., 2000) and is increasingly used as the primary endpoint in clinical trials for acute strokes (Wilson et al., 2005). Its reliability for defining outcome in stroke patients, according to investigations in extensive but fragmented literature, has been shown to be satisfactory and valid (van Swieten et al., 1988, Uyttenboogaart et al., 2005, Banks and Marotta, 2007, Haan et al., 1995).

The Modified Ranking Scale (mRS)	
0	No symptoms.
1	No significant disability. Able to carry out all usual activities, despite some symptoms.
2	Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities.
3	Moderate disability. Requires some help, but able to walk unassisted.
4	Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted.
5	Severe disability. Requires constant nursing care and attention, bedridden, incontinent.
6	Dead.

Table 1:
Modified Ranking Scale (Bruno et al., 2010)

2.6 Therapy

One of the most important factors in the treatment of stroke is the overall goal to quickly set up therapy planning and to initiate first steps to manage the insult. Although there are few medical or surgical treatment options after diagnosis, these require rapid action within the narrow time window, because correct and fast examination of patients with stroke symptoms can reduce disabilities and help avoid reinsult (Yew and Cheng, 2015, Hakimi and Garg, 2016).

This task begins with the first onset of symptoms and extends from the first responder to the paramedic to the interdisciplinary team at the hospital.

The American Heart Association provides guidelines for first responder behaviour in ischemic stroke that should also be followed in spontaneous hemorrhagic stroke. Primarily, clear airways must be ensured, cardiovascular measures must be taken to assist if necessary, and the patient must be transported to a well-equipped hospital as soon as possible.

Upon arrival at the hospital, details regarding the onset of initial symptoms and the patient's medical history must be documented (Hemphill et al., 2015).

If possible, every stroke patient, unless they are subject to intensive care, should be treated in a stroke unit. Treatment in this specialised unit has been shown to reduce mortality and disability and lead to a better functional outcome. Through rapid recognition and treatment of complications, multimodal rehabilitation therapy and early secondary prophylaxis, prognosis is improved (Epple et al., 2015, Weber and Nordmeyer, 2015).

Standard care for all patients with symptoms of a stroke is for them to undergo neuroimaging as soon as possible. This serves primarily to exclude non-ischemic lesions in the central nervous system and to distinguish between ischemic and hemorrhagic strokes (Yew and Cheng, 2015).

After a diagnosis has been established by patient history, physical examination and neuroimaging, further routine examinations follow. These include obtaining an electrocardiogram, determining serum and urine values and checking glucose levels (Hemphill et al., 2015). In addition, early logopedic examination is advisable. Approximately 50% of all patients with an acute stroke suffer from swallowing disorders. Early clarification and assessment of swallowing can prevent a possible risk of aspiration (Weber and Nordmeyer, 2015). Antibiotic therapy has been shown to be effective only after the appearance of clinical and laboratory signs of infection. After three months, patients treated with prophylactic antibiotics did not differ in their functional neurological status from the placebo-treated patient group (Weber and Nordmeyer, 2015). Earlier and more intensive

mobilisation within 24 hours also did not prove effective. It was inferior to later mobilisation. Patients with ICH had a particularly worse outcome (Weber and Nordmeyer, 2015).

Acute treatment of ischemic stroke aims to restore blood flow to the brain tissue. Restoration is achieved by medical treatment with a thrombolytic agent or by endovascular treatment, or both. All disabling strokes should be considered for immediate treatment. Approximately 25% of all ischemic strokes are eligible for drug thrombolysis and 10-12% are eligible for endovascular treatment (Zerna et al., 2018). Achieving successful reperfusion was associated with favorable outcome, functional independence, reduced mortality and a lower rate of symptomatic ICH. The risk of symptomatic ICH was not increased, when successful reperfusion was achieved by mechanical thrombectomy (Kaesmacher et al., 2019). In general, faster reperfusion was associated with better clinical outcome compared with slower reperfusion. Studies demonstrate benefit of endovascular treatment in patients with moderate to severe ischemic stroke (Goyal et al., 2015). In mild stroke, the risk-benefit trade-off is currently the impetus for ongoing randomised clinical trials of thrombolysis (Zerna et al., 2018). The results of the studies however do not yet allow a general treatment recommendation. More studies should be performed on the advantages of endovascular treatment over the medical treatment (Kaesmacher et al., 2019).

Therapy for hemorrhagic strokes focuses on preventing rebleeding and secondary brain damage (Epple et al., 2015). Since important prognostic factors, such as ICH volume, neurologic status and age, are often already established, therapeutic approaches to ICH management focus on prevention and treatment of modifiable factors, such as hematoma progression, PHE, acute hydrocephalus in intraventricular hemorrhage (IVH) and intracranial pressure elevations (Sembill and Kuramatsu, 2019). Aggressive blood pressure management is one therapeutic approach, with the goal of reducing cerebral perfusion. Despite discordant results from the INTERACT-II Study and the ATACH-2 Study, meta-analyses advocate intensive blood pressure reduction because of the associated low hematoma progression, with the possible impact on mortality and functional outcome. Hypotension (<100-120mmHg systolic) should be avoided (Boulouis et al., 2017). Considering our aging society in Germany, blood pressure and blood coagulation of patients are increasingly adjusted by medical treatment options. Therefore, depending on the pre-existing condition, it is necessary, in terms of hemostasis management, to start appropriate replacement therapy of clotting factors or platelets to normalise the International Normalised Ratio (INR) (Hemphill et al., 2015) or to antagonise drug-induced clotting effects. The primary goal is to stop the bleeding. This can be achieved, for example, by intravenous administration of prothrombin complex concentrate or vitamin K (Weber and Nordmeyer, 2015).

Another therapeutic approach is hematoma evacuation. The idea is to reduce the space-occupying effect, the development of edema and the formation of toxic blood degradation products, which in turn can lead to inflammation and local ischemia (Sembill and Kuramatsu, 2019). The large-scale STICH I and II studies showed that hematoma evacuation via open craniotomy did not improve functional outcome in patients (Mendelow et al., 2005, Mendelow et al., 2013). The later published MISTIE III study investigated the minimally invasive surgical techniques of catheter-based hematoma evacuation followed by repeated local thrombolysis. The difficulty of this therapeutic approach is to achieve the desired volume reduction. When the volume reduction reached a value below 15ml, a subgroup analysis showed a more than 10% higher probability for a favourable functional outcome after one year. The outcome depends on the effectiveness and execution of the surgical procedures (Hanley et al., 2019).

The development of PHE is associated with early neurologic deterioration and poor functional outcome. Experimental therapeutic approaches with the application of osmotic substances are available in this setting, but controlled data on these approaches are not yet available (Volbers et al., 2018, Sembill and Kuramatsu, 2019).

IVH is a common complication and a poor predictor of ICH. If hydrocephalus develops, it is treated immediately with an external ventricular drain (EVD) and liquor drainage. Intraventricular fibrinolysis is used to restore communication between the inner and outer cerebrospinal fluid (CSF) spaces. For this purpose, recombinant tissue plasminogen activator is given via the EVD. According to the CLEAR-III study, this therapeutic approach had no effect on functional outcome at 180 days (Hanley et al., 2017). Nevertheless a subgroup analysis showed a correlation between the degree of clot dissolution and a favourable functional outcome, as well as a reduction in mortality. Overlapping lumbar drainage is supposed to reduce shunt dependency with assured communication between the inner and outer CSF spaces, which represents an important treatment goal: with shunt dependence, the probability of disability (mRS 4-5) doubles after six months (Sembill and Kuramatsu, 2019, Hanley et al., 2017). The procedure cannot yet be recommended for routine use, but further studies should evaluate the effect of reducing bleeding volume as much as possible (Neurologie, 2019).

2.7 Imaging

Since imaging can reliably distinguish between an ischemic and hemorrhagic stroke it plays a pivotal role in further treatment triage and is essential for the therapeutic procedure (Yew and Cheng, 2015). Furthermore it can distinguish between the different subtypes of hemorrhage, because clinical signs and scales are usually not sufficient enough to do so

(Hakimi and Garg, 2016). Neuroimaging is expected to reveal the pathophysiology, elucidate the mechanisms and open further options of effective treatment of hemorrhagic stroke (Liebeskind, 2010). It is essential to distinguish the different hemorrhage subtypes from each other. This is also evident within the parenchyma, for example, to distinguish hemorrhagic transformation of ischemia from primary hemorrhage. The mechanisms behind ICH, such as injury to the structural integrity of the cerebrovasculature, abnormalities in coagulation and alteration in blood flow, and also the exclusion of other possible factors, can thus be considered more closely (Hakimi and Garg, 2016, Yew and Cheng, 2015). Therefore, the goal behind imaging is moving further away from the obvious detection of a hematoma to the detection of adjacent features that may provide clues to understanding the pathophysiology (Hakimi and Garg, 2016).

The type of imaging procedure used depends largely on the availability of the imaging modality, eligibility criteria and contraindications of the patient (Yew and Cheng, 2015). Rapid imaging is essential as some patients are at risk of an acute hematoma expansion (HE) which leads to clinically severe neurological deterioration (Hakimi and Garg, 2016). The two well-known technologies Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are predominantly used for imaging in the medical field, with CT being considered the gold standard (Weber and Nordmeyer, 2015).

Computed Tomography

In the past the technique of sequential recording was used. For the sequential imaging of individual axial layers, the patient table had to be advanced step by step between the layers. Today, multi-layer CT enables the examination of larger volume ranges. It also shortens the acquisition time and can acquire thinner layers. The reconstruction layer thickness of the images can be selected retrospectively and is thus independent of collimation. An isotropic resolution, i.e., the same spatial resolution in all three spatial directions, can be achieved. A classic CT consists of a rotating gantry. It contains the essential components, such as the X-ray source, the generator, the detector, the orifice plate on the tube side and the corresponding electronic components and systems for transmitting measurement and control signals between the stationary and rotating parts of the gantry. Other important components are the computer systems for controlling the CT device and image reconstruction, the cooling device and the patient table (Hatem Alkadhi, 2011).

Brain CT is the most common examination in neuroradiology (Hatem Alkadhi, 2011). During the imaging process, the patient is continuously moved through the measuring field on the examination table in the z-direction. The gantry, X-ray tube and detector circle around the patient. The movement of the gantry and the examination table results in a spiral scan of the patient through the X-ray fan, which exposes the individual detector elements. The

acquired data describe a volume composed of three-dimensional image elements, the voxels. With sequential data acquisition layer images are created by recording transversal projections. Tube and detector rotate around the patient while a single table position is maintained. The recording is repeated according to the length of the examination volume. The image is then calculated from the resulting projection data.

The CT density values are expressed in the density scale Hounsfield Unit (HU). The unit is named after Sir Godfrey Newbold Hounsfield, the inventor of CT scans. Water is assigned the fixed value of 0 HU on the scale (Lev and Gonzalez, 2002). All other numbers in the CT can then be converted after a simple calibration (Brooks, 1977). Air for example is set to - 1000 HU and bones can vary between values of several hundred to several thousand HU (Arbabshirani et al., 2018). Thus, the density of tissue is indicated on this scale. Structures that are denser than water have a positive value, structures with lower density have a negative value (Broder and Preston, 2011). The HU values are converted into a digital image for each individual pixel by assigning a grayscale intensity to each value (Lev and Gonzalez, 2002). The lower the density, the darker the tissue appears. The higher the density of the structure, the brighter the color appears. However, the human eye can only distinguish a certain number of shades of gray, which is why the entire range of density values is typically not displayed for an image (Broder and Preston, 2011).

The CT image shows ICH as a hyperdense area (Fiebach and Ebinger, 2008). The peracute bleeding shows up on CT in the first days as a space-occupying lesion with a hypointense margin corresponding to a perifocal edema. Initially, the density of the hemorrhage on the CT may even increase. In the first three days, serum is extruded and the hemorrhage begins to transform. After about three days the density of the hemorrhage starts to decrease. It becomes more and more hypodense until it becomes isodense around the eighth day. At this point it can be difficult to detect the bleeding by CT (Ritter and Schulte-Altedorneburg, 2008).

In addition, a computed tomographic angiography (CTA) is often made. It serves to exclude other pathophysiological processes. Using iodine-containing contrast agents, it is possible to perform it in identical sectioning. The timing of the recording after injection of the contrast medium can be used to control whether an arteriography is performed to detect an aneurysm, a venography to display sinus and cerebral venous thromboses or a display of an arteriovenous malformation (Ritter and Schulte-Altedorneburg, 2008).

