DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEEP LEARNING CLASSIFICATION OF TERRITORIAL ISCHEMIC CEREBROVASCULAR ACCIDENT IN DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGING

by

İlker Özgür KOSKA

February, 2024

İZMİR

DEEP LEARNING CLASSIFICATION OF TERRITORIAL ISCHEMIC CEREBROVASCULAR ACCIDENT IN DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGING

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Department of Biomedical Technologies, Advanced Biomedical Technologies Program

by

İlker Özgür KOSKA

February, 2024

İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"DEEP LEARNING CLASSIFICATION OF TERRITORIAL ISCHEMIC CEREBROVASCULAR ACCIDENT IN DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGING"** completed by **İLKER ÖZGÜR KOSKA** under supervision **PROF. DR. ALPER SELVER** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof.Dr. Alper Selver

Supervisor

Prof. Dr. Oğuz Dicle

Thesis Committee Member

Assoc. Prof. Dr. Mehmet Emre Çek

Thesis Committee Member

Prof. Dr. Mustafa Fazıl Gelal

Examining Committee Member

Prof. Dr. Bilge Karaçalı

Examining Committee Member

Prof. Dr. Okan FISTIKOĞLU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENT

First and foremost, I would like to thank my advisor Prof. Dr. Alper Selver. He was always very encouraging and motivating in every aspect. Most importantly he gave me a lot of freedom to discover my own way and explore different areas, while always being ready to help and give valuable advice when necessary. Many many thanks!!! I would like to thank Prof. Dr. Hüdaver Alper, my mentor, one of my dearest friends and my light on my way with my whole heart. Nothing could start if he did not let me go to find my own destiny and encourage me all the way. Great many thanks!!!

I am also very grateful to my PhD committee: Prof. Dr. Oğuz Dicle and Asoc. Prof. Dr. Emre Çek who gave important feedback about the general direction of the project.

I would like to thank Prof. Dr. Bram van Ginneken, who has given the opportunity to observe their workflow and to work as a researcher for 6 months in DIAG Nijmegen, Radboud University and Prof. Dr. Ender Konukoğlu, who hosted me in ETH Zurich BMIC group as a researcher for 6 months. I would especially like to thank to Dr. Krishna Chaitanya who helped me to sort out the difficulties of working with cluster and on terminal with bash scripting and Prof. Dr. Roman Guggenberger, who has included me in his team in University Hospital of Zurich for AI projects and trusted me as a project developer for their AI research. I also would like to thank Prof. Dr. Irmak Durur Subaşı who pushed me forward when I was discouraged and injected me hopes when I have felt exhausted.

I was very lucky to be the son of the greatest parents ever. Great many thanks to my father Nedim Koska and my mother Semra Koska. They are great alchemists, who turned a deaf teenager to a medical doctor with two subspecialties and a medical image analysis PhD. I would also like to thank to the mother of my son, my life companion Ece Eker for her endless contagious life passion, to my dear son Çağan Koska for his sweet heart, for his deep analysis of life, and for his great effort to realize one of my proudest moments while he was taking the best computing project award from the president at his tenth grade, and to my lovely puppy, Goflet for his great

companionship. Their support, love and being always my side gave me the required fuel to fight the obstacles, difficulties, and uphills of this journey. Finally great many thanks to countless of people who had intersections with my life and behaved kindly to the life and to me. World will be a good place with good mannered people.

İlker Özgür KOSKA

DEEP LEARNING CLASSIFICATION OF TERRITORIAL ISCHEMIC CEREBROVASCULAR ACCIDENT IN DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGING

ABSTRACT

Stroke is a common pathology on a global scale and has a high probability of death or permanent loss of function. Many artificial intelligence applications were developed to shed light on the etiology of stroke, direct treatment and predict prognosis. However, a tool that can determine which arterial supply area has been affected by the signal change in DWI, the most applied imaging method, has not yet been developed. Within the scope of this thesis, my aim was to develop a model that can automatically detect which arterial supply area has been affected by using the signal change in DWI images and provide automatic triage in the absence of a radiologist. For this purpose, 624 volumetric DWI examinations were collected from 3 different centers and an end-to-end fully automatic classifier model was created, which included 5 separate classes: anterior cerebral artery, middle cerebral artery, and posterior circulation strokes, which were the main arterial supply areas, and border zone strokes between these areas and normal subjects. The problem was addressed step by step and cross-section-based, voxel-based and patient-based approaches were developed. The resulting workflow and classifier model performs at the level of a radiologist, thus developing a tool that can be integrated into the clinic and used for triage purposes in the absence of a radiologist is promising.

Keywords: Deep learning, diffusion weighted imaging, stroke, vascular territory classification, image processing, dense prediction, time distributed layer.

DİFÜZYON AĞIRLIKLI MANYETİK REZONANS GÖRÜNTÜLEMEDE İSKEMİK SEREBROVASKULER OLAY BÜYÜK DAMAR SINIFLANDIRILMASINDA DERİN ÖĞRENME

ÖZ

İnme küresel ölçekte yaygın görülen bir patolojidir ve ölüm ya da kalıcı fonksiyon kaybı ile sonuçlanma olasılığı yüksektir. İnmenin etiyolojisine ışık tutmak, tedaviyi yönlendirmek ve prognozu tahmin etmek amacıyla birçok yapay zekâ uygulaması geliştirilmiştir. Ancak en çok uygulanan görüntüleme yöntemi olan difüzyon ağırlıklı manyetik rezonans görüntülemede hangi arteriyel beslenme alanının etkilendiğini sinyal değişikliğinden belirleyebilecek bir araç henüz geliştirilmemiştir. Bu tez kapsamında amacım hangi arteriyel besleme alanının felçten etkilendiğini difüzyon ağırlıklı manyetik rezonans görüntülerindeki sinyal değişikliğinden otomatik olarak tespit edebilecek ve radyolog yokluğunda otomatik triyaj sağlayabilecek bir model geliştirmekti. Bu amaçla 3 farklı merkezden 624 adet hacimsel difüzyon ağırlıklı manyetik rezonans incelemesi toplanmış ve ön serebral arter, orta serebral arter ve arka dolaşım felçleri ile bu alanlar arasında yer alan sınır zon felçleri ve normal örnekler olmak üzere 5 ayrı sınıfı içeren uçtan uca tam otomatik bir sınıflandırıcı model oluşturulmuştur. Problem adım adım ele alınarak kesit bazlı, voksel bazlı ve hasta bazlı yaklaşımlar geliştirilmiştir. Ortaya çıkan iş akışı ve sınıflandırıcı model bir radyolog seviyesinde çalışmaktadır, dolayısıyla kliniğe entegre edilebilecek ve bir radyoloğun yokluğunda triyaj amacıyla kullanılabilecek bir araç geliştirmek açsından umut vericidir.

Anahtar kelimeler: Derin öğrenme, difüzyon ağırlıklı görüntüleme, inme, vasküler sulama sahası sınıflandırması, görüntü işleme, yoğun tahmin, zaman dağıtık tabaka.

CONTENTS

Page

Ph.D. THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGMENT iii
ABSTRACT v
ÖZ vi
CONTENTS vii
LIST OF FIGURES ix
LIST OF TABLES xi
LIST OF ABBREVIATIONS xii

CHAPTER ONE - INTRODUCTION 1

CHAPTER TWO - BACKGROUND 2

2.1 Stroke definition 2
2.2 Ischemic stroke physiopathology 2
2.3 Cerebrovascular anatomy 3
2.4 Brain anatomy and general organization 4
2.5 Vascular territories of brain 5
2.6 Radiologic imaging in stroke 7

CHAPTER THREE - PREVIOUS STUDIES 11

3.1 Early stroke detection 15
3.2 Large vessel occlusion 17
3.3 ASPECTS score 17
3.4 Ischemic core and penumbra evaluation 18
3.5 Prognosis prediction 20
3.6 Challenges 20

CHAPTER FOUR - METHODOLOGY &DATA ANALYSIS 22

4.1 Slice-based classification task 22
4.2 Voxel-based classification task..... 27
4.3 Patient wise 3D end to end classification task 33
4.4 Statistics..... 43

CHAPTER FIVE - RESULTS 45

5.1 Results for slice wise classification approach 45
5.2 Results for voxel wise classification approach..... 47
5.3 Results for patient wise 3D end to end classification approach 52

CHAPTER SIX – DISCUSSION AND CONCLUSION 60

6.1 Discussion for slice wise approach..... 60
6.2 Discussion for voxel wise approach 62
6.3 Discussion for patient wise 3D end to end approach..... 66
6.4 Conclusion 70

REFERENCES..... 72

LIST OF FIGURES

Page

Figure 2.1 Lobar and vascular anatomy of the brain 4

Figure 2.2 Vascular supply zones of the brain 6

Figure 4.1 Case selection process for slice-based classification task 23

Figure 4.2 Applied transformations 25

Figure 4.3 Edge profile of normal and infarct regions 27

Figure 4.4 Model architectures for slice-based classification task 28

Figure 4.5 Projection of the highlighted festures on the native image 28

Figure 4.6 Overview of patient selection for voxel wise classification task 29

Figure 4.7 Model architecture and study workflow for voxel wise task 30

Figure 4.8 Overview of mask building strategy for voxel wise task 32

Figure 4.9 Dataset building and patient flow for patient wise approach 36

Figure 4.10 Supervised deep models for patient wise classification approach 43

Figure 4.11 Autoencoder as the unsupervised model for latent vector 44

Figure 5.1 ROC plots of the augmented input models 47

Figure 5.2 Bar plots of performance metrics for slice-based models 49

Figure 5.3 Histogram of stroke volumes 51

Figure 5.4 Histogram of Dice scores 52

Figure 5.5 Predictions of the models before and after bias field correction and registration 53

Figure 5.6 ROC curves of deep models 54

Figure 5.7 t-SNE and PCA plots of parallel stream model embeddings..... 58

Figure 5.8 Test accuracy plots of classical machine learning models with deep image embeddings..... 58

Figure 5.9 Bar plot of test accuracy of external validation experiments 59

Figure 5.10 Error analysis of predictions..... 59

Figure 6.1 Effects of rotation 65

LIST OF TABLES

Page

Table 4.1 DWI sequences and imaging parameters. 24

Table 4.2 Gender and age distribution of the cases for slice wise task. 25

Table 4.3 Patient demographics of voxel wise classification approach. 31

Table 4.4 Patient demographics for patient wise classification task 35

Table 4.5 Diffusion acquisition protocol and data distirbution for all centers 37

Table 4.6 Center 1 patients' clinical information. 38

Table 5.1 Evaluation of pretrained classifiers with native images for vascular territory prediction. 46

Table 5.2 Evaluation of augmented input pretrained classifiers for vascular territory prediction 48

Table 5.3 Stroke volumes. 50

Table 5.4 Dice scores of ground truth segmentations and model predictions for ordinary single class.model for the test set 50

Table 5.5 Classification metircs of the proposed approach. 51

Table 5.6 Performance metrics of deep networks. 55

Table 5.7 Performance metriccs of pocket algorithm with the embeddings of deep networks. 57

LIST OF ABBREVIATIONS

DWI	: Diffusion weighted imaging
AHA	: American Heart Association
CBF	: Cerebral blood flow
ICA	: Internal carotid artery
ACA	: Anterior cerebral artery
MCA	: Middle cerebral artery
Acom	: Anterior communicating artery
Pcom	: Posterior communicating artery
PCA	: Posterior cerebral artery
AICA	: Anterior inferior cerebellar artery
PICA	: Posterior inferior cerebellar artery
SCA	: Superior cerebellar artery
IV	: Intra venous
CBV	: Cerebral blood volume
TTP:	: Time to peak
MTT	: Mean transit time
CTP	: Computed tomography perfusion
SVM	: Support vector machine
RF	: Random Forest

LR	: Logistic regression
ANN:	: Artificial neural network
RSNA:	: Radiological Society of North America
CNN:	: Convolutional neural network
ReLU:	: Rectified linear unit
AE:	: Auto encoder
ASPECTS:	: Alberta Stroke Program Early CT score
LVO:	: Large vessel occlusion
CCA:	: Common carotid artery
AUC:	: Area under curve
ADC:	: Apparent diffusion coefficient
PC:	: Posterior circulation
WSZ:	: Watershed zone
MNI:	: Montreal Neurology Institute
MLP:	: Multi-layer perceptron
LSTM:	: Long short-term memory
CBAM:	: Convolutional block attention module
t-SNE:	: t distributed stochastic nearest neighbor embedding
PCA:	: Principal component analysis
GT:	: Ground truth
WS:	: Watershed
TDL:	: Time distributed layer

CHAPTER ONE
INTRODUCTION

Stroke is one of the leading causes of mortality and morbidity globally. Rapid treatment of stroke is mandatory to keep functional integrity as much as possible. This is summarized with a famous quote: 'Time is brain'. Actually, acute cessation of blood flow may lead to 2 million neuron loss at every second (Osborn, 2015). Both due to its high prevalence and great benefit potential in the absence of domain experts, automatization in stroke management is highly desired. Imaging studies stay in the corner stone of stroke management decisions. Diffusion weighted magnetic resonance imaging (DWI MRI) is the most sensitive modality to detect and confirm ischemic stroke in hyperacute period which can serve as a rapid triage and stratification tool by using automatization methods. Currently, great many researchers are working on deep learning techniques to develop end to end stroke assessment models. Of the 220 FDA or CE approved AI models in radiology so far, 73 are related to brain assessment and one half of these are related to stroke imaging (van Leeuwen, Schalekamp, Rutten, van Ginneken, & de Rooij, 2021). Large vessel occlusion, collateral status assessment, infarct core and penumbra computation, stroke confirmation, hemorrhagic or ischemic mechanism differentiation and hyperacute window detection are among these target subjects for these commercial products (van Leeuwen et al., 2021). To the date there is no end-to-end automatic tool to classify the affected region by ischemic stroke in DWI images. However, DWI is one of the most frequently utilized modality for stroke work up. In this thesis my motivation was building end to end stroke territory classifier in DWI by using machine learning and particularly deep learning techniques. For this aim, data from three centers from different regions of Turkey was collected. Four different radiologists annotated the ground truths to understand the baseline human agreement about the class labels. Slice-level, voxel-level and patient-level classification approaches were explored using classical machine learning, supervised and unsupervised deep learning approaches. Parallel and sequential image encoding strategies were coupled with attention mechanisms to build more robust stroke territory classifier.

CHAPTER TWO
BACKGROUND

In this chapter through review of literature about machine learning in stroke imaging will be discussed.

2.1 Stroke definition

Stroke refers to acute onset neurological deficits. WHO defines stroke as rapidly progressing clinical symptoms caused by focal or global cerebral dysfunction that lasts more than 24 hours or leads to death and has no apparent cause than vascular etiology (Aho et al., 1980). American Heart Association (AHA) updated this term in 2013 including silent infarcts and silent hemorrhages demonstrated by imaging studies (Sacco et al., 2013).

Stroke can be classified as ischemic or hemorrhagic according to its mechanism. More than 80% of stroke is related to ischemia. Hemorrhagic infarcts can be divided into 2 subgroups as intraparenchymal hemorrhage which accounts 15% and subarachnoid hemorrhage which accounts remaining 5%. Ischemic stroke can be classified as large vessel occlusion (45%), small vessel occlusion (15-30%), embolic (15-25%) and non-occlusive hypoperfusion types. The management strategy, expected prognosis and outcome are completely different for each subgroup, therefore, this distinction is important (Donnan, Fisher, Macleod, & Davis, 2008; Goldstein et al., 2001). Some classification systems also include dissection as a distinct group, since its treatment strategy, etiology, prognosis, and target population are completely different. In dissection the intimal layer of the vessel detaches from the wall and occlusion of the vessel with cessation of blood flow develops.

2.2 Ischemic stroke physiopathology

Brain is highly vulnerable to acute cessation of blood flow since it doesn’t have replacement storage for energy supply. Normal mean cerebral blood flow (CBF) to the brain is 60ml/100gr/min. If the flow decreases to 35 ml/100gr/min, anaerobic glycolysis starts, and brain lactate levels increase which can be demonstrated by MR

spectroscopy. This can trigger a cascade of events that may disturb ionic homeostasis and hydrogen ion starts to accumulate in the extracellular space leading to metabolic acidosis. If the blood flow decreases to less than 15-25 ml/100 gr/min, neuronal dysfunction starts. In that phase the neurons are still viable and the tissue in that state is called as penumbra which can be rescued following early intervention. Rescuing the penumbra is the main stay of stroke management. If the blood flow decreases to less than 6-8 ml/100 gr/min, the oxygen and the glucose are rapidly consumed and the ion pump which regulates the ion homeostasis and membrane integrity is destroyed which eventually leads to cell death and tissue necrosis. This type of tissue which is irreversibly damages is called ischemic core (Daroff, Fenichel, Jankovic, & Mazziotta, 2012; Osborn, 2015; Warlow, Dennis, Gijn, Hankey, & Barnett, 2012). Stroke can be classified into hyper acute, acute, subacute, and chronic according to the time to the event onset. In the hyperacute window which is defined as 6 hours from the start of event, most of the tissue can be salvageable. Therefore, the one of the most important aims of strong management is to reach the patient in hyperacute window and initiate the correct management strategy as early as possible. First 1-2 weeks after the stroke is acute, 2-4 weeks is subacute, and more than 4 weeks is the chronic phase of stroke (Daroff et al., 2012; Warlow et al., 2012).

2.3 Cerebrovascular anatomy

Intracranial arterial network is classically divided into anterior and posterior circulation. Anterior circulation is composed of a pair of internal carotid artery (ICA), anterior cerebral artery (ACA) and middle cerebral artery (MCA) and their branches. Right and left hemispheric large vascular structures are connected to each other with anterior communicating artery (Acom) which is between right and left ACA and posterior communicating artery (Pcom) which is between right and left posterior cerebral artery (PCA). Acom is assumed to be a part of anterior circulation, Pcom is assumed to be a part of posterior circulation. There is an exception for this classification, which is encountered in 11-46 % of the cases which is called as fetal type PCA. In fetal type PCA, PCA is directly originated from terminal ICA (Harnsberger & Macdonald, 2006; Osborn, 2015). In this situation, PCA technically becomes a part of anterior circulation, however its feeding zone is the same as the

regular PCA. Components of posterior circulation are posterior inferior cerebellar artery (PICA), anterior inferior cerebellar artery (AICA), penetrating pontine arteries, superior cerebellar artery (SCA) and PCA which are branches of vertebrobasilar arterial system. Vertebrobasilar arterial system starts with a pair of vertebral arteries originating from arcus aorta and ends with a common midline basilary artery which is formed by joining of two vertebral arteries into a single trunk (Bouthillier, van Loveren, & Keller, 2001).

2.4 Brain anatomy and general organization:

Brain has five lobes which are frontal, temporal, parietal, occipital and limbic lobes (Figure 2.1). All lobes are organized as pairs and the structures in each hemisphere are connected to contralateral part with commissural bundles. Falk cerebri is a fibrofatty structure that divides the brain into left and right. Tentorium cerebelli is the homolog of falx which has a central opening, and which divides the intracranial structures into superior and inferior. Supratentorial structures are brain lobes, thalamus and basal ganglia and the commissural fibers. Normally gray matter which includes the neuronal bodies and measures 3-5 mm thickness resides in the periphery of the lobes which is also called as cortex and white matter which includes the axonal bodies of the neurons are situated in the deep portions of the brain.

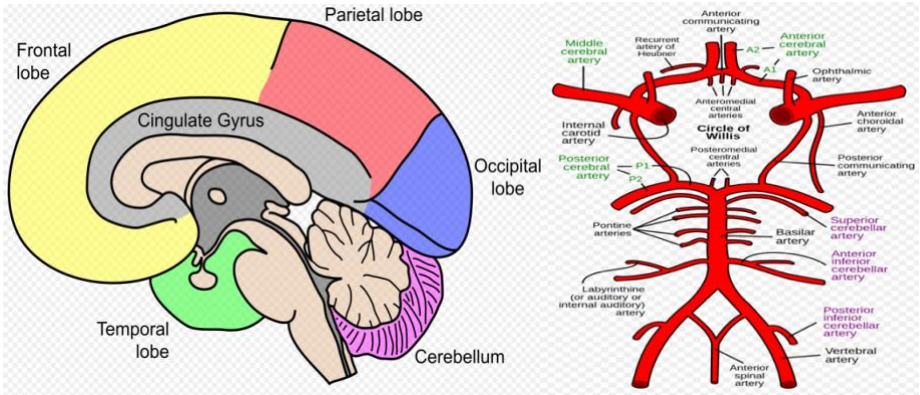


Figure 2.1 a) Lobar (By Rhcastilhos) and b) Vascular anatomy (By NEUROtiker)

Thalamus is the main relay station for sensory input before transmission to the cortex. Thalamus and basal ganglia are pair of gray matter structures which include neuronal bodies but located in deep portions of the brain. Basal ganglia and thalamus mainly function as organizer of motor and sensory input and outputs. The largest

commissural white matter structure is corpus callosum which connects the right and left hemispheres under the falx and above the tentorium. The other important white matter tracts which connect two halves of supratentorial brain are anterior commissure posterior commissure and habenular commissure. There are also ascending and descending white matter tracts which connects the inferior and superior parts of the brain and association fibers which connects the anterior and posterior parts of the brain. Internal capsule is the major descending motor pathway which transmits the signal to the lower neural structures and spinal cord (Osborn, 2015).

2.5 Vascular territories of brain

2.5.1 Anterior circulation

ACA supplies anterior 2/3 of medial cerebral hemispheres, anterior corpus callosum and inferomedial part of frontal lobe. Medial lenticulostriate arteries are a set of small penetrating arteries which supplies the medial part of basal ganglia, genu of corpus callosum and anteroinferior part of internal capsules. MCA has the largest feeding area among other terminal intracranial vascular structures. MCA supplies all the supratentorial parenchyma excluding medial aspect of frontal and parietal lobes, occipital lobes, postero inferior of parietal lobes and inferior of temporal lobes (Figure 2.2). Lateral lenticulostriate arteries which originates from MCA supplies basal ganglia and anterosuperior of internal capsule (Osborn, 2015).

2.5.2 Posterior circulation

PCA, supplies most of the occipital and temporal lobes and posterior of corpus callosum. Most of the thalamus excluding anterior part is supplied by penetrating arteries originated from PCA (Schmahmann, 2003). Vertebrobasilar system supplies all of the infratentorial structures. AICA supplies anteroinferior part and PICA supplies posteroinferior part of cerebellum, SCA supplies most of the remaining hemispheres.

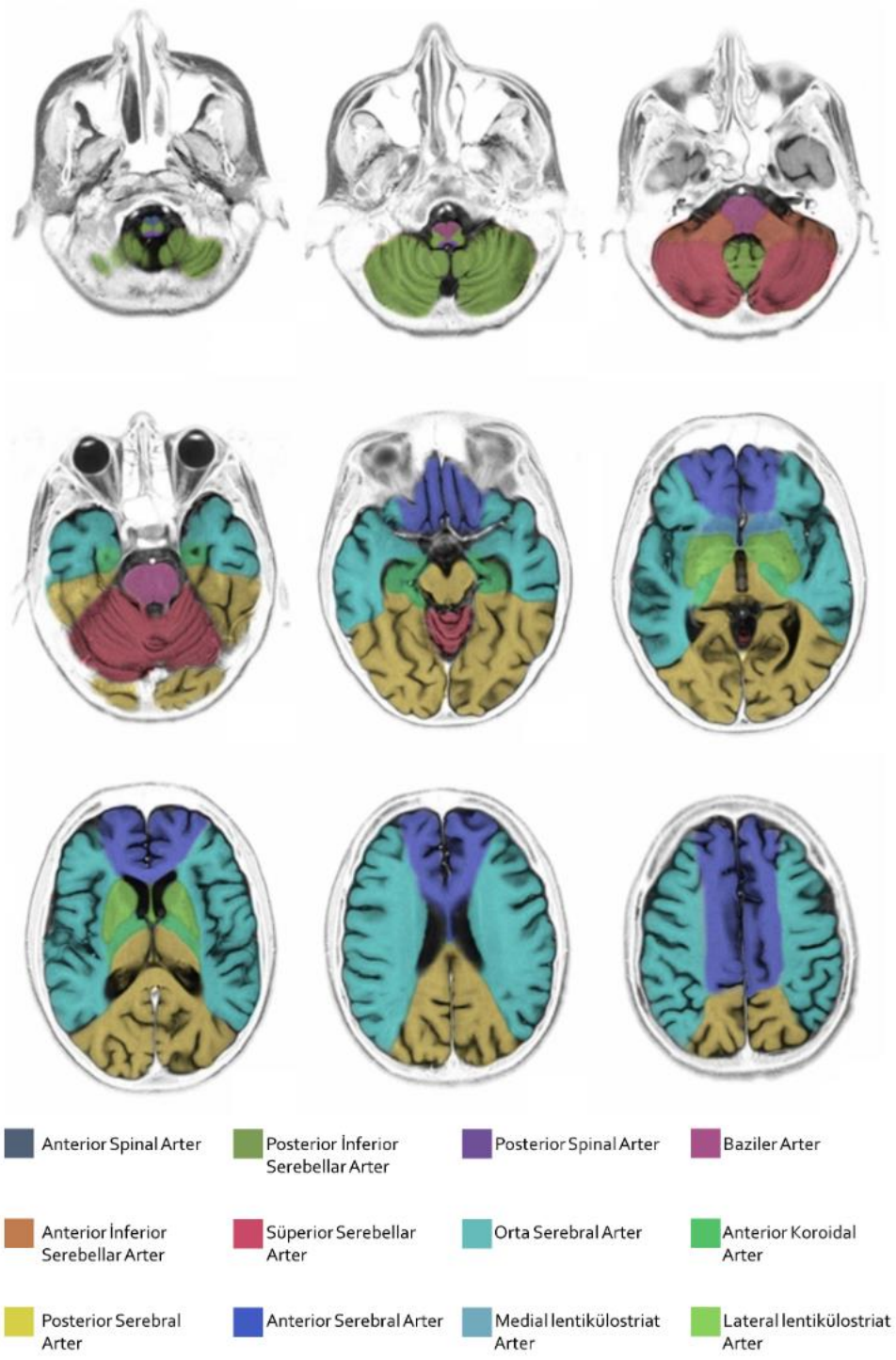


Figure 2.2 Vascular supply zones of the brain (Gaillard, 2024)

2.5.3 Watershed zones

The borders between the arterial territories represent the most distal part of each arterial network. When perfusion pressure drops due to hypotension, vasodilation or cardiac failure, these end regions lose blood flow first. If there is a stenotic lesion proximal to the watershed zones this perfusion decrease becomes more accentuated. The mechanism is completely different from the large or small vessel occlusion type ischemia; therefore, the management strategy is different. Watershed infarct is generally ipsilateral with the lesion site, in high grade ICA stenosis whereas it is generally bilateral when the etiology is cardiac systemic arterial tension related (Moustafa et al., 2010). Watershed infarcts are categorized into 2 groups. Cortical (external watershed infracts) and deep (internal watershed infarcts). Cortical watershed zones are between ACA and MCA or MCA and PCA. Deep watershed regions are between lenticulostriate arteries and deep penetrating branches of MCA.

