

Computer-aided diagnosis for CT based clinical
triage of ischemic and hemorrhagic stroke patients:
a deep learning and quantitative image analysis
approach

THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE IN TECHNICAL MEDICINE

FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF TWENTE
FEBRUARY 2019

BY

R.H.J. BERGMANS

**UNIVERSITY
OF TWENTE.**



MASSACHUSETTS
GENERAL HOSPITAL

©2019 – R.H.J. BERGMANS
ALL RIGHTS RESERVED.

GRADUATION COMMITTEE

| | |
|--------------------------------------|--|
| Chair and technical supervisor | prof. dr. ir. C.H. Slump Faculty of Electrical Engineering, Mathematics & Computer Science (EWD), Robotics and Mechatronics (RAM) University of Twente, Enschede, The Netherlands |
| Medical supervisor | R. Gupta, MD, PhD Department of Radiology, Massachusetts General Hospital Advanced X-ray Imaging Sciences Center Harvard Medical School, Boston, USA |
| Technical supervisor (additional) | C.O. Tan, PhD Cerebrovascular Research Lab, Spaulding Rehab Hospital Department of Radiology, Massachusetts General Hospital Harvard Medical School, Boston, USA |
| Process supervisor | R.J. Haarman, MSc Faculty of Science and Technology (TNW) Technical Medicine University of Twente, Enschede, The Netherlands |
| External committee member | dr. E. Groot Jebbink, MSc Faculty of Science and Technology (TNW) Multi-Modality Medical Imaging (M3I) University of Twente, Enschede, The Netherlands |

Computer-aided diagnosis for CT based clinical triage of ischemic and hemorrhagic stroke patients: a deep learning and quantitative image analysis approach

SUMMARY

Stroke is a sudden development in cerebral disturbance due to insufficient blood supply and imposes one of highest and still increasing health related socioeconomic burdens. In ischemic stroke, large vessel occlusions caused by thromboembolism, lead to a necrotic core (and salvageable penumbra) in the area supplied by the occluded artery. Patients having a small core at the time of imaging potentially benefit from endovascular clot removal in addition to administration of anticoagulants. Although computed-tomography (CT) is the diagnostic imaging workhorse used in emergency radiology, it is not able to detect ischemic lesions in the early hours post ictus. Only diffusion weighted imaging (DWI) is highly sensitive for critically ischemic brain tissue, but it usually not available for emergency situations. Ideally, DWI infarct core lesions should be derivable from conventional non-contrast CT (NCCT) and (multi-phase) CT angiography (CTA), the latter potentially capturing the underlying physiology leading to ischemic stroke.

The first part of this research describes the development of a deep learning system that is able to predict DWI stroke lesions from 3-phase (dual-energy) CT data (NCCT, early and delayed phase CTA). Data from 293 patients have been collected and pre-processed. Results showed that predictions from a deep learning system based on the DeepMedic architecture reached a better Dice Similarity Coefficient (DSC) with respect to segmented DWI ground truth lesions than a U-Net based architecture. Moreover, the predicted infarct core lesion volumes on individual patient basis led to low incorrect EVT decision rates, implying that the system might be useful for triaging AIS patients.

In hemorrhagic stroke, blood leaks into the brain parenchyma due to vessel rupture, causing a hematoma and leading to serious complications, such as neurologic deterioration, increased intracranial pressure or death. Especially hematoma expansion in the early hours post ictus is associated with high mortality, so imaging based assessment of likelihood of expansion is important to select the appropriate therapy for individual patients (drug treatment or highly invasive craniectomy). Dual-energy CT (DECT) angiography allows quantification of iodine in contrast medium, and thus potentially represents the (ongoing) bleeding pattern causing hematoma expansion more accurately than conventional single-energy CTA. Recently, the I2-score, combining two iodine features, has been proposed as new DECT imaging marker to identify expanders with high sensitivity and specificity. Also, quantitative texture features from NCCT readily have been described as potential predictive imaging markers to predict hematoma expansion.

The second part of this research explored if a logistic regression model including both iodine and texture features is able to outperform the I2-score in predicting hematoma expansion. Processing of NCCT and DECT angiography scans of 100 patients (38 training, 62 test) identified 6 iodine and 82 texture features. An extensive feature selection and ranking process showed that NCCT texture features performed inferior to DECT iodine features. A logistic regression model combining the three highest ranked features (two iodine, one texture) was not able to outperform the I2-score on the independent test set. Moreover, performance of both the I2-score and the combined *iodine-texture* model dropped in response to slight data modifications in this test set, indicating required validation of existing models.

Future work should focus on standardizing data acquisitions and incorporating clinical parameters in the prediction models used in both studies. For the deep learning part on ischemic stroke, a tradeoff between number of CTA acquisition phases and the ability to predict DWI lesions or separating the problem in smaller sub steps should be investigated. For hemorrhagic stroke, the I2-score should be adjusted and validated on a larger cohort to avoid overfitting. Finally, the ability of deep learning to predict hematoma expansion based on combined NCCT and DECT images should be explored.

Contents

| | |
|--|----|
| CHAPTER 1 | |
| CLINICAL BACKGROUND | 1 |
| 1.1 Acute ischemic stroke | 1 |
| 1.1.1 Treatment | 1 |
| 1.1.2 Imaging in AIS. | 3 |
| 1.1.3 Imaging of collateral circulation. | 4 |
| 1.1.4 The Massachusetts General Hospital acute stroke imaging algorithm. | 5 |
| 1.1.5 Dual-energy CT. | 6 |
| 1.2 Intracerebral hemorrhage | 6 |
| 1.2.1 Treatment | 6 |
| 1.2.2 Imaging in ICH. | 7 |
| | |
| CHAPTER 2 | |
| TECHNICAL BACKGROUND | 9 |
| 2.1 Quantitative image texture analysis | 9 |
| 2.2 Deep learning | 10 |
| 2.2.1 Convolutional neural networks | 11 |
| 2.2.2 CNN architectures | 12 |
| 2.2.3 Training CNNs. | 12 |
| | |
| CHAPTER 3 | |
| RESEARCH OBJECTIVE | 15 |
| 3.1 Acute ischemic stroke | 15 |
| 3.2 Intracerebral hemorrhage | 16 |
| | |
| CHAPTER 4 | |
| PREDICTING ACUTE ISCHEMIC STROKE DWI LESIONS FROM 3-PHASE CT DATA USING DEEP LEARNING | 17 |
| 4.1 Introduction | 17 |
| 4.2 Methods | 18 |
| 4.2.1 Subject selection and data collection | 18 |
| 4.2.2 Data pre-processing. | 18 |
| 4.2.3 Deep learning architectures | 19 |
| 4.2.4 Training models. | 22 |
| 4.2.5 Experiments | 24 |
| 4.2.6 Performance metric | 24 |
| 4.2.7 Data analysis. | 24 |
| 4.3 Results | 25 |
| 4.3.1 Data characteristics | 25 |
| 4.3.2 Stepwise configuration selection | 26 |
| 4.3.3 Subgroup analysis | 26 |

| | |
|--|----|
| 4.3.4 ROC and Precision–recall analyses | 26 |
| 4.3.5 Bland–Altman analysis | 27 |
| 4.3.6 Visualization of results | 29 |
| 4.4 Discussion | 29 |
| 4.5 Conclusion | 31 |
| CHAPTER 5 | |
| INVESTIGATING THE PREDICTIVE ABILITY OF NCCT TEXTURE FEATURES IN COMBINATION WITH DECT IODINE FEATURES FOR INTRACEREBRAL HEMATOMA EXPANSION | 33 |
| 5.1 Introduction | 33 |
| 5.2 Methods | 34 |
| 5.2.1 Patient selection and data collection. | 34 |
| 5.2.2 Image analysis. | 34 |
| 5.2.3 Statistical analysis | 36 |
| 5.3 Results | 37 |
| 5.3.1 Training and test set. | 37 |
| 5.3.2 Feature selection and ranking. | 37 |
| 5.3.3 Final model selection and performance on training set | 38 |
| 5.3.4 Performance on independent test set: sensitivity analysis. | 39 |
| 5.4 Discussion | 40 |
| 5.5 Conclusion | 41 |
| CHAPTER 6 | |
| CONCLUSION | 43 |
| REFERENCES | 45 |
| APPENDIX | 53 |
| A. Dual–Energy CT post processing | 53 |
| B. Computation of texture features from GLCMs | 54 |
| C. Laplacian of a Gaussian filtering | 54 |
| D. Basic CNN operations | 55 |
| E. DSC optimization | 57 |
| F. Overview of tested CNN configurations | 58 |

Listing of figures

| | | |
|------|--|----|
| 1.1: | Concept of ischemic core and penumbra | 2 |
| 1.2: | Example of final stroke lesion volume depending on reperfusion time | 3 |
| 1.3: | Collateral scoring on axial MIPs | 4 |
| 1.4: | The MGH acute stroke imaging algorithm | 5 |
| 1.5: | Data present after imaging with MGH acute ischemic stroke protocol | 6 |
| 1.6: | Early hematoma growth after ICH. | 7 |
| | | |
| 2.1: | Computation of gray level co-occurrence matrices (GLCM) | 10 |
| 2.2: | Convolution operation as sliding filter | 11 |
| | | |
| 4.1: | Overview of the data pre-processing pipeline | 18 |
| 4.2: | Example of global overview of the implemented U-Net architecture | 20 |
| 4.3: | Implementation of ResDense block | 21 |
| 4.4: | Implementation of ResNeXt block | 21 |
| 4.5: | Implementation of DeepMedic architecture | 22 |
| 4.6: | ROC and Precision-Recall curves from voxel-wise analyses | 27 |
| 4.7: | Correlation and Bland-Altman plots of predicted lesion volumes | 28 |
| 4.8: | Visualization of CT input data, DWI, segmented ground truth and prediction | 29 |
| | | |
| 5.1: | Overview of the images used throughout the processing pipeline | 35 |
| 5.2: | Overview of feature selection process | 36 |
| 5.3: | Cumulative feature count of survived selection simulations on the training set | 39 |
| 5.4: | Akaike Information Criterion (AIC) for top ranked feature models. | 39 |
| 5.5: | Noise sensitivity analysis of I2-score and <i>iodine-texture</i> model | 40 |
| | | |
| A.1: | DECT three material decomposition post processing | 53 |
| D.1: | The concept of stride and pooling | 55 |
| D.2: | The concept of depthwise separable convolutions | 56 |
| D.3: | Effect of a 50% dropout layer | 57 |
| D.4: | Non-linear ReLU function applied to matrix | 57 |
| E.1: | Sensitivity of final DSC to chosen probability threshold | 58 |

Listing of abbreviations

| | |
|---------|--|
| 2D | Two dimensional |
| 3D | Three dimensional |
| ACO | Anterior circulation occlusion |
| ADC | Apparent diffusion coefficient |
| AIS | Acute ischemic stroke |
| AIC | Akaike information criterion |
| AP | Average precision |
| ASPECTS | Alberta stroke program early CT score |
| AUC | Area under the curve |
| BBB | Blood-brain barrier |
| BN | Batch normalization |
| CBF | Cerebral blood flow |
| CI | Confidence interval |
| CNN | Convolutional neural network |
| CT | Computed tomography |
| CTA | Computed tomography angiography |
| CTP | Computed tomography perfusion |
| DECT | Dual-energy computed tomography |
| DICOM | Digital Imaging and Communications in Medicine |
| DL | Dice loss |
| DSC | Dice similarity coefficient |
| DWI | Diffusion weighted imaging |
| EVT | Endovascular thrombectomy |
| F/U | Follow-up |
| FCN | Fully convolutional network |
| FGBG | Foreground-background |
| FLAIR | Fluid-attenuated inversion recovery |
| FOV | Field of view |
| GLCM | Gray level co-occurrence matrix |
| GPU | Graphics processing unit |
| HE | Hematoma expansion |
| HU | Hounsfield Unit |
| IAT | Intra-arterial thrombolysis |
| ICH | Intracerebral hemorrhage |
| ICP | Intracranial pressure |
| INR | International normalized ratio |
| IPH | Intraparenchymal hemorrhage |
| IV-tPA | Intravenous tissue plasminogen activator |
| IVH | Intraventricular hemorrhage |
| LoG | Laplacian of a Gaussian |
| LR | Logistic regression |
| LVO | Large vessel occlusion |
| MCA | Middle cerebral artery |
| MGH | Massachusetts General Hospital |
| MIP | Maximum intensity projection |
| MR | Magnetic resonance |
| MTT | Mean transit time |
| NCCT | Non-contrast computed tomography |
| NifTI | Neuroimaging Informatics Technology Initiative |

| | |
|-------|---|
| NIHSS | National Institute of Health Stroke Scale |
| PR | Precision-recall |
| ReLU | Rectified Linear Unit |
| ROC | Receiver operating characteristic |
| ROI | Region of interest |
| SAH | Subarachnoid hemorrhage |
| SBP | Systolic blood pressure |
| SDH | Subdural hemorrhage |
| SDL | Soft dice loss |
| SDSC | Soft dice similarity coefficient |
| SECT | Single-energy computed tomography |
| SW | Sliding window |
| SWI | Susceptibility weighted imaging |
| TIA | Transient ischemic attack |
| VOI | Volume of interest |
| VNC | Virtual non-contrast |
| VNCa | Virtual non-calcium |

1

Clinical background

Stroke is a sudden development of disturbance in cerebral function caused by insufficient blood supply. In the United States, stroke is responsible for about 1 of every 20 deaths, and worldwide it imposes a huge socioeconomic burden, being persistently among the top three causes of disability and premature mortality. Over 85% of all strokes are ischemic strokes, while only 10% are caused by intracerebral hemorrhage (ICH), and another 3% by subarachnoid hemorrhage.^{1,2}

1.1 ACUTE ISCHEMIC STROKE

In acute ischemic stroke (AIS), reduced regional cerebral blood flow (CBF) leads to a series of functional, biochemical and structural tissue changes and eventually to irreversible neuronal death if the tissue is not reperfused quickly enough. Affected brain tissue can be classified into several areas. The ischemic core represents tissue that is irreversibly damaged. The ischemic penumbra is the area that is functionally impaired but structurally intact and, as such, potentially salvageable. Another compartment is called *benign oligemia*, which represents mildly hypoperfused tissue that is not at risk of infarction under normal circumstances.^{3,4}

Two major mechanisms are responsible for ischemia in acute stroke: thromboembolism and hemodynamic failure. The former usually occurs as a result of embolism or in situ thrombosis, resulting in a large vessel occlusion (LVO) that leads to an abrupt disruption in CBF and which is the cause of nearly half of all acute ischemic strokes.⁵ Hemodynamic failure usually occurs with a present arterial occlusion or stenosis, when collateral blood supply maintains CBF at levels that are sufficient for preservation of brain function under normal circumstances. Ischemia is then triggered by conditions that decrease perfusion proximally to the arterial lesion (systemic hypotension or low cardiac output) or increased metabolic demands (fever, acidosis).⁶ Subsequent stroke pathophysiology is extremely complex and involves numerous processes, including disruption of the blood-brain barrier (BBB), inflammation, necrosis or apoptosis.⁷

1.1.1 TREATMENT

The aim of AIS treatment is to reduce infarct core volume (dead tissue) and to salvage penumbra in the first 48 hours after stroke symptom onset by restoring blood flow to non-ischemic levels.^{3,8} Several studies have shown that final infarct volume plays a critical role in long term functional outcome.^{9,10} Intravenous thrombolysis with tissue plasminogen activator (IV-tPA) became the first evidence-based short-term treatment for improving outcomes after AIS over 20 years ago.¹¹ Since then, multiple randomized-controlled trials showed confirming evidence that IV-tPA is effective within 4.5 hours of stroke onset.¹²⁻¹⁴

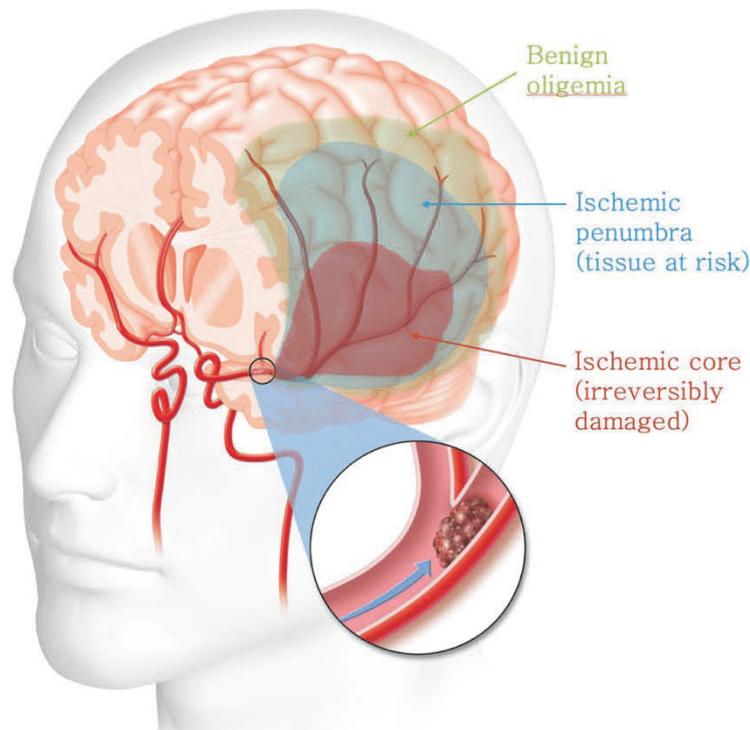


Figure 1.1: Concept of ischemic core and penumbra presented after an occlusion of the middle cerebral artery (MCA).
Image adapted from a neuro course presentation by M. Lev.¹⁵

The most serious complication from IV-tPA is ICH. It often occurs in the (necrotic) ischemic core and is caused by reperfusion injury. Although most reperfusion hemorrhages are asymptomatic, some can provoke neurologic decline and can be fatal. Seet et al. showed that the risk of symptomatic ICH is 5% to 6%.¹⁶ Others showed that there is a positive correlation between symptomatic ICH risk and baseline infarct volume,^{17,18} indicating that it mainly occurs in patients who already had an unfavorable prognosis. The strict therapeutic time window of 4.5 hours limits the efficacy of IV-tPA due to the fact that only a moderate rate of reperfusion by clot dissolving could be achieved in patients with an LVO. Moreover, the increased risk for ICH when using anticoagulants further decreases the rate of eligible patients for this treatment to less than 10% of all ischemic stroke patients reaching the emergency department.¹⁹⁻²²

Although up until 2013 IV-tPA has been shown the only proven treatment to be effective in AIS,²³ intra-arterial thrombolysis (IAT) was tested in the meanwhile in the Prolyse in Acute Cerebral Thromboembolism II (PROACT II) study, which found potential safety and efficacy of IAT for middle cerebral artery (MCA) occlusions within 6 hours of stroke symptom onset.²⁴ Later, imaging studies found that a major predictor for worse IAT outcome was large baseline infarct volume (i.e. > 70-100 mL).²⁵⁻²⁸

After exploring IAT, the Interventional Management of Stroke (IMS) trial investigated the feasibility and safety of combined IV and intra-arterial therapy in AIS.²⁹ In subsequent years, injecting thrombolytic agents or saline into the thrombus evolved to endovascular thrombectomy (EVT). The first trials using EVT in combination with IV-tPA, such as SYNTHESIS and MR RESCUE, failed to demonstrate a beneficial clinical effect. This was mainly caused by early-generation thrombectomy devices, a long period between onset of stroke symptoms and groin puncture and a lack of adequate vessel imaging to confirm patients' eligibility for these therapies.²⁹⁻³¹ In 2015, the MR CLEAN trial in the Netherlands was the first that has been able to demonstrate a beneficial clinical effect of EVT in selected patients by wielding more strict inclusion criteria, using third-generation mechanical thrombectomy devices and ensuring imaging confirmed occlusion of the anterior circulation.³² These results shifted focus of triage in AIS to selecting the specific subset of patients who would have most clinical benefit from EVT combined with IV-tPA. Trials such as ESCAPE, EXTEND-IA and SWIFT PRIME showed that patients selected based on assessment of the collateral circulation on computed tomography angiography (CTA) or estimated size of the ischemic core and penumbra from computed tomography perfusion (CTP), had the benefit of an even larger clinical effect.³³⁻³⁵

Although EVT is more effective in the early time period after stroke onset,³⁶ two recent trials (DAWN and DEFUSE III) have shown that even after 6 hours of stroke onset (up to 24 hours in the DAWN trial), patients with a significant clinical penumbra and small ischemic core, still benefit from EVT in addition to IV-tPA if selected correctly.^{37,38} These findings further highlight the need for individual imaging-based patient selection for successful and safe endovascular stroke therapy.

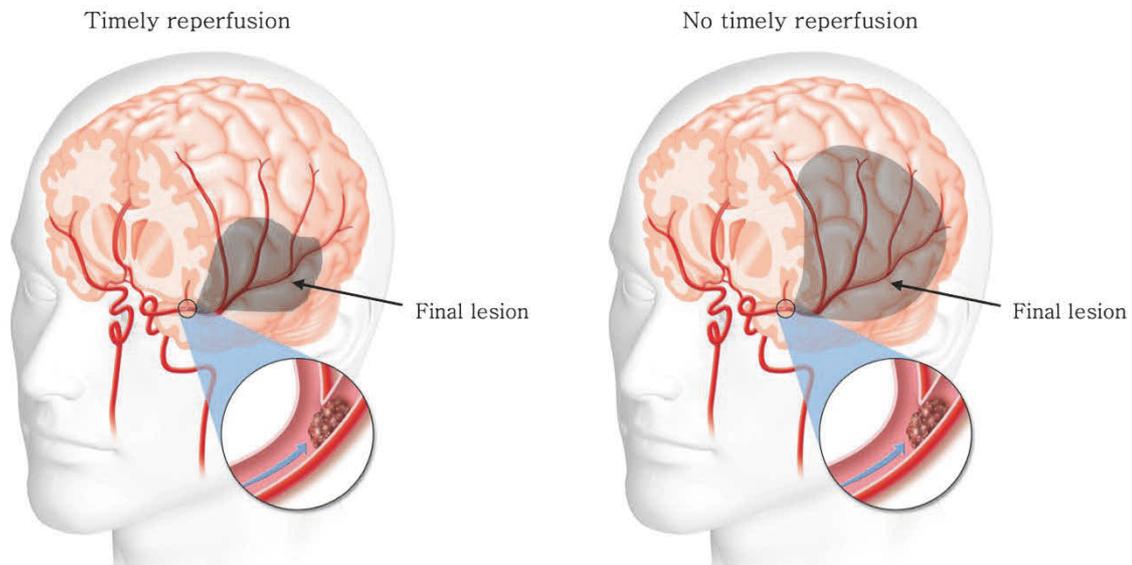


Figure 1.2: Example of final lesion volume depending on reperfusion time. In case of timely reperfusion, the lesion is limited to the initial core, and in case of no timely reperfusion, the penumbra is also included. Images are adapted from a neuro course presentation by M. Lev.¹⁵

1.1.2 IMAGING IN AIS

Patients with a treatable occlusion confirmed on imaging depend on infarct size to determine their eligibility for EVT. Imaging should quickly, accurately and reliably define infarct core and ideally penumbra. The triage in AIS patients is based on three major imaging features: (1) exclusion of ICH and stroke like mimics, (2) detection of the site of arterial occlusion and (3) the determination of the extent of the parenchymal lesion.

For exclusion of hemorrhage or stroke mimics (clinical symptoms that are difficult to distinguish from real strokes), conventional non-contrast computed tomography (CT) is the standard of care. It allows fast and accurate detection of ICH, with near perfect sensitivities and specificities.^{39,40} To assess whether the site of occlusion is accessible for EVT, CTA is performed, which is considered as level 1 evidence for the rapid assessment of a LVO.^{23,41,42} Finally, the lesion extent should be evaluated. However, there is a lot of controversy about which method should be used to assess the infarct.

The Alberta Stroke Program Early CT Score (ASPECTS) is a 10-point scoring system for qualitative assessment of early ischemic changes in different brain areas that are supplied by the MCA.⁴³ It is the most readily available method since a non-contrast CT (NCCT) is the standard first step in AIS patient triage. Since ASPECTS only roughly assesses brain areas, it does not allow for accurate baseline infarct volume estimation, especially in the acute phase when ischemic changes are hardly visible.⁴⁴

CTP utilizes dynamic CTA data consisting of multiple repeated head CT scans during IV administration of iodinated contrast material. Some studies claim that this technique can identify core and penumbra and that it is useful for identifying eligible patients with major anterior circulation occlusions (ACOs) for interventional therapy.⁴⁵ Especially the parameter 'mean transit time' (MTT) of the iodinated contrast was found to be most predictive of at-risk tissue.⁴⁶ Several trials showed a benefit in outcome in CTP based selected patients for EVT,^{34,35,47} while others have found conflicting results regarding the clinical usability of this imaging modality.⁴⁸⁻⁵⁰ The found beneficial effects could be caused by the exclusion of patients with less favorable perfusion profiles. Differences in CT scanners and post-processing algorithms also result in varying performance of CTP.⁵¹⁻⁵³ Another major disadvantage of CTP is its high burden of radiation exposure. In addition, CT perfusion has limited brain coverage, is highly susceptible

to z-axis motion and it lacks clear guidelines for indication, acquisition methods and interpretation.⁵⁴⁻⁵⁶ Software to automate output on extent of baseline infarct and mismatch between baseline infarct and salvageable penumbra could address these concerns.⁵⁷

Although many EVT trials relied on CTP in screening patients, magnetic resonance (MR) diffusion-weighted imaging (DWI) is the near-perfect method for detection of infarct core in AIS.^{58,59} Although some studies state that DWI lesions incorporate both types of ischemia and therefore cannot be solely considered as ischemic core,⁶⁰⁻⁶² others claim that DWI is more precise than CTP in measuring infarct core extent.^{63,64} Apparent diffusion coefficient (ADC) maps can further help interpreting ischemic core from DWI images by providing a standardized unit of diffusion for every voxel⁶⁵, while not suffering from *T2 shine through* caused by other pathologies.⁶⁶ Leslie-Mazwi et al. showed that using DWI as a tool for selection EVT eligible AIS patients results in an 'optimal' tradeoff between screened and treated patients.⁶⁷ Nevertheless, DWI is used in only a few major stroke centers across the globe as standard of care. Although, work-flow related reasons and time delays are constraints to implementing MR in an acute setting, the costs and availability are in general the main reasons that DWI cannot be used at hospitals emergency departments.

