

*Juuso Ketola*

IMAGE RECONSTRUCTION  
AND MACHINE LEARNING  
APPROACHES FOR  
ENHANCED MEDICAL  
IMAGING

CASES IN COMPUTED TOMOGRAPHY AND  
MAGNETIC RESONANCE IMAGING

UNIVERSITY OF OULU GRADUATE SCHOOL;  
UNIVERSITY OF OULU,  
FACULTY OF MEDICINE





ACTA UNIVERSITATIS OULUENSIS  
D Medica 1648

*JUUSO KETOLA*

**IMAGE RECONSTRUCTION AND  
MACHINE LEARNING APPROACHES  
FOR ENHANCED MEDICAL IMAGING**

Cases in computed tomography and magnetic  
resonance imaging

Academic dissertation to be presented with the assent of  
the Doctoral Training Committee of Health and  
Biosciences of the University of Oulu for public defence in  
Auditorium P117 (Aapistie 5B), on 19 November 2021, at  
12 noon

UNIVERSITY OF OULU, OULU 2021

Copyright © 2021  
Acta Univ. Oul. D 1648, 2021

Supervised by  
Professor Miika Nieminen  
Professor Samuli Siltanen  
Professor Simo Saarakkala

Reviewed by  
Professor Elena Loli Piccolomini  
Professor Harrie Weinans

Opponent  
Professor Tanja Tarvainen

ISBN 978-952-62-3100-6 (Paperback)  
ISBN 978-952-62-3101-3 (PDF)

ISSN 0355-3221 (Printed)  
ISSN 1796-2234 (Online)

Cover Design  
Raimo Ahonen

PUNAMUSTA  
TAMPERE 2021

# **Ketola, Juuso, Image reconstruction and machine learning approaches for enhanced medical imaging. Cases in computed tomography and magnetic resonance imaging**

University of Oulu Graduate School; University of Oulu, Faculty of Medicine

*Acta Univ. Oul. D 1648, 2021*

University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

## ***Abstract***

Medical imaging encompasses many imaging modalities with different background physics and technological aspects, but they are united under the same premise: measuring data that can be expressed in image format. Furthermore, imaging protocols are designed to measure features that are archetypical for certain diseases, based on which a diagnosis can be made. Indeed, medical images are far more than just pictures, they contain a substantial amount of information. Some features are not visible to the naked eye, and computational means must be used to analyze them. The extraction and analysis of those features to serve as complementary information for diagnostic decision-making is called radiomics.

This doctoral dissertation focuses on applying algorithms in medical imaging use-tasks for an improved workflow. In the first use-case, iterative reconstruction (IR) algorithms were applied for micro-computed tomography ( $\mu$ CT) reconstruction using a limited amount of measured data. Doing so could facilitate the imaging workflow in  $\mu$ CT laboratories, as samples could be imaged more efficiently as less data are measured. The second use-case considered the interpretation of low back pain (LBP) from lumbar spine magnetic resonance imaging (MRI), which is a challenging task as subjects with degenerative features seen in MRI can be completely asymptomatic. To complement the MRI findings, a texture analysis (TA) pipeline employing radiomics methodology was developed for LBP classification. Finally, in the third use-case a deep learning (DL) -based image reconstruction method was developed for interior computed tomography (CT) angiography. CT is a major source of populational radiation strain, so reducing the dose by scanning only an internal region of the body has the potential to achieve dose-efficiency.

The results showed that a quantitative feature analysis in a  $\mu$ CT study could be performed with much less projection data than measured with a scanner traditionally. In addition, the developed TA method was able to differentiate subjects suffering from LBP from those who were asymptomatic, and an anatomical correspondence to pain manifestation was found. Finally, the developed DL reconstruction network was able to produce excellent image quality, and even increase the reconstructed field-of-view.

*Keywords:* artificial intelligence, computed tomography, deep learning, generative adversarial network, image reconstruction, low back pain, machine learning, magnetic resonance imaging, micro-computed tomography, osteoarthritis, radiation dose



# **Ketola, Juuso, Lääketieteellisen kuvantamisen tehostaminen kuvarekonstruktio- ja koneoppimismenetelmillä. Käyttötapauksia tietokonetomografiassa ja magneettikuvauksessa**

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta

*Acta Univ. Oul. D 1648, 2021*

Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

## ***Tiivistelmä***

Lääketieteellinen kuvantaminen käsittää useita kuvantamismodaliteetteja, joiden taustalla oleva fysiikka ja teknologiset yksityiskohdat poikkeavat merkittävästi. Kuvantamisen lähtökohta on kuitenkin kaikilla menetelmillä sama: Ne mittaavat dataa potilaasta ja esittävät mittauksen kuvan muodossa. Kuvausprotokollat suunnitellaan siten, että niissä näkyvät piirteet paljastavat tietyille sairauksille tyypillisiä ominaisuuksia, joiden perusteella voidaan päätyä diagnoosiin. Visuaalisen sisällön lisäksi lääketieteelliset kuvat sisältävät valtavan määrän informaatiota, jota kaikkea ei voi prosessoida paljain silmin. Tällöin avuksi tarvitaan laskennallisia menetelmiä, kuten kvantitatiivisia piirteitä analysoivaa radiomiikkaa.

Tämä väitöskirja käsittelee kuvarekonstruktio- ja koneoppimisalgoritmien soveltamista kolmeen kuvantamisen käyttötapaukseen. Ensimmäisessä käyttötapauksessa hyödynnettiin iteratiivisia kuvarekonstruktioalgoritmeja rekonstruoimaan harvasti näytteistettyä projektiodataa mikro-tietokonetomografiassa ( $\mu$ TT). Tämä voisi olla hyödyllistä  $\mu$ TT-kuvauksen optimoinnissa, koska kuvausaikojen ollessa pitkiä vain hyvin rajallinen määrä näytteitä on mahdollista kuvata prekliinisissä tutkimuksissa. Toinen käyttötapaus käsitteli lannerangan magneettikuvien (MRI) tulkintaa. Alaselkävivun yhteydessä MRI-kuvissa voi näkyä degeneratiivisia muutoksia, mutta nämä muutokset ovat tyypillisiä myös oireettomilla potilailla. Tutkimuksessa kehitettiin tekstuurianalyysimenetelmä (TA) hyödyntäen radiomiikkaa, jonka avulla pystyttiin luokittelemaan aineistossa olleet koehenkilöt kivuliaisiin ja oireettomiin. Viimeisessä käyttötapauksessa kehitettiin syväoppiva neuroverkko kardiologisen tietokonetomografian rekonstruktioon (TT) ns. sisätomografiaongelmassa, jossa katetaan vain potilaan sisäinen mielenkiintoalue röntgenkeilalla. TT:stä aiheutuu merkittävä osa populaation säteilyrasituksesta, joten annoksen vähentämiseen tähtäävillä menetelmillä on selkeä kansanterveydellinen hyöty.

Tulosten pohjalta selvisi, että vaihtamalla rekonstruktioalgoritmia kvantitatiivista  $\mu$ TT:tä voitaisiin tehdä pienemmällä määrällä projektiodataa kuin yleensä mitataan. Lisäksi kehitetty TA-menetelmä pystyi erottelamaan kivuliaat koehenkilöt oireettomista, ja luokittelun onnistumisen havaittiin riippuvan käytetystä anatomisesta mielenkiintoalueesta. Lopuksi, kehitetyllä syväoppivalla rekonstruktio menetelmällä saavutettiin selkeä kuvanlaadullinen parannus, ja rekonstruotua aluetta voitiin jopa kasvattaa.

*Asiasanat:* alaselkäkipu, generatiivinen kilpaileva verkosto, koneoppiminen, kuvarekonstruktio, magneettikuvaus, mikro-tietokonetomografia, nivelrikko, syväoppiminen, sädeannos, tekoäly, tietokonetomografia



*To my family*



## Acknowledgements

This doctoral research was carried out in the Research Unit of Medical Imaging, Physics and Technology (MIPT), Faculty of Medicine, at the University of Oulu during the years 2017-2021. It was funded by Business Finland, Jane & Aatos Erkko Foundation, the Technology Industries of Finland Centennial Foundation, and Tauno Tönning foundation.

I owe my gratitude to my principal supervisor, Professor Miika Nieminen, Ph.D., for his guidance and mentorship all the way from the stages of study design and endless funding applications to the writing of this thesis. I could always stop by his office for a chat if I had something on my mind. Though I think more often than not we ended those sessions overtime, discussing important aspects of guitar and bass playing gear.

I am of course also grateful to my secondary supervisors, Professor Samuli Siltanen, Ph.D., for providing his expertise in the field of inverse mathematics that is in the core of this dissertation, and Professor Simo Saarakkala, Ph.D., for his insight on the micro-computed tomography and machine learning parts of this study. It was a privilege to be under the supervision of some of the most renowned researchers in these fields.

Satu Inkinen, Ph.D., deserves a special thank you for her significant contribution to this research. With her, we spent countless hours studying the secrets of artificial intelligence and experimenting with different machine learning models and training strategies. And not to forget our little building projects that, in the end, worked out pretty well despite all the chaos that was involved.

I want to thank the official pre-examiners of this doctoral dissertation, Professor Elena Loli Piccolomini, Ph.D., and Professor Harrie Weinans, Ph.D., for providing extremely valuable feedback and criticism that improved this thesis.

I also want to thank my follow-up group, Adjunct Professor Timo Liimatainen, Ph.D., Adjunct Professor Kyösti Heimonen, Ph.D., and Sari Lasanen, Ph.D., for supervising the progress of my research and providing valuable feedback to my doctoral study plan.

I further thank Professor Matti Weckström, M.D., Ph.D., who is unfortunately no longer with us, for seeing potential in me in such an early stage of my biophysics studies. Working in his electrophysiology laboratory awakened the young scientist in me that ultimately led to the writing of this thesis.

My sincere thanks and appreciation also belong to my team, my co-authors, and my collaborators. Specifically, Mikael Juntunen, Ph.D., Tiina Ihme, B.Sc., Professor Jaro Karppinen, M.D., Ph.D., Adjunct Professor Mikko Finnilä, Ph.D., Sakari Karhula, Ph.D., Helinä Heino, B.Sc., and Management Assistant Meri Ojakangas all had a central

role in this research, and without them this dissertation would not exist. I also want to thank the rest of MIT for peer support and after-work discussions, it truly was a pleasure to work with such a gifted group of scientists.

I also want to thank Adjunct Professor Jari Heikkinen, Ph.D., and Jussi Aarnio, Ph.Lic., for giving me the opportunity to apply what I have learned about medical physics in a clinical environment, and for allowing me to finalize this thesis in Mikkeli.

Finally, my deepest thanks belong to my friends and family. I want to thank Topi, Jussi, Jali, and Emppu for all the unforgettable moments we've shared. Never stop the madness. I want to thank everybody who has been involved in my musical endeavours. I want to thank my parents Outi and Heikki, and my brother Tiitus, for always believing in me, and for always supporting me in everything I have done. Shine on you, crazy diamonds!

Most importantly, I want to thank my fiancée, my life-companion, and my best friend Jasmina for always being there for me. You lifted me up when I was on the brink of giving up, and for that I am eternally grateful. You are the strangest and loveliest person I've ever met. Never change.

Mikkeli, October 2021

Juuso Ketola

## List of abbreviations

2D	two-dimensional
3D	three-dimensional
ACLT	anterior cruciate ligament transection
Adam	adaptive moment estimation
ADMIRE	advanced modeled iterative reconstruction
AI	artificial intelligence
AIDR 3D	adaptive iterative dose reduction 3D
ALARA	as low as reasonably achievable
ASiR	adaptive statistical iterative reconstruction
AUC	area-under-curve
BN	batch normalization
BCE	binary cross entropy
BV/TV	bone volume fraction
CBCT	cone-beam computed tomography
CNN	convolutional neural network
CS	compressed sensing
CT	computed tomography
CTA	computed tomography angiography
DL	deep learning
FID	free induction decay
FSE	fast spin echo
FT	Fourier transform
ETL	echo train length
eV	electron volt
FBP	filtered backprojection
FOV	field-of-view
GAN	generative adversarial network
GLCM	gray-level co-occurrence matrix
Gy	Gray, J/kg
IMR	iterative model reconstruction
IR	iterative reconstruction
IRIS	iterative reconstruction in image space
IVD	intervertebral disc
kVp	peak kilovoltage

LBP	low back pain, local binary pattern
MBIR	model-based iterative reconstruction
$\mu$ CT	micro-computed tomography
ML	machine learning
MRI	magnetic resonance imaging
NFBC1966	Northern Finland Birth Cohort 1966
NMR	nuclear magnetic resonance
NPV	negative predictive value
OA	osteoarthritis
OP	osteoporosis
ODD	object-to-detector distance
PCA	principal component analysis
PSNR	peak signal-to-noise ratio
QBM	quantitative bone morphometry
ReLU	rectified linear unit
RF	radiofrequency
RLM	run-length matrix
RMSE	root mean squared error
ROC	receiver-operating characteristic
ROI	region-of-interest
SAFIRE	sinogram-affirmed iterative reconstruction
SPECT	single photon emission computed tomography
SOD	source-to-object distance
SNR	signal-to-noise ratio
SSIM	structural similarity index
Sv	Sievert, J/kg
$T_1$	longitudinal relaxation time constant
$T_2$	transverse relaxation time constant
TA	texture analysis
Tb.N	trabecular number
Tb.Th	trabecular thickness
Tb.Sp	trabecular separation
TE	time to echo
TR	repetition time
TV	total variation
VOI	volume-of-interest

## List of original publications

This dissertation is based on the following publications, which are referred to throughout the text by their Roman numerals (I-III):

- I Ketola JH, Karhula SS, Finnilä MAJ, Korhonen RK, Herzog W, Siltanen S, Nieminen MT & Saarakkala S (2018). Iterative and discrete reconstruction in the evaluation of the rabbit model of osteoarthritis. *Scientific Reports*, 8(1): 12051. DOI: 10.1038/s41598-018-30334-8
- II Ketola JHJ, Inkinen SI, Karppinen J, Niinimäki J, Tervonen O & Nieminen MT (2021). T<sub>2</sub>-weighted magnetic resonance imaging texture as predictor of low back pain: A texture analysis-based classification pipeline to symptomatic and asymptomatic cases. *J Orthop Res*. 2021; 1-11. DOI: 10.1002/jor.24973
- III Ketola JHJ, Heino H, Juntunen MAK, Siltanen S, Nieminen MT & Inkinen SI. Generative adversarial networks improve interior computed tomography reconstruction. *Manuscript submitted for publication*.

This dissertation also contains previously unpublished data.



# Contents

<b>Abstract</b>	
<b>Tiivistelmä</b>	
<b>Acknowledgements</b>	<b>9</b>
<b>List of abbreviations</b>	<b>11</b>
<b>List of original publications</b>	<b>13</b>
<b>Contents</b>	<b>15</b>
<b>1 Introduction</b>	<b>19</b>
<b>2 Machine learning in diagnostic imaging</b>	<b>21</b>
2.1 Radiomics . . . . .	21
2.1.1 Feature extraction and texture analysis . . . . .	22
2.2 Basics of machine learning . . . . .	24
2.2.1 Statistical learning . . . . .	25
2.2.2 Classification . . . . .	26
2.2.3 Model training . . . . .	27
2.2.4 Model evaluation . . . . .	28
2.3 Deep learning models . . . . .	28
2.3.1 Convolutional neural networks . . . . .	29
2.3.2 Convolutional encoder and decoder networks . . . . .	31
2.3.3 Generative adversarial networks . . . . .	32
<b>3 Computed tomography</b>	<b>33</b>
3.1 Principles of computed tomography . . . . .	33
3.1.1 Imaging geometry . . . . .	34
3.1.2 Computed tomography angiography . . . . .	35
3.2 Low-dose computed tomography . . . . .	35
3.2.1 Sparse-view computed tomography . . . . .	36
3.2.2 Interior computed tomography . . . . .	36
3.3 Computed tomography reconstruction . . . . .	37
3.3.1 Inverse problems . . . . .	37
3.3.2 Filtered backprojection . . . . .	38
3.3.3 Feldkamp, Davis and Kress algorithm . . . . .	39
3.3.4 Discretization of the measurement model . . . . .	40
3.3.5 Iterative reconstruction . . . . .	41
3.3.6 Compressed sensing . . . . .	43
3.3.7 Deep learning reconstruction . . . . .	44

3.3.8	Interior reconstruction	45
3.4	Micro-computed tomography	46
3.4.1	Quantitative bone morphometry	47
<b>4</b>	<b>Magnetic resonance imaging</b>	<b>49</b>
4.1	Basic physics of magnetic resonance imaging	49
4.1.1	Relaxation	50
4.2	Image formation	51
4.2.1	Gradient fields and encoding	51
4.2.2	Pulse sequences and signal weighting	51
4.2.3	Image reconstruction	53
4.3	Low back pain diagnostics with magnetic resonance imaging	53
4.3.1	Physiology of low back pain	53
4.3.2	Degenerative imaging findings	54
<b>5</b>	<b>Aims of the thesis</b>	<b>57</b>
<b>6</b>	<b>Materials and methods</b>	<b>59</b>
6.1	Data and imaging protocols	59
6.1.1	Rabbit samples and micro-computed tomography (I)	59
6.1.2	North Finland Birth Cohort 1966 lumbar MRI (II)	61
6.1.3	Computed tomography angiography data collection (III)	63
6.2	Reconstruction algorithms (I, III)	63
6.3	Quantitative bone morphometry (I)	66
6.3.1	Segmentation	66
6.3.2	Feature computation	66
6.3.3	Statistical analyses	67
6.4	Texture analysis and logistic regression classification (II)	67
6.5	Deep learning (II, III)	69
6.5.1	U-Net for segmentation (II)	69
6.5.2	Generator U-Net architecture (III)	69
6.5.3	PatchGAN discriminator (III)	69
6.5.4	Double generative adversarial network model (III)	71
6.6	Image quality metrics (III)	73
<b>7</b>	<b>Results</b>	<b>75</b>
7.1	Quantitative bone morphometry with sparse projection data (I)	75
7.1.1	Evaluation of the rabbit model of osteoarthritis with full data	75
7.1.2	Evaluation using iterative reconstruction methods	75
7.2	Machine learning -enhanced low-back pain diagnostics (II)	81
7.2.1	Classification	81

7.2.2	Sensitivity analysis to non-specific sources of pain . . . . .	83
7.3	Deep learning interior computed tomography angiography (III) . . . . .	85
7.3.1	Image quality analysis . . . . .	85
7.3.2	Extended field-of-view results . . . . .	86
<b>8</b>	<b>Discussion</b>	<b>89</b>
8.1	Key findings . . . . .	89
8.2	Sparse data micro-computed tomography (I) . . . . .	89
8.2.1	Rabbit model obtained with the reference reconstruction . . . . .	89
8.2.2	Iterative reconstruction algorithms . . . . .	89
8.2.3	Imaging time considerations . . . . .	90
8.3	Texture analysis classification of lumbar MRI (II) . . . . .	91
8.3.1	Added value of textural information . . . . .	91
8.3.2	Classification outcome . . . . .	91
8.3.3	Diagnostic interpretations . . . . .	92
8.3.4	Robustness to non-specific sources of pain . . . . .	93
8.3.5	Image reconstruction in the age of deep learning . . . . .	93
8.4	Deep learning interior computed tomography angiography (III) . . . . .	94
8.4.1	Image quality and the potential for extended field-of-view imaging . . . . .	94
8.4.2	Similar studies in the literature . . . . .	94
8.4.3	Choice of model architecture . . . . .	95
8.4.4	Challenges in training generative adversarial networks . . . . .	95
8.5	Added value for current imaging practices . . . . .	98
8.5.1	Improved efficiency in preclinical animal studies . . . . .	98
8.5.2	The potential to reduce radiation dose in cardiac imaging . . . . .	98
8.5.3	Promoting the use of medical artificial intelligence . . . . .	98
8.5.4	Projection to other imaging modalities . . . . .	99
8.6	Research outcomes . . . . .	100
8.7	Limitations . . . . .	100
8.8	Directions for future research . . . . .	101
<b>9</b>	<b>Conclusions</b>	<b>103</b>
	<b>References</b>	<b>105</b>
	<b>Original publications</b>	<b>119</b>



# 1 Introduction

"Images are more than pictures, they are data" (Gillies, Kinahan, & Hricak, 2016). Indeed, medical images contain a great number of features in them. Some features that are quantitative by nature are based on higher-order statistics and cannot be seen with the naked eye. Radiomics is the study of extracting such features, providing additional data and decision support for clinicians (Gillies et al., 2016; Kumar et al., 2012; Lambin et al., 2017). Artificial intelligence (AI), such as machine learning (ML) can further utilize these data in building predictive, prognostic, and decision-making models for diagnostic radiology (Erickson, Korfiatis, Akkus, & Kline, 2017; Wernick, Yang, Brankov, Yourganov, & Strother, 2010). Furthermore, deep learning (DL) methods that employ deep neural networks can be applied to extremely difficult ML problems. Of special interest are convolutional neural networks (CNNs) that can learn complex imaging features in filters that are used as building blocks for the networks. CNNs have been revolutionary in diagnostic imaging research, offering state-of-the-art methods for image segmentation, reconstruction, and computer-aided diagnostics (Erickson et al., 2017; Greenspan, Van Ginneken, & Summers, 2016; J.-G. Lee et al., 2017; Shen, Wu, & Suk, 2017; G. Wang, 2016).

Diagnostic imaging encompasses many imaging modalities, both methods that utilize ionizing radiation, and those that measure data by other means. While the physics and technology have fundamental differences between imaging modalities, the premise is that they measure data that are arranged into images. Structural and functional features corresponding to specific anatomy and diseases can then be interpreted by clinicians and computer programs. Imaging examinations are a vitally important part of diagnostics for many diseases.

Computed tomography (CT) is an imaging modality that produces three-dimensional volumes of the patient by measuring X-ray projections from many angles. As CT utilizes ionizing X-ray radiation, strategies for performing the imaging procedure dose-efficiently have been under heavy research since its invention. Modern CT scanners already utilize technology that involve a fraction of the dose of the original systems (McCollough et al., 2012). However, as the popularity of CT has increased significantly since the early 2000s, it is a major contributor to the medical radiation strain per capita (Bellolio et al., 2017; Bly, Järvinen, Kajaluoto, & Ruonala, 2020; Mettler Jr, 2019). A recent study in Finland shows that the contribution of CT to the collective dose in X-ray imaging was ~70% in 2018, and the estimated effective dose from X-ray imaging per capita had increased by ~60% within 10 years (Bly et al., 2020).

One of the major advances in low-dose CT has been the development of iterative image reconstruction (IR) techniques that are able to reduce the radiation dose to a fraction of the original while preserving sufficient image quality (Beister, Kolditz, & Kalender, 2012; Willeminck, de Jong, et al., 2013; Willeminck, Leiner, et al., 2013). Furthermore, in pre-clinical micro-CT ( $\mu$ CT) scanners, using IR would be advantageous as the imaging protocols are very long so a limited number of samples can be imaged. IR is based on measuring less (or lower quality) projection data and exploiting inversion mathematics to produce good reconstruction quality that would not be possible with traditional methods. With the advent of DL, CNN-based reconstruction algorithms have emerged providing additional dose-savings and speed (G. Wang, Ye, & De Man, 2020; G. Wang, Ye, Mueller, & Fessler, 2018).

In contrast to CT, magnetic resonance imaging (MRI) is an essential imaging modality due to its exceptional soft tissue contrast and the fact that ionizing radiation is not produced in the scan. One important application for MRI are musculoskeletal examinations, such as imaging of the lumbar spine of people who have pain in their lower back (Sheehan, 2010). A major problem of spinal MRI is the insensitivity of the imaging findings to the symptoms, as cohort studies have revealed that MRI abnormalities are in fact common in asymptomatic people as well (Boden et al., 1990; Boos et al., 2000; Jarvik, Hollingworth, Heagerty, Haynor, & Deyo, 2001). Radiomics and ML could be a source of complementary information to the clinician that could help to make the correct diagnosis based on imaging in the future.

The present doctoral dissertation is focused on applying iterative reconstruction techniques and ML methodology on use cases from different imaging modalities. In study I, the applicability of iterative reconstruction techniques was evaluated in a  $\mu$ CT study using sparse projection data. A rabbit model of osteoarthritis was quantified with traditional and iterative algorithms, and the correspondence to a baseline analysis was statistically evaluated. In study II, a ML pipeline was developed for the automated binary classification of  $T_2$ -weighted lumbar spine MRI for individuals experiencing low back pain and asymptomatic subjects. The algorithm's sensitivity to non-specific sources of pain was also investigated. Finally, in study III, a CNN-based image reconstruction method was developed for interior CTA. The possibility of extending the reconstructed field-of-view was also explored.

## 2 Machine learning in diagnostic imaging

Machine learning (ML) is an important branch of artificial intelligence (AI) that studies data-driven algorithms capable of improving their performance by using sample data. ML algorithms build prediction and decision models without explicitly programmed rules by adapting their function based on the data they process (Goodfellow, Bengio, Courville, & Bengio, 2016; Mohri, Rostamizadeh, & Talwalkar, 2018). Statistics and ML are very closely related, but while the premise of statistics is to infer properties from data distributions, ML aims at finding generalized patterns from data that can be modeled to make predictions (Murphy, 2012).

Diagnostic imaging offers a plethora of application areas for ML. With more and more data available every day, ML methods have become state-of-the-art in medical image analysis. With large amounts of data, patterns can be extracted to form complex prediction or decision models that can be used in image enhancement, reconstruction, and segmentation, as well as computed-aided detection, diagnosis, and decision-making (Erickson et al., 2017; Wernick et al., 2010).

More recently, deep learning (DL) methods have emerged with revolutionary results (Greenspan et al., 2016; J.-G. Lee et al., 2017; Shen et al., 2017). DL is a subclass of ML that utilizes deep neural networks to learn very complex and powerful models. DL algorithms are layered, with each layer typically learning more abstract features from the data. As such, DL methods are robust even for extremely difficult ML problems, such as image segmentation (Minaee et al., 2021), image-based cancer diagnosis (Hu et al., 2018), and low-dose computed tomography (CT) image reconstruction (G. Wang et al., 2020, 2018).

### 2.1 Radiomics

Radiomics refers to the large-scale extraction of imaging-based features to databases and subsequent data analysis (Gillies et al., 2016; Kumar et al., 2012; Lambin et al., 2017). Extracted features are used to build ML models for diagnostic, prognostic, and predictive tasks, serving as a complementary source of information and decision support for clinicians. In contrast to traditional medical imaging practice where images are viewed qualitatively as pictures, radiomics promotes quantitative imaging and data mining.

The radiomics workflow can be represented with the following sub-processes (Gillies et al., 2016):

1. Acquisition and reconstruction of high-quality (tomographic) digital medical images
2. Identification and segmentation of a region-of-interest (ROI) from which quantitative data are extracted
3. Extraction of high-dimensional statistical and textural feature data using methods such as texture analysis
4. Model implementation and validation by mining the data to find predictive patterns for selected use-cases.

### **2.1.1 Feature extraction and texture analysis**

The process of extracting features that describe the imaging data as well as possible is at the heart of radiomics. Radiomics features can be roughly classified into two categories. The first category are semantic features. Semantic features are features used in the 'radiology lexicon', such as various scoring systems that are used to describe various lesions (Gillies et al., 2016). For example, the Pfirrmann scoring system is used to assess the condition of intervertebral discs in the spine based on their appearance in magnetic resonance imaging (MRI) (Pfirrmann, Metzdorf, Zanetti, Hodler, & Boos, 2001). Semantic features are mostly qualitative by nature.

The second category are agnostic features. These are mathematically computed from the images and innately quantitative (Gillies et al., 2016). Agnostic features can be further categorized as first-, second- and higher-order statistics. Histogram-derived features are based on first-order statistics that describe data without addressing spatial relationships. In contrast, second- and higher-order statistics are referred to as textural features as they contain information about the spatial relationships between voxels.

The characterization of data based on their textural content is often called texture analysis (TA) (Mirmehdi, 2008). Most commonly used feature categories in TA approaches are histogram features, gradient features (histogram features obtained from the absolute gradient and gradient angle images), co-occurrence features, run-length features, autoregressive model parameters, and wavelets (Table 1) (Castellano, Bonilha, Li, & Cendes, 2004). Examples of TA applications in diagnostic imaging include tumor classification and cancer survival rate prediction in CT (Lubner, Smith, Sandrasegaran, Sahani, & Pickhardt, 2017), as well as brain tumor characterization, epilepsy seizure prediction, and multiple sclerosis disease applications in MRI (Kassner & Thornhill, 2010).

Haralick features from the gray-level co-occurrence matrix (GLCM) describe texture content by counting co-occurrences of grey levels in neighboring voxels (Haralick, Shanmugam, & Dinstein, 1973). The GLCM is computed from image  $I = I(x, y) \in \mathbf{R}^{M \times N}$

**Table 1. Texture analysis feature types.**

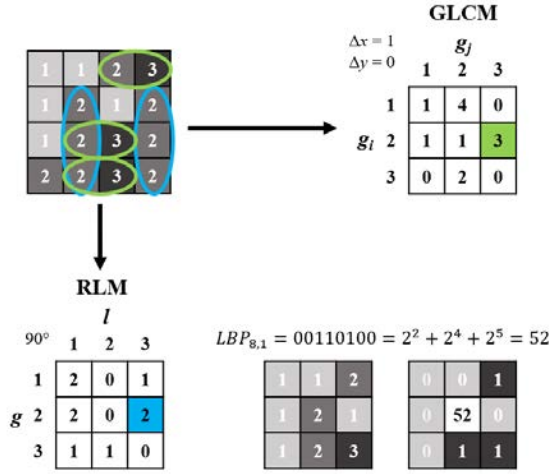
Feature type	Features
Grayscale histogram	Maximum, minimum, mean, variance, skewness, kurtosis, percentiles
Absolute gradient	# non-zero values, maximum, minimum, mean, variance, skewness, kurtosis, percentiles
Gradient angle	Maximum, minimum, mean, variance, skewness, kurtosis, percentiles
Co-occurrence (Haralick) features	Angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation, maximal correlation coefficient
Run-length matrix	Short run emphasis, long run emphasis, gray-level non-uniformity, run-length non-uniformity, run percentage
Autoregressive model	Parameters $\phi_i, \sigma$ from the model $x_j = \sigma + \sum_{i=1}^p \phi_i x_{j-i}$
Wavelet transform	Wavelet energy in low- and high-frequency sub-bands
Local binary patterns	Histogram bins, mean

by thresholding the image in  $G$  values and computing

$$C_{\Delta x, \Delta y}(g_i, g_j) = \sum_{x=1}^M \sum_{y=1}^N \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where  $g_i, g_j = \{0, \dots, G\}$  and  $\Delta x$  and  $\Delta y$  are offsets for which the spatial dependency is computed (Fig. 1). In other words, an element in the GLCM is the sum of voxel pairs that have gray-level values  $i$  and  $j$  with offsets  $\Delta x$  and  $\Delta y$ , respectively. First- and second-order statistics can then be computed from the GLCM (Table 1).

