AALBORG UNIVERSITY
STUDENT REPORT

An Automatic Algorithm for Quality Assurance of MRI Scanners using a DWI Phantom

Aalborg University
Biomedical Engineering and informatics
School of Medicine and Health
Master's Thesis

By

Andreas Ormstrup
Mathias Faulkner

AALBORG UNIVERSITY
STUDENT REPORT

School of Medicine and Health
Fredrik Bajers Vej 7
DK-9220 Aalborg Ø
http://smh.aau.dk

Title:
An Automatic Algorithm for Quality Assurance of MRI Scanners using a DWI Phantom

Theme:
Master's Thesis

Project Period:
Spring Semester 2018

Project Group:
Group 10415

Participant(s):
Andreas Ormstrup
Mathias Faulkner

Supervisor(s):
AAU supervisor:
Ph.D. Lasse Riis Østergaard
AAU CO-supervisor:
Ph.D. Maciej Plocharski

Copies: 5

Page Numbers: 119

Date of Completion:
June 07, 2018

Abstract:

Functional diffusion maps (fDM) evaluates the treatment response with an apparent diffusion coefficient (ADC) over time. Diffusion weighted (DW) images may contain artifacts due to the low signal to noise ratio. Low-quality DW images may impair diagnosis and therefore quality assurance (QA) of the MR scanners has to be conducted. An automatic algorithm for QA of MR scanners was developed in Python, that based on DWI images of a DWI phantom, quantified the accuracy and precision of the ADC measurements of a ADC insert, and the achieved sharpness of a Siemens star insert. The algorithm had a 100 % accuracy for detection of the correct slices of the ADC ramp and the Siemens star within a consistency data set. The sharpness quantified by the algorithm had a connection to the visual analog scale (VAS) score evaluated by a radiologist. A region of interest (ROI) was determined within the ADC ramp to quantify the mean ADC, which had minor differences from the quantified ADC by a radiologist. The day-to-day variance between different ADC measurements was investigated, and temperature correction of the ADC was performed, to achieve more consistency. The necessity of QA was furthermore proven, when an error in the post-processed ADC map of a 3 tesla MR scanner was found.

Table of Contents

Preface

The authors would like to thank Ph.D. Lasse Riis Østergaard and Ph.D. Maciej Plocharski for their supervision and comments during this Master's Thesis.
Furthermore, we would like to thank Dr. Michael Bach because of his information about the DWI phantom from HQ Imaging and preliminary project ideas.
Last but not least, we would like to give a special thanks to the Department of Radiology at Vejle Hospital for data acquisition of the DWI phantom and contribution of helpful information regarding daily routines in the department.

Aalborg University, June 7, 2018

Andreas Ormstrup
<Andreas.Ormstrup@gmail.com>

Mathias Faulkner
<mathiasfaulkner@hotmail.dk>

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is a diagnostic imaging tool developed to generate cross-sectional images of the anatomy of the human body. MRI can both be used to generate images of healthy structures or diseases as cancer. [1; 2]
An addition MRI is diffusion weighted imaging (DWI), which is a method for quantitative and qualitative investigation into the diffusion of the hydrogen protons in water. This diffusion of protons gives additional information to the conventional MRI. In diseases as ischemia or malignant tissue the tumor cellularity (tumor cell density) increases, which results into a restriction of the diffusion of protons. The diffusion can be quantified by an apparent diffusion coefficient (ADC) and ADC maps. [3; 4; 5]
ADC maps could be used in evaluation of cancer treatment response, as there is an inverse relation between ADC and tumor cellularity, meaning the higher cellularity, the lower ADC. [6]
Functional diffusion maps (fDM) is a method for evaluating the change in the ADC maps over time, which provides early information regarding the cancer treatment response. [7] When monitoring the treatment response of a patient with multiple scans over different days, it is important to have the right scanner settings in order to reduce variation of the scans, and to obtain high-quality images. [4; 8]
Therefore to obtain high-quality images and evaluate the consistency of scanners, quality assurance (QA) can be performed. QA is widely used in diagnostic imaging as for instance in CT scans, to prevent low-quality images that can impair diagnosis, increase unnecessary costs of ineffective treatment and avoid unnecessary exposure to radiation. [9]

This Master's Thesis investigates QA of MR scanners in order to document consistently high-quality diffusion weighted images and functional diffusion maps.

Chapter 2

Background

The purpose of this chapter is to provide the background information regarding magnetic resonance imaging (MRI) and diffusion weighted imaging (DWI). Here the MR spin echo is explained and how DWI includes an addition to the standard sequence. Apparent diffusion coefficient (ADC) and ADC maps are furthermore explained in order to provide information on how ADC maps can be used as a biomarker for treatment response. Lastly, it is explained how quality assurance (QA) can be performed on MR scanners in order to achieve high-quality DW images.

2.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a diagnostic tool developed to generate cross-sectional images of the body without radiation or being harmful. MRI uses a strong magnetic field to alter the orientation of the protons in the hydrogen molecules and collects information about the orientation. The proton have a south and north pole, the strong magnetic field in the MR scanner makes the protons align. Hereafter a pulse of radio frequency is applied to alter the orientation of the protons by disrupting the magnetic field. The radio frequency (RF) pulse is only applied for a short amount of time, so the now misaligned protons are in a high energy state. To get into a low energy state the protons must align with the magnetic field again. The time for the protons to align is the concept used for creating the image from the emission of low energy radio frequency photons, which is called an MR spin echo sequence. [1; 2]

MR spin echo sequence consists of three periods, an excitation pulse of 90^o, followed by a refocusing pulse of 180^o and the echo assembled by the MR scanner. To define the spin echo sequence the echo time (TE) and the repetition time (TR) must be explained. The TE is the time between the RF pulse and the echo, while TR is the time between the RF pulses. TE can have multiple entries in the TR, if this is the case the echoes will be called TE1, TE2 etc. [10] The spin echo sequence is illustrated in Figure 2.1.

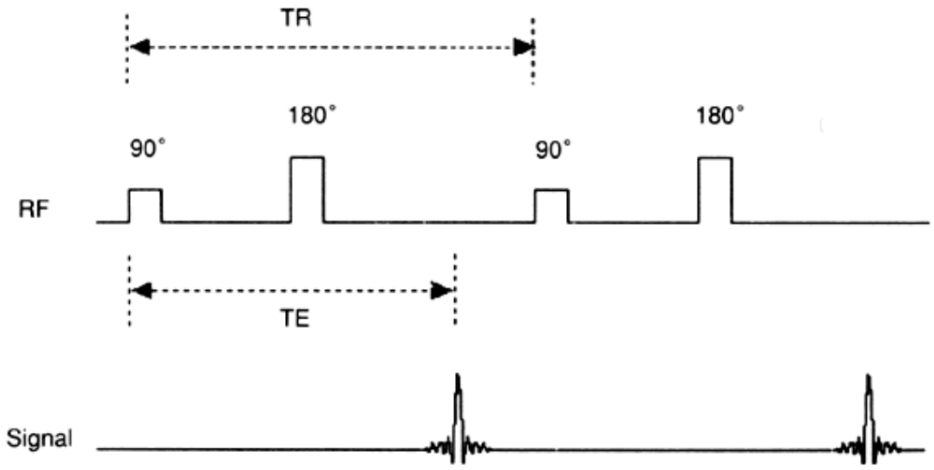


Figure 2.1: The MR spin echo sequence over time. First a pulse of 90^o is excited, followed by a refocusing pulse of 180^o and the resulting echo signal is assembled by the MR scanner.

Multiple types of sequences can be used for creating MR images. In MRI the sequence is created from anatomical structure of the body, and in functional MRI (fMRI) the images is created from the metabolic changes in the body. MRI has different kinds of sequences each providing T1 and T2 weighted images. T1 weighted images are created with a short TE and TR, while longer TE and TR are used to create T2 weighted images. fMRI is as earlier mentioned generated by the metabolic changes in the body. A sequence of fMRI is for an example perfusion weighted MRI or diffusion weighted MRI. [11]

2.2 Diffusion Weighted Imaging

Diffusion Weighted Imaging (DWI) is based on the Brownian motion, which is defined as molecules (eg. water and hydrogen protons), when under influence of thermal energy, are randomly displaced or diffused. The resulting echo signal from the protons can be quantified by a apparent diffusion coefficient (ADC) and thereby be used as an evaluation tool for eg. treatment response or disease progression. The rate of diffusion of protons changes in different tissues, where the diffusion in eg. malignant areas the diffusion may be restricted. Therefore the image contrast in DWI changes in relation to the diffusion of protons in different tissues. DWI today has multiple applications including thoracic malignancies, brain tumors and acute brain ischemia, where DWI is the preferred imaging modality for stroke patients. Regarding brain tumors the ADC

values have shown to be decreased at areas with tumors, as the cell density is higher and therefore restrict the diffusion of protons. [5; 3]

In order to generate a DW image the readout echo signal from the protons are dependent on two diffusion pulsed gradients. Therefore, in addition to a standard MR spin echo (SE) sequence, the pulsed gradient SE sequence in DWI contains two additional diffusion pulsed gradients, as seen in Figure 2.2. [4]

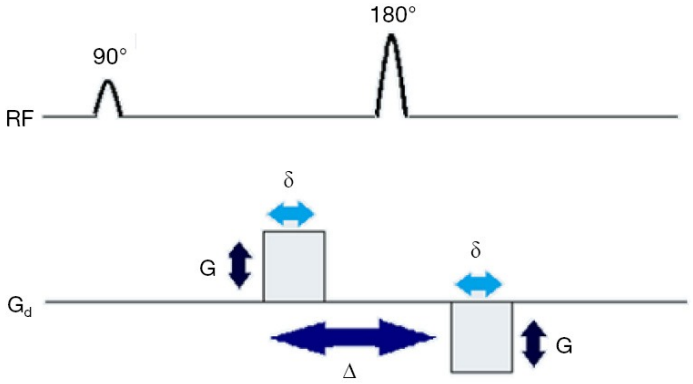


Figure 2.2: Diffusion pulsed gradient SE sequence in DWI. In addition to standard MR SE sequence, two diffusion gradient pulses (G_d) was added to the sequence, where G is the magnitude of the applied gradient, δ is the duration of the diffusion gradient and Δ is the time interval for the diffusion gradients. RF is the radio frequency pulse.

The diffusion pulsed gradient sequence in DWI are added to the standard MR spin echo sequence in order to generate the diffusion weighting in an MR image. As seen in Figure 2.2 the purpose of the first gradient is to induce a phase shift to the protons depending on the position of the proton. This means that the applied phase shift to the protons is different depending on the position. The purpose of the second gradient is to reverse the changes induced by the first gradient on the protons. This is achieved by inducing a gradient pulse with exactly the same duration and magnitude although with opposite direction. Due to the pause between the applied gradient pulses, there is a window of time for the protons to diffuse. This means that if the protons have diffused before the application of the second gradient, the protons will not be able to return to the same phase as before the application of the first gradient. As the protons now spin out of phase the resulting net magnetization of the protons decreases, meaning that the resulting echo signal is decreased achieving signal attenuation. This signal loss is given by Equation 2.1. [4]

$$S(b) = S_0 e^{-bD} \tag{2.1}$$

Where $S(b)$ is defined as the signal received for a particular gradient value defined as b (or b-value). S_o is defined as the signal strength where there has not been applied any diffusion weighting. D is the diffusion coefficient, which is defined as the rate of diffusion of protons in a medium. The b-value reflects the diffusion time interval and the strength of the diffusion gradients, as expressed in Equation 2.2. [4]

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{2.2}$$

Where γ is a constant of 42.58 $^{MHz}/_{T}$, that describes the gyromagnetic ratio of a hydrogen proton. As seen in Figure 2.2, G is defined as the magnitude of the diffusion gradient, δ is the duration of the gradient and Δ is defined as the time interval between the two diffusion gradients. [4]

2.2.1 ADC and Cell Density

In addition to the contrast caused by diffusion, the DW images also have T1 and T2 contrast, as the DWI sequence is added to the standard MR sequence. DW images are therefore prone to be influenced by a T2 echo signal, called T2 shine-through effect, which is similar to the contrast of restricted diffusion and therefore an artifact on the DW image. Qualitative measures such as diffusion maps or ADC maps can be used to avoid these artifacts. ADC maps are created with two DW images with different b-values, usually images obtained at b-values 0 and $1,000$ $^{s}/_{mm^2}$. Each voxel on the ADC map can be quantitatively evaluated by the value of ADC. The ADC therefore quantifies the diffusion of each voxel as expressed with Equation 2.3. [4]

$$ADC\ [^{mm^2}/_{s}] \cong \left[\ln \tfrac{S(b_2)}{S(b_1)}\right] \cdot \frac{1}{b_1 - b_2} \tag{2.3}$$

Where b_1 is the b-value of 0 $^{s}/_{mm^2}$ and b_2 is the b-value of $1,000$ $^{s}/_{mm^2}$ in this example. $S(b_1)$ and $S(b_2)$ is the corresponding measured signal for b_1 and b_2.

The signal intensity of a voxel in the ADC map is directly related to the changes in diffusion inside different types of tissue. A study [6] by Hayashida et al. investigated the correlation between ADC and tumor cellularity (tumor cell density). Cancerous or malignant tissue in the peripheral zone of the prostate has a mean ADC of 986 $^{\mu m^2}/_{s}$, with a 95 % confidence interval from $895-1,076$ $^{\mu m^2}/_{s}$. In contrast the noncancerous or benign tissue in the same area has a mean ADC of $1,572$ $^{\mu m^2}/_{s}$, with a 95 % confidence interval from $1,471 - 1,672$ $^{\mu m^2}/_{s}$. [12] Figure 2.3 shows the results concerning the correlation of ADC and cellularity.

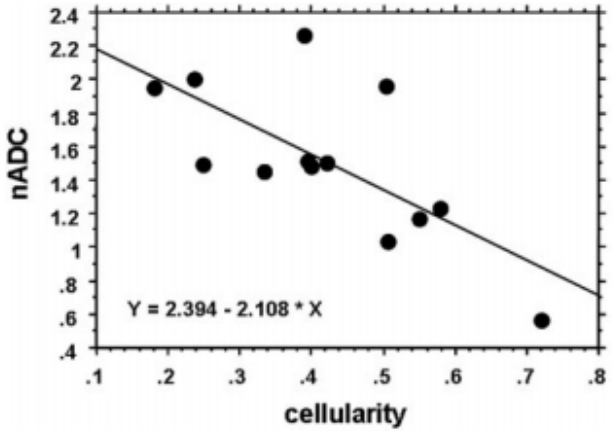


Figure 2.3: Inversely correlation between nADC (normalized ADC) and tumor cellularity. The lower tumor cellularity the higher ADC.

As seen in Figure 2.3, there was found a inversely correlation between ADC and tumor cellularity. Moreover there was not found a correlation between T2 weighted images and tumor cellularity. This inversely correlation means that the ADC decreases with higher cell density. Therefore, if diffusion is low in a tissue, indicating high cellularity, the voxel in the ADC map will have a low or dark contrast in a grayscale image. If the diffusion is high, and thereby low cellularity, the ADC map will have a high or white contrast. Figure 2.4 shows a poorly differentiated adenocarcinoma with different image modalities. A poorly differentiated adenocarcinoma is a description of a tumor that has an abnormal cell structure compared to normal cells, and has a higher spread rate than well differentiated adenocarcinoma, which has a cell structure closer to normal cells. [6]

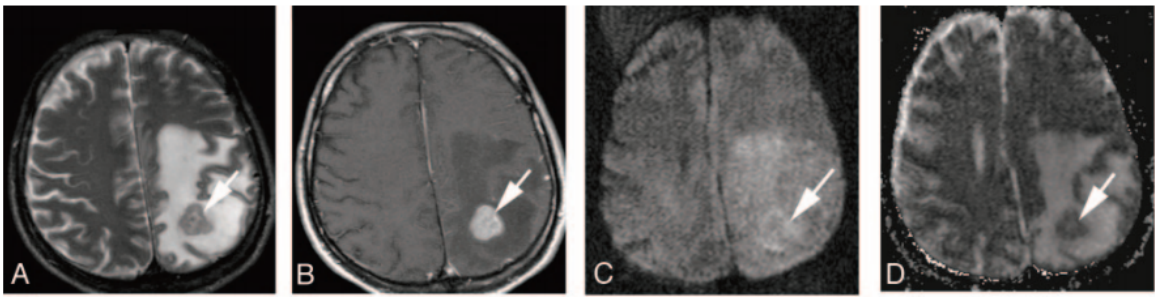


Figure 2.4: Image A shows a T2 weighted image of left parietal subcortical lesion, indicated by the arrow, and image B is the T1 weighted image. Image C is a DW image where the lesion is slightly hyperintense, meaning it has a slightly whiter contrast than the souring tissue. Image D shows the ADC map, where the tumor is hypointense, having a darker contrast than the surroundings. [6]

As seen in Image C in Figure 2.4 the hyperintense contrast indicates that the tumor has a restricted diffusion due to the higher cellularity of the tumor. Image D shows the ADC map, where the potential T2-weighted shine through is removed. The tumor is hypointense, which indicates the inversely correlation between ADC and tumor cellularity. 2.4

2.2.2 ADC as a Biomarker for Treatment Response

In MRI a contrast agent called Gadolinium is often used in order to enhance signal from lesions, by shortening the relaxation rate of the protons. In normal tissue the blood-brain barrier (BBB) effectively stops the contrast from accumulating. However, in lesions the BBB may be leaking, meaning that it may accumulate the contrast which enhances the tumor or lesion. This enhancement of the tumor is today traditionally evaluated by the MacDonald criterion in order to assess the brain tumor response to treatment. However pseudoprogression may occur, which is defined as an increase in enhancement of a lesion or tumor size, but without tumor progression. Meaning it could falsely indicate tumor progression. The pseudoprogression is dismissed or confirmed relatively late, if the tumor enhancement has stabilized or decreased, or by investigating tissue removed with surgery. [13]

RECIST (response evaluation criteria in solid tumors) is a guideline for an assessment of tumor treatment response. As suggested by RECIST 1.1 todays functional methods for treatment response assessment includes PET and PET-CT, MR spectroscopy and contrast enhanced MRI. However these methods have several limitations. A disadvantage with PET and PET-CT is the expensive cost, the difficulties differentiating tumors and inflammation and the exposure to radiation. MR spectroscopy has low spatial resolution and contrast enhanced MR is not suited for daily clinical routines due to the complicity of the image analysis. [14; 7]

DWI and ADC maps could be used as an early evaluation of cancer treatment response. As mentioned in Section 2.2.1, there is an inverse relation between ADC and cellularity. This means that cancer treatment with radiation or chemotherapy with resulting necrosis of tissue in the tumor would result in a lower cellularity of the tumor. Lower cellularity would mean less diffusion restriction inside that tissue resulting in a higher ADC value. Therefore an increase in the ADC value as a result of treatment would indicate a positive therapeutic response. [7]

Functional diffusion maps (fDM) are a method to evaluate the change in ADC value over time. fDM is a voxel-by-voxel based method that quantifies the change in diffusion for each voxel over time. These diffusion changes in each voxel can be superimposed on anatomic MRI in order to obtain spatial information concerning the tumor. The ADC

changes in the fDM are shown as a color map where blue indicates decreased ADC values, green indicates no change and red indicates increased ADC values. Advantages with DWI and fDM are that it provides early information regarding the cancer treatment response and there is no radiation involved. The acquisition time for DWI is a few minutes. This might give suggestions in a clinical routine for individual treatment and increased long term survival. Moreover cost for the department for ineffective treatments could be reduced. [7]

2.3 Quality Assurance of MR Scanners

In order to obtain high-quality DW images, it is important to optimize the scanner settings in a routine clinical practice. This is especially important when monitoring the disease progression of a patient with multiple scans over different days. DWI may also contain artifacts due to the long scanning time compared to T2 weighted images and due to the low signal to noise ratio. DWI makes several assumptions in the acquisition of images, as a perfectly shaped RF pulse, infinitely fast changes of the gradients and a perfect magnetic field homogeneity. These assumptions are not feasible in reality and are therefore a limitation of DWI, which results into artifacts as distortion, noise and low image quality. Artifacts as distortion in the images could occur because of differences in the magnetic field. Different patient related artifacts in DWI may also occur which includes movements of the patient, breathing, and metallic artifacts, which may cause distortion in the images. In order to document consistently high-quality DW images, it is mandatory to perform quality assurance (QA) of the MR scanners. Today, most of the QA of DWI are performed manually by visually inspecting the DW images, slice by slice in a scanning sequence. However this examination often results in low consistency between data and inter-user subjectivity of the examinations. [4; 8]

In order to overcome these qualitative examinations of the DW images, MRI phantoms has increasingly being used for QA of DW images. MRI phantoms have known properties similar to human tissue and can be used for research, testing reproducibility and as quality assurance of scanners. The main focus in the creation of an accurate and stable phantom is using predefined apparent diffusion coefficients (ADC). [15]

2.3.1 DWI Phantom

HQ Imaging, a spin-off of the German Cancer Research Center (DKFZ) [16], has developed a DWI phantom for use in QA of diffusion weighted imaging. Their developed DWI phantom contains an ADC region simulating a non-restricting environment with

high diffusion. Figure 2.5 shows the spherical DWI phantom from HQ Imaging.

Figure 2.5: DWI phantom from HQ Imaging. Spherical phantom container filled with an aqueous polyvinylpyrrolidone (PVP) solution. The container is a sphere with a diameter of 150 mm.

The spherical phantom has a diameter of 150 mm. The phantom is filled with an aqueous polyvinylpyrrolidone (PVP) solution. The PVP solution is approximately 10 % $^{w}/_{w}$ (weight-by-weight) concentration in water, giving an ADC of 1599 $^{\mu m^2}/_{s}$ at 20° Celsius, simulating a non-restricting or benign environment. [16; 17]

PVP is often used for differentiating ADC values as well as T1 and T2 relaxation time in an MRI phantom. Different concentration of PVP in water gives several different T1, T2 and ADC values in the phantom, which can be seen in Table 2.1. [18]

Table 2.1: Controlling values in phantom by differentiating concentration of PVP in water. [18]

PVP	T1 ± SD (ms)	T2 ± SD (ms)	ADC ± SD (20.16 ^{o}C)	ADC ± SD (22.11 ^{o}C)
10 %	2300 ± 5.2	1953.1 ± 0.9	1630 ± 12	1714 ± 6
14 %	2010 ± 3.8	1692 ± 1.3	1439 ± 11	1522 ± 6
18 %	1770 ± 3.4	1479.3 ± 0.9	1318 ± 8	1389 ± 7
25 %	1200 ± 3.4	1017.3 ± 0.8	927 ± 9	983 ± 12
40 %	714 ± 1.7	584.8 ± 0.2	602 ± 10	640 ± 9

The ADC values are dependent of the temperature, therefore a method for controlling the temperature is using an ice bath. However the DWI phantom from HQ Imaging has an integrated thermometer which allows for temperature correction of the phantom. Besides the thermometer, the DWI phantom includes two other regions; the ADC ramp insert and the resolution insert, see Figure 2.6.

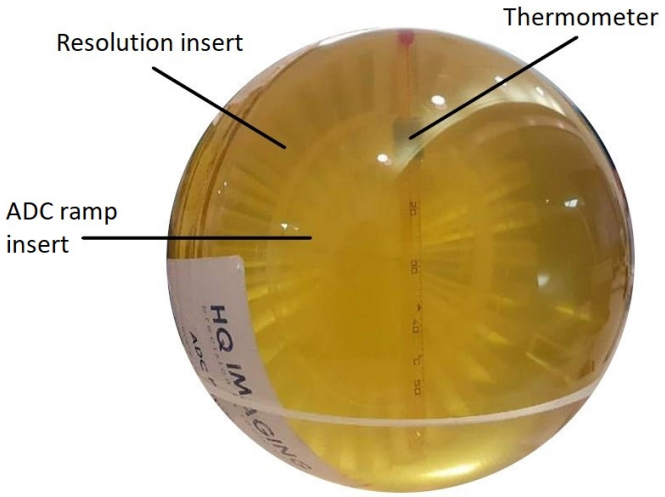


Figure 2.6: DWI phantom from HQ Imaging with indication of thermometer, ADC benign region, and resolution insert.

Cross-sectional images are taken throughout the phantom in order to obtain images of the regions inside.

The ADC ramp insert is a plastic ramp, which around the ramp goes from a maximum of plastic to a maximum of PVP solution. The PVP solution is called the ADC region and enables the assessment of accuracy and precision of the ADC measurements. Figure 2.7 shows a T2 weighted cross-sectional image of the ADC ramp insert within the DWI phantom. The plastic is black because it gives no signal as there is no diffusion of water protons, and the ADC region is white.

Figure 2.7: T2 weighted image of the ADC ramp insert of the DWI phantom.

The resolution insert in the DWI phantom allows the assessment of the resolution or sharpness achieved in the examination of the phantom. Figure 2.8 shows a T2 weighted cross-sectional image of the resolution insert within the DWI phantom.

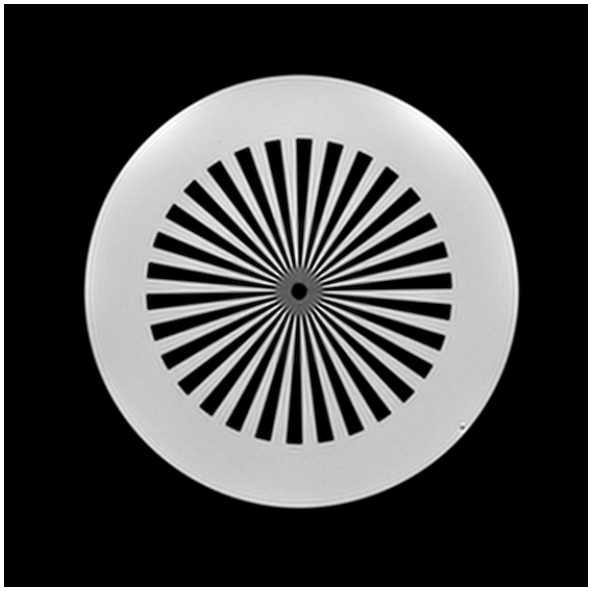


Figure 2.8: T2 weighted resolution insert of the DWI phantom.

The resolution insert, also called a Sinusoidal Siemens Star, is implemented to evaluate the sharpness of the image from the MR scanner. Sharpness is defined as the ability to reproduce fine details in an image. The diameter of the Siemens star is 120 mm. As seen in Figure 2.9, the Siemens star has 30 spokes or contrast bars with an opening angle of 6°. [16]

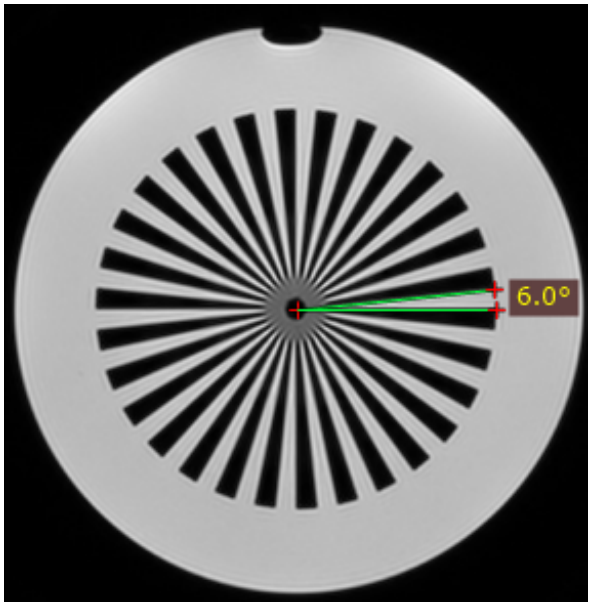


Figure 2.9: T2 weighted Siemens star with indication of the opening angle of 6° each of the 30 contrast bars.

Siemens stars has been used for regular cameras in order to investigate the sharpness of the camera. [19] Similar to cameras, the sharpness of MR scanners can be conducted based on the Siemens star. To evaluate the MRI scanners an inspection of the sharpness of the Siemens star is done by observing the differentiation between black and white to the center of the Star. The further towards the center this differentiation is visible, the better sharpness of the image. The lines which differentiate between black and white can be used to see the sharpness of the image, where a perfect sharpness is obtained when the edge between black and white changes from lowest to highest in intensity value.

2.3.2 Routine Clinical Practice

As explained in Section 2.2.2, there is a need for early assessment of cancer treatment response. Functional diffusion maps (fDM) quantify the treatment response with ADC over time, earlier than traditional RECIST methods. In order to implement fDM in daily clinical routines, a QA of the performance of the MR scanners has to be conducted, where an investigation of the day-to-day variance of MR machines along with the variance between MR machines is needed if a patient is scanned on different machines. This QA is necessary in the evaluation of treatment response in order to guarantee the progression in treatment is a treatment response of a patient and not

just a variance from scanner to scanner or from day to day. This gives value for the patient as an earlier assessment of treatment response would result in an avoidance of unnecessary side effects from radiation- or chemo-therapy. Moreover this gives economical value for the health care department as a earlier assessment minimizes the unnecessary cost of ineffective therapy.

Department of Radiology at Vejle Hospital

This section includes information provided by the Department of Radiology at Vejle Hospital, hereafter referred to as the Department of Radiology.
The Department of Radiology have obtained a DWI phantom from HQ Imaging in the pursuit of performing QA on their MR scanners. However manual QA of their MR scanners can be time consuming and takes up time that could be used on patients. A manual QA of an MR scanner takes approximately 5-7 minutes to complete.

Different automatic methods for QA are available today. [16] HQ Imaging provides an automatic web-based analysis of the images of their DWI phantom. These reports can be available within 24 hours but the department would prefer to be able to define the obtained values and to share the values digitally on the individual workstations without delay. Moreover, according to HQ Imaging, the web-based analysis does not include an automatic assessment of the sharpness of the Siemens star. Therefore there is a need for an automatic algorithm for QA of MR scanners that automatically evaluates the sharpness of the Siemens star.

The Department of Radiology needs QA of their MR scanners with minimal time consumption as possible, therefore the first need for such algorithm is that it has to be automatic. Therefore there is a need for an automatic selection of the ADC ramp insert and the Siemens star in the phantom. Moreover a need is to automatically detect a region of interest (ROI) for the ADC region within the ADC ramp insert and assess the mean ADC values. An automatic quantification of the sharpness in the image is also a need for the Department of Radiology, because to this present it is qualitatively evaluated by eye and thereby have a risk of user-subjectivity in the assessment of the sharpness. Another need for the Department of Radiology is that the current daily values from the QA of the latest consistency data set must be visualized on the first or main page of the algorithm. Moreover a day to day graph from all the latest consistency data sets needs to be available with a visualization of the mean and the standard deviation (SD) in order to investigate the variance. Further all QA tests which deviates from the SD needs to be marked red.

Chapter 3

Project Aim

Quality assurance (QA) of MR scanners is important in order to document consistently high-quality MR images. QA of MR scanners can be performed with a DWI phantom, which consists of an ADC region and a Siemens star for sharpness assessment. Manual QA of MR scanners with a DWI phantom can be time consuming for the professionals at a Department of Radiology. Different automatic QA algorithms are available today, however an automatic assessment of the sharpness is not yet available. Therefore there is a need for an automatic algorithm for QA of MR scanners based on images of a DWI phantom. This leads to the following project aim:

Develop an automatic algorithm to perform QA of MR scanners, by assessment of the accuracy and precision of the ADC measurements and the image sharpness, based on images of the ADC region and the Siemens star of the DWI phantom.

3.1 Objectives

- Develop an automatic algorithm to detect the cross sectional slice of the Siemens star.
- Develop an automatic algorithm to detect a ROI of the Siemens star and quantify sharpness in the image.
- Develop an automatic algorithm to detect the cross sectional slice of the ADC region.
- Develop an automatic algorithm to detect a ROI of the ADC region and quantify a mean ADC.
- Investigate the day to day variance of ADC and sharpness, and the variance between MR scanners using statistical analysis.
- Validate and test the performance of the algorithm with comparison of the algorithms quantification of ADC values and sharpness, with an evaluation performed by a professional from the Department of Radiology.

Chapter 4

Methods

The purpose of this chapter is to describe how the data for the project was acquired and how the algorithm was developed. The development of the algorithm is divided into different blocks, where the purpose of each part of the algorithm is described and the development is explained.