Magnetic Resonance Imaging

MRI is a non-invasive technique that makes the structure of the brain and other tissues visible (Kyberkinetik, 2018). The great advantage of MRI is that it does not require ionising

radiation. This is especially useful for patients who need follow-ups over a longer period of time (Sundin et al., 2015).

The image signal results from the oxygen and water content in the tissue and its distribution, as well as the special electromagnetic properties of the hydrogen molecule and its environment. During the examination a very strong homogeneous magnetic field, 30,000-330,000 times stronger than the earth's magnetic field, is required. In addition, excitation energy in the form of electromagnetic waves in the radio wave range of 64-700 MHz is necessary (Kyberkinetik, 2018). These are irradiated via a transmitting antenna for a short time. The excitation frequency of the water molecule corresponds to a very specific magnetic field interference. Through the excitation, the molecule can absorb energy for a short time, which it quickly releases again. With a sensitive antenna the radiated radio frequencies can be observed and even frequency differences of 1 to 0.1 Hz can be measured (Kyberkinetik, 2018).

In order to be able to evaluate the measurements in three-dimensional space the magnetic field is briefly varied in the different spatial directions - in a controlled manner with the help of the gradient coil. This changes the resonance frequency of the water molecule at different locations differently and with the help of computer analysis the location coding is successful. The different frequency signals are calculated back and combined to an image of the examined object. The resulting image represents the different information from the measurement as a light-dark contrast (Kyberkinetik, 2018). During this location encoding a magnetic field is generated with very strong electric currents. It causes vibrational movements which are perceived as loud noises (Kyberkinetik, 2018).

A desired high-image resolution often extends the scanning time. At the same time a high-image resolution also has a high sensitivity to motion artifacts. The increased scanning time however can worsen the motion artifacts. They affect image quality and can interfere with interpretation especially when imaging small lesions (Havsteen et al., 2017). Since the early days of MRI technology motion artifacts have been an obstacle to the development of ever faster sequences and finer detection devices (Havsteen et al., 2017). Movement artifacts can be minimised to a certain extent through optimal receiver coils, fast procedures and sequences, careful positioning of the patient and accurate instructions. However, the physiological sources must always be taken into account. These include respiration, flow and pulse of cardiac cycles, swallowing reflex and spontaneous head movements (Havsteen et al., 2017).

MRI is an excellent tool for determining the presence, size, localisation and extent of hyperacute ischemia. It is also very sensitive for the detection of subacute blood and is even more precise than CT in detecting chronic ICH (Kidwell et al., 2004). However, MRI is less

sensitive to parenchymal bleeding during the first six hours after the onset of stroke symptoms and the reliability of detecting early parenchymal bleeding has not been proven (Kidwell et al., 2004).

Method of Choice

Between CT and MRI, CT has emerged as the standard diagnostic measure. It offers the advantages of being faster, cheaper and usually more available. It can also be used on patients with implanted devices (Weber and Nordmeyer, 2015) and people suffering from claustrophobia (Yew and Cheng, 2015).

The sensitivity of CT is sufficient to detect mass lesions and acute hemorrhages. For the detection of small or posterior fossa strokes, the sensitivity is sometimes insufficient. MRI has a better resolution and therefore a higher sensitivity. For hemorrhagic strokes, however, both imaging methods are equally sensitive (Yew and Cheng, 2015).

2.8 Prognostic Factors

ICH-Score and New Image Markers

In general, there are clinical factors that indicate poor outcome. These prognostic factors include a large volume of hematoma, progressive expansion of hematoma, blood penetration into the ventricle system, hemorrhage location at an infratentorial site, old age of the patient and diseases such as chronic kidney disease, diabetes or coagulation disorders that require anticoagulant medication (An et al., 2017).

Despite these clinical factors, for a long time there was no standard or widely-accepted prognosis model or grading scale for acute spontaneous hemorrhagic strokes analogous to those used for an ischemic stroke, SAH or trauma until the ICH-Score was introduced. It helps to establish a quantification that allows easy communication of the severity of the event (Hemphill et al., 2001). The aim of the ICH-Score is to use criteria that predict outcome and can be quickly and correctly recorded even if the staff is not properly trained. The criteria included are Glasgow Coma Scale (GCS), IVH, age, ICH volume and glucose levels. GCS is weighted the most and divided into three subgroups to correctly reflect its strong weighting in the influence on the ICH-Score (Hemphill et al., 2001). Within the ICH-Score the GCS takes over the quantification of the degree of consciousness. The opening of the eyes, motor reactions and verbal response are independently scored and added together. The characteristics range from spontaneous reaction or reaction on demand to complete lack of reaction (Hemphill et al., 2001). The higher the point total, the better the condition of the patient (Schwenzer and Ehrenfeld, 2008).

In addition to clinical factors and the well-known image markers, such as hematoma volume, new image markers have been investigated more intensively. Although spot sign has already been established in CTA as a radiological marker for HE and thus as a marker for poor outcome, only a small proportion of patients with ICH undergo examination with contrast injection. Moreover, there is growing evidence that various new imaging markers in NECT have great value for outcome prediction of patients with ICH (Sporns et al., 2018b, Morotti et al., 2019). Especially in low- to middle-income countries, patients may benefit from markers that can be found in a NECT (Morotti et al., 2019). Markers that have already been evaluated include for example blend sign, black hole sign (BHS), heterogeneous densities, the appearance of hypodensities and the island sign (IS) (Morotti et al., 2019).

Blend sign is defined as a mixture of a relatively hypoattenuating region with an adjacent hyperattenuating region within the hematoma. Just as the BHS, which is a very hypodense site that appears black on the scan, it was found to effectively predict HE (Li and Yang, 2017, Morotti et al., 2019). Heterogeneous densities and the appearance of hypodensities compared to the hematoma can also stand out from the scan. They are other predictors of poor outcome and can help assess the patient's risk stratification (Morotti et al., 2019, Boulouis et al., 2016, Barras et al., 2009). The IS consists of scattered hematomas, separate from the main hematoma and corresponds to the predictive value of the other markers (Morotti et al., 2019, Zhang et al., 2018).

Radiomics

Different types of features can be obtained from clinical image data. Normally qualitative semantic features are used to describe lesions. However, there are also quantitative features that can be extracted by mathematical algorithms and serve as descriptors. They are usually divided into the following subgroups: shape features, first-order statistics features, second-order statistics features and higher-order statistics features (Rizzo et al., 2018). Radiomics describes the process of extracting quantitative features that enable the conversion of images into exploitable data. This data can then be analysed by the computer and can significantly support and facilitate the decision on further procedure (Gillies et al., 2016).

The features of the data are so special because they cannot be seen with the human eye. There are numerous radiomic features, such as descriptors of the relationship between the voxels, e.g., size zone matrix, neighbor cap gray tone difference matrix, grayscale co-invasion matrix or run-layer matrix, and textures extracted from wavelet and lapplacias of Gaussian feathered images. There are also size- and shape-based features and descriptors of the image intensity histogram (Yip and Aerts, 2016).

First order statistics deal, for example, with energy, which is a measure of the magnitude of voxel values in an image, or with entropy, which indicates the uncertainty and randomness in the image values. They include everything from minimum to maximum, from standard deviation to range, skewness and many other parameters (Pyradiomics, 2016).

The three-dimensional design features form a separate group of radiomics. They contain descriptors for the three-dimensional size and shape of the Region of Interest (ROI). These are characteristics that are independent of the grayscale intensity distribution. As an example the mesh volume, voxel volume, sphericity or surface area can be named here (Pyradiomics, 2016).

Another group of radiomics, also independent of gray level intensity distribution, deals only with two-dimensional size and shape descriptors such as pixel surface, perimeter and elongation (Pyradiomics, 2016).

Another superordinate group of radiomics is the grayscale cooperation matrix. It describes the common second order probability function of the image area bounded by the mask. Elements of this matrix can for example represent the frequency of the combination of layers. For this they include autocorrelations, joint averages, cluster prominence and contrasts.

The other higher-level groups are also based on gray levels. The gray level size zone matrix quantifies grayscale zones in an image, where a grayscale zone is defined as the number of connected voxels that have the same grayscale intensity. The gray level run length matrix describes the number of pixels in the length that successive pixels with the same grayscale value have together. The quantification of the difference in distance between a gray value and the average gray value of its neighbours is called the gray tone difference matrix. And then there is the gray level dependence matrix that describes the number of connected voxels within a certain distance that depend upon particular center voxel (Pyradiomics, 2016).

Automated Pattern Recognition in Images

Clearly, several indicators of the type and severity of ICH exist that can be discerned from imaging data. However, these indicators cannot necessarily be detected visually but only through an analysis of the digital imaging data.

In recent years, data record sizes have increased sharply just as the number of pattern recognition tools has. These have led to further development of methods (Gillies et al., 2016). To exploit these features from image data, we rely on radiomics and computer-based artificial intelligence (AI), namely a ML approach. ML is a rapidly developing technology

whose benefits are being used in many different areas, including the medical field, to optimise and automate processes (El Naqa I., 2015).

There are a number of algorithms for classification that are differently suited for the different tasks of pattern recognition in images. The most commonly used algorithms include Linear Regression, Logistic Regression, Linear Discriminant Analysis, Classification and Regression Trees, Naive Bayes, K-Nearest Neighbors, Learning Vector Quantization, Support Vector Machines, Random Forest, Boosting and AdaBoost (Le, 2019).

Out of these algorithms Random Forest is one of the most popular and powerful ML algorithms. It is a type of ensemble algorithm for ML consisting of many decision trees. To generate a number of uncorrelated decision trees (forest) whose prediction by common voting is more accurate than that of any individual tree the algorithm uses "bagging" and "feature randomness" as the essential methods. Bagging (Bootstrap Aggregation) means that the training data for each of the individual trees is randomly sampled from the original dataset. The training datasets are all of the same size because random selection of the data also allows for replacement. Feature randomness assigns a randomly selected subset of features to each individual tree for the classification task. Combining the two methods produces an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree (Yiu, 2019).

A Random Forest combines several advantages over other classification methods. For example, the classifier trains very fast because the training time increases only linearly with the number of trees. It is especially efficient for large datasets with many features and training data. Due to these advantageous properties, a Random Forest algorithm was also used in this research.

A distinction is made between the strategies of supervised and unsupervised learning. In unsupervised learning, the principle is that the algorithm itself describes inputs and results and can thus, for example, recognize correlations. In supervised learning, on the other hand, the algorithm is given the results for the inputs by an expert. In this way, the algorithm learns to make associations and later to apply what it has learned to unknown data without prior instructions (Pierson, 2016, Reitmaier, 2015).

3 Hypothesis

Since acute primary ICH is particularly time-sensitive in diagnosis and associated with high morbidity and mortality rates, early detection of high-risk patients is extremely important for further therapy strategies and good management of care. With this in mind, we formulate three hypotheses to provide evidence for further evaluation of rapid detection methods in clinical practice:

– We hypothesise that radiomic imaging features obtained from NECT brain scans can be used to predict clinical outcome in ICH patients.

– Furthermore, we believe that the ML approach based on radiomic imaging features has higher predictive power than the conventional approach based on the ICH Score.

– Finally, we hypothesise that a ML approach that incorporates the ICH Score and radiomic imaging features has the highest predictive power.

With respect to processing time requirements, we anticipate that the ML approach could be easily integrated into a clinical routine, thus facilitating early detection of high-risk patients and serving as a general grading score for acute primary ICH.

To test our hypotheses, we designed a study to obtain ground-truth-data to examine the agreement between prediction and outcome. For the outcome prediction we employed a radiomics based ML approach with NECT brain scans of patients with acute primary ICH.

4 Methods

4.1 Study Design

Our study was designed retrospectively. Inclusion criteria were primary spontaneous ICH, confirmed on a NECT and CTA taken no more than six hours after symptom onset. Exclusion criteria were head traumas, vascular malformations, secondary ICH from hemorrhagic transformation of ischemic infarction, brain tumor or primary IVH. Baseline patient characteristics, including mRS, and vascular risk factors such as diabetes mellitus and hypertension were obtained and taken into consideration. Additionally, follow-up procedures, such as craniectomy or intraventricular drainage placement were retrieved from patients´ clinical records and follow-up CT. The study is based on data from 295 patients with ICH aged 18 years or older treated at our university hospital, a tertiary high-volume stroke centre, between January 2010 and April 2019.