Another widely used TA method is the run-length encoding matrix (RLM) that counts homogeneous run lengths for each gray-level in a certain direction ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  in 2D) (Tang, 1998). An element in the RLM  $R(g, l)$  corresponds to number of



**Fig. 1. Computation of the gray-scale co-occurrence matrix (GLCM), the run-length matrix (RLM) and local binary patterns (LBP) in certain directions. Directions and radii can be changed to extract more features.**

homogeneous run-lengths of  $l$  voxels with gray-level  $g$  (Fig. 1). Statistics can be used to extract features from the RLM.

Higher-order textural features can be computed via mathematical transforms. Applying the wavelet transform to the image before feature extraction allows the analysis of different scales of textures due to the tree-like structure of the wavelet transform (Chang & Kuo, 1993). This allows TA to quantify features from a desired frequency range. Using local binary patterns (LBP) is another more advanced TA method employing binary encoding and weighting based on neighboring voxel gray-levels (Ahonen, Hadid, & Pietikainen, 2006; Guo, Zhang, & Zhang, 2010). The LBP of a voxel is defined by

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \quad s(g_p - g_c) = \begin{cases} 1, & g_p - g_c \geq 0, \\ 0, & g_p - g_c < 0, \end{cases} \quad (2)$$

where  $P$  is the number of neighboring voxels,  $R$  is the radius,  $g_p$  is the gray value of a neighboring voxel, and  $g_c$  is the gray level of the central pixel (Fig. 1).

## 2.2 Basics of machine learning

ML can be defined as the study of computer programs that use applied statistics to estimate complicated functions by learning from experience with respect to given tasks

and a performance measure (Goodfellow et al., 2016; Mohri et al., 2018). Most common machine learning tasks are classifications, in which the program assigns a category to an input, and regressions, in which the program predicts a numerical value for the input. The performance of an ML program with respect to the task can be assessed by quantitative measures such as accuracy, log-loss, and receiver operator characteristics (ROC). During the learning process, the program accumulates experience through a dataset. A dataset is a collection of samples described with a set of features, and can either have a label (supervised learning) or not (unsupervised learning). Learning is defined to occur when improvement in the performance measure for the task is achieved through experience (Goodfellow et al., 2016).

Mathematically, ML models are functionals  $f$  defined with a set of hyperparameters  $\theta$  that control the performance of the model. The dataset is contained in a matrix  $\mathbf{D}$  that contains samples  $\mathbf{x}_i \in \mathbb{R}^d$  and, in the case of supervised learning, the labels  $y_i \in \mathbb{R}$ . For example, in linear regression the ML model is simply given by

$$y_i = f(\theta, \mathbf{x}_i) = \theta^T \mathbf{x}_i, \quad (3)$$

where  $y_i$  is the predicted value,  $\theta = (\theta_0, \theta_1, \dots, \theta_{d-1})^T$  are the weights and bias of the regression model, and  $\mathbf{x}_i = (1, x_1, \dots, x_{d-1})^T$ . Practically, the goal of learning is to find a set of hyperparameters that maximize a loss functional computed over the dataset by means of numerical optimization (Goodfellow et al., 2016).

## 2.2.1 Statistical learning

### *Maximum likelihood estimation*

Statistical learning refers to using probability theory to find the optimal hyperparameters for a model. The likelihood function  $p(\mathbf{D}|\theta)$  measures how probable the observations in the dataset  $\mathbf{D}$  are given model hyperparameters  $\theta$  as a conditional probability. If samples  $(\mathbf{x}_i, y_i)$  in the dataset are assumed independent and identically distributed, the maximum likelihood can be written as

$$\operatorname{argmax}_{\theta} p(\mathbf{D}|\theta) = \operatorname{argmax}_{\theta} \prod_i p(y_i|\mathbf{x}_i, \theta). \quad (4)$$

In practice, it is more convenient to formulate the optimization problem differently. Because the logarithm is a monotonically increasing function, the log-likelihood function is usually optimized to remove local maxima that may be present in the likelihood function. Maximizing the log-likelihood is equivalent to minimizing the negative log-likelihood, which is preferred in ML as most programming libraries only implement

minimization algorithms (Murphy, 2012). Thus, the optimization problem in maximum likelihood estimation is often written as

$$\operatorname{argmin}_{\theta} \sum_i -\log p(y_i | \mathbf{x}_i, \theta). \quad (5)$$

### *Maximum a posteriori estimation*

Maximum a posteriori estimation is another popular estimation method, where Bayesian statistics are used to infer a posterior distribution  $p(\theta | \mathbf{D})$ . Bayes' theorem can be used to write the posterior in terms of the likelihood function, a prior distribution  $p(\theta)$  that contains beliefs about the hyperparameters beforehand, and the marginal data distribution  $p(\mathbf{D})$

$$p(\theta | \mathbf{D}) = \frac{p(\mathbf{D} | \theta) p(\theta)}{p(\mathbf{D})} \propto p(\mathbf{D} | \theta) p(\theta). \quad (6)$$

Consequently, the maximum a posteriori estimate is found by solving

$$\operatorname{argmin}_{\theta} \sum_i -[\log p(y_i | \mathbf{x}_i, \theta) + \log p(\theta)]. \quad (7)$$

The maximum a posteriori estimation can be seen as a maximum likelihood estimation that is regularized by a prior. In the special case of a constant prior, it reduces to a maximum likelihood estimation.

## **2.2.2 Classification**

Classification is defined as a task that identifies the category  $y_i \in \{0, \dots, C-1\}$  of a sample  $\mathbf{x}_i \in \mathbb{R}^d$  by finding a mapping  $f: \mathbb{R}^d \rightarrow \{0, \dots, C-1\}$  from a vector input to a categorical output such that  $y_i = f(\theta, \mathbf{x}_i)$  (Goodfellow et al., 2016).

Logistic regression is a common ML algorithm that works in the case of binary classification ( $C = 2$ ) (Murphy, 2012). The probability of one sample point belonging to either of the classes can be written as a Bernoulli probability mass function  $p(y_i | \mathbf{x}_i, \theta)$  and the likelihood function can be written as the product

$$L(\theta) = \prod_i p(y_i | \mathbf{x}_i, \theta) = \prod_i \sigma(\theta^T \mathbf{x}_i)^{y_i} \cdot [1 - \sigma(\theta^T \mathbf{x}_i)]^{1-y_i}, \quad (8)$$

where  $y_i \in \{0, 1\}$  and  $\sigma(z)$  is the sigmoid function

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (9)$$

The optimization problem in logistic regression is finding the negative log-likelihood

$$\operatorname{argmin}_{\theta} \sum_i \{-y_i \log \sigma(\theta^T \mathbf{x}_i) - (1 - y_i) \log [1 - \sigma(\theta^T \mathbf{x}_i)]\} = \operatorname{argmin}_{\theta} BCE(\theta), \quad (10)$$

where  $BCE(\theta)$  stands for binary cross entropy (also known as log loss).

In multi-class classification, each class has its own parameters  $\theta_c$  and classification is done in a one-versus-rest fashion (Goodfellow et al., 2016; Murphy, 2012). The multi-class model is described with the softmax function

$$p(y_i = c | \mathbf{x}_i, \theta) = \frac{e^{\theta_c^T \mathbf{x}_i}}{\sum_k e^{\theta_k^T \mathbf{x}_i}}. \quad (11)$$

Using maximum likelihood estimation, the optimization problem in multi-class classification is

$$\operatorname{argmin}_{\theta} - \sum_i [(\sum_k y_{ik} \theta_k^T \mathbf{x}_i) - \log(\sum_k e^{\theta_k^T \mathbf{x}_i})], \quad (12)$$

where  $y_{ik}$  are the one-hot encodings of labels  $y_i$  (Murphy, 2012).

### 2.2.3 Model training

ML models are trained by minimizing the loss function  $L(\theta)$  with numerical optimization methods such as gradient descent -based algorithms. Gradient descent is an iterative algorithm that updates  $\theta$  in the direction opposite to the gradient of the loss function  $g = \nabla_{\theta} L(\theta)$  with regard to model parameters via

$$\theta^{(k+1)} = \theta^{(k)} - \eta \nabla_{\theta} L(\theta^{(k)}) = \theta^{(k)} - \eta g^{(k)}, \quad (13)$$

where  $k$  is the iteration number and  $\eta$  is the learning rate. The goal of optimization in ML is to find the global minimum of the loss function in the parameter space.

Gradient descent is guaranteed to converge in convex optimization problems. Most ML problems are however not convex and require improvements to ensure convergence to the global minimum. Adaptive moment estimation (Adam) is a popular extension to the gradient descent algorithm that includes an exponentially decaying average of past gradients

$$m^{(k)} = \beta_1 m^{(k-1)} + (1 - \beta_1) g^{(k)}, \quad (14)$$

and past squared gradients

$$v^{(k)} = \beta_2 v^{(k-1)} + (1 - \beta_2) (g^{(k)})^2, \quad (15)$$

where  $\beta_1$  and  $\beta_2$  are decay rates (Kingma & Ba, 2014). These extensions help the algorithm to get past local minima and saddle points. Both  $m^{(k)}$  and  $v^{(k)}$  are initialized as zero vectors, and thus are biased towards zero, so additional bias-correction

$$\hat{m}^{(k)} = \frac{m^{(k)}}{1 - \beta_1^k} \quad (16)$$

$$\hat{v}^{(k)} = \frac{v^{(k)}}{1 - \beta_2^k} \quad (17)$$

is done to counteract this. An iteration of Adam is computed with

$$\theta^{(k+1)} = \theta^{(k)} - \frac{\eta}{\sqrt{\hat{v}^{(k)} + \epsilon}} \hat{m}^{(k)}. \quad (18)$$

### 2.2.4 Model evaluation

An important aspect of ML is that the models should be generalizable for unseen data. Overfitting refers to training the model too closely to a dataset so that it fails to make predictions for data that are not part of it. A validation dataset is used during the learning process to evaluate the model performance on data not used in training. Overfitting occurs if the loss of the validation set is considerably higher than the loss of the training set. Based on the model performance on the validation set, overfitting can be controlled by tuning model complexity and training hyperparameters such as the learning rate and regularization (Goodfellow et al., 2016; Murphy, 2012). Finally, a separate test set is used after training to evaluate the model performance on completely unseen data that have not been used in training or validation.

## 2.3 Deep learning models

Deep learning (DL) is a subcategory of ML that is based on artificial neural networks. An artificial neuron is a function that receives input signals, computes a weighted sum of those inputs, and passes the sum through a nonlinear activation function  $a(z)$  to produce output

$$o = a(\theta^T \mathbf{x}), \quad (19)$$

where  $\theta$  contains the weights of the neuron. In most artificial neural networks, the nonlinear activation is usually the rectified linear unit (ReLU)

$$a(z) = \max\{0, z\}, \quad (20)$$

the sigmoid  $\sigma(z)$ , or the hyperbolic tangent  $\tanh(z)$  (Goodfellow et al., 2016).

An artificial neural network is defined as an interconnected system of artificial neurons that form layers through which the information extracted from the dataset propagates. Subsequent layers extract higher-order features from the data. A deep neural network consists of an input layer, hidden layers that extract features from the input, and an output layer. Mathematically, a deep neural network can be modeled with the functional

$$f(\mathbf{x}, \Theta) = a^{(L)}(\theta^{(L)T} a^{(L-1)}(\dots \theta^{(l)T} a^{(l)}(\dots a^{(1)}(\theta^{(1)T} \mathbf{x}))), \quad (21)$$

where  $l$  is the layer number and  $\Theta = \{\theta^{(1)}, \dots, \theta^{(L)}\}$  contains all the parameters in the model.

### 2.3.1 Convolutional neural networks

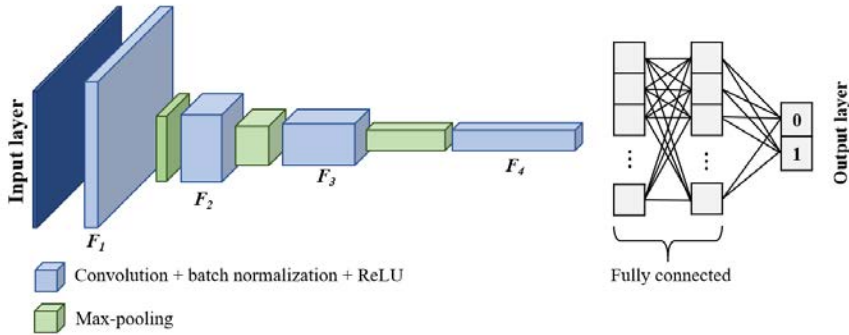
Convolutional neural networks (CNN) are a subclass of a deep neural networks in which the neurons are defined as convolution kernels. CNNs have been widely applied in diagnostic imaging in tasks such as segmentation (Minaee et al., 2021), image-based cancer diagnostics (Hu et al., 2018), and low-dose CT image reconstruction (G. Wang et al., 2020, 2018). Although the term convolution is used, most DL programming libraries use the cross-correlation operation

$$O(i, j) = (k * I)(i, j) = \sum_{m=-M}^M \sum_{n=-N}^N I(i+m, j+n)k(m, n), \quad (22)$$

to extract feature maps from images, where  $k \in \mathbb{R}^{(2M+1) \times (2N+1)}$  is the convolution kernel and  $I \in \mathbb{R}^{I \times J}$  is the input image (Goodfellow et al., 2016). Compared to traditional ML methods, CNNs have the inherent benefit of not having to compute a pre-defined set of features. Instead, learning the features is a part of the learning process (Erickson et al., 2017).

#### Convolutional layers

Most of the computation in CNNs happens in convolutional layers, that are essentially stacks of filters with learnable kernel weights (Fig. 2). Each filter is cross-correlated with the input volume, producing a feature map as the filter response. These feature maps are stacked producing an output volume that propagates to the next layer. If the input volume to a convolutional layer  $l$  is  $W^{(l-1)} \times H^{(l-1)} \times C^{(l-1)}$ , where  $W$  and  $H$  refer to the width and height of the feature maps and  $C$  refers to the number of filters, a convolution layer with  $C^{(l)}$  filters produces an output volume of  $W^{(l)} \times H^{(l)} \times C^{(l)}$ . The size of the output



**Fig. 2.** Convolutional neural networks typically include convolutional layers and max-pooling layers. Fully-connected layers can be used at the end of the network for *e.g.* classification purposes. In this example network, each consecutive convolutional layer  $F_i$  has twice the number of filters than the previous one, and the max-pooling layers reduce the spatial dimensions of the feature maps by a factor of two. Binary classification is performed in the output layer.

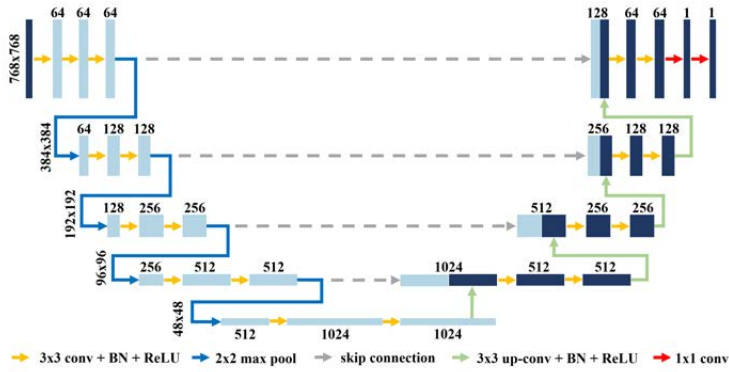
volume is controlled by the kernel size  $K$ , convolution stride  $S$ , and amount of zero-padding  $P$  with  $W^{(l)} = (W^{(l-1)} - K + 2P)/S + 1$  and  $H^{(l)} = (H^{(l-1)} - K + 2P)/S + 1$ . The tensor size of the convolutional layer is  $K \times K \times C^{(l-1)} \times C^{(l)}$ .

### *Pooling layers*

It is common to decrease the spatial size of the feature maps between successive convolutional layers to reduce the amount of learnable parameters (Fig. 2). This helps control the model complexity, computational resources, and overfitting. Usually,  $2 \times 2$  max-pooling is used which outputs the maximum of a  $2 \times 2$  volume in a reduced feature map. Consequently, an input size of  $W^{(l-1)} \times H^{(l-1)} \times C^{(l-1)}$  is reduced to  $W^{(l)} \times H^{(l)} \times C^{(l)}$ , where  $W^{(l)} = (W^{(l-1)} - K)/S + 1$ ,  $H^{(l)} = (H^{(l-1)} - K)/S + 1$ , and  $C^{(l)} = C^{(l-1)}$ .

### *Batch normalization layers*

The normalization of feature maps in-between convolutional layers makes training of CNNs faster and stabler (Goodfellow et al., 2016). Batch normalization (BN) layers normalize the data with the mean  $\mu_B$  and variance  $\sigma_B^2$  of a training mini-batch



**Fig. 3.** The U-Net architecture used for the segmentation of MR images in Study II. The MR image is given in the input layer, and the network computes a segmentation mask in the output layer. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit. (Figure reissued [modified] under CC BY 4.0 license from Publication II © 2021 Authors).

$B = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  and apply the affine transformation

$$\mathbf{o}_i = \gamma \frac{\mathbf{x}_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta, \quad (23)$$

where  $\gamma$  and  $\beta$  are learnable parameters and  $\varepsilon$  is a small constant applied for numerical stability.

### 2.3.2 Convolutional encoder and decoder networks

Encoder CNNs are networks that compute a feature representation of the image input. Usually encoder output is spatially compressed and contains a large number of feature channels. For example, in classification problems fully connected layers can be applied after the encoder to output a class prediction for the input image. Decoder CNNs work the opposite way — they take a feature encoding as input and reconstruct an image from it. Encoders and decoders can be combined to create encoder-decoder networks that can be used for image-to-image transformation tasks such as segmentation (Minaee et al., 2021). U-Net is an extremely popular network that has symmetrical encoding and decoding pathways with skip-connections in-between (Fig. 3) (Ronneberger, Fischer, & Brox, 2015). U-Net has become a standard in biomedical image segmentation (Minaee et al., 2021), but is also powerful for other image processing tasks such as CT image reconstruction and denoising as well (Jin, McCann, Froustey, & Unser, 2017; G. Wang et al., 2020, 2018).

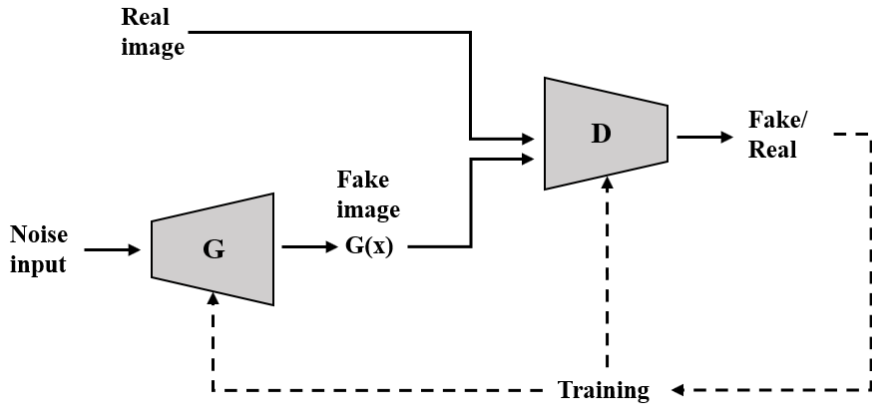


Fig. 4. Schematic of the vanilla GAN. The generator network ( $G$ ) generates a fake image from noise input, that the discriminator ( $D$ ) classifies as fake or real. The goal of training is to simultaneously make  $G$  produce good enough images to fool  $D$  while  $D$  is trained to identify the fake images as well as possible.

### 2.3.3 Generative adversarial networks

A generative adversarial network (GAN) is a specific type of DL framework that employs two CNNs that compete with each other in a zero-sum game (Creswell et al., 2018). A generator CNN,  $G$ , is trained to output images  $G(x)$  that a discriminator network,  $D$ , attempts to classify as fake or real (Fig. 4). These networks are trained simultaneously with two tasks in mind

1.  $G$  should be able to produce images that are so realistic that  $D$  would classify them as real images.
2.  $D$  should be able to distinguish between real and fake images as well as possible.

As such, improving one network has a direct adversarial effect on the other. In diagnostic imaging, GANs have been applied in many tasks, including classification, detection, segmentation, and reconstruction (Yi, Walia, & Babyn, 2019).

## 3 Computed tomography

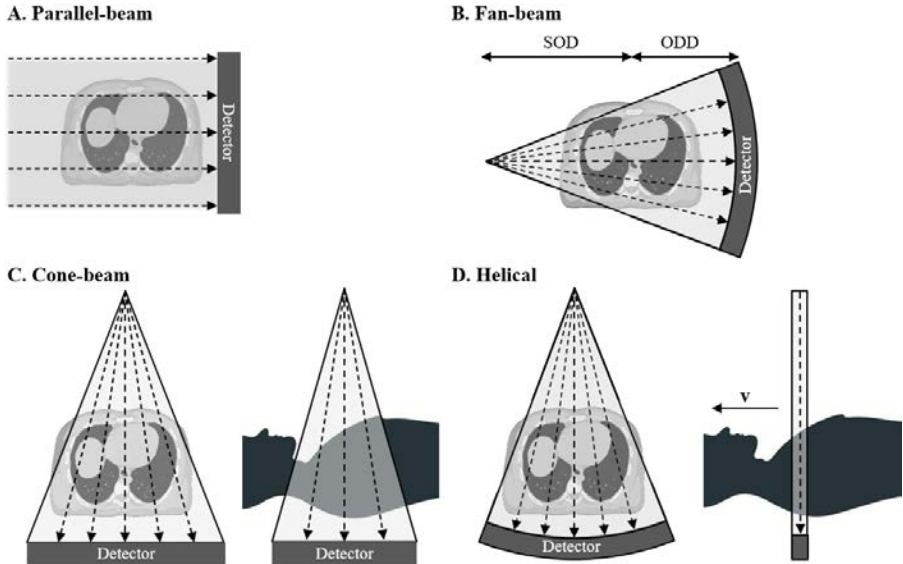
### 3.1 Principles of computed tomography

CT is an important diagnostic imaging method that combines X-ray attenuation data measured from hundreds or thousands of directions into cross-sectional images (slices) of the patient anatomy (Fig. 5). The basic working principle of CT is simple: An X-ray tube and a detector are located on the opposite sides of the patient, and the system rotates while collecting X-ray projections. After the measurement, the projection data are computationally processed to slice images in a mathematical process known as reconstruction. Projection data used for the reconstruction of a single slice image are called the sinogram.

Contrast in X-ray images is formed due to varying attenuation properties of electromagnetic radiation in different tissues. As X-rays pass through the patient, they can interact with the atoms of the tissues via photoelectric effect, elastic scattering, Compton scattering, and pair production. The most important interaction mechanisms are the photoelectric effect and Compton scattering. Elastic scattering is only relevant for very low photon energies (up to 20 keV) and pair production requires photon energies higher than 1.02 MeV (Bushberg, Seibert, Leinholdt, & Boone, 2012; Dowsett, Kenny, & Johnston, 2006).

In the photoelectric effect, the incident photon is absorbed by the atom, ejecting a photoelectron. The energy that remains after the binding energy of the electron is transferred to its kinetic energy. The photoelectron leaves a vacancy that is filled by an outer-shell electron, producing characteristic X-rays or ejecting Auger electrons (Dowsett et al., 2006). The photoelectric effect is the main source of contrast in X-ray images (Bushberg et al., 2012). This is because the probability of the photoelectric effect is higher in dense materials and for low X-ray energies, and because it does not involve scattering that would degrade image quality. Consequently, dense tissues undergo more attenuation than soft tissues, giving rise to the distinctive contrast between bone and soft tissue of X-ray methods. Furthermore, the higher the peak kilovoltage (kVp) of the tube selected, the lower the contrast becomes.

Compton scattering is another important X-ray interaction mechanism, where the incident X-ray interacts with an outer-shell electron, ejecting it from the atom. Consequently, the photon loses some of its energy and scatters changing direction. The contribution of the Compton effect to contrast is significantly lower than that of photoelectric effect. More importantly, the scattered photons hit the detector at



**Fig. 5. Illustration of different CT imaging geometries. The source-detector pair rotates around the patient in all cases. A. Parallel-beam geometry. B. Fan-beam geometry introduces geometrical magnification dependent on the source-to-object distance (SOD) and object-to-detector distance (ODD). C. Cone-beam geometry includes a flatpanel detector that captures divergent X-rays in 2D projections. D. Helical geometry employs fan-beam geometry while the table moves at velocity  $v$ , resulting in a helical sampling trajectory.**

incorrect locations, degrading image quality and increasing noise (Bushberg et al., 2012). Anti-scatter grids and software are often used to mitigate these effects.

### **3.1.1 Imaging geometry**

The imaging geometry of CT systems has evolved in the last decades. The first CT systems employed parallel-beam geometry, in which all the X-rays in a given projection angle are parallel to each other (Fig. 5A). While simple in principle, a source of parallel X-rays is difficult to produce (Kalender, 2011). A more practical approach is the fan-beam geometry, in which a point source emits a fan of X-rays (Fig. 5B). This results in a geometrical magnification dependent on the source-to-object distance (SOD) and object-to-detector distance (ODD). Parallel-beam and fan-beam approaches are 2D geometries capable of measuring one slice image at a time.

Modern scanners can perform 3D imaging with cone-beam CT (CBCT) and helical geometries. CBCT allows the beam to diverge like in projection radiography, measuring

2D projection images from angular increments with a flatpanel detector (Fig. 5C). CBCT results in an isotropic voxel size but is more prone to geometrical artefacts. In helical CT, fan-beam geometry is employed while the patient table simultaneously moves in the axial direction (Fig. 5D). The resulting projection data are interpolated in the axial direction to yield sinograms of the desired slice thickness that are reconstructed into slice images (Kalender, 2011). Multi-slice helical scanners include several detector arrays to capture multiple spiral trajectories simultaneously, allowing extremely fast scans.

### **3.1.2 Computed tomography angiography**

CT angiography (CTA) is an imaging technique used to visualize blood vessels in the body by injecting a highly attenuating contrast medium bolus into the bloodstream. Consequently, the otherwise poor contrast between soft tissue and blood is enhanced and the arteries and veins are visualized. CTA has many application areas in diagnostic CT, most importantly in cardiac, pulmonary, renal, and brain imaging.

## **3.2 Low-dose computed tomography**

Compared to planar or projection radiography, in which a single projection image is acquired, CT naturally yields a much greater radiation dose. Effective doses in most planar radiography examinations are in the range of decimals of mSv, but in some typical CT protocols it can be as high as 8 mSv (Kalender, 2011). While CT accounts for a relatively low percentage of all radiological examinations, it is a major contributor to the populational dose in Western countries (Kalender, 2011).

As the popularity and demand for X-ray examinations increases, the role of dose optimization in imaging protocols is emphasized. As low as reasonably achievable (ALARA) is a radiation protection principle adapted for medical imaging to address these considerations. ALARA means that ionizing radiation exposure from medical imaging should be minimized as much as practically and reasonably possible while not compromising image quality. Reductions in the CT dose can be achieved in many ways, for example by limiting the exposure time, number of projection views, and angular coverage, tightening beam collimation, or by using tube-current modulation, beam filtering, and advanced detector technologies.

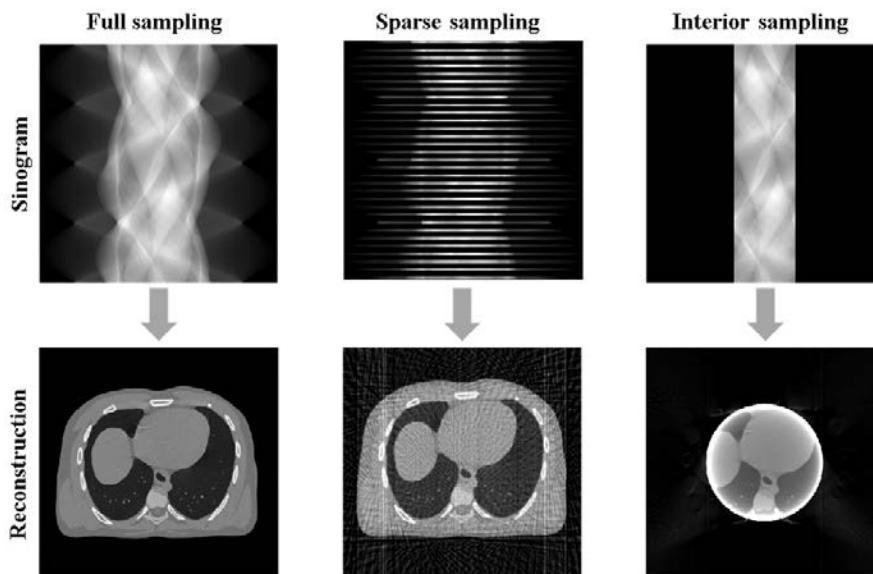


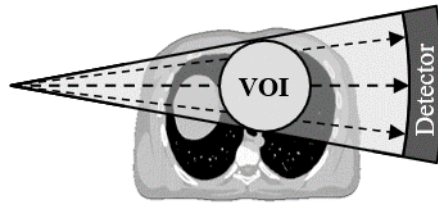
Fig. 6. Illustration of sparse and interior sampling and the resulting image artifacts. In sparse sampling, the angular sampling is decreased resulting in streaking and noise. In interior sampling, the field-of-views is limited by beam collimation, leading to truncation cupping.

### 3.2.1 Sparse-view computed tomography

Sparse-view CT refers to the collection of fewer projection views resulting in decreased exposure compared to complete sampling. As less data are collected, image artifacts such as streaks and noise are enhanced in the reconstruction (Fig. 6). Iterative reconstruction algorithms and compressive sensing have been successful at reducing these artifacts even with very high down-sampling factors (Bian et al., 2010; Liu, Ma, Fan, & Liang, 2012; Niu et al., 2014). Recently, the power of DL has been demonstrated in sparse CT. Several studies have shown that DL is capable of producing sparse-view reconstructions with an image quality that is comparable to full-dose reconstructions (Adler & Öktem, 2018; Chen et al., 2018; Han & Ye, 2018; Jin et al., 2017).

