A THESIS
Presented to the Department of Medicine
of the University of Padua

In Partial Fulfillment
of the Requirements for the
Master's Degree in Medicine and Surgery

CT Effective Dose in Critical Patients: Comparison between Deep Learning Image Reconstruction (DLIR), Filtered Back Projection (FBP) and Iterative Algorithms

Under Supervision of
Prof. Quaia Emilio, MD
Chief of Radiology Department

and Examination of
Prof. Navalesi Paolo, MD
Director of the School of Anaesthesia and Intensive Care

By Lanza de Cristoforis Elena Kiyomi
June 2023

What came before the Big Bang? According to the no-boundary proposal, asking what came before the Big Bang is meaningless—like asking what is south of the South Pole—because there is no notion of time available to refer to. The concept of time only exists within our universe.

— Stephen Hawking, Brief Answers to the Big Questions

Table of Contents

I Abstract ix

List of Figures xiii

List of Tables xv

II Introduction 1

1 Principles of Computed Tomography (CT) 3

1.1 Acquisition 3

1.2 Reconstruction 5

1.3 Visualization 7

1.3.1 Hounsfield Scale and Hounsfield Units: a Visual Interpretation . . . 8

1.3.2 Lambert–Beer's Law, Linear Attenuation Coefficient and CT-Value 9

1.3.3 Image Noise 13

1.3.4 Signal to Noise Ratio (SNR) 15

1.4 Dosimetry 15

1.4.1 X-ray Beam Energy and Electronvolts (eV) 16

1.4.2 Absorbed Dose and Gray (Gy) 17

1.4.3 Computed Tomography Dose Index (CTDI) 17

1.4.4 Dose Length Product (DLP) 19

1.4.5 Effective Dose (E), Equivalent Dose (H_T) and Sieverts (Sv) 20

1.4.6 Effective Dose Conversion Factor E_{DLP} 22

2 Fundamentals of Reconstruction 25

2.1 Main Categories of Algorithms 25

2.2 Filtered Back Projection (FBP) 27
2.3 Iterative Algorithms 29
2.3.1 AIDR3D 32
2.3.2 ADMIRE 32
2.4 DLIR 32

3 DLIR 33
3.1 Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) 33
3.2 Deep Learning Image Reconstruction (DLIR) 40
3.2.1 TrueFidelity 41

III Main Body of the Study 43

4 Ojectives 45

5 Materials and Methods 47
5.1 Sudy Design 47
5.2 Patient Population 50
5.3 CT Protocol 50
5.4 Quantitative Dose Analysis 51
5.5 Quantitative Image Quality Analysis 51
5.6 Data Analyses 51
5.7 Statistical Analyses 52
5.7.1 Saphiro-Wilk Test 52
5.7.2 Paired t-Test 53
5.7.3 Wilcoxon Signed-Rank Test 53
5.7.4 Cohen’s d 53

6 Results 55
6.1 Patient Demographics 55
6.2 Quantitative Dose Analyses 55
6.3 Quantitative Image Quality Analyses 58

7 Discussion 59

8 Conclusion 63

IV Appendices 65

A Delve Deeper into Data 67
A.1 Results stratified by acquisition phase 67
A.1.1 Results in FBP vs DLIR 67
A.1.2 Results in AIDR3D (iterative) vs DLIR 68
A.1.3 Results in ADMIRE (iterative) vs DLIR 69
A.2 Patient Demographics 70
A.2.1 Patient Demographics in FBP vs DLIR 70
A.2.2 Patient Demographics in AIDR3D (iterative) vs DLIR 70
A.2.3 Patient Demographics in ADMIRE (iterative) vs DLIR 70
A.3 Assessing Normality of the Distribution 71
A.3.1 Assessing Normality in FBP vs DLIR 71
A.3.2 Assessing Normality in AIDR3D (iterative) vs DLIR 71
A.3.3 Assessing Normality in ADMIRE (iterative) vs DLIR 72

B Foreword to the Source Code Appendices 73

C BP Simulation in MATLAB 75

D FBP Simulation in MATLAB 77

E Global Noise Level in MATLAB 79
E.1 For a single image 79
E.2 Iterative version 84

F Cohen's d computation in Python 87

Part I

Abstract

Abstract

Background: Since the introduction of computed tomography (CT) imaging in 1971, different algorithms have been developed to reconstruct images starting from the countless attenuation values acquired by the rotating detector. Such algorithms significantly impact radiation exposure, since a reconstruction that improves image quality at a given dose translates into the possibility of achieving the same baseline quality required for diagnosis at a lower dose. At the time of writing, three main categories of reconstruction methods have been developed: filtered back projection (FBP) in 1972, iterative reconstruction (IR) in 2008 and deep learning image reconstruction (DLIR) in 2018. DLIR has been proven effective in reducing radiation dose compared with the previous techniques in both phantom studies and studies where patients were randomly assigned to be examined using either one method or the other. However, studies regarding how far this reduction may have an impact on the same individual in real clinical settings are still limited.

Objective: To evaluate if there is radiation dose reduction and image quality improvement in the same acute patient when using DLIR in comparison to FBP and IR in daily clinical practice.

Methods: This retrospective study included 83 critical care patients who underwent CT imaging of the same anatomical region multiple times within a period of 30 days using both DLIR (TrueFidelity) and FBP or IR (AIDR3D and ADMIRE). Regions included were chest, abdomen and trunk (chest + abdomen). All examinations were performed using automatic exposure control (AEC) which modulates the tube current and hence radiation exposure according to the algorithm applied. Radiation dose was assessed using CT dose index volume (CTDI volume), dose-length product (DLP) and Effective Dose. For the quantification of image quality, Noise and Signal to Noise Ratio (SNR) were used. All parameters were compared across the different reconstruction methods for each patient using both parametric and non-parametric testing. In cases of contrast-enhanced CT (CECT), all parameters were retrieved for every acquisition phase (direct, arterial, venous or delayed) as well as for their total value as stated in the patient protocol.

Results: Our analysis suggested that DLIR majorly and consistently outperformed the traditional techniques with regards to image quality, and to a lesser extent, it correlated with dose reduction.

Specifically on average, the total values for FBP were 24.67 ± 61.01 mGy for CTDI,

1350.62 ± 1191.68 mGy * cm for DLP, 20.19 ± 17.91 mSv for Effective Dose, image noise was 28.85 ± 32.77 HU, and SNR was 3.99 ± 1.23 HU.

Those values were improved in DLIR: 9.56 ± 5.86 mGy for CTDI, 1085.33 ± 626.30 mGy * cm for DLP, 16.13 ± 9.55 mSv for Effective Dose, image noise was 8.45 ± 3.24 HU, and SNR was 11.53 ± 9.28 HU.

Regarding IR, the total dose was not found to be affected by the use of DLIR, but image quality was improved. The mean values for examinations with IR were 14.00 ± 12.46 mGy for CTDI, 1235.53 ± 873.67 mGy * cm for DLP, 18.45 ± 13.16 mSv for Effective Dose, image noise was 14.85 ± 2.73 HU, and SNR was 4.84 ± 2.74 HU.

Conclusion: According to our study, DLIR provides benefits in terms of dose and image quality over the traditional FBP. It also outperforms IR methods for image quality, but not for dose. Further research is needed to see if those improvements translate into safer imaging practices, higher diagnostic confidence, and ultimately better patient care.

Keypoints:

- Compared to FBP, DLIR both reduces radiation dose and improves image quality;
- Compared to IR, DLIR doesn't necessarily reduce radiation dose, but it improves image quality.

List of Figures

1.1 Basics of a CT scan 4
1.2 Basics of Scan Modes 5
1.3 Workflow of a CT scan 5
1.4 Projection of a 2 x 2 matrix 6
1.5 Forward and Inverse Problem 7
1.6 Transverse Section of the Abdomen 8
1.7 Hounsfield scale 9
1.8 Transverse Section of the Abdomen, with HU values 10
1.9 Lambert-Beer's Law 11
1.10 Noisy vs Denoised Image of a cat 13
1.11 Workflow for calculating noise and SNR 15
1.12 Electronvolt eV 16
1.13 Structure of a typical PMMA Phantom 18
1.14 Pitch 19
1.15 Overview of Dosimetry Parameters 23

2.1 Algebraic Reconstruction Technique 26
2.2 Sinogram 28
2.3 BP simulation 28
2.4 FBP simulation 29
2.5 FBP simulation with doubled projections 29
2.6 Iterative Reconstruction simulation 31

3.1 Subcomponents of Artificial Intelligence 34
3.2 Classical Programming vs Machine Learning 34

3.3 Evaluation of classifiers 35
3.4 Classic Machine Learning vs Deep Learning 37
3.5 Number of parameters in Classic Programming and Deep Learning 39
3.6 ANN, CNN and RNN 39
3.7 TrueFidelity Algorithm by GE Healthcare 41
3.8 Image reconstruction with FBP vs IR vs DLIR 42

5.1 Study design 47
5.2 Cohen's d 54

List of Tables

I HU Cut-Off Values for Different Tissue Types 9

II Linear Attenuation Coefficients μ of Human Tissues, Air and Water for Photon Energies of 60–150 keV. 12

III Tissue Weighting Factors w_T 21

IV Effective Dose Conversion Factors E_{DLP} 22

V CT scanners at the University Hospital of Padua 48

VI Dose Analysis Results 57

VII Image Quality Analysis Results 58

VIII Dose reduction results for FBP vs DLIR: an alternative quantification . . . 67

IX Quality improvement results in FBP vs DLIR: an alternative quantification . 68

X Dose reduction results for AIDR3D vs DLIR: an alternative quantification . 68

XI Quality improvement in AIDR3D vs DLIR: an alternative quantification . . 69

XII Dose reduction results for ADMIRE vs DLIR: an alternative quantification . 69

XIII Quality improvement in ADMIRE vs DLIR: an alternative quantification . . 69

Part II

Introduction

Chapter 1

Principles of Computed Tomography (CT)

1.1 Acquisition

A Computed Tomography scan, more commonly referred to as a CT scan, is a medical imaging method used to visualize the internal structures of a body by reconstructing its cross-sectional images in a non-invasive way. It was developed by South African-American physicist Allan M. Cormack [1] and British engineer Godfrey N. Hounsfield [2], who shared the 1979 Nobel Prize in Physiology or Medicine for their invention [3].

In a CT scan, as shown in Figure 1.1, we have an X-ray tube, or X-ray source, that sends X-rays towards the body. Positioned on the opposite side of the body there is a row of detectors that measure how many of the sent rays reach the other side, and to which extent they are attenuated by the body. The X-ray tube then rotates slightly, together with the detectors, thanks to some rails inside the gantry, which is the circular structure around the body. This slight rotation allows to acquire a new set of values from a different angle. This process is repeated in a definite pattern of movement until the whole circumference of the body is covered. The data obtained from one complete rotation allows to reconstruct the image of that specific slice of the body, so if one wishes to examine a wider range of the patient's height, their body has to be moved with regards to the gantry to make new acquisitions for every slice, until the whole volume of interest is covered [5]. In modern CT scans, there is a motorized platform that continuously moves the patient during the rotation

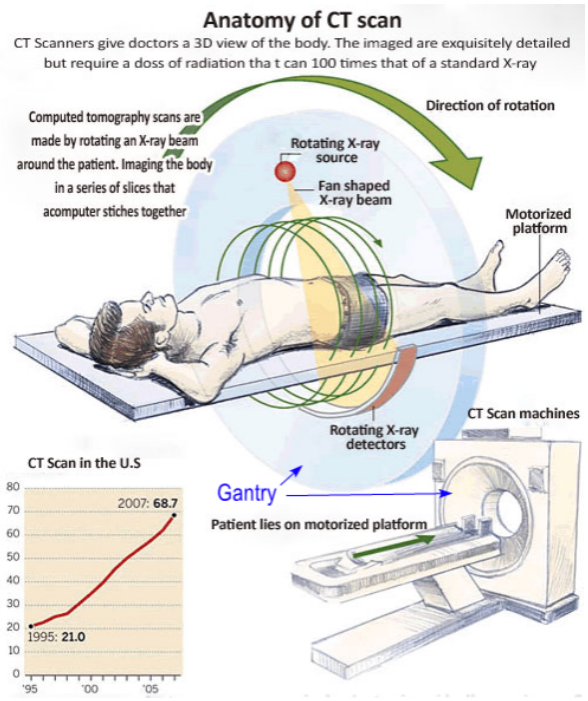

Figure 1.1: Basics of a CT scan. Image reproduced from mepids.com [4].

of the x-rays and detector, and as a result, the collected data has a helical pattern, as depicted in Figure 1.2.

Once the CT values are obtained, they can be projected to form what is formally known as a sinogram. The sinogram is then processed by an algorithm to generate the reconstructed, or tomographic, images, as shown in Figure 1.3. They are a representation of the slices of the body.

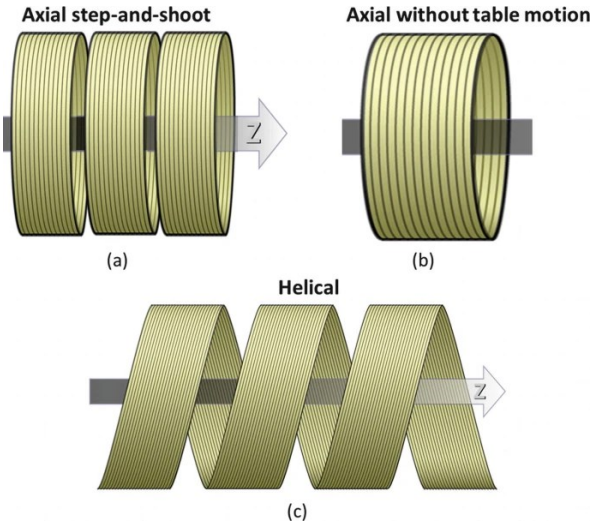

Figure 1.2: Basics of Scan Modes: (a) axial step-and-shoot scan with table motion, (b) axial scan without table motion, and (c) helical scan. Reproduced from Yu [6].

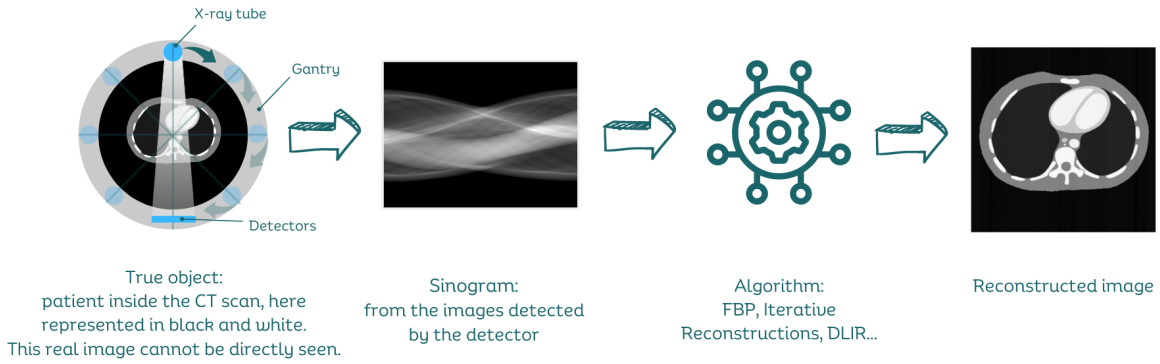

Figure 1.3: Workflow of a CT scan. Once the projections of the true object are made, it is the algorithm that is charged to analyse the sinogram to generate the reconstructed image.

1.2 Reconstruction

Mathematically speaking, to perform an image reconstruction is to compute the spatial structure of an object that casts shadows around it using exclusively the shadows themselves as a starting point. Being each shadow two-dimensional, it is inherently insufficient, by itself, for determining the spatial distribution of a three-dimensional object. This is why there is the necessity of casting shadows in multiple directions, i.e. to have the X-ray tube and the detectors rotate around the object [7], as exemplified in Figure 1.4.

In mathematics, this type of challenge posed by CT scans to reconstruct the original object distribution is known as the inverse problem, as opposed to the forward problem. Their difference is illustrated in figure 1.5.

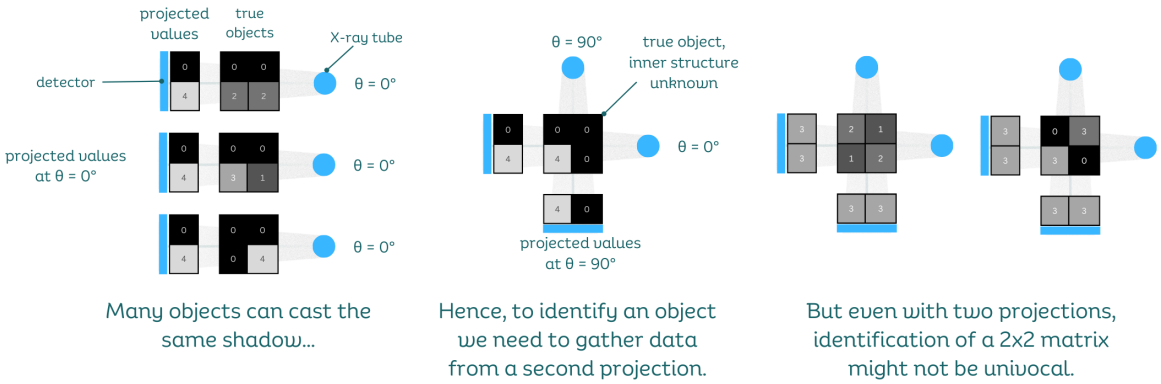

Figure 1.4: Projection of a matrix. A matrix is a set of numbers or equations ordered in rows and columns. The shadow loses one dimension when it is projected, so for a 2 x 2 matrix, a second projection from θ = 90° is needed. However, even in this simple case, two projections might not be sufficient: in the last example we see two different matrices that are characterised by the same projections. In this case, we would need a third projection from θ = 45° to univocally identify the object: the values projected from this additional angle would be [2, 2, 2] in the left image, and [0, 6, 0] in the right image, giving the opportunity to differentiate the two matrices. This model may seem trivial, but it illustrates how easily the number of needed projections increases when the complexity of the image also increases.

More specifically, in a forward problem, the goal is to make predictions of some measurements when given a comprehensive description of a system, so in our case, to predict projections B when given the model A. A forward problem is in a sense very convenient because in a deterministic system, it is characterized by a unique solution, however the same does not hold true for its counterpart.

In an inverse problem, the aim is to infer the values of the parameters that characterize the same deterministic system when provided with the actual results of some measurements, so in our case, to predict model A while only having projections B. As foreshadowed by the description of Figure 1.5, this type of problem can have a set of multiple, and at times infinite, solutions [8]. For this reason, reconstructing a unique, sharp and close-to-the-true body image starting from a limited number of CT attenuation values is a particularly complex and daunting problem, and throughout the years, multiple algorithms have been developed to improve the outcome. Of course, increasing the number of information gathered would simplify the solution, however, the radiation dose to the patient would concomitantly increase.

A curious interpretation of this is that regardless of the specific algorithm employed, the solution to the problem (i.e., the reconstructed image), is nothing more than the most probable solution among a set of many, selected by means of the algorithm, using the available

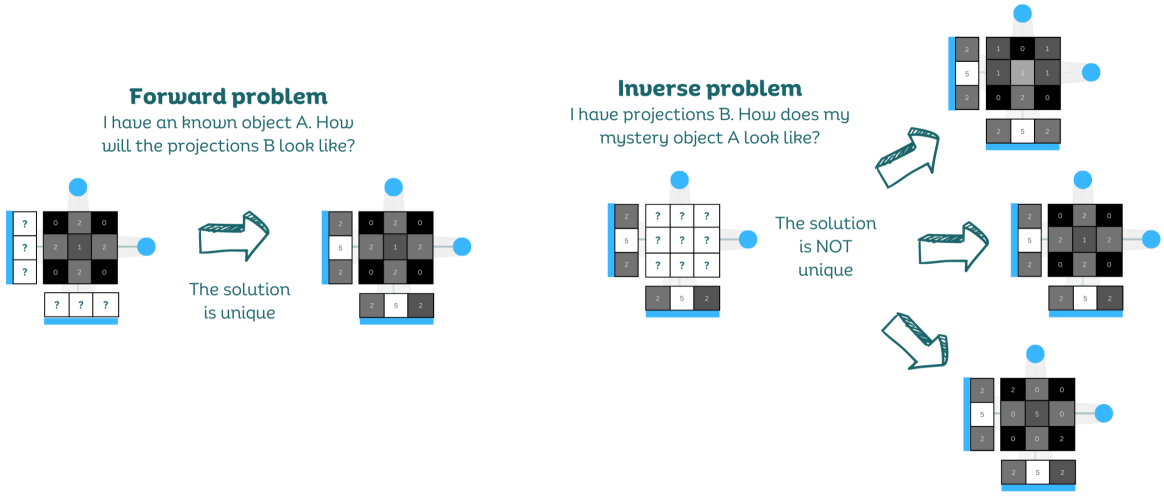

Figure 1.5: Forward and inverse problem. Some inverse problems even have infinite solutions.

information, i.e., the CT attenuation values. In other words, a CT scan is not a photograph in the traditional sense, or as an X-ray image might be, but it is rather a conjecture. To reconstruct an image is to predict, or to make an educated guess of, how the structures of the body would look like if we did indeed slice it open to take direct pictures of the sections. The accuracy of those predictions largely relies on the quantity of data we can start with, namely the CT attenuation values, and on the performance of our predictor, the algorithm. This is the reason why some lesions are more easily detected using some algorithms than others [9].