1.1.3 IMAGING OF COLLATERAL CIRCULATION

Collateral vessels are tiny arterioles that connect the distal branches of the cerebral arteries and they supply the ischemic region beyond an occlusion by retrograde filling of distal arteries.⁶⁸ Collaterals play an important role in time from vascular occlusion to symptom onset.⁶⁹ Since patients with a better collateral circulation have smaller infarct core volumes with probably relatively large penumbra, a different approach to estimate infarct size is to evaluate the quality of the collaterals.^{70,71} The presence of collaterals in AIS patients has been studied since the early 2000s on interventional digital subtraction angiography and have shown to be an independent predictor of good clinical outcome after stroke therapy.⁷²⁻⁷⁴ Later, collateral assessment with CTA using maximum intensity projections (MIP) confirmed these early results.⁷⁵ Moreover, CTA collateral scoring showed a strong correlation with DWI infarct size⁷⁶, especially when combined with NCCT ASPECTS.⁷⁷

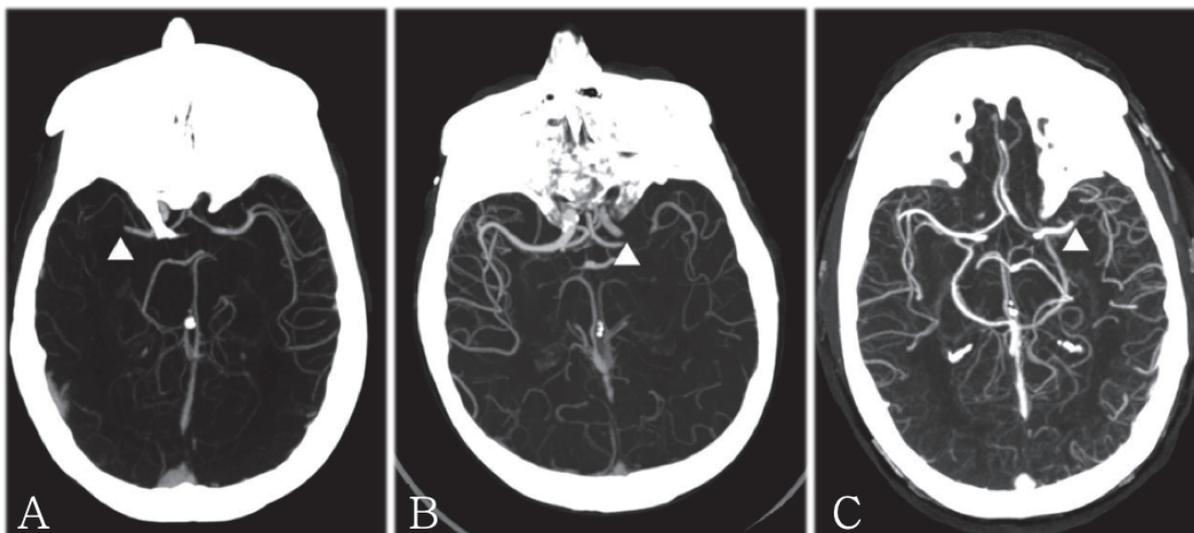


Figure 1.3: Collateral scoring: examples of axial MIPs from CTAs with (A) poor, (B) intermediate, and (C) good collaterals. Arrowheads denote the site of occlusion.⁷⁸

To avoid mislabeling of collateral status using single-phase CTA, multiphase CTA is used to better capture collaterals in a more delayed phase of contrast enhancement, since it takes longer for collaterals to fill up.⁷⁹ Absence of collaterals at delayed phase CTA is therefore a predictor for large DWI infarct lesion and poor clinical outcome.^{80,81} Both the ESCAPE and MR CLEAN trials found that patients with poor collaterals on baseline CTA receive no additional benefit from EVT.^{32,33} Others state that multiphase CTA is at least equal or even better than CTP in predicting tissue fate and to use as selection tool for EVT in AIS.^{79,82} Even more advanced methods to capture the collateral circulation is by using raw CTP data and reconstruct these to a dynamic CTA or to a timing invariant CTA.⁸³⁻⁸⁵ Because of all these

different image acquisition methods, multiple collateral scoring systems exist.⁸⁶ Automated collateral scoring using algorithms have the potential to replace these qualitatively assessed scoring systems.⁸⁷

1.1.4 THE MASSACHUSETTS GENERAL HOSPITAL ACUTE STROKE IMAGING ALGORITHM

Clinical trials have shown that estimation of (the ratio/mismatch between) ischemic core and penumbra are accurate tools in selecting AIS patients that are eligible for EVT.^{34,35} The DAWN and DEFUSE trials showed that even up to 24 hours after stroke symptom onset, patients can still benefit from EVT when selected properly based on small ischemic core (and significant clinical penumbra).^{37,38} Therefore, ischemic core (and penumbra) volume should be considered in final clinical decision making for EVT. The other two important imaging features are: (1) exclusion of ICH and stroke like mimics and (2) detection of the site of arterial occlusion.

The Massachusetts General Hospital (MGH) acute stroke imaging algorithm is designed to address all of these important features. It distinguishes between patients that arrive at the emergency department within and after 6 hours of stroke symptom onset. Within 6 hours, NCCT is used to assess (1) above. Subsequently, early and delayed phase CTA is acquired to capture both the presence and location of vessel occlusion (early / arterial phase) and to visualize the collaterals (delayed phase). If a patient is eligible for MR imaging, DWI is performed, which is the most accurate imaging modality to estimate infarct core in the acute phase.^{63,64,88} In case of retrievable occlusion and a small ischemic core (<70 mL), the risk of hemorrhagic conversion (due treatment) is smaller than the potential benefit of EVT, and thus EVT is performed. For an ischemic core between 70 and 100 mL, it is uncertain if EVT will have benefit, and thus will this decision depend on clinical factors. Above 100 mL, EVT is not considered as beneficial because of the increased risk of hemorrhagic conversion.

After 6 hours, patients directly go to MR for DWI assessment to save time on workflow. Susceptibility-weighted imaging (SWI) is then used to assess (1) above and to localize thrombus. Also, for every patient the National Institute of Health Stroke Scale (NIHSS) is derived from physical examination before imaging. NIHSS is a primary clinical diagnostic tool to assess stroke severity and global localization. In this imaging algorithm, it also helps estimating the core/penumbra ratio, by combining functional impairment with precise localization of the infarct on DWI. Perfusion imaging is only performed if a patient is not eligible for MR imaging, or if perfusion data might be desirable to evaluate the full clinical picture of a patient.⁵⁴

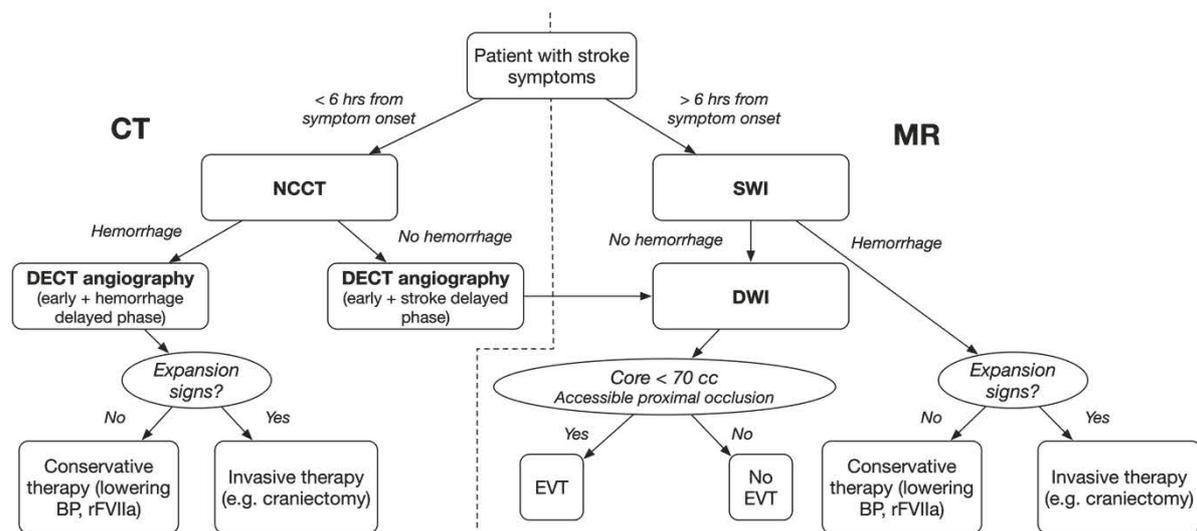


Figure 1.4: The MGH acute stroke imaging algorithm. BP = blood pressure, rFVIIa = recombinant activated factor VII. Illustration adapted from González et al.⁵⁴

Although DWI is most sensitive for imaging of the ischemic core in acute stroke, no 24/7 availability (costs) and delayed workflow are reasons for many hospitals not to use this modality for emergency cases. This raises the need for individual patient selection for EVT using fewer and still reliable resources, such as validated CT modalities. Ideally, quantitative information on the ischemic lesion size that is captured by DWI should be derivable from information that is available in CT. Quantification of iodine in dual-energy CT (DECT) angiography could serve a first step in this approach.

1.1.5 DUAL-ENERGY CT

In 2017, MGH replaced conventional CTA with DECT in its AIS imaging protocol as standard of care. DECT utilizes scanning at two different energies to decompose tissue into three different materials based on differences in attribution of photoelectric and Compton scattering to X-ray attenuation.^{89,90} Using three-material decomposition, a DECT source volume can be decomposed into virtual non-contrast (VNC) images and iodine-only images. VNC can reliably capture any hemorrhage, while early and delayed-phase iodine images contain valuable information about both thrombus location, collaterals and (local) tissue perfusion status.⁹¹ Since collateral function has previously been shown to be related to ischemic lesion size,^{70,71} and perfusion data has been used to predict infarct volumes,⁹² DECT images theoretically hold all information required to assess the three important imaging features for AIS, ischemic lesion size potentially included.

Another application where DECT is useful is in case of ICH. Here, quantification of iodine holds valuable information on the likelihood of hematoma expansion, as described in section 1.2. More background information on DECT post-processing is presented in Appendix A.

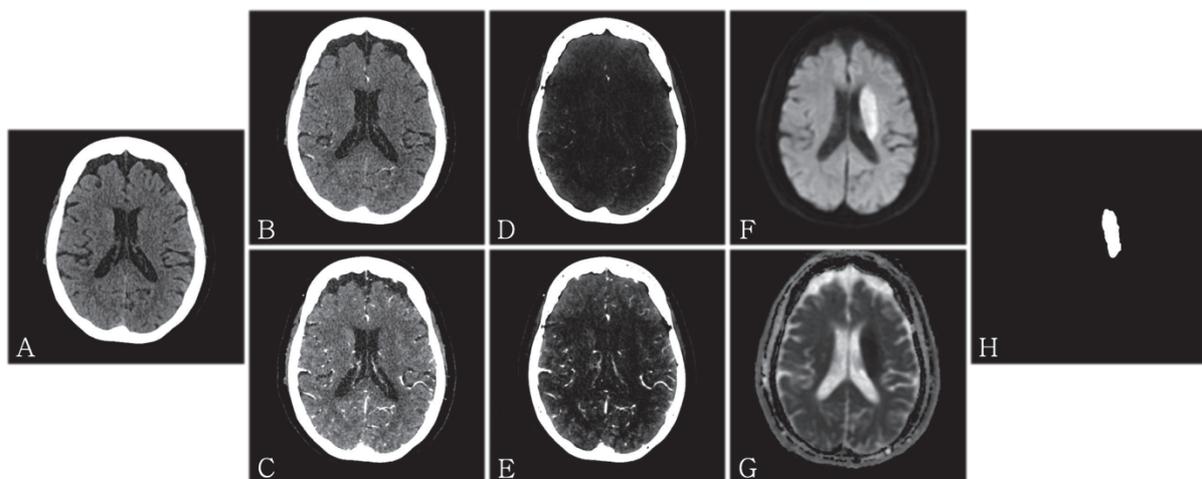


Figure 1.5: Data present after imaging with MGH acute ischemic stroke protocol. Left M1 occlusion. MR 39 mins after CT. Delay time = 15 sec. A) NCCT; B) Simulated single-energy CTA (120 kV); C) Simulated single-energy delayed CTA (120 kV); D) DECT iodine image; E) DECT delayed iodine image; F) DWI; G) ADC; H) Ground truth lesion segmentation

1.2 INTRACEREBRAL HEMORRHAGE

Only 10% of all strokes are caused by ICH, which is an acute and spontaneous bleeding into the brain parenchyma. The overall incidence of ICH counts 24.6 cases per 100,000 people per year, with approximately 63,000 cases occurring in the United States annually.^{93,94} Most occurrences of ICH (75%–85%) are classified as primary ICH and include mainly spontaneous rupture of vessels, while the remaining occurrences are classified as secondary and include cerebral hemorrhage driven by other causes such as trauma.⁹⁵ ICH is commonly caused by hypertension or cerebral amyloid angiopathy.^{96,97} Artery rupture comprises the primary ICH phase, while a subsequent inflammatory response triggered by the released blood in the parenchyma is considered as the secondary phase. The inflammatory response leads to brain tissue damage, blood clot degradation, edema and BBB disruption, all driving hematoma growth.⁹⁸ Complications including hydrocephalus, severe tissue shifts, increased intracranial pressure (ICP), and mass effect may occur, resulting in catastrophic tissue damage and neurologic deterioration.⁹⁹ Hematoma expansion by 33% has been noted within the first 3 hours in ~38% of patients presenting with ICH.¹⁰⁰ Each millimeter hematoma expansion has been associated with as 7% increased risk of long-term dependence.¹⁰¹

1.2.1 TREATMENT

A critical component in treatment of ICH is achieving hemostasis to avoid rapidly increasing ICP, and thus neurologic deterioration fatality. Only the procoagulant recombinant activated factor VII (rFVIIa) has been shown in a randomized, double-blind, placebo-controlled trial to reduce mortality (by 38%) and it was associated with an improvement in functional outcome at 90 days,¹⁰² although the subsequent FAST trial was only semi-successful, because of the risk of arterial thromboembolic events, among others.¹⁰³

Lowering blood pressure is another approach to decrease the amount of blood leaking into the brain parenchyma. However, two trials (INTERACT II and ATACH II) failed to demonstrate improved clinical outcomes as a result of this treatment.^{104,105} Also early surgical treatment (decompressive craniotomy) versus conservative treatment was studied in the STICH trial without success for the intervention group,¹⁰⁶ however, later studies showed that craniectomy may reduce mortality.^{107,108} The clinical effect of minimally invasive evacuation of the hematoma is currently still being researched in the MISTIE III trial.¹⁰⁹

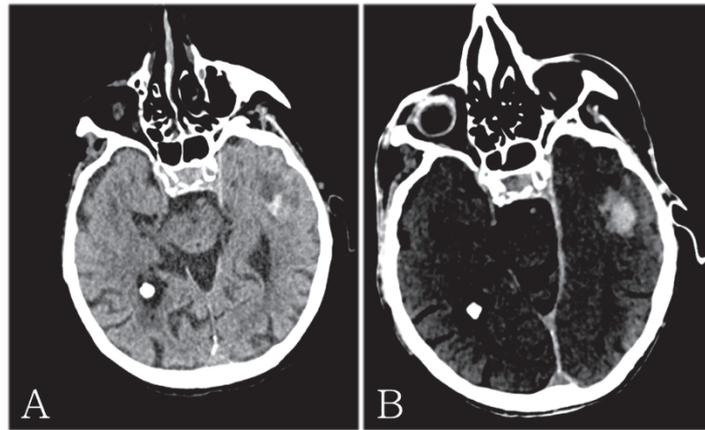


Figure 1.6: Early hematoma growth after ICH. A) Initial CT scan; B) Follow-up CT scan after 6 hours showing 4 cc of hematoma expansion.

1.2.2 IMAGING IN ICH

Imaging should be expected to effectively confirm the diagnosis of ICH, for which NCCT is considered as the gold standard. Next, hematomas should be assessed for likelihood of expansion, since hematoma expansion in the early hours post-ictus is strongly associated with mortality.¹⁰¹ The use of CTA has also been introduced as a potentially viable way to characterize risk of hematoma expansion in acute presentations and especially ‘spot sign’ has been proposed as radiographic marker for this phenomena.^{110,111} The spot sign is defined as foci of enhancement within a hematoma, thought to represent areas of contrast extravasation and active bleeding.⁹⁰ Although the spot sign is highly specific (80% - 93%) for hematoma expansion, it has relatively low sensitivity (50%-60%).¹¹²⁻¹¹⁵ In fact, results from the large PREDICT trial demonstrate a sensitivity only slightly greater than 50% when the spot sign is assessed on arterial phase imaging.¹¹⁶ Timing of angiographic imaging plays an important role in this.^{117,118} One of the difficulties in spot sign reading is differentiating hyperdense hemorrhage from contrast staining of the brain parenchyma due to spotty or diffuse contrast extravasation from leaky blood vessels or secondary vessel disruption, contributing to further hematoma expansion.¹¹⁹⁻¹²² Although quantitative analysis on the brightness of spot-sign on conventional single-energy CT (SECT) angiography to define a spot-sign score was not successful,^{111,123} recent work showed that quantitative iodine features derived from DECT can be combined to a so called ‘I2-score’ to improve prediction of hematoma expansion.¹²⁴ The I2-score uses the total iodine content in the hematoma and total iodine content in the brightest spot as quantitative predictors to estimate the likelihood of expansion. However, despite extensive research, still 42% to 50% of patients with ICH die within the first 30 days of their events.^{125,126} Thus, early identification of patients at risk for hematoma expansion remains a topic of interest to select candidates for early targeted medical or surgical intervention.¹¹⁴

2

Technical background

Medical imaging plays an important role in clinical decision making regarding diagnosis and therapy. With technological advancements and growth of imaging volumes, radiologists need to read more images and interpret them faster, increasing their workload.¹²⁷ On one hand, this gives rise to the desire for automated reading and interpretation of images using computer-aided diagnosis software, to help radiologists in facing their potential fatigue and the tremendous variations in pathologies they have to assess.¹²⁸ On the other hand, it offers the possibility to augment the knowledge of the radiologist with information from many scans together, hidden for the naked eye and human interpretation capabilities.¹²⁹ In fact, “*Images are more than pictures, they are data*”.¹³⁰ The hidden ‘features’ in this data can be extracted by engineered image analysis algorithms, or learned by advanced learning algorithms. Both hematoma expansion prediction and AIS detection on CT may benefit from such augmented radiology, and so the patient will do in the end.

2.1 QUANTITATIVE IMAGE TEXTURE ANALYSIS

Texture is a form of appearance determined by the local structural composition of an object. For an image, this means that a group of pixels (or voxels) can have different appearances depending on their pixel values and spatial arrangement. When first- or higher order statistics are applied to such an image region, certain texture features can be computed that quantify properties of this appearance. These features are *agnostic* compared to the *semantic* features that radiologists use to describe lesions.¹³⁰ Nevertheless, they can hold valuable information. In oncology, for example, they are increasingly used in clinical decision support to quantify tumor heterogeneity, and thus classifying them as benign or malignant.¹³¹ Recently, texture features have also been proposed to predict hematoma expansion, claiming good results.¹³²

Hematoma appearance can provide radiologists with information about the bleeding pattern and the cause of ICH, giving them a grasp on the likelihood whether a hematoma might expand or not. Many radiographic signs of morphologic hematoma appearance have been proposed to predict hematoma expansion, none of them being able to with both a high sensitivity and specificity. The variance of interpretation between radiologists makes it difficult to standardize assessment of hematomas. Quantitative texture analysis may have the potential to replace these ‘qualitative’ semantic signs in a more abstract form by computing mathematically derived numbers. However, in order to be able to compare these numbers, it is desired to standardize both image acquisition and region of interest (ROI) selection methods.

There are different methods to perform texture analysis.¹³³ First order statistics do not address spatial relationships between pixels, but solely rely on properties of the image histogram, such as mean pixel intensity, variance, skewness and kurtosis of the histogram. Second order statistics include spatial

relationships in different ways, depending on the used method. Texture can be calculated along one direction, using so called *run length matrices* or in multiple directions, using *gray level co-occurrence matrices* (GLCM). The latter method counts how often adjacent pixel or voxel pairs having a specific combination of gray levels in a specific spatial arrangement occur in a certain region (Figure 2.1), allowing features such as entropy, energy, homogeneity, contrast and dissimilarity to be computed. First, for every specified direction and combination of adjacent pixels, a separate GLCM is computed. After having normalized the sum of the separate GLCMs, different texture features can be computed, such as energy, entropy and dissimilarity (Appendix B).

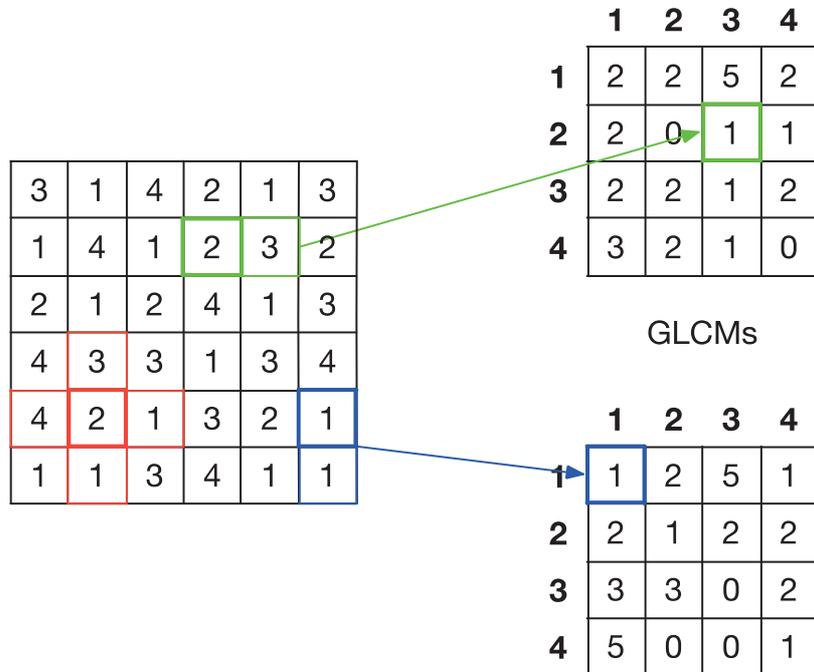


Figure 2.1: Computation of gray level co-occurrence matrices (GLCM) from a 6 × 6 image grid containing 4 gray levels. Red highlighted areas denote the considered spatial orientations of the pixel pairs for a 2D GLCM (4-neighborhood), resulting in 4 individual GLCMs (only 2 presented). For 3D GLCM analysis, this is extended to a 6-neighborhood.

Pre-processing images using a *Laplacian of a Gaussian* (LoG) filter before performing texture analysis, allow both fine and coarse texture features to be computed. The Laplacian component is sensitive to rapid spatial changes in gray levels, enhancing texture, while the Gaussian component transforms images to different scale spaces, depending on the width of the filter. Moreover, it smoothens undesirable CT quantum noise, and makes the texture analysis less dependent of image reconstruction parameters.¹³¹

Computing different texture features on different scales, rapidly results in a huge set of features that can be used as potential predictors, in our case, to predict hematoma expansion. Since many of these features will be correlated with each other, feature selection is an important aspect in the process of composing a predictive model. In case of a binary classification problem, the number of used features is ideally far less than the number of positive cases to avoid overfitting.¹³⁴ Stepwise selection using regression models are commonly used methods to reduce these high dimensional datasets to a lower dimensional feature space.¹³⁵

2.2 DEEP LEARNING

Predictive models can also be *learned* from a dataset. This is the field of supervised machine learning. Advanced learning algorithms (such as support-vector machines, k-nearest neighbors, and random forests) are trained on big datasets containing many data samples with their corresponding (often multi-dimensional) feature space, trying to find an optimal decision boundary that is able to classify each sample in its ground truth category (e.g. expanding vs. non-expanding hematomas). It is the aim to learn from the training data in such a way, that generalizable predictions can be made to new, unseen data samples.

Deep learning goes one step further: instead of learning from a readily available set of features, the feature extraction step from the raw data is incorporated in the learning process. Features no longer need to be human designed, such as the texture features discussed in section 2.1. This allows to extract even more abstract features than humans are able to design. Deep learning is able to do this by processing structured, large, raw datasets, in a layered architecture, together forming an artificial neural network, inspired by the structure of neurons in the human brain. Each layer consists of multiple densely connected individual neurons, that are able to process input data points by multiplying it with a *learned* weight, and by adding a bias, followed by a non-linear activation operation. When sequentially passing processed data to deeper layers, this results in a hierarchical representation of learned features with increasing levels of abstractness.¹³⁶ Supervising this process with every data sample's ground truth ensures that the learned features can be used for tailored tasks (e.g. classification, regression). The learning step itself takes place by iteratively comparing the network's predicted output with the ground truth using a loss function, and backpropagating the computed error through all network parameters to adjust the parameters in a direction that the loss function is minimized. Difficulties in minimizing this loss function are uniqueness, existence and robustness of it's minimum.

2.2.1 CONVOLUTIONAL NEURAL NETWORKS

When spatial information in raw data plays an important role, such as pixel values in medical image data, artificial neurons should not process all image data points individually as one long vector. This makes them sensitive to subtle translations or rotations in the image input data. Instead, neighboring pixel information should be incorporated, preferably on different scales. Spatially grouping artificial neurons sharing their *learned* weights and bias terms, let them form sliding convolutional filters. These layers of convolutional filters process the image local contextual information, and form the basis of what is called a convolutional neural network (CNN). Combining these convolutional operations with subsequent non-linear activation operations, allow the filtered input to be transformed into more abstract feature maps. Next, subsampling operations, increase the receptive field of the convolutional filters in subsequent layers, also allowing features to be extracted from different scales of the original input image. The combination of these operations in a CNN, is inspired by the organization of neurons in the visual cortex.

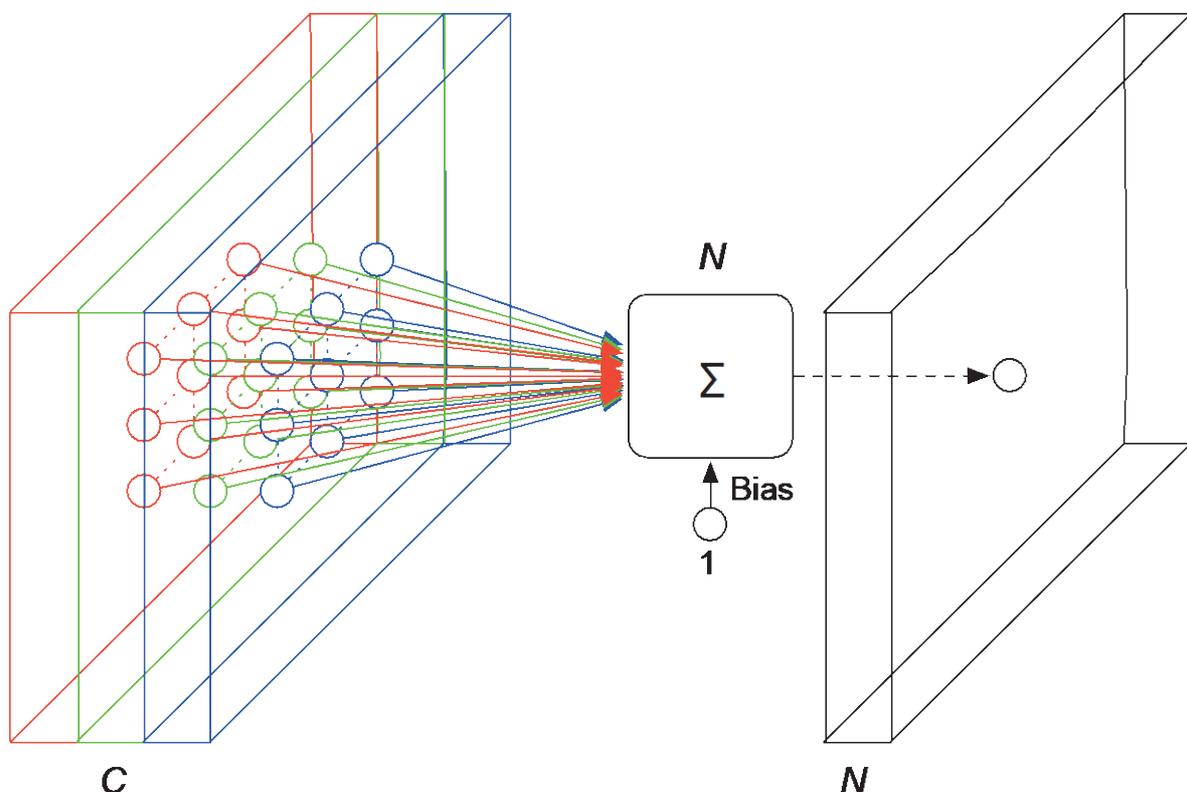


Figure 2.2: Convolution operation acting as sliding filter (size 3×3) on a C -channel input image/feature map ($C = 3$ in this example). Arrows denote learnable weights. Dashed arrow transfers result of convolutional operation to destination in output feature map. Note that this basic convolutional operation is performed in both the spatial and the channel domain. Number of filters is denoted by N , producing N output feature maps ($N = 1$ in this example).