### 3.2.2 Interior computed tomography

Sometimes the layout of the imaging task does not require full exposure of the patient. For example in cardiac or dentomaxillofacial imaging, data from the organs surrounding the volume-of-interest (VOI) may be redundant. Interior CT refers to collimating the



**Fig. 7. Interior tomography refers to the collection of projection data from a limited volume-of-interest (VOI) by collimating the X-ray beam to expose less of the surrounding organs.**

X-ray beam to a smaller field-of-view (FOV) so that only the VOI is exposed, limiting the dose absorbed by the surrounding anatomy (Fig. 7). It has been estimated that the total dose could be reduced by 58% with interior CT (McCollough et al., 2012), and that the adjacent organ doses could be reduced by 60.5% in cardiac interior CT (Juntunen et al., 2020). However, this sampling leads to truncation artifacts with traditional reconstruction methods that manifest as a severe cupping profile in the image (Fig. 6). Many reconstruction algorithms have been developed to suppress this effect, but they include heavy restrictions (Clackdoyle, Noo, Guo, & Roberts, 2004; Faridani, Finch, Ritman, & Smith, 1997; Faridani, Ritman, & Smith, 1992; Kudo, Courdurier, Noo, & Defrise, 2008; G. Wang & Yu, 2013; Xu et al., 2011; Yu & Wang, 2009).

### **3.3 Computed tomography reconstruction**

#### **3.3.1 Inverse problems**

Image reconstruction in CT is an archetypical inverse problem. Inverse problems are counterparts of direct problems, where a model is used to calculate observations for a known object. In CT, indirect measurements of the patient (projections) are measured and then used to recover the attenuation properties of the patient. Thus, the mathematical problem of image reconstruction is equivalent to finding a solution to the inverse problem associated with it by computing an approximation of the patient with a reconstruction algorithm.

An inverse problem is considered well-posed if the Hadamard conditions apply:

1. A solution exists,
2. The solution is unique,

3. The solution depends continuously on the input.

In practice, all CT measurements have noise and other measurement instabilities, thus the inverse problem is ill-posed (Mueller & Siltanen, 2012). A large number of projection views have to be collected to yield a reconstruction of good image quality. If incomplete data are available, constraints must be applied to the reconstruction task with regularization techniques.

### 3.3.2 Filtered backprojection

#### Parallel-beam filtered backprojection

X-ray attenuation through an object can be approximately modelled with the monochromatic Beer-Lambert law

$$I = I_0 e^{-\int_{l(s,\beta)} \mu(x,y) ds}, \quad (24)$$

where  $I$  denotes the attenuated intensity,  $I_0$  denotes the original intensity,  $\mu(x,y)$  is a map of the linear attenuation coefficients representing the CT image, and

$$l(s,\beta) = \{(x,y) \in \mathbb{R}^2 : -x \sin \beta + y \cos \beta = s\} \quad (25)$$

is the line along which the integral operation is performed, where  $s$  is the orthogonal distance between the X-ray and center of rotation and  $\beta$  is the angle between the X-ray and the x-axis. Following linearization and reparametrization, the sinogram  $p(s,\beta)$  can be expressed with:

$$p(s,\beta) =: -\ln \frac{I}{I_0} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x,y) \delta(-x \sin \beta + y \cos \beta - s) dx dy, \quad (26)$$

where  $\delta$  is the Dirac's delta and the double integral is also known as the Radon transform (Radon, 1986).

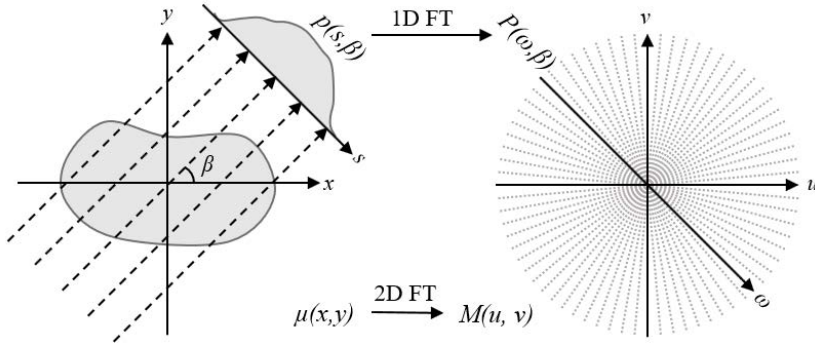
The Fourier central slice theorem proves that the Fourier transform (FT) of a projection line,  $P(\omega,\beta)$ , is equivalent to a central slice parallel to the projection line in the 2D FT of the image (Fig. 8) (Kalender, 2011):

$$P(\omega,\beta) = \int_{-\infty}^{\infty} p(s,\beta) e^{-2\pi i \omega s} ds \quad (27)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x,y) e^{-2\pi i \omega (-x \sin \beta + y \cos \beta)} dx dy. \quad (28)$$

Consequently, the image can be reconstructed with a parallel-beam filtered backprojection (FBP) algorithm (Kalender, 2011):

$$\mu(x,y) = \int_0^{\pi} \int_{-\infty}^{\infty} P(\omega,\beta) |\omega| e^{2\pi i \omega s} d\omega d\beta, \quad (29)$$



**Fig. 8. Schematic illustration of the central Fourier slice theorem. Taking the Fourier transform (FT) of a projection  $p(s, \beta)$  is equivalent to a central slice in the 2D FT of the image.**

where  $|\omega|$  is the (Ram-Lak) ramp filter with an impulse response  $h(s) = \int_{-\infty}^{\infty} |\omega| e^{2\pi i \omega s} d\omega$ . Practically, the FBP algorithm consists of two steps: (1) ramp-filtering the projections in the frequency space and computing the inverse FT, and (2) computing the inverse Radon transform of the filtered projections. Step (1) can also be done by convolving the sinogram with  $h(s)$ :

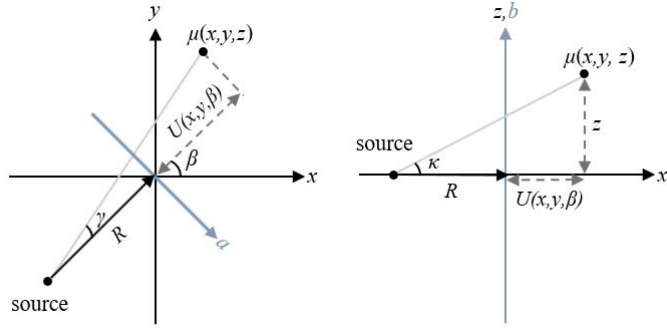
$$\mu(x, y) = \int_0^{\pi} h(s) * p(s, \beta) d\beta. \quad (30)$$

### 3.3.3 Feldkamp, Davis and Kress algorithm

The Feldkamp, Davis and Kress (FDK) algorithm is an approximative 3D reconstruction algorithm developed for CBCT (Feldkamp, Davis, & Kress, 1984). The cone beam geometry introduces the fan angle  $\gamma$  and the cone angle  $\kappa$ , whose geometrical effects are taken into account by pre-weighting the projections by  $\cos \gamma \cos \kappa = R / (\sqrt{R^2 + a^2 + b^2})$  during row-by-row ramp filtering

$$q(a, b, \beta) = h(a) * \left[ \frac{R}{\sqrt{R^2 + a^2 + b^2}} p(a, b, \beta) \right], \quad (31)$$

where  $a$  and  $b$  are the flatpanel detector pixel locations,  $\beta$  is the projection angle, and  $R$  is the SOD (Turbell, 2001). In addition, FDK assumes the detector is in the center of rotation, which leads to scaling the data with the distance between the source and the reconstruction point  $U(x, y, \beta) = R + x \cos \beta + y \sin \beta$  (Fig. 9) (Turbell, 2001).



**Fig. 9. Geometry of the FDK algorithm.** The cone beam diverges with fan angle  $\gamma$  and cone angle  $\kappa$ . FDK assumes the detector  $(a, b)$  is at the center of rotation, so the data are scaled by the distance  $U(x, y, \beta)$ .

Consequently, the FDK reconstruction formula is

$$\mu_{FDK}(x, y, z) = \int_0^{2\pi} \frac{R^2}{U(x, y, \beta)^2} q(a(x, y, \beta), b(x, y, \beta), \beta) d\beta, \quad (32)$$

where  $a(x, y, \beta) = \frac{R}{U(x, y, \beta)}(-x \sin \beta + y \cos \beta)$  and  $b = \frac{R}{U(x, y, \beta)}z$ . FDK gives an exact reconstruction in the  $z = 0$  plane, but the cone angle causes an increasing geometrical distortion along  $b$ , however for small cone angles this effect remains at acceptable levels (Kalender, 2011).

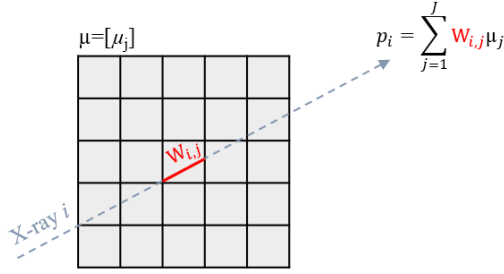
### 3.3.4 Discretization of the measurement model

In practice, the CT measurement model must be discretized for numerical computation. In the discrete model, sinogram data  $\mathbf{p} = [p_i] \in \mathbb{R}^I$  and image data  $\boldsymbol{\mu} = [\mu_j] \in \mathbb{R}^J$  are discrete, and sampled with  $I$  and  $J$  data points, respectively.  $I$  is the product of detector pixels and projection angles, *i.e.* the amount of measurements, and  $J$  is the amount of voxels in the image. The measurement geometry is characterized by computing the system matrix  $\mathbf{W} = [w_{i,j}] \in \mathbb{R}^{I \times J}$ , in which the weights  $w_{i,j}$  correspond to the ray-lengths of X-rays  $i$  in voxels  $\mu_j$  (Fig. 10). With these definitions, a projection measurement  $p_i$  can be represented by

$$p_i = \sum_{j=1}^J w_{i,j} \mu_j. \quad (33)$$

In matrix form, the measurement model becomes

$$\mathbf{p} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad (34)$$



**Fig. 10. Discretized CT measurement model.** The system matrix  $\mathbf{W}$  contains the imaging geometry given by the ray lengths  $W_{i,j}$ , and the projection measurements  $p_i$  can then be computed as the weighted sum of attenuation coefficients.

where  $\varepsilon \in \mathbb{R}^I$  is the measurement noise.

### 3.3.5 Iterative reconstruction

Iterative reconstruction (IR) is a popular branch of low-dose CT reconstruction algorithms. In contrast to analytical methods, IR employs an iterative optimization algorithm that updates the reconstruction estimate  $\mu^k$  with each step (Beister et al., 2012). The inverse problem in IR can be formulated as follows:

$$\underset{\mu}{\operatorname{argmin}} [\|\mathbf{W}\mu - \mathbf{p}\|_p^2 + \lambda R(\mu)] = \underset{\mu}{\operatorname{argmin}} [f(\mu)], \quad (35)$$

where  $\|\cdot\|_p$  is the  $L^p$ -norm,  $R(\mu)$  is the regularization term containing *a priori* information to constrain the problem, and  $\lambda > 0$  is the regularization parameter controlling the amount of regularization in the solution. Numerical optimization is used to find the best estimate of  $\mu$  iteratively, for example by employing gradient descent algorithms

$$\mu^{k+1} = \mu^k - \alpha \nabla f(\mu^k), \quad (36)$$

where  $k$  is the current iteration number,  $\alpha$  is the step size, and  $\nabla$  is the gradient operator. Iteration is continued until the residual  $\mu^{k+1} - \mu^k$  becomes small enough to fulfill the given converge conditions. For example in the case of  $L^2$ -norm, the recursive loop becomes

$$\mu^{k+1} = \mu^k - \alpha \left[ 2\mathbf{W}^T(\mathbf{W}\mu^k - \mathbf{p}) + \lambda \nabla R(\mu^k) \right], \quad (37)$$

where  $\mathbf{W}^T$  is the transpose of system matrix  $\mathbf{W}$  (Fig. 11).

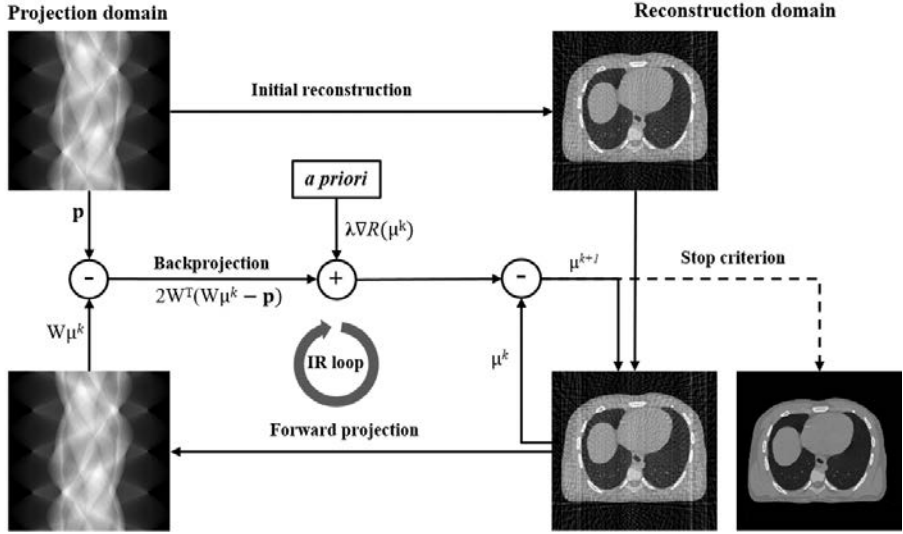


Fig. 11. Schematic chart of a common iterative reconstruction (IR) optimization loop. The optimization is iterated until a stop criterion is achieved.

The role of regularization is emphasized when the inverse problem becomes severely ill-conditioned. Using any constraints known *a priori* will help the optimization converge better or faster. For example, Tikhonov regularization employs the term

$$\Gamma(\mu) = \|\mathbf{I}\mu^k\|_2^2, \quad (38)$$

where  $\mathbf{I}$  is the identity matrix, to imply smoothness and continuity in the reconstructed image (Calvetti, Morigi, Reichel, & Sgallari, 2000; Mueller & Siltanen, 2012). Consequently, noise and artifacts in the reconstruction are suppressed but in turn the reconstruction is blurred.

### *Iterative reconstruction in clinical scanners*

Following the success of IR in low-dose imaging, all major CT vendors have developed their own IR algorithms that are available in their products. Benefits of IR in the clinical setting are twofold: IR can improve image quality with normal radiation dose, and IR can preserve image quality with reduced radiation dose (Beister et al., 2012; Willeminck, de Jong, et al., 2013; Willeminck, Leiner, et al., 2013). Reported dose reductions with IR range from 23% to 75% compared to FBP (Willeminck, de Jong, et al., 2013; Willeminck, Leiner, et al., 2013).

Three main categories of IR in clinical scanners have been identified (Willeminck, de Jong, et al., 2013):

1. Full IR methods alternate between forward and backprojection steps, comparing measured and modelled projection data in the projection space. Adaptive statistical iterative reconstruction (ASiR) (GE Healthcare, Waukesha, WI, USA) and model-based iterative reconstruction (MBIR) algorithms such as Veo (GE Healthcare, Waukesha, WI, USA), iterative model reconstruction (IMR) (Philips Healthcare, Best, the Netherlands), and advanced modeled iterative reconstruction (ADMIRE) (Siemens Healthcare, Forchheim, Germany) are examples of such algorithms. MBIR algorithms also model the acquisition physics such as the focal spot size, geometry, and photon statistics that are used as *a priori* information (Beister et al., 2012).
2. Image-domain methods, such as iterative reconstruction in image space (IRIS) (Siemens Healthcare, Forchheim, Germany), apply de-noising algorithms to images reconstructed by *i.e.* FBP, and are essentially post-processing techniques.
3. Some algorithms pre-process data in the projection domain and then post-process the image after reconstruction. Adaptive iterative dose reduction 3D (AIDR 3D) (Canon Medical Systems, Otawara, Japan), iDose<sup>4</sup> (Philips Healthcare, Best, the Netherlands), and sinogram-affirmed iterative reconstruction (SAFIRE) (Siemens Healthcare, Forchheim, Germany) are examples of such algorithms.

The texture in images produced by IR varies from images reconstructed with FBP. While numerical image quality may be higher with IR, clinicians often prefer the slightly granular structure of FBP over the paint-brushed impression given by IR. Therefore, IR is often mixed with FBP to a certain level that is optimized per protocol.

### **3.3.6 Compressed sensing**

Compressed sensing (CS) is a signal processing method developed for reconstructing signals acquired with very few samples by exploiting sparsity conditions. If the signal has a sparse (compressed) representation in some domain (many zero coefficients), the uncompressed signal can be recovered with optimization methods (Donoho, 2006). In medical imaging, CS theory has widely been used to solve severely ill-posed reconstruction problems in CT, MRI, PET, and SPECT (G. Wang, Bresler, & Ntziachristos, 2011).

Total variation (TV) -based methods are widely used in the field of CS (Rudin, Osher, & Fatemi, 1992). The TV of an image in the reconstruction domain  $\Omega$  is defined

by

$$TV(\mu) = \int_{\Omega} \|\nabla\mu\|_1 dx, \quad (39)$$

where  $\nabla$  is the gradient operator. As the TV is defined as the integral of an absolute gradient, sparsely sampled images have high total variation because of the noise and artifacts present in the images. Aiming to minimize the TV will suppress these unwanted details in the image while the edges are preserved, but in turn this will make the image appear blotchy. TV is among the most studied IR algorithms and has demonstrated good performance in several low-dose conditions (Ritschl, Bergner, Fleischmann, & Kachelrieß, 2011; Sidky & Pan, 2008; Tian, Jia, Yuan, Pan, & Jiang, 2011; Yu & Wang, 2009).

### **3.3.7 Deep learning reconstruction**

In recent years, DL-based image reconstruction methods have emerged. DL not only has the potential for substantial dose reductions by improving the quality of low-dose images, but also to speed up the long reconstruction times of iterative techniques because the CNN output is quick and cheap to compute after training (G. Wang, 2016; G. Wang et al., 2020, 2018; Willemink & Noël, 2019).

DL can be applied in many ways during image reconstruction (Table 2) (G. Wang et al., 2020). It can be used to pre-process an incomplete sinogram, for example by inpainting missing projection views in sparse CT (Dong, Vekhande, & Cao, 2019; Ghani & Karl, 2018), or it can post-process sparse reconstructions to yield an image quality comparable to full-dose images with U-Net (Jin et al., 2017). Regularization in iterative reconstruction can be formulated with CNNs, and instead of limiting the regularization to certain functionals it can be learnt directly from the data (Chen et al., 2018). Several iterative reconstruction algorithms have been ‘unrolled’ into a convolutional neural network (CNN) form, showing promising results (Adler & Öktem, 2018; Gupta, Jin, Nguyen, McCann, & Unser, 2018; Wu, Kim, & Li, 2019). Model-based DL reconstruction methods use CNN-based priors for CS reconstruction (Aggarwal, Mani, & Jacob, 2018). Direct mapping from projection data to reconstructed images is also possible with domain transform approaches (Y. Li, Li, Zhang, Montoya, & Chen, 2019; Ye, Buzzard, Ruby, & Bouman, 2018). Among the more advanced reconstruction networks are generative adversarial networks (GANs). These have been able to estimate high-dose images from low-dose images (Wolterink, Leiner, Viergever, & Išgum, 2017), and match image quality of low-dose phases to high-dose phases in multiphase CTA (Kang, Koo, Yang, Seo, & Ye, 2019).

**Table 2. Examples of parameters that DL can learn during different steps of the CT scan.**

Step	Learnable parameters
Data acquisition	Scan parameters and geometry
Pre-processing	Imaging physics and photon statistics
Reconstruction	Optimization mathematics and regularization
Post-processing	Artifacts and noise

### 3.3.8 Interior reconstruction

Interior reconstruction refers to reconstruction of interior tomography projection data. It is a challenging inverse problem, and theoretically does not have a unique solution. Because the sinogram is truncated, a prominent cupping artefact is present at the edges of the VOI. Several algorithms have been suggested to solve this problem, but most of them require prior information that is very restrictive (G. Wang & Yu, 2013; Yu & Wang, 2009). For example, the differentiated backprojection algorithm can solve the problem when a untruncated projection is available (Clackdoyle et al., 2004), or a small region-of-interest (ROI) is exactly known *a priori* (Kudo et al., 2008). In lambda tomography, it is possible to recover an image that has jumps (boundaries between different attenuation areas) in the right locations (Faridani et al., 1997, 1992). TV-based techniques have been able to find exact solutions when a small ROI is known (Taguchi et al., 2011), when the imaged ROI or VOI is piecewise constant (Cong, Yang, & Wang, 2012), or piecewise polynomial (Ward, Lee, Ye, & Unser, 2015; Yang, Yu, Jiang, & Wang, 2010). CS has also been applied to the problem in wavelet (Klann, Quinto, & Ramlau, 2015) and shearlet (Bubba, Labate, Zanghirati, & Bonettini, 2018) bases.

A different approach of interior reconstruction is to pre-process the sinogram to extend it, and then reconstruct the untruncated sinogram. Most sinogram extension techniques approximate the sinogram edges mathematically, for example with symmetric mirroring of the projection data (Ohnesorge, Flohr, Schwarz, Heiken, & Bae, 2000), convex hull estimation (Kolditz, Meyer, Kyriakou, & Kalender, 2010; Sourbelle, Kachelrieß, & Kalender, 2005), near-boundary value extrapolation (Kyrieleis, Titarenko, Ibison, Connolley, & Withers, 2011), water-cylinder fitting (Hsieh et al., 2004), and sinogram decomposition (Zamyatin & Nakanishi, 2007).

DL approaches for interior reconstruction include formulation of the differentiated backprojection algorithm with CNNs (Han & Ye, 2019), and computing untruncated prior images for reconstruction (Huang, Preuhs, Manhart, Lauritsch, & Maier, 2021).

### 3.4 Micro-computed tomography

Micro-computed tomography ( $\mu$ CT) is essentially CT performed in the micrometer range. It has a very broad range of applications, including but not limited to disease model phenotyping by *in vitro* and *in vivo* small animal and tissue sample imaging, evaluating pre-clinical study outcomes, developing drug and treatment interventions (Holdsworth & Thornton, 2002; Ritman, 2011; Schambach, Bag, Schilling, Groden, & Brockmann, 2010), *in situ* mechanical testing (Brault et al., 2013), geoscience studies (Cnudde et al., 2006), palaeontology and archeology (Abel et al., 2011; Suwa, Kono, Katoh, Asfaw, & Beyene, 2007).

$\mu$ CT imaging systems usually employ microfocus X-ray tubes, flatpanel detectors, cone-beam geometry, 'step-and-shoot' acquisition, and the FDK reconstruction algorithm (Holdsworth & Thornton, 2002; H. Li, Zhang, Tang, & Hu, 2008). Various scanner designs have been developed for different uses. *In vivo*  $\mu$ CT is usually performed with the rotating gantry design, where the tube-detector pair rotates around the small animal like in a scaled-down clinical scanner, allowing longitudinal, gated, and dynamic imaging as well as angiography (Badea, Drangova, Holdsworth, & Johnson, 2008). *In vitro* studies are performed with a scanner in which the specimen rotates while the tube-detector pair remains stationary (Holdsworth & Thornton, 2002). Furthermore, extremely high resolution (in the nanometer range) can be achieved with synchrotron  $\mu$ CT systems employing nearly monochromatic high-brilliance X-ray beams produced from a synchrotron facility beamline (Bonse & Busch, 1996; Metscher, 2009).

The physical properties of microfocus X-ray tubes and geometrical magnification from cone-beam geometry combined with long exposure times and dense angular sampling of projections are all factors that contribute to the high resolution of  $\mu$ CT. Consequently, the imaging times for  $\mu$ CT can range from a few minutes for *in vivo* scans to several hours for *in vitro* scans. The long scan times can result in severe movement-induced artifacts if the animal or sample moves and the reconstruction algorithm does not take the motion into account (Boas & Fleischmann, 2012; Schulze et al., 2011). Furthermore, the radiation dose in  $\mu$ CT is relatively high. Organ doses between 0.28 to 0.76 Gy and surface doses up to 5.5 Gy have been measured in mice, while LD50/30 values (a dose that would be lethal to 50% of population within 30 days of exposure) range from 5.0 to 7.6 Gy and the rate of life shortening is estimated to be 7.2%/Gy (Paulus, Gleason, Kennel, Hunsicker, & Johnson, 2000; Schambach et al., 2009; Willekens et al., 2010). Adverse radiation-induced effects are naturally of ethical concern in small animal imaging, but may also alter tissue microstructure affecting

experiment outcomes in longitudinal studies (Laperre et al., 2011; Willekens et al., 2010).

### **3.4.1 Quantitative bone morphometry**

$\mu$ CT can offer extremely high (one to a few microns) spatial resolution and excellent contrast between bone and soft tissue. Thus, it can be used to extract quantitative data on the trabecular bone micro-structure, making it a widely-used research tool for musculoskeletal diseases such as osteoporosis and osteoarthritis (OA) (Boerckel, Mason, McDermott, & Alsberg, 2014). Osteoporosis is a systemic skeletal disease characterized by low bone mass and changes in the bone micro-architecture, leading to bone fragility and increased risk of fracture (Rachner, Khosla, & Hofbauer, 2011). OA is a progressive joint disease, where cartilage degeneration and changes in subchondral bone micro-architecture lead to pain, stiffness, and disability in the synovial joints (Buckwalter & Mankin, 1998).

Both clinical and preclinical studies have been conducted to understand the pathophysiology of these diseases and develop new treatment methods. As human clinical studies encompass several limitations, most importantly variations in symptoms and difficulties in obtaining bone and cartilage samples, several animal models have been developed. The most important animal models used for OA include natural degeneration and post-traumatic OA induced by surgical injury such as anterior cruciate ligament transection (ACLT) in rodents, dogs, goats, and horses (Kuyinu, Narayanan, Nair, & Laurencin, 2016; Lampropoulou-Adamidou et al., 2014). To simulate osteoporotic conditions in animals, an ovariectomy is commonly performed on rats to induce postmenopausal trabecular bone loss (Egermann, Goldhahn, & Schneider, 2005).

Subchondral bone morphometry is a standard method to extract quantitative features from  $\mu$ CT data for further analyses such as statistical comparison between diseased and control groups. After a segmentation of bone and soft tissue volumes has been obtained, morphometric variables describing the bone micro-structure are calculated in image processing software (Fig. 12). The standard set of variables includes the bone volume fraction (BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), and trabecular separation (Tb.S) (Bouxsein et al., 2010). Several studies have found significant alterations in these parameters in diseased bone (Bobinac et al., 2013; Finnilä et al., 2017; Florea et al., 2015; Sniekers et al., 2008; Z.-M. Zhang, Li, Jiang, Jiang, & Dai, 2010).

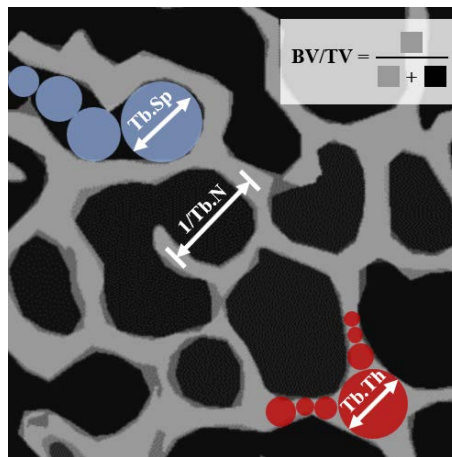


Fig. 12. Computation of common quantitative bone morphometry features. BV/TV is computed by dividing the bone volume by the entire tissue volume. Tb.Sp and Tb.Th are computed by fitting spheres between and inside trabeculae, and Tb.N is computed as the inverse of the distance between trabeculae.

## 4 Magnetic resonance imaging

### 4.1 Basic physics of magnetic resonance imaging

Magnetic resonance imaging (MRI) is an important diagnostic imaging method, that employs strong magnetic fields and radiofrequency (RF) waves to image patient anatomy and physiology. As such, MRI does not expose the patients to ionizing radiation like X-ray based methods.

Nuclei with non-zero intrinsic angular momentum, spin  $m$ , are able to absorb and emit RF energy in external magnetic fields in a process called nuclear magnetic resonance (NMR). MRI is commonly perceived as the medical application of hydrogen ( $^1\text{H}$ ) NMR, although MRI methods for other nuclei such as  $^3\text{He}$ ,  $^{13}\text{C}$ ,  $^{23}\text{Na}$ ,  $^{31}\text{P}$ , and  $^{129}\text{Xe}$  exist as well (Brown et al., 2014; McRobbie, Moore, Graves, & Prince, 2017). Hydrogen nuclei that only contain a single proton have two linearly independent spin states ( $m = \frac{1}{2}$  or  $m = -\frac{1}{2}$ ) that have the same energy in the absence of a magnetic field. When placed in a static external field ( $\mathbf{B}_0$ -field), the energies of these states become

$$E_m = -\gamma m \hbar B_0, \quad (40)$$

where  $\gamma$  is the gyromagnetic ratio,  $\hbar = 6.626 \times 10^{-34}$  Js is the reduced Planck constant, and  $B_0$  is the magnitude of the static external field. As the  $m = \frac{1}{2}$  state is less energetic, in thermal equilibrium there is a slight population bias on this state. Most nuclei do not however exist exclusively on either of these states, but in a weighted superposition (Brown et al., 2014; McRobbie et al., 2017). Yet on average, this population discrepancy causes a net magnetization  $\mathbf{M}$  along  $\mathbf{B}_0$  (z-axis) which is the sum of the magnetic properties of the spin system. In addition, because the nuclei have a nonzero spin, they precess around the local magnetic field with angular frequency

$$\omega_0 = -\gamma \mathbf{B}_0, \quad (41)$$

that is called the Larmor frequency.

If the magnetization is exposed to another field produced with a RF coil,  $\mathbf{B}_1$ , that is perpendicular to  $\mathbf{B}_0$  and precessing at  $\|\omega_1\| \approx \|\omega_0\|$ , the magnetization flips from its initial alignment until the  $\mathbf{B}_1$ -field is switched off. These excitation RF pulses can be applied for arbitrarily long times depending on the desired flip angle. After excitation, the magnetization returns back along  $\mathbf{B}_0$  while producing an oscillating magnetic field that induces small but measurable signals in the receiver coils.

### 4.1.1 Relaxation

After flipping the magnetization, recovery of magnetization along  $\mathbf{B}_0$  follows via energy transfer from the spin system to its molecular environment in a process called  $T_1$  relaxation (also called longitudinal and spin-lattice relaxation). The energy gained from the RF pulse is dissipated in vibration and rotation motion in the molecular structure. This process follows exponential growth

$$M_z(t) = M_z^0(1 - e^{-\frac{t}{T_1}}), \quad (42)$$

where  $M_z$  is the amount of magnetization along  $\mathbf{B}_0$  (z-axis),  $M_z^0$  is the initial maximum value of  $M_z$  before excitation, and  $T_1$  is a first-order time constant.