2.6 Radiologic imaging in stroke:

Imaging in acute ischemic stroke serves many aspects for stroke management. First aim of imaging is confirming stroke diagnosis and differentiating from stroke mimics. Then classification of the mechanism as ischemic and hemorrhagic comes. If there is no hemorrhage, then there is a possibility that the patient may benefit either intravenous or intraarterial revascularization attempts. Then determination of time to the event and the volume of potential salvageable tissue comes. Etiology as large vessel occlusion, small vessel occlusion, cardio embolism or watershed infarcts are also important determinants of management strategy. The extent of the affected region and the functional and anatomic classification of the affected region are other important pieces of the puzzle for decision-making process. Different imaging modalities display different properties of diseased tissue and highlight different aspects of stroke management decisions.

2.6.1 Non-contrast CT

Non-contrast CT is the simplest, fastest, and most easily accessible method for stroke imaging. The most important strength of non-contrast CT is its great ability to differentiate the mechanism of the stroke as ischemic or hemorrhagic whereas the most important weakness is its relatively insensitivity to detect stroke in hyperacute window. There are some ancillary findings in non-contrast CT related to stroke evolution. Dense MCA sign or dot sign, insular ribbon sign and lentiform nucleus sign are among them. Dens MCA sign refers to visualization of thrombus directly in M1 segment of MCA as a hyperdense dot. In 30% of cases, dense MCA sign can be demonstrated (Presman, Tourie, & Thompson, 1987). Insular ribbon sign and lentiform nucleus sign can be attributed to the loss of gray-white matter differentiation due to cytotoxic edema. The earliest time to observe these signs are after 1-3 hours of the onset of the event (Nakano et al., 2001). After hyperacute phase the water content of intracellular compartment gradually increases. This is reflected by sulcal effacement in 6-12 hours and attenuation decrease in 12-24 hours after stroke. 6-24 hours of stroke is known as late hyper acute phase. In 3-10 days, cell swelling, and mass effect increases which is the main mechanism of secondary damage in large infarcts. After 7-10 days the mass effect starts to decrease and finally after 4 weeks completely disappears. When the mass effect decreases the disrupted blood brain barrier may lead to small bleedings and patchy contrast enhancement in imaging (Becker, Desch, Hacker, & Pencz, 1979).

2.6.2 CT angiography (CTA)

Main aim with CTA is confirming large vessel occlusion and demonstrating the occlusion site and thrombosis extent. Intravenous (IV) contrast agents are utilized for CTA with acquisition at the time window when peak enhancement in the vessel lumen occurs. If there is considerable thrombosis load, intraarterial thrombolysis or mechanical thrombectomy might be a better choice for revascularization. Additionally tandem lesions in ICA and aortic arch can be demonstrated which can highlight possible etiology and guide long-term management strategy (Barnwell et al., 1994).

2.6.3 CT perfusion (CTP)

CTP is critical in the decision of stratifying patients for reperfusion therapy. CTP and CTA can be performed sequentially at the same session with the same contrast by using high end CT machines. In CTP first pass of contrast agent from the intraluminal space to the capillary bed is evaluated. Considering arterial contrast density curve and using some mathematical operations cerebral blood flow (CBF), cerebral blood volume (CBV), and time to peak (TTP) contrast density in the tissue as well as mean transit time (MTT) of contrast agent from the tissue can be computed. CBF, CBV and MTT are interrelated to each other by the following equation:

$$CBF = CBV / MTT$$

The main purpose of CTP is determining the extent of infarct core and penumbra. (Srinivasan, Goyal, Al Azri, & Lum, 2006; de Lucas et al., 2008). In infarct core both CBV and MTT decreases whereas in penumbra just MTT decreases and CBV is either normal or increased due to autoregulatory mechanisms. Penumbra, which reflects potential salvageable tissue can be computed by subtracting the area with decreased CBV from the area with decreased CBF (Hopyan et al., 2010).

2.6.4 MRI

Similar to non-contrast CT, the findings of stroke in MRI is also time dependent. The most sensitive conventional sequence is FLAIR. In FLAIR hyperintensity develops in 4-6 hours whereas in T1 images this can be delayed up to 12-24 hours. The first clue before apparent signal change is blurring of gray-white matter interface which is due to edema. Intraarterial high signal in FLAIR may be observed due to either slow antegrade or collateral retrograde flow. In normal arteries there should be no signal which is referred as flow void due to dephasing of protons in rapidly flowing blood. T2* imaging or susceptibility weighted imaging can demonstrate intraarterial thrombus as blooming artefact. SWI and T2* sequences additionally can demonstrate small hemorrhage better than CT or other MR sequences due to their great sensitivity to dephasing of protons in the presence of deoxy hemoglobulin (Wycliffe et al., 2004). DWI is the most important sequence to confirm ischemic stroke in the early hyperacute period (Osborn, 2015). Ion pumps become dysfunctional in the cells in stroke area due

to unmet demand of oxygen and glucose which eventually leads to accumulation of water in the cell. Cytotoxic edema in a region decreases the extracellular free space and free diffusion of protons in that space decreases. DWI can measure this diffusion restriction by applying a pair of opposite directional gradients in a window comparable to water diffusion time. DWI coupled with FLAIR can also estimate the age of infarct in hyperacute window when unwitnessed stroke episode is encountered. If there is hyperintensity in DWI but there is no signal hyperintensity in FLAIR for a stroke patient this can be assumed as hyperacute stroke which is called as FLAIR DWI mismatch sign (Marks et al., 1996). Similar to CT, MR can also be used for vascular lumen and tissue perfusion assessment. In contrast to CT, MR angiography can be performed without contrast material administration by using special sequences such as time of flight or phase contrast.

CHAPTER THREE
PREVIOUS STUDIES

Stroke is one of the leading causes of death globally. 6.1% of total 565.594 deaths in Turkey in 2021was related to acute stroke (Topçuoğlu, & Özdemir, 2023). Every 15 minutes one person is dying in Turkey due to stroke (Topçuoğlu, & Özdemir, 2023). Furthermore, the cost of stroke to the healthcare system is great as well. According to a statistic in USA, the annual cost of stroke to the healthcare system was $56.5 billion between 2018-2019 (Ovbiagele & Nguyen-Hyunh, 2011). Due to these facts and the potential benefit from early intervention there is great interest in medical image analysis community to develop automatic stroke analysis systems from imaging (van Leeuwen et al., 2021). Digital healthcare is a rising topic among researchers (Aydın et al., 2023). 214 applications for radiological imaging AI were approved by FDA and or CE so far. Of these 214 applications 73 are related to neuroimaging and more than half of these 73 applications are related to stroke assessment which is evidence of great interest of researchers on stroke management automatization tasks (van Leeuwan et al., 2021). Classical machine learning techniques such as support vector machines (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998), random forests (Breiman, 2001), decision trees (Rokach & Maimon, 2005), logistic regression (Kleinbaum, Dietz, Gail, & Klein, 2002), linear discriminant analysis (Xanthopoulos, Pardalos, & Trafalis, 2013), and artificial neural networks (Bishop, 1994) as well as deep learning methods were utilized for these tasks. The exclusion of artefacts from the image is also an important preprocessing step since it may interfere with the quality of quantitative methods. Classical machine learning methods use hand crafted features and data science principles which hierarchically evaluate the feature subsets to find the best subset respecting to the curse of dimensionality (Nixon & Aguado, 2012). Curse of dimensionality refers to the importance of dimension of representation vector considering the number of samples in order to prevent overfitting (El Aboudi & Benhlima, 2016). Overfitting refers to learning the specific characteristics of training samples rather than general characteristics that can be applied to unseen data. An

overfit model performs greatly on the training samples however fails greatly also on the unseen external data.

Radiomics which evaluates high throughput features obtained from the images and constructs classification models based on these features extracts features basically using three types of computational methods (van Griethuysen et al., 2017). One of these methods is shape-based features which computes the boundary pixels and summarizes the characteristics of these boundary pixel distribution related to the shape of the lesion. Second set of methods is histogram based. Histogram counts the number of pixels with each intensity level (Haralick, Shanmugam, & Dinstein, 1973). The last method is second and higher order image statistics based which are gray level co-occurance matrix, gray level run length matrix, gray level size zone matrix and gray level dependence matrix (Zwanenburg et al., 2020). Each of these methods computes a specific property of the image and transforms the image into a set of numbers in the orders of thousands. These number set represents the underlying image and therefore is called feature representation vector. The feature vectors of all samples collectively form the feature space. For a classifier to work properly, the number of features (the dimension of feature space) should not be too high compared to the number of samples. Ideally for each feature there should be at least 10 samples (Lambin et al., 2017). Therefore, the number of the extracted features should be diminished before feeding it into classical machine learning classifiers. Feature selection methods are employed to discard nonrelevant features and to keep the best ones (Chandrashekar & Şahin, 2014). For this purpose, initially the features which can be reproduced well by different observers are kept. These features are called robust features (Kumar & Minz, 2014). Typically, correlation scores among observers are used for this selection. The second step is elimination of redundant features. Since the features are obtained by computing different but interrelated properties of the image they can greatly overlap or can be obtained by linear combination of the others (Kumar & Minz, 2014). These will give little information together, instead they will harm the model due to dimensionality increase. Therefore, the redundant features should be discarded typically based on intra class correlation coefficients. And finally, the most relevant a handful subset should be remained after supervised feature selection. At the end, robust, non-redundant and relevant feature subset is fed into the classifiers.

There are various studies based on classical machine learning models for stroke imaging research. A support vector machine (SVM) based classifier was used for carotid atherosclerosis detection and random forest (RF) based classifier was used to detect brain edema (Bento et al., 2019; Chen at al., 2016; Thornhill et al., 2014). Logistic regression (LR) was used for thrombus detection in CTA and artificial neural network (ANN) was used for perfusion defects in CTP (Thornhill et al., 2014; Vargas, Spiotta, & Chatterjee, 2019). Varying degrees of accuracies ranging from 85% to 97% were reached in these early studies. However, the main problem was small sample size and lack of benchmarking opportunity against a well curated public dataset. To mitigate this problem various efforts are combined to publish public datasets for serving as benchmarking datasets. Furthermore, global challenges with money prizes were arranged to accelerate the solutions. RSNA Intracranial Hemorrhage Detection challenge and dataset, ISLES ischemic stroke lesion segmentation challenge and dataset and ATLAS challenge are some examples of global outcome of these efforts (Liew et al., 2018; Winzeck et al., 2018; Flanders et al., 2020). In Turkey, TUSEB provided a CT dataset focusing on slice wise ischemic and hemorrhagic lesion classification task (Koç et al., 2022). These efforts have accelerated the production of AI applications for stroke. However, the main architecture for these applications were deep learning based rather than classical machine learning.

In deep learning a set of layers specialized in certain tasks are stacked and relevant features are extracted by this architecture (Goodfellow, Bengio, & Courville, 2016). The convolution, maximum pooling and rectified linear unit (ReLU) activation functions are backbone of deep vision models. Therefore, the models specialized on vision tasks which are heavily dependent on convolution are called convolutional neural networks (CNN) (Le Cun, Bengio, & Hinton, 2015). Convolution operation is based on matrix multiplication. A small matrix which is usually 3x3 for 2D images or 3x3x3 in 3D setting and which is called as filter traverses the image in a sliding window fashion. This filter is multiplied with the image patch under it elementwise and the results are summed. Therefore, it transforms the underlying patch based on its values. By this way edges, corners, edges and other high-level features could be captured which progressively turned into low level abstract features that represent the underlying image. Maximum pooling operation is used to find the highest peaks and

keep it in order to reduce the dimension of the feature maps, and ReLU activation function adds nonlinearity to the model to further assist complex classification tasks. 2D CNN evaluates slices and 3D CNN evaluates the entire image volume. However, in some instances for example when there is not enough samples or computational resources, the image volume may require slice by slice encoding rather than 3D encoding. Then recurrent neural networks (RNN) are utilized (He, Zhang, Ren, & Sun, 2015). RNN evaluates each slice like CNN and keeps that encoding to add to the next slice. Therefore, entire image volume can be encoded slice by slice. In these architectures a gold standard ground truth is required. These can be stroke territory classes or stroke mechanism classes among others (Lee, Kim, Kim, & Kang, 2017). This ground truth supervises the model while learning, therefore this type of training is called as supervised deep learning. Sometimes the ground truths cannot be at hand, therefore supervised training cannot be applied. Autoencoders (AE) are self-supervised deep learning architectures (Zhai, Zhang, Chen, & He, 2018). Autoencoders have two parts which are encoder and decoder. Encoder is the same as supervised CNN which encodes the image into representative vector which is called bottleneck layer. Since there is no explicit supervision signal, the input itself may behave as ground truth. For this purpose, a decoder part that converts the bottleneck features into the original image is added. Another task for medical image analysis is classification. For the classification tasks the whole image volume is predicted as a certain class. However, sometimes the pathologic region may be the region of interest and the main aim might be to segment that area from the remaining image. This type of tasks is called dense prediction tasks since each voxel in the image is predicted either a member of the region of interest or not. U-Net is the usual architecture for segmentation tasks (Ronneberger, Fischer, & Brox, 2015). For training U-Net which is a supervised network, a segmentation map is prepared for guiding the model. ITKSnap or 3D Slicer are free software to prepare these segmentation masks (Fedorov et al., 2012; Yushkevich et al., 2006). U-Net architecture is like AE since it includes an encoder and a decoder part. However, each down sampled layer is reflected to the corresponding decoder level and concatenated with its output to keep the spatial information.

Deep learning methods are widely used for stroke AI research. Choi et. al. used RNN to predict stroke with 0.94 accuracy (Choi et al., 2021). Do et. al. used a RNN with DWI to predict MR ASPECTS score with AUC of 0.94 (Do et al., 2020). Praveen, Agrawal, Sundaram & Sardesai used AE for accurate segmentation of stroke region with 0.90 accuracy (Praveen, Agrawal, Sundaram & Sardesai, 2018). Various aspects of stroke clinical management can benefit from AI automatization. LVO detection, LVO location, ICH detection, CT Perfusion analysis, Collateral assessment, ASPECT scoring, ischemic region segmentation, aneurysm detection, hemorrhage detection and classification are the main targets for AI stroke research and commercialization (Soun et al., 2021).

3.1 Early stroke detection

When a patient admits the emergency room with stroke suspicion the usual first step is acquisition of non-contrast CT to rule out hemorrhagic infarct (Demaerschalk et al., 2016).

Hemorrhage detection is an important target for stroke management. Dawud, Yurtkan, & Öztoprak conducted a study with 12.635 CT images to explore binary prediction of hemorrhage of brain and reached 93% accuracy (Dawud, Yurtkan & Öztoprak, 2019). Perreira et. al. utilized a CNN to classify 300 CT slices into normal, ischemic, and hemorrhagic stroke and reached 99% accuracy (Pereira et al., 2018).

Some studies focused on ischemic stroke detection in non-contrast CT with varying degrees of performance. However, in hyperacute setting, in less than 6 hours, CT signs are subtle for ischemic infarct diagnosis. In (Chin et al., 2017), the authors reached 90% accuracy with patch wise input and in (Marbun, Seniman, & Andayani, 2018) the authors reached 90% accuracy with slice wise input for ischemic stroke prediction. Both studies were conducted on small sample size which questions their generalization ability in larger cohorts. However, in middle cerebral artery occlusions the thrombus inside the vessel can be seen in hyperacute window (Tomsick et al., 1996). Hyperdense MCA sign is a good candidate for detection as a proxy of acute stroke. In a study reached 0.87 AUC to detect hyperdense MCA sign (Lisowska, Beveridge, Muir, & Poole, 2017). MRI, particularly diffusion weighted imaging (DWI) can identify

ischemic lesions in hyperacute phase (Yeo et al., 2021). Usual routine in Turkey is rapid acquisition of DWI for confirming or ruling out stroke and then either triage to a dedicated stroke center or onsite treatment that is based on the detected pathology. The sensitivity of DWI for the diagnosis of acute stroke is 73%-92% within 3 hours of onset (Baron et al., 2015). Time is critical for stroke management. If the treatment is initiated within 6 hours of stroke onset the best possible outcome can be obtained. Therefore, time to the stroke onset is another important criterion while assessing stroke. In radiological evaluation DWI positive FLAIR negative lesions can be regarded as in hyperacute stage (Campbell et al., 2012). If the lesion is well delineated in FLAIR, then odds are high that the onset of the target lesion is older than 6 hours. This is known as DWI-FLAIR mismatch. However, more objective, and reliable criteria are required. Shinohara et al conducted a study to detect stroke onset time and classified the patients who are within 4.5 hours of event onset with 92% precision (Shinohara et al., 2020). Zhang et. al. built a CNN to detect time after stroke onset and classified the patients within 4.5-hour window with a predictive AUC of 74% which may help for stratifying the patients with unwitnessed stroke episode (Zhang et al., 2021). Medical images have some unique properties compared to natural images. Symmetry property of the organs is among them. The symmetric organization of the brain was benefitted by researchers to diagnose acute ischemic stroke in DWI and ADC maps with AUC of 0.85 (Abedi et al., 2017, Cui et al., 2021).

Litjens et. al. built a model leveraging contralateral information to extract MCA and achieved 96% AUC (Litjens et al., 2017).

Stroke has different subtypes that require different management strategies. Therefore, the affected territory has also implications regarding the selection of treatment strategies. In another study the authors analyzed DW images with SVM and classified stroke territory according to total anterior, partial anterior and lacunar infarcts with 92% accuracy (Subudhi, Dash, & Sabut, 2018). The authors utilized transfer learning with EfficientNet B0 and MobileNetV2 to classify DWI images into MCA, watershed, and posterior circulation infarcts (Çetinoğlu, Koska, Uluç, & Gelal, 2021). They reached %93 overall accuracy. Lee et. al. designed a similar study and classified DWI images into anterior vascular zone infarcts, posterior vascular zone

infarcts and normal slices, and they reached 86% accuracy with transfer learning (Lee et al, 2023).

3.2 Large vessel occlusion (LVO):

Possible commonly known etiologies of ischemic stroke are large vessel occlusion, small vessel occlusion and cardio embolism according to TOAST classification (Goldstein et al., 2001). Large vessel occlusion can be treated by using either intravenous or intraarterial thrombolysis, particularly earlier than 6 hours window after event onset (Sweid et al., 2020). Although LVO accounts 38% of acute ischemic stroke, it is responsible for 60% of all stroke related disabilities and 90% of stroke related deaths (Powers et al., 2019). CTA is the main tool for diagnosing LVO. Amukotuwa et. al. utilized fast CTA to detect anterior circulation occlusions and reached 0.94 sensitivity and 0.76 specifity (Amukotuwa et al., 2019). Chatterjee, Somayii, & Kabakis used CNN to detect LVO and Shaham & Lederman used RNN for the same task. They reached 0.82 sensitivity and 0.94 AUC respectively to detect LVO (Chatterjee, Somayii, & Kabakis, 2019; Shaham & Lederman, 2018). Brainomix (Brainomix Ltd.), Rapid (iSchemaView), Viz (Viz.ai, California, USA) and CINA (Avicenna.ai, La Ciotat, France) are commercial products which can predict LVO. These tools also detect perfusion deficits and estimate stroke core and penumbra when they are utilized with CT perfusion images. Rapid also predicts collateral status after occlusion based on symmetry of vessel density of images (Li, Hong, Wang, & Jiang, 2021). Mc Louth et. al. used CINA to validate LVO (Mc Louth et al., 2021) and obtained 98% accuracy.

3.3 Alberta Stroke Program Early CT Score (ASPECTS SCORE)

While assessing the acute ischemia in CT, the exact extent of the lesion is difficult to discern due to poor margins initially. ASPECTS score is developed to mitigate this difficulty and to estimate severity of affection (Barber, Demchuk, Zhang, & Buchan, 2000). MCA territory infarcts account for 50% of all ischemic infarcts (Ng, Stein, Ning, & Black-Schaffer, 2007). ASPECTS score evaluates the extent of MCA infarct area by marking 10 regions of MCA feeding zone. 6 of these are hemispheric regions divided into two compartments as cranial and caudal. The three hemispheric regions

in either cranial or caudal division is further divided into 3 regions indicating anterior, middle, or posterior one third. The remaining 4 zones are caudate nucleus, lentiform nucleus, insular ribbon, and internal capsule (Aviv et al., 2007). The best score is 10 and 1 point is subtracted from the best score for each affected area to determine the final score. ASPECTS score is key to define the suitability of the patient for reperfusion therapy. Currently AHA recommends reperfusion therapy for MCA strokes with >5 ASPECTS score (Barber et al., 2000). ASPECTS score is initially defined for non-contrast CT examinations however it is extended to MRI as DWI-ASPECTS score. There are important studies in the literature and commercial products to automatically predict the ASPECTS score both from CT and DWI images. e-ASPECTS (Brainomix, Oxford, UK) and RAPID-ASPECTS (Siemens Gmbh, USA) are among them. Nagel et. al. designed a benchmark study utilizing RAPID-ASPECTS and e-ASPECTS against 2 radiologists. E-ASPECTS performed better than human in that study (Nagel et al., 2017). In contrast to the findings of Nagel et al, Goebel et. al. found that expert consensus of radiologists performed better than e-ASPECTS which was validated by follow-up imaging (Goebel et al., 2018). On the other hand, Guberina et. al. benchmarked RAPID-ASPECTS against human and found that the performance of the software is better than the human consensus (Guberina et al. 2018). The conflicting result among these studies reflects the bias in the studies which should be decreased in multi center extensive validation cohorts. The CT protocol, the age of the infarct, the experience of the radiologists, the homogeneity of imaging protocols can affect the results. However, these AI tools exhibit considerable performances which can be adopted to the workflow with secondary supervision for increasing productivity and efficiency. In a study, authors built a CNN model to calculate the CT ASPECTS score and reached 0.94 AUC for a dichotomized task that predicts ASPECTS score >5 (Naganuma et al., 2021). Fahed et. al. used RNN to predict DWI-ASPECTS and reached AUC of 0.94 (Fahed et al., 2018). Cheng et. al. compared AI DWI-ASPECTS score and consensus of senior radiologist against junior and senior radiologists and found that software performed better than junior radiologist to correlate with the labels derived from the consensus of senior radiologists (Cheng et al., 2021).