We defined a binary outcome. A mRS <3 at discharge was defined as good outcome versus a bad outcome with a mRS >3. In consideration of the inclusion criteria 295 patients were included.

The layout of the study was viewed and approved by the ethics committee (Ethik-Kommission der Ärztekammer Hamburg, WF-035/18) and written informed consent was waived by the institutional review board. All study protocols and procedures were conducted in accordance with the Declaration of Helsinki.

4.2 Image Acquisition

Patients and the corresponding CT scans were acquired from the Radiology Information System (RIS) lights off database. RIS is a large database that is very useful for various functions and serves an efficient workflow in radiology departments. Patient data and images can be stored, distributed and processed in the database (Medicoengineering, 2019). The search mask in RIS was defined based on the inclusion criteria for the study. The relevant time period was set and the selection was limited to head CTs only. All hits displayed were then reviewed individually to verify if the patient met the requirement for inclusion in the study.

All CT scans were performed on 256-slice scanners (Philips iCT 256) with the following imaging parameters: NECT at 120 kV, 280-320 mA, 5.0-mm slice reconstruction; Data sets were checked for quality and excluded in case of strong motion artifacts. After precise selection, all relevant information stored in RIS was reviewed and noted in an Excel

spreadsheet (Microsoft Excel 15.21.1, Microsoft Corporation, One Microsoft Way, Redmond WA, USA). The corresponding native CT, follow-up CT and CTA were loaded and stored in the msPACS system. msPACS is an image archiving and communication system. Within it, medical imaging can be stored and retrieved from many different modalities. To make the data compatible, the Digital Imaging and Communications in Medicine (DICOM) data type was used. DICOM is a universal format for image data and is also compatible for image storage and transmission in msPACS (Medicoengineering, 2019). The different programmes and databases are built to allow physicians to access, interpret and share medical images and reports across multiple platforms (PaxeraHealth Cooperation, 2021). In the next step the DICOM data were sent to the ClearCanvas programme, where the anonymisation and delivery of the data took place. ClearCanvas is a workstation viewer that integrates the RIS client and DICOM PACS and is available via open source access on Github (Synaptive Medical, 2019).

4.3 Image Analysis

The ICH volumes of the anonymised NECT were segmented with Analyze (Analyze 11.0 AnalyzeDirect Inc, Overland Park). We have therefore extracted the DICOM data from CT scans, converted it into Analyze format and implemented it into the software.

The coarser scan series with 4/4 images was used for the analysis. The location of the bleeding was defined and documented. The different regions of bleeding were coded accordingly in an Excel Table (Microsoft Excel 15.21.1, Microsoft Corporation, One Microsoft Way, Redmond WA, USA). We have differentiated between 0 =lobar, 1 =basal ganglia, 2 = thalamic, 3 = brainstem/ Pons, 4 = cerebellar bleeding location.

In the Analyze program the respective file had to be selected and loaded. The contrast window was set correctly from 0 to 80 HU. By clicking on the ROI button we were shown the tool selection. The auto trace tool or manual trace tool were best suited to circle the bleeding. With the slice slider we could navigate through the volume and edit the individual layers until the hematoma was completely captured. In the end we checked if the auto trace tool had leaked into other structures and corrected it if necessary with the "add limit" or "eraser" function. The hematoma volume was then based on the measured area and the corresponding slice thickness automatically calculated for each slice.

4.4 Machine Learning

The subfield of ML used in this study is supervised learning. In this process, an AI is supposed to reproduce regularities that have been taught to it by an expert by specifying

results. For this purpose, a certain output is always assigned to an input. Initially, the result is predetermined and therefore known by the AI. After sufficient training, the algorithm is then supposed to apply what it has learned to unknown data. The learning process can be compared with the correct results and thus be supervised (Rostamizadeh, 2012, Guido and Rother, 2017).

First the classifier was trained on the basic clinical information such as sex and age. In the next step it was fed the image features. The predictive value of the image features was compared to the predictive value of the basic clinical information. In the third step the model was trained with both sources of information in order to reach the full potential of ML based prediction. This process is repeated several times. In the end, the trained ML approach is able to make the desired prediction for test data which it has not been trained with (Arbabshirani et al., 2018).

In detail the following procedure was followed: The consensus ROIs derived from the overlapping segmentations of both readers were converted to an isotropic resolution of 1mm x 1mm x 1mm using sitk BSpline Interolaroren. Then the radiomic features were defined with the PyRadiomics Python package v2.1.0 10 for all data sets.

A total of 1218 quantitative image features were extracted from the ICH ROIs. The group of 252 first order characteristics included 18 radiomics based on unfiltered images, 144 based on wavelet decompositions and 90 radiomics based on log-sigma-Laplacian of Gaussian filters. Further a group of 902 radiomics, assigned to texture elements, was extracted. Of these, 68 were based on unfiltered images, 544 were based on wavelet decomposition and 290 were based on log-sigma-Laplacian of Gaussian filters. Additionally, 14 form elements were extracted. The information gained through radiomics was read into the ML algorithm. A ML based classification was performed using Random Forest algorithms (Python scikit-learn environment v0.20.314). The feature selection, based on the ML algorithm, was performed separately for each training data set considering Gini impurity measures.

Random Forest, the ML technique used for classification, uses multiple decision trees to improve the stability of the result and reduce overfitting of the algorithm (Großekathöfer and Lingner, 2005). They have a comparably low tendency to overfit and are well suited for a large number of heterogeneous predictors (Kniep et al., 2019).
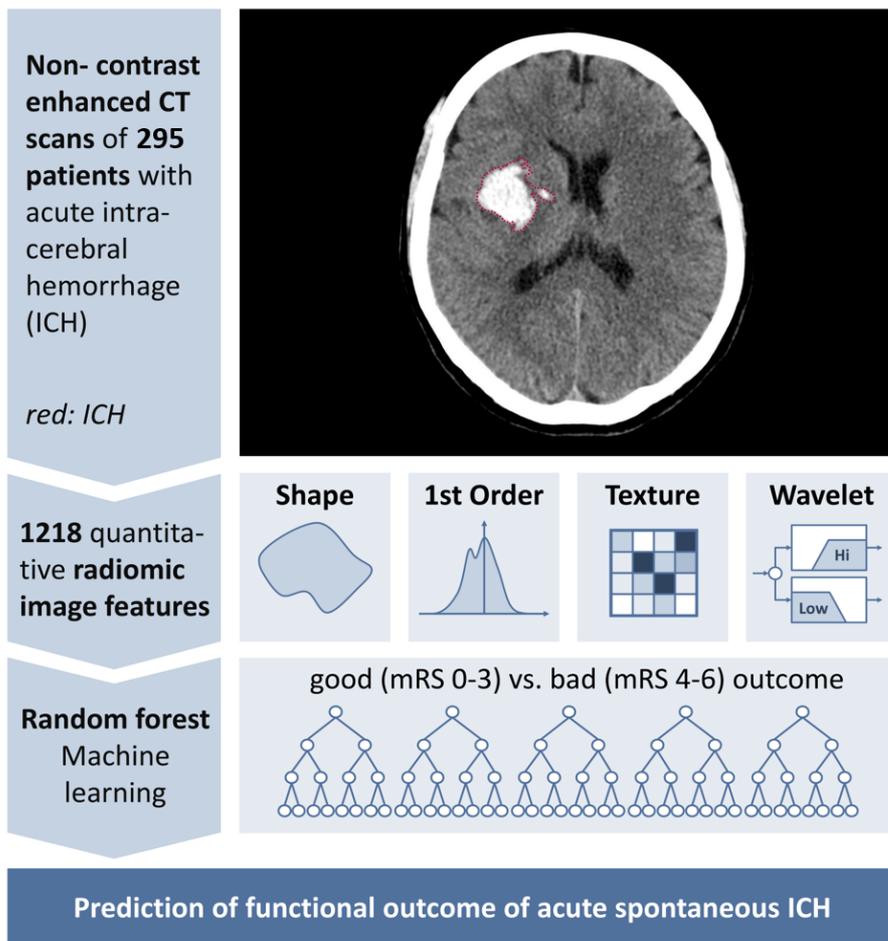
Figure 2:
Conceptual overview of the proposed machine learning approach for intracerebral hemorrhage outcome prediction showing the major processing steps: CT based image acquisition and segmentation, feature extraction (n= 1218), and statistical learning (random forest algorithm). NECT: Noncontrast-Enhanced Computed Tomography; ICH: Intracerebral hemorrhage. CT: Computed tomography; mRS: Modified Rankin Scale. Modified from Nawabi et al. Imaging-Based Outcome Prediction of Acute Intracerebral Hemorrhage. Translational Stroke Research. 2021. CC BY 4.0
https://doi.org/10.1007/s12975-021-00891-8

Decision trees are an approach to construct classification rules. They consist of a sequence of binary decisions represented as a tree. In each node, a yes-no decision is made and in each end node, a decision is made to assign the tree to the class that occurs there most often. Random Forests classify on the basis of many classification trees, the formation of the classification trees required for this purpose is random (Morik and Ligges, 2013).

Validation is achieved by independent validation sets in a model-external fivefold cross-validation. The advantage is that the external model validation is unaffected by cluster correlated data compared to the internal model validation. The stability of the results can be verified by a comparative analysis with randomly selected cross validation sets. In order to evaluate the accuracy of the ML models, we used the native and follow-up CT and the physician's letter to establish the mRS. Cross-validation ensures the generalisability of the model. The generalisation capability of a model is of utmost importance to be able to

24

estimate the predictive quality of unknown data. It must be ensured that the model does not learn the data by heart, a phenomenon known as overfitting. This classically leads to good training performance but extremely poor performance on unknown data.

The principle behind cross-validation is that the training data is randomly divided into approximately equal-sized folds. Always one-fold acts as the test set and the remaining folds are used to train the model. Then the test error of the retained test set is calculated. This procedure is repeated with all folds. The total error is averaged over the test errors of the folds (Großekathöfer and Lingner, 2005).

4.5 Statistics

We performed a multivariate logistic regression analysis to relate the dichotomous dependent variable to the independent variables and predict a good functional outcome. This allowed us to investigate the probability of whether an independent variable contributes to a good or bad outcome.

To do this, a backward-selection algorithm was used with IBM SPSS Statistics v.23.0.0.2. Stepwise regression parameters were set to an entry-threshold of 0.05, a removal-threshold of 0.10, and a maximum number of iterations of 20. To validate our model and to check and confirm its stability a 5-fold cross-validation of the algorithm was performed with independent training and model external validation sets.

With the results of the regression analysis of all cross-validation sets the Receiver Operating Characteristics (ROC) curve was generated. In the following the Area Under the Curve (AUC) was determined. The AUCs are a valid estimate of model classification performance in a generalised environment because each of the analysed Random Forest learners was trained with a unique training data set and tested with a unique external-model validation set. To compare the statistical significance of the independent variables, the P-values were calculated according to the Mann-Whitney/Wilcoxon U Statistic, using the verification R-Package v1.42. We assumed a statistical significance of the independent variable if the P-value was < 0.05.

The confidence interval (CI) for sensitivities and specificities was set to 95%. CIs were bootstrapped with pROC v1.1021 and qwraps2 v0.3.0 R-packages and 2000 repetitions. To control for alpha error inflation Bonferroni adjustments were applied. The statistically significant differences in specificities were determined with a CI analysis using the psychometric v2.2 R-package.
The Matthews Correlation Coefficient (MCC) was calculated to describe the established confidence matrix. For the unbiased comparison of binary classifiers, MCC is considered

the favoured measure (Chicco, 2017). According to its maximum value the classifiers were analysed at operating points. The correlation between reality and the decision of the ML-based algorithm is given in the formula by the covariance in the numerator and the maximum covariance in the denominator. With TP: true positives, TN: true negatives, FP: false positives and FN: false negatives in the confusion matrix, MCC is defined as follows:

$$\text{MCC} = \frac{\text{TP x TN} - \text{FP x FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

5   Results

The study we conducted included data from NECT images of 295 patients. The group was divided into patients with a good outcome, who had a mRS score between 0-3 and patients with a poor outcome, who had a mRS score between 4-6. A total of 75 patients (25%) were discharged with a good mRS score. The remaining 220 patients (75%) belonged to the group with a poor outcome.

