Deep Learning Applications for Finding Geometric Constraints of Chronic Subdural Hematomas

Rosalie Voorend
s2134861

Supervisors:
Dr. E. Thibeau-Sutre
Dr. J.M. Wolterink
Dr. ir. A.A.M. Spill

January, 2024

Abstract

This thesis explores the use of deep learning in the medical field, specifically in the diagnosis and treatment of chronic subdural hematoma (cSDH). The research focuses on developing automated methods to identify geometric constraints in brain CT images, a critical step in standardizing and enhancing the treatment of cSDH. The study involves brain stripping techniques, ideal midplane detection, and the development of a slice selection algorithm. It utilizes the Design Science Research Methodology for a structured approach and evaluates its solutions through numerical results and visual inspections. The brain stripping method can successfully identify skull masks. The midplanes are detected with a 3.07-degree angle difference from the ground truth.

Contents

Chapter 1

Introduction

Every year, up to 130 out of 100.000 elderly people are diagnosed with a chronic subdural hematoma (cSDH) [34]. cSDH is a pervasive neurological condition where blood is collected between the brain and the skull, resulting in constant pressure on the brain. While sometimes a cSDH can go unnoticed, in severe cases it can lead to neurological deficits, coma, and even death. In fact, for 16.7 percent of patients over the age of 65, cSDH is fatal [27]. Treatment of cSDH differs per case but typically includes undergoing drainage to remove accumulated blood or receiving follow-up care. The choice of whether a patient should be operated on or receive follow-up care is not standardized and differs not only per hospital but also per surgeon [38].

1.1 Background

1.1.1 Chronic Subdural Hematoma

A hematoma is an intracorporeal hemorrhage, typically caused by physical trauma, resulting in bleeding that is localized outside blood vessels. A bruise is a common form of a hematoma. These hematomas can also occur inside the brain or between the skull and the brain. Subarachnoid, subdural, and intracranial hematomas are examples of hematomas inside the brain, located in different brain regions. The focus of this report is on subdural hematomas [1], which are located between the dura mater and the arachnoid mater, see Figure 1.1. Subdural hematomas cause an increase in pressure on the brain, which may lead to compression of the brain tissue. If such a hematoma grows over several weeks, it is called a cSDH [28]. Unlike acute SDH, which grows immediately after trauma.

A cSDH is one of the most frequent neurosurgical conditions [48]. It is encountered more commonly amongst the elderly, with 130 incidents per 100,000 persons per year (measured from a population in Finland). Among adults younger than 70 years this number is lower, namely 8.2 to 17.6 incidents per 100,000 persons per year. [33].

CSDH can grow over weeks or months, which can increase intracranial pressure that could result in deforming and damaging regions of the brain. This can affect a patient with cSDH with light symptoms, such as headaches and nausea [1], or more serious neurological deficits. As the symptoms of cSDH can be non-specific to the disease, it can be hard to diagnose a cSDH. Besides neurological assessment, medical imaging can help in the identification of cSDH in a patient. For example, with CT scans, as shown in Figure 1.2. Figure 1.2 also shows the effects of the intracranial pressure caused by the hematoma.

In contrast to acute subdural hematomas, the prognosis of cSDH is more manageable and can be treated. Treatments include close observation of the patient and the growth

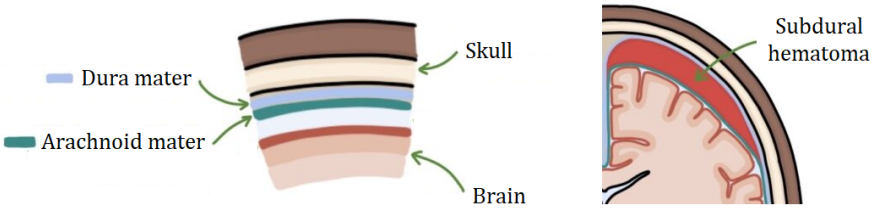

FIGURE 1.1: Subdural hematoma localization in the brain. The hematoma is located between the dura mater and the arachnoid mater. Meaning, it is between the outer most layer of the brain and the skull.

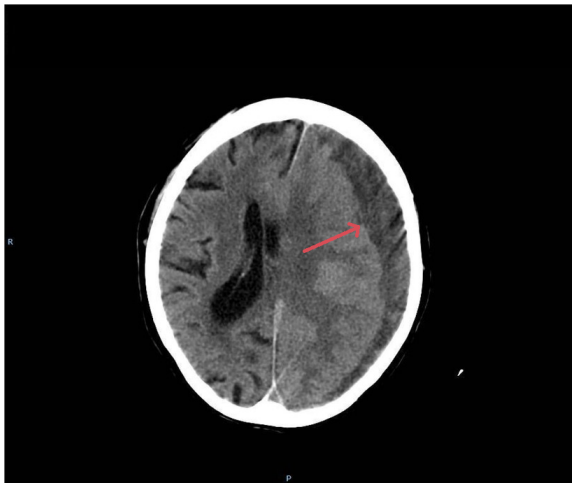

FIGURE 1.2: Chronic subdural hematoma in a CT scan. The darker matter is the hematoma, which is indicated by the arrow. The dark intensity of the hematoma suggests that this is not a new hematoma.

of the hematoma or surgical intervention, such as drainage or craniotomy for more severe cases [1]. The latter treatment should release the blood pressure in the brain. The decision boundary for such treatments depends heavily on the neurologist assessing the patient, which forms the basis of the clinical problem.

1.1.2 Clinical Problem

Current Protocol

When a patient with cSDH checks in to a hospital, the hematoma first has to be assessed. This is done through a Computed Tomography (CT) scan. In this scan, a radiologist can find accumulated blood between the dura mater and the arachnoid mater. Then, the radiologist picks a single slice in which they identify the hematoma to be the largest. Then they investigate how much the mass is deforming the brain. If a brain is severely deformed on one side, this might be an easy task. The brain is likely deformed asymmetrically. This is assessed by first drawing the ideal midline. This is done by connecting the frontal and occipital crest with a straight line. Then the deformation is measured by finding the membrane between the left and right ventricle. The closest distance from this membrane to the midline is called the midline shift. As shown in Figure 1.3. Most hospitals have a threshold of this midline shift that decides whether surgery is required.

Besides the midline shift, no quantified measures of the deformation are measured,

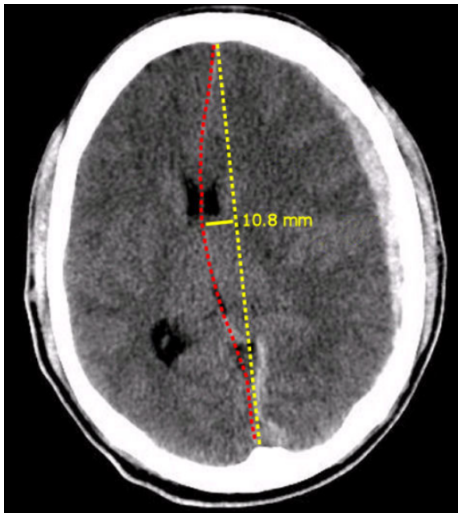

FIGURE 1.3: Ideal midline (yellow) and shifted midline (red). Also the midline shift of 10.8 mm has been indicated. [3]

except that occasionally the size of the hematoma is measured. Some other aspects are investigated too. The compression of the ventricles could indicate the pressure of the hematoma, but sometimes the fluid in the ventricles cannot drain properly anymore, hence the ventricles could grow. Furthermore, the intensity of the hematoma is checked, which indicates how long the hematoma has been present. A high intensity implies a more recent hematoma and vice versa.

If the radiologist decides that surgery might be needed, then a neurologist assesses the CT scan similarly. Lastly, if the neurologist decides surgery might be needed, a neurosurgeon also takes the same steps to make the final decision.

Besides looking at the cSDH scan, the neurosurgeon from MST emphasizes that they operate based on patients and not only based on the scan. The evaluation of a patient's cognitive functions is also taken into account.

This procedure is based on interviews with one neurosurgeon from MST. It is thus not certain if the same procedure applies to all hospitals or even all neurosurgeons at MST. This also leads us to the limitations of the current procedure.

Limitations of the Current Protocol

The first and most prevalent limitation is the lack of consistency in the current assessment. Each of the measurements in the current protocol is done by hand and is thus subjective to the physician. Drawing the midline, finding the spot between the left and right ventricle, and finding the shortest distance to the ideal midline are all subjective steps. The only quantified measure is the size of the midline shift in millimeters.

The second limitation is that the measurement is done on a single 2D slice. The CT image that the radiologist obtains is a 3D image and thus contains much more information than what might be visible on one slice. For example, if a patient's head is tilted during the making of the scan, the measured midline shift could result in a smaller or larger number than it is.

Another limitation regards bilateral cSDHs. This is when a patient has a cSDH on both sides of the brain. This would result in pressure coming from both sides, which could compress important brain regions, but it might not be visible on the scan. If both sides

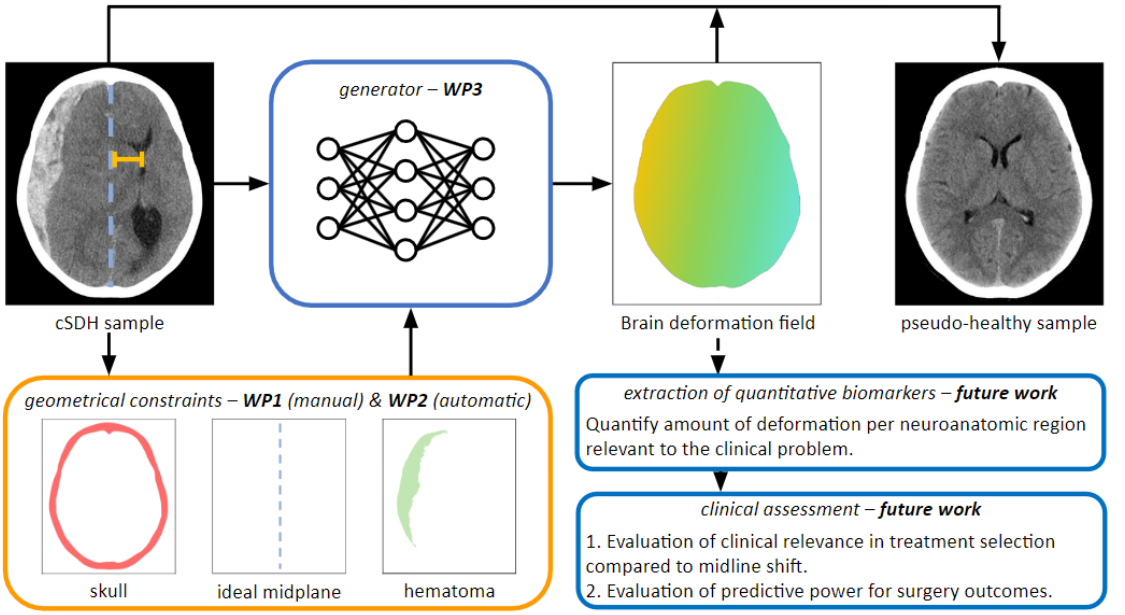

FIGURE 1.4: Pipeline of the Brain-SHIFT project for the computation of the brain deformation field using deep learning methods. The orange block highlights the geometric constraints, which are the focus of this report.

apply the same amount of pressure, there would be no midline shift and strictly speaking, surgery would not be necessary.

The last limitation is in the measurement of the midline shift. This is a very local measurement that decides the pressure of the cSDH on the whole brain. It is very hard to tell which regions of the brain the pressure is most present. For neurologists, it is important to know which regions are affected since more vital regions would increase the need for surgery. For example, pressure on the brain stem, which could lead to unconsciousness, should be prioritized over pressure on other regions of the brain. There is much more data available in the 3D CT image as a whole, yet only a threshold of 5 mm is taken as a guideline [38].

1.1.3 Brain-SHIFT Project

Due to the lack of standardization and global measurements, the Brain-SHIFT project is established. The Brain-SHIFT project aims to standardize the current assessment with the use of deep learning. It does so, by finding a deformation field that can map a a CT image with cSDH to a pseudo-healthy brain, as displayed in Figure 1.4. In this deformation field, the resulting pressure from the hematoma can be found in each part of the brain. This also aids the neurologist in deciding whether essential parts of the brain are affected, such as speech or consciousness. Furthermore, the Brain-SHIFT project assesses the CT scan in a 3D manner. Where clinicians base their decisions on one local measurement, the deformation field can provide a global assessment. Additionally, deep learning allows for the decision process to be fully automated. It can form a precise, consistent guideline that can greatly aid clinicians.

Considering the lack of available data, mainly the lack of scans with hematomas and healthy scans of the same patient, the available data must be used to its full potential.

Without having a healthy version of a brain, it is unknown what the deformation field will look like, but it is known that there are several constraints that the deformation field must adhere to.

Brain Stripping

The first geometric constraints should ensure that the skull does not change and that the brain stays inside the skull. For this, the skull, and only the skull, should be extracted. This means that the brain and the scalp should be removed and only a clean skull is left, hence the name brain stripping. Also, calcification that may be present inside the brain should be removed. Furthermore, no traces of the CT machine should be included in the skull extraction. This can be challenging as the machine may have the same intensity as the skull in many images. If the skull is deformed, e.g. due to trauma, this should not be affected.

Ideal Midplane

A healthy brain is nearly symmetric and can be divided in half with an imaginary line [43]. This imaginary line is the second geometric constraint and is called the ideal midline. The ideal midline can be drawn on CT slices, as was shown in yellow in Figure 1.3. With 3D images, it can be done in 3D manner. This would be a midplane that could be defined as the plane that delineates the left hemisphere from the right hemisphere.

cSDH Segmentation Mask

The segmentation mask delineates the precise boundaries of the cSDH in the CT scan. This assigns a unique pixel or voxel value to each part of the image that contains the cSDH.

1.2 Research Question

Automatically obtaining these geometric constraints that can aid the finding of a deformation field, is the focus of this report. Thus finding the skull and the midplane. The segmentation mask is beyond the scope, due to a lack of available data at the time of writing. Because the deformation field should be found fully automatically, also the constraints should be found fully automatically. To find these, the following research questions have been established: *To what extent can deep learning be used to detect geometric constraints of brain CT images automatically?* The following sub-questions aid to answer the main question:

1. To what extent can the skull be obtained by applying a brain-stripping algorithm to brain CT images?
2. To what degree can the ideal midplane be obtained automatically from brain CT images?
3. What is the proposed value of the Brain-SHIFT project for the clinical decision-making process for treating cSDH?