3.4 Ischemic core and penumbra evaluation:

Perfusion is an important characteristic to assess health of tissue both for brain and other organs in the body. Ischemic core is the irreversibly damaged tissue and penumbra is the potentially salvageable tissue in the setting of acute ischemic infarct (Hopyan et al., 2010). One of the main purposes of revascularization therapy is rescuing the penumbra as much as possible without further compromising the status of the patients due to reperfusion injury. Therefore, the determination of core and penumbra location and the volumes are important for patient management. In radiology practice there are two largely utilized tools to determine the ischemia core and penumbra. One is diffusion perfusion mismatch. Both diffusion restricted and perfusion deficit areas correspond to core and perfusion deficit areas without diffusion restriction correspond to penumbra. Second tool is solely perfusion imaging based. In CTP imaging the areas with delayed time to maximum peak more than 6 seconds correspond to penumbra or core. With CTP imaging just the second tool and with MR imaging both tools can be used. Since CT imaging with CTA for LVO and CTP for perfusion assessment is the main working horse of stroke assessment most of the conducted research about ischemia core and penumbra evaluation is CT based. This assessment requires segmentation of affected areas, so it is very labor intensive. The AI segmentation tools can take place here and greatly reduce the time required for segmentation compared to manual work. Furthermore, with manual segmentation the interobserver correlation is low which hampers its utilization for further downstream tasks (Yeo et al., 2021). Rapid, F-Stroke, E-Stroke, CINNA and Vitra are some commercial software for core and penumbra evaluation. Significant differences were encountered with these softwares (Koopman et al., 2019). Nevertheless, they are useful as a complementary tool. Chen, Bentley, & Rueckart used two stage cascade CNN with DWI and reached 0.67 Dice score for stroke core segmentation which is comparable to manual segmentation (Chen, Bentley, & Rueckart, 2017). In another study the authors leveraged self-supervision with AE to locate stroke core and penumbra on perfusion MRI and reached AUC of 0.68 (Ho, Speier, El-Saden, & Arnold, 2017). The predictive ability of deep networks for future unseen data is also a hot topic for AI research. Stroke is a dynamic process which evaluates in different directions in time. The final diseased area is one of the most important determinants

of patient prognosis. Therefore, some researchers attempted to predict the final stroke volume based on the initial appearances. Nishi et. al. used 3D U-Net on DWI images and predicted final stroke volume with AUC of 0.88 (Nishi et al., 2020). Yu et. al. approached the problem as a 2.5 D task. In 2.5 D approach three orthogonal slices are used to simulate 3D extent of the lesion. In their work the authors predicted the 3-7 days later infarct volume by using initial MRIs and achieved a median AUC of 0.92 (Yu et al, 2020).

3.5 Prognosis prediction

For the stroke patients and their families, the most important question is expected outcome. This information is vital for the managing clinician as well since they can advise them more confidently about expected outcome.

Modified Rankin scale (mRS) is a clinical tool which assess functional status of the patient (Bruno, Close, Gomadam, Akinwuntan, & Switzer, 2013). It is a 5-point scale with worsening status while the score is increasing (de Haan, Horn, Limburg, van Der Meulen, & Bossuyt, 1993). Heo et. al. built a CNN to predict mRS as a dichotomized target with cut off score as 2 on DWI. They reached AUC of 0.88 (Heo et al., 2019). Ding et. al. developed a CNN to predict functional outcome based on infarct volume on DWI and reached AUC of 0.97 (Ding et al., 2022).

3.6 Challenges:

The main assumption of machine learning and deep learning studies is that the training and testing data came from the same distribution. Therefore, when a model learns the relevant features of the training set it can be generalized to the test set. This is a strong assumption and requires a large, diverse, and comprehensive data collection strategy which covers the characteristics of the population as much as possible. However, the data regulation and privacy rules obviate to collect such data. Most of the research conducted so far are carried out on the data from single or a few centers with a few hundred patients at most. Therefore, all studies have bias conditioned on the distribution of the dataset at hand to some extent. This is the most important challenge that should be overcome. The solution is either collection of large samples from different regions of the world, obtained from different scanners and different

acquisition parameters or federated learning which implies using the data where it is acquired without taking them outside that center. The explainability and the trustworthiness of the models which consider the inherent uncertainty of medicine are also important challenges. The models may rely on spurious correlations other than relevant features while modelling the underlying data distribution. These models may work well with that particular data, however when data drift is encountered, they cannot produce reliable predictions. However, physicians take the responsibility of the patient, and their management decision may drive the prognosis to a favorable or less favorable direction. Therefore, they require to justify their decisions in a rational reason-outcome model. However deep models are mainly black box models with little logic to the human while making their predictions. Currently saliency maps and gradient back propagation-based solutions may help to some extent. However better designed more reliable and trusted explanation tools should be developed in order to accelerate full adoption of AI tools in radiology clinics.

CHAPTER FOUR

METHODOLOGY AND DATA ANALYSIS

Three centers involved in the model development process of this thesis. The data from each center could be collected in different time frames. Also, different strategies required different sets of data elements. Basically, three different approaches to the problem were developed, slice wise, voxel wise and patient wise approaches. Therefore, three different dataset building strategies were required which were disclosed in the following paragraphs. Institutional Review Board approval was obtained from the local ethics committee for this retrospective study. Informed patient consent was waived by the ethical committee. Center 1 was İzmir Katip Çelebi University Atatürk Training and Research hospital radiology department, center 2 was Kahramanmaraş Sütçü İmam University radiology department, and center 3 was Hatay Training and Research hospital radiology department.

4.1 Slice-based classification task

4.1.1 Patient Selection for slice-based classification task

Slice-based classification task was the product of first data collection episode. Data from center 1 was used for this task. DWIs of consecutive 291 acute ischemic stroke patients from January 2017 to April 2020 in center 1 were reviewed. 11 cases were excluded due to severe artefacts. Only 9 patients had ACA infarct and were excluded to avoid severe dataset imbalance (Figure 4.1). Remaining 271 patients with acute stroke were analyzed. Images were acquired at 1.5 T MR (MAGNETOM, Siemens Healthcare, Erlangen, Germany). Two different MR diffusion acquisition protocols were utilized (Table 4.1).

4.1.2 Dataset Construction

Stroke slices and classes of these slices were annotated by two radiologists one with 5 years and the other with 27 years of experience independently. In the cases of disagreement consensus was obtained by reevaluation. Three target classes were

determined: Middle cerebral artery (MCA), posterior circulation (PC) and watershed (WSZ) infarcts. Posterior circulation infarcts included all the infarcts arising from the vertebro-basilary system failure. Watershed infarcts included anterior, posterior, internal and external border zones between major arterial territories. 1717 stroke slices from 271 patients including 883 MCA slices from 129 MCA infarcts, 416 WSZ slices from 65 WSZ infarcts and 418 PC slices from 77 PC infarcts were obtained.

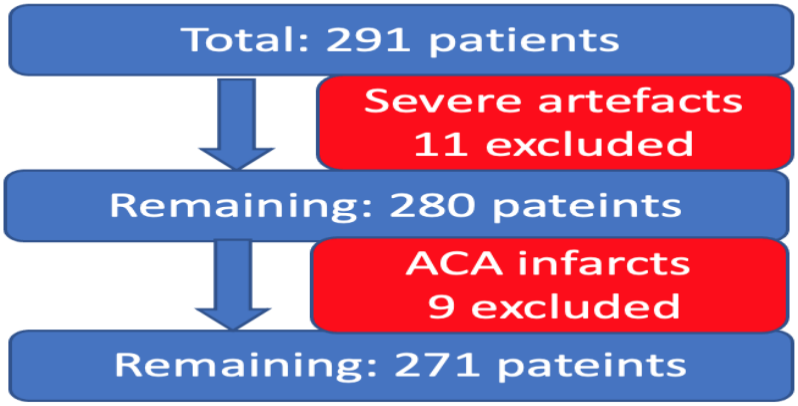


Figure 4.1 Case selection process for slice-based classification task. Of the 291 patients having DWI images with stroke between January 2017 to April 2020 in center1, 11 were excluded due to severe artefacts and 9 anterior cerebral infarcts were excluded in order to avoid severe data imbalance

Google Colab environment was used for training the models. All 1717 slices were utilized for replicating our previous study for benchmarking purposes. In accordance with our limited computational resources, as the dataset size grew too much due to applied transformations, the dataset was down sampled to a balanced 900 slices subset to fit into the memory of Colab. 10 percent of each class of dataset was reserved for validation and 10 percent for test.

4.1.3 Image preprocessing:

We have studied binary prediction of stroke presence and multiclass prediction of affected territory in a previous work (Çetinoğlu et al., 2021). We replicated the models constructed for that previous work as a benchmark. Thus, three channels of the input

were constructed by replicating the native images three times. MobileNetV2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018) and EfficientNetB0 (Tan & Le, 2019) are two of the state of the art pretrained networks which can be utilized for medical image analysis. MobileNetV2 is a light-weight architecture which is designed to be used in low computing capability mobile devices. Each slice was resampled to 224x224 and normalized to [0-1] range,

Table 4.1 DWI sequences and imaging parameters

DWI sequence name	**ep_b1000t**	**re_b1000t**
Parameters		
TR (ms)	6600	5880
TE (ms)	89	61
b-value (s/mm^2)	1000	1000
FOV (mm x mm)	230 x 230	221 x 221
Matrix	192 x 192	168 x 168
Slice thickness (mm)	5	5
NEX	3	1

TR: Repetition time, TE: Echo time, FOV: Field of view, NEX: Number of excitations

to conform to input requirements for transfer learning. Then we applied four additional transformations to construct three channels of the current study's input. First transformation involved cropping the lesion in a bounding box preserving the spatial

position (Figure 4.2). This was indeed a more explicitly encoded version of hard attention (Serra J et al., 2018) injecting prior knowledge to the network.

Second transformation was similar to the first one but included all horizontal and vertical span of the slice as a continuation of the bounding box therefore exhibiting a cross sign crop. The crops were obtained automatically using the segmentation maps as the constructor for the map. For the first transformation, minimum bounding box covering the segmentation map is cropped. For the second transformation minimum coverage of the x axis and whole y axis span and minimum coverage of the y axis and

Table 4.2 Gender and age distribution of the cases for slice wise classification task

	MCA strokes	**Posterior strokes**	**Watershed strokes**
Male	68 (52.7%)	43 (55.8%)	40 (61.5%)
Female	61 (47.3%)	34 (44.2%)	25 (38.2%)
Total	129 (100%)	77 (100%)	65 (100%)
Age (mean, range)	67.66 (32-94)	69.7 (42-89)	68.35 (37-92)

MCA: Middle cerebral artery



a) b) c) d) e)

Figure 4.2 Applied transformations a) Vertical b) Cross sign c) Binary threshold d) Transvers crop e) Edge Map

whole x axis span was used. Stroke territorial distribution prediction problem is actually a spatial location prediction problem. The spatial location of the area could be best described by its neighborhood. Therefore, we hypothesized that a model which is aware of the neighborhood might be helpful (Figure 4.2). Third transformation involved obtaining the edge map of the slices. Medical images are relatively predictable repetitive patterns of signals. Considering that, classically a radiology resident learns normal anatomy first and then determines the pathology as deflections from the normal background. Most of the pixels in a normal brain slice represent the homogenous architecture of either tissue or cerebrospinal fluid. Therefore, regarding classification of the pathology, most of the pixels are just noise for the model. High variations in the signal determines change from one tissue type to another. Therefore, edges are critical for exact spatial position description. We hypothesized that, edge map of the slices as an additional input might help the model for the correct prediction (Figure 4.3). Fourth transformation involved thresholding the image by ninety percent of the maximum value of the image and binarizing according to that threshold. As acute stroke in DWI is always represented by high signal, we hypothesized that highlighting the high signal areas and nulling other parts preserving spatial position might help the model focus on high signal components for prediction (Figure 4.2). After extensive experiments with different input combinations, we constructed two types of inputs which were the most successful for the classification of territory classes for benchmarking purposes. These were [Native DW, bounding box cropped, Cross sign cropped] and [Native DW, Edge Map, Thresholded]

4.1.4 Model architecture and training

Transfer learning was employed for our experiments. Compared to similar CNN architectures, EfficientNetB0 and MobileNetV2 are lighter models with limited entropic capacity, thus we thought they were more suitable choices for our relatively small dataset. Fully connected layers were replaced by a custom top which consisted of two layers with 512 and 256 nodes respectively and followed by drop out with 0.5 value. Final layer was a 3 node softmax layer. All the experiments were carried out with that modified architecture (Figure 4.4). After training was completed, the

additional augmentation channels were stripped, and only native images were used in test time. Therefore, the targeted augmentations helped the model learn where to focus by providing additional supervision, however test was performed with native images for fair comparison (Figure 4.5)

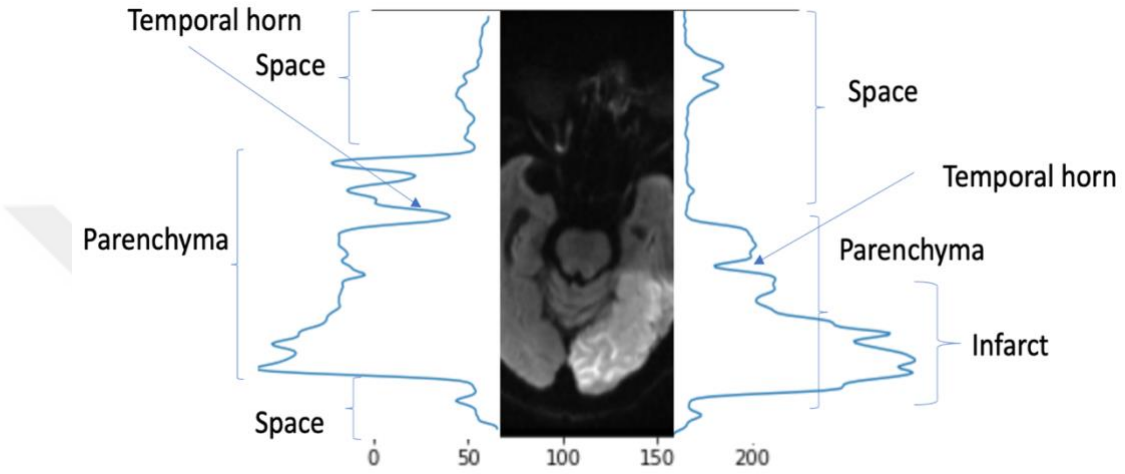


Figure 4.3 Edge profile of normal and infarct regions. 1D intensity profile of both sites of the image. Front and back ends of the image did not include any relevant anatomy and the pixels.were residing on empty space there. Therefore, relatively flat and low-level intensities were encountered. Right of the hemisphere was normal, represented by relatively monotonic signal. Note that temporal horn created sharp deflection because of the abrupt intensity change in both sites. Also note that posterior aspect of the normal site had some signal increase representing the magnetic field bias closer to the receive coil. On the other hand, left hemisphere was affected by posterior cerebral artery infarct which was represented by a sharp increase of the 1D signal posteriorly. Note that pathologic high signal exhibit minor deflections representing relatively preserved signal in cortical subcortical interfaces

4.2 Voxel-based classification task

4.2.1 Patient Selection for voxel-based classification task

The same dataset from center 1 was used for voxel-based classification task. This task required segmentation ground truth masks of the stroke areas. Only center 1 dataset included these segmentation maps. Very small lesions with volume of less than

1 cm^3 could pose great difficulties for a segmentation target with no a priori shape, color, or texture. Therefore, we decided to discard these very small lesions. Watershed region was the class with great uncertainty. Since watershed strokes reside at the border zones of major arterial territories and these territories had anatomical variability with no fixed localization, we also decided to discard watershed lesions from the dataset. DWIs of consecutive 291 acute ischemic stroke patients from January 2017 to April 2020 in center 1 were reviewed. After excluding cases with severe artefact, with < 1 cm^3 lesion volume and with watershed infarcts, 218 patients were remained (Figure 4.6). Basic demographics was provided in table 4.3.

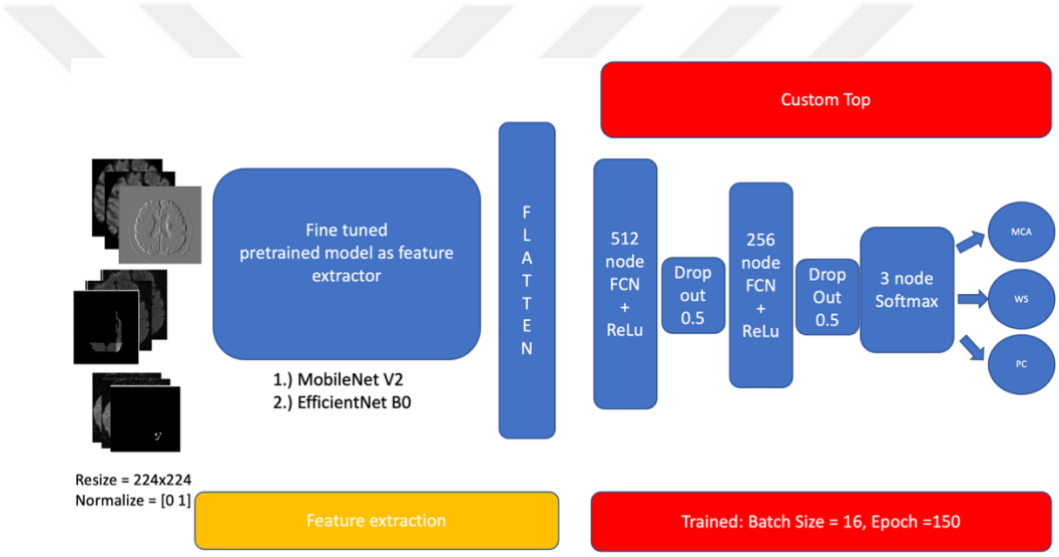
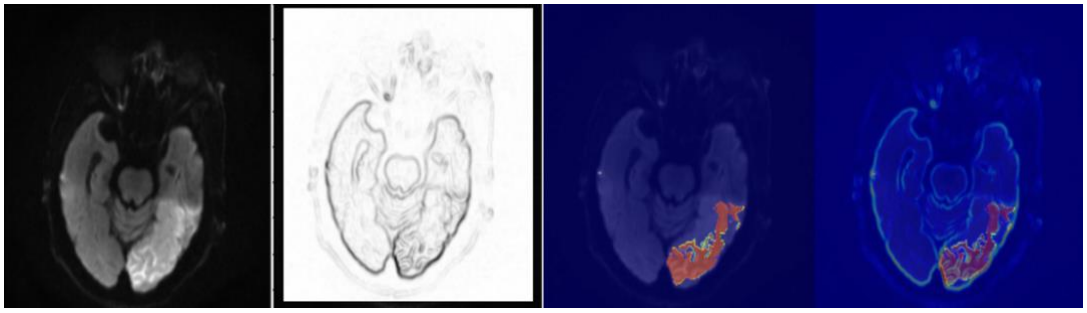


Figure 4.4 Model architectures for slice-based classification task. Pretrained MobileNetV2 and EfficientNetB0 architectures were used. Last fully convolutional layers were dropped and replaced with 3-layer custom top. Convolutional part of the pretrained models was used as feature extractor and fine-tuned. Extracted features were flattened after taking global average pooling of each feature map. 3-layer custom top was constructed by stacking 512 and 256 node fully connected layers followed by 3 node SoftMax layer. 0.5 Drop Out was applied after both fully convolutional layers FCN: Fully connected network MCA: Middle cerebral artery zone PC: Posterior circulation zone WS: Watershed zone



a) b) c) d)

Figure 4.5 Projection of the highlighted features on the native image. a) Native image b) Edge transform excluded all the monotonic data leaving only high transitions of the signal c) High signal data accentuated the relevant portions and guided the model to look for high intensity regions d) Both edge transform and relevant region projections on the image

In that task, 3D volumetric assessment of the patients instead of 2D slice-wise classification approach and voxel wise dense classification that converted the multiclass segmentation results to multiclass classification results instead of weak slice-wise classification of the patients from the previous task were used as the principal method (Figure 4.7).

4.2.2 Image preprocessing and mask preparation:

3D Slicer (version 4.10.2) was used for preparing segmentation masks (Fedarov et al., 2012). Infarct areas in the images were segmented using threshold and brush functions of 3D Slicer slice by slice. All 3D extent of the infarct zones was annotated volumetrically.

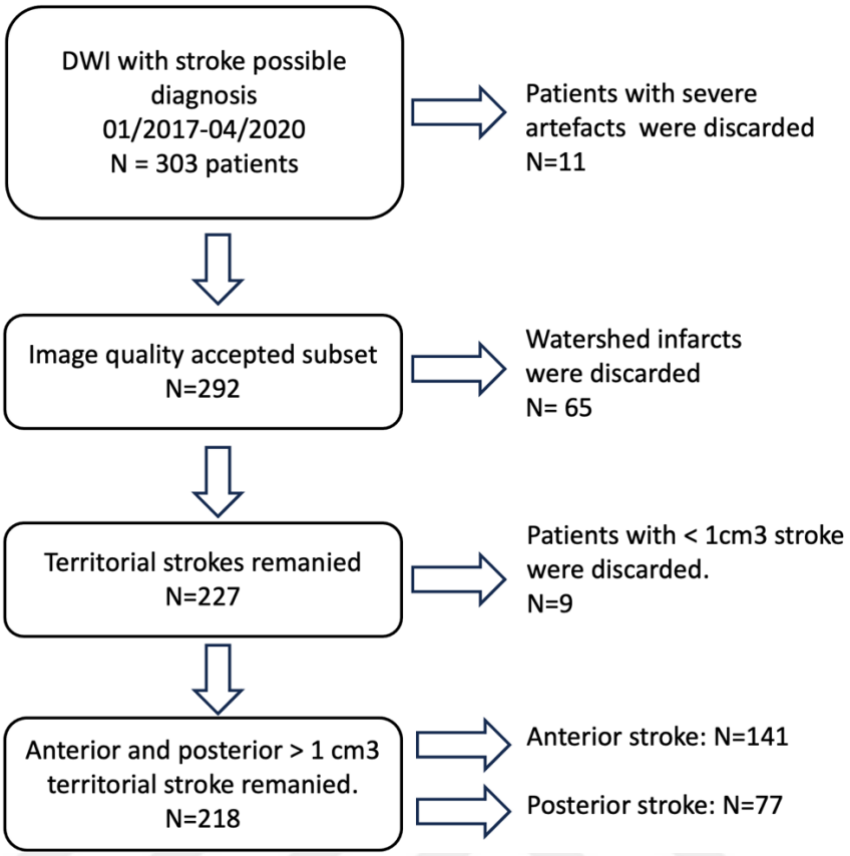


Figure 4.6 Overview of patient selection for voxel wise classification task

A radiologist with 5 years of neuroradiology experience segmented the images, labeled the territorial class and a senior neuroradiologist with 27 years of neuroradiology experience made a second evaluation. The accepted segmentation masks and final ground truth class label assignments were determined by consensus. Two sets of masks were constructed. Initially the mask labels were annotated as a single class denoted by 1. For the main objective of this study which was voxel wise dense classification of stroke territory based on segmentation annotations, a mask transformation approach was utilized. The mask annotations of anterior territory infarcts were remained as they were, and the stroke mask annotations of posterior territory infarcts were transformed into integer 2. By this way, the single segmentation masks of infarct zones were essentially turned into distinct mask objects representing different stroke regions (Figure 4.8).

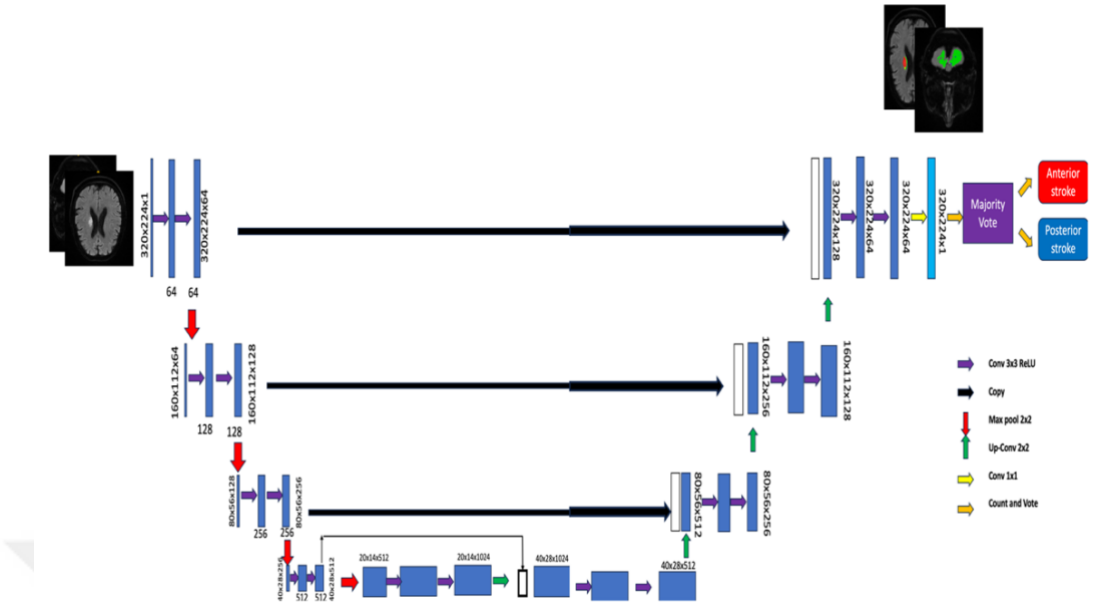


Figure 4.7 Model architecture and study workflow for voxel wise classification task

Three sets of images were obtained for benchmarking purposes. First set was the original images. Second set was N4 Bias corrected version, and third set was registered version of N4 Bias corrected images to a common template of MNI (Tustison et al., 2010, Mazziotta et al., 1995). All three image sets were exposed to the same preprocessing steps.