4.1 Data Acquisition

The data used in this project were MR consistency data sets of a DWI phantom, provided by the Department of Radiology at Vejle Hospital. A consistency data set contains first a survey and then four different MR scanning sequences, T2 weighted images, DWI with b0 and b1600 and lastly the ADC map. Each scanning sequence contains 25 cross sectional DICOM images of the phantom. The Department of Radiology has three MR scanners from Philips, where two of them are 1.5 T scanners, named MR1 and MR2, and one is a 3 T scanner, named MR4. For all three scanners, a consistency data set was conducted of the phantom every working day over a period of approximately 7 weeks, provided 30, 29 and 30 consistency data sets for respectively MR1, MR2 and MR4. The consistency data sets were conducted on a DWI phantom from HQ Imaging. The phantom is further described in Section 2.3.1.

The protocol used for data acquisition and conductance of the consistency data sets were as following:

- Place the phantom in the MR scanner and rotate the phantom so the thermometer is located on the top.
- Adjust the angle of the slices for acquisition of sequences with arrows located inside the phantom as seen on Figure 4.1. This gives consistency in the imaging of the phantom, but the algorithm needs to take account for further small changes.

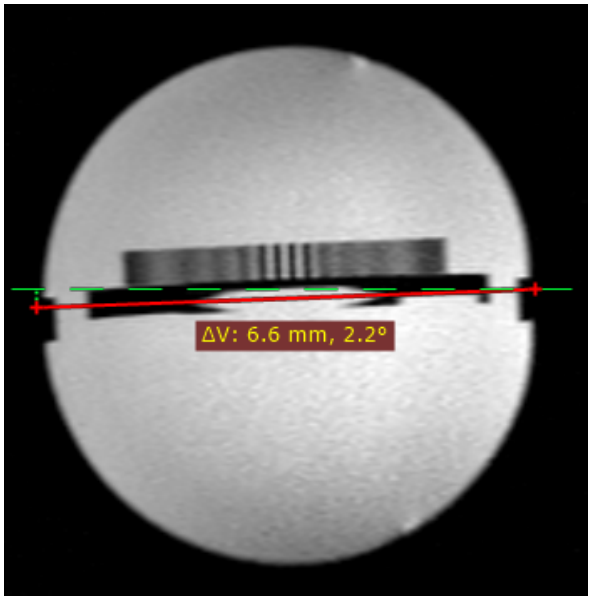


Figure 4.1: A survey sequence of the phantom where the angle for the cross-sectional slices is determined.

- Acquire the four MR scanning sequences T2 weighted, b0, b1600 and ADC map.
- Save the consistency data set for the radiologist to evaluate.
- The radiologist determines the accuracy of the ADC measurements with a region of interest (ROI) in the ADC region. The ROI is defined as a circle in the ADC region of the ADC ramp insert in each sequence, see Figure 4.2. The sharpness is evaluated of the T2 weighted image and ADC map of the Siemens star with a score from 1 to 5. Lastly the temperature of the thermometer in the phantom is noted.

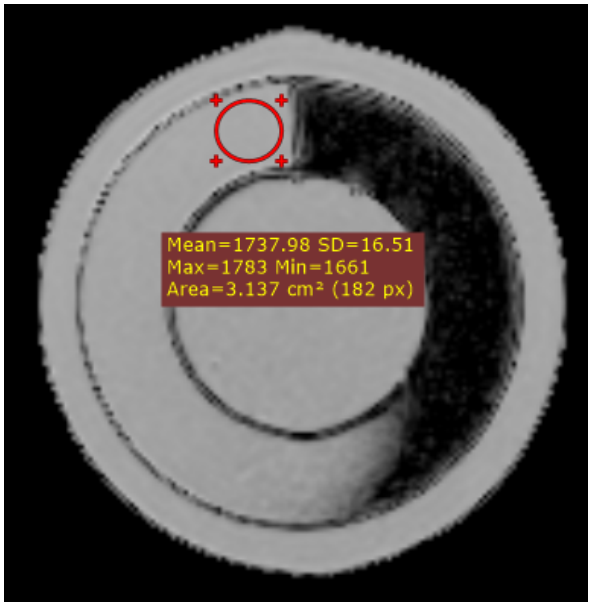


Figure 4.2: ADC map of the ADC ramp insert. A ROI is defined in the ADC region.

- The consistency data sets are uploaded to an external server with a picture archiving and communication system (PACS).

A PACS system that the authors had access to was needed, to enable download of the consistency data sets to use in the project. Therefore an external server was provided for the authors at Vejle Hospital. A PACS system was installed on the external server: Orthanc-Server, Version: 1.3.2, which is a standalone DICOM server developed by Sébastien Jodogne at the CHU of Liège (Belgium), in the Department of Medical Physics. [20]
The radiologists were then able to upload the consistency data sets to this PACS on the external server. From the PACS server the authors were able to download the consistency data sets locally to a PC.

4.2 Algorithm Development

An algorithm was developed in Python from Python Software Foundation, Version: 3.6.5. [21] A Graphical User Interface (GUI) was developed to ease the quality assurance and create a convenient and easy to use application for the radiologists. Image analysis was performed on two specific slices in the phantom, the ADC ramp insert and the Siemens star as described in Section 2.3.1.

The algorithm development was divided into six different blocks. Each block describes the purpose of the part of the algorithm and how it was developed. Block 1 describes the initial setup for the QA. This includes the development of the GUI and how the scanning sequences and DICOM images were chosen for the analysis. Block 2 describes how the Siemens star was detected among all 25 DICOM images within a sequence. Block 3 explains how the sharpness of the Siemens star was quantified. Block 4 describes how the ADC ramp insert was detected among all 25 DICOM images within a sequence. Block 5 describes how the ROI of the ADC region within the insert was detected and quantified. Lastly an explanation is provided in Block 6 regarding how the QA was visually presented. An overview of the pipeline of the algorithm is illustrated in Figure 4.3.

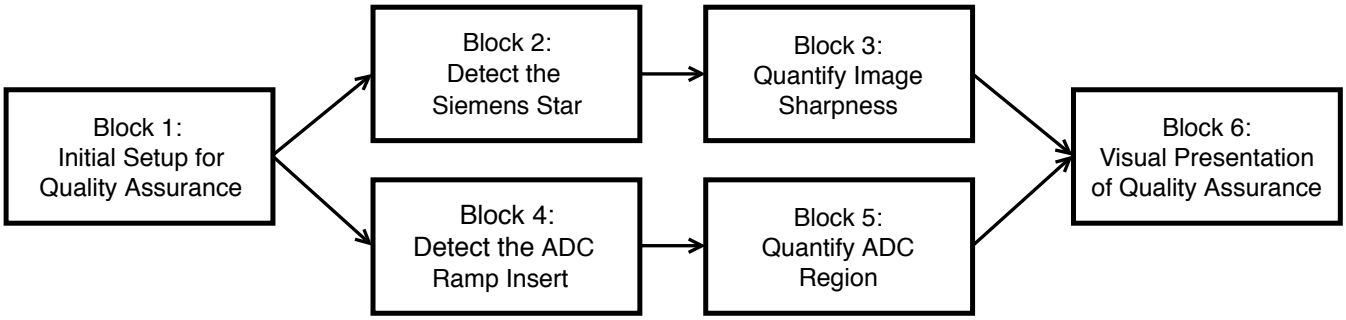


Figure 4.3: An overview of the different blocks in the algorithm development. First the initial setup is explained in Block 1, next Block 2 and 4 explains the detection of respectively the ADC ramp insert and the Siemens star. Block 3 describes the quantification of the Siemens star and Block 5 describes the detection of the ADC benign ROI and the quantification of the ADC region. Lastly Block 6 describes the visual presentation of the QA.

4.2.1 Block 1: Initial Setup for Quality Assurance

This section provides information of the purpose of the initial setup and how it was developed. The purpose of the initial setup was to enable a user friendly and fast selection of consistency data set for the quality assurance. Moreover the initial setup ensured that the correct scanning sequences of the consistency data set were chosen for the analysis. As the scanning sequences consisted of DICOM images, the purpose of the initial setup was also to read the DICOM images and Meta data, in order to enable analysis of the images. The DICOM images consisted of 16 bit images, meaning that the initial setup ensured that the DICOM images were read in 16 bit, in order to avoid loss of information in the images.

Tab for Loading Consistency Data Set and Start of the Algorithm

The Load Consistency Tab in the GUI was developed in order to ease the initial quality assurance, see Figure 4.4. The tab enables the load of one or multiple consistency data sets. The button "Load Single Consistency Data Set" is selected when loading one consistency data set of the phantom. The button "Load Multiple Consistency Data Sets" loads an entire folder with all the consistency data sets within.

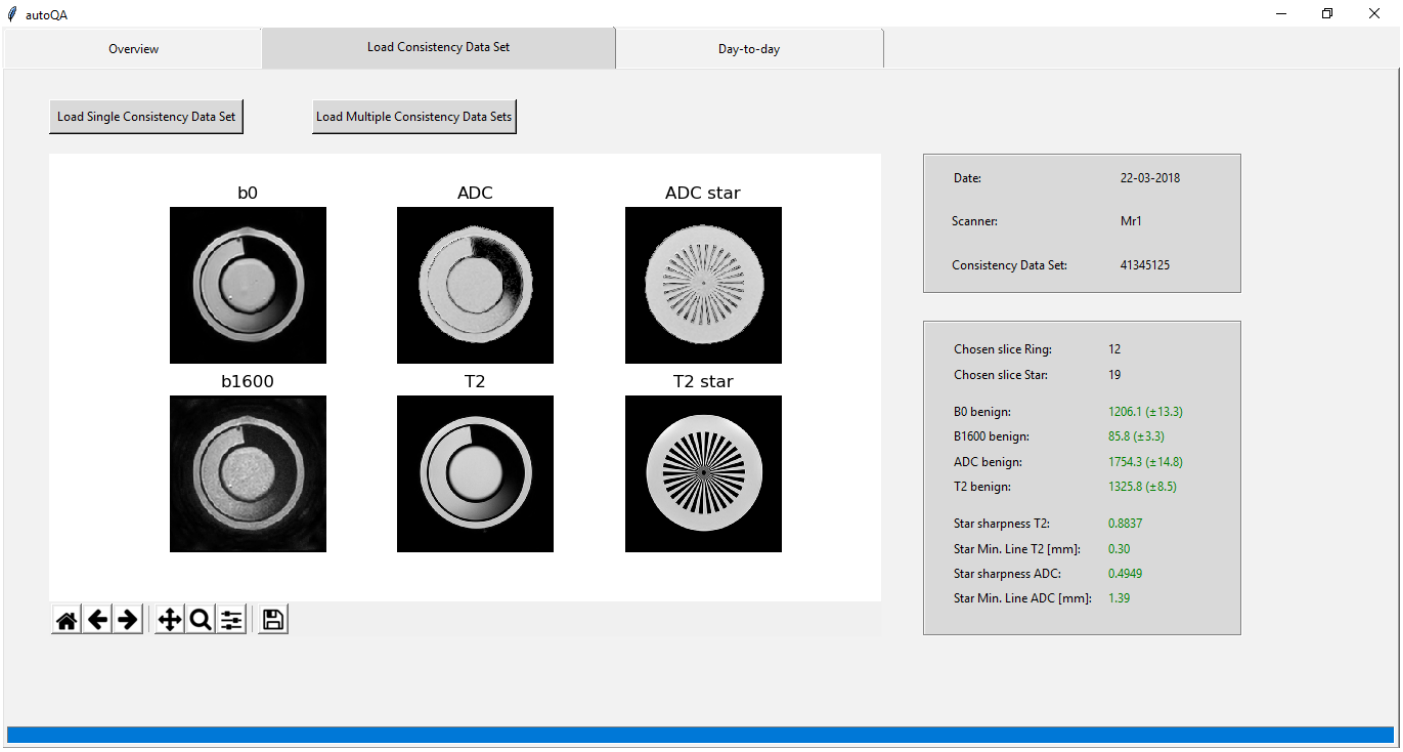


Figure 4.4: The tab "Load consistency data set" is shown, where a single consistency data set is loaded and the used images are shown in the canvas. The labels in the top right of the figure are information from the scanner and the labels bottom right are the results from the loaded consistency data set.

The canvas shown on Fig 4.4 is a presentation of the used images after the consistency data set had been loaded into the program. Information about the test is shown on the top right of the tab, which is the date of the scanning, name of the MR scanner and the ID of the consistency data set. The labels in the bottom right are the results from the consistency data set. As a scanning sequence consisted of 25 cross-sectional images of the phantom, the two chosen slices are shown, which in this specific consistency data set were slice 12 for the ADC ramp insert and 18 for the Siemens star. The other

results are the extracted values from the ADC region from three sequences b0, b1600 and ADC map, respectively. The results from the sharpness of the star are also shown for both the T2 weighted image and the ADC map.

DICOM Images

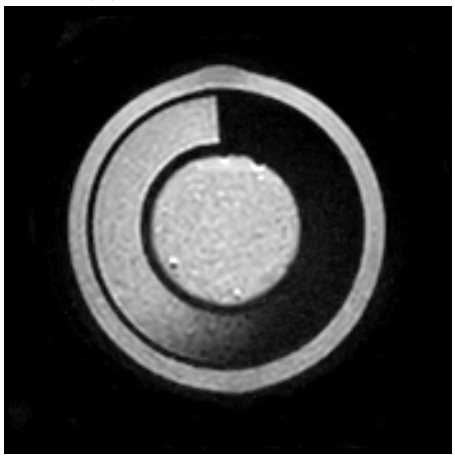
DICOM is the golden standard for images in health care. [22] The images contain information about the patient, modality type, sequence type etc. This contained information is called Meta Data, and is used so the data of the patient never can be separated from the image. [22] The data acquired from the Department of Radiology were 16 bit DICOM images. The main difference between 8 and 16 bit images is the value of the pixel in the image. An 8 bit image has 256 numeric values in grayscale, while a 16 bit image can have 65536 numeric values. This means that the automatic analysis of the benign ADC was conducted on the 16 bit images in order to obtain ADC values in ranges from 0-65536 and therefore preserve the most information in the image.

Scanning Sequences

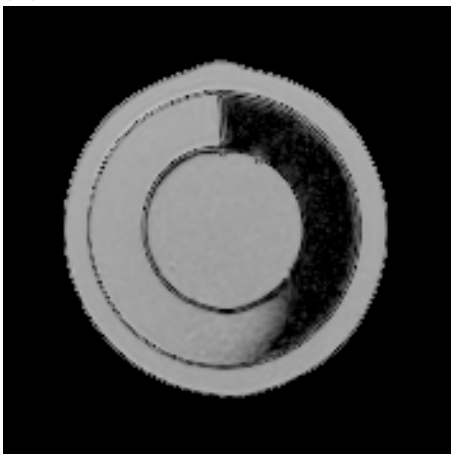
Each consistency data set of the phantom consisted of four scanning sequences; T2 weighted, b0, b1600 and ADC map. The sequences can be seen in Figure 4.5.

(a) T2 weighted image

(b) DWI image with a B value of 0



(c) DWI image with a B value of 1600



(d) ADC map obtained from the DW images

Figure 4.5: All four scanning sequences of the DWI phantom.

The four scanning sequences consisted of 25 cross-sectional images of the phantom. The resolution of the T2 weighted images was 560x560 pixels for the two 1.5 T MR scanners and 576x576 pixels for the 3 T MR scanner. The resolution of the diffusion weighted images, b0, b1600 and the ADC map was 176x176 pixels for the two 1.5 T MR scanners and 192x192 pixels for the 3 T MR scanner.

Save Results from a Consistency Data Set

The results from a consistency data set consisted of a quantification of the signal intensities of the ADC ramp insert for the four scanning sequences, T2, b0, b1600 and

ADC map. Moreover the results consisted of the quantification of the sharpness of the Siemens star for T2 and the ADC map. In order to investigate the day-to-day variance, the results from each consistency data set needed to be saved. The results was saved with a dictionary-like object, called Shelve. [23] Shelve functions as a database, where a key is selected for the underlying values. The key for the results of each consistency data set was chosen as the accession number of the consistency test. The accession number is unique for each consistency data set.

4.2.2 Block 2: Detect the Siemens Star

This section provides information about the automatic slice selection of the Siemens star. In order to develop an automatic algorithm, the Siemens star had to be detected automatically. As each scanning sequence consisted of 25 cross-sectional images, the correct slice of the Siemens star, according to Section 2.3.1, had to be detected. In order to detect the correct slice, a reference image of the Siemens star was selected. The T2 weighted scanning sequence was selected for the automatic slice detection, as the T2 weighted images consisted of high quality images of the phantom. Image similarity was then calculated between the reference image and all cross-sectional images in the scanning sequence.

Object Detection

The phantom is a circular object and the structures within are defined around the center. The structures in the phantom in this case is the contrast bars defining the Siemens star. In order to achieve image similarity to the reference image, the object defining the Siemens star had to be detected. To detect the structures within the phantom, the most appropriate method was to determine the center. There are multiple methods for determining the center of objects, one of them is using the outer contour [24] and extract the coordinates. The coordinates of the outer contour are the edge of the outer most structure within the image. The mean of all the coordinates was calculated and from this a center of the object could be defined. But because of deformities on the edge of the phantom the center was misaligned to the wanted point of center in the phantom, as indicated by a red dot in Figure 4.6.

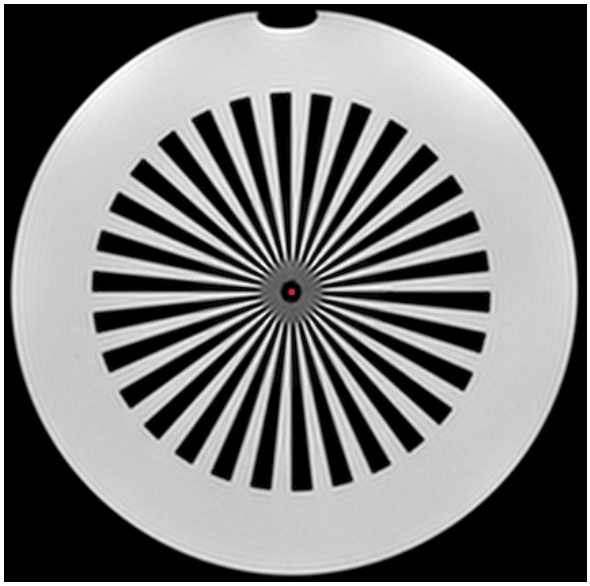


Figure 4.6: Wanted point of center within the phantom indicated by a red dot. A deformity can be seen in the top of the phantom image.

The wanted center was right in the middle of the Siemens star and therefore the center needed to be calculated from the edge without taking the deformities into account. If the phantom is considered to be a perfect circle without the deformities Hough circle transform can be used. [25] Hough transform was used to find the center and area of the phantom. The Hough transform finds edges in the image and draws circles with a specific radius, which iterates to find the most fitting circle in the image. The center of the circle is defined as the point in the image where the most circles overlaps, which is illustrated in Figure 4.7. [25]

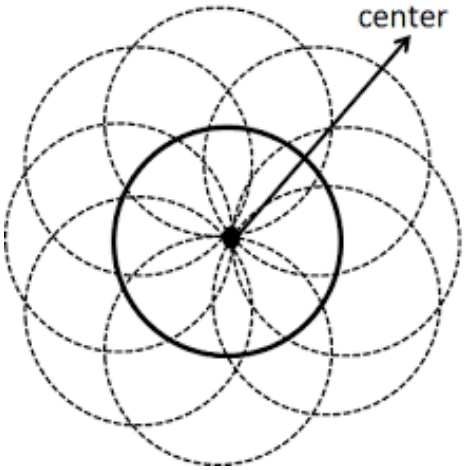


Figure 4.7: Hough circle transform. The center is defined as the point of most overlaps in the circle.

This is a 2D version of Hough Circle Transform, the transform used in this is a 3D model, where the radius is not defined but changes iteratively. The point in the 3D area with the most overlaps is the center of the circle, when iterated over the radius.

The circles drawn in the image are described by following term:

$$(x - a_x)^2 + (y - a_y)^2 = r^2 \tag{4.1}$$

Where a_x and a_y defines the center of the circle, r is the radius while x and y are the outer perimeters defined of the edge detection. [25]

The result of the Hough circle transform on a T2 weighted image can be seen on Figure 4.8.

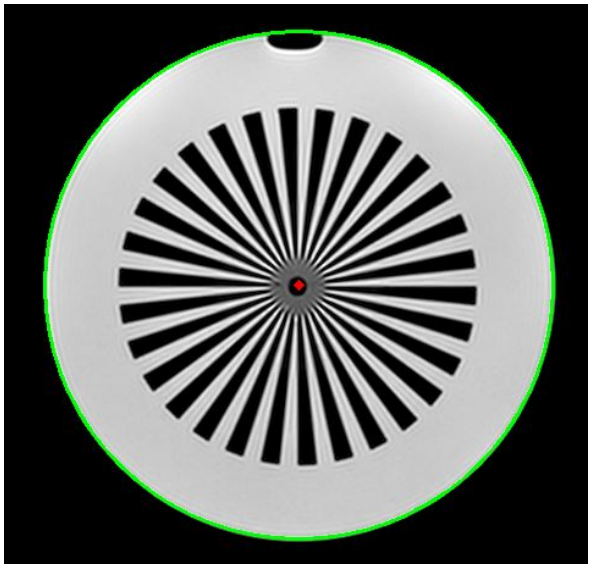


Figure 4.8: Hough circle transform on a T2 weighted image of the Siemens star. The found center and circumference are marked in red and green, respectively.

Image Similarity

Several different methods exists for image similarity. Mean square error (MSE) is a popular and easy to implement method for image similarity, and have been used for more than 50 years. MSE is a measure that calculates the error between two signals, by averaging the squared intensity differences in the images. However MSE has been proven to not be highly indicative of human perception of image quality. [26] Therefore other methods were developed in order to take advantage of how the human visual system (HVS) perceives image quality. Structural Similarity Index (SSIM) is a method that was developed based on that the HVS extracts structural information. SSIM calculates the similarity of images based on three different features in the images: the similarity of the local brightness $l(x, y)$, local contrasts $c(x, y)$ and local structures $s(x, y)$ as expressed in Equation 4.2.2. [27]

$$\begin{aligned} SSIM(x,y) &= l(x,y) \cdot c(x,y) \cdot s(x,y) \\ &= \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right) \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right) \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right) \end{aligned}$$

Where x and y are local areas in the same places in the images that are compared, meaning that multiple sliding windows in the images are compared. The local sample

means of the area x and y are defined as μ_x and μ_y, the local standard deviations are defined as σ_x, σ_y and σ_{xy} are the cross correlation. [27]

Image similarity of the Siemens stars with SSIM was conducted. However SSIM was not able to detect the correct Siemens star according to the reference image. The bad performance of SSIM could be due to rotation of the images. The placement of the contrast bars in the Siemens star compared to the reference image may vary due to rotation. Another method for image similarity is histogram similarity. Histogram similarity gives information concerning the distribution of intensities in an image. As histograms are not based on local areas in the images, but distributes the intensities of all pixels, histogram similarity is therefore not sensitive to rotation of the images. [28] Therefore image similarity of the Siemens star was based on histogram similarity. However unnecessary information concerning the image with the Siemens star was obtained, due to the black background and the outer perimeter of the phantom, as seen in Figure 4.8. Therefore in order to obtain information exclusively concerning the Siemens star, a mask was determined of the Siemens star. The mask used in this case is a spatial filter, where the spatial filtering is performed directly on the pixels in the image. Spatial filters are created in different sizes and can be used both for smoothing of pixels and for sharpening of edges. [29]

The purpose of the mask in this case was to segment the Siemens star. Therefore a binary image was created with the mask, where all pixels within the mask were assigned to 255 and all outside the mask were assigned to 0. The mask created in this case was generated from the found center, radius and circumference by Hough Circle Transform. In order to exclude the outer perimeter of the phantom, the mask was generated from 70 % of the radius. The intensity values of the images within the mask were extracted and a histogram was calculated of the intensity values. The same procedure was conducted on the reference image of the Siemens star.

Image similarity was then conducted where the similarity was determined by comparing the histograms of all 25 cross-sectional images to the reference image of the Siemens star. A method for histogram similarity is calculating the bin-by-bin distance between the images. The histogram distances compare corresponding histogram bins and does not take the correlation between neighbour bins into account. [28] Six different methods for histogram distance comparison were selected to determine the best performing distance method for image similarity. The six methods which were selected from the Python OpenCV package (Version 3.0.0) [30] were: Chi-square (Equation 4.2), Chi-square alternative (Equation 4.3), Correlation (Equation 4.4), Hellinger (Equation 4.5), Intersection (Equation 4.6) and lastly the Kullback-Leibler divergence (Equation 4.7).

$$\text{Chi-square} = d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)} \tag{4.2}$$

$$\text{Chi-square alternative} = d(H_1, H_2) = 2 \cdot \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I) + H_2(I)} \tag{4.3}$$

$$\text{Correlation} = d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H_1})(H_2(I) - \bar{H_2})}{\sqrt{\sum_I (H_1(I) - \bar{H_1})^2 \sum_I (H_2(I) - \bar{H_2})^2}} \tag{4.4}$$

$$\text{Hellinger} = d(H_1, H_2) = \sqrt{1 \frac{-1}{\sqrt{H_1 H_2 N^2}} \sum_I \sqrt{H_1(I) \cdot H_2(I)}} \tag{4.5}$$

$$\text{Intersection} = d(H_1, H_2) = \sum_I min(H_1(I), H_2(I)) \tag{4.6}$$

$$\text{Kullback-Leibler divergence} = d(H_1, H_2) = \sum_I H_1(I) log\left(\frac{H_1(I)}{H_2(I)}\right) \tag{4.7}$$

Where H_1 is the first compared histogram and H_2 is the other histogram, but needs to be the same size as H_1. $\bar{H_k}$ is the mean of the histogram k and lastly N is the number of histogram bins.

Each method was tested with 15 consistency data sets to determine the best performing histogram similarity method, for selecting the correct Siemens star image within all 25 cross-sectional images. The results from the test can be seen in Table 4.1.

Table 4.1: Comparing histogram similarity methods for highest accuracy in the automatic slice selection for the Siemens star.

Comparison method	Accuracy Siemens star (%)
Chi-square	100
Chi-square alternative	93.3
Correlation	100
Hellinger	93.3
Intersection	86.67
Kullback-Leibler divergence	93.3

The histogram distance methods Chi-square and Correlation had an accuracy of 100 % when determining the Siemens star from all the 25 cross-sectional slices. The other

histogram methods had a relative high accuracy as well. Therefore, as not to depend solely on one distance measure, a majority vote of all the different distance measures were created. The majority vote enabled that at least two methods should determine the same cross-sectional image as the Siemens star. If the vote was a tie and thereby inconclusive, the detection of Chi-square was selected.

After the method for comparing the histograms was determined, the histogram of the reference image of the Siemens star was compared to all the cross-sectional images histograms. The histograms with highest similarity for the Siemens star can be seen in Figure 4.9.

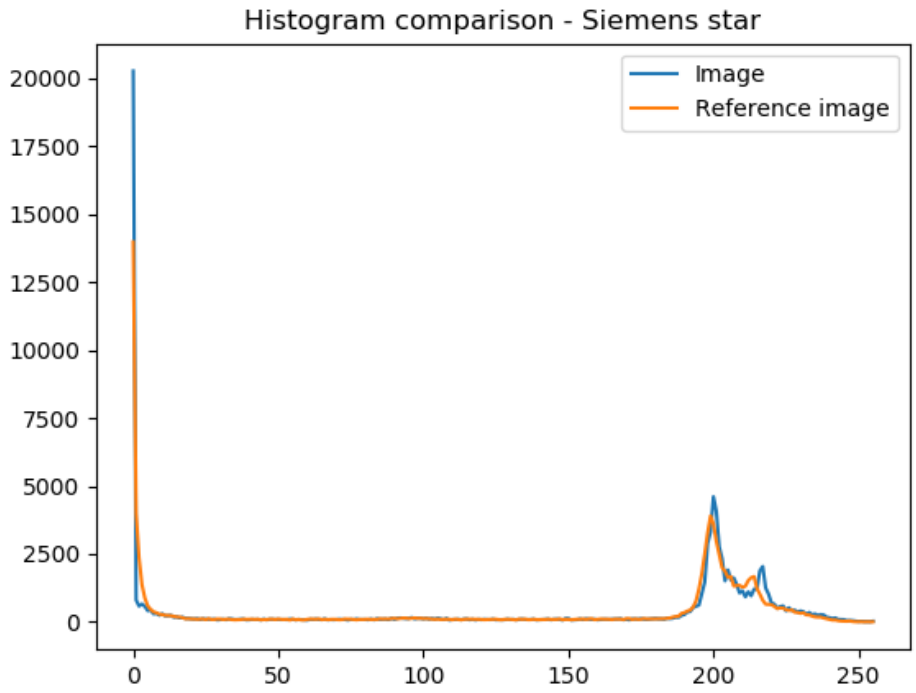


Figure 4.9: Comparing histogram of reference image and image of the Siemens star.

Slice Selection

The image shown on 4.10 is the found result of using the histogram similarity method with majority vote and the reference image of the Siemens star.

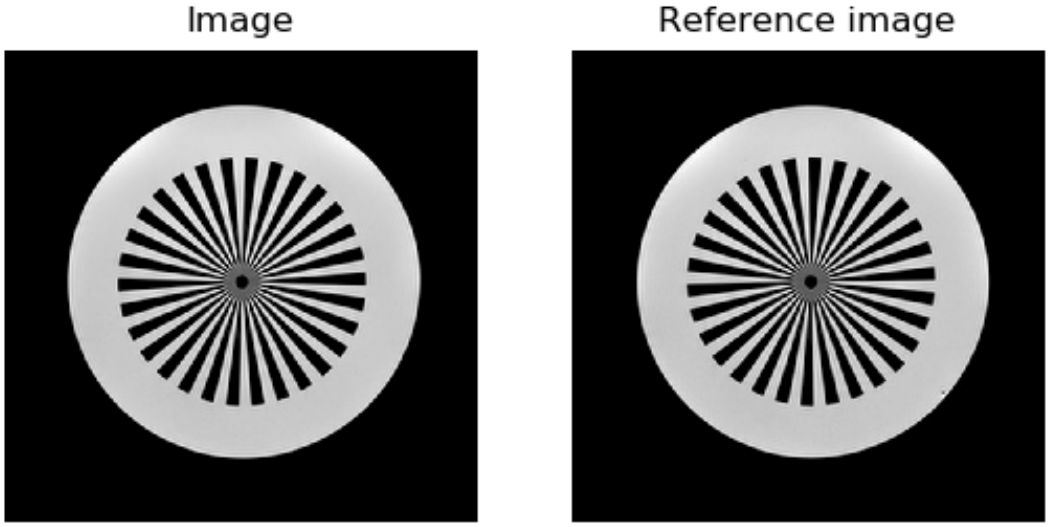


Figure 4.10: Image similarity of the Siemens star.

4.2.3 Block 3: Quantify Image Sharpness

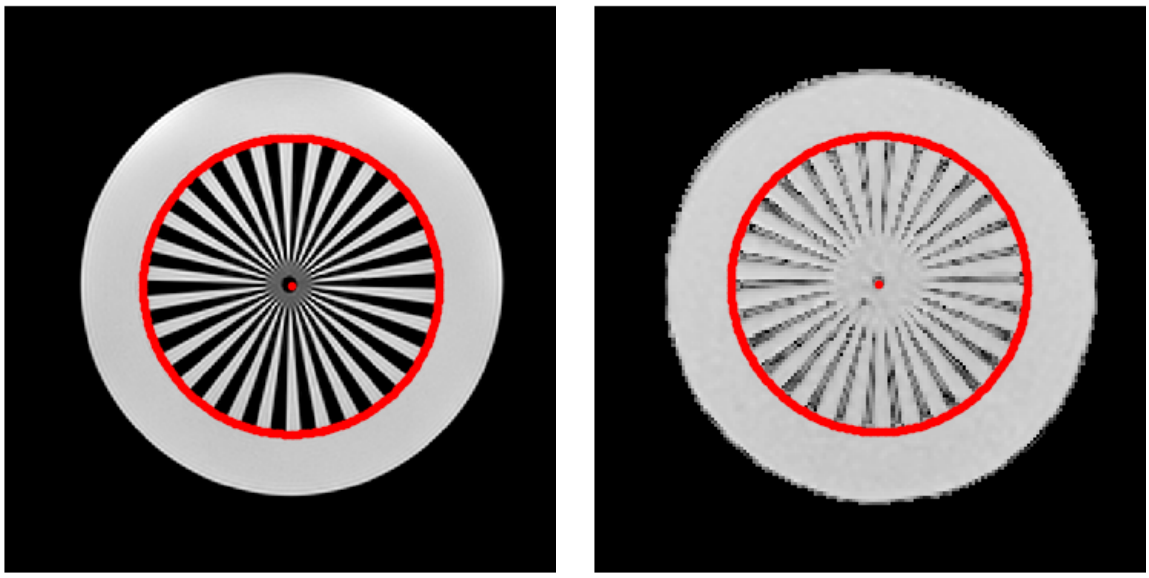
This section provides information concerning the quantification of image sharpness of the MRI scanner. Manual assessment of the image sharpness can be time consuming and an automatic assessment of the sharpness of the MR scanner and the quality of the DW images is not yet available. Therefore the purpose of this section is to explain how the image sharpness was automatically quantified. The image sharpness was based on both the T2 weighted image and the ADC map of the Siemens star.