The first approach with the conventional predictors met a classifier performance of ROC AUC of 0.77 (95% CI [0.72; 0.82]) and an MCC of 0.31 (95% CI [0.26; 0.36]).

The ML approach based on the new high-end image markers alone yielded a performance of ROC AUC 0.81 (95% CI [0.77; 0.86]) with an MCC of 0.46 [0.42; 0.50] in the random forest classifier. By the cut-off values chosen, the classifier yielded sensitivities of > 80% and specificities > 70% and is therefore superior to the performance based on conventional data.

The combination of the conventional predictors and the high-end image features in the ML approach did not improve the predictive power. Since the values of the combined approach are very similar to those of the evaluation with only high-end image features, figures of this third approach will not be discussed in further detail.

Not all clinical parameters that were included were relevant. There were no statistically significant differences in the clinical parameters of hypertension (P-value = 0.46), diabetes mellitus (P-value = 0.50), sex (P-value = 0.29) or age (P-value = 0.13).

Statistically significant differences were found in the group of patients with a poor outcome. Their images showed more IVH. While only 21% in the first group showed IVH, 58% in the second group did. This resulted in a P-value < 0.001 in our evaluations.

In the group with the poor outcome, higher ICH volumes were found. While the bleeding in that group averaged an area of 35.2 cm^2, the group with the good outcome only filled an area of 8.4 cm^2, which resulted in a statistically significant P-value < 0.001. Supratentorial craniectomies occurred more frequently in the group with a mRS between 4-6. In the group with good outcome, only 8% of patients needed to be relieved with supratentorial craniectomy, while in the group with the worse outcomes 21% had to be relieved. Since this parameter with a P-value of < 0.02 indicates a significant statistical difference, it was also included in the analyses.
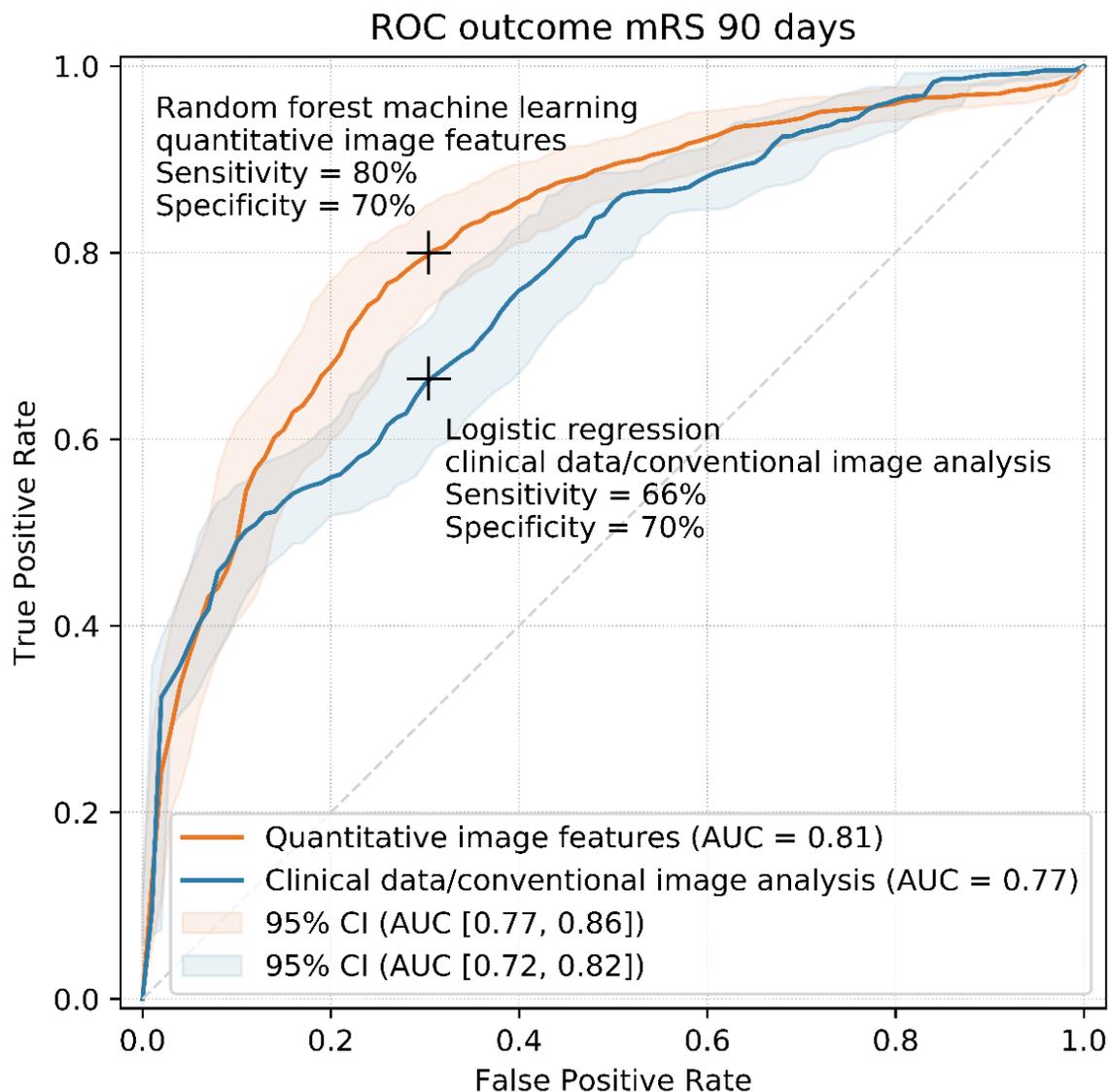
Fig. 3:
Receiver Operating Characteristics (ROC) curves for clinic outcome prediction of logistic regression with conventional predictors (Model 1), the proposed machine learning classifier based on quantitative image features (Model 2). AUC: Area under the curve; CI: confidence interval
Modified from Nawabi et al. Imaging-Based Outcome Prediction of Acute Intracerebral Hemorrhage. Translational Stroke Research. 2021. CC BY 4.0 https://doi.org/10.1007/s12975-021-00891-8

Not included were the clinical parameters without significant statistical difference. These included the age of the patients (P-value = 0.29), gender (P-value = 0.29) and previous diseases, such as high blood pressure (P-value = 0.46) or diabetes mellitus (P-value = 0.50). The location of the bleeding was also not decisive and was therefore not considered further.

The predictive power of the 15 most important characteristics was determined. Based on that information it could be concluded that the second order is generally superior to the first order. Only three predictors, derived from unfiltered images, were included for the outcome prediction. They are the exception. After calculating all metrics of the first 15 parameters, it

was found that, with the 3 exceptions, all metrics were calculated on the basis of texture (fig. 3). Among the 3 exceptions of the predictors derived from unfiltered images are the maximum diameter of the shape (9.1%), the length of the minor axis (8.1%) and the sphericity of the shape (5.8%).

The unfiltered original images contributed only 16% of the total predictive power. That the other predictors with the highest predictive power were mainly derived from images filtered with Log-Sigma (47%) or Wavelet (32%) is shown by our importance analyses. However, the texture metrics are leading in predictive power within the features, which dominated with 69% share of predictive power.

In summary, we know that filter-based texture metrics that have been extracted have the largest share of predictive power.
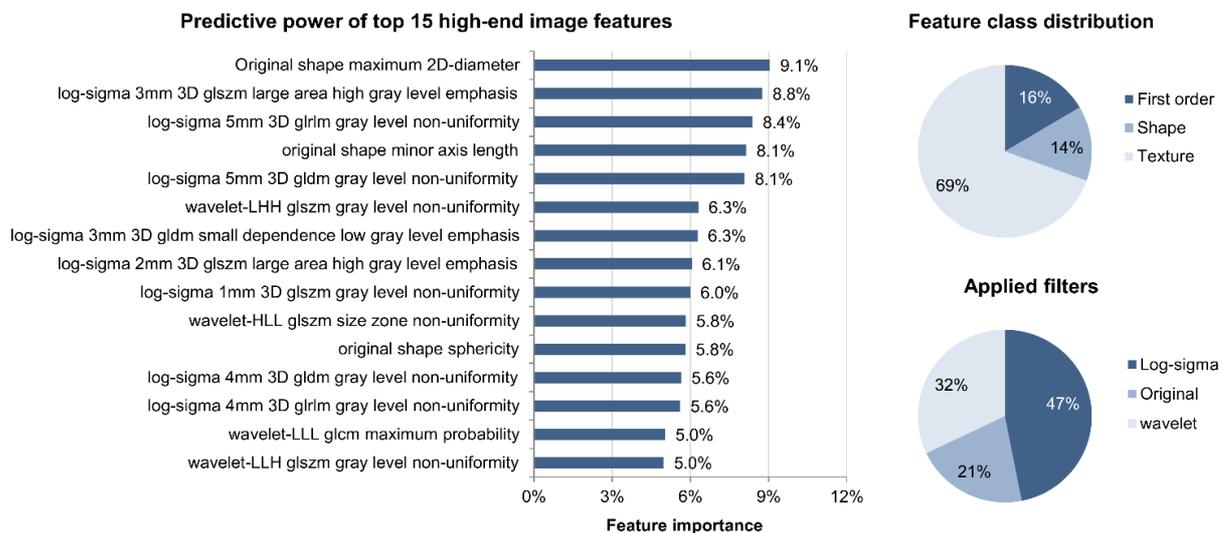

Fig. 4:
Predictive value of high-end image features. Bar charts show feature importance for prediction of the top- 15 high-end image features utilized in model 2 based on Gini impurity measures. Pie charts show distribution of feature classes and applied filters in utilized top-100 predictors. Analyses are based on mean feature importance of all cross-validation training sets. Glcm: gray level co-occurrence matrix; gldm: gray level dependence matrix; glrlm: gray level run length matrix; glszm: gray level size zone; H: high-pass wavelet decomposition; L: low-pass wavelet decomposition.
Modified from Nawabi et al. Imaging-Based Outcome Prediction of Acute Intracerebral Hemorrhage. Translational Stroke Research. 2021. CC BY 4.0 https://doi.org/10.1007/s12975-021-00891-8

6   Discussion

We developed a radiomics based ML algorithm that can predict the outcome of patients after acute ICH at admission. The algorithm distinguishes between an outcome between 0-3 or 4-6 on the mRS. A decision is then made based on the information contained in the image markers in each patient's imaging scans. The ML algorithm has been trained on these predictors and has learned to solely use the diagnostic power of the features, without controlling for patient conditions, image acquisition parameters or a priori provided additional information, such as location of bleeding. Only to account for the effects of therapeutic interventions, decompressive craniectomy was included as a parameter and the algorithm was adjusted for the variable.

The importance of this study is reflected in the results. Quantitative features of routine head CT scans used in the ML classifiers provided high discriminatory accuracy in prediction of functional outcome after ICH. After statistical analysis of the study, the results of the AUC confirm a high diagnostic value. In order to ensure the stability of the study, patients with previous diseases or disease progressions like vascular malformations that could bias the results were excluded from the study beforehand.

Furthermore, our investigations are based on a heterogeneous patient set of 295 patients and their evaluated results. This allows the assumption that our approach is generalisable and thus applicable in a situation such as an admission due to ICH not overlooking important information based on gender or age for example. It is therefore promising for further studies and everyday practice. The mRS is used extensively to evaluate the state of recovery from stroke and thus its validity and reliability are supported by a large body of evidence (Banks and Marotta, 2007). Superiority of mRS compared to other scores and scales has been shown because its assessment is not only based on physical and motor aspects but also on factors that play an essential role for the patient's well-being and self-maintenance. These include cognition and language, mood swings, post stroke depression and the ability to function and interact socially. Also, characteristics of daily life, such as preparing food or managing money are included in the assessment (Banks and Marotta, 2007). In addition to the sensitive selection of criteria, the great experience and security of the readers is also an important aspect of being able to use the mRS sensibly in our study. In the case of retest-reliability, it was found that certification and training of the raters greatly enhanced the reliability of the score.

It is important to note that comorbidities, which are very common in stroke patients, have a direct influence on the mRS. Cardiovascular diseases, metabolic disorders or operations can result in a misapplication of the score (Banks and Marotta, 2007). To avoid misinterpretation, the attributes hypertension, diabetes and surgical relief were noted and

taken into account for each patient. The mRS is also used as the standard score in many other studies. This allows a simplified comparison of the studies. In addition to foreign studies such as the results of the Austrian stroke unit register (Eckhardt R., 2008), the German guideline for neurology also refers to the mRS (Neurologie, 2008). The mRS is also frequently used cross-disciplinary in neurology, neurosurgery and psychiatry (Staykov, 2018).