1.3 Overview

The document starts with assessing the clinical value of the Brain-SHIFT project. Then, a relevant literature review for the geometric constraints is assessed. The methodology then describes the dataset used and the steps taken to find the geometric constraints. Next, the results are given and discussed. Lastly, limitations and future work are discussed and a conclusion is drawn.

Chapter 2

Clinical Value

This chapter elaborates on the clinical value of the project. It is evaluated whether the goals of the project are feasible and tackling the correct problems that are in the ad-hoc procedure. A risk analysis, a swot analysis, and finally, a value proposition are applied to constitute the clinical relevance.

2.1 Opportunity of the Project

The main drive behind the project is to aid neurologists in deciding if they should operate on a patient or not. This should not only help standardize the decision process but also make it more efficient. Another drive behind the project is in the ease of use because the product identifies pressure on the brain which is otherwise not visible from a CT scan. Also, time-effectiveness is a drive, because the tool can identify parts of the brain automatically which saves time compared to having to do so manually. However, implementing an AI medical tool comes with a lot of risks and thus the final product should be evaluated strictly before it is deployed. To assess these risks, a risk analysis is done. Additionally, a SWOT analysis is done to evaluate the project's position regarding opportunities and possible threats. Finally, a value proposition proposes the project as a whole.

2.1.1 Risk Analysis

A risk analysis is produced to assess the possible risks of the project. As the project is already a work in progress, the risk assessment is concentrated on the deployment of the project. Specifically, the adoption of the project regarding the acceptance of clinicians, ethical decisions, and privacy regulations. Furthermore, the risks of the daily usage and maintenance of the project in clinics are discussed. Although the funding is beyond the scope of this research, some financial risks regarding the potential obsoleteness of the project are briefly touched on. For each risk owner, the roles are very general as there is currently not a large team behind the project so creating very specific roles sets unrealistic expectations.

The risks and mitigation or contingency actions are defined in the risk analysis in the appendix in section 8.6. They can be categorized into the following categories: *Information Security Risks*, *Technology Risks*, *Technological Environment Risks*, *Human Risks*, *Project Planning Risks*, and *Financial Risk*s. The first three can be linked under a broader umbrella term *Technological Risks*.

Technological Risks

These risks are all revolved around the data and trained model. The mitigation actions for all of these risks constitute prevention measures such as periodic reviews of the source of risk. The data-specific risks can be mitigated by thoroughly checking the data, and collecting data from other sources. The privacy regulation risks can be mitigated with careful handling of the data, as well as firewalls, encryption, and security protocols to prevent possible attacks. The environmental risks can be mitigated by obtaining a clear understanding of the technical environment and adapting the project accordingly. Alternatively, the technical environment can be transferred, but this comes with all the same other risks.

Human Risk

The next risk category is *Human Risks*. This concerns both the clinicians, as well as the patients who potentially undergo surgery. The first can be mitigated by familiarizing clinicians with the tool through clear introductions and instructions. It is also important to allow them to provide feedback. The risk regarding the patients must be mitigated by emphasizing that the neurologist always exercises ultimate authority over the decision. A contingency plan would include strict consideration of the use of the tool and again weighing its benefits against the risks.

Project Planning Risk

The next category is *Project Planning Risks*. These risks stress the clarity of the objective of the project. The appropriate mitigation action would be to ensure consistent and effective communication, for example by meeting regularly with all disciplines involved. These meetings should ensure clear definitions of roles and responsibilities. Taking unforeseen circumstances into account and drafting contingency plans accordingly is important too.

Financial Risk

The last category is *Financial Risks*. The project could fail if the estimated budget is exceeded. To mitigate this, planning the finances correctly, not only for the production process but also for the maintenance is needed. A detailed budget plan for such scenarios is necessary as a contingency plan. Nevertheless, budgeting is beyond the scope of this report.

2.1.2 SWOT Analysis

A SWOT analysis [40] was done to identify the project's Strengths, Weaknesses, Opportunities, and Threats. It is a business strategy that assesses an organization's or business product's position on the market. It is also useful to differentiate between internal versus external influences. In the case of the project, it can give insight into the strengths of the project and how these strengths can be affected when upscaling the project to other hospitals, for example. Figure 2.1 shows the SWOT analysis. Below, the four categories are discussed.

Strengths

The first strength of the project is the innovative approach. The ad-hoc process of deciding the severeness of the cSDH is done fully manually. Applying DL is a cutting-edge approach for medical imaging and might give new insights, undetectable by the human eye.

The second strength focuses on clinical relevance. The project is requested based on a neurosurgeon's experience, who needed enhancement and standardization in deciding whether a cSDH requires operating. Although one neurosurgeon is not representative of all neurologists, it does point out a clinical need.

The third strength addresses the standardized precision that comes with DL. Where neurologists now estimate midline shifts and hematoma sizes, the DL model does this systematically. This improves the precision of the current click-and-drag method.

Next is data-driven insights, which allow for faster processing of CT scans. This also gives the ability to analyze more CT scans, which can lead to new insights.

The last strength, interdisciplinary knowledge collaboration, addresses the collaboration between clinicians and researchers. By combining skills and knowledge, the full scope of the project can be tackled effectively.

Weaknesses

The project handles patient data, which may be a weakness if not addressed properly. It requires careful and strict guidelines on privacy.

Furthermore, the project is data-dependent. The data that is fed to train the final product defines the outcome. The project will be trained on data from the MST, which should not create a problem when deploying it in the MST, but it does raise a problem for upscaling.

Another weakness is the complexity of the project. Although the final tool should be user-friendly, there is still a possibility that the tool will break. To counteract this, specific DL knowledge is required.

Lastly, the production of the tool requires computational resources.

Opportunities

The first opportunity is in the technological advancement. It uses an innovative approach, which gives the opportunity to easily implement even newer DL approaches, compared to the manual approach.

The project could result in a growing interest in AI in healthcare. Clinicians have the opportunity to use the tool and see its benefits in practice, without feeling threatened that it might replace their job, as it is only an aiding tool. This does assume the product works as expected.

The final project leaves room for publication, which can attract interest from outside MST. This leads to the possibility of upscaling and thus deploying the project in more hospitals. This would also lead to the availability of more data that can improve the model.

The last opportunity is future research pathways. When deployed, more people come in touch with AI in healthcare, which may lead to new perspectives and opportunities that could be researched.

Threats

The first threat is regarding data privacy, which was also shown to be a weakness. Strictly handling this data according to appropriate privacy laws and ethical considerations is very important. An attack on privacy or a mistake in data handling can have a very large negative impact.

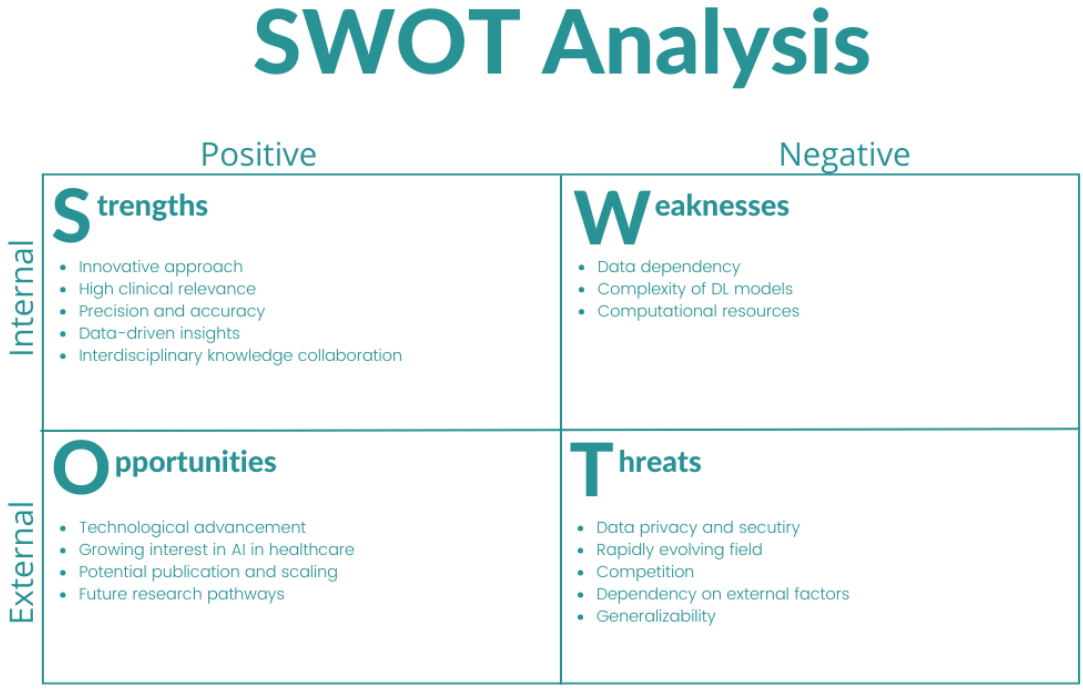

FIGURE 2.1: SWOT Analysis of the project

The next threat is the rapidly evolving field that the project positions itself in, which forms competition. Although this showed to be a strength, it also forms a threat. Newer models might evolve that easily outperform the project, or even make the project obsolete.

Another threat is the dependency on external factors. These include software availabilities and collaborations with MST. The latter could form a threat when neurologists lack trust in AI. Then the deployment would be obsolete.

The last threat is in generalizability. The final product is trained on data from MST, which forms a bias towards those specific images. CT scans that are made with different machines from different hospitals might lead to worse results.

2.1.3 Value Proposition

To ensure the final project meets the needs of the clinicians, a value proposition canvas is created. The left side of the canvas on Figure 2.2 outlines the products and services, gain creators, and pain relievers offered by the AI tool. It emphasizes benefits such as the AI's ability to estimate the deformation field of a brain with cSDH, automatically detect certain features, and process multiple scans efficiently without the need for manual annotation. This leads to more accurate insights and efficient processing, thus creating value for clinicians. The right side of the canvas focuses on the customer segment, in this case, clinicians, and details their jobs, pains, and gains associated with diagnosing cSDH. The AI tool aims to relieve pains such as the lack of standardized protocols and the tediousness of manually annotating CT scans while enhancing gains like standardized protocols and the ability to make informed operating decisions.

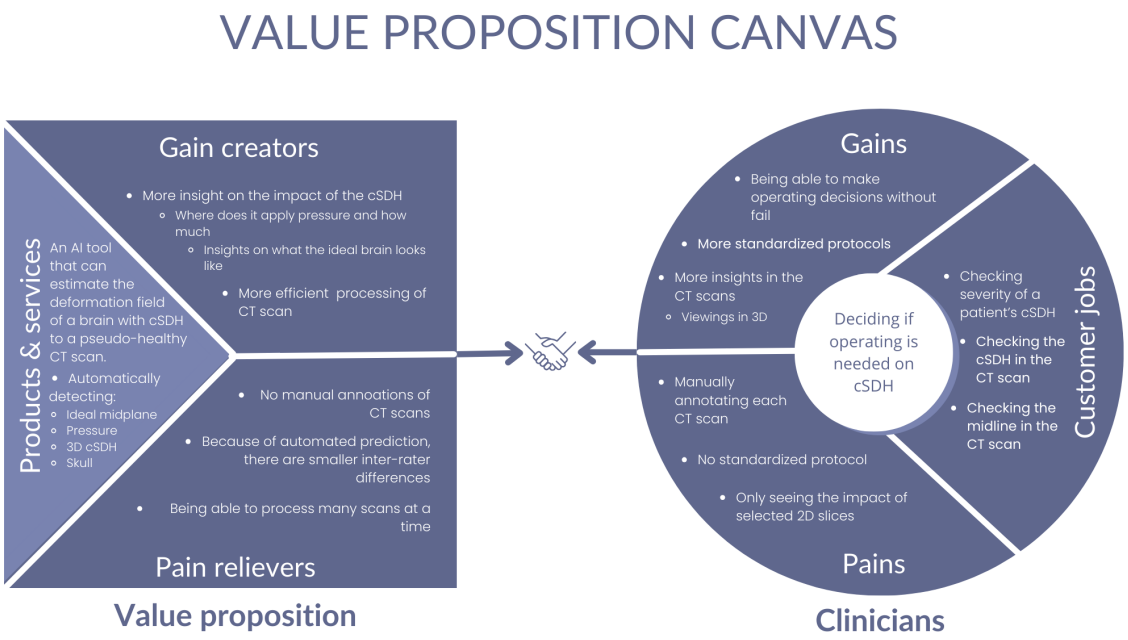

FIGURE 2.2: Value proposition canvas of the project

Higher efficiency

In the current protocol, each scan has to be assessed by three clinicians who all annotate a scan themselves and have to measure their annotations manually. This process takes each clinician approximately five minutes, according to the clinicians at MST. As a result, they only have a midline shift measure, which is an indication of deformation, but gives no insight into what regions of the brain are affected and how much they are affected. Retrieving the deformation field is done automatically and provides global insights. It would likely take some time for the deformation field to load, but because it is done automatically a clinician can do something else in the meantime. Also not each clinician in the protocol has to do this, but if the radiologist does it once, it is also available for the other clinicians.

Standardized Protocol

Currently, three clinicians look at one scan with no baseline. Assessing the presence of a midline shift does not always give the same outcome. Three clinicians annotating the same dataset only agreed on a midline shift 88% of the time [8]. Also, the presence of a subdural hematoma was agreed on only 87% of the time. Assessing the image with AI can reach higher values than that [49][42][21]. Although the deformation field does not replace clinicians, it definitely provides valuable and consistent insights.

Application Beyond cSDH

Although the Brain-SHIFT project is aimed at cSDH, the deformation field provides insight into the whole head, based on a single mass removal. Currently, the project is developed for a cSDH, but it could also be trained for other types of hematomas, hemorrhages, and tumors. The detection of these with AI has already shown to be effective [49][42][21], so expanding the deformation field would also be plausible.

This Brain-SHIFT project not only streamlines the current protocols, saving valuable time but also ensures consistency in results, which has been a significant issue with manual annotations where agreement on midline shifts and SDH detection is far from perfect. The deformation field fosters standardized procedures and empowers clinicians with more accurate and consistent insights for informed decision-making in the diagnosis and treatment of cSDH.

