

AI-based Analysis of Cardiovascular Disease Risk in CT of Breast Cancer Patients



Sanne G.M. van Velzen

AI-based Analysis of Cardiovascular Disease Risk in CT of Breast Cancer Patients

Sanne G.M. van Velzen

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

ISBN: 978-94-6419-173-8

Copyright © S.G.M. van Velzen 2021. All rights reserved.

No part of this publication may be reproduced or transmitted in any form by any means without prior permission from the copyright owner. The copyright of the articles that have been published has been transferred to the respective journals.

Printed by Gildeprint.

AI-based Analysis of Cardiovascular Disease Risk in CT of Breast Cancer Patients

Analyse van Cardiovasculair Risico op CT Beelden
van Borstkankerpatiënten met behulp van Kunstmatige Intelligentie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H. R. B. M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op donderdag 8 april 2021 des middags te 2.30 uur

door

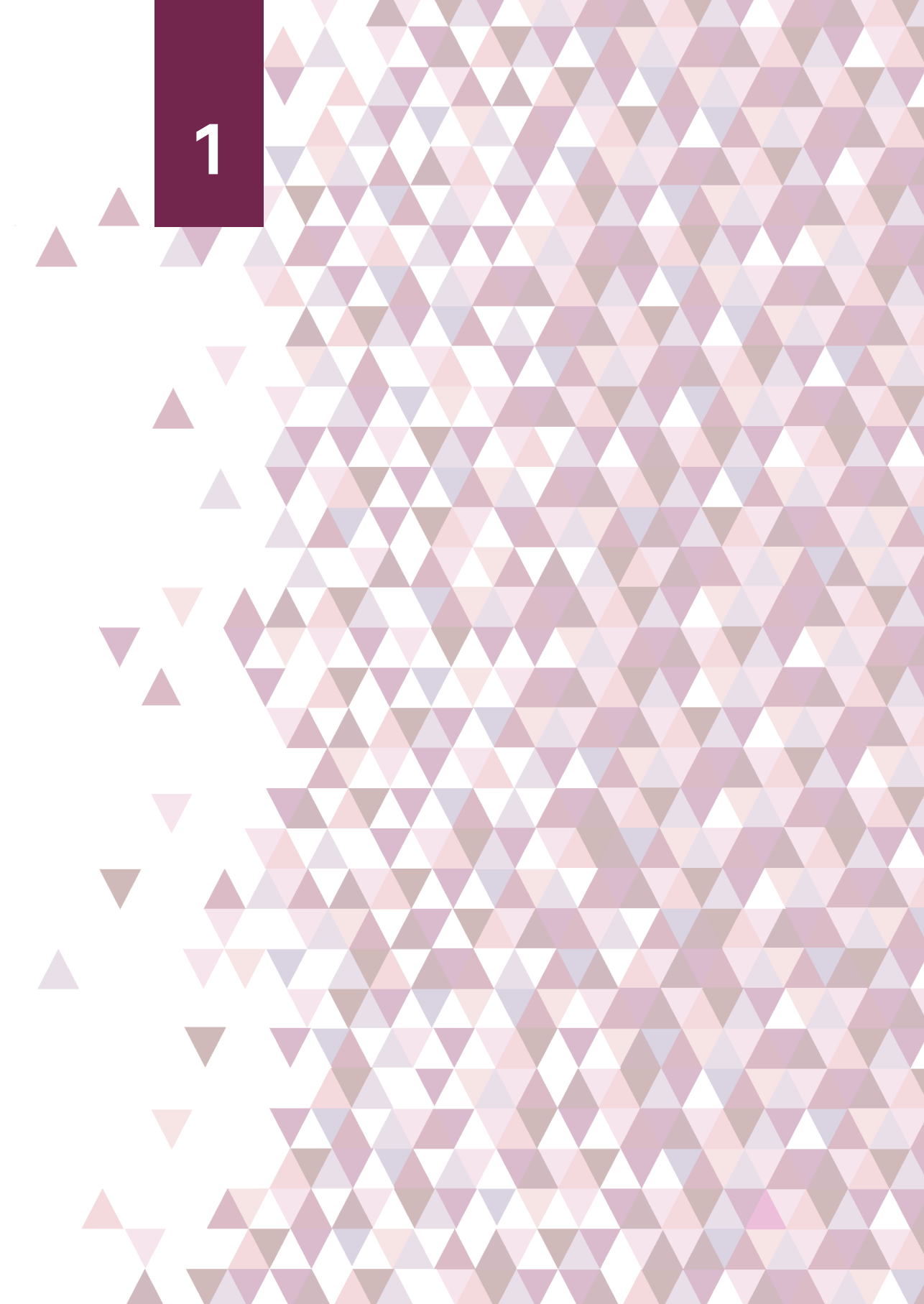
Sanne Geertruida Margaretha van Velzen
geboren op 29 december 1992
te Castricum

Promotoren: Prof. dr. I. Išgum
Prof. dr. H. M. Verkooijen
Prof. dr. ir. M. A. Viergever

Contents

CHAPTER 1	
Introduction	9
CHAPTER 2	
Deep learning for automatic calcium scoring in CT	17
CHAPTER 3	
Arterial calcifications in Western and Asian breast cancer patients	39
CHAPTER 4	
Identification of breast cancer patients at risk of cardiovascular disease from radiotherapy planning CT scans	53
CHAPTER 5	
Generative models for reproducible coronary calcium scoring	73
CHAPTER 6	
Direct prediction of cardiovascular mortality from low-dose chest CT	95
CHAPTER 7	
AI for estimation of radiotherapy-related risk of ischemic heart disease in breast cancer survivors	105
CHAPTER 8	
Summary and discussion	129
Nederlandse samenvatting	139
Bibliography	147
Acknowledgments	167
Publications	171
Biography	175

1



CHAPTER 1

Introduction

1.1 Cardiovascular disease in breast cancer patients

Rising incidence of breast cancer, early diagnosis due to screening and advances in treatment have led to an increasing number of breast cancer survivors. Accordingly, awareness regarding importance of quality of life after breast cancer is increasing. Cardiovascular disease (CVD) is common in cancer survivors^{1,2} and in older women diagnosed with breast cancer it is the primary cause of death.³ In relation to this, the risk of CVD events is increased by breast cancer treatments. For instance, patients that received treatment with anthracyclines have a five time higher risk of chronic heart failure than breast cancer patients that were treated with chemotherapy without anthracyclines.⁴ Patients that received radiotherapy are at increased risk of CVD events compared with patients that did not receive radiotherapy.^{5,6} Moreover, patients with pre-existing CVD risk factors have the highest risk of cardiotoxicity due to breast cancer treatment.⁷ Therefore, identification of patients at increased risk may provide options for cardio-prevention, such as lifestyle intervention, pharmaco-prevention or opting for less cardiotoxic treatment.

A frequent underlying cause of CVD is atherosclerosis.⁸ Atherosclerosis is the process of thickening of the innermost layer of the artery wall. This thickening consists of fat build-up in the artery wall and an accompanying inflammation reaction. In this process fatty plaques are formed that, in a later stadium, are transformed into fibrotic tissue and calcium deposits, i.e. calcifications. Subclinical atherosclerosis can smolder for years, without symptoms. However, plaques can become unstable, rupture and induce thrombosis, which can cause life-threatening events, such as heart attack and stroke.^{8,9}

CVD risk is determined based on the presence of traditional risk factors. For instance, several large studies carried out in the general population showed that based on age, blood pressure, cholesterol levels, presence of diabetes and smoking history, a risk profile can be determined.^{10,11} The risk profile is directly linked to the risk of experiencing a CVD event within the next 10 year. However, one in three breast cancer patients with subclinical atherosclerosis does not have any traditional risk factor and wouldn't be identified as high risk according to these criteria.¹² Therefore, identification of patients at increased risk of developing CVD events, may be done more accurately by detecting (subclinical) atherosclerosis. Imaging with computed tomography (CT) provides an opportunity for precise identification of patients at risk, because it visualizes calcified atherosclerotic plaque.

1.2 Visualization and quantification of atherosclerosis

CT images visualize calcifications in the arteries and can be used to identify patients at risk. Quantification of calcifications in the arteries is also known as calcium scoring. Of all arterial calcifications throughout the body, calcification in the coronary arteries (CAC)

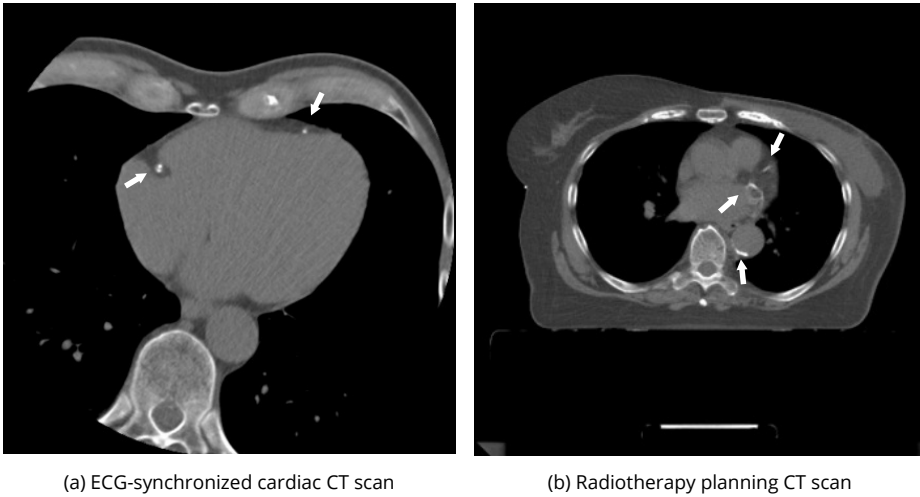


Figure 1.1: Examples of CT scans that visualize calcified atherosclerotic plaque. Even though the CT scans are made for entirely different purposes, both visualize the heart and large vessels. In both scans calcified lesions on the coronary arteries are visible.

is most predictive for CVD events.^{13,14} Therefore, quantification is typically focused on CAC. In clinical practice the amount of CAC is measured by identifying lesions in the coronary arteries with an intensity of > 130 Hounsfield units.¹⁵ Subsequently, the amount of CAC can be quantified into volume, calcium mass or Agatston score. The Agatston score is most widely applied and can be used to assign patients to a cardiovascular risk category.^{16,17} CVD risk categories range from very low CVD risk, indicated by an Agatston score of 0, to a very high risk of CVD, with an Agatston score of >400 .

In clinical practice, cardiac CT without contrast enhancement is used for quantification of CAC, because it provides a high-resolution and non-invasive view of the coronary arteries. Acquisition of these scans is synchronized with electrocardiography (ECG). ECG-triggering during scanning minimizes the effect of cardiac motion on the scan and minimizes motion artefact in the scan, which allows optimal visualization of the coronary arteries. Given the clinical relevance, calcium scoring is increasingly applied to non-cardiac CTs showing the heart.¹⁸ This includes CT scans that are made for other purposes than calcium scoring and are often not acquired with ECG-synchronization. Because the heart is moving during acquisition of these scans, they often contain cardiac motion artefacts. Moreover, because these scans are made for a wide variety of clinical indications, the field of view often contains the whole chest or body, which leads to a lower resolution. These limitations provide extra challenges for CVD risk prediction.

Clinically, calcium scoring is done manually by a medical expert who identifies

lesions in the CT scan using dedicated interactive software. Although this is not considered a difficult task for experts, it is tedious and time consuming, especially when large numbers of scans are involved. Therefore, automation of CVD risk prediction may be preferred for large scale studies and wide application in clinic.

1.3 Radiotherapy treatment planning

The majority of patients with breast cancer is treated with adjuvant radiotherapy. During clinical routine of radiotherapy treatment, a non-ECG-synchronized CT scan is made to determine the 3D irradiation dose distribution in 3D CT radiation planning. On this treatment planning CT scan the tumor is delineated by a clinician to evaluate the tumor coverage. Also, several organs in the radiation field are delineated to monitor the potential harm to healthy tissue that should not be irradiated. Subsequently, using dedicated irradiation planning software, the optimal plan is calculated for each patient individually. This plan is a careful balance between optimal tumor coverage and minimal harm to healthy tissue.

Because radiation exposure to the heart is known to increase the risk of CVD and ischemic heart disease,¹⁹ several measures are taken to minimize the dose to the heart. For instance, the heart is delineated on the planning CT to take it into account as organ at risk for calculating the irradiation plan. Moreover, for patients with breast cancer on the left side the deep inspiration breath-hold technique is applied. This means that during irradiation, the patient holds her breath to maximize the distance between the tumor and the heart. Left sided breast cancer patients, therefore, also receive a second CT scan for planning where the deep-inspiration breath-hold technique is applied.

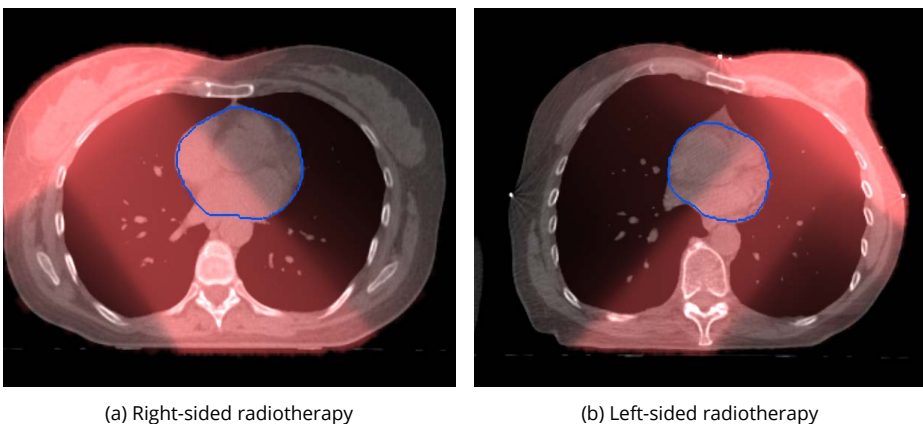


Figure 1.2: Examples of planning CT scans with an illustrative heatmap of the planned dose in a patient with a right-sided tumor (a) and a patient with a left-sided tumor (b). The heart is delineated as organ at risk in blue.

The routine CT scans made for 3D CT planning provide a way to monitor the planned radiation exposure of the heart in radiotherapy. Moreover, because CT visualizes atherosclerosis, they provide an opportunity for identification of patients at increased CVD risk by means of calcium scoring.¹² Furthermore, besides CAC, more information is present in CT images that is possibly predictive for CVD events.

1.4 Machine learning

In this thesis machine learning techniques are used for automatic analysis of CT images. In machine learning a model is trained with examples (data) to map an input to an output value. Mapping is done based on information that is extracted from the input. In conventional machine learning the information representations, i.e. features, are designed to describe the information in the input that is relevant for the problem. This manner of handcrafting features is based on expert domain knowledge and requires considerable human effort. After training the model with examples and corresponding labels (outputs), the model can be applied to new cases.

In deep learning, complex models are used that are able to intrinsically learn from the data to extract relevant information directly from the input and, thereby, skipping the feature computation.²⁰ The basis of current deep learning methods lies in artificial neural networks (ANNs). ANNs are inspired by the way the human brain processes information and consist of a number of artificial neurons connected to form a certain architecture. The value of a neuron is determined by the sum of the connected previous neurons, each multiplied by a trainable weight. Thereafter the value is passed through a nonlinear function. By combining multiple layers of nonlinear neurons, a highly nonlinear relationship between input and output can be modeled.

The type of ANN that is most used and most researched in medical image analysis, are convolutional neural networks (CNNs). CNNs share the weights at each location of the image. This way the layers of the network form trainable convolution kernels that move over the image. The activation patterns in the deeper layers form a complex feature representation of the input that is used for prediction.

A CNN can be trained in two manners: supervised and unsupervised. In supervised learning an often manually created label is provided for each training input example. The CNN is trained perform to a certain task with examples of input and label pairs. In supervised learning the final performance of the CNN is highly dependent on the quality of the training data and labels.²¹ In unsupervised learning no label is provided for the input. For instance the input image itself is used as target for training the model to recognize patterns and clusters in the dataset.

It is not only possible with CNNs to output one label per input image, but whole images can be predicted, containing either discrete values like segmentation labels or continuous values like intensity values. An example of the latter are CNNs that predict realistic medical images. Very popular over the past years are generative adversarial

networks (GANs).²² GANs combine a generator CNN with a discriminator CNN: the generator outputs an image and the discriminator distinguishes generated images from real images. This system is trained in such a way that the CNNs compete with each other: the discriminator tries to spot the generated images among the real images and the generator tries to 'fool' the discriminator by predicting very realistic images.

1.5 Outline of the thesis

This thesis presents automatic methods for identification of breast cancer patients at increased risk of CVD events. The methods are based on analysis of CT scans made during radiotherapy treatment planning. These images are used for automatic detection of atherosclerosis and prediction of CVD risk, but also for investigating the effect of breast cancer treatments on the risk of CVD.

CHAPTER 2 evaluates a deep learning method for automatic calcium scoring. In this chapter the performance of the method is extensively evaluated in a large set containing five different types of CT scans. The performance of the originally trained deep learning method was evaluated on other unseen types of CT and the effect of selection of training data on the performance in the five types of CT scans was assessed.

In **CHAPTER 3** we applied the automatic calcium scoring method described in Chapter 2 to CT scans of two cohorts of breast cancer patients: Western breast cancer patients and Asian breast cancer patients. In this chapter we evaluate the differences in calcium burden between these two groups of patients.

CHAPTER 4 evaluates whether CAC quantified on planning CT scans is associated with increased risk of CVD in breast cancer patients. The method described and evaluated in Chapter 2 is used to automatically measure CAC on CT scans of a large cohort of 18,316 breast cancer patients from three hospitals. Subsequently, CAC scores were related to risk of CVD events using survival analysis.

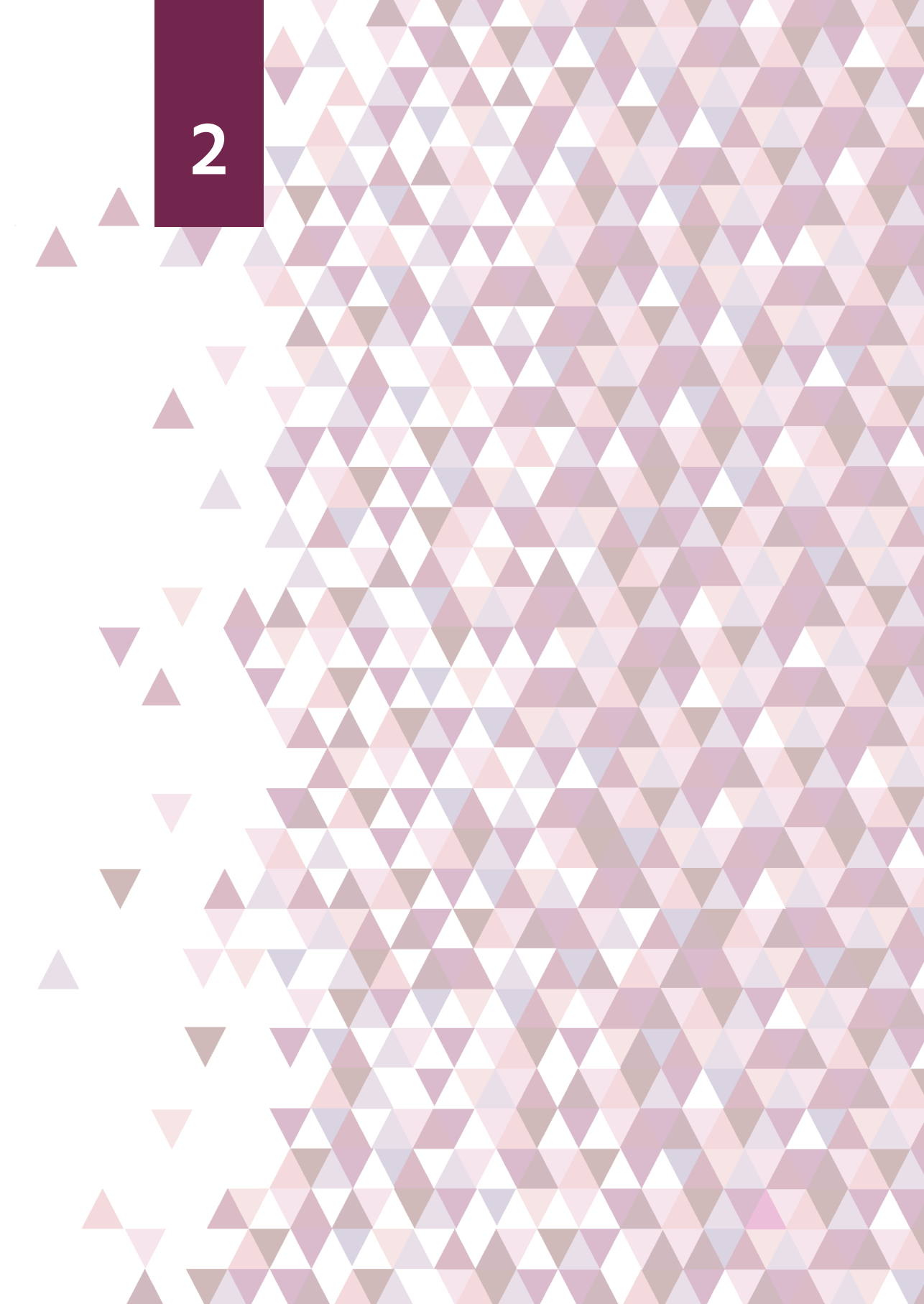
Because the interscan reproducibility of clinically used calcium scores is limited, **CHAPTER 5** describes an automatic method for quantification of CAC in non-ECG-synchronized CT scans with increased interscan reproducibility compared to clinical calcium scoring methods. This method uses a generative model to decompose an image containing CAC into an image without CAC and an image visualizing only CAC. The method is able to segment whole lesions, including parts below the clinical intensity level threshold.

CHAPTER 6 describes a machine learning method for direct prediction of cardiovascular mortality from an CT image. The method uses an autoencoder CNN to extract image features in an unsupervised manner, without hand-crafting them. Subsequently, 5-year

cardiovascular mortality is predicted based on the extracted image features.

Because the radiation exposure is not homogenous in radiotherapy for breast cancer, **CHAPTER 7** describes a deep learning method for calculating radiation exposure of cardiac substructures and coronary arteries during radiotherapy. Additionally, the association of dose delivered to the cardiac substructures and arteries, and ischemic heart disease is evaluated.