The decision to dichotomise the score for our study was made to ensure comparability with other studies. Risks can be better represented and odd rations can be calculated (Sandset, 2017). Concerns about dichotomising the score often revolve around the risk of not recognising treatment effects (Banks and Marotta, 2007). Treatment progress in therapy would have to be enormous to be captured and recognised as such by the dichotomised score (Weisscher et al., 2008). Since treatment effects are not the focus of this study but rather outcome prognosis, such concerns are therefore irrelevant.

However, there is disagreement about the dividing line of the mRS between a good outcome and a poor outcome. Overall, there is only a scarce amount of long-term data that relates the quality of life of the different stages to each other. The comparison of individual levels showed that the mRS categories 2 and 3 are very similar. Despite the fact that there was a great heterogeneity in the Level 3 group, it was clearly more similar to the Level 2 than to the Level 4 group (Rangaraju et al., 2017). We decided, contrary to many other studies that define 0-2 as a good and 3-6 as a bad outcome (Sandset, 2017), to separate our classification with 0-3 as a good and 4-6 as a bad outcome, as previously recommended by the relevant study (Rangaraju et al., 2017).

So far, conventional image markers have shown that the size of the volume of ICH plays an important role in the prognosis of the course of a hemorrhagic stroke (Broderick et al., 1993). Among other things, this image feature is also included in the ICH Score in conventional therapy approaches. But markers such as BHS or IS, which were mentioned before as prognostic factors for a poor outcome in ICH patients have further moved into focus of research and led us to suspect that the NECT scans contain significantly more information than what we can see with the human eye. While an experienced and routine radiologist can still identify some of those markers, most of them are only detectable by AI.

With our approach to outcome prediction, we especially want to take into account all the important features not visible to us, thus obtaining more information and preventing incorrect treatment. This allowed us to achieve a higher diagnostic significance than with conventional methods, which are based on the determination of the GCS, among other things.

However, the use of the algorithm does not mean that the information provided by conventional image markers is superfluous. On the contrary, the ML algorithm is a useful addition for clinicians in a situation where quick decisions have to be made under confederation of many aspects and can thus serve as a clinical decision support tool. Early identification of ICH is important to initiate proper therapeutic interventions and assess prognosis. Correct treatment in a stroke unit is an important independent prognostic factor for non-traumatic ICHs. However, patients with non-traumatic ICHs are still very often cared for on neurosurgical wards and not, as recommended, in a stroke unit or neurological intensive care unit. If, contrary to recommendations, stroke patients cannot be transferred to such ward, the algorithm offers the possibility of supporting professionals who are not specialised in the subject field in prognosis and therapy planning.

The algorithm is codependent of the treating clinician. After manually defining the ROI, one of the great advantages of the ML approach is that it automatically extracts features from the images in a data-driven manner (Arbabshirani et al., 2018).

The two readers, who were important for the study, had the knowledge and many years of experience in the respective field to ensure the qualitative standard of the evaluations. Furthermore, they were both blinded to all additional information, like ground truth, the classifier prediction and the other reader's opinion, to ensure objectivity.

Other studies also addressed the predictive value of radiomics in ICH. Here, the results also indicated that radiomics can aid risk assessment, treatment strategy and prognosis in ICH patients. Likewise, positive results were obtained when comparing the predictive power of radiomics and visual ICH markers. There were notable limitations, such as the location of the ICH, the size of the hematoma, or the patient's systolic blood pressure on admission (Haider, 2021). These and other studies suggest that the new method and its potential should be further investigated: A great advantage of this approach is that the ML based version is independent of the patient's state of consciousness. The until now most widely used ICH-score includes the GCS and ICH volume next to other factors (Hemphill et al., 2001). This score is not always well established, as it depends on the investigating clinician or it cannot be established at all due to the patient's state of consciousness. In addition, the collection of the ICH-score requires more time. The reproducibility of the score is more difficult since it is dependent on the investigator. The GCS that is most heavily weighted in the ICH score fails to include brainstem reflexes. Difficulties with the score also arise with intubated patients (Sternbach, 2000). The highest error rates in the score are found at the intermediate level of consciousness, for which the detection of changes in the condition is essential (Rowley and Fielding, 1991). The prognostic value of GCS should be carefully reconsidered when using it (Balestreri et al., 2004).

Spontaneous acute ICHs always require a quick procedure and quick therapy decisions. Often a compromise must be made between simplicity and accuracy of evaluation (Hemphill et al., 2001). Our method again proves to be a better alternative because it is time-saving, reliable, can easily be integrated into clinical routine and requires no compromises since the algorithm is thorough, accurate and fast. Furthermore, it supports the clinician in charge, so outcome prediction and therapy decisions based on it are less dependent on the clinician's level of training. In addition to this support and the time and energy savings there is also no need for costly tests or scores. The ML algorithm serves as a kind of reassurance for the clinician and lends itself to be used as a kind of second opinion. Whenever a new solution is feasible, i.e., time and cost efficient, and can demonstrate a reduction in error rates, its use makes sense, even if the solution is not yet perfect (acatech, 2020).

In order to offer patients the best possible care and keep both primary and secondary neurological damage as little as possible, intensive care therapy demands a multimodal graduated approach (Huge, 2018). This requires consideration and execution of many activities simultaneously. Here too, the algorithm can bring its supporting function to bear.

With more stroke patients, there are also more therapy decisions that need to be made. Especially the most frequently represented group, the elderly patients, must be treated particularly carefully. In addition to a rising prevalence of arteriosclerotic risk factors, such as hypertension, they also frequently show increased fragility of small arteriovenous brain vessels, which incites the probability of stem ganglia bleeding and an increase in degenerative vascular changes in the context of CAA. These age-associated cerebral changes affect the prognosis, outcome and rehabilitation ability of patients in different and diverse ways. It should be noted that for example the decrease in brain volume and neurocognitive deficits also offer less neuroplastic compensation possibilities in the event of a stroke. At the same time however, the risk of critical cerebral pressure development, which could result in neurosurgical relief, decreases. These are all factors that require quick action by the clinician and are difficult to examine and evaluate in the acute situation (Erbguth, 2020). The outcome prediction of the algorithm can serve as auxiliaries here.

In principle the use of ML algorithms can be expected in all phases of patient care: From prevention, screening, diagnosis, therapy planning and therapy to aftercare and rehabilitation. As a rule, ML algorithms should only be used as a guide, comparable to a second opinion. With only a few exceptions, the decision-making authority lies with humans (acatech, 2020).

Despite the promising results, several limitations deserve comment. There are very classical limitations, which almost always occur in association with radiomics based analyses and classifications. These include the misclassification of the ground truth

definitions and the under- or overfitting of the algorithm. This would make the results less generalisable.

This problem could also arise from more study specific limitations, such as a limited number of patient records. Although we are satisfied with the stability of the model, as it allows us to assume sufficient robustness, a study with more patients would allow a greater generalisation of the results. However, it should be taken into account that we have relied on a large, highly heterogeneous data set of 295 patients obtained over a period of almost a decade. The fivefold cross-validation is sufficient in such cases to guarantee the feasibility and performance of the method. Additionally, the interobserver agreement could have been determined to support the objectivity of the study.

Furthermore, our study is a single-center study. A multicenter study with different patient cohorts would further validate the results we have presented. A sufficient stability of the classifier performance is guaranteed by the standardised and calibrated image parameters. A further limitation results from the retrospectivity of the study. A similar study with a prospective data set would result in a higher scientific statement.

As mentioned earlier, there are disagreements about the dividing line of a good or a poor outcome when dichotomising the mRS. Since we decided to separate the score at 0-3 and 4-6, contrary to many other studies, this complicates to some extent the comparability with other studies. Basically, it is difficult to find a consensus on how to assess the outcome situation. However, many studies give more weight to factors such as activities of daily living as an assessment measure and thus refer more to sub-items such as the possibility of unassisted walking. In our study, the focus is less on aspects such as the ability to participate in outdoor activities, but much more on survival and later coping in life under assessment of moderate or severe disability and quality of life, so that a separation between 3 and 4 makes sense, since severe dependence is often scored from 4 and 4-5 stands for severe disability (Ferro JM, 2011, Zeltzer, 2008).

To train the ML algorithm and prepare it to predict outcome, the ROIs have to be read into the AI system. The ROIs were previously defined manually by experienced clinicians. Here the performed semiautomatic segmentation of ROIs still implies an observer-dependent possible source of error within the ML process. The human being is fallible and the results of the study are based on an initially manually trained AI. To minimise this source of error, the consensus segmentations were performed by two readers. A semi-automatic demarcation of the ROIs was chosen. Furthermore, it is shown that radiomic features are comparably stable with regard to variations in segmentations. It could also be argued that it is very time consuming to insert the many variables into the ML model. However, the ML model can teach itself with the additional data (Arbabshirani et al., 2018).

Hence, more studies with even larger and prospective datasets on this approach could further validate this objective tool for outcome prediction. Taking into account other clinical variables as presented in the study of Wang et al (Wang et al., 2019) or including markers like ICH expansion and PHE which are known for increasing the mass effect and causing a rapid deterioration of the situation, as presented by Morotti et al (Morotti et al., 2019), offers interesting possibilities to further explore options. These future projects are useful to further investigate our approach and to compare it with other studies.

In conclusion, it can be said that the model can serve as a useful addition to everyday hospital life. It is easy to use even by inexperienced or untrained clinicians. It is almost always available and can therefore be easily integrated into clinical routine and thus also prevent misjudgement of severity based on only visually presented information (Fig. 5). Especially because of its reliability and speed it offers a good prognosis estimation tool. This provides the opportunity to establish a uniform score and that will help to respond quickly and reliably to the patient's current state with coordinated therapy planning, reprioritisation and action.
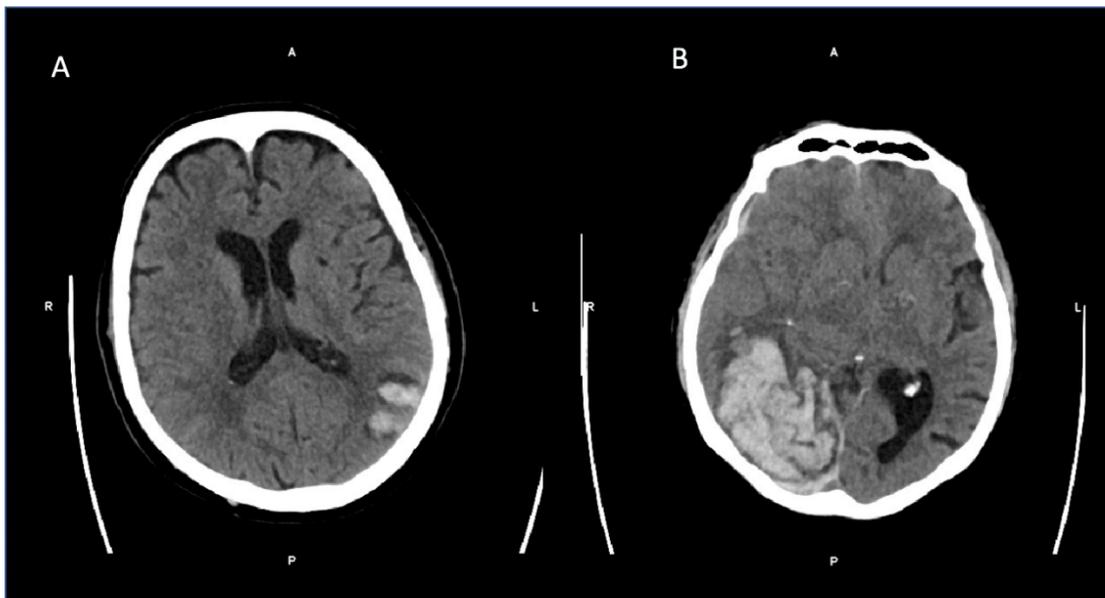

Fig. 5
A: Nonenhanced CT with a left lobar ICH with an ICH volume of 10.935 mL and an outcome on the mRS of 4
B: Nonenhanced CT with a right lobar ICH with an ICH volume of 218.559mL and an outcome on the mRS of 3

7 Summary

This study focuses on hemorrhagic strokes, the smaller but more fatal group of strokes (Yew and Cheng, 2015). Early symptoms include vomiting and coma with rapid deterioration of symptoms also commonly seen in hemorrhagic strokes. Further neurological deterioration is usually accompanied by an increase in hematoma volume or the progressions of perihematomal edema. Consequently, CBF may be reduced and disturbed leading to acute CSF circulation disorder. This imbalance can only be bridged for a short time by autoregulatory mechanisms before a potentially fatal dysregulation occurs. Given this scenario, a set of cascades and immunological processes get in motion that contribute to cell death and necrosis, entering a vicious circle (Xiao et al., 2017). Uptodate, standardized therapies for patients with ICH have not been established yet.