State-of-the-art deep CNNs advanced the performance of computer vision systems, outperforming humans in many image recognition, object localization and structural segmentation tasks. But why do these networks need to be ‘deep’? The word deep refers to the number of hierarchical layers used in a neural network. It is well-known that a finite number of artificial neurons is able to approximate any continuous function. The more neurons are involved, the more complex functions can be mapped. However, with increasing number of parameters (neuron’s weights and biases), the computational costs to find this complex mapping increases, because all parameters need to be optimized individually.¹³⁷ Similar performance can be achieved by using fewer neurons, organized in layers, successively processing the input data. In this way, the amount of trainable parameters in a network can be reduced, while better being able to optimize all parameters and especially to *generalize* to new, unseen data.¹²⁸

In medical imaging, often highly complex, local features need to be learned from relatively few variations in training data. And it is mainly the number of variations in a dataset to be learned that matters for generalization, rather than the number of parameters in a network. That is why deep learning is *data hungry*, and often data augmentation techniques need to be used. It is both the number of data samples, variations and the amount of learnable parameters that need to be balanced for every specific application in order to avoid *overfitting* to training data. Ideally, big datasets, containing all of the variations that need to be learned, should be used to train deep, complex CNNs, in order to achieve a desirable performance and generalizability. To do so, computational power in the form of graphical processing units (GPUs) is necessary to efficiently compute the many required parallel matrix multiplication operations (convolutions) in reasonable amount of time.

2.2.2 CNN ARCHITECTURES

Both the structure of input data (n -dimensional matrices) and desired output determine the global architecture of a CNN. In case of image classification tasks, a CNN architecture, acting as feature extractor, is combined with a subsequent classifier. This is often a conventional artificial neural network, consisting of (a few layers of) densely connected neurons on top of the CNN, but in theory, any kind of classifier could be used. However, when a spatial output is desired, a CNN in general turns into a fully convolutional network (FCN), where the top is replaced by (de)convolutional layers and/or upsampling operations, to reconstruct the extracted features back to the original image space. In this way, a CNN is able to perform segmentation or localization tasks. A very popular architecture in biomedical image segmentation is the so called U-Net, named after its U-shaped architecture.¹³⁸

While a CNN’s global architecture mainly determines the format how input data is mapped to desired output format, local architecture plays an important role in optimizing performance. Designing a CNN’s architecture can be very complicated given the endless possibilities of combining and connecting convolutional, subsampling, activation and normalization layers. Several concepts have been worked out that improve CNN performances on big image recognition datasets, such as MNIST, CIFAR-10 and ImageNet.¹³⁹⁻¹⁴¹ These concepts can in fact be used as separate building blocks for CNNs in general, and are mainly focused on effectively optimizing extremely deep networks (> 30 convolutional layers). One of the most famous and commonly used concepts is the so called ‘residual’ connection, which is a shortcut connection between a few stacked layers using identity mappings.¹⁴² This makes it easier to optimize parameters in deep networks, enabling to reach higher performances. Another likewise concept is that of densely connected convolutional layers.¹⁴³ Other approaches to increase performance focus on reducing the number of parameters and computational complexity, like Google’s Inception module, and its extreme Xception version.¹⁴⁴⁻¹⁴⁶ Also combinations of different concepts have been described, such as ResNeXt, which is a combination of the residual and Xception concepts.¹⁴⁷ Other, more general concepts are those of dropout, randomly dropping out neurons while training a CNN (acting as regularization), and batch normalization, accelerating training of CNNs (Appendix D).^{148,149}

2.2.3 TRAINING CNNs

As mentioned earlier, the advantage of deep learning and CNNs is that features no longer need to be designed by humans, but are learned by the computer itself, to be used in making predictions. This is done by using gradient descent algorithms to optimize the learnable network parameters. When comparing the ground truth (y) with the predicted output of the network (y_p), an error can be computed using a loss function g . The learnable set of network parameters (\mathbf{W}) are adjusted in the opposite direction to the gradient of this loss function (∇g). The size of the adjustment steps to be taken is determined by the

learning rate (η). A commonly used algorithm for optimizing these weights is called *stochastic gradient descent* (SGD), in which the loss function is not computed over all data samples in the training set (epoch), but after a prespecified number of samples, together forming a *batch*. Updating the weights takes place iteratively after every batch, using the following equation:

$$\mathbf{W} = \mathbf{W} - \eta \nabla g(\mathbf{W}) \quad (2.1)$$

in which $g(\mathbf{W})$ is the loss over a batch as function of the current set of weights. The gradient of the loss function is then backpropagated through all network layers to update the weight parameters.

Training schemes

Data samples used to compose a batch for training CNNs can take different forms. The most straightforward method is to consider data samples as whole (multi-channel) 2D images or 3D image volumes in case of a network architecture utilizing 2D and 3D convolutional operations respectively (This is what is called a *dense* training scheme. The network is then able to use the complete contextual information of an image or volume to make its predictions. An alternative strategy is to use small patches of the original data samples, only providing the network with local contextual information (*patch*-based training scheme). Although multiple patches are then required to predict the whole original data sample, it may have the advantage that the network is forced to learn local instead of global image features to make predictions, potentially avoiding overfitting on non-relevant global features. Architectures such as DeepMedic have been developed that try to implement the benefits of both dense and patch-based training schemes by converting input patches to multiple image scales and by predicting patches having a smaller field of view (FOV) than the used input patches, effectively using more (local) contextual information than covered by the predicted output patch.¹⁵⁰

3

Research objective

In both ischemic and hemorrhagic stroke, diagnostics play an essential role in adequate therapy selection and thus potential outcome for patients. Serious adverse events could occur if patients are not selected properly.

3.1 ACUTE ISCHEMIC STROKE

For ischemic stroke, this means that the risk of hemorrhagic conversion is considered as larger than the potential benefit of EVT in patients having a large (>100 cc) ischemic core, and thus, these patients should be treated conservatively. Although DWI is able to identify patients (ischemic core < 70 mL) who may benefit from EVT, this modality is generally not available in the acute setting at the emergency department. In major stroke centers, CTP is often used as a second best option, having limited capabilities of identifying infarct core, and having a high radiation burden for patients. Therefore there is a need for a readily available imaging modality using less radiation, that is able to identify patients eligible for EVT by detecting and estimating stroke lesion core volumes. With MGH having one of the largest bodies of radiological images in the world, and having access to MR and DECT imaging at the emergency department, there is potential data available to train a deep learning system that is able to detect infarct core lesions from the initially acquired CT imaging data. In previous work, a processing pipeline including a deep learning algorithm based on U-Net has been developed to process NCCT, early- and delayed phase (3-phase CT) DECT angiography images with the aim to detect the ischemic core lesion as visible on DWI. Although the algorithm reached a promising performance on the training dataset in terms of Dice similarity coefficient (DSC), the DSC on the independent test set was practically 0. It suffered severely from generalization problems due to overfitting.¹⁵¹ Having more training data available, and improving the deep learning architecture may potentially close this generalization gap and thus improve performance on the independent test dataset.

Research question AIS:

To what extent is a deep learning system able to make clinically useful DWI infarct core lesion predictions based on 3-phase CT data on an individual patient basis to be used for triaging AIS patients?

Sub-questions:

What is the combination of deep learning architecture and (processed) 3-phase CT data that results in the highest DSC on the independent test set?

- In terms of input 3-phase CT input data:
 - o SECT vs. DECT angiography;
 - o Augmented data vs. original data only

- Using a dense or patch-based training scheme;
- In terms of network architecture:
 - U-Net based vs. DeepMedic based network
 - 2D vs 3D convolutional operations
 - Separate vs. mixed modality processing

3.2 INTRACEREBRAL HEMORRHAGE

For ICH, the main decision in terms of therapy selection is to choose between targeted medical therapy (rFVIIa or lowering blood pressure) or to perform decompressive craniectomy, all with their own level of invasiveness and risk profile. Because hematoma expansion is associated with increased neurological deterioration, poor outcome and mortality, it is desired to prevent expansion or to tackle the consequences of it, by selecting a treatment strategy that fits with the patient's (clinical) baseline situation and hematoma expansion likelihood. This clinical situation is readily known at the time of imaging, while the hematoma expansion likelihood is difficult to assess. Many radiographic imaging markers have been proposed to predict hematoma expansion, such as CTA spot sign, all having good specificity, but only moderate sensitivity. Recent work showed that quantitatively assessing spot signs using DECT iodine features, combined to an I2-score model, can increase sensitivity in predicting hematoma expansion. Alternative strategies rely on extracting quantitative hematoma texture features derived from NCCT. Since texture features seem not to be directly correlated to DECT iodine features, combining texture and iodine features, potentially brings in new information (synergy) to predict hematoma expansion.

Research question ICH:

Are NCCT texture features combined with DECT iodine features able to outperform the I2-score model in predicting hematoma expansion on an independent test set, without decreasing specificity?

Sub-questions:

- How do texture features rank when combined with a data set containing DECT iodine features during a feature selection process?
- What is the sensitivity to noise for both the I2-score model and the combined iodine/texture model on the final model performance?

The next two chapters of this thesis are focused on answering above mentioned research questions. Chapter 4 starts with the description of a pipeline that processes 3-phase CT and corresponding DWI data to be used for a deep learning system. The design of both U-Net and DeepMedic based architectures are highlighted and the experimental setup to find the best combination of (processed) 3-phase CT data and network architecture is presented. Finally, the results of the top performing configurations on the independent test set are analyzed to see if the system is able to make clinically useful predictions for triaging AIS patients.

In chapter 5, a short recap is given on the recently proposed I2-score based on iodine features from DECT scans to predict hematoma expansion. An extension of this processing pipeline incorporating texture analysis on NCCT is then described, followed by the design of an extensive feature selection and ranking method. A prediction model is constructed using a combination of top-ranked iodine and texture features, and is compared to the performance of the I2-score model in a noise sensitivity analysis on the independent test set.

Finally, chapter 6 concludes this thesis by providing an answer on above mentioned research questions for both AIS and ICH clinical problems.

4

Predicting acute ischemic stroke DWI lesions from 3-phase CT data using deep learning

4.1 INTRODUCTION

Imaging plays an important role in therapy selection for patients with AIS to maximize clinical benefits and minimize adverse events, such as hemorrhagic conversion. Recent large-scale clinical trials have shown the benefit of EVT in combination with IV-tPA for patients selected on infarct size or quality of the collateral circulation, even up to 24 hours after stroke symptom onset.^{32-35,37,38} Since DWI has a near-perfect sensitivity to detect AIS lesions in the acute phase, it is considered as the gold standard to assess ischemic core volume.^{58,59} However, DWI is not widely available, and is in general constrained due to workflow issues. Alternatively, CTP is used by many major stroke centers, suffering from difficulties in standardized core volume assessment and a high radiation dose for the patient.⁵⁰ This emerges the need to assess the ischemic core using readily available imaging modalities and less radiation.

Other than DWI and CTP, also NCCT and (multi-phase) CTA have been studied extensively to derive predictive features for infarct core volume, and thus eligibility for EVT. These features are mainly ASPECTS, collateral scoring, or a combination of both.^{43,70-72,76,77,79} As an alternative to conventional single-energy CTA, DECT angiography is becoming more available as a diagnostic tool in the emergency department. It has the possibility to generate iodine-only images, that potentially contain information on local tissue perfusion status, and thus to detect core infarct.⁹¹

The introduction of deep learning into the medical image analysis domain opened up new possibilities combine abstract information and insights from many scans and different modalities together to predict tailored information, such as infarcted brain tissue. Recently, an initial image processing pipeline including a deep convolutional neural network (CNN) based on the U-Net architecture has been developed to detect DWI lesions from 3-phase CT data (NCCT, early- and delayed phase DECT iodine-only images). Although the predicted DWI lesions on the training set reached a Dice similarity coefficient (DSC) of approximately 0.6, indicating good performance for clinical usability, the DSC on the independent test set was practically 0 due to severe generalization problems.¹⁵¹ The aim of this study is to address and overcome this issue by adjusting the initial CNN architecture, collecting and presenting more varied CT data to the CNN and find the best working combination between CNN and 3-phase input CT data that results in the highest DSC on the independent test set.

4.2 METHODS

4.2.1 SUBJECT SELECTION AND DATA COLLECTION

This retrospective study was approved by the local institutional review board. Required informed consent was waived. All patients who were referred for head CT imaging and were suspected for stroke, transient ischemic attack (TIA) or cerebral ischemia between January 2015 and August 2018 were evaluated on the availability of NCCT, early and delayed phase CTA (SECT or DECT), DWI, ADC and T2-weighted imaging. Patients were included only when scans from all modalities were available. In case of partially missing data (e.g. a scan not covering the complete brain), and when there was more than a 24 hour time difference between CT and MR acquisition, patients were excluded.

Scans were acquired with scanners from multiple vendors. All CT images were acquired at thin slices (0.6 mm) and were reconstructed using a standard filtered backprojection reconstruction kernel. For the NCCT and conventional single-energy CTA acquisitions, the peak beam energy was 120 kV and 100 kV respectively. All DECT images were acquired using a Somatom Definition Force scanner (Siemens Healthcare, Forchheim, Germany), with tube A operating at 100 kVp and tube B at 150 kVp (including Sn-filter). The delay time between early and delayed acquisitions was strictly 15 s in this protocol. Simulated single-energy early and delayed phase images were reconstructed by taking a weighted average (factor 0.6) between the low and high energy acquisitions. MR images were acquired with vendor specific imaging protocols of the desired sequences (DWI/ADC and T2/T2 FLAIR).

4.2.2 DATA PRE-PROCESSING

All DICOM (Digital Imaging and Communications in Medicine) files were de-identified and all three-dimensional (3D) volumes were converted to the NIfTI (Neuroimaging Informatics Technology Initiative) format. MATLAB R2018a (The Mathworks, Natick, Massachusetts, USA) was used as basic environment for all data pre-processing. An overview of the complete data pre-processing pipeline is depicted in Figure 4.1. An example of both pre-processed SECT, DECT, MR and corresponding lesion segmentation is presented in Figure 1.5.

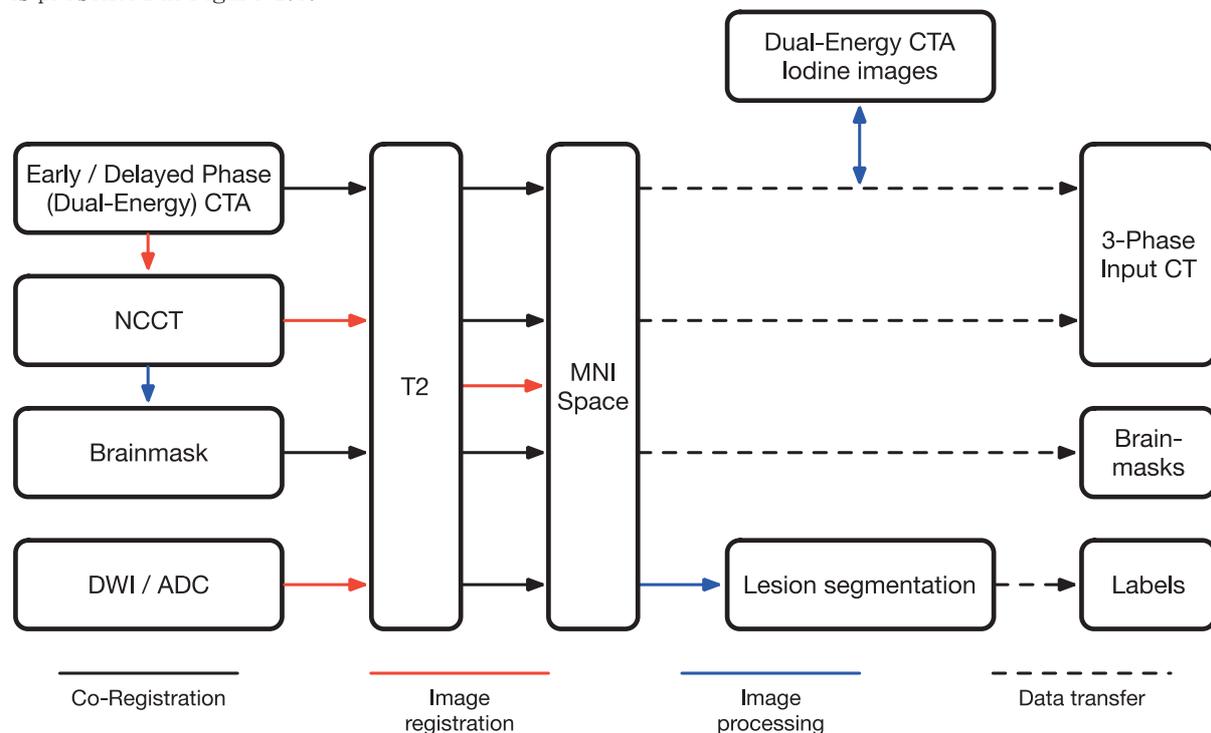


Figure 4.1: Overview of the data pre-processing pipeline.

Creating brain masks

On the NCCT, brain extraction was performed based using an algorithm of Muschelli et al.¹⁵². Images were first thresholded between 0–100 Hounsfield Units (HU) and then smoothed using a 3D Gaussian

filter kernel with a sigma of 1 mm^3 . The results were used as input for the FSL¹⁵³ (Nuffield Department of Clinical Neurosciences, University of Oxford, UK) brain extraction tool, operated at a fractional intensity setting of 0.01 to create the brain masks. To make the result more robust, the resulting masks were post-processed by extracting the largest 3D connected component and subsequently performing a 3D hole-filling operation.

Image registration

Subsequently, all volumes were registered to a standard Montreal Neurological Institute (MNI) space using SPM12 (Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom) package for MATLAB. The order of operations was; (1) Registration of DWI/ADC to T2; (2) Registration of (DE)CT angiography to NCCT, and subsequently to T2; (3) Normalization of T2 to MNI space using a template, and co-transforming all other modalities. The gray value distributions of the 3D volumes to be registered were smoothed before all registration steps, and the origin was set to image volume the center of gravity. Registration steps were performed using an iterative affine registration algorithm based on Collignon et al.¹⁵⁴ to minimize the normalized mutual information criterion. The normalization step adopts a two-stage procedure, by first using the registration method mentioned before, followed by an iterative non-linear transformation (warping) step. Output volumes had a field of view (FOV) of $230 \times 230 \times 136 \text{ mm}$ centered at the anterior commissure of the used MNI template, resliced at voxel dimensions of $0.45 \times 0.45 \times 1 \text{ mm}$ using trilinear interpolation, resulting in an image grid of $512 \times 512 \times 136$ voxels.

DECT Iodine-only images

For all DECT acquisitions, the normalized 3D volumes were converted back to DICOM files and post-processed using syngo.via (Dual-Energy CT module, Head (CA) application) to compute iodine-only images. The used iodine ratio (dual-energy slope) was 2.12 and the reconstruction resolution (smoothing) was 4 mm. For a more comprehensive explanation of DECT post-processing, see Appendix A.

Stroke lesion segmentation

Ground truth labels were created by manually segmenting stroke lesions from DWI and ADC volumes in MNI space. First, a volume of interest (VOI) was selected from the DWI using a slice-by-slice ROI selection method. Second, a patient specific threshold was used that led to a visually good segmentation result. All segmentations were performed by the same observer. Finally, the segmentation result was improved by transferring the initial segmentation mask to the ADC volume, and adding all voxels with an ADC value $< 500 \times 10^{-6} \text{ mm}^2/\text{s}$ that were connected to the readily existing segmentation mask.

Input pre-processing

NCCT, early and delayed phase CTA were used as 3-phase input data. All volumes were resampled to $128 \times 128 \times 64$ voxels, having approximately cubic shaped voxel dimensions of 2 mm^3 . A multi window-level conversion was applied to enhance the visibility of stroke lesions on NCCT (0 – 80 HU), and (collaterals on) both CTA acquisitions (0 – 100 HU). Conventionally, data is normalized to a zero mean and unit variance distribution for deep learning purposes, but since all our voxels are already standardized in HU, we only divide all voxel values by 1000, just bringing the them to the same order of magnitude of this distribution.

Two types of datasets were created for training, validation and testing: a SECT and a DECT set. The SECT set contained both SECT and DECT patients, with the simulated SECT angiography derived from DECT replacing the conventional single-energy CTA. For the DECT set, early and delayed phase iodine only images were used instead of the simulated single-energy CTAs. Both SECT and DECT sets were split into a training, validation and independent testing part according to the following ratio: 0.6/0.2/0.2. To minimize differences in stroke positive voxels between the sets, patients were sorted based on increasingly ground truth stroke volume and. Per group of 5 consecutive patients, they were assigned randomly to the training (3), validation (1) and test set (1).

4.2.3 DEEP LEARNING ARCHITECTURES

Two types of fully convolutional deep-learning models have been implemented using the Keras (F. Chollet et al., www.keras.io) and Tensorflow (M. Adabi et al, www.tensorflow.org) libraries based on the Python programming language (Python Software Foundation, www.python.org). Both models were implemented in a two-dimensional (2D) and 3D fashion, using 2D and 3D operations respectively. The designed models

predict for every input voxel the likelihood of being infarcted. Models were trained on a NVIDIA GTX 1080 Ti GPU with 11 GB of memory.

U-Net model

The main idea of the U-net architecture is that it can perform localization and segmentation tasks by combining an encoding (feature extracting) part with a successive decoding (feature reconstruction) part, in which high resolution features from the encoding part are reused.¹³⁸ The conventional U-Net is adapted in several ways: (1) by replacing the convolutional layers with so-called *convolutional blocks*, consisting of a set of sequential operations and a specific connection pattern between different layers in the block; (2) by replacing all the pooling and up sampling operations with different (more advanced) transition layers. The network's depth D is defined by the number of transition layers. The set of levels then becomes L with $l = 0, \dots, D$ with the transition layers T transforming the feature maps from one level to the other. The global architecture of the adopted U-Net then becomes as depicted in Figure 4.2.

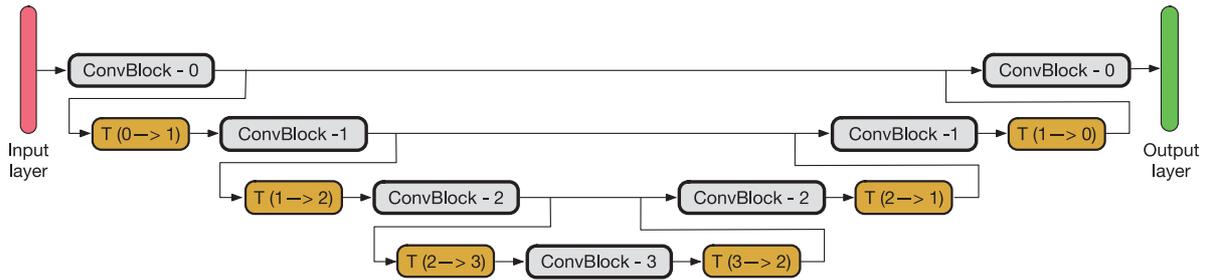


Figure 4.2: Example of global overview of the implemented U-Net architecture for $D = 3$. Numbers represent levels l of a convolutional block or transition (T) between levels of convolutional blocks.

Convolutional block design

Two types of convolutional blocks have been designed that use the residual, dense and Xception concepts, the latter using depthwise separable convolutions (Appendix D).^{142,143,146} The *ResDense* block combines the first two concepts, while the *ResNeXt* block concept combines the first and the last and is adopted from the block-design described in Xie et al.¹⁴⁷ An implementation of both *ResDense* and *ResNeXt* blocks at an arbitrary level l in the network is depicted in Figure 4.3 and Figure 4.4 respectively. In all blocks, the Rectified Linear Unit (ReLU) is used as non-linear activation function. Dropout (50%) has been implemented to avoid co-adaptation of neurons, and batch normalization (BN) is applied for accelerated learning.^{148,149} Feature maps are zero padded before every convolutional operation to let the input and the output feature maps have the same size. The order of operations within a convolutional block is Dropout - ReLU - Conv - BN, where dropout is only used in the Dense and Xception parts of the blocks. Also throughout the rest of the network, this order of operations and is kept the same. Filters in the convolutional layers were equally shaped in all directions having size 3, and the number of filters is $2 \times ((l \times c_{in}) + c_{in})$, with c_{in} being the number of input modalities (3 in our case). In the decoding part of the network, the *ResNeXt* configuration uses regular residual blocks without depthwise separable convolutions, while the *ResDense* configuration uses the same block types as in the encoding part.

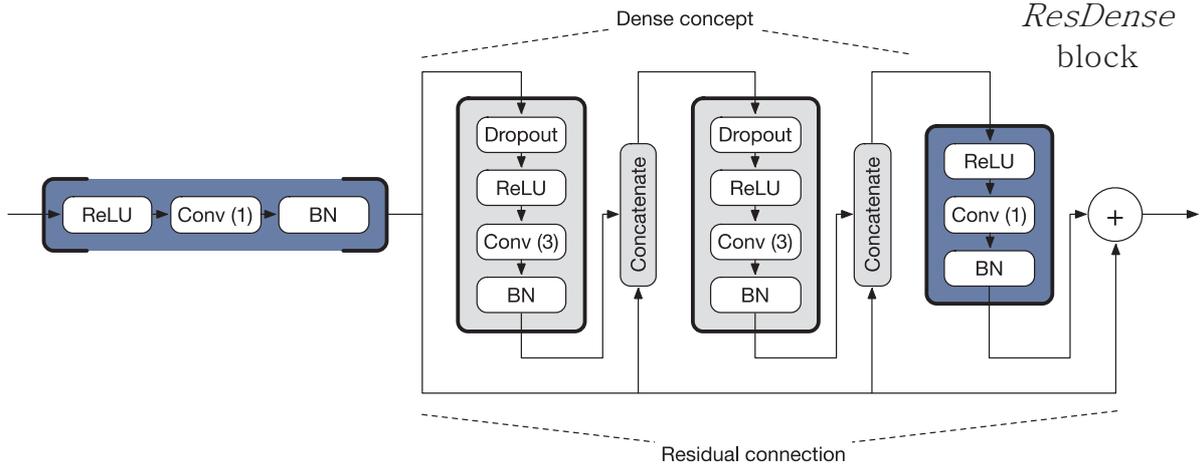


Figure 4.3: Implementation of ResDense block. The subblocks highlighted in gray present the Dense concept, while the granite highlighted subblocks (spatially separable depthwise convolutions) are required for the residual connection. The left granite subblock is only present in ConvBlocks in the decoding part of the network.