Following excitation, the magnetization also has a transverse component  $M_{xy}$  that is perpendicular to  $\mathbf{B}_0$ . After the flip, a small number of spins are in transverse phase coherence initially. This coherence is lost in a process called  $T_2$  relaxation, that occurs due to two mechanisms (Brown et al., 2014; McRobbie et al., 2017). First, the energy exchange in  $T_1$  relaxation causes spins contributing to  $M_{xy}$  to realign. Thus,  $T_1$  relaxation always results in  $T_2$  relaxation as well. Second, changes in the chemical environment close to individual spins causes small fluctuations in the  $B_0$  field. Therefore spins in different environments are subject to varying static field strengths and precess with slightly different frequencies. These processes cause  $M_{xy}$  to decay exponentially with

$$M_{xy}(t) = M_{xy}^0 e^{-\frac{t}{T_2}}, \quad (43)$$

where  $M_{xy}^0$  is the amount of initial transverse magnetization after the flip, and  $T_2$  is a first-order time constant.

$T_2$  relaxation is defined as the natural relaxation due to atomic and molecular interaction.  $\mathbf{B}_0$  fields are never completely homogeneous due to intrinsic magnetic nonidealities and other field distortions, causing additional magnetization decay with the time constant  $T_2'$  (Brown et al., 2014; McRobbie et al., 2017). The observed signal decays with the time constant  $T_2^*$  that includes contributions from both relaxation mechanisms and follows the equation

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} \quad (44)$$

Because the transverse magnetization precesses around  $\mathbf{B}_0$ , it will produce an oscillating magnetic field that induces oscillating currents in nearby receiver coils according to the Faraday-Lenz law. The amplitude of this signal decays with  $T_2^*$  relaxation. The resulting free induction decay (FID) signal has a shape of a damped sine wave.

## 4.2 Image formation

### 4.2.1 Gradient fields and encoding

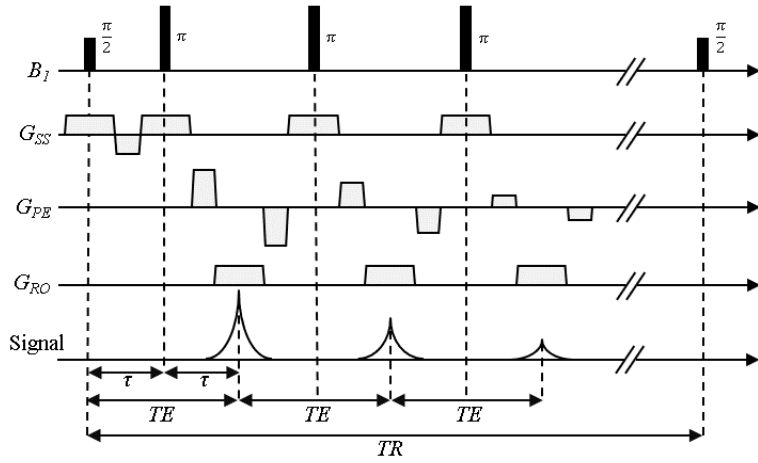
Multidimensional data in MRI are acquired by employing gradient fields. A slice selection gradient  $G_{SS}$  is applied along  $\mathbf{B}_0$  to create a spatial dependency on the Larmor frequency. Selective RF pulses can then be used to excite spins in the desired position of the z-axis. The slice thickness of the excited space depends on the RF pulse bandwidth and the strength of the  $G_{SS}$  (Brown et al., 2014; McRobbie et al., 2017). The slice selection gradient accumulates a phase in the spin system, so often a negative refocusing gradient lobe is used to correct for this effect. After excitation, a phase-encoding gradient  $G_{PE}$  can be used along an orthogonal dimension. Usually the y-direction is phase encoded in conventional sequences (Brown et al., 2014; McRobbie et al., 2017). Applying  $G_{PE}$  for a set time accumulates a phase that is dependent on the location along y-axis. After phase encoding, a readout gradient  $G_{RO}$  can be used in the x-direction, which frequency encodes the spins. These spatial encoding steps result in an excited slice volume whose spins precess with varying Larmor frequencies in the readout direction, and have a varying phase in the phase encoding direction. The FID signal is collected during the application of  $G_{RO}$ .

### 4.2.2 Pulse sequences and signal weighting

An MRI pulse sequence is a programmed chain of gradient changes and RF pulses over time. Most sequences start with an excitation pulse flipping the magnetization along transverse axes. Excitation is followed by a sequence of gradient applications and RF pulses in different directions and these contribute to signal manipulation and sampling. The sequence is repeated after repetition time  $TR$  after all of the longitudinal magnetization has recovered.

One of the most popular MRI sequences is the spin echo sequence (Brown et al., 2014; McRobbie et al., 2017). Additional spin dephasing due to  $T_2'$  relaxation results in accelerated decay of signal. However, this dephasing is reversible by applying a  $\pi$ -pulse about the x or y-axis at time  $t = \tau$ . This pulse inverts the phase of individual spins, and they begin to accumulate an additional phase but in the opposite direction. As a result, a spin echo is observed at time-to-echo  $TE = 2\tau$ , at which time the signal amplitude is related to  $T_2$ -relaxation.

Multiple echoes can be collected in one excitation in fast spin echo (FSE) sequences (also known as turbo spin echo sequences). After the first refocusing pulse, subsequent ones are applied every  $TE$  (Fig. 13). The number of echoes collected per  $TR$  is called



**Fig. 13. Pulse diagram of a fast spin echo sequence. A refocusing slice selection gradient lobe is used after excitation, and the slice selection gradient lobes during refocusing pulses are self-refocusing. Negative phase-encoding lobes are used to rewind the phase after signal readout to allow incremental sampling along  $k_y$  in this sequence.**

the echo train length (ETL). FSE can significantly reduce imaging times, however as the echo amplitude decreases with subsequent echoes, the signal will have a lower signal-to-noise ratio (SNR) (Brown et al., 2014; McRobbie et al., 2017).

In spin echo sequences,  $TR$  and  $TE$  can be used to control the contrast in the resulting images. The MRI signal  $S$  is approximately related to  $TR$ ,  $TE$ ,  $T_1$ , and  $T_2$  with

$$S \propto \rho(1 - e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}}, \quad (45)$$

where  $\rho$  is the proton density. When  $TE$  is short (compared to  $T_2$ ),  $T_2$  relaxation effects will not have had time to take place before the echo, while sequences with longer  $TE$ s exhibit differences in  $T_2$  relaxation effects in different tissues. Sequences that highlight  $T_2$  effects are called  $T_2$ -weighted. Similar interaction can be seen with  $TR$  and  $T_1$ . If  $TR$  is long (multiple of  $T_1$ ), most of the longitudinal magnetization has recovered, essentially removing  $T_1$  effects from the signal. Conversely,  $T_1$ -weighted signals have a short  $TR$ . If both the  $TE$  is short and the  $TR$  is long, both relaxation effects disappear and the signal is weighted by the proton density.

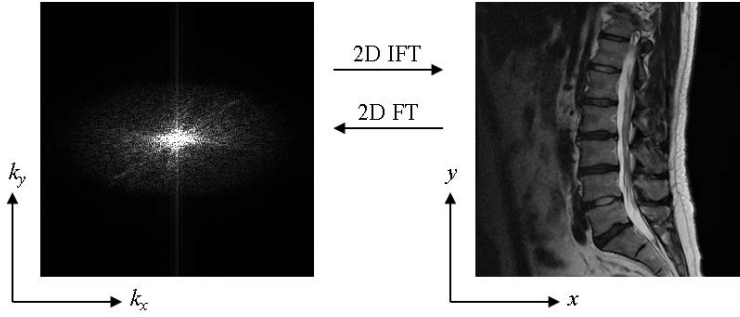


Fig. 14. Reconstruction of k-space data is done by applying the discrete inverse Fourier transform (IFT).

### 4.2.3 Image reconstruction

The recording signal from a spin system will result in a 1D FID signal sampled with  $N = 2n$  samples. Collection of the MRI signals is done in k-space which is a spatial frequency domain with complex coordinates  $k_x$  and  $k_y$ . Recording a 1D MRI signal as described above refers to collecting a line in the k-space along  $k_x$  (Brown et al., 2014; McRobbie et al., 2017). By varying the phase-encoding gradient strength, lines in varying  $k_y$  positions are obtained. By repeating the sampling process  $M = 2m$  times, a  $N \times M$  k-space is obtained. Many other k-space sampling strategies, such as radial and projection sampling exist as well. Mathematically, the k-space is the 2D Fourier transform of the MRI slice. Therefore, the MRI slice  $f(x, y)$  can be reconstructed simply by applying the discrete inverse 2D Fourier transform (Brown et al., 2014; McRobbie et al., 2017) (Fig. 14)

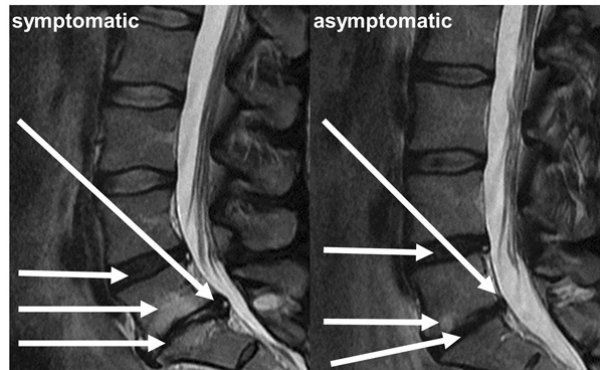
$$f(x, y) = \frac{1}{MN} \sum_{k_x=-m}^{m-1} \sum_{k_y=-n}^{n-1} F(k_x, k_y) e^{i\pi(\frac{k_x x}{m} + \frac{k_y y}{n})}, \quad (46)$$

where  $k_x$  and  $k_y$  are the coordinates of the k-space.

## 4.3 Low back pain diagnostics with magnetic resonance imaging

### 4.3.1 Physiology of low back pain

Low back pain (LBP) is a leading cause of disability and sick leave as well as a major economic burden worldwide (Dagenais, Caro, & Haldeman, 2008; Hartvigsen et al., 2018). LBP is a complex condition with anatomical, physiological, psychological and social aspects that influence the experience of pain (Hartvigsen et al., 2018). Most



**Fig. 15. Degenerative features seen in lumbar MRI in both symptomatic and asymptomatic subjects. Arrows correspond to (from top to bottom): IVD herniation, IVD degeneration, vertebral endplate changes, and collapsed IVD. (Figure reissued under CC BY 4.0 licence from Publication II © 2021 Authors).**

commonly LBP is the result of an injury or degeneration of lumbar innervated tissues, such as facet joints, intervertebral discs (IVDs), ligaments, or muscles. Annular tearing (Videman et al., 2003; Videman & Nurminen, 2004), IVD height narrowing (Videman et al., 2003), facet joint degeneration (Boswell, Singh, Staats, & Hirsch, 2003), and vertebral endplate lesions (Y. Wang, Videman, & Battié, 2012), are conditions that have been related to LBP. Certain diseases also cause LBP, including disc herniations and ruptures, sciatica, arthritis, infections, and spinal cancer (R. Chou, Fu, Carrino, & Deyo, 2009). LBP that cannot be explained with any known pathology is called non-specific LBP (Balagué, Mannion, Pellisé, & Cedraschi, 2012).

### **4.3.2 Degenerative imaging findings**

MRI allows good visualization of the spinal anatomy. Therefore, MRI is the preferred modality for spinal imaging (Sheehan, 2010), and is often used to confirm suspicion of IVD herniation, nerve root entrapment, spinal stenosis, trauma, or tumors (R. Chou et al., 2009). Degenerative changes such as vertebral endplate changes and IVD degeneration are also visualized in MRI (Carragee, 2005; Toyone et al., 1994). However, most of these MRI abnormalities can be found in completely asymptomatic people as well (Fig. 15). Up to 87% of asymptomatic people can have degenerative findings or lesions in their lumbar MRI (Boden et al., 1990; Boos et al., 2000; Jarvik et al., 2001). Therefore, degeneration seen in MRI can not be attributed to LBP symptoms alone (D. Chou et al., 2011; Jensen et al., 1994).

**Table 3. Pfirrmann grading system for IVD degeneration seen in T<sub>2</sub>-weighted MRI.**

Grade	Structure	Nucleus	Signal intensity	Height of IVD
I	Homogeneous, bright	Clear	Hyperintense, isointense to CSF	Normal
II	Inhomogeneous or banded	Clear	Hyperintense, isointense to CSF	Normal
III	Inhomogeneous, grey	Unclear	Intermediate	Normal to slightly decreased
IV	Inhomogeneous, grey or black	Lost	Intermediate to hypointense	Normal to moderately decreased
V	Inhomogeneous, black	Lost	Hypointense	Collapsed

**Table 4. Definition of Modic changes characterized by signal intensity in T<sub>1</sub> and T<sub>2</sub> -weighted sequences and histopathology.**

Type	T <sub>1</sub> -weighted signal	T <sub>2</sub> -weighted signal	Histopathology
1	Hypointense	Hyperintense	Oedema, inflammation
2	Hyperintense	Iso- or hyperintense	Fatty degeneration
3	Hypointense	Hypointense	Bony sclerosis

Finding MRI features that are specific to LBP is an active area of research. IVD degeneration is often categorized into five Pfirrmann grades (Table 3) (Pfirrmann et al., 2001). Degenerative changes in vertebral endplates are categorized with Modic types 1-3 (Table 4) (Modic, Steinberg, Ross, Masaryk, & Carter, 1988). These stages of degeneration have been associated with the severity of LBP (Kim et al., 2012; Luoma et al., 2000; Määttä, Wadge, MacGregor, Karppinen, & Williams, 2012; Mok et al., 2016; Takatalo et al., 2012, 2011). Disc herniation and fat infiltration in the multifidus muscle seen in MRI have been related to sciatica symptoms (Kader, Wardlaw, & Smith, 2000; Kjaer, Bendix, Sorensen, Korsholm, & Leboeuf-Yde, 2007). Despite extensive research, a specific MRI marker for non-specific LBP has not been found.



## 5 Aims of the thesis

The purpose of this dissertation was to study iterative reconstruction, machine learning, and deep learning algorithms in CT and MRI use cases described in the publications. Specifically, the aims of the studies were to:

1. investigate the applicability and performance of different iterative reconstruction algorithms in quantitative bone morphometry performed with sparse  $\mu$ CT projection data, and to investigate if a rabbit model of osteoarthritis could be quantitatively analyzed even with extreme data sparsity;
2. develop a texture analysis and machine learning pipeline for the prediction of low back pain from  $T_2$ -weighted MRI images, and to compare the predictive ability between intervertebral discs and vertebral bodies as well as between upper and lower lumbar levels;
3. develop a deep learning -based image reconstruction method based on generative adversarial networks for interior tomography, and to investigate if the method could be used for extended field-of-view imaging.



## 6 Materials and methods

This thesis consists of three studies performed with different modalities and data that are presented in Table 5. The studies are referred to with Roman numerals I-III according to the list of original publications.

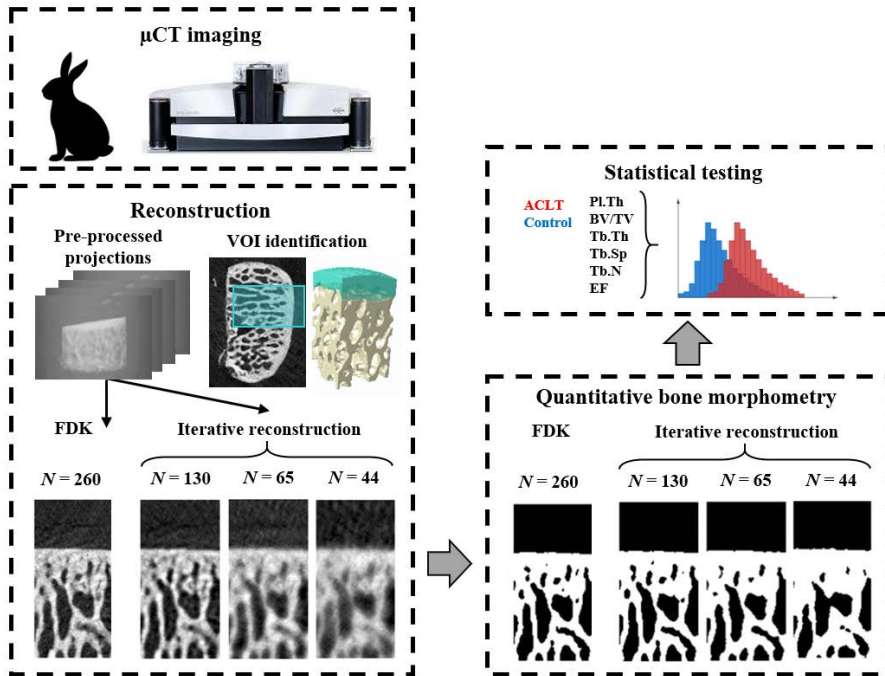
**Table 5. Summary of the imaging modalities and data used in the original publications.  $\mu$ CT, micro-computed tomography; MRI, magnetic resonance imaging; CTA, computed tomography angiography; ACLT, anterior cruciate ligament transection.**

Study	Modality	Data
I	$\mu$ CT	Osteochondral rabbit samples ( $N = 12$ in the ACLT group and $N = 16$ in the control group)
II	MRI	Northern Finland Birth Cohort 1966 lumbar MRI and pain questionnaires ( $N = 110$ in the symptomatic group and $N = 408$ in the asymptomatic group)
III	CTA	CTA slices from the Oulu University Hospital picture archiving and communication system ( $N = 10,500$ )

### 6.1 Data and imaging protocols

#### 6.1.1 Rabbit samples and micro-computed tomography (I)

In the first study, osteochondral rabbit samples were used. The rabbit ACLT model is a well-established and reproducible animal model commonly used in studying degenerative OA (Lampropoulou-Adamidou et al., 2014). Randomized knees of six female New Zealand white rabbits (*Oryctolagus cuniculus*, age 14 months at the end of the experiments, weight  $4.5 \pm 0.3$  kg) were operated on with ACLT surgery to onset OA. Four animals (age 14 months at the end of the experiments, weight  $4.8 \pm 0.3$  kg) were left unoperated as the control group. The animals were euthanized after 8 weeks from the ACLT surgery, and the medial and lateral femoral condyles were harvested from the operated knee (ACLT group) and both knees (control group). All procedures were approved by the Animal Ethics Committee at the University of Calgary and were conducted according to the guidelines of the Canadian Council on Animal Care (Certificate of Animal Use Protocol Approval #AC11-0035).



**Fig. 16. Design of Study I. Osteochondral rabbit samples imaged with micro-computed tomography ( $\mu$ CT) were reconstructed at varying sparsity levels ( $N = 260, 130, 65,$  and  $44$  projections) using the Feldkamp, Davis, and Kress algorithm (FDK) and different iterative reconstruction algorithms. Results given by quantitative bone morphometry analysis were compared between volumes-of-interest (VOI) of reconstructions given by the various reconstruction methods statistically using the Mann-Whitney test.**

The osteochondral samples were imaged with SkyScan 1272 high-resolution cone-beam  $\mu$ CT scanner (Bruker micro-CT, Kontich, Belgium). Prior to imaging, samples were wrapped in moisturized tissue paper and placed firmly in a plastic container to prevent drying and movement artifacts. The imaging parameters were 50 kV tube voltage, 200  $\mu$ A tube current, 0.5 mmAl filtering, 2016 ms exposure time per projection view,  $4 \times 4$  binning of projections, 2 frames per projection,  $1008 \times 672$  projection size,  $181.3^\circ$  angular coverage,  $0.7^\circ$  angular step, and isotropic  $9 \mu\text{m}$  camera pixel size. Projections ( $N = 260$ ) were pre-processed for beam hardening, post-alignment, and ring artifact corrections using the NRecon software (Bruker micro-CT, Kontich, Belgium) (Fig. 16).

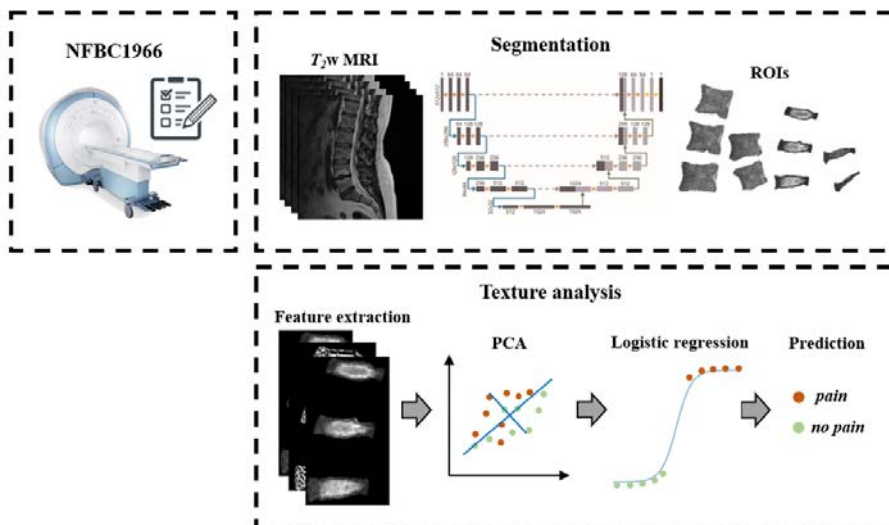


Fig. 17. Design of Study II.  $T_2$ -weighted ( $T_2w$ ) MRI segmented into four regions-of-interest (ROIs) and questionnaire-based pain characteristics from the North Finland Birth Cohort 1966 (NFBC1966) were used for a binary classification using a texture analysis pipeline consisting of feature extraction, principal component analysis (PCA), and logistic regression.

### 6.1.2 North Finland Birth Cohort 1966 lumbar MRI (II)

The Northern Finland Birth Cohort 1966 (NFBC1966) data were used in Study II (Rantakallio, 1988). NFBC1966 is a prospective population-based birth cohort, based on children born in the Oulu and Lapland regions between January 1st and December 31st in 1966. Data on health and lifestyle have been collected from the mothers ( $N = 12,068$ ) and the children ( $N = 12,231$ ) since the onset of the cohort with postal questionnaires and clinical examinations.

The 46-year follow-up study of the NFBC1966 consisted of postal health questionnaires, based on which a subset of cohort members ( $N = 5,861$ , ages 46-48) were invited for clinical examinations. After the clinical examinations, attendees living within 100 km from Oulu underwent an MRI of the lumbar spine ( $N = 1,540$ ). The imaging was performed on a Signa HDxt 1.5 T MRI system (GE Healthcare, Chicago, IL, USA). From the MRI protocol, a  $T_2$ -weighted FSE sequence ( $TE = 112.7$  ms,  $TR = 3,500$  ms,  $ETL = 27$ , slice thickness = 4 mm, matrix size =  $512 \times 512$ , and FOV =  $280$  mm  $\times$   $280$  mm) was chosen for Study II because of the visualization and contrast it offered for IVDs and vertebral bodies (Fig. 17).

**Table 6. Descriptive statistics of the data used in Study II.**

Variable	Symptomatic ( $N = 110$ , 21%)	Asymptomatic ( $N = 408$ , 79%)
	$N$ (%) or $mean \pm STD$	$N$ (%) or $mean \pm STD$
Male <sup>1</sup>	43 (39%)	210 (51%)
Female <sup>1</sup>	67 (61%)	198 (49%)
Modic type 1 <sup>1</sup>	24 (22%)	50 (12%)
Modic type 2 <sup>1</sup>	41 (37%)	99 (24%)
Modic type 3 <sup>1</sup>	19 (17%)	31 (7.6%)
Pfarrmann 1	1 (0.9%)	1 (0.2%)
Pfarrmann 2	103 (94%)	393 (96%)
Pfarrmann 3	80 (73%)	289 (71%)
Pfarrmann 4 <sup>1</sup>	60 (55%)	164 (40%)
Pfarrmann 5 <sup>1</sup>	30 (27%)	54 (13%)
Weight	80.2 $\pm$ 16.1 kg	77.6 $\pm$ 15.1 kg
Height	170.2 $\pm$ 9.2 cm	171.9 $\pm$ 9.2 cm
BMI <sup>2</sup>	27.6 $\pm$ 4.6	26.2 $\pm$ 4.4

*Note: Modic and Pfarrmann grades listed by incidence for all lumbar levels.*

<sup>1</sup> Variable significantly different ( $p < 0.05$ ,  $\chi^2$  test with contingency tables)

<sup>2</sup> Variable significantly different ( $p < 0.05$ , independent  $t$ -test, distribution normality tested with the Kolmogorov-Smirnov test)

Before imaging, the prevalence of low back pain (LBP) was elicited using a pain questionnaire. The questions posed in the questionnaire were "Have you had any aches or pains in your low back" (options: no (1) or yes (2)) and "How often have you had aches or pains during the last 12 months (options: (1) 1-7 days, (2) 8-30 days, (3) more than 30 days, or (4) daily)". Furthermore, those individuals who experienced pain graded the intensity of the pain using a scale from 0 (no pain) to 10 (extremely severe or bothersome pain). Based on the questionnaires, two groups were identified for Study II: the symptomatic group who had clinically-relevant pain symptoms (frequency  $\geq 30$  days and intensity  $\geq 6/10$ ) and the asymptomatic group that had no or very mild pain experiences (frequency  $\leq 7$  days and intensity  $\leq 3/10$ ). The used definition of clinically relevant LBP was motivated by previous LBP studies (Määttä et al., 2016; Saukkonen et al., 2020). The final data used in Study II consisted of 110 individuals in the symptomatic group and 408 individuals in the asymptomatic group. Table 6 presents descriptive statistics of the identified groups.

### 6.1.3 Computed tomography angiography data collection (III)

A registry of 500 CTA examinations that covered the entire heart craniocaudally was collected from the Oulu University Hospital picture archiving and communication system for Study III. These data were acquired on a SOMATOM Definition Flash CT scanner (Siemens Healthineers, Erlangen, Germany) with a CTA protocol (80-120 kVp depending on patient size, 256 mean mAs,  $0.73 \times 0.73 \text{ mm}^2$  mean pixel size, 2 mm slice thickness,  $512 \times 512$  image size).

The CTA data were pre-processed with zero padding, resized to  $768 \times 768$ , and cropped in the z-direction. This resulted in a dataset with 29,249 images where the heart was centered. From the full dataset, 10,000 images were randomly selected from 495 patients for the training data in Study III due to computational constraints. These data were split to training ( $N = 9,500$ ) and validation ( $N = 500$ ) sets. The test set was obtained randomly from the remaining 5 patients ( $N = 500$ ).

From the selected slices, four images were simulated for the final training data using the ASTRA tomography toolbox (v1.8, imec-Vision Lab, University of Antwerp, CWI, Amsterdam, the Netherlands) (Palenstijn, Batenburg, & Sijbers, 2011; Van Aarle et al., 2016, 2015) and the Spot toolbox (v1.2, University of British Columbia, Vancouver, CA) in the MATLAB (R2016b, MathWorks, Natick, MA, USA) programming environment (Fig. 18), as follows:

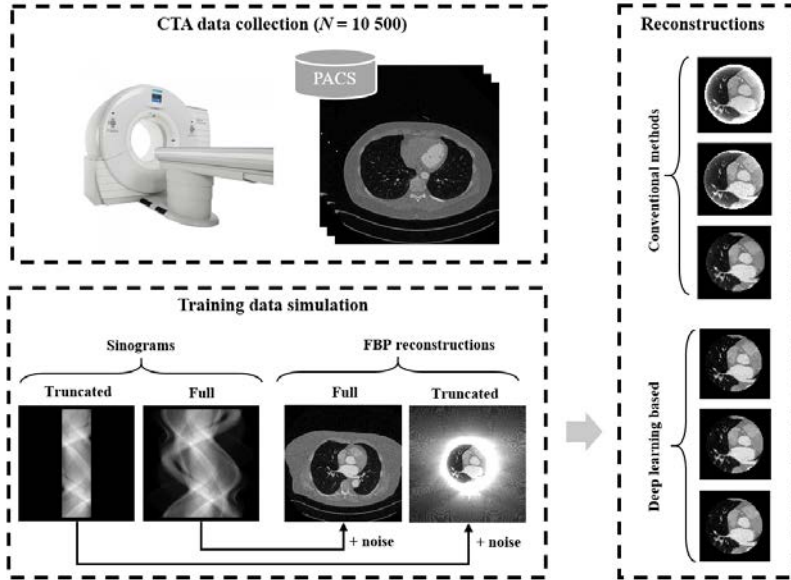
1. Full sinograms with a 768 pixel wide virtual detector with a pixel width equivalent to the reconstruction pixel size with parallel-beam geometry and 720 projection angles ( $360^\circ$  angular coverage);
2. Truncated sinograms with a 192 pixel wide virtual detector with same the specifications as above;
3. FBP (with Ram-Lak filter) reconstruction using the full sinogram; and
4. FBP (with Ram-Lak filter) reconstruction using the truncated sinogram.

In addition, Gaussian noise (with an amplitude equivalent to 1% of the full sinogram maximum value) was simulated for the sinograms to avoid inverse crime.

## 6.2 Reconstruction algorithms (I, III)

Various reconstruction algorithms were used in Studies I and III. All reconstructions in Studies I and III (except for FDK in Study I) were computed with the ASTRA tomography toolbox (v1.8) and the Spot toolbox (v1.2) in MATLAB (R2016b).

*Study I:* Pre-processed  $\mu$ CT projections were reconstructed in the cone-beam geometry, varying the sparsity of the dataset by leaving a portion of the projection views



**Fig. 18. Design of Study III.** Training data ( $N = 9,500$ ,  $N = 500$ , and  $N = 500$  slices in training, validation, and test sets, respectively) were used to train a deep learning -based network developed for interior computed tomography angiography (CTA) reconstruction. The developed method was compared with several state-of-the-art methods, both deep learning-based and conventional (iterative and sinogram pre-processing -based) methods in terms of several image quality characteristics.

out of the reconstruction process (Fig. 16). The projection data were down-sampled to 130, 65, and 44 projections (half, one-fourth, and one-sixth of the original number of projections) before reconstruction using IR methods. An FDK reconstruction (Hamming filter,  $\alpha = 0.54$ ) of the full dataset was obtained as a baseline using the NRecon software (Bruker micro-CT, Kontich, Belgium). The IR algorithms compared in this study were the conjugate gradient least squares (CGLS) method, TV regularization, and the discrete algebraic reconstruction technique (DART), and 25 iterations were used for each algorithm.