This life-threatening scenario makes hemorrhagic stroke one of the three leading causes of death and the leading cause of disability (Starostka-Tatar et al., 2017). The high morbidity comes along with emotional, financial and physical impairments contributing to a much wider burden (Paolucci et al., 2003).

Since medical and surgical treatment options are limited and quick initiation of procedures is mandatory, imaging plays a particularly important role in the early management of ICH (Hakimi and Garg, 2016, Yew and Cheng, 2015). In brief, early imaging aims to identify high-risk patients at an early stage by predicting their clinical outcome as this sets the course for the further patient's triage. Prior to this triage reliable differentiation between stroke subtypes and other primary causes of hemorrhage need to be ruled out.

Next to imaging, clinical factors can also be used contributing to a more precise clinical outcome prediction. In this regard, the ICH-score is the most clinically established score to combine both clinical and imaging features (Hemphill et al., 2001). Imaging features included in the ICH Score are limited to semantic feature attributes. High-quantitative imaging markers that can be extracted by mathematic algorithms may therefore serve as additional prognostic value. These markers may be summarized as radiomics and contain even more nuanced information about the bleeding. Some features for example are size- or shape-based while others are descriptors of the image intensity histogram or the relationship between voxels (Pyradiomics, 2016). Few of these markers can be directly recognized and evaluated by the reader but most are invisible to the human perception (Yip and Aerts, 2016). Algorithms are already used in neuroradiology but are mainly applied to conditions such as ischemic stroke or traumatic brain injury (Heo et al., 2019, Rau CS, 2018). ML-approaches in ICHs are mainly focused on automatic volume determination and rapid diagnosis (Arbabshirani et al., 2018, Scherer et al., 2016). Wang et al were the first to integrate laboratory data, initial clinical situation and imaging information into a ML-

approach to predict disease progression. However, the approach focused on ICH volume and location, presence of IVH, compression of the ventricles and midline shift (Wang et al., 2019). In conclusion, the integration of high-quantity features into an ML-approach could improve the diagnostic performance in predicting functional outcome.

We hypothesized that radiomic features in patients with ICH may predict functional outcome of patients with ICH. Secondly, we hypothesized that our proposed approach may have a higher predictive power than the above-mentioned ICH score.

For this purpose, a radiomics based ML-approach was designed in which the algorithm was trained by supervised learning. For the ML-based classification of clinical outcome, a random forest method was implemented. This ensures stability of the algorithm's results and reduces over-adaptation. Random forest has a relatively small tendency to overfit and is therefore suitable for large amounts of heterogeneous data sets. Validation was achieved by model-external fivefold cross-validation. In contrast to model-internal validation, it remains unaffected by cluster correlations. By applying the random forest technique and model-external five-fold cross validation, the approach is generalizable and applicable to unknown data.

Conventional predictors summed up in the ICH-Score met a classifier performance of ROC AUC of 0.77 and a MCC of 0.31. The second approach based on the new high-end image markers met a performance of ROC AUC of 0.81 and a MCC of 0.46 in the random forest classifier. With a sensitivity of over 80% and a specificity of over 70% the classifier is superior to the approach based on the only the conventional data. Combining the two approaches did not further improve the predictive power.

The model was adjusted for factors that were significantly associated with clinical outcome. These included average area of hemorrhage, supratentorial craniectomy and presence of an IVH.

The study demonstrates that the extracted quantitative features from the CT head scan have a high discriminatory accuracy in predicting the functional outcome after ICH. Statistical analysis confirms the diagnostic power, with the filter-based texture metrics having the greatest predictive power. Ultimately, our proposed radiomics based ML-approach has a higher diagnostic performance than qualitative semantic features and the ICH score with the great advantage of being independent of the patient's state of consciousness.

Thus, our proposed method could serve as an additional clinical decision support tool. This translational approach is mainly limited as imaging segmentation is performed semi-manually and therefore lacks practicability in the clinical emergency setting. Further

improvement of our approach may be achieved by including of a fully automated segmentation method with deep learning algorithms.

Bibliography

ACATECH 2020. acatech POSITION — Machine learning in der Medizintechnik.

AN, S. J., KIM, T. J. & YOON, B. W. 2017. Epidemiology, Risk Factors, and Clinical Features of Intracerebral Hemorrhage: An Update. *J Stroke,* 19**,** 3-10.

ARBABSHIRANI, M. R., FORNWALT, B. K., MONGELLUZZO, G. J., SUEVER, J. D., GEISE, B. D., PATEL, A. A. & MOORE, G. J. 2018. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med,* 1**,** 9.

BALESTRERI, M., CZOSNYKA, M., CHATFIELD, D. A., STEINER, L. A., SCHMIDT, E. A., SMIELEWSKI, P., MATTA, B. & PICKARD, J. D. 2004. Predictive value of Glasgow coma scale after brain trauma: change in trend over the past ten years. *Journal of Neurology, Neurosurgery &amp; Psychiatry,* 75**,** 161.

BANKS, J. L. & MAROTTA, C. A. 2007. Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials. *Stroke,* 38**,** 1091-1096.

BARRAS, C. D., TRESS, B. M., CHRISTENSEN, S., MACGREGOR, L., COLLINS, M., DESMOND, P. M., SKOLNICK, B. E., MAYER, S. A., BRODERICK, J., DIRINGER, M. N., STEINER, T. & DAVIS, S. 2009. Density and Shape as CT Predictors of Intracerebral Hemorrhage Growth. *AHA Journal,* 40.

BOULOUIS, G., MOROTTI, A., BROUWERS, H. B., CHARIDIMOU, A., JESSEL, M. J., AURIEL, E., PONTES-NETO, O., AYRES, A., VASHKEVICH, A., SCHWAB, K. M., ROSAND, J., VISWANATHAN, A., GUROL, M. E., GREENBERG, S. M. & GOLDSTEIN, J. N. 2016. Association Between Hypodensities Detected by Computed Tomography and Hematoma Expansion in Patients With Intracerebral Hemorrhage. *JAMA Neurol,* 73**,** 961-8.

BOULOUIS, G., MOROTTI, A., GOLDSTEIN, J. N. & CHARIDIMOU, A. 2017. Intensive blood pressure lowering in patients with acute intracerebral haemorrhage: clinical outcomes and haemorrhage expansion. Systematic review and meta-analysis of randomised trials. *J Neurol Neurosurg Psychiatry*.

BRODER, J. & PRESTON, R. 2011. Imaging the Head and Brain *Science Direct***,** 45.

BRODERICK, J. P., BROTT, T. G., DULDNER, J. E., TOMSICK, T. & HUSTER, G. 1993. Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. *Stroke,* 24**,** 987-93.

BROOKS, R. A. 1977. A quantitative theory of the Hounsfield unit and its application to dual energy scanning. *J Comput Assist Tomogr,* 1**,** 487-93.

BRUNO, A., SHAH, N., LIN, C., CLOSE, B., HESS, D. C., DAVIS, K., BAUTE, V., SWITZER, J. A., WALLER, J. L. & NICHOLS, F. T. 2010. Improving Modified Rankin Scale Assessment With a Simplified Questionnaire. *Stroke,* 41**,** 1048-1050.

CHEN, Q. X., T.; ZHANG, M.; XIA, N.; LIU, J.; YANG, Y. 2021. Radiomics in Stroke Neuroimaging: Techniques, Applications, and Challenges. *Aging and Disease*.

CHICCO, D. 2017. Ten quick tips for machine learning in computational biology. *BioData Mining*.

CLARKE, P., MARSHALL, V., BLACK, S. E. & COLANTONIO, A. 2002. Well-Being After Stroke in Canadian Seniors. *AHA Journal,* 33.

DIENER, H. & WEIMAR, C. 2012. Akuttherapie des ischämischen Schlaganfalls. *Thieme Verlag*.

DUNCAN, P. W., JORGENSEN, H. S. & WADE, D. T. 2000. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. *Stroke,* 31**,** 1429-38.

ECKHARDT R., S. S., BRAININ M. 2008. Management von Patienten mit intrazerebralen Blutungen an Österreichischen Stroke Units: Ergebnisse des GÖG-BIQG Österreichischen Stroke Unit Registers 2003-2007. *Wiener Medizinische Wochenzeitschrift*

EL NAQA I., M. M. J. 2015. *What Is Machine Learning?*, Springer, Chem.

EPPLE, C., KURAMATSU, J., HUTTNER, J., STEINER, T. & SCHWAB, S. 2015. Update on Therapy of Intracerebral Hemorrhage.

ERBGUTH, F. 2020. [Acute stroke treatment in old age]. *Med Klin Intensivmed Notfmed,* 115**,** 351-366.

FERRO JM, C. I., COUTINHO JM, CANHÃO P, BARINAGARREMENTERIA F, CUCCHIARA B, DEREX L, LICHY C, MASJUAN J, MASSARO A, MATAMALA G, POLI S, SAADATNIA M, STOLZ E, VIANA-BAPTISTA M, STAM J, BOUSSER MG;. 2011. *Decompressive surgery in cerebrovenous thrombosis: a multicenter registry and a systematic review of individual patient data* [Online]. [Accessed].

FIEBACH, J. B. & EBINGER, M. 2008. Imaging diagnostics in apoplectic stroke: CT and MRT - Indications and limitations. *Georg Thieme Verlag KG*.

GILLIES, R. J., KINAHAN, P. E. & HRICAK, H. 2016. Radiomics: Images Are More than Pictures, They Are Data. *Radiology,* 278**,** 563-77.

GOYAL, M., DEMCHUK, A. M., MENON, B. K., EESA, M., REMPEL, J. L., THORNTON, J., ROY, D., JOVIN, T. G., WILLINSKY, R. A., SAPKOTA, B. L., DOWLATSHAHI, D., FREI, D. F., KAMAL, N. R., MONTANERA, W. J., POPPE, A. Y., RYCKBORST, K. J., SILVER, F. L., SHUAIB, A., TAMPIERI, D., WILLIAMS, D., BANG, O. Y., BAXTER, B. W., BURNS, P. A., CHOE, H., HEO, J. H., HOLMSTEDT, C. A., JANKOWITZ, B., KELLY, M., LINARES, G., MANDZIA, J. L., SHANKAR, J., SOHN, S. I., SWARTZ, R. H., BARBER, P. A., COUTTS, S. B., SMITH, E. E., MORRISH, W. F., WEILL, A., SUBRAMANIAM, S., MITHA, A. P., WONG, J. H., LOWERISON, M. W., SAJOBI, T. T. & HILL, M. D. 2015. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med,* 372**,** 1019-30.

GROSSEKATHÖFE, U. & LINGNER, T. 2005. Neue Ansätze zum maschinellen Lernen von Alignments. *AG Neuroinformatik*

GRUNWALD, Z., BESLOW, L. A., URDAY, S., VASHKEVICH, A., AYRES, A., GREENBERG, S. M., GOLDSTEIN, J. N., LEASURE, A., SHI, F. D., KAHLE, K. T., BATTEY, T. W., SIMARD, J. M., ROSAND, J., KIMBERLY, W. T. & SHETH, K. N. 2017. Perihematomal Edema Expansion Rates and Patient Outcomes in Deep and Lobar Intracerebral Hemorrhage. *Neurocrit Care,* 26**,** 205-212.

GUIDO, S. & ROTHER, K. 2017. *Einführung in Machine Learning mit Python Praxiswissen Data Science,* Heidelberg, O'Reilly.

HAACKE, C., ALTHAUS, A., SPOTTKE, A., SIEBERT, U., BACK, T. & DODEL, R. 2006. Long-Term Outcome After Stroke. *Stroke,* 37**,** 193-198.

HAAN, R. D., LIMBURG, M., BOSSUYT, P., MEULEN, J. V. D. & AARONSON, N. 1995. The Clinical Meaning of Rankin ‘Handicap’ Grades After Stroke. *Stroke,* 26**,** 2027-2030.