1.3 Visualization

The visualization of the final image plays a crucial role in interpreting and extracting meaningful information in clinical settings. This subsection aims to provide a comprehensive yet accessible overview of the terms and concepts employed in this study when assessing and comparing the quality of the images produced by the different algorithms:

1. Hounsfield Scale and Hounsfield Units: a Visual Interpretation;
2. Linear Attenuation Coefficient and CT-Value;
3. Image Noise;
4. Signal to Noise Ratio (SNR).

1.3.1 Hounsfield Scale and Hounsfield Units: a Visual Interpretation

To understand what Hounsfield Scale and Hounsfield Units are and why they are important, it's useful to first observe Figure 1.6:

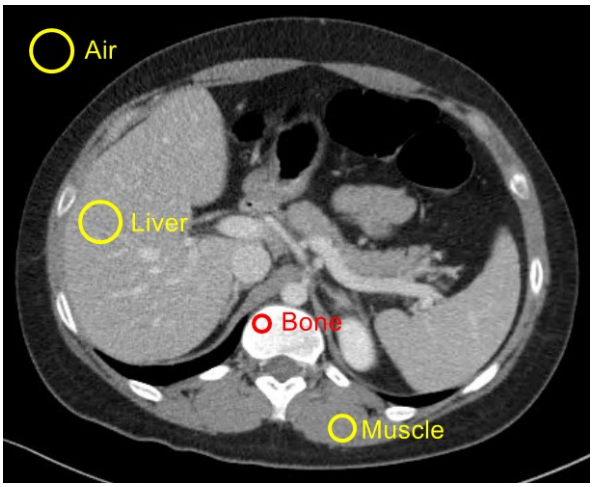

Figure 1.6: Transverse Section of the Abdomen. How would you describe the colors of the air, liver, bones and muscles?

How could the colors of the structures in the image be described? Of course, an idea would be to employ adjectives such as white, black and grey, or to state that a structure is darker or lighter when compared to another one. Those descriptive terms surely have the advantage of being easily understood by a wide population, but being categorical variables, they lack the measurability that comes with a quantitative variable. In other words, looking at the image we see that both the liver and muscles are grey. But how grey are they? Do they share the same shade of grey?

To answer these questions, the Hounsfield scale in Figure 1.7 was introduced. It is a scale that conventionally sets 0HU, or Hounsfield Units, to the shade of distilled water and -1000 HU to that of the air, both acquired under standard temperature and pressure. Everything else is then measured in relationship to them.

In theory, this scale is open ended, so reach beyond -1000 and 0. It is very understandable, since we can both have materials that are less dense than air (for example, if we were to measure a helium balloon inside of a CT), and materials that are denser than water, such a bone. In clinical practice, most values range between -1000HU and +3000HU, covering about 4000 different shades. It is an enormously precise scale: for a comparison, the human eye can perceive no more than around 900 shades on a grayscale, even when using high

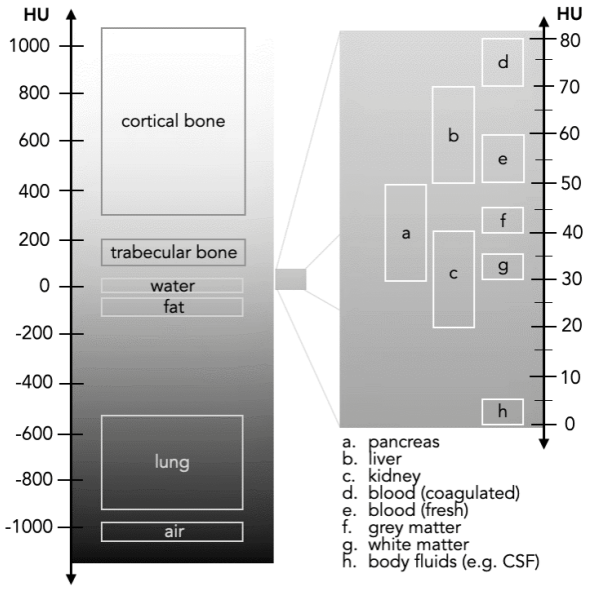

Figure 1.7: Hounsfield scale: a visual interpretation with corresponding tissues. Reproduced from Hartung and Cadogan [10].

brightness and high contrast displays [11]. Machines are inherently better judges of color variation than we are.

HU values are commonly used to differentiate between different types of tissue. In the code presented in Appendix E, the cut-off values proposed by Christianson et al. [12] and shown in Table I were taken into account:

Table I: HU Cut-Off Values for Different Tissue Types

HU values	Tissue Type	
<800	HU	aerated tissue
300 – 0	HU	fat tissue
0 – 100	HU	soft tissue
>300	HU	bone

1.3.2 Lambert–Beer's Law, Linear Attenuation Coefficient and CT-Value

An X-ray beam is composed of photons with higher energy and shorter wavelengths than visible light, properties that allow them to penetrate deeper into materials and thus be used to create diagnostic images. As these photons travel towards the detector on a pathway of

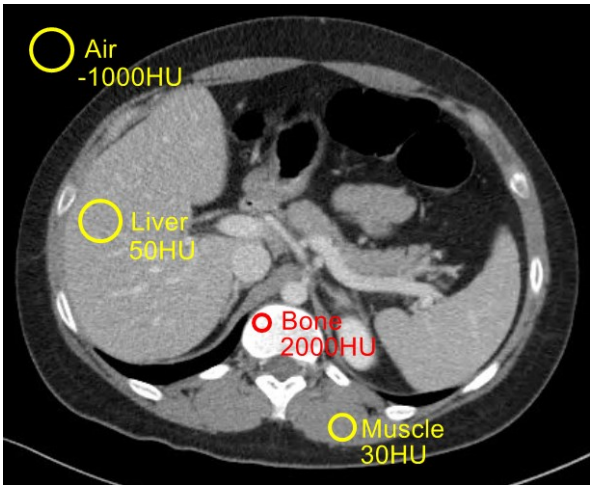

Figure 1.8: Transverse Section of the Abdomen, with HU values. How was your guess?

length η, they interact with the matter so that the initial intensity of the beam, generally referred to as $I_{(0)}$, is attenuated to a final intensity $I_{(\eta)}$. The relationship between η, $I_{(0)}$ and $I_{(\eta)}$ illustrated in 1.9 is described by the Lambert–Beer's law of attenuation [13]:

$$I_{(\eta)} = I_{(0)} \cdot e^{-\mu\eta}$$

where

- $I_{(\eta)}$ is the intensity of the beam when it reaches the detector, usually measured in gray (Gy) or roentgen (R);
- $I_{(0)}$ is the original intensity of the beam when it was generated by the X-ray tube, measured in Gy or R;
- e is the Euler's number (approximately 2.718);
- η is the length of the pathway, usually in cm;
- μ is the linear attenuation coefficient, usually in $cm^{-}1$. It describes the inherent ability of a material to attenuate X-ray radiation.

Note that this formula only applies for monochromatic beams, namely beams that consist of photons with the same energy level, as it is the case for CT scans by virtue of the filter [14].

In this equation, $e^{-\mu\eta}$ represents the exponential function of $-\mu\eta$, and since it is raised to a negative power, it indicates rapid decay. In other words, the initial intensity of the beam

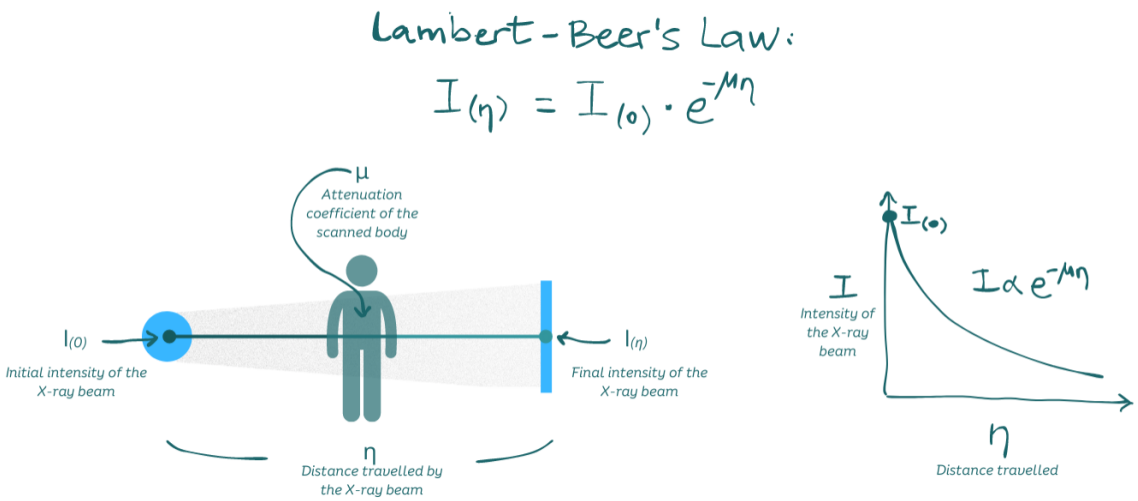

Figure 1.9: Lambert-Beer's Law: this equation describes how the intensity of a beam changes when it travels through a body. In a CT scan, the distance η is always the same (= the diameter of the gantry), furthermore we know the values of $I_{(0)}$ and $I_{(\eta)}$ because we can arbitrarily set the former, and read the latter at the detector. We can therefore calculate the only unknown variable, μ, which describes how much it attenuates the beam, or how "dense" it is.

is exponentially reduced as the distance increases and as the material it travels through is denser.

In a CT scan, μ is easily computed at each acquisition point, since $I_{(0)}$ and η are preset values, and the detector measures the $I_{(\eta)}$. However, even if we know that the higher μ is, the denser the material is [15], it's not particularly convenient to think of density in terms of μ, because of its narrow range of values that reach into many digits after the decimal point, with different tissues having a similar value to each other and to water, as exemplified in Table II.

A much more practical way to grasp the density information carried by μ is to project μ into the Hounsfield scale. On the one hand it creates a visual feedback in form of a grayscale that spans from white (very dense) to black (virtually intangible), as well as an easier-to-navigate metric scale, keeping in mind that the value -1000HU is assigned to air and 0HU to water. The value of μ converted into HU is referred to as the CT attenuation value. It's calculated using the following calibration formula. It centers the value of μ_{water} to 0HU, and sets the value range between μ_{water} and μ_{air} to 1000 [18]:

Table II: Linear Attenuation Coefficients μ of Human Tissues, Air and Water for Photon Energies of 60–150 keV. Adapted from Hamideen et al. [16] and National Institute of Standards and Technology [17].

Energy	60keV	80keV	110keV	150keV
Air	0.028	0.011	0.006	0.001
Adipose tissue	0.173	0.161	0.149	0.137
Soft tissue	0.190	0.174	0.160	0.147
Brain	0.199	0.181	0.166	0.152
Muscle	0.199	0.181	0.166	0.153
Lung	0.201	0.183	0.168	0.154
Blood	0.203	0.185	0.169	0.156
Water	0.206	0.184	0.171	0.151
Skin	0.208	0.191	0.175	0.161
Bone, compact	0.468	0.361	0.303	0.268
Bone, cortical	0.523	0.380	0.307	0.267

$$CT\,Value = \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \cdot 1000$$

where

$CT\,Value$	is the number of HU that the particular site will have;
μ	is the linear attenuation coefficient computed at that particular site;
μ_{water}	is the linear attenuation coefficient of water;
μ_{air}	is the linear attenuation coefficient of air.

Note that the linear attenuation coefficient of air μ_{air} is so small that is negligible and often omitted:

$$CT\,Value = \frac{\mu - \mu_{water}}{\mu_{water}} \cdot 1000$$

In brief, a CT Value is a number that correlates both with the density of a material (the higher the CT Value, the denser the tissue) and the Hounsfield scale (the lighter the hue, the denser the tissue).

It is noteworthy that the choice of the numbers 0 and -1000 in the Hounsfield scale is arbitrary, but carefully chosen. Since practically all organs, with the exception of the bones, have attenuation values that are relatively comparable to those of water, with the difference being of permille, it was sensible to choose 1000 as a scaling factor [7].

1.3.3 Image Noise

The term image noise refers to the impurities of a digital image that are unrelated to the actual image itself, and accounts for the grainy texture. Over the years, several denoising algorithms have been developed. While they surely improve the image visibility, they tend to deliver a patchy appearance, as appreciable in Figure 1.10.

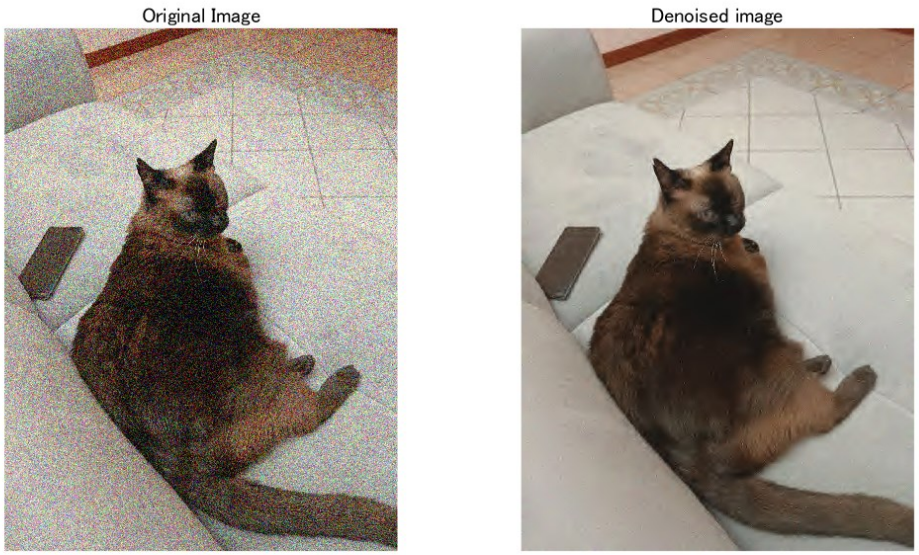

Figure 1.10: Image of a cat before and after denoising. The image was processed with the "wdenoise2" function from MATLAB. While the original image has a grainy appearance, its denoised version is clearer but patchier.

To assess the quality of an image, it is always necessary to consider the extent to which the detected signal can be regarded as true (i.e. the real image) and the degree to which it is attributed to random background events (i.e. the grains) arising from either the detection or transmission process. These random signals fall under the broad category of noise, and no image is exempt [19]. Noise can result from a vast variety of sources, but in CT scans the principal types are [18]:

- Quantum noise: it describs the randomness in the signal derived from the randomness of the number of photons travelling in the beam toward the patient. Since these photons are produced by accelerated electrons that are chaotically hitting a target material, their number inevitably fluctuates. In other words, this noise is due to the fact that the X-ray beam is not perfectly uniform throughout the entire examination;

- Anatomical noise: it describes disturbances of the signal generated from the patients themselves, hampering a clear visualization of the other structures of interest. For example, adipose patients absorb more radiation due to their excess fat tissue, and fewer photons will reach the detector, generating a blurry image under the same conditions. This is why radiation dose is higher in havier patients;

- Electronic noise: it is due to electrons originated from external events. They are a cause of contamination because their electrical flow generates a current that deviates the signal from their predetermined destination, transporting it to another point of the imaging chain. Those streams of deviating electrons can arise from anywhere, but particulary from thermal sources and shot noise, representing a serious challenge in clinical settings, where overheating and loud devices are bread and butter. This type of noise affects both digital and analog devices;

- Structured noise: also known as fixed pattern noise. It is due to the construction pattern of the detector itself: usually detectors can read multiple points, or pixels, at the same time. The information from those arrays is then conveyed through parallel channels, and each of them has its own amplifier circuit. These circuits can't be perfectly matched or tuned with respect to each other, so some detectors might read out slightly differently than others. The noise due to this mismatch has a fixed pattern, so it is consistent across multiple images taken with the same machine. This allows to correct for it by taking two calibration images: a gain image, taken by exposing the detector to radiation in the absence of an object, and an offset image, taken without any radiation hitting the detector. The two images are then used to adjust the readout, removing the structured noise. Because the specific pattern can vary over time, it is important to routinely carry out recalibrations.

Mathematically speaking, the noise in a given region of interest (ROI) is quantified as the standard deviation σ_{ROI} of the CT Values therein. It is a logical correspondence, since standard deviation is a measure of the dispersion of a set of values.

In this thesis, noise was calculated on a ROI selected in the abdominal wall fat (σ_{fat}, or SD_{fat} in Figure ??), following the indications presented by Nam et al. [20].

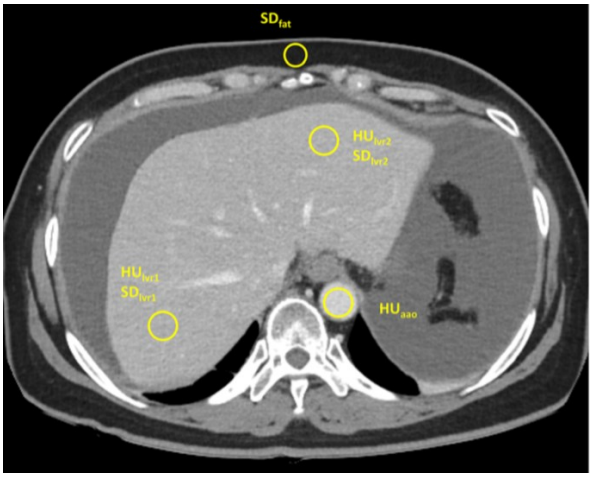

Figure 1.11: Workflow for calculating noise and SNR. 1eV is the energy that the electron has when it reaches the positively charge plate. Image reproduced from Nam et al. [20].

1.3.4 Signal to Noise Ratio (SNR)

The Signal to Noise Ratio (SNR) is a key indicator of system performance. The SNR measures the relationship between the strength of the signal—which represents the detected photons on a particular pixel—and the strength of the noise. [21]. The strength of the signal in a ROI can be mathematically described as the mean CT value therein μ_{ROI}. Consequently, the SNR can be calculated with the following equation [18]:

$$SNR = \frac{\mu_{ROI}}{\sigma_{ROI}}$$

In this study, for the compuation of SNR we followed again the indications by Nam et al. [20], selecting two ROI in the liver, making sure that each comprised an area $\geq$ 1 cm^2, as shown in Figure 1.11. The values were then averaged:

$$SNR = \frac{\mu_{\text{right liver lobe}}}{\sigma_{\text{right liver lobe}}} + \frac{\mu_{\text{left liver lobe}}}{\sigma_{\text{left liver lobe}}}$$

1.4 Dosimetry

Dosimetry is the study of the radiation doses delivered to patients, and the evaluation of the potential risks associated with them, notably of radiation-induced cancers. In CT scans, it is a particularly relevant matter: despite accounting for only 12% of all imaging procedures, they contributed to over 50% of the total radiation dose received in the United States from medical imaging in 2009, and these numbers are alarmingly growing [22]. In this section, we will cover the following basic concepts relevant to this study:

1. X-ray Beam Energy and Electronvolts (eV);
2. Absorbed Dose and Gray (Gy);
3. Computed Tomography Dose Index (CTDI);
4. Dose Length Product (DLP);
5. Effective Dose (E), Equivalent Dose (H_T) and Sieverts (Sv);
6. Effective Dose Conversion Factor (E_{DLP}).