Weiß et al. [31] investigated how the sharpness of CT scans could be examined with the Siemens star in a phantom suitable for X-ray CT. According to Weiß et al., the Siemens star can be examined with different methods: (1) Visual assessment where the quality of the contrast bars are investigated. (2) A slanted-edge modulation transfer function (MTF) method, which extracts the edge spread function (ESF) of the contrast bars. The ESF is used to calculate the line spread function that with a Fourier transform enables the calculation of the MTF. The MTF investigates the contrast of the edges over the frequency spectrum of the image. In order to obtain a single quantification of the image sharpness, the frequency at which the MTF image contrast has dropped to 10 % is specified. (3) And last the image sharpness is evaluated using the structure width of the contrast bars as a function of image contrast. As explained in Section 2.3.1, the further towards the center the differentiation between black and white is visible, the better sharpness of the image. This method was implemented in order to

obtain a quantification of image sharpness in milimeters (mm). The purpose of this method is to determine at which width of the contrast bars towards the center, there is close to no image contrast. Meaning, determine at which minimum line width in the image the scanner can reproduce details.

Determine Center and Circumference

To ultimately determine the sharpness of the Siemens star, the center and circumference of the Siemens star were found with Hough circle, as explained in Block 2 (Section 4.2.2). In order to obtain the outer perimeter of the contrast bars the radius was set to be 70 % of the radius found by the Hough circle transform, and hereby only the Siemens star was extracted. The center and outer perimeter of the contrast bars found in the T2 weighted image and the ADC map are shown in Figure 4.11.

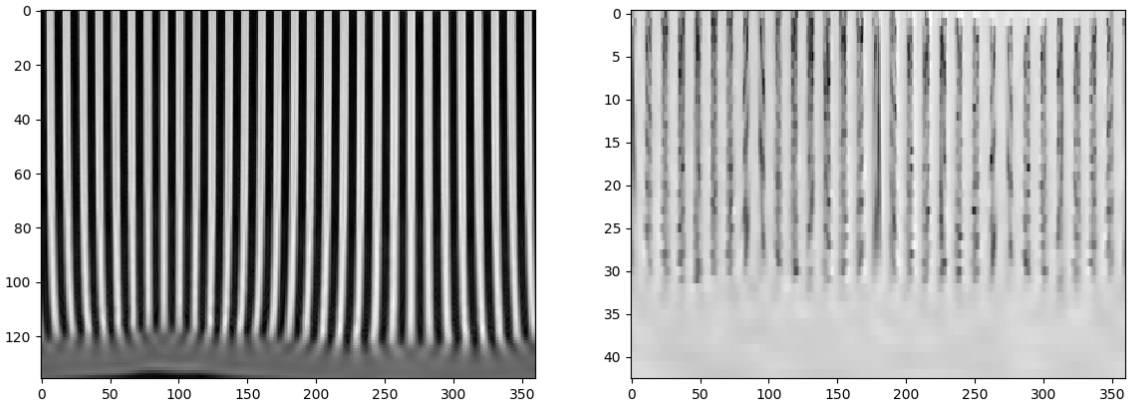


(a) T2 weighted image. **(b)** ADC map.

Figure 4.11: The center of the Siemens star for the T2 weighted image and the ADC map is found and the circumference for the outer perimeter of the contrast bars was calculated.

Polar Coordinate System

To evaluate the information in the Siemens star a polar coordinate system was created. A polar coordinate system is a graphical explanation of a circle, the bottom of the graphical visualization is the middle of the circle, while the top is the outer radius defined beforehand, see Figure 4.12.



(a) Polar coordinate system of the T2 weighted image.

(b) Polar coordinate system of the ADC map.

Figure 4.12: Polar coordinate system of the T2 weighted Siemens star and the ADC map. The width of the image is the 360° around the Siemens star and the height is the pixels from the outer perimeter of the Siemens star to the center.

The polar coordinate systems, as seen in Figure 4.12, the angles around the circle from 0° to 359° are placed in the x-direction. In the y-direction a number of methods can be used, as an example Weiß et al. [31] converted the Siemens star to a polar coordinate system by using the width of the image from inside the circle and outwards. In this project the polar coordinate system was created by profile lines that extract intensity values from the angles around the Siemens star from the center to the circumference. This was completed using the parametric equation of a circle to examine the Siemens star, see equation 4.8 and 4.9. [32]

$$x = a + R \cdot cos(\theta) \tag{4.8}$$

$$y = b + R \cdot sin(\theta) \tag{4.9}$$

Where θ is the angle, R is the radius, a and b defines the center of the circle and x and y defines the circumference.

Sliding Window

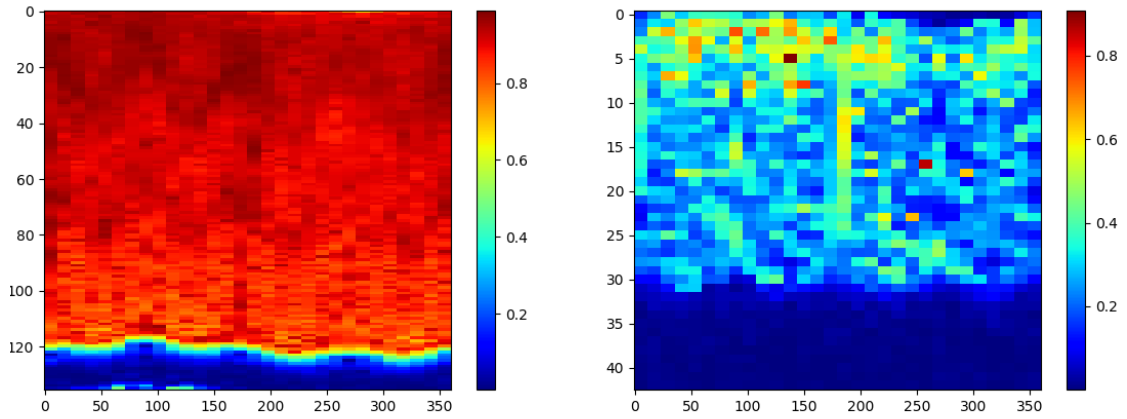
To measure the image contrast locally, the contrast was calculated within sliding windows on the polar coordinate system. As explained in Section 2.3.1, each contrast bar have an opening angle of 6° and therefore the size of the sliding windows needs to be twice that size in the x-direction to contain both the white and black contrast bar in the same area of the sliding window. Thus, the height and width of the sliding windows were determined to be respectively one row and 3.3 % of the image size. The height of one row was selected in order to achieve the maximum number of quantifications in the height of the polar coordinate system. The sliding windows were chosen not to overlap, as to calculate a contrast one window size after another. The image contrast was calculated within each sliding window.

Michelson Contrast

The image contrast was calculated with the Michelson contrast formula, see Equation 4.10. [31]

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{4.10}$$

Where C is the Contrast, I_{max} is the maximum intensity within the window and I_{min} is the minimum value. Michelson contrast was chosen, as it is used for images with repeating patterns, as seen on the polar coordinate system. The Michelson contrast was calculated within each sliding window, providing a plot over all sliding windows with the Michelson contrast. Each window was quantified with a value between 0 and 1, where 0 explains no image contrast and 1 explains high image contrast. To visualize the difference a colormap was applied to the plot, where blue is 0 and red is 1. The image with the Michelson Contrast equation applied can be seen on Figure 4.13, where Figure 4.13a shows the image contrast of the T2 weighted image and Figure 4.13b shows the image contrast of the ADC map.



(a) Michelson Contrast calculation of the polar coordinate system of the T2 weighted image.

(b) Michelson Contrast calculation of the polar coordinate system of the ADC map.

Figure 4.13: Polar coordinate system of the T2 weighted Siemens star and the ADC map.

In order to obtain a single value for quantification of the sharpness of the Siemens star, the minimum detectable line width of the contrast bars was determined. It is known from Section 2.3.1, that the Siemens star has a diameter of 120 mm and the opening angle of each spoke is 6°. 3 mm was subtracted from the radius of 60 mm, as the black center of the Siemens star was removed, see Figure 4.14.

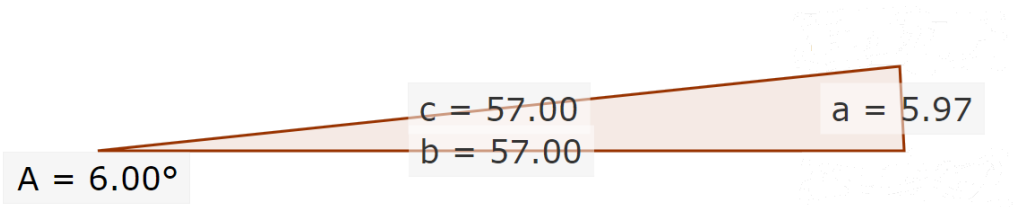
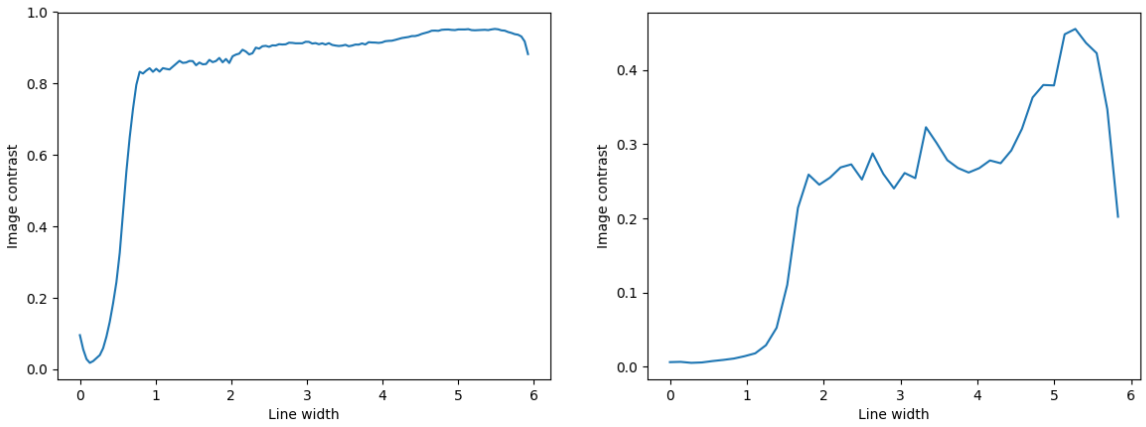


Figure 4.14: Illustration of a contrast bar of the Siemens star. A is defined as the opening angle of 6°, c and b is defined as the length of the bar (and the radius of the circle) of 57 mm and a is defined as the largest width of the contrast bar of 5.97 mm.

Where A is the opening angle of 6°, c and b are defined as the length of the bar of 57 mm and a is defined as the largest width of the contrast bar. With this information the largest line width of the contrast bars can be calculated with Equation 4.11. [33]

$$a = \sqrt{b^2 + c^2 - 2bc \cdot cos(A)} \tag{4.11}$$

The calculation of the line width of the contrast bars enabled the investigation of the image contrast as a function of structure width, see Figure 4.15.



(a) T2 weighted image. The image contrast degrades at a line width at approximately 0.5 mm.

(b) ADC map. The image contrast degrades at a line width at approximately 1.5 mm.

Figure 4.15: Image contrast as a function of line width for the T2 weighted image and the ADC map.

In order to specify at which radius the minimum detectable line width was measured, an image contrast that was degraded to near zero needed to be defined. As seen in Figure 4.15a, the declining slope of image contrast ends approximately at 10 % of image contrast. Therefore the threshold for minimum detectable line width was set to 10 % of image contrast for the T2 weighted image. As seen in Figure 4.15b, the image contrast had a plateau immediately after 5 % of image contrast. Therefore the selected threshold in image contrast for the ADC map was 5 % of image contrast.

The selected radius for T2 and ADC map was used in Equation 4.11 for b and c to calculate the minimum detectable line width. This minimum detectable line width was therefore used to quantify sharpness in the image and to determine at which minimum line width the scanner could reproduce details in the images. As the maximum line width is 5.97 mm, a minimum line width ratio was calculated with Equation 4.12.

$$c = \frac{a - b}{a} \tag{4.12}$$

Where c is the minimum line width ratio, a is the maximum line width of 5.97 mm and b is the measured minimum line width.

Besides the minimum line width, a calculation of the mean contrast of the Siemens star was conducted, in order to quantify an overall image contrast. Therefore a mean image contrast was calculated of the Michelson plot (Figure 4.13). In order to obtain an overall quantification of the image sharpness, a mean value of the minimum line width ratio and the mean image contrast was calculated.

Finally, three different values of the Siemens star were obtained. A minimum line width in mm that gave a quantification of the ability to reproduce fine details in the image. A fine detail is defined as the size of a single pixel in the image, which is 0.411 and 1.307 mm for respectively T2 weighted images and the ADC maps of the phantom. A mean image contrast of the Siemens star, that gave a quantification of the overall image contrast. And a mean value of the two, that gave a single overall quantification of the image sharpness.

4.2.4 Block 4: Detect the ADC Ramp Insert

This section provides information about the methods for automatically detecting the ADC ramp insert from the other slices in the phantom. Similar to Block 2 in Section 4.2.2, the ADC ramp insert had to be automatically detected within the 25 cross-sectional images of the phantom. The T2 weighted scanning sequence included high quality images of the phantom and was therefore used for image similarity measures. A reference image for the ADC ramp insert, according to Section 2.3.1, was selected for the image similarity.

Determine Center and Circumference

The center was found as in Block 2 (Section 4.2.2) by using Hough circle transform. An iteration was implemented to find the maximum circle in the area for each iteration by changing the minimum radius. This gave the outer contour of the out most circle. The found center and circumference can be seen in Figure 4.16

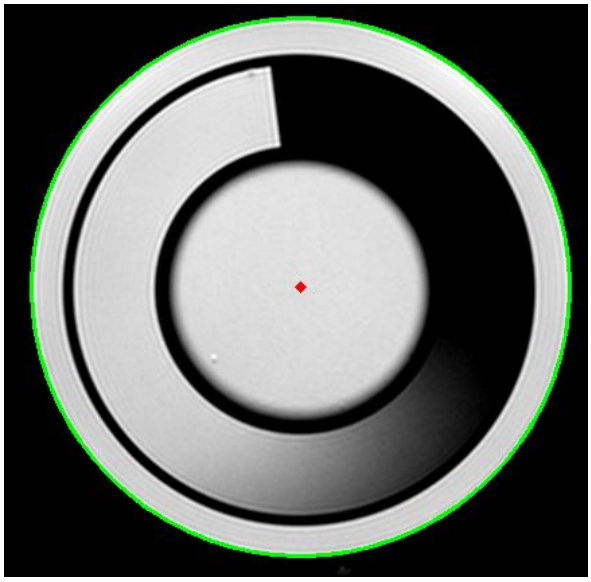


Figure 4.16: Hough circle transform on a T2 weighted image of the ADC ramp insert. The found center and circumference are marked in red and green, respectively.

Image Similarity

Similar to Section 4.2.2 there are slight differences in rotation of the phantom and therefore the histogram similarity method for the image similarity was used. The mask used for this part of the phantom was between 85 % and 65 % of the radius, where the intensities are extracted. This ensured only the data of the ADC ramp insert was extracted and used for histogram similarity with the reference image of the ADC ramp insert. The two circles defined by the two radius where the region in between was extracted can be seen in Figure 4.17.

Figure 4.17: The two red circles illustrates the region which was extracted for the histogram similarity.

Image intensities were extracted with the mask for all the 25 cross-sectional images and the reference image of the ADC ramp insert. For every iteration histograms were calculated and compared in order to detect the ADC ramp insert of the cross-sectional images. The histograms distance measures were compared to determine the best performing method for image similarity. The six histogram correlation methods explained in Section 4.2.2 was used to find the ADC ramp insert on 15 consistency data sets. The results from the correlation method can be seen in Table 4.2.

Table 4.2: Comparing histogram correlation methods for highest accuracy in the automatic slice selection.

Comparison method	Accuracy ADC ramp insert (%)
Chi-square	73.3
Chi-square alternative	86.6
Correlation	60
Hellinger	86.6
Intersection	93.3
Kullback - Leibler divergence	86.6

The results from the test showed a highest accuracy of the algorithm of 93.3 % with intersection, which was not sufficient. A limitation of the histogram similarity methods is that they ignore rotation in the image. Therefore the histogram method was prone to make errors, when only basing the detection on the distribution of intensities. The criteria for finding the ADC ramp insert is the change from high to low intensities around the ring. Therefore an array of intensities around the ring, with 70 % of the radius to obtain values in the middle of the ring, was extracted for both the cross-sectional images and the reference image. The similarity of these arrays was then investigated with a distance error. However due to rotation of the phantom, the arrays had to be aligned in order to calculate the difference between corresponding pixels. The Procrustes method is an optimization method for curve-fit which uses scaling, translation and rotation. It was used to minimize the sum of squares differences between landmarks in the two intensity arrays, by Equation 4.13. [34]

$$M^2 = \sum (Data1 - Data2)^2 \tag{4.13}$$

Where M is the sum of squares, Data1 is the intensities within the reference image and Data2 is the intensities within the cross-sectional image, where the scaling, translation and rotation are performed.

If the disparity found by the minimizing of sum of squares differences were under 0.3, an additional nine iteration were performed from 71 to 80 % of the radius found by the Hough circles. The iterations were performed to give consistency to the algorithm so no incorrect slices were found. The automatic slice detection was 100 % accurate when using the Procrustes method.

Slice Selection

The reference image and the detected image with the Procrustes method can be seen in Figure 4.18.

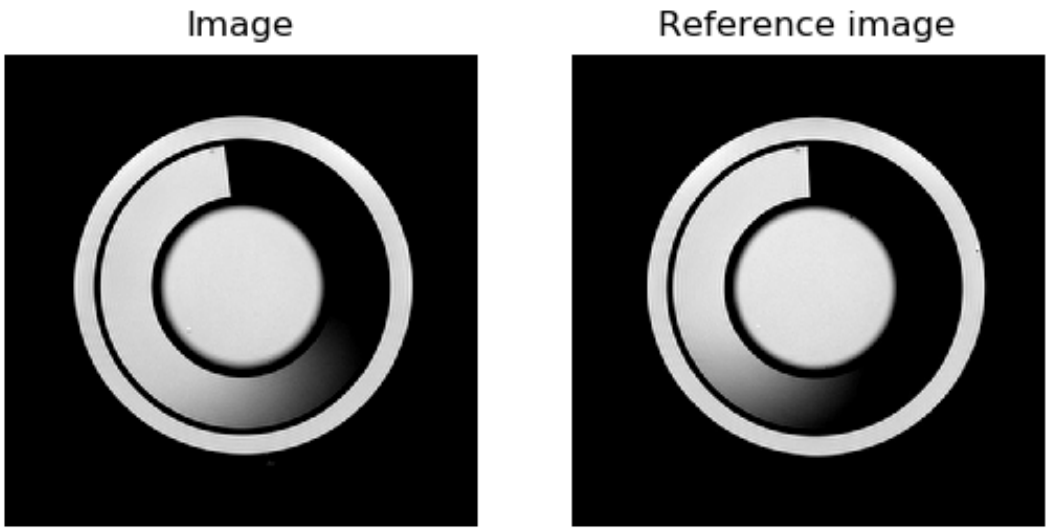


Figure 4.18: Image similarity of the ADC ramp insert.

4.2.5 Block 5: Quantify ADC Region

To quantify the ADC region of the ADC ramp insert, the right part of the ADC ramp insert needs to be found. First of all, the center and circumference of the phantom were found in Block 4 in Section 4.2.4. Afterwards the coordinates were determined, where the biggest difference in intensities was present within the ADC ramp. When the difference was found a step of 5 % in the direction of the maximum intensities was performed to extract intensities from a ROI, containing the highest intensity values.

Difference

The radius and circumference of the slice with the ADC ramp were found in Block 4 in Section 4.2.4. A percentage of the radius of the circumference was used to find the intensities within the ADC ramp insert, where the ADC benign region is present. The percentage were set to be 70 % of the radius, so the circle were in the middle of the

ADC ramp insert in the phantom. The circle at which the intensities was extracted, is shown in Figure 4.19.

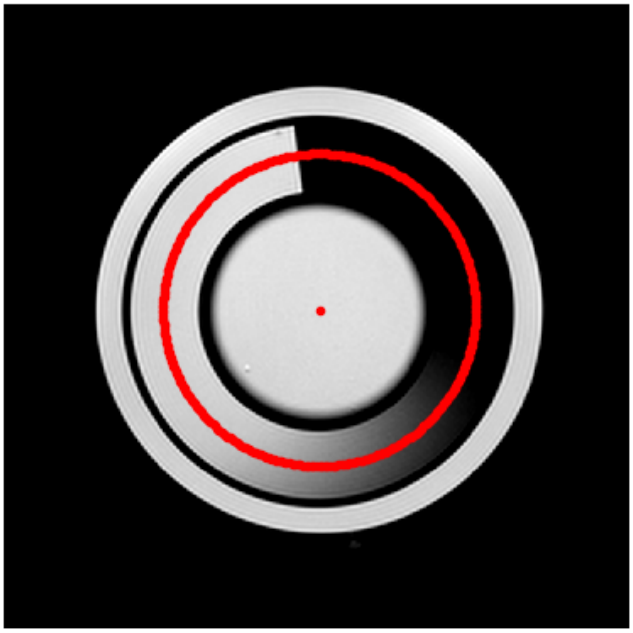


Figure 4.19: The center of the ADC ramp insert for the T2 weighted image is found and the coordinates for the circumference is calculated.

The difference in intensity values that was extracted around the ADC ramp insert is shown in Figure 4.20.

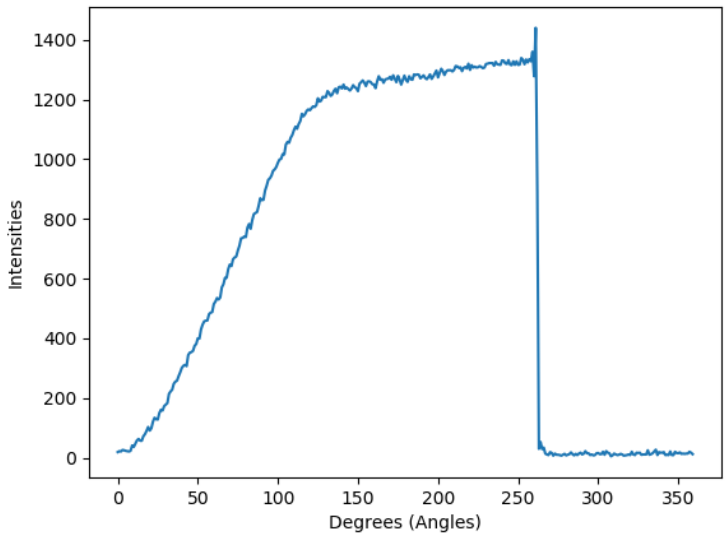


Figure 4.20: The intensities around the ADC ramp insert in the phantom.

The angle of the maximum difference shown in Figure 4.20 was determined in order to determine the placement of the ROI in the ADC region. A ROI was therefore defined to extract image data of the ADC benign region. The ROI was selected to be a circle in the ADC region, therefore a center of the circle was determined, see Figure 4.21.

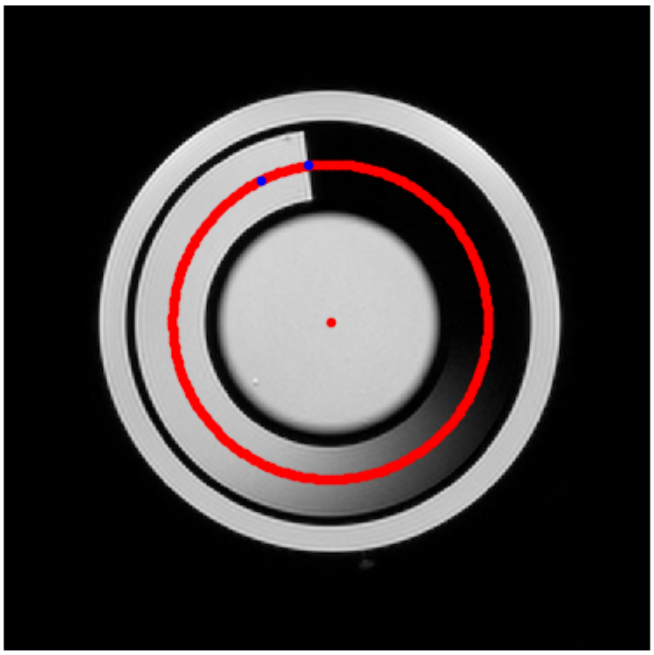


Figure 4.21: The center of area where intensities were extracted.

As the ADC ramp insert may be flipped, due to the orientation of the phantom when scanned, the area of the ADC region was needed to be determined. Therefore 5 % of the angle of the maximum difference was calculated in the direction towards the maximum intensity values. This way a ROI in the ADC region was obtained. The circle of the ROI was defined with a radius of 20 pixels to obtain a ROI within the ADC region, see Figure 4.22. The radius was chosen so the correct area were chosen even if small differences in the size or placement of the Hough Circle were present.

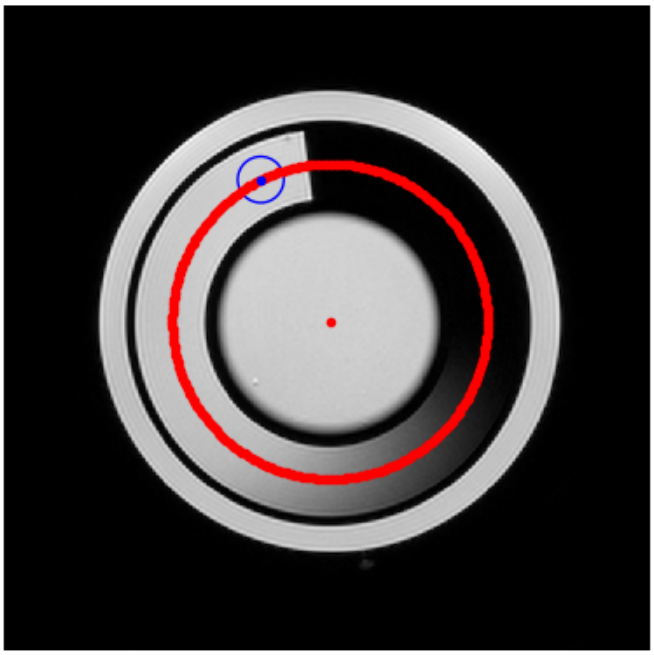


Figure 4.22: The ROI where intensities were extracted.

Image Masking

In order to obtain the intensity values of the ROI shown in Figure 4.22 a mask was calculated. The mask provided the coordinates of the ROI in the image. These coordinates were used to extract the intensities of the ROI in the original image. The mask used to extract the intensities of the ROI is shown in Figure 4.23.

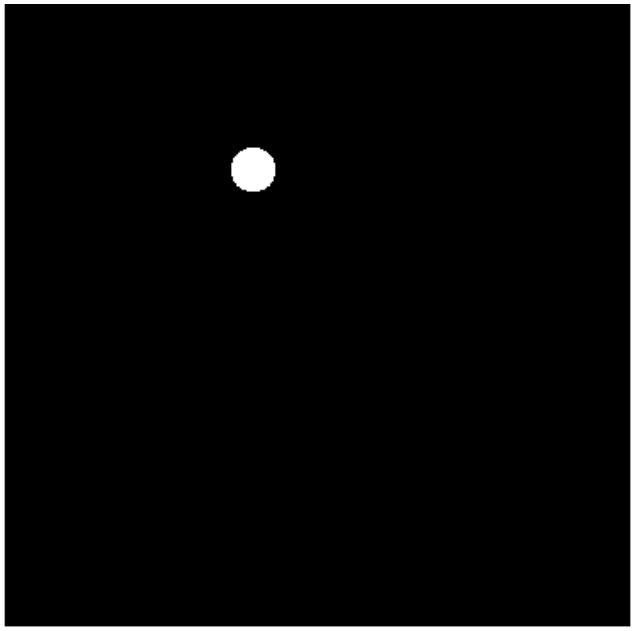


Figure 4.23: The image mask from the ROI in the ADC benign region.

The mask was implemented to extract intensities of the ADC benign region. The ROI from the ADC benign region is shown in the T2 weighted image in Figure 4.24.

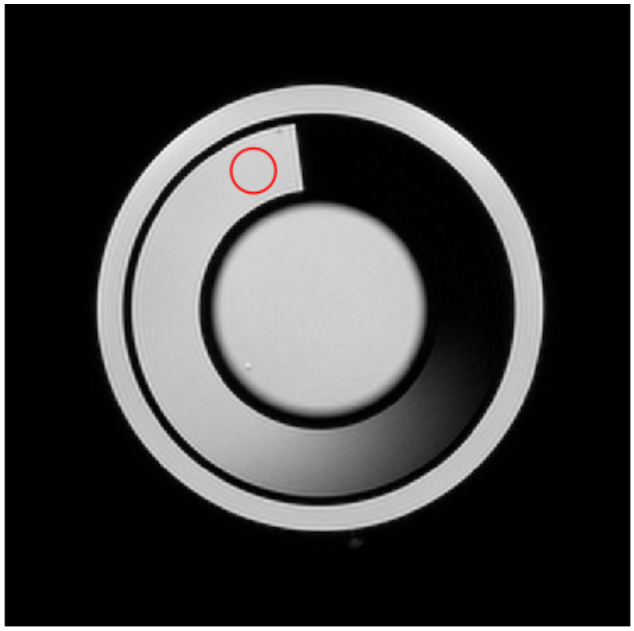


Figure 4.24: ROI in the ADC region of the T2 weighted image.

4.2.6 Block 6: Visual Presentation of Quality Assurance

According to Section 2.3, it is mandatory to perform QA on MR scanners in order to document consistently high-quality DW images. High-quality images are important in a clinical practice when monitoring the disease progression of a patient. A variance in the measured ADC may occur from day to day, therefore this variance needs to be examined in order to know that the measured variance is not a progression in the patients disease. Moreover QA of the sharpness of the DW images are needed, in order investigate the ability of the scanner to reproduce fine details in the images. The day to day variance in sharpness of the scanner is important to examine in order to investigate if the scanners ability to reproduce fine details has changed from one day to another. Lastly, it is important to examine the variance of the ADC measurements between scanners. When monitoring the disease of a patient the patient may be scanned on different MR scanners. Therefore the variance between scanners needs to be examined, in order to differentiate if the variance between scans of the ADC measurements, is a disease progression or a variance between scanners. Thus, the purpose of this section is to describe how the visual presentation of the QA was presented for the user of the algorithm.

Tab for overview

The overview tab of the GUI was developed to show the radiologist at the Department of Radiology the current daily values for each MR scanner, see Figure 4.25.