Finally, **CHAPTER 8** summarizes the presented methods and findings and discusses the results as well as future directions.



CHAPTER 2

Deep learning for automatic calcium scoring in CT

BASED ON: S. G. M. van Velzen, N. Lessmann, B. K. Velthuis, I. E. Bank, D. H. van den Bongard, T. Leiner, P. A. de Jong, W. B. Veldhuis, A. Correa, J. G. Terry, J. J. Carr, M. A. Viergever, H. M. Verkooijen, and I. Išgum. "Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols", *Radiology*, vol. 295 (2020), pp. 66–79.

Abstract

BACKGROUND Although several deep learning (DL) calcium scoring methods achieve excellent performance for specific CT protocols, their performance on a range of CT exams is unknown.

PURPOSE To evaluate the performance of a DL method for automatic calcium scoring across a wide range of CT and to investigate whether the method can adapt to different types of CT when representative images are added to the existing training dataset.

MATERIALS AND METHODS The study included 7240 subjects who received various types of non-contrast CT including the heart: coronary artery calcium (CAC)-scoring, diagnostic chest, PET attenuation correction, radiotherapy treatment planning, CAC-screening and low-dose chest CT. CAC and thoracic aorta calcifications (TAC) were quantified using a convolutional neural network trained with (1) 1181 low-dose chest CT (baseline); (2) a small set of the respective exam type supplement to the baseline (data-specific); and (3) a combination of all available exam types (combined). Supplement training sets contained 199-568 CTs, depending on the populations calcium burden. The performance was evaluated with intraclass correlation coefficients (ICC) between DL and reference (Agatston) CAC and (volume) TAC scores and linearly weighted kappa for cardiovascular risk categories (Agatston score, CVD risk categories: 0,1-10,11-100,101-400,>400).

RESULTS At baseline, the DL model achieved an ICC between 0.79-0.97 for CAC and 0.66-0.98 for TAC over the different types of CT exams. ICCs improved to 0.84-0.99 and 0.92-0.99 for data-specific training, and to 0.85-0.99 and 0.96-0.99 for combined training, for CAC and TAC respectively. For assignment of CVD risk category, kappa (95% CI) for all test CT scans was 0.90 (0.89-0.91) for the baseline training. It increased to 0.92 (0.91-0.93) for both data-specific and combined training.

CONCLUSIONS A deep learning calcium scoring method to measure calcified atherosclerotic plaque is robust to substantial differences in image acquisition protocol and subject population. Augmenting the DL training data with a set of data-specific images improved the performance to that of current calcium scoring methods.

2.1 Introduction

Coronary artery calcium (CAC) scoring in dedicated CT examinations is frequently performed for measuring coronary atherosclerotic plaque burden and prediction of cardiovascular disease (CVD) risk. The CAC score may have increasing application as noted in the 2018-2019 AHA/ACC Guidelines for Cholesterol and Prevention where CAC score is a tool to refine the 10-year risk of atherosclerotic cardiovascular disease when the CVD risk may be uncertain.²³ CAC scoring is performed using non-contrast enhanced ECG-synchronized cardiac CT, but, multiple reports indicate that application of coronary calcium scoring in CT scans made with non-dedicated protocols, allows CVD risk prediction.²⁴ Accordingly, in addition to automatic methods for calcium scoring in non-contrast enhanced ECG-synchronized cardiac CT,²⁵⁻²⁸ methods were also developed for automatic scoring in low-dose chest CT for lung cancer screening,²⁹⁻³² radiotherapy treatment planning CT^{12,33} and PET attenuation correction CT.³⁴ These CT exams all include the heart; however vary widely in resolution, field of view, reconstruction kernels, noise characteristics and the presence of ECG-synchronization (Figure 2.1). Moreover, calcifications in the thoracic aorta (TAC) are predictors of CVD³⁵ and hence, several methods for automatic quantification of TAC in CT have been developed,^{32,36} as well. Although a number of these methods achieved close to human level performance in the type of CT they were developed for, their performance in different types of CT has not been established.

In spite of many efforts supporting the development of robust DL software,^{37,38} an established limitation of DL algorithms is that performance can rapidly degrade and fail with what appears to humans as only slight variations in the input data.³⁹ Therefore, the purpose of this study was to evaluate the performance of a DL method for automatic calcium scoring³² across a wide range of CT exams that visualize the heart and the thoracic aorta. Furthermore, we investigated whether addition of a low number of representative examples to the baseline training data is sufficient for the DL method to adapt to different types of CT exams.

2.2 Materials and Methods

2.2.1 Image datasets

We included six existing datasets with 7240 subjects who received a non-contrast CT exam visualizing the heart (Table 2.1). These were 902 consecutive ECG-triggered cardiac CTs for calcium scoring (CAC-CT)²⁸ used in clinical practice, 399 cardiac PET attenuation correction (AC_{PET}) CTs from the MYOMARKER study⁴⁰ acquired in rest condition during radionuclide myocardial perfusion imaging, 1409 consecutive radiotherapy treatment planning (RadTherapy) CTs of breast cancer patients from the UMBRELLA cohort,³³ 470 consecutive diagnostic chest CTs, 2879 ECG-gated CAC

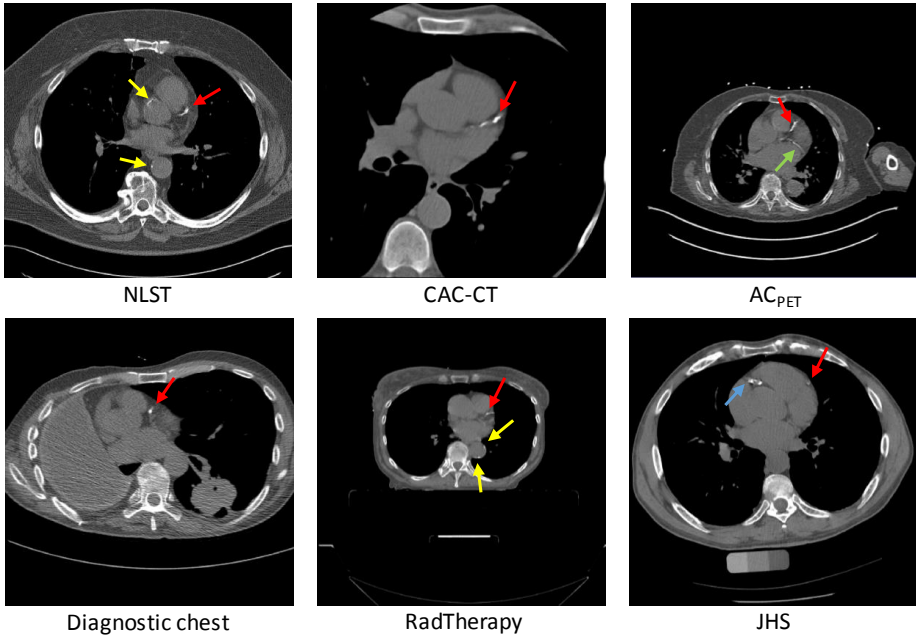


Figure 2.1: Examples of lung screening CT from the National Lung Screening Trial (NLST), coronary artery calcification (CAC) scoring CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT, radiotherapy treatment planning (RadTherapy) and CT exams from the Jackson Heart Study (JHS). Indication of CAC in the LAD (red), LCX (green), RCA (blue) and aorta (yellow).

screening CTs from the Jackson Heart Study (JHS)^{41,42} and 1181 lung screening low-dose chest CTs from the National Lung Screening Trial (NLST).⁴³

The AC_{PET} , JHS, RadTherapy and NLST examinations were acquired as part of Ethics Committee approved studies where informed consent was obtained from all subjects. The CAC-CT and diagnostic chest CT exams originate from the University Medical Center Utrecht and were retrospectively and anonymously collected, the need for informed consent was therefore waived by our institutional review board.

Table 2.1: Characteristics of the included datasets of lung screening CT from the National Lung Screening Trial (NLST), Coronary Artery Calcification (CAC) scoring CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT, radiotherapy treatment planning (RadTherapy) CT and CT exams from the Jackson Heart Study (JHS).

	NLST CT	CAC-CT	AC _{PET}	Diagnostic chest CT	RadTherapy CT	JHS CT
Number of scans	1181	902	399	470	1409	2879
Training set	1012	323	179	159	529	361
Men (%)	610 (60)	243 (65)	97 (54)	84 (53)	1 (0)	135 (37)
Women (%)	402 (40)	130 (35)	82 (46)	75 (47)	528 (100)	226 (63)
Age (years)	61 ± 4	51 ± 6	68 ± 7	67 ± 8	56 ± 7	57 ± 9
Validation set	169	50	20	20	39	37
Men (%)	114 (67)	31 (62)	12 (60)	15 (75)	0 (0)	20 (54)
Women (%)	55 (33)	19 (38)	8 (40)	5 (25)	39 (100)	17 (46)
Age (years)	61 ± 4	51 ± 7	68 ± 5	65 ± 5	54 ± 10	57 ± 10
Test set	-	529	200	291†	841	2479
Men (%)	-	383 (72)	124 (62)	169 (58)	1 (0)	893 (36)
Women (%)	-	146 (28)	76 (38)	122 (42)	840 (100)	1586 (64)
Age (years)	-	52 ± 7	67 ± 7	64 ± 9	56 ± 7	60 ± 9
Acquisition						
Hospital	31 hospitals	University Medical Center Utrecht	Meander Medical Center Amersfoort	University Medical Center Utrecht	University Medical Center Utrecht	Jackson Medical Mall
scanner type	All major CT vendors	Philips Brilliance iCT 256	Siemens Biograph 40 PET/CT	Philips Brilliance 16P/64/iCT 256/ MX8000 IDT 16	Philips Brilliance Big Bore CT	GE Healthcare Lightspeed 16Pro
ECG triggering	no	yes	no	no	no	yes
Tube voltage (kVp)	120/140	120	100/120/140	100/120/140	120	120
Reconstruction						
Slice thickness (mm)	1.0-3.0	3.0	3.0	0.9/1.0	3.0	2.5
Increment (mm)	0.6-3.0	1.5	2.	0.7	3.0	2.5
In-plane resolution (mm)	0.49-0.98	0.29-0.49	0.71-0.98	0.52-0.89	0.78-1.37	0.68
Prevalence and extent†						
CAC >0 (%)	873 (74)	256 (48)	168 (84)	178 (61)	128 (15)	1211 (49)
Median (IQR) CAC Agatston *	172 (40-624)	80 (13-315)	671 (224-2058)	369 (85-1086)	18 (6-87)	93 (27-374)
TAC >0 (%)	1027 (87)	173 (33)	183 (92)	223 (77)	462 (55)	-
Median (IQR) TAC (mm3)*	549 (260-1850)	142 (35-576)**	1522 (383-3703)	1249 (238-4493)	174 (53-631)	-

† 16 scans were excluded from the test set, because calcium scoring was not feasible due to large metal implants.

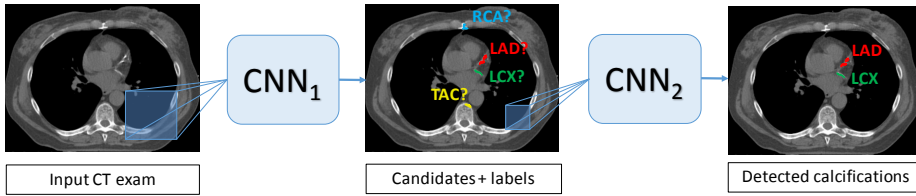


Figure 2.2: Architecture of the DL calcium scoring algorithm. The algorithm consists of two convolutional neural networks (CNN). The first CNN has a large field of view and detects candidate calcifications (voxels) in the image and labels them according to their anatomical location. The second CNN has a smaller field of view and detects true calcified voxels among the candidates resulting from the first CNN.

2.2.2 Reference calcium scores

CAC and TAC were semi-automatically labelled by four trained observers, each with prior experience in over 500 exams, without overlap in the sets between observers. The observers were supervised by an expert radiologist (PdJ, 10 years of experience), who was consulted in case of doubt. In-house custom-built software for semi-automatic calcium scoring was used.^{12,44} In brief, the software for calcium scoring shows all regions of ≥ 3 adjacent voxels with attenuation above 130HU with a colored overlay. The observer manually identifies lesions with a mouse click and labels them according to their anatomical location i.e. left anterior descending (LAD, including the left main coronary), left circumflex (LCX), right coronary artery (RCA) and thoracic aorta (TAC). Subsequently, three-dimensional component labelling is performed to mark all connected voxels in the lesion as calcification. Calcifications in the Diagonal and Obtuse Marginal branches were included in the LAD and LCX, respectively. Excessive image noise and large metal implants interfered with semi-automatic segmentation of calcium in 16/291 (5%) diagnostic chest CTs. These exams were therefore excluded from the test set. In the JHS, reference CAC scores were obtained by semi-automatically determining the artery trajectory and segmenting calcifications within this trajectory⁽²¹⁾. TAC was not quantified in the JHS and reference TAC scores were therefore not available for this dataset. As representative examples for training the automatic method, 400 exams were re-annotated using the standard protocol described above.

In the NLST CT examinations, we used the manual calcium annotations as previously published.³² In these exams, there was poor signal-to-noise ratio, making semi-automatic segmentation infeasible.

2.2.3 Automatic calcium scoring using deep learning

The DL network (Figure 2.2) for automatic calcium scoring³² consisted of two consecutive convolutional neural networks (CNN) that detected, quantified and labeled calcifications according to their anatomical location (LAD, RCA, LCX and TAC). Off-target calcifications in the heart such as valvular calcification were assigned anatomic locations but were not included in this study. The code is publicly available at <https://github.com/sgmvanvelzen/calcium-scoring>.

Two modifications were made to the previously published DL method:³² To standardize the field of view, exams with a field of view larger than originally used NLST exams (RadTherapy, AC_{PET} and diagnostic chest CT) were preprocessed with a CNN-based localization method.⁴⁵ Next, to match the standard calcium scoring protocol in which calcifications are defined as lesions above 130 HU,¹³ we postprocessed the results by identifying 3D connected clusters of voxels above 130 HU. We discarded lesions with less than 25% of the voxels labeled as calcium by the network, as these were likely false positive findings. Moreover, lesions larger than 10,000 mm³ were discarded as these were likely connected to bone or foreign objects such as pacemaker wires.

CAC and thoracic aorta calcifications (TAC) were quantified using the DL method trained with (I) 1181 low-dose chest CT from the NLST (baseline); (II) a small set of the respective exam type supplemented to the baseline (data-specific); and (III) a combination of all available exam types (combined, Figure 2.3)¹. The baseline training set was used for the original development of the DL algorithm for calcium scoring in NLST CT.³² For data-specific training a number of randomly selected exams in each dataset was reserved (Table 2.1). To ensure sufficient CAC and TAC examples from the target CT type are available for training and validation, the number of reserved exams was based on the calcium burden in the respective population (Table 2.1). For each dataset, a data-specific network training was performed combining the baseline with the training exams from the respective dataset. To evaluate whether a single DL network could be applied to all types of CT, a combined training was performed with all available training exams except the diagnostic chest exams, since these often contained severe pathologies that did not appear in other exams. To allow for direct comparison, the performance of the baseline, data-specific and combined network was evaluated using the same test set. There was no overlap between training and test data.

The network was trained using parameters as described by Lessmann et al.³² In brief, CNN1 was trained for 1000 iterations (batch size of 32) and CNN2 was trained for 750 iterations (batch size of 64), after which training was converged in all cases. During training, batches of examples were randomly selected from the training exams. Owing to the very low calcium burden of women undergoing radiotherapy for breast cancer,

¹Although the architecture of the network remained unchanged, we refer to the network with baseline training, data-specific training and combined training as 'baseline', 'data-specific' and 'combined' network, respectively.

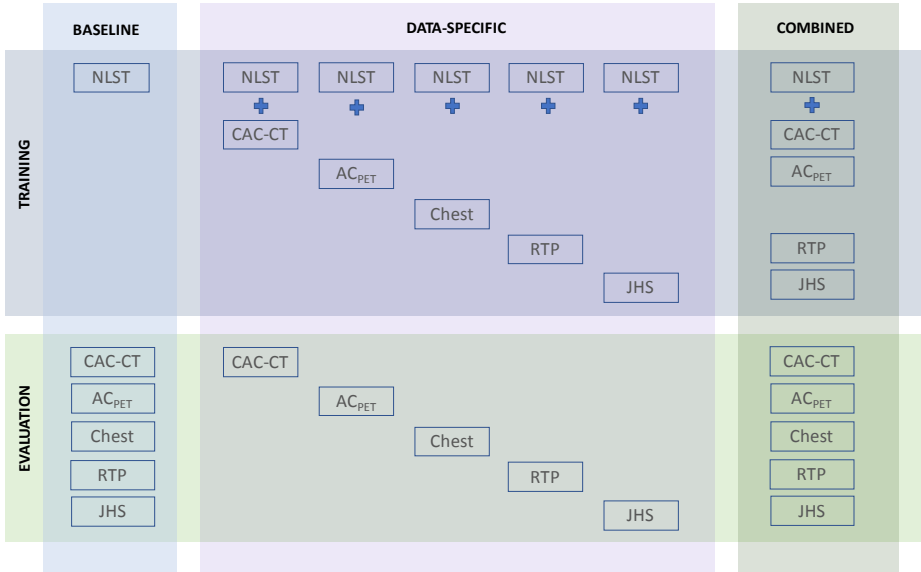


Figure 2.3: Training and evaluation of the baseline, data-specific and combined algorithms. The baseline is trained with NLST scans and its performance is evaluated in each CT protocol type. Five data-specific algorithms are trained, one specifically for each CT protocol type, and evaluated in the respective CT type. The combined algorithm is trained with a combination of all available CT protocol types (excluding diagnostic chest CT) and its performance is evaluated in all available CT protocol types. The CTs used for training are National Lung Screening Trial (NLST), coronary artery calcification (CAC) scoring CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT, radiotherapy treatment planning (RTP) and CT exams from the Jackson Heart Study (JHS)

we enforced 20% RadTherapy examples per batch during RadTherapy-specific training

2.2.4 Statistical analysis

The volume of calcification was determined (mm^3). The calcium score was computed using the Agatston method.⁴⁶ Based on the Agatston score, subjects were stratified into five commonly used CVD risk groups: 0, 1-10, 11-100, 101-400, >400.⁴⁷

For each network, average sensitivity and average false positive calcium volume per CT type were calculated. Reliability of volume and Agatston scores was assessed with the intraclass correlation coefficient (ICC) between automatically and manually obtained scores and their agreement by examining Bland-Altman plots with 95% limits of agreement. Since the errors tend to increase with increasing CAC, in the Bland-Altman plots regression for non-uniform differences was used to model the variation of the absolute differences between manual and automatic scoring.⁴⁸ To calculate the 95% limits of agreement the predicted absolute differences were multiplied by $1.96 * (\pi/2)^{0.5}$, since the absolute differences have a half-normal distribution. Reliability of the CVD risk categorization was assessed with Cohen's linearly weighted kappa, and agreement

was assessed by proportion of subjects assigned to the same category manually and automatically. To test whether differences in absolute errors in quantified per label calcification volume and Agatston score between algorithms are significant, a Wilcoxon signed rank test for paired samples was used with a significance level of 0.05. Analyses were performed with IBM SPSS statistics version 23 and an online statistical tool.⁴⁹

2.3 Results

2.3.1 Demographics, prevalence and extent of coronary and thoracic calcifications

Median age in test sets ranged from 52 years for CAC-CT to 67 years for AC_{PET} CT (Table 2.1). Percentage of women in the test populations ranged from 28% in CAC-CT to 100% in RadTherapy CT. The prevalence (CAC 15-92%, TAC 33-91%) and extent of CAC and TAC (CAC: median Agatston score 18-671m3, TAC: median volume 174-1522m3) differed substantially across the data sets.

2.3.2 Deep learning evaluation of Agatston score and CAC volume

For CAC volume, the baseline network achieved an ICC (95% CI) of 0.97 (0.96-0.97) for CAC-CT, 0.84 (0.80-0.88) for AC_{PET}, 0.90 (0.88-0.92) for diagnostic chest and 0.85 (0.83-0.87) for RadTherapy exams (Table 2.2). ICCs (95% CI) of the dataset-specific network improved to 0.98 (0.98-0.99) for CAC-CT, 0.97 (0.96-0.98) for AC_{PET}, 0.98 (0.97-0.98) for diagnostic chest and 0.92 (0.91-0.93) for RadTherapy exams. The combined network showed similar detection performance as the data-specific network. In line with this, Bland-Altman plots in Figure 2.4 show improved agreement of Agatston scores for the combined network with respect to the baseline network.

On average, the total computation time ranged between 2 – 7 minutes, depending on the exam type, image volume and the extent of calcium burden in the exam. The agreement of the automatically detected TAC volumes improved in 3 out of 4 datasets when adding representative data to the training. The baseline network had an ICC (95% CI) for TAC volume of 0.85 (0.83-0.87) for CAC-CT, 0.66 (0.57-0.73) for AC_{PET}, 0.96 (0.95-0.97) for diagnostic chest and 0.98 (0.97-0.98) for RadTherapy exams (Table 2.2). ICCs (95% CI) for CAC-CT, AC_{PET} and RadTherapy increased to 0.96 (0.95-0.96), 0.96 (0.94-0.97) and 0.99 (0.98-0.99) for the data-specific network and to similar ICCs for the combined network. ICC (95% CI) for diagnostic chest CT was lower than baseline for the data-specific network, but similar to baseline for the combined network, with 0.97 (0.96-0.98).

Bland-Altman plots (Figure 2.5) show that the limits of agreement for TAC are narrower for the combined network with respect to the baseline for CAC-CT, RadTherapy and AC_{PET} exams. However, the 95% limits of agreement for the combined training do not visibly differ from the baseline for diagnostic chest CT.

Table 2.2: Volume-wise performance evaluation of the baseline, data-specific and combined training in Coronary Artery Calcification (CAC) scoring CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT and radiotherapy treatment planning (RadTherapy) CT.