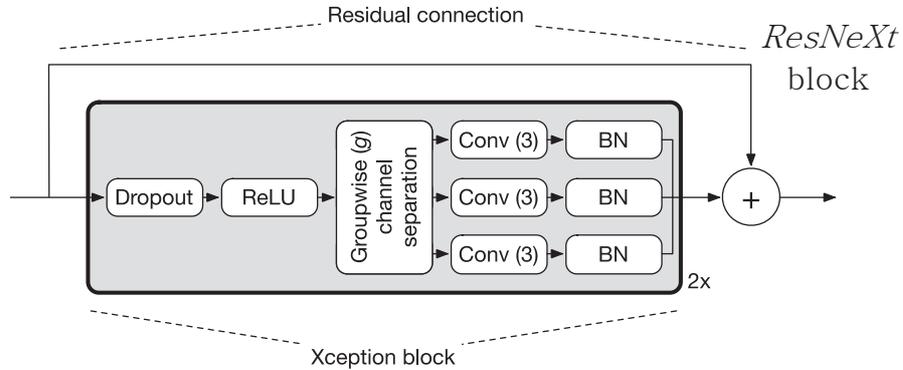


Figure 4.4: Implementation of ResNeXt block (only present in encoding part of network). The groupwise channel separation operation splits the stack of featuremaps in groups of size g in a depthwise manner. Subsequently, convolution operations act separately on these groups, without blending information between groups. Groups size $g = \text{number of filters} / 3$.

Transition layers

All transition layers operate between convolutional blocks of different levels. No dropout is used in the transition layers. Except for the first level ($l = 0$), every convolutional block at level l comes with transition layer. In the encoding part, down sampling of the output feature maps of the convolutional block in level l going to level $l + 1$ with a factor 2 is realized by transition layer $T_{l \rightarrow l+1}$ using a convolutional operation with filter size 2 and stride 2 (Appendix D). This method of down sampling includes a learning step instead of using a sub-sampling scheme with fixed rules (e.g. max pooling). In the *ResNeXt* configuration, depthwise separable convolutions (Appendix D) are used in the transition layers.

The opposite operation (so-called *deconvolution*) could be used to up sample the feature maps coming out of level $l + 1$ going to level l in the decoding part, however they usually suffer from checkerboard artifacts in the final output.¹⁵⁵ Another approach to still have a learning step involved in these layers ($T_{l+1 \rightarrow l}$) is to use a conventional nearest neighbor up sampling operation followed by a conventional convolutional layer, which is the method used in this architecture.

The number of convolutional filters in a transition layer is the same as in the subsequent convolutional block. The order of operations is (Up) – ReLU – Conv – BN.

Input and output layers

Both input and output layers follow a different order of operations as described above. Input data is pre-processed by a convolutional layer and batch normalized before entering the first convolutional block (Conv – BN). The output of the final convolutional block is processed by a ReLU activation layer and then passed to a convolutional layer having one filter of size 1 (spatially separable convolution). Finally, a sigmoid activation function is used to convert the output pixelwise to a likelihood map of being infarcted.

DeepMedic model

A different approach using a fully convolutional network is adopted from the DeepMedic architecture.¹⁵⁰ Here, the feature extracting part consists of two parallel pathways, processing image patches of a different image scales and corresponding FOVs. Feature maps were not zero-padded prior to convolutional layers, thus leading to a decrease in size of the feature maps after every convolutional layer. Although no subsampling operations are used in this architecture (which increases the spatial localization accuracy)¹³⁸, the receptive field of the filters in successive convolutional layers still increases (e.g. two successive 3×3 filters have a receptive field of 5×5). The combination of these aspects ensures that features can be learned based on local contextual information of the image area that is being predicted.

An overview of the DeepMedic architecture is depicted in Figure 4.5. Let f be the shape of the filters in a certain dimension (x, y for 2D; x, y, z for 3D) used in the convolutional layers, P_h the input shape of the normal (high) resolution patch, and o the shape of the predicted output image patch, then the depth (# convolutional layers) of the feature extracting path is $D = (P_h - o) / (f - 1)$, with $d = 1, \dots, D$. The input shape of the low resolution path then becomes $P_l = (o / 2) + (D \times (f - 1))$. With $P_h = 32$, $o = 16$ and $f = 5$ (all based on the $128 \times 128 \times 64$ input volumes), $D = 4$ and $P_l = 24$, and thus $P_l < P_h$. The dual-scale aspect is achieved by defining the FOV of P_l to be $2 \times P_l$, which requires down sampling of the original image input patch with a factor 2. Then the receptive field at depth d of the convolutional filters in the high resolution path becomes $R_h = f + d \times (f - 1)$ and $R_l = (P_h / P_l) \times (f + d \times (f - 1))$ in the low resolution path. The number of filters in each convolutional layer is $3 + (f - 1) \times d$. After D convolutional layers, the feature maps of the low-resolution path are up sampled to match the feature dimensions of the high resolution path, and are then concatenated before entering the pixel/voxel wise prediction part. This second part exists of two fully connected layers implemented as convolutions with filter kernels of size 1. All convolutional layers are succeeded by a dropout (50%), batch-normalization and ReLU activation layer, except for the output convolutional layer (sigmoid activation). For the 3D configuration, f_z is decreased by a factor 2, and thus also the z -dimension of all other above mentioned shapes.

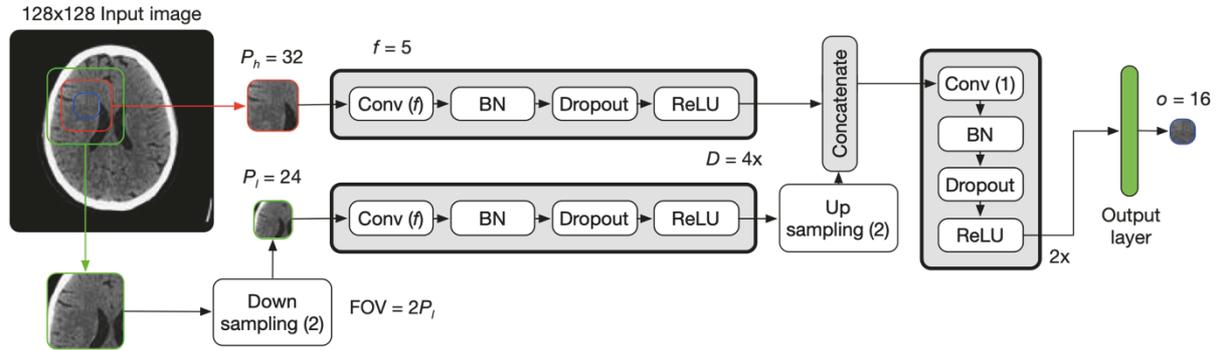


Figure 4.5: Implementation of DeepMedic architecture. The presented configuration ($D = 4$, $f = 5$, $P_h = 32$, $P_l = 24$ and $o = 16$) is the one used in the experiments. An example of the patch-extraction process is presented on the left.

Separate modality processing

Since a convolutional operation acts both in the spatial and channel domain, it cancels out separate channel information. When separate imaging modalities are encoded as image channels, as in our case, this means that already after the first convolutional layer, the extracted features contain mixed information of all input modalities. However, extracting features solely from individual modalities and combine them at output end of the network is more similar to a radiologist’s approach to analyze these different scans. To mimic this approach, both models have also been implemented using parallel pathways of the same architecture for every input modality. Output feature maps of the parallel pathways were concatenated at the end of the feature extracting part in the DeepMedic architecture and right before the final convolutional layer in the U-Net architecture.

4.2.4 TRAINING MODELS

Although both models (U-Net/DeepMedic) have been implemented in several configurations (2D/3D operational dimensionality and mixed/separate modality processing), all models used the same loss function and hyperparameters.

Loss function

After every feed forward predicted batch, a loss can be computed by comparing the pixel values p_n (each individually varying between 0 and 1) of the predicted likelihood maps P (also called soft-labels) with the ground truth values r_n from the corresponding binary ground truth segmentations R in that batch of N voxels in total, using a loss function. The gradient of this loss function determines in which direction the weights in the convolutional filters in the network should be adjusted to minimize this loss function iteratively and let the network fit the data properly over time.

In our case, only relatively few pixels are stroke positive (1), which requires a loss function that takes the huge class imbalance with respect to the stroke negative pixels (0) into account. A commonly used loss function to deal with this problem is the so-called *Dice Loss* (DL), because it implicitly leaves out true negatives. It can be computed directly from the *Dice Similarity Coefficient* (DSC). Since our predictions are probabilities (soft-labels), the DSC and DL turn into their soft variants SDSC and SDL:

$$\text{SDSC} = \frac{2 \sum_{n=1}^N p_n r_n}{\sum_{n=1}^N (p_n + r_n) + \varepsilon} \quad (4.1) \quad \text{SDL} = 1 - \text{SDSC} \quad (4.2)$$

The ε term (10^{-7}) is required to avoid instability of the loss function in the case when both P and R are empty. This necessity is also directly the downside of the SDL, since it has no smooth gradients for being used in the stochastic gradient descent weight optimization algorithm. Another popular loss function is the (*binary*) *cross-entropy* loss function that is also used in logistic regression and tries to maximize the log-likelihood of P being the same as R for both classes. Since we have a class imbalance problem, we want to penalize false positives more than false negatives, and thus we want to have a weighted binary cross-entropy (WBCE) loss:

$$\text{WBCE} = -\frac{1}{N} \sum_{n=1}^N (w p_n r_n \log(p_n) + (1 - r_n) \log(1 - p_n)) \quad (4.3)$$

with N the total number of pixels within a batch and w the weight attributed to the positive class and defined as

$$w = \frac{N - \sum_n p_n}{\sum_n p_n} \quad (4.4)$$

In order to have the benefit of both loss functions, the final implemented loss function is weighted version of both:

$$\text{Loss} = 0.5 \times \text{SDSC} + 0.5 \times \text{WBCE} \quad (4.5)$$

Hyperparameters

All models were trained (i.e. iteratively minimizing the loss function along its gradient) with the Adam optimizer, using individual adaptive learning rates for different weight parameters.¹⁵⁶ The base learning rate of the optimizer was set to 0.001 (default). The loss on the validation set was monitored to save the best model weights after every epoch where a new minimum was reached. Moreover, the loss monitor triggered the following actions after *not* reaching a new minimum after a predefined number of epochs: (1) the learning rate decreased with a factor 0.25 after 10 epochs; (2) training was stopped after 20 epochs. The used weight initialization method was according to He et al.¹⁵⁷ Batch size was maximized for every configuration based on the available GPU memory, with a maximum of 256 samples.

Training schemes

Two basic training schemes have been implemented: dense training and patch-based training. During dense training a batch consists of multiple full sized input samples, i.e. $128 \times 128 \times 3$ images for the 2D case, and $128 \times 128 \times 64 \times 3$ for the 3D case. In a patch-based training scheme, local patches of a

predefined shape are extracted from these individual input samples to fill a batch. Two different patch-sampling strategies have been employed: sliding-window (SW) and foreground-background (FGBG). In the SW approach, every input sample is split up in equally shaped patches, together covering the whole area of the input sample, without having overlap. In the FGBG approach, patches of a predefined shape were sampled with the central pixel/voxel (data points) having an equiprobability of being lesion positive or negative. Central data points were randomly chosen from both classes per input sample. Patches were padded when the central data point was near the border of the input sample. This sampling strategy further minimizes the class-imbalance between lesion positive and negative voxels in a batch, and it has the aim to more effectively learn to detect stroke. The shape of the used patches was 32×32 for 2D configurations and $32 \times 32 \times 16$ for 3D configurations.

In addition, real-time data augmentation was implemented by applying geometric transformations (rotations, scales, translations) to the input samples to avoid overfitting to the training data and improve generalizability of the models.

4.2.5 EXPERIMENTS

The following *configuration parameters* of deep learning architectures and forms of 3-phase CT data were combined in total 128 different configurations:

- Deep learning architecture
 - o CNN type (U-Net / DeepMedic)
 - U-Net (ResDense / ResNeXt)
 - o Operational dimensionality (2D / 3D)
 - o Modality processing (mixed / separate)
- Data
 - o Input modality (SECT / DECT)
 - o Data augmentation (yes / no)
 - o Training scheme (dense / patch-based)
 - Patch based (SW / FGBG)

The SECT datasets contain both SECT and DECT cases, where the latter cases are converted to simulated SECT images. A complete overview of all configurations is presented in the Appendix F.

4.2.6 PERFORMANCE METRIC

Since all trained models generate likelihood maps for voxels being infarcted, the outputs were binarized to get a final prediction and to be able to calculate the DSC:

$$\text{DSC} = \frac{2(X \cap Y)}{X + Y} \quad (4.6)$$

with X the binarized predictions and Y the ground truth voxels. The DSC ranges from 0 (no similarity) to 1 (complete similarity). After every experiment, DSC on the independent test set was calculated using the model weights with the minimum loss on the validation set, assuming this set of weights leads to best generalization to the test set.

Since the final computed DSC depends on the chosen threshold for binarization, an algorithm similar to expectation maximization has been implemented to estimate the threshold resulting in the highest DSC for every trained model. Threshold optimization was performed on the training set using the same model weights (minimum loss on validation set). Subsequently, the found threshold was applied to the predictions on the test set to calculate the final DSC. An example of the DSC optimization process is presented in Appendix E.

4.2.7 DATA ANALYSIS

All statistical tests were implemented in the R programming language. Significant differences between groups were tested using the (Welch) Two-Sample t Test or the Mann-Whitney U test, depending on the normality of the input data (Shapiro-Wilk test) and equality of variances (F-test).

Data characteristics

Common data acquisition parameters (CT to MR time and CTA delay time) between SECT and DECT datasets were assessed for significant differences.

Stepwise configuration selection

Best performing configurations were selected based on a stepwise configuration selection. In every step, the results of all (remaining) experiments were considered in two groups for every configuration parameter and groups were tested on a statistical significant difference ($p < 0.05$) in DSC on the independent test set. The configuration parameter with the lowest significant p -value was selected to be fixed for the next step, and thus only the configuration results of best performing alternative of that parameter were kept in the next step. Selection stopped at the step where there were no more significant differences between groups of remaining configuration parameters.

Additional analyses

For the best result of the remaining configurations, voxel-wise receiver operating characteristics (ROC) analysis and precision-recall (PR) analyses were done. To correct as much as possible for the huge class imbalance, only voxels covering the brain area in the NCCT were considered using the pre-processed brain masks. For all patients' binarized output maps, the predicted stroke volume was calculated, which is the most clinically relevant output metric. Bland-Altman analysis was performed to compare predicted and ground truth lesion volumes.

Subgroup analysis

Finally, analyses for two clinically relevant subgroups were done to explore performance. The first group contained only patients where a stroke lesion was already seen/suspected based on the radiologist's report of the NCCT. These cases were considered as easier cases and therefore might show an increased performance. The second group contained only patients that had a confirmed anterior circulation occlusion (ACO) based on CTA, since it is assumed that these patients are more likely to have a large stroke lesion. For both groups, models were trained on these datasets separately, only using the best configurations according to the stepwise configuration selection. Results between groups were compared using a pairwise t-test.

4.3 RESULTS

4.3.1 DATA CHARACTERISTICS

Scans of in total 293 patients (109 DECT, 184 SECT) were included and successfully pre-processed. Among those, 121 patients were identified as easy cases, and 106 patients had a confirmed ACO. The distribution of the delay between the two CTA acquisitions and the time between CT and MR acquisitions are described in Table 4.1. There was a statistical significant difference between SECT and DECT angiography delay times in all datasets ($p < 0.001$).

Table 4.1: Distribution of acquisition parameters of different (sub)datasets. All values are presented as median and interquartile range (25% - 75%).

| | All | | | Easy | | | ACO | | |
|----------------------|----------------|----------------|---------|----------------|----------------|---------|----------------|---------------|---------|
| | SECT | DECT | p | SECT | DECT | p | SECT | DECT | P |
| CT to MR time (min)* | 59 (25-308) | 73 (32-298) | 0.628 | 54 (26-215) | 42 (17-174) | 0.394 | 44 (25-108) | 42 (26-80) | 0.704 |
| CTA delay time (s) | 69 (35- 81) | 14 (14-15) | < 0.001 | 66 (15-78) | 14 (14-15) | < 0.001 | 61 (15-79) | 15 (14-15) | < 0.001 |
| Total # patients | 293 | 109 | - | 121 | 31 | - | 106 | 30 | - |

* Note that CT to MR times are skewed due to the presence of MR prior to CT acquisitions (negative times).

4.3.2 STEPWISE CONFIGURATION SELECTION

The stepwise selection process identified the following significant configuration parameters 1) SECT ($p < 0.001$); 2) DeepMedic ($p < 0.001$); 3) 2D ($p < 0.05$). With these configuration parameters fixed, there was no additional significant improvement in performance considering the patch sampling strategy (SW/FGBG), with/without data augmentation and mixed/separate modality processing subgroups. The results (p-values) of the stepwise configuration selection process are presented in Table 4.2.

Table 4.2: Results of the stepwise configuration selection. All values are presented as p-values of (Welch) Two-Sample t Tests or Mann-Whitney U tests. Sampling strategy was considered in the analysis after the set of configurations was constrained to a patch-based training scheme.

| Total configs | Modality processing | Data augmentation | Training scheme | Patch sampling strategy | Operational dimensionality | CNN type | Input modality |
|---------------|---------------------|-------------------|----------------------|-------------------------|----------------------------|----------------------|-----------------------|
| 128 | 0.509 | 0.971 | $9.60 \cdot 10^{-2}$ | - | $9.97 \cdot 10^{-2}$ | $1.08 \cdot 10^{-4}$ | $1.42 \cdot 10^{-13}$ |
| 64 | 0.242 | 0.734 | 0.192 | - | 0.196 | $6.25 \cdot 10^{-7}$ | SECT |
| 16 | 0.309 | 0.783 | Patch | 0.378 | $4.73 \cdot 10^{-2}$ | DeepMedic | SECT |
| 8 | 0.236 | 0.973 | Patch | 0.654 | 2D | DeepMedic | SECT |

4.3.3 SUBGROUP ANALYSIS

Performance of the remaining configurations on all patients and on both subgroup datasets are presented in Table 4.3. All configurations have the above mentioned top performing parameters fixed. Although the 'Easy' dataset has the highest DSC in configuration 3, a pairwise t-test showed a statistical significant difference in DSC between the 'All' and 'ACO' groups ($p = 0.014$), with a higher performance for the 'All' dataset. Since for both the 'All' and 'Easy' datasets, the DSC for configuration 3 is highest, with a higher DSC for the 'Easy' dataset, performance on these two datasets using this configuration is further explored in ROC, PR and Bland-Altman analyses.

Table 4.3: Performance of remaining configurations on the independent test set of all cases and on both subgroups. Top performing configuration parameters (patch-based, 2D, DeepMedic, SW sampling strategy) were fixed. Values are presented as DSC.

| Configuration | All | Easy | ACO | Patch sampling strategy | Modality processing | Data augmentation |
|---------------|--------------|--------------|--------------|-------------------------|---------------------|-------------------|
| 1 | 0.213 | 0.301 | 0.204 | SW | Mixed | No |
| 2 | 0.229 | 0.306 | 0.190 | SW | Mixed | Yes |
| 3 | 0.254 | 0.330 | 0.194 | SW | Separate | No |
| 4 | 0.199 | 0.192 | 0.143 | SW | Separate | Yes |
| 5 | 0.192 | 0.184 | 0.198 | FGBG | Mixed | No |
| 6 | 0.208 | 0.089 | 0.165 | FGBG | Mixed | Yes |
| 7 | 0.220 | 0.245 | 0.196 | FGBG | Separate | No |
| 8 | 0.245 | 0.129 | 0.186 | FGBG | Separate | Yes |

4.3.4 ROC AND PRECISION-RECALL ANALYSES

A voxel-wise ROC analysis of the top performing configuration on all cases (configuration 3, Table 4.3) is presented in Figure 4.6. ROC analysis is limited to the voxels of the brain area in the NCCT, based on the generated NCCT brain masks. Optimal threshold cutoff points (indicated with *) were chosen based on maximization of the DSC on the training set. In the 'All' dataset, this threshold resulted in a sensitivity, specificity and precision of 0.31, 0.99 and 0.32 respectively for the training set and 0.33, 0.99 and 0.21 for the test set. In the 'Easy' dataset, these numbers were as follows: 0.38, 0.99 and 0.38 for the training set and 0.33, 0.99 and 0.34 for the test set.

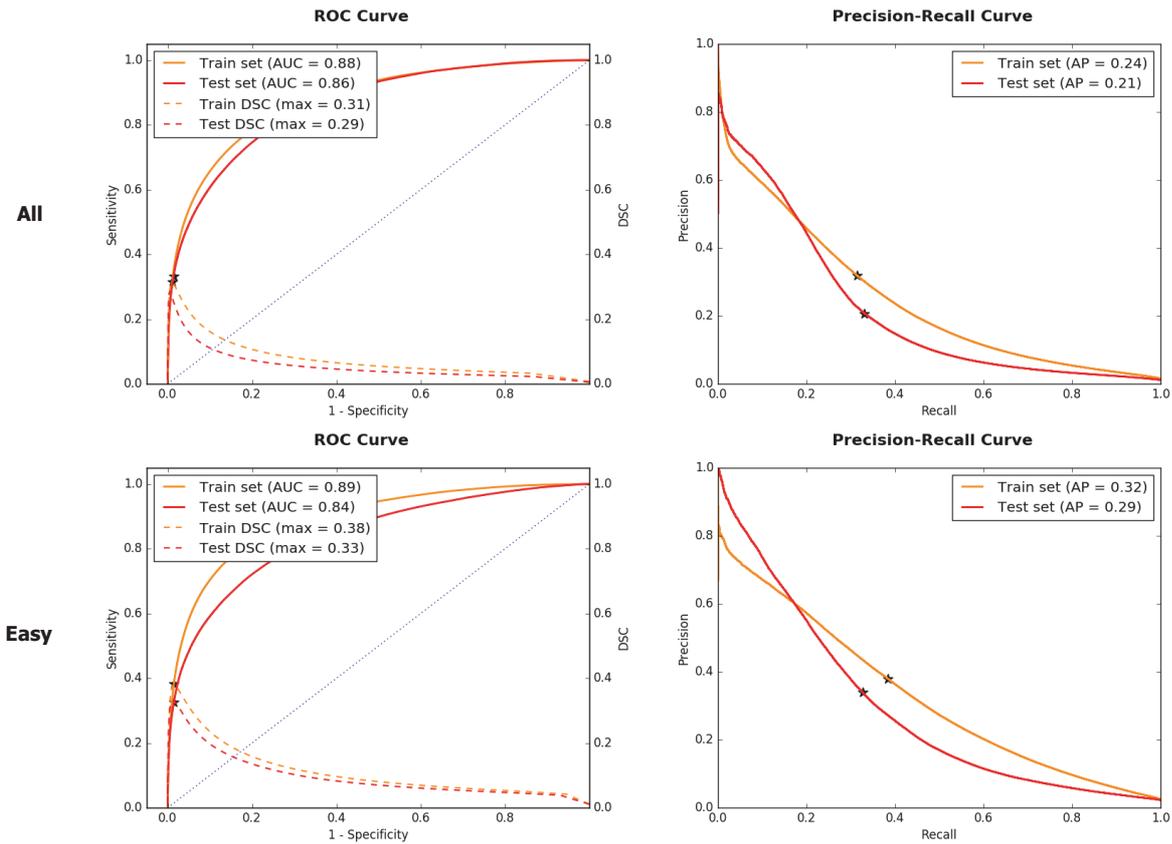


Figure 4.6: ROC curves (left) and Precision-Recall curves (right) from voxel-wise analysis limited to the brain region. The top row presents the performance on the independent test set on ‘All’ cases, the bottom row the ‘Easy’ cases. The * on the curves indicates the used threshold to binarize predicted outputs where DSC on the training set is maximized. Note that at ‘All’ cases the maximum DSC on the test set presented by the dashed curve is higher than the DSC in Table 4.3, since the binarization threshold maximizing DSC on the test set is slightly different than for the training set, which is used in Table 4.3. AP = average precision.

4.3.5 BLAND-ALTMAN ANALYSIS

For the same configuration as above and again from both ‘All’ and ‘Easy’ datasets, predicted stroke lesion volumes were compared with DWI (ground truth) lesion volumes on individual patient basis, using a Bland-Altman analysis. The corresponding plots are presented in Figure 4.7. Median and 75% interquartile range (IQR) values of differences between predicted and ground truth DWI lesions were all positive, while the 25% IQR values were negative. In the ‘All’ dataset, the correlation coefficient between the ground truth and predicted lesion volumes was higher in the training set ($r = 0.71$) than in the test set ($r = 0.58$), while in the ‘Easy’ dataset, there was only a small difference between both coefficients ($r = 0.65$ and $r = 0.67$ respectively). Areas regarding the correctness in clinical decision making for EVT based on the predicted and ground truth stroke volumes are defined as follows: correct (green) in case both predicted and ground truth volume are < 70 cc or > 100 cc; incorrect (blue) in case predicted volume > 100 cc and ground truth < 70 cc or vice versa; uncertain (blue), in case predicted and/or ground truth lesion volume were between 70 cc and 100 cc, since in this lesion volume range one should rely on clinical information whether to perform EVT or not. The incorrect EVT decision are rates are as follows: 4% on the ‘All’ train dataset, 5% on the ‘All’ test dataset, 7% on the ‘Easy’ train dataset and 8% on the ‘Easy’ test dataset.