4.2.3 Preprocessing details

The images were clipped into 1%-99% signal range since the highest and lowest 1% might represent noise. Then they were resampled into 1.5 mm resolution which rendered the sizes of images into different dimensions, therefore the resampled images were cropped or padded into 224*320 dimension based on their transformed sizes in order to fit into fixed dimension of the built network. Same spatial operations were applied to masks keeping the affine matrices the same in order to keep matching.

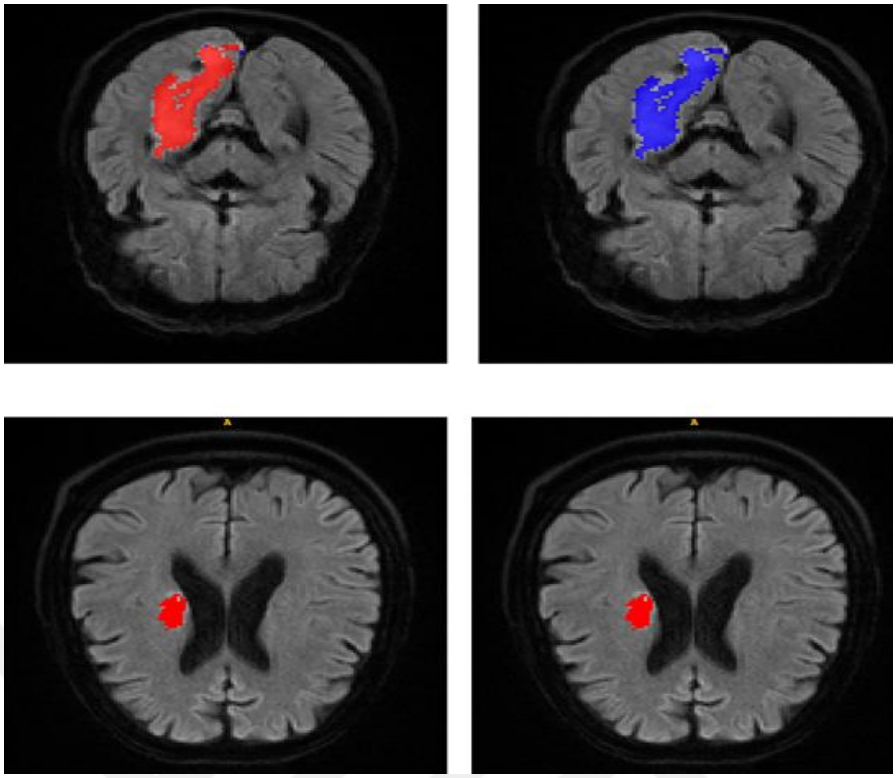
Table 4.3 Patient demographics of voxel wise classification approach

	Anterior circulation	**Posterior circulation**	**p value**
Number of patients	141 (65%)	77 (35%)	
Gender: (Number, %)			
Male	74 (52%)	43 (55%)	0.73
Female	67 (48%)	34 (45%)	
Age: Mean/SD	67.65 (15.20)	69.70 (10.89)	0.44

Finally, the signal range of the images were rescaled into 0-1 range by min-max normalization.

4.2.4 Model architecture and training:

2D U-Net like architecture which consists of an encoder part that transforms the image into low level feature vector as a bottleneck layer and a decoder which takes the bottleneck and transforms it into ground truth segmentation maps was employed (Ronneberger et al., 2015). 4 down sampling and 4 up sampling steps with 64 initial filter numbers were used. Xavier initialization with a pair of convolutions and rectified linear unit activation function followed by max pooling layer were used (Glorot & Bengio, 2018). Data augmentation consisted of just elastic deformation. Dice coefficient with background inclusion was the chosen metric and 1- Dice coefficient was the loss function for training the model. Adam optimizer with 1e-3 learning rate was utilized for training (Diederik, Kingma, & Ba, 2014). Random axial slices were selected from the dataset in each training turn (every slice of a patient remained in either train, validation, or test set) and fed into the model by batches of 20 slices with 5000 epochs.



a) b)

Figure 4.8 Overview of mask building strategy for voxel wise classification task. Mask transformations. Upper row: Patient from posterior stroke dataset; Lower row: Patient from anterior stroke dataset a) Single class mask for all stroke lesions b) Multi class mask based on territory of stroke lesion. The conversion of masks into different classes based on the stroke location which it represents, rendered them as different segmentation targets

10% of the dataset was reserved for validation and 10% for testing (19 for each). The remaining 180 volumes were used for both single class stroke region segmentation and multiclass stroke territory-based segmentation objects with separate models. Dice scores were reported as a measure of segmentation efficiency of single class model. After getting 3D multiclass segmentation maps for each patient, majority voting to the predictions was applied to evaluate the most frequent class in each volume. The most frequent class which was denoted by the largest mask was assigned as final territory class of each patient. Accuracy, sensitivity, precision and f1 scores were reported for the classification.

4.3 Patient wise 3D end to end classification task.

4.3.1 Patient selection for patient wise 3D end to end classification tasks:

Model building for patient wise 3D end to end classification task was carried out after all the data collected. Three centers from different regions of Turkey and with different MRI machines participated in the study. Basic demographic information was provided in table 4.4. DWIs of consecutive 292 adult acute ischemic stroke patients from January 2017 to April 2020 from the first center were obtained. Additional one patient was added which was missing in previous data building periods. 15 patients due to more than one territory involvement, 25 patients due to severe artefacts and 10 patients due to label uncertainty were excluded. More patients due to artefacts were excluded in this approach than previous approaches. Previous tasks were slice or voxel based. Therefore, if the diseased slices or involved voxels were artefact free, they were included in the previous tasks. However, in this task all 3D span of the imaging volume was encoded. Therefore, if any part of the image volume including the disease-free slices had severe artefacts they were excluded from the analysis. Remaining 242 patients were included in this approach. Images from the center 1 were acquired at 1.5 T MR (MAGNETOM, Siemens Healthcare, Erlangen, Germany). Two different DWI MR acquisition protocols were utilized (Table 4.5). Additionally, randomly selected 150 DWIs with normal radiological reports were added. For this dataset, an excel sheet including the clinical data was also prepared along with class labels indicating the affected territories from the ischemia.

From the second center, DWIs of consecutive 141 adult acute ischemic stroke patients from June 2020 to December 2020 were obtained. 6 patients due to severe artefacts, 23 patients due to label uncertainty, 11 patients due to large infarcts encompassing more than one territory and 14 patients due to having embolic stroke were excluded. Remaining 87 patients were included in the study. Images were acquired at 1.5 T MR (Ingenia, Philips, Eindhoven, the Netherlands) with one standard acquisition protocol (Table 4.5).

From the third center, DWIs of consecutive 151 adult acute ischemic stroke patients from November 2019 to April 2020 were obtained. 17 patients due to severe artefacts,

29 patients due to label uncertainty and 6 patients due to more than one territory involvement were excluded. Remaining 99 patients were included in the study. Additionally, 46 randomly selected DWIs with normal radiological reports were included. Images were acquired at 1.5 T MR (MAGNETOM, Siemens Healthcare, Erlangen, Germany) with one standard acquisition protocol (Table 4.5). For the second and third centers just class labels of the affected territory from ischemia were noted, and clinical information was not collected. Label uncertainty was defined by the annotating radiologist. If the annotating radiologist could not assign one of posterior circulation, watershed region, anterior cerebral artery region or middle cerebral artery region labels confidently to the image volume, this image volume was excluded due to label uncertainty. Patient selection process was summarized in Figure 4.9.

4.3.2 Class label annotation:

We determined five target classes for this approach. Stroke affecting anterior cerebral artery territory, middle cerebral artery territory, posterior circulation territory, watershed region including anterior, posterior and deep watershed zones and finally normal images were our target classes.

Two radiologists with 5 years and 27 years of neuroradiology experience (YKÇ and MFG) annotated the dataset from the first center. Additionally, the clinical signs and symptoms were collected from electronic patient records for center 1.

32 different signs such as hemiparesis or symptoms such as dizziness were recorded for each patient. Recorded symptoms and signs and their distribution were given in Table 4.6. Territory classes were annotated by a radiologist with 12 years of neuroradiology experience (ND) for the dataset from the second center and by a radiologist with 10 years of neuroradiology experience (MS) for the dataset from the third center.

Table 4.4 Patient demographics for patient wise classification task

	Center 1	**Center 2**	**Center 3**
Sex: **Count and percentage**			
Male: Stroke: Normal:	136 (56%) 70 (46%)	52 (59%) NA	61 (61%) 26 (56%)
Female: Stroke: Normal:	106 (44%) 80 (54%)	35 (41%) NA	38 (39%) 20 (44%)
Age: Mean and Interval			
Stroke: Normal:	68. 15 (32-94) 64.33 (32-95)	68.52(29-93) NA	68.14 (37-91) 65.31 (33-89)

4.3.3 Human reader study:

Clinical data was not obtained for the second and third center datasets. Two datasets among three which they didn't annotate initially were sent to each radiologist for secondary annotation of territory classes for exploring interobserver correlation.

4.3.4 Preprocessing:

The quality, size and variability of the sample dataset which should reflect the target population is one of the most important prerequisites for building performant models. In building radiology applications, it is difficult to collect adequate dataset which is the main limiting factor of the success of those models. Fortunately, there are some work arounds for the sufficient data availability problem. Data augmentation, synthetic data generation, metric learning, and heavy regularization are some important

techniques to mitigate small data problem (Candemir, Nguyen, Folio, & Prevedello, 2021). They all fall under two broad categories: Either reducing the variation in the dataset or playing with the model and the training parameters to increase their abilities to capture relevant features. In this study the variation in the dataset was reduced by using spatial and signal normalization techniques. All three image sets were exposed to the same preprocessing steps. The dataset was resampled into 1*1*1 mm resolution and cropped or padded to obtain fixed dimension of 128*128*24. The dimension of the input was selected by exploring the dimensions of the images and their mean values in the dataset. The bias induced by the magnetic field inhomogeneities were eliminated by N4ITK filter (Tustison et al., 2010). The skull was stripped by dipy median otsu method and the images were denoised (Garyfallidis et al., 2014). NYUL signal normalization was used (Reinhold, Dewey, Carass, & Prince, 2019).

NYUL is a linear piecewise histogram equalization technique. It computes the mean histogram of the training set and equalizes the histogram of each image volume depending on that mean histogram. Mean histogram of the training set was computed, and all other volumes were transformed conforming to the mean histogram of the training set. Finally, the images were spatially aligned to a common space. Montreal neurology institute (MNI 152) brain MRI template was used for spatial normalization (Mazziotta et al., 1995). And the resulting signal and spatial normalized arrays were rescaled into [0 1] interval.

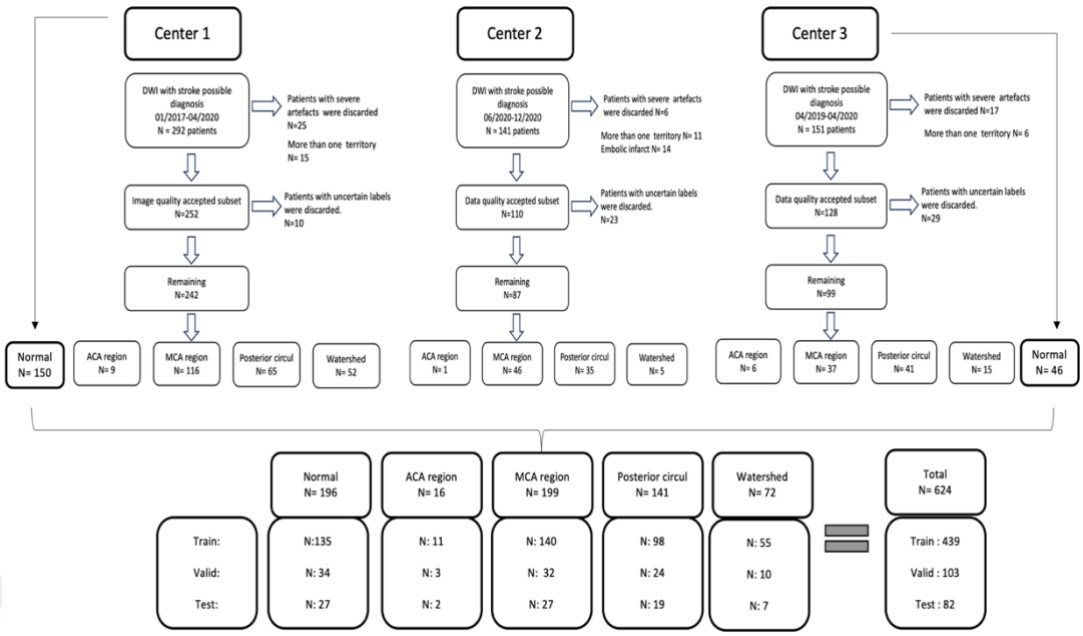


Figure 4.9 Dataset building and patient flow for patient wise approach. Adult patients from 3 centers were included in this study. The excluded patients and the reason for exclusion were given

4.3.5 Dataset building for deep networks:

All deep learning models benefit from the increase of dataset size (Cheng et al., 2021). Therefore, we designed two step approach to explore the best strategy to encode the stroke data in DWI. In the first step we used all data from three centers. We split the data into training, validation and test sets considering the class distributions. We developed two stage approach with this dataset distribution. In the first stage we explored six different deep learning approaches. Five of them were supervised models and we noted the classification performance of them. The last deep learning model was a self-supervised model which was solely used for feature extraction from its bottleneck layer.

Table 4.5 Diffusion acquisition protocol and data distribution for all centers

Center	**Center 1**		**Center 2**	**Center 3**
Machine	Siemens Magnetom		Philips Ingenia	Siemens Magnetom
Sequence	*ep_b1000t	*re_b1000t	*ep2d_diff_3scan	*re_b1000t
Parameters	6600	5800	6200	5900
TR (ms)				
TE (ms)	89	61	69	65
B values (s/mm2)	1000	1000	1000	1000
FOV (mm x mm)	230 x 230	221 x 221	229 x 229	240 x 240
Matrix	192 x 192	168 x 168	176 x 176	192 x192
Slice thickness	5	5	5	4.5
NEX	3	1	1	1
Stroke patients	91	151	87	99
ACA	3	6	1	6
MCA	39	77	46	37
Posterior Circulation	25	40	35	41
Watershed	24	28	5	15
Normal	32	118	NA	46

Table 4.6 Center 1 patients' clinical information

Symptom and Signs	Training-set: 164	Validation-set: 38	Test-set: 31	Total
Hemiparesis	88	21	14	123
Dysarthria	42	8	11	61
Facial asymmetry	33	6	1	40
Hemiplegia	14	2	2	18
Sensory aphasia	1	0	0	1
Motor aphasia	1	0	0	1
Headache	2	0	1	3
Vertigo	7	3	2	12
Global aphasia	18	6	7	31
Seizure	12	2	1	15
Babinski (+)	10	4	1	15
Ataxia	4	1	1	6
Hemi hypesthesia	14	2	5	21
Agitation	1	0	0	1
Syncope	6	1	1	8
Lethargy	4	1	0	5
Motor slowness	1	0	0	1
Diplopia	3	1	0	4
Paresthesia	1	0	0	1
Hemianopsia	10	3	3	16
Mono paresis	13	2	2	17
Ptosis	2	0	0	2
Dysmetria	3	0	1	3
Anisocoria	4	0	0	4
Quadriparesis	1	0	0	1
Vertical gaze palsy	1	0	0	1
Vision loss	0	1	0	1
Dysdiadochokinesia	0	0	1	1
Nystagmus	1	1	0	2
Gaze asymmetry	2	0	0	2
Mono hypesthesia	0	0	1	1
Global hypesthesia	0	0	0	0

4.3.6 Dataset building for pocket algorithm:

In the second stage we used all six deep learning models including the self-supervised model to extract features from their last convolutional layer. We obtained the representation vector for each sample from these models by applying global average pooling to the last feature maps of their convolutional part. In that stage we applied pocket algorithm to the representation vector of the samples. Finally, we selected best performing model and approach among these 11 different approaches (5 from the predictions of deep learning models and 6 from the predictions of pocket

algorithm which was trained using representation vectors of deep models) based on accuracy and f1 scores.

4.3.7 Dataset building for subgroup analysis:

We also explored the effect of number of classes on the performance of models. For this purpose, we applied ablation studies by reducing the number of classes to 3 and 2 sequentially by excluding normal patients initially, then combining anterior cerebral and middle cerebral artery region strokes into one class as anterior circulation infarcts, and finally excluding the watershed infarcts which was the most uncertain class and keeping just anterior and posterior circulation infarcts.

4.3.8 Dataset building for external validation:

In the second step we desired to observe the generalization ability of our best model on external dataset. For this aim we built another model with the best performing model architecture and pipeline from previous step. In this step, training, validation, and test sets were solely from center 1 with the same distribution as the previous steps.

The trained model was used for inference for center 2 and center 3 datasets in order to estimate external validation performances on two distinct datasets.

4.3.9 Dataset building for clinical information based and clinical information integrated models:

Finally, we explored the effect of clinical knowledge on the predictions. The clinical information was available for just the dataset from the first center. Therefore, we integrated the clinical vector of the patients from center 1 to the feature embeddings of the images from center 1 in a late fusion approach and trained a multi-layer perceptron (MLP) with the same architecture as the fully connected layers of previous deep models which was composed of 512 and 256 node layers. We additionally explored support vector machine (SVM), random forest (RF), Adaboost and Gaussian process learning with these embeddings and clinical vectors (Hearst et al., 1998; Breiman, 2001; Schapire, 2013; Li, 2014). For this step we used LSTM, parallel stream

and multi attention model embeddings which had the highest performances on the previous step (Hochreiter & Schmidhuber, 1997).

4.3.10 Model architectures:

The information of volumetric image stack could be encoded into latent space by several ways. Either taking the whole volume as a single tensor and directly encoding it or taking each slice as a separate instance and encoding them successively into a final representation are more commonly used approaches (Soffer et al., 2019). However, all the slices or activation maps don't contribute to the final prediction task equally. Therefore, incorporating some attention mechanisms is a plausible solution (Ba, Mnih, & Kavukçuoğlu, 2014). Considering these principles, six different end to end model building approaches were employed (Figure 4.10, 4.11). These approaches can be broadly classified under three categories, directly 3D encoding the image volume to obtain the representation vector, sequential 2D encoding of individual slices to obtain the final 3D encoded representation, and self-supervised models. Three types of directly 3D encoded models were built. One was a vanilla 3D CNN classifier, the other two were attention modules integrated version of this 3D CNN. For attention module integration two approaches were employed. In the first approach, attention module was inserted just after the input layer and in the second approach multi-level attention modules were inserted to the various parts of the 3D CNN. Two types of models which leveraged sequential encoding of individual slices to obtain final 3D encoding were built. One was a parallel stream model that encoded each slice in parallel and the other one was LSTM-CNN model that encoded the slices sequentially and integrated the distilled information to the representation vector before moving to the subsequent slice. And finally, 3D autoencoder was employed to leverage self-supervision (Zhai, Zhang, Chen, & He, 2018). Supervised models were used both for classification task and for extraction of relevant features from their convolutional part. Self-supervised model was solely used for feature extraction from the bottleneck layer. The representation feature vector obtained from either last convolutional layer of supervised models or bottleneck layer of self-supervised model were further processed by pocket algorithm (Gallant, 1990). Pocket algorithm keeps the best result which is determined by number of misclassifications in its pocket. If subsequent iterations yield

better results, the algorithm replaces its weights with the new weights, otherwise keeps the previous weights.

Convolutional block attention module (CBAM) was employed as the attention mechanism in our models. CBAM is an important method to increase the performance of the models for various computer vision tasks (Woo, Park, Lee, & Kweon, 2018). This module has two successive operations. In the first operation the channels are reweighted according to their attention score and in the second operation the distinct parts of the activation maps are reweighted. This can be assumed as slice selection and focusing on the region of interest in the selected slices. CBAM was used as attention mechanism for our models.

4.3.11 Details of the models:

The first explored architecture was vanilla 3D classifier model which we called as bland 3D model. In that architecture the whole image volume was fed into the model at once. This model had 4 down sampling steps with 3*3*3 size kernels resulting 8*8*1 dimensional latent representation followed by two fully connected dense unit with 512 and 256 nodes and 0.3 dropout factor.

The second explored architecture was what we called as multi attention model. This model was similar to the previous model however we incorporated CBAM blocks after first two down sampling steps. By this way we expected the model to focus on more relevant activation maps and more relevant region of interests among all activation maps which could help the model converge better and faster. The original CBAM block was built for activation maps of 2D images. We converted this module to 3D version by subclassing.

The third explored architecture was what we called as initial attention model. Considering that a particular region in a particular slice might harbor information about the stroke and the other slices or regions do not have that information, starting the selection process before the information mixing due to convolution operation might be plausible. For this purpose, we utilized CBAM module as a first step before any layers of vanilla classifier, hoping to start with a more refined search space.

The fourth explored model was what we called as parallel stream model. In that set up we fed the model with individual slices by the help of "Time Distributed Layer" of Keras (Chollet, 2015). By this way each slice was encoded separately by a 2D CNN. Just after the 2D encoding of each slice of a volume, CBAM block was utilized to the latent vectors of each slice and the reweighed latent vectors were further encoded by an additional convolutional layer.

The fifth explored model was Bidirectional LSTM model. In that set up the individual slices of an image volume were encoded by a 2D CNN, and the latent vectors were further encoded into their final representation by a 64-unit bidirectional LSTM layer.

And the last explored model was 3D CNN AE. The encoder part of that model was the same as the convolutional part of 3D vanilla CNN. Bilinear up sampling was used for the decoder part.

4.3.12 Training:

ADAM optimizer with 1e-3 learning rate was used for training deep models. Categorical cross entropy was the loss function and accuracy was the evaluation metric. Batch size was 8 and save best only callback which was monitored by the validation accuracy was used to ensure that the non-overfitted best model was saved. Maximum number of epochs was 10.000 however all the models converged after a few hundred epochs. Hold out training scheme was employed with stratified sampling of the patients for each split in order to preserve the class ratios in all splits.

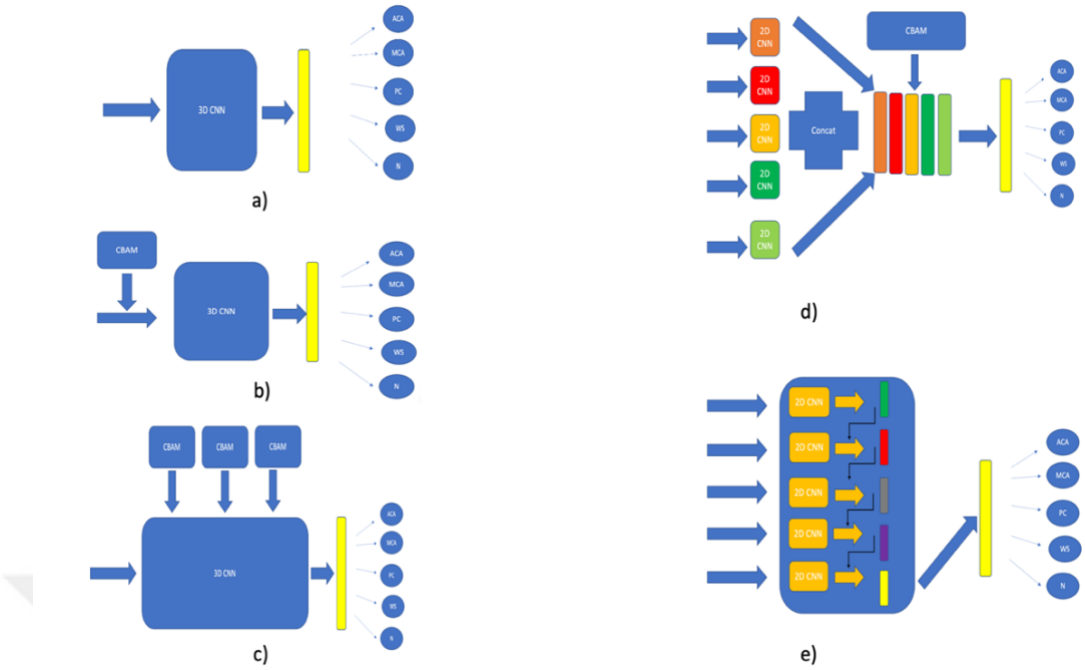


Figure 4.10 Supervised deep models for patient wise classification approach. a) Vanilla 3D CNN b) Initial attention model c) Multi attention model d) Parallel stream model e) LSTM CNN

4.4 Statistics:

Phyton (Version 3.7.3 https://www.python.org) with standard modules and matplotlib package for visualization were used. Numpy package for matrix manipulations, pandas package for dataset manipulation and scikit learn package for statistics were used. TensorFlow coupled with Keras wrapper were used for deep learning implementation. Student's t test was used for comparison of population means of ages and infarct volumes among two classes with $p<0.05$ as threshold for statistical significance. Dice scores of ground truth masks and segmentation masks were reported. Classification report function of scikit-learn package was used for reporting the performance of classification model. Cohen's kappa was used to explore interobserver agreement for human reader study. ROC curves were plotted, and AUC scores were computed where appropriate.