1.4.1 X-ray Beam Energy and Electronvolts (eV)

In the International System of Units (SI), the unit of measure used to quantify energy is the Joule (J) [23]. In the specific case of ionizing radiations, however, it is an established practice to use electronvolts (eV), whereby

$$1\ \text{eV} \approx 1.602 \times 10^{-19}\ \text{J}.$$

One eV represents the kinetic energy that an electron gains when moving across a potential difference of 1 Volt in vacuum, and it can be better visualized with the help of Figure 1.12. eV are more commonly used than J in this setting, because the order of magnitude of the X-ray beam energy needed to generate satisfactory imaging is closer to 10^{-19} J than to 1 J, with the most common values ranging between 20 keV and 140 keV [24].

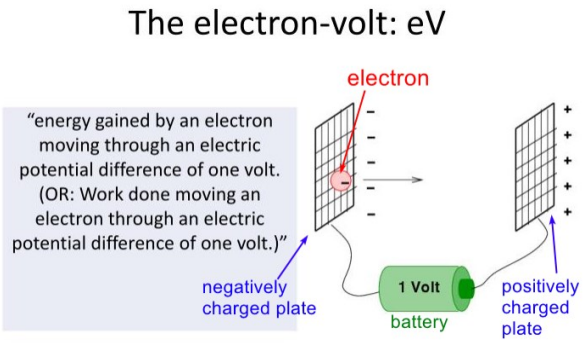

Figure 1.12: Visualisation of an Electronvolt eV. 1eV is the energy that the electron has when it reaches the positively charge plate. Image adapted from slideserve.com [25].

The necessary amount of energy required to produce conclusive medical images is influenced by various factors, including scan parameters such as tube current, tube voltage, and scan length, as well as patient size and imaging protocols [18].

1.4.2 Absorbed Dose and Gray (Gy)

Absorbed dose refers to the amount of X-ray energy that is absorbed by a patient when exposed to radiation, and its SI unit is Gray (Gy) [23]. If a given body absorbed 1 Gy, it would mean that on average, each kg of it acquired 1 J of energy.

$$\text{Gy} = \frac{\text{J}}{\text{kg}}$$

Typically, we find absorbed doses measured in one thousands of Gy, the milligray (mGy).

1.4.3 Computed Tomography Dose Index (CTDI)

To estimate the quantity of absorbed dose, Shope et al. [26] introduced the concept of CT Dose Index (CTDI) in 1981. They proposed to calculate the dose absorbed by a phantom, and use it as an estimate for the dose acquired by the patient. Such phantoms, as illustrated in Figure 1.13, are commonly made of Polymethyl Methacrylate (PMMA), are cylindrical and are found in two dimensions: one measures 32 cm in diameter and is a fairly good surrogate of an adult body, the other measures 16 cm in diameter and is used both in pediatric settings and to mimic the adult head.

A PMMA phantom typically has a central hole and one or multiple peripheral holes. These holes are designed to accommodate either a pencil with an ion chamber that can measure the radiation dose received, or a simple PMMA plug to close the hole. Only one pencil is inserted at a time, and the other holes are plugged to assure homogeneity across the rest of the phantom. The phantom is then layed inside the machine and an acquisition is made, so the pencil registers the dose received on that particular position. This value is then related to the beamwidth to calculate the dose absorbed by the ion chamber, and this dose is sometimes referred to as CTDI_{100}, since the pencil's ion chamber is 100 mm long. This entire process is then repeated moving the pencil to a hole in another position.

The central and peripheral measurements of CTDI_{100} are then combined to estimate the dose absorbed by the entire phantom. There are different methodologies to achieve this goal, but a simple, well-known equation uses the following weighting scheme [18]:

$$\text{CTDI}_{\text{w}} = \frac{1}{3}\,\text{CTDI}_{\text{center}} + \frac{2}{3}\,\text{CTDI}_{\text{periphery}}$$

where

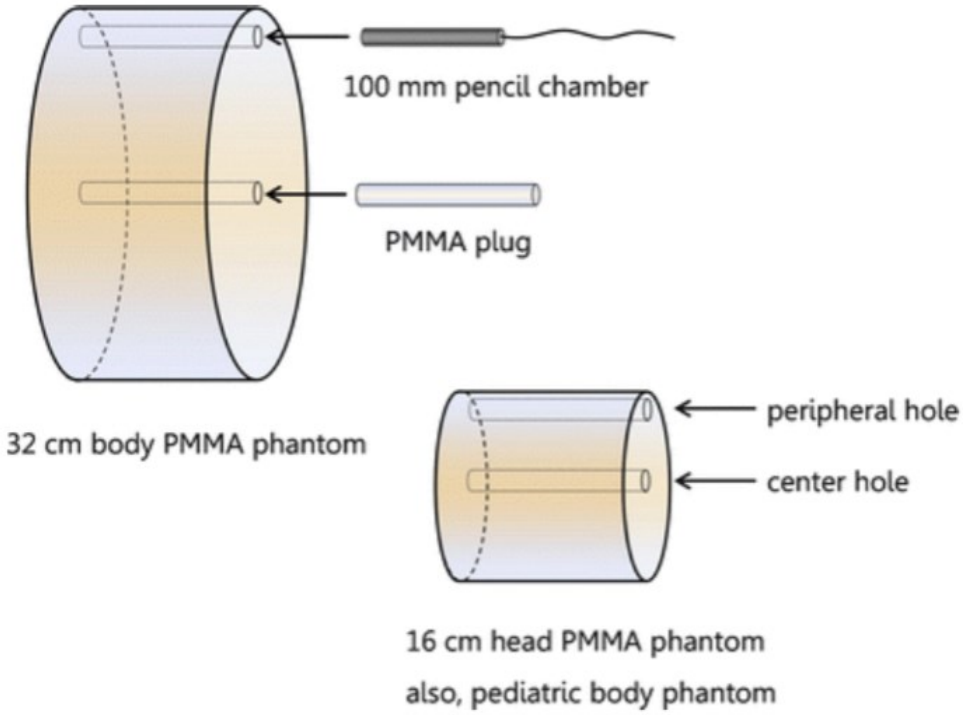

Figure 1.13: Structure of a typical PMMA Phantom. PMMA (Polymethyl Methacrylate) is a transparent material, also known as acrylic or acrylic glass. It is commonly used as a replacement for glass, since it is shatterproof. Other advantages are resistance to UV light and excellent light transmission [27]. The image is reproduced from Bushberg et al. [18].

$CTDI_w$	is the overall CTDI of the phantom, calculated through a weighting system;
$CTDI_{center}$	is the $CTDI_{100}$ determined by placing the pencil in the central hole of the phantom;
$CTDI_{periphery}$	is the $CTDI_{100}$ determined by placing the pencil in a peripheral hole of the phantom.

Note that since the pencil's ion chamber is 100 mm long, the $CTDI_w$ only measures the dose absorbed by a slice of the same thickness.

In helical scans, this $CTDI_w$ value needs to be lastly adjusted with regards to the contraction or dilation in the helical pattern, as shown in Figure 1.14:

$$CTDI_{vol} = \frac{CTDI_{center}}{pitch}$$

where

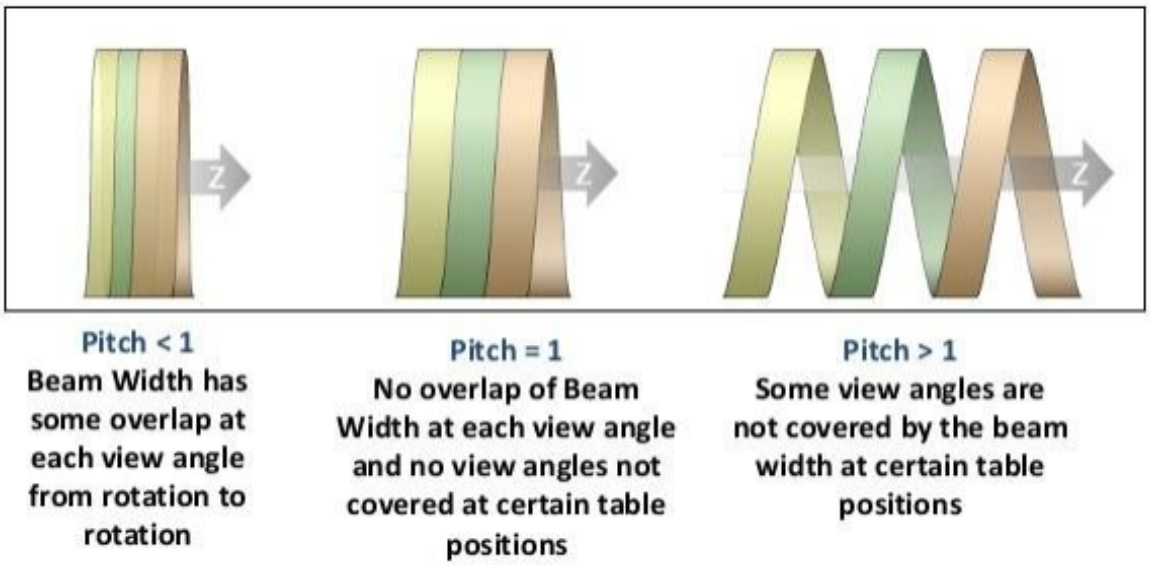

Figure 1.14: Pitch: in a helical scanner, the gantry and the motorized table continuously move with regard to each other, without waiting for a whole rotation of the x-ray beam. This results in a helical pattern, and its contraction or dilation in length is measured by the pitch. Reproduced from Doaa et al. [28].

$CTDI_{vol}$	is the volume CTDI of the phantom. It is a modified $CTDI_w$ that accounts for eventual overlaps or rarifications of the X-ray beam;
$CTDI_w$	is the $CTDI_w$ determined after a complete, single and monoplanar rotation of the X-ray tube;
pitch	a measure of the spread of the helix.

Since the phantoms used to calculate the CTDI are standardized, they also enable a reliable comparison between different scanners and scan protocols. Nonetheless, it is important to remember that CTDI is not equivalent to the actual dose absorbed by a patient, but just an estimation of it. In fact, the true patient dose is influenced by various factors, including the patient's size, body composition and scan region [29].

CTDI is measured in Gy.

1.4.4 Dose Length Product (DLP)

As previously stated, the standard $CTDI_w$ (and hence, the adjusted $CTDI_{vol}$) only measure the radiation dose absorbed by a cylindrical slice that measures 1cm in thickness, and either

32 cm or 16 cm of diameter. To estimate the dose absorbed by a longer body, it is necessary to consider the Dose Length Product (DLP) [18]:

$$DLP = CTDI \times L$$

where

- DLP is an estimate of the total radiation absorbed by a patient during the exam;
- CTDI is either the $CTDI_w$ determined after a complete, single and monoplanar rotation of the X-ray tube, or the $CTDI_{vol}$ for helical scans;
- L is the length of the irradiated body.

DLP correlates better with the patient's radiation risk, since it accounts for the whole irradiated length, and is not standardized to 1 cm.

1.4.5 Effective Dose (E), Equivalent Dose (H_T) and Sieverts (Sv)

When different tissues are exposed to radiation, they are not affected by it to the same extent, so their unique sensitivities must be considered when evaluating the potential risks. To ad-dress this variability, it is helpful to consider the effective dose (E) [30]. It is a quantity that takes into account both the amount of radiation absorbed and the sensitivity for each tissue, then sums those values to assess the effect on the whole body:

$$E = \sum_T w_T \cdot H_T$$

where

E is the effective dose in Sv;

T is the subscript that represents the different tissues in the body;

w_T is the weighting factor for tissue T. Its value depends on the particular tissue, examples are found in Table III;

H_T is the equivalent dose. For radiation composed of photons, such as X-rays, its numerical value is identical to that of the absorbed dose in Gy, with just the measuring unit changed to Sv. Indeed, the equivalent dose was developed to compare different radiation types (such as neutrons, protons, alpha particles) to the dose of photons, so it makes sense that the numerical value remains unchanged.

Since we are still considering a radiation dose, it would be correct to measure the effective dose it in J/kg. However, as there is the fear of mistaking it with the absorbed dose (measured in Gy, also equivalent to J/kg), the SI recommends using the unit Sievert (Sv) instead, to stress that it indicates a biological effect [23]:

$$\mathrm{Sv} = \frac{\mathrm{J}}{\mathrm{kg}}$$

Approximatively, 1 Sv corresponds to a probability augmented of about 5% to develop fatal cancer, when compared to the risk of the general population [31]. It is important to keep in mind that this is a generic estimate of an age- and sex-averaged model of the human body, and it does not represent the risk of any true individual [32].

Table III: Tissue Weighting Factors w_T of different tissues and organs. Adapted from Valentin et al. [31]

w_T	Tissue or Organ
0.01	Bone surface, Brain, Salivary glands, Skin
0.04	Bladder, Oesophagus, Liver, Thyroid
0.08	Gonads
0.12	Bone-marrow (red), Colon, Lung, Stomach, Breast, Adrenals, Extrathoracic region, Gall bladder, Heart, Kidneys, Lymphatic nodes, Muscle, Oral mucosa, Pancreas, Prostate, Small intestine, Spleen, Thymus, Uterus/cervix

1.4.6 Effective Dose Conversion Factor E_{DLP}

The effective dose conversion factor E_{DLP} serves to estimate the effective dose when given the DLP and the anatomical region covered. The effective dose is calculated with the formula [33]:

$$E = E_{DLP} \cdot DLP$$

where

- E is the estimated effective dose, previously explained;
- E_{DLP} is the effective dose conversion factor. It is specific for every anatomical region, values are shown in Table IV;
- DLP is the dose length product, previously explained.

Table IV: Effective Dose Conversion Factors E_{DLP} of different regions, values for adults. Adapted from McCollough et al. [33]

Anatomical Region	E_{DLP}
Head and neck	0.0031
Head	0.0021
Neck	0.0059
Chest	0.014
Abdomen and pelvis	0.015
Trunk	0.015

Figure 1.15 provides an overview of the various dosimetry parameters.

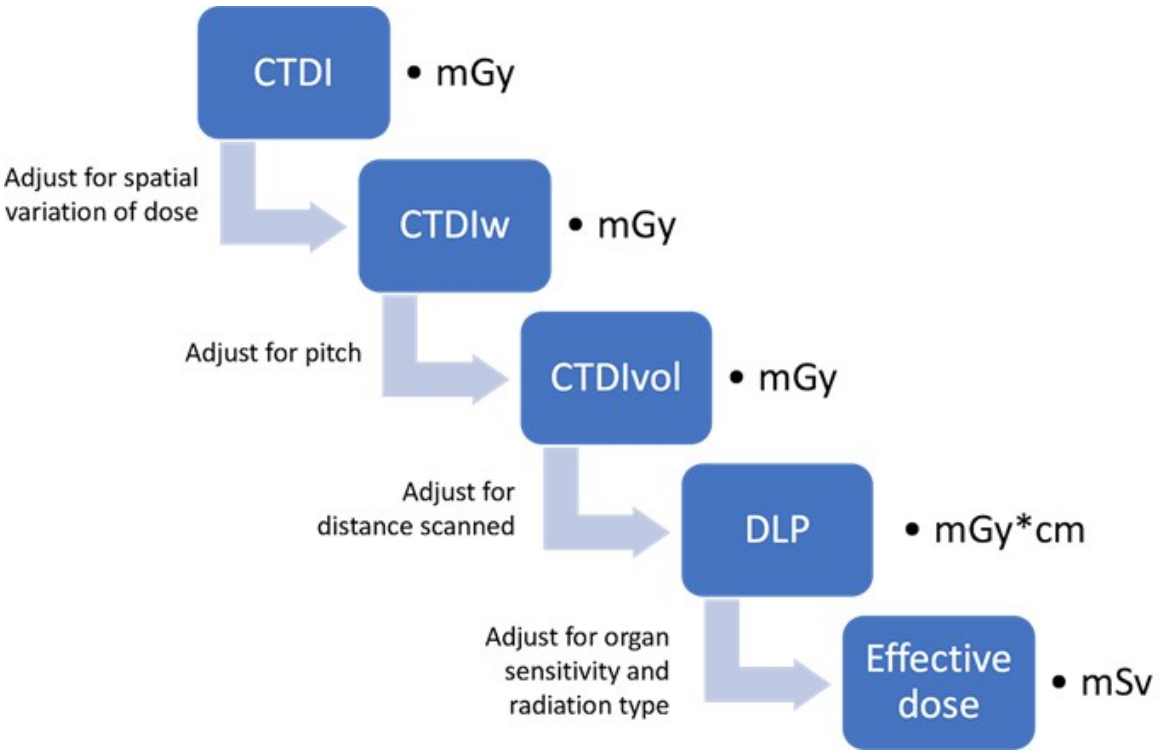

Figure 1.15: Overview of Dosimetry Parameters: Image reproduced from Abdulla et al. [34].

Chapter 2

Fundamentals of Reconstruction

2.1 Main Categories of Algorithms

At the time of writing (2023), three main categories of reconstruction methods have been developed: Fourier-based methods, algebraic and statistical methods, and deep neuronal network-based methods [7]:

- Fourier-based methods owe their name to the Fourier Slice Theorem, which they rely upon. The most notable among them is Filtered Back Projection (FBP), and it has been the method of choice for decades due to its computational efficiency. Fourier-based methods suffer from various limitations such as sensitivity to noise and artefacts, which inevitably translate to the necessity of higher radiation doses [35].

- Algebraic and statistical methods treat the problem as a system of linear equations, where each equation represents a ray from the source to the detector, as represented in figure 2.1. The first method to appear was the Algebraic Reconstruction Technique (ART), and it was the one present in the very first CT scan. However, since it requires a computational effort that is too heavy for the computers that were available at the time, it wasn't suitable for clinical application, so it was replaced by Fourier-based methods. With recent advancements in computational capabilities, algebraic methods have found renewed interest.

  Statistical methods, on the other hand, make use of a statistical viewpoint. The underlying principle is that they attempt to find an image that once projected mathematically

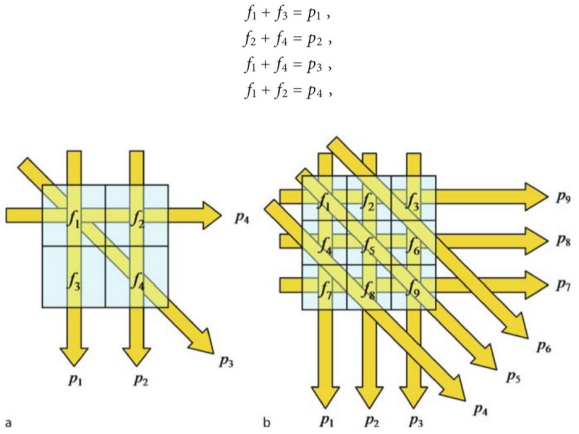

Figure 2.1: Algebraic Reconstruction Technique (ART): The set of projections results in a linear system of equations. f represents the real, unknown attenuation value in the object space, and p represents the projected value measured by the detector. In the image on the left, all four attenuation values can be exactly computed using four projections from three projection angles θ. If the matrix that is to be reconstructed is larger, more projections are required. Image reproduced from [7].

in a simulation of the CT scan, will produce projected values that are as close as possible to the ones that were really measured. One well-known technique among those is the Maximum Likelihood Method, often applied when too few photons reach the detectors and the signal to noise ratio decreases as a result. In these situations, direct approaches such as the Fourier-based or algebraic methods may deliver images that are unacceptably noisy, whereas a statistical method is more likely to succeed. Needless to say, the computational effort is accordingly more intense. In the retrospective study presented in part II of this thesis, two of the CT scans used a statistical approach to reconstruct the images. One method was developed by Toshiba (now Canon Med-ical Systems) and is called Adaptive Iterative Dose Reduction (AIDR3D), whereas the other one was developed by Siemens and is called Advanced Modeled Iterative Reconstruction (ADMIRE). It's important to note that while these methods fall under this category, they also incorporate aspects of model-based and machine learning approaches in their design, making them more advanced and versatile than the first techniques such as ART.