*Study III:* Simulated interior and full sinograms were reconstructed in parallel-beam geometry. An FBP reconstruction using the fully sampled sinogram was used as the ground truth (GT) to which the reconstructions were compared. A GAN-based reconstruction method was developed in the study, and details of the method are described later. The developed network was compared with FBP using truncated sinogram (TS-FBP), TV regularization (100 iterations) using a truncated sinogram (TS-TV), and an adaptive detruncation -based sinogram extension followed by FBP

(ES-ADT), as well as a U-Net based reconstruction post-processing network applied on TS-FBP (FBPConvNet) (Jin et al., 2017), and a U-Net based sinogram extension followed by FBP (ES-UNet) that was developed as an intermediate step for this research line (Ketola, Heino, Juntunen, Nieminen, & Inkinen, 2021).

### *Conjugate gradient least-squares (I)*

The ASTRA toolbox implementation of the CGLS algorithm was used. It aims to minimize the least-squares problem

$$\operatorname{argmin}_{\mu \in \mathcal{K}_k(\mathbf{W}^T \mathbf{W}, \mathbf{W}^T \mathbf{p})} \|\mathbf{W}\mu - \mathbf{p}\|_2^2 \Leftrightarrow \mathbf{W}^T \mathbf{W}\mu = \mathbf{W}^T \mathbf{p}, \quad (47)$$

performing sequential linear searches along the  $\mathbf{W}^T \mathbf{W}$ -conjugate directions that span the Krylov subspace  $\mathcal{K}_k = \operatorname{span}\{\mathbf{W}^T \mathbf{p}, (\mathbf{W}^T \mathbf{W})\mathbf{W}^T \mathbf{p}, \dots, (\mathbf{W}^T \mathbf{W})^{k-1} \mathbf{W}^T \mathbf{p}\}$  (Simoncini & Szyld, 2007).

### *Total variation regularization (I, III)*

TV regularization was used to solve the inverse problem

$$\operatorname{argmin}_{\mu} [\|\mathbf{W}\mu - \mathbf{p}\|_2^2 + \alpha \|\nabla \mu\|_1], \quad (48)$$

where  $\alpha$  was the regularization parameter tuned with the L-curve method (Hansen, 1992). The Barzilai-Borwein minimization method was used in the computation (Barzilai & Borwein, 1988; Hämläinen et al., 2014).

### *Discrete algebraic reconstruction technique (I)*

DART is an algorithm that incorporates prior knowledge of the attenuation distribution in the imaged sample (Batenburg & Sijbers, 2011). It alternates between optimization via a continuous IR algorithm for a set number of iterations and segmentation steps where voxels are classified to grayscale values which are given *a priori*. In addition, a certain number of pixels controlled by a parameter are fixed in each segmentation step that are not updated in the iterative reconstruction. CGLS with 10 iterations was used as the continuous IR algorithm and the grayscale values and thresholds for bone, soft tissue, and background were determined from an initial CGLS reconstruction.

### *Adaptive de-truncation (III)*

The ADT algorithm was applied to extend truncated sinograms by approximating the patient shape with a water-filled ellipse using a square root fit  $\sqrt{a\xi^2 + n\xi + c}$ , where  $a$ ,  $b$ , and  $c$  are the fit parameters, and  $\xi$  refers to the detector pixels (Juntunen et al., 2020; Kolditz et al., 2010; Sourbelle et al., 2005). The fitted square-root function was used to estimate the attenuation profile for the truncation areas in each projection angle. Following the extension, the extended sinograms were reconstructed with FBP.

## **6.3 Quantitative bone morphometry (I)**

In Study I, quantitative morphometry (QBM) was performed on the full data FDK reconstruction and the down-sampled iterative reconstructions (Fig. 16). The accuracy of the QBM analysis performed on the iterative reconstructions was compared against the FDK reconstructions. CT Analyzer software (CTAn, v.1.16, Bruker micro-CT, Kontich, Belgium) was used in the QBM analysis and preprocessing.

### **6.3.1 Segmentation**

Prior to the QBM, each reconstruction was segmented into bone and soft tissue. A 2 mm  $\times$  2 mm  $\times$  4 mm VOI was first identified in the weight-bearing regions of the femoral condyles, as inspired by a previous rabbit ACLT model study (Florea et al., 2015). Data from the VOI were filtered with a median (radius = 1) and unsharpening (radius = 1, amount = 50%) filter. An automatic Otsu thresholding algorithm was used for the segmentation (Otsu, 1979). After segmentation, a de-speckling filter was used to remove any noise-related regions mistaken for bone from the data. Furthermore, the trabecular bone and subchondral bone plate regions were manually segmented for feature analysis.

### **6.3.2 Feature computation**

Quantitative features for trabecular bone and the subchondral bone plate that were computed are presented in Table 7. In the Pl.Th computation an additional despeckling filter was run to remove spores in the plate, and morphological closing was used to remove pores in plate edges. The ellipsoid factor (EF) was calculated with the BoneJ plugin (v1.4.2) for the ImageJ software (v.1.51n, National Institutes of Health, Bethesda, Maryland, USA) (Abràmoff, Magalhães, & Ram, 2004; Doube, 2015; Doube et al., 2010).

**Table 7. Quantitative bone morphometry features computed for Study I.**

Feature	Unit	Explanation
<b>Subchondral bone plate</b>		
Plate thickness (Pl.Th)	$\mu\text{m}$	Average thickness of the subchondral bone plate
<b>Trabecular bone</b>		
Bone volume fraction (BV/TV)	%	Bone volume per total volume
Trabecular thickness (Tb.Th)	$\mu\text{m}$	Mean thickness of trabeculae
Trabecular separation (Tb.Sp)	$\mu\text{m}$	Mean spacing between trabeculae
Trabecular number (Tb.N)	$\text{mm}^{-1}$	Amount of trabeculae per unit length
Ellipsoid factor (EF)	a.u.	Anisotropy; how prolate (negative EF) or oblate (positive EF) the trabeculae are

### 6.3.3 Statistical analyses

QBM features in the ACLT and control groups were compared statistically with the Mann-Whitney test for two independent datasets (Mann & Whitney, 1947) in the SPSS software (v.24, IBM Analytics, New York, USA). The statistical testing method was chosen based on the sample size and absence of distribution normality requirements.

## 6.4 Texture analysis and logistic regression classification (II)

In Study II, a TA pipeline was developed for identifying symptomatic individuals from the NFBC1966 lumbar MRI data. U-Net was used to segment vertebrae and IVDs from the mid-sagittal MR images. Segmentation masks obtained with U-Net were categorized into four ROIs, as follows:

1. upper vertebrae L1, L2, and L3;
2. lower vertebrae L4 and L5;
3. upper IVDs L1-L2, L2-L3, and L3-L4; and
4. lower IVDs L4-L5 and L5-S1.

Data within the segmentation masks were subject to a custom-programmed TA module developed in MATLAB (R2019b) that extracted a total of  $N = 603$  textural features from each ROI (Table 8). In addition, Modic and Pfirrmann scores were graded by two experienced clinicians (only the grades corresponding to the lumbar levels in the ROI

**Table 8. Textural and semantic features used for binary classification in Study II. Co-occurrence and run-length features were computed in four directions (0°, 45°, 90°, and 135°) using 8-bit gray-level discretization. Co-occurrence features were computed using radii 1-5 voxels and local binary patterns were computed with radius of 1 voxel.**

Feature type	Feature names	<i>N</i>
Histogram	Maximum, minimum, mean, variance, skewness, kurtosis, percentiles	12
Co-occurrence	Angular second moment, contrast, correlation, sum of squares variance, inverse difference moment, sum average, sum entropy, sum variance, entropy, difference variance, difference entropy, information measures of correlation, maximal correlation coefficient	280
Run-length	Long and short run emphasis, run-length and grayscale non-uniformity, run percentage	20
Gradient	Mean, variance, skewness and kurtosis of absolute gradient and gradient angle images and non-zero values of absolute gradient image	9
Autoregression	Model parameters $\phi_1 - \phi_4, \sigma$	5
Wavelet	Wavelet energy of approximation and horizontal, vertical, and diagonal details using five decomposition levels	20
Local binary patterns	Histogram bins, mean	257
Modic	Modic grade in each vertebrae, mean Modic grade	6
Pfirschmann	Pfirschman grade in each intervertebral disc, mean Pfirschmann grade	6
Demographic	Gender, weight, height, body mass index	4

were used), and some demographic variables were mined from the questionnaire data (Table 6).

The TA feature space was reduced using a principal component analysis (PCA) with an 80% variance criterion. The final data for classification consisted of the principal components (PCs), Modic and Pfirschmann features, and demographic features for each ROI. These data were split into training (80%) and test (20%) sets with equal class distribution in both sets. Finally, a logistic regression classifier using the ‘liblinear’

solver was trained using the Sklearn ML library (v.0.21.2).  $L_2$  regularization was added to the classifier and the amount was tuned using five-fold cross-validation and a grid search.

## **6.5 Deep learning (II, III)**

The deep learning networks used throughout this thesis were developed in Python (v3.7) using the PyTorch DL framework (v1.5).

### **6.5.1 U-Net for segmentation (II)**

A U-Net architecture encompassing five encoding and four decoding layers was used in Study II for image segmentation (Fig. 19). Lumbar vertebrae L1...L5 and IVDs L1-L2...L5-S1 were manually segmented from 200 samples that were used to train the U-Net. A combined loss functional using BCE and Jaccard index with equal weights was used and separate models were trained for vertebrae and IVDs. The models were trained for 300 epochs using a 15% data subset for validation. The rest of the MRI data were segmented with the U-Net model.

### **6.5.2 Generator U-Net architecture (III)**

In Study III, a variation of U-Net was used as the generator network of the developed double GAN (DGAN) method (Fig. 20). Three more encoding and decoding layers were included, and a dropout was included in the bottleneck layers (Isola, Zhu, Zhou, & Efros, 2017). Two models were trained, one for sinogram extension and the other to improve the image quality in the reconstruction space.

### **6.5.3 PatchGAN discriminator (III)**

PatchGAN is an encoder architecture used as a discriminator in GAN models that encodes images as real or fake in patches whose size is determined by the receptive field size of the encoder output (Isola et al., 2017). In Study III, a  $70 \times 70$  PatchGAN architecture was used as the discriminator of the GAN model (Fig. 21). The output of PatchGAN contains predictions for  $70 \times 70$  receptive fields. As the output of PatchGAN is a tensor, the average of the output values was used in the loss calculation.

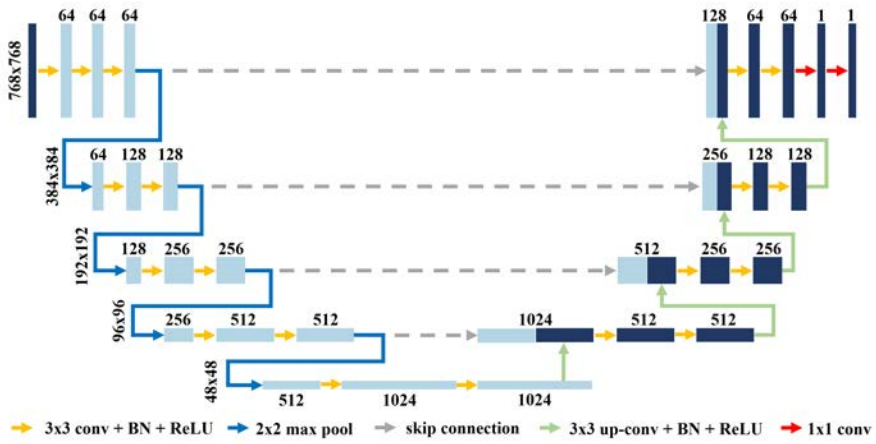


Fig. 19. The U-Net architecture used for the segmentation of MRI images in Study II. An MRI image is given in the input layer, and the network computes a segmentation mask in the output layer. Numbers correspond to the tensor sizes for an example image size of  $512 \times 512$  voxels. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit. (Figure reissued [modified] under CC BY 4.0 license from Publication II © 2021 Authors).

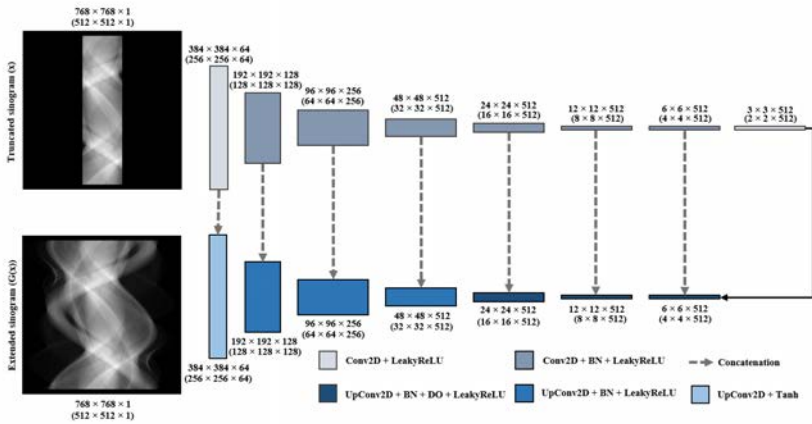


Fig. 20. The U-Net architecture used as the generator network for the GAN model in Study III. The numbers in brackets correspond to the version of the network where reconstructed images were used as input. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit; UpConv, up-sampling convolution; DO, dropout; Tanh, hyperbolic tangent. (Figure modified from Publication III).

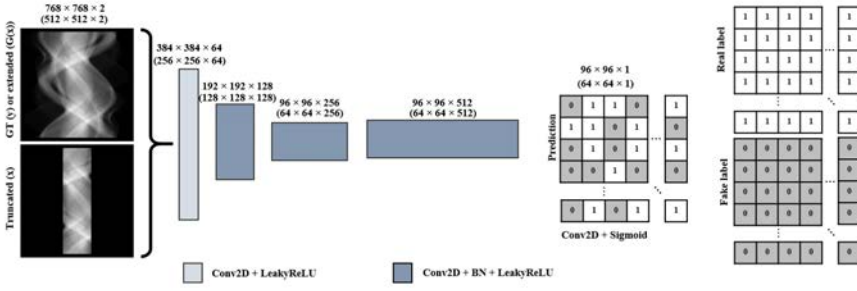


Fig. 21. The PatchGAN architecture used as the discriminator network in Study III. Each element in the network output corresponds to a  $70 \times 70$  patch in the original imaged classified as real or fake. The numbers in brackets correspond to the version of the network where reconstructed images were used as input. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit. (Figure modified from Publication III).

### 6.5.4 Double generative adversarial network model (III)

In the DGAN method developed for Study III, separate GANs were trained for sinogram extension in the projection space and image quality improvement in the reconstruction space. Both GANs consisted of a U-Net generator and a PatchGAN discriminator. Compared to the vanilla GAN model, a conditional GAN was used in which the input is not random noise but rather an image. During the training of the sinogram extension GAN, image pairs of truncated and full sinograms were used as inputs. Correspondingly, image pairs of suboptimal reconstructions and ground truth (GT) reconstructions were used for training the reconstruction space GAN.

Details of the training of the GAN models were as follows (Fig. 22). A step in discriminator ( $D$ ) optimization was done first. Image pairs of truncated sinograms  $x$  and full sinograms  $y$  were fed through  $D$  and the loss was computed as

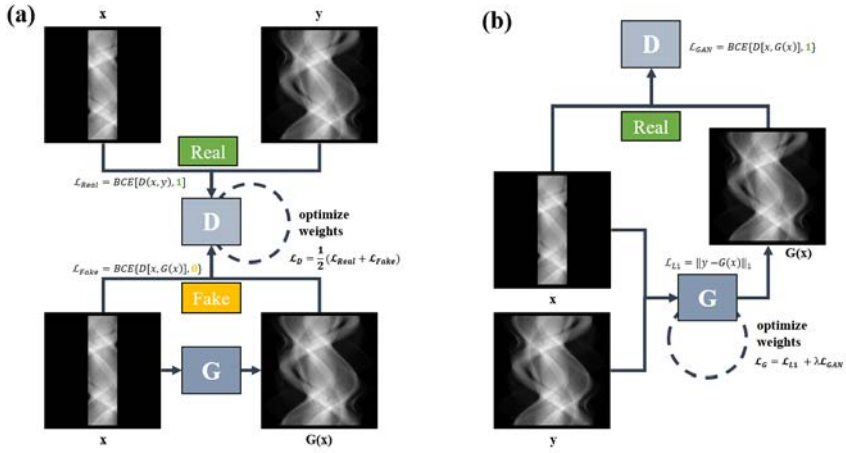
$$\mathcal{L}_{Real} = BCE[D(x, y), 1] = -\sum_i \log d_i, \quad (49)$$

where  $d_i$  are the patch outputs of  $D$ . Similarly, the loss was computed for image pairs of truncated sinograms  $x$  and extended sinograms  $G(x)$ , corresponding to the ‘fake’ case

$$\mathcal{L}_{Fake} = BCE[D(x, G(x)), 0] = -\sum_i \log(1 - d_i). \quad (50)$$

The final loss for  $D$  was obtained with

$$\mathcal{L}_D = \frac{1}{2}(\mathcal{L}_{Real} + \mathcal{L}_{Fake}). \quad (51)$$



**Fig. 22. Schematic figure of the training process used for the GAN models in Study III. BCE refers to the binary cross entropy (log) loss. (Figure modified from Publication III).**

Afterwards an optimization step was taken for the generator  $G$ . The  $L_1$  loss between the generated image  $G(x)$  and ground truth  $y$

$$\mathcal{L}_{L_1} = \|y - G(x)\|_1 = \sum_i |y_i - g_i|, \quad (52)$$

where  $y_i$  and  $g_i$  refer to voxels in the images, was combined with the loss given by the output of  $D(x, G(x))$  given the 'real' label

$$\mathcal{L}_{GAN} = BCE[D(x, G(x)), 1] = -\sum_i \log d_i \quad (53)$$

with

$$\mathcal{L}_G = \mathcal{L}_{L_1} + \lambda \mathcal{L}_{GAN}, \quad (54)$$

where  $\lambda$  was tuned to 0.01.

The developed reconstruction method using the DGAN approach consisted of six steps:

1. Extension of the truncated sinogram (TS) with the sinogram extension GAN
2. Superimposing the truncated sinogram to the extended sinogram to ensure any measured data were not changed by the GAN
3. FBP to obtain an initial reconstruction
4. Post-processing with the reconstruction space GAN

5. Forward projection of the intermediate reconstruction and superimposing of the original truncated sinogram
6. FBP to obtain the final reconstruction.

Training the networks was done with Adam using the default parameters. To improve the training outcome, label smoothing was used to reduce model overconfidence (Müller, Kornblith, & Hinton, 2019), different learning rates were used for  $G$  (0.01) and  $D$  (0.04) (Heusel, Ramsauer, Unterthiner, Nessler, & Hochreiter, 2017), and the learning rates were decayed by a factor of 10 after 100 epochs.

## 6.6 Image quality metrics (III)

The image quality in Study III was assessed using the root mean squared error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). The following definitions were used:

$$RMSE = \sqrt{\sum_i^N \frac{(I_i - GT_i)^2}{N}}; \quad (55)$$

$$PSNR = 20 \log_{10} \left( \frac{L}{RMSE} \right); \text{ and} \quad (56)$$

$$SSIM = \frac{(2\mu_I \mu_{GT} + c_1)(2\sigma_{I,GT} + c_2)}{(\mu_I^2 + \mu_{GT}^2 + c_1)(\sigma_I^2 + \sigma_{GT}^2 + c_2)}, \quad (57)$$

where  $I$  is the analyzed image,  $GT$  is the ground truth image (full data FBP in Study III),  $L = 2^{16} - 1$  is the dynamic range of the images,  $\mu$  denotes the average,  $\sigma_{I,GT}$  is the covariance of  $I$  and  $GT$ ,  $\sigma$  denotes the standard deviation,  $c_1 = (0.01 \cdot L)^2$ , and  $c_2 = (0.01 \cdot L)^2$ .



## 7 Results

### 7.1 Quantitative bone morphometry with sparse projection data (I)

#### 7.1.1 Evaluation of the rabbit model of osteoarthritis with full data

With the reference reconstruction (FDK with full 260 projection views), significantly lower BV/TV and Tb.Th, and significantly higher EF, were measured in the ACLT group ( $p < 0.05$ , non-parametric Mann-Whitney test) (Table 9). Furthermore, the Pl.Th and Tb.Sp were higher, and the Tb.N was lower in the ACLT group but these changes were not statistically significant (Table 9).

#### 7.1.2 Evaluation using iterative reconstruction methods

Example reconstructions using CGLS, TV, and DART exhibited visually degraded levels of image quality as the number of used projections was reduced (Fig. 23b). After segmentation, all algorithms preserved similar features to FDK using 130 projections, additionally DART was able to capture those features even with 65 and 44 projections based on visual inspection (Fig. 23c).

**Table 9. Statistics of the rabbit model features in the anterior cruciate ligament transection (ACLT) ( $N = 12$ ) and control ( $N = 16$ ) groups using the full data ( $N = 260$  projection views) FDK (Feldkamp, Davis, and Kress algorithm) reconstruction in quantitative bone morphometry. The statistical significance is indicated with a one-tailed (two-tailed)  $p$ -value. Pl.Th, plate thickness; BV/TV, bone volume fraction; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.N, trabecular number; EF, ellipsoid factor.**

Feature	ACLT	Control	$p$
<b>Subchondral bone plate</b>			
Pl.Th ( $\mu\text{m}$ )	$513.8 \pm 90.7$	$485.0 \pm 99.7$	0.268 (0.537)
<b>Trabecular bone</b>			
BV/TV (%)	$47.7 \pm 4.4^2$	$51.9 \pm 3.4$	0.010 (0.020)
Tb.Th ( $\mu\text{m}$ )	$202.9 \pm 18.2^1$	$214.7 \pm 16.1$	0.045 (0.090)
Tb.Sp ( $\mu\text{m}$ )	$311.6 \pm 45.0$	$285.6 \pm 25.9$	0.080 (0.159)
Tb.N ( $\text{mm}^{-1}$ )	$2.36 \pm 0.19$	$2.42 \pm 0.12$	0.254 (0.507)
EF (a.u.)	$-0.189 \pm 0.04^2$	$-0.266 \pm 0.04$	0.001 (0.002)

<sup>1</sup>Values significantly different from the control group (one-tailed  $p < 0.05$ ).

<sup>2</sup>Values significantly different from the control group (two-tailed  $p < 0.05$ ).

Non-parametric Mann-Whitney test (exact).

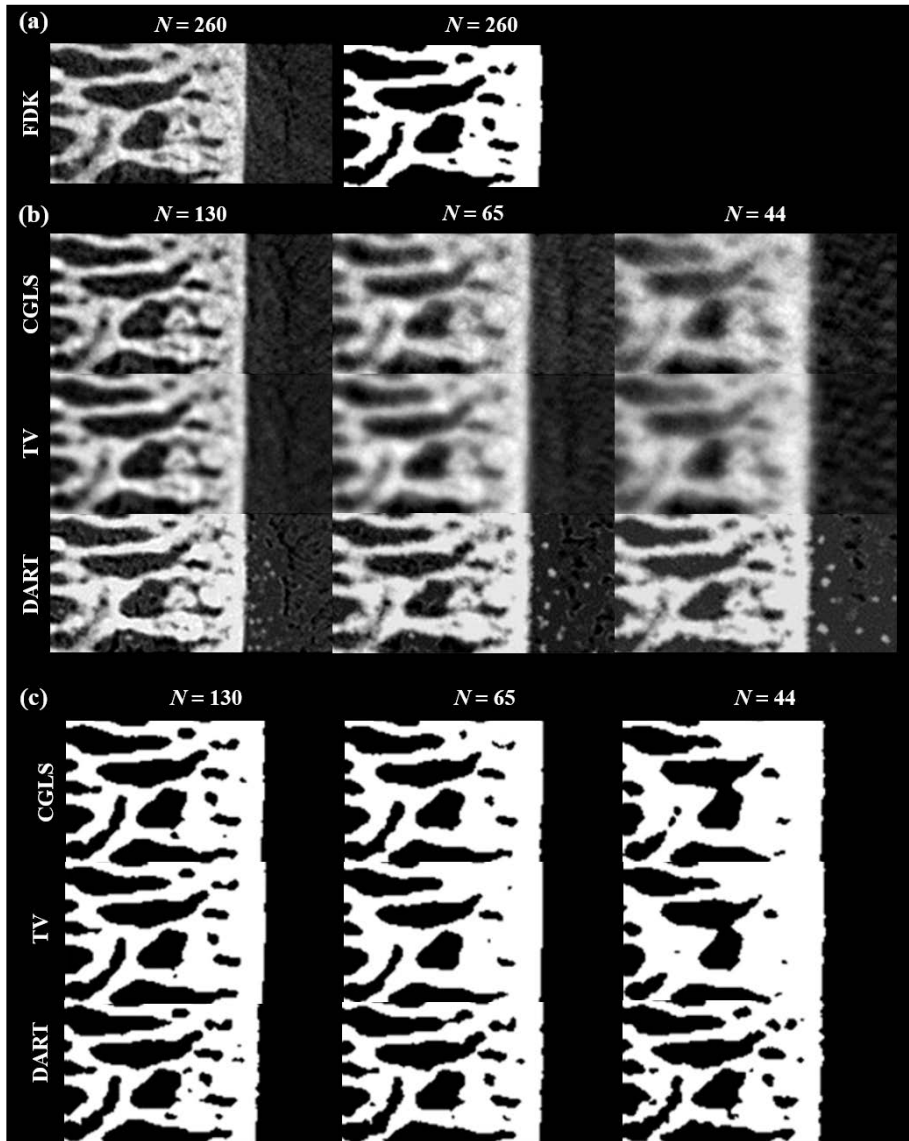
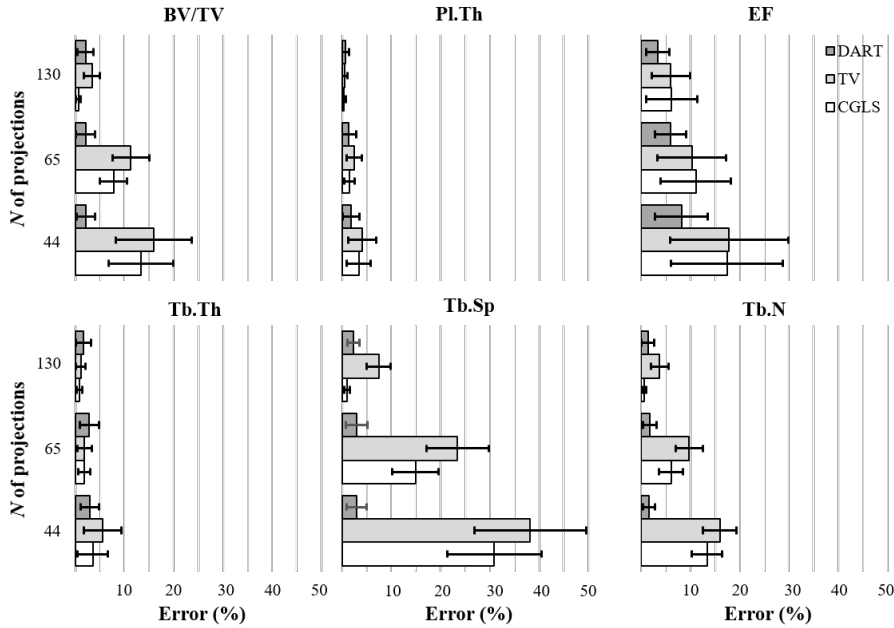


Fig. 23. (a) Example reconstruction slice and segmentation mask using a full data FDK (Feldkamp, Davis, and Kress algorithm) reconstruction (image size  $80 \times 160$  voxels). (b) Example reconstruction slices. (c) Example segmentation masks obtained with varying numbers of projections and different iterative reconstruction algorithms. CGLS, conjugate gradient least squares; TV, total variation regularization; DART, discrete algebraic reconstruction technique. (Figure reissued [modified] under CC BY 4.0 license from Publication I © 2018 Authors).



**Fig. 24.** Mean error % (bars) and standard deviation (lines) of individual quantitative bone morphometry parameters for the iteratively reconstructed data compared to the full data. Mean error was calculated as the mean of the errors in individual samples. Pl.Th, plate thickness; BV/TV, bone volume fraction; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.N, trabecular number; EF, ellipsoid factor; CGLS, conjugate gradient least squares; TV, total variation regularization; DART, discrete algebraic reconstruction technique. (Figure reissued [modified] under CC BY 4.0 license from Publication I © 2018 Authors).

Statistically significant differences in BV/TV and EF were preserved with all the methods in case of 130 projection images, additionally DART also preserved the statistically significant reduction in Tb.Th (Table 10). When 65 projections were used for reconstruction, DART preserved all the statistical significances, while only CGLS was able to correctly capture the significant difference in BV/TV (Table 11). In addition, TV generated a falsely significant difference in Tb.N. Similarly in the case of 44 projections, DART preserved all the significances while CGLS and TV generated a falsely significant difference in Tb.N (Table 12).

Using DART in reconstruction resulted in the smallest deviations (maximally 8%) from the reference FDK data for most QBM parameters at all sparsity levels (Fig. 24). In contrast with CGLS and TV, the error approached 40% with Tb.Sp using 44 projections, for example. Overall, DART was the best-performing IR algorithm both in terms of statistical analysis and QBM errors.

**Table 10. Statistics of the rabbit model features in the anterior cruciate ligament transection (ACLT) ( $N = 12$ ) and control ( $N = 16$ ) groups using half of the projection data ( $N = 130$  projections) and iterative reconstruction methods. A statistical significance is indicated with a one-tailed (two-tailed)  $p$ -value. PI.Th, plate thickness; BV/TV, bone volume fraction; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.N, trabecular number; EF, ellipsoid factor; CGLS, conjugate gradient least squares; TV, total variation regularization; DART, discrete algebraic reconstruction technique.**

50% of full data (130 projections)				
Feature	Algorithm	ACLT	Control	$p$
<b>Subchondral bone plate</b>				
PI.Th ( $\mu\text{m}$ )	CGLS	$531.5 \pm 92.6$	$483.3 \pm 99.6$	0.254 (0.508)
	TV	$533.2 \pm 93.3$	$486.2 \pm 97.9$	0.254 (0.508)
	DART	$527.6 \pm 90.9$	$484.6 \pm 99.4$	0.316 (0.631)
<b>Trabecular bone</b>				
BV/TV (%)	CGLS	$47.9 \pm 4.3^2$	$51.7 \pm 3.4$	0.019 (0.037)
	TV	$49.8 \pm 4.3^1$	$53.2 \pm 3.2$	0.026 (0.053)
	DART	$47.6 \pm 4.8^2$	$51.9 \pm 3.7$	0.015 (0.029)
Tb.Th ( $\mu\text{m}$ )	CGLS	$202.8 \pm 16.4$	$213.7 \pm 16.4$	0.061 (0.121)
	TV	$220.5 \pm 16.8$	$228.4 \pm 17.2$	0.140 (0.280)
	DART	$199.2 \pm 18.5^1$	$213.4 \pm 17.6$	0.033 (0.066)
Tb.Sp ( $\mu\text{m}$ )	CGLS	$308.2 \pm 45.8$	$283.8 \pm 25.7$	0.111 (0.223)
	TV	$311.7 \pm 45.2$	$287.9 \pm 24.0$	0.087 (0.174)
	DART	$305.8 \pm 41.5$	$281.7 \pm 25.9$	0.095 (0.189)
Tb.N ( $\text{mm}^{-1}$ )	CGLS	$2.37 \pm 0.18$	$2.42 \pm 0.12$	0.265 (0.529)
	TV	$2.26 \pm 0.17$	$2.34 \pm 0.12$	0.148 (0.296)
	DART	$2.40 \pm 0.18$	$2.44 \pm 0.11$	0.353 (0.706)
EF (a.u.)	CGLS	$-0.201 \pm 0.06^2$	$-0.268 \pm 0.05$	0.002 (0.004)
	TV	$-0.203 \pm 0.05^2$	$-0.277 \pm 0.05$	0.002 (0.004)
	DART	$-0.189 \pm 0.05^2$	$-0.263 \pm 0.05$	0.001 (0.002)

<sup>1</sup>Values significantly different from the control group (one-tailed  $p < 0.05$ ).