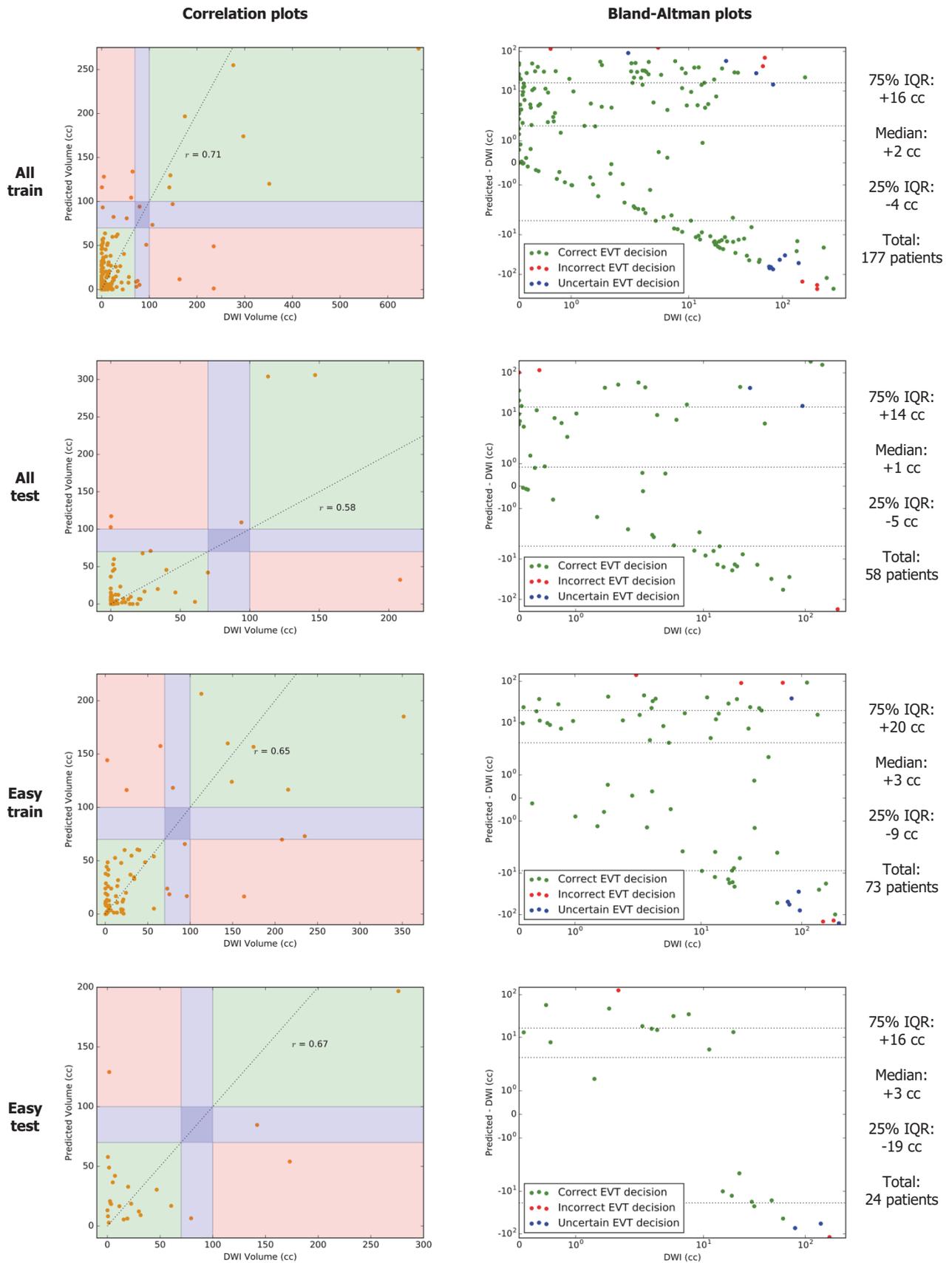


Figure 4.7: Correlation plots (left) with reference line (dotted) and Bland-Altman plot (right) of predicted results from the ‘All’ and ‘Easy’ training and test datasets to present the coherence between predicted and ground truth DWI lesion volumes. In the correlation plots, the green, red and blue shaded areas correspond with correct, incorrect and clinical EVT decisions in the Bland-Altman plots. Median, 25th and 75th quantiles of differences between predicted and ground truth lesions are indicated by dotted horizontal lines. Note that there is a difference in axes between all plots, with the Bland-Altman plots having logarithmic axes.

4.3.6 VISUALIZATION OF RESULTS

The predicted output of one patient from the independent test set is presented in Figure 4.8, together with the 3-phase input CT data (multi window-levelled), DWI and ground truth segmentation. Data is presented as an example. The DSC of the presented patient is 0.69 and of the presented slice 0.79. The ground truth volume for this patient was 94 cc and the predicted volume 109 cc.

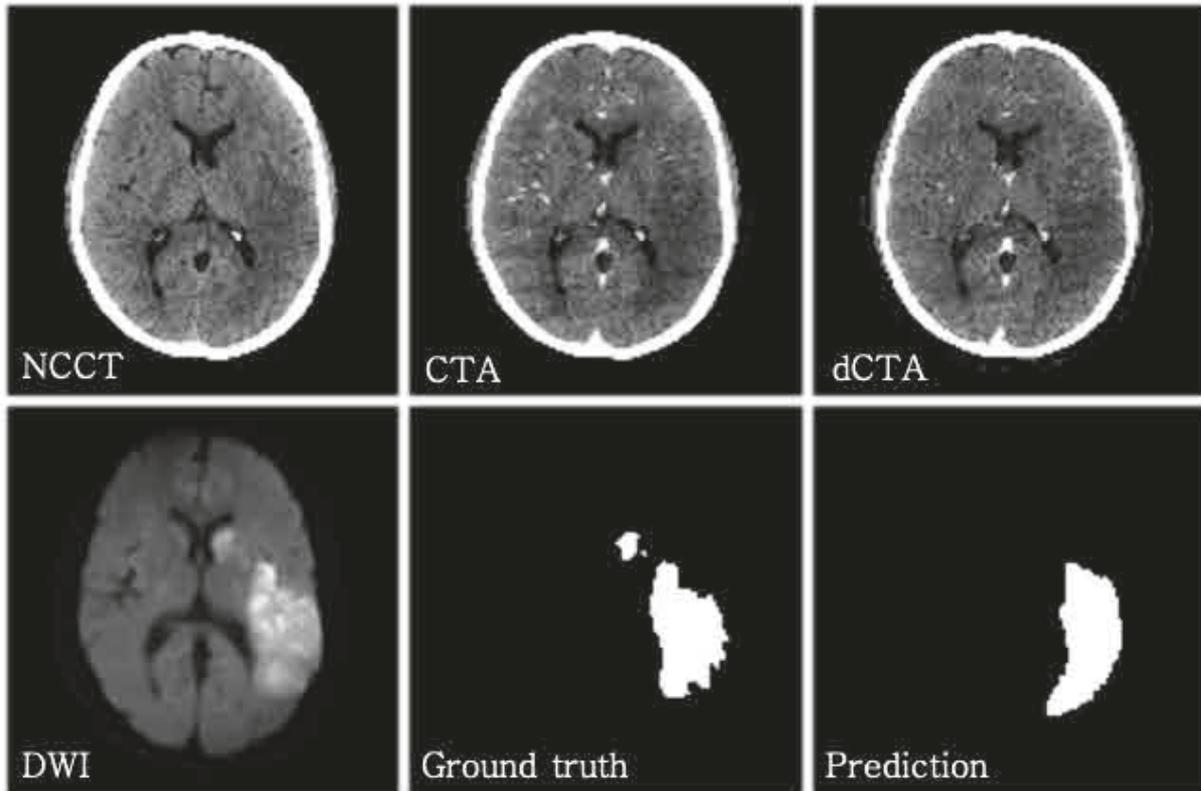


Figure 4.8: Visualization of CT input data (top row), DWI, segmented ground truth (label) and prediction (bottom row) of an independent test case (SECT input). Used CT window levels are the same as applied to the CNN's input data. The patient had a double left M2 inferior division occlusion. DWI was acquired 45 minutes after CT and shows a stroke lesion in the left MCA territory.

4.4 DISCUSSION

In this study, multiple deep learning architectures based on two CNN types have been implemented and tested in several configurations to find the best performing system to detect acute stroke lesions from 3-phase CT input data. The 2D DeepMedic architecture using SECT data and a sliding-window patch-sampling strategy performed better than other tested configurations on the available data. Although these top performing configurations showed generalization in performance between the training and independent test set, the DSC was still limited (0.254), even on 'easy' cases (0.330). However, Bland-Altman analysis showed that the system might be useful to make clinically relevant stroke lesion volume predictions to be used for triaging AIS patients.

The SECT configuration parameter showed the most significant difference in performance compared to all others ($p = 1.42 \cdot 10^{-13}$). Since deep learning is very data hungry, a simple explanation of this finding is that more SECT than DECT scanned patients were available. This is in line with the other significant configuration parameters: DeepMedic, and thus a patch-based training scheme; and 2D images instead of 3D volumes. These configuration parameters all have the advantage that they lead to more data samples than their counterpart. Another explanation for the superiority of SECT above DECT could be that the voxel values of the iodine-only images are not that well standardized compared to original SECT Hounsfield Units, making them less suitable to be window-levelled in using as standard window-level of 0 - 100 HU. On random visual inspection of the window-levelled iodine images, there seems to be quite some difference between the generated voxel gray intensities by the DECT post-processing software. Further, patches force the CNN more to extract local image features that are expected to be relevant for the detection of AIS, such as subtle local gray-white matter washout in NCCT images, instead of

irrelevant global features, such as shape of the brain/head. The DeepMedic network is specifically designed to process contextual information of small image patches in a multi-scale approach. Spatial localization accuracy of this network is more maintained since it uses no transition (up/down sampling) layers, like U-Net does.¹³⁸ Finally, although above mentioned configuration parameters had significant better results than others, the stochastic nature of the used gradient descent optimization algorithms during training lead to outlying results. For example, there were still two other configurations (3D DeepMedic SECT with separate modality processing, and 3D U-Net ResNeXt SECT with data augmentation) that reached higher DSCs (0.278 and 0.294 resp.) than the highest reported DSC for *all* cases in Table 4.3 (0.254).

Results from the ROC analysis are difficult to interpret in terms of absolute numbers because there is a huge class imbalance between stroke positive and negative voxels although only voxels in the brain area were considered. The precision-recall (PR) curves implicitly correct for this class imbalance by considering only stroke positive voxels and positively predicted voxels. The curves show that the performance on the training is fairly generalizable to the test set, since there is only a minimal drop AUC, AP and DSC.

Correlation plots of the Bland-Altman analysis showed moderate to good correlations between ground truth and predicted lesion volumes. The median and IQR values in the Bland-Altman plots tend to an overestimation of predicted lesion volumes, since both are positive. However, when taking the clinical decision margins into account, the number of incorrect EVT decisions due to a too high or too low predicted lesion volume is practically the same ('All' train: 4 and 3 out of 177 patients; 'All' test: 2 and 1 out of 58 patients; 'Easy' train: 3 and 2 out of 78 patients; 'Easy' test: 1 and 1 out of 24 patients). The higher incorrect EVT decision rate in the 'Easy' dataset compared to the 'All' dataset may be caused by the higher average ground truth lesion volume in the 'Easy' dataset, closer to the EVT decision range of 70 to 100 cc. Nevertheless, the low numbers of incorrect EVT decision rates in combination with the moderate to good correlation coefficients are promising, and might have implications to further study the effects of these decision models towards clinical application. Finally, all Bland-Altman plots show a 'linear' pattern (logarithmic axes) in the lower part of the graphs, starting from 0 difference between predicted and ground truth lesion volumes. Probably these datapoints represent false negative 'detected' lesions.

CTP is the most commonly used method to determine ischemic core (and penumbra) in acute stroke patients in clinical practice. There are two other studies in literature that compare identification of ischemic core using CTP with DWI in a comparable way to this study. Cereda et al. reached a sensitivity and specificity of 67% and 87% respectively in a voxel-based ROC analysis when using a threshold of 38% relative cerebral blood flow (rCBF), only considering voxels in the according to CTP hypo-perfused brain region. Correlation coefficient between predicted lesion volumes and DWI volumes was 0.83 and mean difference between CTP and DWI lesion volumes was only 0.3 cc.¹⁵⁸ Copen et al. reached a voxel-wise sensitivity of 79% to 81% and a specificity of 83% to 85% depending on the chosen cerebral blood flow threshold, considering the brain voxels on the side where the lesion was suspected.⁶⁴ Both show a superior performance compared the results presented in this study and suggest that four dimensional (4D) CTP data is richer in information compared to 3-phase CT.

A limitation of this study might be the quality of the input data. More specific selection of cases that all have the same image acquisition parameters would be required to achieve more standardized data. These acquisition parameters include standardization of CTA delay times, contrast injection rates and used image reconstruction kernels. In this way, for example imaging of the collateral physiology could be improved, of which it is known that it holds predictive information on stroke lesion volume. Other aspects include narrowing the allowed time window between CT and MR acquisition and add parameters, such as time from stroke onset to CT, so that the deep learning architecture might able to learn how to interpret the input data with respect to these parameter values. Also ADC data could be useful to determine several levels of restricted diffusion in the brain to penalize a deep learning system more for inaccurate predictions of areas with severe restricted diffusion.

Two other interesting ideas are as follows. First, to use richer CTP data as input, which in fact consists of many more phases (and thus radiation dose and contrast) compared to the currently used 3-phase data in this study. A CNN potentially only needs a part of these phases to make accurate stroke lesion predictions. Ideally, an optimum should be found between the minimum required number of contrast-

enhanced CT phases and the CNN's ability to predict acute stroke lesions. This could potentially lead to implications to change future stroke CT imaging protocols. Second, to go along with the current 3-phase CT data, but to split up the problem in multiple individual prediction problems. For example, deep learning could be used to predict ASPECTS from NCCT data, while a second (deep learning) pipeline could process CTA data to quantify the collateral circulation in the early and delayed phase. The combined outputs of these systems could be used in a final classification or regression model to predict DWI lesion volumes, with the ASPECTS also indicating the global location of the infarct.¹⁵⁹

Further research should mainly focus on further improving input data by standardization and adding clinical parameters, rather than improving CNN architectures. It is clear when comparing this study to other studies where CNNs have shown outstanding results in detection and segmentation tasks in the field of medical imaging, that the right data and the way how data is presented to the network (data augmentation, patches, sampling) is key to set a good baseline result to potentially solve this hard, but truly relevant problem of acute stroke detection on CT images.

4.5 CONCLUSION

This study has shown that a deep learning system might be useful to predict DWI infract core lesions based on 3-phase CT data for triaging AIS patients. Systems based on a 2D DeepMedic architecture and processing SECT images perform better on the independent test set than systems based on the U-Net architecture. Although the performance in terms of DSCs of these 2D DeepMedic based systems still seems limited, the predicted lesion volumes result in a promising low number of incorrect EVT decisions for triaging AIS patients. Improvements may be achieved by adding more CTA phases to the input data, enriching the available information on the underlying vascular stroke physiology, by standardizing CT acquisitions and by including clinical parameters to the prediction model. Future research should focus on these aspects, for example by finding a tradeoff between required CTA acquisitions (radiation and contrast dose) and on the implications of the predicted lesion volumes in clinical practice.

5

Investigating the predictive ability of NCCT texture features in combination with DECT iodine features for intracerebral hematoma expansion

5.1 INTRODUCTION

Hematoma expansion (HE) is a common and serious complication after intracerebral hemorrhage (ICH), strongly associated with increased mortality, worse functional outcome and neurologic deterioration.^{100,101,160} The exact mechanism driving HE is not clear, possibly being caused by the inflammatory response after initial bleeding, hemostatic dysregulation, breakdown of the blood-brain barrier or local tissue distortion.^{98,161,162} Early recognition of patients at risk for HE helps to guide selection of ICH patients for targeted intervention.

Many radiographic markers for HE, based on qualitative assessment of non-contrast computed tomography (NCCT) and computed tomography angiography (CTA) scans, have been identified. On NCCT, these include hematoma location¹⁶³ and morphologic appearance of the hematoma, such as shape and density variation, which are both considered as independent predictors of HE.^{164,165} Given this, several radiographic markers were proposed to qualitatively describe shape or density variations, including island sign¹⁶⁶, satellite sign¹⁶⁷, black hole sign¹⁶⁸, blend sign¹⁶⁹, swirl sign¹⁷⁰, margin irregularity¹⁷¹ and hematoma hypodensities.¹⁷² All of these markers are thought to reflect active hemorrhage, its time course or its multifocal nature. On CTA, the presence of foci of hyperintensities in the hematoma, known as the spot sign, is thought to represent active contrast extravasation, and has been shown to be independently associated with an increased risk of HE.^{111,116,173}

Although a high specificity is important for these markers to not indicate aggressive/invasive therapy (e.g. craniotomy) in vain, the general problem of most of these individual markers is that they lack good sensitivity in order to make them clinically valuable. Also standardized assessment and thus automated quantification is desired to seamlessly incorporate them in the clinical workflow. Recently, a study on HE using dual-energy CT (DECT) showed that quantified iodine in the hematoma and spot signs can be combined to a so-called I2-score to predict HE with a higher sensitivity than conventional, qualitatively extracted image features.¹²⁴ Other quantitative features that have been studied focused on morphologic appearance of hematomas on NCCT by computing different texture features based on image histogram

analysis of the hematoma region.^{132,174} Since both iodine and texture features assess different hematoma aspects, combining them in a prediction model would potentially lead to synergy. The aim of this study is to explore if a prediction model that combines iodine and texture can further improve the sensitivity of the I2 score without decreasing specificity.

5.2 METHODS

5.2.1 PATIENT SELECTION AND DATA COLLECTION

Data from patients who were referred to DECT between October 2014 and February 2018 and who had primary intraparenchymal ICH were analyzed retrospectively (approved by institutional review board, informed consent waived). Exclusion criteria were: (1) no follow-up NCCT within 48 hours for assessment of actual HE; (2) surgical treatment (craniotomy or minimal invasive hematoma evacuation) between initial DECT and follow-up scan; (3) (mainly) intraventricular hemorrhage; (4) presence of an extraventricular drain to decrease intracranial pressure at baseline scan; (5) inadequate image quality or image analysis. Patients scanned before February 2017 were used for training, while the remaining patients were used as independent test set.

All scans were acquired with a dual-source CT scanner (Siemens Somatom Definition Force™) at the emergency radiology department. Series included an initial thin-sliced (0.6 mm) non-contrast head CT, followed by an early and delayed phase DECT angiography. The early phase was timed using contrast bolus tracking. The NCCT was acquired at 120 kV and the DECT at 80 kV and 140 kV. Images were reconstructed using a standard filtered backprojection reconstruction kernel. Follow-up NCCT images were reconstructed at a 3 mm slice interval using a soft reconstruction kernel

5.2.2 IMAGE ANALYSIS

All image processing algorithms were implemented using MATLAB R2018a (The Mathworks, Natick, Massachusetts, USA), unless otherwise specified. An illustration of the global image processing pipeline is presented in Figure 5.1.

Hematoma segmentation and expansion assessment

Dual-energy CT images were converted in to virtual non-contrast (VNC) and iodine images, using syngo.via (Dual Energy Brain Hemorrhage, Siemens Healthcare), sliced at 3 mm intervals for computational cost reasons in the subsequent hematoma segmentation algorithm. On the VNC and on the follow-up NCCT images, the tissue was semi-automatically segmented into skull, parenchyma and hematoma. Patients were assessed for HE between initial VNC and earliest NCCT follow-up scan, and were classified as expander if the hematoma growth was >25% or >3 cm³. If the automated volume measurement on the follow-up NCCT was unreliable for expansion measurement due to ventriculostomy, hematomas were classified for expansion by neuroradiologic assessment by two experienced neuroradiologists, blinded to data analysis, until consensus was reached.

Iodine features

Hematoma masks from the initial 3 mm sliced VNC were transferred to the iodine images. Automatic detection of regional maxima using a quantile filtered mixture separation algorithm was used to detect spots in the iodine images.¹⁷⁵ Total iodine content within the hematoma (I_h) and, if any spots present, total iodine in the brightest spot (I_{bs}) were quantified (both in mg iodine) and combined to the I2-score as presented in Tan et al.¹²⁴

$$I_2 = \sqrt{I_h^2 + (125 \times I_{bs})^2} \quad (5.1)$$

Other computed iodine features were total iodine content in all spots (I_s , mg), mean iodine concentration in the hematoma (mean I_h), mean iodine concentration in spots (mean I_s), and maximum iodine concentration in all spots (max I_s), the latter three all in mg/ml. These bring the total number of iodine features to 6 per hematoma.

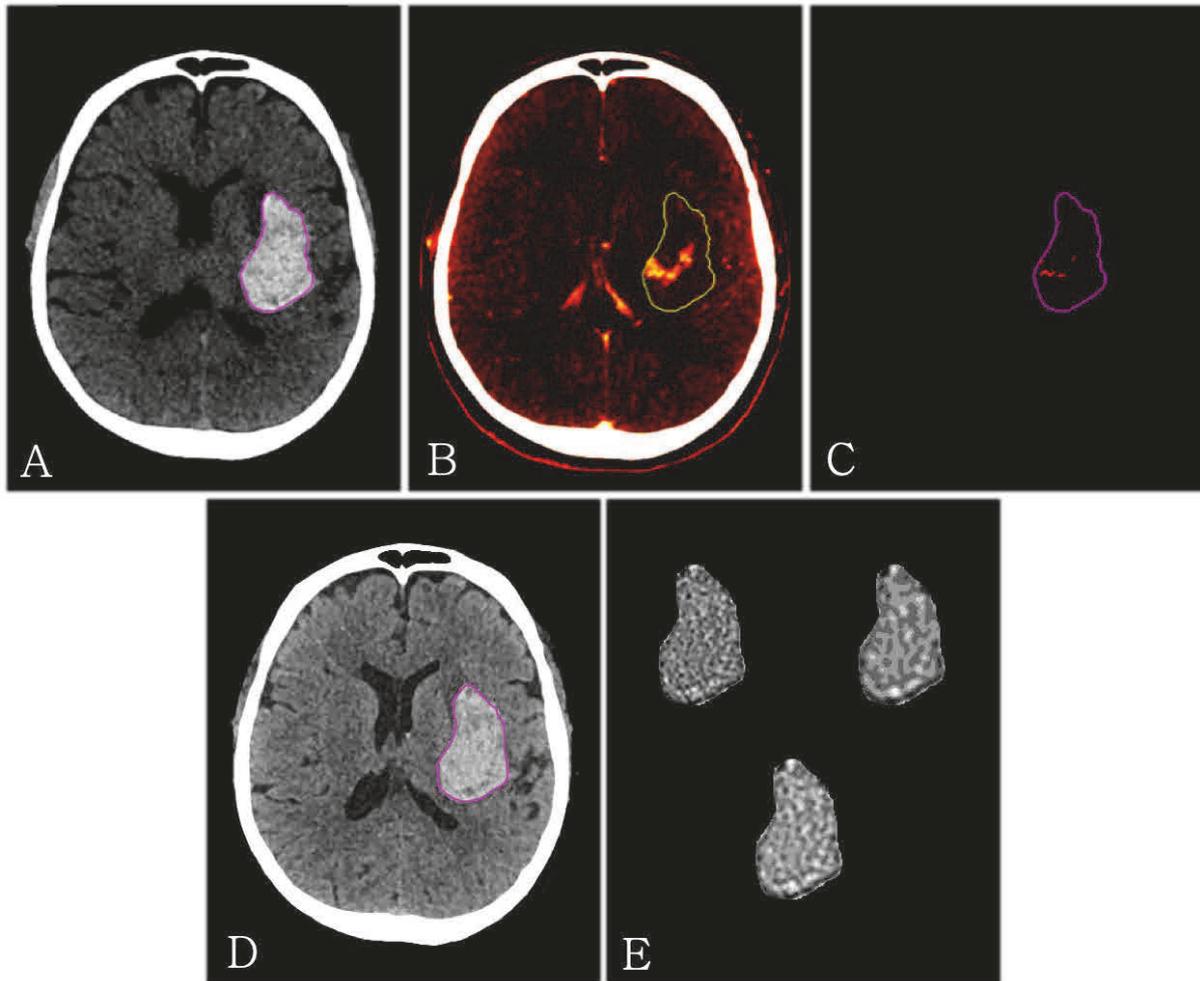


Figure 5.1: Overview of the images used throughout the processing pipeline. A) DECT Virtual non-contrast image used for hematoma segmentation. B) Hematoma mask is transferred to iodine only image. Two iodine features are computed on this image (I_h , mean I_h). C) Automatic spot detection using a quantile filtered mixture separation algorithm. Iodine features I_s , I_{bs} , mean I_s and max I_s . D) Hematoma mask transferred to NCCT for texture analysis. E) LoG filtered hematomas at different filter widths ($\sigma = 0.5, 0.9$ and 1.25 mm respectively).

Texture features

Subsequently, the VNC images and corresponding three-dimensional (3D) hematoma masks were registered and resliced to the thin sliced initial non-contrast head CT. Texture features were computed based on the masked 3D hematoma regions on the NCCT using different methods: (1) based on the image histogram; (2) based on two-dimensional (2D) gray level co-occurrence matrices (GLCMs); (3) based on 3D GLCMs. Image histogram analysis allowed computation of the following texture features: mean voxel intensity, standard deviation, variance, uniformity, skewness and kurtosis. Analysis of the GLCMs resulted in the following other texture features: entropy, energy, homogeneity, contrast and dissimilarity. A more detailed description of these methods is presented in section 2.1 and Appendix B. All GLCMs were computed on original Hounsfield Units of the NCCT images, without rescaling image intensities.

Moreover, NCCT images were pre-processed with *Laplacian of a Gaussian* (LoG) filters using three different filter widths ($\sigma = 0.5, 0.9$ and 1.25 mm, based on Shen et al.¹³²) and two different filter dimensions (2D and 3D). While the Laplacian component enhances texture, the Gaussian component allows extraction of texture features in different scale spaces, and thus assessment of both fine and coarse texture. Moreover, it stabilizes texture features with respect to CT image noise and possible differences in acquisition parameters. More details about the LoG filtering process is presented in Appendix C.

Features based on 2D GLCMs were only computed on 2D LoG filtered images, while 3D GLCM based features were computed on 3D LoG filtered images. Both 2D and 3D GLCM based features were computed on non-filtered images as well. Histogram based texture features were computed on non-filtered and all

types of filtered images. All filtering and texture computation methods combined resulted in 82 different texture features per hematoma.

5.2.3 STATISTICAL ANALYSIS

Statistical analyses were performed using R (R: A Language and Environment for Statistical Computing, www.r-project.org). Patient characteristics were described using median and interquartile range (IQR) for non-normal distributed variables, and mean and standard deviation for normal distributed variables. Significance in differences ($p < 0.05$) was assessed with a Mann-Whitney U Test or (Welch) Two-Sample t Test respectively.

Feature selection

An overview of the complete feature selection process is presented in Figure 5.2. Since we have a total set of 88 features (6 iodine, 82 texture), feature reduction is necessary to avoid model overfitting to training data. For a binary classification problem, the number of features should be (far) less than the number of positive cases in the training set to avoid model overfitting.¹³⁴ An extensive feature selection process has been designed to accomplish this. The method is based on how often features survive the selection process using three different selection methods with an increasing level of Gaussian noise applied to the training data, ranging from no noise to 1 standard deviation (11 levels in total). The feature survival aspect was chosen to assess the stability of features to slight variations in the training data. Three pre-selection methods were applied at all 11 noise levels, leading to a total of 33 feature selection simulations. Features were finally ranked on the number of survived selection simulations.