Chapter 3

Literature Review

This chapter discusses relevant literature regarding the three geometrical constraints. For each constraint, methods are discussed and compared. The research was conducted by looking up academic literature and selecting the most relevant ones. The essence of the selected papers is integrated below. This not only provides the state of the art but also some background information on methods used later in this report.

3.1 Brain Stripping

Brain stripping has been done in different ways, leveraging windowing, thresholding, and later deep learning techniques. Because medical images often have different imaging parameters [15], different methods can tackle different challenges that come with brain stripping. These include the range of intensity of images [16], varying brain structures [14], and motion artifacts [51]. Moreover, the outcome of applying different methods can diverge, even when applied to the same input images.

For the task at hand, the brain stripping method mustn't be too computationally expensive, as it is only a small step before being fed into a larger network.

Kalavathi and Prasath already compared different techniques [19] on skull stripping, where the brain is extracted from images. Although this is a different goal, it does aim to separate the brain from the skull region and other structures present in the image, hence it is a very similar process to brain stripping. Thus, the most relevant techniques are presented below.

3.1.1 Intensity-Based Methods

Intensity-based skull stripping methods rely on the intensity of pixel values of the image. The distribution of intensities can be modeled with histogram-based, edge-based, or region-growing methods. Then, by classifying which intensity values belong to which brain structure, the region of interest (ROI) can be found automatically. This is often done with a statistical analysis of the intensity distribution. Often, methods are also enhanced with additional cleaning of the resulting stripped image [19].

An early intensity-based method was created by Ward, using existing software that automatically detects intracranial regions [45][10]. It applies the Nelder-Mead method to the image, which is a search optimization method to converge to a local minimum or maximum of a multidimensional space, similar to gradient descent. Then for different brain structures, upper and lower bound intensity values are set to exclude non-brain pixels. Afterward, a connected-component analysis is performed to check whether the resulting

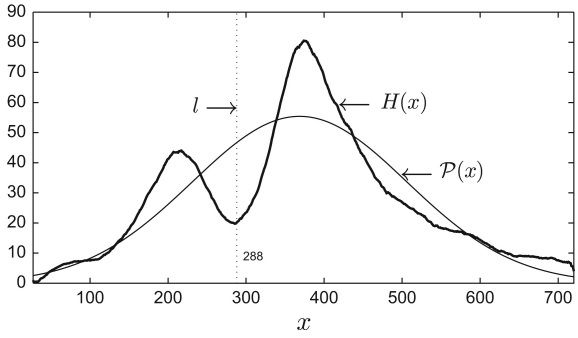

FIGURE 3.1: Histogram partitioning (l) using the maximum distance between histogram distribution ($H(x)$) of an image and a normal Gaussian distribution ($P(x)$)[5]

object is one component, and if not, other components are excluded. This method was shown to be highly dependent on the setting of the upper and lower bounds of different brain structures.

Balan et al. introduced a smart way of analyzing the histogram that is derived from the pixel values of the image [5]. Existing methods that they compare to, simply take the value in between to turning points as a threshold, but this is not always an accurate delineation between different brain structures. Instead, the intensity distribution of the image is compared to a Gaussian normal distribution. At the point of maximum divergence between the two, the threshold value between two brain structures is set, as is shown in Figure 3.1. This resulted in very accurate partitionings of brain structures in multiple datasets.

3.1.2 Morphology-Based Methods

Morphology-based methods use a combination of thresholding and edge detection to find the ROI. Early methods depended on histogram thresholding and the anatomy of the brain [6]. By applying a set of morphological operations they can identify different brain structures and thus find the ROI. The histogram thresholding allows for automatic detection of intensity values of the image, realizing a fully automated process. In this case, the morphological operations are image-processing techniques that process images based on detecting and altering shapes in the images.

Shanthi et al. introduced a method that first applies histogram thresholding and then employs seed growth to find more specific brain regions [36] [18]. The seed growth uses values from the histogram to identify different brain structures, which are set as seeds. The pixels with these seed values are found on the image and by looking at the values of neighboring pixels the ROI grows until it reaches the boundary of that specific brain region.

Besides thresholding and morphological operation, a more recent skull stripping algorithm developed by Swiebocka-Wiek uses filtration to improve edge detection and thus provide better tissue separation [41]. However, the filtration sharpens the edges of the output image, which leads to the brain not being recreated ideally.

With histogram thresholding, morphology-based methods are limited, because it can be difficult to decide the threshold if the histogram does not indicate a clear distinction. Also, the resolution of the image could limit the thresholding as well as the morphological operations.

Rahmad et al. [32] use a hybrid method combining morphology- and intensity-based methods. The skull is separated from the brain using intensity thresholding. Next, a set of morphological operations is applied to acquire the region inside the skull and convert it to a mask. The skull is then subtracted from this mask, ensuring cavities inside the skull are not mistaken as part of the skull or brain. They provide an efficient method with an emphasis on the creation of accurate masks of the brain and keeping the cavities in the anterior part of the brain (Figure **??** shows such a cavity).

Besides the presented methods, there are many more brain-stripping methods. However, the biggest difference is found in setting the threshold values and deciding the order of morphological operations. Also, more hybrid methods are available, but not always more efficient. Furthermore, more computationally expensive methods are not discussed, as they are less relevant to the task at hand.

3.1.3 Deep Learning Methods

Identifying the skull can also be seen as a segmentation task that can be performed by a deep learning model. As this method is a segmentation task, more on this is explained in section 3.2.2.

3.2 Ideal Midplane Detection

Many physicians still manually estimate the ideal midline on a CT scan, to find out how big the midline shift is [13]. Research shows that automatically finding this ideal midline or midplane has been done in different manners already. However, the current research mainly tackles this problem in a 2D manner, using slices of brain CT scans, instead of the whole image.

Finding ideal midlines is less common in the available literature than finding the deformed midlines. Therefore, the presented methods below are not only about finding the ideal midlines but the most relevant works are presented and discussed below.

3.2.1 Early Fully Automated Methods

Liu et al. introduced one of the first fully automated methods using anatomical markers to delineate the deformed midline [26]. First, the most ideal slice of the CT image is selected based on where the midline shift is most visible. Then, markers on the brain are detected, as shown in Figure 3.2. Points A and B are found by following the inner skull and finding the largest turning points. The ideal midline is also extracted by connecting these points. Points C and D are found as centroids of the lateral ventricle and the occipital horn respectively. Points E and F are found using a vectorization algorithm that finds the directional single connected chain of pixels. Finally, the deformed midline is formed by connecting all the best candidates for points A, B, C, D, E, and F.

Chen et al. introduced an automatic method for identifying the midline shift in brain magnetic resonance (MR) images [7]. First, the ideal midline is found by connecting the frontal and lateral falces. The deformed midline is then estimated based on the size and spatial information of brain lesions, like hematomas, compared to the ideal midline. This is then enhanced by checking local symmetry and intensity gradient symmetry of areas near the estimated midline.

Liao et al. introduced a method that looks at three markers on the brain and draws a Bezier curve through them [25] [24]. First, the best slice is selected, scalp stripped, and

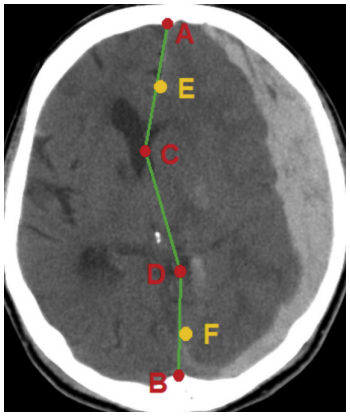

FIGURE 3.2: Markers of a deformed midline from [26]

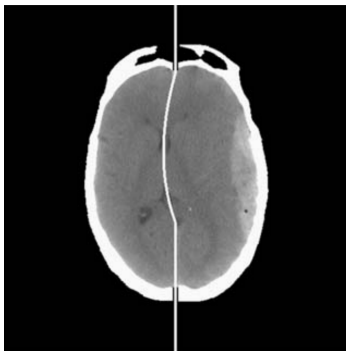

FIGURE 3.3: Estimated deformed midline from [25]

its position is adjusted according to the ideal midline. The first two points are the upper and lower straight segments that represent parts of the longitudinal fissure that divide the left and right hemispheres, like A-E and F-B in Figure 3.2 or as seen at the bottom of Figure 3.3. The third set of markers consists of three control points: the lowest point of the upper falx, the highest point of the lower falx, and the point at where there are free margins in the falx (approximately points E, F, and C in Figure 3.2). Finally, these points are connected with a Bezier curve, resulting in an approximation of the deformed midline looking like Figure 3.3.

Although these methods are (fully) automated, they are based on initial midline estimations that leave margins for mistakes. They are also very dependent on the quality of the image, so results could change for CT images taken on different machines.

3.2.2 Deep Learning Methods

Literature shows multiple methods for midline detection. Again, most methods are for real midlines and not ideal midlines. The methods are also in 2D and not in 3D, but the most relevant works are presented below.

Gupta et al. used deep learning to predict ideal midlines, deformed midlines, and midline shifts on non-contrasted brain CT images [22]. After skull stripping and brain extraction, they used a fully convolutional UNet [35] to segment the image into three classes: the

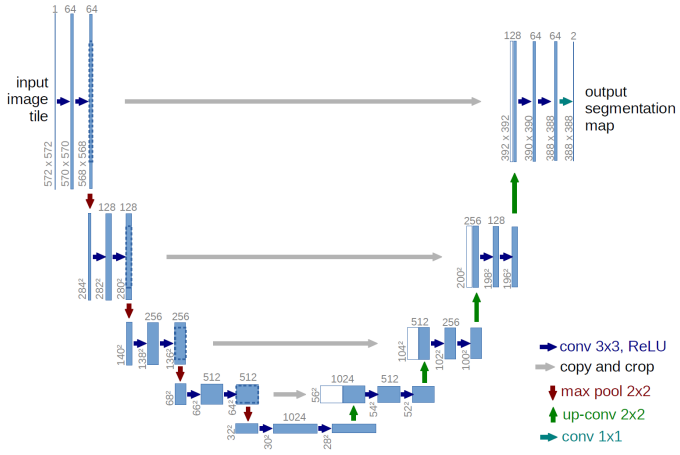

FIGURE 3.4: The UNet architecture [35]

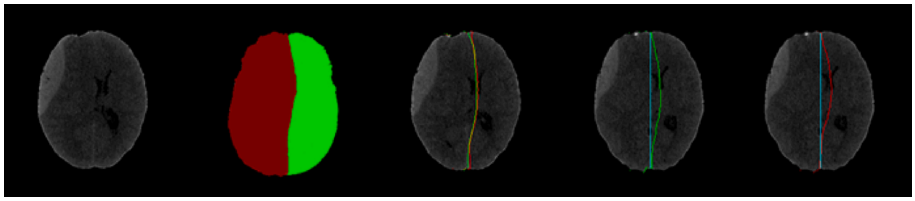

FIGURE 3.5: Midline detection with a UNet [22] From left to right: Input image; segmented hemispheres; predicted (red) and ground truth (green) midline; ideal (cyan) and ground truth (green) midline; ideal (cyan) and predicted (red) midline

left and right hemispheres and the background. With the predicted segmentations, the junction between the left and right hemispheres was detected and taken as the deformed midline. Then, the top-most and bottom-most coordinate points of the deformed midline were automatically detected by looking at minimum and maximum row indices. These points were connected with a straight line to form the ideal midline. Figure 3.5 shows how the ideal and deformed midlines are found from the predicted segmentation.

Wu et al. created the first 3D midline surface delineation for 3D CT images where hemorrhages are present [47]. They introduce a hemisphere segmentation network that delineates the midplane in three steps. The first step is a UNet that segments the whole brain, the left and right hemispheres, and the hematoma. By applying an edge detector to these segmentations, the initial midplane is found by looking at the edge between the left and right hemisphere, similar to the method above [22]. A distance-weighted map is introduced to train the UNet, which applies a gradient to the predicted and ground truth midplane, allowing for more efficient learning. The second step applies rectification learning to align the axes of the brain with the axes of the image. The last step is a ResNet for midline correction for cases with bilateral intraventricular hemorrhages since these hemorrhages tend to cross the midplane and thus mess up the prediction.

Since the available data and created annotations allow for a segmentation task, segmentation networks for medical imaging are investigated. The UNet, as was mentioned before, is a successful variant of this [50][2][37]. Over time, variations on the UNet have emerged that promise good segmentation results as well [4].

The literature on medical segmentations is not always for brain images, but they are for medical images nonetheless. The task of segmenting parts of the brain can be similar

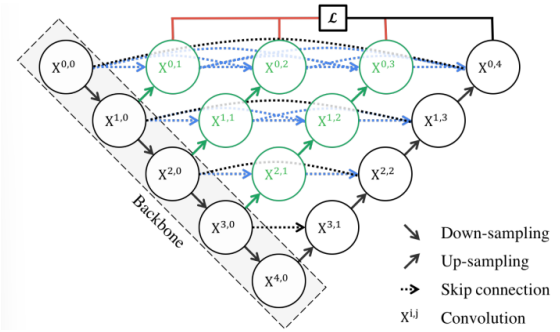

FIGURE 3.6: The UNet++ architecture [52]

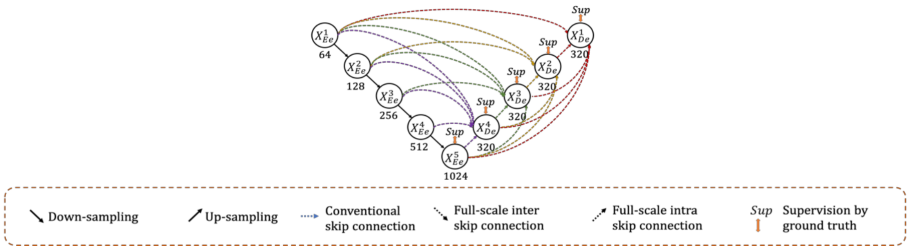

FIGURE 3.7: The UNet3+ architecture [17]

to the task of segmenting other body parts. Relevant literature on medical segmentations is shown and discussed below.