<sup>2</sup>Values significantly different from the control group (two-tailed  $p < 0.05$ ).

Non-parametric Mann-Whitney test (exact).

**Table 11. Statistics of the rabbit model features in the anterior cruciate ligament transection (ACLT) ( $N = 12$ ) and control ( $N = 16$ ) groups using one-fourth of the projection data ( $N = 130$  projections) and iterative reconstruction methods. A statistical significance is indicated with a one-tailed (two-tailed)  $p$ -value. Pl.Th, plate thickness; BV/TV, bone volume fraction; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.N, trabecular number; EF, ellipsoid factor; CGLS, conjugate gradient least squares; TV, total variation regularization; DART, discrete algebraic reconstruction technique.**

25% of full data (65 projections)				
Feature	Algorithm	ACLT	Control	$p$
<b>Subchondral bone plate</b>				
Pl.Th ( $\mu\text{m}$ )	CGLS	540.2 $\pm$ 93.6	490.2 $\pm$ 98.1	0.211 (0.423)
	TV	543.3 $\pm$ 93.9	495.2 $\pm$ 96.5	0.186 (0.371)
	DART	524.6 $\pm$ 94.8	486.86 $\pm$ 97.9	0.332 (0.664)
<b>Trabecular bone</b>				
BV/TV (%)	CGLS	52.0 $\pm$ 4.3 <sup>2</sup>	55.4 $\pm$ 3.3	0.024 (0.047)
	TV	53.8 $\pm$ 4.8	57.1 $\pm$ 3.5	0.055 (0.110)
	DART	47.4 $\pm$ 4.2 <sup>2</sup>	52.6 $\pm$ 4.2	0.002 (0.004)
Tb.Th ( $\mu\text{m}$ )	CGLS	236.7 $\pm$ 16.0	243.1 $\pm$ 17.4	0.186 (0.371)
	TV	255.4 $\pm$ 19.9	260.2 $\pm$ 18.5	0.316 (0.631)
	DART	202.8 $\pm$ 18.3 <sup>2</sup>	219.3 $\pm$ 17.9	0.021 (0.042)
Tb.Sp ( $\mu\text{m}$ )	CGLS	306.4 $\pm$ 45.9	281.6 $\pm$ 24.6	0.073 (0.146)
	TV	314.9 $\pm$ 46.0	289.1 $\pm$ 25.2	0.103 (0.205)
	DART	304.5 $\pm$ 46.1	276.4 $\pm$ 27.4	0.061 (0.121)
Tb.N ( $\text{mm}^{-1}$ )	CGLS	2.20 $\pm$ 0.14	2.28 $\pm$ 0.11	0.080 (0.159)
	TV	2.11 $\pm$ 0.14 <sup>1</sup>	2.20 $\pm$ 0.12	0.045 (0.090)
	DART	2.34 $\pm$ 0.19	2.40 $\pm$ 0.11	0.198 (0.397)
EF (a.u.)	CGLS	-0.205 $\pm$ 0.04 <sup>2</sup>	-0.280 $\pm$ 0.04	0.001 (0.002)
	TV	-0.190 $\pm$ 0.05 <sup>2</sup>	-0.280 $\pm$ 0.04	0.001 (0.002)
	DART	-0.187 $\pm$ 0.04 <sup>2</sup>	-0.253 $\pm$ 0.04	0.001 (0.002)

<sup>1</sup>Values significantly different from the control group (one-tailed  $p < 0.05$ ).

<sup>2</sup>Values significantly different from the control group (two-tailed  $p < 0.05$ ).

Non-parametric Mann-Whitney test (exact).

**Table 12. Statistics of the rabbit model features in the anterior cruciate ligament transection (ACLT) ( $N = 12$ ) and control ( $N = 16$ ) groups using one-sixth of the projection data ( $N = 130$  projections) and iterative reconstruction methods. A statistical significance is indicated with a one-tailed (two-tailed)  $p$ -value. PI.Th, plate thickness; BV/TV, bone volume fraction; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.N, trabecular number; EF, ellipsoid factor; CGLS, conjugate gradient least squares; TV, total variation regularization; DART, discrete algebraic reconstruction technique.**

17% of full data (44 projections)				
Feature	Algorithm	ACLT	Control	$p$
<b>Subchondral bone plate</b>				
PI.Th ( $\mu\text{m}$ )	CGLS	$548.4 \pm 96.3$	$498.3 \pm 93.7$	0.130 (0.260)
	TV	$551.9 \pm 96.8$	$501.4 \pm 92.9$	0.130 (0.260)
	DART	$524.2 \pm 95.7$	$488.9 \pm 97.1$	0.349 (0.698)
<b>Trabecular bone</b>				
BV/TV (%)	CGLS	$54.4 \pm 5.8$	$58.5 \pm 5.0$	0.055 (0.110)
	TV	$55.6 \pm 6.4$	$59.9 \pm 5.5$	0.061 (0.121)
	DART	$47.5 \pm 4.0^2$	$52.6 \pm 4.2$	0.002 (0.004)
Tb.Th ( $\mu\text{m}$ )	CGLS	$270.3 \pm 24.4$	$276.5 \pm 47.1$	0.332 (0.664)
	TV	$285.6 \pm 28.2$	$291.5 \pm 27.5$	0.437 (0.873)
	DART	$202.1 \pm 15.4^2$	$219.2 \pm 18.3$	0.010 (0.020)
Tb.Sp ( $\mu\text{m}$ )	CGLS	$321.0 \pm 47.1$	$294.1 \pm 27.9$	0.130 (0.260)
	TV	$328.1 \pm 48.4$	$299.8 \pm 29.9$	0.120 (0.241)
	DART	$301.8 \pm 43.6$	$277.9 \pm 28.2$	0.095 (0.189)
Tb.N ( $\text{mm}^{-1}$ )	CGLS	$2.02 \pm 0.15^1$	$2.12 \pm 0.13$	0.041 (0.082)
	TV	$1.95 \pm 0.15^2$	$2.05 \pm 0.12$	0.021 (0.042)
	DART	$2.35 \pm 0.18$	$2.40 \pm 0.12$	0.268 (0.537)
EF (a.u.)	CGLS	$-0.179 \pm 0.04^2$	$-0.234 \pm 0.05^2$	0.001 (0.002)
	TV	$-0.180 \pm 0.05^2$	$-0.252 \pm 0.04$	0.001 (0.002)
	DART	$-0.182 \pm 0.04^2$	$-0.240 \pm 0.05$	0.002 (0.004)

<sup>1</sup>Values significantly different from the control group (one-tailed  $p < 0.05$ ).

<sup>2</sup>Values significantly different from the control group (two-tailed  $p < 0.05$ ).

Non-parametric Mann-Whitney test (exact).

**Table 13. Characterization of classifier performance in Study II using a semantic classification based on grading features (Modic and Pfirrmann) with demographic variables. IVD, intervertebral disc; NPV, negative predictive value; ROC-AUC, receiver-operating characteristic area-under-curve.**

Parameter	Semantic classification			
	Modic grading		Pfirrmann grading	
	L1-L2...L3-L4	L4-L5, L5-S1	L1-L2...L3-L4	L4-L5, L5-S1
Specificity	0.62	0.61	0.56	0.55
Sensitivity	0.54	0.59	0.64	0.64
NPV	0.84	0.85	0.85	0.85
Precision	0.28	0.29	0.28	0.27
Accuracy	0.60	0.61	0.58	0.57
ROC-AUC	0.57	0.64	0.60	0.62

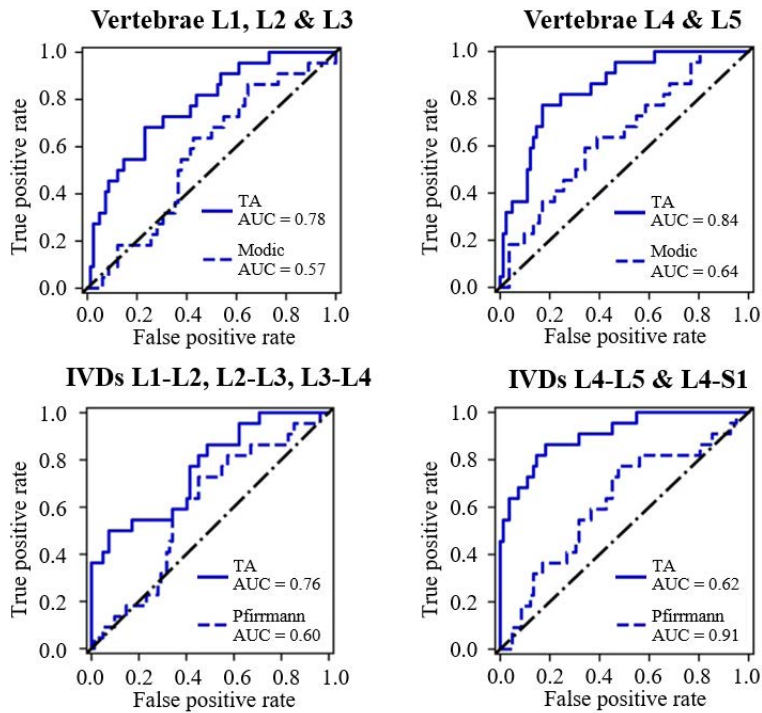
**Table 14. Characterization of classifier performance in Study II when textural features were additionally included in the analysis. IVD, intervertebral disc; NPV, negative predictive value; ROC-AUC, receiver-operating characteristic area-under-curve.**

Parameter	Texture based classification			
	Lumbar vertebrae		Lumbar IVDs	
	L1, L2, L3	L4, L5	L1-L2...L3-L4	L4-L5, L5-S1
Specificity	0.84	0.77	0.79	0.83
Sensitivity	0.55	0.77	0.55	0.82
NPV	0.87	0.93	0.87	0.94
Precision	0.48	0.47	0.41	0.56
Accuracy	0.78	0.77	0.74	0.83
ROC-AUC	0.78	0.84	0.76	0.91

## 7.2 Machine learning -enhanced low-back pain diagnostics (II)

### 7.2.1 Classification

Using semantic and demographic features in the classification of symptomatic and asymptomatic subjects resulted in inferior classifier performance (Table 13). When textural features were added, the results improved substantially in all of the four ROIs (Table 14). Notably, in the two lowest IVDs (L4-L5 and L5-S1), the classification was the most accurate in terms of all of the analyzed scores (specificity = 83%, sensitivity = 82%, negative predictive value (NPV) = 94%, precision = 56%, accuracy = 83%, ROC-AUC (receiver-operating characteristic area-under-curve) = 0.91). The ROC-AUC scores for the other ROIs were 0.78 for the three uppermost vertebrae, 0.84 for the two lowest vertebrae, and 0.76 for the three uppermost IVDs (Table 14, Fig. 25).



**Fig. 25. Receiver-operating characteristic curves for the semantic (Modic grading for vertebrae and Pfirrmann grading for intervertebral discs) and texture based classification (including semantic features) in the analyzed regions-of-interest. TA, texture analysis; AUC, area-under-curve. (Figure reissued [modified] under CC BY 4.0 license from Publication II © 2021 Authors).**

**Table 15. Classification performance characteristics using the first definition for non-specific low back pain. IVD, intervertebral disc; NPV, negative predictive value; ROC-AUC, receiver-operating characteristic area-under-curve.**

Parameter	Case 1: No subjects with sciatica symptoms ( $N = 69$ )			
	Lumbar vertebrae		Lumbar IVDs	
	L1, L2, L3	L4, L5	L1-L2...L3-L4	L4-L5, L5-S1
Specificity	0.88	0.87	0.77	0.93
Sensitivity	0.64	0.50	0.79	0.71
NPV	0.94	0.91	0.95	0.95
Precision	0.47	0.39	0.37	0.63
Accuracy	0.84	0.81	0.77	0.90
ROC-AUC	0.83	0.90	0.91	0.94

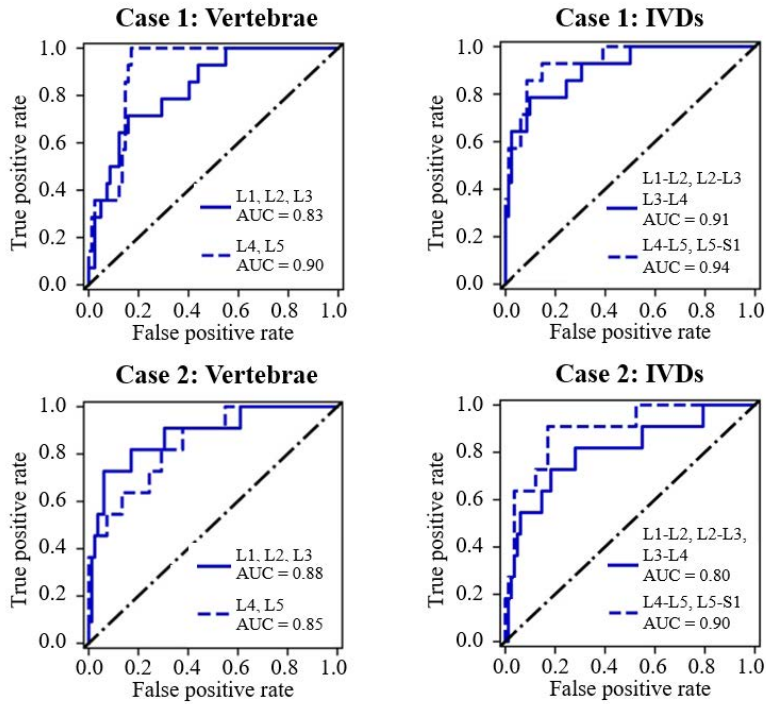
**Table 16. Classification performance characteristics using the second definition for non-specific low back pain. IVD, intervertebral disc; NPV, negative predictive value; ROC-AUC, receiver-operating characteristic area-under-curve.**

Parameter	Case 2: Additional removal of large Modic 1/2 ( $N = 54$ )			
	Lumbar vertebrae		Lumbar IVDs	
	L1, L2, L3	L4, L5	L1-L2...L3-L4	L4-L5, L5-S1
Specificity	0.85	0.93	0.93	0.83
Sensitivity	0.73	0.45	0.55	0.82
NPV	0.96	0.93	0.94	0.97
Precision	0.40	0.45	0.50	0.40
Accuracy	0.84	0.87	0.88	0.83
ROC-AUC	0.88	0.85	0.80	0.90

### 7.2.2 Sensitivity analysis to non-specific sources of pain

To further assess the effectiveness of the model used in Study II, two cases describing non-specific sources of LBP were analyzed. In the first case, subjects experiencing sciatica symptoms (radiating pain in the leg) were discarded from the data. The removal of such subjects resulted in 69 cases left to analyze. In this case, the classifier preserved its ability to distinguish symptomatic from asymptomatic subjects. The results even improved slightly for most scores in all ROIs (Table 15). The best ROC-AUC (0.94) was obtained with the lower IVDs (Fig. 26). For the other ROIs, the ROC-AUC scores were 0.83 for upper vertebrae, 0.90 for lower vertebrae, and 0.91 for upper IVDs (Fig. 26).

The second case involved removing subjects with sciatica symptoms and with large Modic changes as graded by two clinicians. A ‘large’ change was defined as a degenerative change corresponding to Modic types 1 or 2 in the vertebral endplate that



**Fig. 26. Receiver-operating characteristic curves for the two non-specific cases. Case 1: Subjects with sciatica symptoms removed from the dataset. Case 2: Subjects with sciatica symptoms and large Modic class 1 or 2 changes removed from the dataset. IVD, intervertebral disc; AUC, area-under-curve. (Figure reissued [modified] under CC BY 4.0 license from Publication II © 2021 Authors).**

exceeded 25% of the height of vertebra. Doing so resulted in a further reduction to 54 cases. In this case, the results worsened in all of the ROIs except for the upper vertebrae (Table 16). However, the highest ROC-AUC was again observed in the lowest IVDs (0.90) (Fig. 26). For the other ROIs, the ROC-AUC scores were 0.88 for upper vertebrae, 0.85 for lower vertebrae, and 0.80 for upper IVDs (Fig. 26). Overall, the results in case two were comparable to the original classification results (Table 14, Fig. 25).

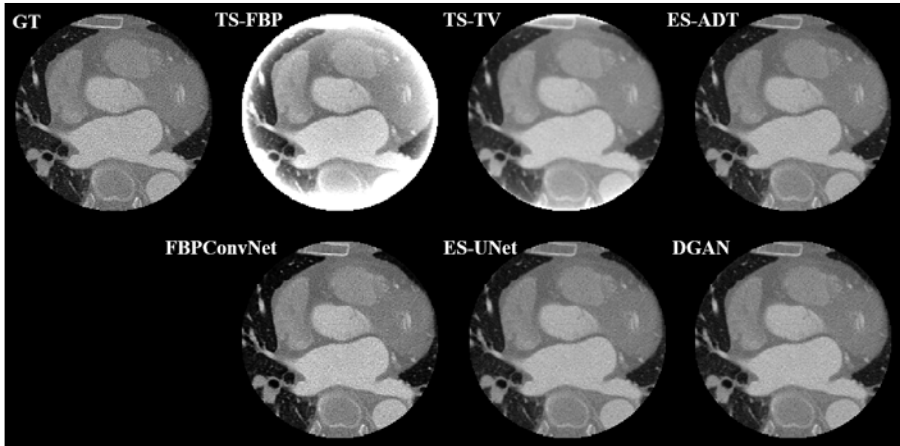


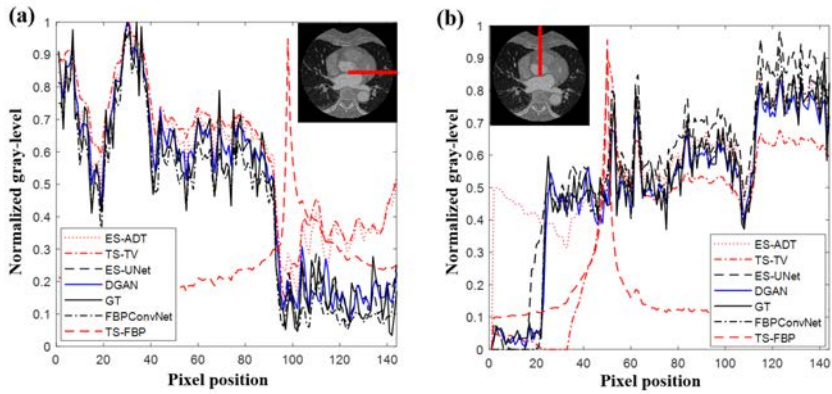
Fig. 27. Example reconstructions using the different image reconstruction approaches in Study III visualized in the original circular reconstruction field-of-view of 96 voxel in radius. The images are normalized and windowed with the following bounds: Ground truth (GT) [0,1]; truncated sinogram filtered backprojection (TS-FBP) [0,0.1]; truncated sinogram total variation regularization (TS-TV) [0,0.75]; extended sinogram with adaptive detruncation (ES-ADT) [0,1.1]; FBPCovNet [0,1]; extended sinogram with U-Net (ES-UNet) [0,1.1]; double generative adversarial network (DGAN) [0,0.95]. (Figure obtained from Publication III).

## 7.3 Deep learning interior computed tomography angiography (III)

### 7.3.1 Image quality analysis

An example interior CTA reconstruction slice image obtained with the developed DGAN method developed in Study III is visually compared with the reference methods in Fig. 27. In terms of image quality metrics, DGAN resulted in the best RMSE ( $0.02 \pm 0.01$ ), PSNR ( $34.5 \pm 2.6$  dB), and SSIM ( $0.93 \pm 0.03$ ) values for the original circular reconstruction FOV of 96 voxel radius (Table 17). Nearly as good results were observed for the ES-UNet ( $0.03 \pm 0.01$  RMSE,  $34.1 \pm 2.7$  dB PSNR,  $0.90 \pm 0.03$  SSIM) and ES-ADT (worse RMSE and PSNR but a similar SSIM of  $0.93 \pm 0.04$ ).

In the original FOV, all of the DL-based methods were in excellent agreement with the ground truth FBP reconstruction, whereas TS-FBP, TS-TV and ES-ADT deviated from it when line profiles in the horizontal and vertical directions were investigated in an example slice (Fig. 28).



**Fig. 28.** Normalized line profiles in the horizontal (a) and vertical (b) directions in the original field-of-view (FOV) and the 50% extended FOV for each algorithm investigated in Study III. The original FOV is highlighted in the plot and image. GT, ground truth; TS-FBP, truncated sinogram FBP; TS-TV, truncated sinogram total variation regularization; ES-ADT, extended sinogram with adaptive detruncation; ES-UNet, extended sinogram with U-Net; DGAN, double generative adversarial network. (Figure modified from Publication III).

### 7.3.2 Extended field-of-view results

DGAN yielded the best results for a 10% FOV extension ( $0.02 \pm 0.01$  RMSE,  $32.9 \pm 2.7$  PSNR, and  $0.91 \pm 0.03$  SSIM), 20% FOV extension ( $0.03 \pm 0.01$  RMSE,  $31.6 \pm 2.6$  dB PSNR, and  $0.90 \pm 0.03$  SSIM), and 50% FOV extension ( $0.04 \pm 0.01$  RMSE,  $29.0 \pm 2.6$  dB PSNR, and  $0.84 \pm 0.03$  SSIM) (Table 17). Visually, DGAN was able to remove the truncation artifacts completely and also recover structures in the extended reconstruction FOV (Fig. 29). In the example line profile, DGAN was in excellent agreement with the GT profile even in the extended regions. Out of the reference algorithms, ES-UNet was the second best, but slightly inferior compared to DGAN in the extended areas in terms of the analyzed image quality metrics (Table 17).

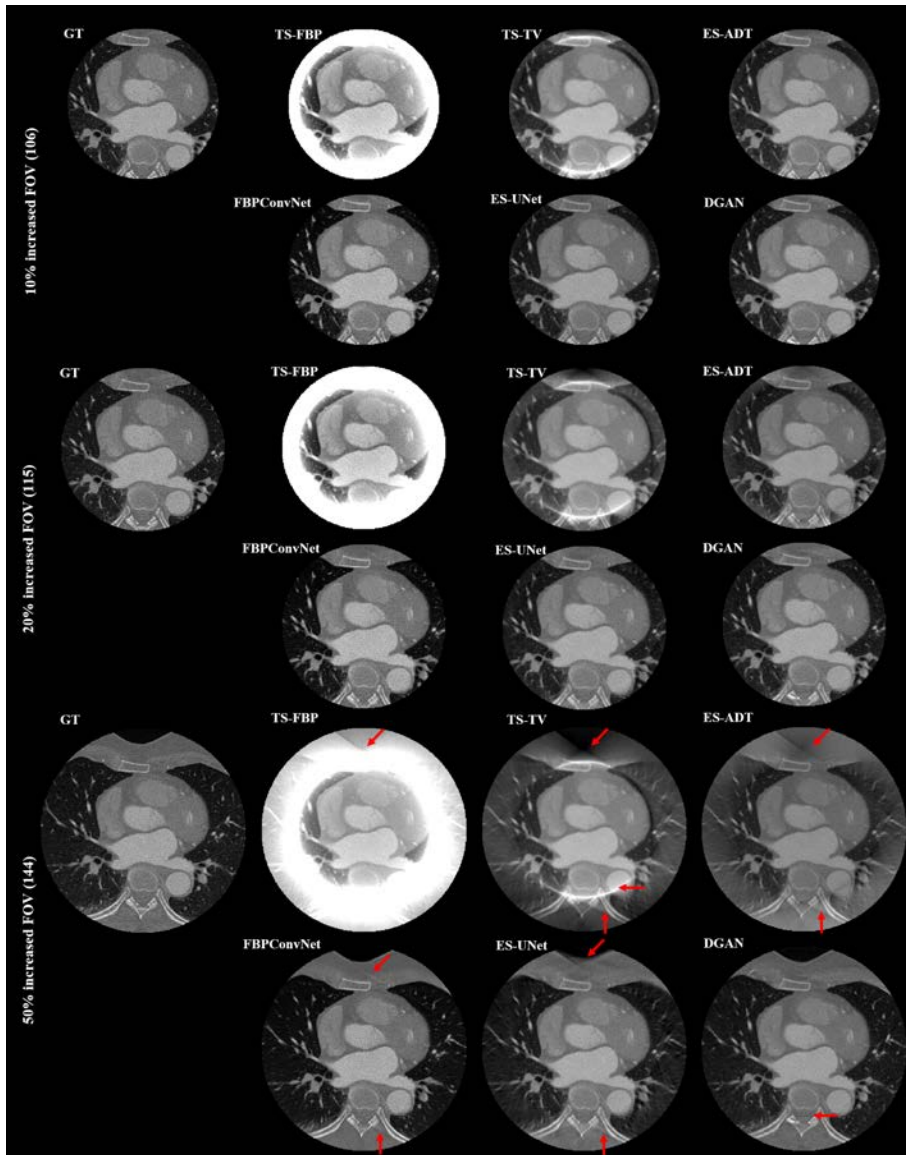


Fig. 29. Example reconstructions with enlarged FOVs using the image reconstruction methods in Study III. The images are normalized and windowed with the following bounds: Ground truth (GT) [0,1]; truncated sinogram filtered backprojection (TS-FBP) [0,0.1]; truncated sinogram total variation regularization (TS-TV) [0,0.75]; extended sinogram with adaptive detruncation (ES-ADT) [0,1.1]; FBPCConvNet [0,1]; extended sinogram with U-Net (ES-UNet) [0,1.1]; double generative adversarial network (DGAN) [0,0.95]. Arrows highlight some of the subtle artifacts in the images. (Figure modified from Publication III).

**Table 17. Image quality metrics computed for the reconstruction methods used in Study III with different reconstruction fields-of-view (FOVs). TS-FBP, truncated sinogram FBP; TS-TV, truncated sinogram total variation regularization; ES-ADT, extended sinogram with adaptive detruncation; FCN, FBPCovNet; ES-UNet, extended sinogram with U-Net; DGAN, double generative adversarial network; RMSE, root mean squared error; PSNR, peak signal-to-noise-ratio; SSIM, structural similarity index.**

Param	TS-FBP	TS-TV	ES-ADT	FCN	ES-UNet	DGAN
RMSE <sup>1</sup>	0.37±0.09	0.12±0.05	0.05±0.04	0.06±0.02	0.03±0.01	0.02±0.01
RMSE <sup>2</sup>	0.34±0.08	0.11±0.04	0.06±0.03	0.06±0.02	0.03±0.02	0.02±0.01
RMSE <sup>3</sup>	0.30±0.07	0.11±0.04	0.07±0.03	0.06±0.02	0.03±0.02	0.03±0.01
RMSE <sup>4</sup>	0.26±0.06	0.13±0.03	0.13±0.04	0.06±0.02	0.05±0.02	0.04±0.01
PSNR <sup>1</sup>	8.9±2.4	19.1±4.0	27.5±4.0	24.9±3.3	34.1±2.7	34.5±2.6
PSNR <sup>2</sup>	9.7±2.1	19.4±3.1	24.7±3.0	24.9±3.3	32.3±3.3	32.9±2.7
PSNR <sup>3</sup>	10.6±2.0	19.4±2.7	23.0±2.8	25.0±3.1	30.6±3.3	31.6±2.6
PSNR <sup>4</sup>	11.6±2.0	18.2±2.0	18.2±2.7	25.3±2.8	27.4±3.4	29.0±2.6
SSIM <sup>1</sup>	0.24±0.09	0.74±0.09	0.93±0.04	0.83±0.03	0.90±0.03	0.93±0.03
SSIM <sup>2</sup>	0.25±0.07	0.71±0.07	0.90±0.04	0.82±0.03	0.91±0.03	0.91±0.03
SSIM <sup>3</sup>	0.28±0.07	0.70±0.06	0.86±0.05	0.79±0.04	0.89±0.03	0.90±0.03
SSIM <sup>4</sup>	0.33±0.07	0.64±0.05	0.74±0.08	0.76±0.04	0.84±0.03	0.84±0.03

<sup>1</sup>Original FOV (96 voxel radius).

<sup>2</sup>10% increased FOV (106 voxel radius).

<sup>3</sup>20% increased FOV (115 voxel radius).

<sup>4</sup>50% increased FOV (144 voxel radius).

## 8 Discussion

### 8.1 Key findings

To summarize, the key findings of the studies were that:

1. The required  $\mu$ CT projection data for performing a QBM analysis evaluating the rabbit model of OA could be reduced to a fraction of the original amount when the reconstruction was done iteratively, while keeping the statistical outcome the same as the reference scan using full data.
2. Subjects with LBP could be differentiated from asymptomatic subjects based on textural features extracted from lumbar MRI. The classification was robust to non-specific sources of LBP, and IVDs at the lower lumbar levels were more sensitive than other ROIs.
3. A method for interior CTA reconstruction based on GANs was developed. The method was able to produce excellent image quality in the reconstructions, and extend the reconstruction FOV by 20%.

### 8.2 Sparse data micro-computed tomography (I)

#### 8.2.1 *Rabbit model obtained with the reference reconstruction*

Significant changes were observed in the trabecular bone microstructure after 8 weeks in rabbits operated on with ACLT, which indicates OA-related degeneration. In another 8-week rabbit model study comparing ACLT and control groups, significant reductions in both BV/TV and Tb.Th have been observed (S. X. Wang, Lavery, Dumitriu, Plaas, & Grynepas, 2007). Consistently, in a similar study comparing knees operated on with ACLT to contralateral knees after 4 weeks of surgery, analogous changes were reported in BV/TV and Tb.Th (Florea et al., 2015). However, in the 4-week study, subchondral bone plate was also significantly thinner in the ACLT group while the structural model index (an alternative measure of anisotropy to EF) was not significantly different.