	Reference Volume (mm3)	Baseline			Data-specific			BD _{p-value} †
		Sensitivity (mm3) %, 95%CI	FP vol mm3	Volume ICC	Sensitivity (mm3) %, 95%CI	FP vol mm3	Volume ICC	
CAC-CT								
CAC	140	122 (88, 85-91)	8	0.97 (0.96-0.97)	130 (93, 91-95)	4	0.98 (0.98-0.99)	<0.01
LAD	69	67 (97, 96-99)	17	0.91 (0.89-0.93)	67 (97, 96-99)	5	0.97 (0.96-0.97)	<0.01
LCX	21	9 (41, 37-45)	0	0.69 (0.63-0.74)	14 (66, 62-70)	1	0.64 (0.59-0.69)	0.2
RCA	50	37 (74, 70-78)	1	0.91 (0.90-0.93)	45 (91, 89-94)	2	0.97 (0.97-0.98)	<0.01
Aorta	253	175 (69, 65-73)	23	0.85 (0.83-0.87)	226 (89, 86-92)	24	0.96 (0.95-0.96)	<0.01
ACPET								
CAC	1093	837 (77, 71-82)	116	0.84 (0.80-0.88)	1011 (93, 89-96)	155	0.97 (0.96-0.98)	0.9
LAD	420	395 (94, 91-97)	155	0.80 (0.71-0.85)	390 (93, 89-96)	63	0.89 (0.86-0.92)	<0.01
LCX	256	118 (46, 39-53)	15	0.69 (0.58-0.77)	197 (77, 71-83)	40	0.89 (0.85-0.91)	<0.01
RCA	416	241 (58, 51-65)	29	0.61 (0.51-0.70)	397 (95, 93-98)	79	0.96 (0.95-0.97)	0.91
Aorta	3373	2206 (65, 59-72)	270	0.66 (0.57-0.73)	2766 (82, 77-87)	453	0.96 (0.94-0.97)	<0.01
Diagnostic chest								
CAC	477	404 (86, 81-90)	86	0.90 (0.88-0.92)	439 (93, 90-96)	84	0.98 (0.97-0.98)	0.93
LAD	203	191 (95, 92-98)	93	0.89 (0.86-0.91)	186 (93, 90-96)	75	0.93 (0.91-0.95)	<0.01
LCX	106	45 (43, 37-49)	5	0.46 (0.36-0.55)	76 (73, 68-78)	14	0.91 (0.89-0.93)	<0.01
RCA	169	137 (82, 78-87)	19	0.84 (0.81-0.88)	153 (92, 88-95)	19	0.97 (0.96-0.97)	<0.01
Aorta	1985	1786 (90, 86-93)	301	0.96 (0.95-0.97)	1797 (90, 87-94)	414	0.92 (0.90-0.94)	0.95
RadTherapy								
CAC	19	14 (75, 72-78)	2	0.85 (0.83-0.87)	16 (86, 84-89)	1	0.92 (0.91-0.93)	0.03
LAD	11	10 (98, 97-99)	1	0.98 (0.97-0.98)	10 (98, 97-99)	1	0.97 (0.97-0.98)	0.93
LCX	4	1 (29, 26-32)	0	0.34 (0.27-0.39)	3 (87, 85-89)	0	0.99 (0.99-0.99)	0.01
RCA	5	2 (44, 40-47)	1	0.52 (0.47-0.57)	3 (54, 51-56)	0	0.59 (0.54-0.63)	<0.01
Aorta	407	381 (94, 92-95)	66	0.98 (0.97-0.98)	386 (95, 93-96)	47	0.99 (0.98-0.99)	<0.01

For the CT exams from the Jackson Heart Study this analysis was not possible because only total Agatston scores were available. Volume, sensitivity and false positive (FP) volume are averages of the respective dataset.

P-values of difference in absolute errors of per label calcium volumes between:

† baseline and data-specific

‡ baseline and combined

§ data-specific and combined

continued...

Sensitivity (mm3) (%, 95% CI)	FP vol (mm3)	Combined		
		Volume ICC	BC _{p-value} [‡]	DC _{p-value} [§]
130 (93, 91-95)	2	0.99 (0.99-0.99)	<0.01	0.58
64 (93, 91-96)	1	0.97 (0.97-0.98)	<0.01	<0.01
18 (87, 84-90)	2	0.91 (0.89-0.92)	<0.01	<0.01
45 (91, 89-93)	1	0.98 (0.98-0.99)	<0.01	<0.01
218 (86, 83-89)	26	0.96 (0.96-0.97)	0.11	0.53
1010 (92, 89-96)	124	0.97 (0.97-0.98)	0.04	0.07
395 (94, 91-97)	60	0.90 (0.89-0.93)	<0.01	0.73
200 (78, 72-84)	50	0.82 (0.78-0.87)	0.06	0.67
382 (92, 88-96)	47	0.97 (0.97-0.98)	0.02	<0.01
2838 (84, 79-89)	509	0.97 (0.96-0.98)	0.28	0.22
439 (92, 90-96)	89	0.98 (0.97-0.98)	0.27	0.71
188 (93, 91-96)	71	0.93 (0.92-0.95)	0.03	<0.01
77 (72, 96-79)	25	0.91 (0.89-0.93)	0.34	<0.01
153 (90, 88-95)	20	0.97 (0.96-0.98)	0.82	<0.01
1785 (90, 86-93)	426	0.97 (0.96-0.98)	0.24	0.71
16 (86, 83-88)	1	0.92 (0.91-0.93)	0.06	0.59
10 (95, 93-96)	0	0.99 (0.99-0.99)	0.22	0.15
3 (92, 90-94)	0	0.99 (0.99-0.99)	0.28	0.81
3 (57, 54-60)	1	0.59 (0.54-0.63)	0.06	1
388 (95, 94-97)	39	0.99 (0.98-0.99)	0.05	0.06

Table 2.3: Reliability of continuous Agatston scores and risk category assignment based on Agatston scores for the baseline, data-specific and combined training in Coronary Artery Calcification (CAC) scoring CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT, radiotherapy treatment planning (RadTherapy) CT and CT exams from the Jackson Heart Study (JHS).

	Baseline		Data-specific		Combined	
	Categorical κ _{lw}	Continuous ICC	Categorical κ _{lw}	Continuous ICC	Categorical κ _{lw}	Continuous ICC
CAC-CT	0.95 (0.93-0.97)	0.97 (0.97-0.98)	0.98 (0.96-0.99)	0.99 (0.98-0.99)	0.97 (0.96-0.98)	0.99 (0.99-0.99)
AC_{PET} CT	0.88 (0.82-0.93)	0.83 (0.78-0.87)	0.84 (0.82-0.90)	0.83 (0.96-0.97)	0.92 (0.88-0.96)	0.91 (0.97-0.98)
Diagnostic chest CT	0.90 (0.86-0.94)	0.90 (0.87-0.92)	0.91 (0.88-0.95)	0.89 (0.97-0.98)	0.91 (0.88-0.95)	0.90 (0.97-0.98)
RadTherapy CT	0.85 (0.80-0.90)	0.79 (0.76-0.81)	0.89 (0.85-0.93)	0.84 (0.82-0.86)	0.91 (0.88-0.95)	0.96 (0.83-0.87)
JHS CT	0.85 (0.83-0.86)	0.90 (0.89-0.91)	0.90 (0.88-0.91)	0.94 (0.94-0.95)	0.89 (0.88-0.90)	0.89 (0.94-0.95)

95% confidence intervals are reported for linearly weighted kappa (κ_{lw}) and ICC values.

P-values of difference in absolute errors of Agatston scores between:

† baseline and data-specific

‡ baseline and combined

§ data-specific and combined

BC_p-value[‡]

BD_p-value[†]

DC_p-value[§]

ACC

κ_{lw}

ICC

κ_{lw}

ACC

κ_{lw}

ICC

ACC

κ_{lw}

ICC

2.3.3 Risk category assignment

The reliability and accuracy of the CVD risk categories was high (>0.81) for all three networks (Table 2.3). The baseline network achieved a linearly weighted κ of 0.95, 0.88, 0.90, 0.85, and 0.85 for CAC-CT, AC_{PET} , diagnostic chest, RadTherapy and JHS exams, respectively. For all test sets combined the κ (95% CI) was 0.90 (0.89-0.91). Reliability increased with the data-specific network to 0.98, 0.91, 0.89 and 0.90 for CAC-CT, diagnostic chest, RadTherapy and JHS exams, respectively. These values are similar to the reported reliability in NLST (0.91)(10). However, reliability of the data-specific network in AC_{PET} was slightly lower than baseline (0.84). For the combined network reliability was excellent in all exam types with a κ of 0.97 for CAC-CT, 0.92 for AC_{PET} , 0.91 for diagnostic chest, 0.91 for RadTherapy, and 0.89 for JHS exams. Overall, κ (95% CI) was 0.92 (0.91-0.93) for all test sets combined for both the data-specific and combined network.

The majority of patients were assigned to the correct CVD risk category (2194/2479 (89%) - 507/529 (96%)), and 14/529 (3%) - 172/2479 (7%) ended up in the neighboring risk group (Figure 2.6). Points along the axes present zero-score subjects for which the algorithm falsely detected CAC lesion or subjects with a non-zero score for which the algorithm missed a CAC lesion. These errors mostly comprised lesions in the coronary ostia labeled as TAC or mitral valve calcifications labeled as LCX.

2.3.4 Detection of presence versus absence of CAC

Many studies have demonstrated the negative predictive value of a zero CAC score.^{50,51} Accuracy (% , 95% CI) of detecting zero-score patients was 263/273 (96%, 94-99), 23/32 (72%, 56-87), 95/113 (84%, 77-91), 690/713 (97%, 95-98) and 1068/1268 (84%, 82-86) for the baseline algorithm in CAC-CT, AC_{PET} , diagnostic chest, RadTherapy and JHS exams, respectively (Table 2.4). For the data-specific algorithm accuracy (% , 95% CI) was 268/273 (98%, 97-100), 12/32 (38%, 21-54), 94/113 (83%, 76-90), 691/713 (97%, 96-98) and 1168/1268 (92%, 91-94) for the baseline algorithm in CAC-CT, AC_{PET} , diagnostic chest, RadTherapy and JHS exams, respectively. For the combined algorithm accuracy (% , 95% CI) was 269/273 (99%, 97-100), 22/32 (69%, 53-85), 98/113 (87%, 80-93), 701/713 (98%, 97-99) and 1149/1268 (91%, 89-92) for the baseline algorithm in CAC-CT, AC_{PET} , diagnostic chest, RadTherapy and JHS exams, respectively. Errors that assigned a subject with zero-score to the neighboring risk category, or vice versa, occurred mostly in voxels representing noise in the direct proximity of a coronary artery, or in small lesions affected by motion. Larger errors that assigned a zero-score subject to a higher risk category are described above.

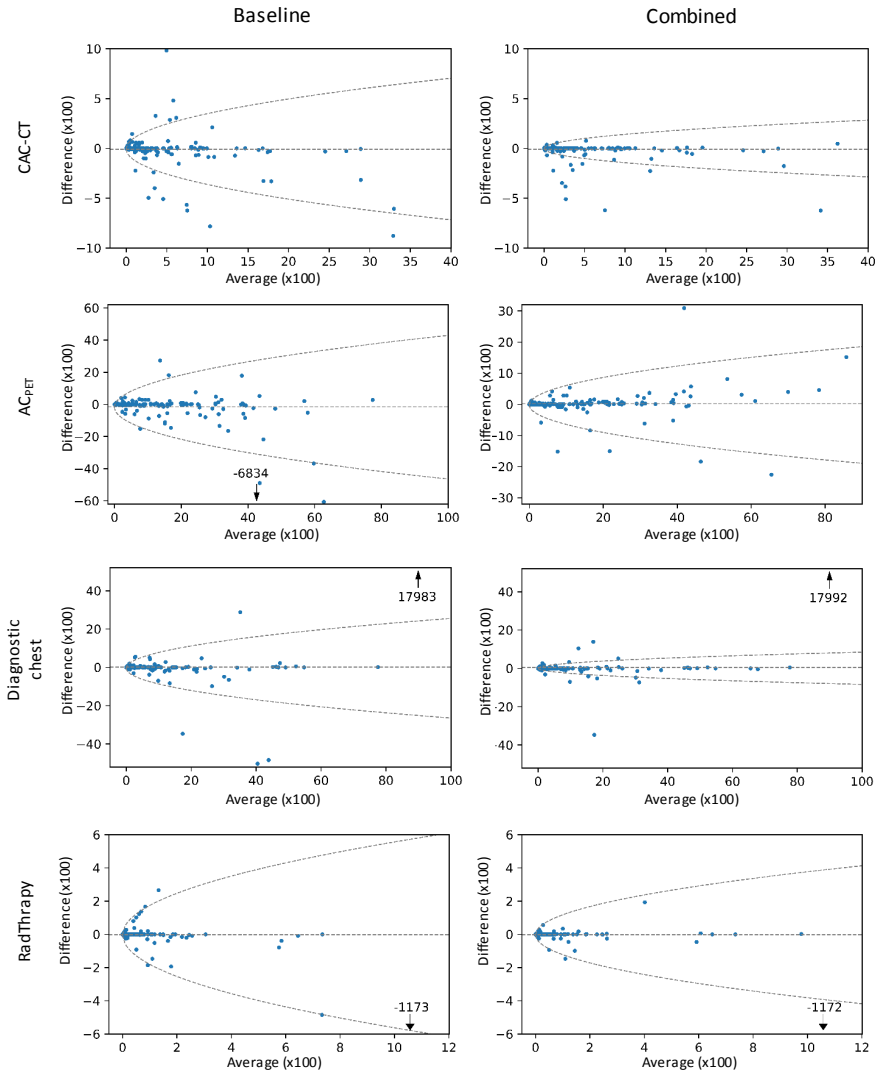
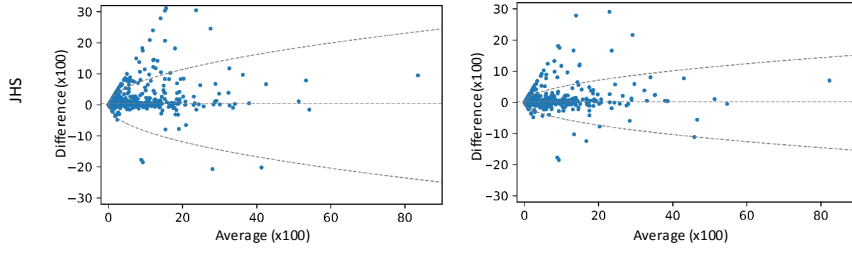


Figure 2.4: Bland-Altman plots of coronary artery calcification (CAC) Agatston scores with 95% limits of agreement (dotted lines) comparing manual scoring with automatic scoring in CAC-CT, PET attenuation correction (AC_{PET}) CT, clinical chest, radiotherapy treatment planning (RadTherapy) and CAC research CT from the Jackson Heart Study (JHS). Outliers are indicated by an arrow with the difference. 95% limits of agreement are represented by the following formula: $\text{Difference} = \pm 1.96(\pi/2)^{0.5}(b+a\text{Mean}^{0.5})$. For the baseline algorithm coefficients a and b were 4.6 and -7.1 for CAC-CT, 18.2 and -178.0 for AC_{PET} CT, 10.6 and -46.5 for diagnostic chest, 7.2 and -3.2 for RadTherapy, and 10.6 and -24.3 for JHS CT exams. For the combined algorithm coefficients a and b were 1.8 and -1.9 for CAC-CT, 7.9 and -76.2 for AC_{PET} CT, 3.4 and 2.7 for diagnostic chest CT, 4.8 and -2.4 for RadTherapy, and 6.6 and 10.8 for JHS exams.

continued...



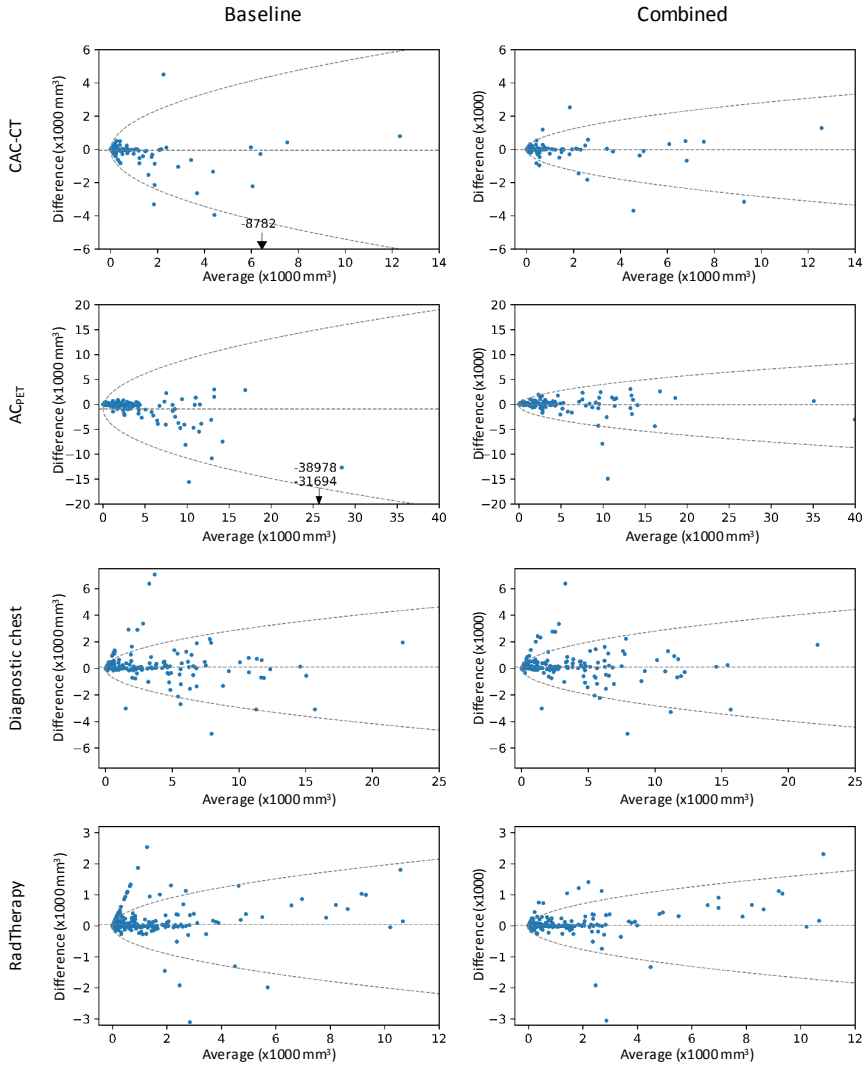


Figure 2.5: Bland-Altman plots of thoracic aorta calcification volumes (mm³) with 95% limits of agreement (dotted lines) comparing manual scoring with automatic scoring in CAC-CT, attenuation correction CT (AC_{PET}), clinical chest and radiotherapy treatment planning (RadTherapy). Outliers are indicated by an arrow with the difference. 95% limits of agreement are represented by the following formula: $\text{Difference} = \pm 1.96(\pi/2)^{0.5}(b + a\text{Mean})^{0.5}$. For the baseline algorithm coefficients a and b were 21.9 and -38.3 for CAC-CT, 40.4 and -859.2 for AC_{PET} CT, 11.9 and -18.0 for diagnostic chest, and 8.1 and 17.5 for RadTherapy exams. For the combined algorithm coefficients a and b were 11.5 and -20.5 for CAC-CT, 17.2 and -226.1 for AC_{PET} CT, 11.4 and -3.9 for diagnostic chest CT, and 6.7 and -27.1 for RadTherapy exams.

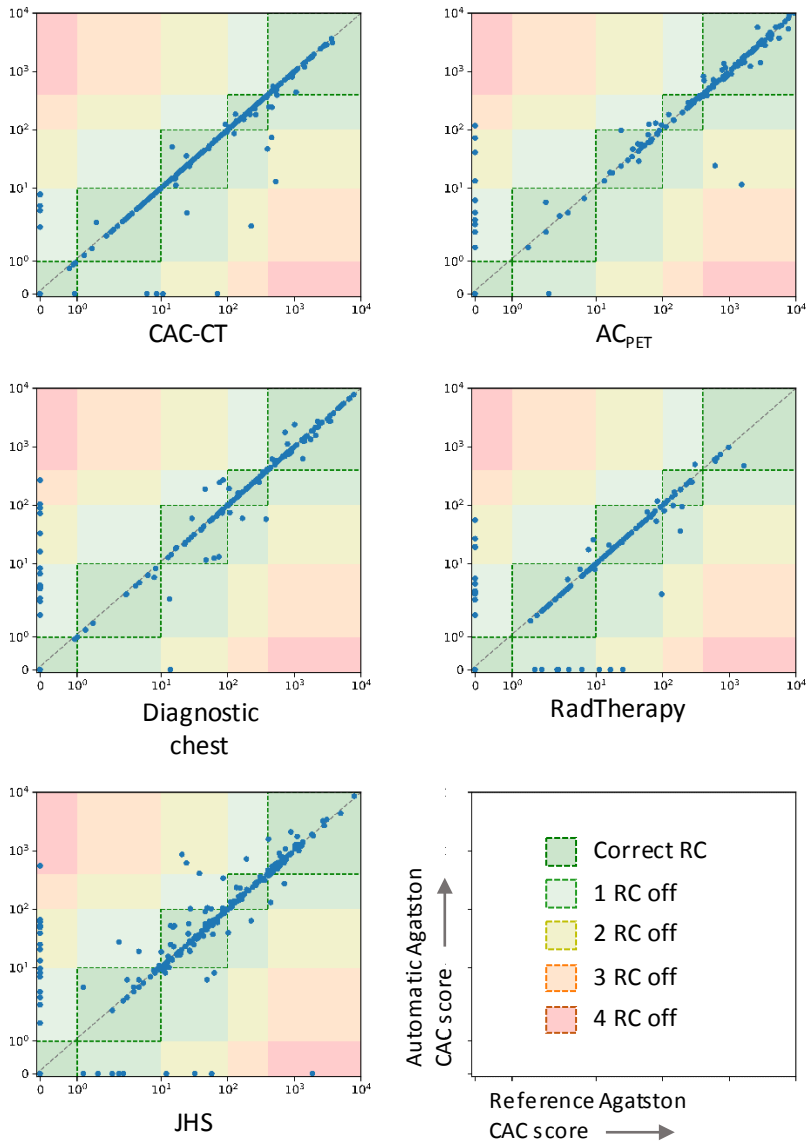


Figure 2.6: Agatston scores calculated automatically with the combined algorithm plotted against manually calculated Agatston scores for scoring in coronary artery calcification (CAC) scoring CT, attenuation correction CT (AC_{PET}), clinical chest, radiotherapy treatment planning (RadTherapy) and CAC research CT from the Jackson Heart Study (JHS). Difference between risk categories (RC) assigned by manual and automatic calcium scoring is indicated by colored blocks. CVD risk categories: 0, 1-10, 11-100, 101-400, >400. For JHS exams a random selection of 500 exams is shown for visualization purposes. Note that the scale is a log scale.