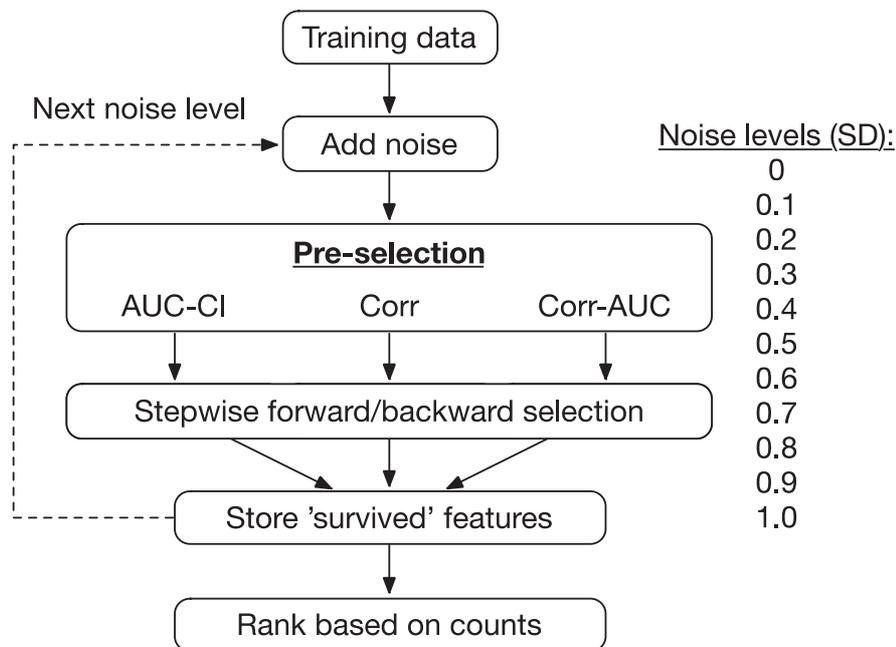


Figure 5.2: Overview of feature selection process.

Pre-selection method 1 (AUC-CI)

First, for all of the 88 individual features, receiver operating characteristic (ROC) analysis was performed and the area under the curve (AUC) was computed with its corresponding 95% confidence interval (CI). If the lower end of this interval was below 0.5, indicating that this feature might have no individual predictive value, the feature was dropped.

Pre-selection method 2 (Correlation)

Correlation between every pair of features was calculated first. For all pairs with a correlation coefficient above 0.8 and starting with the highest correlated pair, the feature with the highest mean absolute correlation coefficient with all other (remaining) features was dropped.

Pre-selection method 3 (Correlation – AUC)

Both correlations between every pair of features and AUC for every individual feature were computed. Again, for all pairs with a correlation coefficient above 0.8 and starting with the highest correlated pair, the feature with the highest mean absolute correlation coefficient with all other (remaining) features was dropped.

Subsequently, a second selection step was performed using a stepwise forward/backward parameter selection method. Starting with an empty logistic regression model, parameters were added/removed based on maximization of the likelihood ratio test between two possible candidate models, while the total number of model parameters was kept to a minimum. The Akaike Information Criterion (AIC) combines both these aspects by maximizing likelihood ratios and assigning penalties for an increasing number of model parameters. Therefore, the stepwise selection process aims to minimize AIC in every step.

Feature ranking

After the 33 feature selection simulations, selected features for all selection methods at all noise levels were collected and counted cumulatively. Features were ranked based on counts, with the highest ranking feature having survived the most feature selection simulations, indicating to be the most stable and most predictive feature.

Final model selection

Logistic regression models were fit to the training data with in increasing number of features, starting with the highest ranked feature only. For every model, AIC was computed, and the model with the minimum AIC value was selected as final combined *iodine–texture* model.

Iodine–texture model vs. I2–model

The performance of the combined iodine–texture model on the independent test set is compared to the I2–score, with a classification border for expansion at $I2 > 20$.¹²⁴ To assess stability of both models, a noise sensitivity analysis is performed, evaluating the performance of both models with an increasing level of noise applied to the data of the test set (again ranging from 0 to 1 standard deviation).

5.3 RESULTS

5.3.1 TRAINING AND TEST SET

Between October 2014 and February 2017, 38 patients with in total 42 hematomas (13 expanders) were included, successfully processed and were considered as the training set. From February 2017 to February 2018, another 64 patients with in total 69 hematomas (14 expanders) were included. The same datasets as in Tan et al.¹²⁴ were used with the following number of dropped patients due to extra required image acquisition and processing steps: 3 in the training set (failed registration of VNC to NCCT) and 2 in the test set (no baseline NCCT available). Characteristics of the two groups are presented in Table 5.1.

5.3.2 FEATURE SELECTION AND RANKING

Total survived selection simulations of features varied between 0 and 29. Two iodine features (I_h and I_{bs}) survived the most simulations, followed by image histogram mean after 3D LoG filtering ($\sigma = 0.9$ mm). I_s and 3D GLCM based dissimilarity (also 3D LoG filtered, $\sigma = 0.9$ mm) were the next two in the ranking, completing the group of features that survived a minimum of 5 simulations. This threshold of 5 survived simulations was chosen given the limited number of expanders (13) in the training set. Other features (up to 3 simulation survivals) are presented in Figure 5.3. There was only one texture feature that has been selected first in any of the simulations during the stepwise forward/backward feature selection process: 3D GLCM based dissimilarity (3D LoG filtered, $\sigma = 0.9$ mm). For the iodine features, I_h has been selected first 28 times, and I_{bs} and I_s both 2 times.

Table 5.1: Patient characteristics

| Variable | Training set ($n = 38$) | Test set ($n = 64$) | p -value |
|---|---------------------------|-----------------------|------------|
| Age (y)* | | | |
| Male | 66.6 \pm 15.5 (21) | 64.7 \pm 16.0 (35) | - |
| Female | 66.5 \pm 14.5 (17) | 66.8 \pm 18.7 (29) | - |
| Time to F/U CT scan (h) [†] | 7 [5 – 17] | 6 [4 – 13] | 0.16 |
| Time between early and delayed CTA (s) [†] | 70 [67 – 74] | 69 [59 – 74] | 0.18 |
| Admission SBP (mmHg) [†] | 164 [142 – 190] | 150 [135 – 174] | 0.06 |
| Admission INR [†] | 1.2 [1.1 – 1.3] | 1.0 [1.0 – 1.2] | <0.05 |
| Anticoagulant use (other than aspirin) | 11 (29) | 13 (20) | - |
| Last seen well (h) | | | |
| <3 | 9 (24) | 23 (36) | - |
| 3–6 | 6 (16) | 7 (11) | - |
| >6 | 23 (60) | 31 (48) | - |
| Unknown | 0 (0) | 3 (5) | - |
| Type of ICH | | | |
| IPH | 26 (68) | 44 (68) | - |
| IPH with SAH/SDH | 5 (13) | 10 (16) | - |
| IPH with IVH | 7 (19) | 10 (16) | - |
| Cause of ICH | | | |
| Hypertension | 26 (67) | 30 (48) | - |
| Cerebral amyloid angiopathy or malignancy | 4 (11) | 4 (6) | - |
| Trauma | 3 (8) | 9 (14) | - |
| Other | 1 (3) | 6 (9) | - |
| Unknown or undetermined | 4 (11) | 15 (23) | - |

Note: Unless otherwise specified, data are number of patients, with percentages in parentheses.

*Data are means \pm standard deviations, with the number of patients between parentheses.

[†]Data are medians with interquartile range between block parentheses.

F/U = follow-up, SBP = systolic blood pressure, INR = international normalized ratio, ICH = intracerebral hemorrhage

IPH = intraparenchymal hemorrhage, SAH = subarachnoid hemorrhage, SDH = subdural hemorrhage

IVH = intraventricular hemorrhage

5.3.3 FINAL MODEL SELECTION AND PERFORMANCE ON TRAINING SET

For the group of 5 features that survived a minimum of 5 feature selection simulations (I_h , I_{bs} , $Mean_{3D0.9}$, I_s , $Dissimilarity_{3D0.9}$) a logistic regression model was fit to the training data including an increasing number of features, ordered by ranking. The model that included 3 features (I_h , I_{bs} , $Mean_{3D0.9}$) had the lowest AIC, and was selected as final combined *iodine-texture* logistic regression (LR) model (Figure 5.4). Sensitivity and specificity on the training set were 100% and 100% respectively, compared to 85% and 90% for the I2-score.

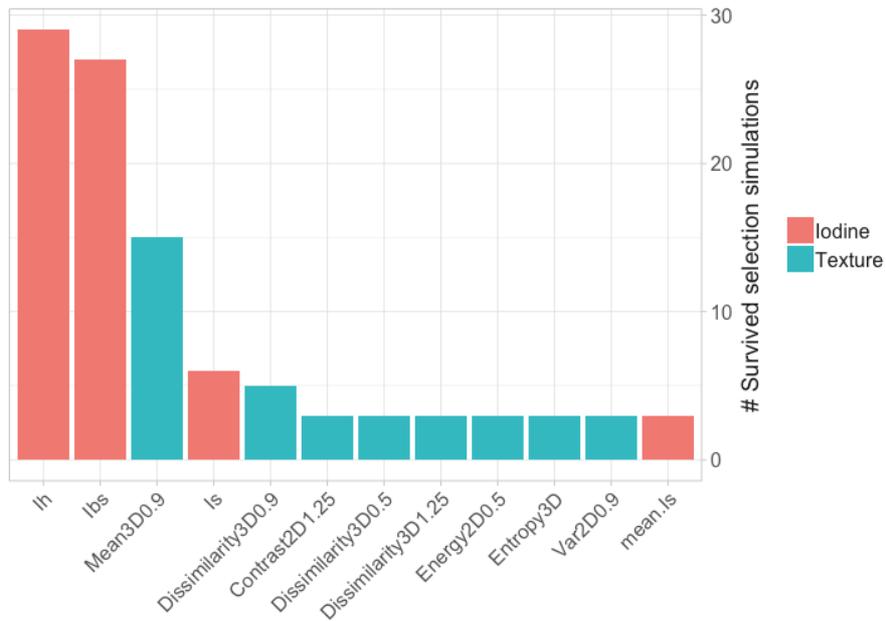


Figure 5.3: Cumulative feature count of survived selection simulations on the training set.

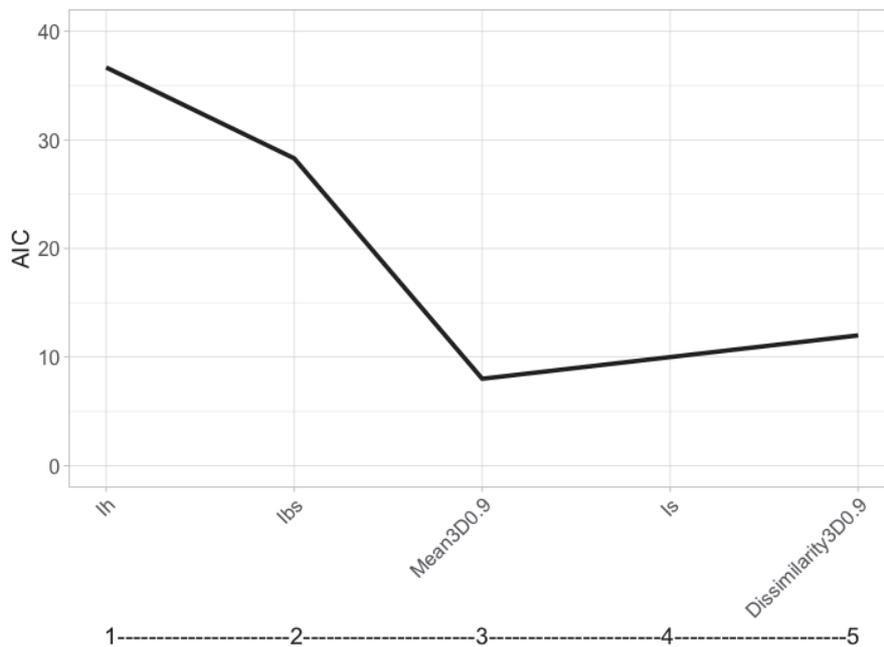


Figure 5.4: Akaike Information Criterion (AIC) for different logistic regression models fit on the training data, with an increasing number of features (from left to right) that survived the feature selection process at least 5 times.

5.3.4 PERFORMANCE ON INDEPENDENT TEST SET: SENSITIVITY ANALYSIS

On the independent test set at baseline (without noise), the I2-score had a sensitivity and specificity of 50% and 87% respectively and the combined *iodine-texture* LR model 43% and 84%. With an increasing noise level applied to the test data, sensitivities of both models increased, while specificities decreased. The results of the noise sensitivity analysis are presented in Figure 5.5.

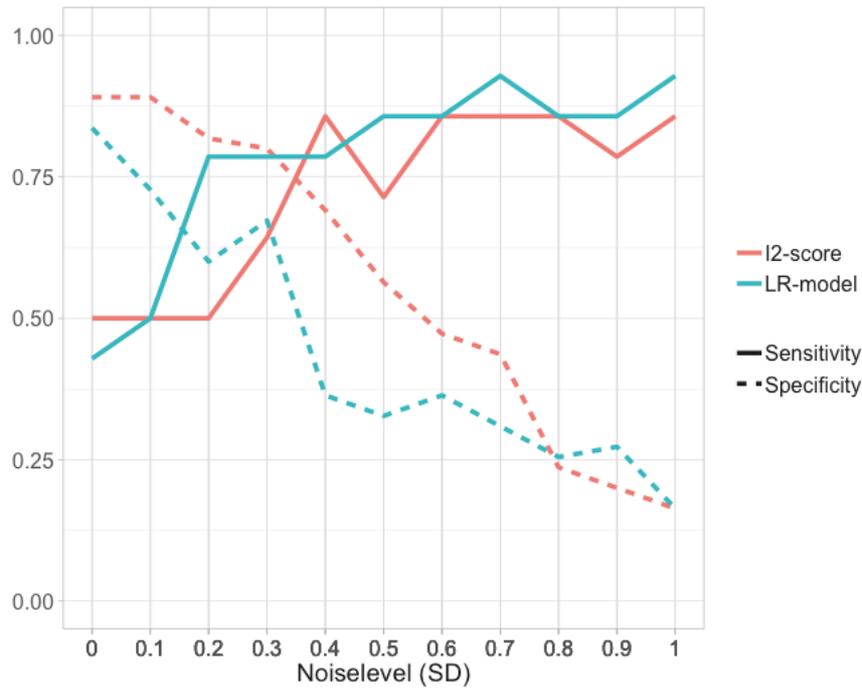


Figure 5.5: Noise sensitivity analysis of I2-score model and combined *iodine-texture* logistic regression (LR) model on the independent test set.

5.4 DISCUSSION

In this study, multiple texture parameters were explored on their predictive ability for HE in combination with DECT iodine features. Extensive feature selection has shown that the iodine features I_h and I_{bs} are the most stable from all assessed features, and especially I_h has shown its predictive value with 28 times been selected as first, and thus as most predictive feature. A combined *iodine-texture* logistic regression model was not able to outperform the recently published I2-score. However, sensitivity analysis of performance on the independent test set demonstrated that nor the I2-score model, nor the combined *iodine-texture* LR model was robust to slight modifications (± 0.2 SD) to the data.

The feature selection and ranking process clearly showed that texture features performed worse than iodine features (especially I_h and I_{bs}), also when different random seeds for were used for the noise generation (not reported in results). Other studies, however, reported positively about the predictive value of texture parameters for HE. Shen et al. reached sensitivities up to 85% for histogram based texture uniformity, while Connor et al. reports significant differences for image histogram standard deviation of the hematoma region.^{132,174} It is known that texture features are in general sensitive to differences in acquisition parameters. The only texture parameters that survived the feature selection process were the texture features after LoG filtering of the scans, and consisted of both image histogram and GLCM based texture features. Apparently, stabilizing texture features by LoG filtering of the data was more important than the exact used texture computation method in this dataset. However, even after LoG filtering, the quantitative texture features were not able to capture the (ongoing) physiologic bleeding process that is described by other qualitative hematoma characteristics on NCCT, such as blend sign and swirl sign.^{169,170} It might be that texture parameters are better able to capture the final result of this bleeding pattern, when hematoma/tissue has readily been reorganized, and thus may be better linked to patient outcome metrics instead of potential or ongoing HE.

Iodine features, however, can provide a better representation of the actual, *ongoing* bleeding process. This process could be divided in initial vessel rupture, leading to focal bleeding, a potential (bright) spot sign (I_{bs}) and initial hematoma growth. Subsequently, secondary vessel leakage, leads more slowly to diffusely accumulated iodine in the hematoma (I_h), where the initial focal bleeding also contributes to the total iodine content in the hematoma. This may explain the superior performance of I_h in the feature selection process. The total iodine in all spots (I_s) was selected only 5/33 times, suggesting that it is mainly the intensity of one bleeding (I_{bs}) that dominates expansion probability rather than a possible multifocal bleeding pattern (I_s). This finding is contradicting the scoring system of spots signs by intensity and number, described in Romero et al.¹¹⁴

It is remarkable that the performance of the combined *iodine-texture* LR model drops considerably from the training set to the test set. Although the AIC penalizes models containing many features, the selected model with 3 features based on the minimal AIC still clearly overfits on the training data (100% sensitivity and specificity). Ideally, the number for model features should be far less than the number of positive cases in the training data to avoid overfitting.¹³⁴ The number of expanders in the used training dataset is only 13, which might be too few for a model based on 3 features. Fitting a model on a training dataset containing more positive cases would potentially overcome this problem and let the results on the training set be more generalizable to the test set. A different cause of the non-generalizability may be that other acquisition settings than delay time (only recorded), such as contrast injection rate, timing of the early phase CTA or used reconstruction kernels, are significantly different between the training and the test set.

Other interesting findings are that also the performance of the I2-score drops from the training set to the test set (sens/spec 85%/90% to 50%/87%), and that its performance on the test set is remarkably worse compared to Tan et al. (sens/spec 71%/93%), while having used almost the same database. An explanation for the difference in performance could be that this study required a redo of the hematoma segmentations. Since there was no focus on texture analysis in Tan et al., VNC images from DECT were generated as (gray) RGB images rather than in HUs and were only used for hematoma segmentation purposes. However, it was not possible to use these RGB images for registration to the conventional NCCT, that was meant to be used for texture analysis, requiring VNC images to be regenerated using syngo.via. Because the user has to define a boxed field-of-view of the 3D DECT volume to be converted to VNC, the original VNC images could not be replicated, and thus a redo of the segmentations was required. Therefore, used hematoma masks in this study were different compared to Tan et al. and could have caused the different performance on the test set. This fits with the fact that the performed sensitivity analysis also showed that the I2-score is sensitive to slight data modifications, indicating that this I_h/I_{bs} model might have been fit on a too small training dataset and needs some recalibration on a larger data set. Moreover, the accuracy of the used semi-automatic segmentation algorithm (based on max-flow/min-cut 3D region growing) has never been validated on intracerebral hematomas, although the results look good on visual inspection. Studies on larger cohorts are required to validate the initial reported performance of the I2-score.

This study has several limitations. First, data came from only one DECT scanner type from one hospital. Although this works as an advantage for the performance of this study, generalizability of the identified features compared to other centers and scanner types is limited, and requires studies with larger cohorts, multiple centers and DECT scanner types as well. Moreover, the relation between these quantitative image features and clinical outcome measures, such as modified Rankin scale (mRS) or mortality has not been assessed. Third, the used prediction models in this study only used quantitative radiographic features and did not include clinical parameters such as anticoagulant usage, time from symptom onset to CT scan, blood pressure, National Institute of Health Stroke Scale (NIHSS), Glasgow Coma Scale (GCS) or other features.¹⁷⁶

Future research should focus on validating the I2-score by standardizing image acquisition, studying larger cohorts and selecting patients more strictly on solely intraparenchymal hemorrhages without intraventricular or subarachnoid component. Further, incorporating clinical parameters in the model may increase model performance. Another interesting but different approach to reveal the potential synergy between image features from NCCT and DECT iodine images is to develop a deep learning system that takes segmented hematoma regions of both modalities as input to learn combined predictive features, and use these to classify hematomas as expanders or non-expanders.

5.5 CONCLUSION

This study has shown that a logistic regression model combining NCCT texture features and DECT iodine features was not able to outperform recently proposed I2-score on an independent test set in prediction hematoma expansion. Moreover, performances of both the I2-score model and combined *iodine-texture* model were not robust to slight data modifications. Although the results from an extensive feature selection process clearly showed that DECT iodine features have more predictive value than NCCT texture features to, future research should validate and/or update the I2-score model based on a larger,

strictly selected study cohort using standardized image acquisition protocols, to make its performance generalizable to other datasets.

6

Conclusion

The research performed for this thesis has attempted to develop models able to assist in clinical triage of hemorrhagic and acute ischemic stroke patients by identifying subgroups relevant for therapy selection. Both studies were based on analysis of CT scans, which is the most relevant and widespread imaging modality used in emergency radiology. By acquiring both non-contrast CT images, early and delayed (dual-energy) contrast-enhanced phases, these scans are able to capture relevant underlying vascular physiology in the brain leading to intracerebral hemorrhage expansion or development of necrotic tissue in ischemic brain areas, referred as the ischemic core.

Identifying the extent of the ischemic core is important to select AIS patients eligible for EVT. This study has shown that a deep learning architecture based on a 2D DeepMedic architecture, taking SECT image patches from above described 3-phase CT as input, performed better than other U-Net based architectures in predicting ground truth DWI ischemic core lesions. Moreover, the predicted infarct core lesion volumes on individual patient basis led to low incorrect EVT decision rates, implying that the system might be useful for clinically triaging AIS patients. Standardizing CT acquisitions, including clinical parameters to the prediction model and adding CTA phases, enriching the information on the underlying vascular stroke physiology, may be required to further improve the performance of this deep learning system. Future work should focus on these aspects, for example by finding a tradeoff between required CTA acquisitions (radiation and contrast dose) and on the implications of the predicted lesion volumes in clinical practice.

To predict hematoma expansion in hemorrhagic stroke patients, quantitative texture features based on NCCT images have been identified and combined with DECT iodine features in the development of a logistic regression model. The results have shown that a logistic regression model combining iodine and texture features was not able to outperform the recently proposed I2-score on an independent test set. Results from an extensive feature selection process have shown that DECT iodine features have more predictive value than texture features from NCCT for predicting hematoma expansion. Future research should validate and/or update the I2-score model based on a larger, strictly selected study cohort using standardized image acquisition protocols, to make its performance generalizable to other datasets.

As more and more medical imaging data is stored and annotated in a structured way, more quantitative imaging features can be extracted and/or learned by self-learning systems, to be used in applications that are able to aid in daily clinical practice. Today, we are just at the beginning of the computer-aided diagnosis revolution in healthcare.

References

1. Benjamin EJ, Virani SS, Callaway CW, et al. Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. *Circulation* 2018 Jan 1. [Epub ahead of print].
2. Murray CJL, Barber RM, Foreman KJ, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: Quantifying the epidemiological transition. *Lancet* 2015;386:2145–91.
3. Jovin TG, Demchuk AM, Gupta R. Pathophysiology of Acute Ischemic Stroke. *Contin Lifelong Learn Neurol* 2008;14:28–45.
4. Baron J-C. Perfusion Thresholds in Human Cerebral Ischemia: Historical Perspective and Therapeutic Implications. *Cerebrovasc Dis* 2001;11:2–8.
5. Smith WS, Lev MH, English JD, et al. Significance of large vessel intracranial occlusion causing acute ischemic stroke and tia. *Stroke* 2009;40:3834–40.
6. Alexandrov A V., Sharma VK, Lao AY, et al. Reversed Robin Hood syndrome in acute ischemic stroke patients. *Stroke* 2007;38:3045–8.
7. Guo Y, Li P, Guo Q, et al. Pathophysiology and Biomarkers in Acute Ischemic Stroke – A Review. *Trop J Pharm Res* 2014;12:1097.
8. Hasan TF, Rabinstein AA, Middlebrooks EH, et al. Diagnosis and Management of Acute Ischemic Stroke. *Mayo Clin Proc* 2018;93:523–38.
9. Yoo AJ, Chaudhry ZA, Nogueira RG, et al. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. *Stroke* 2012;43:1323–30.
10. Zaidi SF, Aghaebrahim A, Urra X, et al. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. *Stroke* 2012;43:3238–44.
11. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue Plasminogen Activator for Acute Ischemic Stroke. *N Engl J Med* 1995;333:1581–8.
12. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med* 2008;359:1317–29.
13. Wahlgren N, Ahmed N, Dávalos A, et al. Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke–Monitoring Study (SITS–MOST): an observational study. *Lancet* 2007;369:275–82.
14. Wardlaw JM, Murray V, Berge E, et al. Thrombolysis for acute ischaemic stroke. *Cochrane database Syst Rev* 2014:CD000213.
15. Lev MH, Bergmans RHJ, Kuppens D, et al. Neuro Course 2018 – Dual-Energy CT for Clinical Triage of Hemorrhagic and Ischemic Stroke. 2018. [Epub ahead of print].
16. Seet RCS, Rabinstein AA. Symptomatic intracranial hemorrhage following intravenous thrombolysis for acute ischemic stroke: A critical review of case definitions. *Cerebrovasc Dis* 2012;34:106–14.
17. Lansberg MG, Thijs VN, Bammer R, et al. Risk factors of symptomatic intracerebral hemorrhage after tPA therapy for acute stroke. *Stroke* 2007;38:2275–8.
18. Singer OC, Humpich MC, Fiehler J, et al. Risk for symptomatic intracerebral hemorrhage after thrombolysis assessed by diffusion-weighted magnetic resonance imaging. *Ann Neurol* 2008;63:52–60.
19. del Zoppo GJ, Poeck K, Pessin MS, et al. Recombinant tissue plasminogen activator in acute thrombotic and embolic stroke. *Ann Neurol* 1992;32:78–86.
20. Kleindorfer D, Lindsell CJ, Brass L, et al. National US estimates of recombinant tissue plasminogen activator use: ICD-9 codes substantially underestimate. *Stroke* 2008;39:924–8.
21. Bhatia R, Hill MD, Shobha N, et al. Low rates of acute recanalization with intravenous recombinant tissue plasminogen activator in ischemic stroke: Real-world experience and a call for action. *Stroke* 2010;41:2254–8.