- Deep Learning approaches: in recent years, there has been a growing interest in the application of deep learning, specifically convolutional neural networks (CNNs), for CT image reconstruction. These methods can be trained to learn complex features and

structures in the data, leading to improved image quality, reduced noise and artefacts, and potential reduction in radiation dose. However, the performance of these methods relies heavily on the quality and quantity of the training data.

2.2 Filtered Back Projection (FBP)

To fully understand this method, it is first important to understand what a sinogram is. Its concept is fairly simple and exemplified in Figure 2.2. A CT scan acquires projections from various angles (θ) in a circular motion, and a way to represent those projections is to put them sequentially next to each other. This process produces a new image commonly known as the sinogram. This term comes from the fact that a circular object in the original image casts a sinusoidal curve in this sequential representation. Because the process of producing the sinogram is called Radon transform, sometimes the sinogram is also referred to as the Radon space.

Since a sinogram is acquired through convolution, the simplest idea to reconstruct the original image would be to project it backwards, rotating it in the opposite direction. This method is known as back projection (BP), and a simulation is shown in Figure 2.3.

Although backprojection provides an insight into the true content in the object space, its limits are evident at first glance. The produced image is extremely blurry, with halos occupying areas where nothing should be: after all, even without truly dissecting the body in the object space, we can see with our eyes that nothing is present in the machine on the outside of the body. This recognition leads to the idea that it must be possible to develop some kind of filter that can better predict the true location of the body and its components in the object space.

The first filters that were developed are based on the Fourier transform, a mathematical technique that breaks down complex waveform signals into a set of simple sine and cosine waves of varying frequencies. By applying the Fourier transform to the sinogram, which is also made of waves, we can analyze the meaningful frequencies in the set of data and backproject only those, effectively reconstructing the image. This method is called Filtered Back Projection (FBP), since we apply a filter to the sinogram before backprojecting it. A simulation of FBP is shown in Figure 2.4.

FBP performs visibly better than BP, yet it still has some limits: an attentive eye will

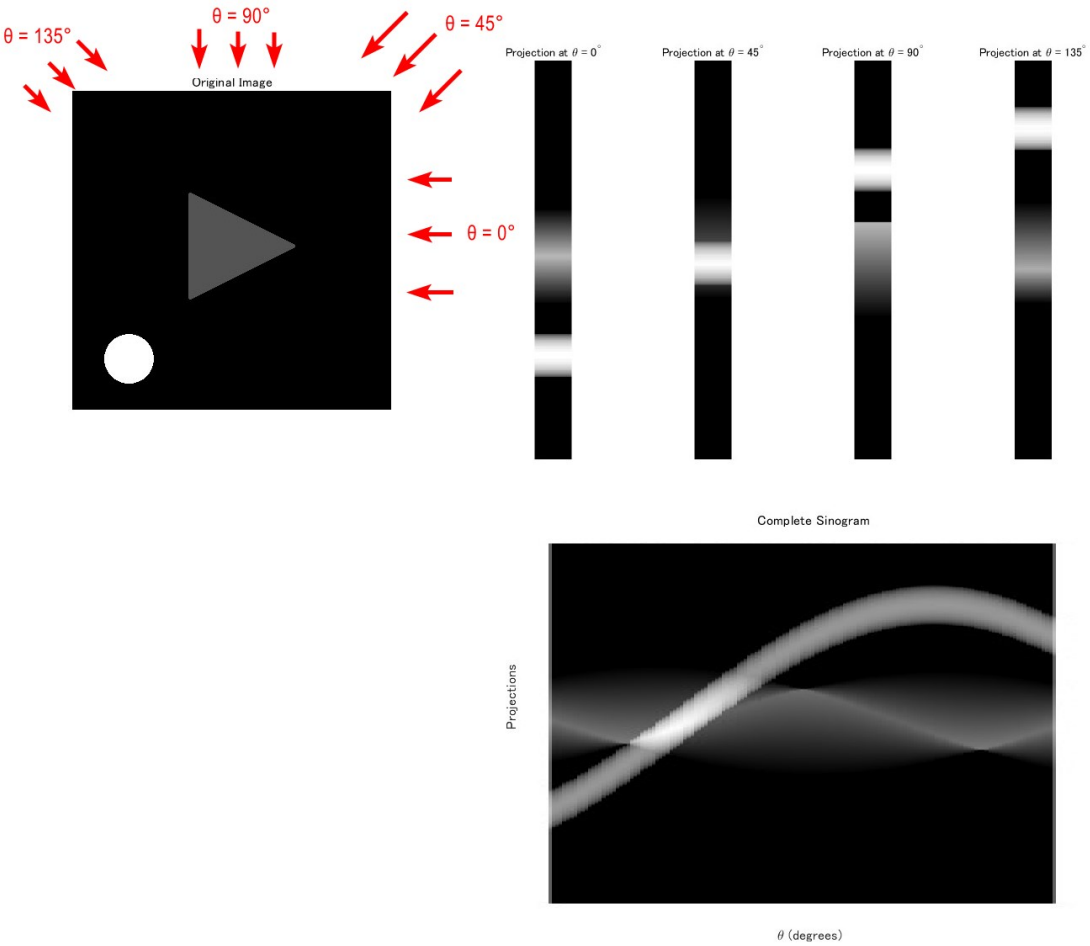

Figure 2.2: Making of a Sinogram: the projections of the white circle in the original image (also called object space) draw a sinusoidal line when placed adjacent to each other, hence the name sinogram to refer to the Radon space. In the sinogram, one can observe three spots where the projection of the triangle (grey area in the middle) draws a sharp, perfectly horizontal line. These correspond to the projections where the angle θ was perfectly perpendicular to one of the sides of the triangle, as appreciable in the singled-out projection θ = 90. The projections of the object and the sinogram were realized with MATLAB.

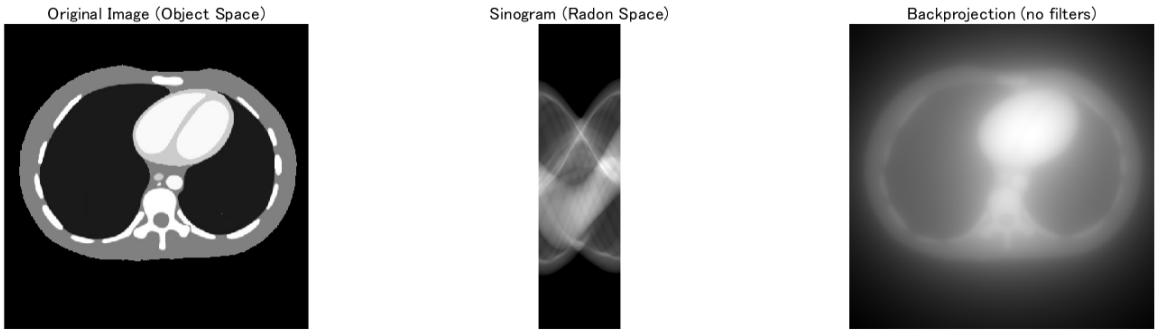

Figure 2.3: BP Simulation. The image in the object space is adapted from AMBOSS [36], the simulation was conducted with MATLAB.

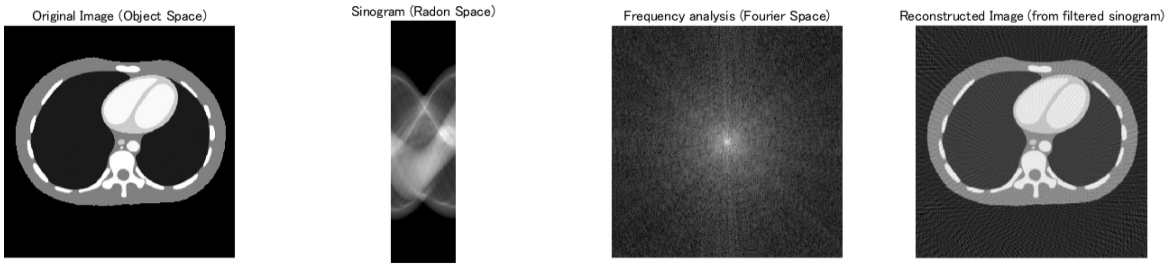

Figure 2.4: FBP Simulation: for the making of this image, 90 projections were simulated around the phantom in the object space. The image in the object space is adapted from AMBOSS [36], the simulation was conducted with MATLAB.

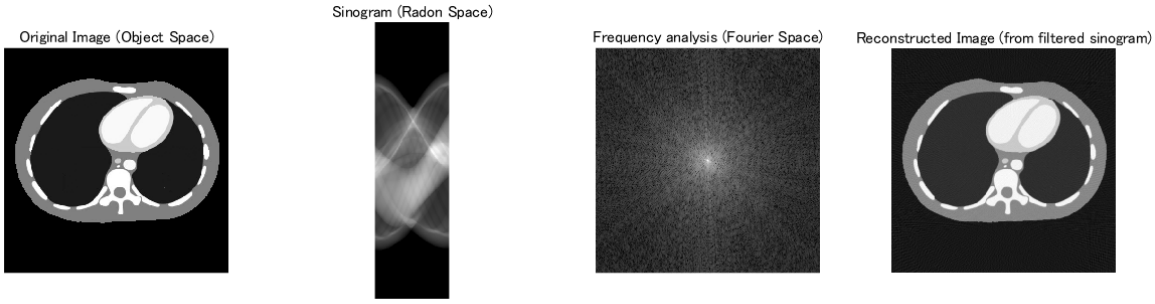

Figure 2.5: FBP Simulation with doubled projections: for the making of this image, 180 projections were simulated around the phantom in the object space. The noise is visibly decreased when compared to the previous reconstruction in Figure 2.4. To achieve the same quality improvement in real life, the total radiation dose would have to be doubled. The image in the object space is adapted from AMBOSS [36], the simulation was conducted with MATLAB.

notice that the hue of black in the lungs of the reconstructed image in Figure 2.4 is slightly different than the one found in the original image, meaning that the density of its tissue was slightly miscalculated. This issue may not seem particularly relevant in everyday life, but in clinical settings can lead to misdiagnosis. The reconstructed image is also remarkably noisier, having a grainy texture that might impede the visualization of smaller structures.

Image noise is a major problem with FBP, and it can be significantly reduced by increasing the number of the acquired projections, as illustrated in Figure 2.5. On the other hand, this inevitably translates into an increased patient dose.

2.3 Iterative Algorithms

Whereas BP and FBP take the acquired projection data and project it back into an image in one step, iterative algorithms approach the problem in a fundamentally different way:

they involve an estimation process that seeks to find an image that, when projected, closely matches the measured projection data. In other words, once they reconstruct the first image, they execute a simulation to see how the reconstructed image would be projected in a sinogram, if it was in the object space. If the simulated projection and the real sinogram differ, they adjust the reconstructed image, and simulate another projection, until the comparison between the two sinograms is satisfactory.

Since the process of reconstructing, projecting and comparing is iteratively repeated, they are named "iterative algorithms".

One of the earliest and simplest IR methods employed was the Algebraic Reconstruction Technique (ART), and it was actually the method present in the very first CT scan. We can roughly describe its working in the following steps, also depicted in Figure 2.6.:

1. Initialization: the start point of the algorithm is an initial guess for image A. In the first machines, it was a blank image, but in more sophisticated approaches, the initial image might incorporate prior information based on patient-specific data, such as a first reconstruction with FBP;

2. Forward Projection: project the current estimate A to create a simulation of how the sinogram would look like, if A was the real image in the object space;

3. Comparison: compare the simulated sinogram of A to the actual sinogram B that was measured on the patient by the CT scan. The difference between the sinogram A and B is the error, or residual;

4. Backprojection and Update: backproject the error into the image A, to update image A. At this stage, a noise model can be incorporated to weigh the residuals based on the expected noise level in the measurements.

   As a result of this step, image A is somewhat closer to how the object in the CT scan looked;

5. Iteration: the processes of forward projection, comparison, backprojection and update are repeated until image A produces a sinogram A that is indistinguishable from the real sinogram B;

6. Production of final image: when the quality of the image is satisfactory. After a certain number of iteration that is variable from specific algorithm to algorithm, the reconstructed image doesn't significantly improve anymore from iteration to iteration.

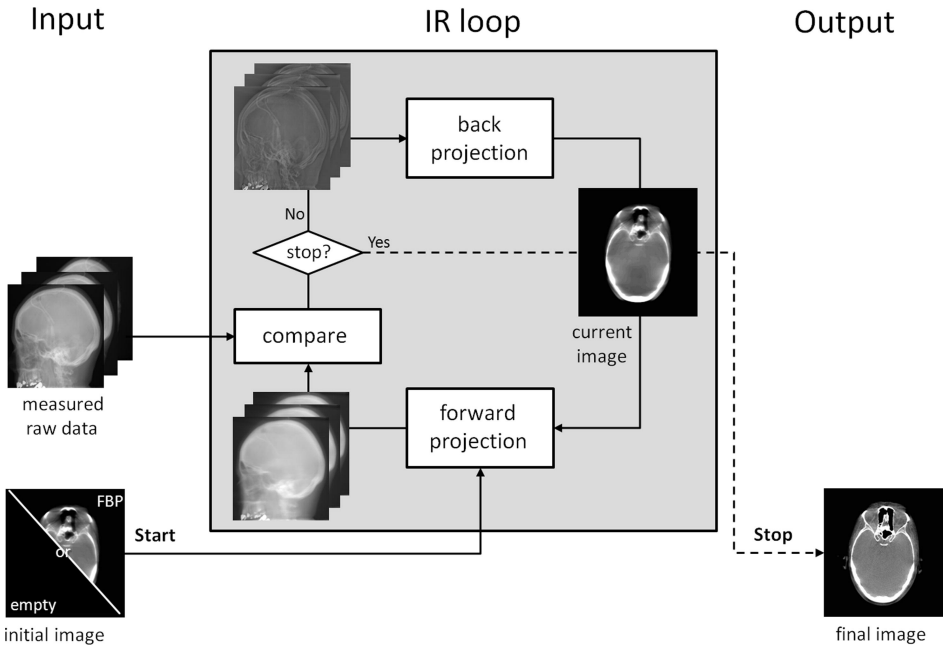

Figure 2.6: Iterative Reconstruction simulation. Reproduced from Beister et al. [37].

The power of iterative methods lies in their flexibility and robustness. They can incorporate sophisticated models at various steps of the imaging process, including considerations about noise characteristics, scanner geometries, and even patient characteristics, leading to more accurate and potentially lower-dose images.

On the other hand, their main disadvantage is the computational complexity. Each iteration requires multiple projections and backprojections of the entire image, and this process has to be repeated for every slice that makes up the CT scan. It is intrinsically bound to be computationally intensive. Fortunately, advances in computing power, including the use of graphics processing units (GPUs) and parallel processing techniques, have made iterative methods increasingly accessible for clinical use.

In this study, two specific iterative algorithms were employed since present at the university hospital of Padua: AIDR3D and ADMIRE.

2.3.1 AIDR3D

Adaptive Iterative Dose Reduction 3D (AIDR3D) is an iterative reconstruction method developed by Canon Medical Systems, previously Toshiba Medical Systems. What distinguishes it from other iterative reconstruction methods is the use of the "3D" or tridimensional approach, by considering information from adjacent slices when estimating each image slice. This can lead to more accurate image reconstruction and better noise suppression [38].

2.3.2 ADMIRE

Analogally, Advanced Modeled Iterative Reconstruction (ADMIRE) is an iterative reconstruction method developed by Siemens Healthineers. Its distinguishing feature is, among others, that its mathematical models allow for reducing the number of iterations, resulting in a substantially higher acquisition speed [39].

2.4 DLIR

DLIR is the newest category of reconstruction algorithms, and as we will see in the next chapter, they are based on artificial intelligence and machine learning methods.

Chapter 3

DLIR

3.1 Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL)

We live in an era where likely everyone has, at least once, come across terms such as Artificial Intelligence or Machine Learning. But what are they exactly, and what is Deep Learning? To answer those questions, it is helpful to step back in history.

Humans have been creating machines to eliminate physical labour since the dawn of civilisation, and today the world is filled with machines that tend to our needs: some cultivate our crops, others build our homes, prepare and package our food, aid in the medical field, or even amuse us. The bulk of these devices operates by carrying out a series of predetermined steps that can also be referred to as algorithms.

In recent years, taking advantage of the increasing computing power, people have been trying to develop more complex steps, or algorithms, that are capable of mimicking the human intellect in their problem-solving skills, language comprehension or creativity. This field that aims to bestow human intelligence to machines is known as Artificial Intelligence (AI) [40].

Since AI aims to reproduce the capabilities of the human mind, it has different branches that reflect our diversity. The sector that tries to replicate the ability to learn new things when presented with data is called Machine Learning (ML) [40].

To understand this concept, let us consider a case where we have a stack of pictures, some of them representing cats and some of them representing dogs. In classical programming,

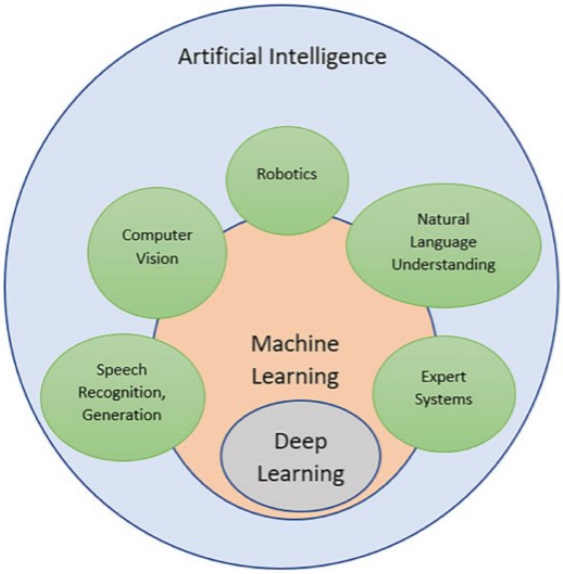

Figure 3.1: Subcomponents of Artificial Intelligence. Reproduced from Tamoghna et al. [40].

if we wanted the computer to automatically parse those images and assign them the right label ("cat" or "dog") as an output, we would have to input two things: (i) the images to be evaluated 0, i.e. the data, and (ii) a set of hard-coded, immutable rules to make the decision. In our example, those rules could take into account the differences in size (dogs tend to be bigger), ear shape (cats often have pointy ears), snout size (usually bigger in dogs), other objects in the picture (if the animal is sitting on a tree, probably it's a cat, or if there is a bone-shaped toy, probably it's a dog) and so on. Then the program would be executed, and the accuracy of its output (assigning the label "cat" to a cat image, and the label "dog" to a dog image) would entirely depend on the slickness of the human that designed the set of rules. For some classes of problems, the development of such rules might be simple, but when there are a lot of potential parameters, it becomes nearly impossible for the human mind to identify them all in order to develop a reliable algorithm.

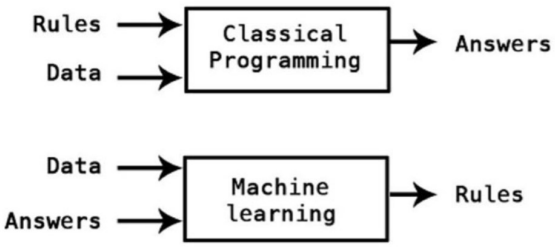

Figure 3.2: Classical Programming vs Machine Learning. Reproduced from Tamoghna et al. [40].

In the case of machine learning, we still have two inputs and one output, but their nature is different. As an input, we still need (i) the images to be evaluated, i.e. the data, but as a second input (ii) we feed the answers themselves, so we already tell the algorithm that image 1 features a cat, image 2 a dog, image 3 a cat again, and so forth. The output that we expect is an efficient set of rules to differentiate between the two animals, as shown in Figure 3.2. Once this set of rules is produced, we can utilize them to differentiate future images and label them correctly without human intervention.

In the process of finding an effective set of rules, two steps are particularly relevant. One is feature extraction, and the other one is model evaluation. Feature extraction refers to the process of identifying features that are highly intercorrelated (like pointy ears for cats, or bone-shaped toys for dogs), and discarding features that are not relevant (the presence of fur might for example be irrelevant, since both animals display it). Then the machine generates several models that take into account all features, weighing them to a different extent: for example, in the concomitant presence of pointy ears and a bone-shaped toy, the pointy ears could be more relevant. After a bunch of hypothetical models are generated, a model evaluation is performed using metrics such as true positive rate and false positive rate to assess which model best assigns the right label to the right image, and is consequently the best classifier.