HAIDER, S. Q., A.; JAIN, A.; THARMASEELAN, H.; BERSON, E.; ZEEVI, T.; MAJIDI, S.; FILIPPI, C.; ISEKE, S.; GROSS, M.; ACOSTA, J.; MALHOTRA, A.; KIM, J.; SANSING, L.; FALCONE, G.; SHETH, K.; PAYABVASH, S. 2021. *european journal of neurology*.

HAKIMI, R. & GARG, A. 2016. Imaging of Hemorrhagic Stroke. *Continuum (Minneap Minn),* 22**,** 1424-1450.

HANLEY, D. F., LANE, K., MCBEE, N., ZIAI, W., TUHRIM, S., LEES, K. R., DAWSON, J., GANDHI, D., ULLMAN, N., MOULD, W. A., MAYO, S. W., MENDELOW, A. D., GREGSON, B. A., BUTCHER, K., VESPA, P., WRIGHT, D. W., KASE, C. S., CARHUAPOMA, J. R., KEYL, P. M., DIENER-WEST, M., MUSCHELLI, J., BETZ, J., THOMPSON, C., SUGAR, E., YENOKYAN, G., JANIS, S., JOHN, S., HARNOF, S., LOPEZ, G., ALDRICH, E. F., HARRIGAN, M., ANSARI, S., JALLO, J., CARON, J., LEDOUX, D., ADEOYE, O., ZUCCARELLO, M., ADAMS, H., ROSENBLUM, M., THOMPSON, R. E. & AWAD, I. 2017. Thrombolytic removal of intraventricular haemorrhage in treatment of severe stroke: results of the randomised, multicentre, multiregion, placebo-controlled CLEAR III trial. *The Lancet*.

HANLEY, D. F., THOMPSON, R. E., ROSENBLUM, M., YENOKYAN, G., LANE, K., MCBEE, N., MAYO, S. W., BISTRAN-HALL, A. J., GANDHI, D., MOULD, W. A., ULLMAN, N., HASAN ALI, M., J RICARDO CARHUAPOMA, M., PROF CARLOS S KASE, M., PROF KENNEDY R LEES, M., JESSE DAWSON, M., ALASTAIR WILSON, P., JOSHUA F BETZ, M., ELIZABETH A SUGAR, P., YI HAO, D., RADHIKA AVADHANI, M., PROF JEAN-LOUIS CARON, M., MARK R HARRIGAN, M., ANDREW P CARLSON, M., DIEDERIK BULTERS, F., DAVID LEDOUX, M., PROF JUDY HUANG, M., CULLY COBB, M., GAURAV GUPTA, M., RYAN KITAGAWA, M., MICHAEL R CHICOINE, M., HIREN PATEL, P., ROBERT DODD, M., PAUL J CAMARATA, M., STACEY WOLFE, M., AGNIESZKA STADNIK, M., P LYNN MONEY, B., PATRICK MITCHELL, P., ROSARIO SARABIA, M., PROF SAGI HARNOF, M., PROF PAL BARZO, P., PROF ANDREAS UNTERBERG, M., JEANNE S TEITELBAUM, M., PROF WEIMIN WANG, M., PROF CRAIG S ANDERSON, M., PROF A DAVID MENDELOW, F., BARBARA GREGSON, P., SCOTT JANIS, P., PROF PAUL VESPA, M., WENDY ZIAI, M., PROF MARIO ZUCCARELLO, M. & PROF ISSAM A AWAD, M. 2019. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. *The Lancet,* 393.

HATEM ALKADHI, S. L., PAUL STOLZMANN, HANS SCHEFFEL 2011. Wie funktioniert CT?

HAVSTEEN, I., OHLHUES, A., MADSEN, K. H., NYBING, J. D., CHRISTENSEN, H. & CHRISTENSEN, A. 2017. Are Movement Artifacts in Magnetic Resonance Imaging a Real Problem?—A Narrative Review. *Frontiers in Neurology,* 8.

HEMPHILL, J. C., BONOVICH, D. C., BESMERTIS, L., MANLEY, G. T. & JOHNSTON, S. C. 2001. The ICH Score. *Stroke,* 32**,** 891-897.

HEMPHILL, J. C., GREENBERG, S. M., ANDERSON, C. S., BECKER, K., BENDOK, B. R., CUSHMAN, M., FUNG, G. L., GOLDSTEIN, J. N., MACDONALD, R. L., MITCHELL, P. H., SCOTT, P. A., SELIM, M. H. & WOO, D. 2015. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage. *Stroke,* 46**,** 2032-2060.

HEO, J., YOON, J. G., PARK, H., KIM, Y. D., NAM, H. S. & HEO, J. H. 2019. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke,* 50**,** 1263-1265.

HUGE, V. 2018. Intensivmedizinische Therapie intrazerebraler Blutungen. *Medizinische Klinik - Intensivmedizin und Notfallmedizin,* 113**,** 164-173.

IRONSIDE, N., CHEN, C.-J., DING, D., MAYER, S. A. & CONNOLLY, E. S. 2019. Perihematomal Edema After Spontaneous Intracerebral Hemorrhage. *Stroke,* 50**,** 1626-1633.

KAESMACHER, J., CHALOULOS-IAKOVIDIS, P., PANOS, L., MORDASINI, P., MICHEL, P., HAJDU, S. D., RIBO, M., REQUENA, M., MAEGERLEIN, C., FRIEDRICH, B., COSTALAT, V., BENALI, A., PIEROT, L., GAWLITZA, M., SCHAAFSMA, J., PEREIRA, V. M., GRALLA, J. & FISCHER, U. 2019. Mechanical Thrombectomy in Ischemic Stroke Patients With Alberta Stroke Program Early Computed Tomography Score 0–5. *Stroke,* 50**,** 880-888.

KIDWELL, C. S., CHALELA, J. A., SAVER, J. L., STARKMAN, S., HILL, M. D., DEMCHUK, A. M., BUTMAN, J. A., PATRONAS, N., ALGER, J. R., LATOUR, L. L., LUBY, M. L., BAIRD, A. E., LEARY, M. C., TREMWEL, M., OVBIAGELE, B., FREDIEU, A., SUZUKI, S., VILLABLANCA, J. P., DAVIS, S., DUNN, B., TODD, J. W., EZZEDDINE, M. A., HAYMORE, J., LYNCH, J. K., DAVIS, L. & WARACH, S. 2004. Comparison of MRI and CT for Detection of Acute Intracerebral Hemorrhage. *JAMA,* 292**,** 1823-1830.

KNIEP, H. C., MADESTA, F., SCHNEIDER, T., HANNING, U., SCHONFELD, M. H., SCHON, G., FIEHLER, J., GAUER, T., WERNER, R. & GELLISSEN, S. 2019. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology,* 290**,** 479-487.

KNÖS, N., JANSEN, O., BRENCK, J. & DIENER, H. 2012. Lokale Therapie des ischämischen Insultes: mechanische Thrombektomie. *Georg Thieme Verlag KG*.

KURAMATSU, J., GERNER, S., HUTTNER, H. & SCHWAB, S. 2017. Acute Management of Anticoagulation-Associated Intracerebral Hemorrhage. *Neurology International Open*.

KYBERKINETIK, M.-P.-I. F. B. 2018. Funktionsweise von Magnetresonancetomographie (MRT).

LE, J. 2019. *The Top 10 Machine Learning Algorithms Every Beginner Should Know* [Online]. Available: https://builtin.com/data-science/tour-top-10-algorithms-machine-learning-newbies? [Accessed 21.07.2021].

LEV, M. H. & GONZALEZ, R. G. 2002. CT Angiography and CT Perfusion Imaging. *Science Direct*.

LI, R. & YANG, M. 2017. A comparative study of the blend sign and the black hole sign on CT as a predictor of hematoma expansion in spontaneous intracerebral hemorrhage.

LIEBESKIND, D. S. 2010. Imaging the future of stroke: II. Hemorrhage. *Ann Neurol,* 68**,** 581-92.

MAYER, S. A., BRUN, N. C., BEGTRUP, K., BRODERICK, J., DAVIS, S., DIRINGER, M. N., SKOLNICK, B. E. & STEINER, T. 2008. Efficacy and Safety of Recombinant Activated Factor VII for Acute Intracerebral Hemorrhage. *New England Journal of Medicine,* 358**,** 2127-2137.

MEDICOENGINEERING, D. O. O. 2019. *Picture archiving and communication system (PACS, RIS)* [Online]. Available: https://www.medico-eng.com/en/products.php?id=71 [Accessed 18.07.2021 2021].

MENDELOW, A. D., GREGSON, B. A., FERNANDES, H. M., MURRAY, G. D., TEASDALE, G. M., HOPE, D. T., KARIMI, A., SHAW, M. D. M. & BARER, D. H. 2005. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *The Lancet*.

MENDELOW, A. D., GREGSON, B. A., ROWAN, E. N., MURRAY, G. D., GHOLKAR, A. & MITCHELL, P. M. 2013. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial. *The Lancet*.

MORIK, K. & LIGGES, U. 2013. Wissensentdeckung in Datenbänken- Entscheidungsbäume. 377.

MOROTTI, A., BOULOUIS, G., DOWLATSHAHI, D., LI, Q., BARRAS, C. D., DELCOURT, C., YU, Z., ZHENG, J., ZHOU, Z., AVIV, R. I., SHOAMANESH, A., SPORNS, P. B., ROSAND, J., GREENBERG, S. M., AL-SHAHI SALMAN, R., QURESHI, A. I., DEMCHUK, A. M., ANDERSON, C. S., GOLDSTEIN, J. N. & CHARIDIMOU, A. 2019. Standards for Detecting, Interpreting, and Reporting Noncontrast Computed Tomographic Markers of Intracerebral Hemorrhage Expansion. *Ann Neurol,* 86**,** 480-492.

NEUROLOGIE, D. G. F. 2008. Intrazerebrale Blutungen. 26.

NEUROLOGIE, D. G. F. 2019. *MISTIE III* [Online]. Available: https://dgn.org/neuronews/neuronews/mistie-iii/ [Accessed 17.02.2021].

PAOLUCCI, S., ANTONUCCI, G., GRASSO, M. G., BRAGONI, M., COIRO, P., ANGELIS, D. D., FUSCO, F. R., MORELLI, D., VENTURIERO, V., TROISI, E. & PRATESI, L. 2003. Functional Outcome of Ischemic and Hemorrhagic Stroke Patients After Inpatient Rehabilitation. *Stroke,* 34**,** 2861-2865.

PAXERAHEALTH COOPERATION. 2021. *PaxeraHealth Products* [Online]. Available: https://paxerahealth.com/products/ [Accessed 18.07.2021 2021].

PIERSON, L. 2016. *Data Sience für Dummies,* Weinheim, Wiley-VCH Verlag.

PYRADIOMICS 2016. Radiomic Features. *Github*.

RANGARAJU, S., HAUSSEN, D., NOGUEIRA, R. G., NAHAB, F. & FRANKEL, M. 2017. Comparison of 3-Month Stroke Disability and Quality of Life across Modified Rankin Scale Categories. *Interventional Neurology,* 6**,** 36-41.

RAU CS, K. P., CHIEN PC, HUANG CY, HSIEH HY, ET AL. 2018. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models.

REITMAIER, T. 2015. *Aktives Lernen für Klassifikationsprobleme unter der Nutzung von Strukturinformationen,* Kassel, kassel university press.

RITTER, M. A. & SCHULTE-ALTEDORNEBURG, G. 2008. Imaging of Intracerebral Haemorrhage - CT, MRI or Both? *Georg Thieme Verlag KG*.

RIZZO, S., BOTTA, F., RAIMONDI, S., ORIGGI, D., FANCIULLO, C., MORGANTI, A. G. & BELLOMI, M. 2018. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp,* 2**,** 36.

ROSTAMIZADEH, A. 2012. *Foundations of machine learning,* Cambridge, MIT Press.

ROWLEY, G. & FIELDING, K. 1991. Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *The Lancet,* 337**,** 535-538.

SANDSET, E. C. 2017. Outcome measures in Stroke *European Stroke Organisation*

SCHERER, M., CORDES, J., YOUNSI, A., SAHIN, Y.-A., GÖTZ, M., MÖHLENBRUCH, M., STOCK, C., BÖSEL, J., UNTERBERG, A., MAIER-HEIN, K. & ORAKCIOGLU, B. 2016. Development and Validation of an Automatic Segmentation Algorithm for Quantification of Intracerebral Hemorrhage. *Stroke,* 47**,** 2776-2782.