UNet++

Zhou et al. introduced the UNet++ [52], which has a similar architecture to the original UNet, depicted in Figure 3.4 [35]. The UNet++ was designed because according to Zhou et al., the UNet did not meet the strict segmentation requirements that play a large role when dealing with medical imaging. UNet++ differs from the original UNet in its skip connections: The aim was to lessen the semantic gap between the feature maps of the encoder and decoder part of the original UNet. The UNet++ uses a series of nested, dense, convolutional blocks as skip connections, as depicted in Figure 3.6. The number of convolutional layers of the convolution block depends on the pyramid level of the UNet++. Each convolutional layer is followed by a concatenation layer that merges the output from the convolutional layer above with the corresponding upsampled output of the lower convolutional block. With deep supervision [23] the model's performance is also enhanced, because this enables the use of only one segmentation branch as the final output.

UNet3+

The UNet3+, as introduced by Huang et al. [17], follows the development from UNet [35] to UNet++ [52] and builds upon this. Also the UNet3+ only differs in its implementation of the skip connections by introducing full-scale skip connections. This allows lower-level decoder layers to receive information from higher-level encoder layers. These higher-level encoder layers are downsampled with MaxPooling and the different levels of encodings are concatenated into one. This is also shown in Figure 3.8. Then, a UNet3+ also uses full-scale deep supervision, similar to the UNet++. However, in the UNet3+, the architecture yields a side output from each decoder stage, supervised by the ground truths.

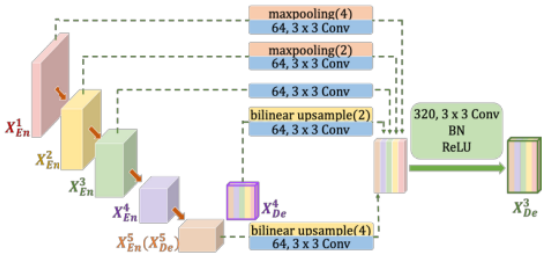

FIGURE 3.8: Full scale aggregated feature map from UNet3+ [17]

Each of the UNet adaptations has been shown to improve an image segmentation task for medical images. However, each segmentation task is different and although results are promising, none tackle the identification of the ideal midplane. Also, the midline detection networks are in a 2D manner and do not provide a method for 3D midplane identification.

Chapter 4

Methodology

This chapter explains the research approach. From the design methodology to an explanation of the dataset, followed by the steps taken for the geometric constraints, as well as their technical implementation.

4.1 Design Science Research Methodology

The methodology of this research follows Peffers' Design Science Research Methodology (DSRM) [31]. It is a commonly accepted framework for researching and implementing information systems and it includes six steps:

1. Problem identification and motivation
2. Defining objectives of the solution
3. Designing and developing the solution
4. Demonstrating the solution
5. Evaluating the solution
6. Communicating the solution

The six steps are also shown in Figure 4.1. This Figure also shows that the DSRM framework is a systematic, iterative approach to problem and context identification, development, evaluation, and communication. This study can ensure a structured process for effectively constructing the geometrical constraints. It allows for prototyping, testing, evaluating, and discussing. Evaluations are based on numerical results, visual inspections, and discussions will be held with team members and neurologists. New iterations will again be prototyped and implemented. This aids in making to final product not only functional but also clinically relevant.

Accompanying Peffer's DSRM, Wieringa's Design Science Cycle (DSC) [46] is incorporated. This also emphasizes the iterative process of the design process of an information system. Although this cycle is essentially very similar to Peffers DSRM, Figure 4.2 shows that this cycle has a focus on the interaction between the problem context and the solution domain. This stresses that the developed geometric constraints will be continuously adapted based on empirical evaluations. Both frameworks help create an accurate model for automatically recognizing the geometrical constraints.

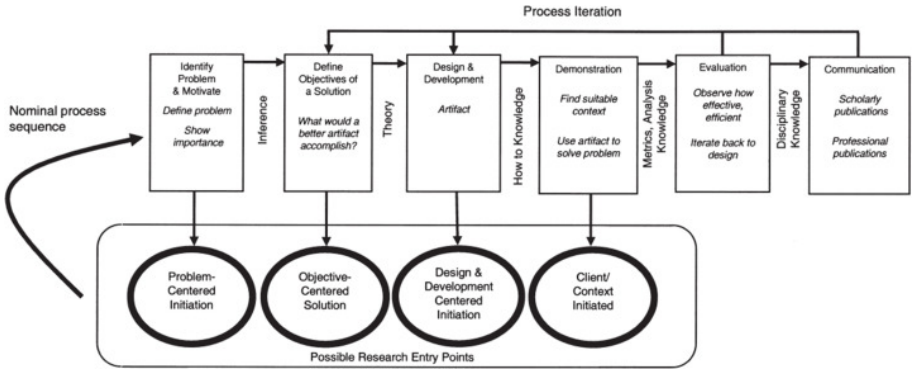

FIGURE 4.1: Peffers' DSRM framework [31]

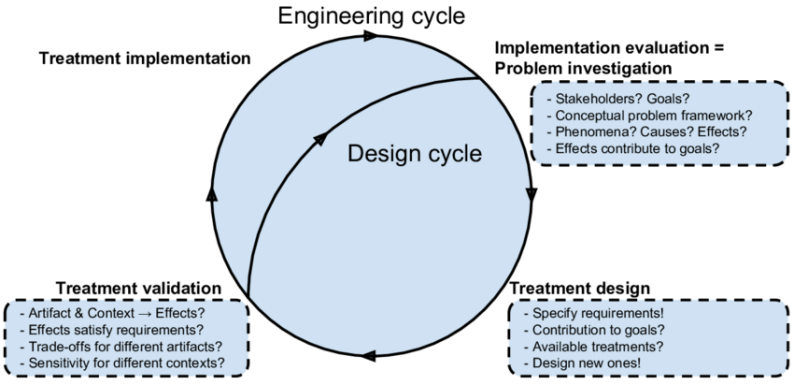

FIGURE 4.2: Wieringa's DSC framework [46]

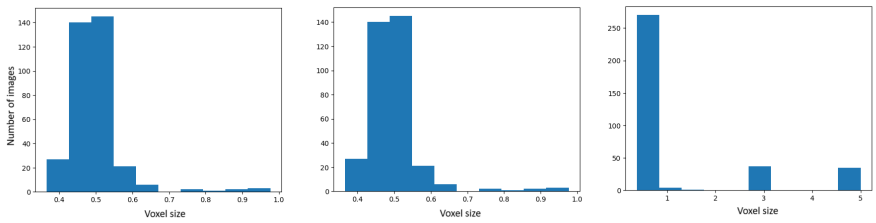

FIGURE 4.3: Voxel shapes of scans in the CQ-500 dataset in coronal, sagittal, and axial direction

To correctly apply the DSRM and DSC, it is important to emphasize the evaluation and communication phase. As the Brain-SHIFT project is located in the medical field, it is crucial to incorporate physicians in this phase. They can provide valuable insight from a clinical point of view. Asking questions like "Does this improve the current protocol?", "What is the level of trust in the deformation field?" and "Which factors of the deformation field do you like and which ones do you not like?", could greatly improve this evaluation phase. It can steer the next iteration in a direction that meets the expectations of the physicians.

4.2 The Dataset

4.2.1 CQ-500

Throughout most of the project, the dataset from MST was not yet available. Hence, the CQ-500 [8] was used for development. This dataset contains 491 3-dimensional CT scans from 3 different types of CT machines and the images are reviewed by physicians. There are scans of patients with healthy brains and several types of hemorrhages and hematomas, including subdural hematomas. The scans are annotated with their type of hematoma and the binary presence of a midline shift larger than 5 mm. These annotations were made by three senior radiologists, and are not always in agreement with each other. 65 Patients in the dataset were labeled with a midline shift and 53 patients were labeled with an SDH, which is 13% and 11% respectively.

The pixel values, image dimensions, and metadata of this dataset were explored. There is a large range of number of slices in the dataset. Most images have between 150 and 200 slices. Images with very few slices have been taken out of consideration, as the resolution of these images is very low. Also, pictures with more than 500 slices have not been considered, for some consistency in image sizes. The images in the dataset have a pixel value range from -3024 until 1692. A majority has a value smaller than 0. These are background pixels. The other values shown showcase different types of brain or head tissues and perhaps other noises present in the image. Lastly, Figure 4.3 shows the number of voxels of all images in the dataset. The voxel sizes in coronal and sagittal directions are identical because all slices of all images in the dataset are squared. As for the slices, most images have a voxel size smaller than 1. These sizes probably differ based on the type of CT machine the images were taken on.

Annotations

For predicting ideal midplanes, the data had to be annotated. This annotation process is done in a 2D manner due to time and medical expertise constraints. Initially, CT images of 50 patients were annotated by the author with their ideal midlines. Only 50 patients were chosen because the dataset is only a pseudo-dataset and the annotation process is very time-consuming. To stay consistent with the dataset as a whole, 13% of these 50 patients' images were labeled with a midline shift and 11% were labeled with an SDH. The rest of the patients were selected randomly. The annotations are done with a MeVisLab script that is depicted in Appendix 8.1. Later on, an additional 20 images were annotated for a test set.

The midline was defined as a straight line passing through the frontal crest and occipital crest of the skull, from an axial view. This method is execuTable on every slice of the brain above the eyes, no matter the state of the grey matter, ensuring consistency and reproducibility. Hence, the annotated slices were all picked to be above the eyes. For each image, 5 slices are annotated:

1. The lowest slice is the first slice in which the orbit of the eyes is no longer visible in the skull.
2. The middle slice would be the slice where the left and right lateral ventricles appear to be the largest.
3. The top slice is the highest slice in which the grey matter is still clearly visible and textured enough to distinguish possibly present hematomas.
4. Slice 4 is placed in between the bottom and the middle slice.
5. Slice 5 is placed in between the top and the middle slice.

This protocol was established jointly with a clinician from MST, so it should be taken into account that this is not a fully professional or globally used method. It is merely to ensure consistency in the labels. The same protocol is followed by another annotator to investigate inter-rater differences. These annotations will be called annotation 1 and annotation 2 respectively.

4.2.2 Data Preprocessing

Before using both datasets to obtain the geometric constraints, the data is preprocessed. This should ensure that all images are cleaned from the head fix of the machine and that the data is in the correct shape and orientation to put into a deep learning model. Further enhancements, such as rotating and cropping the images to center the brain are also implemented. Moreover, some data augmentations have been applied. Each of the following steps is performed in Python 3.8.10.

Train, validation and test set

The 50 annotated images were split 75%, 12.5%, and 12.5% respectively [12]. These all have the same distribution of SDHs and midline shifts. They were split at random. The test set was later extended with 20 more images.

Preprocessing

Both datasets are preprocessed. The images are windowed [30], with a window center of 40 and a window width of 80. This means the Houndsfield Units (HU) of the image pixels are cut off if they are smaller than 0 and larger than 80. These values were chosen because the neurosurgeon from MST worked with these values.

Next, noise from the images is removed. This includes random particles that are present on the CT images, such as removing the head fix of the CT scanner, or parts of skin or ears that are visible in the image. An example is shown in Appendix **??**. The connected areas are turned into masks with morphological dilation and filled in where needed. The largest mask, which is a mask of the head, is kept and the other parts are removed.

To prepare the images for further processing, the brain is centered. To do so, three steps are applied to enhance the brain images. The brains are all set to the same orientation angle, they are placed in the middle of the image and the images are normalized.

First, all brains are given the same orientation. So if a brain is oriented horizontally, it is tilted to be oriented vertically, to ensure all brains are oriented the same. To do so, a mask is created of the brain on a single slice. The angle of this mask is calculated and a rotation matrix is created. This rotation matrix is then applied to all the slices in the image to ensure the full 3D image is tilted as a whole.

After tilting, the brain is not yet centered in the image. Cropping and padding are applied accordingly. The cropping function finds the top left corner and the bottom right corner of the skull in the 3D images and crops to the borders of this. This results in a smaller image. Then, equal padding is added to each border so the image size is $(n_{slices}, 512, 512)$.

Finally, the data was normalized so all values are between 0 and 1.

Label Enhancement

The midline annotations are lines of one-pixel width. These lines are thickened and a Gaussian smoothing function is added to create a gradient from the annotation. For some trainings, only the beginning and end markers of the annotation are taken. This is done by simply taking the minimum and maximum array indices. Also, these are thickened and smoothed.

Data Augmentations

Data augmentations have shown to be very effective when training UNets [9]. The chosen transforms are random flips in the vertical direction. Horizontal flips would change the orientation of the brain and are thus not applied. Some random rotations were also used. Lastly, a small amount of random Gaussian noise was also added to make the network training more robust [9]. Considering each of these transforms is random, their probability is set to $p = 0.5$.

4.3 Geometric Constraints

The geometric constraints are all identified after cleaning and preprocessing the data.

4.3.1 Brain Stripping

Thresholding in combination with the preprocessing steps, is the chosen method to do the brainstripping. This is inspired by the intensity-based brain stripping methods, mentioned in Section 3.1.1. Figure 3.1, showed a gap between a peek of voxel values around 20 HU and

a second peek starting at 80. Combining this with the literature research from section 3.1, a thresholding brainstripping technique seemed fitting. After inspection of thresholding, the machine's head fix seemed to have similar voxel values as the skull. Hence, besides brainstripping, the machine is removed by keeping only the largest component. A before and after of this is shown in Appendix 8.2.

Implementation

The brain stripping is applied as a transform following the preprocessing transforms and before the data augmentation transforms. To do the brainstripping only pixel values above the threshold are kept. In some images, the preprocessing did not remove the machine from the images perfectly. Hence, only the largest component was kept; the skull.

4.3.2 Ideal Midplane

The ideal midplane is estimated as a symmetry measure in a pseudo-healthy brain scan. To estimate the ideal midplane, a U-Net architecture is used. This deep learning method is chosen, rather than e.g. symmetry spacing because it can be more robust against irregular brain shapes or other image artifacts that are likely to occur in real-life practices. Furthermore, in 3D manner, there might be more contextual information in the brain that may be relevant other than symmetry. Before estimating the ideal midplanes, some ideal midlines were predicted, from which a 3D plane was estimated. The reason for this was that it was too time-consuming to annotate brains in 3D manner.

Implementation

First, ideal midlines are predicted from the annotated slices. These predictions are made by training a U-Net for 150 epochs. The model was also validated during training. The UNet architecture uses an Adam [20] and a linear learning rate scheduler [29]. Batch normalization and dropout with $p = 0.2$ were also implemented. A Dice loss [39] was used to compare the output images to the label images. This measures the overlap between two segmented volumes. For perfect alignment, the Dice score is 1, while no overlap gives a score of 0. The loss function of this is $1 - DSC$. It is often used to evaluate image segmentation.