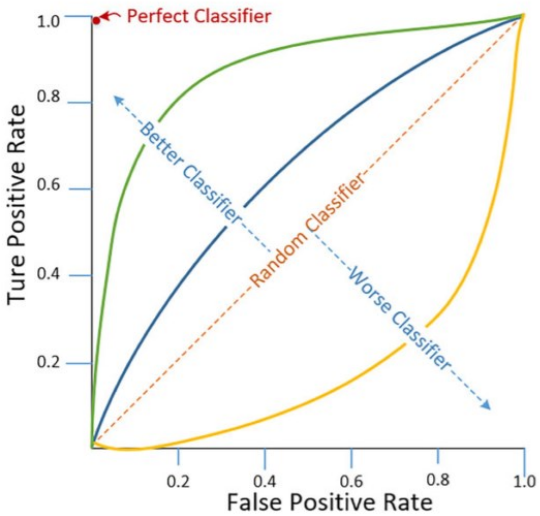

Figure 3.3: Evaluation of classifiers: to compare the efficacy of different classifiers, the true positive and false positive rates are considered. A perfect classifier would display a 100% true positive rate, and a 0% false positive rate. Reproduced from Tamoghna et al. [40].

It is important to note that during the training process, when the different features are

identified and weighted differently, a machine generates all kinds of models, so some of the classifiers will mainly or consistently predict the opposite of the desired result (i.e. predicting mainly cats when dogs are in the image). In this case, they are referred to as bad classifiers. There may be also lots of random classifier suggestions, where the predicted just randomly assigns labels, because the features it identified are not relevant at all. It is only through model evaluation that good classifiers are identified.

During the training period, a machine often divides the input data into two random groups. One group is used for the training, where different models are generated, and the second group is used for model evaluation, where all models are first applied to the images, and then their result is compared to the true labels, to see which model correctly assigns the most labels. In other words, first the machine tries the different methods on the second group, and then "peeks" the answer that we initially provided to see if the model resulted in the true answer. This process in two steps characterises classic machine learning.

A major difficulty faced by classic machine learning in many real-world situations is the complexity of every single piece of data. Depending on the angle of view, the ears of the cat might not seem so pointy, and the shape of the tail might be difficult to be identified when partially tucked under the body. In a picture taken at night, it might be even challenging to identify the very silhouette of the animal against the background. Many of these sources of variation are extremely hard to be picked upon at such a high level from the raw, input data: it would require nearly human-level understanding. Starting from this consideration, it was thought to construct a system inspired by the structure of the neurons in the human brain, and the concept of Deep Learning (DL) was born [41]. In deep learning, the problem of finding a rule is tackled by an artificial neural network, as depicted in Figure 3.4. To mimic our interneurons, these artificial networks have multiple interconnected layers, hence the term "deep". Oftentimes they are referred to as Deep Neuronal Networks (DNN).

To obtain a general grasp of how deep learning works, it is useful to understand its basic physical structure. Each circle in Figure 3.4 represents an artificial neuron, generally simply called "neuron", and it consists of a small computation unit. The first set of neurons is charged with analysing the input data, and is thus referred to as the visible layer, because we can physically see the input data. In this first layer, each neuron analyses a different region of the image, making a computation, and as a result it outputs a single real number. This number is then sent to each neuron in the following layer, which is the first of the hidden

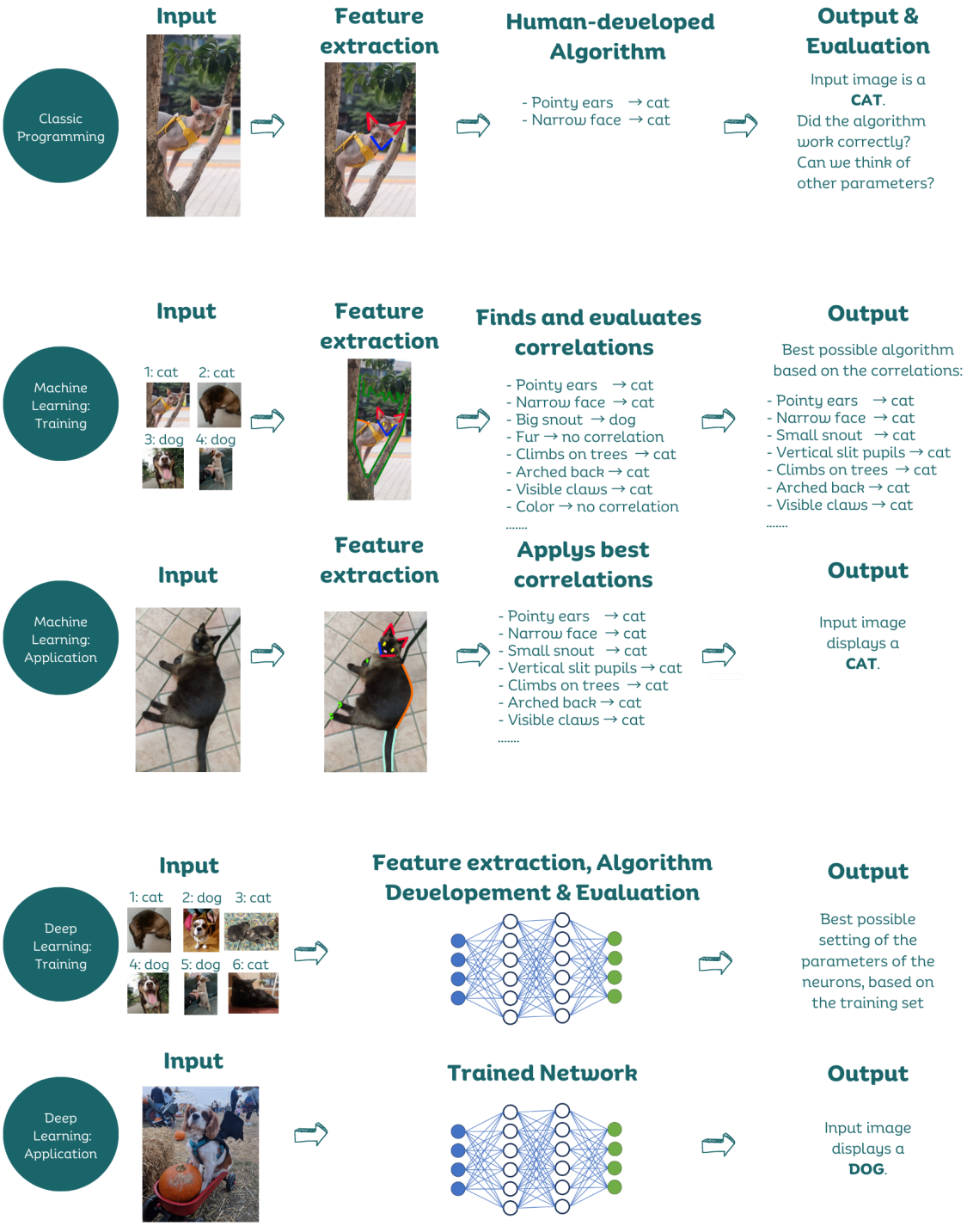

Figure 3.4: Classic Machine Learning vs Deep Learning: In classic machine learning, model evaluation follows feature extraction. In Deep Learning, the two steps are carried out simultaneously.

layers. The term “hidden” refers to the fact that their values are not present in the original data itself. If we concentrate on a single neuron in this hidden layer, we see that it receives an input, i.e. a number, from each of the neurons in the visible layer. Based on those inputs, it also makes a random calculation, obtaining a single real number, and then sends it to every single neuron in the next hidden layer. This process of receiving, computing and sending continues through the available layer: the more layers available, the “deeper” the neuronal network is. Finally, the last layer produces an object identification, such as the labels “cat” or “dog”. A point to notice is that this system enables a hierarchical understanding of the data: higher levels (deeper layers) are the result of the composition of lower levels, and most importantly, this hierarchy is generated completely automatically.

One may ask, what is the point of computing so many random calculations? How is the right sequence of calculations found? The answer lies in the sheer power of their numerousness. Just in our diagram, we have $4 \times 6 \times 6 \times 2 = 288$ unique pathways that start from the input layer to arrive at the output. Among those many pathways, some of them will consistently reproduce a result that is relatively close to the truth. Those pathways will be majorly taken into account, i.e. they will weigh more, and the other pathways that produce random or opposite results will weigh less and thus be ignored and subsequently replaced during the process of learning. The computations carried out by each neuron are namely the result of adaptative, parametrized, non-linear basis functions, and their parameters are the ones that are modified during training of the neural network, until the best fitting parameters are found. Since it is a trial-and-error approach, feature extraction and model evaluation occur at the same time.

As we can see in Figure 3.5, deep learning allows us to take into account many more parameters than classical programming can.

Currently, we dispose of different arrangements of the layers in the neuron, and depending on their architecture, we distinguish different topologies. Depending on the real-world problem to solve, it might be in fact sensible to choose a different number of layers, or neurons per layer, or the basis functions carried out in these layers, not to mention the parameters in it. The main types of neuronal networks are illustrated in Figure 3.6 and are [40]:

1. Artificial Neural Networks (ANN): the name of this category might be misleading, since many people interpret the term ANN as a synonym for neural networks that are

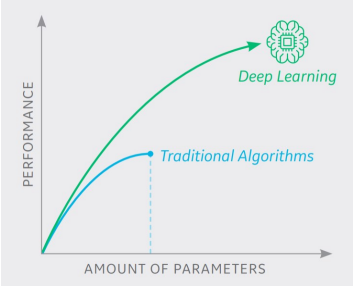

Figure 3.5: Number of parameters in ML and DL: In Classic Programming, the number of parameters is limited by the human capability of finding efficient rules and coding them. In Deep Learning, this task is automatised and much more efficient. Reproduced from Hsieh et al. [42].

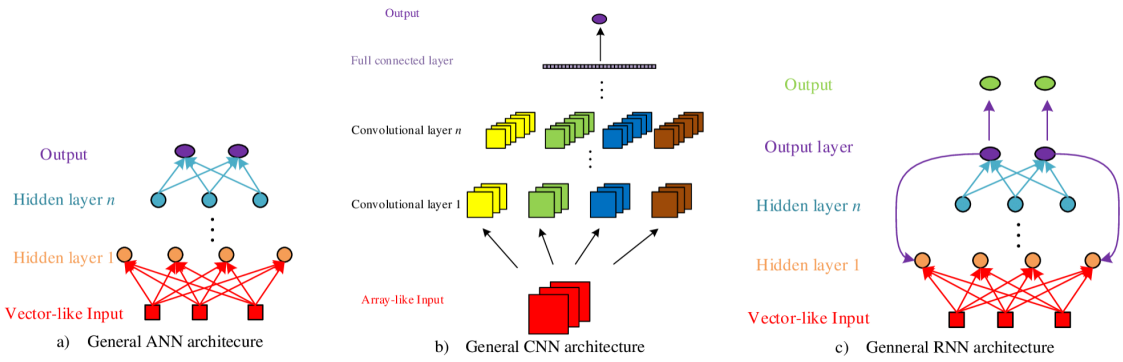

Figure 3.6: ANN, CNN and RNN: In ANN, the information travels forward from the input to the output. In CNN, each neuron analyzes a small neighbourhood of data, called kernel, and slides its window until the whole data is covered. The information stays unidirectional. In RNN, neurons also backproject information creating a looping mechanism. Reproduced from Xiang et al. [43].

artificial. But ANNs are in fact a specific type of those, characterised by a feed-forward system: every layer receives an input, processes it, and outputs it to the next layer, with the information proceeding from one layer to the other without passing twice through the same node. Previously, in our diagram we considered an ANN.

Due to this tiered layering, ANNs excel in fields where the main focus is pattern recognition, such as the interpreting of handwriting in pictures to automatically convert it into machine-written text, or weather prediction. A disadvantage is that it typically requires an enormous amount of data to be effectively trained;

2. Convolutional Neural Network (CNN): their key feature is the presence of convolutional layers that performs convolutional operations. In this special type of operation, the neuron works by applying a filter, also known as a kernel, to an input matrix such

as an image. This filter consists of a small transformation matrix, made up of weights, and when it is applied to a single neighbourhood of the input matrix, it produces a number that is momentaneously stored by the neuron. The filter then slides across the entire input matrix, covering its width and height, computing the result at each position and saving it as well, before moving to the next point. Once the whole input matrix is covered, all these momentaneously saved products are summed up to give a single output value, which is the final rational number produced by that single neuron in the convolutional layer, and that is to be sent to the neurons in the next layer.

But how are the weights in the kernel chosen? Those are also learnable parameters. During the training process, the network learns by trial and error the most useful filters for the task at hand.

CNNs are excellent for tasks like image and video recognition, segmentation and generation. It can be for example used to identify suspicious items like tumours in medical imaging. The downside of CNN is mainly that it is computationally very demanding, and that it has difficulties processing an object's orientation in space;

3. Recurrent Neural Nets (RNN): they are characterized by a looping mechanism that retains and feeds information to the next analysis. This gives RNNs a kind of memory that allows them to be effective in tasks like language modelling, speech recognition and translation.

3.2 Deep Learning Image Reconstruction (DLIR)

Deep Learning Image Reconstruction (DLIR) is a type of reconstruction algorithm that uses deep learning. At the time of writing, three popular DLIR methods have been developed by leading manufacturers [44]:

- TrueFidelity, developed by GE Healthcare [45];
- AiCE (Advanced intelligent Clear-IQ Engine), developed by Canon Medical Systems [46];
- Pixelshine, developed by AlgoMedica [47].

3.2.1 TrueFidelity

In this thesis, the specific DLIR algorithm in analysis is TrueFidelity, since it was the one installed in the CT scan at the university hospital of Padua.

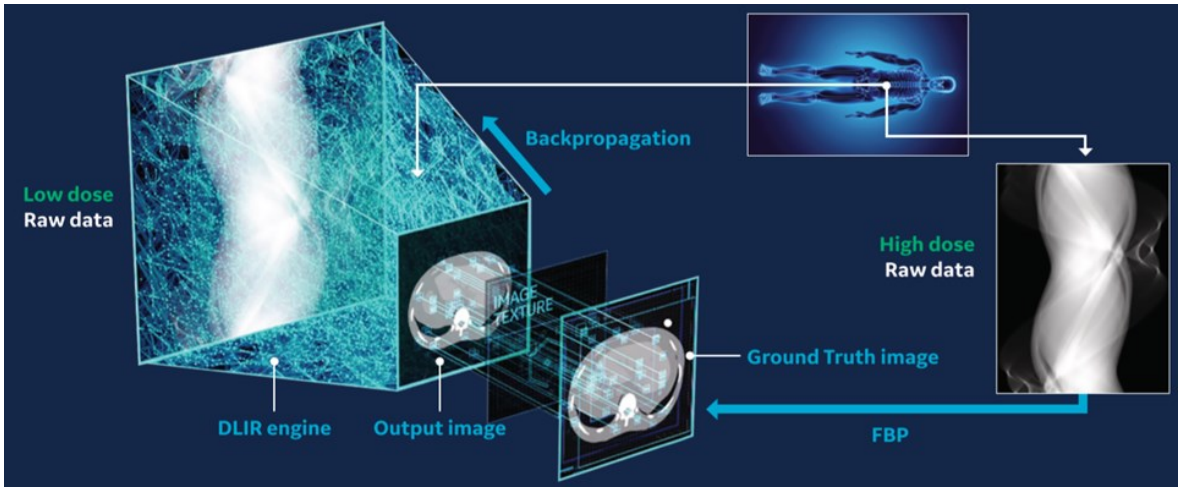

Figure 3.7: TrueFidelity Algorithm by GE Healthcare. Reproduced from Hsieh et al. [42].

Its workflow is depicted in Figure 3.7. The Ground Truth images are the training data sets, and were selected from high-dose, low-noise FBP images for supervised training. They come together with an input, low-dose sinogram, labelled in the Figure as raw data.

First, the DLIR engine, which is made of a CNN, generates a CT image from a low-dose, raw-data sinogram. The generated image is indicated as output CT image. Then, this output image is compared mathematically to the ground truth image to find any differences in terms of image noise, texture, spatial resolution and other metrics. These differences are used to fine-tune the CNN in a process known as backpropagation, with the goal to reduce the difference between the DLIR output and the ground truth, until the two virtually match. During the learning period, backpropagation commonly uses the gradient descent optimization algorithm to adjust the weight of neurons.

This training process on a single image is then repeated on millions of ground truth training data sets, all displaying FBP images that faithfully represent the scanned object.

After the training process, the DLIR engine undergoes an extensive testing phase of inferencing, where the trained neuronal network is used to check its performance on a large validation dataset that was not previously used. During this process, real-world conditions are tested, so the step of backpropagation is not included [42].

The difference in image quality that results from reconstructing the same sinogram with FBP, an iterative algorithm and DLIR is shown in Figure 3.8.

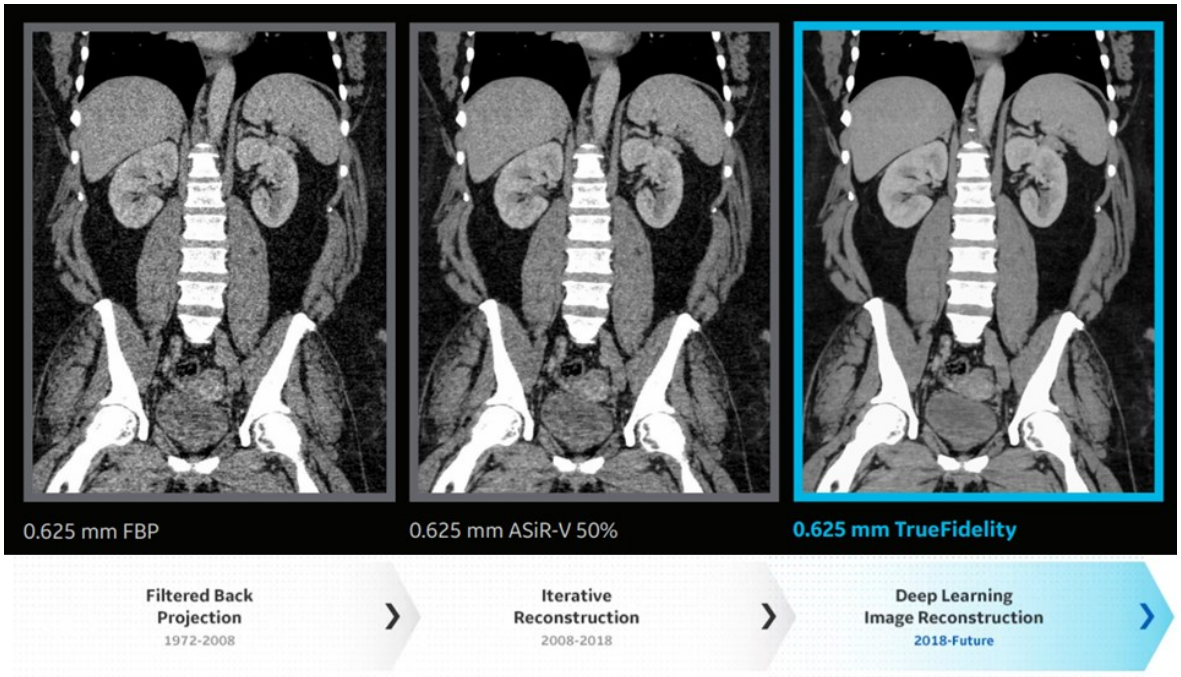

Figure 3.8: Image reconstruction with FBP vs IR vs DLIR. The grainy appearance, or image noise, is visibly reduced when switching from one algorithm to the next. Reproduced from GE Healthcare [45].