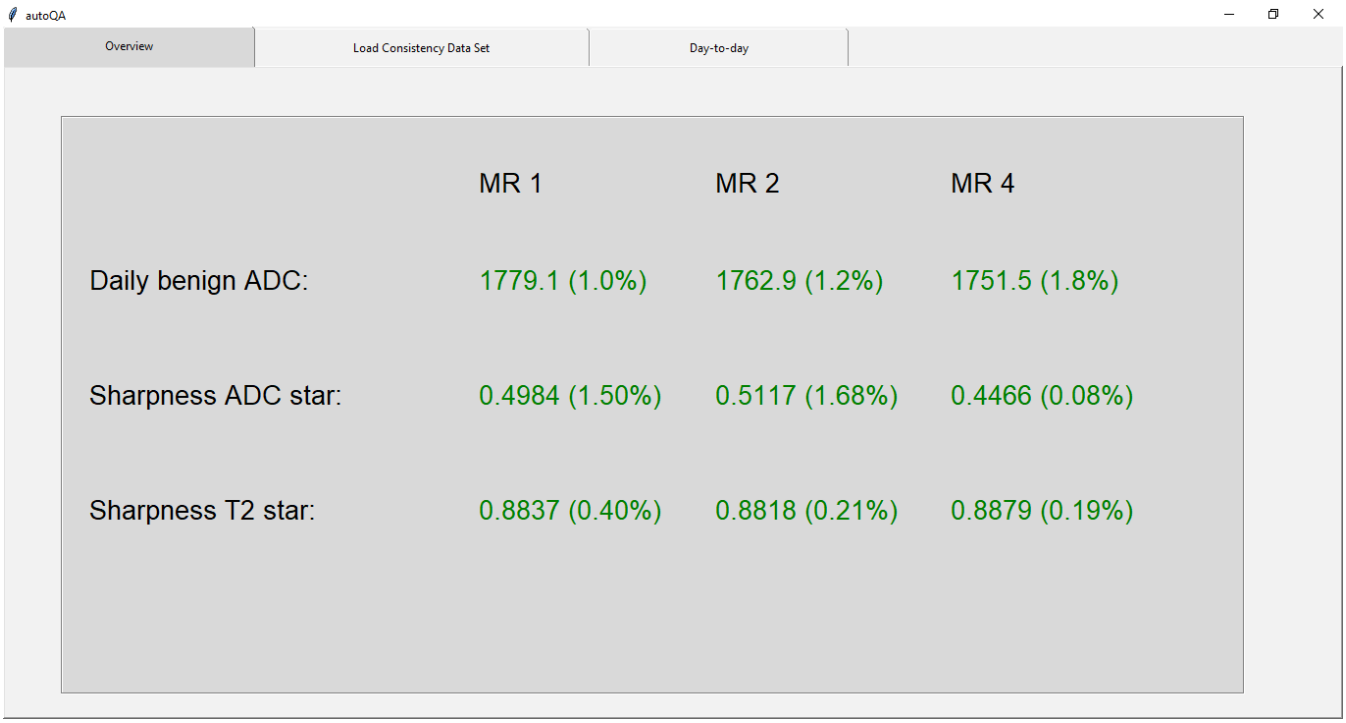


Figure 4.25: First tab for an overview of latest consistency data set compared to the general mean of all consistency data sets.

The daily values are defined as the ADC measurements and the image sharpness, including a difference in percentage from the mean of all consistency data sets. The daily values of all three scanners are presented from the latest performed consistency data set. The "Daily benign ADC" shows the results from the ROI in the ADC sequence in the ADC benign region. The "Sharpness T2 star" label shows the sharpness in the Siemens star for the T2 weighted image, while "Sharpness ADC star" shows the sharpness from the ADC sequence. Each value is featured with an explanation, for an example the label "Daily value benign" have a text box embedded showing "ADC of the latest consistency data set and the difference in percentage from the mean of all tests", when the mouse is hovering over the value. For the values of the sharpness the text illustrated in a box is "Sharpness of the latest consistency data set (0 is worst and 1 is best) and the difference in percentage from the mean of all tests".

Tab for Day-to-day Variance

The day-to-day tab was developed as a QA to visually present the day to day variance of the data options and the variance between scanners. Data options were defined

as the values from the ADC benign region and the image sharpness found from the Siemens star. These data options can be chosen on the tab to be illustrated in the graph. The benign values can be chosen separately for all four scanning sequences respectively b0, b1600, ADC map and T2. On the tab illustrated in Figure 4.26, the data options can be chosen by selecting the check boxes "Benign b0", "Benign b1600", "Benign ADC" and "Benign T2", when the check boxes are selected they can be plotted by pressing the button "Plot". The image sharpness can be selected for the two sequences, the T2 weighted sequence and the ADC map, by choosing the check boxes "T2 star" and "ADC star".

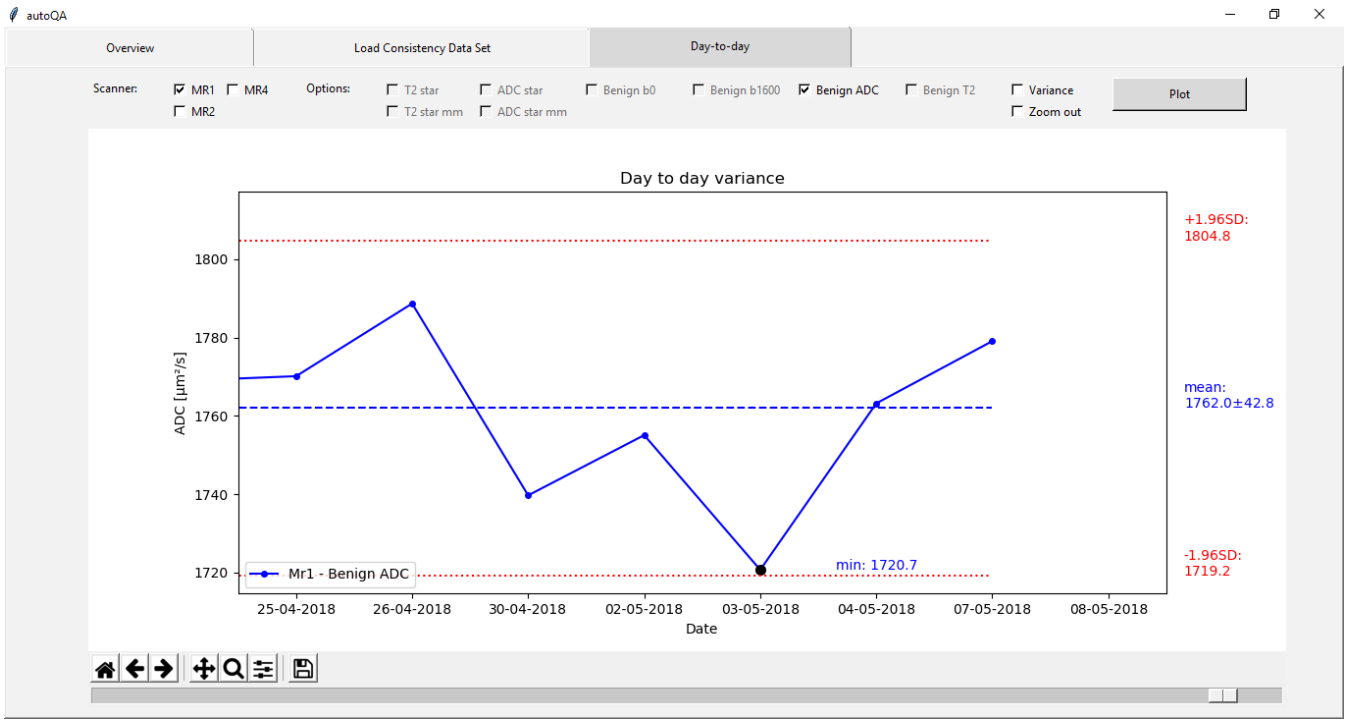


Figure 4.26: The tab "Day-to-day" gives the visual presentation of the latest consistency, but can be scroll back to all the consistency data sets performed. The day-to-day graph can show all the results from the consistency data sets and show the difference between the scanners.

Furthermore it should be possible to see the differences between scanners, hereby multiple scanners can be loaded at the same time so the difference between the scanners can be seen, by selecting the check boxes "MR1", "MR2" and "MR4", which are the three scanners currently located in the Department of Radiology. When only one scanner at the time is chosen the 1.96 SD are shown in the plot. The SD is removed when multiple scanners are chosen so it gives a better overview without being crowded, which is illustrated in Figure 4.27. If a point is outside the SD, the mark is colored red,

this function works whether only one or multiple scanners are chosen, so the outliers can be noticed.

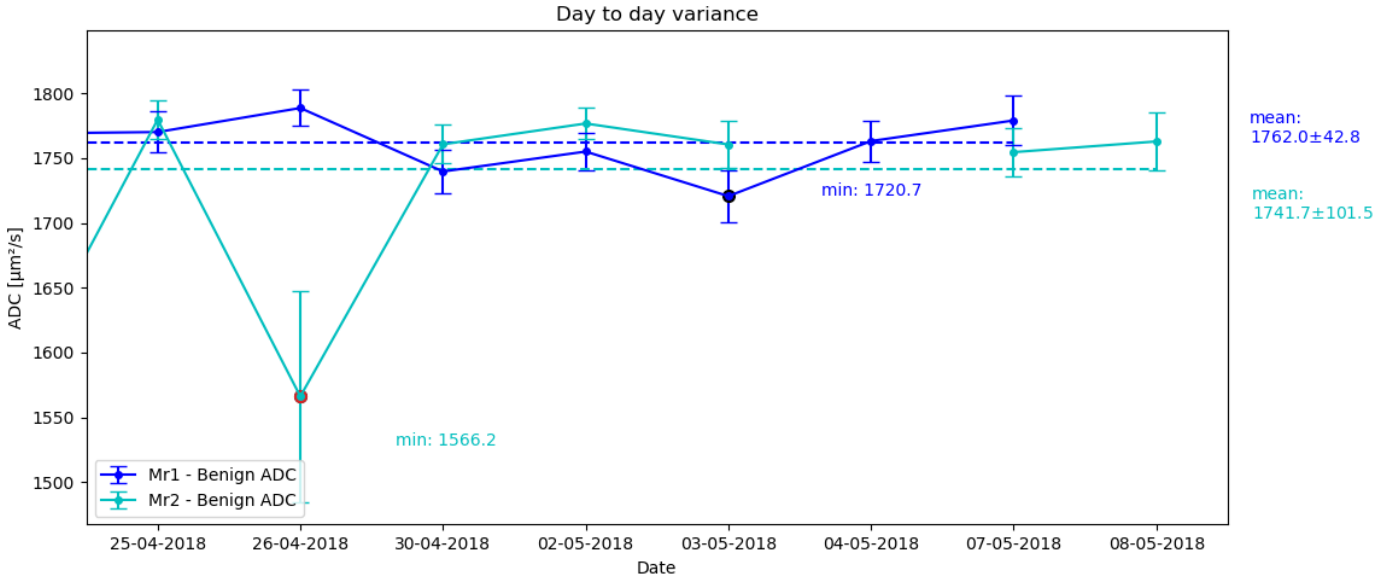


Figure 4.27: The day-to-day graph with the two scanners "MR1" and "MR2" selected, the SD was hereby removed to make the graph less crowded. The check box "Variance" was chosen and two outliers were marked with red.

The check box "Variance" can be chosen and will show the variance within the ROI of the currently chosen sequence of the ADC benign region. The variance is also illustrated in Figure 4.27

4.3 Algorithm Testing

Algorithm testing is required in order to evaluate the performance of the algorithm. This section will therefore describe how each of the objectives were tested. The algorithm was tested on 30 consistency data sets for MR1 and MR4 and 29 for MR2, providing a total of 89 consistency data sets.

4.3.1 Automatic Slice Detection

As each scanning sequence consisted of 25 cross sectional DICOM images, both the ADC ramp insert and the Siemens star should be automatically detected. The refer-

ence image of the ADC ramp insert is shown in Section 2.3.1 in Figure 2.7 and the reference image of the Siemens star in Figure 2.8.

In order to determine the accuracy of the algorithm to detect the cross-sectional images, which corresponds to the two reference images, image quality assessment (IQA) was used. IQA is a method that aims to efficiently evaluate the quality of images. [35]

The subjects used in the IQA were chosen to be the two authors. Different specific criteria of the reference images were selected when determining if the algorithm had chosen the correct image. The criteria of the Siemens star were:

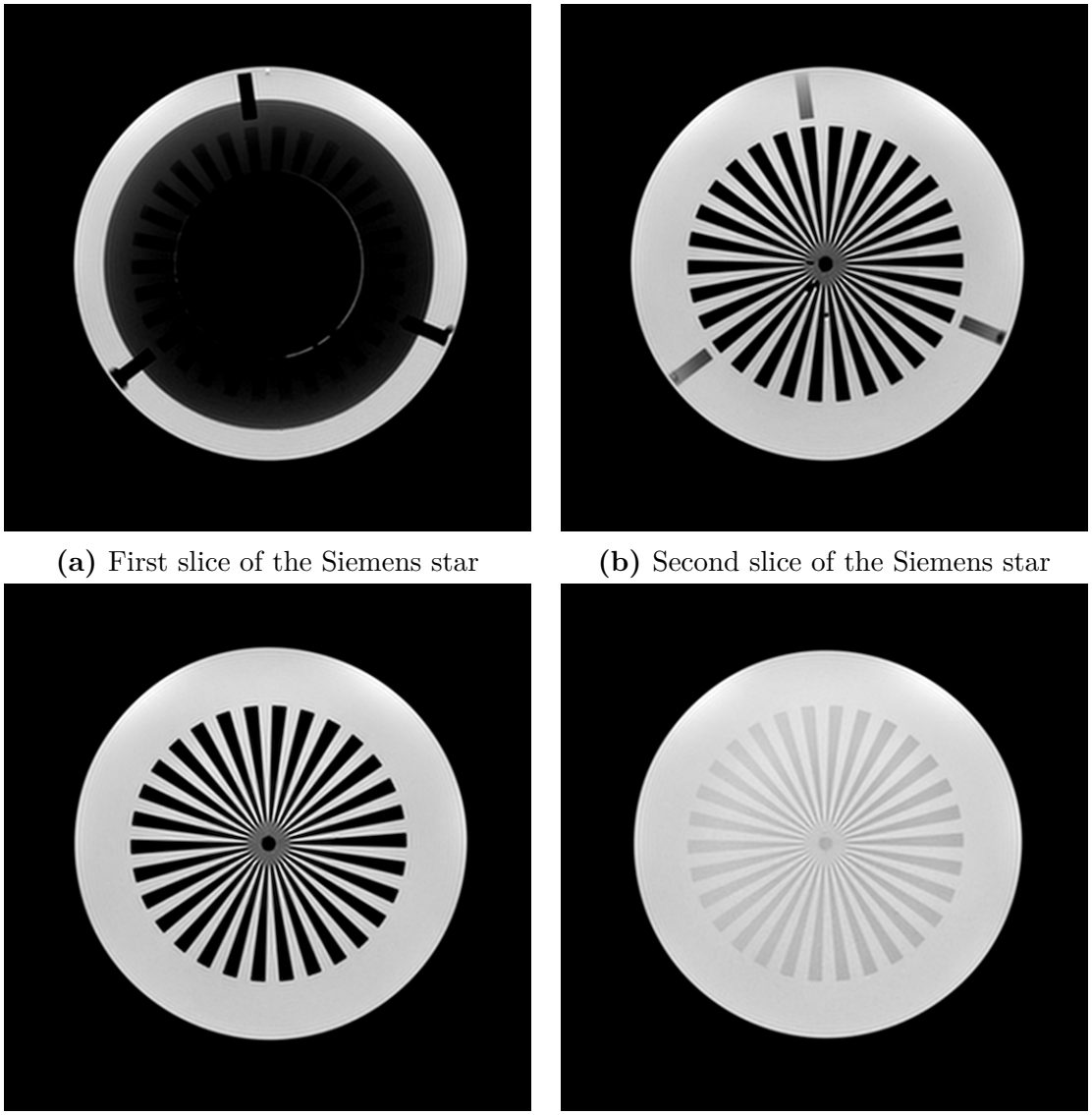
1. A high contrast between the black and white contrast bars around the circle.
2. Minor quantities of black spots between the contrast bars.

The criteria for the ADC ramp insert were:

1. A high contrast between black and white in the ADC ramp insert.
2. The gray area of the image was opposite of the high difference in contrast between black and white.

The detected image of the algorithm was evaluated as correct or wrong, and the resulting mean detection accuracy was calculated based on all consistency data sets for both the ADC ramp insert and the Siemens star.

For an example of IQA, four images from each of the two sections in the phantom were chosen to illustrate the differences and the choices for accepting or not accepting the slices. The images from the Siemens star can be seen in Figure 4.28.



(a) First slice of the Siemens star

(b) Second slice of the Siemens star

(c) Third slice of the Siemens star

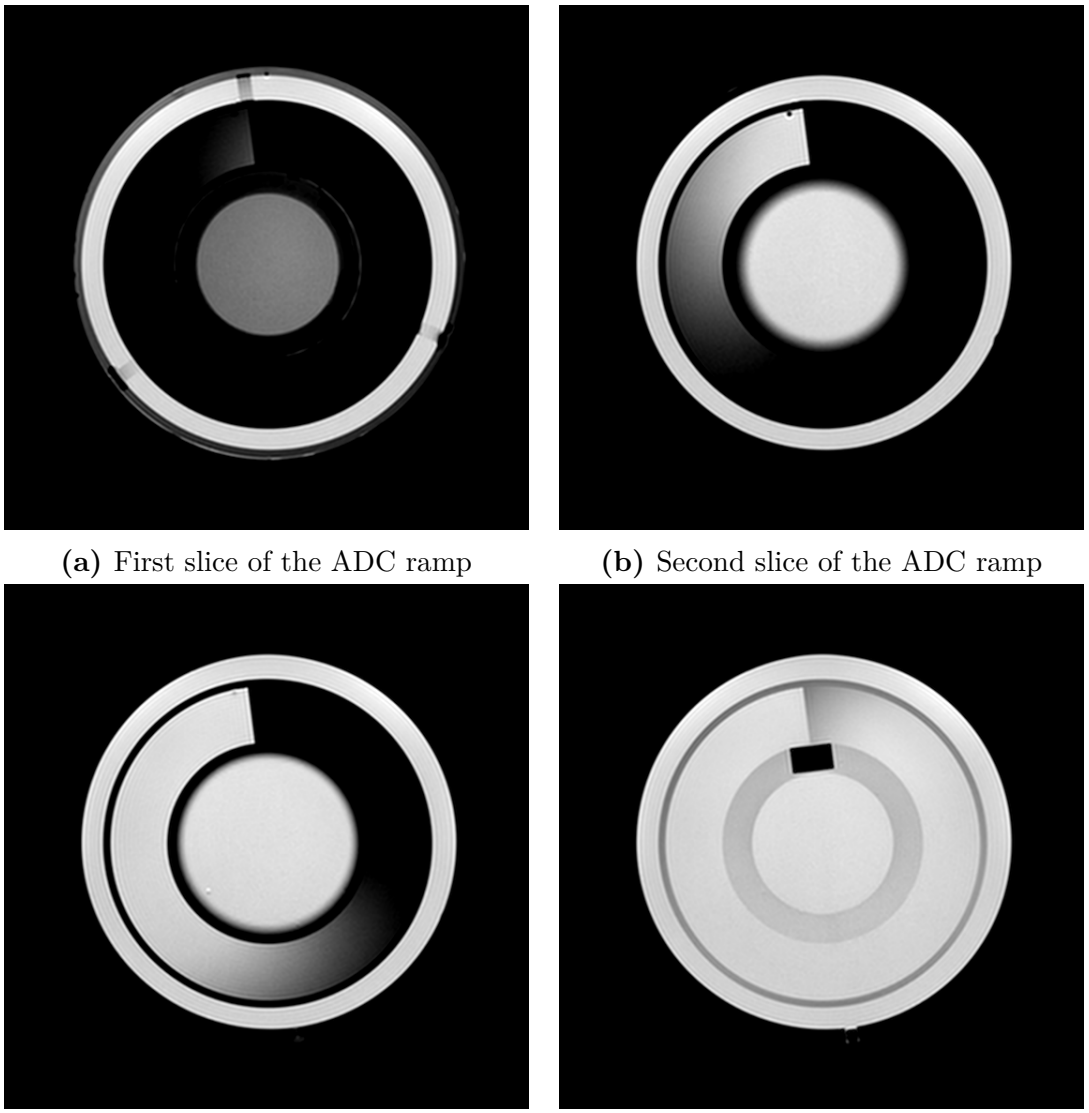
(d) Fourth slice of the Siemens star

Figure 4.28: The slices of the Siemens star where the correct slices are the second and third slice.

IQA was used with the criteria to qualitatively select the ideal cross-sectional image of the Siemens star. The first slice seen in Figure 4.28a is the Siemens star barely visible and would therefore not be chosen as an acceptable image. Slice 2 seen in Figure 4.28b would be accepted, because there is a low quantity of black spots and the Siemens star is full visible. Figure 4.28c shows the ideal image as slice 3 and would therefore be accepted. Lastly Figure 4.28d shows the fourth slice which would not be accepted

because the Siemens star contains a low difference in intensities between the contrast bars.

Next the slices from the ADC ramp insert were evaluated, the four slices chosen for illustration can be seen in Figure 4.29.



(a) First slice of the ADC ramp

(b) Second slice of the ADC ramp

(c) Third slice of the ADC ramp

(d) Fourth slice of the ADC ramp

Figure 4.29: The slices of the ADC ramp, where the correct slice is the third slice.

IQA was used with the criteria chosen to qualitatively select the ideal cross-sectional

image of the ADC ramp. The first slice seen in Figure 4.29a there are no white area and is therefore not chosen to be acceptable. Figure 4.29b shows the second slice where more white is present, but the values within are still affected of the transition from black to white intensities, therefore this slice is also not accepted. The third slice seen in Figure 4.29c is the ideal image and would therefore be accepted. Lastly the fourth slice where a transition to the next part of the phantom is in process, this can give varying values, and therefore Figure 4.29d would not be accepted.

4.3.2 Quantification of the ADC Benign Values

The purpose of this test was to investigate the ability of the algorithm to detect the correct ROI for the ADC region and thus quantify the correct ADC values. The quantified mean ADC benign values of the algorithm were compared to the manually quantified ADC values performed by a radiologist. Therefore the quantification by the radiologist was chosen to be ground truth.

The ADC quantification will be visually presented in a plot over all the consistency data sets. The purpose of this is to compare the quantification of the algorithm to the ground truth and to evaluate the change from consistency data set to consistency data set. The ADC benign quantification for the three scanning sequences b0, b1600 and ADC will be evaluated. Moreover a mean ADC difference and a mean difference in percentage for all the consistency data sets will be calculated. However, in order to evaluate the bias between the two ADC quantification, a Bland-Altman plot will be used in the evaluation. A Bland-Altman plot is method for evaluating the bias between the mean differences and is used to estimate limits of agreements (LoA). The agreements are constructed at which 95 % of the differences between the algorithms quantification and the ground truth will fall within. If most of the differences lie outside the limits of agreement, it could indicate that there is no agreement between the methods. The agreements does however not explain whether those agreements are acceptable or not, as acceptable limits must be assigned based on clinical significance. The Bland-Altman plot is plotted on an XY axis, where X is the mean of the two ADC quantifications and Y is defined as the difference between the two ADC quantifications. The limits of agreement are defined as $\pm 1.96SD$ of the mean difference. [36]

4.3.3 Quantification of Sharpness

The purpose of this test was to investigate the ability of the algorithm to detect the ROI of the Siemens star and quantify the sharpness of the image. The quantified sharpness of the algorithm was compared to a sharpness evaluation performed by a

radiologist. The evaluations performed by the radiologist were conducted with IQA, where the Siemens star was quantified with a visual analogue scale (VAS) spanning from 1-5, where 1 is defined as no similarity to the reference image of a Siemens star and 5 is defined as a strong similarity to the reference image. A VAS is a rating system that assists in subjective measurements, and were typically used in a clinical setting. [37] The five scores where defined by the radiologist as:

1. The star could be seen, but the black center was not recognized. The contrast bars were blurred, noisy or with double contours. The gray center was over 4 cm.
2. Black center could be seen. Noisy but delimited contrast bars. Gray center was over 4 cm.
3. Black center could be seen. The contrast bars were seen and delimited. Gray center was around 3-4 cm.
4. Black center could be seen. The contrast bars were well defined. Gray center was around 2-3 cm.
5. Black center was sharp. The contrast bars were well defined. Gray center was around 1-2 cm.

The quantification of the algorithm will be based on three different values as explained in Section 4.2.3. The sharpness of the Siemens star was quantified by a minimum line width, a mean image contrast and a mean of the two. Three T2 weighted images and three low-quality T2 weighted images of the Siemens star will be presented, in order to show differentiating results. As described in Section 2.3, DWI has a low signal to noise ratio, which means that the images may contain artifacts as blurring, ghosting or change in image intensities. Blurring artifacts appear as reduced visual details in the image and ghosting appear as repetitive patterns that are similar to structures in the image. [38] Therefore in order to simulate these artifacts, two of the low-quality T2 weighted images of the Siemens star were created by blurring of the original image of the Siemens star. Blurring was achieved by convolving the image with a low-pass filter, which removes high frequencies in the image such as edges and noise. Another low-quality T2 weighted image was created by changing the intensities of pixels in the image. The quantification of the T2 weighted images will be compared to the quantification of a radiologist of the Department of Radiology. Moreover the quantifications of the algorithm on ADC maps of the Siemens star will be presented.

4.3.4 Day-to-day Variance

The purpose of this test was to visualize the day-to-day variances of the ADC values from the ADC map and test for temperature correlation in the measurements. A difference in ADC values were observed between the 1.5 T scanners and the 3 T scanner, and from an article of Wagner et al. [17] it was found that the ADC values were independent of magnetic field strength. Therefore a calculation of the ADC map was performed with Equation 2.3 from Section 2.2.1 with the DW images from the b0 and b1600 sequences. To test for the variance in the scanner of the ADC map a temperature correction of the ADC values needed to be performed and the difference between the non-corrected and corrected ADC values were addressed. The used method for the temperature correction can be seen in Equation 4.14. [17]

$$ADC_{20C^o} = ADC_{measured} \cdot e^{c_2 \cdot (T_0 - T)} \tag{4.14}$$

Where ADC_{20C^o} is the corrected ADC value at 20 C°, $ADC_{measured}$ is the measured ADC, c_2 is described as the constant for temperature dependence at a specific PVP solution, in this case a value of 0.02453. T is the measured temperature noted from the thermometer in the phantom before and after the scanning and T_0 is 20 C°. [17]

To show the difference between the non-corrected and the corrected ADC value a regression line was calculated with Equation 4.15 to show the increase in ADC value due to temperature of the data. [39]

$$Y_i = \alpha + \beta X_i \tag{4.15}$$

Where α and β are population parameters of the data that the regression line is fitted to. Y_i is the calculated ADC as a function of the temperature X_i. A coefficient of determination, R^2, was calculated in order to describe the proportion of the total variance in ADC (Y_i), that is explained by the regression line. R^2 is used to express the goodness of fit of the regression line and is calculated by Equation 4.16. [39]

$$R^2 = \frac{\text{regression SS}}{\text{total SS}} \tag{4.16}$$

Where regression SS is the linear regression sum of squares, which explains the amount of variability among the ADC values (Y_i) that explains a linear regression. The total sum of squares (total SS) explains the overall variability of the ADC values. Regression SS is calculated with Equation 4.17 and total SS is calculated with Equation 4.18. [39]

$$\text{regression SS} = \sum \left(\hat{Y}_i - \bar{Y}\right)^2 \tag{4.17}$$

Where $\bar{Y}$ is the mean ADC value and $\hat{Y}_i$ is the predicted ADC values based on the regression line.

$$\text{total SS} = \sum \left(Y_i - \bar{Y}\right)^2 \tag{4.18}$$

Where $\bar{Y}$ is the mean ADC value calculated of the observed ADC values Y_i.

Moreover 95 % confidence limits were calculated in order to illustrate at which interval there is a 95 % probability of the mean of new ADC values will lie within. Meaning, that the confidence limits provides an interval of which the regression line will lie within. The standard error in Equation 4.19 of the regression line is used to calculate the confidence limits.[39]

$$S_{\hat{Y}_i} = \sqrt{S^2_{Y \cdot X} \left[\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x^2}\right]} \tag{4.19}$$

Where X_i is the values of temperature, $\bar{X}$ is the mean temperature, $\sum x^2$ is the sum of squares of temperature, n is the number of temperatures and $S^2_{Y \cdot X}$ is the sample residual mean square. The standard error of the regression line, $S_{\hat{Y}_i}$, can be used to calculate the 95 % confidence interval with Equation 4.20. [39]

$$CI = \hat{Y}_i \pm t_{0.05(2)} \cdot S_{\hat{Y}_i} \tag{4.20}$$

Where $t_{0.05(2)}$ is the critical value of the t distribution at 95 % probability.

A prediction interval is moreover calculated in order to illustrate at which interval around the regression line, there is a 95 % probability a new ADC sample will lie within. Meaning, 95 % of all new ADC samples will with 95 % probability be within the prediction limits. This gives an indication of the variance of values within the scanner. The standard error for estimating a single new ADC value is used for calculating the prediction intervals as expressed by Equation 4.21. [39]

$$(S_{\hat{Y}_i})_1 = \sqrt{S^2_{Y \cdot X} \left[1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x^2}\right]} \tag{4.21}$$

The standard error for estimating a single new ADC value, $(S_{\hat{Y}_i})_1$, is then used to calculate the 95 % prediction interval with Equation 4.22. [39]

$$PI = \hat{Y}_i \pm t_{0.05(2)} \cdot (S_{\hat{Y}_i})_1 \tag{4.22}$$

The plot including the regression line, confidence interval and prediction interval will have the temperature in the x-direction while having the measured and corrected ADC in the y-direction.

4.3.5 Completion Time of the Algorithm

The completion time of the algorithm was investigated, as time is an important factor in a clinical setting. The completion time was defined as the time it took for a QA of a single consistency data set. As explained in Section 2.3.2, the manual completion time of a QA of a single consistency data set per scanner is approximately 5-7 minutes. To investigate the algorithm completion time of QA of a single consistency data set, the mean completion time of all 30 consistency data set per scanner was calculated. Moreover the time reduced by the algorithm to perform a QA of a consistency data set compared to the Department of Radiology is calculated. The completion time test was conducted on a Laptop PC with Windows 10 Home, 64 bits operating system, with an AMD A8-5545M APU 1.7 GHz CPU, AMD Radeon HD 8510G 750 MB GPU and 8 GB ram.

Chapter 5

Results

This chapter presents the results of the algorithm testing. Firstly, the results of the automatic slice detection of the ADC benign region and the Siemens star will be described. Secondly, the quantification of the ADC benign ROI is presented. Next, the results of the quantification of the Siemens star is shown. Afterwards, the results of the investigation into the day-to-day variance of the measured ADC and the variance between scanners are presented. Lastly, the completion time of the algorithm to perform QA is compared to the completion time of a radiologist at the Department of Radiology at Vejle Hospital.

5.1 Automatic Slice Detection

The following section provides information about the automatic slice selection, where the results of the methods used are displayed. The displayed results are the six histogram correlation methods for the Siemens star and the disparity extracted from the Procrustes method for detection of the ADC ramp.

The six histogram distance measures are shown in two plots. The two methods Correlation and Intersection, where the highest value of the distance measurement corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, is shown in Figure 5.1.

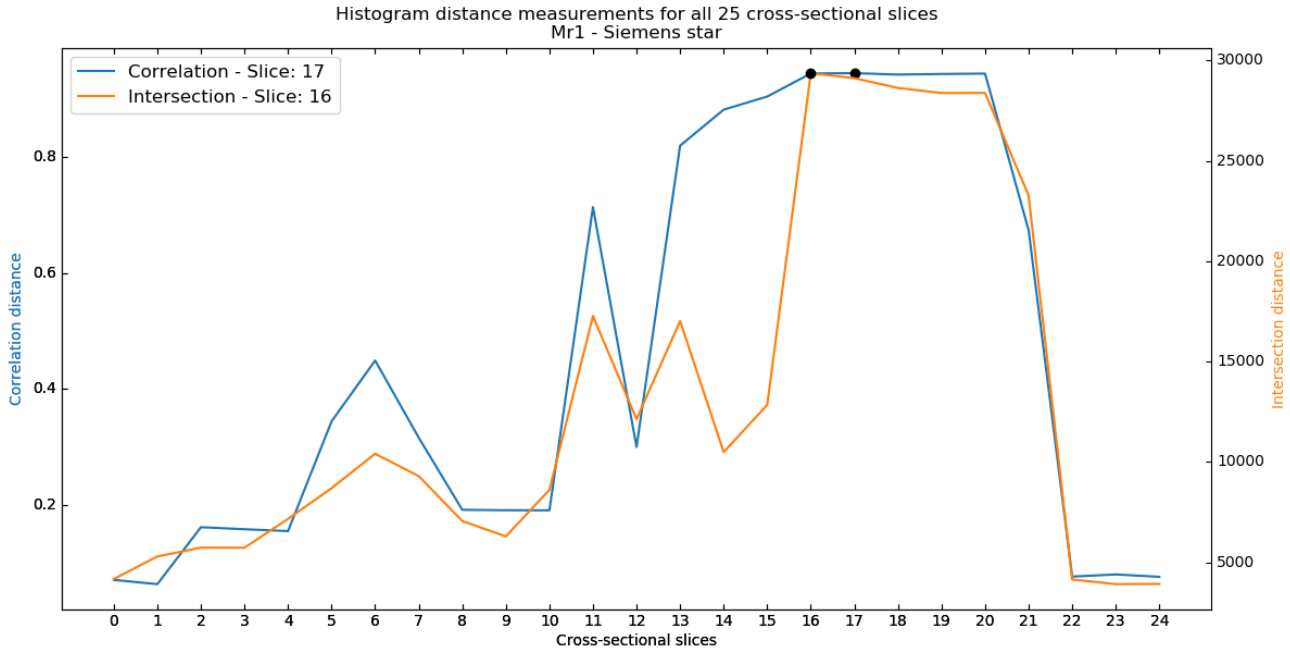


Figure 5.1: Results of the two distance measurements correlation and intersection of a consistency data set from MR1, where 25 cross-sectional images were investigated to detect the Siemens star.