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{4.1}$$

Furthermore, a vanilla UNet++ was implemented. The input was varied, by training based on single slices or by all annotated slices as one input. Also changing the input from the full brain and skull, to only the skull or only the brain was tested. Table 4.1 shows an overview of this ablation study.

The UNets predict in 2D and thus predict a midline. Because marker points are given as a label, also marker points are predicted. To decide which point to take from the predictions, the highest intesity value is used. The predicted markers are trained on the labels that have a gradient, with the highest intensity as the most accurate location. Hence, this method is used. For each patient, five slices were annotated, thus also the midline of five slices is predicted. From these five midline predictions, a plane was estimated to constitute the 3D midplane. This estimation is done based on a singular value decomposition. Given matrix A of dimensions $m \times n$, the singular value decomposition is given:

TABLE 4.1: Overview of the different networks trained

ID	Network	Input	Gradient	Number of slices
1	UNet	Skull	Yes	1
2	UNet	Brain	Yes	1
3	UNet	Skull and brain	Yes	1
4	UNet++	Skull	Yes	1
5	UNet++	Brain	Yes	1
6	UNet++	Skull and brain	Yes	1
7	UNet	Skull	No	1
8	UNet	Brain	No	1
9	UNet	Skull and brain	No	1
10	UNet	Skull	Yes	5
11	UNet	Brain	Yes	5
12	UNet	Skull and brain	Yes	5

$$A = U\Sigma V^T \tag{4.2}$$

In 4.2, U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ diagonal matrix and V^T is the transpose of an $n \times n$ orthogonal matrix.

From this, the normal vector from the plane was identified. The y and z parameters were set to the maximum image shape, so only the x-coordinates are estimated to the best possible fit. This results in a 3D plane. The planes are evaluated as described in the evaluation.

4.3.3 Slice Selection Algorithm

Additionally, a slice selection algorithm is implemented that decides which slices from the CT scan are used. The CQ-500 dataset is annotated for 5 slices that are handpicked based on the protocol. However, to fully automate the process, it is important that also the choosing of these slices is automated. This is done partially before predicting and partially after predicting, as showcased in Figure 4.4. Before predicting, slices are filtered so that only slices above the eye sockets are left. Slices lower than the eye sockets contain a lot of parts of the skull that make it harder to predict the midline. These parts also contain more facial features than brain matter, hence these slices are not relevant. According to facial ratios [11], this starts at the 60th percentile of the face ideally. An extra five percent is taken because not every skull is ideal. Also, the very top slices of the skull are less relevant and contain less information for ideal midlines. Therefore, the 55th until the 85th percentile of the skull is taken from the input image. Appendix 8.3 highlights this part of the skull for clarity.

After predicting the midline, a Structural Similarity Index Measure (SSIM) [44] is performed on the results. By comparing the left side of the image to the right side of the image, it can be determined whether both markers are present in the prediction. If both markers are present and predicted at the right coordinates, the SSIM is larger than 0.9. If one is missing, the similarity index will be lower. When both markers are missing, the predictions also had a low SSIM, so this measure would also work for those cases.

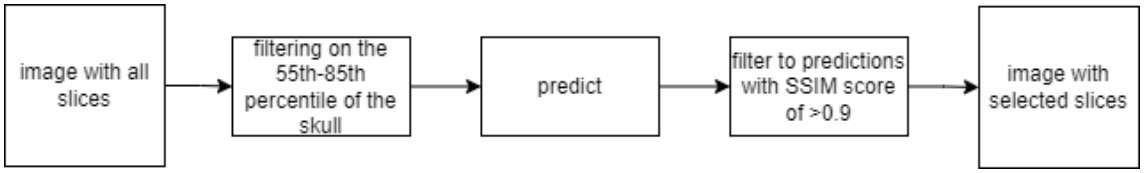

FIGURE 4.4: The slice selection algorithm pipeline

4.4 Evaluation

To evaluate the performance of the final result, relevant metrics should be applied. As the scope of this project focuses on the geometrical constraints, only the metrics to evaluate these are discussed. These metrics should indicate to what extent the output was performed correctly. Thus, whether the deformation field can estimated correctly based on the automatically generated geometric constraints. However, as this network is still in progress, the geometric constraint predictions can only be compared to the available annotations. Considering the lack of medical background, not all these methods can rely on manual checks and thus other metrics should be implemented.

4.4.1 Brain Stripping

For brain stripping, there are no ground truths available. To inspect the resulting images, the presence of machinery or brain tissues other than the skull is checked manually.

4.4.2 Ideal Midplane

The evaluation of the ideal midplane consists of 2D evaluations of midlines as well as 3D evaluations of the midplanes.

2D Midline Evalution

The midline evaluation compares the predicted midlines to the ground truth midlines. When comparing these, the relative position and angle are of most importance. For this, the angle will be calculated between the two lines in the middle of the skull. This angle difference is set relative to the image size. A small angle difference on a small image can be a good prediction, yet a small angle difference on a larger image can result in a very off prediction. The distance between both lines is also calculated between both lines. The distance is measured at several places inside the skull and the largest distance is taken as the distance measure.

3D Midplane Evalution

The constructed midplane from the labels is compared to the constructed midplane from the predicted lines. These planes are also compared by angle and distance, the same way as the 2D manner. Furthermore, the distance between the midlines and the estimated planes is calculated for each point, and the average of these will be taken to find out to what extent the estimated plane resembles the ground truth.

Chapter 5

Results

In this section, the results of the geometric constraints are showcased. They are discussed later in Chapter 6.

5.1 Brain Stripping

Figure 5.1 shows the brain stripped images of four patients. Of each patient, four slices are taken each at a different level. These levels are custom for visualization purposes. The first level still contains eyesockets and other facial features. The second level shows the cavity in the anterior part of the skull. The last two levels are "normal" levels, that are also targeted during the annotation process. For each level, it can be deduced that only the brain is present and no other noise or machinery is visible. For the patient on the far right side, it is noticeable on the second level that the front of the brain is slightly cropped at the anterior side. Appendix 8.4 also shows this. This is because the images in Figure 5.1 are after preprocessing. The original image of the CQ-500 dataset does not have the head centered and is partially cropped out of the image. The patient was likely not centered when taking the scan.

5.2 Ideal Midlines and Midplanes

The ideal midplanes are estimated from the ideal midlines that are predicted by the network. Therefore, the following results are split into three parts. First, the ideal midline detection results are given and then the plane estimation results are given. Lastly, the results of the ablation study are given. All the results showcased before the ablation study, are based on the network that performed the best.

5.2.1 2D Midline Detection

Figure 5.2 shows the input image, the label, an overlay of the prediction and the label, and the raw prediction. In the far right image, it can be seen that there are two spots. These are the predicted marker points. These spots have different intensities. The darkest spot is taken as the predicted marker coordinates. In the middle right image, this darkest spot is indicated in red. A straight line between these red markers is what makes the ideal midline prediction. In this same image, there are white spots visible, which are from the labels. The colored parts belong to the prediction. It can be seen that the prediction matches the label almost fully, except for the darker blue spots on the outside.

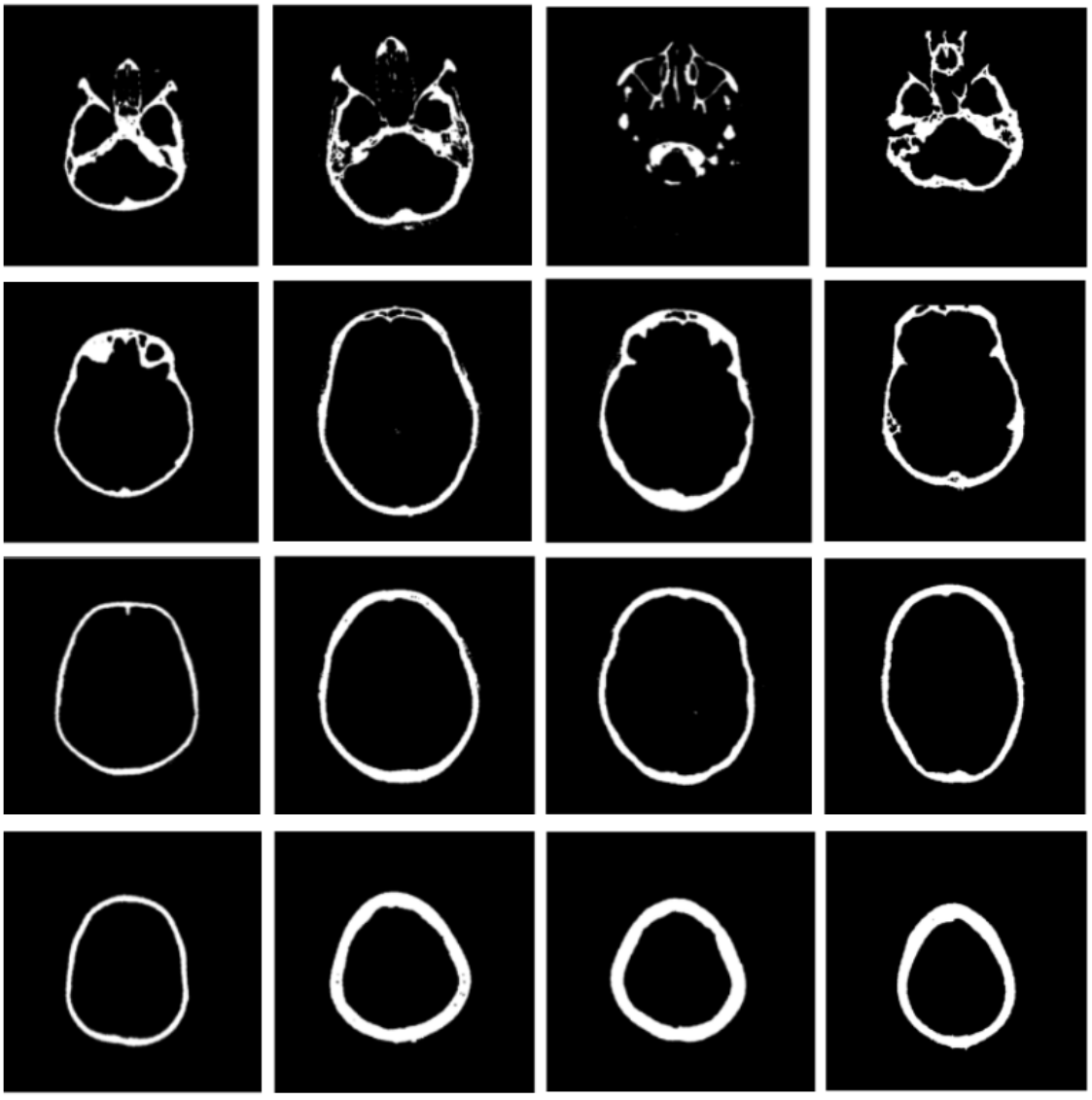

FIGURE 5.1: Four slices at custom-picked levels of four patients' brain stripped image. Each column is one patient and each row is a different level. The first level still contains eyesockets and other facial features. The second level shows the cavity in the anterior part of the skull. The last two levels are "normal" levels, which are also targeted during the annotation process.

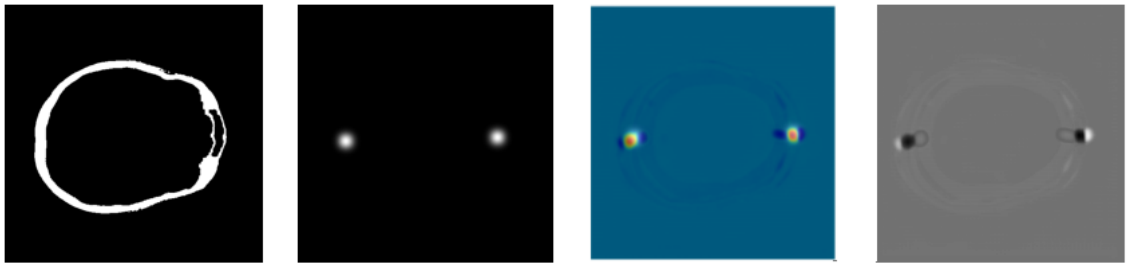

FIGURE 5.2: Results of UNet trained with markers. From left to right: The input image, the ground truth markers, and overlay of the grount truth and the prediction, and the raw predicted markers

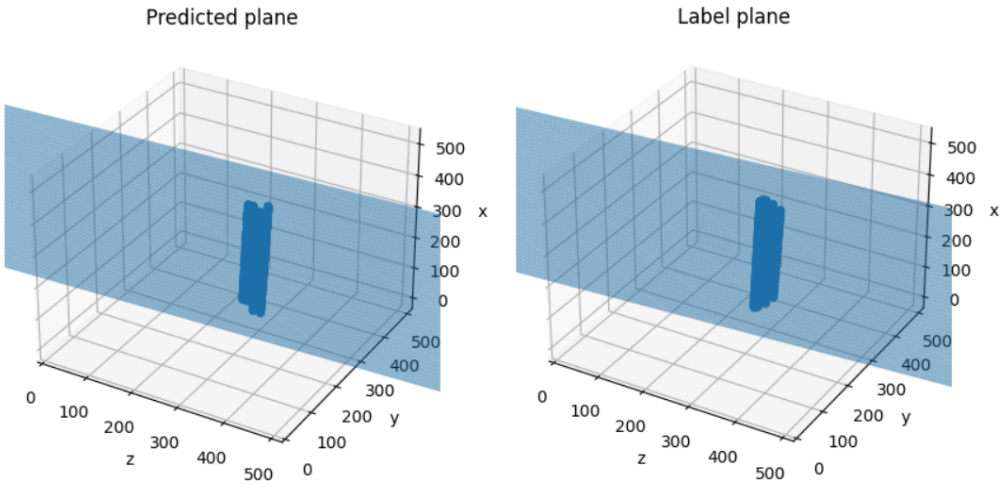

FIGURE 5.3: Midplane estimations from the predicted and annotated midlines

5.2.2 Plane Estimation from Midlines

From the midlines, a midplane is estimated. For each patient, a plane is fitted to the 5 lines. This is done both for the annotated and predicted midlines. Figure 5.3 shows such an estimation. The blue lines are the midlines and the rectangle is the estimated midplane. When comparing the predictions to the labels, it can be seen that the x, y, and z orientations resemble similar values. When looking more closely, it can be seen that the length of the predicted midlines is slightly different than the label midlines. Especially for a larger z value, thus higher slices, the predicted midlines seem larger. The shading of the rectangles, which resembles the angle, also differs slightly.