Part III

Main Body of the Study

Chapter 4

Ojectives

As illustrated in the introduction, from a theoretical point of view, DLIR should perform better than both FBP and Iterative Algorithms in reconstructing an image when given the same attenuation values. We therefore expect two main consequences:

- Image quality should be improved when dose exposition is the same. Evidence supporting this notion was found both in a phantom study by Zhong et al. [48], and in a retrospective study by Alagic et al. [49];

- Radiation dose should automatically be reduced when using DLIR. This reduction is carried out by automated exposure control systems. They are meant to automati-cally modulate the tube current to ensure constant diagnostic image quality across all anatomical regions of patients of different sizes, all the while guaranteeing a minimization of the radiation dose. Their calibration is usually carried out by obtaining a traditional 2-dimensional X-ray image [50]. Evidence supporting this dose reduction was found for example in a phantom study by Greffier et al. [51] and in a prospective study by Benz et al. [52].

In the abovementioned studies, the reconstructions were either carried out on phantoms, or reconstructed applying different algorithms but on the very same data, or the CT scans were purposely taken a second time immediately after, but using a lower dose. In all those cases, a real-world situation is not necessarily recreated. In this study, we aim to assess the impact of DLIR on daily clinical practice.

This impact will be evaluated in terms of:

- radiation dose reduction, quantified as CTDI, DLP and Effective dose;
- image noise reduction, quantified as Image noise and SNR.

Chapter 5

Materials and Methods

This retrospective study has been approved by the regional ethics committee.

5.1 Sudy Design

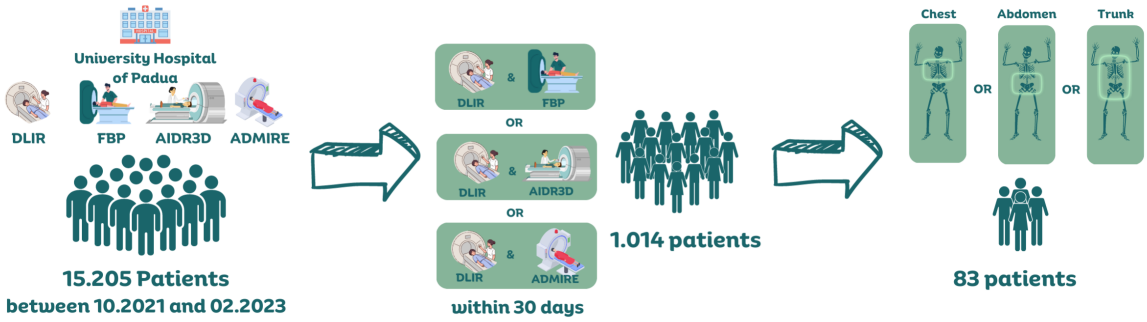

Figure 5.1: Study Design. Probably due to preferential patterns of mobilisation of patients within the hospital (for example, patients in the ICU tend to be scanned with the same machine), the number of patients included in the study is limited when compared to the initial pool.

As illustrated in Figure 5.1, we retrospectively pooled 15.205 unique patients who collectively underwent a total of 35.497 CT examinations at the University Hospital of Padua in the timeframe between 16 October 2021 and 28 February 2023, covering an observational period of 15 months, or 501 days. This initial pool of patients was selected because it was known that the imaging was performed by one of the 4 scanners of interest, 1 of them with DLIR, 1 with FBP, and 2 with an iterative algorithm (AIDR3D or ADMIRE), as illustrated in Table V.

Table V: CT scanners at the University Hospital of Padua

Vendor	Scanner Name	Algorithm Name	Algorithm Type	Total Pat.	Also DLIR within 30 Days
GE Healthcare	Revolution EVO	TrueFidelity	DLIR	3.666	—
Siemens Healthineers	Somatom Sensation 64	FBP	Fourier Based Method	1.904	142
Canon Medical Systems (Previously Toshiba)	Aquillion ONE	AIDR 3D	Iterative Reconstruction (Hybrid)	5.596	239
Siemens Healthineers	Somatom Definition Edge	ADMIRE (Advanced Modeled Iterative REconstruction)	Iterative Reconstruction (Hybrid)	7.741	633

Using Python 3 [53] running on a Jupyter Notebook [54], it was determined which patients had been examined with both DLIR and one of the other algorithms within a timeframe of 30 days, to ensure that they did not significantly alter their physical characteristics such as weight in this short period.

Data regarding all those 1.014 patients were manually retrieved on the hospital computers to ensure that they covered the same anatomical region in both examinations, and if it was the case, the patient was included in the study. For each included patient, the following information was registered in an Excel sheet [55] using the following nomenclature:

- Patient ID, abbreviated as ID;
- Date of birth, abbreviated as DoB;
- Gender, abbreviated as sex;
- Body region scanned, abbreviated as reg;
- Use of contrast medium, abbreviated as wcm;

- Date of the scan specific to the algorithm, abbreviated as FBP_date, Aq_date, DE_date, DLIR_date;

- Number of days between the two dates, abbreviated as diff;

- CTDI and DLP relative to each acquisition phase for each algorithm, abbreviated as dir_CTDI_FBP, art_CTDI_FBP, ven_CTDI_FBP, art_CTDI_FBP; dir_DLP_FBP, art_DLP_FBP, ven_DLP_FBP, art_DLP_FBP; dir_CTDI_Aq, art_CTDI_Aq, ven_CTDI_Aq, art_CTDI_Aq; dir_DLP_Aq, art_DLP_Aq, ven_DLP_Aq, art_DLP_Aq; dir_CTDI_DLIR, art_CTDI_DLIR, ven_CTDI_DLIR, art_CTDI_DLIR; dir_DLP_DLIR, art_DLP_DLIR, ven_DLP_DLIR, art_DLP_DLIR;

- Total dose resulted from the CT examination, abbreviated as tot_FBP, tot_Aq, tot_DE, tot_DLIR;

- μ (mean in HU) and σ (Standard Deviation) in abdominal wall fat, right liver lobe (1) and left liver lobe (2), for each acquisition phase of each algorithm, abbreviated as dir_SD_fat_FBP, dir_HU_lvr1_FBP, dir_SD_lvr1_FBP, dir_HU_lvr2_FBP, dir_SD_lvr2_FBP, art_SD_fat_FBP, art_HU_lvr1_FBP, art_SD_lvr1_FBP, art_HU_lvr2_FBP, art_SD_lvr2_FBP, ven_SD_fat_FBP, ven_HU_lvr1_FBP, ven_SD_lvr1_FBP, ven_HU_lvr2_FBP, ven_SD_lvr2_FBP, tard_SD_fat_FBP, tard_HU_lvr1_FBP, tard_SD_lvr1_FBP, tard_HU_lvr2_FBP, tard_SD_lvr2_FBP; dir_SD_fat_Aq, dir_HU_lvr1_Aq, dir_SD_lvr1_Aq, dir_HU_lvr2_Aq, dir_SD_lvr2_Aq, art_SD_fat_Aq, art_HU_lvr1_Aq, art_SD_lvr1_Aq, art_HU_lvr2_Aq, art_SD_lvr2_Aq, ven_SD_fat_Aq, ven_HU_lvr1_Aq, ven_SD_lvr1_Aq, ven_HU_lvr2_Aq, ven_SD_lvr2_Aq, tard_SD_fat_Aq, tard_HU_lvr1_Aq, tard_SD_lvr1_Aq, tard_HU_lvr2_Aq, tard_SD_lvr2_Aq; dir_SD_fat_DE, dir_HU_lvr1_DE, dir_SD_lvr1_DE, dir_HU_lvr2_DE, dir_SD_lvr2_DE, art_SD_fat_DE, art_HU_lvr1_DE, art_SD_lvr1_DE, art_HU_lvr2_DE, art_SD_lvr2_DE, ven_SD_fat_DE, ven_HU_lvr1_DE, ven_SD_lvr1_DE, ven_HU_lvr2_DE, ven_SD_lvr2_DE, tard_SD_fat_DE, tard_HU_lvr1_DE, tard_SD_lvr1_DE, tard_HU_lvr2_DE, tard_SD_lvr2_DE; dir_SD_fat_DLIR, dir_HU_lvr1_DLIR, dir_SD_lvr1_DLIR, dir_HU_lvr2_DLIR, dir_SD_lvr2_DLIR art_SD_fat_DLIR, art_HU_lvr1_DLIR, art_SD_lvr1_DLIR, art_HU_lvr2_DLIR, art_SD_lvr2_DLIR ven_SD_fat_DLIR, ven_HU_lvr1_DLIR, ven_SD_lvr1_DLIR, ven_HU_lvr2_DLIR, ven_SD_lvr2_DLIR, tard_SD_fat_DLIR, tard_HU_lvr1_DLIR, tard_SD_lvr1_DLIR, tard_HU_lvr2_DLIR, tard_SD_lvr2_DLIR.

For the body region scanned, only the chest, abdomen and trunk (chest + abdomen) were included. Other regions such as the head, neck or limbs were excluded.

The total dose resulting using DLIR, includes also the dose needed to create the topogram and the monitoring, so it is not a pure addition of the DLP in the various phases for that examination.

After the acquisition of this data, it was decided to transform all values ending in "_Aq" (Aquillion ONE using AIDR3D) and those ending in "_DE" (Definition Edge, using ADMIRE) to the same ending "_IR" (Iterative Reconstruction) in order to lose information about the difference between the two machines.

If you are curious about the data kept split between these two methods, delve deeper into data in Appendix A.

5.2 Patient Population

As a result of these selections, a total of 83 patients were included in the study. All patients are unique and appear only once in the database. For patients who had repeated the scans multiple times (for example, DLIR on day 1, FBP on day 3, AIDR3D on day 4), only the comparison between the two closest dates was considered (in our example, DLIR on Day 1 and FBP on day 3).

5.3 CT Protocol

The CT scans were performed using four different scanners:

- GE Healthcare: Revolution Evo (DLIR). The following imaging parameters were used: Tube voltage: 100.00 kV, Tube current: 190.00 mA, Slice thickness: 1.25 mm, Spiral Pitch Factor: 0.5156, Matrix size: 512 x 512.
- Siemens Healthineers: Somatom Sensation 64 (FBP). The following imaging parameters were used: Tube voltage: 120.00 kV, Tube current: 359.00 mA, Slice thickness: 3.00 mm, Spiral Pitch Factor: 0.8000, Matrix size: 512 x 512.
- Canon Medical Systems (Previously Toshiba): Aquillion ONE (AIDR3D). The following imaging parameters were used: Tube voltage: 120.00 kV, Tube current: 88.00

mA, Slice thickness: 3.00 mm, Spiral Pitch Factor: 0.8130, Matrix size: 512 x 512.

- Siemens Healthineers: Somatom Definition Edge (ADMIRE). The following imaging parameters were used: Tube voltage: 120.00 kV, Tube current: 200.00 mA, Slice thickness: 3.00 mm, Spiral Pitch Factor: 0.6000, Matrix size: 512 x 512.

5.4 Quantitative Dose Analysis

They were assessed through CTDI, DLP and total dose, as shown in the Patient Protocol. The effective dose was calculated on DLP using the conversion factor illustrated in subsection 1.4.6.

5.5 Quantitative Image Quality Analysis

Initially, a program with MATLAB [56] was written to automatically compute image noise con DICOM images with the Global Noise Level method presented by Christianson et al. [57], following the instructions in the paper by Malkus et al. [58]. However, due to local regulations that prohibit the exportation of DICOM data, even when anonymised, it was not possible to apply it to the patients in this study.

It was therefore chosen to follow the calculations also used by Nam et al. [20].

5.6 Data Analyses

Data analyses were performed using Python 3 [53] running within a Jupyter Notebook [54] environment. The libraries employed for data manipulation and statistical testing were Pandas [59], NumPy [60], and SciPy [61]. Additionally, the datetime module [62] was employed.

Pandas was used for data structuring and preprocessing, while NumPy provided support for numerical operations. SciPy was utilized for statistical testing, and the datetime module was used to handle date and time data. These allowed for robust and reproducible analyses.

5.7 Statistical Analyses

In the main body of the study, the only applied statistics were mean, standard deviation and Wilcoxon signed-rank test, calculated on an overall mean of all parameters computed across each phase.

The more detailed approach that is described in this section was originally applied but then eliminated due to the belief that the results lacked readability. The results of these analyses are still present in Appendix A.

Please feel free to skip the analytical part if you think that it is too complicated or that you have never heard of those tests. They are not relevant to the main body of the study, and only appear in Appendix A. If you know what mean, standard deviation and signed-rank test are, you are good to skip to the next section.

For statistical analyses, we first employed the Saphiro-Wilk test to see if our data was normally distributed, and consequently decide if it was best to proceed with parametric tests (that is, tests that assume a normal distribution) or non-parametric tests (that is, tests that don't assume a normal distribution) to assess if there was a statistically significant difference between the paired groups (FBP vs DLIR, AIDR3Dvs DLIR, ADMIRE vs DLIR).

When the distribution of our data was found not to be normal, we executed the Wilcoxon signed-rank test.

If the distribution of the data was found to be normal, we executed both the Paired t-Test and the Wilcoxon signed-rank test, since the sample size was limited.

For the effect sizes, we computed the Cohen's d in both cases.

5.7.1 Saphiro-Wilk Test

The Shapiro-Wilk test is one of the most powerful tests for assessing normality, particularly when dealing with small sample sizes [63].

In this test, the null hypothesis is that the data comes from a normally distributed population. If the calculated p-value is lower than the chosen α (in our case 0.05), we reject the null hypothesis, indicating that the data isn't, probably, normally distributed.

For the computation, the function scipy.stats.shapiro [64] was used.

5.7.2 Paired t-Test

The paired t-test [65], also known as the dependent t-test, investigates whether there is a significant difference in the mean values of the same group on two different occasions. This test is used when the samples are dependent; that is, when there is a natural pairing of observations in the data, and the data is normally distributed.

In this test, the null hypothesis states that there is no significant difference between the paired observations. In our study, we used the one-sided version of the test, under the alternative hypothesis that one group (FBP or iterative algorithms) had larger values than the other group (DLIR). Therefore, if the p-value is lower than the chosen α (in our case, again 0.05), we reject the null hypothesis, suggesting a significant difference.

5.7.3 Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test [65] is a non-parametric test, meaning that it makes no assumptions about a population's parameters, such as normal distribution. It is used to determine whether there are differences between two sets of paired observations when the conditions to conduct a paired t-test are not met.

In our case, an example of paired observation was an examination with DLIR paired with an examination with ADMIRE by means of the same patient ID.

The test first computes the absolute values of the differences between paired observations, then ranks them and tests whether the distribution of these ranks is symmetric about zero, in which case it would imply that there is no observable difference between the paired groups. In our study, we used the one-sided version of the test, under the alternative hypothesis that one group (FBP or iterative algorithms) typically had larger values than the other (DLIR). The significance level α for all tests was set to 0.05, again under the null hypothesis that there is no significant difference between the paired observations.

For the computation, the function scipy.stats.wilcoxon [66] was used.

5.7.4 Cohen's d

Cohen's d [65] is a measure of effect size that quantifies the difference between two means in standard deviation units. In other words, it gives a measure of how far two distributions

are with regard to each other, as illustrated in Figure 5.2. It is calculated by subtracting the mean of one group from the mean of another and dividing the result by the pooled standard deviation of the two groups.

There are no hard rules to interpret the Cohen's d, but a commonly accepted rule of thumb is that values smaller than 0.20 indicate a small effect size, those between 0.20 and 0.75 show a moderate effect, and those over 0.75 are considered to be large.

For the computation, the code presented in Appendix F was used.

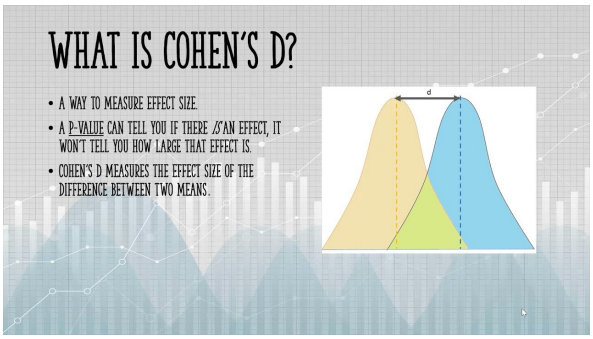

Figure 5.2: Cohen's d. Image reproduced from statisticshowto.com [67].

Chapter 6

Results

6.1 Patient Demographics

Of the 83 selected patients, 56 are males (67%), and 27 are females (33%). The difference in prevalence between the two genders is maybe not so surprising: notably, males more often receive critical care than women [68], and all our patients were in a situation where they had to be scanned multiple times within a short period. Of those patients, 14 underwent examination of the chest, 51 of the abdomen, and 18 of the trunk (chest + abdomen).

The mean age of the patient when they underwent the DLIR was 58.9 years old, with a standard deviation of 15.8 years. The youngest patient was 31 years old, and the oldest was 84.

The mean number of days that intercurred between the two examinations was 10.8 days, with a standard deviation of 8.6 days. The minimum delay was of 1 day, and the maximum difference was 30 days.

6.2 Quantitative Dose Analyses

Quantitative Dose Analyses are reported in Table VI. The study as it was conducted only supported a dose reduction for FBP when stratification according to body region was not undertaken. This may be due to the very limited patients' number in this study group (n = 12). Consequently, some regions were heavily underrepresented, for example patients for the trunk group were just 2.

The study couldn’t find any improvement in dose with regard to iterative algorithms.

Table VI: Dose Analysis Results. * indicates a p value < 0.05 for that cell. when compared to the same values in DLIR. Therefore, we expect the p-value in DLIR not to be significant, since the comparison is with its own population.

		CTDI in mGy			DLP in mGy*cm			Effective Dose in mSv		
		Mean ± SD	Range	p	Mean ± SD	Range	p	Mean ± SD	Range	p
FBP	Total	14.00 ± 12.46	(6.41, 51.79)	*	1085.33 ± 626.30	(284.00, 2105.00)	*	16.13 ± 9.55	(3.98, 31.57)	*
	Chest	13.22 ± 4.43	(7.55, 18.35)		461.50 ± 156.90	(284.00, 666.00)		6.46 ± 2.20	(3.98, 9.32)	
	Abdomen	9.58 ± 3.24	(6.41, 13.90)		1633.80 ± 458.34	(881.00, 2105.00)		24.51 ± 6.88	(13.21, 31.57)	
	Trunk	22.40 ± 25.45	(7.31, 51.79)		1003.00 ± 399.68	(555.00, 1323.00)		15.04 ± 6.00	(8.32, 19.84)	
DLIR	Total	24.67 ± 61.01	(3.45, 355.42)		1235.53 ± 873.67	(145.89, 4528.42)		18.45 ± 13.16	(2.04, 67.93)	
	Chest	22.50 ± 43.45	(4.30, 171.56)		485.15 ± 352.73	(145.89, 1295.67)		6.79 ± 4.94	(2.04, 18.14)	
	Abdomen	16.63 ± 41.46	(3.82, 236.76)		1278.43 ± 829.01	(199.67, 4528.42)		19.18 ± 12.44	(3.00, 67.93)	
	Trunk	49.14 ± 103.07	(3.45, 355.42)		1697.62 ± 932.73	(253.03, 3273.26)		25.46 ± 13.99	(3.80, 49.10)	
IR	Total	9.56 ± 5.86	(3.52, 35.27)		1350.62 ± 1191.68	(106.00, 6778.00)		20.19 ± 17.91	(1.48, 101.67)	
	Chest	10.00 ± 5.90	(3.52, 20.69)		502.50 ± 433.16	(106.00, 1489.00)		7.04 ± 6.06	(1.48, 20.85)	
	Abdomen	8.85 ± 5.11	(3.60, 28.70)		1178.28 ± 833.98	(201.00, 4752.00)		17.67 ± 12.51	(3.01, 71.28)	
	Trunk	11.46 ± 7.74	(4.25, 35.27)		2444.53 ± 1688.44	(564.00, 6778.00)		36.67 ± 25.33	(8.46, 101.67)	

6.3 Quantitative Image Quality Analyses

Table VII: Image Quality Results. * indicates a p-value < 0.05 for that cell, when compared to the same values in DLIR. Therefore, we expect the p-value in DLIR not to be significant, since the comparison is with its own population.

	Noise in HU			SNR in HU		
	Mean ± SD	Range	p	Mean ± SD	Range	p
FBP	14.85 ± 2.73	(11.50, 18.94)	*	3.99 ± 1.23	(2.37, 6.15)	*
DLIR	8.45 ± 3.24	(4.29, 18.19)		11.53 ± 9.28	(6.55, 30.30)	
IR	28.85 ± 32.77	(7.33, 105.50)	*	4.84 ± 2.74	(0.21, 8.71)	*

Noise and SNR were significantly improved with regards to both FBP and IR, as shown in Table VII. We therefore conclude that the advantage offered by DLIR was mainly invested in improving image quality, rather than lowering radiation dose.