22. Zivin JA. Acute stroke therapy with tissue plasminogen activator (tPA) since it was approved by the U.S. Food and Drug Administration (FDA). *Ann Neurol* 2009;66:6–10.
23. Jauch EC, Saver JL, Adams HP, et al. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2013;44:870–947.
24. del Zoppo GJ, Higashida RT, Furlan AJ, et al. PROACT: A Phase II Randomized Trial of Recombinant Pro-Urokinase by Direct Arterial Delivery in Acute Middle Cerebral Artery Stroke. *Stroke* 1998;29:4–11.
25. Hill MD, Rowley HA, Adler F, et al. Selection of acute ischemic stroke patients for intra-arterial thrombolysis with pro-urokinase by using ASPECTS. *Stroke* 2003;34:1925–31.
26. Hill MD, Demchuk AM, Tomsick TA, et al. Using the baseline CT scan to select acute stroke patients for IV-IA therapy. *Am J Neuroradiol* 2006;27:1612–6.
27. Yoo AJ, Verduzco LA, Schaefer PW, et al. MRI-based selection for intra-arterial stroke therapy: Value of pretreatment diffusion-weighted imaging lesion volume in selecting patients with acute stroke who will benefit from early recanalization. *Stroke* 2009;40:2046–54.
28. Lansberg MG, Straka M, Kemp S, et al. Magnetic Resonance Imaging Profile and Response to Endovascular Reperfusion: Results of the DEFUSE 2 Prospective Cohort Study. *Lancet Neurol* 2012;11:860–7.
29. Broderick JP, Palesch YY, Demchuk AM, et al. Endovascular Therapy after Intravenous t-PA versus t-PA Alone for Stroke. *N Engl J Med* 2013;368:893–903.
30. Ciccone A, Valvassori L, Nichelatti M, et al. Endovascular Treatment for Acute Ischemic Stroke. *N Engl J Med* 2013;368:904–13.
31. Kidwell CS, Jahan R, Gornbein J, et al. A Trial of Imaging Selection and Endovascular Treatment for Ischemic Stroke. *N Engl J Med* 2013;368:914–23.
32. Berkhemer OA, Fransen PSS, Beumer D, et al. A Randomized Trial of Intraarterial Treatment for Acute Ischemic Stroke. *N Engl J Med* 2015;372:11–20.
33. Goyal M, Demchuk AM, Menon BK, et al. Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke. *N Engl J Med* 2015;372:1019–30.
34. Campbell BCV, Mitchell PJ, Kleinig TJ, et al. Endovascular Therapy for Ischemic Stroke with Perfusion-Imaging Selection. *N Engl J Med* 2015;372:1009–18.
35. Saver JL, Goyal M, Bonafe A, et al. Stent-Retriever Thrombectomy after Intravenous t-PA vs. t-PA Alone in Stroke. *N Engl J Med* 2015;372:2285–95.
36. Saver JL, Goyal M, Van Der Lugt A, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: Ameta-analysis. *JAMA - J Am Med Assoc* 2016;316:1279–88.
37. Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *N Engl J Med* <https://doi.org/10.1056/NEJMoa1706442>.
38. Albers GW, Marks MP, Kemp S, et al. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med* 2018;378:708–18.
39. Wintermark M, Rowley HA, Lev MH. Acute stroke triage to intravenous thrombolysis and other therapies with advanced CT or MR imaging: Pro CT. *Radiology* 2009;251:619–26.
40. Latchaw RE, Alberts MJ, Lev MH, et al. Recommendations for imaging of acute ischemic stroke: A scientific statement from the American Heart Association. *Stroke* 2009;40:3646–78.
41. Torres-Mozqueda F, He J, Yeh IB, et al. An acute ischemic stroke classification instrument that includes CT or MR angiography: The Boston acute stroke imaging scale. *Am J Neuroradiol* 2008;29:1111–7.
42. González RG, Lev MH, Goldmacher G V., et al. Improved outcome prediction using CT angiography in addition to standard ischemic stroke assessment: Results from the STOPStroke study. *PLoS One* 2012;7:7–12.
43. Pexman JHW, Barber PA, Hill MD, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke. *Am J Neuroradiol* 2001;22:1534 LP–1542.
44. Bal S, Bhatia R, Menon BK, et al. Time dependence of reliability of noncontrast computed tomography in comparison to computed tomography angiography source image in acute ischemic stroke. *Int J Stroke* 2015;10:55–60.
45. Wannamaker R, Guinand T, Menon BK, et al. Computed Tomographic Perfusion Predicts Poor Outcomes in a Randomized Trial of Endovascular Therapy. *Stroke* 2018;49:1426–33.
46. Wintermark M, Flanders AE, Velthuis B, et al. Perfusion-CT assessment of infarct core and penumbra: Receiver operating characteristic curve analysis in 130 patients suspected of acute

- hemispheric stroke. *Stroke* 2006;37:979–85.
47. Lansberg MG, Christensen S, Kemp S, et al. Computed tomographic perfusion to Predict Response to Recanalization in ischemic stroke. *Ann Neurol* 2017;81:849–56.
 48. Hassan AE, Zacharatos H, Rodriguez GJ, et al. A comparison of computed tomography perfusion-guided and time-guided endovascular treatments for patients with acute ischemic stroke. *Stroke* 2010;41:1673–8.
 49. Sheth KN, Terry JB, Nogueira RG, et al. Advanced modality imaging evaluation in acute ischemic stroke may lead to delayed endovascular reperfusion therapy without improvement in clinical outcomes. *J Neurointerv Surg* 2013;5:62–6.
 50. Borst J, Berkhemer OA, Roos YB, et al. Value of Computed Tomographic Perfusion-Based Patient Selection for Intra-Arterial Acute Ischemic Stroke Treatment. *Stroke* 2015;46:3375–82.
 51. Heit JJ, Wintermark M. Perfusion computed tomography for the evaluation of acute ischemic stroke strengths and pitfalls. *Stroke* 2016;47:1153–8.
 52. Kamalian S, Kamalian S, Maas MB, et al. CT cerebral blood flow maps optimally correlate with admission diffusion-weighted imaging in acute stroke but thresholds vary by postprocessing platform. *Stroke* 2011;42:1923–8.
 53. Haussen DC, Dehkharghani S, Rangaraju S, et al. Automated CT Perfusion Ischemic Core Volume and Noncontrast CT ASPECTS (Alberta Stroke Program Early CT Score): Correlation and Clinical Outcome Prediction in Large Vessel Stroke. *Stroke* 2016;47:2318–22.
 54. González RG, Copen WA, Schaefer PW, et al. The Massachusetts General Hospital acute stroke imaging algorithm: an experience and evidence based approach. *J Neurointerv Surg* 2013;5:i7–12.
 55. Menon BK, Goyal M. Imaging Paradigms in Acute Ischemic Stroke: A Pragmatic Evidence-based Approach. *Radiology* 2015;277:7–12.
 56. Goyal M, Menon BK, Derdeyn CP. Perfusion Imaging in Acute Ischemic Stroke: Let Us Improve the Science before Changing Clinical Practice. *Radiology* 2013;266:16–21.
 57. Campbell BC V, Yassi N, Ma H, et al. Imaging selection in ischemic stroke: Feasibility of automated CT-perfusion analysis. *Int J Stroke* 2015;10:51–4.
 58. Fiebach JB, Schellinger PD, Jansen O, et al. CT and diffusion-weighted MR imaging in randomized order: Diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. *Stroke* 2002;33:2206–10.
 59. Schaefer PW, Hunter GJ, He J, et al. Predicting cerebral ischemic infarct volume with diffusion and perfusion MR imaging. *AJNR Am J Neuroradiol* 2002;23:1785–94.
 60. Guadagno J V., Warburton EA, Aigbirhio FI, et al. Does the acute diffusion-weighted imaging lesion represent penumbra as well as core? A combined quantitative PET/MRI voxel-based study. *J Cereb Blood Flow Metab* 2004;24:1249–54.
 61. Guadagno J V., Warburton EA, Jones PS, et al. The diffusion-weighted lesion in acute stroke: Heterogeneous patterns of flow/metabolism uncoupling as assessed by quantitative positron emission tomography. *Cerebrovasc Dis* 2005;19:239–46.
 62. Sobesky J, Weber OZ, Lehnhardt FG, et al. Does the mismatch match the penumbra? Magnetic resonance imaging and positron emission tomography in early ischemic stroke. *Stroke* 2005;36:980–5.
 63. Schaefer PW, Souza L, Kamalian S, et al. Limited reliability of computed tomographic perfusion acute infarct volume measurements compared with diffusion-weighted imaging in anterior circulation stroke. *Stroke* 2015;46:419–24.
 64. Copen WA, Yoo AJ, Rost NS, et al. In patients with suspected acute stroke, CT perfusion-based cerebral blood flow maps cannot substitute for DWI in measuring the ischemic core. *PLoS One* 2017;12:1–20.
 65. Purushotham A, Campbell BC V, Straka M, et al. Apparent diffusion coefficient threshold for delineation of ischemic core. *Int J Stroke* 2015;10:348–53.
 66. Abbott AL, Silvestrini M, Topakian R, et al. Optimizing the definitions of stroke, transient ischemic attack, and infarction for research and application in clinical practice. *Front Neurol* 2017;8:1–14.
 67. Leslie-Mazwi TM, Hirsch JA, Falcone GJ, et al. Endovascular stroke treatment outcomes after patient selection based on magnetic resonance imaging and clinical criteria. *JAMA Neurol* 2016;73:43–9.
 68. Liebeskind DS. Collateral circulation. *Stroke* 2003;34:2279–84.
 69. Liebeskind DS. Collateral perfusion: Time for novel paradigms in cerebral ischemia. *Int J Stroke* 2012;7:309–10.

70. Elijevich L, Goyal N, Mainali S, et al. CTA collateral score predicts infarct volume and clinical outcome after endovascular therapy for acute ischemic stroke: A retrospective chart review. *J Neurointerv Surg* 2016;8:559–62.
71. Miteff F, Levi CR, Bateman GA, et al. The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke. *Brain* 2009;132:2231–8.
72. Kucinski T, Koch C, Eckert B, et al. Collateral circulation is an independent radiological predictor of outcome after thrombolysis in acute ischaemic stroke. *Neuroradiology* 2003;45:11–8.
73. Christoforidis GA, Mohammad Y, Kehagias D, et al. Angiographic assessment of pial collaterals as a prognostic indicator following intra-arterial thrombolysis for acute ischemic stroke. *Am J Neuroradiol* 2005;26:1789–97.
74. Bang OY, Saver JL, Kim SJ, et al. Collateral Flow Predicts Response to Endovascular Therapy for Acute Ischemic Stroke. *Stroke* 2011;42:693–9.
75. Nambiar V, Sohn SI, Almekhlafi MA, et al. CTA collateral status and response to recanalization in patients with acute ischemic stroke. *Am J Neuroradiol* 2014;39:884–90.
76. Higazi MM, Abdel-Gawad EA. CT angiography collateral scoring: Correlation with DWI infarct size in proximal middle cerebral artery occlusion stroke within 12 h onset. *Egypt J Radiol Nucl Med* 2016;47:991–7.
77. Mehrkhani F, Berkhemer OA, Majoie CBLM, et al. Combined Evaluation of Noncontrast CT ASPECTS and CT Angiography Collaterals Improves Detection of Large Infarcts in Proximal Artery Occlusive Stroke. *J Neuroimaging* 2018;28:524–9.
78. Raymond SB, Schaefer PW. Imaging Brain Collaterals: Quantification, Scoring, and Potential Significance. *Top Magn Reson Imaging* 2017;26:67–75.
79. Menon BK, d’Esterre CD, Qazi EM, et al. Multiphase CT Angiography: A New Tool for the Imaging Triage of Patients with Acute Ischemic Stroke. *Radiology* 2015;275:510–20.
80. Souza LCS, Yoo AJ, Chaudhry ZA, et al. Malignant CTA Collateral Profile Is Highly Specific for Large Admission DWI Infarct Core and Poor Outcome in Acute Stroke. *Am J Neuroradiol* 2012;33:1331–6.
81. Maas MB, Lev MH, Ay H, et al. Collateral vessels on CT angiography predict outcome in acute ischemic stroke. *Stroke* 2009;40:3001–5.
82. D’Esterre CD, Trivedi A, Pordeli P, et al. Regional Comparison of Multiphase Computed Tomographic Angiography and Computed Tomographic Perfusion for Prediction of Tissue Fate in Ischemic Stroke. *Stroke* 2017;48:939–45.
83. Beyer SE, Thierfelder KM, Von Baumgarten L, et al. Strategies of collateral blood flow assessment in ischemic stroke: Prediction of the follow-up infarct volume in conventional and dynamic CTA. *Am J Neuroradiol* 2015;36:488–94.
84. Van Den Wijngaard IR, Holswilder G, Wermer MJH, et al. Assessment of collateral status by dynamic ct angiography in acute mca stroke: Timing of acquisition and relationship with final infarct volume. *Am J Neuroradiol* 2016;37:1231–6.
85. Smit EJ, Vonken EJ, Van Seeters T, et al. Timing-invariant imaging of collateral vessels in acute ischemic stroke. *Stroke* 2013;44:2194–9.
86. McVerry F, Liebeskind DS, Muir KW. Systematic review of methods for assessing leptomeningeal collateral flow. *AJNR Am J Neuroradiol* 2012;33:576–82.
87. Boers AMM, Sales Barros R, Jansen IGH, et al. Value of Quantitative Collateral Scoring on CT Angiography in Patients with Acute Ischemic Stroke. *AJNR Am J Neuroradiol* 2018;39:1074–82.
88. Campbell BC V, Purushotham A, Christensen S, et al. The infarct core is well represented by the acute diffusion lesion: Sustained reversal is infrequent. *J Cereb Blood Flow Metab* 2012;32:50–6.
89. Gupta R, Phan CM, Leidecker C, et al. Evaluation of dual-energy CT for differentiating intracerebral hemorrhage from iodinated contrast material staining. *Radiology* 2010;257:205–11.
90. Phan CM, Yoo AJ, Hirsch JA, et al. Differentiation of hemorrhage from iodinated contrast in different intracranial compartments using dual-energy head CT. *Am J Neuroradiol* 2012;33:1088–94.
91. Skornitzke S, Fritz F, Mayer P, et al. Dual-energy CT iodine maps as an alternative quantitative imaging biomarker to abdominal CT perfusion: determination of appropriate trigger delays for acquisition using bolus tracking. *Br J Radiol* 2018;91:20170351.
92. Flottmann F, Broocks G, Faizy TD, et al. CT-perfusion stroke imaging: A threshold free probabilistic approach to predict infarct volume compared to traditional ischemic thresholds. *Sci Rep* 2017;7:1–10.

93. Poon MTC, Bell SM, Al-Shahi Salman R. Epidemiology of Intracerebral Haemorrhage. *Front Neurol Neurosci* 2015;37:1–12.
94. Rincon F, Mayer SA. The epidemiology of intracerebral hemorrhage in the United States from 1979 to 2008. *Neurocrit Care* 2013;19:95–102.
95. Ferro JM. Update on intracerebral haemorrhage. *J Neurol* 2006;253:985–99.
96. Brott T, Thalinger K, Hertzberg V. Hypertension as a risk factor for spontaneous intracerebral hemorrhage. *Stroke* 1986;17:1078–83.
97. Thrift AG, McNeil JJ, Forbes A, et al. Three important subgroups of hypertensive persons at greater risk of intracerebral hemorrhage. *Hypertension* 1998;31:1223–9.
98. Silva Y, Leira R, Tejada J, et al. Molecular Signatures of Vascular Injury Are Associated With Early Growth of Intracerebral Hemorrhage. *Stroke* 2005;36:86–91.
99. Leasure A, Kimberly WT, Sansing LH, et al. Treatment of Edema Associated With Intracerebral Hemorrhage. *Curr Treat Options Neurol* 2016;18:1–14.
100. Brott T, Broderick J, Kothari R, et al. Early Hemorrhage Growth in Patients With Intracerebral Hemorrhage. *Stroke* 1997;28:1 LP–5.
101. Davis SM, Broderick J, Hennerici M, et al. Hematoma growth is a determinant of mortality and poor outcome after intracerebral hemorrhage. *Neurology* 2006;66:1175–81.
102. Mayer SA, Brun NC, Begtrup K, et al. Recombinant Activated Factor VII for Acute Intracerebral Hemorrhage. *N Engl J Med* 2005;352:777–85.
103. Mayer SA, Brun NC, Begtrup K, et al. Efficacy and safety of recombinant activated factor VII for acute intracerebral hemorrhage. *N Engl J Med* 2008;358:2127–37.
104. Anderson CS, Heeley E, Huang Y, et al. Rapid Blood-Pressure Lowering in Patients with Acute Intracerebral Hemorrhage. *N Engl J Med* 2013;368:2355–65.
105. Qureshi AI, Palesch YY, Barsan WG, et al. Intensive Blood-Pressure Lowering in Patients with Acute Cerebral Hemorrhage. *N Engl J Med* 2016;375:1033–43.
106. Mendelow AD, Gregson BA, Fernandes HM, et al. 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. *Lancet* 2005;365:387–97.
107. Fung C, Murek M, Z'Graggen WJ, et al. Decompressive hemicraniectomy in patients with supratentorial intracerebral hemorrhage. *Stroke* 2012;43:3207–11.
108. Yao Z, Ma L, You C, et al. Decompressive Craniectomy for Spontaneous Intracerebral Hemorrhage: A Systematic Review and Meta-analysis. *World Neurosurg* 2018;110:121–8.
109. Fam MD, Hanley D, Stadnik A, et al. Surgical Performance in Minimally Invasive Surgery Plus Recombinant Tissue Plasminogen Activator for Intracerebral Hemorrhage Evacuation Phase III Clinical Trial. *Neurosurgery* 2017;00:1–7.
110. Murai Y, Takagi R, Ikeda Y, et al. Three-dimensional computerized tomography angiography in patients with hyperacute intracerebral hemorrhage. *J Neurosurg* 1999;91:424–31.
111. Delgado Almandoz JE, Yoo AJ, Stone MJ, et al. Systematic characterization of the computed tomography angiography spot sign in primary intracerebral hemorrhage identifies patients at highest risk for hematoma expansion: the spot sign score. *Stroke* 2009;40:2994–3000.
112. Del Giudice A, D'Amico D, Sobesky J, et al. Accuracy of the spot sign on computed tomography angiography as a predictor of haematoma enlargement after acute spontaneous intracerebral haemorrhage: A systematic review. *Cerebrovasc Dis* 2014;37:268–76.
113. Du FZ, Jiang R, Gu M, et al. The accuracy of spot sign in predicting hematoma expansion after intracerebral hemorrhage: A systematic review and meta-analysis. *PLoS One* 2014;9:1–16.
114. Romero JM, Bart Brouwers H, Lu J, et al. Prospective validation of the computed tomographic angiography spot sign score for Intracerebral hemorrhage. *Stroke* 2013;44:3097–102.
115. Rizos T, Dörner N, Jenetzky E, et al. Spot signs in intracerebral hemorrhage: Useful for identifying patients at risk for hematoma enlargement? *Cerebrovasc Dis* 2013;35:582–9.
116. Demchuk AM, Dowlatshahi D, Rodriguez-Luna D, et al. Prediction of haematoma growth and outcome in patients with intracerebral haemorrhage using the CT-angiography spot sign (PREDICT): A prospective observational study. *Lancet Neurol* 2012;11:307–14.
117. Ederies A, Demchuk A, Chia T, et al. Postcontrast CT extravasation is associated with hematoma expansion in cta spot negative patients. *Stroke* 2009;40:1672–6.
118. Chakraborty S, Blacchiere D, Lum C, et al. Dynamic nature of the CT angiographic “spot sign.” *Br J Radiol* 2010;83:216–20.
119. Mericle RA, Lopes DK, Fronckowiak MD, et al. A grading scale to predict outcomes after intra-arterial thrombolysis for stroke complicated by contrast extravasation. *Neurosurgery*

- 2000;46:1305–7.
120. Greer DM, Koroshetz WJ, Cullen S, et al. Magnetic Resonance Imaging Improves Detection of Intracerebral Hemorrhage over Computed Tomography after Intra-Arterial Thrombolysis. *Stroke* 2004;35:491–5.
 121. Wang B, Yan S, Xu M, et al. Timing of Occurrence Is the Most Important Characteristic of Spot Sign. *Stroke* 2016;47:1233–8.
 122. Orito K, Hirohata M, Nakamura Y, et al. Leakage sign for primary intracerebral hemorrhage a novel predictor of hematoma growth. *Stroke* 2016;47:958–63.
 123. Almandoz JED, Yoo AJ, Stone MJ, et al. The spot sign score in primary intracerebral hemorrhage identifies patients at highest risk of in-hospital mortality and poor outcome among survivors. *Stroke* 2010;41:54–60.
 124. Tan CO, Lam S, Kuppens D, et al. Spot and diffuse signs: Quantitative markers of intracranial hematoma expansion at dual-energy CT. *Radiology* 2019;290.
 125. Hardie K, Jamrozik K, Hankey GJ, et al. Trends in five-year survival and risk of recurrent stroke after first-ever stroke in the Perth Community Stroke Study. *Cerebrovasc Dis* 2005;19:179–85.
 126. Vermeer SE, Algra A, Franke CL, et al. Long-term prognosis after recovery from primary intracerebral hemorrhage. *Neurology* 2002;59:205 LP–209.
 127. McDonald RJ, Schwartz KM, Eckel LJ, et al. and Technological Advancements of Cross-Sectional Imaging on Radiologist Workload. *Acad Radiol* 2010;22:1191–8.
 128. Akdeniz EG. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017;19:221–48.
 129. Liew C. The future of radiology augmented with Artificial Intelligence: A strategy for success. *Eur J Radiol* 2018;102:152–6.
 130. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–77.
 131. Lubner MG, Smith AD, Sandrasegaran K, et al. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *Radiographics* 2017;37:1483–503.
 132. Shen Q, Shan Y, Hu Z, et al. Quantitative parameters of CT texture analysis as potential markers for early prediction of spontaneous intracranial hemorrhage enlargement. *Eur Radiol* 2018;28:4389–96.
 133. Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. *Clin Radiol* 2004;59:1061–9.
 134. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *AJNR Am J Neuroradiol* 2010;31:809–16.
 135. Mao WHNE-Y. Texture Analysis Methods for Medical Image Characterisation. In: Rijeka: IntechOpen; 2010:Ch. 4.
 136. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
 137. Bengio Y. Learning Deep Architectures for AI. *Found Trends@ Mach Learn* 2009;2:199–200.
 138. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.1007/978-3-319-24574-4_28.
 139. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–324.
 140. Krizhevsky A. *Learning Multiple Layers of Features from Tiny Images.*; 2012.
 141. Deng J, Dong W, Socher R, et al. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09.*; 2009.
 142. Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimed Tools Appl* <https://doi.org/10.1007/s11042-017-4440-4>.
 143. Huang G, Liu Z, van der Maaten L, et al. Densely Connected Convolutional Networks. <https://doi.org/10.1109/CVPR.2017.243>.
 144. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2015;07–12–June:1–9.
 145. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision. <https://doi.org/10.1109/CVPR.2016.308>.
 146. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017* 2017;2017–Janua:1800–7.
 147. Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017* 2017;2017–Janua:5987–95.
 148. Hinton G. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. 2014;15:1929–58.

149. Ioffe S, Szegedy C. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift.
150. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61–78.
151. Kuppens D. Machine Learning for CT-based clinical triage in hemorrhagic and ischemic stroke. 2018 Sep. [Epub ahead of print].
152. Muschelli J, Ullman NL, Mould WA, et al. Validated automatic brain extraction of head CT images. *Neuroimage* 2015;114:379–85.
153. Jenkinson M, Beckmann CF, Behrens TEJ, et al. FSL. *Neuroimage* 2012;62:782–90.
154. Collignon, Maes F, Delaere D, et al. Automated multi-modality image registration based on information theory. *Imaging* 1995;3:263–74.
155. Aitken A, Ledig C, Theis L, et al. Checkerboard artifact free sub-pixel convolution.pdf.
156. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. <https://doi.org/http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>.
157. He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proc IEEE Int Conf Comput Vis* 2015;2015 Inter:1026–34.
158. Cereda CW, Christensen S, Campbell BCV, et al. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *J Cereb Blood Flow Metab* 2016;36:1780–9.
159. Dehkharghani S, Bammer R, Straka M, et al. Performance of CT ASPECTS and collateral score in risk stratification: Can target perfusion profiles be predicted without perfusion imaging? *Am J Neuroradiol* 2016;37:1399–404.
160. Mayer SA. Recombinant activated factor VII for acute intracerebral hemorrhage. *Stroke* 2007;38:763–7.
161. Balami JS, Buchan AM. Complications of intracerebral haemorrhage. *Lancet Neurol* 2012;11:101–18.
162. Wartenberg KE, Mayer SA. Ultra-Early Hemostatic Therapy for Intracerebral Hemorrhage: Future Directions. *Front Neurol Neurosci* 2015;37:107–29.
163. Yogendrakumar V, Demchuk AM, Aviv RI, et al. Location of intracerebral haemorrhage predicts haematoma expansion. *Eur Stroke J* 2017;2:257–63.
164. Barras CD, Tress BM, Christensen S, et al. Density and shape as CT predictors of intracerebral hemorrhage growth. *Stroke* 2009;40:1325–31.
165. Morotti A, Boulouis G, Romero JM, et al. Blood pressure reduction and noncontrast CT markers of intracerebral hemorrhage expansion. *Neurology* 2017;89:548–54.
166. Li Q, Liu QJ, Yang WS, et al. Island sign: An imaging predictor for early hematoma expansion and poor outcome in patients with intracerebral hemorrhage. *Stroke* 2017;48:3019–25.
167. Yu Z, Zheng J, Ali H, et al. Significance of satellite sign and spot sign in predicting hematoma expansion in spontaneous intracerebral hemorrhage. *Clin Neurol Neurosurg* 2017;162:67–71.
168. Yu Z, Zheng J, Ma L, et al. The predictive accuracy of the black hole sign and the spot sign for hematoma expansion in patients with spontaneous intracerebral hemorrhage. *Neurol Sci* 2017;38:1591–7.
169. Li Q, Yang W-S, Wang X-C, et al. Blend sign predicts poor outcome in patients with intracerebral hemorrhage. *PLoS One* 2017;12:e0183082.
170. Selariu E, Zia E, Brizzi M, et al. Swirl sign in intracerebral haemorrhage: definition, prevalence, reliability and prognostic value. *BMC Neurol* 2012;12.
171. Blacquiere D, Demchuk AM, Al-Hazzaa M, et al. Intracerebral Hematoma Morphologic Appearance on Noncontrast Computed Tomography Predicts Significant Hematoma Expansion. *Stroke* 2015;46:3111–6.
172. Boulouis G, Morotti A, Bart Brouwers H, et al. Association between hypodensities detected by computed tomography and hematoma expansion in patients with intracerebral hemorrhage. *JAMA Neurol* 2016;73:961–8.
173. Wada R, Aviv RI, Fox AJ, et al. CT angiography “spot sign” predicts hematoma expansion in acute intracerebral hemorrhage. *Stroke* 2007;38:1257–62.
174. Connor D, Huynh TJ, Demchuk AM, et al. Swirls and spots: relationship between qualitative and quantitative hematoma heterogeneity, hematoma expansion, and the spot sign. *Neurovascular Imaging* 2015;1:8.
175. Tan CO, Taylor JA, Ler ASH, et al. Detection of multifiber neuronal firings: A mixture separation model applied to sympathetic recordings. *IEEE Trans Biomed Eng* 2009;56:147–58.
176. Yaghi S, Dibu J, Achi E, et al. Hematoma expansion in spontaneous intracerebral hemorrhage:

Predictors and outcome. *Int J Neurosci* 2014;124:890-3.

Appendix

A. DUAL-ENERGY CT POST-PROCESSING

The fundamental principle of dual-energy CT (DECT) is that certain materials show a clear difference in x-ray attenuation profile for varying clinically used x-ray beam energies. Most of our body's soft tissue is composed of C, H, N and O atoms. The electron binding energy of the innermost electron shell, called k -edge, in these materials is very low (< 1 keV). Because the average photon energy used in medical x-rays (approximately $1/3$ of the maximum beam energy) is far higher than this k -edge energy, the attenuation of x-rays is dominated by the Compton effect, which is directly proportional to the physical density of materials. However, materials with a higher atomic number (Z) have a higher k -edge, such as calcium (4 keV) and iodine (33 keV). Here, the photoelectric effect still plays an important role in x-ray attenuation, causing an increase in attenuation, especially when the average photon energy is close to the material's k -edge. At higher energies, the photoelectric effect decreases with the used peak kilovoltage (kVp), proportionally to kVp^{-3} .