Table 2.4: Detection of zero coronary artery calcium (CAC) score scans in CAC-CT, PET attenuation correction (AC_{PET}) CT, diagnostic chest CT, radiotherapy treatment planning (RadTherapy) CT and CT exams from the Jackson Heart Study (JHS).

	Reference Score	Automatic Score									
		Baseline		Data-specific		Combined					
		0	1-10	>10	0	1-10	>10	0	1-10	>10	
CAC-CT	0	273	263	7	3	268	4	1	269	5	0
	1-10	49	4	42	3	1	47	1	2	47	0
	>10	203	0	1	202	0	1	202	2	2	199
AC_{PET} CT	0	32	23	5	4	12	12	8	22	6	4
	1-10	8	0	8	0	0	7	1	1	7	0
	>10	160	1	1	158	0	0	160	0	0	160
Diagnostic chest CT	0	113	95	8	10	94	11	8	98	8	7
	1-10	11	0	11	0	0	11	0	0	11	0
	>10	150	0	2	148	1	1	148	2	1	147
RadTherapy CT	0	713	690	11	12	691	13	9	701	8	4
	1-10	51	6	45	0	5	46	0	6	43	2
	>10	77	5	4	68	3	1	73	3	1	73
JHS CT	0	1268	1068	78	122	1168	44	56	1149	49	70
	1-10	171	27	106	38	49	109	13	43	107	21
	>10	1036	14	7	1015	27	15	994	70	21	1004

2.4 Discussion

Calcium scoring is commonly used for CVD risk prediction from non-contrast enhanced ECG-triggered cardiac CT. Alternatively, calcium scores can be derived from routine chest CT obtained for other purposes, such as lung cancer screening. In this study, we demonstrated that a deep learning method for automatic calcium scoring is robust to substantial differences in CT type (e.g., PET attenuation correction vs. low-dose chest CT) and patient population (e.g., breast cancer patients vs. heavy smokers older than 55 years) without additional training. The DL method achieved an ICC between automatic and manual reference scores of 0.79-0.97 for CAC and 0.66-0.98 for TAC over different CT protocols. Additionally, by augmenting the baseline training data with a relatively small data-specific set improved the performance to the level achieved on the data the network was originally developed for (ICC range 0.84-0.99 for CAC and 0.92-0.99 for TAC). Furthermore, training one instance with a combination of all included image types resulted in a network with similar performance (ICC range 0.85-0.99 for CAC and 0.96-0.99 for TAC).

Calcium scoring is a challenging task in several of the included image types. While in literature interobserver agreement is high in dedicated CAC-CT (Agatston ICC=0.99 and risk category $\kappa=0.99$),⁴⁴ it is lower in AC_{PET} ($\kappa=0.94$ for 4 risk categories).⁵² Similarly, the performance of the baseline network varied over the datasets. We demonstrated that in various types of CTs, the agreement of automatically and manually assigned CVD risk categories was high (κ range 0.85-0.95) and it approached interobserver agreement.^{44,52,53}

After CT protocol-specific training, the DL method performed similarly to previous methods (κ 0.84 vs. 0.85 for AC_{PET} ³⁴) or outperformed those (κ 0.98 vs. 0.91 for CAC-CT, 0.89 vs 0.80 for RadTherapy, 0.91 vs 0.80 for chest CT^{12,44,54}). Note that due to different datasets a direct comparison cannot be made, except for CAC-CT, in which the same test set was used.⁴⁴ While the use of CT protocol-specific training resulted in improved risk category assessment for CAC-CT, RadTherapy and JHS exams, for AC_{PET} , agreement of risk categories slightly decreased. Visual inspection of the results revealed that the AC_{PET} -specific network segmented small false positive lesions, mostly representing noise in the vicinity of the coronaries, more often than the baseline, resulting in the incorrect CVD risk categorization in the lowest risk categories. The AC_{PET} exams suffered more from noise than the other types of evaluated exams. Moreover, since for AC_{PET} only 199 representative exams were added to the training set, a possible explanation of degrading performance could be insufficient variation in the representative training examples. By combining different types of exams into one training set the network is able to score all included types of CTs with high reliability compared with manual scoring.

For assessment of CVD risk, it is relevant to distinguish presence and absence of CAC.⁵¹ Since the accuracy for detecting zero-score patients is excellent (268/273

(98% for the data-specific algorithm) and the false positive rate is low (1/256 (0%)), the presented method would be suited to rule out presence of calcifications in patients in dedicated CAC-CT. In other CT protocols, errors occurred more frequently (accuracy from 12/32 (38%) in AC_{PET} to 691/713 (97%) in RadTherapy), but one should note that in these more challenging exams, such errors are not uncommon for human readers either.

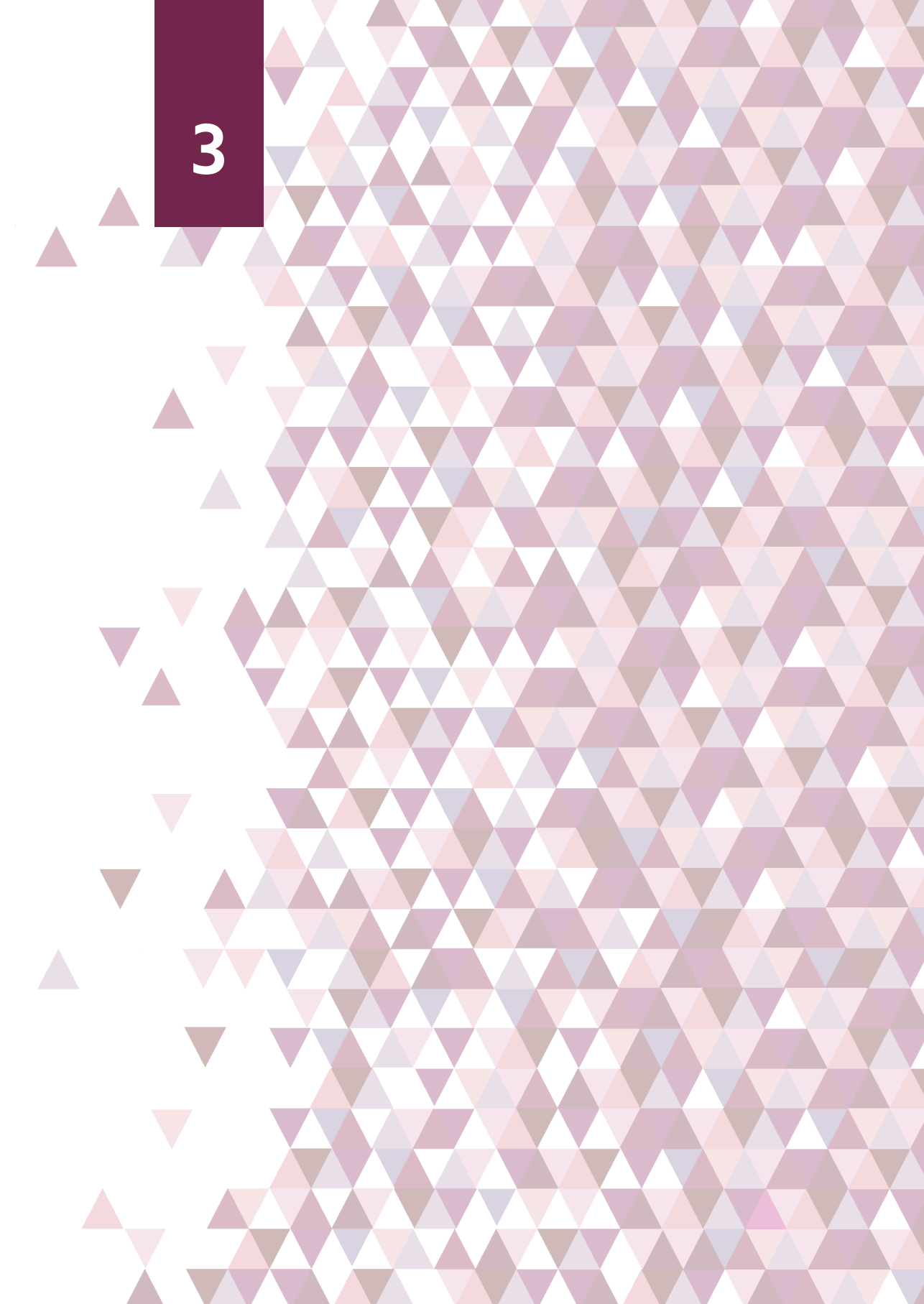
Multiple strategies have been proposed for adapting DL networks to a different task or different input data. One well-known and well-performing approach is fine-tuning a pretrained network.^{55,56} In our study the baseline training set was available, therefore, instead we retrained the networks from initialization with a combination of CT protocol-specific data and the baseline. This approach improved the performance of the network compared with the baseline. In situations where the baseline dataset is not available, fine tuning could be an interesting alternative.

This study has several limitations. Firstly, instead of using pixel-wise annotation, as in the NLST exams,³² we annotated lesions by identifying 3D clusters of voxels above 130 HU matching standard calcium scoring. Since the training set mostly consists of NLST scans, this difference in training and test references might have influenced the performance of the networks. However, given the automatically obtained segmentations are postprocessed matching standard calcium scoring, we expect this influence to be marginal. Secondly, reference calcium scores in the JHS exams were obtained with a different protocol.⁴² However, since both manual and automatic scoring methods identified 3D clusters of voxels above 130 HU, we expect the discrepancy to be negligible. Thirdly, since most CTs were acquired for other purposes than calcium scoring, not all scans were acquired according to SCCT/SRT guidelines for calcium scoring in non-cardiac chest CT.⁵⁷ However, updated AHA/ACC guidelines²³ focus on presence vs. absence and risk categorization. Relatively small shifts in absolute score are of uncertain clinical significance. Moreover, since the performance of the DL method is excellent, it can potentially help radiologists report CAC in non-contrast chest CT as SCCT/SRT guidelines advice. Lastly, we did not investigate how many representative exams should be added to the training set to ensure optimal performance, but estimated the amount of data-specific exams based on the calcium burden in the relevant population.

In conclusion, this study presents the validation of a deep learning algorithm in large and diverse sets of CT exams. The results show that the method adapted well to a previously unseen CT type when a few representative training examples were added to the large training set. A single combined deep learning model trained with all CT types performed as well as specialized models, indicating potential for use in clinical practice.

Acknowledgments

The authors thank the Dutch Cancer Society for supporting this research (NCT03206333). We are grateful to the United States National Cancer Institute (NCI) for providing access to NCI's data collected by the National Lung Screening Trial. The statements contained in this paper are solely ours and do not represent or imply concurrence or endorsement by NCI. The authors also wish to thank the staffs and participants of the JHS. The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.



CHAPTER 3

Arterial calcifications in Western and Asian breast cancer patients

BASED ON: S. A. M. Gernaat*, S. G. M. van Velzen*, V. Koh, M. J. Emaus, I. Išgum, N. Lessmann, S. Moes, A. Jacobson, P. W. Tan, D. E. Grobbee, D. H. J. G. van den Bongard, J. I. Tang, and H. M. Verkooijen. "Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of western and asian breast cancer patients", *Radiotherapy and Oncology*, vol. 127 (2018), 487–492 *Both authors contributed equally.

Abstract

PURPOSE This study automatically quantified calcifications in coronary arteries (CAC) and thoracic aorta (TAC) on breast planning computed tomography (CT) scans and assessed its reproducibility compared to manual scoring.

MATERIAL AND METHODS Dutch ($n = 1199$) and Singaporean ($n = 1090$) breast cancer patients with radiotherapy planning CT scan were included. CAC and TAC were automatically scored using deep learning algorithm. CVD risk categories were based on Agatston CAC: 0, 1–10, 11–100, 101–400 and >400 . Reliability between automatic and manual scoring was assessed in 120 randomly selected CT scans from each population, with linearly weighted kappa for CAC categories and intraclass correlation coefficient for TAC.

RESULTS Median age was higher in Dutch patients than Singaporean patients: 57 versus 52 years. CAC and TAC increased with age and were more present in Dutch patients than Singaporean patients: 24.2% versus 17.3% and 73.0% versus 62.2%, respectively. Reliability of CAC categories and TAC was excellent in the Netherlands (0.85 (95% confidence interval (CI) = 0.77–0.93) and 0.98 (95% CI = 0.96–0.98) respectively) and Singapore (0.90 (95% CI = 0.84–0.96) and 0.99 (95% CI = 0.98–0.99) respectively).

CONCLUSIONS CAC and TAC prevalence was considerable and increased with age. Deep learning software is a reliable method to automatically measure CAC and TAC on radiotherapy breast CT scans.

3.1 Introduction

Cardiovascular disease (CVD) is an important cause of death in women with breast cancer, with 24% of patients over 65 years dying of this disease.^{58,59} The risk of CVD following breast cancer is increased by exposure to cardiotoxic breast cancer therapies such as radiation therapy,^{60–62} anthracycline-based chemotherapy,⁴ and trastuzumab.^{63,64} Breast cancer patients with pre-existing CVD risk factors have the highest risk of treatment induced cardio toxicity.⁶⁵

Coronary artery calcification (CAC) and thoracic aorta calcification (TAC) are markers of atherosclerosis.^{66–69} In clinical practice, these calcifications are measured on cardiac computed tomography (CT) scans.^{70,71} CAC, measured on cardiac CT scans, is an independent predictor of CVD events and mortality.^{72,73} Similarly, TAC, measured on cardiac CT scans, is associated with an increased risk of coronary heart disease, independent of CAC and other CVD risk factors.^{74,75} The majority (>60%) of breast cancer patients undergoes radiation therapy after surgery and receives a non-contrast enhanced CT scan of the breasts including the heart for individual radiation treatment planning. CAC and TAC can be visualized and quantified on these scans.¹² Information on CAC and TAC can be used to estimate the risk of CVD in the individual breast cancer patient, and thus to indicate who could benefit from cardiac monitoring, cancer therapy with reduced cardiotoxicity and/or cardioprotective medication.^{2,76}

In the present study, we applied a new deep learning algorithm for automated quantification of CAC and TAC on planning CT scans.³² Differences in prevalence of (subclinical) atherosclerosis, coronary artery calcifications, and CVD risk factors have been observed among Western and Asian women without breast cancer, with the highest prevalence in the Western population.^{77–79} This study aimed to investigate the prevalence and quantity of CAC and TAC, as well as CAC based CVD risk in Western (i.e. the Netherlands) and Asian (i.e. Singapore) breast cancer patients.

3.2 Methods

3.2.1 Design and study population

This cross-sectional study included women diagnosed with in situ or invasive breast cancer who underwent a radiotherapy planning CT scan at the Radiation Oncology Departments of the University Medical Center Utrecht (UMCU) in the Netherlands or the National University Hospital of Singapore (NUHS) in Singapore. At the UMCU, breast cancer patients were included in the context of the Utrecht cohort for Multiple BReast cancer intErvention studies and Long-term evaLuAtion (UMBRELLA) study, which includes all breast cancer patients referred to the Radiation Oncology Department of the UMCU since 2013.⁸⁰ Here, patients gave informed consent for use of their routine clinical, imaging, pathology, and follow up data. 1199 patients, who were enrolled in

UMBRELLA between 2013 and 2016, were included in the current study. At NUHS, 1090 breast cancer patients with a radiotherapy planning CT scan were included from January 2005 (date of introduction of digital planning CT scan) to September 2015. We excluded one patient's CT scan as severe anatomical deformities of the chest limited the calcification detection. Information on date of birth was extracted from the patient's CT scan. The Medical Ethical Review Board waived the need for informed consent of patients in Singapore.

3.2.2 Imaging data

At the UMCU, planning CT scans were conducted with a Philips Brilliance Big Bore CT scanner using 16 x 0.75 mm collimation, 120 kVp, 3 mm slice thickness. At NUHS, all planning CT scans were conducted with a Philips Brilliance Big Bore CT scanner using 16 x 1.5 mm collimation, 120 kVp, 5 mm slice thickness. All breast cancer patients received a CT scan during free breathing, without contrast enhancement or ECG-synchronization. At the UMCU, the majority of patients planned for left-sided and/or parasternal irradiation underwent an additional deep inspiration breath-hold scan. For patients with multiple CT scans (49.5% at UMCU and 2.1% at NUHS, the first CT scan was used. In the UMCU, if two scans were made on the same day (i.e. one with and one without deep inspiration breath-hold technique), the scan with breath-hold technique was used.

3.2.3 Manual annotation

To define the reference for training and evaluation of the automatic calcium scoring algorithm, CAC and TAC were manually annotated in 803 planning CT scans (i.e. 683 CT scans from UMCU and 120 CT scans from NUHS). A random selection of 240 CT scans (i.e. 120 CT scans from UMCU and 120 CT scans from NUHS) was done for comparing manual annotation with automatic annotation. Manual annotation was performed using 3D region growing of voxels representing calcifications above the standard threshold of 130 Hounsfield Units (HU). The reference annotation was conducted by trained medical students (AJ and SM) and PhD students (SGMvV and SAMG), each with experience in over 500 CT scans prior to performing the automatic segmentations.

For training of the CNN based localization algorithm bounding boxes around the chest, similar to the field of view in low-dose chest CT scans, 45 CT scans from the UMCU were manually annotated. These training scans were not part of the test data for calcium scoring in this study and selected to have a variety in field of view representative for the whole dataset.

3.2.4 Automatic calcification quantification

All CT scans were automatically scored for atherosclerotic calcifications in the thoracic aorta and coronary arteries, i.e. left coronary artery (LAD) including the left main coronary artery, left circumflex (LCX), and right coronary artery (RCA), using a method based on deep learning described by Lessmann et al.³²

Briefly, the algorithm consists of two consecutive convolutional neural networks (CNN) labeling voxels as calcium in the LAD, LCX, RCA, the aorta or cardiac valves. To identify calcified voxels and label them according to their anatomical location, the network needs to learn from larger spatial context, but it does not necessarily require analysis of detailed local information. Therefore, the first CNN exploits a large field of view (105 mm x 105 mm). To keep the number of trainable parameters limited, dilated convolutions are used. To accurately determine which of the identified voxels represent calcifications, the second network analyzes detailed local texture. Hence, the receptive field of this CNN is 52 mm x 52 mm. This network performs two-class classification to identify true calcifications among the potential calcifications detected by the first network. This second CNN does not alter anatomical labels of calcified voxels. Because the accuracy of calcium detection in the cardiac valves was not highly accurate in the original work, it was not utilized in the present study.

This algorithm was developed to analyze low-dose chest CT scans. Therefore, to apply the method to the data used in this study, all radiotherapy treatment planning scans were cropped to a standardized field of view similar to the field of view of chest CT scans. This was performed using CNN based localization described by de Vos et al.⁸¹ This localization algorithm was trained with 45 UMBRELLA scans (40 for training and 5 for validation), with manually annotated bounding boxes.

Thereafter, the CAC scoring algorithm was trained with data originally used to develop the algorithm (1181 low-dose chest CT scans from the National Lung Screening Trial)⁸² and with 563 cropped radiotherapy planning CT scans of breast cancer patients within UMBRELLA). The first CNN was trained using 340 epochs (5 batches, 50 minibatches of size 86) and the second CNN using 80 epochs (10 batches, 100 minibatches of size 512), with a learning rate of 0.0005 and 0.001 respectively. During training a batch ratio of 20% UMBRELLA and 80% NLST examples was enforced. Note that the algorithm identified voxels representing calcifications and not lesions as standardly performed in clinic. Hence, to identify calcified lesions following clinical standard and our manual annotation protocol, 3D region growing with 130 HU threshold was applied to all automatically identified calcification voxels. For lesions with multiple labels, the label with the largest volumewise contribution was chosen for the whole lesions.

CAC was expressed in the Agatston score, calculated by multiplying density score (corresponding to the highest HU of the calcification area) and calcification area (mm²), and each patient was assigned to one of five CVD risk categories: 0, 1–10, 11–100,

101–400, >400.^{46,47,53,83} TAC was expressed in volume scores (mm^3). Scans with extremely high automatically detected CAC score (>1000) ($n = 10$) and TAC volume (>10,000 mm^3) ($n = 4$) were manually inspected and corrected if needed.

To quantitatively evaluate reproducibility of the automatic and manual calcium scoring methods, 120 CT scans were randomly selected from the UMCU and NUHS. As such, the total test set included 240 CT scans. Note that test scans were not used for training the calcium scoring methods in any way.

Recommended slice thickness for CAC scoring is 3 mm, which was used for planning CT scans from the Netherlands. However, planning CT scans from Singapore were conducted with 5 mm slice thickness. To evaluate the influence of slice thickness on the automatic CAC score, we resampled the CT scans from UMCU to 5 mm slice thickness with weighted average to increase approximate image acquisition. To ensure that there were no differences between the identified calcifications in the acquired and resampled CT scans, calcium maps were resampled accordingly. Only calcification voxels exceeding 130 HU threshold in the 5 mm CT scans were retained. Subsequently, CAC scores and corresponding CVD risk categories were calculated for the CT scans with 5 mm slice thickness and compared with the scores of the original CT scans with 3 mm slice thickness.

3.2.5 Statistical analysis

Median (interquartile range (IQR)) was used to describe continuous variables with skewed distributions. The number of breast cancer patients with CAC Agatston scores above zero and above 100 was calculated per age category for Dutch and Singaporean patients separately. In addition, median (IQR) volume (in mm^3) of TAC was calculated per age category. Differences between breast cancer patients from the Netherlands and Singapore were tested with Chi-Square Test (or Fisher's exact test when the cell count was less than five) in case of categorical variables, and with Mann-Whitney U Test in case of continuous variables with skewed distributions. To maximize the number of images, and thereby obtain most reliable insight in calcification prevalence, this analysis was performed utilizing the complete set of available images.

The performance of the software, i.e. reproducibility, was evaluated with reliability and agreement measures using 240 CT scans from the test set. Reliability (agreement beyond chance) of CAC categories was assessed with Cohen's linearly weighted kappa.⁸⁴ Reliability of continuous CAC Agatston score and TAC volume was assessed with intraclass correlation coefficient (ICC). The two way random effects and absolute agreement ICC was used to assess reliability between automatic and manual calcification scoring, taking into account the variance between patients and structural differences between automatic and manual calcification scoring. Agreement (degree to which CAC categories are identical between automatic versus manual CAC scoring method) was assessed with proportional agreement. Analyses were performed with IBM SPSS

Table 3.1: Distribution of age and cardiac calcification in 2288 breast cancer patients by country of residence.