Figure 5.4 shows an estimated plane from the top view. This view shows that each point of the midlines has a different distance from the plane, although this distance is still very small. The average distance of the lines to the plane in Figure 5.4 is 0.35 mm. Table 5.1 shows the average distance to the estimated planes from the ground truths of annotation 1 and annotation 2.

Distance to plane 1	**Distance to plane 2**
0.42	0.44

TABLE 5.1: The average distances (mm) to the planes of annotation 1 and annotation 2

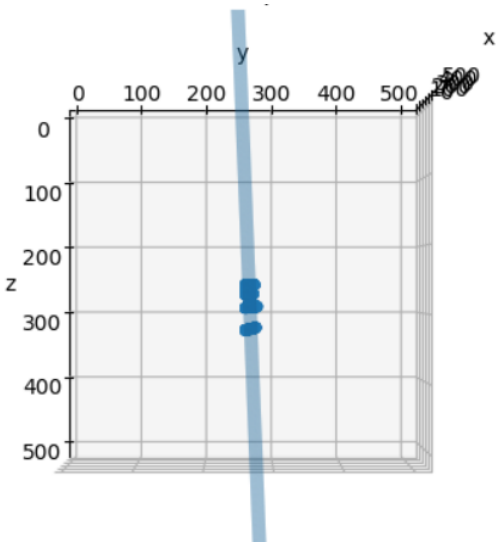

FIGURE 5.4: Top view of the midplane estimation. This shows that the plane does not perfectly match each point in the midline predictions.

5.2.3 Ablation study

Detecting the midlines and estimating the midplanes accordingly has been done for multiple models. These are compared to each other to find out which network and types of input combinations optimize best.

From the ablation study, all of the models in Table 4.1 that contained a single slice as input did not converge during training. Table 5.2 shows the final validation loss after 150 epochs. None of these losses reach a lower value than 0.9. The UNet++ models (ID = 4, 5, and 6) all reached a slightly lower loss than the others.

ID	1	2	3	4	5	6	7	8	9
Dice loss	0.99	0.99	0.99	0.93	0.91	0.97	0.99	0.98	0.98

TABLE 5.2: Final validation loss of UNet ablation study

The other three models did converge and have been evaluated further. These are networks with IDs 10, 11, and 12. They all have an input of five slices and only differ in their input images. From Table 5.3 it can be deduced that all angles are rather small. Most are less than 3 degrees. The angle between the predictions and the symmetry line is smaller than the other angles, all of these are smaller than 1.7 degrees.

What also stands out from Table 5.3 is that the distance is much smaller for the model that was trained with skull images. The distance of these predictions is only 2 mm. However, the distance to the symmetry line is slightly larger than the distance to the annotations. For models 11 and 12 however, the distance ranges from 22 mm to 25 mm. This is more than ten times as large as the distances from model 10.

Figure 5.5 shows the results of the three different models visually. The first row shows the input images. The lines in all other images display the estimated plane in 2D. This is not the predicted midline, but the 3D plane visualized in one slice. For each other image, the red line showcases the predicted plane. The blue lines show the plane of annotation 1, the green lines show the plane of annotation 2, and lastly, the yellow line is the symmetry line. Here, it is clear that the model trained with skull images can estimate the ideal

		Annotation 1		Annotation 2		Symmetry line	
ID	**Input**	Angle	Distance	Angle	Distance	Angle	Distance
10	**skull**	3.07	2.68	2.89	2.12	1.32	5.03
11	**brain**	3.64	24.80	3.16	25.00	1.66	23.71
12	**head**	2.95	23.87	2.69	24.07	1.52	22.61

TABLE 5.3: These are models with IDs 10, 11, and 12. They all have an input of five slices and only differ in their input images. The predictions are compared to annotations 1, annotations 2, and a symmetry line. The angle (degrees) and distance (mm) are the average values of the test set.

	Predictions	**Annotations 2**	**Symmetry line**
Angle	3.07	1.11	2.64
Distance	2.68	1.42	3.67

TABLE 5.4: Annotation 1 compared to predictions of the best model, annotation 2, and the symmetry line. The angle (degrees) and distance (mm) are the average values of the test set.

midplane the closest to the ground truth. This is the only column where the two lines cross. For models 11 and 12 there is a clear distance between the prediction and ground truths.

All in all, it can be deduced that model 10, where images with only the skull are used as an input, performs the best.

Notably, the yellow symmetry line does divide the image of the head in half quite neatly. It goes through the frontal crest of the skull and nearly goes through the occipital crest. This matches the annotation protocol.

This leads to comparing the annotations to the other ground truths to assess how well the annotations are done. Table 5.4 shows the angles and distances between annotation 1 and the other ground truths. Also, the annotations are compared to the predictions again, for a clear overview, but this is the same as in Table 5.3. In Table 5.4 it can be seen that the angle and distance are smallest when comparing annotation 1 to annotation 2. There is almost only a one-degree angle difference and a distance of 1,4 mm. Furthermore, it can be seen that the angle difference between annotation 1 and the prediction is larger than the angle difference between annotation 1 and the symmetry line. However, there is only a very small difference between the two, namely 0.43 degrees. The distance to the symmetry line is the largest, whereas the distance to annotation 2 is the smallest. Appendix 8.5 also showcases these results visually.

5.3 Slice Selection Algorithm

	Only labeled slices	**With SSA**	**Without SSA**
Angle	3.07	3.11	3.81
Distance	2.68	2.47	15.04

TABLE 5.5: Average angle (degrees) and distance (mm) of the test set when running with and without the slice selection algorithm (SSA) for model 10.

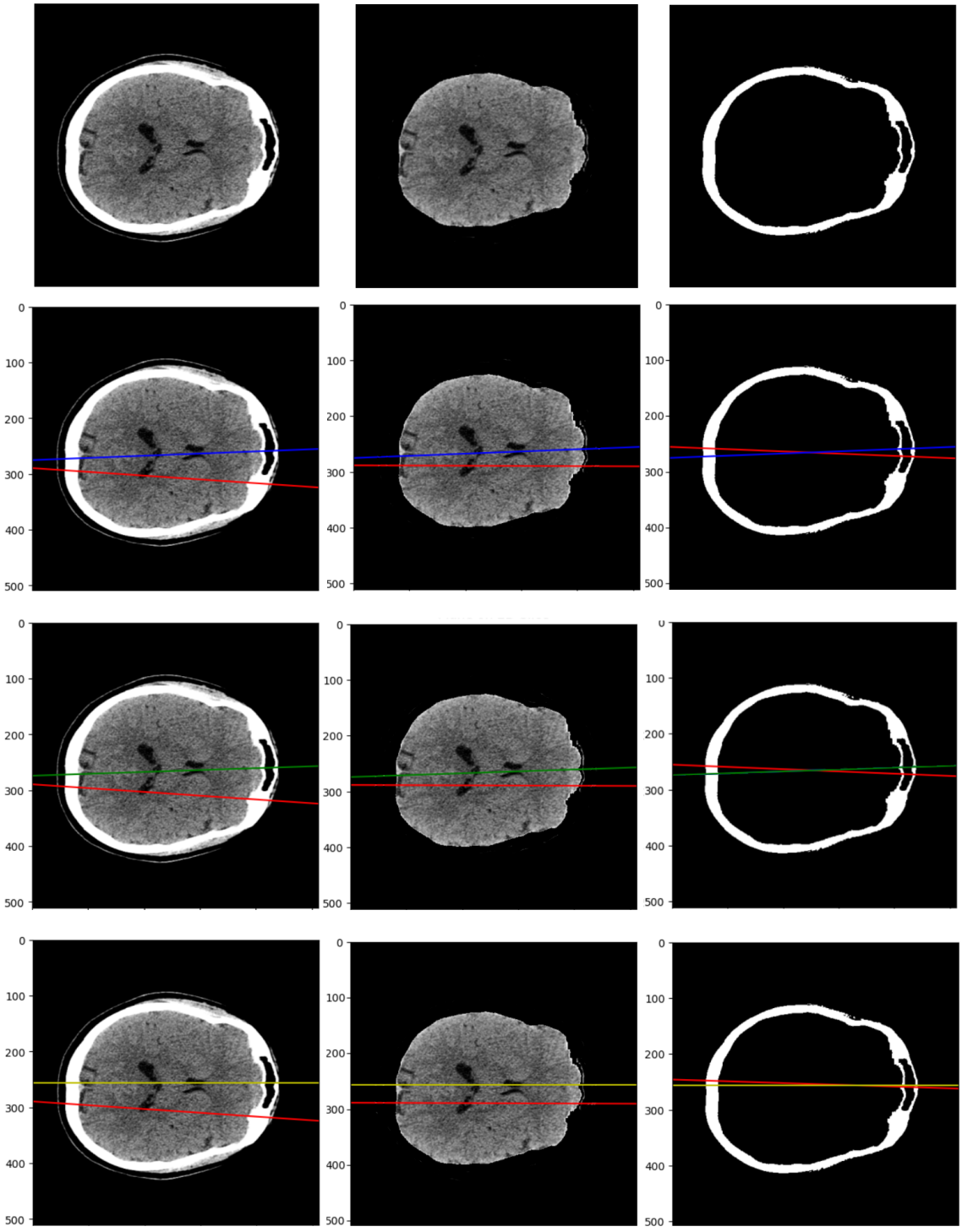

FIGURE 5.5: The visual results of the ideal midplane estimations of models 10, 11, and 12. Each column is one model, so from left to right, it is model 12. 11, and 10. Each row shows the prediction (red) compared to a different ground truth. The second row is compared to annotation 1 (blue). The third row is compared to annotation 2 (green). The bottom row is compared to the symmetry line (yellow). The right column is from the model that performed the best overall.

The implementation of the slice selection algorithm is tested for model 10. When using the model for an entire brain, 512 slices are given to the network. After applying the slice selection algorithm to the test set, on average only 20 good slices were left. Table 5.5 shows the results with and without the slice selection algorithm. Again the baseline results are given for easier comparisons. As can be deduced, the angle improves with 0.8 degrees. The distance decreases by 12.57 mm.

Chapter 6

Discussion

This section discusses the results from chapter 5. These results are analyzed and evaluated. Furthermore, the limitations of the findings are addressed.

6.1 Brain Stripping

The threshold brain stripping showed promising results to delineate the skull as a geometric constraint, which can be used as an input for finding the deformation field. The method resulted in clean extractions at various levels in the skull, as shown in Figure 5.1. All other components, such as the brain, skin, ears, and noise, were filtered out after applying the brain stripping method.

Sometimes the machine or calcification was still present after brain stripping because these artifacts have similar voxel values to the skull. However, keeping only the largest component successfully removes these artifacts. While this does remove unwanted artifacts, there might be a possibility that it also removes parts of the skull that are not perfectly connected to the skull. It is important to ensure no essential parts are lost when removing. In Appendix 8.4, it can be seen that a part of the skull where the eyesockets are is removed. While this is not relevant as a constraint for the removal of the hematoma, it does indicate the possibility that parts can be removed that should not be removed.

Another limitation is the noticeable cropping that was reflected in the right side of Appendix 8.4, where the frontal part of the skull was cropped. This is likely due to the patient not being positioned correctly when taking the CT scan. This showcases that this method only brain strips existing data and does not enhance the resulting skull. However, when applying the deformation field, in the pseudo-healthy image, the skull should not be cropped. Considering the skull constraint is there to also ensure brain matter does not move beyond the skull, this particular case would not affect the final results negatively. Because the inside of the skull remained intact and the brain was removed correctly.

6.2 Midlines and Midplanes

6.2.1 2D Midline Detection

Figure 5.2 showed results of raw marker predictions. The predicted markers are positioned in the same spot as the ground truth markers, showing the model can locate the frontal and occipital crest. However, there is a lot of blur around the markers. This blur is on the left and right side of the prediction so in the direction of the midline. Considering the marker labels are taken from midline annotations, not all markers may be exactly at a consistent

position in the crest. This could cause the model uncertainty about the exact location. An improvement could thus be to use more accurate labels, such as marker points from actual marker annotations, instead of from midlines. However, as long as the predicted ideal midline reaches the skull, a slight difference in the length of the midline does not affect a good midline prediction.

6.2.2 Plane Estimation from Midlines

Based on five predicted midlines per image, a midplane was estimated. Figure 5.3 shows an estimation of a midplane from the annotated midlines and the predicted midlines. The midplanes are similar in orientation along the x, y, and z coordinates. However, the midline predictions differ in length from the annotated midlines. Although this might not have a large effect in 2D, because a longer midline is not necessarily wrong, it might affect the estimation of the midplane. It suggests that the 2D model is overestimating the midlines. Nevertheless, this is an aspect that can be tweaked in post-processing for better estimations of the midplane.

Furthermore, in Figure 5.4 the distance of the midlines to the estimated planes is evaluated. This consistently led to distances smaller than 0.5 mm. This is promising for using the model in a clinical environment. It was also noticed that larger deviations in the distance also tend to have larger inter-rater errors. This suggests that some brains are harder to annotate systematically. This emphasizes the subjectivity of the annotations, which the model fits to. Using more annotations for the training could resolve this problem. Especially annotations from clinical practitioners would be beneficial.

6.2.3 Ablation study

The ablation study showed that models 1 to 9 did not converge in validation loss. All these models were trained with a single slice, implying that this is not sufficient. There might not be enough context available in the input images, to predict the marker points. However, it might be interesting to investigate a different type of network for single slices. Instead of reconstructing the full image, finding only the coordinates of the marker points with a convolutional network could lead to better results. There was also a slight improvement in the dice loss in the UNet++. This indicates that with further hyperparameter optimization, this network might be able to find the marker points.

From the models that were trained with all five slices as input, the model with the skull input images performed the best. The main reason why this performed better, is because it can predict the midlines with a small distance to the ground truth, as was indicated in Table 5.3. Figure 5.5 also shows that the predicted lines match the ground truths better for the skull images than the others. When looking at the predicted line in red, it can be deduced that it does go through the frontal and occipital crest, but not exactly through the thickest parts. This could be because Figure 5.5 shows the plane estimations, so although it might not be a perfect line for this slice, it could be an overall correct estimation. It could also be caused by the way the marker points are extracted from the raw markers that were shown in Figure 5.2. For example, instead of taking the highest intensity, finding a mean value could improve results.