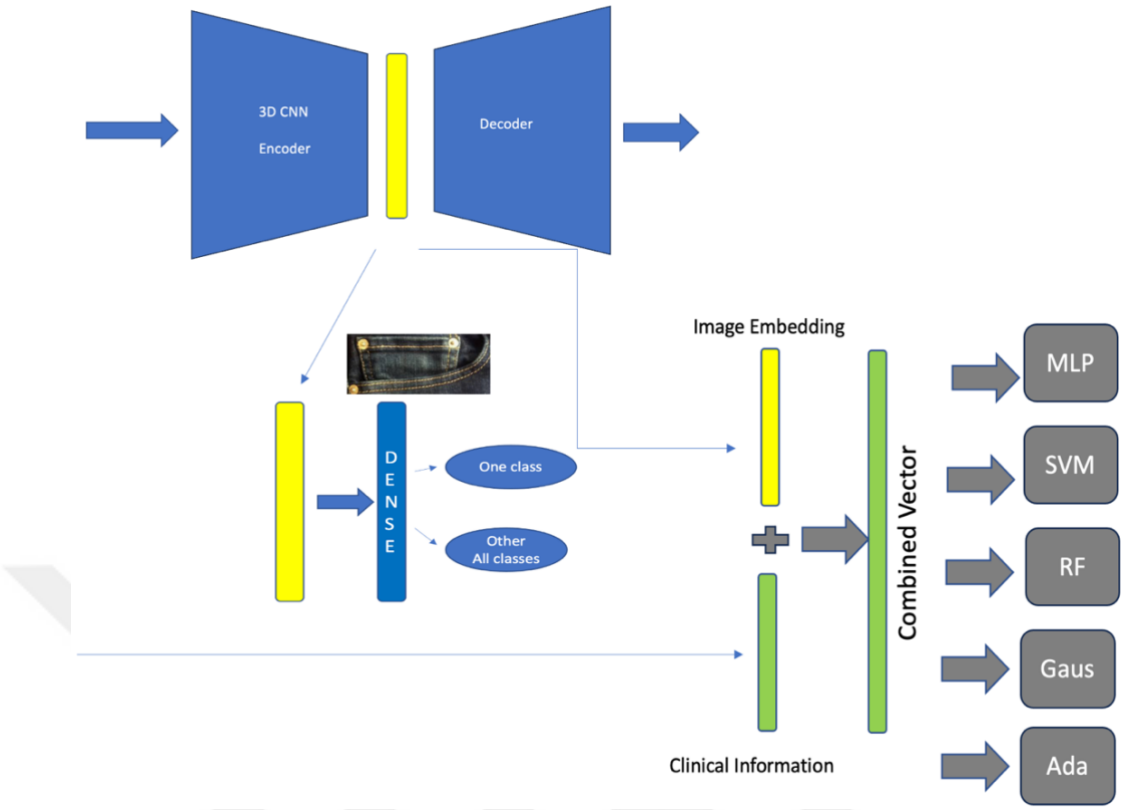


Figure 4.11 Autoencoder as the unsupervised model for latent vector (yellow rectangle) extraction.

CHAPTER FIVE

RESULTS

The results section was organized conforming to the three different approaches which were discussed in detail in Chapter 4.

5.1 Results for slice-wise classification approach

Of the 271 stroke patients included for this task, 151 (55.7 %) were male and 120 (44.3 %) were female. 129 patients (47.6 %) had MCA, 77 patients had posterior circulation (28.4%) and 65 patients had watershed (24 %) infarcts. The mean age of MCA patients was 67.66, PC patients was 69.7 and watershed zone patients was 68.35 (Table 4.2). In our previously reported multiclass stroke territory prediction with three channel native input, overall accuracy of modified MobileNetV2 model was 93 %. This model achieved 95% accuracy, and 0.93 F1 score for MCA stroke, 96.3 % accuracy, 0.94 F1-score for PC stroke, 94.7 % accuracy and 0.92 F1-score for WSZ stroke. Modified EfficientNetB0 had slightly inferior performance values, with 87 % overall accuracy. This model achieved 90.7 % accuracy and 0.87 F1-score for MCA stroke, 94.7 % accuracy, and 0.92 F1-score for PC stroke, and 88.7 % accuracy and 0.83 F1-score for WSZ stroke. Besides, both models had excellent concordance between predictions and true labels with 0.90 kappa value for MobileNetV2 and 0.81 kappa value for EfficientNetB0 (Table 5.1)

Multiclass stroke territory prediction with novel input augmentation strategy was the main contribution of the current study. For this purpose, [Native DW, Bounding box crop, Cross sign crop] prepared input produced impressive results with 94.4 % overall accuracy with modified MobileNetV2 model. This model achieved 95 %, 94 % and 95 % accuracy with 0.95, 0.94 and 0.95 F1 scores for MCA, posterior and watershed infarct zones respectively.

Modified EfficientNetB0 had slightly inferior performance values, with 86 % overall accuracy. This model achieved 86 % accuracy and F1 scores for MCA and posterior zone prediction, and 0.85 F1 score for watershed zone prediction.

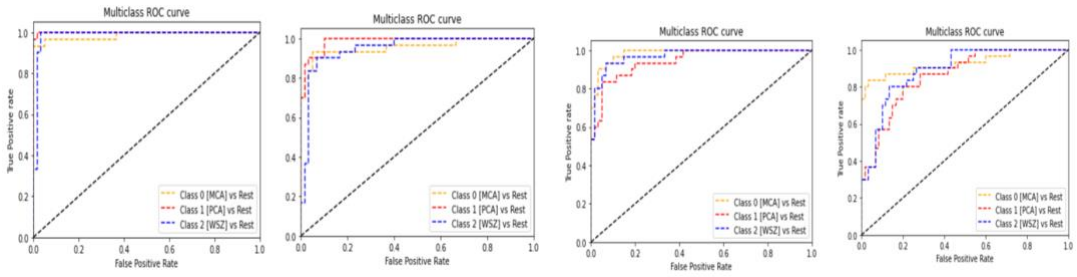
Table 5.1 Evaluation of pretrained classifiers with native images for vascular territory prediction

Model	**MobileNetV2**			**EfficientNet-B0**		
Territory	MCA	PC	WS	MCA	PC	WS
Sensitivity	98%	92%	89%	91%	91%	79%
Specifity	93.5%	98.5%	97.5%	90.5%	96.5%	93.5%
PPV	88.3%	96.9%	94.7%	82.7%	92.8%	85.9%
NPV	98.9%	96.1%	94.7%	95.2%	95.5%	90%
Accuracy	95%	96.3%	94.7%	90.1%	94.7%	88.7%
F1-score	0.93	0.94	0.92	0.87	0.92	0.82
Mean accuracy	93%			87%		

MCA: Middle cerebral artery**,** PC: Posterior circulation, WS: Watershed region, PPV: Positive predictive value, NPV: Negative predictive value

On the other hand, multiclass stroke territory classifier with [Native DW, Edge map, Thresholded] input, contrary to our expectation produced moderate results. Overall accuracy with modified MobileNetV2 model was 82 %. This model achieved 85 % accuracy, for MCA and 82 % accuracy for posterior and watershed zones. F1 scores for MCA, posterior and watershed zones were 0.90, 0.75 and 0.82 respectively. Besides, modified EfficientNetB0 with this type of input had the lowest overall accuracy. This model achieved 76 % accuracy for overall, posterior and watershed infarcts and 77 % accuracy for MCA infarct. F1 scores were 0.86, 0.72 and 0.69 respectively. We presented performance metrics as a table (Table 5.2). We performed

ROC-analysis as well to evaluate how performant our models were. (Figure 5.1) and displayed performance metrics as bar plot as well (Figure 5.2). ROC-AUC were 0.98 for MCA and watershed infarct predictions and 0.99 for posterior zone prediction with MobileNetV2 classifier with [Native DW, Bounding box crop, Cross sign crop] input. Lowest AUC values were obtained with EfficientNetB0 classifier with [Native DW, Edge map, Thresholded] input. (AUC = 0.86 for posterior infarct zone territory). Furthermore, we applied McNemar test to check if the models behaved in a similar manner while making mistakes. McNemar test for both MobileNetV2 and EfficientNetBo classifiers with [Native DW, Bounding box crop, Cross sign crop] input and MobileNetV2 for either input types yielded statistically significant p value ($p < 0.05$).



a) b) c) d)

Figure 5.1 ROC plots of the augmented input models. […] denotes input type a) MobilenetV2 [N, C, Cs]. b) MobilenetV2 [N, E, Th] c) EfficientNetB0 [N, C, Cs]. d)EfficientNetB0 [N, E, Th] N: Native image C: Bounding Box Cs: Cross sign E: Edge map Th: Thresholded image

5. 2. Results for voxel-wise classification approach

5.2.1 Patients:

Of the 218 patients, 141 (65 %) were anterior stroke, and 77 were posterior stroke (35 %). Of the 141 cases in the anterior stroke group, 74 (52%) were male and 67 (48%) were female and in the posterior stroke group 43 (55%) were male and 34 (35%) were female. The mean age of the patients in the anterior stroke group was 67.65 (SD:

Table 5.2 Evaluation of augmented input pretrained classifiers for territory prediction

Model	Territory	AUC	F1	Accuracy	PPV	Sensitivity
MobileCrop	MCA	0.98	0.95	0.95	0.97	0.93
	Posterior	0.99	0.95	0.94	0.97	0.93
	Watershed	0.98	0.94	0.94	0.91	0.97
MobileEdge	MCA	0.98	0.90	0.85	0.93	0.87
	Posterior	0.94	0.75	0.82	0.90	0.63
	Watershed	0.97	0.82	0.82	0.71	0.97
EfficientCrop	MCA	0.95	0.86	0.86	0.89	0.83
	Posterior	0.98	0.86	0.86	0.82	0.90
	Watershed	0.95	0.85	0.86	0.86	0.83
EfficientEdge	MCA	0.92	0.86	0.77	0.89	0.83
	Posterior	0.86	0.72	0.76	0.65	0.80
	Watershed	0.89	0.69	0.76	0.76	0.63

MCA: Middle cerebral artery, AUC: Area under curve, PPV: Positive predictive value

MobileCrop: MobileNetV2 with [Native DW, Bounding box crop, Cross sign crop]

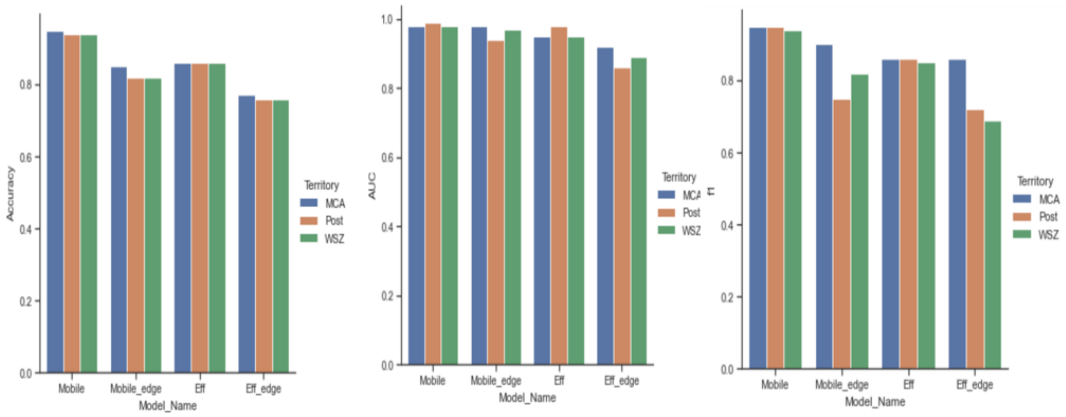
MobileEdge: MobileNetV2 with [Native DW, Edge map, Thresholded]

EfficientCrop: EfficientNetB0 with [Native DW, Bounding box crop, Cross sign crop]

EfficientEdge: EfficientNetB0 with [Native DW, Edge map, Thresholded]

Mobile: MobileNetV2 with [Native image, Bounding box crop, Cross sign crop] input

Mobile_edge: MobileNetV2 with [Native image, Thresholded image, Edge map] input



a) b) c)

Figure 5.2 Bar plots of performance metrics for slice-based models. a) Accuracy of the models for MCA, posterior circulation, or watershed zones b) AUC of the models for MCA, posterior circulation, or watershed zones c) F1 score of the models for MCA, posterior circulation, or watershed zones

Eff: EfficientNetB0 with [Native image, Bounding box crop, Cross sign crop] input

Eff_edge: EfficientNetB0 with [Native image, Thresholded image, Edge map] input

MCA: Middle cerebral artery, Post: Posterior circulation, WSZ: Watershed zone

15.20, the lowest: 32 - the highest: 94) and in the posterior stroke group was 69.70 (SD:10.89, the lowest: 42- the highest: 89). There were no statistical differences in two groups either for age or gender (p values were 0.73 and 0.44 respectively). All the demographic data were shown in Table 4.3.

5.2.2 Segmentation models:

The mean dice score of the segmentation model was 0.71 with a slightly better performance in anterior stroke segmentations (Mean dice: 0.73) than posterior stroke segmentations (Mean dice:0.68), however without statistically significant difference (p= 0.2) (Table 5.4, Figure 5.4). The mean stroke volume of the population was 32.29 cc (SD: 41.25) with greater volumes in anterior circulation (Mean: 41.32, SD: 46.75) than posterior circulation (Mean:15.63, SD:19.65) (p< 0.01).

For the multiclass segmentation the overall dice scores were lower, 0.46 for original images, 0.49 for N4 Bias corrected version and 0.51 for N4 Bias correction followed

by registration to a common space version. However, the majority of pixel wise predictions of all three models were still sufficient to infer the territorial class.

5.2.3 Classification models based on aggregation of voxel wise predictions:

After converting the pixel wise dense predictions of multi class segmentation results into territorial class labels by applying majority voting to the masks, we computed confusion matrices of final predictions. The model built with original images reached.

Table 5.3 Stroke volumes. The values are cm^3

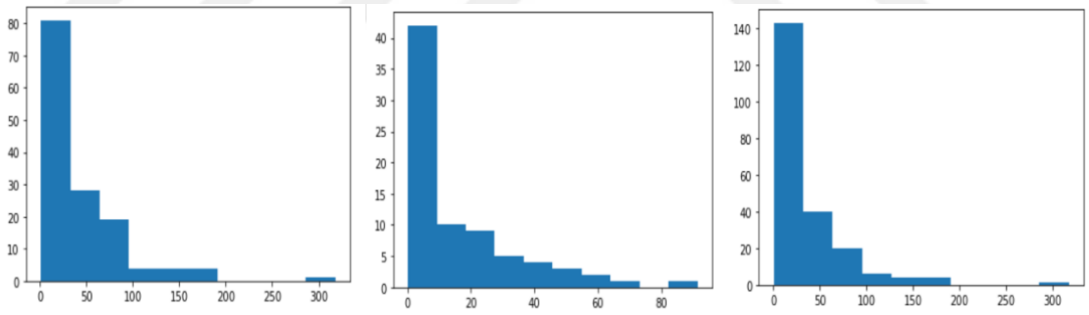
	Mean:	**SD:**	**Maximum:**	**Median:**	**Number of patients >50 cc stroke volume**
Anterior:	41.39	46.75	317.71	13.77	41
Posterior:	15.63	19.65	91.31	4.48	7
Total:	32.29	41.25	317.71	17.97	48

Table 5.4 Dice scores of ground truth segmentations and model predictions for ordinary single class model for the test set.

	Mean	**Std**	**Median**	**Max**	**Number >0.7 Dice**	**Number <0.5 Dice**	**Number 0.5<Dice<0.7**
Anterior	0.73	0.13	0.74	0.93	92	9	40
Posterior	0.68	0.18	0.73	0.91	39	10	28
Total	0.71	0.15	0.74	0.93	131	19	68

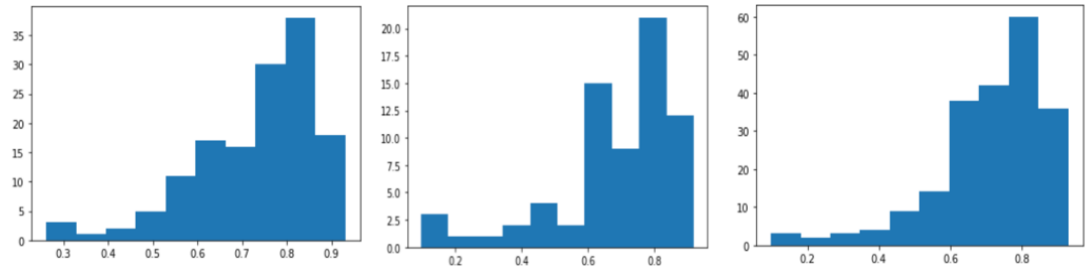
Table 5.5 Classification metrics of the proposed approach. Tr: Train Val: Validation. Ts: Test

	Precision **Tr/Val/Ts**	**Recall** **Tr/Val/Ts**	**F1 score** **Tr/Val/Ts**	**Accuracy** **Tr/Val/Ts**
Original Images	0.81/ 0.80/ 0.76	0.99/ 0.99/ 0.98	0.89/ 0.88/ 0.86	0.82/ 0.81/ 0.77
N4 Bias Field **Corrected Images**	0.84/ 0.85/ 0.79	0.99/ 0.97 /0.97	0.91/ 0.90/ 0.87	0.84/ 0.84 0.80
N4 Bias Field Corrected **and Registered Images**	0.87/ 0.85/ 0.81	0.99/ 0.99/ 0.98	0.93/ 0.90/ .89	0.87/ 0.86/ 0.83



a) Anterior stroke b) Posterior stroke c) Whole dataset

Figure 5.3 Histogram of stroke volumes in anterior stroke, posterior stroke and total dataset. The anterior stroke lesions had statistically significantly larger mean volume compared to posterior stroke lesions



a) Anterior stroke b) Posterior stroke c) Whole dataset

Figure 5.4 Histogram of Dice scores of ordinary segmentation models for anterior stroke, posterior stroke and total dataset

0.76 precision, 0.98 recall, 0.86 f1 score and 0.77 accuracy whereas the model built with N4 bias corrected images reached 0.79, 0.97, 0.87, 0.80 and the model built with images which were N4 bias corrected and then registered into a common space reached 0.81, 0.98, 0.89, and 0.83 values respectively. The performance metrics of the built models were provided in Table 5.5. Bias field correction greatly increased the segmentation and subsequent classification performance (Figure 5.5)

5.3 Results for patient wise 3D end to end classification approach:

5.3.1 Patients:

624 patients from three centers were included in this approach which were split into training, validation and test sets as 439, 103 and 82 respectively. Of 392 samples from center 1, 242 were stroke patients (61%) and 150 were normal (39%). Of 242 stroke patients 136 were male (56%) and 106 were female (44%) and of 150 normal subjects 70 were male (46%) and 80 were female (54%). The mean age for the patients with stroke was 68.15 (Range: 32-94) and for the normal samples mean age was 64.33 (Range: 32-95). Of the 87 patients from center 2, all were stroke patients with 52 male(59%) and 35 female (41%). The mean age of center 2 patients was 68.52 (Range: 29-93). Of the 145 samples from center 3, 99 were stroke patients (68%) and 46 were normal subjects (32%). Of the 99 stroke patients from center 3, 61 were male (61%) and 38 were female (39%) while of the 46 normal subjects 26 were male (56%), 20 were female (44%). The mean age of patients with stroke from center 3 was 68.14

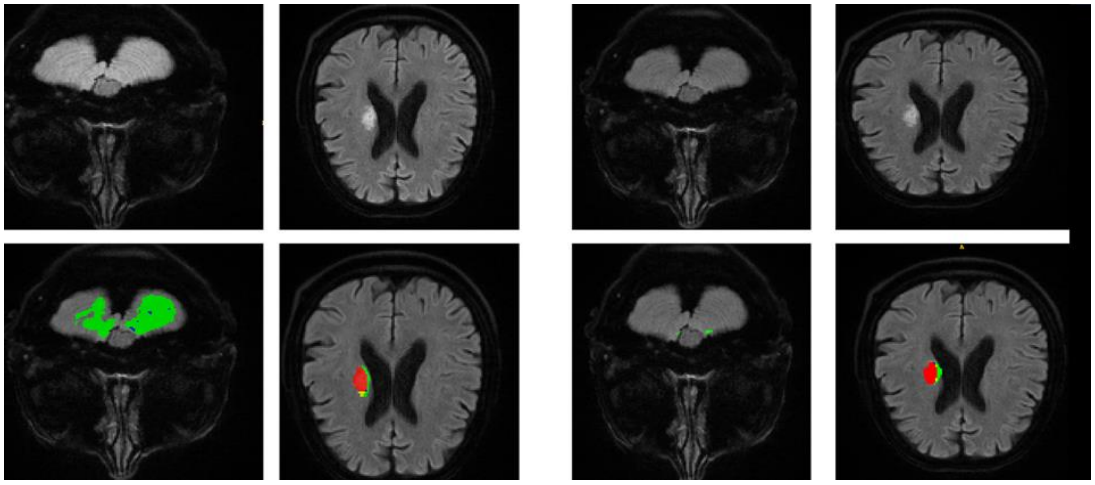
(Range: 37-91) and that of normal subjects was 65.31 (Range: 33-89). The basic demographic information was provided in Table 4.4 and the clinical information for center 1 patients was provided in Table 4.6.

5.3.2 Human reader study:

The images were labelled in the center of acquisition by the staff radiologists and in order to explore the interobserver agreement they were sent to the staff radiologists of other two centers. Pairwise agreement was good between center 1 and center 2 radiologists with 0.87, center 1 and center 3 radiologists with 0.89 and center 2 and center 3 radiologists with 0.86 kappa values.

5.3.3 Deep models:

The best test accuracies were obtained with LSTM model and parallel stream model. The test accuracies for these models were 0.84 and 0.88 respectively. Mean f1 scores of both models were 0.80. AUC were also calculated for parallel stream, LSTM and multi attention models which were the most successful ones for five class classification tasks (Figure 5.6) AUC scores were 0.98, 0.95 and 0.92 for parallel stream, LSTM and multi attention models respectively. Both LSTM and parallel stream models were based on slice by slice encoding strategies either sequential or in parallel. In general, directly 3D encoding strategies either using attention modules in various levels or without attention gates did not work well for our five-class dataset. 3D model without attention gates could only reach 0.56 test accuracy.



a) b)

Figure 5.5 Predictions of the models before and after bias field correction and registration. Upper row: images, Lower row: Mask overlays on top of images. a) Anterior stroke patient. The predictions which were displayed on the lower row correctly predicted most of the voxels for anterior stroke segmentation. However, there were great many false positive predictions for the posterior circulation due to high signal induced by the bias field. b) Same patient after bias field correction. The bias field corrected images display darker signal now which is particularly evident in both cerebellar areas

Interestingly adding attention gates in the initial layers affected the performance of the model worse, however adding multilevel attention units after initial layer added considerable performance gain with 0.77 test accuracy. LSTM and parallel stream models learnt the MCA and posterior circulation infarcts much better than watershed and ACA infarcts. The test set f1 scores for posterior circulation infarcts in parallel stream model and LSTM were 0.94 and 0.88 respectively (Table 5.6). T-SNE and principal component analysis (PCA) are dimensionality reduction techniques which are also useful to visualize class boundaries in lower dimensional feaure space. T-SNE and PCA plots of parallel stream model clearly demonstrated 3 distinct clusters which was interpreted as evidence of quality of embeddings (Figure 5.7).

5.3.4 Subgroup analysis with three class data and deep models:

Multi attention model reached 0.84 test accuracy with three class data. F1 scores for anterior and posterior circulation were 0.86 and 0.90 with this model with good

performance, however f1 score for watershed infarcts with this model was 0.57. Similar trend was observed in all models except LSTM, with 0.70-0.80 f1 scores for anterior and posterior circulation infarct prediction, and 0.35-0.60 f1 scores for watershed infarcts. LSTM model could not learn this dataset at all which showed low performance metrics (Table 5.6).

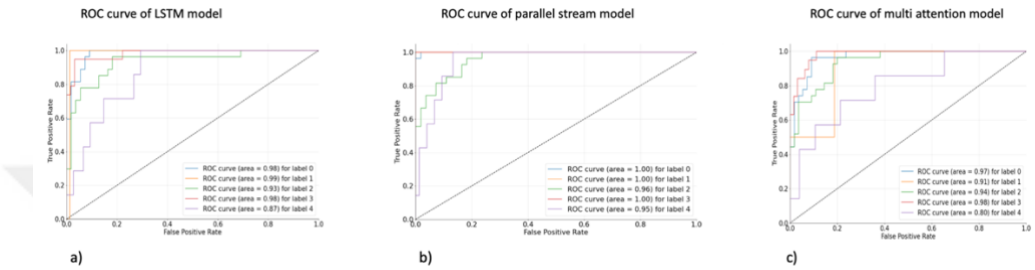


Figure 5.6 ROC curves of a) LSTM b) Parallel stream and c) Multi attention models

5.3.5 Subgroup analysis with two class data and deep models:

Adding attention gates to the 3D model both in multiple levels and in initial layer increased the performance of the models for anterior or posterior circulation infarcts classification. Without attention gates, the test accuracy was 0.81 for 3D network which eventually increased after attention gates integration. The performance of LSTM was moderate with 0.83 accuracy, while parallel stream model had the lowest accuracy with 0.75 with this dataset (Table 5.6).