#### 8.2.2 *Iterative reconstruction algorithms*

When comparing the results of the statistical analysis performed on the QBM features obtained from IR reconstructions to those of the reference data, DART was clearly the best-performing algorithm. DART was the only algorithm that was able to recover the

statistically significant differences between the ACLT and control groups even with the most dramatic sparsity level of 44 projections (17% of original data). Furthermore, falsely significant differences were not observed with the DART reconstructions, but they were present in Tb.N using CGLS and TV reconstructions at the two lowest sparsity levels. In the case of Tb.Th, however, some significance using a two-tailed Mann-Whitney test was observed for the two lowest sparsity levels with DART, while in the reference analysis Tb.Th was only significant with the one-tailed test.

A similar trend was seen when percentual errors of QBM parameters relative to the reference parameters obtained with FDK were analyzed. When DART was used in the reconstruction, the error was maximally at  $\sim 8\%$  (EF). In contrast, up to  $\sim 40\%$  errors were observed in Tb.Sp using the TV method and  $\sim 30\%$  with CGLS in the sparsest case, while DART remained below 5%.

Most likely, DART owes its excellent performance to the discrete a priori information it incorporates in the reconstruction. Being able to classify voxels into bone or soft tissue during the iterative process appears to make delineation of the tissue interface more exact. This in turn is advantageous in the segmentation process as ROI delineation is easier with sharp borders. Very few, if any, other tomographic reconstruction algorithms implement regularization in this way. On the other hand, the discreteness of the reconstruction would not be a desirable quality when the image is visualized without a segmentation procedure. In clinical imaging, this might not be feasible at least with only two grayscale bins. Nevertheless, the applicability of DART for purely quantitative studies was demonstrated in Study I.

In contrast, CGLS does not imply such restrictions. It is a very basic optimization scheme for large-scale linear systems. TV on the other hand is a classical regularization technique and is widely used in tomography problems. TV implies the resulting image is smooth, which was an undesired property due to the segmentation step in the analysis of Study I. With these considerations in mind, the problem layout of Study I might have been inherently biased towards DART.

### **8.2.3 *Imaging time considerations***

Using the full data, the  $\mu$ CT protocol used in Study I took roughly 17.5 minutes per sample (260 projections, two projections per view, and 2016 ms per projection). This is considered relatively fast for  $\mu$ CT, as high-resolution scans can take several hours per sample. Clearly, time can be a limiting factor in research where  $\mu$ CT is used in imaging. Using only half of the projection data, for example, would roughly halve the needed time for imaging. In contrast, CGLS took  $\sim 45$  seconds, TV took  $\sim 17$  minutes,

and DART took ~18 minutes to reconstruct the image with 130 projections with the equipment available at the time of study. Thus, TV and DART took twice the time to reconstruct that it would have taken to acquire half of the projections. However, the reconstruction time is restricted by the amount of available GPU memory, and only an 8 GB GPU was available for Study I. For Study III, the GPU was upgraded to 24 GB. With modern GPUs, the reconstruction times would be much less and would be at more tolerable levels.

Implementing IR as an option in  $\mu$ CT scanners could be feasible if a sufficiently good GPU was installed in the reconstruction unit. Saving time in the imaging process would mean more samples could be imaged in the same time. Many  $\mu$ CT studies are subject to statistical inaccuracies because of the limited sample size. If  $\mu$ CT was not a bottleneck in the data acquisition workflow, perhaps larger sample sizes could be used for statistical assertiveness. Imaging time is not the sole concern, however, as there are ethical considerations in animal studies that must be taken into account as well. Another benefit of a shorter scan time is that the risk of motion artifacts can be minimized. Furthermore, especially in long high-resolution scans, adverse radiation-induced effects in the tissues can become an issue that could be alleviated with lower doses.

## **8.3 Texture analysis classification of lumbar MRI (II)**

### ***8.3.1 Added value of textural information***

Classification based on semantic and demographic features yielded inferior results with the best ROC-AUC of 0.64 in the upper lumbar levels (Table 13, Fig. 25). Incorporating textural features in the analysis improved the results greatly, with 0.91 ROC-AUC in the lower IVDs (Table 13, Fig. 25). These results suggest there are features in lumbar MRI that can offer complementary information to the degeneration seen by eye. In future, TA could be used as an additional tool for the clinician interpreting MRI.

### ***8.3.2 Classification outcome***

The outcome of the classification in terms of false positives and false negatives can be further discussed by examining the confusion matrix -derived scores in Table 13. For the best-performing ROI, the two lowest IVDs, the ROC-AUC of 0.91 suggests a good overall classification result, which is backed up by the accuracy of 83%. An NPV score of 94% means that the negatives were identified excellently, thus only a small number of false negatives were present in the test set. A precision score of 56% indicates a larger

proportion of false positives, which is also indicated by lower sensitivity (82%) and specificity (83%) scores than NPV, as NPV reveals the number of true negatives is high and the number of false negatives is low.

False positives were expected in the classification outcome due to the nature of the problem itself. It is well-known that MRI-related degeneration can be observed in completely asymptomatic subjects (Boden et al., 1990; Boos et al., 2000). A longitudinal study suggests up to 87% of asymptomatic cases could be false positives based on MRI readings solely (Jarvik et al., 2001). This would translate to 13% specificity which is much lower than the 83% obtained with TA. If 87% of the 408 asymptomatic subjects had been false positives, the number of false positives would have been 355 and the precision score would be 23% if all the symptomatic cases were classified correctly. Thus even in terms of the worst score, the classification outcome was much better than if classification had been done based on MRI degeneration. This is backed up by the baseline classification performed with Modic and Pfirrmann grades in Table 13.

### **8.3.3 Diagnostic interpretations**

As the ROI containing the two lowest IVDs outperformed the rest in terms of most classification metrics, it suggests there are regional differences in pain manifestation. Interestingly, the results obtained with the three uppermost IVDs were worse than the results from the vertebral ROIs. The results between the vertebral ROIs were similar, with the two lowest vertebrae slightly outperforming the three uppermost vertebrae.

The role of IVDs in pain manifestation is discussed in the literature. A strong genetic correlation has been found between IVDs and manifestation of LBP (Freidin et al., 2019). In addition, two different phenotypes of IVD degeneration in the upper and lower lumbar levels have been found that originate from annulus and endplate-related injuries (Adams & Dolan, 2012; Y. Li et al., 2016). The anatomical dependence on the classification outcome found in Study II could be interpreted as complementary evidence for these findings, although no phenotyping was done in the study. Nevertheless, the large variation in results and clear superiority of the lower IVD ROI suggests such a distinction might exist.

An open question remains whether the change in texture seen in Study II is only a biomarker or in fact a part of a disease mechanism. More studies connecting carefully selected textural features with biological and biochemical data are certainly needed to make interpretations on this. Furthermore, it would be extremely interesting to conduct a longitudinal study to see if any treatment options can in fact reverse the observed change in the textural features.

### **8.3.4 Robustness to non-specific sources of pain**

In the two cases formulated for non-specific sources of pain, the classification results were similar to the original classification performed on the full data (Table 15, Fig. 26). Interestingly, the results improved slightly when the subjects who experienced symptoms of sciatica (radiating pain in the leg) were removed from the data. With the additional removal of large Modic changes, the results for the IVDs were worsened while the results for the vertebrae were improved slightly. Modic changes are related to the MRI signal intensity in the vertebral endplates. As the number of removed cases was low (15 cases), in the already small dataset the extreme Modic changes could have been outliers adversely affecting the classification. The interpretability of these results is hindered by the low sample size in the symptomatic group, as it was merely 69 for the first case and 54 for the second case. However, as the classification results did not considerably worsen in either cases, it could be interpreted that the developed model was robust to non-specific pain as well.

### **8.3.5 Image reconstruction in the age of deep learning**

In Studies I and II, data analysis was done purely quantitatively. In such applications, it could be questioned whether or not the process of image reconstruction is required as an intermediate data processing step at all (Chung, Kalpathy-Cramer, Knopp, & Jaffray, 2021). Certainly, when medical images are viewed by medical professionals, reconstruction is needed to process the data to an interpretable form, but in the case of using machines as image interpreters it might not be necessary. This is an extremely interesting novel ML concept, as removing the image reconstruction step could streamline the radiological data workflow accelerating data processing speed and removing estimation errors introduced by the image reconstruction algorithm. Several studies have already been conducted on CT sinograms, performing body part identification and intracranial hemorrhage detection (H. Lee et al., 2019), vessel detection and luminal cross-section estimation (De Man et al., 2019), and pulmonary nodule detection (Gao, Tan, Liang, Li, & Huo, 2019), with similar or better results as by using reconstructed images.

In the age of deep learning, such tools might prove useful and be adopted in clinical practice as well. Such development calls for improved visualization tools if the traditional cross-sectional images are not required anymore. In the end, representing imaging data in an interpretable way will be required for diagnosis and clinical decision-making, as raw data in the sinogram or k-space domain is hardly interpretable by eye.

## **8.4 Deep learning interior computed tomography angiography (III)**

### **8.4.1 *Image quality and the potential for extended field-of-view imaging***

The truncation artifact was removed by all of the studied algorithms except for TS-FBP and TS-TV, although for TS-TV the artifact was greatly reduced as well (Fig. 27). In the analyzed line profile, only minor differences were observed between the DL methods and ES-ADT (Fig. 28). Overall, in the original FOV, ES-ADT, FBPCovNet, ES-UNet, and DGAN resulted in good image quality both visually and in terms of the numerical analysis. TS-FBP was not expected to perform well, as it was only included to set a baseline result for the scenario where no effort would be made to correct for the problems present in interior reconstruction. TS-TV, although more sophisticated than TS-FBP, inherently produces smooth images due to the formulation of the total variation prior. For this reason, TS-TV lagged behind the other algorithms in performance.

The best numerical image quality was obtained by the DGAN method for the original FOV of 96 voxel radius and for all of the extended FOVs (Table 17). The other DL methods also exhibited good results. While a 50% FOV extension seems unrealistic for any clinical applications, the potential for extending the reconstruction FOV slightly by 10-20% as demonstrated in Fig. 29 could allow better dose-efficiency in *e.g.* cardiac, dentomaxillofacial, and extremity scans. The image quality could also potentially be improved *e.g.* when scanning obese patients that do not fit into the reconstruction FOV.

Subtle artifacts in the extended areas can be seen even in the DL reconstructions (Fig. 29). In the case of DGAN, the artifact highlighted in the figure originated from the last steps where the original truncated sinogram is superimposed on the extended sinogram. In some images, this was found to generate a subtle ring where the original FOV limits were located. Several filtering techniques were attempted, but while they removed the artifact, they also affected the spatial resolution of the reconstructions negatively. Removal of this artifact is a task for future studies.

### **8.4.2 *Similar studies in the literature***

Recently, another method for pre-processing truncated CT data with the Pix2Pix GAN architecture was published (Huang et al., 2021). The relevant differences to Study III are that the method processes CBCT data, so it is a 3D method, but optimization is performed only in the projection space. DGAN additionally improves the reconstruction obtained from the extended sinogram. Additionally, more severe truncation was simulated, and an extended FOV analysis was performed in Study III.

### **8.4.3 Choice of model architecture**

The process of developing the reconstruction workflow for Study III was initiated by studying U-Net based models. As an intermediate result, the ES-UNet method was reported in a conference proceeding (Ketola et al., 2021). The premise for Study III was indeed the idea of exploiting DL to produce more precise and realistic sinogram extensions than with an ADT-type methodology based on square root fits.

After obtaining encouraging results with ES-UNet, the method was extended to a GAN-based sinogram extension. GANs are known for their ability to generate extremely realistic images, for example from descriptive text input (H. Zhang et al., 2017). Finally, the Pix2Pix conditional GAN architecture was chosen. In the original publication describing Pix2Pix, several image-to-image transformations were carried out, including synthesizing realistic RGB images from drawn contours only (Isola et al., 2017). As truncated sinograms have partial data in input images much like the examples described in the Pix2Pix article, it was hypothesized to apply well to the interior reconstruction problem as well.

Both the sinogram extension and improvement of TS-FBP reconstructions were briefly investigated using Pix2Pix models, but they offered no substantial improvement in terms of numerical image quality to their U-Net counterparts, ES-UNet and FBPCnvNet. Thus, the final reconstruction workflow nicknamed DGAN consisted of two Pix2Pix models for optimization in both sinogram and reconstruction spaces, and a final forward- and backprojection step where the true measured data was superimposed in the middle of the processed sinogram.

Using U-Nets instead of GANs in the optimization would probably result in similar numerical image quality. However, the observed value of using GANs was twofold: (1) a minor improvement in RMSE, PSNR, and SSIM, and (2) the generation of more realistic image appearance in the extended FOV. In other words, U-Net based methods seemed to visually produce artifacts and anatomical distortions in the extended FOV, whereas the GANs were able to recover more realistic structures through the use of the discriminator. Granted, the visual inspection was based only on a small subset of the full test data and radiologists were not used for grading the image quality, as the clinical image quality was considered beyond the scope of this proof-of-concept study.

### **8.4.4 Challenges in training generative adversarial networks**

In the DL community, GANs are notorious for being difficult to train. The author echoes these thoughts. The difficulty arises from the employment of two networks competing

against each other in a zero-sum game. Improvement in one network results in an adversarial effect in the other. In addition, training individual models for the training data of 10,000 images took approximately 50 hours per model.

One aspect of GAN training is that the generator loss should converge smoothly towards zero, as in essence it is a measure of image similarity between the training samples and their respective ground truth. The other aspect is that the discriminator should be able to distinguish the ground truth from generated images. While convergence of the generator is inherently important in image-to-image networks, disregarding the discriminator can quickly lead to over-classification as either true or fake, offering no value to training the generator. On the other hand, training the discriminator too well will result in divergence of generator loss.

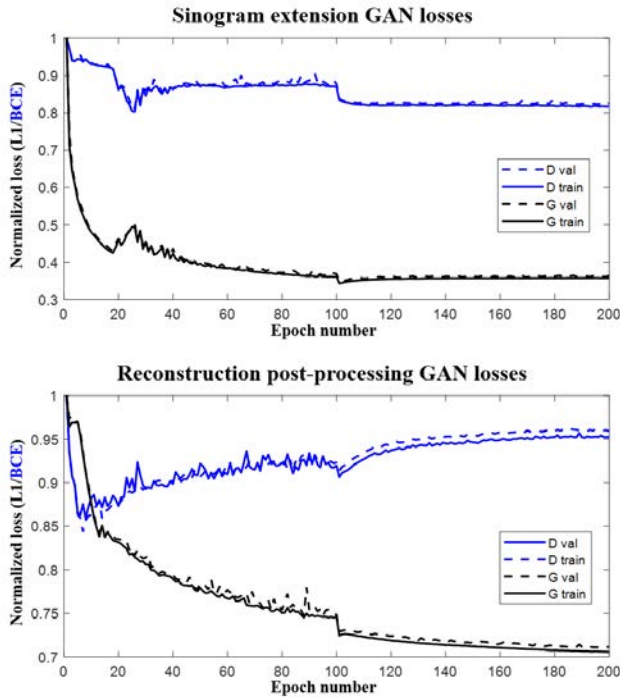
Due to the challenges present in training DL models and GANs in particular, several strategies were used based on empirical evidence in the literature and media.

### *Learning rate decay*

A well-known procedure for training DL networks is to tune the learning rate of the optimization algorithm during training. Usually, a larger learning rate is chosen at the beginning of the training, and it is reduced by a factor of ten in regular intervals. In Study III, the learning rate was reduced by a factor of 10 at 100 epochs. In Fig. 30, a jump in the loss of the generator can be observed as the learning rate is reduced. This effect can be intuitively explained by imagining a convex surface with a minimum. If the step size of the gradient optimization is too large, the optimization step might go past the minimum. On the other hand, if the initial step size is too small, it might take very long for the algorithm to reach the minimum or it may stop at earlier local minima. However, if the desired minimum is first reached by a coarser search, and afterwards the step size is decreased, convergence can improve.

### *Label smoothing*

As described above, one problem of training GANs is that the discriminator might become over-confident in classifying the samples, resulting in over-classification into either real or fake classes (for example, predicting every sample as fake). Label smoothing introduces slight uncertainty in the classification by assigning *e.g.* values of 0.1 and 0.9 to true and fake classes otherwise represented as 0 and 1, like was done in Study III. Introducing this uncertainty to the training process resulted in substantial improvements in the training outcome.



**Fig. 30.** Normalized training curves of the GANs in Study III. Training and validation losses for the generator (G) and discriminator (D) are plotted for the sinogram extension GAN and reconstruction post-processing GAN. The learning rate was adjusted at 100 epochs. (Figure modified from Publication III).

### *Two time-scale update rule*

The two time-scale update rule, *i.e.* using a higher learning rate on the discriminator, was the final training strategy implemented in Study III. This resulted in larger step sizes and suboptimal convergence of the discriminator, giving bias to the convergence of the generator. In the final model, an initial learning rate of 0.04 was assigned to the discriminator and 0.01 to the generator. Doing so improved the generator convergence as the discriminator was otherwise too quick to learn the classification in the initial epochs.

## **8.5 Added value for current imaging practices**

### **8.5.1 Improved efficiency in preclinical animal studies**

Research on many disease models includes preclinical  $\mu$ CT imaging as part of the workflow. As the imaging times are long, the number of samples that can be imaged in  $\mu$ CT facilities can become a limiting factor in the statistical credibility of study outcomes. The author sees utilizing IR or DL in  $\mu$ CT reconstruction as a possibility to increase the time-efficiency of  $\mu$ CT while keeping the image quality good enough for quantitative studies. Naturally, the radiation dose inflicted on the samples would be greatly reduced as well, which would be beneficial for *in vivo* studies and high-resolution longitudinal *ex vivo* studies. In Study I, relatively extreme cases of projection data downsampling were presented. Perhaps, a less dramatic downsampling factor could be utilized in commercial products to achieve advantageous results without compromising qualitative image quality.

### **8.5.2 The potential to reduce radiation dose in cardiac imaging**

While radiation dose measurements were not performed in Study III, it is clear that reducing exposure by measuring lower-quality data than normally results in higher dose-efficiency. With IR, dose reductions ranging from 23% to 75% compared to FBP have been estimated in the literature (Willeminck, de Jong, et al., 2013; Willeminck, Leiner, et al., 2013). With DL, the effect could potentially be even more dramatic (G. Wang et al., 2020). Specifically for cardiac interior tomography, a recent study employing photon counting technology and IR estimated up to  $\sim 61\%$  dose reductions on adjacent organs based on dose measurements (Juntunen et al., 2020). While current CT systems are well developed in terms of dose efficiency, the popularity and demand for CT has grown to such a high level that most of the estimated population dose from X-ray imaging comes from CT (Bly et al., 2020). Thus, it remains important to develop new low-dose imaging techniques.

### **8.5.3 Promoting the use of medical artificial intelligence**

While impressive results have been reported in many fields of medical AI, the general atmosphere regarding AI-enhanced methods in the clinic remains cautious. Studies II and III aimed to apply ML and DL methods on concrete use-cases to show the applicability of the methodology to real imaging problems. In the long run, publishing high-quality medical AI research in large quantities will hopefully help to alleviate

the concerns of the public. The popularization of science via social media, blog texts, podcasts, television, etc. is also an important aspect in spreading knowledge and quashing false rumors and dystopic visions of future AI.

As a reconstruction technique, the power of DL is partly due to the fact that it does not require any pre-set regularization criteria. Rather, all of the prior information available in the training data can be learned in the filters of the network. This has positive and negative effects. When no strict implications for regularization are set, some of the adverse effects of regularization such as TV blurring can be avoided. However, the ‘black box’ nature of CNNs can be problematic when no mathematical grounds for regularization are defined. This can lead to problems in clinical validation of CNN-based reconstruction methods. Some of the more advanced CT reconstruction algorithms combine the robustness of DL with well-known mathematical properties of IR *e.g.* by unrolling iterative algorithms and replacing operations with learnable filters (G. Wang, 2016; G. Wang et al., 2020, 2018).

#### **8.5.4 Projection to other imaging modalities**

The image reconstruction investigated in Studies I and III were developed for CT imaging. However, while the physics and specific implementation details might differ in other modalities, the studied techniques are essentially universal. For example, nuclear medicine modalities such as single-photon emission computed tomography (SPECT) and positron emission tomography (PET) have employed IR methods for decades, as the imaging systems capability to measure good quality data are limited, resulting in severe ill-posedness of the reconstruction inverse problems. In MRI, parallel imaging methods utilizing CS methodology in reconstruction have been under heavy research (Jaspan, Fleysler, & Lipton, 2015). Other non-ionizing methods such as ultrasound and electrical impedance tomography also employ these methods (Hua, Woo, Webster, & Tompkins, 1991; Liebgott, Basarab, Kouame, Bernard, & Friboulet, 2012).

Specifically, DL-based reconstruction approaches can be seen as inherently universal. In essence, everything a supervised DL reconstruction model needs is a well annotated comprehensive training dataset and a robust CNN architecture developed with the problem setting in mind. A good starting point for DL reconstruction research can be achieved with models that exist in the literature, and alterations based on the physics or available *a priori* information can be used to improve the outcome. This mindset was the premise for Study III. Specifically in the area of interior CT, the reconstruction method developed for interior CTA could be applicable to other CT protocols that could benefit from interior scanning, such as dentomaxillofacial and extremity imaging.

Likewise, ML or DL -assisted diagnostics such as the TA method for LBP characterization developed in Study II can be expanded to other modalities as well. Indeed, an extensive number of computer-aided diagnostics models and radiomics-enhanced data mining software have been developed in the literature for a plethora of applications outside MRI and LBP identification.

## **8.6 Research outcomes**

The aim of Study I was to investigate if a quantitative animal model could be analyzed by including only a limited subset of projection data in the reconstruction. From a purely quantitative perspective, the amount of projection data could be reduced to 17% when the reconstruction was performed with DART. However, the image quality of the reconstructions prior to segmentation were not applicable for qualitative analyses. Even with a less dramatic level of sparsity, advantages in imaging time and radiation dose could be achieved while keeping the image quality at an acceptable level.

In Study II, the aim was to develop a predictive model for LBP using lumbar MRI as input. The developed model was able to classify data with good accuracy, and was also robust to non-specific sources of LBP, but the classification was binary using a conservative division of symptomatic and asymptomatic subjects leaving mildly symptomatic people out. Furthermore, an anatomical dependence on classification accuracy was found, with IVDs L4-L5 and L5-S1 being most sensitive to LBP classification.

Finally, in Study III the aim was to develop a DL-based reconstruction technique for interior tomography capable of producing good image quality even in this mathematically challenging reconstruction task. The DGAN model developed in the study yielded excellent results and was also able to recover anatomical features from outside the imaged FOV.

## **8.7 Limitations**

In Study I, FDK was used as the reference data. However, applying FDK reconstruction results in geometrical distortions that increase the further slices are from the isocenter. As such, an FDK reconstruction cannot be called a true 'ground truth'. FDK is, however, the most commonly used reconstruction method and implemented routinely in all commercial  $\mu$ CT scanners. Furthermore, the imaging geometry modeled for the reconstruction process was inaccurate as the X-ray source was estimated as an infinitesimal point source whereas in reality it has a finite diameter.

In Study II, a binary pain target variable was defined based on a conservative definition of asymptomatic subjects, excluding subjects with mild pain symptoms from the analysis. Because the data were from a birth cohort, all subjects were of a similar age (46-48 years old). Therefore, the results cannot be projected onto the population at large as very specific inclusion criteria were met by the analyzed population subset. The added value of ML could have been further demonstrated by including a traditional (non-ML) regression analysis. Further anatomical information could have been extracted by including other spinal structures such as apophyseal joints and musculature instead of focusing solely on IVDs and vertebrae. Only the mid-sagittal plane was included in the TA, and including volumetric data with more MRI slices would have allowed the inclusion of posterolateral lesions in the IVD annulus in the data, for example. Furthermore, intra-IVD analysis comparing the nucleus pulposus and annulus fibrosus would have yielded more specific information about the IVDs. However, including a larger MRI volume or additional ROIs would have increased the computational costs remarkably, so it was left as an area for future investigation.

In Study III, the main limitation was that the study was based on simulations, as raw data in the CT scanner could not be accessed. In the forward model of the simulation, X-ray radiation was assumed to be monochromatic and originating from a point source. Axial simulations were done for each slice instead of employing a helical imaging geometry and interpolating the data. Challenges in training CNNs can lead to reproducibility issues, making a potential clinical validation of DL methods difficult. Finally, the voxel size of the slices was relatively large, but higher-resolution CTA images in the available data did not include the torso to simulate the interior geometry.

Lastly, while studies II and III both used a separate test set to account for over-fitting that were not used for training the models, those test sets were subsets of the same cohort or data collection the training set belonged to. It would be impactful to also characterize the model performance on completely unrelated and independent data with different statistical characteristics.

## **8.8 Directions for future research**

Regarding LBP predictions from lumbar MRIs, several studies are planned for future investigations. A DL pipeline capable of analyzing a 3D volume containing all the spinal structures with a CNN architecture is in development. Inclusion of more MRI contrasts as image channels is also an interesting idea to investigate. This model will be trained for non-binary pain classification as well. The output of the CNN could output 'heat maps' directing the clinician's attention to areas in the spine the network thinks are

important in terms of pain manifestation. Furthermore, including genomic data and more precise differentiation of anatomies that are likely to contribute to pain symptoms made by a physiatrist in radiomics analysis would be an interesting research direction.

In DL reconstruction research, future steps include applying the developed methodology in other imaging protocols such as dentomaxillofacial and extremity scans. For clinical assertiveness, it would be interesting to explore the limits of the method with real projection data acquired from a scanner instead of simulations. Preliminary tests applying DL reconstruction on  $\mu$ CT data have also been done. Furthermore, research projects applying DL in multi energy CT reconstruction, metal artifact reduction techniques, and MRI reconstruction and fingerprinting methods have been planned. In these studies, clinical image quality validation with the help of radiologists with expertise in the image modality in question would be a great asset.

## 9 Conclusions

This doctoral dissertation can be summarized as follows:

1. IR algorithms have the potential to substantially reduce the imaging time and radiation dose associated with preclinical quantitative  $\mu$ CT studies.
2. TA and ML can be used in predicting LBP from  $T_2$ -weighted lumbar MRI, even in the case where no explanatory symptoms are present. IVDs L4-L5 and L5-S1 were the most sensitive to this classification when different ROIs were compared, suggesting that anatomical differences in pain manifestation can be seen in MRI.
3. GANs can be used to generate missing data in truncated sinograms and to substantially improve interior reconstruction image quality when compared to non-DL based methods. Furthermore, the reconstruction FOV can potentially be increased by 20% beyond the limits of the sampling FOV.



## References

- Abel, R. L., Parfitt, S., Ashton, N., Lewis, S. G., Scott, B., & Stringer, C. (2011). Digital preservation and dissemination of ancient lithic technology with modern micro-CT. *Computers & Graphics*, 35(4), 878–884.
- Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, 11(7), 36–42.
- Adams, M. A., & Dolan, P. (2012). Intervertebral disc degeneration: evidence for two distinct phenotypes. *Journal of anatomy*, 221(6), 497–506.
- Adler, J., & Öktem, O. (2018). Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6), 1322–1332.
- Aggarwal, H. K., Mani, M. P., & Jacob, M. (2018). MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2), 394–405.
- Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037–2041.
- Badea, C., Drangova, M., Holdsworth, D. W., & Johnson, G. (2008). In vivo small-animal imaging using micro-CT and digital subtraction angiography. *Physics in Medicine & Biology*, 53(19), R319–R350.
- Balagué, F., Mannion, A. F., Pellisé, F., & Cedraschi, C. (2012). Non-specific low back pain. *The Lancet*, 379(9814), 482–491.
- Barzilai, J., & Borwein, J. M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1), 141–148.
- Batenburg, K. J., & Sijbers, J. (2011). DART: a practical reconstruction algorithm for discrete tomography. *IEEE Transactions on Image Processing*, 20(9), 2542–2553.
- Beister, M., Kolditz, D., & Kalender, W. A. (2012). Iterative reconstruction methods in X-ray CT. *Physica Medica*, 28(2), 94–108.
- Bellolio, M. F., Heien, H. C., Sangaralingham, L. R., Jeffery, M. M., Campbell, R. L., Cabrera, D., . . . Hess, E. P. (2017). Increased computed tomography utilization in the emergency department and its association with hospital admission. *Western Journal of Emergency Medicine*, 18(5), 835.
- Bian, J., Siewerdsen, J. H., Han, X., Sidky, E. Y., Prince, J. L., Pelizzari, C. A., & Pan, X. (2010). Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Physics in Medicine & Biology*, 55(22), 6575.
- Bly, R., Järvinen, H., Kaijaluoto, S., & Ruonala, V. (2020). Contemporary collective

- effective dose to the population from X-ray and nuclear medicine examinations—changes over last 10 years in Finland. *Radiation Protection Dosimetry*, 189(3), 318–322.
- Boas, F. E., & Fleischmann, D. (2012). CT artifacts: causes and reduction techniques. *Imaging Med*, 4(2), 229–240.
- Bobinac, D., Marinovic, M., Bazdulj, E., Cvijanovic, O., Celic, T., Maric, I., . . . Cicvaric, T. (2013). Microstructural alterations of femoral head articular cartilage and subchondral bone in osteoarthritis and osteoporosis. *Osteoarthritis and Cartilage*, 21(11), 1724–1730.
- Boden, S. D., McCowin, P., Davis, D., Dina, T., Mark, A., & Wiesel, S. (1990). Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *The Journal of Bone and Joint Surgery. American volume*, 72(8), 1178–1184.
- Boerckel, J. D., Mason, D. E., McDermott, A. M., & Alsberg, E. (2014). Microcomputed tomography: approaches and applications in bioengineering. *Stem Cell Research & Therapy*, 5, 144.
- Bonse, U., & Busch, F. (1996). X-ray computed microtomography ( $\mu$ CT) using synchrotron radiation (SR). *Progress in Biophysics and Molecular Biology*, 65(1-2), 133–169.
- Boos, N., Semmer, N., Elfering, A., Schade, V., Gal, I., Zanetti, M., . . . Main, C. J. (2000). Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. *Spine*, 25(12), 1484–1492.
- Boswell, M. V., Singh, V., Staats, P. S., & Hirsch, J. A. (2003). Accuracy of precision diagnostic blocks in the diagnosis of chronic spinal pain of facet or zygapophysial joint origin: A systematic review. *Pain Physician*, 6(4), 449–456.
- Bouxein, M. L., Boyd, S. K., Christiansen, B. A., Guldberg, R. E., Jepsen, K. J., & Müller, R. (2010). Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. *Journal of Bone and Mineral Research*, 25(7), 1468–1486.
- Brault, R., Germaneau, A., Dupré, J.-C., Doumalin, P., Mistou, S., & Fazzini, M. (2013). In-situ analysis of laminated composite materials by X-ray micro-computed tomography and digital volume correlation. *Experimental Mechanics*, 53(7), 1143–1151.
- Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., , Thompson, M. R., & Venkatesan, R. (2014). *Magnetic resonance imaging: physical principles and sequence design*. Wiley-liss New York:.