The histogram distance measures Chi-squared, Hellinger, Alternative chi-squared and the Kullback-Leibler divergence, where the minimum values corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, as seen in Figure 5.2.

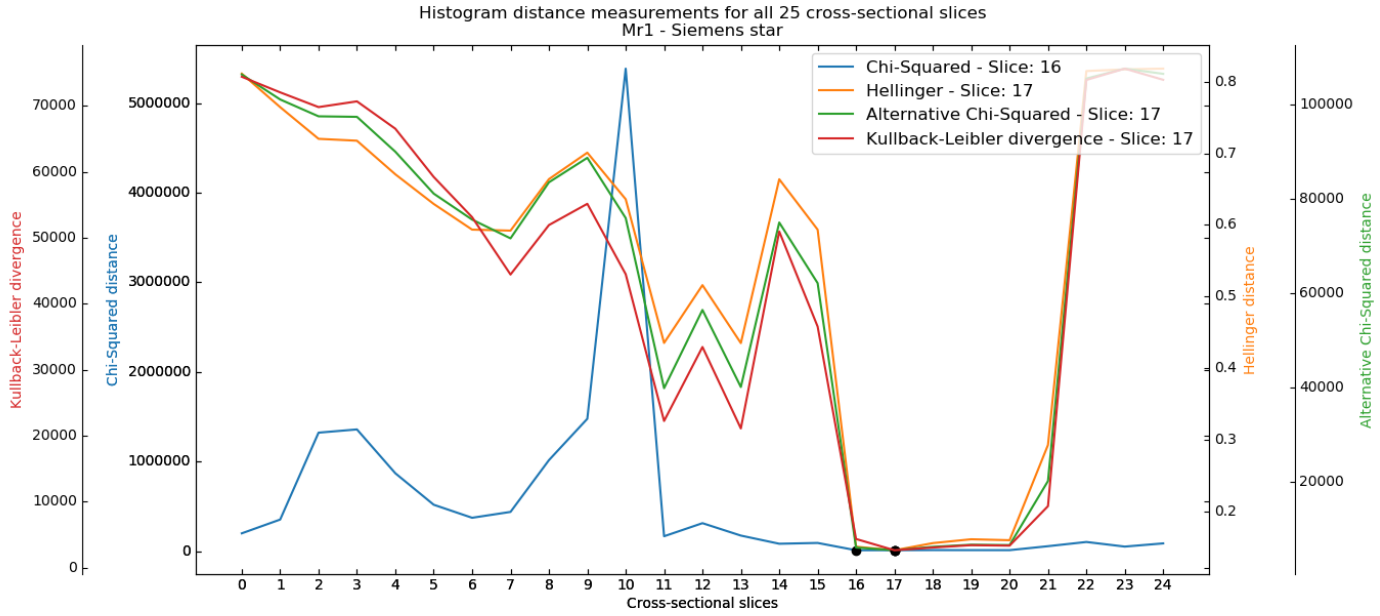


Figure 5.2: Results of the four distance measurements Chi-Squared, Hellinger, Alternative Chi-Squared and Kullback-Leibler divergence of a consistency data set from MR1, where 25 cross-sectional images were investigated to detect the Siemens star.

In this example the majority vote ended by choosing slice number 17, where 4 out of 6 voted for the selected cross-sectional slice. The histogram distance measurements which voted for the slice were Correlation, Hellinger, Alternative Chi-squared and the Kullback-Leibler divergence, while the last two methods Intersection and Chi-squared voted for slice number 16.

Slice number 17 which were chosen by the majority vote can be seen in Figure 5.3.

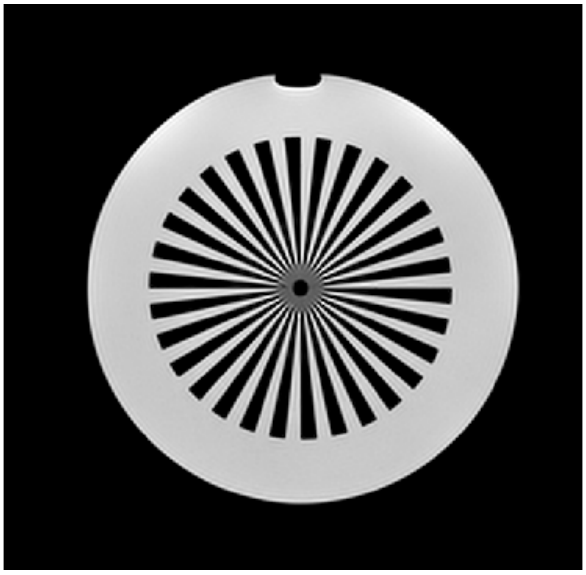


Figure 5.3: The result of the automatic slice detection found to be slice 17.

The results from the six histogram distance measurements and the corresponding selected slice of the Siemens star for MR2 and MR4 can be seen in Appendix A.

The Procrustes method resulted in a disparity value, where the lowest disparity corresponds to the cross-sectional slice with the highest similarity to the reference image of the ADC ramp. The disparity value is visualized in Figure 5.4 for MR1.

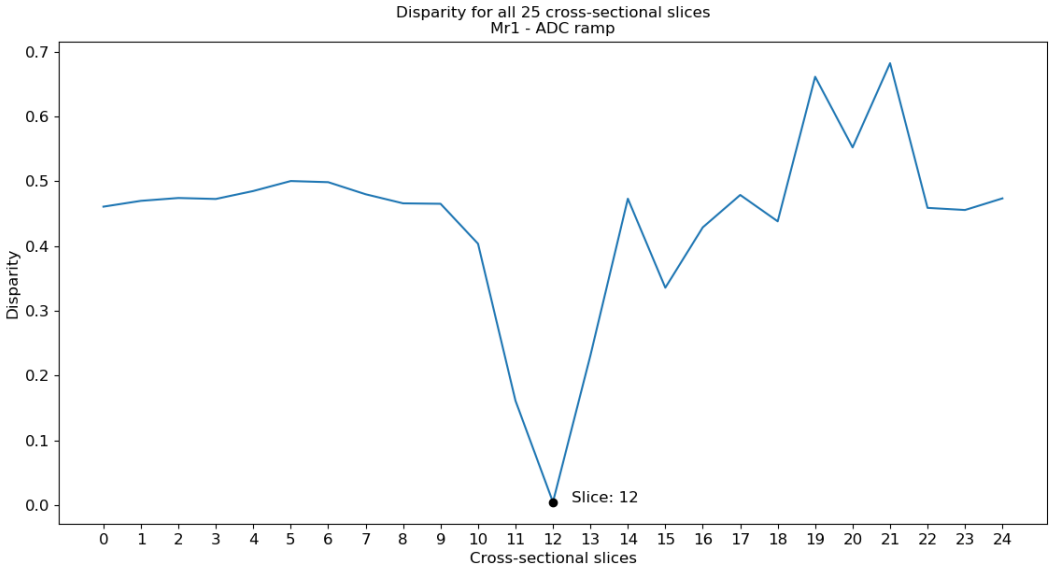


Figure 5.4: Results from the disparity value of a consistency data set from MR1, where 25 cross-sectional images were investigated to detect the ADC ramp.

The results from the automatic slice selection of the ADC ramp showed a disparity value of 0.0046. The found image of the ADC ramp can be seen in Figure 5.5, which were selected by the lowest disparity of the Procrustes method.

Figure 5.5: The result of the automatic slice detection found to be slice 12 by the lowest disparity.

The disparities from the Procrustes method with the corresponding selected slice of the ADC ramp for MR2 and MR4 can be seen in Appendix A.

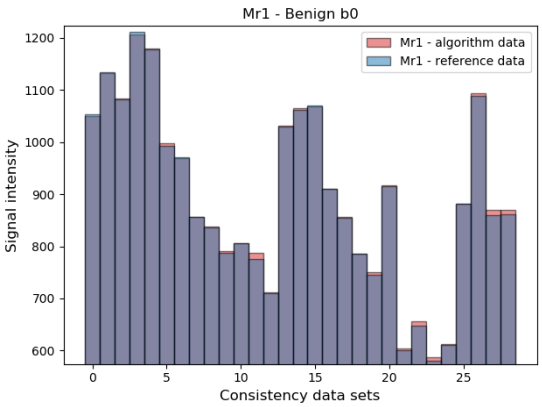
After the IQA explained in Section 4.3.1 was performed the automatic slice detection of 30 consistency data sets for MR1 and MR4 and 29 for MR2 all images chosen by the algorithm were chosen to be acceptable which concludes the automatic slice detection with an accuracy of 100 % for both the ADC ramp and the Siemens star.

5.2 Quantification of the ADC Benign Values

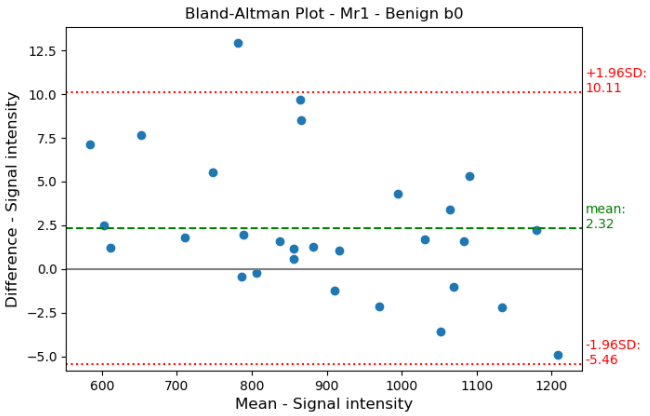
This section includes the results of the comparison of the algorithms quantification of the ADC benign values compared to the values from the Department of Radiology, defined as reference data.

The results contain the data from the algorithm in 29 consistency data sets for MR1. One consistency data set from MR1 was removed because of missing data from the Department of Radiology. The sequences used in this result section are b0, b1600 and the ADC map, which are the sequences the Department of Radiology provided results for.

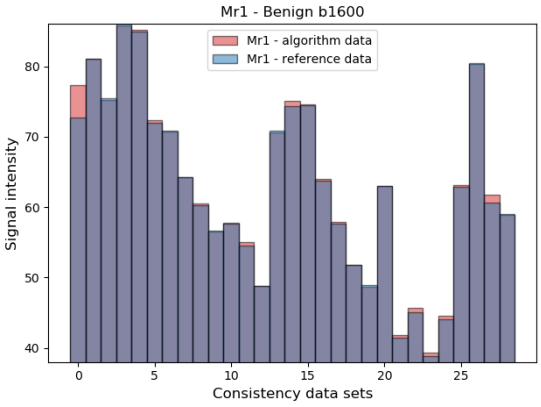
Figure 5.6a illustrates the results of the b0 sequence from MR1 with the benign values.



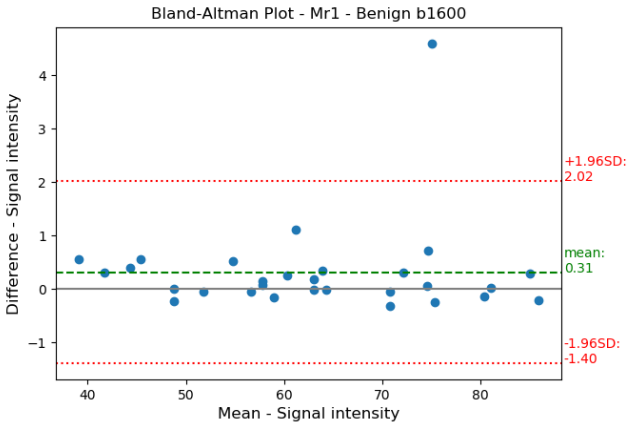
(a) Results from MR1 of b0.



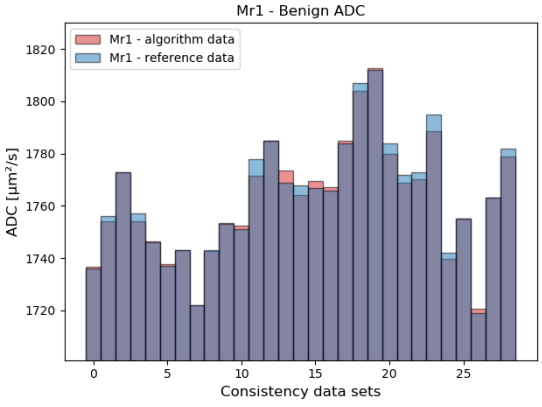
(b) Bland-Altman plot for b0 from MR1.



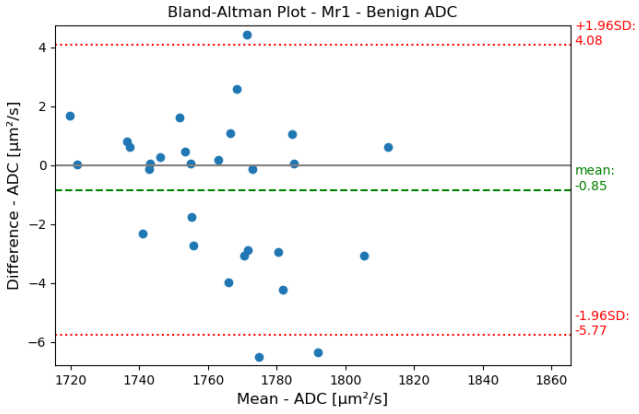
(c) Results from MR1 of b1600.



(d) Bland-Altman plot for b1600 from MR1.



(e) Results from MR1 of ADC.



(f) Bland-Altman plot for ADC from MR1.

Figure 5.6: Results from MR1 of b0, b1600 and ADC of the algorithm compared to the reference data for 29 consistency data sets and the difference between the signal intensities and ADC are illustrated in BA plots.

As seen in Figure 5.6a the benign values for the b0 varied between 580 and 1211 in signal intensity, and had a mean of 893.14 in the reference data. The algorithm data varied between 587 and 1206 with a mean value of 895.46, which gives a difference of 2.32 in the mean between the two groups. The Bland-Altman plot seen in Figure 5.6b illustrates the differences in values between the two methods of quantifying the benign b0 values. The maximum difference between the two methods of quantifying data was 12.96, which was the only data point outside the limits of agreement (LoA). The mean of the difference was 2.32, where the 95 % lower and upper LoA were between -5.46 and 10.11.

Next the b1600 values for MR1 are illustrated in Figure 5.6c, where the signal intensities varied between 38.8 and 86.0 in the reference data, which had a mean of 62.83. The algorithm data varied between 39.4 and 86.8 with a mean value of 63.14, which gives a difference of 0.31 in the mean between the two groups. The Bland-Altman plot seen in Figure 5.6d illustrates the differences in b1600 values between the two methods of quantifying the benign b1600 values. The maximum difference between the two methods of quantifying data was 4.60, which where the only data point outside the limits of agreement. As seen on the image the mean of the difference was 0.31, where the 95 % lower and upper LoA were between -1.4 and 2.02.

Lastly as seen in Figure 5.6e the quantified ADC values varied between 1719 and 1812 $\mu m^2/s$ in the reference data. The mean of the ADC values where 1763.38 $\mu m^2/s$. The algorithm data varied between 1720 and 1812 $\mu m^2/s$ with a mean value of 1762.53 $\mu m^2/s$, which gives a difference of 0.85 in the mean between the two groups. The Bland-Altman plot seen in Figure 5.6f illustrates the differences in ADC between the two methods of quantifying the ADC benign values. The maximum differences between the two methods of quantifying data were three values outside of the LoA which were 4.42, -6.51 and -6.35 $\mu m^2/s$. As seen on the image the mean of the difference was -0.85 $\mu m^2/s$, where the 95 % lower and upper LoA were between -5.77 and 4.02 $\mu m^2/s$.

The results of the comparison of quantified values for MR2 and MR4 can be seen in Appendix B.

The results of the comparisons for all three MR scanner, MR1, MR2 and MR4 can be seen in Table 5.1.

Table 5.1: Results of the comparison of quantification of benign values from the sequences b0, b1600 and the ADC map.

Quantification	b0 mean diff. ($\pm 1.96SD$)	b1600 mean diff. ($\pm 1.96SD$)	ADC mean diff. ($\pm 1.96SD$)
MR1 benign	2.3250 (± 7.7849)	0.3112 (± 1.7086)	-0.8465 (± 4.9218)
- Percentage	0.3096 (± 0.9741)	0.5092 (± 2.4283)	-0.0473 (± 0.2771)
MR2 benign	5.4415 (± 9.9385)	0.7360 (± 1.7210)	0.3761 (± 14.9084)
- Percentage	0.6266 (± 1.2543)	1.1499 (± 2.4212)	0.0359 (± 0.9414)
MR4 benign	-1.8560 (± 50.9480)	-0.7077 (± 4.3329)	2.4672 (± 27.5695)
- Percentage	-0.1103 (± 4.8646)	-1.2107 (± 5.9045)	0.0808 (± 0.8628)

5.3 Quantification of Sharpness

This section includes the results of the consistency data sets and altered images of the Siemens star. The sharpness of the Siemens star was quantified by a minimum line width, a mean image contrast and a mean of the two as explained in Section 4.3.3. Both the results of the algorithm and evaluations of the images performed by a radiologist from the Department of Radiology are described. First three quantifications of three images of the T2 weighted Siemens star are described, next three quantifications of three low-quality images of the T2 weighted Siemens star. Lastly three quantifications of the ADC Siemens star are described.

5.3.1 T2 Weighted Siemens Star

A Siemens star from a T2 weighted sequence from the scanner "MR1" can be seen in Figure 5.7. The radiologist from the Department of Radiology rated the T2 weighted images with a VAS score of 4, for all 3 MR scanners.

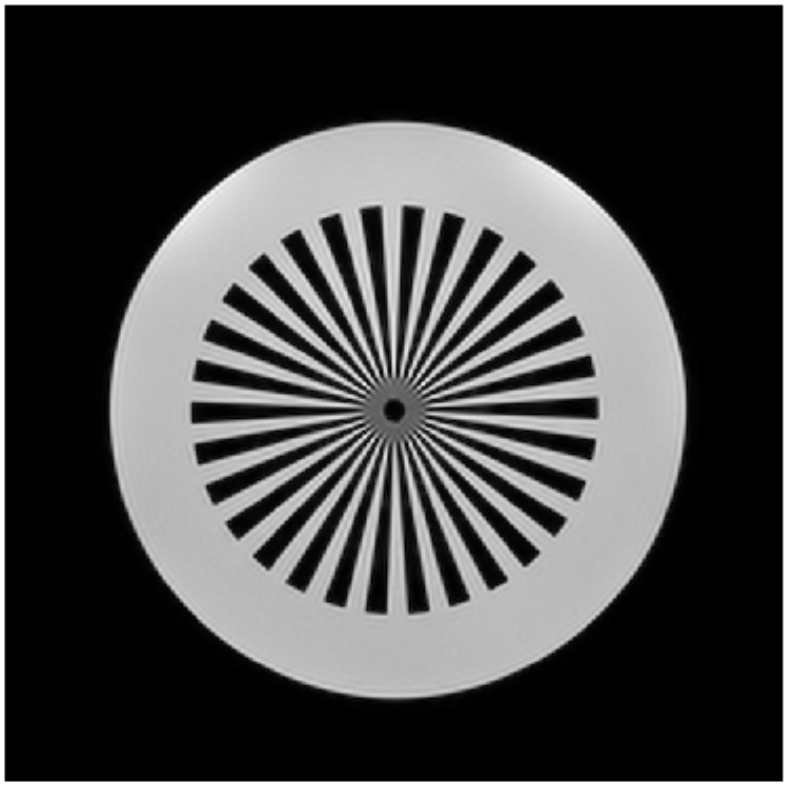


Figure 5.7: T2 weighted image of the Siemens star from MR1.

The minimum line width measured of the Siemens star in Figure 5.7 was 0.394 mm with a ratio of 0.934. The mean image contrast was calculated to 0.818. The resulting overall quantification of image sharpness was then 0.876.

A similar image where found in the scanner "MR2" which can be seen in Figure 5.8.

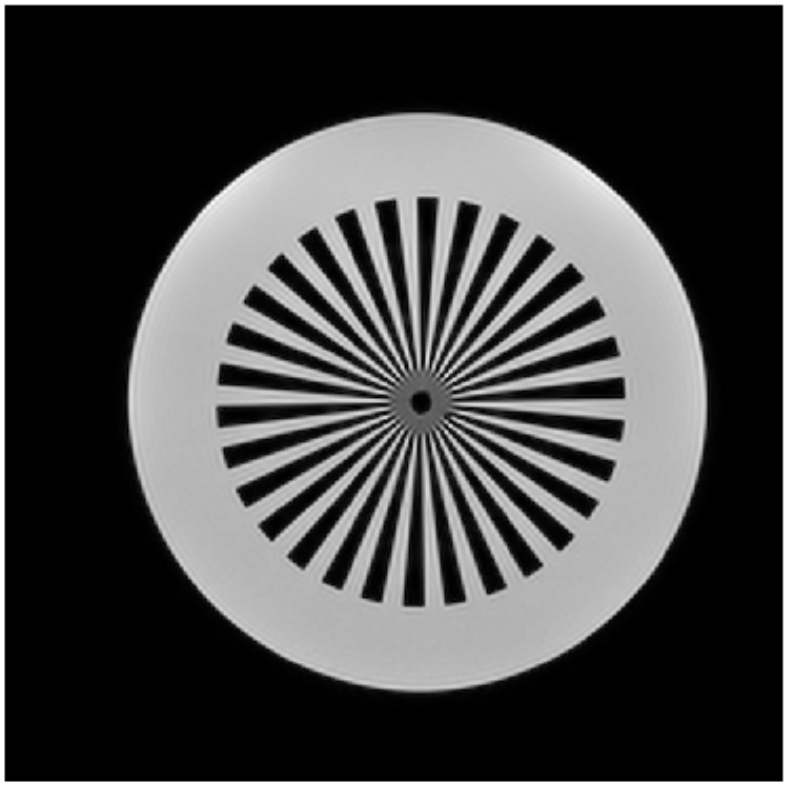


Figure 5.8: T2 weighted image of the Siemens star from MR2.

The minimum line width measured of the Siemens star in Figure 5.8 was 0.386 mm with a ratio of 0.935. The mean image contrast was calculated to 0.826. The resulting overall quantification of image sharpness was then 0.881.

Lastly an image where extracted from the scanner "MR4" for the T2 weighted sequence which can be seen in Figure 5.9.

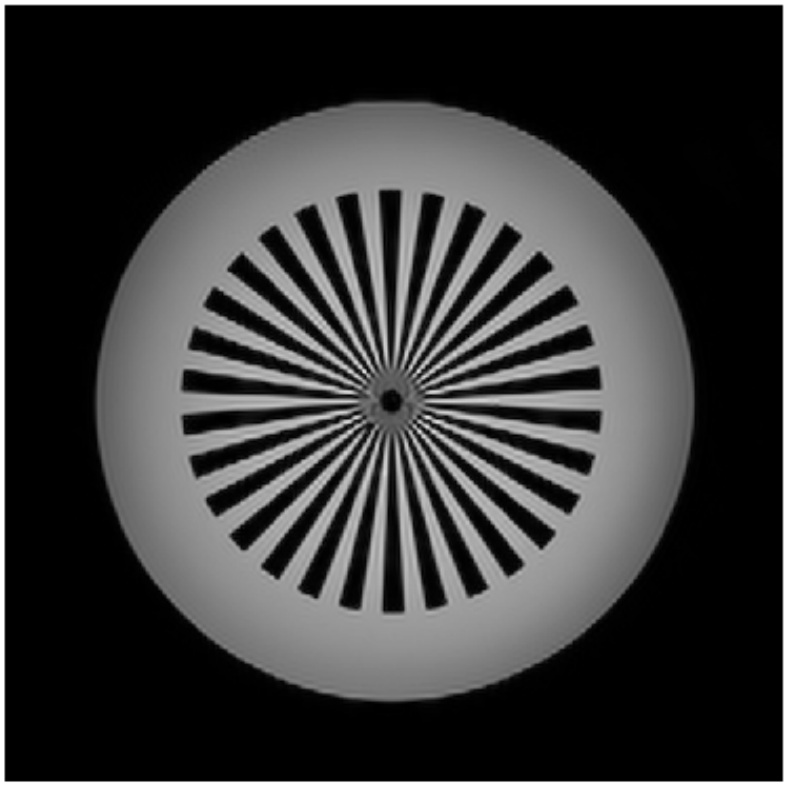


Figure 5.9: T2 weighted image of the Siemens star from MR4.

The minimum line width measured of the Siemens star in Figure 5.9 was 0.292 mm with a ratio of 0.951. The mean image contrast was calculated to 0.829. The resulting overall quantification of image sharpness was then 0.890.

The evaluation shown above was performed on all the consistency data sets from the 3 MR scanners, which were 30 consistency data sets from "MR1", 29 from "MR2" and lastly 30 from "MR4". The results from the evaluation of the Siemens star of the T2 weighed images can be seen in Table 5.2

Table 5.2: The evaluation of the Siemens star for T2 weighted images, with the number (N) of consistency data sets, minimum line width both in millimeter and converted to a ratio from 0 to 1. Image contrast and the quantified sharpness of the algorithm and lastly the VAS score performed by a professional from the Department of Radiology is shown.

Scanner	N	Contrast($\pm 1.96SD$)	Min. line width($\pm 1.96SD$)		Sharpness($\pm 1.96SD$)	VAS
		Ratio [0-1]	Width [mm]	Ratio [0-1]	Ratio [0-1]	Ratio [1-5]
MR1	30	0.821($\pm$0.009)	0.359($\pm$0.05)	0.940($\pm$0.009)	0.880($\pm$0.007)	4
MR2	29	0.828($\pm$0.005)	0.358($\pm$0.063)	0.940($\pm$0.011)	0.884($\pm$0.007)	4
MR4	30	0.829($\pm$0.005)	0.298($\pm$0.114)	0.950($\pm$0.019)	0.890 ($\pm$0.010)	4

The results from Table 5.2 showed a sharpness from the Siemens star of 0.880$\pm$0.007, 0.884$\pm$0.007 and 0.890 $\pm$0.010 for MR1, MR2 and MR4, respectively. MR1 had a mean minimum line width of 0.359$\pm$0.05 which is equal to ratio of 0.940$\pm$0.009 and lastly a contrast of 0.821$\pm$0.009 for the Siemens stars of the 30 consistency data sets. The Department of Radiology evaluated the images to a VAS score of 4 for all three scanners. The minimum line width found for MR2 was 0.358($\pm$0.063), which was a ratio of 0.940$\pm$0.011. The Michelson contrast was found to be 0.828$\pm$0.005 for MR2 for the Siemens stars of the 29 consistency data sets. MR4 had a mean line width of 0.298$\pm$0.114, which translates to a ratio of 0.950$\pm$0.019. Lastly MR4 had a Michelson contrast of 0.829$\pm$0.005 for the Siemens stars from the 30 consistency data sets.

5.3.2 Low-quality T2 Weighted Siemens Star

Low quality images were created to show the difference in the contrast, line width and the mean sharpness when the test is performed on low quality images. On Figure 5.10 a Gaussian filter have been added with a mask of with 15 pixels in height and width. A Gaussian filter was used to simulate noise in a Siemens star.

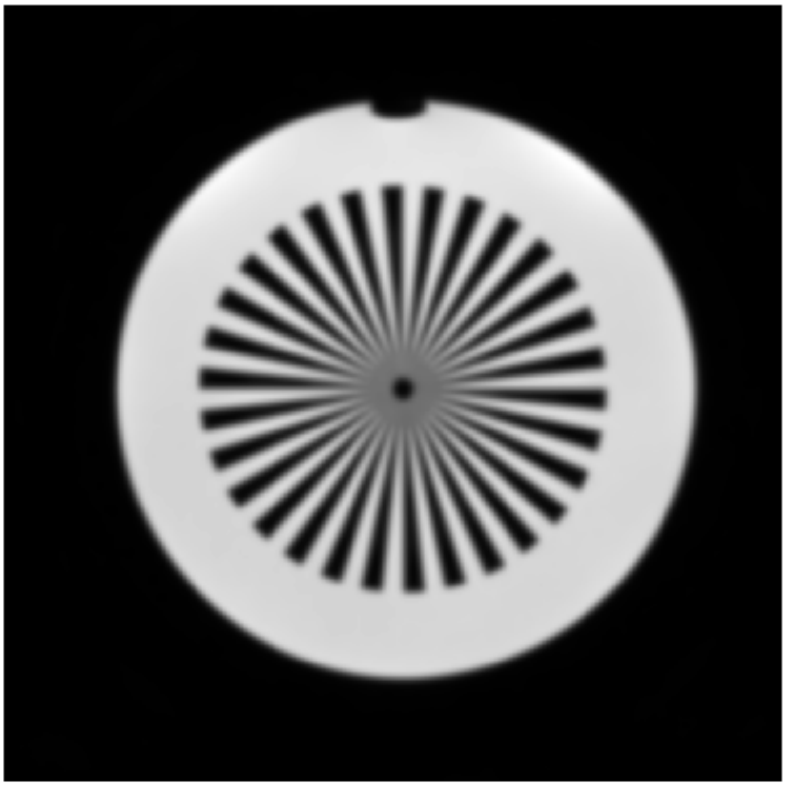


Figure 5.10: Low-quality Siemens star with Gaussian filter of 15x15 was applied.

The minimum line width measured of the Siemens star in Figure 5.10 was 1.245 mm with a ratio of 0.791. The mean image contrast was calculated to 0.524. The resulting overall quantification of image sharpness was then 0.658. The Department of Radiology rated this Siemens star with a VAS of 2.

A Gaussian filter was also applied to the second image, where the filter size was increased to 25 pixels in height and width to simulate more noise in a Siemens star. The image can be seen in Figure 5.11.

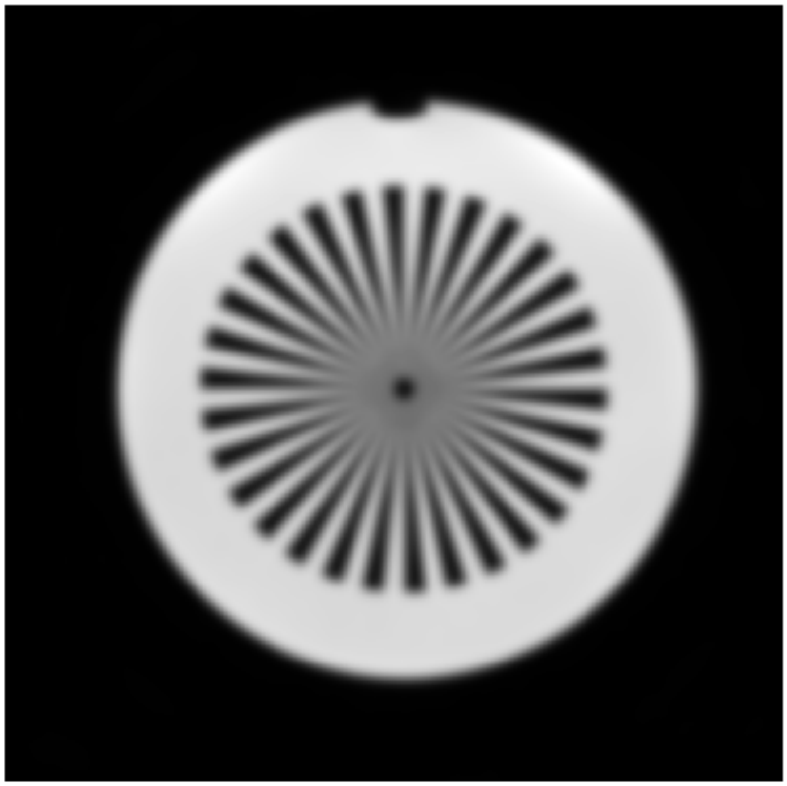


Figure 5.11: Low-quality Siemens star with Gaussian filter of 25x25 was applied.

The minimum line width measured of the Siemens star in Figure 5.11 was 2.103 mm with a ratio of 0.647. The mean image contrast was calculated to 0.318. The resulting overall quantification of image sharpness was then 0.483. The Department of Radiology rated this Siemens star with a VAS of 1.

The last low quality image was an image where lower intensities were added, to simulate a smaller contrast difference between the black and white contrast bars. The image of the low quality Siemens star is shown in Figure 5.12.