Chapter 7

Discussion

In this retrospective study, we have performed a thorough comparison of different image reconstruction methods (DLIR versus FBP, AIDR3D or ADMIRE), each implemented on a different CT machine, across four different phases of acquisition (direct, arterial, venous, and delayed), for patients in critical care who needed multiple CT scans within a short period of time. The theory behind DLIR supposes an improvement in both dose parameters (CTDI, DLP, and Effective Dose) and image quality metrics (noise and SNR) when compared to the more traditional methods. Our goal was to evaluate if and to which extent this new technology has impacted the imaging of critical care patients who had to undergo multiple CT scans within a brief period of time (30 days).

From the analyzed data, we noticed that over the past year and a half, ever since the acquisition of the DLIR machine at our hospital, there has been a significant decrease in the use of FBP. This limited the number of patients included in this category (n= 12). Nonetheless, we found that DLIR's performance significantly surpassed FBP: image noise was reduced, on average, by 14.85 ± 2.73 HU (specifically by 50.97% in the direct phase, 48.54% in ar-terial phase, 52.74% in venous phase). The value of SNR was, on average, increased by 3.99 ± 1.23 HU, which translated to an astonishing improvement of 107.41% in the direct phase, 137.83% in the arterial phase, 129.37% in the venous phase, and 108.82% in the tar-dive phase. We could also describe a substantial reduction in radiation dose: overall, CTDI volume was reduced by 14.00 ± 12.46 mGy, DLP by 1085.33 ± 626.30 mmGy, and Effective Dose by 16.13 ± 9.55 mSv. Specifically, in the direct phase, CTDI was reduced by 40.94%, DLP by 38.1%, and the effective dose by 36.52%. A radiation dose reduction for the phases

with contrast medium could not be confirmed, but the results in the direct phase were significant enough to have an impact on the total dose of the examination, which was lowered by 19.58% on average. These results mirror the efficiency of DLIR over FBP reported by other studies.

Comparing DLIR with the general group of iterative reconstructions also resulted in a significant improvement in image quality: image noise was improved on average by 28.85 ± 32.77 HU, and SNR by 4.84 ± 2.74 HU. For this general group however, our study could not observe a significant dose reduction.

On the other hand, if we consider just the group of the iterative algorithm AIDR3D versus DLIR, we have noticed a similar pattern as FBP. It was calculated on a subset of patients that coincidentally had the same number (n= 12). Specifically, DLIR yielded a 31.12% reduction in image noise in the direct phase, 25.14% in the arterial phase and 40.63% in the venous phase. It also increased the signal-to-noise ratio (SNR) by 52.57% in the direct phase and 85.65% in the venous phase. For this patient group, we could describe a radiation dose reduction in three phases. Specifically, in the direct phase, the reductions were 41.77% in CTDI, 38.19% in DLP, and 38.19% in effective dose. For the arterial phase we saw a reduction of 34.10% in CTDI, 33.31% in DLP, and 33.31% in effective dose. For the venous phase the extent was somewhat smaller: 18.64% in CTDI, 14.98% in DLP, and effective dose 14.98%. These reductions accounted for a reduction of the total dose of examination by an average of 25.13%. These findings also align with the results of other studies.

Lastly, for the comparison group (n=59) between ADMIRE and DLIR, an improvement of the dose was found but to a much more moderate degree, and only in the direct phase: CTDI was reduced by 21.83%, DLP by 19.89% and effective dose by 19.26%. However, image quality improvement had a consistent impact across all phases: image noise was reduced by 68.28%, 73.49%, 79.35% and 78.19% respectively for direct, arterial, venous and tardive phase, and SNR was improved by 134.14%, 100.76%, 119.49% and 163.70% each.

Overall we can say that we found consistency in noise reduction, and above all in image quality improvement, which emphasizes the robustness of the DLIR approach and its potential in daily clinical practice. However, further evaluation is needed to investigate potential differences in other anatomical locations such as the head or limbs, or specific to a given pathology. It would be also interesting to investigate the extent to which these quantitative improvements reflect on clinicians, and particularly radiologists, in daily clinical practice,

to quantify if the image quality significantly affects parameters such as subjective quality, or the impact on workflow and efficiency, for example in terms of time needed to read the images or correctly diagnose pathologies.

In conclusion, our study contributed to the growing body of evidence that deep learning-based CT reconstruction methods like TrueFidelity can not only provide significant benefits in terms of dose reduction and image quality over the traditional FBP, but also outperform other iterative reconstruction methods. The expectation is that those advantages will trans-late into safer imaging practices, higher diagnostic confidence, and ultimately better patient care. Considering the increasing computational power and the growing wealth of healthcare institutions, we anticipate a broader implementation of DLIR in the foreseeable future.

Chapter 8

Conclusion

In conclusion, we found evidence that the reconstruction power of DLIR is mainly used to improve image quality, and to a lesser extent, to lower CT dose.

Part IV

Appendices

Appendix A

Delve Deeper into Data

During the drafting of the thesis, a more complex analysis was performed. In this section, first we present the main findings that had a significant value, with a $p < 0.05$. Every other non-reported parameter was found not to be significant, with a $p > 0.05$. At the end of the section are then listed patient demographics and analysis of normality results. If you wish for more specific results of the tests, please contact the author. They were abridged to provide a better overview.

A.1 Results stratified by acquisition phase

A.1.1 Results in FBP vs DLIR

Our alternative study could find a dose reduction due to DLIR for parameters in the direct and arterial phases.

Table VIII: Dose reduction results for FBP vs DLIR: an alternative quantification

Phase	Estimated % Reduction	Mean Reduction	Cohen's d
dir_CTDI	40.94%	4.39 ± 5.64 mGy	1.05 (large effect)
dir_DLP	38.01%	178.56 ± 285.14 mGy * cm	0.84 (large effect)
Effective Dose in art	36.52%	2.51 ± 3.33 mSv	1.02 (large effect)
art_CTDI	26.63%	2.73 ± 3.98 mGy	0.87 (large effect)
tot	19.58%	212.54 ± 929.73 mGy	0.31 (moderate effect)

Table IX: Quality improvement results in FBP vs DLIR: an alternative quantification

Phase	Parameter	Estimated % Improvement	Mean Improvement	Cohen's d
dir	Noise	50.97%	7.09 ± 3.31 HU	2.81 (large effect)
	SNR	107.41%	2.87 ± 1.69 HU	2.22 (large effect)
art	Noise	48.54%	7.35 ± 3.23 HU	2.87 (large effect)
	SNR	137.83%	5.12 ± 2.35 HU	2.76 (large effect)
ven	Noise	52.74%	8.03 ± 3.33 HU	3.16 (large effect)
	SNR	129.37%	6.99 ± 3.40 HU	2.69 (large effect)
tard	Noise	–	–	–
	SNR	108.82%	5.90 ± 1.69 HU	4.03 (large effect)

A.1.2 Results in AIDR3D (iterative) vs DLIR

Under both parametric and non-parametric testing, the DLIR superiority in terms of radiation dose could be detected for both CTDI and DLP in direct and arterial phases, CTDI in the venous phase, and total dose derived from the examination. Cohen's d values often above 0.75 suggest that this reduction had a remarkable value, and it was enough to impact the total dose as well.

Furthermore, parametric testing suggested that there is a reduction of dose for the DLP during the venous phase as well. Since data regarding this phase was determined to be normally distributed, we can confidently accept this result.

Table X: Dose reduction results for AIDR3D vs DLIR: an alternative quantification

Phase	Estimated % Reduction	Mean Reduction	Cohen's d
dir_CTDI	41.77%	4.61 ± 6.44 mGy	0.97 (large effect)
dir_DLP	38.19%	205.67 ± 334.62 mGy * cm	0.83 (large effect)
Effective Dose in dir	38.19%	3.08 ± 4.67 mSv	0.89 (large effect)
art_CTDI	34.10%	3.75 ± 3.60 mGy	1.36 (large effect)
art_DLP	33.31%	176.46 ± 206.52 mGy * cm	1.12 (large effect)
Effective Dose in art	33.31%	2.65 mSv	1.57 (large effect)
ven_CTDI	18.64%	1.99 ± 2.78 mGy	0.96 (large effect)
ven_DLP	14.98%	86.06 ± 179.25 mGy * cm	0.64 (moderate effect)
Effective Dose in ven	14.98%	1.29 ± 2.06 mSv	0.84 (large effect)
tot	25.13%	369.90 ± 906.98 mGy	0.55 (moderate effect)

About the quality improvement:

Table XI: Quality improvement in AIDR3D vs DLIR: an alternative quantification

Phase	Parameter	Estimated % Improvement	Mean Improvement	Cohen's d
dir	Noise	31.12%	2.98 ± 1.72 HU	2.34 (large effect)
	SNR	52.57%	2.43 ± 1.27 HU	2.58 (large effect)
art	Noise	25.14%	2.45 ± 1.98 HU	1.62 (large effect)
	SNR	–	–	–
ven	Noise	40.63%	4.67 ± 3.35 HU	1.86 (large effect)
	SNR	85.65%	7.45 ± 13.67 HU	0.73 (moderate effect)

A.1.3 Results in ADMIRE (iterative) vs DLIR

Under non-parametric testing, the DLIR superiority in terms of radiation dose could only be detected in the direct phase. Cohen's d value slightly greater than 0.2 suggests that this reduction had a moderate effect.

Again for this patients' group, the results of the testing did not change even when stratification was made according to the body region (chest, abdomen or trunk).

Table XII: Dose reduction results for ADMIRE vs DLIR: an alternative quantification

Phase	Estimated % Reduction	Mean Reduction	Cohen's d
dir_CTDI	21.83%	2.30 ± 8.60 mGy	0.37 (moderate effect)
dir_DLP	19.89%	108.44 ± 486.08 mGy * cm	0.31 (moderate effect)
Effective Dose	19.26%	1.56 ± 5.97 mSv	0.37 (moderate effect)

Table XIII: Quality improvement in ADMIRE vs DLIR: an alternative quantification

Phase	Parameter	Estimated % Improvement	Mean Improvement	Cohen's d
dir	Noise	68.28%	19.66 ± 23.68 HU	1.14 (large effect)
	SNR	134.14%	3.75 ± 2.78 HU	1.86 (large effect)
art	Noise	73.49%	28.84 ± 35.18 HU	1.13 (large effect)
	SNR	100.76%	3.96 ± 3.65 HU	1.50 (large effect)
ven	Noise	79.35%	38.16 ± 50.63 HU	1.04 (large effect)
	SNR	119.49%	6.49 ± 5.13 HU	1.74 (large effect)
tard	Noise	78.19%	34.33 ± 40.89 HUU	1.16 (large effect)
	SNR	163.70%	6.51 ± 4.24 HU	2.12 (large effect)

A.2 Patient Demographics

A.2.1 Patient Demographics in FBP vs DLIR

Of all the patients included, if we consider only those that were selected to compare FBP and DLIR, we found 12 patients. 9 of them are males (75%), and 3 females (25%).

Of those, 4 underwent examination of the chest, 5 of the abdomen, and 3 of the complete trunk (chest + abdomen).

The mean age of the patient when they underwent the DLIR was 58.0 years old, with a standard deviation of 8.48 years. The youngest patient was 52 years old, and the oldest was 64.

The mean number of days that intercurred between the two examinations was 17.0 days, with a standard deviation of 9.7 days. The minimum delay was of 2 days, and the maximum difference was 29 days.

A.2.2 Patient Demographics in AIDR3D (iterative) vs DLIR

Similarly, if out of all the included patients we consider only those that were selected to compare AIDR3D and DLIR, we found 12 patients. 10 of them are males (83%), and 2 females (17%).

Of those, none underwent examination of the chest, 11 of the abdomen, and just 1 of the complete trunk (chest + abdomen).

The mean age of the patient when they underwent the DLIR was 50.5 years old, with a standard deviation of 27.5 years. The youngest patient was 31 years old, and the oldest was 70.

The mean number of days that intercurred between the two examinations was 11.3 days, with a standard deviation of 8.7 days. The minimum delay was of 2 days, and the maximum difference was 27 days.

A.2.3 Patient Demographics in ADMIRE (iterative) vs DLIR

Finally, if out of all the included patients we consider only those that were selected to compare ADMIRE and DLIR, we found 59 patients. 37 of them are males (63%), and 22 are females (37%).

Of those, 10 underwent examination of the chest, 35 of the abdomen, and 14 of the complete trunk (chest + abdomen).

The mean age of the patient when they underwent the DLIR was 61.3 years old, with a standard deviation of 15.8 years. The youngest patient was 34 years old, and the oldest was 84.

The mean number of days that intercurred between the two examinations was 9.5 days, with a standard deviation of 8.0 days. The minimum delay was of 1 day, and the maximum difference was 30 days.

A.3 Assessing Normality of the Distribution

A.3.1 Assessing Normality in FBP vs DLIR

Overall, in FBP vs DLIR the distribution seemed to be normal, so we proceeded to apply parametric testing and accept its results. The only exceptions to the normal distribution were:

- venous CTDI;
- venous DLP;
- venous noise;
- tardive noise.

For those parameters, we only considered results from non-parametric testing.

A.3.2 Assessing Normality in AIDR3D (iterative) vs DLIR

Overall, the data seemed to be normally distributed with the exception of:

- direct DLP;
- venous noise;
- venous SNR.

Only in those cases, we considered results from non-parametric testing.

A.3.3 Assessing Normality in ADMIRE (iterative) vs DLIR

Overall, the Saphiro-Walk test's results strongly suggested a departure from a normal distribution, with regard to each parameter in each acquisition phase. Therefore, for this patient group the study was continued with non-parametric analyses.

The only exception was SNR, which was found to have a normal distribution in all phases, so we used parametric testing.

Appendix B

Foreword to the Source Code Appendices

Coming from a medical background, it may seem challenging to gather up the courage to approach coding, especially since it can be, at times, an incredibly frustrating experience. Nonetheless, it is an extraordinarily powerful and fascinating tool that enables one to conduct hands-on experiments and gain a better insight into what and how various things are computed.

In the hope of rendering service to at least one other fellow amateur coder, some of the source code written during the drafting of this thesis is listed here on the following pages.

Kindly note that this work was authored by a non-professional, and there is therefore no guarantee of the correctness or efficacy of the computations.

Appendix C

BP Simulation in MATLAB

```
%% BP Simulation

% Load the image into the object space and convert it to
    grayscale
phantom_image = imread('C:\Users\Documents\MATLAB\
    Creating_pictures\transversalschnitt.png');
phantom_image = rgb2gray(phantom_image);
phantom_image = im2double(phantom_image);

% Specify the angle range for the projections
theta = 0:179; % For example in thsi case, we consider theta
    from 0 to 179 in steps of one (0, 1, 2, 3...)

% Compute the Radon Transform to generate the sinogram
[R, xp] = radon(phantom_image, theta);


% Perform the inverse radon transform without filtering
reconstructed_image = iradon(R, theta, 'linear', 'none', 1.0,
     size(phantom_image, 1));

```

```
18 % Display the original phantom image
19 subplot(1, 3, 1);
20 imshow(phantom_image, []);
21 title('Original Image (Object Space)');
22
23 % Display the sinogram
24 subplot(1, 3, 2);
25 imshow(R, []), title('Sinogram (Radon Space)');
26
27 % Display the reconstructed image
28 subplot(1, 3, 3);
29 imshow(reconstructed_image, []), title('Backprojection (no
       filters)');
30
31 % Save the figure
32 saveas(gcf, 'BP_simulation.png')
```

Appendix D

FBP Simulation in MATLAB

```
%% FBP simulation

% Load the image into the object space and convert it to
    grayscale
phantom_image = imread('C:\Users\Documents\MATLAB\
    Creating_pictures\transversalschnitt.png');
phantom_image = rgb2gray(phantom_image);
phantom_image = im2double(phantom_image);

% Specify the angle range for the projections
theta = 0:2:179; % For example in this case, we consider
    theta from 0 to 179 in steps of two (0, 2, 4, 6...)

% Compute the Radon Transform to generate the sinogram
[R, xp] = radon(phantom_image, theta);

% Perform 2D Fourier Transform
F = fft2(phantom_image);

% Shift the zero-frequency component to the center of the
    spectrum
```

```
18 F_shifted = fftshift(F);
19
20 % Compute the inverse Radon transform from the filtered
       sinogram
21 reconstruction_wf = iradon(R, theta, 'linear', 'Ram-Lak',
       1.0, size(phantom_image, 1));
22
23 % Display the original phantom image
24 subplot(1, 4, 1);
25 imshow(phantom_image, []);
26 title('Original Image (Object Space)');
27
28 % Display the sinogram
29 subplot(1, 4, 2);
30 imshow(R, [], 'Xdata',theta,'Ydata',xp,'InitialMagnification'
       ,'fit');
31 title('Sinogram (Radon Space)');
32
33 % Display the absolute value of the Fourier Transform (log
       scale)
34 subplot(1, 4, 3);
35 imshow(log(abs(F_shifted) + 1), []), title('Frequency
       analysis (Fourier Space)')
36
37 % Display the reconstructed image
38 subplot(1, 4, 4);
39 imshow(reconstruction_wf, []);
40 title('Reconstructed Image (from filtered sinogram)');
41
42 % Save the figure
43 saveas(gcf, 'FBP_simulation.png')
```

Appendix E

Global Noise Level in MATLAB

In this section, two sets of code are presented.

The first set illustrates, in a step-by-step approach, how to determine the noise level in a single image, calculated on the soft tissue as illustrated by Christianson et al. [57], inspired by the explanations presented in the paper by Malkus et al. [58]. During this stuy, it was chosen to calculate the noise level on air, as it is the usual place of choice. However, as it provides better visual feedback to calculate it on soft tissue first, the code for soft tissue was included. To switch from soft tissue to air, it will suffice to change the segmentation range in the mask (as stated in the code).

The second set illustrates how to automatically iterate through the whole dicom folder, determining the GNL in a more precise way, since noise is not equally distributed across the CT scan. When running this second code, it is recommended to disable the displaying of the images shown in the step-by-step approach to speed up the computing process.

Sadly, due to local regulations that prohibit the exportation of dicom files from the hospital computers, even when anonymized, and at the same time prohibits the installation of programs such as MATLAB on their computers, these codes could never be executed on the real patient data in this thesis. It was nonetheless entertaining to develop the code, and it was tried on a training set kindly provided at dicomlibrary.com [69].