5.3.6 Pocket algorithm:

Contrary to deep models for five classes, direct 3D encoding strategy worked better with pocket algorithm. The best performing model was bland 3D model with 90% mean test accuracy, followed by multi-level attention and initial attention models both with 89% accuracy. Mean test accuracy of LSTM and parallel stream models were 88% and 87% respectively. Additionally, a pocket algorithm with autoencoder derived

features was built which reached 89% accuracy. Mean f1 scores of all 3D encoding models were 0.93 while slice based encoding models were 0.92 (Table 5.7).

Table 5.6 Performance metrics of deep networks

Deep Networks		Accuracy						F1 score					
		Train		Valid		Test		Train		Valid		Test	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Bland 3D	5 class	0.74	0.05	0.52	0.06	0.56	0.06	0.77	0.17	0.45	0.35	0.44	0.32
	3 class	0.99	0.01	0.64	0.06	0.71	0.05	0.99	0.01	0.57	0.19	0.63	0.24
	2 class	0.95	0.02	0.68	0.06	0.81	0.05	0.94	0.02	0.69	0.07	0.78	0.10
Multi attention	5 class	0.99	0.01	0.75	0.05	0.77	0.06	0.98	0.01	0.68	0.14	0.67	0.19
	3 class	0.99	0.01	0.65	0.09	0.84	0.06	0.99	0.01	0.61	0.21	0.77	0.18
	2 class	0.96	0.02	0.83	0.07	0.96	0.02	0.96	0.02	0.81	0.08	0.96	0.02
Initial attention	5 class	0.59	0.08	0.37	0.21	0.38	0.19	0.56	0.37	0.30	0.31	0.31	0.36
	3 class	0.99	0.01	0.60	0.09	0.71	0.08	0.98	0.01	0.54	0.19	0.64	0.18
	2 class	0.99	0.01	0.75	0.07	0.79	0.06	0.99	0.01	0.79	0.03	0.78	0.03
LSTM	5 class	0.96	0.02	0.84	0.05	0.84	0.05	0.95	0.03	0.79	0.07	0.80	0.12
	3 class	0.62	0.10	0.58	0.11	0.56	0.12	0.61	0.03	0.50	0.23	0.50	0.16
	2 class	0.85	0.05	0.73	0.08	0.83	0.06	0.82	0.09	0.67	0.19	0.80	0.10
Parallel stream	5 class	0.98	0.01	0.82	0.06	0.88	0.04	0.99	0.01	0.75	0.07	0.80	0.20
	3 class	0.94	0.03	0.62	0.12	0.73	0.09	0.93	0.05	0.60	0.17	0.67	0.16
	2 class	0.87	0.06	0.70	0.08	0.75	0.07	0.85	0.06	0.69	0.12	0.76	0.08

5.3.7 Subgroup analysis with pocket algorithm:

In three class experiments overall performance was not better than whole dataset. 0.76, 0.85 and 0.77 mean test accuracies were obtained with 3D bland, multi attention and initial attention models' embeddings respectively. 0.74, 0.73 and 0.69 mean test accuracy rates could be reached with the embeddings of LSTM, parallel stream and autoencoder models respectively. Interestingly, direct 3D encoding models were more successful for posterior circulation infarcts with 0.91 accuracy as the highest rate with

multi attention model while slice by slice encoding models were more successful at predicting watershed infarcts with 0.87 accuracy with both LSTM and parallel stream models. For the two class experiments, autoencoder based model was the most successful model with 0.87 accuracy followed by initial attention model and multi attention model which had 0.85 and 0.81 accuracies respectively. Similar trend of decreasing performance with slice by slice encoding models while using pocket algorithm was observed in two class experiments as well. The accuracy with embeddings of LSTM model was 0.69 and that of parallel stream model was 0.59 (Table 5.7)

5.3.8 Clinical information based and clinical information integrated models:

We trained a MLP model with the clinical information driven data. The training accuracy was 0.73 and the test accuracy was 0.70 with this model. When we combined the clinical vector with image embeddings the performance of the models increased. With a MLP model and LSTM based embeddings 0.90 test accuracy was reached while with the parallel stream model embeddings 0.80 and multi attention model embeddings 0.77 test accuracies were reached. Gaussian-process learning model reached 0.90, 0.80 and 0.87 test accuracies with LSTM, parallel stream and multi attention model embeddings respectively. Random Forest model reached 0.80, 0.87 and 0.84 test accuracies while SVM model reached 0.93, 0.87 and 0.87 test accuracies with LSTM, parallel stream and multi attention model embeddings respectively (Figure 5.8).

5.3.9 External validation with best performing model and pipeline:

We also explored the generalization ability of our model for unseen data. For this purpose, we trained a new model with best performing architecture and the data from center 1. All the data from center 2 and center 3 were used for inference.

Parallel stream model as the winner of full data approach with five classes, LSTM model as the second winner and multi attention model as the winner of three class predictions were selected for retraining. The training data which was exclusively from center 1 included 164 samples, and the validation and test data included 38 and 31

samples respectively. The external test data from center 2 had 86 samples and from center 3 had 93 samples.

Both the performance metrics of internal and external validation data were poor with 0.68 accuracy for internal test set and 0.58 and 0.53 accuracy for center 2 and center 3 data respectively (Figure 5.9).

Error analysis of predictions showed that majority of the mispredictions were due to small lesions with low interrater agreement (Figure 5.10).

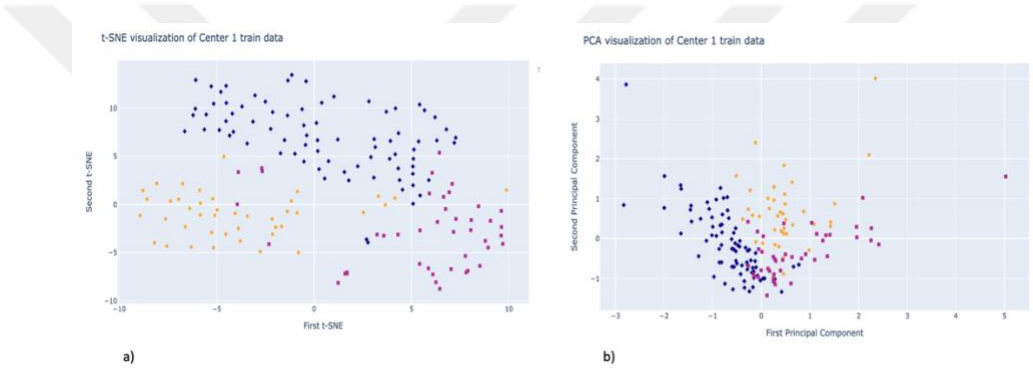


Figure 5.7 t-SNE and PCA plots of parallel stream model embeddings

Table 5.7 Performance metrics of pocket algorithm with the embeddings of deep models

Pocket algorithm		Accuracy						F1 score					
		Train		Valid		Test		Train		Valid		Test	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Bland 3D	5 class	0.98	0.01	0.88	0.08	0.90	0.05	0.98	0.01	0.91	0.06	0.93	0.03
	3 class	0.98	0.01	0.76	0.04	0.76	0.05	0.98	0.01	0.81	0.07	0.82	0.08
	2 class	0.98	0.01	0.75	0.04	0.78	0.04	0.98	0.01	0.72	0.04	0.78	0.04
Multi attention	5 class	0.96	0.03	0.88	0.06	0.89	0.06	0.98	0.01	0.93	0.04	0.93	0.04
	3 class	0.94	0.05	0.82	0.06	0.85	0.08	0.96	0.02	0.86	0.07	0.86	0.10
	2 class	0.97	0.01	0.75	0.03	0.81	0.03	0.97	0.01	0.68	0.3	0.82	0.03
Initial attention	5 class	0.98	0.01	0.87	0.06	0.89	0.05	0.98	0.01	0.91	0.04	0.93	0.03
	3 class	0.98	0.01	0.73	0.02	0.77	0.02	0.98	0.01	0.79	0.07	0.82	0.06
	2 class	0.98	0.01	0.73	0.04	0.85	0.04	0.98	0.01	0.79	0.04	0.80	0.04
LSTM	5 class	0.93	0.07	0.87	0.07	0.88	0.08	0.96	0.04	0.92	0.04	0.92	0.05
	3 class	0.72	0.08	0.73	0.10	0.74	0.11	0.82	0.07	0.83	0.08	0.88	0.09
	2 class	0.79	0.06	0.63	0.05	0.69	0.05	0.79	0.05	0.68	0.06	0.72	0.06
Parallel stream	5 class	0.92	0.13	0.86	0.11	0.87	0.13	0.95	0.07	0.91	0.07	0.92	0.07
	3 class	0.70	0.11	0.72	0.11	0.73	0.12	0.78	0.12	0.76	0.14	0.75	0.15
	2 class	0.61	0.06	0.58	0.08	0.59	0.06	0.71	0.07	0.70	0.07	0.66	0.08
Autoencoder	5 class	0.88	0.09	0.86	0.09	0.89	0.07	0.88	0.07	0.86	0.08	0.86	0.08
	3 class	0.69	0.02	0.68	0.03	0.69	0.01	0.77	0.03	0.74	0.03	0.75	0.04
	2 class	0.87	0.03	0.85	0.04	0.87	0.03	0.87	0.04	0.85	0.05	0.84	0.04

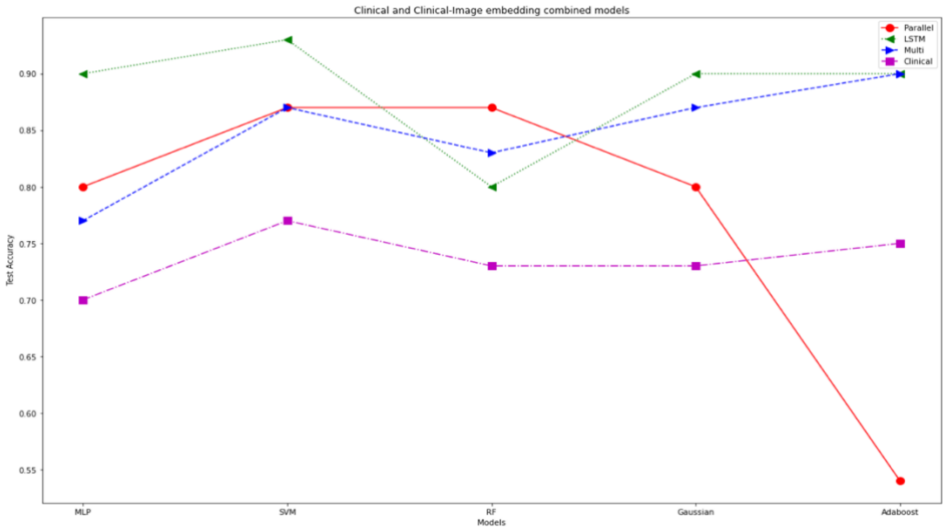


Figure 5.8 Test accuracy plots of classical machine learning models with image embeddings

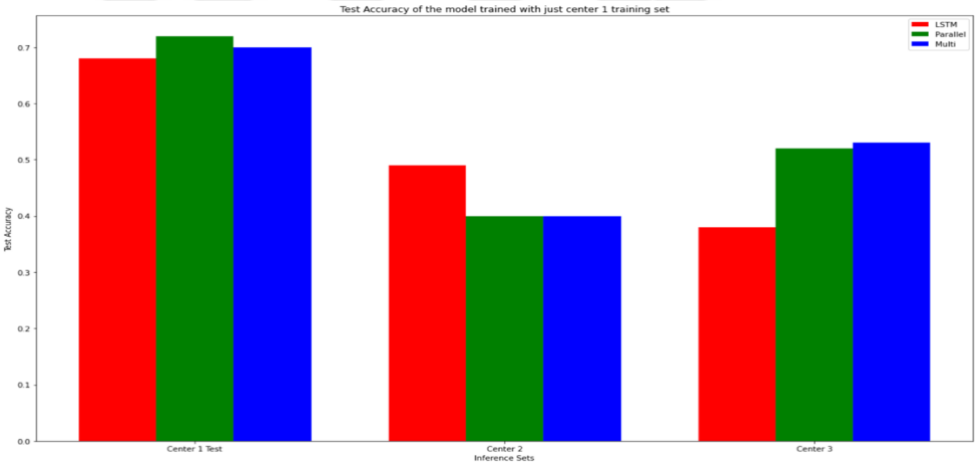


Figure 5.9 Bar plot of test accuracy of external validation experiments

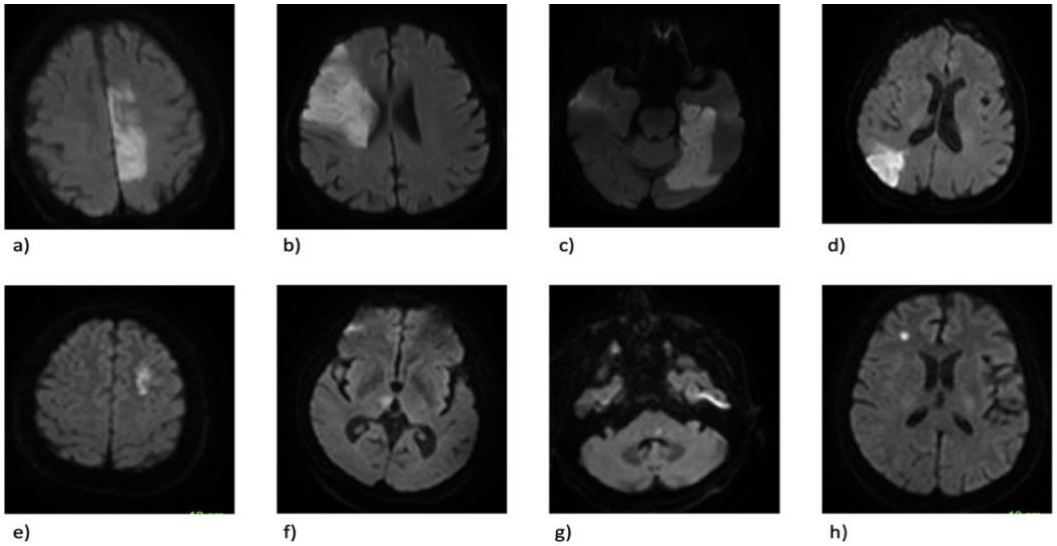


Figure 5.10 Error analysis of predictions. Upper row true positive predictions for a) ACA b) MCA c) PC and d) WS region infarcts. Lower row false positive predictions for d) ACA (GT: WS) e) MCA (GT: PC) f) Normal (GT: PC) e) ACA (GT:WS)

CHAPTER SIX

DISCUSSION AND CONCLUSION

6.1 Discussion for slice-wise approach

In this approach, the feasibility of performance improvement of stroke classification by using transformed images were evaluated. MobileNetV2 with [Native DW, Bounding box crop, Cross sign crop] input achieved 94.4 % accuracy in predicting stroke territory type with approximately half of the training data compared with the model trained with native images. Utilization of more data while training improves performance of deep learning models. Better performance with less sample size was an important aspect accentuating the benefit of constraining the search space of the model while making decisions.

Chin et al. reached over 90 % accuracy for detection of ischemic stroke with CT (Chin et al., 2017) and Marbun et al. reached 90% accuracy for multiclass stroke prediction with CNN (Marbun et al., 2018). However, the best performance with CNN on detection of stroke with CT was achieved by Pereira et al (Pereira et al., 2018) with 99% accuracy. Their dataset included 100 ischemic stroke, 100 hemorrhagic stroke and 100 normal cases. The density difference between these classes on CT is large and leads to wide interclass variance which might help the model for accurate prediction.

In a study to detect basal ganglia hemorrhagic stroke GoogleNet achieved AUC=1 and 100 % sensitivity with 80 % specifity (Desai, Flanders, & Lakhani., 2017). Arabbhsirani et al utilized 3D CNN to reprioritize head CT scans for intracranial hemorrhage detection (Arabshirani et al., 2018). Although there is no research about territorial classification of stroke, 0.95 F1 score along with AUC=0.99 of our best model is comparable even better than morphological classification of stroke which seems to be a clear indicator of its proof of concept.

MCA is the most often affected artery in acute stroke and responsible of 50 % of all ischemic strokes (Ng et al., 2007) PC and watershed infarcts are responsible of 20

% and 5-10 % respectively (Merwick & Werring, 2014; Naidich, Castillo, Cha, & Smirniotopoulos, 2012).

In randomized studies it is shown that, intravenous thrombolysis should be employed earlier than 4.5 hours from the onset of the event to be effective for anterior circulation stroke (Nogueira et al., 2018). Lee et al designed a study to predict stroke onset time less than 4.5 hours relying on DWI-FLAIR mismatch. Their logistic regression and random forest (RF) models achieved 75.8 % sensitivity (Lee et al., 2020) There is no strict rule obtained from randomized studies for exact available period to employ IV thrombolysis in posterior circulation stroke (Powers et al., 2018). Therefore, it is an important requirement to differentiate MCA infarcts which constitute majority of anterior circulation infarcts from posterior circulation infarcts. In one study designed to differentiate stroke and its mimics under 4.5 hours from event onset with CNN Abedi et al, achieved 80 % sensitivity and 86.2% specifity (Abedi et al., 2017). On the other hand, Kassasbeh et al used CT perfusion for core infarct determination to guide early management. Their model achieved AUC:0.85 and 90 % sensitivity and 62 % specifity for core infarct volume determination. In the largest cohort for stroke detection (Kasabeh et al., 2019) Oman et al tackled MCA stroke detection problem utilizing CTA source images (Oman et al., 2019) and Takahashi et al utilizing support vector machines (SVM) to identify dot sign as thromboembolic candidate (Takahashi et al., 2014). They reached 93 % sensitivity and 82 % specifity with DeepMedic model and 97.5 % sensitivity with their SVM classifier respectively. In our model we achieved 93 % sensitivity and specifity and 95 % accuracy along with 0.97 positive predictive value (PPV) to differentiate MCA stroke from other classes.

In a meta-analysis, posterior circulation infarcts exhibited more probability for recurrence in acute stage than anterior circulation stroke (Flossman & Rothwell, 2003). Our model reached 93 % sensitivity and specifity and 94 % accuracy along with 0.97 PPV for detection of PC infarct.

Our initial model without augmented input achieved 89 % sensitivity to predict watershed infarcts. This relatively lower sensitivity compared to other classes might be attributed to specific spatial location of watershed zones which are periphery of major territories. In daily practice radiologist has the advantage of evaluating

consecutive slices in order to reach a final diagnosis. Particularly small MCA infarcts near the vertex and watershed infarcts might be indistinguishable while evaluating a single slice even for an experienced radiologist. However due to design of our study, our models analyze a single slice each time and final decision of the model is made thereafter. This might explain some overlap of watershed infarcts with MCA infarct. However with the same architecture constructed as modified MobileNetV2 and modification of the input as [Native DW, Bounding box crop, Cross sign crop] increased the sensitivity to 97 % along with higher accuracy and F1 scores (0.94) and AUC value (0.98) despite utilization of less training data. This performance gain highlights the contribution of our approach.

On the other hand, our study has several limitations. Our dataset did not have sufficient labelled images for TOAST classification. Also, we could not include all major vascular territories due to available data constraints. However, we could cover %80-85 of all major arterial stroke territories under 3 classes. Further studies with larger datasets including all stroke subtypes may address this limitation. Our models analyzed each slice as a separate instance, however patient level predictions are more relevant to integrate into clinical workflows. Another limitation of our study was the limited variance of image acquisition conditions. Although two different protocols of acquisition schemes are utilized, all the images are obtained from a single MR machine.

6.2 Discussion for voxel wise approach

We developed a fully automated, highly accurate, deep learning network for determining the territory of acute stroke as anterior or posterior circulation in DWI with this approach. Our network could determine the stroke territory end to end. This eliminates potential failures from necessary user interactions in semiautomatic models.

Previous reports were based on slice wise prediction which might introduce some problems into the system. For the slice wise approach, the pathologic slices should be selected by the user which makes the system impractical for triage purpose. A fully automatic system obviates this requirement which makes the system more suitable into clinical translation. Additionally in slice wise systems the model can predict different

classes for individual slices of the same patient. This can introduce additional uncertainties into the final predictions derived from different set of selected slices by different users. A fully automated end to end system can prevent this uncertainty and make its predictions more robust. To the best of our knowledge there are only 2 studies on prediction of stroke territory class in DWI including our previous study (Lee et al., 2023; Çetinoğlu et al., 2021). Both studies approached the problem slice-wise and applied transfer learning by using off the shelf pretrained networks. Actually, in the previous task we tried to enhance this strategy by applying image transformations and accentuating the features for differential diagnosis with a clinician perspective. Transfer learning with slice-based data expansion strategy is a necessity to leverage the already learnt image prior features embedded in the weights of these networks with small medical imaging datasets. The authors reached 93% overall accuracy for prediction of middle cerebral artery, posterior circulation, or watershed infarcts in one of these and 86% for prediction of anterior circulation, posterior circulation or normal slices in the other one (Lee et al., 2023; Çetinoğlu et al., 2021). However, in the study of Lee et al, since the majority of slices were usually derived from normal class even in already diseased patients, it is unclear how would they aggregate the prediction results for final patient level classification for clinically meaningful applications (Lee et al., 2023). Actually our 94.3% accuracy with our previous approach was a evidence of the benefit gathered by using image processing methods to distill the more relevant features.

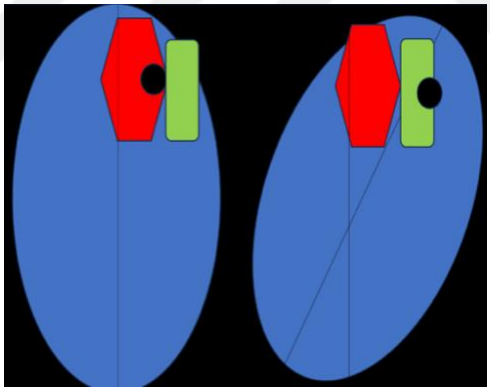
In this study we used a voxel wise prediction approach to derive final patient level prediction with the accuracy of 0.83 with our best model. In medical imaging tasks, different levels of predictions might be possible as voxel wise, ROI wise, slice wise, modality wise and patient wise. Voxel wise prediction is the densest approach which decreases along the increasing patch size that is used for the prediction. To the best of our knowledge the only example of patient level prediction based on voxel wise predictions was used by Yogananda et al. for the prediction of MGMT methylation status of the glial tumors (Yogananda et al., 2021). The authors reached 0.93 mean ROC_AUC which reflected the success of this approach over slice wise and modality wise approaches by more than 0.10 % margin. Actually, using the full volume of the images of the patient to reach a final class is not straightforward for an image element

which has no fixed shape, no fixed boundary properties, no fixed color or no distinct heterogeneity properties driven by underlying complex biology. The only determining feature for an acute stroke lesion in DWI is the location of the high signal intensity voxels conforming to a priori vascular feeding zone. Therefore, additional guidance driven by neighborhood voxels and collective behavior of voxels located in a similar region helped the voxel wise approach to perform better.

In convolutional neural networks, the training actually means learning the correct weights of the convolution filters in a broader sense. Although in lifelong learning approaches there is some dynamism of the values of these weights, in usual applications these convolutional filters are fixed structures, which implies that they affect every part of the image in a similar way. In a dataset like ours when the target could be anywhere, but the spatial location dictates the actual class label, a great challenge is introduced to the neural network. One of the main strengths of neural networks is the translation invariance of the convolution filters, that is the target object can be found in anywhere in the image. However, if the spatial location shift changes the class label, this property could be detrimental for the performance of the model. This was evident for the lower performance of our model before N4 Bias correction. The local magnetic field inhomogeneities and the magnetic properties of the scanned tissue might introduce some noise which is also affected by the distance to the isocenter of the receiving elements of the coil. This is known as bias field and usually evident as brighter regions in the scanned tissue which is not related to actual bright tissue instead is the consequence of the bias field. There are numerous studies which did not care on the effects of bias field, since the neural networks can learn to eliminate it, while focusing on the target. However, in a task like ours, when the target is just high signal in some location and the location dictates the class label, this effect could be detrimental as shown in Figure 5.5. The model learned to predict both anterior and posterior circulation infarcts and had filters which would be activated when it has encountered bright signal at the corresponding locations. Therefore, although the model predicted correct regions as the anterior circulation infarct, the bright signal derived from bias field in the cerebellum lead to activation of posterior circulation infarct detector filters and a large false positive area for posterior circulation infarct was predicted which was eliminated after bias field correction (Figure 5.5). This point

is very import according to us, because it emphasizes strongly that knowing the target domain and the target data while developing these systems has utmost importance.