When a CT scan is acquired at two energies, the difference in contribution of the Compton and photoelectric effect in both beams, is bigger for higher Z materials compared to soft tissue materials. For example, when an iodinated contrast-enhanced CT scan of the head is acquired at 100 kVp and 150 kVp, the attenuation due to the presence of iodine will be much higher in the low kV scan compared to the high kV scan, because at the low energy scan the photoelectric effect has a bigger contribution to the total attenuation (average beam energy close to iodine k -edge). Thus with increasing iodine density, the computed Hounsfield Units (HU) on the low kV image will increase more rapidly than on the high kV image. For soft tissues, such as brain parenchyma (BP) and intracerebral hemorrhage (ICH), the difference in contribution of the photoelectric effect is much less, since x-ray attenuation is dominated by the Compton effect in both scans. In this way, the difference in increase in HUs with increasing soft tissue density between the low and high kV images is less.

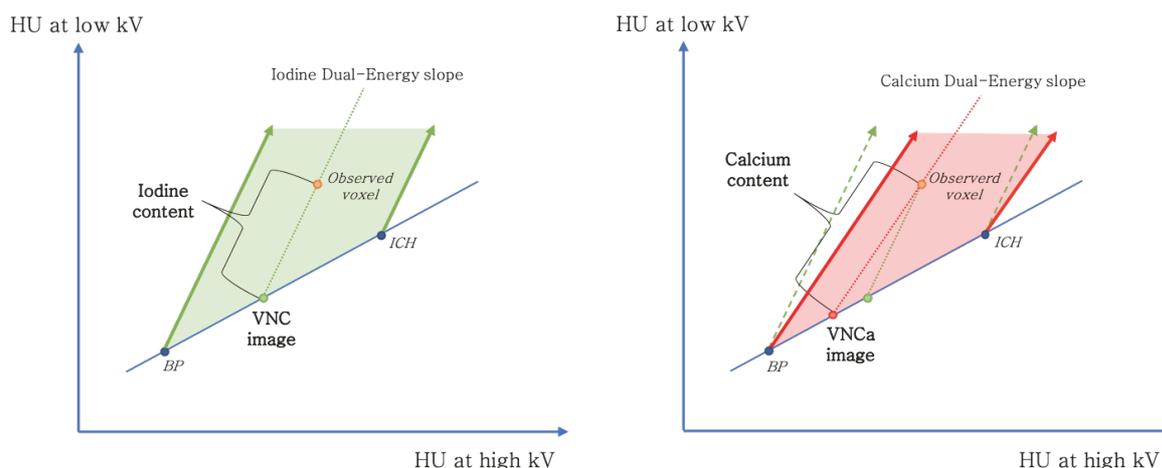


Figure A.1: Post-processing principle of three material decomposition in DECT for the generation of VNC / iodine-only images (left) and VNCa / calcium-only images (right). Note that the dual-energy slope of calcium is lower compared to iodine, because calcium has a lower k -edge, and thus the photoelectric effect contributes less to the total attenuation.

Both material types can be characterized with dual-energy slopes (Figure A.1), one considered as the (soft tissue) base-material pair slope (BP, ICH) and the other as the high Z material slope (e.g. iodine or calcium). The HUs of an observed voxel are then interpreted as a displacement from the base-material

slope along the high Z material slope, representing the content of that material in the observed voxel. In this way, three-material decomposition is achieved by virtually decomposing every voxel into a base-material part and high Z material part. In turn, virtual non-contrast (VNC) / iodine only and virtual non-calcium (VNCa) / calcium-only images can be computed when an iodine and calcium slope are used respectively. Image reconstruction parameters and selection of base-material pair depend on clinical application. Final material quantification can be achieved by doing calibration experiments.

B. COMPUTATION OF TEXTURE FEATURES FROM GLCMS

As discussed in section 2.1, GLCMs can be computed for adjacent pixels or voxels (hereafter elements) pairs in a specific directions with respect to the central assessed element. When multiple (often symmetrical) directions are considered, a separate GLCM is computed for every individual direction. After summing up these GLCMs and normalizing the result, texture features such as entropy, energy, homogeneity, contrast and dissimilarity can be computed using the following equations:

$$\text{Entropy} = \sum_{i,j} p(i,j) \log(p(i,j)) \quad (\text{B.1})$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \quad (\text{B.2})$$

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (\text{B.3})$$

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (\text{B.4})$$

$$\text{Dissimilarity} = \sum_{i,j} |i-j| p(i,j) \quad (\text{B.5})$$

with p representing the GLCM, and i and j the GLCM row and column elements respectively.

C. LAPLACIAN OF A GAUSSIAN FILTERING

In order to enhance texture of the segmented hematomas on different scales, CT images were filtered using a *Laplacian of a Gaussian* (LoG) kernel with varying filter width (σ). The details of the used convolutional kernels are presented below.

2D LoG filtering

$$\text{Gauss kernel: } g(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{y^2}{2\sigma_y^2}} \quad (\text{C.1})$$

$$\text{LoG kernel: } \text{LoG}(x,y) = g_{xx} + g_{yy} \quad (\text{C.2})$$

$$\text{Filter size: } H = \left[\left[\frac{-h_x}{2}, \frac{h_x}{2} \right], \left[\frac{-h_y}{2}, \frac{h_y}{2} \right] \right] \quad (\text{C.3})$$

3D LoG filtering

$$\text{Gauss kernel: } g(x,y,z) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{y^2}{2\sigma_y^2}} \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{z^2}{2\sigma_z^2}} \quad (\text{C.4})$$

$$\text{LoG kernel: } \text{LoG}(x,y,z) = g_{xx} + g_{yy} + g_{zz} \quad (\text{C.5})$$

$$\text{Filter size: } H = \left[\left[\frac{-h_x}{2}, \frac{h_x}{2} \right], \left[\frac{-h_y}{2}, \frac{h_y}{2} \right], \left[\frac{-h_z}{2}, \frac{h_z}{2} \right] \right] \quad (\text{C.6})$$

$$h_x = 2 \lceil 2\sigma_x \rceil \quad h_y = 2 \lceil 2\sigma_y \rceil \quad h_z = 2 \lceil 2\sigma_z \rceil \quad (\text{C.7})$$

$$\sigma_x = \frac{\sigma}{v_x} \quad \sigma_y = \frac{\sigma}{v_y} \quad \sigma_z = \frac{\sigma}{v_z} \quad (\text{C.8})$$

with v_x , v_y , v_z the voxel dimensions of the CT data.

D. BASIC CNN OPERATIONS

The basic operations in a CNN are structured in layers to ensure a certain order. The most commonly used order of layers is as follows: convolution, batch-normalization, dropout, non-linear activation function. While the convolution and non-linear activation layers are considered as standard, batch-normalization and dropout are optional.

The basic operation in a convolutional layer is presented in Figure 2.2. It shows that N sliding filters of size 3×3 process a C -channel input to an N -channel output feature space that is transferred to the next layer. Every individual convolution operation thus has only one output value that is transferred to the position of the central element of the considered input space area. When the filters slide in a stepwise manner over the complete grid of the input data, a new value is computed for all data elements except the border elements. To avoid shrinking of the feature space in the next layer, the input data is usually zero-padded.

Stride and pooling

The size of the steps that a convolutional filter takes to slide over the image grid is determined by the stride s . It represents the number of data elements the filter moves to perform the next convolution. Both the filter shape f and the filter stride s can be chosen arbitrarily for all dimensions. If f and s are represented by a scalar, the shape in all dimensions is equal. Figure D.1 (left) presents an example of stride when $s = 2$ for the filter moving in the x direction. The number of possible convolution operations is halved for both directions, resulting in a down sampled output space (not depicted) with the same factor $s = 2$. In this way, the convolutional layer acts as a learnable down sampling operation. This in contrast to down sampling using pooling operations (Figure D.1 right), in which fixed rules are used, such as in maximum and average pooling.

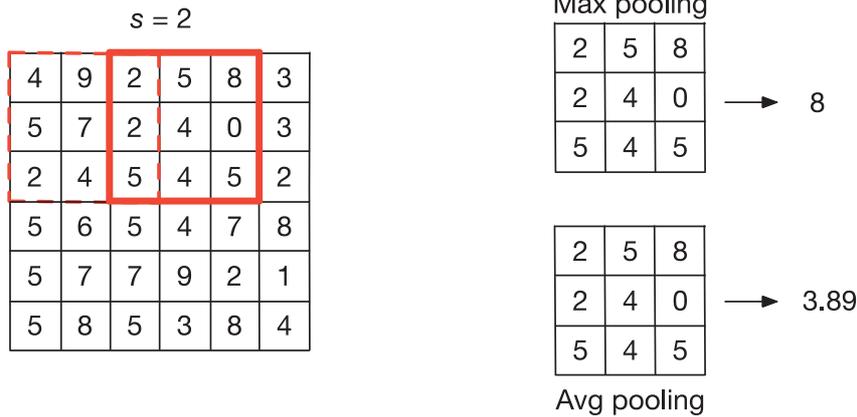


Figure D.1: The concept of stride (left) with $s = 2$ and pooling (right). Both convolution and pooling operations have a single value as output for every considered input space area (red). For the sake of this example, the pooling size is taken the same as the filter size (3×3). Note that the pooling size is usually equal to the stride, and thus would be 2×2 instead of 3×3 in practice.

Depth wise separable convolutions

The basic convolution operation presented in Figure 2.2 is a cross-channel operation, using data from all channels to compute an output. However, in some cases it might be intuitively undesirable to mix channel information, for example when different input channels represent different images, such as in 3-phase

CT data. To overcome this problem, individual convolution operations should only act on a group (g) of or even individual input channels. This is what is called a depth wise separable convolution. In addition to avoiding mixed channel information, the computational costs of this operation are lower when computing the same number of output feature maps compared to a regular convolution, since the number of required weights for every produced output channel is reduced by the number of input channels C or C/g for the grouped case. The concept of the depth wise separable convolution is presented in Figure D.2.

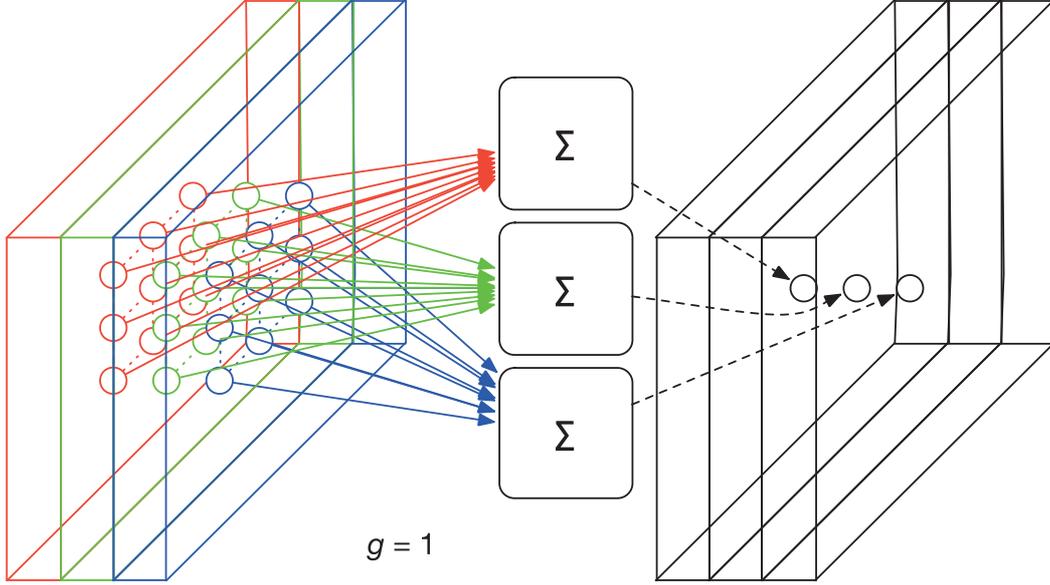


Figure D.2: Depthwise separable convolution acting on individual channels ($g = 1$) of a multi-channel input space ($C = 3$). Three filters compute three output channels using the same number of weights as a single filter in a regular convolution.

Batch normalization

The output of the convolutional layers is sent to a batch normalization layer, where the data is normalized according to learned batch statistics. Since backpropagation takes place after all data samples (m) of one batch have been processed, it is important that the gradient of the backpropagated error through the network should not vanish or explode to avoid ineffective adjustments and thus learning of the weights and biases in the convolutional layers. A batch normalization layer acts on the feature space (x) of a specific layer from one batch by bringing its mean (μ) to 0 and its variance (σ^2) to 1. Subsequently, the normalized output is scaled with γ and is shifted with β to an output space y . Both these scaling and shifting parameters can be learned using the stochastic gradient descent algorithm. Batch normalization allows to use higher learning rates, accelerating training of CNNs, and be less careful about weight initialization. An overview of the steps in batch normalization (BN) is presented below.¹⁴⁹

$$\text{Batch mean:} \quad \mu = \frac{1}{m} \sum_{i=1}^m x_i \quad (\text{D.1})$$

$$\text{Batch variance:} \quad \sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2 \quad (\text{D.2})$$

$$\text{Normalize:} \quad \hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (\text{D.3})$$

$$\text{Scale and shift:} \quad y_i = \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad (\text{D.4})$$

Dropout

A regularization method to prevent overfitting of the learned weights in the convolutional layers is *dropout*. A dropout layer randomly drops input elements of the feature space with a given probability by setting them to 0. This effectively drops neurons in the subsequent convolutional layer and thus avoids co-adaptation of learned weights that work well on the training set, while these are far more likely to fail on an independent test set. The effect of a dropout layer is demonstrated in Figure D.3.

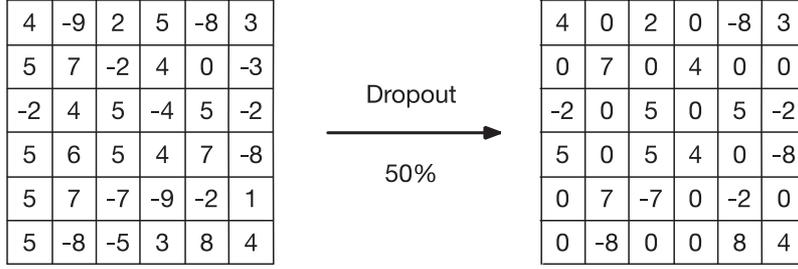


Figure D.3: Effect of a 50% dropout layer.

Non-linear activation functions

To allow a neural network in general to learn complex mappings from input to output, non-linearities, called activation functions, are required that transform the data between successive other layers. The most commonly used activation is the Rectified Linear Unit (ReLU) function, and is defined as follows:

$$y = \max(0, x) \quad (D.5)$$

The effect of the ReLU function on a certain input is presented in Figure D.4.

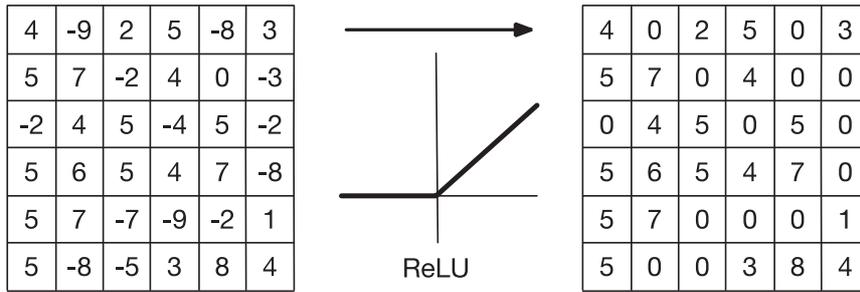


Figure D.4: Non-linear ReLU function applied to certain input matrix.

The ReLU function has a few advantages. First is that the computational costs are very low. Second, it solves the vanishing gradient problem compared to other activation functions (such as sigmoid), because the slope does not saturate when x gets large. The latter especially occurs in early used layers in the network, since here the actual gradient is a multiplication of gradients in later layers. If gradients are very small or even zero, the weight updates computed by the stochastic gradient descent algorithm are also (close to) zero, and thus no learning takes place. The third advantage of the ReLU function is that it is sparsely activated, since it is zero for all negative inputs. This is desirable, because it only activates neurons that have learned specific features relevant to the presented input data. It makes a network faster, as fewer things need to be computed.

For final classification layers, the earlier mentioned sigmoid function is still very useful, since it is only able to produce output values between 0 and 1, representing likelihoods of output data points belonging to one or the other class (in our case whether a voxel belongs to the negative (no infarct) or positive (infarcted) class). The sigmoid function is defined as follows:

$$y = \frac{1}{1 + e^{-x}} \quad (D.6)$$

E. DSC OPTIMIZATION

The output of both CNNs used in this thesis is generated by a final output layer, giving a voxel wise likelihood of being infarcted using above mentioned sigmoid activation function (D.6). A thresholding operation is required to go from this output likelihood map to a final binary output (infarct vs. no infarct). Subsequently, the DSC can be computed, which is used as final performance metric. Given the definition of the sigmoid function with output values between 0 and 1, it seems intuitively to use a threshold of 0.5

for final voxel classification. However, trial-and-error showed that the final DSC is strongly dependent on the chosen threshold. An explanation for this sensitivity is the huge class imbalance in the used dataset, with relatively few lesion positive voxels. The predictions need to be very specific to not end up with a near-zero DSC, because false positives decrease the DSC heavily. In order to optimize the used threshold that maximizes DSC, an algorithm similar to expectation maximization has been implemented, assuming that the DSC only increases when stepwise lowering the threshold from 1 to 0 (thus starting with zero false positives) until a maximum has been reached. Figure E.1 demonstrates the influence of the probability threshold on the final DSC for a test configuration of the U-Net architecture after being trained to segment stroke lesions from DWI and ADC input data. It shows that the optimal probability threshold is close to 1 and that the difference in DSC with respect to a threshold of 0.5 is 0.46 in on the DSC scale.

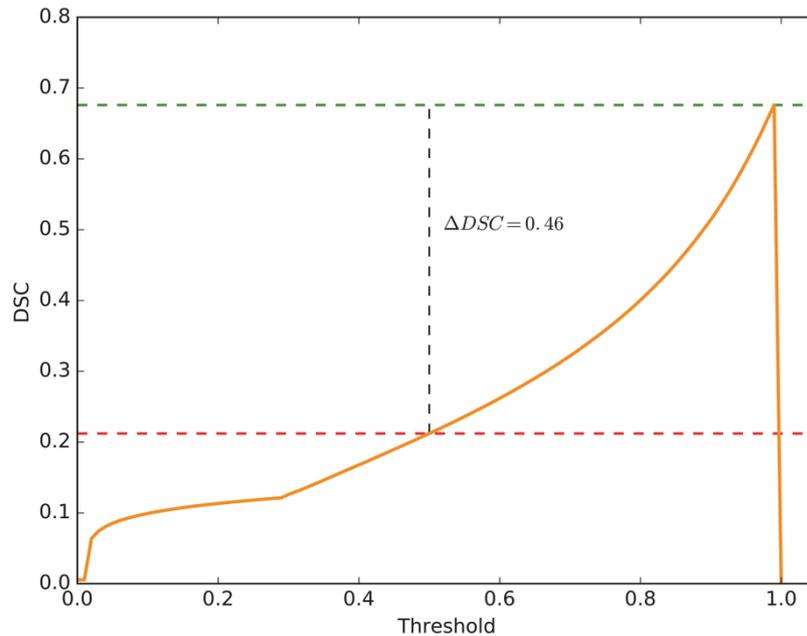


Figure E.1: Example of the sensitivity of the final computed DSC with respect to the chosen probability threshold to binarize the CNN's output.

F. OVERVIEW OF TESTED CNN CONFIGURATIONS

| CNN Type | Block type | Operational dimensionality | Modality processing | Training scheme | Sampling strategy | Input modality | Data augmentation | Network depth | Batch size |
|-----------|------------|----------------------------|---------------------|-----------------|-------------------|----------------|-------------------|---------------|------------|
| DeepMedic | - | 2D | Mixed | Patch | SW | SECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | SW | SECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | SW | SECT | No | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | SW | SECT | No | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | SW | DECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | SW | DECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | SW | DECT | No | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | SW | DECT | No | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | FGBG | SECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | FGBG | SECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | FGBG | SECT | No | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | FGBG | SECT | No | 4 | 128 |
| DeepMedic | - | 2D | Mixed | Patch | FGBG | DECT | No | 4 | 128 |

| | | | | | | | | | |
|-----------|----------|----|----------|-------|------|------|-----|---|-----|
| DeepMedic | - | 2D | Mixed | Patch | FGBG | DECT | Yes | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | FGBG | DECT | No | 4 | 128 |
| DeepMedic | - | 2D | Separate | Patch | FGBG | DECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | SW | SECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | SW | DECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | SW | SECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | SW | DECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | SW | SECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | SW | SECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | SW | DECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | SW | DECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | SW | SECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | SW | SECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | SW | SECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | SW | SECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | SW | DECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | SW | DECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | SW | DECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | SW | DECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | FGBG | SECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | FGBG | DECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | FGBG | SECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | FGBG | DECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | FGBG | SECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | FGBG | SECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Mixed | Patch | FGBG | DECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Mixed | Patch | FGBG | DECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | FGBG | SECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | FGBG | SECT | Yes | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | FGBG | SECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | FGBG | DECT | No | 4 | 128 |
| U-Net | ResDense | 2D | Separate | Patch | FGBG | DECT | Yes | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | FGBG | DECT | No | 4 | 128 |
| U-Net | ResNeXt | 2D | Separate | Patch | FGBG | DECT | Yes | 4 | 128 |
| DeepMedic | - | 3D | Mixed | Patch | SW | SECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | SW | SECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | SW | SECT | No | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | SW | SECT | No | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | SW | DECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | SW | DECT | Yes | 4 | 2 |

| | | | | | | | | | |
|-----------|----------|----|----------|-------|------|------|-----|---|---|
| DeepMedic | - | 3D | Mixed | Patch | SW | DECT | No | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | SW | DECT | No | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | FGBG | SECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | FGBG | SECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | FGBG | SECT | No | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | FGBG | SECT | No | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | FGBG | DECT | No | 4 | 2 |
| DeepMedic | - | 3D | Mixed | Patch | FGBG | DECT | Yes | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | FGBG | DECT | No | 4 | 2 |
| DeepMedic | - | 3D | Separate | Patch | FGBG | DECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | FGBG | SECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | FGBG | SECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | FGBG | SECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | FGBG | SECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | FGBG | DECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | FGBG | DECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | FGBG | DECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | FGBG | DECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | FGBG | SECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | FGBG | SECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | FGBG | SECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | FGBG | SECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | FGBG | DECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | FGBG | DECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | FGBG | DECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | FGBG | DECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | SW | SECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | SW | SECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | SW | SECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | SW | SECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | SW | DECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Mixed | Patch | SW | DECT | Yes | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | SW | DECT | No | 4 | 2 |
| U-Net | ResDense | 3D | Separate | Patch | SW | DECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | SW | SECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | SW | SECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | SW | SECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | SW | SECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | SW | DECT | No | 4 | 2 |
| U-Net | ResNeXt | 3D | Mixed | Patch | SW | DECT | Yes | 4 | 2 |
| U-Net | ResNeXt | 3D | Separate | Patch | SW | DECT | No | 4 | 2 |

| | | | | | | | | | |
|-------|----------|----|----------|-------|----|------|-----|---|-----|
| U-Net | ResNeXt | 3D | Separate | Patch | SW | DECT | Yes | 4 | 2 |
| U-Net | ResDense | 2D | Mixed | Dense | - | SECT | No | 6 | 256 |
| U-Net | ResDense | 2D | Mixed | Dense | - | SECT | Yes | 6 | 256 |
| U-Net | ResDense | 2D | Separate | Dense | - | SECT | No | 6 | 256 |
| U-Net | ResDense | 2D | Separate | Dense | - | SECT | Yes | 6 | 256 |
| U-Net | ResDense | 2D | Mixed | Dense | - | DECT | No | 6 | 256 |
| U-Net | ResDense | 2D | Mixed | Dense | - | DECT | Yes | 6 | 256 |
| U-Net | ResDense | 2D | Separate | Dense | - | DECT | No | 6 | 256 |
| U-Net | ResDense | 2D | Separate | Dense | - | DECT | Yes | 6 | 256 |
| U-Net | ResNeXt | 2D | Mixed | Dense | - | SECT | No | 6 | 256 |
| U-Net | ResNeXt | 2D | Mixed | Dense | - | SECT | Yes | 6 | 256 |
| U-Net | ResNeXt | 2D | Separate | Dense | - | SECT | No | 6 | 256 |
| U-Net | ResNeXt | 2D | Separate | Dense | - | SECT | Yes | 6 | 256 |
| U-Net | ResNeXt | 2D | Mixed | Dense | - | DECT | No | 6 | 256 |
| U-Net | ResNeXt | 2D | Mixed | Dense | - | DECT | Yes | 6 | 256 |
| U-Net | ResNeXt | 2D | Separate | Dense | - | DECT | No | 6 | 256 |
| U-Net | ResNeXt | 2D | Separate | Dense | - | DECT | Yes | 6 | 256 |
| U-Net | ResDense | 3D | Mixed | Dense | - | SECT | No | 6 | 4 |
| U-Net | ResDense | 3D | Mixed | Dense | - | SECT | Yes | 6 | 4 |
| U-Net | ResDense | 3D | Separate | Dense | - | SECT | No | 6 | 4 |
| U-Net | ResDense | 3D | Separate | Dense | - | SECT | Yes | 6 | 4 |
| U-Net | ResDense | 3D | Mixed | Dense | - | DECT | No | 6 | 4 |
| U-Net | ResDense | 3D | Mixed | Dense | - | DECT | Yes | 6 | 4 |
| U-Net | ResDense | 3D | Separate | Dense | - | DECT | No | 6 | 4 |
| U-Net | ResDense | 3D | Separate | Dense | - | DECT | Yes | 6 | 4 |
| U-Net | ResNeXt | 3D | Mixed | Dense | - | SECT | No | 6 | 4 |
| U-Net | ResNeXt | 3D | Mixed | Dense | - | SECT | Yes | 6 | 4 |
| U-Net | ResNeXt | 3D | Separate | Dense | - | SECT | No | 6 | 4 |
| U-Net | ResNeXt | 3D | Separate | Dense | - | SECT | Yes | 6 | 4 |
| U-Net | ResNeXt | 3D | Mixed | Dense | - | DECT | No | 6 | 4 |
| U-Net | ResNeXt | 3D | Mixed | Dense | - | DECT | Yes | 6 | 4 |
| U-Net | ResNeXt | 3D | Separate | Dense | - | DECT | No | 6 | 4 |
| U-Net | ResNeXt | 3D | Separate | Dense | - | DECT | Yes | 6 | 4 |