When investigating the other models' performances in Table 5.3 and Figure 5.5, it is clear that these predictions are not good enough to be taken as an ideal midplane. For model 11 with only the brain as an input, there is a rather good angle estimation, yet the distance to the ground truth is too large. The prediction does not resemble a good symmetry of the brain. For model 12, with the full head as an input, both the angle and

the distances are off. In Table 5.3 these angles and distances are a bit smaller than for model 11, but the distances are still too large to resemble a good midplane. Especially since the clinical threshold for a midline shift is 5 mm, having a distance of 25 mm would truly be too large of a difference to be clinically relevant for the ideal midplane.

Table 5.4 shows a comparison of annotation 1 to the other ground truths. There is a very small angle and distance difference between annotation 1 and annotation 2. It makes sense since the same protocol was used when retrieving these annotations. The small deviation does show that the protocol is not a golden standard and that it still deviates per person. It also confirms the clinical problem, where different annotators create different annotations.

Furthermore, there is an angle between annotation 1 and the symmetry line. The symmetry line could be used as an ideal midplane, given that the preprocessing corrects the tilt of the brain images correctly. Table 5.4 shows both this angle and the distance are larger than the angle and distance between both annotations. The angle might be somewhat correct since the head is rotated in preprocessing. However, using the symmetry line as an ideal midplane likely leads to larger distances, because the symmetry of a brain in a CT image could be tilted. Even a tilt of a few millimeters would already lead to a larger distance. The distance of the prediction is smaller because it is based on the actual CT images and not on the idea that a brain is symmetrical. That being said, the angle and distance of the symmetry lines are better than the performance of models 12 and 11. This indicates that investigating whether there is room for optimization in the tilt correction could also be a good way to find the ideal midline.

All in all, the ablation study shows that the models trained in five slices resulted in better midplane predictions. It suggests that annotating more slices per patient and training models fully in 3D could improve the results even more. Moreover, the markers of the midlines were predicted well, but there was still some blur that created uncertainty. Therefore, training a network to find coordinates, would also be relevant to future work. It should also be noted that the provided findings are all based on a model trained with annotations that are made only jointly with a clinician. To ensure clinical correctness, a model trained on annotations made with medical expertise would be crucial.

6.2.4 Slice Selection Algorithm

When applying the Brain-SHIFT project in practice, the deformation field should be found fully automatically. Clinicians should not have to handpick slices beforehand. Therefore, also the geometric constraints, especially the midplane, should not rely on handpicked slices. The slice selection algorithm should ensure this.

The slice selection algorithm showed to improve the results, as seen in Table 5.5. When applying the slice selection algorithm, there is a very slight increase in angle compared to using the five handpicked slices, but this difference is negligible. There is a slight improvement in distance to the ground truth. The slice selection algorithm keeps only the very good predictions, meaning that it removes outliers. When estimating the plane, it makes sense that the distance to the plane decreases if only the good midlines are kept. A smaller distance to the plane ensures more consistency in the predictions, likely indicating the predictions are also better. Hence, the smaller distance after applying the slice selection algorithm.

There is a much larger difference between the predictions without a slice selection algorithm and with the slice selection algorithm. The model is trained for the slices that are labeled, which are handpicked according to a protocol. These are five good slices. However, a full image has 512 slices, of which some only contain background and others

contain facial features that the model is not trained for. It also leads to wrong predictions, as is indicated in Table 5.5. When applying the slice selection, the results are similar to the results with handpicked slices. This proves that the slice selection algorithm is a good solution for picking slices from a full image. This is especially important for the clinical implementation.

6.3 Overall Clinical Relevance

The previous sections discussed the results of the geometric constraints. However, they should still be placed into the broader picture. The research question asked to what extent the geometric constraints could be found using deep learning for finding the deformation field. First of all, the intensity-based brain-stripping method showed that the skull was obtained. The skull was obtained without leaving traces of the brain or other noises. However, there was no deep learning used for this, as there were no ground truths available. The midplane was shown to be found with an angle difference of 3.07 degrees to the ground truth and a distance of 2.68 mm. These are better estimations of midplanes than using the symmetry axis. However, when comparing two annotations, a smaller angle and distance are found, so the predictions do not outperform human inter-rater differences. It does however provide a midplane that matches the actual midplane rather well. This could still benefit clinicians, as it provided a 3D view, rather than a 2D view of a single slice. The slice selection algorithm also took the clinical use of the findings into account, by ensuring that only relevant slices were used for the plane estimation.

To know how well the retrieved skull and midplane can aid the finding of the deformation field, they would have to be tested in the full pipeline of the Brain-SHIFT project. As this is still a work in progress at the time of writing, that is not possible. It would give valuable insight, nonetheless. However, all constraints were found automatically and accurately and therefore meet the expectations of geometric constraints for the deformation field network.

The clinical value of the Brain-SHIFT process was also assessed. Given that the deformation field can be found correctly, the project relieves pains in the currently used protocol and implements gains for clinicians. It streamlines the current protocols and aids the standardization of the current procedures. It enables more informed decision-making in the diagnosis and treatment of SDHs. It also has the potential to be expanded to other masses in the head if trained for a wider range of masses.

Chapter 7

Conclusion

This report showcased the development and evaluation of a deep learning-based approach to identify geometric constraints within brain CT images, which are crucial for estimating the deformation field in patients with a cSDH. A method for automatically detecting the skull and ideal midplanes was developed, optimized, and evaluated. The results show that the methods developed can successfully identify these geometric constraints. The brain stripping method resulted in clean skull masks. The midplane detection resulted in 3.07 degree difference from the ground truth. The results are promising to use as an input for the deformation field calculation.

Furthermore, the clinical value of the Brain-SHIFT project was investigated. From a SWOT analysis, risk analysis and value proposition, it can be concluded that the project is substantial, as it streamlines current protocols, aids in standardizing procedures, and enables more informed decision-making in the diagnosis and treatment of SDHs. The potential for this tool extends beyond cSDH to other masses in the head, should it be trained for a wider range of pathologies.

For future research, several recommendations can be made. Firstly, further exploration of networks trained with more slices as input could improve the prediction of marker points. Secondly, involvement of clinical practitioners in the annotation process could ensure medical accuracy.

Bibliography

[1] Aug 2021. URL: https://www.nhs.uk/conditions/subdural-haematoma/.

[2] Parvez Ahmad, Hai Jin, Roobaea Alroobaea, Saqib Qamar, Ran Zheng, Fady Alnajjar, and Fathia Aboudi. Mh unet: A multi-scale hierarchical based architecture for medical image segmentation. *IEEE Access*, 9:148384–148408, 2021. doi:10.1109/ACCESS.2021.3122543.

[3] Khalid Alhusseiny. Midline shift: Radiology reference article, Apr 2023. URL: https://radiopaedia.org/articles/midline-shift.

[4] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net, Nov 2022. URL: https://arxiv.org/abs/2211.14830.

[5] Andre Balan, Agma Traina, Marcela Ribeiro, Paulo Marques, and Caetano Traina Jr. Smart histogram analysis applied to the skull-stripping problem in t1-weighted mri, Feb 2012. URL: https://www.sciencedirect.com/science/article/pii/S0010482512000157.

[6] M E Brummer, R M Mersereau, R L Eisner, and R J Lewine. Automatic detection of brain contours in mri data sets, 1993. URL: https://pubmed.ncbi.nlm.nih.gov/18218403/.

[7] M Chen, A Elazab, F Jia, and J Wu. Automatic estimation of midline shift in patients with cerebral glioma ..., Aug 2015. URL: https://www.researchgate.net/publication/281363395_Automatic_estimation_of_midline_shift_in_patients_with_cerebral_glioma_based_on_enhanced_voigt_model_and_local_symmetry.

[8] Sasank Chilamkurthy, Rohit Ghosh, Swetha Tanamala, Mustafa Biviji, Norbert G. Campeau, Vasantha Kumar Venugopal, Vidur Mahajan, Pooja Rao, and Prashant Warier. Development and validation of deep learning algorithms for detection of critical findings in head ct scans, Apr 2018. URL: https://arxiv.org/abs/1803.05854.

[9] P Chlap, H Min, N Vandenberg, J Dowling, L Holloway, and A Haworth. A review of medical image data augmentation techniques for deep learning applications, Jun 2021. URL: https://pubmed.ncbi.nlm.nih.gov/34145766/.

[10] RW Cox. Afni: Software for analysis and visualization of functional magnetic resonance neuroimages, Jun 1996. URL: https://pubmed.ncbi.nlm.nih.gov/8812068/.

[11] Mohammad Mahdi Dehshibi and Azam Bastanfard. Unsupervised feature based facial family similarity recognition. 04 2011.

[12] Rosie Dunford, Quanrong Su, and Ekraj Tamang. The pareto principle. *The Race*, 2021. URL: https://api.semanticscholar.org/CorpusID:15925174.

[13] P Gruen. Surgical management of head trauma, May 2002. URL: https://pubmed.ncbi.nlm.nih.gov/12391640/.

[14] Jenny Gu and Ryota Kanai. What contributes to individual differences in brain structure?, Apr 2014. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4009419/.

[15] Akifumi Hagiwara, Shohei Fujita, Yoshiharu Ohno, and Shigeki Aoki. Variability and standardization of quantitative imaging: Monoparametric to multiparametric quantification, radiomics, and artificial intelligence, Sep 2020. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7413678/.

[16] Shady Hermena and Michael Yound. Ct-scan image production procedures. URL: https://pubmed.ncbi.nlm.nih.gov/34662062/.

[17] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation, Apr 2020. URL: https://arxiv.org/abs/2004.08790.

[18] eSAT Journals. 3d segmentation of glioma from brain mr images using seeded region growing and fuzzy c-means clustering, Feb 2016. URL: https://www.academia.edu/21613706/3D_SEGMENTATION_OF_GLIOMA_FROM_BRAIN_MR_IMAGES_USING_SEEDED_REGION_GROWING_AND_FUZZY_C_MEANS_CLUSTERING.

[19] P Kalavathi and V B Surya Prasath. Methods on skull stripping of mri head scan images-a review, Jun 2016. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4879034/.

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. arXiv:1412.6980.

[21] Yong En Kok, Stefan Pszczolkowski, Zhe Kang Law, Azlinawati Ali, Kailash Krishnan, Philip M. Bath, Nikola Sprigg, Robert A. Dineen, and Andrew P. French. Semantic segmentation of spontaneous intracerebral hemorrhage, intraventricular hemorrhage, and associated edema on ct images using deep learning. *Radiology: Artificial Intelligence*, 4(6):e220096, 2022. PMID: 36523645. arXiv:https://doi.org/10.1148/ryai.220096, doi:10.1148/ryai.220096.

[22] M Kumar Nag, A Gupta, A S Hariharasudhan, A Kumar Sadhu, A Das, and N Ghosh. Quantitative analysis of brain herniation from non-contrast ct images using deep learning, Dec 2020. URL: https://pubmed.ncbi.nlm.nih.gov/33316319/.

[23] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets, 2014. arXiv:1409.5185.

[24] C-C Liao, Y-F Chen, and F Xiao. Brain midline shift measurement and its automation: A review of ..., Apr 2018. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5925103/.

[25] CC Liao, F Xiao, JM Wong, and IJ Chiang. Automatic recognition of midline shift on brain ct images, Feb 2010. URL: https://pubmed.ncbi.nlm.nih.gov/20132930/.

[26] R Liu, S Li, B Su, C L Tan, TY Leong, B C Pang, CC T Lim, and C K Lee. Automatic detection and quantification of brain midline shift using anatomical marker model, Nov 2013. URL: https://pubmed.ncbi.nlm.nih.gov/24332442/.

[27] Giancarlo Mignucci-Jiménez, Alejandro J Matos-Cruz, Irakliy Abramov, Sahin Hanalioglu, Melissa S Kovacs, Mark C Preul, and Caleb E Feliciano-Valls. Puerto rico recurrence scale: Predicting chronic subdural hematoma recurrence risk after initial surgical drainage, Jun 2022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9282733/.

[28] JD Miller and R Nader. Acute subdural hematoma from bridging vein rupture: A potential mechanism for growth, Jun 2014. URL: https://pubmed.ncbi.nlm.nih.gov/24313607/.

[29] Koyel Mukherjee, Alind Khare, and Ashish Verma. A simple dynamic learning rate tuning algorithm for automated training of dnns, 2019. arXiv:1910.11605.

[30] Andrew Murphy. Windowing (ct): Radiology reference article, Mar 2023. URL: https://radiopaedia.org/articles/windowing-ct.

[31] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24:45–77, 01 2007.

[32] Fakhrul Razan Rahmad, Wan Nurshazwani Wan Zakaria, Ain Nazari, Mohd Razali Md Tomari, Nik Farhan Nik Fuad, and Anis Azwani Muhd Suberi. Hybrid skull stripping method for brain ct images, Jan 1970. URL: https://link.springer.com/chapter/10.1007/978-981-15-5281-6_44.

[33] M Rauhala, TM Luoto, H Huhtala, G Iverson, T Niskakangas, J Öhman, and P Helén. The incidence of chronic subdural hematomas from 1990 to 2015 in a defined finnish population, Mar 2019. URL: https://pubmed.ncbi.nlm.nih.gov/30901751/.

[34] Minna Rauhala, Pauli Helén, Heini Huhtala, Paula Heikkilä, Grant L Iverson, Tero Niskakangas, Juha Öhman, and Teemu M Luoto. Chronic subdural hematoma-incidence, complications, and financial impact, Sep 2020. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7415035/#:~:text=The%20incidence%20of%20CSDH%20among,100%2C000%2Fyear)%20treated%20cases.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, May 2015. URL: https://arxiv.org/abs/1505.04597.

[36] K J Shanthi and M Sasi Kumar, 2007. URL: https://www.semanticscholar.org/paper/Skull-stripping-and-automatic-segmentation-of-brain-Shanthi-Kumar/ff4786bd1d3733ae0efc858574dd7f1d82fca89c.

[37] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. doi:10.1109/ACCESS.2021.3086020.

[38] Jehuda Soleman, Philipp Taussky, Javier Fandino, and Carl Muroi. Evidence-based treatment of chronic subdural hematoma. In Farid Sadaka, editor, *Traumatic Brain Injury*, chapter 12. IntechOpen, Rijeka, 2014. doi:10.5772/57336.