	The Netherlands Western population n = 1,199	Singapore Asian population n = 1,089	P-value
Age at time of CT scan, years, median (IQR)	57 (50–66)	52 (45–60)	<0.01
Age at time of CT scan in categories, years, % (n)			<0.01†
<41	5.7 (68)	12.6 (137)	
41-50	20.9 (250)	32.2 (351)	
51-60	32.8 (393)	32.4 (353)	
61-70	28.9 (346)	16.1 (174)	
>70	11.8 (142)	6.7 (73)	
Coronary artery calcification in Agatston score categories, % (n)			<0.01‡
0	75.8 (909)	82.7 (901)	
1-10	8.8 (106)	3.9 (43)	
11-100	8.8 (106)	7.1 (77)	
101-400	4.4 (53)	3.9 (43)	
>400	2.1 (25)	2.3 (25)	
Thoracic aortic calcification			
Median (IQR) in volume (mm ³)	91 (0–692)	41 (0–495)	<0.01
Patients with calcifications, % (n)	73.0 (875)	62.2 (677)	<0.01

Abbreviations: CT = computed tomography, IQR = interquartile range.

* Statistically significant differences between countries were tested with the Chi-Square Test in case of a categorical variable (or Fisher's exact test when the cell count was less than five) and with the Mann-Whitney U Test in case of a continuous variable.

† All categories except for 51–60.

‡ Categories 0 and 1–10.

statistics version 23 and an online statistical tool.⁴⁹

3.3 Results

Median age at time of planning CT scan was higher in breast cancer patients from the Netherlands than from Singapore: 57(IQR = 50–66) versus 52 years (IQR = 45–60), $p < 0.01$ (Table 3.1). The prevalence of CAC was higher in patients from the Netherlands than from Singapore: 24.2% versus 17.3% ($p < 0.01$). Among these, 6.5% and 6.2% of patients from the Netherlands and from Singapore had a CAC score above 100. Similarly, the prevalence of TAC was higher among patients from the Netherlands than from Singapore (73.0% versus 62.2%, $p < 0.01$). Three CT scans, with an automatically determined CAC score exceeding 1000 were corrected. Two were corrected to a higher CAC score (from 4761 to 5473 and from 1243 to 1310) and one CT scan was corrected to a lower CAC score (from 1191 to 721).

The prevalence of CAC increased with higher age from 3% for patients younger than 41 years to 38% for patients between 41 and 70 years (Figure 3.1). The prevalence

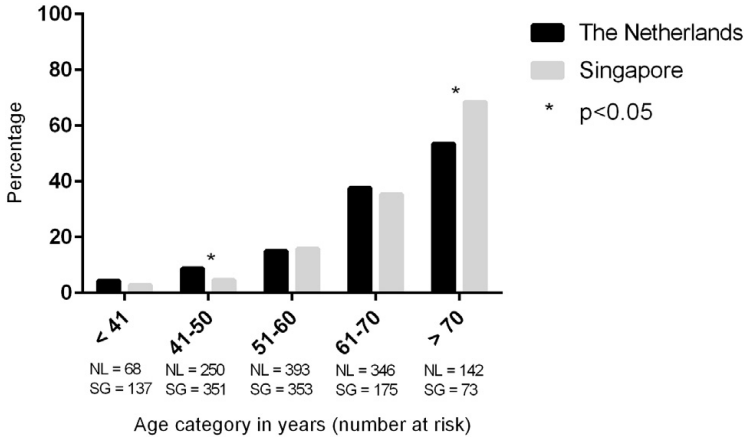


Figure 3.1: Distribution of breast cancer patients with a coronary calcification (Agatston) score above zero by age category and country of residence.

of CAC was higher in breast cancer patients of 70 years or older from Singapore than the Netherlands (70% versus 55%, $p < 0.05$). CAC scores over 100 were present in 10% to 15% for patients between 61 and 70 years, and in 18% to 29% for patients of 70 years and older (Figure 3.2). Median TAC volumes (mm^3) increased with age from 41 mm^3 for patients between 51 to 60 years to almost 2000 mm^3 for patients of 70 years or older (Figure 3.3).

3.3.1 Impact of difference in slice thickness

Using the resampled CT scans from the Netherlands (5 mm), the prevalence of CAC was 13.3% and 4.1% of breast cancer patients had a CAC score above 100 compared to 24.4% and 6.5% respectively, using the originally acquired CT scans from the Netherlands (3 mm).

3.3.2 Performance of automatic analysis

Using the test set from the Dutch population, reliability of CAC categories assessed with linearly weighted kappa was 0.85 (95% confidence interval (CI) = 0.77–0.93) and slightly higher for scans with deep inspiration breath-hold (Table 3.2). Proportion of agreement for CAC categories was 0.87 (95% CI = 0.79–0.92) and slightly higher for scans with deep inspiration breath-hold. For continuous CAC score, the ICC was 0.95 (95% CI = 0.93–0.97) and slightly higher for scans with deep inspiration breath-hold. For TAC volume, the ICC was 0.98 (95% CI = 0.96–0.98). Using the test set from the Singaporean population, linearly weighted kappa was 0.90 (95% CI = 0.84–0.96) and proportion of agreement was 0.90 (95% CI = 0.84–0.95) for CAC categories. The ICC was 0.99 (95% CI = 0.98–0.99) for TAC volume.

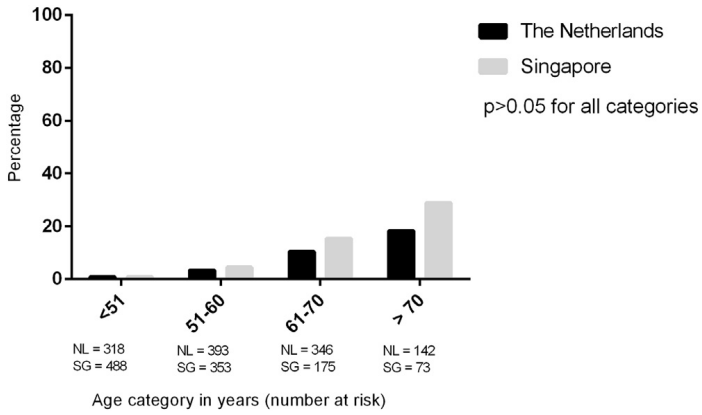


Figure 3.2: Distribution of breast cancer patients with a coronary calcification (Agatston) score above 100 by age category and country of residence.

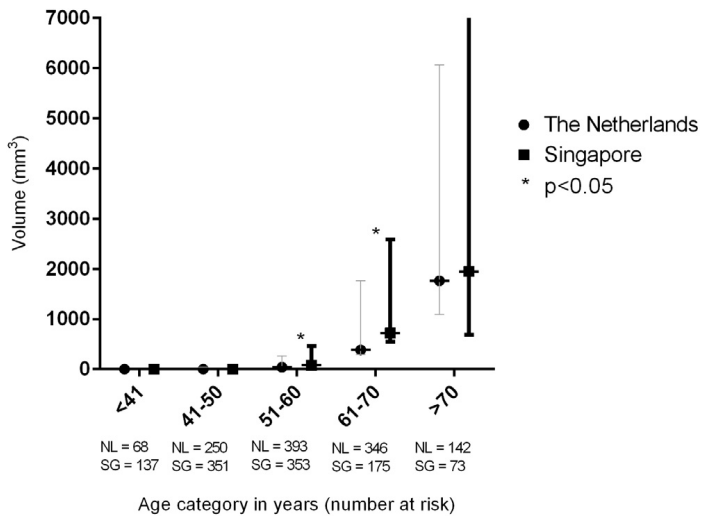


Figure 3.3: Median thoracic aortic calcification (interquartile range) in breast cancer patients by age category and country of residence.

Table 3.2: Reproducibility of calcifications in coronary arteries and thoracic aorta comparing automatic versus manual scoring on radiotherapy planning CT scans of breast cancer patients stratified by country of residence.

	CAC (Agatston) scores				TAC volume (mm ³)	
	Categorical		Continuous		Continuous	
	kappat(95% CI)	Proportion of agreement (95% CI)	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)
The Netherlands						
Automatic vs. manual (n = 120)	0.85 (0.77–0.93)	0.87 (0.79–0.92)	0.95 (0.93–0.97)	0.98 (0.96–0.98)		
Breath-hold (n = 40)	0.88 (0.76–1.00)	0.90 (0.75–0.97)	1.00 (0.99–1.00)	0.97 (0.96–0.98)		
Without breath-hold (n = 80)	0.84 (0.75–0.94)	0.85 (0.75–0.92)	0.94 (0.90–0.96)	1.00 (0.99–1.00)		
Singapore						
Automatic vs. manual						
Without breath-hold (n = 120)	0.90 (0.84–0.96)	0.90 (0.84–0.95)	0.99 (0.99–1.00)	0.99 (0.98–0.99)		

Abbreviation: CI = confidence interval, CAC = coronary artery calcification, TAC = thoracic aorta calcification, ICC = intraclass correlation coefficient. Cohen's linearly weighted kappa.

3.4 Discussion

The prevalence of CAC and TAC in breast cancer patients from the Netherlands and Singapore is considerable and increases with age. Up to one third of breast cancer patients aged under 70 years has CAC. More than half of patients aged over 70 years has CAC with higher prevalence in patients from Singapore than the Netherlands (70% versus 55%). TAC is present in two thirds of patients from the Netherlands and Singapore. This study also showed that the new evaluated automatic scoring method based on deep learning algorithm is a reliable method to measure CAC and TAC on radiotherapy planning CT scans.

Knowing a breast cancer patient's CVD risk is important, especially when treatment with cardiotoxic chemotherapy is considered, which occurs mainly among patients aged under 70 years. In these cases, together with the patient, physicians may decide to timely start with cardioprotective medication or opt for cancer therapy with reduced cardiotoxicity.^{2,76,85} Patients with a high risk of CVD can be monitored before, during, and after treatment, to reduce their risk of CVD by treatment with antihypertensives, statin therapy or adopting a healthier lifestyle.^{2,76} The information can be used for patient counseling prior to radiotherapy and would help patients understand their individual CVD risk as compared to their risk of breast cancer recurrence. This would be especially significant in elderly patients with early stage breast cancer.⁸⁶

When using originally acquired CT scans, the present study found a lower prevalence of CAC in Singaporean patients compared to Dutch patients which was mainly originating from differences in CVD risk categories 0 and 1–10. This may be explained by the difference in slice thickness, which was larger in scans from Singapore (5 mm) than the Netherlands (3 mm). Studies, both in phantoms and in patients, showed that the CAC score decreased with increasing slice thickness.^{87,88} An estimation of the prevalence of CAC in Dutch patients showed that the prevalence of CAC would likely be lower when scans would have been acquired with 5 mm thick slices. This indicates that the prevalence of CAC in Singaporean breast cancer patients might have been underestimated in this study.

The present study shows high reproducibility of automatic compared to manual CAC and TAC scoring on breast planning CT scans of both populations. As such, the deep learning method can be used to measure CAC and TAC in different populations with different disease characteristics and scan protocols. The automatic CAC scoring algorithm using deep learning performed better than the previous software based on supervised pattern recognition: reliability between automatic and manual CAC scoring ranged from 0.94 to 1.00 in the present study and from 0.66 to 0.95 in a previous study using radiotherapy planning CT scans.¹² Planning CT scans performed with deep inspiration breath-hold have a higher reproducibility of automatic calcification detection, probably due to the fact that these scans have less motion artifacts compared with scans performed during free breathing. The use of deep inspiration breath-hold

is a technique that significantly reduces the irradiated volume and dose to the heart compared to free breathing, especially for patients with left-sided local or internal mammary lymph node radiotherapy.⁸⁹

Both the software based on supervised pattern recognition (İşgum et al.⁹⁰) and the deep learning algorithm (Lessmann et al.³²) were developed for low-dose, non-dedicated chest CT scans acquired in a lung cancer screening trial. However, there is a major difference in the employed technology. The method described by İşgum et al.⁹⁰ exploits a conventional machine learning approach where the discriminative features describing potential calcifications are handcrafted and subsequently used by a classifier that discriminates true coronary artery calcifications from other potential calcifications. In contrast, the method proposed by Lessmann et al.³² employs a deep learning approach where a network learns discriminative features directly from the data itself. Moreover, the two methods differ in tasks. While the method proposed by İşgum et al.⁹⁰ detects only coronary calcifications, the algorithm proposed by Lessmann et al.³² identifies coronary calcifications as well as calcifications in the thoracic aorta. Since TAC is a stronger predictor for non-coronary cardiovascular events than CAC, this is valuable additional information.

Typically, a large set of CT scans is needed to train deep learning algorithms. As manual calcium scoring is cumbersome and time consuming due to image acquisition without ECG-triggering, it was not feasible in the present study to manually annotate all planning CT scans. The deep learning algorithm, however, identified CAC and TAC on all available planning CT scans to increase evaluation of the prevalence and distribution of CAC and TAC in breast cancer patients from the Netherlands and Singapore. This may have introduced a small bias toward manual annotations in the CT scans used for training.

Our results are in contrast with those of another Dutch study looking at CAC in patients with noninvasive or invasive breast cancer treated with breast-conserving surgery between 2008 and 2010.⁹¹ This study reported a CAC prevalence of 47% (unadjusted for age), which is almost double the prevalence reported in the current study. Both study populations were of comparable age. Unlike the present study, Mast et al. (2012) manually detected CAC.⁹¹ It is, however, unlikely that the difference in CAC scoring (automatic versus manual) explains the difference in CAC prevalence between the current study and those reported by Mast et al.,⁹¹ as we showed high reproducibility of automatic versus manual CAC scoring on radiotherapy planning CT scans. As the breast cancer population of Mast et al. consisted only of 80 patients, their reported CAC estimates are less precise than those CAC estimates reported in the current study.⁹¹

The Multi-Ethnic Study of Atherosclerosis (MESA) investigated CAC and TAC in women free from clinically apparent CVD with different ethnic origins living in the United States.⁷⁰ MESA used electrocardiogram (ECG)-triggered cardiac CT scans to detect calcifications.⁷⁰ CAC was more present in white women (44.6%) than in Chinese women (36.6%) (unadjusted for age).⁷⁸ Compared to our study, this study

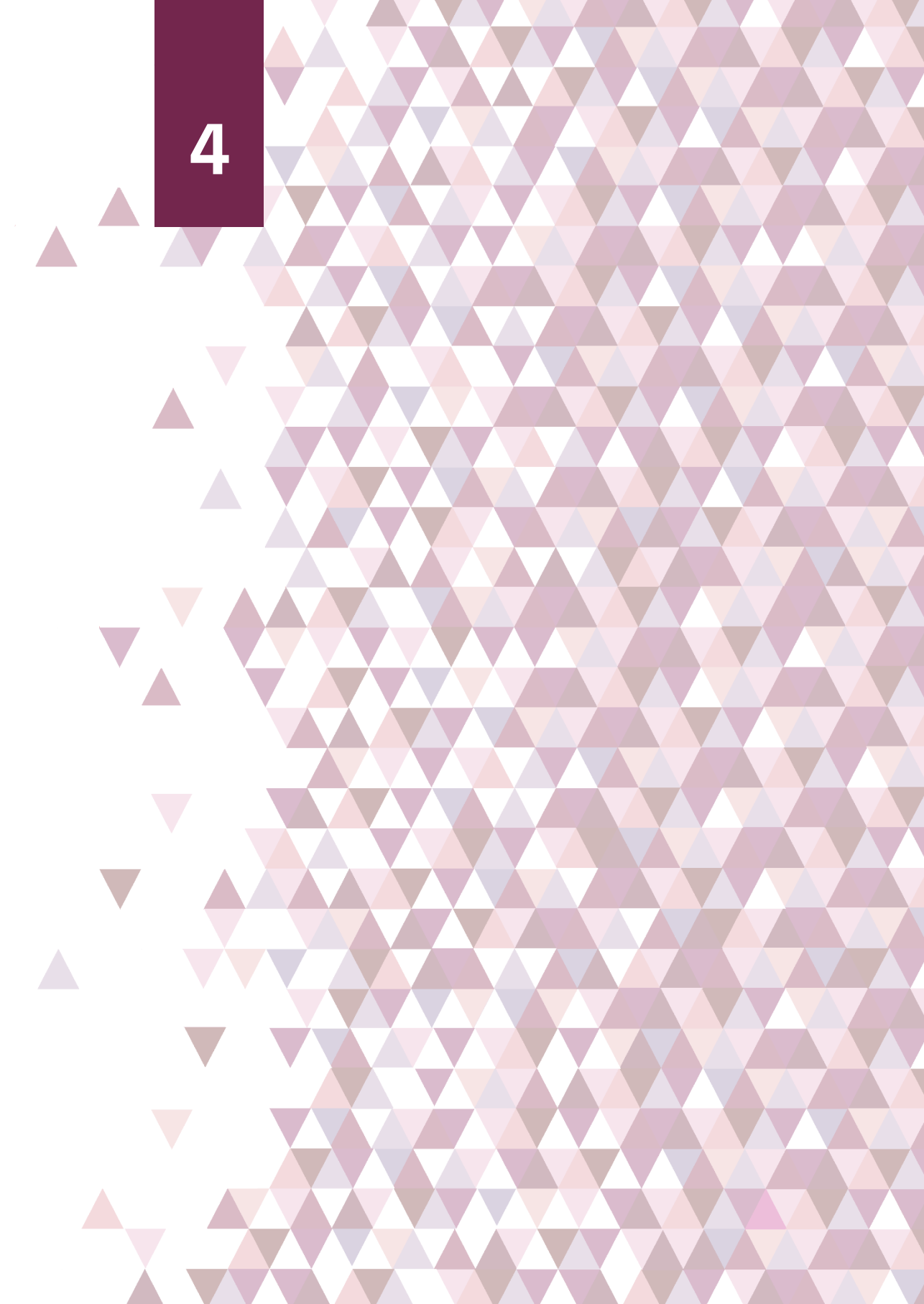
population was older which may (partly) explain the higher CAC prevalence.⁷⁸ Kanaya et al. compared prevalence of CAC in white women and in South Asian women using data from a community-based cohort of asymptomatic women in the United States.⁹² Non-contrast enhanced cardiac CT scans were used to assess the CAC.⁹² Similarly to our study, South Asian women aged over 70 years had higher CAC scores than white women aged over 70 years. Furthermore, the prevalence of TAC has been evaluated in women aged between 42 and 50 years participating in the Healthy Women Study in the United States.⁹³ CAC was measured on electron beam tomographic scans.⁹³ This study reported a prevalence of TAC (volume not specified) of 78% which is comparable to the prevalence of TAC reported in the current study (73%).⁹³

We acknowledge that the association between CAC and TAC (automatically) detected on radiotherapy planning CT scans and the risk of cardiovascular events has not been established. Studies that have investigated CAC scores detected in ungated low-dose chest CT scans acquired in screening of heavy smokers, found that these CAC scores were independently associated with CVD events and all-cause mortality.^{94–96} The Bragatston study is a study on Automated Quantification of Coronary Artery Calcifications on Radiotherapy Planning CTs for Cardiovascular Risk Prediction in Breast Cancer Patients (NCT03206333). Bragatston is an ongoing multicenter study led by the University Medical Center Utrecht,⁹⁷ with the aim to quantify the association between TAC and CAC and risk of CVD, and to optimize and validate the newly UMCU developed automated software based on deep learning.³²

To conclude, the prevalence of CAC and TAC is considerable in breast cancer patients both from the Netherlands and Singapore, and increases with age. Up to one third of breast cancer patients aged under 70 years has CAC and these patients often receive cardiotoxic breast cancer treatments. Early identification of breast cancer patients at high risk of CVD may help physicians find a good balance between optimal tumor control and minimal treatment induced CVD.

Acknowledgments

This research was supported by the Dutch Cancer Society (NCT03206333).



CHAPTER 4

Identification of breast cancer patients at risk of cardiovascular disease from radiotherapy planning CT scans

BASED ON: R. Gal, S. G. M. van Velzen, M. J. Hooning, M. J. Emaus, F. van der Leij, M. L. Gregorowitsch, E. L. Blezer, S. A. Gernaat, N. Lessmann, M. G. Sattler, T. Leiner, P. A. de Jong, A. J. Teske, J. Verloop, J. J. Penninkhof, V. Ilonca, H. Meijer, J. J. van Tol-Geerdink, J.-P. Pignol, D. H. van den Bongard, I. Išgum, and H. M. Verkooijen. "Identification of breast cancer patients at risk of cardiovascular disease by automatic quantification of coronary artery calcifications on radiotherapy planning CT scans". Submitted.

Abstract

IMPORTANCE Cardiovascular disease (CVD) is common in patients treated for breast cancer, especially in patients treated with systemic treatment and radiotherapy and in those with pre-existing CVD risk factors. Coronary artery calcium (CAC), a strong independent CVD risk factor, can be automatically quantified on radiotherapy planning computed tomography (CT) scans and may help identify patients at increased CVD risk.

OBJECTIVE To evaluate the association of CAC with CVD and coronary artery disease (CAD) in breast cancer patients.

DESIGN, SETTING, AND PARTICIPANTS Multicenter cohort study of 15,915 breast cancer patients receiving radiotherapy between 2005-2016 who were followed up through December 31, 2018. Age, calendar time and treatment adjusted Cox proportional hazard models were used to evaluate the association of CAC with CVD and CAD.

EXPOSURES CAC scores were automatically extracted from planning CT scans using a deep learning algorithm. Patients were classified into Agatston risk categories (0,1-10,11-100,101-399,>400 units).

MAIN OUTCOMES AND MEASURES Occurrence of (non-)fatal CVD and CAD were obtained from national registries.

RESULTS Seventy percent (n=11,179) had no CAC. CAC scores of 1-10, 11-100, 101-400 and >400 were present in 10.0% (n=1,584), 11.5% (n=1,825), 5.2% (n=830) and 3.1% (n=497) respectively. After a median follow-up of 51.2 months, CVD risks increased from 5.2% in patients with no CAC to 28.2% in patients with CAC>400. After adjustment, CVD risk increased with higher CAC score: $HR_{CAC=1-10}=1.2$ (95%CI=1.0-1.5), $HR_{CAC=11-100}=1.8$ (1.5-2.1), $HR_{CAC=101-400}=2.2$ (1.8-2.7) and $HR_{CAC>400}=3.6$ (3.0-4.4). CAC was particularly strongly associated with CAD ($HR_{CAC>400}=7.8$, 95%CI=5.5-11.2). The association between CAC and CVD was strongest in patients treated with anthracyclines ($HR_{CAC>400}=5.8$, 95%CI=3.0-11.4).