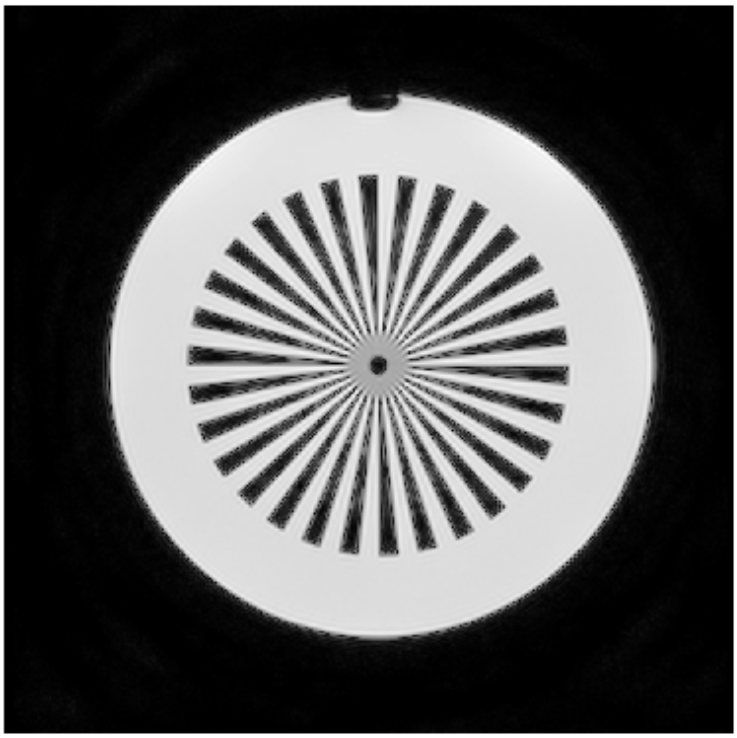


Figure 5.12: Low-quality Siemens star where high intensities were lowered.

The minimum line width measured of the Siemens star in Figure 5.12 was 0.558 mm with a ratio of 0.906. The mean image contrast was calculated to 0.569. The resulting overall quantification of image sharpness was then 0.738. The Department of Radiology rated this Siemens star with a VAS of 2.

A summary of the three altered images with the Michelson contrast, minimum line width in millimeters and ratio, sharpness and the VAS score can be seen in table 5.3.

Table 5.3: The evaluation of the low quality Siemens star, minimum line width both in millimeter and converted to a ratio from 0 to 1. Image contrast and the quantified sharpness of the algorithm and lastly the VAS score performed by a professional from the Department of Radiology is shown.

Altered image	Contrast	Min. line width		Sharpness	VAS
	Ratio [0-1]	Width [mm]	Ratio [0-1]	Ratio [0-1]	Ratio [1-5]
Gaussian 15x15	0.524	1.244	0.791	0.658	2
Gaussian 25x25	0.318	2.103	0.647	0.482	1
Lowered intensities	0.569	0.558	0.906	0.738	2

5.3.3 ADC Map Siemens Star

The Siemens star was also evaluated for the ADC map, where the Radiologist from the Department of Radiology rated the Siemens star quality with a VAS of 2 for the two scanner with 1.5 T and a VAS score of 1 on the 3 T scanner. The Siemens star was evaluated on one slice from each of the three MR scanners. The slice from the MR scanner MR1 can be seen in Figure 5.13.

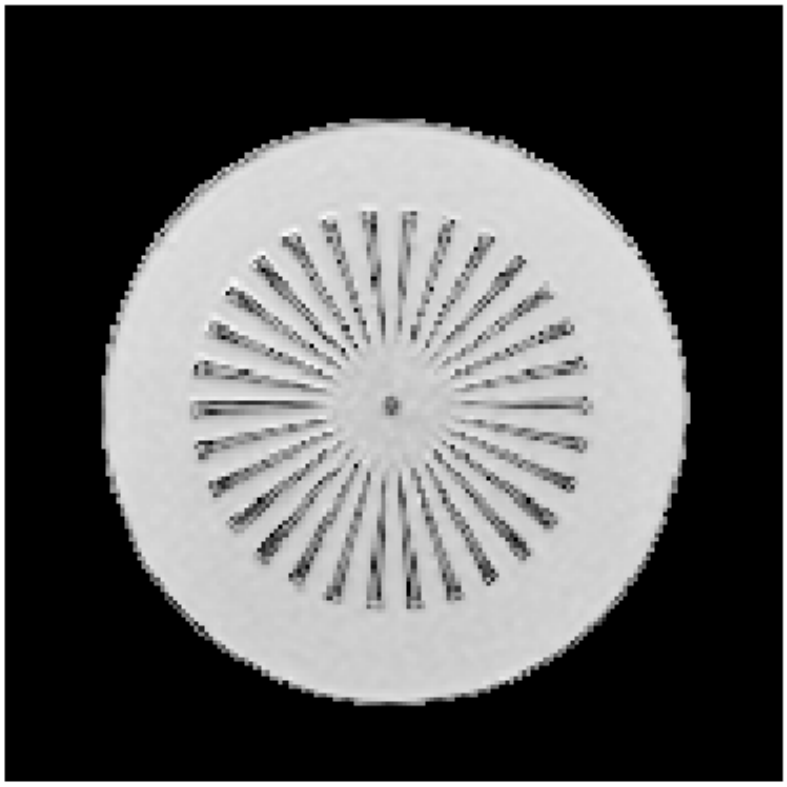


Figure 5.13: ADC Siemens star of MR1.

The minimum line width measured of the Siemens star in Figure 5.13 was 1.388 mm with a ratio of 0.767. The mean image contrast was calculated to 0.224. The resulting overall quantification of image sharpness was then 0.496.

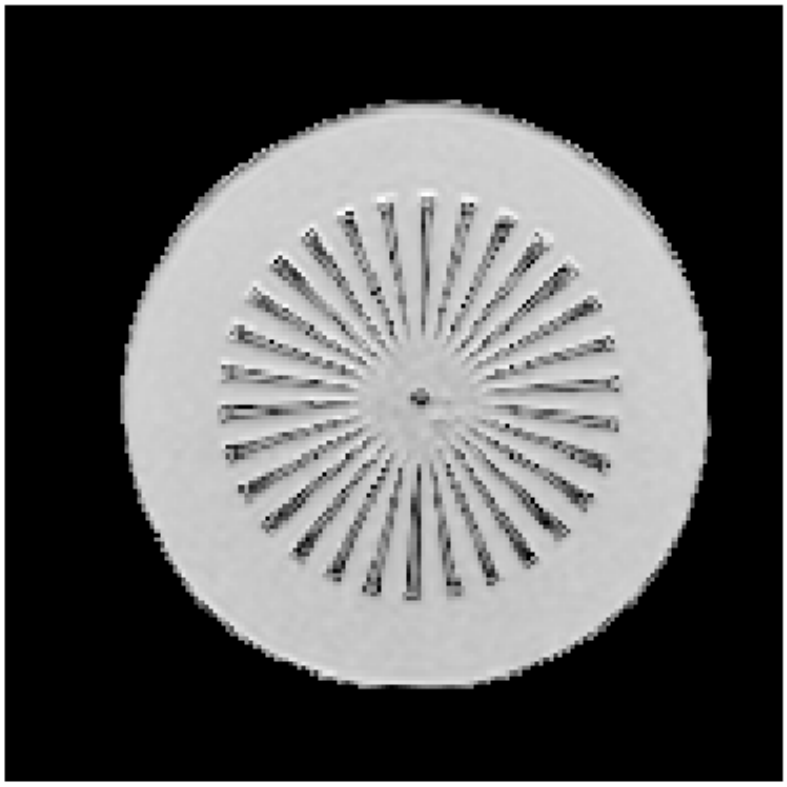


Figure 5.14: ADC Siemens star of MR2.

The minimum line width measured of the Siemens star in Figure 5.14 was 1.248 mm with a ratio of 0.790. The mean image contrast was calculated to 0.242. The resulting overall quantification of image sharpness was then 0.516.

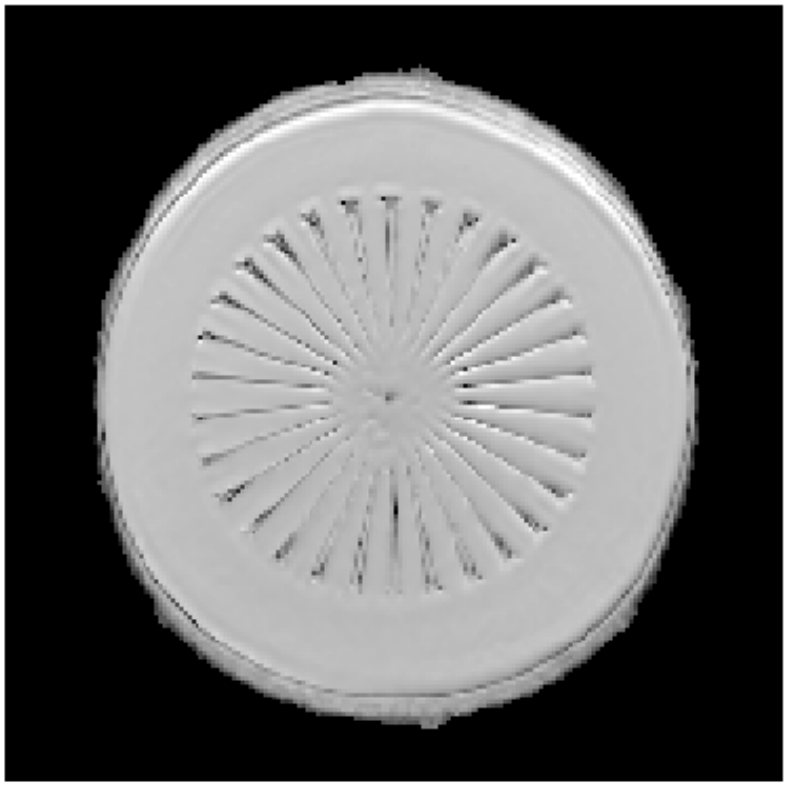


Figure 5.15: ADC Siemens star of MR4.

The minimum line width measured of the Siemens star in Figure 5.15 was 1.218 mm with a ratio of 0.796. The mean image contrast was calculated to 0.099. The resulting overall quantification of image sharpness was then 0.447.

A summary table can be seen in Table 5.4 where the Michelson contrast, the width in millimeters and ratio of the minimum line width and the mean sharpness are displayed for the 3 MR scanners.

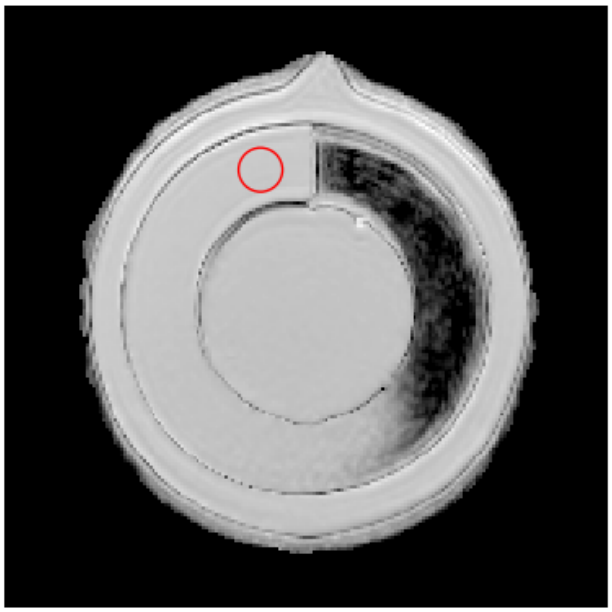
Table 5.4: The evaluation of the Siemens star from the ADC maps, with the number (N) of consistency data sets, minimum line width both in millimeter and converted to a ratio from 0 to 1. Image contrast and the quantified sharpness of the algorithm and lastly the VAS score performed by a professional from the Department of Radiology is shown.

Scanner	N	Contrast($\pm 1.96SD$)	Min. line width($\pm 1.96SD$)		Sharpness($\pm 1.96SD$)	VAS
		Ratio [0-1]	Width [mm]	Ratio [0-1]	Ratio [0-1]	Ratio [1-5]
MR1	30	0.231($\pm$0.019)	1.647($\pm$1.576)	0.724($\pm$0.264)	0.478($\pm$0.135)	2
MR2	29	0.242($\pm$0.090)	1.201($\pm$0.343)	0.799 ($\pm$0.058)	0.521($\pm$0.065)	2
MR4	30	0.100($\pm$0.020)	1.229($\pm$0.213)	0.794($\pm$0.036)	0.447($\pm$0.025)	1

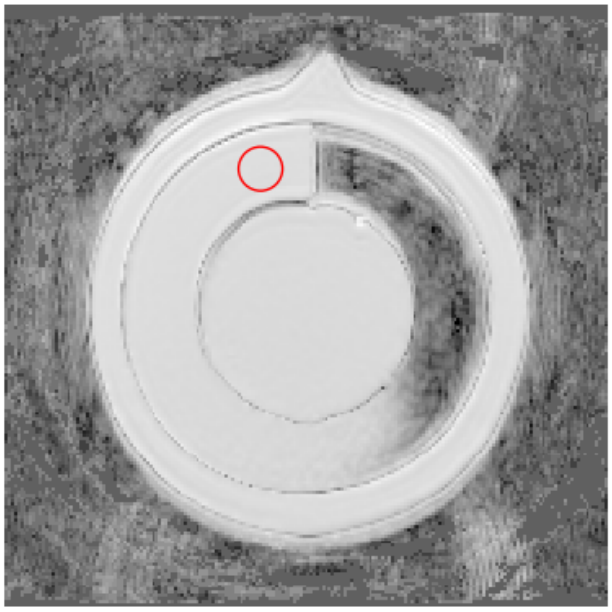
The mean sharpness of the 3 scanners were found to be 0.478$\pm$0.135, 0.521$\pm$0.065 and 0.447$\pm$0.025 for MR1, MR2 and MR4, respectively. The minimum line width was in MR1 1.647$\pm$1.576, which was a ratio of 0.724$\pm$0.264. The Michelson contrast was 0.231$\pm$0.019 for the Siemens stars in the 30 consistency data sets. For MR2 the Siemens stars for the 29 consistency data sets ended with a minimum line width of 1.201$\pm$0.343 that translates to a ratio of 0.799$\pm$0.058. The Michelson contrast for MR2 was 0.242$\pm$0.090. The minimum line width found for MR4 was 1.229$\pm$0.213, with a calculated ratio of 0.794$\pm$0.036. The Michelson contrast for MR4 was 0.100$\pm$0.020 for the Siemens stars of the 30 consistency data sets.

5.4 Day-to-day Variance

This section explains the results of the day to day variance of ADC and sharpness and the variance between MR scanners. Regarding the variance between scanners, the calculated mean ADC for all consistency data sets for MR1, MR2 and MR4 were respectively; $1,762.53$ (± 42.40), $1,742.23$ (± 101.27) and $3,191.47$ (± 82.82) $\mu m^2/s$. A minor difference of 1.15 % was found between the two 1.5 Tesla scanners, MR1 and MR2. However a larger difference was found between the 1.5 Tesla scanners and the 3 Tesla scanner, MR4, of 45.09 %. Moreover the simulated ADC of the PVP solution in the phantom was 1599 $\mu m^2/s$, according to Section 2.3.1. To investigate this, an ADC map of the MR4 scanner was calculated based on the DW images, b0 and b1600. The calculated ADC map was compared to the ADC map created by the Philips MR scanner, referred to as the post-processed ADC map, see Figure 5.16.



(a) ADC map post-processed by the Philips MR4 scanner. The ADC ROI had a mean ADC of $3,110.1$ (± 23.4) $\mu m^2/s$.



(b) Calculated ADC map of the b0 and b1600 DWI from MR4, of the ADC benign region. The calculated ADC ROI had a mean ADC of $1,681.1$ (± 12.6) $\mu m^2/s$

Figure 5.16: Both the post-processed ADC map and the calculated ADC map from a consistency data set of MR4.

The post-processed ADC map had a mean ADC of $3,110.1$ (± 23.4) $\mu m^2/s$, the ADC map can be seen in Figure 5.16a . The calculated ADC map of the two DW images of MR4, as seen in Figure 5.16b had a mean ADC of $1,681.1$ (± 12.6) $\mu m^2/s$.

The calculated mean ADC for all consistency data sets for MR1, MR2 and MR4 were then respectively; $1,762.53$ (± 42.40), $1,742.23$ (± 101.27) and $1,725.06$ (± 44.76) $\mu m^2/s$. These mean values was however still different from the simulated ADC within the phantom of 1599 $\mu m^2/s$. Therefore an investigation into the correlation between ADC and temperature was conducted as explained in Section 4.3.4.

Figure 5.17 shows the correlation between the measured ADC for MR1 and the temperature of the measurement.

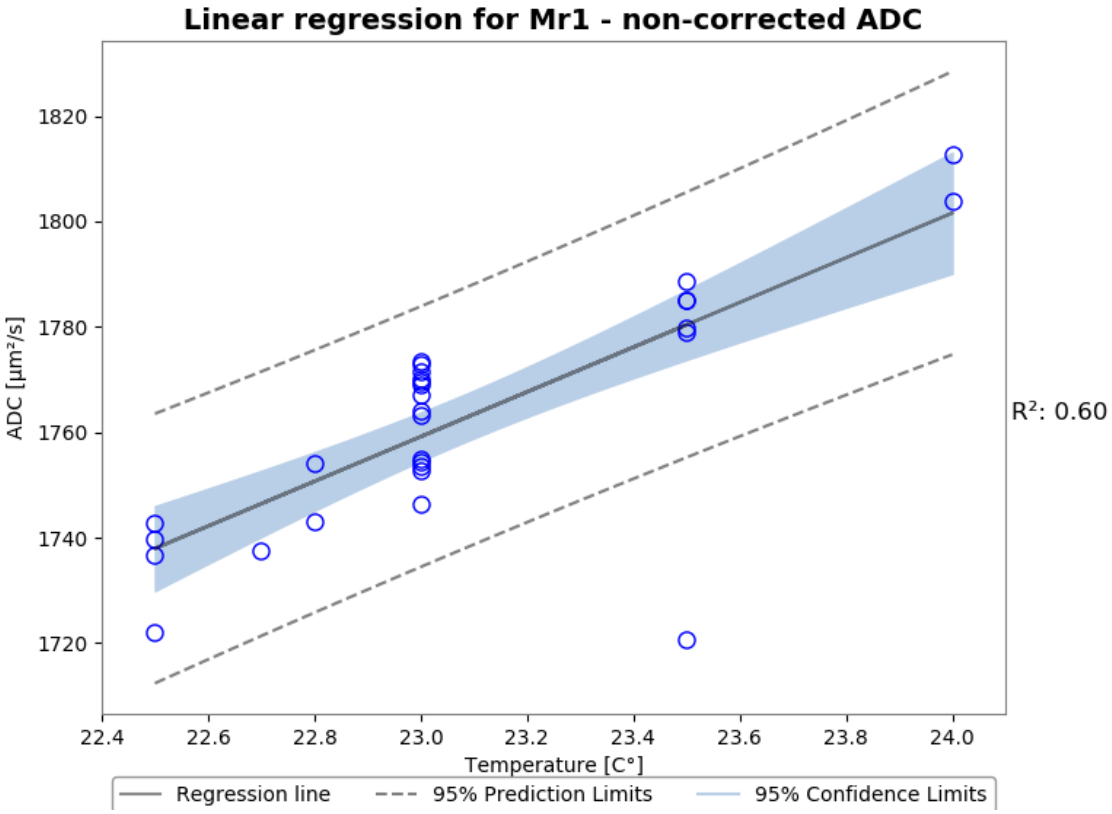


Figure 5.17: Results of non-corrected ADC temperature dependency of MR1.

As seen in Figure 5.17 a high linear correlation was found with a R^2 value of 0.6 between ADC and temperature. The minimum and maximum ADC were calculated to be respectively $1,720.67$ and $1,812.61$ $\mu m^2/s$. Therefore in order to correct this

dependency of temperature in the calculation of ADC, the ADCs were corrected for the temperature with Equation 4.14 in Section 4.3.4. The resulting correlation between ADC and temperature is shown in Figure 5.18.

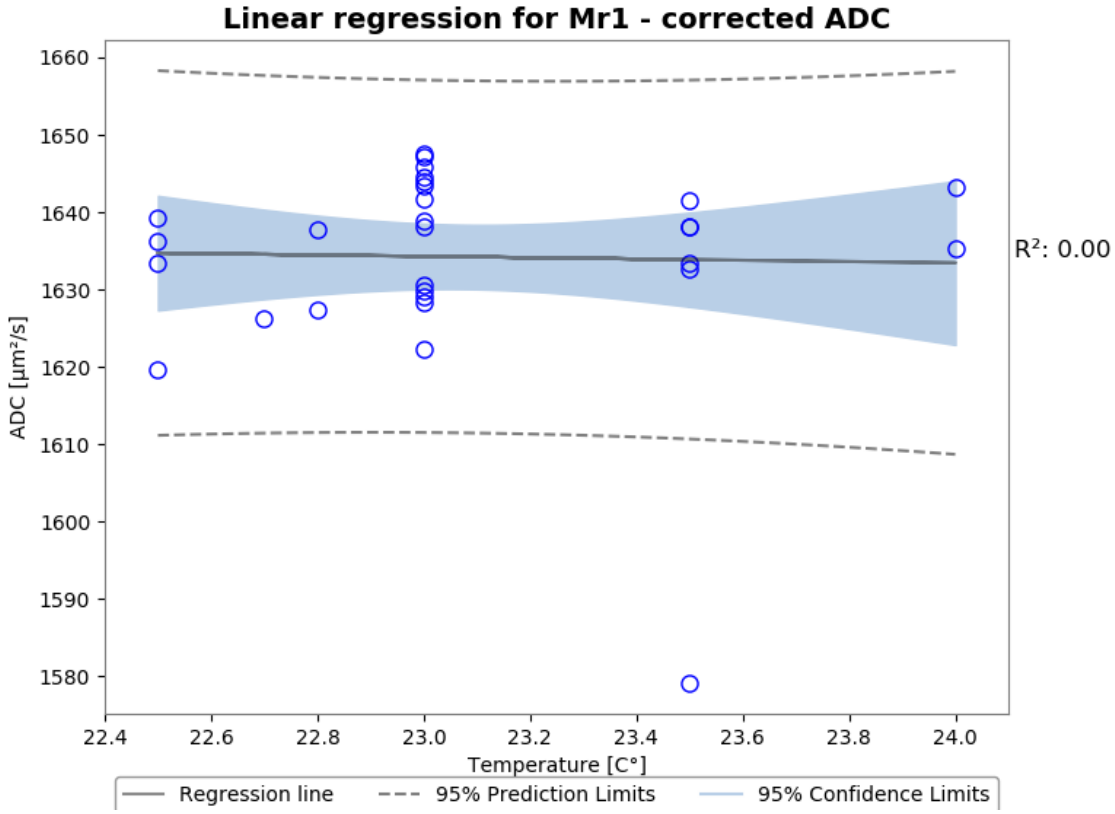


Figure 5.18: Results of temperature corrected ADC of MR1.

As seen in Figure 5.18 the linear correlation between ADC and temperature is no longer present with a R^2 value of 0.0. The mean ADC for the corrected ADC measurements of MR1 was calculated to $1,634.23$ (± 24.87) $\mu m^2/s$. The minimum and maximum ADC were calculated to be respectively $1,579.11$ and $1,647.60$ $\mu m^2/s$.

Figure 5.19 shows the correlation between the measured ADC for MR2 and the temperature of the measurement.

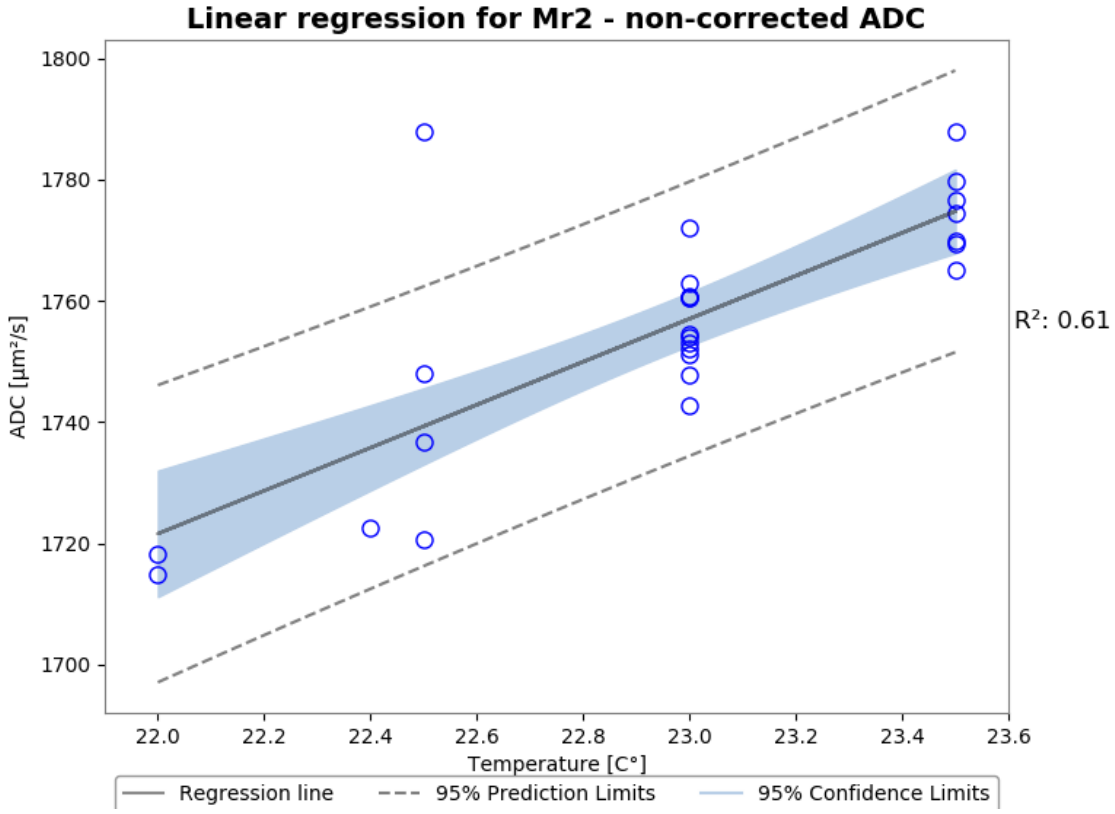


Figure 5.19: Results of non-corrected ADC temperature dependency of MR2.

As seen in Figure 5.19 a high linear correlation was found with a R^2 value of 0.61 between ADC and temperature. The minimum and maximum ADC were calculated to be respectively $1,566.19$ and $1,787.79$ $\mu m^2/s$. Two outliers, as seen in Appendix B in Figure B.1f with an ADC of $1,566.19$ and $1,572.60$ $\mu m^2/s$, were removed for the investigation into temperature correlation, as these ADC measurements were conducted on low-quality consistency data sets. The dependency of temperature in the calculation of ADC was corrected with Equation 4.14. The resulting correlation between the corrected ADC and temperature is shown in Figure 5.20.

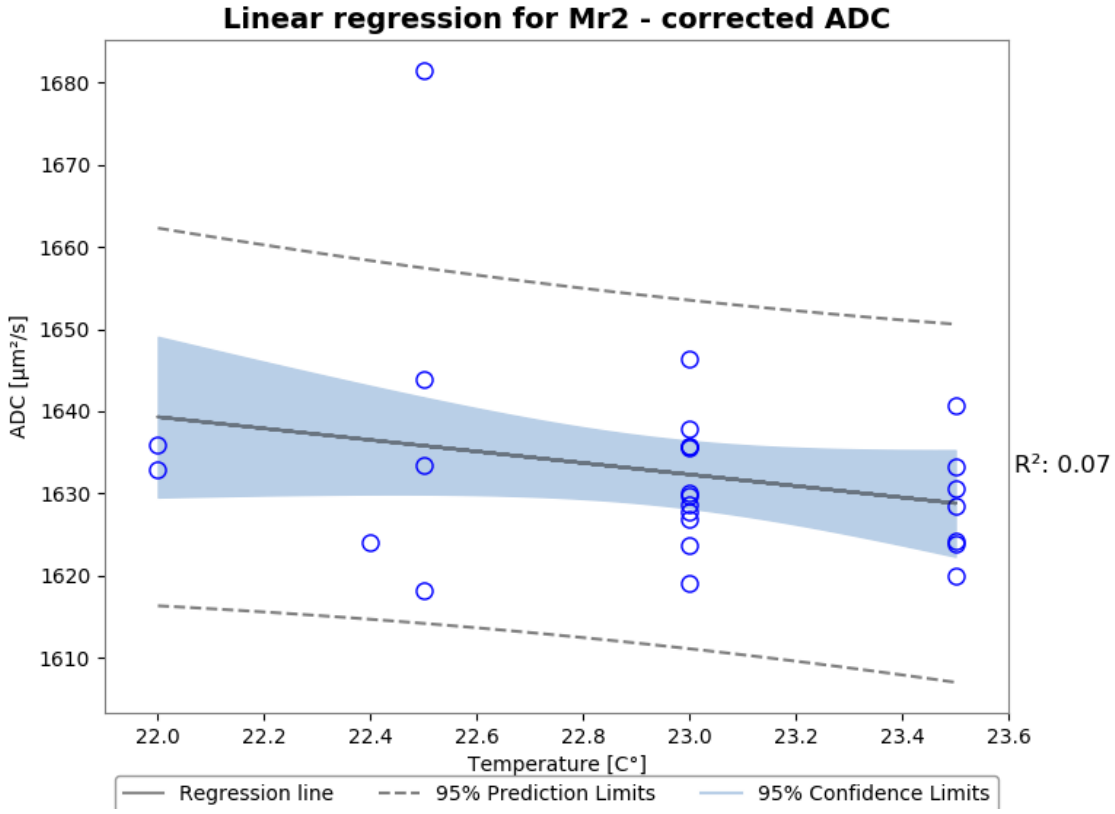


Figure 5.20: Results of temperature corrected ADC of MR2.

As seen in Figure 5.20 the linear correlation between ADC and temperature is lowered to a R^2 value of 0.07. The mean ADC for the corrected ADC measurements of MR2 was calculated to $1,632.61$ (± 23.68) $\mu m^2/s$. The minimum and maximum ADC were calculated to be respectively $1,618.20$ and $1,681.45$ $\mu m^2/s$. The regression line indicates a minor decrease in ADC as a function of temperature. The minor decrease could be caused by an outlier in ADC of approximately $1,680$ $\mu m^2/s$ with 22.4 C°.

Figure 5.21 shows the correlation between the measured ADC for MR4 and the temperature of the measurement.

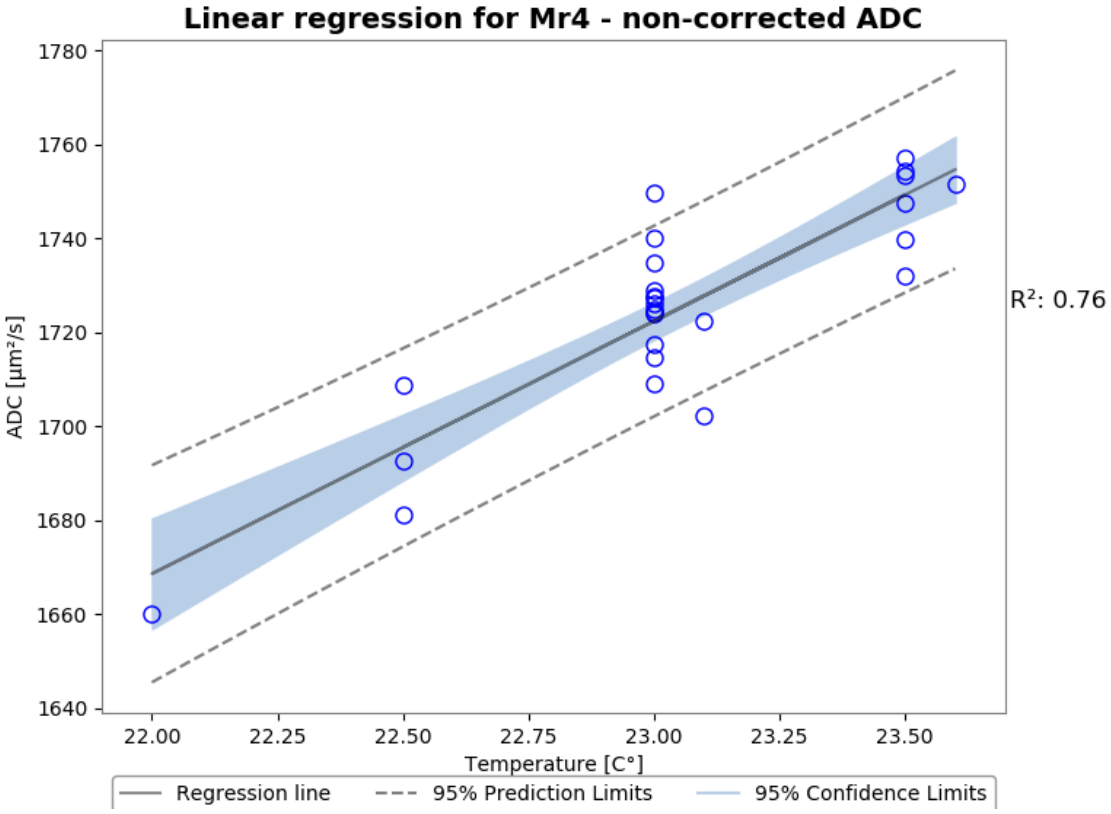


Figure 5.21: Results of non-corrected ADC temperature dependency of MR4.