[39] T. Sørenson. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. Biologiske skrifter. I kommission hos E. Munksgaard, 1948. URL: https://books.google.nl/books?id=rpS8GAAACAAJ.

[40] RF Stewart, OJ Benepe, and A Mitchell. Formal planning: the staff planner's role at start up (no. 250). *California: Stanford Research Institute*, 1965.

[41] Joanna Swiebocka-Wiek. Skull stripping for mri images using morphological operators, Jan 1970. URL: https://link.springer.com/chapter/10.1007/978-3-319-45378-1_16.

[42] G. Tong, X. Wang, H. Jiang, A. Wu, W. Cheng, X. Cui, L. Bao, R. Cai, and W. Cai. A deep learning model for automatic segmentation of intraparenchymal and intraventricular hemorrhage for catheter puncture path planning. *IEEE journal of biomedical and health informatics*, 27(9):4454–4465, 2023. doi:10.1109/JBHI.2023.3285809.

[43] Bin Wan, Şeyma Bayrak, Ting Xu, H Lina Schaare, Richard AI Bethlehem, Boris C Bernhardt, and Sofie L Valk. Heritability and cross-species comparisons of human cortical functional organization asymmetry. *eLife*, 11:e77215, jul 2022. doi:10.7554/eLife.77215.

[44] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi:10.1109/TIP.2003.819861.

[45] B D Ward. Afni program: 3dintracranial, Jun 1999. URL: https://afni.nimh.nih.gov/afni/doc/help/3dIntracranial.html.

[46] Roel J. Wieringa. Design science methodology for information systems and software engineering, 2014. URL: https://link.springer.com/book/10.1007/978-3-662-43839-8.

[47] D Wu, H Li, J Chang, C Qin, Y Chen, Y Liu, Q Zhang, B Huang, M Feng, R Wang, and et al. Automatic brain midline surface delineation on 3d ct images with intracranial hemorrhage, Sep 2022. URL: https://pubmed.ncbi.nlm.nih.gov/35298377/.

[48] Yad R Yadav, Vijay Parihar, Hemant Namdev, and Jitin Bajaj. Chronic subdural hematoma, 2016. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4974954/.

[49] Heming Yao, Craig Williamson, Jonathan Gryak, and Kayvan Najarian. Automated hematoma segmentation and outcome prediction for patients with traumatic brain injury. *Artificial Intelligence in Medicine*, 107:101910, 2020. URL: https://www.sciencedirect.com/science/article/pii/S093336571931317X, doi:10.1016/j.artmed.2020.101910.

[50] Xiao-Xia Yin, Le Sun, Yuhan Fu, Ruiliang Lu, and Yanchun Zhang. U-net-based medical image segmentation, Apr 2022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9033381/.

[51] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. Motion artifacts in mri: A complex problem with many partial solutions, Oct 2015. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4517972/.

[52] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. arXiv:1807.10165.

Chapter 8

Appendix

8.1 Annotation Tool

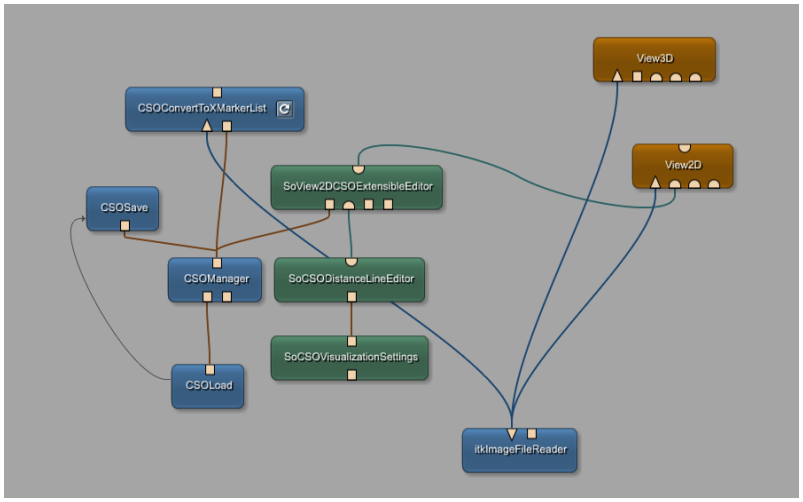

FIGURE 8.1: MeVisLab network for inspecting and annotation brain CT scans

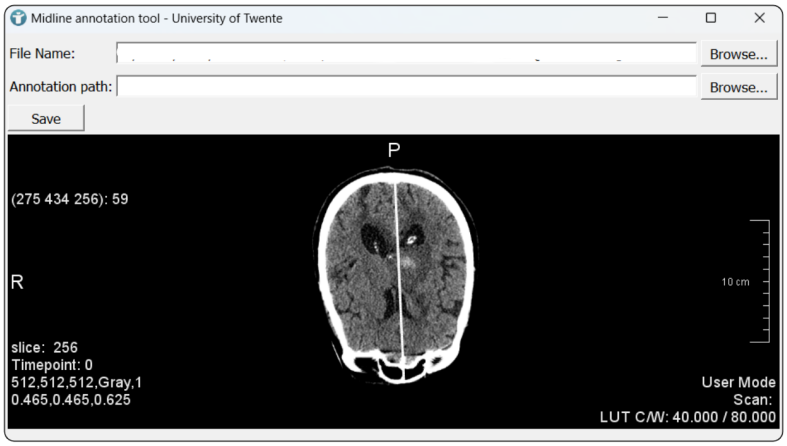

FIGURE 8.2: Annotation tool for annotation midlines on CT slices

8.2 Removing Noise

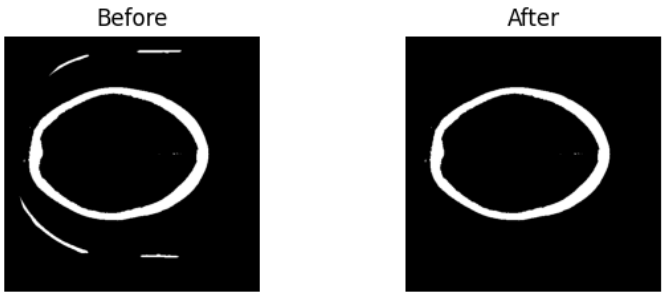

FIGURE 8.3: Before and after keeping the largest component. The left side shows the head fix of the CT machine, which is removed on the right side.

8.3 Slice Selection

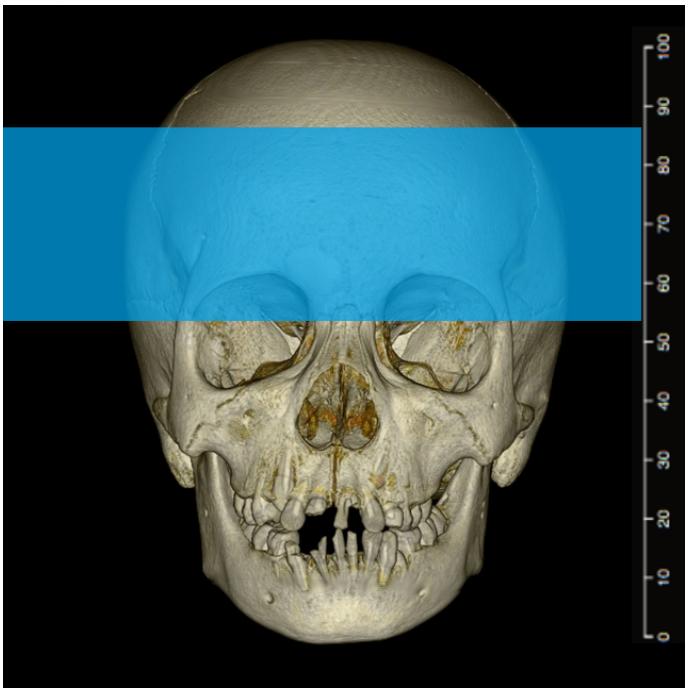

FIGURE 8.4: Taking 55th to 85th percentile of the skull

8.4 Brain stripping results in 3D

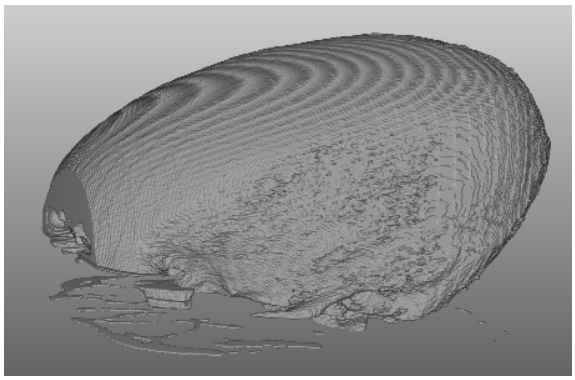

FIGURE 8.5: Brainstripped image in 3D

8.5 Annotation 1 compared to other ground truths

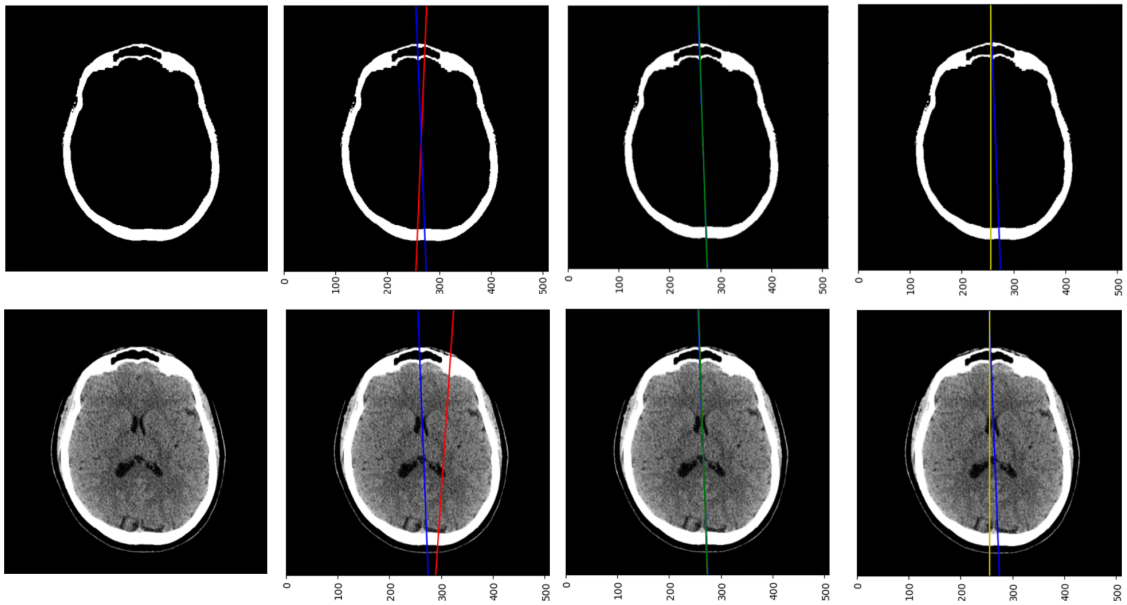

FIGURE 8.6: Annotation 1 compared to predictions of the best model (red), annotation 2 (green), and the symmetry line (yellow).

8.6 Risk Analysis

ID	Risk Description	Risk Category	Likelihood of Risk	Impact if Occurs	Severity Rating	Risk Owner	Risk Response	Mitigation Action	Contingency Plan
1	Someone unauthorized gets access to the data	Information Security Risk	Medium	High	High	Security Manager	Mitigate	Implement 2FA, encryption and firewalls. Regular security checks.	
2	General regulatory and compliance risk - GDPR complience, privacy policies, personal data processing	Information Security Risk, Technology Risk	Medium	High	High	Security Manager	Mitigate	Regular checks if the data is according to GDPR standards	
3	Not enough data available to train a good model	Technology Risk	High	Medium	Medium	Data Science Team	Mitigate	Find more suitable data	
4	Not high-quality enough data available to train a good model	Technology Risk	Medium	Low	Low	Data Science Team	Mitigate	Find more or other suitable data	
5	Model does not generalize well to unseen data	Technology Risk	Medium	Low	Low	Data Science Team	Mitigate	Find more suitable data to train with and apply more generalization techniques	
6	Not enough computation power	Technological Environment Risk	Low	Low	Low	Data Science Team	Transfer		
7	Bad integration with current software at MST	Technological Environment Risk	Low	Medium	Low	Project Manager	Mitigate	Checking available hardware and software at MST and make design choices accordingly. Communicate regularly.	
8	Updating or adapting the model is too complex for MST	Technological Environment Risk	Low	Medium	Low	Project Manager	Transfer		
9	Clinicians do not adapt well to the model	Human Risk	Medium	High	High	Clinical Lead, Project Manager	Mitigate	Introduction talks and videos with clear instructions of the use of the product. Allow clinicians to provide feedback.	
10	Model detects a hematoma when there is none	Human Risk	Low	High	Medium	Clinical Lead	Mitigate, contingency plan	Validating the model with experts and improving its accuracy. Involving clincians in the development, emphasizing the tool is not leading.	Establish a protocol for manual review of all positive detections by a medical expert.
11	Model detects no hematoma when there is one	Human Risk	Low	High	Medium	Clinical Lead	Mitigate, contingency plan	Validating the model with experts and improving its accuracy. Involving clincians in the development, emphasizing the tool is not leading.	Establish a protocol for manual review of all positive detections by a medical expert. Adding urgent care for missed detections.
12	Interdisciplinary lack of communication, causing lack of clairty and confusion	Project Planning Risk	Medium	Low	Low	Project Manager	Mitigate	Regular meetings and shared communication platforms	
13	Business Case becomes obsolete or is undermined by external or internal changes	Project Planning Risk	Low	High	Medium	Project Manager, Business Development manager	Mitigate, contingency plan	Regular reviews of the goals and current market	Develop a strategy to shift the project to the right focus
14	The product does not significantly advance current procedure	Financial Risk	Low	Medium	Low	Project Manager, Business Development manager	Mitigate, contingency plan	Conduct market research and ensure unique value propositions	Reasses the projects features and strategies
15	Lack of funds that would support the recovery of the product in case of malfunctioning	Financial Risk	Low	Medium	Low	Finance Manager	Mitigate, contingency plan	Implement strict budgeting and check regularly	Make use of a pre-approved financial emergency plan or identify priority features.