- Bubba, T. A., Labate, D., Zanghirati, G., & Bonettini, S. (2018). Shearlet-based regularized reconstruction in region-of-interest computed tomography. *Mathematical Modelling of Natural Phenomena*, 13(4), 34.
- Buckwalter, J., & Mankin, H. (1998). Articular cartilage: degeneration and osteoarthritis, repair, regeneration, and transplantation. *Instructional Course Lectures*, 47, 487–504.
- Bushberg, J. T., Seibert, A. J., Levinholdt, E. M., & Boone, J. M. (2012). *The essential physics of medical imaging, 3rd edition*. Lippincott Williams & Wilkins.
- Calvetti, D., Morigi, S., Reichel, L., & Sgallari, F. (2000). Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, 123, 423–446.
- Carragee, E. J. (2005). Persistent low back pain. *New England Journal of Medicine*, 352(18), 1891–1898.
- Castellano, G., Bonilha, L., Li, L., & Cendes, F. (2004). Texture analysis of medical images. *Clinical Radiology*, 59(12), 1061–1069.
- Chang, T., & Kuo, C.-C. (1993). Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing*, 2(4), 429–441.
- Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., . . . Wang, G. (2018). LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE Transactions on Medical Imaging*, 37(6), 1333–1347.
- Chou, D., Samartzis, D., Bellabarba, C., Patel, A., Luk, K. D., Kisser, J. M. S., & Skelly, A. C. (2011). Degenerative magnetic resonance imaging changes in patients with chronic low back pain: a systematic review. *Spine*, 36, S43–S53.
- Chou, R., Fu, R., Carrino, J. A., & Deyo, R. A. (2009). Imaging strategies for low-back pain: systematic review and meta-analysis. *The Lancet*, 373(9662), 463–472.
- Chung, C., Kalpathy-Cramer, J., Knopp, M. V., & Jaffray, D. A. (2021). In the era of deep learning, why reconstruct an image at all? *Journal of the American College of Radiology*, 18(1, Part B), 170-173. (Special Issue: Provocative Issue)
- Clackdoyle, R., Noo, F., Guo, J., & Roberts, J. A. (2004). Quantitative reconstruction from truncated projections in classical tomography. *IEEE Transactions on Nuclear Science*, 51(5), 2570–2578.
- Cnudde, V., Masschaele, B., Dierick, M., Vlassenbroeck, J., Van Hoorebeke, L., & Jacobs, P. (2006). Recent progress in X-ray CT as a geosciences tool. *Applied Geochemistry*, 21(5), 826–832.
- Cong, W., Yang, J., & Wang, G. (2012). Differential phase-contrast interior tomography. *Physics in Medicine & Biology*, 57(10), 2905.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A.

- (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65.
- Dagenais, S., Caro, J., & Haldeman, S. (2008). A systematic review of low back pain cost of illness studies in the United States and internationally. *The Spine Journal*, 8(1), 8–20.
- De Man, Q., Haneda, E., Claus, B., Fitzgerald, P., De Man, B., Qian, G., . . . Wang, G. (2019). A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms. *Medical Physics*, 46(12), e790–e800.
- Dong, X., Vekhande, S., & Cao, G. (2019). Sinogram interpolation for sparse-view micro-CT with deep learning neural network. In *Medical Imaging 2019: Physics of Medical Imaging* (Vol. 10948, p. 109482O).
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289–1306.
- Doube, M. (2015). The ellipsoid factor for quantification of rods, plates, and intermediate forms in 3D geometries. *Frontiers in Endocrinology*, 6, 15.
- Doube, M., Kłosowski, M. M., Arganda-Carreras, I., Cordelières, F. P., Dougherty, R. P., Jackson, J. S., . . . Shefelbine, S. J. (2010). BoneJ: free and extensible bone image analysis in ImageJ. *Bone*, 47(6), 1076–1079.
- Dowsett, D. J., Kenny, P. A., & Johnston, R. E. (2006). *Physics of diagnostic imaging, the second edition*. Hodder Education.
- Egermann, M., Goldhahn, J., & Schneider, E. (2005). Animal models for fracture treatment in osteoporosis. *Osteoporosis International*, 16, S129–S138.
- Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *RadioGraphics*, 37(2), 505–515.
- Faridani, A., Finch, D. V., Ritman, E. L., & Smith, K. T. (1997). Local tomography II. *SIAM Journal on Applied Mathematics*, 57(4), 1095–1127.
- Faridani, A., Ritman, E. L., & Smith, K. T. (1992). Local tomography. *SIAM Journal on Applied Mathematics*, 52(2), 459–484.
- Feldkamp, L. A., Davis, L. C., & Kress, J. W. (1984). Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6), 612–619.
- Finnilä, M. A., Thevenot, J., Aho, O.-M., Tiitu, V., Rautiainen, J., Kauppinen, S., . . . others (2017). Association between subchondral bone structure and osteoarthritis histopathological grade. *Journal of Orthopaedic Research*, 35(4), 785–792.
- Florea, C., Malo, M., Rautiainen, J., Mäkelä, J., Fick, J., Nieminen, M., . . . Korhonen, R. (2015). Alterations in subchondral bone plate, trabecular bone and articular cartilage properties of rabbit femoral condyles at 4 weeks after anterior cruciate

- ligament transection. *Osteoarthritis and Cartilage*, 23(3), 414–422.
- Freidin, M. B., Tsepilov, Y. A., Palmer, M., Karssen, L. C., Suri, P., Aulchenko, Y. S., . . . others (2019). Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. *Pain*, 160(6), 1361–1373.
- Gao, Y., Tan, J., Liang, Z., Li, L., & Huo, Y. (2019). Improved computer-aided detection of pulmonary nodules via deep learning in the sinogram domain. *Visual Computing for Industry, Biomedicine, and Art*, 2(1), 1–9.
- Ghani, M. U., & Karl, W. C. (2018). Deep learning-based sinogram completion for low-dose CT. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)* (pp. 1–5).
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, 278(2), 563–577.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1) (No. 2). MIT press Cambridge.
- Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5), 1153–1159.
- Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6), 1657–1663.
- Gupta, H., Jin, K. H., Nguyen, H. Q., McCann, M. T., & Unser, M. (2018). CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE Transactions on Medical Imaging*, 37(6), 1440–1453.
- Hämäläinen, K., Harhanen, L., Hauptmann, A., Kallonen, A., Niemi, E., & Siltanen, S. (2014). Total variation regularization for large-scale X-ray tomography. *International Journal of Tomography and Simulation*, 25(1), 1–25.
- Han, Y., & Ye, J. C. (2018). Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE Transactions on Medical Imaging*, 37(6), 1418–1429.
- Han, Y., & Ye, J. C. (2019). One network to solve all ROIs: Deep learning CT for any ROI using differentiated backprojection. *Medical Physics*, 46(12), e855–e872.
- Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4), 561–580.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*(6), 610–621.
- Hartvigsen, J., Hancock, M. J., Kongsted, A., Louw, Q., Ferreira, M. L., Genevay, S., . . . others (2018). What low back pain is and why we need to pay attention. *The*

- Lancet*, 391(10137), 2356–2367.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv preprint arXiv:1706.08500*.
- Holdsworth, D. W., & Thornton, M. M. (2002). Micro-CT in small animal and specimen imaging. *Trends in Biotechnology*, 20(8), S34–S39.
- Hsieh, J., Chao, E., Thibault, J., Grekowitz, B., Horst, A., McOlash, S., & Myers, T. (2004). A novel reconstruction algorithm to extend the CT scan field-of-view. *Medical Physics*, 31(9), 2385–2391.
- Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis - a survey. *Pattern Recognition*, 83, 134–149.
- Hua, P., Woo, E. J., Webster, J. G., & Tompkins, W. J. (1991). Iterative reconstruction methods using regularization and optimal current patterns in electrical impedance tomography. *IEEE Transactions on Medical Imaging*, 10(4), 621–628.
- Huang, Y., Preuhs, A., Manhart, M., Lauritsch, G., & Maier, A. (2021). Data extrapolation from learned prior images for truncation correction in computed tomography. *IEEE Transactions on Medical Imaging*.
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1125–1134).
- Jarvik, J. J., Hollingworth, W., Heagerty, P., Haynor, D. R., & Deyo, R. A. (2001). The longitudinal assessment of imaging and disability of the back (LAIDBack) study: baseline data. *Spine*, 26(10), 1158–1166.
- Jaspan, O. N., Fleysher, R., & Lipton, M. L. (2015). Compressed sensing MRI: a review of the clinical literature. *The British Journal of Radiology*, 88(1056).
- Jensen, M. C., Brant-Zawadzki, M. N., Obuchowski, N., Modic, M. T., Malkasian, D., & Ross, J. S. (1994). Magnetic resonance imaging of the lumbar spine in people without back pain. *New England Journal of Medicine*, 331, 69–73.
- Jin, K. H., McCann, M. T., Froustey, E., & Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9), 4509–4522.
- Juntunen, M. A., Sepponen, P., Korhonen, K., Pohjanen, V.-M., Ketola, J., Kotiaho, A., . . . Inkinen, S. I. (2020). Interior photon counting computed tomography for quantification of coronary artery calcium: pre-clinical phantom study. *Biomedical Physics & Engineering Express*, 6(5), 055011.
- Kader, D., Wardlaw, D., & Smith, F. (2000). Correlation between the MRI changes in

- the lumbar multifidus muscles and leg pain. *Clinical Radiology*, 55, 145–149.
- Kalender, W. A. (2011). *Computed tomography: Fundamentals, system technology, image quality, applications*. John Wiley & Sons.
- Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., & Ye, J. C. (2019). Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Medical Physics*, 46(2), 550–562.
- Kassner, A., & Thornhill, R. (2010). Texture analysis: a review of neurologic MR imaging applications. *American Journal of Neuroradiology*, 31(5), 809–816.
- Ketola, J. H., Heino, H., Juntunen, M. A., Nieminen, M. T., & Inkinen, S. I. (2021). Deep learning-based sinogram extension method for interior computed tomography. In *Medical Imaging 2021: Physics of Medical Imaging* (Vol. 11595, p. 115953Q).
- Kim, S.-Y., Lee, I.-S., Kim, B.-R., Lim, J.-H., Lee, J., Koh, S.-E., . . . Park, S. L. (2012). Magnetic resonance findings of acute severe lower back pain. *Annals of Rehabilitation Medicine*, 36, 47-54.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kjaer, P., Bendix, T., Sorensen, J. S., Korsholm, L., & Leboeuf-Yde, C. (2007). Are MRI-defined fat infiltrations in the multifidus muscles associated with low back pain? *BMC Medicine*, 5(2).
- Klann, E., Quinto, E. T., & Ramlau, R. (2015). Wavelet methods for a weighted sparsity penalty for region of interest tomography. *Inverse Problems*, 31(2), 025001.
- Kolditz, D., Meyer, M., Kyriakou, Y., & Kalender, W. A. (2010). Comparison of extended field-of-view reconstructions in c-arm flat-detector CT using patient size, shape or attenuation information. *Physics in Medicine & Biology*, 56(1), 39.
- Kudo, H., Courdurier, M., Noo, F., Noo, & Defrise, M. (2008). Tiny a priori knowledge solves the interior problem in computed tomography. *Physics in Medicine & Biology*, 53(9), 2207.
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., . . . others (2012). Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9), 1234–1248.
- Kuyinu, E. L., Narayanan, G., Nair, L. S., & Laurencin, C. T. (2016). Animal models of osteoarthritis: classification, update, and measurement of outcomes. *Journal of Orthopaedic Surgery and Research*, 11, 19.
- Kyrieleis, A., Titarenko, V., Ibison, M., Connolley, T., & Withers, P. (2011). Region-of-interest tomography using filtered backprojection: assessing the practical limits. *Journal of Microscopy*, 241(1), 69–82.
- Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., De Jong, E. E., Van Timmeren,

- J., ... Walsh, S. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, *14*, 749–762.
- Lampropoulou-Adamidou, K., Lelovas, P., Karadimas, E. V., Liakou, C., Triantafilopoulos, I. K., Dontas, I., & Papaioannou, N. A. (2014). Useful animal models for the research of osteoarthritis. *European Journal of Orthopaedic Surgery & Traumatology*, *24*, 263–271.
- Laperre, K., Depypere, M., van Gastel, N., Torrekens, S., Moermans, K., Bogaerts, R., ... Carmeliet, G. (2011). Development of micro-CT protocols for in vivo follow-up of mouse bone architecture without major radiation side effects. *Bone*, *49*(4), 613–622.
- Lee, H., Huang, C., Yune, S., Tajmir, S. H., Kim, M., & Do, S. (2019). Machine friendly machine learning: interpretation of computed tomography without image reconstruction. *Scientific reports*, *9*(1), 1–9.
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: General overview. *Korean Journal of Radiology*, *18*(4), 570.
- Li, H., Zhang, H., Tang, Z., & Hu, G. (2008). Micro-computed tomography for small animal imaging: Technological details. *Progress in Natural Science*, *18*(5), 513–521.
- Li, Y., Li, K., Zhang, C., Montoya, J., & Chen, G.-H. (2019). Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. *IEEE Transactions on Medical Imaging*, *38*(10), 2469–2481.
- Li, Y., Samartzis, D., Campbell, D. D., Cherny, S. S., Cheung, K. M., Luk, K. D., ... others (2016). Two subtypes of intervertebral disc degeneration distinguished by large-scale population-based study. *The Spine Journal*, *16*(9), 1079–1089.
- Liebgott, H., Basarab, A., Kouame, D., Bernard, O., & Friboulet, D. (2012). Compressive sensing in medical ultrasound. In *2012 IEEE International Ultrasonics Symposium* (pp. 1–6).
- Liu, Y., Ma, J., Fan, Y., & Liang, Z. (2012). Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in Medicine & Biology*, *57*(23), 7923.
- Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V., & Pickhardt, P. J. (2017). CT texture analysis: definitions, applications, biologic correlates, and challenges. *RadioGraphics*, *37*(5), 1483–1503.
- Luoma, K., Riihimäki, H., Luukkonen, R., Raininko, R., Viikari-Juntura, E., & Lamminen, A. (2000). Low back pain in relation to lumbar disc degeneration. *Spine*,

- 25(4), 487–492.
- Määttä, J. H., Karppinen, J., Paananen, M., Bow, C., Luk, K. D., Cheung, K. M., & Samartzis, D. (2016). Refined phenotyping of Modic changes: imaging biomarkers of prolonged severe low back pain and disability. *Medicine*, 95(22), e3495.
- Määttä, J. H., Wadge, S., MacGregor, A., Karppinen, J., & Williams, F. M. K. (2012). Vertebral endplate (Modic) change is an independent risk factor for episodes of severe and disabling low back pain. *Spine*, 37(14), 1231–1239.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60.
- McCullough, C. H., Chen, G. H., Kalender, W., Leng, S., Samei, E., Taguchi, K., . . . Pettigrew, R. I. (2012). Achieving routine submillisievert CT scanning: report from the summit on management of radiation dose in CT. *Radiology*, 264(2), 567–580.
- McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2017). *MRI: from picture to proton*. Cambridge University Press.
- Metscher, B. D. (2009). MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. *BMC physiology*, 9(11), 11.
- Mettler Jr, F. A. (2019). Medical radiation exposure in the United States: 2006–2016 trends. *Health Physics*, 116(2), 126–128.
- Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Mirmehdi, M. (2008). *Handbook of texture analysis*. Imperial College Press.
- Modic, M. T., Steinberg, P. M., Ross, J. S., Masaryk, T. J., & Carter, J. R. (1988). Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. *Radiology*, 166(1), 193–199.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Mok, F. P., Samartzis, D., Karppinen, J., Fong, D. Y., Luk, K. D., & Cheung, K. M. (2016). Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. *The Spine Journal*, 16, 32–41.
- Mueller, J. L., & Siltanen, S. (2012). *Linear and nonlinear inverse problems with practical applications*. SIAM.

- Müller, R., Kornblith, S., & Hinton, G. (2019). When does label smoothing help? *arXiv preprint arXiv:1906.02629*.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Niu, S., Gao, Y., Bian, Z., Huang, J., Chen, W., Yu, G., . . . Ma, J. (2014). Sparse-view x-ray CT reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12), 2997.
- Ohnesorge, B., Flohr, T., Schwarz, K., Heiken, J., & Bae, K. (2000). Efficient correction for CT image artifacts caused by objects extending outside the scan field of view. *Medical physics*, 27(1), 39–46.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.
- Palenstijn, W. J., Batenburg, K. J., & Sijbers, J. (2011). Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs). *Journal of Structural Biology*, 176(2), 250–253.
- Paulus, M. J., Gleason, S. S., Kennel, S. J., Hunsicker, P. R., & Johnson, D. K. (2000). High resolution X-ray computed tomography: an emerging tool for small animal cancer research. *Neoplasia*, 2(1-2), 62–70.
- Pfirschmann, C. W., Metzendorf, A., Zanetti, M., Hodler, J., & Boos, N. (2001). Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine*, 26(17), 1873–1878.
- Rachner, T. D., Khosla, S., & Hofbauer, L. C. (2011). Osteoporosis: now and the future. *The Lancet*, 377(9773), 1276–1287.
- Radon, J. (1986). On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4), 170–176.
- Rantakallio, P. (1988). The longitudinal study of the Northern Finland Birth Cohort of 1966. *Paediatric and Perinatal Epidemiology*, 2(1), 59–88.
- Ritman, E. L. (2011). Current status of developments and applications of micro-CT. *Annual Review of Biomedical Engineering*, 13, 531–552.
- Ritschl, L., Bergner, F., Fleischmann, C., & Kachelrieß, M. (2011). Improved total variation-based CT image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6), 1545-1561.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241).
- Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4), 259–268.
- Saukkonen, J., Määttä, J., Oura, P., Kyllönen, E., Tervonen, O., Niinimäki, J., . . .

- Karppinen, J. (2020). Association between Modic changes and low back pain in middle age: a northern Finland birth cohort study. *Spine*, 45(19), 1360–1367.
- Schambach, S. J., Bag, S., Schilling, L., Groden, C., & Brockmann, M. A. (2010). Application of micro-CT in small animal imaging. *Methods*, 50(1), 2–13.
- Schambach, S. J., Bag, S., Steil, V., Isaza, C., Schilling, L., Groden, C., & Brockmann, M. A. (2009). Ultrafast high-resolution in vivo volume-CTA of mice cerebral vessels. *Stroke*, 40(4), 1444–1450.
- Schulze, R., Heil, U., Groß, D., Bruellmann, D., Dranischnikow, E., Schwanecke, U., & Schoemer, E. (2011). Artefacts in CBCT: a review. *Dentomaxillofacial Radiology*, 40(5), 265–273.
- Sheehan, N. (2010). Magnetic resonance imaging for low back pain: indications and limitations. *Annals of the Rheumatic Diseases*, 69, 7–11.
- Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248.
- Sidky, E. Y., & Pan, X. (2008). Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17), 4777–4807.
- Simoncini, V., & Szyld, D. B. (2007). Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14(1), 1–59.
- Sniekers, Y. H., Intema, F., Lafeber, F. P., van Osch, G. J., van Leeuwen, J. P., Weinans, H., & Mastbergen, S. C. (2008). A role for subchondral bone changes in the process of osteoarthritis; a micro-CT study of two canine models. *BMC Musculoskeletal Disorders*, 9, 1–11.
- Sourbelle, K., Kachelrieß, M., & Kalender, W. A. (2005). Reconstruction from truncated projections in CT using adaptive detruncation. *European Radiology*, 15(5), 1008–1014.
- Suwa, G., Kono, R. T., Katoh, S., Asfaw, B., & Beyene, Y. (2007). A new species of great ape from the late miocene epoch in ethiopia. *Nature*, 448(7156), 921–924.
- Taguchi, K., Xu, J., Srivastava, S., Tsui, B. M., Cammin, J., & Tang, Q. (2011). Interior region-of-interest reconstruction using a small, nearly piecewise constant subregion. *Medical Physics*, 38(3), 1307–1312.
- Takatalo, J., Karppinen, J., Niinimäki, J., Taimela, S., Mutanen, P., Sequeiros, R. B., ... Tervonen, O. (2012). Association of Modic changes, Schmorl's nodes, spondylolytic defects, high-intensity zone lesions, disc herniations, and radial tears with low back symptom severity among young Finnish adults. *Spine*, 37(14), 1231–1239.

- Takatalo, J., Karppinen, J., Niinimäki, J., Taimela, S., Näyhä, S., Mutanen, P., ... Tervonen, O. (2011). Does lumbar disc degeneration on magnetic resonance imaging associate with low back symptom severity in young Finnish adults? *Spine*, 36(25), 2180–2189.
- Tang, X. (1998). Texture information in run-length matrices. *IEEE Transactions on Image Processing*, 7(11), 1602–1609.
- Tian, Z., Jia, X., Yuan, K., Pan, T., & Jiang, S. B. (2011). Low-dose CT reconstruction via edge-preserving total variation regularization. *Physics in Medicine & Biology*, 56(18), 5949-5967.
- Toyone, T., Takahashi, K., Kitahara, H., Yamagata, M., Murakami, M., & Moriya, H. (1994). Vertebral bone-marrow changes in degenerative lumbar disc disease. an MRI study of 74 patients with low back pain. *The Journal of Bone and Joint Surgery. British volume*, 76-B(5), 757–764.
- Turbell, H. (2001). *Cone-beam reconstruction using filtered backprojection* (Unpublished doctoral dissertation). Linköping University Electronic Press.
- Van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabrovolski, A., ... Sijbers, J. (2016). Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics Express*, 24(22), 25129–25147.
- Van Aarle, W., Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J., & Sijbers, J. (2015). The ASTRA toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157, 35–47.
- Videman, T., Battié, M. C., Gibbons, L. E., Maravilla, K., Manninen, H., & Kaprio, J. (2003). Associations between back pain history and lumbar MRI findings. *Spine*, 28(6), 582–588.
- Videman, T., & Nurminen, M. (2004). The occurrence of anular tears and their relation to lifetime back pain history: a cadaveric study using barium sulfate discography. *Spine*, 29(23), 2668–2676.
- Wang, G. (2016). A perspective on deep imaging. *IEEE Access*, 4, 8914–8924.
- Wang, G., Bresler, Y., & Ntziachristos, V. (2011). Guest editorial compressive sensing for biomedical imaging. *IEEE Transactions on Medical Imaging*, 30(5), 1013–1016.
- Wang, G., Ye, J. C., & De Man, B. (2020). Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2(12), 737–748.
- Wang, G., Ye, J. C., Mueller, K., & Fessler, J. A. (2018). Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging*, 37(6), 1289–1296.
- Wang, G., & Yu, H. (2013). The meaning of interior tomography. *Physics in Medicine & Biology*, 58(16), R161.

- Wang, S. X., Laverty, S., Dumitriu, M., Plaas, A., & Gryn timer, M. D. (2007). The effects of glucosamine hydrochloride on subchondral bone changes in an animal model of osteoarthritis. *Arthritis & Rheumatism*, 56(5), 1537–1548.
- Wang, Y., Videman, T., & Battié, M. C. (2012). ISSLS prize winner: Lumbar vertebral endplate lesions: Associations with disc degeneration and back pain history. *Spine*, 37(17), 1490–1496.
- Ward, J. P., Lee, M., Ye, J. C., & Unser, M. (2015). Interior tomography using 1D generalized total variation. Part I: Mathematical foundation. *SIAM Journal on Imaging Sciences*, 8(1), 226–247.
- Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine learning in medical imaging. *IEEE Signal Processing Magazine*, 27(4), 25–38.
- Willekens, I., Buls, N., Lahoutte, T., Baeyens, L., Vanhove, C., Caveliers, V., . . . de Mey, J. (2010). Evaluation of the radiation dose in micro-CT with optimization of the scan protocol. *Contrast Media & Molecular Imaging*, 5(4), 201–207.
- Willemink, M. J., de Jong, P. A., Leiner, T., de Heer, L. M., Nievelstein, R. A., Budde, R. P., & Schilham, A. M. (2013). Iterative reconstruction techniques for computed tomography part 1: technical principles. *European Radiology*, 23, 1623–1631.
- Willemink, M. J., Leiner, T., de Jong, P. A., de Heer, L. M., Nievelstein, R. A., Schilham, A. M., & Budde, R. P. (2013). Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality. *European Radiology*, 23, 1632–1642.
- Willemink, M. J., & Noël, P. B. (2019). The evolution of image reconstruction for CT — from filtered back projection to artificial intelligence. *European Radiology*, 29, 2185–2195.
- Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, 36(12), 2536–2545.
- Wu, D., Kim, K., & Li, Q. (2019). Computationally efficient deep neural network for computed tomography image reconstruction. *Medical Physics*, 46(11), 4763–4776.
- Xu, Q., Mou, X., Wang, G., Sieren, J., Hoffman, E. A., & Yu, H. (2011). Statistical interior tomography. *IEEE Transactions on Medical Imaging*, 30(5), 1116–1128.
- Yang, J., Yu, H., Jiang, M., & Wang, G. (2010). High-order total variation minimization for interior tomography. *Inverse Problems*, 26(3), 035013.
- Ye, D. H., Buzzard, G. T., Ruby, M., & Bouman, C. A. (2018). Deep back projection for sparse-view CT reconstruction. In *2018 IEEE Global Conference on Signal and*

*Information Processing (GlobalSIP)* (pp. 1–5).

- Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, *58*, 101552.
- Yu, H., & Wang, G. (2009). Compressed sensing based interior tomography. *Physics in Medicine & Biology*, *54*(9), 2791-2805.
- Zamyatin, A. A., & Nakanishi, S. (2007). Extension of the reconstruction field of view and truncation correction using sinogram decomposition. *Medical Physics*, *34*(5), 1593–1604.
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5907–5915).
- Zhang, Z.-M., Li, Z.-C., Jiang, L.-S., Jiang, S.-D., & Dai, L.-Y. (2010). Micro-CT and mechanical evaluation of subchondral trabecular bone structure between postmenopausal women with osteoarthritis and osteoporosis. *Osteoporosis International*, *21*(8), 1383–1390.

## Original publications

- I Ketola JH, Karhula SS, Finnilä MAJ, Korhonen RK, Herzog W, Siltanen S, Nieminen MT & Saarakkala S (2018). Iterative and discrete reconstruction in the evaluation of the rabbit model of osteoarthritis. *Scientific Reports*, 8(1): 12051. DOI: 10.1038/s41598-018-30334-8
- II Ketola JHJ, Inkinen SI, Karppinen J, Niinimäki J, Tervonen O & Nieminen MT (2021). T<sub>2</sub>-weighted magnetic resonance imaging texture as predictor of low back pain: A texture analysis-based classification pipeline to symptomatic and asymptomatic cases. *J Orthop Res*. 2021; 1-11. DOI: 10.1002/jor.24973
- III Ketola JHJ, Heino H, Juntunen MAK, Siltanen S, Nieminen MT & Inkinen SI. Generative adversarial networks improve interior computed tomography reconstruction. *Manuscript submitted for publication*.

Articles I and II have been published under the Creative Commons BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

Original publications are not included in the electronic version of the dissertation.



1632. Tahvonen, Pirita (2021) Appropriateness of radiological examinations exposing to ionizing radiation: the effect of active referral guideline implementation
1633. Turunen, Suvi (2021) Maternal thyroid dysfunction in pregnancy : prevalence and effect on perinatal outcome
1634. Modarress Julin, Mahsa Marella (2021) Physical activity, vertebral size and lumbar Modic changes in midlife
1635. Mustonen, Caius (2021) Cerebral protection in aortic arch surgery with a special reference to Acute Type A Aortic Dissection
1636. Väisänen, Janne (2021) Head and neck cancer: patient characteristics, medical care seeking and long-term prognosis
1637. Salonen, Johanna (2021) Acute exacerbations of fibrosing interstitial lung diseases: clinical features, prognosis and associations of inflammatory cells with disease course
1638. Hirsikangas, Sari (2021) Hoitoon sitoutuminen, siihen yhteydessä olevat tekijät ja palveluohjausintervention vaikutukset terveystalvaeluja paljon käyttävillä asiakkailla
1639. Tokola, Susanna (2021) Optimization of treatment in DLBCL patients
1640. Elsayed, Hassaan (2021) Mapping of  $T_{1\rho}$  dispersion parameters in articular cartilage
1641. Lahti, Anna-Maija (2021) Long-term survival and post-stroke epilepsy after primary intracerebral hemorrhage
1642. Huhtela, Outi (2021) Temporomandibular disorders and bruxism in university students – prevalence, association with psychosocial background factors and effectiveness of applied relaxation
1643. Kananen, Janne (2021) Increased effect of physiological respiratory brain pulsations in focal-onset epilepsy
1644. Rantala, Arja (2021) Pelillinen mobiilisovellusinterventio kouluikäisten lasten päiväkirurgisella hoitopolulla
1645. Kielenniva, Katja (2021) Microbiota of the first stool after birth
1646. Pelli, Ari (2021) Electrocardiogram and diabetes mellitus as predictors of mortality benefit from ICD therapy in primary prophylactic patients
1647. Tohmola, Anniina (2021) Gerontologisen hoitotyön osaamisalueet ja valmistumisvaiheessa olevien sairaanhoitajaopiskelijoiden itsearvioitu gerontologisen hoitotyön osaaminen

Book orders:

Virtual book store

<http://verkkokauppa.juvenesprint.fi>

S E R I E S E D I T O R S

**A**  
**SCIENTIAE RERUM NATURALIUM**

*University Lecturer Tuomo Glumoff*

**B**  
**HUMANIORA**  
*University Lecturer Santeri Palviainen*

**C**  
**TECHNICA**  
*Postdoctoral researcher Jani Peräntie*

**D**  
**MEDICA**  
*University Lecturer Anne Tuomisto*

**E**  
**SCIENTIAE RERUM SOCIALIUM**  
*University Lecturer Veli-Matti Ulvinen*

**E**  
**SCRIPTA ACADEMICA**  
*Planning Director Pertti Tikkanen*

**G**  
**OECONOMICA**  
*Professor Jari Juga*

**H**  
**ARCHITECTONICA**  
*Associate Professor (tenure) Anu Soikkeli*

**EDITOR IN CHIEF**  
*University Lecturer Santeri Palviainen*

**PUBLICATIONS EDITOR**  
*Publications Editor Kirsti Nurkkala*

ISBN 978-952-62-3100-6 (Paperback)

ISBN 978-952-62-3101-3 (PDF)

ISSN 0355-3221 (Print)

ISSN 1796-2234 (Online)