As seen in Figure 5.21 a high linear correlation was found with a R^2 value of 0.76 between ADC and temperature. The minimum and maximum ADC were calculated to be respectively $1,660.02$ and $1,757.17$ $\mu m^2/s$. The dependency of temperature in the calculation of ADC was corrected with Equation 4.14. The resulting correlation between ADC and temperature is shown in Figure 5.22.

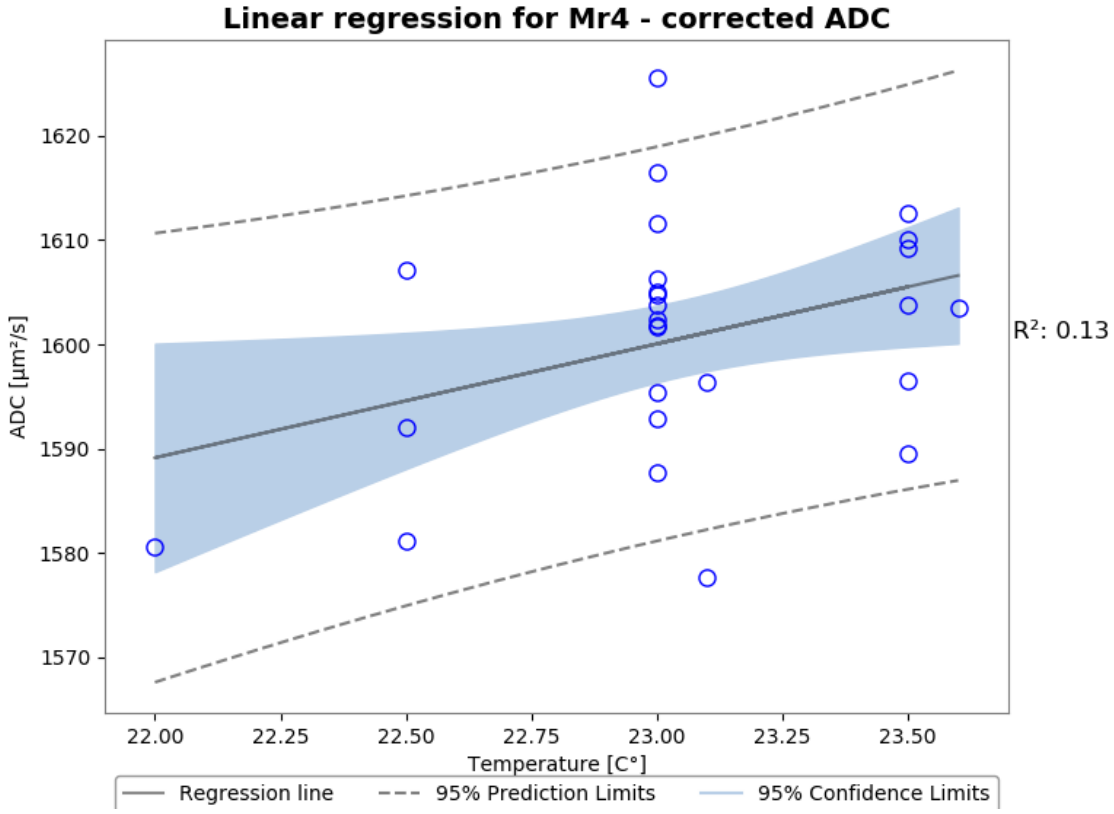


Figure 5.22: Results of temperature corrected ADC of MR4.

As seen in Figure 5.22 the linear correlation between ADC and temperature is lowered to a R^2 value of 0.13. The mean ADC for the corrected ADC measurements of MR4 was calculated to $1,600.60$ (± 21.91) $\mu m^2/s$. The minimum and maximum ADC were calculated to be respectively $1,577.61$ and $1,625.56$ $\mu m^2/s$. The regression line indicates a minor increase in ADC as a function of temperature.

5.5 Completion Time of the Algorithm

The algorithm was tested for its completion time of a single consistency data set by running the algorithm for all three MR scanners. The mean completion time was calculated with 1.96 SD for all consistency data sets. The completion time for running the algorithm for each scanner compared to the completion time of the Department of Radiology, according to Section 2.3.2, can be seen in Table 5.5.

Table 5.5: Results of completion time in seconds of the algorithm and the Department of Radiology, for each MR scanner with all consistency data sets. N is the number of consistency data sets.

		Mean completion time ($\pm 1.96SD$) [s]		
Scanner	N	Algorithm	Department of Radiology	Time reduction (%)
MR1	30	14.76 (± 1.45)	360	95.90
MR2	29	15.17 (± 2.06)	360	95.79
MR4	30	16.60 (± 3.10)	360	95.39

As seen in the table the mean completion time for "MR1" was found to be 14.76±0.74 seconds, which were almost the same completion time for "MR2" with a mean completion time of 15.17±1.05 seconds and lastly the scanner "MR4" had a mean completion time of 16.60±1.58 seconds. The time reduced by the algorithm compared to the Department of Radiology was approximately 95 % for all three scanners.

Chapter 6

Discussion

In order document consistently high-quality DW images and fDM, it is mandatory to conduct QA of MR scanners. QA of the MR scanners was performed with a DWI phantom, with a scanning protocol consisting of T2 weighted, DWI and ADC scanning sequences of an ADC region and a Siemens star for sharpness assessment. Different automatic QA algorithms are available today, however an automatic assessment of the scanner sharpness is not yet available. Therefore an automatic algorithm for QA of MR scanners using DW images from the DWI phantom was developed.

The automatic slice detection had an accuracy of 100 % for the 89 consistency data sets for both the ADC ramp and the Siemens star. Concerning the slice detection of the Siemens star, low accuracy was achieved when only using a single method for histogram similarity. Therefore it was chosen to base the slice detection on a majority vote of all methods for histogram similarity. However the algorithm might not always select the correct slice. If the consistency data set is of too low quality, then there will not be a high similarity between the reference images of the Siemens star, and the images within the consistency data set. The same restriction could be present for low quality images of the ADC ramp when using the Procrustes method in the automatic slice detection.

The sharpness of the cross-sectional images of the Siemens star in the phantom was investigated with a method that quantified the minimum detectable structure line width, as a function of image contrast. An overall quantification of the image sharpness was conducted with a mean value of the minimum detectable line width and the mean contrast of the image. The algorithms quantification of image sharpness was compared to the quantification of the radiologist by the Department of Radiology. Concerning the image sharpness of T2 weighted images of the Siemens star, Section 5.3, a connection was seen between the algorithm and the VAS score performed by the radiologist. For MR1 the mean sharpness was 0.880 and the mean VAS score was 4. This indicates that when quantifying T2 weighted images, the VAS score relatively high with a score of 4 out of 5, whereas the algorithm quantification was relatively high with a sharpness of 0.880 with a maximum value of 1. The same connection was found for MR2 and MR4 with a mean sharpness of respectively 0.884 and 0.890 and a mean

VAS of 4 for both. The investigation of the minimum detectable line width gave a quantification of the ability of the scanner to reproduce fine details in the images. The mean minimum detectable line width was relatively small for MR1, MR2 and MR4, which were able to detect fine details down to respectively 0.359, 0.358 and 0.298 mm, before the details became blurred. For both the overall sharpness and the minimum detectable line width, the 3 T MR scanner, had a slight indication of providing higher image quality and sharpness compared to the two 1.5 T MR scanners. These findings are consistent with Wolfsberger et al. [40], who found that 3 T MRI was superior to 1.5 T MRI in higher resolution and image quality. The advantages of 3 T MRI are due to the higher signal to noise ratio (SNR), which gives approximately twice the SNR resulting in faster scanning time or higher resolution. [41]
Three low-quality T2 weighted images of the Siemens star were created in order to investigate the performance of the algorithm to quantify low-quality sharpness in the images. To simulate relevant MRI artifacts, two images of the Siemens star were blurred with two different degrees and the intensities in the third image were changed. The results showed that the algorithm rated the low-quality images of the Siemens star of respectively 0.658, 0.482 and 0.738, which was lower than the quantification of the unchanged T2 weighted images. Similar to the unchanged T2 weighted images there was found a connection between the algorithm and the radiologist for the low-quality images, as the VAS scores was respectively 2, 1 and 2. These VAS scores indicates that blurring of the Siemens star is regarded worse than a change in intensities. This relation is found to be consistent with the quantification of the algorithm, as the image with a change in intensities was rated higher than the blurred images.
These findings were also consistent with the quantification of image sharpness conducted on the ADC maps. As seen on the ADC maps, both blurring and change in image intensities were present, which also was indicated by the mean VAS scores of 2, 2 and 1 for respectively MR1, MR2 and MR4. A connection was seen between the algorithm and the VAS score performed by the radiologist, as the algorithms quantification of image sharpness for the ADC maps were respectively 0.478, 0.521 and 0.447. The blurring of the MR4 ADC map was more significant compared to the ADC map of MR1 and MR2, which is also indicated by the sharpness quantification of the algorithm and the VAS score. These findings suggests that the DW images of the 3 T MR scanner are of lower quality compared to 1.5 T MR scanners, which is opposite of the findings based on T2 weighted images and the findings reported by Wolfsberger et al. [40]. Lavdas et al. [42] compared DW images and ADC maps of a phantom acquired of a 1.5 T and a 3 T scanner. They found that 1.5 T DW images resulted in high-quality images, and when acquiring DW images with 3 T scanner with the same protocol as for the 1.5 T, the resulting images was of low-quality. The images included severe artifacts as distortion and signal loss. A protocol optimization which fitted to the 3 T scanner was then conducted, which included a minimization of the echo time

(TE) of the scanning sequence. The optimization resulted into a significant reduction in signal losses in the images, it was however still questionably if the image quality of the DW images was of as high quality, as the images from the 1.5T scanner. [42] These findings by Lavdas et al. and the findings of this Master's thesis are however contradicted by Sánchez-González et al. [43], who reports that the higher SNR of the 3T scanner benefits DWI.

The method for quantification of image sharpness by the algorithm could be elaborated with directional information in the vertical and horizontal direction within the Siemens star. This elaboration could be investigated by dividing the Siemens star into segments in both directions. The quantification of the minimum line width was conducted on the mean image contrast, but with the division into segments, information regarding the dependence on direction could be achieved. This could potentially give quality assurance on specific directional gradients of the MR scanner.

The benign values for the scanning sequences b0, b1600 and the ADC map were extracted and compared to the values quantified by the Department of Radiology. The mean differences between the algorithm data and the reference data were between -1.856 to 5.44 when observing the signal intensities and size of ADC values varied greatly. As an example the signal intensities for b1600 for MR1 were between 38.8 and 86 likewise the ADC values varied between 1719 to 1812 $\mu m^2/s$. So the difference in percentage was found to explain the varying size. When examining the percentage difference the mean of the values varied between -1.21 and 1.15 percentage points (p.p.), while the lowest 1.96 SD were ± 0.2271 p.p. and the highest were ± 5.9045 p.p. The difference between the algorithm data and the reference data could be caused by the size of the ROI used in the algorithm compared to the radiologist. The ROI used in the algorithm is the largest possible while being completely certain that no errors were present due to reaching the edge of the ADC ramp. But the values could also have varied due to the specific placement of the ROI where the radiologist avoided all black spots and other noise sources, the algorithm used a specific placement within the ADC ramp to determine a ROI. The minor difference between the algorithm data and the reference data could be explained by the difference in placement and size of the ROI.

A significant difference between the values in the 3 T scanner and the 1.5 T scanners was found, where the mean ADC values for MR1, MR2 and MR4 was respectively; $1,762.53$ (± 42.40), $1,742.23$ (± 101.27) and $3,191.47$ (± 82.82) $\mu m^2/s$. According to an article by Wagner et al. [17] ADC values should be independent of the strength of the magnetic field. This indicated problems with the post-processed ADC maps of the consistency data sets from the 3 T MR scanner. Therefore a calculation of the ADC map was performed from the two b-values of 0 and 1600 s/mm^2 from the 3 T MR scanner. The correction of the ADC values resulted into values between 1660 and

1755 $\mu m^2/s$ as expected based on the article by Wagner et al. This correction indicated an error in the post-processing of the ADC map from the scanner. This was reported to and acknowledged by the Department of Radiology, which shows the necessity and benefits of QA of MR scanners, when working with diffusion weighted imaging.

The temperature correction of the ADC values was performed with an equation from Wagner et al. [17], where the ADC value was changed to the expected values at 20 C^o. The equation changed the R^2 value for the regression line from 0.6, 0.61 and 0.76 to a corrected R^2 value of 0.0, 0.07 and 0.13 for MR1, MR2 and MR4, respectively. After the temperature correction was conducted on the ADC values, the mean ADC for MR1, MR2 and MR4 was respectively $1,634.23$ (± 24.87), $1,632.61$ (± 23.68) and $1,600.60$ (± 21.91) $\mu m^2/s$. These results imply that the temperature corrected mean ADC values are closer to the expected ADC of 1599 $\mu m^2/s$. The 1.96 SD was also lowered by 17.53, 77.59 and 22.85 after the temperature correction for respectively MR1, MR2 and MR4, which indicates a higher precision of the ADC measurements. This means that the temperature correction could be a potential implementation in the algorithm for more precision in the calculation of ADC values, and thereby show a more consistent change in ADC values.

A further algorithm development objective could be to implement manual selection of the correct cross-sectional slices of the ADC ramp insert and the Siemens star. As mentioned before, if the consistency data set is of low quality, the cross-sectional slices would not have a high similarity to the reference images. On the other hand, this might not be considered as a major problem, as a wrongly selected slice of the ADC ramp, because of low quality of the concerned consistency data set, would in one way or another provide a quantification of the mean ADC that is significantly different from the mean and standard deviation of all the other ADC quantifications.
An implementation of temperature corrected ADC calculations in the algorithm could be conducted. This would provide more consistency between calculations of the ADC, as less day to day variance. The temperature of the consistency data set is however manually noted from the thermometer in the phantom and cannot be saved in the metadata of the DICOM images. This would therefore mean that the temperature for the concerned consistency data set would have to be manually typed into the GUI. This might not be a limitation to the algorithm, as the temperature could be typed in at the same time the concerned consistency data set is chosen to be loaded. However when loading multiple consistency data sets, the temperature would have to be typed in for each one of them, which would be time consuming.
An investigation into the versatility of the algorithm to perform QA on different types of phantoms could be of interest, as there is several types of DWI phantoms. Moreover, QA performed on a DWI phantom with different PVP solutions simulating both benign and malignant regions would be of interest, to investigate the day to day variance of

simulated ADC benign and malignant regions.

6.1 Conclusion

Functional diffusion maps (fDM), created by DWI, assess the treatment response with ADC over time, earlier than traditional RECIST methods. DW images may however contain artifacts due to the low signal to noise ratio. Low-quality DW images may impair diagnosis and the assessment of treatment response and therefore QA of the MR scanners has to be conducted to investigate performance. QA is today mostly performed manually by visually inspecting DW images, which may result in low consistency between the examinations. MRI and DWI phantoms are therefore increasingly being used for QA of MR scanners. DWI phantoms includes an ADC region for assessment of the accuracy and precision of ADC measurements, and a Siemens star for assessment of the achieved sharpness of the images. Manual QA of MR scanners with a DWI phantom can be time consuming for the professionals at a hospital department, and an automatic assessment of the achieved sharpness is not yet available. Therefore an automatic algorithm for QA of MR scanners was developed, which quantified the accuracy and precision of the ADC measurements and the achieved sharpness, based on DW images of the DWI phantom.
The automatic algorithm included the development of a GUI, where an automatic slice detection was performed to detect a Siemens star and an ADC ramp within a consistency data set. The algorithm had a 100 % accuracy for detection of the corrects slices from the consistency data sets. The sharpness of the Siemens star was quantified by the Michelson contrast and the minimum line length. The sharpness quantified by the algorithm had a connection to the VAS score evaluated by a radiologist, and was therefore found to be a promising scale for the sharpness of the MR images. A ROI was determined within the benign region to extract the signal intensities of the DW images and the mean ADC of the ADC map. The obtained values were found to have minor differences from the values quantified by a radiologist from the Department of Radiology.
The day-to-day variance was presented in the GUI, so the Department of Radiology can observe the quality of their scanners over time. The algorithm could potentially aid the radiologists to have an overview of the variance between the scanners and the variance from day to day. In addition, a lower completion time for QA of their MR scanners could be achieved, as the completion time of the algorithm was about 15 seconds, while the radiologists use 5 to 7 minutes per QA. The necessity of QA was furthermore proven, when an error in the post-processed ADC map of the 3 T MR scanner was found. The error was reported to the Department of Radiology, that has performed immediate actions in order to investigate and resolve the issue.

Bibliography

[1] Katti G, Ara SA, Shireen A. Magnetic Resonance Imaging (MRI) – A Review. International Journal of Dental Clinics. 2011;3(1):65–70. Available from: https://doi.org/10.1016/j.mrrev.2015.02.002.

[2] Priyanthi AC, Sumalatha BN, Naik V, Sodhi A, Rajshekhar S. Magnetic resonance imaging: Refined to re-find. Journal of Advanced Clinical Research Insights. 2017;4(4):139–142. Available from: https://doi.org/10.15713/ins.jcri.178.

[3] Hagmann P, Jonasson L, Maeder P, Thiran JP, Wedeen VJ, Meuli R. Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond. RadioGraphics. 2006;26(suppl_1):S205–S223. PMID: 17050517. Available from: https://doi.org/10.1148/rg.26si065510.

[4] Chilla GS, Tan CH, Xu C, Poh CL. Diffusion weighted magnetic resonance imaging and its recent trend—a survey. Quantitative Imaging in Medicine and Surgery. 2015 06;5:407–422. Available from: https://doi.org/10.3978/j.issn.2223-4292.2015.03.01.

[5] Baliyan V, Das C, Sharma R, Gupta A. Diffusion weighted imaging: Technique and applications. world journal of radiology. 2016 09;8:785–798. Available from: https://doi.org/10.4329/wjr.v8.i9.785.

[6] Hayashida Y, Hirai T, Morishita S, Kitajima M, Murakami R, Korogi Y, et al. Diffusion-weighted imaging of metastatic brain tumors: Comparison with histologic type and tumor cellularity. AJNR American journal of neuroradiology. 2006 09;27:1419–25. Available from: http://www.ajnr.org/content/27/7/1419.long.

[7] Thoeny H, D Ross B. Predicting and Monitoring Cancer Treatment Response with Diffusion-Weighted MRI. Journal of magnetic resonance imaging : JMRI. 2010 07;32:2–16. Available from: http://dx.doi.org/10.1002/jmri.22167.

[8] Liu Z, Wang Y, Gerig G, Gouttard S, Tao R, Fletcher T, et al.. Quality control of diffusion weighted images; 2010. Available from: https://doi.org/10.1117/12.844748.

[9] Périard MA, Chaloner P. Diagnostic X-Ray Imaging Quality Assurance: An Overview. The Canadian Journal of Medical Radiation Technology. 1996 10;p. 171–177. Available from: https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/radiation/diagnostic-imaging-quality-assurance-overview.html.

[10] Jung BA, Weigel M. Spin echo magnetic resonance imaging. Journal of Magnetic Resonance Imaging. 2013;37(4):805–817. Available from: http://dx.doi.org/10.1002/jmri.24068.

[11] Holdsworth J Samantha, Bammer J Roland. Magnetic Resonance Imaging Techniques: fMRI, DWI, and PWI. Seminars in neurology. 2008;28(4):395–406. Available from: http://dx.doi.org/10.1055/s-0028-1083697.

[12] Kim JH, Kim JK, Park B, Kim N, Cho K. Apparent diffusion coefficient: Prostate cancer versus noncancerous tissue according to anatomical region. Journal of Magnetic Resonance Imaging. 2008 11;28(5):1173–1179. Available from: http: https://doi.org/10.1002/jmri.21513.

[13] Clarke JL, Chang S. Pseudoprogression and pseudoresponse: Challenges in brain tumor imaging. Current Neurology and Neuroscience Reports. 2009 May;9(3):241–246. Available from: https://doi.org/10.1007/s11910-009-0035-4.

[14] Eisenhauer A, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: revised RECIST guideline (version 1.1). European journal of cancer (Oxford, England : 1990). 2008 12;45:228–47. Available from: http://dx.doi.org/doi:10.1016/j.ejca.2008.10.026.

[15] Gatidis S, Schmidt H, Martirosian P, Schwenzer NF. Development of an MRI phantom for diffusion-weighted imaging with independent adjustment of apparent diffusion coefficient values and T2 relaxation times. Magnetic Resonance in Medicine. 2014;72(2):459–463. Available from: http://dx.doi.org/10.1002/mrm.24944.

[16] HQ Imaging;. [Accessed: 24.04.18]. http://hq-imaging.com/.

[17] Wagner F, Laun FB, Kuder TA, Mlynarska A, Maier F, Faust J, et al. Temperature and concentration calibration of aqueous polyvinylpyrrolidone (PVP) solutions for isotropic diffusion MRI phantoms. PLOS ONE. 2017 06;12(6):1–13. Available from: https://doi.org/10.1371/journal.pone.0179276.

[18] Keenan KE, Wilmes LJ, Aliu SO, Newitt DC, Jones EF, Boss MA, et al. Design of a Breast Phantom for Quantitative Magnetic Resonance Imaging. Journal

of Magnetic Resonance Imaging. 2016;44(3):610 – 619. Available from: https://doi.org/10.1002/jmri.25214.

[19] Loebich C, Wuellera D, Klingen B, Jaeger A. Digital camera resolution measurement using sinusoidal siemens stars. SPIE Proceedings. 2007;6502(1):6502 – 6513. Available from: https://doi.org/10.1117/12.703817.

[20] Jodogne S, Bernard C, Devillers M, Lenaerts E, Coucke P. Orthanc – A Lightweight, RESTful DICOM Server for Healthcare and Medical Research. In: Biomedical Imaging (ISBI), IEEE 10th International Symposium on. San Francisco, CA, USA; 2013. p. 190–193. Available from: http://orbi.ulg.ac.be/handle/2268/163949.

[21] Python Software Foundation;. [Accessed: 24.04.18]. http://www.python.org.

[22] Gibaud B. The DICOM standard: a brief overview. 2008 01;.

[23] Python Software Foundation. Shelve;.

[24] Malik J, Belongie S, Leung T, Shi J. Contour and Texture Analysis for Image Segmentation. International Journal of Computer Vision. 2001 Jun;43(1):7–27. Available from: https://doi.org/10.1023/A:1011174803800.

[25] Atherton TJ, Kerbyson DJ. The coherent circle Hough transform. In: in: British Machine Vision Conference, BMVA. Press; 1993. p. 269–278.

[26] Wang Z, Bovik AC. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. IEEE Signal Processing Magazine. 2009 Jan;26(1):98–117. Available from: https://doi.org/10.1109/MSP.2008.930649.

[27] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing. 2004 April;13(4):600–612. Available from: https://doi.org/10.1109/TIP.2003.819861.

[28] Ma Y, Gu X, Wang Y. Histogram similarity measure using variable bin size distance. Computer Vision and Image Understanding. 2010;114(8):981 – 989. Available from: https://doi.org/10.1016/j.cviu.2010.03.006.

[29] Gonzalez RC, Woods RE. Digital Image Processing (3rd Edition). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.; 2006.

[30] OpenCV. Histograms;. [Accessed: 12.05.18]. Available from: https://docs.opencv.org/2.4/modules/imgproc/doc/histograms.html.

[31] Weiss D, Shi Q, Kuhn C. Measuring the 3D resolution of a micro-focus X-ray CT setup. 2012 09;Available from: http://www.ndt.net/article/ctc2012/papers/365.pdf.

[32] Stewart J. Calculus. 3rd ed. Pacific Grove, Calif. : Brooks/Cole Pub. Co; 1995.

[33] Molokach J. Law of Cosines—A Proof Without Words. The American Mathematical Monthly. 2014;121(8):722–722. Available from: http://www.jstor.org/stable/10.4169/amer.math.monthly.121.08.722.

[34] Gower JC. Generalized procrustes analysis. Psychometrika. 1975 Mar;40(1):33–51. Available from: https://doi.org/10.1007/BF02291478.

[35] Ding Y. Visual Quality Assessment for Natural and Medical Image. Berlin, Heidelberg: Springer; 2018. Available from: http://dx.doi.org/10.1007/978-3-662-56497-4.

[36] Kalra A. Decoding the Bland–Altman plot: Basic review. Journal of the Practice of Cardiovascular Sciences. 2017;3:36 – 38. Available from: http://dx.doi.org/10.4103/jpcs.jpcs_11_17.

[37] Klimek L, Bergmann KC, Biedermann T, et al. Visual analogue scales (VAS): Measuring instruments for the documentation of symptoms and therapy monitoring in cases of allergic rhinitis in everyday health care. Allergo Journal International. 2017;26:16 – 24. Available from: http://dx.doi.org/10.1007/s40629-016-0006-7.

[38] Zaitsev M, Maclaren J, Herbst M. Motion Artefacts in MRI: a Complex Problem with Many Partial Solutions. 2015 01;42. Available from: https://doi.org/10.1002/jmri.24850.

[39] Zar JH. Biostatistical Analysis. 5th ed. Pearson Education; 2013.

[40] Wolfsberger S, Ba-Ssalamah A, Pinker K, Mlynárik V, Czech T, Knosp E, et al. Application of three-tesla magnetic resonance imaging for diagnosis and surgery of sellar lesions. Journal of Neurosurgery. 2004;100(2):278–286. PMID: 15086236. Available from: https://doi.org/10.3171/jns.2004.100.2.0278.

[41] Lauren Shapiro BA, Ernesto Staroswiecki MS, Garry Gold MD. MRI of the Knee: Optimizing 3T Imaging. Seminars in roentgenology. 2010;p. 238–249. PMID: 20727453. Available from: https://doi.org/10.1053/j.ro.2009.12.007.

[42] Lavdas I, Miquel ME, McRobbie DW, Aboagye EO. Comparison between diffusion-weighted MRI (DW-MRI) at 1.5 and 3 tesla: A phantom study. Journal of Magnetic Resonance Imaging;40(3):682–690. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.24397.

[43] Sánchez-González J, Luna A. DWI at 3 T: Advantages, Disadvantages, Pitfalls, and Advanced Clinical Applications. 2012 01;p. 51–73. Available from: https://doi.org/10.1007/978-3-642-21052-5_4.

Appendix A

Automatic slice detection

This appendix provides information about the automatic slice selection for MR2 and MR4, where the results of the methods used are displayed. The displayed results are the six histogram correlation method for the Siemens star and the disparity extracted from the Procrustes method for detection of the ADC ramp.

The results for MR2 will be explained, were firstly the six histogram distance measures are shown in two plots. The two methods Correlation and Intersection, where the highest value of the distance measurement corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, is shown in Figure A.1.

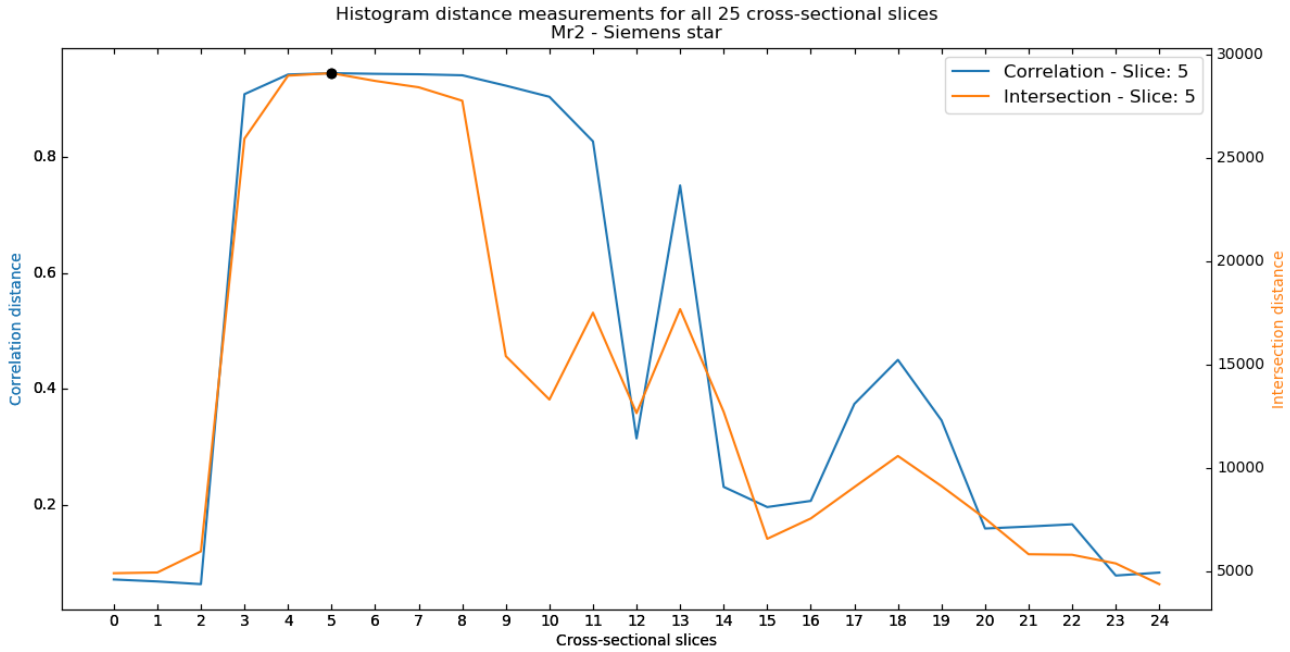


Figure A.1: Results of the two distance measurements correlation and intersection of a consistency data set from MR2, where 25 cross-sectional images were investigated to detect the Siemens star.

The histogram distance measures Chi-squared, Hellinger, Alternative chi-squared and the Kullback-Leibler divergence, where the minimum values corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, as seen in Figure A.2.

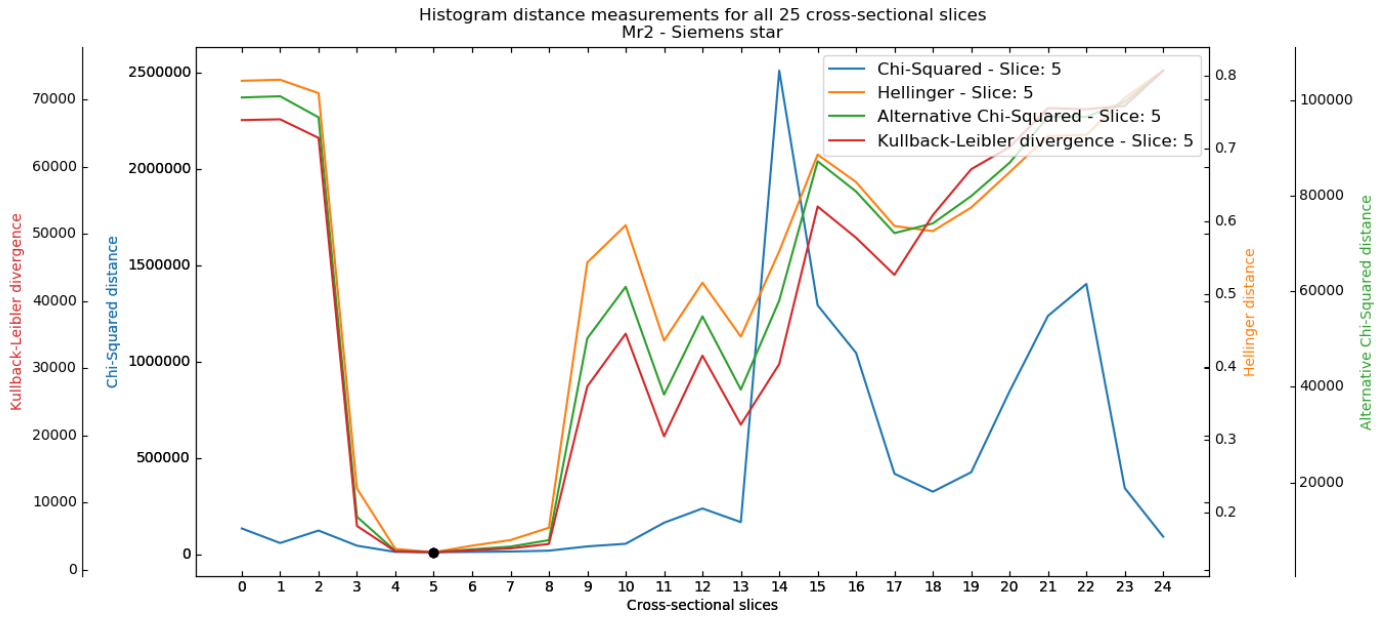


Figure A.2: Results of the four distance measurements Chi-Squared, Hellinger, Alternative Chi-Squared and Kullback-Leibler divergence of a consistency data set from MR2, where 25 cross-sectional images were investigated to detect the Siemens star.