SCHWENZER, N. & EHRENFELD, M. 2008. *Chirurgische Grundlagen,* Stuttgart, Georg Thieme Verlag KG.

SEMBILL, J. & KURAMATSU, J. 2019. *Acute treatment of intracerebral hemorrhage*.

SMITH, E. E. & ROMERO, J. R. 2018. Mixed emotions: What to do with patients who have lobar and deep hemorrhages on MRI? *Neurology,* 90**,** 55-56.

SPORNS, P. B., KEMMLING, A., MINNERUP, J., HANNING, U. & HEINDEL, W. 2018a. Imaging-based outcome prediction in patients with intracerebral hemorrhage. *Acta Neurochir (Wien),* 160**,** 1663-1670.

SPORNS, P. B., KEMMLING, A., SCHWAKE, M., MINNERUP, J., NAWABI, J., BROOCKS, G., WILDGRUBER, M., FIEHLER, J., HEINDEL, W. & HANNING, U. 2018b. Triage of 5 Noncontrast Computed Tomography Markers and Spot Sign for Outcome Prediction After Intracerebral Hemorrhage. *Stroke,* 49**,** 2317-2322.

SPRIGG N, F. K., APPLETON JP, AL-SHAHI SALMAN R, BERECZKI D, BERIDZE M, ET AL.; TICH-2 INVESTIGATORS 2018. Tranexamic acid for hyperacute primary IntraCerebral Haemorrhage (TICH-2): an international randomised, placebo-controlled, phase 3 superiority trial. *Lancet*.

STAROSTKA-TATAR, A., LABUZ-ROSZAK, B., SKRZYPEK, M., GASIOR, M. & GIERLOTKA, M. 2017. [Definition and treatment of stroke over the centuries]. *Wiad Lek,* 70**,** 982-987.

STAYKOV, D. 2018. What`s new in the management of intracerebral hemorrhage. *Journal für Neurologie, Neurochirurgie und Psychiatrie*

STERNBACH, G. L. 2000. The Glasgow Coma Scale11Medical Classics is coordinated by George Sternbach, MD, of Stanford University Medical Center, Stanford, California. *The Journal of Emergency Medicine,* 19**,** 67-71.

SULTER, G., STEEN, C. & KEYSER, J. D. 1999. Use of the Barthel Index and Modified Rankin Scale in Acute Stroke Trials. *Stroke,* 30**,** 1538-1541.

SUNDIN, A., WILLS, M. & ROCKALL, A. 2015. Radiological Imaging: Computed Tomography, Magnetic Resonance Imaging and Ultrasonography. *Front Horm Res,* 44**,** 58-72.

SYNAPTIVE MEDICAL. 2019. *ClearCanvas by Synaptive Medical* [Online]. Available: https://www.clearcanvas.ca [Accessed 18.07.2021 2021].

TRUELSEN, T., BEGG, S. & MATHERS, C. 2006. *The global burden of cerebrovascular disease* [Online]. World Health Organisation. Available: https://www.who.int/healthinfo/statistics/bod_cerebrovasculardiseasestroke.pdf [Accessed 01.10.219 2019].

UYTTENBOOGAART, M., STEWART, R. E., VROOMEN, P. C. A. J., KEYSER, J. D. & LUIJCKX, G.-J. 2005. Optimizing Cutoff Scores for the Barthel Index and the Modified Rankin Scale for Defining Outcome in Acute Stroke Trials. *Stroke,* 36**,** 1984-1987.

VAN SWIETEN, J. C., KOUDSTAAL, P. J., VISSER, M. C., SCHOUTEN, H. J. & VAN GIJN, J. 1988. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke,* 19**,** 604-7.

VOLBERS, B., GIEDE-JEPPE, A., GERNER, S., SEMBILL, J., KURAMATSU, J., LANG, S., LÜCKING, H., STAYKOV, D. & HUTTNER, H. 2018. Peak perihemorrhagic edema correlates with functional outcome in intracerebral hemorrhage. *Neurology*.

WANG, H. L., HSU, W. Y., LEE, M. H., WENG, H. H., CHANG, S. W., YANG, J. T. & TSAI, Y. H. 2019. Automatic Machine-Learning-Based Outcome Prediction in Patients With Primary Intracerebral Hemorrhage. *Front Neurol,* 10**,** 910.

WEBER, R. & NORDMEYER, H. 2015. [State-of-the-art Treatment of Acute Stroke]. *Fortschr Neurol Psychiatr,* 83**,** 641-50; quiz 651-2.

WEISSCHER, N., VERMEULEN, M., ROOS, Y. B. & DE HAAN, R. J. 2008. What should be defined as good outcome in stroke trials; a modified Rankin score of 0-1 or 0-2? *J Neurol,* 255**,** 867-74.

WILSON, J. T. L., HAREENDRAN, A., HENDRY, A., POTTER, J., BONE, I. & MUIR, K. W. 2005. Reliability of the Modified Rankin Scale Across Multiple Raters. *Stroke,* 36**,** 777-781.

XIAO, M., LI, Q., FENG, H., ZHANG, L. & CHEN, Y. 2017. Neural Vascular Mechanism for the Cerebral Blood Flow Autoregulation after Hemorrhagic Stroke. *Neural Plast,* 2017**,** 5819514.

XU, X. Z., J.; YANG, K.; WANG, Q.; CHEN, X.; XU, B. 2021. Prognostic prediction of hypertensive intracerebral hemorrhage using CT radiomics and machine learning. *Brain and Behavior*.

YEW, K. S. & CHENG, E. M. 2015. Diagnosis of acute stroke. *Am Fam Physician,* 91**,** 528-36.

YIP, S. S. & AERTS, H. J. 2016. Applications and limitations of radiomics. *Phys Med Biol,* 61**,** R150-66.

YIU, T. 2019. *Understanding Random Forest* [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 [Accessed 21.07.2021].

ZELTZER, L. 2008. *Modified Rankin Scale (MRS)* [Online]. Stroke Engine. Available: https://strokengine.ca/en/assessments/modified-rankin-scale-mrs/ [Accessed 23.07.2021 2021].

ZENG, X., DENG, A. & DING, Y. 2017. The INTERSTROKE study on risk factors for stroke. *The Lancet,* 389**,** 35.

ZERNA, C., G., T., CAMPBELL, B. R., J. & HILL, M. 2018. Current practice and future directions in the diagnosis and acute treatment of ischaemic stroke. *Lancet,* 392.

ZHANG, F., LI, H., QIAN, J., ZHANG, S., TAO, C., YOU, C. & YANG, M. 2018. Island Sign Predicts Long-Term Poor Outcome and Mortality in Patients with Intracerebral Hemorrhage.

Attachments

Baseline characteristics	All (n=295)	mRS 0-3 (n=75)	mRS 4-6 (n=220)	*P*-Value
Clinical parameters				
Age [years], **median (IQR)**	74 (63;81)	74 (60;78)	73.5 (65;81)	0.13
Female, n (%)	134 (45)	38 (51)	96 (44)	0.29
Hypertension, n (%)	227 (77)	60 (80)	157 (76)	0.46
Diabetes mellitus, n (%)	44 (15)	13 (17)	31 (14)	0.50
CT parameters				
Bleeding location, n (%)				
• **Lobar**	135 (46)	39 (52)	196 (44)	0.21
• **Basal Ganglia**	103 (35)	23 (31)	80 (36)	0.37
• **Thalamus**	14 (5)	4 (5)	10 (5)	0.78
• **Brainstem and Pons**	12 (4)	1 (1)	11 (5)	0.17
• **Cerebellar**	31 (11)	8 (11)	23 (10)	0.96
Intraventricular hemorrhage, n (%)	143 (48)	16 (21)	127 (58)	<0.001
ICH volume [cm3]**, median (IQR)**	23.5 (8.9;60.7)	8.4 (3.5;22.3)	35.2 (13.6;76.2)	<0.001
Procedure process and results				
Craniectomy				
• **supratentorial**	50 (17)	6 (8)	49 (20.9)	0.02
• **infratentorial**	5 (2)	1 (1)	4 (1.7)	0.78
mRS, n (%)				
• **0-3**	75 (25)	75 (100)	-	
• **4-6**	220 (75)	-	235 (100)	

Table 2:
Comparison of baseline demographic, clinical and radiological characteristics between ICH patients with good clinical outcome (modified Rankin Scale (mRS) 0-3) versus poor clinical outcome (mRS 4-6).
Modified from Nawabi et al. Modified from Nawabi et al. Imaging-Based Outcome Prediction. Translational Stroke Research. 2021. CC BY 4.0 https://doi.org/10.1007/s12975-021-00891-8

	Regression coefficient			OR 95% CI	
Predictor	B	P-value	OR	Lower	Upper
Intraventricular hemorrhage (1 = yes)	-1.248	<0.001	0.287	0.147	0.559
Location supra- vs infra-tentorial (1 = supra-tentorial)	0.968	0.049	2.633	1.005	6.896
Location basal ganglia (1 = yes)	-0.761	0.033	0.467	0.232	0.940
Location brainstem/pons (1 = yes)	-1.975	0.085	0.139	0.015	1.315
ICH volume (l)	-0.046	<0.001	0.955	0.938	0.972
Constant	0.098	0.831	1.103		

Table 3:
Prediction of functional outcome (mRS 0-3): Logistic regression model using basic clinical data and conventional image markers. B: Logistic regression coefficient; CI: Confidence interval
Modified from Nawabi et al. Modified from Nawabi et al. Imaging-Based Outcome Prediction. Translational Stroke Research. 2021. CC BY 4.0 https://doi.org/10.1007/s12975-021-00891-8

	Regression coefficient			OR 95% CI	
Predictor	B	P-value	OR	Lower	Upper
Intraventricular hemorrhage (1 = yes)	-1.248	<0.001	0.287	0.147	0.559
Location supra- vs infra-tentorial (1 = supra-tentorial)	0.968	0.049	2.633	1.005	6.896
Location basal ganglia (1 = yes)	-0.761	0.033	0.467	0.232	0.940
Location brainstem/pons (1 = yes)	-1.975	0.085	0.139	0.015	1.315
ICH volume (l)	-0.046	<0.001	0.955	0.938	0.972
Constant	0.098	0.831	1.103		

Table 4:
Classification performance metrics of logistic regression model based on clinical data and conventional image markers compared to random forest ML algorithm using high-end image features. Metrics are shown at 70% specificity cut-off points. Bonferroni corrections have been applied to account for alpha spending error. PPV: Positive predictive value; NPV: Negative predictive value; MCC: Matthews correlation coefficient; CI: Confidence interval.
Modified from Nawabi et al. Imaging-Based Outcome Prediction. Translational Stroke Research. 2021. CC BY 4.0 https://doi.org/10.1007/s12975-021-00891-8

Note of Thanks

First I want to thank Prof. Dr. med. Jens Fiehler for the acceptance as a doctoral student at his institute, the provision of a topic and the good supervision.

I also would like to thank my doctoral mother Priv.-Doz. Dr. med. Uta Hanning for the review of the dissertation, the continuous good supervision and support.

A special thanks goes to my mentor Dr. med. Jawed Nawabi for the support, the proofreading, the good advice and motivation, the nice cooperation and especially the guidance and mentoring even when we were no longer working at the same university.

Furthermore, I would like to thank Dr. med. Helge Kniep for his support and explanations, especially regarding the technical and computer science aspects.

I would also like to thank the rest of the team of neuroradiology at the UKE for the friendly reception in the offices, the broad-minded support and answering of all questions.

Also I am thanking my former fellow student Henriette Scholz for the nice coffee breaks, the many long nights and weekends spent analyzing our data together for both of our projects.

For the encouraging support and motivation during the writing process, I especially thank my family and close ones: my father for helping me to structure my work, for new ideas and suggestions for improvement, and for frequent proofreading. My mother, my uncle Matthias and Kai for the constant motivation and reassurance when I had little drive myself.

And last but not least, I would like to thank Jennifer Meiners and Austin Hugenberg who, as native English speakers, turned my text into a more readable one.

Resume (Lebenslauf)

Entfällt aus datenschutzrechtlichen Gründen.

Affidavit

Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe 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.

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

Unterschrift: ....