CONCLUSIONS AND RELEVANCE CAC on breast cancer radiotherapy planning CT scans was strongly associated with CVD, especially CAD. Automated CAC scoring on radiotherapy planning CT scans can be used as a fast and low-cost tool to identify breast cancer patients at increased risk of CVD, allowing implementing CVD risk mitigating strategies with the aim to reduce the risk of CVD burden after breast cancer.

4.1 Background

The risk of cardiovascular disease (CVD) in patients with cancer is increasingly being acknowledged.⁹⁸ CVD is common in patients diagnosed with breast cancer, where survival is relatively favorable.^{1,2,98} In older women treated for breast cancer, CVD competes with breast cancer as the leading cause of death.³ Some treatments, such as anthracyclines, targeted therapy and radiotherapy are associated with an increased risk of CVD.^{99,100} Patients treated with anthracyclines have a five times higher risk of developing chronic heart failure or reduced left ventricular ejection fraction (LVEF) compared with those treated without.⁴ The risks of coronary artery disease and myocardial infarction are increased in patients who received radiotherapy.^{5,6} Breast cancer patients with pre-existing CVD risk factors (e.g. hypertension, obesity, diabetes) have the highest risk of treatment-induced cardiotoxicity.⁷

The presence and amount of coronary artery calcium (CAC), as quantified on CT scans, is an independent risk factor for CVD.^{101,102} The majority of breast cancer patients are treated with radiotherapy and undergo planning CT scans of the chest. A study including 939 breast cancer patients found an association between CAC on radiotherapy planning CT scans and acute coronary events.¹⁰³ CAC scoring is usually performed manually and therefore time-consuming. Our research group developed a deep-learning calcium scoring algorithm for automated CAC scoring on dedicated cardiac computed tomography (CT) scans.^{15,104} Automatic quantification of CAC offers the opportunity for routine, fast, and low-cost identification of breast cancer patients with an increased CVD risk without additional diagnostic radiation exposure.

We conducted the Bragatston study to evaluate whether automatically measured CAC on radiotherapy planning CT scans is associated with an increased risk of CVD and death in patients treated for breast cancer.⁹⁷

4.2 Methods

4.2.1 Study population

Radiotherapy planning CT scans and clinical data were collected for patients who were irradiated for primary breast cancer between January 2005 and December 2016 in the University Medical Center Utrecht, Utrecht, the Erasmus MC, Rotterdam, and the Radboudumc, Nijmegen, the Netherlands. Patients were excluded when they had metastasis at presentation or when planning CT scans were obtained ≥ 1 year after breast cancer diagnosis. For patients with multiple planning CT scans, the first scan was used. When two scans were made on the same day, the scan with breath-hold technique was used because of higher reproducibility of CAC scoring (i.e., less motion artifacts).³³

Through linkage with the Netherlands Cancer Registry (NCR), information was

obtained on laterality, TNM stage, grade, receptor status, type of surgery and chemotherapy, endocrine therapy, targeted therapy and occurrence of other cancers.¹⁰⁵

The Institutional Review Boards of the participating hospitals decided that the Medical Research Involving Human Subjects Act did not apply. Requirement for informed consent was waived.

4.2.2 Automatic calcification quantification

Planning CT scans without contrast enhancement and without ECG-synchronization were acquired as part of clinical routine (typical parameters: 120 kVp, in-plane resolution 0.51–1.56 mm, and 1.5–5.0 mm slice thickness). CT scans were reconstructed to 3.0 mm slice thickness and 1.5 mm increment. CAC was measured automatically using a deep-learning calcium scoring algorithm which detects and quantifies calcifications in the coronary arteries.^{32,104} Reproducibility of automatically versus manually measured calcium scores was high (linearly weighted kappa value 0.91 and intraclass correlation coefficient 0.85)¹⁰⁴

CAC was expressed as modified Agatston score, calculated by multiplying the calcification area (in mm²) by the weight (1, 130–199 Hounsfield Units (HU); 2, 200–299 HU; 3, 300–399 HU; 4, >399 HU) determined by the maximum density of the area and summing the lesion scores over slices. A minimal lesion definition of 1.5 mm³ was maintained to eliminate noise. Patients were categorized into five categories: 0, 1–10, 11–100, 101–400, >400 Agatston units.¹⁵

4.2.3 Outcome assessment

The primary outcome was incidence of (non-)fatal CVD events, i.e., hospitalization or death from CVD, between date of the CT scan and end of follow-up (December 31, 2017). Other outcomes were hospitalization or death from coronary artery disease (a subset of CVD), and all-cause mortality (including CVD deaths; follow-up until December 31, 2018).

For each patient, data on hospitalization for CVD and cancer (recurrence), and death was collected through linkage with national registries by Statistics Netherlands. Each hospital sent individual patient data to Statistics Netherlands and a pseudonym of the combination of national personal identification number, sex, date of birth, and postal code was used for linkage with national registries. Data on hospital admissions, including inpatient hospital care, day care and long-standing observation (i.e., ≥four hours), were obtained by linkage to the national basic registration hospital care (Landelijke Basisregistratie Ziekenhuiszorg, LBZ; and before 2013 Landelijke Medische Registratie, LMR), managed by Dutch Hospital Data (DHD). DHD used the International Classification of Disease 9th revision (ICD-9) before 2013 and the 10th revision (ICD-10) from 2013 onwards for disease classification.¹⁰⁶ Diagnoses registered according to ICD-9 were converted to ICD-10 using the conversion table of the Dutch WHO-FIC

Collaborating Centre.¹⁰⁷ CVD was defined by ICD-10 I00-199. Lymphedema (ICD-10: I88-89) and varicose veins (ICD-10: I83-86) were not considered CVD events. Coronary artery disease was defined by ICD-10 codes I20-25, e.g. myocardial infarction and angina pectoris. Cancer, including recurrent disease or metastases, was defined by ICD-10 codes C00-99, and non-melanoma skin cancer (ICD-10: C44) was not considered for analysis. Breast cancer hospitalizations in the LBZ/LMR within one year after the radiotherapy planning CT scan were considered as belonging to the initial breast cancer diagnosis and therefore not considered as recurrence or new breast cancer.

Mortality data were obtained from the National Death Register, which contains information on primary cause of death from all the deceased persons registered in the Netherlands. Causes of death were classified according to the ICD-10.¹⁰⁶

4.2.4 Statistical analysis

Follow-up time started at date of radiotherapy planning CT scan and ended at date of the event, censoring, or end of follow-up, whichever came first. Censored observations included non-cardiovascular death for the CVD analyses and death other than death from coronary artery disease for the coronary artery disease analyses. In the CVD and coronary artery disease analyses, patients were also censored at time of diagnosis of breast cancer recurrence or other cancers, or at end of follow-up (December 31, 2017). For the all-cause mortality analyses, patients who did not die during follow-up were censored at the end of follow-up (December 31, 2018).

CVD, coronary artery disease and all-cause mortality rates per 1000 person years were calculated for each CAC category (0, 1-10, 11-100, 101-400, >400). Cox proportional hazards models were used to examine the association between CAC scores and (non-)fatal CVD events, (non-)fatal coronary artery disease events and all-cause mortality. Results were expressed as hazard ratios (HRs) with corresponding 95% confidence intervals (CIs). Models were stepwise adjusted for age at CT scan, calendar year of CT scan (i.e., before 2012 versus 2012 and later), left- versus right-sided radiotherapy, treated with anthracyclines versus not (i.e., no chemotherapy, chemotherapy without anthracyclines or type of chemotherapy unknown), and treated with trastuzumab yes versus no. The proportional hazards assumption was assessed using visual inspection of Kaplan–Meier curves and Schoenfeld residuals for calendar year of CT scan, radiotherapy, chemotherapy with anthracyclines, and trastuzumab. Because the proportional hazards assumption was not met for calendar year of CT scan, cox proportional hazards models were stratified by calendar year. To test the linearity assumption, Martingale residuals against age at CT scan were plotted. Because of a nonlinear association between age and mortality, natural splines with 3 knots were included in the all-cause mortality model.

The association between CAC scores and CVD was examined for patients with left-versus right-sided breast cancer and with anthracyclines versus without. The number

Table 4.1: Baseline characteristics of the Dutch multicenter cohort of 15,919 patients receiving radiotherapy for breast cancer between 2005-2016

	Total	CAC score (Agatston units)	
		0	1-10
N (%)	15915	11179 (70.2)	1584 (10.0)
Sex, female, n (%)	15879 (99.8)		
Age, mean (SD)	59.0 (11.2)	56.1 (10.4)	62.4 (10.4)
Follow-up time in months, median (IQR)†	51.2 (27.9 - 81.1)	53.0 (29.4 - 81.7)	53.0 (28.1 - 87.5)
Pathological tumor stage, n (%)			
DCIS/T0	2152 (13.5)	1595 (14.3)	200 (12.6)
T1	9540 (59.9)	6561 (58.7)	1010 (63.8)
T2	3436 (21.6)	2456 (22.0)	302 (19.1)
T≥3	567 (3.6)	401 (3.6)	50 (3.2)
Unknown	220 (1.4)	166 (1.5)	22 (1.4)
Pathologic nodes stage, n (%)			
N0	10760 (67.6)	7471 (66.8)	1080 (68.2)
N1m	986 (6.2)	700 (6.3)	101 (6.4)
N1	2369 (14.9)	1728 (15.5)	231 (14.6)
N2	840 (5.3)	605 (5.4)	69 (4.4)
N3	460 (2.9)	320 (2.9)	49 (3.1)
Unknown	500 (3.1)	355 (3.2)	54 (3.4)
Laterality, n (%)			
Left-sided‡	8223 (51.7)	5780 (51.7)	831 (52.5)
Radiotherapy, n (%)			
Local	4179 (26.3)	2657 (23.8)	438 (27.7)
Local with boost	4503 (28.3)	3495 (31.3)	362 (22.9)
Loco regional (with and without boost)	1824 (11.5)	1323 (11.8)	155 (9.8)
Other	67 (0.4)	42 (0.4)	7 (0.4)
Unknown	5342 (33.6)	3662 (32.8)	622 (39.3)
Chemotherapy, n (%)			
None	9680 (60.8)	6116 (54.7)	1078 (68.1)
With anthracyclines	4333 (27.2)	3554 (31.8)	325 (20.5)
Without anthracyclines	1119 (7.0)	903 (8.1)	95 (6.0)
Type of chemotherapy unknown	783 (4.9)	606 (5.4)	86 (5.4)
Hormonal therapy, n (%)			
None	8778 (55.1)	6038 (54.0)	917 (57.9)
Aromatase inhibitor	592 (3.7)	361 (3.2)	68 (4.3)
Tamoxifen	4671 (29.3) §	3442 (30.8)	404 (25.5)
Other or type unknown	1878 (11.8)	1338 (12.0)	195 (12.3)
Trastuzumab, n (%)			
Yes	1257 (7.9)	1034 (9.2)	103 (6.5)

CAC = coronary artery calcium; IQR = interquartile range; SD = standard deviation. item[†] Time from radiotherapy planning CT scan until date of (non-)fatal cardiovascular disease or censoring (non-cardiovascular disease, diagnosis of breast cancer recurrence or other cancers, or end of follow-up (December 31, 2017).

‡ Information about laterality was missing for 3 patients.

§ In combination with tamoxifen, 268 patients were also treated with aromatase inhibitor and 118 patient with other types of hormonal therapy.

continued...

CAC score (Agatston units)		
11-100	101-400	>400
1825 (11.5)	830 (5.2)	497 (3.1)
65.8 (9.5)	69.5 (8.2)	71.0 (7.9)
48.7 (27.0 - 81.5)	42.2 (22.2 - 66.9)	34.9 (18.9 - 60.9)
216 (11.8)	92 (11.1)	49 (9.9)
1157 (63.4)	507 (61.0)	305 (61.4)
379 (20.8)	189 (22.8)	110 (22.1)
55 (3.0)	31 (3.7)	30 (6.0)
18 (1.0)	11 (1.3)	3 (0.6)
1267 (69.4)	594 (71.6)	348 (70.0)
109 (6.0)	48 (5.8)	28 (5.6)
247 (13.5)	104 (12.5)	59 (11.9)
96 (5.3)	42 (5.1)	28 (5.6)
53 (2.9)	22 (2.7)	16 (3.2)
53 (2.9)	20 (2.4)	18 (3.6)
938 (51.4)	426 (51.4)	248 (49.9)
556 (30.5)	321 (38.7)	207 (41.6)
388 (21.3)	175 (21.1)	83 (16.7)
194 (10.6)	91 (11.0)	61 (12.3)
13 (0.7)	3 (0.4)	2 (0.4)
674 (36.9)	240 (28.9)	144 (29.0)
1376 (75.4)	675 (81.3)	435 (87.5)
308 (16.9)	109 (13.1)	37 (7.4)
71 (3.9)	32 (3.9)	18 (3.6)
70 (3.8)	14 (1.7)	7 (1.4)
1050 (57.5)	470 (56.6)	300 (60.4)
92 (5.0)	50 (6.0)	18 (3.6)
461 (25.3)	237 (28.6)	127 (25.6)
222 (12.2)	73 (8.8)	52 (10.5)
75 (4.1)	30 (3.6)	15 (3.0)

of patients treated with trastuzumab was too small for subgroup analysis. Tests for interaction were performed by comparing models with an interaction term of CAC with covariates to the model without the interaction term using the likelihood ratio test.

Treatment of newly diagnosed cancers or recurrence may change the probability of developing CVD. Also, death other than from CVD precludes the occurrence of CVD. To assess the effect of these competing events, sensitivity analyses were conducted comprising competing risk survival analysis.¹⁰⁸ New cancers, recurrences and/or all-cause mortality (excluding death from CVD) were considered as competing risks. Results were presented as subdistribution hazard ratios (sHR) with 95% CIs and represent the instantaneous CVD risk in patients who had not yet experienced the event. Statistical analyses were conducted in SPSS version 25 and R Statistical Software version 3.6.2.

4.3 Results

Between 2005 and 2016, we identified 18,316 patients with 22,477 planning CT scans. After deduplication, linkage with the NCR and Statistics Netherlands, and excluding ineligible patients, 15,915 patients were available for analysis. Mean age at CT scan was 59.0 years (Table 4.1). A total of 11,179 (70.2%) patients had no CAC, 10% (n=1,584), 11.5% (n=1,825), 5.2% (n=830) and 3.1% (n=497) had CAC scores of 1-10, 11-100, 101-400 and >400 respectively. Patients with higher CAC scores were on average older compared to patient without CAC.

4.3.1 CAC and cardiovascular disease

After a median follow-up time of 51.2 months (IQR=27.9-81.7), 1,341 (8.2%) patients were hospitalized for CVD and 113 (0.7%) patients died of CVD (Table 4.2). The risk of (non-)fatal CVD events increased with increasing CAC score (Table 4.3). Five percent of the patients without CAC (n=579/11,179) had a CVD event (hospitalization or death from CVD) during follow-up, while 28.0% of the patients with a CAC score of >400 (n=140/497) had a CVD event. The incidence rate of CVD increased from 10.7 per 1000 person-years in patients without CAC to 77.8 per 1000 person-years in patients with a CAC score of >400.

After adjustment for age at planning CT scan and stratification for calendar year of CT scan, CAC was associated with CVD risk (Table 4.3). Compared to patients without CAC, patients with a CAC score of 1-10, 11-100, 101-400 and >400 had a 1.2, 1.8, 2.2 and 3.6 higher risk of (non-)fatal CVD, respectively (Table 4.4).

When treated with anthracyclines, patients with a CAC score of >400 had a 5.8 times higher risk (95%CI=3.0-11.4) than patients without CAC (Table 4.5). In patients not treated with anthracyclines, this risk was 3.5 times higher (95%CI=2.8-4.3). The interaction between CAC and anthracyclines was significant. We did not observe a

Table 4.2: Number of hospitalizations of cardiovascular disease and coronary heart disease, and deaths by CAC score category

	Total n = 15,915	CAC score (Agatston units)				
		0 n = 11,179	1-10 n = 1,584	11-100 n = 1,825	101-400 n = 830	>400 n = 497
Hospital admission for CVD, n (%)	1,341 (8.4)	643 (5.8)	146 (9.2)	260 (14.3)	154 (18.6)	138 (27.7)
Hospital admission for coronary artery disease, n (%)	336 (2.1)	127 (1.1)	37 (2.3)	70 (3.8)	51 (6.1)	51 (10.3)
All-cause mortality, n (%)	1100 (6.9)	581 (5.2)	146 (9.2)	162 (8.9)	108 (13.0)	103 (20.7)
Death from CVD, n (%)	113 (0.7)	28 (0.3)	16 (1.0)	23 (1.3)	18 (2.2)	28 (5.6)

difference in the association between patients with left- and right-sided breast cancer (HR>400=3.8, 95%CI=2.9-5.0 and HR>400=3.4, 95%CI=2.5-4.5 respectively).

Table 4.3: Number and incidence rates of cardiovascular disease, coronary heart disease, and deaths by CAC score category

CAC score (Agatston units)	n (%)	Cardiovascular disease		Coronary heart disease		All-cause mortality	
		Events, n (%)	Incidence rate†	Events, n (%)	Incidence rate†	Deaths, n (%)	Incidence rate†
0	11,179 (70.2)	579 (5.2)	10.7	121 (1.1)	2.2	760 (6.8)	10.9
1-10	1,584 (10.0)	138 (8.7)	17.4	37 (2.3)	4.5	183 (11.6)	17.7
11-100	1,825 (11.5)	253 (13.9)	29.5	75 (4.1)	8.3	222 (12.2)	19.3
101-400	830 (5.2)	148 (17.8)	43.8	50 (6.0)	13.8	146 (17.6)	30.8
>400	497 (3.1)	140 (28.2)	77.8	53 (10.7)	26.7	139 (28)	52.8

CAC = coronary artery calcium; CVD = cardiovascular disease.

Events included all cardiovascular disease hospitalizations and death (ICD-10 codes I00-199, Lymphedema and varicose veins excluded (I88-89 and I83-86 respectively)).

† Incidence rate per 1000 person-years

Table 4.4: Risk of cardiovascular disease, coronary heart disease and all-cause mortality by CAC score category

CAC score (Agatston units)	Unadjusted	Model 1	Model 2	Model 3
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Cardiovascular disease				
0	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
1-10	1.6 (1.4-2)	1.2 (1.0-1.5)†	1.2 (1.0-1.5)†	1.1 (0.9-1.4)
11-100	2.7 (2.4-3.2)	1.8 (1.5-2.1)	1.8 (1.5-2.1)	1.8 (1.5-2.1)
101-400	4.1 (3.4-4.9)	2.2 (1.8-2.7)	2.2 (1.8-2.7)	2.1 (1.7-2.6)
>400	7.1 (5.9-8.6)	3.6 (2.9-4.4)	3.6 (3-4.4)	3.4 (2.8-4.2)
Coronary heart disease				
0	1.0 (ref)	1.0 (ref)	1.0 (ref)	
1-10	2.1 (1.4-3.0)	1.7 (1.2-2.5)	1.7 (1.2-2.5)	
11-100	3.8 (2.8-5)	2.9 (2.1-3.9)	2.8 (2.1-3.9)	
101-400	6.3 (4.5-8.8)	4.3 (3.0-6.2)	4.3 (3.0-6.1)	
>400	12.2 (8.8-16.8)	7.9 (5.5-11.3)	7.8 (5.5-11.2)	
All-cause mortality				
0	1.0 (ref)	1.0 (ref)	1.0 (ref)	
1-10	1.6 (1.4-1.9)	1.3 (1.1-1.6)	1.4 (1.2-1.7)	
11-100	1.8 (1.5-2.0)	1.3 (1.1-1.5)	1.3 (1.1-1.6)	
101-400	2.9 (2.4-3.5)	1.7 (1.4-2.1)	1.8 (1.5-2.2)	
>400	5.1 (4.2-6.1)	2.7 (2.3-3.3)	2.8 (2.3-3.4)	

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio; ref = reference.

Cardiovascular disease included all cardiovascular disease hospitalizations and death (ICD-10 codes I00-199, Lymphedema and varicose veins excluded (I88-89 and I83-86 respectively))

Coronary heart disease included all coronary heart disease hospitalizations and death (ICD-10 codes I20-25, including angina pectoris and myocardial infarction)

Model 1: adjusted for age at time of CT scan and stratified by calendar year of CT scan

Model 2: adjusted for age at time of CT scan, laterality, treated with anthracyclines and treated with trastuzumab, and stratified by calendar year of CT scan

Model 3: adjusted for age at time of CT scan, laterality, treated with anthracyclines and treated with trastuzumab, and stratified by calendar year of CT scan; model included interaction between CAC and anthracyclines

Mortality model 1 and 2: age was transformed (natural spline with df=3)

† 95% CI lower limit >1.0

Table 4.5: Risk of cardiovascular disease by CAC score category, stratified by treatment

CAC score†	Left-sided radiotherapy		Right-sided radiotherapy		Anthracyclines		No anthracyclines	
	Incidence rate‡	HR (95% CI)	Incidence rate‡	HR (95% CI)	Incidence rate‡	HR (95% CI)	Incidence rate‡	HR (95% CI)
0	11.5	1.0 (ref)	9.9	1.0 (ref)	9.3	1.0 (ref)	11.2	1.0 (ref)
1-10	16.8	1.1 (0.9-1.4)	18.0	1.4 (1-1.8)	20.9	1.7 (1.1-2.7)	16.7	1.1 (0.9-1.4)
11-100	30.9	1.8 (1.4-2.2)	27.9	1.8 (1.4-2.2)	21.0	1.5 (1.0-2.3)§	30.8	1.8 (1.5-2.1)
101-400	41.4	2.0 (1.5-2.6)	46.7	2.5 (1.9-3.3)	53.0	3.1 (1.9-5.1)	42.7	2.1 (1.7-2.6)
>400	89.4	3.8 (2.9-5.0)	66.9	3.4 (2.5-4.5)	110.1	5.8 (3.0-11.4)	76.1	3.5 (2.8-4.3)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio; ref = reference
 Cardiovascular disease included all cardiovascular disease hospitalizations and death (ICD-10 codes I00-199, Lymphedema and varicose veins excluded (I88-89 and I83-86 respectively))

Models were adjusted for age at time of CT scan and stratified by calendar year of CT scan.