E.1 For a single image

```
1 %% Global Noise Level (GNL) is assessed through following
```

```
    steps:
% Step 0: import the image
% Step 1: segment the image to select a specific tissue type
% Step 2: create a noise map
% Step 3: generate a histogram with the standard deviations
% Step 4: compute the GNL


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% STEP 0: IMPORT THE IMAGE (version for one single image)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


% Specify the folder where the DICOM files are stored
folder = "C:\Users\Documents\DLIR\Assessing_image_noise\
    series-00000";

% Read the DICOM file in the folder
info = dicominfo(fullfile(folder, 'image-00115.dcm')); %
    choose an image to calculate the noise
img = dicomread(info);

% Display the image
imshow(img, []);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```
%% STEP 1: SEGMENT THE IMAGE TO SELECT SOFT TISSUE (0-100HU)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Extract the slope and intercept values from the metadata:
slope = info.RescaleSlope;
intercept = info.RescaleIntercept;

% Convert the image data to Hounsfield Units (HU):
img_hu = double(img) * slope + intercept;

% Create a binary mask of the areas between 0 and 100 HU:
mask = (img_hu >= 0) & (img_hu <= 100);
% set ''mask = imb_hu < -800'' to calculate noise level on
    air

% Apply the mask to the original image
img_segmented = img_hu .* mask;

% Find connected components
CC = bwconncomp(mask);

% Remove small objects. Keep only the largest connected
    component (assuming it's the body, to exclude for example
    the cushion. It also works on air, since the largest
    connected region in the picture should be the air itself,
    cushion excluded)
numPixels = cellfun(@numel,CC.PixelIdxList);
[~,idx] = max(numPixels);
mask = false(size(mask));
```

```
56 mask(CC.PixelIdxList{idx}) = true;
57
58 % Apply again the new mask to the original image
59 img_segmented = img_hu .* mask;
60
61 % Display masked image
62 imshowpair(img_hu, img_segmented, 'montage')
63
64
65
66 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
67
68 %% STEP 2: CREATE A NOISE MAP OF THE SOFT TISSUE
69
70 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
71
72 % Define the size of the neighborhood kernel in pixels
73 kernel_size_mm = 7;  % Size of kernel in millimeters
74 pixel_size_mm = 25.4/96;  % Size of pixel in millimeters, if
     96ppi
75 kernel_size_px = round(kernel_size_mm / pixel_size_mm);  %
     Size of kernel in pixels
76 if mod(kernel_size_px, 2) == 0
77     kernel_size_px = kernel_size_px + 1
78 end
79
80 % Apply stdfilt function to the masked image with the defined
      kernel size
81 std_image = stdfilt(img_segmented, true(kernel_size_px));
82 imshow(std_image, [])
83 std_image = std_image .* mask;
```

```
84
85 % Display the heatmap
86 figure;
87 imagesc(std_image);
88 colormap jet;
89 colorbar;
90
91
92
93 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
94
95 %% STEP3: GENERATE A HISTOGRAM WITH THE CORRESPONDING SDs
96
97 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
98
99
100 % Extract std values only in the masked region
101 std_values = std_image(mask);
102
103 % Round them to the nearest integer
104 std_values_int = round(std_values)
105
106 % Plot them in a histogram
107 bins = max(std_values_int)
108 if mod(bins, 2) == 1
109     bins = bins + 1
110 end
111
112 histogram(std_values, bins);
113 title('Standard deviation histogram');
114 xlabel('SD in HU');
```

```
115 ylabel('Frequency');
116
117
118
119 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
120
121 %% STEP 4: DETERMINE GLOBAL NOISE LEVEL
122
123 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
124
125
126 % We do this by identifying the mode of the histogram peak
      corresponding to homogeneous tissue
127 gnl = mode(std_values_int)
128
129 % Print the bin with the highest count
130 fprintf('The Global Noise Level of this single image is %d\n'
      , gnl);
```

E.2 Iterative version

```
1 %% GLOBAL NOISE LEVEL (GNL): ITERATIVE VERSION THROUGH WHOLE
    DICOM FOLDER
2
3 % Feedback of initiated execution
4 fprintf("Code in execution...")
5
6 % STEP 0: IMPORT THE IMAGES (iterative version)
7
8 folder = "C:\Users\Documents\DLIR\Assessing_image_noise\
    series-00000";
```

```
% Get a list of all DICOM files in the directory
filelist = dir(fullfile(folder, '*.dcm'));

% Iterate over all DICOM files
gnl_list = []  % Initialize an empty list to store the image-specific gnls
for i = 1:length(filelist)

  % Read the DICOM files in the folder
    img = dicomread(fullfile(folder,filelist(i).name));
    info = dicominfo(fullfile(folder,filelist(i).name));

    % Iterate to find out GNL within image
    Image_noise_step1(); % STEP 1, if the code we previously saw in Step 1 is saved as ''Image_noise_step1.m'' in the same directory. Consider improving computational speed by disabling the displaying of the single images in the code within the file.
    Image_noise_step2(); % STEP 2, apply same considerations as in Step 1. File name must be ''Image_noise_step2.m''
    Image_noise_step3(); % STEP 3 , apply same considerations as in Step 1. File name must be ''Image_noise_step3.m''
    Image_noise_step4(); % STEP 4, apply same considerations as in Step 1. File name must be ''Image_noise_step4.m''
    gnl_list(end+1) = gnl;
end

mean_gnl = mean(gnl_list)
fprintf('The Global Noise Level of the whole CT scan is %.2f\n HU', mean_gnl);
```

Appendix F

Cohen's d computation in Python

```python
def cohens_d(column):
    data_FBP = df[column + '_FBP'].dropna()
    data_DLIR = df[column + '_DLIR'].dropna()

    # Ensuring paired data (to automatically filter out IDs where either is NaN)
    common_ids = set(data_FBP.index).intersection(set(data_DLIR.index))
    data_FBP = data_FBP.loc[common_ids]
    data_DLIR = data_DLIR.loc[common_ids]

    # Proceed with the calculation if there is data remaining
    if len(common_ids) > 0:
        diff = data_FBP - data_DLIR
        n = len(diff)
        d = diff.mean() / diff.std(ddof=1)
        print(f"Cohen's d for {column}: {d}")
    else:
        print(f"Not enough data for the calculation of Cohen's d for {column}: ")

# Apply the tests for each parameter
parameters = ['dir_CTDI', 'dir_DLP', 'art_CTDI', 'art_DLP', 'ven_CTDI', 'ven_DLP', 'tard_CTDI', 'tard_DLP', 'tot']

for param in parameters:
    cohens_d(param)
```

Bibliography

[1] 1979: Allan MacLeod Cormack (1924-1998) | St John's College, University of Cambridge. https://www.joh.cam.ac.uk/1979-allan-macleod-cormack-1924-1998. (Visited on 05/16/2023).

[2] Caroline Richmond. "Sir Godfrey Hounsfield". In: BMJ : British Medical Journal 329.7467 (Sept. 2004), p. 687. ISSN: 0959-8138. (Visited on 05/18/2023).

[3] The Nobel Prize in Physiology or Medicine 1979. https://www.nobelprize.org/ prizes/medicine/1979/s (Visited on 05/16/2023).

[4] CT Scanning |Biomedical Tutorials |Mepits. https://www.mepits.com/tutorial/416/biomedical/ct-scanning. (Visited on 06/15/2023).

[5] Computed Tomography (CT). https://www.nibib.nih.gov/ science-education/science-topics/computed-tomography-ct. (Visited on 05/16/2023).

[6] Lifeng Yu. "Novel CT Acquisition". In: Computed Tomography : Approaches, Applications, and Operations. Ed. by Ehsan Samei and Norbert J. Pelc. Cham: Springer International Publishing, 2020, pp. 27–43. ISBN: 978-3-030-26957-9. DOI: 10.1007/978-3-030-26957-9_3. (Visited on 06/15/2023).

[7] Thorsten M. Buzug. Computed Tomography: From Photon Statistics to Modern Cone-Beam CT. Berlin: Springer, 2008. ISBN: 978-3-540-39407-5 978-3-540-39408-2.

[8] Albert Tarantola. Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2005. ISBN: 978-0-89871-572-9.

[9] Tormund Njølstad et al. “Low-Contrast Detectability and Potential for Radiation Dose Reduction Using Deep Learning Image Reconstruction—A 20-Reader Study on a Semi-Anthropomorphic Liver Phantom”. In: European Journal of Radiology Open 9 (Jan. 2022), p. 100418. ISSN: 2352-0477. DOI: 10.1016/j.ejro.2022.100418. (Visited on 06/15/2023).

[10] Michael P. Hartung, Mike Cadogan, and Michael P. Hartung and Mike Cadogan. Abdominal CT: Attenuation. Feb. 2023. (Visited on 06/15/2023).

[11] Tom Kimpe and Tom Tuytschaever. “Increasing the Number of Gray Shades in Medical Display Systems–How Much Is Enough?” In: Journal of Digital Imaging 20.4 (Dec. 2007), pp. 422–432. ISSN: 0897-1889. DOI: 10.1007/s10278-006-1052-3.

[12] Olav Christianson et al. “Automated Technique to Measure Noise in Clinical CT Examinations”. In: American Journal of Roentgenology 205.1 (July 2015), W93–W99. ISSN: 0361-803X, 1546-3141. DOI: 10.2214/AJR.14.13613. (Visited on 05/20/2023).

[13] Ervin D. Podgorsak. Radiation Oncology Physics: A Handbook for Teachers and Students. Vienna: International Atomic Energy Agency, 2005. ISBN: 978-1-280-84923-7.

[14] Fuyuhiko Tamanoi et al. “Studies on the Exposure of Gadolinium Containing Nanoparticles with Monochromatic X-rays Drive Advances in Radiation Therapy”. In: Nanomaterials 10.7 (July 2020), p. 1341. ISSN: 2079-4991. DOI: 10.3390/nano10071341. (Visited on 05/19/2023).

[15] Elijah Halliwell et al. “Increase in Linear Attenuation Coefficient by Changing Crystal Structure of Materials for Radiation Shielding and Biomedical Devices Safety”. In: Colloids and Surfaces A: Physicochemical and Engineering Aspects 622 (Aug. 2021), p. 126646. ISSN: 09277757. DOI: 10.1016/j.colsurfa.2021.126646. (Visited on 05/19/2023).

[16] M. S. Hamideen, R.S Abady, and J.M. Sharaf. “Study on Mass and Linear Attenuation Coefficients of PMMA as Human Tissue-Equivalent Material”. In: Latvian Journal of Physics and Technical Sciences 59.2 (Apr. 2022), pp. 64–73. ISSN: 2255-8896. DOI: 10.2478/lpts-2022-0012. (Visited on 05/19/2023).

[17] NIST: X-Ray Mass Attenuation Coefficients - Table 4. https://physics.nist.gov/PhysRefData/XrayMass (Visited on 05/22/2023).

[18] Jerrold T. Bushberg et al. The Essential Physics of Medical Imaging. Fourth edition, international edition. Philadelphia Baltimore New York London Buenos Aires Hong Kong Sydney Tokyo: Wolters Kluwer, 2021. ISBN: 978-1-975103-22-4 978-1-975167-66-0.

[19] WJ Hossack. Topic 5: Noise in Images.

[20] Ju Gang Nam et al. "Deep Learning Reconstruction for Contrast-Enhanced CT of the Upper Abdomen: Similar Image Quality with Lower Radiation Dose in Direct Comparison with Iterative Reconstruction". In: European Radiology 31.8 (Aug. 2021), pp. 5533–5543. ISSN: 1432-1084. DOI: 10.1007/s00330-021-07712-4. (Visited on 06/17/2023).

[21] D. J. Schroeder. Astronomical Optics. 2nd ed. San Diego: Academic Press, 2000. ISBN: 978-0-12-629810-9.

[22] Kalpana M. Kanal et al. "U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations". In: Radiology 284.1 (July 2017), pp. 120–133. ISSN: 0033-8419. DOI: 10.1148/radiol.2017161911. (Visited on 05/23/2023).

[23] BIPM. The International System of Units, 9th Edition.

[24] Joël Greffier et al. "Spectral CT Imaging: Technical Principles of Dual-Energy CT and Multi-Energy Photon-Counting CT". In: Diagnostic and Interventional Imaging 104.4 (Apr. 2023), pp. 167–177. ISSN: 22115684. DOI: 10.1016/j.diii.2022.11.003. (Visited on 05/24/2023).

[25] lowell. The Electron-Volt: eV. https://www.slideserve.com/lowell/the-electron-volt-ev. July 2014. (Visited on 06/17/2023).

[26] Thomas B. Shope, Robert M. Gagne, and Gordon C. Johnson. "A Method for Describing the Doses Delivered by Transmission X-Ray Computed Tomography". In: Medical Physics 8.4 (July 1981), pp. 488–495. ISSN: 00942405. DOI: 10.1118/1.594995. (Visited on 05/24/2023).

[27] Polymethyl Methacrylate (PMMA) - Properties, Uses & Application. https://omnexus.specialchem.com guide/polymethyl-methacrylate-pmma-acrylic-plastic. (Visited on 06/15/2023).

[28] Doaa Awda Hussein, Fanar M. Abed, and Alauldeen A Hasan. “3D OBJECT RECONSTRUCTION USING IMAGE- BASED TECHNIQUES FOR MEDICAL APPLICATIONS”. In: (2019). DOI: 10.13140/RG.2.2.18659.71207. (Visited on 05/25/2023).

[29] Cynthia H. McCollough et al. “CT Dose Index and Patient Dose: They Are Not the Same Thing”. In: Radiology 259.2 (May 2011), pp. 311–316. ISSN: 1527-1315. DOI: 10.1148/radiol.11101800.

[30] H. Smith and International Commission on Radiological Protection. 1990 Recommendations of the International Commission on Radiological Protection. 1. ed. ICRP Publication Radiation Protection 60. Oxford: Pergamon Press, 1991. ISBN: 978-0-08-041144-6.

[31] J. Valentin and International Commission on Radiological Protection, eds. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP Publication 103. Oxford: Elsevier, 2007. ISBN: 978-0-7020-3048-2.

[32] Cynthia H. McCollough, Jodie A. Christner, and James M. Kofler. “How Effective Is Effective Dose as a Predictor of Radiation Risk?” In: American Journal of Roentgenology 194.4 (Apr. 2010), pp. 890–896. ISSN: 0361-803X, 1546-3141. DOI: 10.2214/AJR.09.4179. (Visited on 05/28/2023).

[33] Cynthia McCollough et al. The Measurement, Reporting, and Management of Radiation Dose in CT. Tech. rep. AAPM, Jan. 2008. DOI: 10.37206/97. (Visited on 05/28/2023).

[34] Sarah Abdulla and Cristopher Clarke. FRCR Physics Notes: Medical Imaging Physics for the First FRCR Examination. ISBN: 978-1-9999885-2-4.

[35] Martin J. Willemink and Peter B. Noël. “The Evolution of Image Reconstruction for CT—from Filtered Back Projection to Artificial Intelligence”. In: European Radiology 29.5 (May 2019), pp. 2185–2195. ISSN: 1432-1084. DOI: 10.1007/s00330-018-5810-7. (Visited on 05/28/2023).

[36] AMBOSS GmbH. Darstellung Eines Lateralen Strahlengangs. https://next.amboss.com/de/search?q=tho (Visited on 05/30/2023).

[37] Marcel Beister, Daniel Kolditz, and Willi A. Kalender. “Iterative Reconstruction Methods in X-ray CT”. In: Physica Medica 28.2 (Apr. 2012), pp. 94–108. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2012.01.003. (Visited on 06/20/2023).

[38] I. Hernandez-Giron, W.J.H. Veldkamp, and J. Geleijns. AIDR 3D Enhanced — The Latest Hybrid Model-Based Iterative Dose Reduction Technology from Canon Medical. A Whitepaper.

[39] Juan C. Ramirez-Giraldo, Katharine L. Grant, and Rainer Raupach. ADMIRE: Advanced Modeled Iterative Reconstruction. A Whitepaper.

[40] SHRAVAN KUMAR BELAGAL MATH TAMOGHNA GHOSH. PRACTICAL MATHEMATICS FOR AI AND DEEP LEARNING. S.l.: BPB PUBLICATIONS, 2022. ISBN: 978-93-5551-194-2.

[41] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2016. ISBN: 978-0-262-03561-3.

[42] Jiang Hsieh et al. A New Era of Image Reconstruction: TrueFidelity - Technical White Paper on Deep Learning Image Reconstruction. July 2019.

[43] Xiang Yu et al. “A Systematic Survey of Deep Learning in Breast Cancer”. In: International Journal of Intelligent Systems 37.1 (Jan. 2022), pp. 152–216. ISSN: 0884-8173, 1098-111X. DOI: 10.1002/int.22622. (Visited on 06/07/2023).

[44] Timothy P. Szczykutowicz et al. “A Review of Deep Learning CT Reconstruction: Concepts, Limitations, and Promise in Clinical Practice”. In: Current Radiology Reports 10.9 (Sept. 2022), pp. 101–115. ISSN: 2167-4825. DOI: 10.1007/s40134-022-00399-5. (Visited on 06/14/2023).

[45] TrueFidelity CT. https://www.gehealthcare.com/products/truefidelity. (Visited on 06/14/2023).

[46] AiCE Deep Learning Reconstruction |CT |Canon Medical Systems. https://global.medical.canon/proc tomography/aice_dlr. (Visited on 06/14/2023).

[47] Software et al. Algomedica |Low Radiation CT Scan Medical Software. https://www.algomedica.com/ radation-ct-scans-algomedica. (Visited on 06/14/2023).

[48] Jingyu Zhong et al. “Deep Learning Image Reconstruction Algorithm Reduces Image Noise While Alters Radiomics Features in Dual-Energy CT in Comparison with Conventional Iterative Reconstruction Algorithms: A Phantom Study”. In: European Radiology 33.2 (Feb. 2023), pp. 812–824. ISSN: 1432-1084. DOI: 10.1007/s00330-022-09119-1. (Visited on 06/10/2023).

[49] Zlatan Alagic et al. “Deep Learning versus Iterative Image Reconstruction Algorithm for Head CT in Trauma”. In: Emergency Radiology 29.2 (2022), pp. 339–352. ISSN: 1070-3004. DOI: 10.1007/s10140-021-02012-2. (Visited on 06/10/2023).

[50] Jian Wang and Ma Hao Chen. SOMATOM Emotion 6/16-Slice Configuration: Application Guide. 2007.

[51] Joël Greffier et al. “Image Quality and Dose Reduction Opportunity of Deep Learning Image Reconstruction Algorithm for CT: A Phantom Study”. In: European Radiology 30.7 (July 2020), pp. 3951–3959. ISSN: 1432-1084. DOI: 10.1007/s00330-020-06724-w. (Visited on 06/10/2023).

[52] Dominik C. Benz et al. “Radiation Dose Reduction with Deep-Learning Image Reconstruction for Coronary Computed Tomography Angiography”. In: European Radiology 32.4 (Apr. 2022), pp. 2620–2628. ISSN: 1432-1084. DOI: 10.1007/s00330-021-08367-x. (Visited on 06/10/2023).

[53] Welcome to Python.Org. https://www.python.org/. June 2023. (Visited on 06/13/2023).

[54] Project Jupyter. https://jupyter.org. (Visited on 06/13/2023).

[55] Microsoft Excel Spreadsheet Software |Microsoft 365. https://www.microsoft.com/en-us/microsoft-365/excel. (Visited on 06/20/2023).

[56] MATLAB Home. https://matlab.mathworks.com/. (Visited on 06/14/2023).

[57] Olav Christianson et al. “Automated Technique to Measure Noise in Clinical CT Examinations”. In: American Journal of Roentgenology 205.1 (July 2015), W93–W99. ISSN: 0361-803X, 1546-3141. DOI: 10.2214/AJR.14.13613. (Visited on 06/13/2023).

[58] Annelise Malkus and Timothy P. Szczykutowicz. “A Method to Extract Image Noise Level from Patient Images in CT”. In: Medical Physics 44.6 (June 2017), pp. 2173–2184. ISSN: 00942405. DOI: 10.1002/mp.12240. (Visited on 06/13/2023).

[59] Pandas - Python Data Analysis Library. https://pandas.pydata.org/. (Visited on 06/14/2023).

[60] NumPy. https://numpy.org/. (Visited on 06/14/2023).

[61] SciPy. https://scipy.org/. (Visited on 06/14/2023).

[62] Datetime — Basic Date and Time Types — Python 3.11.4 Documentation. https://docs.python.org/3/lib (Visited on 06/14/2023).

[63] Nornadiah Mohd Razali and Bee Yap. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests". In: J. Stat. Model. Analytics 2 (Jan. 2011).

[64] Scipy.Stats.Shapiro — SciPy v1.10.1 Manual. https://docs.scipy.org/doc/scipy/reference/generated/sci (Visited on 06/16/2023).

[65] Timothy C. Urdan. Statistics in Plain English. Fifth. New York: Routledge, Feb. 2022. ISBN: 978-1-00-300645-9. DOI: 10.4324/9781003006459. (Visited on 06/14/2023).

[66] Scipy.Stats.Wilcoxon — SciPy v1.10.1 Manual. https://docs.scipy.org/doc/scipy/reference/generated/s (Visited on 06/16/2023).

[67] Cohen's D: Definition, Examples, Formulas. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cohens-d/. (Visited on 06/14/2023).

[68] Tasnim I. Lat, Meghan K. McGraw, and Heath D. White. "Gender Differences in Critical Illness and Critical Care Research". In: Clinics in Chest Medicine 42.3 (Sept. 2021), pp. 543–555. ISSN: 0272-5231. DOI: 10.1016/j.ccm.2021.04.012. (Visited on 06/13/2023).

[69] DICOM Library - Anonymize, Share, View DICOM Files ONLINE. https://www.dicomlibrary.com/?m (Visited on 06/13/2023).