The medical image datasets are usually small which necessitates extensive data augmentation to alleviate this scarcity (Goceri, 2023). Actually, all deep learning studies benefit from data augmentation, but this requirement is stronger in medical imaging datasets due to this small dataset problem. The usual minimal set of data augmentation operations are rotation, translation, scaling and left right flip. Elastic deformations are also very helpful however it is not used frequently in medical imaging manuscripts. However similar to the above discussion, if the target class label strongly depends on spatial location, the rotation operation which will change the position of the bright signal in the cartesian coordinate system can change the predicted class label, since the filters activated in that location is preconditioned on the actual class that should reside there (Figure 6.1).



a) b)

Figure 6.1 Effects of rotation. Red and green polygons represent different feeding zones, and the black circle represents the lesion. The lesion displayed in the red territory initially (a), is shifted into the green region territory after rotation (b)

Therefore, minimal rotations in the range of a few degrees could be tolerated however more extensive rotations should be avoided. In a similar manner, the rotated samples in our dataset should be registered to a standard space in order to have better results which was evident by the performance gain of our models after registration.

Other groups have also proposed deep learning methods for noninvasive, image-based stroke lesion segmentation and its extensions in DWI and other modalities, but each of these has several limitations. For ischemic stroke segmentation, Praveen et al. built a stacked sparse autoencoder which performed perfectly on the Ischemic Stroke Lesion Segmentation (ISLES) 2015 dataset, with an average Dice score of 0.94. (Praveen et al., 2018). Pinto et al. built restricted Boltzmann machine to extract features from lesions and perfusion from different sequences to predict the final stroke lesion. They achieved 0.38 Dice score on ISLES 2017 (Pinto et al., 2020). DWI-ASPECTS can be thought as the MRI equivalent of CT-ASPECTS. Do et al. built recurrent residual convolutional neural network for predicting ASPECTS score in acute stroke patients in DWI with an AUC of 0.94 (Do et al., 2020). These studies are the evidence of high performance of deep learning models for segmentation of stroke lesions in DWI which can be translated to class prediction by aggregation into multiclass class labels. Our study is a proof of concept of this idea which translated the multiclass stroke territory segmentation outputs to the classification outputs by employing majority voting.

There were some limitations of this approach as well. One of the limitations was limited type of the data. Although we used the images from 2 different MR machines, the images were from single center. Another limitation was lack of external validation set. The third limitation was the constrained number of classes.

6.3 Discussion for patient wise 3D end to end approach.

Fast treatment of stroke is essential for better clinical outcome. In the absence of domain experts, deep learning-based tools may facilitate fast triage of stroke patients to a dedicated stroke center with confirmation of stroke and determination of the affected territory. In the previous approaches we explored with slice wise approach and voxel wise approach which were particularly important methods under limited data settings. In this study we built an end-to-end high-performance classifier which predicts the affected territory from stroke in DWI. We also explored the effects of different encoding strategies, different number of classes and integration of clinical information. Our best model with deep networks and image-based data reached 0.88 accuracy and 0.80 mean f1 score. The best performing deep model for our study was

parallel stream model for all stroke territories and normal images, while multi attention model performed better for prediction of anterior circulation, posterior circulation, or watershed region strokes.

Two groups attempted to classify the affected territory from stroke in DWI (Lee et al., 2023; Çetinoğlu et al., 2021). Both used slice wise approach and reached 0.86 and 0.93 accuracy respectively. Slice wise approach has the advantage of increasing the number of samples since each slice becomes a distinct training instance, and leveraging 2D encoding which is known to need less samples to learn compared to 3D encoding. However, 2D approach has some drawbacks. First, the affected slices should be selected initially which requires human interaction which is an important obstacle for end-to-end design. Second, the context could not be embedded to the model since each slice becomes an independent instance. However medical images are representation of human body which is inherently 3D in nature. 3D information can be encoded either directly or slice by slice. 3D CNN for direct encoding and LSTM for slice-by-slice encoding are plausible choices for these tasks which were demonstrated in the literature for various organ system and disease classifications (Haq, 2022; Montaha et al., 2022). However, encoding of a small lesion in a large 3D anatomy may suffer from the great noise around the lesion. The hierarchical coarse to fine approach can be a solution to mitigate this obstacle when the anatomy permits. In (Jamaludin, Kadir, & Zisserman, 2017) the authors first isolated the disk from the image then predicted the pathology class. In stroke classification problem this approach encounters difficulties since the lesion can reside anywhere in the brain. Segmentation or detection before classification approach cannot be utilized for stroke classification problem since the anatomy around the lesion determines the class. One possible solution is application of attention mechanisms to the feature maps. Attention mechanisms were successfully used in medical image classification tasks (Liu et al., 2020; Li et al., 2022) and they are actually the backbone of successful transformer architectures. Application of attention mechanism increased the performance of all our 3D encoding models. Dramatic increase of accuracy from 0.81 to 0.96 with two class models and from 0.56 to 0.77 with five class models after integrating attention layers to the network was of particular interest. LSTM models which encode each slice separately and add the previous embeddings sequentially to the subsequent embedding

coupled with some attention mechanisms is a nice solution to focus on the target which was applied to various tasks including brain tumor classification (Montaha et al., 2022). We leveraged this property of LSTM and reached 0.84 accuracy with five class model and 0.83 accuracy with two class model. LSTM encodes each slice sequentially; however medical image data can benefit from parallel encoding which can be built by Time distributed layer (TDL) as well. Time distributed layer is employed for time distributed tasks such as video analysis or gesture prediction from frames (Montaha et al., 2022). To the best of our knowledge TDL was not employed for stroke related classification tasks so far. Encoding each slice of the image volume in parallel and application of CBAM module to the slice embeddings to reweigh the relevant regions could be a plausible alternative to encode the 3D data with keeping the context while reducing the parameters. Indeed, this approach worked very well and parallel stream model was the best performing model for five class stroke classification task with 0.88 accuracy, 0.80 f1 score and 0.98 AUC. We expected the performance increase while we were experimenting with three class and two class models since the number of decision planes would be less. The accuracy of our best model with three class model was 0.83 which was below our expectations. We believe that one reason for this was the decreased number of available training data from 439 to 304 which could potentially affect the training process worse. Some authors demonstrated that when the model is trained with multiple tasks the performance was increased (Hussein et al., 2022) due to multiple supervising signals. Therefore, one other reason for this performance decrease might be the decreasing number of target classes leading to fewer supervising signals which might hamper the performance. Although the number of training data decreased from 304 to 249 with two class models, the accuracy increased to 0.96. We believe that this increase was partly due to the decreasing number of decision planes and partly due to the conversion of the problem to a less conflicting version. Indeed, the watershed class was the most uncertain class and was responsible for majority of the false predictions of the models demonstrated by our error analysis. Excluding it from the classification task possibly provided simpler decision planes with an easier classification task.

Pocket algorithm is a small modification of MLP. Actually, the fully connected layers of a CNN is a type of MLP. In our models this was a two hidden layer

architecture with 512 and 256 nodes. Pocket algorithm adds a new functionality to MLP which evaluates misclassification rate after each iteration and keeps only the weights only if the misclassification rate is lower than previous iteration. We designed pocket algorithm with exactly the same architecture of MLP as the fully connected layers of our deep models for fair comparison. To the best of our knowledge there is no study with application of pocket algorithm to medical imaging data. Interestingly embeddings from 3D encoding models performed better than embeddings from slice-based encoding models with pocket algorithm. The size of the vector produced by 3D encoding models was 256 while that was 128 with slice-based models. Since deep learning models jointly optimize the convolutional part and the fully connected part, they could specifically select the optimal features conditioned on the fully connected layers. Therefore, we believe that slice-based models had better performance due to better encoding of the images. Indeed t-SNE and PCA display of the embeddings of parallel stream model clearly demonstrated three relatively easier separable clusters (Van der Maaten & Hinton, 2008; Pearson, 1901). On the other hand, pocket algorithm was initiated with these embeddings and trained to find their own optimum weights. Therefore, with a larger feature pool as the starting point which was the case with the embeddings of our 3D encoding networks might provide an advantage of exploring a greater feature space compared to starting with the embeddings of slice-based encoding networks.

Some groups tried to integrate clinical knowledge to the image embeddings in order to improve the performance of their models (Xie et al., 2020). Although this may complicate the pipeline since human interaction is required to explore and integrate clinical variables, we also believe that clinical knowledge integration should be an essential part of AI classification models based on medical imaging data. The radiological decision making cannot be reduced to just pattern matching driven by underlying images. For sure the image set has great importance for concluding a final differential diagnosis, however the interpretation process is largely dependent on clinical reasoning based on all the clinical information including demographics, symptoms, signs, laboratory exploration and many times previous or follow up evaluation. Therefore, we tried to integrate the clinical information to our models in order to explore their effects. We also built a model based on just clinical information

in order to assess the rational of solely using clinical data. The performance of clinical information-based model was poor with 0.70 test accuracy. This poor performance could be expected since the clinical signs and symptoms evaluation-based classification (Bamford classification) was demonstrated to poorly correlate with image-based classification (Tao et al., 2012). However, integration of clinical information to the image embeddings substantially improved the performance of the models. The MLP model with LSTM based image embeddings and clinical information reached 0.90 accuracy and the same input with SVM model reached 0.93 accuracy. On the other hand, the set up for testing external validation data yielded poor performance with 0.68 best accuracy. However, we believe that this comparison is not fair since this model is trained with less than one third of the original data. The higher performing deep networks were trained with 439 samples distributed over three centers, three MRI machines and four protocols. Isolating the data from one center for training greatly reduced the number of samples to 164 and the variety of the samples to a single machine.

There were some limitations of this approach also. One of the main limitations was the small number of patients from each center. Although total 624 patients from 3 centers were a considerable number, it would still benefit from larger sample size. Second limitation was limited number of samples for anterior cerebral artery infarcts. Third limitation was the limited number of samples which had clinical information. Actually, we have clinical information for the patients for just center 1. Fourth limitation was small lesions which lead to label uncertainty particularly in posterior fronto-temporo-parietal lesions. The low interobserver correlation with these lesions should alert us for the model predictions biased towards the subjective prediction of the observers. Fifth limitation was the absence of objective measures to define the exact territory class which requires special techniques such as selective angiography or spin tagged MRI perfusion. However, this type of data is extremely difficult to collect since in the routine clinical practice they are rarely employed. We believe the high performance of our model against human observers is a plausible approach for a model which aims fast triage in the absence of field experts.

6.4 Conclusion

In conclusion we built first end to end model which predicts the affected territory from stroke in DWI with high performance The technical importance of this thesis is that it represents the first application of time distributed layers coupled with attention mechanisms in order to encode the volumetric data slice wise in parallel and first application of pocket algorithm to image embeddings from various deep networks in stroke research and in general medical image analysis. The clinical importance of this thesis is the potential of the built model and pipeline for fast triage of stroke patients stratified as anterior cerebral artery, middle cerebral artery, posterior circulation or watershed infarcts based on DWI in the absence of domain experts.

REFERENCES

Abedi, V., Goyal, N., Tsivgoulis, G., Hosseinichimeh, N., Hontecillas, R., Bassaganya-Riera, J., Elijovich, L., Metter, J. E., Alexandrov, A. W., Liebeskind, D. S., Alexandrov, A. V., & Zand, R. (2017). Novel Screening Tool for Stroke Using Artificial Neural Network. *Stroke*, *48*(6), 1678–1681. https://doi.org/10.1161/STROKEAHA.117.017033

Aho, K., Harmsen, P., Hatano, S., Marquardsen, J., Smirnov, V. E., & Strasser, T. (1980). Cerebrovascular disease in the community: results of a WHO collaborative study. *Bulletin of the World Health Organization*, *58*(1), 113–130.

Amukotuwa, S. A., Straka, M., Smith, H., Chandra, R. V., Dehkharghani, S., Fischbein, N. J., & Bammer, R. (2019). Automated Detection of Intracranial Large Vessel Occlusions on Computed Tomography Angiography: A Single Center Experience. *Stroke*, *50*(10), 2790–2798. https://doi.org/10.1161/STROKEAHA.119.026259

Aviv, R. I., Mandelcorn, J., Chakraborty, S., Gladstone, D., Malham, S., Tomlinson, G., Fox, A. J., & Symons, S. (2007). Alberta Stroke Program Early CT Scoring of CT perfusion in early stroke visualization and assessment. *AJNR. American journal of neuroradiology*, *28*(10), 1975–1980. https://doi.org/10.3174/ajnr.A0689

Arbabshirani, M.R., Fornwalt, B.K., Mongelluzzo, G.J., Suever, J.D., Geise, B.D., Patel, A.A., & Moore, G.J. (2018). Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digital Medicine*, *1*.

Aydin, A., Yürük, S. E., Reisoğlu, İ., & Goktas, Y. (2023). Main barriers and possible enablers of academicians while publishing. *Scientometrics*, *128*(1), 623–650. https://doi.org/10.1007/s11192-022-04528-x

Barber, P. A., Demchuk, A. M., Zhang, J., & Buchan, A. M. (2000). Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta

Stroke Programme Early CT Score. *Lancet (London, England)*, *355*(9216), 1670–1674. https://doi.org/10.1016/s0140-6736(00)02237-6

Ba, J., Mnih, V., & Kavukcuoglu, K. (2014). Multiple Object Recognition with Visual Attention. *CoRR, abs/1412.7755*.

Barnwell, S. L., Clark, W. M., Nguyen, T. T., O'Neill, O. R., Wynn, M. L., & Coull, B. M. (1994). Safety and efficacy of delayed intraarterial urokinase therapy with mechanical clot disruption for thromboembolic stroke. *AJNR. American journal of neuroradiology*, *15*(10), 1817–1822.

Baron, C. A., Kate, M., Gioia, L., Butcher, K., Emery, D., Budde, M., & Beaulieu, C. (2015). Reduction of Diffusion-Weighted Imaging Contrast of Acute Ischemic Stroke at Short Diffusion Times. *Stroke*, *46*(8), 2136–2141. https://doi.org/10.1161/STROKEAHA.115.008815

Becker, H., Desch, H., Hacker, H., & Pencz, A. (1979). CT fogging effect with ischemic cerebral infarcts. *Neuroradiology*, *18*(4), 185–192. https://doi.org/10.1007/BF00345723

Bento, M., Souza, R., Salluzzi, M., Rittner, L., Zhang, Y., & Frayne, R. (2019). Automatic identification of atherosclerosis subjects in a heterogeneous MR brain imaging data set. *Magnetic resonance imaging*, *62*, 18–27. https://doi.org/10.1016/j.mri.2019.06.007

Bishop, C.M. (1994). Neural networks and their applications. *Review of Scientific Instruments, 65*, 1803-1832. https://doi.org/10.1063/1.1144830

Bouthillier, A., van Loveren, H. R., & Keller, J. T. (1996). Segments of the internal carotid artery: a new classification. *Neurosurgery*, *38*(3), 425–433. https://doi.org/10.1097/00006123-199603000-00001

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324

Bruno, A., Close, B., Gomadam, A., Akinwuntan, A. E., & Switzer, J. A. (2013). The simplified mRS questionnaire reflects stroke severity. *International journal of stroke : official journal of the International Stroke Society*, *8*(8), E55. https://doi.org/10.1111/ijs.12113

Campbell, B. C., Christensen, S., Levi, C. R., Desmond, P. M., Donnan, G. A., Davis, S. M., & Parsons, M. W. (2012). Comparison of computed tomography perfusion and magnetic resonance imaging perfusion-diffusion mismatch in ischemic stroke. *Stroke*, *43*(10), 2648–2653. https://doi.org/10.1161/STROKEAHA.112.660548

Candemir, S., Nguyen, X. V., Folio, L. R., & Prevedello, L. M. (2021). Training Strategies for Radiology Deep Learning Models in Data-limited Scenarios. *Radiology. Artificial intelligence*, *3*(6), e210014. https://doi.org/10.1148/ryai.2021210014

Cetinoglu, Y. K., Koska, I. O., Uluc, M. E., & Gelal, M. F. (2021). Detection and vascular territorial classification of stroke on diffusion-weighted MRI by deep learning. *European journal of radiology*, *145*, 110050. https://doi.org/10.1016/j.ejrad.2021.110050

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Comput. Electr. Eng.*, *40*, 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Chatterjee, A., Somayaji, N.R., & Kabakis I.M. (2019). Abstract WMP16: artificial intelligence detection of cerebrovascular large vessel occlusion-nine month, 650 patient evaluation of the diagnostic accuracy and performance of the Viz.ai LVO algorithm. *Stroke*, 50, Suppl_1: AWMP16-AWMP16.

Chen, L., Bentley, P., & Rueckert, D. (2017). Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage. Clinical*, *15*, 633–643. https://doi.org/10.1016/j.nicl.2017.06.016

Chen, Y., Dhar, R., Heitsch, L., Ford, A., Fernandez-Cadenas, I., Carrera, C., Montaner, J., Lin, W., Shen, D., An, H., & Lee, J. M. (2016). Automated

quantification of cerebral edema following hemispheric infarction: Application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *NeuroImage. Clinical*, *12*, 673–680. https://doi.org/10.1016/j.nicl.2016.09.018

Cheng, P. M., Montagnon, E., Yamashita, R., Pan, I., Cadrin-Chênevert, A., Perdigón Romero, F., Chartrand, G., Kadoury, S., & Tang, A. (2021). Deep Learning: An Update for Radiologists. *Radiographics : a review publication of the Radiological Society of North America, Inc*, *41*(5), 1427–1445. https://doi.org/10.1148/rg.2021200210

Cheng, X., Su, X., Shi, J., Liu, Q., Zhou, C., Dong, Z., Xing, W., Lu, H., Pan, C., Li, X., Yu, Y., Zhang, L., & Lu, G. (2021). Comparison of automated and manual DWI-ASPECTS in acute ischemic stroke: total and region-specific assessment. *European radiology*, *31*(6), 4130–4137. https://doi.org/10.1007/s00330-020-07493-2

Chin, C., Lin, B., Wu, G., Weng, T., Yang, C., Su, R., & Pan, Y. (2017). An automated early ischemic stroke detection system using CNN deep learning algorithm. *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, 368-372.

Choi, Y. A., Park, S. J., Jun, J. A., Pyo, C. S., Cho, K. H., Lee, H. S., & Yu, J. H. (2021). Deep Learning-Based Stroke Disease Prediction System Using Real-Time Bio Signals. *Sensors (Basel, Switzerland)*, *21*(13), 4269. https://doi.org/10.3390/s21134269

Chollet, F. (2015). Keras. Software available from https://www.keras.io

Cui, L., Han, S., Qi, S., Duan, Y., Kang, Y., & Luo, Y. (2021). Deep symmetric three-dimensional convolutional neural networks for identifying acute ischemic stroke via diffusion-weighted images. *Journal of X-ray science and technology*, *29*(4), 551–566. https://doi.org/10.3233/XST-210861

Daroff, R.B., Fenichel G.M., Jankovic J., & Mazziotta JC. (2012). *Neurology in Clinical Practice.* Elsevier Health Sciences.

Dawud, A.M., Yurtkan, K., & Oztoprak, H. (2019). Application of Deep Learning in Neuroradiology: Brain Haemorrhage Classification Using Transfer Learning. *Computational Intelligence and Neuroscience, 2019*.

De Haan, R., Horn, J., Limburg, M., Van Der Meulen, J., & Bossuyt, P. (1993). A comparison of five stroke scales with measures of disability, handicap, and quality of life. *Stroke*, *24*(8), 1178–1181. https://doi.org/10.1161/01.str.24.8.1178

de Lucas, E. M., Sánchez, E., Gutiérrez, A., Mandly, A. G., Ruiz, E., Flórez, A. F., Izquierdo, J., Arnáiz, J., Piedra, T., Valle, N., Bañales, I., & Quintana, F. (2008). CT protocol for acute stroke: tips and tricks for general radiologists. *Radiographics : a review publication of the Radiological Society of North America, Inc*, *28*(6), 1673–1687. https://doi.org/10.1148/rg.286085502

Demaerschalk, B. M., Kleindorfer, D. O., Adeoye, O. M., Demchuk, A. M., Fugate, J. E., Grotta, J. C., Khalessi, A. A., Levy, E. I., Palesch, Y. Y., Prabhakaran, S., Saposnik, G., Saver, J. L., Smith, E. E., & American Heart Association Stroke Council and Council on Epidemiology and Prevention (2016). Scientific Rationale for the Inclusion and Exclusion Criteria for Intravenous Alteplase in Acute Ischemic Stroke: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*, *47*(2), 581–641. https://doi.org/10.1161/STR.0000000000000086

Desai, V., Flanders, A.E., & Lakhani, P. (2017). Application of Deep Learning in Neuroradiology: Automated Detection of Basal Ganglia Hemorrhage using 2D-Convolutional Neural Networks**.** *ArXiv, abs/1710.03823*.

Diederik P. Kingma, & Ba J. L. (2014). Adam: A method for stochastic optimization. *arXiv:1412.6980v9*

Ding, L., Liu, Z., Mane, R., Wang, S., Jing, J., Fu, H., Wu, Z., Li, H., Jiang, Y., Meng, X., Zhao, X., Liu, T., Wang, Y., & Li, Z. (2022). Predicting functional outcome in patients with acute brainstem infarction using deep neuroimaging

features. *European journal of neurology*, *29*(3), 744–752. https://doi.org/10.1111/ene.15181

Do, L. N., Baek, B. H., Kim, S. K., Yang, H. J., Park, I., & Yoon, W. (2020). Automatic Assessment of ASPECTS Using Diffusion-Weighted Imaging in Acute Ischemic Stroke Using Recurrent Residual Convolutional Neural Network. *Diagnostics (Basel, Switzerland)*, *10*(10), 803. https://doi.org/10.3390/diagnostics10100803

Donnan, G. A., Fisher, M., Macleod, M., & Davis, S. M. (2008). Stroke. *Lancet (London, England)*, *371*(9624), 1612–1623. https://doi.org/10.1016/S0140-6736(08)60694-7

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *2016 International Conference on Engineering & MIS (ICEMIS)*, 1-5.

Fahed, R., Lecler, A., Sabben, C., Khoury, N., Ducroux, C., Chalumeau, V., Botta, D., Kalsoum, E., Boisseau, W., Duron, L., Cabral, D., Koskas, P., Benaïssa, A., Koulakian, H., Obadia, M., Maïer, B., Weisenburger-Lile, D., Lapergue, B., Wang, A., Redjem, H., … Piotin, M. (2018). DWI-ASPECTS (Diffusion-Weighted Imaging-Alberta Stroke Program Early Computed Tomography Scores) and DWI-FLAIR (Diffusion-Weighted Imaging-Fluid Attenuated Inversion Recovery) Mismatch in Thrombectomy Candidates: An Intrarater and Interrater Agreement Study. *Stroke*, *49*(1), 223–227. https://doi.org/10.1161/STROKEAHA.117.019508

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, *30*(9), 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001

Flanders, A. E., Prevedello, L. M., Shih, G., Halabi, S. S., Kalpathy-Cramer, J., Ball, R., Mongan, J. T., Stein, A., Kitamura, F. C., Lungren, M. P., Choudhary, G., Cala, L., Coelho, L., Mogensen, M., Morón, F., Miller, E., Ikuta, I., Zohrabian, V., McDonnell, O., Lincoln, C., … RSNA-ASNR 2019 Brain Hemorrhage CT Annotators (2020). Construction of a Machine Learning Dataset through

Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge. *Radiology. Artificial intelligence*, *2*(3), e190211. https://doi.org/10.1148/ryai.2020190211

Flossmann, E., & Rothwell, P. M. (2003). Prognosis of vertebrobasilar transient ischaemic attack and minor stroke. *Brain : a journal of neurology*, *126*(Pt 9), 1940–1954. https://doi.org/10.1093/brain/awg197

Gallant, S.I. (1990). Perceptron-based learning algorithms. *IEEE transactions on neural networks, 1 2*, 179-91.

Gaillard, F. (Accessed on 16 Jan 2024). *Cerebral vascular territories (illustration). Case study*. Radiopaedia.org https://doi.org/10.53347/rID-10814

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., Nimmo-Smith, I., & Dipy Contributors (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*, *8*, 8. https://doi.org/10.3389/fninf.2014.00008

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *International Conference on Artificial Intelligence and Statistics*.

Goceri E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. *Artificial intelligence review*, 1–45. Advance online publication. https://doi.org/10.1007/s10462-023-10453-z

Goebel, J., Stenzel, E., Guberina, N., Wanke, I., Koehrmann, M., Kleinschnitz, C., Umutlu, L., Forsting, M., Moenninghoff, C., & Radbruch, A. (2018). Automated ASPECT rating: comparison between the Frontier ASPECT Score software and the Brainomix software. *Neuroradiology*, *60*(12), 1267–1272. https://doi.org/10.1007/s00234-018-2098-x

Goldstein, L. B., Jones, M. R., Matchar, D. B., Edwards, L. J., Hoff, J., Chilukuri, V., Armstrong, S. B., & Horner, R. D. (2001). Improving the reliability of stroke

subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke*, *32*(5), 1091–1098. https://doi.org/10.1161/01.str.32.5.1091

Goodfellow I., Bengio Y., & Courville A. (2016). *Deep Learning*. The MIT Press.