† CAC Agatston scores

‡ Incidence rate per 1000 person-years

§ 95% CI lower limit >1.0

4.3.2 CAC and coronary artery disease

Patients with a higher CAC score had a higher risk of coronary artery disease. One percent of the patients without CAC had a coronary artery disease event and 10.3% of the patients with a CAC score of >400 had a coronary artery disease event (Table 4.2). The incidence rate increased from 2.2 for patients without CAC to 26.7 per 1000 person-years for patients with a CAC score of >400 (Table 4.3), corresponding to an age adjusted 8-fold increased risk of coronary artery disease (HR=7.8, 95%CI=5.5-11.2) (Table 4.4). The association between CAC and coronary artery disease was not different for patients with CT scans before 2012 versus 2012 and later, when treated with left- versus right-sided radiotherapy when treated with anthracyclines yes versus no (i.e., interactions were not significant; Table 4.7).

4.3.3 CAC and all-cause mortality

The incidence rate of all-cause mortality increased from 10.9 deaths per 1000 person-years in patients without CAC to 52.8 deaths per 1000 person-years in patients with CAC >400 (Table 4.3). Also in relative terms, patients with a CAC score of >400 had a higher risk of dying than patients without CAC (HR=2.8, 95%CI=2.3-3.4; Table 4.4). The association between CAC and all-cause mortality was not different for patients with CT scans before 2012 versus 2012 and later, when treated with left- versus right-sided radiotherapy when treated with anthracyclines yes versus no (i.e., interactions were not significant). The association between CAC and all-cause mortality was stronger in patients treated with anthracyclines compared to patients treated without (HR>400=4.6, 95%CI=2.4-8.6 and HR>400=2.8, 95%CI=2.2-3.4, respectively; Table 4.8), but test for interaction was not significant.

4.3.4 Competing risks

When taking competing risks into account, CVD risk remained strongly associated with CAC (Table 4.6; Figure 4.1). Results were consistent with the results from the Cox proportional hazards analyses described above.

Table 4.6: Competing risks survival analysis of cardiovascular disease by CAC category

CAC score (Agatston units)	Subdistribution HR (95% CI) [†]		
	Model 1	Model 2	Model 3
0	1.0 (ref)	1.0 (ref)	1.0 (ref)
1-10	1.2 (1.0-1.5) [‡]	1.2 (1.0-1.5) [§]	1.2 (1.0-1.5) [§]
11-100	1.8 (1.5-2.1)	1.8 (1.5-2.1)	1.8 (1.5-2.1)
101-400	2.2 (1.8-2.7)	2.2 (1.8-2.7)	2.2 (1.8-2.6)
>400	3.5 (2.9-4.3)	3.5 (2.8-4.3)	3.4 (2.8-4.2)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio; ref = reference

All models were adjusted for age at time of CT scan and stratified by calendar year of CT scan

Model 1: competing risk was all-cause mortality

Model 2: competing risk was cancer hospitalization

Model 3: competing risks were all-cause mortality and cancer hospitalization

Cardiovascular disease included all cardiovascular disease hospitalizations and death (ICD-10 codes I00-199, Lymphedema and varicose veins excluded (I88-89 and I83-86 respectively))

[†] Competing risk survival analysis as described by Fine and Gray (Fine 1999 J Am Stat Assoc) provides subdistribution hazard ratios.

[‡] 95% CI lower limit <1.0

[§] 95% CI lower limit >1.0

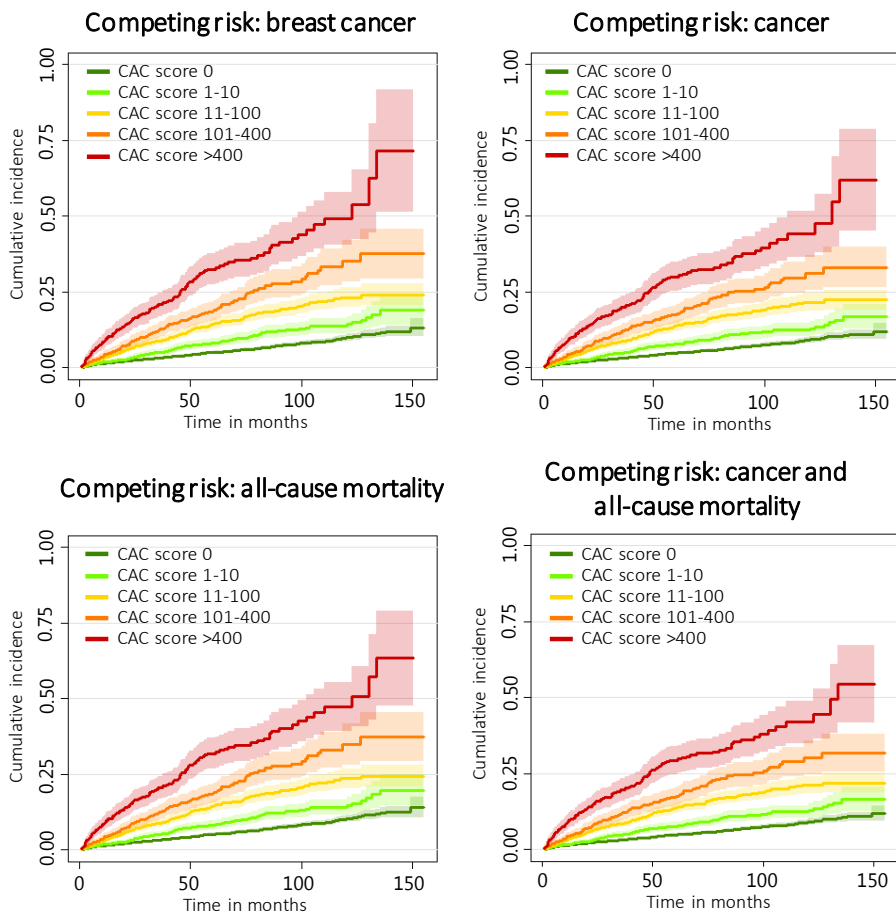


Figure 4.1: Cumulative incidence curves for CVD by CAC category in the presence of competing risks (Aalen-Johansen estimates and 95% intervals). Cardiovascular disease included all cardiovascular disease hospitalizations and death (ICD-10 codes I00-199, Lymphedema and varicose veins excluded (I88-89 and I83-86 respectively)). CAC = coronary artery calcium

Table 4.7: Risk of coronary artery disease by CAC score category, stratified by treatment

CAC score	Left-sided radiotherapy		Right-sided radiotherapy		Anthracyclines		No anthracyclines	
	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)
0	2.3	1.0 (ref)	2.0	1.0 (ref)	1.3	1.0 (ref)	2.5	1.0 (ref)
1-10	5.2	1.9 (1.2-3.1)	3.7	1.4 (0.8-2.6)	3.7	2.4 (0.9-6.6)	4.7	1.6 (1.1-2.4)
11-100	9.7	3.4 (2.3-5.0)	6.7	2.3 (1.4-3.6)	6.4	3.6 (1.5-8.6)	8.5	2.8 (2.0-3.8)
101-400	15.9	5.1 (3.2-8.2)	11.6	3.4 (2.0-5.9)	2.4	1.1 (0.2-8.9)	15.3	4.6 (3.2-6.6)
>400	27.3	8.4 (5.1-13.9)	26.0	7.4 (4.4-12.4)	20.4	9.1 (2-41.8)	27.0	7.8 (5.4-11.3)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio

Coronary heart disease included all coronary heart disease hospitalizations and death (ICD-10 codes I20-25), including angina pectoris and myocardial infarction
Models were adjusted for age at time of CT scan and stratified by calendar year of CT scan

† Incidence rate per 1000 person-years

Table 4.8: Risk of all-cause mortality by CAC score category, stratified by treatment (follow-up until 2018)

CAC score	Left-sided radiotherapy		Right-sided radiotherapy		Anthracyclines		No anthracyclines	
	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)	Incidence rate†	HR (95% CI)
0	11.7	1.0 (ref)	10.0	1.0 (ref)	14.2	1.0 (ref)	9.7	1.0 (ref)
1-10	17.1	1.2 (1.0-1.5)‡	18.3	1.5 (1.2-1.9)	18.0	1.2 (0.8-1.8)	17.6	1.4 (1.1-1.7)
11-100	19.8	1.2 (1.0-1.5)‡	18.6	1.3 (1.0-1.6)§	16.8	1.1 (0.7-1.6)	19.7	1.3 (1.1-1.5)
101-400	31.2	1.6 (1.2-2.0)	30.4	1.9 (1.4-2.5)	23.8	1.4 (0.8-2.5)	31.7	1.8 (1.5-2.2)
>400	59.5	2.9 (2.2-3.7)	46.1	2.6 (2.0-3.5)	76.7	4.6 (2.4-8.6)	51.4	2.8 (2.2-3.4)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio

Models were adjusted for age at time of CT scan (natural spline with df=3) and stratified by calendar year of CT scan

Incidence rate per 1000 person-years

95% CI lower limit <1.0

95% CI lower limit >1.0

4.4 Discussion

CAC measured on radiotherapy planning CT scans of patients with breast cancer is strongly associated with (non-)fatal CVD, especially in patients treated with anthracyclines.¹⁰⁹ The association between CAC and coronary artery disease was even stronger. Because patients with breast cancer planned for radiotherapy routinely undergo a planning CT scan, automated CAC scoring can be used as a fast and low-cost tool to identify patients at increased risk of CVD. Patients with an increased risk may be referred for complete CVD work up, lifestyle advice and/or initiation of cardioprotective pharmacotherapy.

A meta-analysis including five population-based cohorts showed that 63.9% of 6,739 low-risk women had no CAC, which is somewhat lower compared to our cohort.¹⁰² After a median follow-up ranging from 7.0 to 11.6 years, the risk of coronary artery disease was increased in women with a CAC score of >0-100 (HR=1.8) and a CAC score above 100 (HR=4.4). We found slightly stronger associations in our breast cancer population.

Our results are in line with a Dutch study including 939 breast cancer patients (mean age of 58.4 years).¹⁰³ The proportion of patients with CAC (manually quantified on radiotherapy planning CT) was lower than in our cohort (21.1% and 28.8%, respectively). Breast cancer patients with a CAC score ≥ 100 had a five times higher risk (HR=5.0) of acute coronary events compared to patients without CAC. This is in line with our study, where a strong association between CAC and coronary artery disease was found: HR₁₀₁₋₄₀₀=4.3 and HR_{>400}=8.2 compared to patients without CAC.

The association of CAC with CVD and coronary artery disease was similar in patients treated with left- and right-sided radiotherapy. Women treated with left-sided radiotherapy have an increased risk of CVD compared to women treated with right-sided radiotherapy, and this risk increases with time since diagnosis.^{5,19} Our median follow-up of 51.6 months (i.e., 4.3 years) may be too short to draw conclusions on a potential interaction between CAC and left- versus right-sided radiotherapy. Because of increased awareness of the potential damage of radiation dose to the heart and the introduction of the deep inspiration breath-hold technique in left-sided radiotherapy around 2010, the mean heart dose has been reduced and the effect of radiotherapy on the risk of CVD may have decreased over years.¹¹⁰⁻¹¹²

Completeness of the registry on hospital admissions varied over the years. Not all hospital admissions were registered, and missing hospitalizations increased from 10.5% in 2006 to 25% in 2013. From 2014 onward, around 90% of all hospital admissions were covered. As such, the absolute risk of hospitalization was underestimated. However, since the probability of incomplete data on hospitalization was most likely non-differential (i.e., similar for patients with low and high CAC scores), the association between CAC and CVD events is unlikely to be distorted. Second, until 2011, linkage was based on sex, date of birth and postal code and the combination of variables was

not unique. Therefore, 3% of the hospital admissions could not be uniquely linked to one person and these patients had to be excluded. Third, until 2015, CVD events that required hospitalization for \geq four hours without overnight stay were not registered as hospitalization, which may have resulted in an underestimation of CVD incidence.

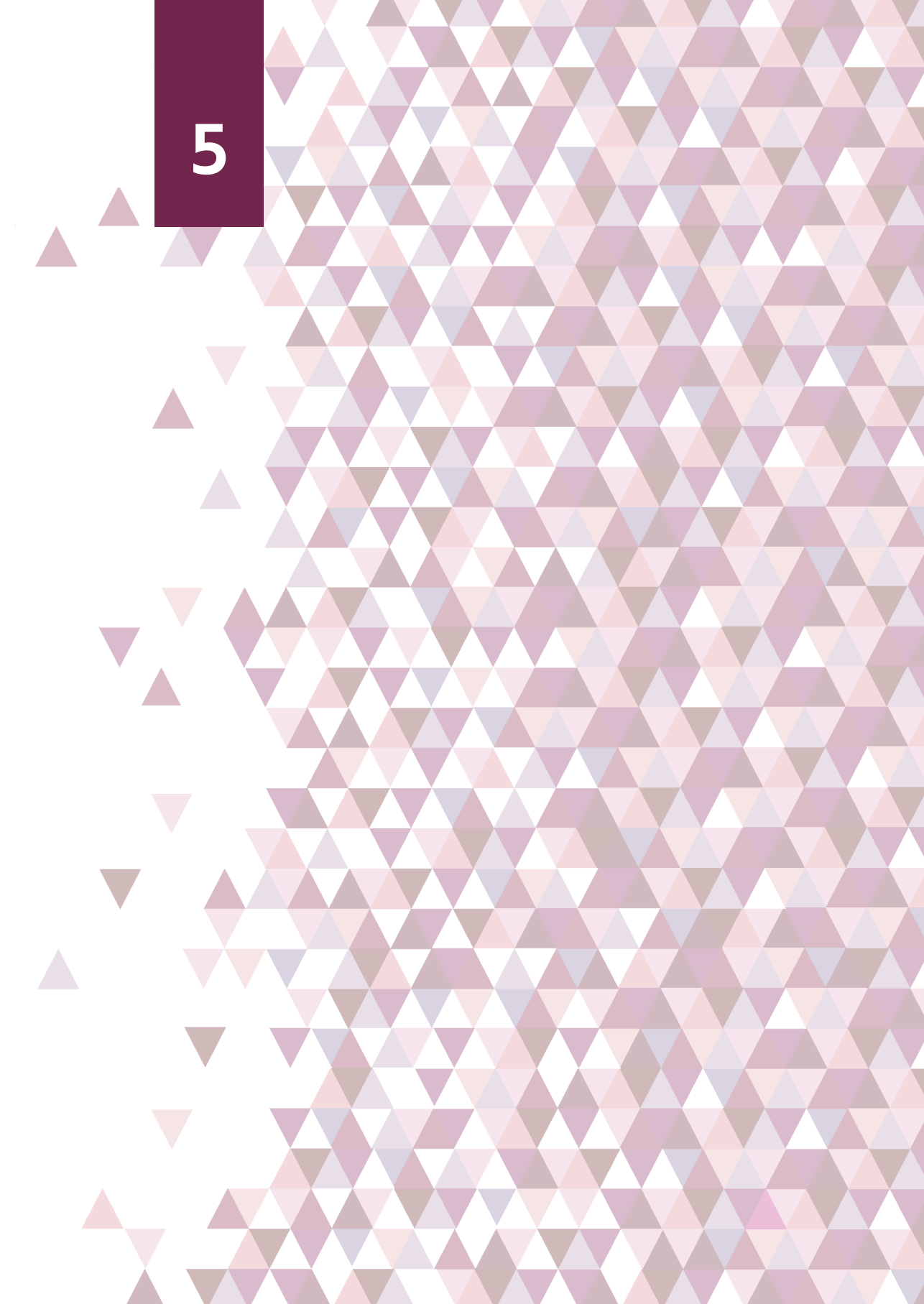
Information on traditional CVD risk factors such as smoking, body mass index, low physical activity, and family history of CVD was lacking. In a previous study, we demonstrated that 28% of the breast cancer patients with a CAC score above 100, had no other traditional CVD risk factors.¹² Based on traditional risk factors, these patients would not have been identified as high risk patients. This implies that CAC scoring may be of added value for the screening of patients at increased risk of CVD.

Given the strong association between automatically measured CAC on radiotherapy planning CT scans and the risk of CVD in patients with breast cancer, future research should focus on how to use CAC quantification in clinical practice, with the ultimate aim to lower CVD burden in (ex) breast cancer patients. Until now, there is no evidence indicating that measuring CAC, and taking action in case of increased risk, improves patient outcomes. We do not know whether cardiovascular evaluation, followed by lifestyle or pharmaceutical interventions, reduces the risk of (death from) CVD. Neither do we know whether the benefits of de-escalation of treatment in order to reduce the risk of CVD (e.g. changing radiotherapy technique or comprising target volumes, or chemotherapy dose reduction) outweigh the potential benefits in terms of tumor control and prevention of breast cancer recurrence.

To conclude, with our in-house developed deep learning algorithm, the presence and extent of CAC can be automatically and accurately quantified on routinely performed radiotherapy planning CT scans. The CAC score is a strong predictor for CVD that can be fast and 'freely' obtained from planning CT scans, which may enable identification of patients at increased or even high risk of CVD. Timely and accurate identification of high-risk patients might allow mitigation of cardiovascular risk, including management of CVD risk factors and switching to less cardiotoxic treatment regimens.

Acknowledgments

This study was funded by the Dutch Cancer Society (grant number UU 2015-7947). The authors thank the members of the Bragatston study group for their collaboration.



CHAPTER 5

Generative models for reproducible coronary calcium scoring

BASED ON: S. G. M. van Velzen, B. D. de Vos, J. M. Noothout, H. M. Verkooijen, M. A. Viergever, and I. Išgum. "Generative models for reproducible coronary calcium scoring". Submitted.

Abstract

Coronary heart disease (CHD) is among the leading causes of death worldwide. A strong and independent predictor of (non-)fatal cardiac events is the amount of coronary artery calcification (CAC) quantified in CT. However, quantification of CAC (calcium scoring) suffers from limited interscan reproducibility, which is mainly due to the use of a fixed intensity level threshold for segmentation. This is especially pronounced in CT exams without ECG-synchronization where lesions are more subject to cardiac motion and partial volume effects. To address the limited interscan reproducibility of CAC quantification, we propose a CAC quantification method that does not require a segmentation threshold. Our method uses a CycleGAN to decompose an image with CAC into an image without CAC and an image showing only CAC. To simplify the task, we analyze only the heart in slices with CAC. The method was trained using 626 low-dose chest CTs and 514 radiotherapy treatment planning CTs. Evaluation was performed in radiotherapy treatment planning CTs of 1,662 patients, each having two scans. A lower relative interscan difference in CAC mass was achieved by the proposed method: 47% compared to 89% for manual clinical calcium scoring and 114% for automatic clinical calcium scoring. The increased interscan reproducibility achieved by our method may lead to increased reliability of CHD risk categorization and, thus, improved accuracy in the prediction of CHD events.

5.1 Introduction

Coronary heart disease (CHD) is among the leading causes of death worldwide.¹¹³ A clear manifestation of CHD is calcification of the coronary arteries. The amount of coronary artery calcifications (CAC), i.e. coronary calcium score, is a strong and independent predictor of cardiac events, e.g. myocardial infarction or heart failure.^{13,14} In CT, CAC is defined as a high density area of ≥ 130 HU¹⁵ in the artery. Using this definition CAC is commonly quantified into CAC volume, mass or Agatston score.¹⁵ The Agatston score is used most because it directly indicates a patient's risk of CHD events.

In a clinical setting the amount of CAC is routinely quantified in non-contrast-enhanced cardiac CT exams with ECG-synchronization. Given its clinical relevance, the guidelines of the Society of Cardiovascular Computed Tomography and the Society of Thoracic Radiology¹⁸ recommend to quantify and report CAC on *all* non-contrast CTs showing the heart, including non-ECG-synchronized scans. However, the interscan reproducibility of calcium scores is limited. For example, the reported mean interscan variability of Agatston scores in cardiac CT with ECG-synchronization ranges from 15% to 41%.^{114–117} Hence, in non-ECG-synchronized scans the limited interscan reproducibility is further emphasized due to increased influence of partial volume effect and cardiac motion. For instance, Jacobs et al.¹¹⁸ reported an average interscan variability of 71% for Agatston scores in non-triggered chest CT scans, that led to assignment of different cardiovascular risk category in 24% of subjects.

The use of a single intensity level threshold has been identified as one of the major causes of the limited reproducibility of CAC quantification,¹¹⁹ which may lead to under- or overestimation of the amount of CAC. This is especially pronounced in CT exams without ECG-synchronization where visualization of the calcifications is often extremely affected by cardiac motion. As a consequence, the calcifications are blurred or completely remain below the threshold. Nevertheless, common to all manual and automatic calcium scoring methods is the use of the clinical definition that identifies voxels in the coronary arteries above the 130 HU threshold as CAC in an explicit manner (through segmentations)^{32,104} or implicit manner (through regression of the CAC amount).^{54,120} To address this issue, automatic methods have been proposed that either use automatic adaptive thresholds or omitted thresholding for segmentation of calcified lesions. Groen et al.¹²¹ proposed using an adaptive threshold dependent on the maximum intensity value of each CAC lesion. Song et al.¹²² proposed to adapt the threshold based on the intensity of the background in the vicinity of lesions in combination with deconvolution of the image, using a scanner-specific point spread function to decrease partial volume effect. Saur et al.¹²³ omitted the intensity threshold for segmentation by employing a mesh-based algorithm that segmented CAC lesions by refining their boundary based on the intensity value profile. A different method that circumvents thresholding for segmentation, was proposed by Šprem et al.,¹²⁴ who built

on work by Dehmeshki et al.¹²⁵ Both used an expectation-maximization algorithm to determine the partial calcium content in each voxel of a CAC lesion and its vicinity.

While the aforementioned methods all show improved interscan reproducibility of CAC quantification, they require the clinically used threshold for initial lesion detection, thereby missing lesions below the intensity level threshold. This is particularly a problem with small lesions and scans with motion artefacts, where complete lesions may remain under the detection threshold. Although motion detection¹²⁶ in CAC lesions has been proposed, this does not include correction that would allow quantification of such CAC lesions.

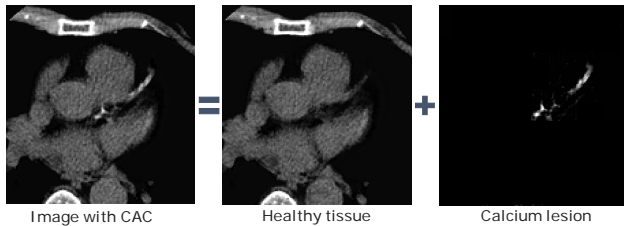


Figure 5.1: An image containing coronary artery calcium (CAC) is decomposed into a healthy tissue image without CAC and an image containing only CAC.