In this example the majority vote ended by choosing slice number 5, where all methods voted for the selected cross-sectional slice.

Slice number 5 which were chosen by the majority vote can be seen in Figure A.3.

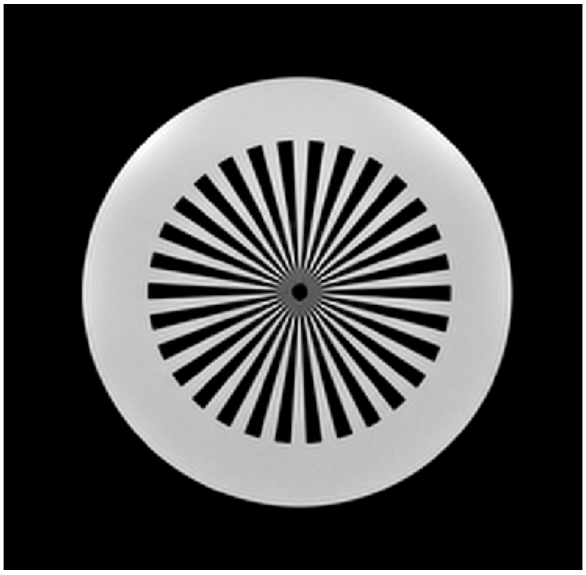


Figure A.3: The result of the automatic slice detection found to be slice 5.

The Procrustes method resulted in a disparity value, where the lowest disparity corresponds to the cross-sectional slice with the highest similarity to the reference image of the ADC ramp. This disparity value is visualized in Figure A.4 for MR2.

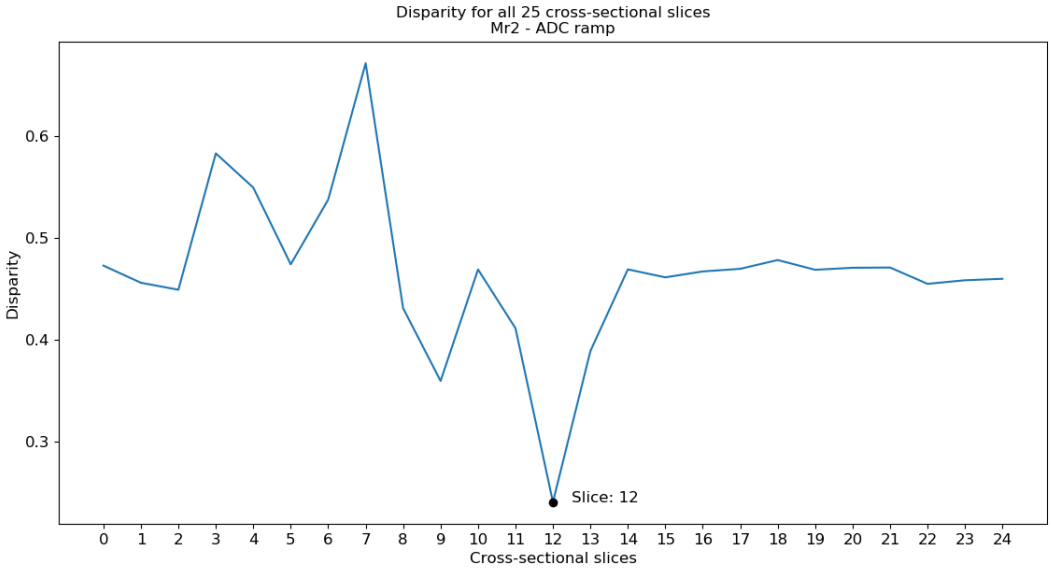


Figure A.4: Results from the disparity value of a consistency data set from MR2, where 25 cross-sectional images were investigated to detect the ADC ramp.

The results from the automatic slice selection of the ADC ramp showed an disparity value of 0.2409. The found image of the ADC ramp can be seen in Figure A.5, which were selected by the lowest disparity of the Procrustes method.

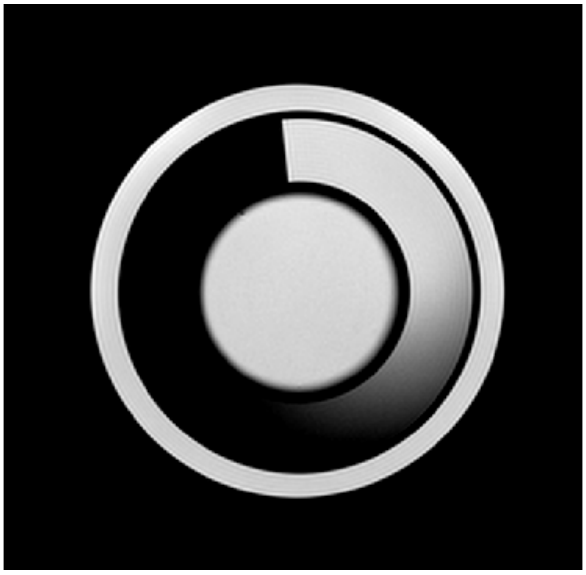


Figure A.5: The result of the automatic slice detection found to be slice 12 by the lowest disparity.

Now the results for MR4 is explained, were firstly the six histogram distance measures is shown in two plots. The two methods Correlation and Intersection, where the highest value of the distance measurement corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, is shown in Figure A.6.

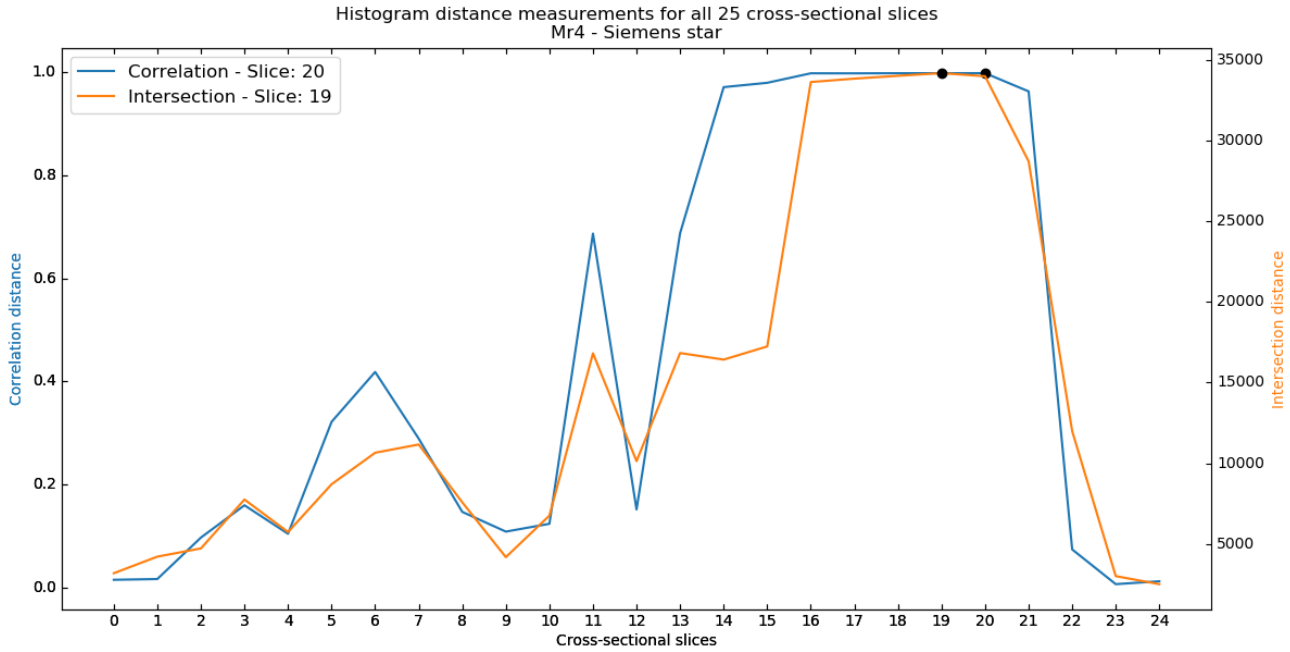


Figure A.6: Results of the two distance measurements correlation and intersection of a consistency data set from MR4, where 25 cross-sectional images were investigated to detect the Siemens star.

The histogram distance measures Chi-squared, Hellinger, Alternative chi-squared and the Kullback-Leibler divergence, where the minimum values corresponds to the cross-sectional image with the highest similarity to the reference image of the Siemens star, as seen in Figure A.7.

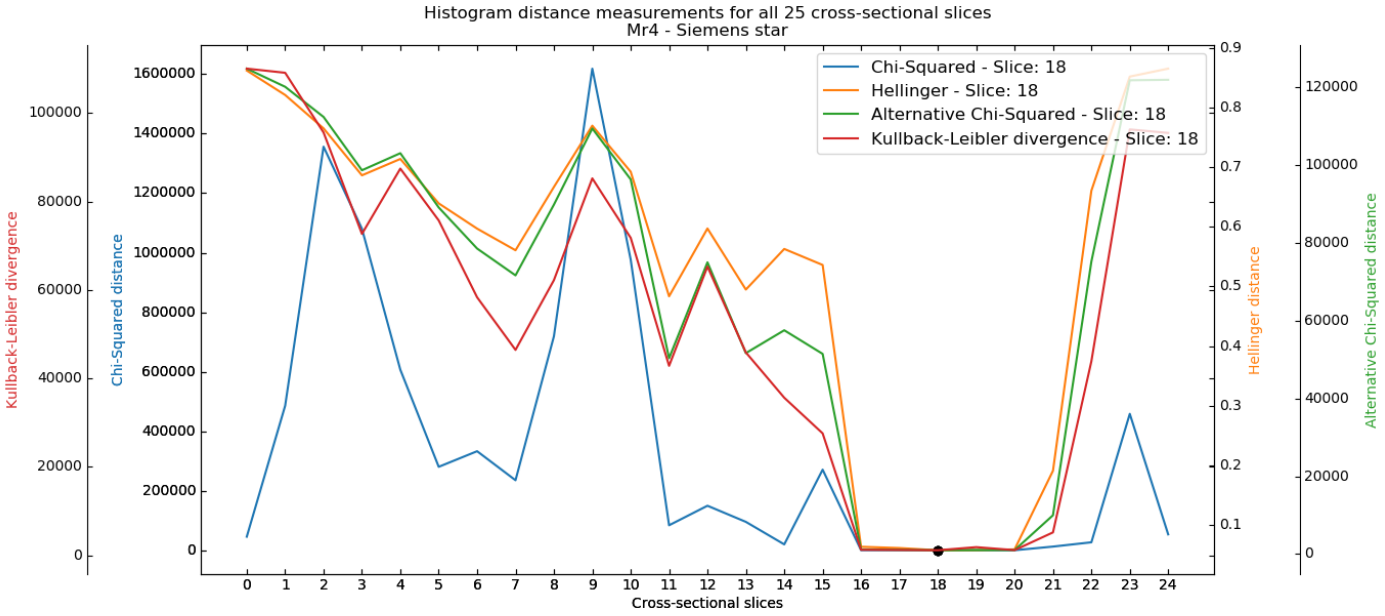


Figure A.7: Results of the four distance measurements Chi-Squared, Hellinger, Alternative Chi-Squared and Kullback-Leibler divergence of a consistency data set from MR4, where 25 cross-sectional images were investigated to detect the Siemens star.

In this example the majority vote ended by choosing slice number 18, where 4 out of 6 voted for the selected cross-sectional slice. The histogram distance measurements which voted for the slice were Chi-Squared, Hellinger, Alternative Chi-squared and the Kullback-Leibler divergence, while the last two methods Correlation and Intersection voted for slice number 20 and 19, respectively.

Slice number 18 which were chosen by the majority vote can be seen in Figure A.8.

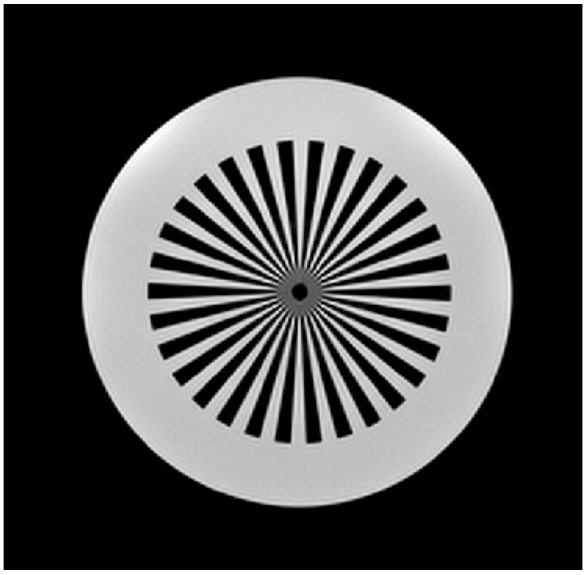


Figure A.8: The result of the automatic slice detection found to be slice 18.

The Procrustes method resulted in a disparity value, where the lowest disparity corresponds to the cross-sectional slice with the highest similarity to the reference image of the ADC ramp. This disparity value is visualized in Figure A.9 for MR4.

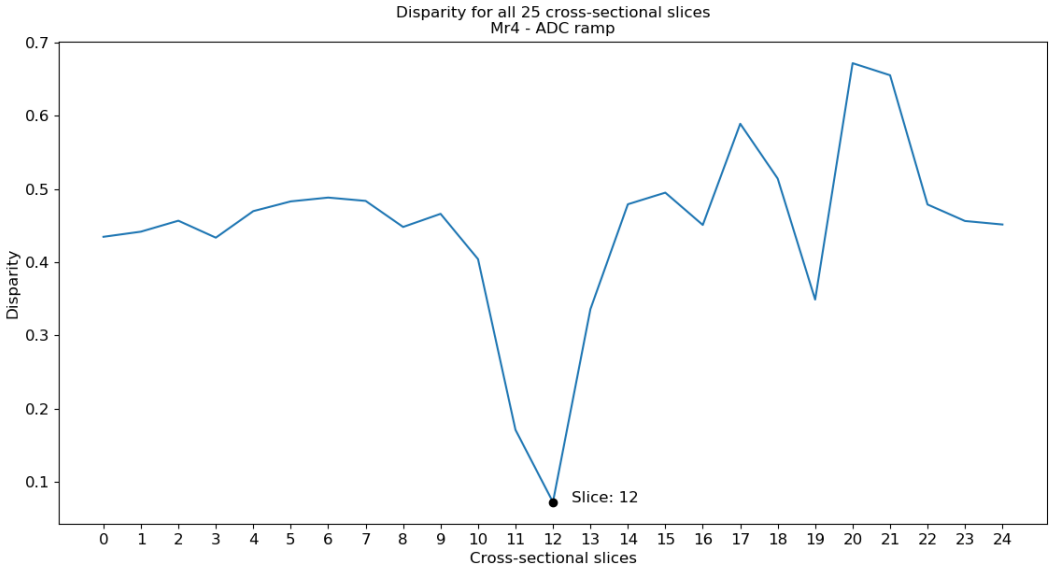


Figure A.9: Results from the disparity value of a consistency data set from MR4, where 25 cross-sectional images were investigated to detect the ADC ramp.

The results from the automatic slice selection of the ADC ramp showed an disparity value of 0.0723. The found image of the ADC ramp can be seen in Figure A.10, which were selected by the lowest disparity of the Procrustes method.

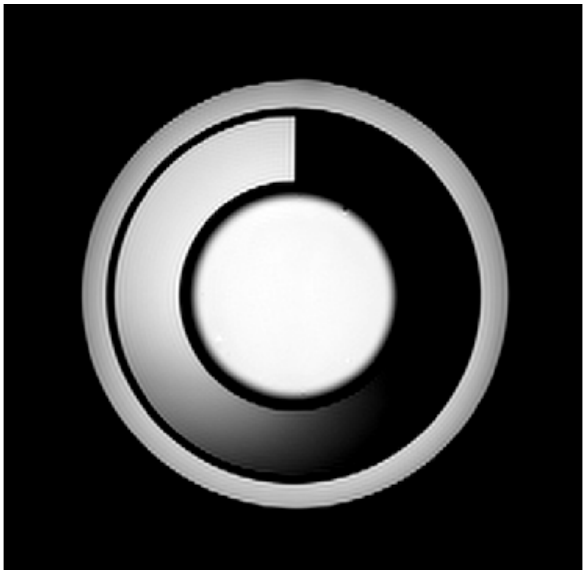


Figure A.10: The result of the automatic slice detection found to be slice 12 by the lowest disparity.

After the IQA explained in Section 4.3.1 was performed the automatic slice detection of 30 consistency data sets for MR1 and MR4 and 29 for MR2 all images chosen by the algorithm were chosen to be acceptable which concludes the automatic slice detection with an accuracy of 100 % for both the ADC ramp and the Siemens star.

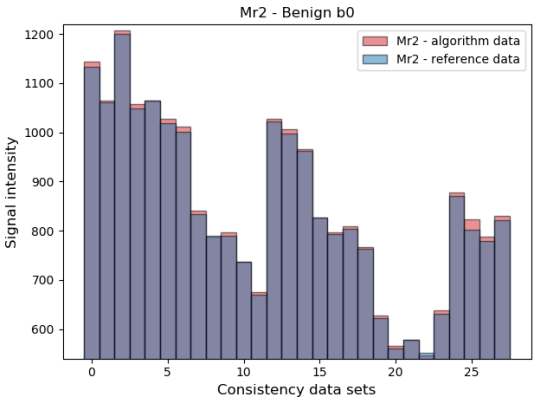
Appendix B

Quantification of the ADC benign values

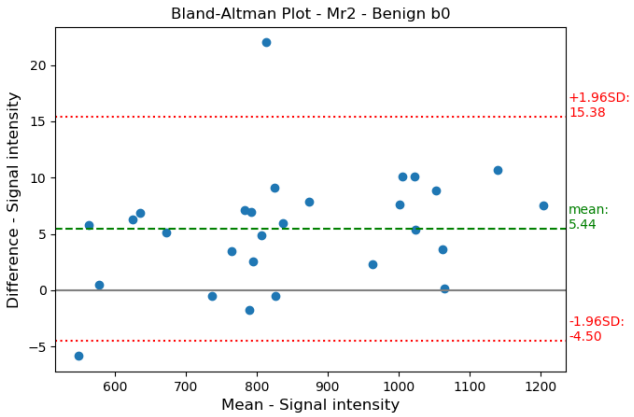
This appendix includes the results of the comparison of the algorithms quantification of the ADC benign values compared to the values from the Department of Radiology, defined as reference data.

The results contain the data from the algorithm in 28 consistency data sets for MR2 and 26 for MR4. One consistency data set from MR2 were removed because of missing data from the Department of Radiology, likewise were four consistency data sets removed from MR4 because of missing data. The sequences used in this result section are b0, b1600 and the ADC map, which are the sequences the department of radiology delivered results for. The values compared to the radiologist are for the ADC benign region.

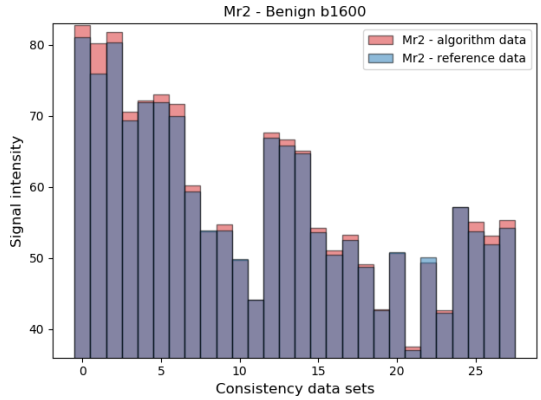
Figure B.1 illustrates the results of the b0, b1600 and ADC sequences from MR2 with the benign values.



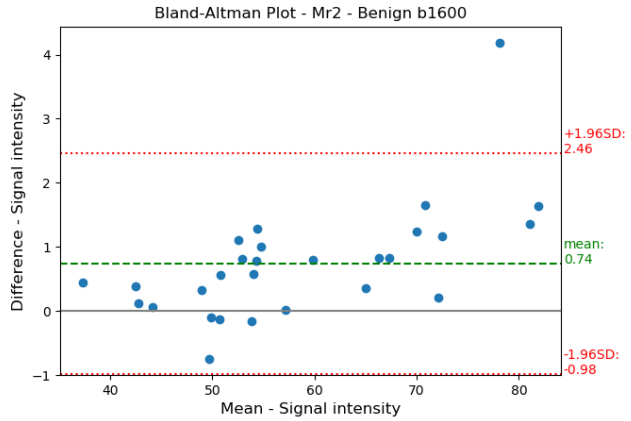
(a) Results from MR2 of b0.



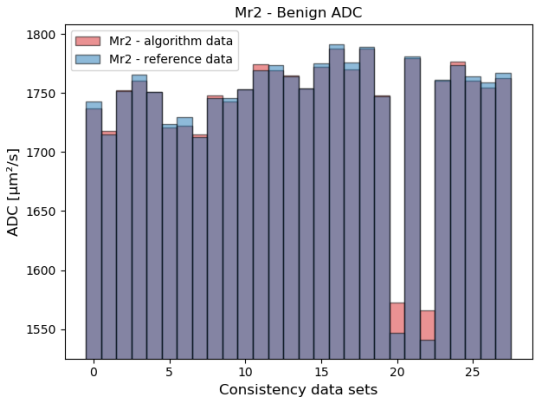
(b) Bland-Altman plot for b0 from MR2.



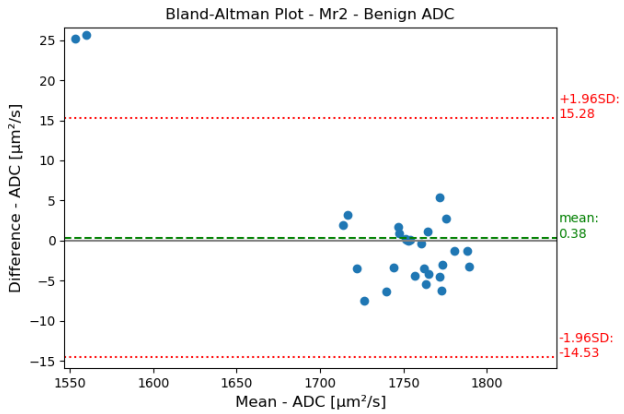
(c) Results from MR2 of b1600.



(d) Bland-Altman plot for b1600 from MR2.



(e) Results from MR2 of ADC.



(f) Bland-Altman plot for ADC from MR2.

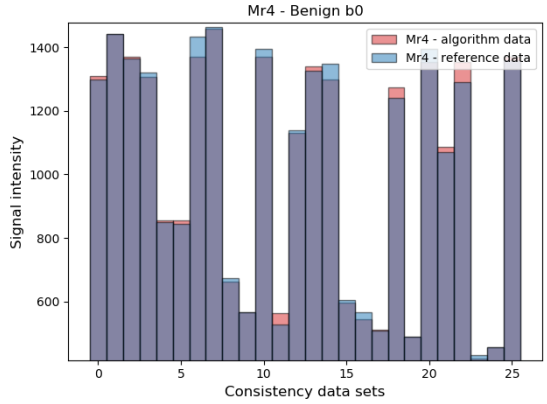
Figure B.1: Results from MR2 of b0, b1600 and ADC of the algorithm compared to the reference data for 28 consistency data sets and the differences between the signal intensities and ADC are illustrated in BA plots.

As seen in Figure B.1a the benign values for the b0 varied between 552 and 1200 in signal intensity, and had a mean of 847.68 in the reference data. The algorithm data varied between 546 and 1208 with a mean value of 853.12, which gives a difference of 5.44 in the mean between the two groups. The Bland-Altman plot seen in Figure B.1b illustrates the differences in values between the two methods of quantifying the benign b0 values. The maximum differences between the two methods of quantifying data were two values outside the limits of agreements (LoA) which were 21.99 and -5.83. The mean of the differences was 5.44, where the 95 % lower and upper LoA were between -4.5 and 15.38.

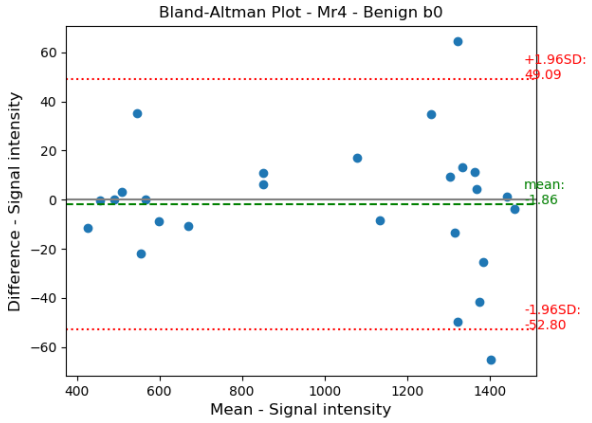
Next the b1600 values for MR2 are illustrated in Figure B.1c, where the signal intensities varied between 37.1 and 81.1 in the reference data, which had a mean of 58.05. The algorithm data varied between 37.5 and 82.7 with a mean value of 58.79, which gives a difference of 0.73 in the mean between the two groups. The Bland-Altman plot seen in Figure B.1d illustrates the differences in b1600 values between the two methods of quantifying the benign b1600 values. The maximum difference between the two methods of quantifying data was 4.19, which where the only data point outside the LoA. As seen on the image the mean of the difference was 0.74, where the 95 % lower and upper LoA were between -0.98 and 2.46.

Lastly Figure B.1e the quantified ADC values varied between 1541 and 1791 $\mu m^2/s$ in the reference data. The mean of the ADC values where 1740.23 $\mu m^2/s$. The algorithm data varied between 1566 and 1787 $\mu m^2/s$ with a mean value of 1742.23 $\mu m^2/s$, which gives a difference of 2.0 in the mean between the two groups. The Bland-Altman plot seen in Figure B.1f illustrates the differences in ADC between the two methods of quantifying the ADC benign values. The maximum differences between the two methods of quantifying data were 25.19 and 25.60, which where the two data points outside the LoA. As seen on the image the mean of the difference was 0.38 $\mu m^2/s$, where the 95 % lower and upper LoA were between -14.53 and 15.28 $\mu m^2/s$.

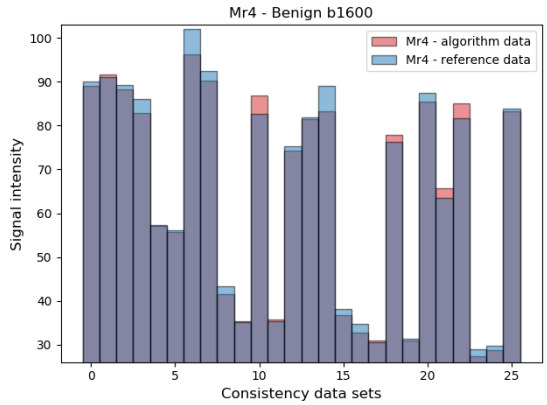
Figure B.2 illustrates the results of the b0, b1600 and ADC sequences from MR4 with the benign values.



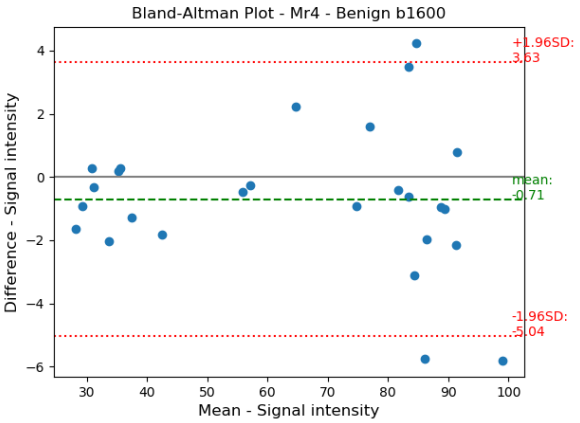
(a) Results from MR4 of b0.



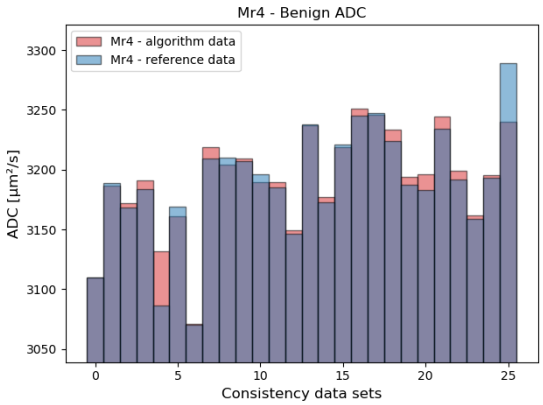
(b) Bland-Altman plot for b0 from MR4.



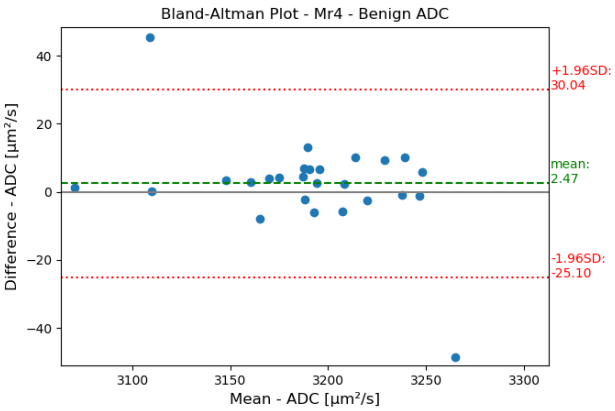
(c) Results from MR4 of b1600.



(d) Bland-Altman plot for b1600 from MR4.



(e) Results from MR4 of ADC.



(f) Bland-Altman plot for ADC from MR4.

Figure B.2: Results from MR4 of b0, b1600 and ADC of the algorithm compared to the reference data for 26 consistency data sets and the differences between the signal intensities and ADC are illustrated in BA plots.

As seen in Figure B.2a the benign values for the b0 varied between 431 and 1463 in signal intensity, and had a mean of 1015.62 in the reference data. The algorithm data varied between 419 and 1459 with a mean value of 1013.76, which gives a difference of 1.85 in the mean between the two groups. The Bland-Altman plot seen in Figure B.2b illustrates the differences in values between the two methods of quantifying the benign b0 values. The maximum differences between the two methods of quantifying data were two values outside the LoA which were -65.23 and 64.38. The mean of the differences was -1.86, where the 95 % lower and upper LoA were between -52.8 and 49.09.

Next the b1600 values for MR4 are illustrated in Figure B.2c, where the signal intensities varied between 28.9 and 102 in the reference data, which had a mean of 65.10. The algorithm data varied between 27.2 and 96.2 with a mean value of 64.39, which gave a difference of 0.71 in the mean between the two groups. The Bland-Altman plot seen in Figure B.2d illustrates the differences in b1600 values between the two methods of quantifying the benign b1600 values. The maximum differences between the two methods of quantifying data were three data points outside the LoA, which were 4.25, -5.77 and -5.82. As seen on the image the mean of the difference was -0.71, where the 95 % lower and upper LoA were between -5.04 and 3.63.

Lastly Figure B.2e the quantified ADC values varied between 3070 and 3289 $\mu m^2/s$ in the reference data. The mean of the ADC values where 3189 $\mu m^2/s$. The algorithm data varied between 3071.16 and 3250.88 $\mu m^2/s$ with a mean value of 3191.47 $\mu m^2/s$, which gives a difference of 2.47 in the mean between the two groups. The Bland-Altman plot seen in Figure B.2f illustrates the differences in ADC between the two methods of quantifying the ADC benign values. The maximum differences between the two methods of quantifying data were two data points outside the LoA, which were -48.64 and 45.51. As seen on the image the mean of the difference was 2.47 $\mu m^2/s$, where the 95 % lower and upper LoA were between -25.1 and 30.04 $\mu m^2/s$.