Guberina, N., Dietrich, U., Radbruch, A., Goebel, J., Deuschl, C., Ringelstein, A., Köhrmann, M., Kleinschnitz, C., Forsting, M., & Mönninghoff, C. (2018). Detection of early infarction signs with machine learning-based diagnosis by means of the Alberta Stroke Program Early CT score (ASPECTS) in the clinical routine. *Neuroradiology*, *60*(9), 889–901. https://doi.org/10.1007/s00234-018-2066-5

Haq, A. U., Li, J. P., Agbley, B. L. Y., Khan, A., Khan, I., Uddin, M. I., & Khan, S. (2022). IIMFCBM: Intelligent Integrated Model for Feature Extraction and Classification of Brain Tumors Using MRI Clinical Imaging Data in IoT-Healthcare. *IEEE journal of biomedical and health informatics*, *26*(10), 5004–5012. https://doi.org/10.1109/JBHI.2022.3171663

Haralick, R.M., Shanmugam, K.S., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern., 3*, 610-621.

Harnsberger H.R., & Macdonald A.J. (2006). *Diagnostic and Surgical Imaging Anatomy: Brain, head & neck, spine*. Amirsys.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

Hearst, M.A., Dumais ,S. T., Osuna, E., Platt, J., & Scholkopf B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, 13, 8–28. https://doi.org/10.1109/5254.708428

Heo, J., Yoon, J. G., Park, H., Kim, Y. D., Nam, H. S., & Heo, J. H. (2019). Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke*, *50*(5), 1263–1265. https://doi.org/10.1161/STROKEAHA.118.024293

Ho, K.H., Speier, W., El-Saden, S., & Arnold, C.W. (2017). *Classifying acute ischemic stroke onset time using deep imaging features* [Conference presentation abstract]. AMIA Annual Symposium Proceedings, p. 892, Washington, DC. Frontal.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

Hopyan, J., Ciarallo, A., Dowlatshahi, D., Howard, P., John, V., Yeung, R., Zhang, L., Kim, J., MacFarlane, G., Lee, T. Y., & Aviv, R. I. (2010). Certainty of stroke diagnosis: incremental benefit with CT perfusion over noncontrast CT and CT angiography. *Radiology*, *255*(1), 142–153. https://doi.org/10.1148/radiol.09091021

Hussein, R., Zhao, M.Y., Shin, D., Guo, J., Chen, K.T., Armindo, R.D., Davidzon, G.A., Moseley, M.E., & Zaharchuk, G. (2022). Multi-task Deep Learning for Cerebrovascular Disease Classification and MRI-to-PET Translation. *2022 26th International Conference on Pattern Recognition (ICPR)*, 4306-4312.

Jamaludin, A., Kadir, T., & Zisserman, A. (2017). SpineNet: Automated classification and evidence visualization in spinal MRIs. *Medical image analysis*, *41*, 63–73. https://doi.org/10.1016/j.media.2017.07.002

Kasasbeh, A. S., Christensen, S., Parsons, M. W., Campbell, B., Albers, G. W., & Lansberg, M. G. (2019). Artificial Neural Network Computer Tomography Perfusion Prediction of Ischemic Core. *Stroke*, *50*(6), 1578–1581. https://doi.org/10.1161/STROKEAHA.118.022649

Kleinbaum, D.G., Dietz, K., Gail, M., & Klein, M. (2002). *Logistic Regression.* Springer-Verlag, New York.

Koç, U., Akçapınar Sezer, E., Özkaya, Y. A., Yarbay, Y., Taydaş, O., Ayyıldız, V. A., Kızıloğlu, H. A., Kesimal, U., Çankaya, İ., Beşler, M. S., Karakaş, E., Karademir, F., Sebik, N. B., Bahadır, M., Sezer, Ö., Yeşilyurt, B., Varlı, S., Akdoğan, E., Ülgü, M. M., & Birinci, Ş. (2022). Artificial Intelligence in Healthcare Competition

(TEKNOFEST-2021): Stroke Data Set. *The Eurasian journal of medicine*, *54*(3), 248–258. https://doi.org/10.5152/eurasianjmed.2022.22096

Koopman, M. S., Berkhemer, O. A., Geuskens, R. R. E. G., Emmer, B. J., van Walderveen, M. A. A., Jenniskens, S. F. M., van Zwam, W. H., van Oostenbrugge, R. J., van der Lugt, A., Dippel, D. W. J., Beenen, L. F., Roos, Y. B. W. E. M., Marquering, H. A., Majoie, C. B. L. M., & MR CLEAN Trial Investigators (2019). Comparison of three commonly used CT perfusion software packages in patients with acute ischemic stroke. *Journal of neurointerventional surgery*, *11*(12), 1249–1256. https://doi.org/10.1136/neurintsurg-2019-014822

Kumar, V., & Minz, S. (2014). Feature Selection: A literature Review. *Smart Comput. Rev.*, *4*, 211-229.

Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., Sanduleanu, S., Larue, R. T. H. M., Even, A. J. G., Jochems, A., van Wijk, Y., Woodruff, H., van Soest, J., Lustberg, T., Roelofs, E., van Elmpt, W., Dekker, A., Mottaghy, F. M., Wildberger, J. E., & Walsh, S. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews. Clinical oncology*, *14*(12), 749–762. https://doi.org/10.1038/nrclinonc.2017.141

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, H., Lee, E. J., Ham, S., Lee, H. B., Lee, J. S., Kwon, S. U., Kim, J. S., Kim, N., & Kang, D. W. (2020). Machine Learning Approach to Identify Stroke Within 4.5 Hours. *Stroke*, *51*(3), 860–866. https://doi.org/10.1161/STROKEAHA.119.027611

Lee, E. J., Kim, Y. H., Kim, N., & Kang, D. W. (2017). Deep into the Brain: Artificial Intelligence in Stroke Imaging. *Journal of stroke*, *19*(3), 277–285. https://doi.org/10.5853/jos.2017.02054

Lee, K. Y., Liu, C. C., Chen, D. Y., Weng, C. L., Chiu, H. W., & Chiang, C. H. (2023). Automatic detection and vascular territory classification of hyperacute staged ischemic stroke on diffusion weighted image using convolutional neural

networks. *Scientific reports*, *13*(1), 404. https://doi.org/10.1038/s41598-023-27621-4

Li, W. (2014). Gaussian Process Learning: A Divide and Conquer Approach. In: Zeng, Z., Li, Y., King, I. (eds) *Advances in Neural Networks* – ISNN 2014. ISNN 2014. Lecture Notes in Computer Science vol 8866. Springer. https://doi: 10.1007/978-3-319-12436-0_29

Li, W., Hong, W., Wang, E., & Jiang, Y. (2021). RAPID Software to the Clinical Application Value of Acute Basilar Artery Occlusion with Endovascular Treatment. *Journal of stroke and cerebrovascular diseases: the official journal of National Stroke Association*, *30*(12), 106147. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.106147

Li, Z., Xing, Q., Li, Y., He, W., Miao, Y., Ji, B., Shi, W., & Jiang, Z. (2023). A Novel Multi-Scale Channel Attention-Guided Neural Network for Brain Stroke Lesion Segmentation**.** *IEEE Access*, 11**,** 66050-66062. https://doi.org/10.1109/ACCESS.2023.3289909

Liew, S. L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., Lefebvre, S., Nakamura, W., Saldana, D., Schmiesing, A., Tran, C., Vo, D., Ard, T., Heydari, P., Kim, B., Aziz-Zadeh, L., … Stroud, A. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific data*, *5*, 180011. https://doi.org/10.1038/sdata.2018.11

Lisowska, A., Beveridge, E., Muir, K., & Poole, I. (2017). *Thrombus detection in CT brain scans using a convolutional neural network* [Conference presentation abstract]. International Conference on Bioimaging, vol. 3, pp. 24–33, Aichi, Japan. https://doi.org/ 10.5220/0006114600240033

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep

learning in medical image analysis. *Medical image analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, L., Kurgan, L., Wu, F. X., & Wang, J. (2020). Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. *Medical image analysis*, *65*, 101791. https://doi.org/10.1016/j.media.2020.101791

Malhotra, K., Gornbein, J., & Saver, J. L. (2017). Ischemic Strokes Due to Large-Vessel Occlusions Contribute Disproportionately to Stroke-Related Dependence and Death: A Review. *Frontiers in neurology*, *8*, 651. https://doi.org/10.3389/fneur.2017.00651

Marbun, J.T., Seniman, Andayani U. (2018). Classification of stroke disease using convolutional neural network. *Journal of Physics*: Conference Series. 978: 012092. https://doi.org/10.1088/1742-6596/978/1/012092

Marks, M. P., de Crespigny, A., Lentz, D., Enzmann, D. R., Albers, G. W., & Moseley, M. E. (1996). Acute and chronic stroke: navigated spin-echo diffusion-weighted MR imaging. *Radiology*, *199*(2), 403–408. https://doi.org/10.1148/radiology.199.2.8668785

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage*, *2*(2), 89–101. https://doi.org/10.1006/nimg.1995.1012

McLouth, J., Elstrott, S., Chaibi, Y., Quenet, S., Chang, P. D., Chow, D. S., & Soun, J. E. (2021). Validation of a Deep Learning Tool in the Detection of Intracranial Hemorrhage and Large Vessel Occlusion. *Frontiers in neurology*, *12*, 656112. https://doi.org/10.3389/fneur.2021.656112

Merwick, Á., & Werring, D. (2014). Posterior circulation ischaemic stroke. *BMJ (Clinical research ed.)*, *348*, g3175. https://doi.org/10.1136/bmj.g3175

Montaha, S., Azam, S., Rafid, A.R., Hasan, M.Z., Karim, A., & Islam, A. (2022). TD-CNN-LSTM: A Hybrid approach combining CNN and LSTM to classify brain tumor on 3D MRI scans performing ablation study. I*EEE Access*, PP**,** 1-1. https://doi.org/10.1109/ACCESS.2022.3179577

Moustafa, R. R., Izquierdo-Garcia, D., Jones, P. S., Graves, M. J., Fryer, T. D., Gillard, J. H., Warburton, E. A., & Baron, J. C. (2010). Watershed infarcts in transient ischemic attack/minor stroke with > or = 50% carotid stenosis: hemodynamic or embolic?. *Stroke*, *41*(7), 1410–1416. https://doi.org/10.1161/STROKEAHA.110.580415

Nagel, S., Sinha, D., Day, D., Reith, W., Chapot, R., Papanagiotou, P., Warburton, E. A., Guyler, P., Tysoe, S., Fassbender, K., Walter, S., Essig, M., Heidenrich, J., Konstas, A. A., Harrison, M., Papadakis, M., Greveson, E., Joly, O., Gerry, S., Maguire, H., … Grunwald, I. Q. (2017). e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. *International journal of stroke : official journal of the International Stroke Society*, *12*(6), 615–622. https://doi.org/10.1177/1747493016681020

Nakano, S., Iseda, T., Kawano, H., Yoneyama, T., Ikeda, T., & Wakisaka, S. (2001). Correlation of early CT signs in the deep middle cerebral artery territories with angiographically confirmed site of arterial occlusion. *AJNR. American journal of neuroradiology*, *22*(4), 654–659.

Naganuma, M., Tachibana, A., Fuchigami, T., Akahori, S., Okumura, S., Yi, K., Matsuo, Y., Ikeno, K., & Yonehara, T. (2021). Alberta Stroke Program Early CT Score Calculation Using the Deep Learning-Based Brain Hemisphere Comparison Algorithm. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association*, *30*(7), 105791. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105791

Naidich, T.P., Castillo, M., Cha, S., & Smirniotopoulos, J.G. (2012). *Imaging of the Brain: Expert Radiology Series*. Elsevier Health Sciences.

Ng, Y. S., Stein, J., Ning, M., & Black-Schaffer, R. M. (2007). Comparison of clinical characteristics and functional outcomes of ischemic stroke in different vascular territories. *Stroke*, *38*(8), 2309–2314. https://doi.org/10.1161/STROKEAHA.106.475483.

Nishi, H., Oishi, N., Ishii, A., Ono, I., Ogura, T., Sunohara, T., Chihara, H., Fukumitsu, R., Okawa, M., Yamana, N., Imamura, H., Sadamasa, N., Hatano, T., Nakahara, I., Sakai, N., & Miyamoto, S. (2020). Deep Learning-Derived High-Level Neuroimaging Features Predict Clinical Outcomes for Large Vessel Occlusion. *Stroke*, *51*(5), 1484–1492. https://doi.org/10.1161/STROKEAHA.119.028101

Nixon, M.S., Aguado, A.S. (2012). *Feature extraction & image processing for computer vision* (3rd ed.).Elsevier: Academic Press.

Nogueira, R. G., Jadhav, A. P., Haussen, D. C., Bonafe, A., Budzik, R. F., Bhuva, P., Yavagal, D. R., Ribo, M., Cognard, C., Hanel, R. A., Sila, C. A., Hassan, A. E., Millan, M., Levy, E. I., Mitchell, P., Chen, M., English, J. D., Shah, Q. A., Silver, F. L., Pereira, V. M., … DAWN Trial Investigators (2018). Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *The New England journal of medicine*, *378*(1), 11–21. https://doi.org/10.1056/NEJMoa1706442

Öman, O., Mäkelä, T., Salli, E., Savolainen, S., & Kangasniemi, M. (2019). 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *European radiology experimental*, *3*(1), 8. https://doi.org/10.1186/s41747-019-0085-6

Osborn, A.G. (2015). *Osborn Beyin Görüntüleme Patoloji ve Anatomi.* Dünya Tıp Kitapevi, Ankara

Ovbiagele, B., & Nguyen-Huynh, M. N. (2011). Stroke epidemiology: advancing our understanding of disease mechanism and therapy. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, *8*(3), 319–329. https://doi.org/10.1007/s13311-011-0053-1

F.R.S., K.P. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 1, 2*, 559-572.

Pereira, D.R., Filho, P., de Rosa, G., Papa, J.P., & Albuquerque, V.H. (2018). Stroke Lesion Detection Using Convolutional Neural Networks. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1-6.

Pinto, A., Pereira, S., Meier, R., Wiest, R., Alves, V., Reyes, M., & Silva, C. A. (2021). Combining unsupervised and supervised learning for predicting the final stroke lesion. *Medical image analysis*, *69*, 101888. https://doi.org/10.1016/j.media.2020.101888

Powers, W. J., Rabinstein, A. A., Ackerson, T., Adeoye, O. M., Bambakidis, N. C., Becker, K., Biller, J., Brown, M., Demaerschalk, B. M., Hoh, B., Jauch, E. C., Kidwell, C. S., Leslie-Mazwi, T. M., Ovbiagele, B., Scott, P. A., Sheth, K. N., Southerland, A. M., Summers, D. V., Tirschwell, D. L., & American Heart Association Stroke Council (2018). 2018 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*, *49*(3), e46–e110. https://doi.org/10.1161/STR.0000000000000158

Praveen, G. B., Agrawal, A., Sundaram, P., & Sardesai, S. (2018). Ischemic stroke lesion segmentation using stacked sparse autoencoder. *Computers in biology and medicine*, *99*, 38–52. https://doi.org/10.1016/j.compbiomed.2018.05.027

Pressman, B. D., Tourje, E. J., & Thompson, J. R. (1987). An early CT sign of ischemic infarction: increased density in a cerebral artery. *AJR. American journal of roentgenology*, *149*(3), 583–586. https://doi.org/10.2214/ajr.149.3.583

Reinhold, J. C., Dewey, B. E., Carass, A., & Prince, J. L. (2019). Evaluating the Impact of Intensity Normalization on MR Image Synthesis. *Proceedings of SPIE--the International Society for Optical Engineering*, *10949*, 109493H. https://doi.org/10.1117/12.2513089

Rokach, L., & Maimon, O. (2005). Decision trees. In *Data Mining and Knowledge Discovery Handbook*, pp. 165–192, Springer, Boston.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv, abs/1505.04597*.

Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J., Culebras, A., Elkind, M. S., George, M. G., Hamdan, A. D., Higashida, R. T., Hoh, B. L., Janis, L. S., Kase, C. S., Kleindorfer, D. O., Lee, J. M., Moseley, M. E., Peterson, E. D., Turan, T. N., Valderrama, A. L., Vinters, H. V., … Council on Nutrition, Physical Activity and Metabolism (2013). An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, *44*(7), 2064–2089. https://doi.org/10.1161/STR.0b013e318296aeca

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510-4520.

Schmahmann J. D. (2003). Vascular syndromes of the thalamus. *Stroke*, *34*(9), 2264–2278. https://doi.org/10.1161/01.STR.0000087786.38997.9E

Serrà, J., Surís, D., Miron, M., & Karatzoglou, A. (2018). Overcoming catastrophic forgetting with hard attention to the task. *International Conference on Machine Learning*.

Shaham, U., & Lederman, R.R. (2018). Learning by coincidence: Siamese networks and common variable learning. *Pattern Recognit., 74*, 52-63.

Schapire, R.E. (2013). Explaining AdaBoost. *Empirical Inference*, 37–52. https://doi.org/ 10.1007/978-3-642-41136-6_5

Shinohara, Y., Takahashi, N., Lee, Y., Ohmura, T., Umetsu, A., Kinoshita, F., Kuya, K., Kato, A., & Kinoshita, T. (2020). Usefulness of deep learning-assisted identification of hyperdense MCA sign in acute ischemic stroke: comparison with

readers' performance. *Japanese journal of radiology*, *38*(9), 870–877. https://doi.org/10.1007/s11604-020-00986-6

Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019). Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology*, *290*(3), 590–606. https://doi.org/10.1148/radiol.2018180547

Soun, J. E., Chow, D. S., Nagamine, M., Takhtawala, R. S., Filippi, C. G., Yu, W., & Chang, P. D. (2021). Artificial Intelligence and Acute Stroke Imaging. *AJNR. American journal of neuroradiology*, *42*(1), 2–11. https://doi.org/10.3174/ajnr.A6883

Srinivasan, A., Goyal, M., Al Azri, F., & Lum, C. (2006). State-of-the-art imaging of acute stroke. *Radiographics : a review publication of the Radiological Society of North America, Inc*, *26 Suppl 1*, S75–S95. https://doi.org/10.1148/rg.26si065501

Subudhi, A.K., Dash, M., & Sabut, S.K. (2020). Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. *Biocybernetics and Biomedical Engineering*. Volume 40, Issue 1, https://doi.org/10.1016/j.bbe.2019.04.004

Sweid, A., Hammoud, B., Ramesh, S., Wong, D., Alexander, T. D., Weinberg, J. H., Deprince, M., Dougherty, J., Maamari, D. J., Tjoumakaris, S., Zarzour, H., Gooch, M. R., Herial, N., Romo, V., Hasan, D. M., Rosenwasser, R. H., & Jabbour, P. (2019). Acute ischaemic stroke interventions: large vessel occlusion and beyond. *Stroke and vascular neurology*, *5*(1), 80–85. https://doi.org/10.1136/svn-2019-000262

Takahashi, N., Lee, Y., Tsai, D. Y., Matsuyama, E., Kinoshita, T., & Ishii, K. (2014). An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiological physics and technology*, *7*(1), 79–88. https://doi.org/10.1007/s12194-013-0234-1

Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ArXiv,* abs/1905.11946.

Tao, W. D., Liu, M., Fisher, M., Wang, D. R., Li, J., Furie, K. L., Hao, Z. L., Lin, S., Zhang, C. F., Zeng, Q. T., & Wu, B. (2012). Posterior versus anterior circulation infarction: how different are the neurological deficits?. *Stroke*, *43*(8), 2060–2065. https://doi.org/10.1161/STROKEAHA.112.652420

Thornhill, R. E., Lum, C., Jaberi, A., Stefanski, P., Torres, C. H., Momoli, F., Petrcich, W., & Dowlatshahi, D. (2014). Can shape analysis differentiate free-floating internal carotid artery thrombus from atherosclerotic plaque in patients evaluated with CTA for stroke or transient ischemic attack?. *Academic radiology*, *21*(3), 345–354. https://doi.org/10.1016/j.acra.2013.11.011

Tomsick, T., Brott, T., Barsan, W., Broderick, J., Haley, E. C., Spilker, J., & Khoury, J. (1996). Prognostic value of the hyperdense middle cerebral artery sign and stroke scale score before ultraearly thrombolytic therapy. *AJNR. American journal of neuroradiology*, *17*(1), 79–85.

Topcuoglu, M. A., & Ozdemir, A. O. (2023). Acute stroke management in Turkey: Current situation and future projection. *European stroke journal*, *8*(1 Suppl), 16–20. https://doi.org/10.1177/23969873221103943

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, *9*(nov), 2579-2605.

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J. C., Pieper, S., & Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer research*, *77*(21), e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339

van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European radiology*, *31*(6), 3797–3804. https://doi.org/10.1007/s00330-021-07892-z

Vargas, J., Spiotta, A., & Chatterjee, A. R. (2018). Initial Experiences with Artificial Neural Networks in the Detection of Computed Tomography Perfusion Deficits. *World neurosurgery*, S1878-8750(18)32382-9. Advance online publication. https://doi.org/10.1016/j.wneu.2018.10.084

Warlow, C.P., Dennis, M.S., Gijn, J.C., Hankey, G.J., & Barnett, H.J. (2001). Stroke : a practical guide to management. ISBN 10: 0865428743 Wiley Blackwell Science Inc.

Winzeck, S., Hakim, A., McKinley, R., Pinto, J. A. A. D. S. R., Alves, V., Silva, C., Pisov, M., Krivov, E., Belyaev, M., Monteiro, M., Oliveira, A., Choi, Y., Paik, M. C., Kwon, Y., Lee, H., Kim, B. J., Won, J. H., Islam, M., Ren, H., Robben, D., … Reyes, M. (2018). ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI. *Frontiers in neurology*, *9*, 679. https://doi.org/10.3389/fneur.2018.00679

Wikimedia commons (2023, Dec 12).*By NEUROtiker - Own work, CC BY-SA 3.0.* https://commons.wikimedia.org/w/index.php?curid=2653584

Wikimedia commons (2023, Dec 12). *By Rhcastilhos Gray519.png, Public Domain.* https://commons.wikimedia.org/w/index.php?curid=1597012

Woo, S., Park, J., Lee, J., & Kweon, I. (2018). CBAM: Convolutional Block Attention Module. *ArXiv, abs/1807.06521***.**

Wycliffe, N. D., Choe, J., Holshouser, B., Oyoyo, U. E., Haacke, E. M., & Kido, D. K. (2004). Reliability in detection of hemorrhage in acute stroke by a new three-dimensional gradient recalled echo susceptibility-weighted imaging technique compared to computed tomography: a retrospective study. *Journal of magnetic resonance imaging : JMRI*, *20*(3), 372–377. https://doi.org/10.1002/jmri.20130

Xanthopoulos, P., Pardalos, P.M., & Trafalis, T.B. (2013). Linear Discriminant Analysis. in Robust Data Mining, pp. 27–33, Springer, New York, NY, 2013.

Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2020). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical image analysis*, *69*, 101985 .

Yeo, M., Kok, H. K., Kutaiba, N., Maingard, J., Thijs, V., Tahayori, B., Russell, J., Jhamb, A., Chandra, R. V., Brooks, M., Barras, C. D., & Asadi, H. (2021). Artificial intelligence in clinical decision support and outcome prediction - applications in stroke. *Journal of medical imaging and radiation oncology*, 10.1111/1754-9485.13193. Advance online publication. https://doi.org/10.1111/1754-9485.13193

Yogananda, C. G. B., Shah, B. R., Nalawade, S. S., Murugesan, G. K., Yu, F. F., Pinho, M. C., Wagner, B. C., Mickey, B., Patel, T. R., Fei, B., Madhuranthakam, A. J., & Maldjian, J. A. (2021). MRI-Based Deep-Learning Method for Determining Glioma *MGMT* Promoter Methylation Status. *AJNR. American journal of neuroradiology*, *42*(5), 845–852. https://doi.org/10.3174/ajnr.A7029

Yu, Y., Xie, Y., Thamm, T., Gong, E., Ouyang, J., Huang, C., Christensen, S., Marks, M. P., Lansberg, M. G., Albers, G. W., & Zaharchuk, G. (2020). Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA network open*, *3*(3), e200772. https://doi.org/10.1001/jamanetworkopen.2020.0772.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015

Zhai, J., Zhang, S., Chen, J., & He, Q. (2018). Autoencoder and Its Various Variants. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**,** 415-419.

Zhang, H., Polson, J. S., Nael, K., Salamon, N., Yoo, B., El-Saden, S., Scalzo, F., Speier, W., & Arnold, C. W. (2021). Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time. *Computerized medical imaging and*

graphics: the official journal of the Computerized Medical Imaging Society, *90*, 101926. https://doi.org/10.1016/j.compmedimag.2021.101926

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., Boellaard, R., Bogowicz, M., Boldrini, L., Buvat, I., Cook, G. J. R., Davatzikos, C., Depeursinge, A., Desseroit, M. C., Dinapoli, N., Dinh, C. V., Echegaray, S., … Löck, S. (2020). The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*, *295*(2), 328–338. https://doi.org/10.1148/radiol.2020191145