To address the limited interscan reproducibility of CAC detection and quantification, we propose a method that does not require an intensity level threshold. Our method takes an image with CAC and decomposes it into an image without CAC and a CAC-map, i.e. an image showing only CAC (Figure 5.1).

Several earlier studies proposed to use an autoencoder for generating healthy counterparts of images with pathology in an unsupervised manner.^{127–129} Subsequently, the pixel-wise intensity difference between the original image and the reconstruction was used to e.g. detect brain lesions in MRI scans,¹²⁷ brain CT,¹²⁹ or detected anomalies in head CT.¹²⁸ Alternative approaches used generative adversarial networks (GAN) for translating an image containing pathologies to a healthy image. Baur et al.¹³⁰ and Seah et al.¹³¹ used conditional GANs to generate healthy images and used the difference with the input image for multiple sclerosis lesion segmentation, and visualizing features of congenital heart failure in chest X-rays, respectively. Furthermore, Baumgartner et al.¹³² proposed to use a Wasserstein GAN to directly generate a difference map that indicates changes in the brain related to Alzheimer’s disease. The map is added to the input image to create a healthy brain image from an image containing Alzheimer’s disease. To tackle the instability and mode collapse problems that often arise during training of GANs, Zhu et al.¹³³ proposed to use a cycle-consistency loss in their CycleGAN for image-to-image translation using natural images. Sun et al.¹³⁴ proposed to use a CycleGAN implementation for translating images containing lesions to healthy images. Subsequently, the pixel-wise intensity difference was used to detect brain lesions in MRI and liver tumors in CT scans.

To decompose an image with CAC into an image without CAC and CAC-map, we employ a CycleGAN to translate images between the *containing* CAC domain and *not containing* CAC domain. To our knowledge, our method is the first to use a generative model for accurate segmentation of CAC lesions. The method builds on our preliminary work,¹³⁵ in which we presented a semi-automatic method for segmentation of CAC, that analyzed the vicinity of manually identified CAC lesions. The method was trained using clinical threshold-based calcium scores as reference and hence, was not able to detect lesions below the threshold. In the current work, we extend the method to perform fully automatic CAC detection and quantification, including CAC lesions below the standard detection threshold. The method is aimed at scans where the low interscan reproducibility is most dire and therefore, experiments were performed in non-ECG-synchronized scans, with and without breathing motion, that often contain excessive cardiac motion artefacts and have a low spatial resolution. To simplify the challenging task and allow the method to focus on the region of interest only, CAC quantification was performed in the heart slices containing CAC. The method is developed and evaluated using 4,038 CT scans of 2,276 breast cancer patients and 626 CT scans of lung screening participants. Moreover, we show that it outperforms clinically used calcium scoring methods in terms of interscan reproducibility of CAC quantification.

5.2 Data

We included 4,038 radiotherapy treatment planning (RTP) CTs of 2,276 breast cancer patients.⁹⁷ In this set 1,762 patients had two scans made on the same day: one with and one without breathhold. Therefore, this set allows evaluation of CAC scoring interscan reproducibility. The remaining 514 patients had one CT scan made without breathhold. Scans were acquired in the University Medical Center Utrecht with a Philips Brilliance Big Bore scanner. Intravenous (IV) contrast was not induced and the acquisition was not ECG-synchronized. The CTs were acquired with 120 kVp and reconstructed to 0.92-1.37 mm in-plane resolution and 3.0 mm slice thickness and increment. The need for informed consent was waived by the local institutional review board.

Additionally, we included 626 low-dose chest CT images of 626 subjects included in the National Lung Screening Trial (NLST). Scans were acquired in 29 different hospitals, on scanners from all major vendors.^{32,82} Tube voltage was either set to 120 or 140 kVp, depending on the subjects weight. IV contrast was not induced and acquisition was not ECG-synchronized. Images were reconstructed to 0.49-0.98 mm in-plane resolution and 1.0-2.5 mm slice thickness and 0.6-2.5 mm increment.

For development and training of the method, RTP CT scans of 514 patients with a single CT and all 626 NLST scans were used. The latter were used to enhance training with a diverse set of scans of heavy smokers with high CAC burden. In method development, a validation set of 100 patients with two available RTP scans (200 scans)

was used. We divided the test scans into two sets with patient scan pairs: Test-set 1, a set consisting of scans of 119 patients totalling 238 RTP CTs with manually segmented CAC allowing detailed evaluation of calcium scoring; and Test-set 2, a set consisting of scans of 1543 patients of in total 3086 RTP CTs without reference CAC scores, used for assessment of interscan reproducibility.

To allow training, an experienced observer labelled axial slices of all training CTs, validation CTs and CTs from Test-set 1 as *containing CAC* or *not containing CAC*, based on visual assessment. Moreover, for a subset of 199 RTP scans of the training set, manual segmentations of the heart were available from clinical practice.

To define reference standard for CAC, lesions were annotated and assigned coronary artery label using manual voxel-painting in Test-set 1. This was only done in the 34 scan pairs in which both scans had CAC, because pairs without CAC and discordant pairs are not informative for measuring interscan reproducibility. Next, to allow comparison with calcium scoring used in clinic, manual annotations from a previous study for Test-set 1 were available.¹⁰⁴

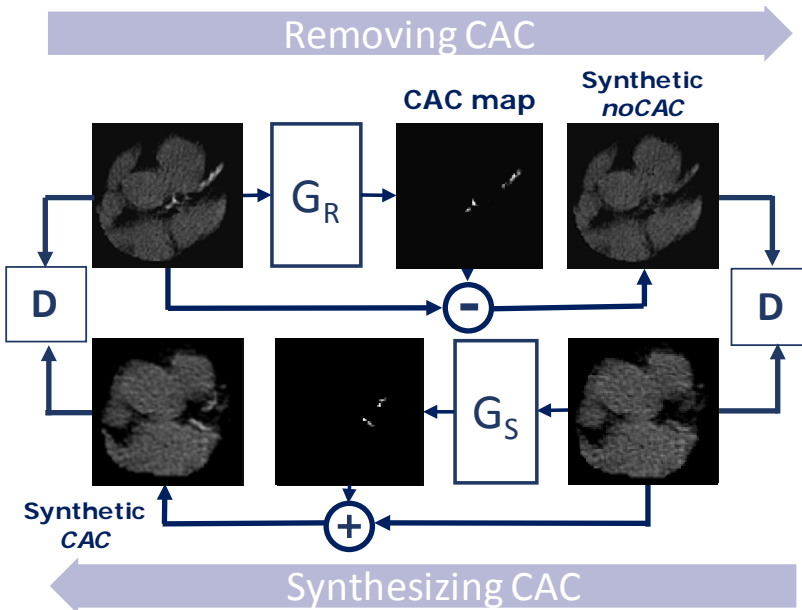


Figure 5.2: Schematic overview of the proposed CycleGAN: In the pathway that removes CAC, a CAC map is predicted by the generator (G_R) and subtracted from the image to obtain a synthetic image without CAC. In the pathway that synthesizes CAC the CAC map predicted by the generator (G_S) is added to the image. Synthetic images are compared to real examples by discriminators (D).

5.3 Method

We propose an automatic method that detects and segments CAC by decomposing a CT slice into an image without CAC and an image showing a CAC map, i.e. only CAC. The method allows segmentation of whole lesions, including parts below the 130 HU threshold, leading to increased interscan reproducibility of CAC quantification.

The need for improved interscan reproducibility of CAC quantification is highest in non-ECG-synchronized scans. However, non-ECG-synchronized scans are typically made for other purposes than calcium scoring, thus, the field of view in these scans is highly variable. Therefore, before generating the CAC-map, we define the region of interest to simplify the CAC detection task by initial heart segmentation, and subsequent identification of CT slices with CAC.

Table 5.1: Architecture of the Heart Segmentation CNN, CAC Classification CNN and the Generators in the proposed CycleGAN

General	Heart Segmentation CNN	CAC Classification CNN	CAC-map generators CycleGAN
Dimensions	3D	2D	2D
Input size	64x64x64 voxels	224x224 voxels	224x224 voxels
Type of padding (layers)	Zero (All)	Reflection (All)	Reflection (All)
Architecture (kernel, output channels, stride)			
1st Layer	Conv (7x7x7, 16)	Conv (7x7, 64)	Conv (7x7, 64)
2nd Layer	Conv (3x3x3, 32, 2)	Conv (3x3, 128, 2)	Conv (3x3, 128, 2)
3rd Layer	Conv (3x3x3, 64, 2)	Conv (3x3, 256, 2)	Conv (3x3, 256, 2)
ResNet Blocks	9 Blocks (3x3x3, 64, 1)	6 Blocks (3x3, 256, 1)	6 Blocks (3x3, 256, 1)
3rd last Layer	DeConv (3x3x3, 32, 2)	-	DeConv (3x3, 128, 2)
2nd last Layer	DeConv (3x3x3, 16, 2)	Global Average Pool	DeConv (3x3, 64, 2)
Last Layer	Conv (7x7x7, 2)	Dense Layer (256x2 channels)	Conv (7x7x7, 1)
Activation last Layer	Sigmoid	Softmax	Sigmoid

Conv = Convolution layer; DeConv = Deconvolution layer or transposed convolution layer; All convolution layers, except the last layer, are followed by a batch normalization layer and rectifier linear unit activation. Resnet Blocks consist of 2 convolution layer pairs with a residual connection.

Heart segmentation is performed with a 3D convolutional neural network (CNN), which, for efficiency, analyses the full image in a patch based manner. The network architecture is based on ResNet¹³⁶ and is detailed in Table 5.1, left. The Dice loss is used during training and Adam optimization algorithm (lr = 0.001) is used for optimization. During inference, overlapping patches were analysed and the results were averaged for final prediction.

We classify each slice in the heart mask according to the presence of visible CAC with a 2D CNN with ResNet¹³⁶ based architecture (Table 5.1, middle). The CNN analyses axial image slices centered and cropped around the heart. For cropping, the center of mass of the heart segmentation is used. The network is trained with per slice labels indicating the presence of CAC, using the crossentropy loss and Adam optimization algorithm (lr = 0.001).

In the identified slices, we generate a CAC-map. Generating the CAC-map in a supervised manner, where a CNN directly predicts the map, is hardly feasible, since identifying voxels with visible CAC is difficult and extremely time consuming. Hence, we propose to obtain a CAC-map using weakly supervised labels, i.e. labels indicating whether CAC is present in the image. In our approach a CAC-map is obtained by subtracting a generated healthy tissue image from an image containing CAC. Because it is not possible to obtain paired CT scans of the same patient, where one would contain CAC and the other not, a paired approach for image generation is not feasible. Therefore, an unpaired approach is used to generate synthetic *noCAC* images from images containing CAC. For this, images are assigned to two domains based on weak labels: one domain consisting of images containing visible CAC and the other domain containing images without visible CAC. Hence, to generate a healthy tissue image for subtraction, a 2D CycleGAN¹³³ for unpaired image-to-image translation is used to translate images from one domain to another. The CycleGAN is a type of GAN²² that improves in training stability and performance over the traditional GAN, by adding cycle-consistency.¹³³

The CycleGAN consists of two generators, and two discriminators that compete with each other. The generators are CNNs that convert an input image from either domain to a realistic image from the other domain. The discriminator CNNs compare the synthesized images against 'real' examples from the other domain. In contrast to a conventional CycleGAN, that directly translates an input image to an image of the target domain, we use an adjusted CycleGAN (Figure 5.2) that adds an extra step: images with CAC are translated to the *noCAC* domain by predicting a CAC map and subtracting it from the input image. In the synthesizing pathway the CAC map is added to the input image to generate a synthetic image containing CAC from an image without CAC. The generator in the pathway that removes CAC is used to obtain the CAC map and, subsequently, the CAC map can be used to quantify the amount of CAC in the image.

The generators used in the CycleGAN for generating the CAC map have a ResNet-

based architecture,¹³⁶ with 6 ResNet blocks using ReLU activation and no downsampling in between blocks (Table 5.1, right). To ensure the CAC maps only contain positive values, the generators have a sigmoid activation in the last layer. For the discriminator networks we use PatchGANs, which classify 70x70 overlapping image patches as real or synthetic, described in detail by.¹³⁷ The PatchGANs are able to capture high-frequency structure and have fewer parameters than a full-image discriminator and are therefore, easier to train.¹³⁷ The patch-level discriminators are applied to the image in a fully convolutional manner, averaging all responses to provide the output of the discriminator. The system is trained using the loss terms traditionally used for training a CycleGAN (discriminator loss, cycle-consistency loss and identity loss¹³³). Additionally, a sparsity (L1) loss on the CAC map is used to encourage the generator to predict sparse CAC maps.

To make the method more robust to noise, during training images are augmented by reducing or amplifying noise. To reduce the noise, images are smoothed with a Gaussian filter with $\sigma = 0.5$ pixels. To amplify noise, a noise image is obtained by subtracting the smoothed image from the original input image that is, subsequently, added to the input image. Moreover, during training random cropping and rotation are used for augmentation.

Despite the sparsity loss, the generated CAC-map may contain low levels of noise, because of excessive image noise. Noise in the CAC map may lead to unrealistically low values for soft tissue in the synthetic noCAC image. Therefore, we mask voxels in the CAC-map that lead to a value of < -10 HU in the synthetic noCAC image, as those are likely noise voxels in the CAC-map.

Finally, CAC was quantified by CAC pseudo-mass and adjusted Agatston score. The pseudo-mass was calculated by multiplying the sum of HU values of the voxels in the lesions by the voxel volume. The adjusted Agatston score was calculated by multiplying the area of a lesion with a density score (1, <199 HU; 2, 200–299 HU; 3, 300–399 HU; 4, >399 HU) determined by the maximum density of the area, and summing over axial slices. Please note that the adjusted Agatston score is equal to the conventional Agatston score for methods using the 130 HU intensity threshold.

5.4 Evaluation

Detection of CAC lesions was evaluated using accuracy, sensitivity, rate of false positive scans and F1-score, between the automatic and reference labels that indicated visible CAC.

The reproducibility of the CAC score was evaluated in scan pairs, using absolute relative differences of quantified CAC pseudo masses. To measure the interscan agreement of adjusted Agatston scores, the two-way random intraclass correlation coefficient (ICC) for absolute agreement was computed and Bland Altman plots with 95% limits of agreement were examined.⁴⁸ Moreover, patients were assigned one of four CHD risk

categories (I-IV: 0–10, 11–100, 101–400, <400) based on their adjusted Agatston score. Agreement between risk categories of each patient determined from the patient’s two scans was evaluated using Cohen’s linearly weighted kappa (κ).

5.5 Experiments and Results

We evaluate the detection performance of visible CAC and the interscan reproducibility of the proposed method. Moreover, we compare the performance of our method with two clinical calcium scoring methods that use the 130 HU threshold for segmentation of CAC: manual calcium scoring and automatic calcium scoring with a validated deep learning method.¹⁰⁴ Finally, we perform an ablation study to evaluate the different components of the method.

Prior to heart segmentation all scans were resampled to 1.5 mm in-plane resolution for standardization across the dataset and 3.0 mm slice spacing, the slice spacing of manual reference annotations. The heart segmentation CNN was trained and evaluated using a subset of 199 radiotherapy treatment planning CT scans of the training set, for which reference heart segmentations were available. Of this subset, 100 scans were used for training, 20 scans for validation and 79 scans for evaluation. The CNN was trained for 250,000 iterations with batches of 10 patches of 64x64x64 voxels. Batches were balanced for presence of the heart during training. After training, the heart segmentation CNN achieved a median Dice score of 0.95 (interquartile range 0.94–0.96).

After the heart segmentation, images were resampled to 1.0 mm in-plane resolution, which was the average resolution of the training set, and 1.5 mm slice spacing, for comparison with clinical calcium scoring.¹⁰⁴ Moreover, slices were clipped between -50 HU and 950 HU, for optimal contrast between soft tissue and CAC, and scaled between 0 and 1. The training set of 626 low dose chest CTs and 314 radiotherapy treatment planning CTs was used. Given that the method is aimed for analysis of RTP CT, the validation set of 200 such scans of 100 breast cancer patients (100 pairs) was used. During training manually defined reference labels indicating the presence of CAC in axial CT slices were used.

The slice classification CNN was trained for 1,500,000 iterations using batches of 20 axial CT slices of 224x224 voxels. Batches were randomly sampled and balanced for presence of visible CAC. After training, the slice classification CNN achieved a per slice accuracy of 0.93, with a sensitivity of 0.78 for visible CAC and a false positive rate of 0.06, on all 17,450 CT slices containing the heart of Test-set 1.

The adjusted CycleGAN for generating the CAC map was trained with the same dataset and preprocessing as the slice classification CNN. The adjusted CycleGAN was trained for 375,000 iterations using batches of 4 axial CT slices of 224x224 voxels. Batches were balanced for both the presence of CAC and relative slice position in the heart, using the heart segmentations. Finally, the predicted CAC maps were resampled to the original in-plane resolution prior to evaluation.

Table 5.2: Per artery detection performance (A) and interscan reproducibility (B) of clinical manual (Clinical_M) and automatic (Clinical_A) calcium scoring, and the proposed method in Test-set 1. CAC detection is shown for pairs and for single scans. Reproducibility is evaluated with the absolute relative difference in CAC pseudo-mass in scan-rescan artery pairs for concordant positive artery pairs ($\Delta_R\text{Pos}$) as well as all artery pairs ($\Delta_R\text{All}$).

	Detection		interscan Reproducibility	
	Single (%)	Pairs (%)	Δ_R All	Δ_R Pos [†]
Proposed				
CAC	81 (79)	34 (67)	0.95	0.55
LAD	57 (92)	26 (84)	0.84	0.61
RCA	16 (57)	5 (36)	1.25	0.35
LCX	8 (67)	3 (50)	0.99	0.33
Clinical_M				
CAC	70 (69)	30 (59)	1.05	0.77
LAD	53 (85)	23 (74)	1.00	0.75
RCA	10 (36)	4 (29)	1.42	1.14
LCX	7 (58)	3 (50)	0.84	0.46
Clinical_A				
CAC	64 (63)	26 (51)	1.15	0.76
LAD	48 (77)	21 (68)	1.07	0.81
RCA	9 (32)	2 (14)	1.63	0.74
LCX	7 (58)	3 (50)	0.86	0.48
	(A)		(B)	

[†] Pairs in which CAC was detected in the artery in both scans by the respective method.

5.5.1 CAC detection per scan

In the clinic CAC is routinely detected on a scan-level and used for CVD risk prediction. To provide insight in the detection performance of visible CAC on a scan-level, Figure 5.4 illustrates the results and Figure 5.3A (proposed) shows the sensitivity, rate of false positive scans (FPR) and F1-score in Test-set 1. The proposed method achieved a sensitivity for visible CAC of 0.83, with an FPR of 0.14, meaning 21 out of 150 scans without visible CAC, and an F1-score of 0.80. In comparison, the sensitivity was 0.69 and 0.67 for manual and automatic clinical calcium scoring, respectively. The F1-score of the clinical calcium scoring methods was comparable with our method, due to a low FPR. Among our false positive cases, in 10 cases a lesion in the aorta was erroneously detected (Figure 5.4, FP Aorta), among which 6 were located in the ostia of the coronary arteries. In 9 scans noise was detected as CAC by the proposed method. In 5 of these 9 scans, the false positive lesion was located near the edge of the heart segmentation (Figure 5.4, FP Noise).

As is shown in Figure 5.4, the proposed method also segments the parts of the calcium lesions that remain below the threshold. Particularly in lesions severely affected by motion, e.g. commonly lesions in the RCA, the difference between the proposed

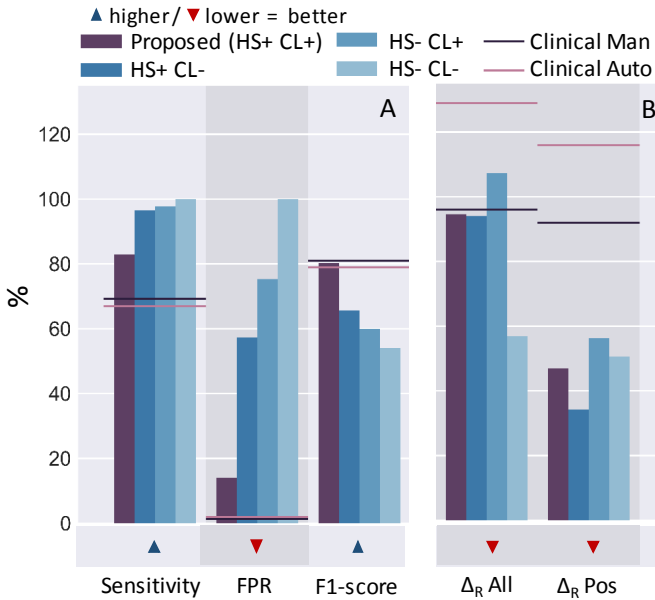


Figure 5.3: The results of the ablation study for the method with and without heart segmentation (HS -/+) and with and without slice classification (CL -/+). We evaluated the detection performance (A) with sensitivity, false positive rate (FPR) and F1-score. We evaluated the interscan reproducibility (B) with absolute relative difference in CAC mass for all pairs (Δ_R All) and concordant positive pairs (Δ_R Pos). For comparison, the performance of clinical threshold-based manual and automatic calcium scoring are shown with horizontal lines.

method and manual clinical calcium scoring is especially pronounced.

5.5.2 CAC detection per artery

Because the impact of cardiac motion is different per location, we additionally evaluate the detection performance per artery. A subset of 34 concordant scan pairs from Test-set 1 that contain CAC with per artery CAC annotations were used for evaluation. Overall, the proposed method detected CAC in 81 out of 102 arteries with lesions (Table 5.2A, left). Out of 51 lesion pairs, the method detected both lesions in 34 pairs (Table 5.2A, right). In comparison, manual clinical calcium scoring detected 70 out of 102 single lesions, and automatic clinical calcium scoring detected 64 out of 102. Manual clinical calcium scoring detected CAC in both lesions in 30 out of 51 lesion pairs and automatic clinical calcium scoring detected 26 pairs out of 51. The achieved improvement of our method over the clinical methods was largest in the RCA, in which 64% of the lesions remained below the 130 HU threshold due to blurring caused by cardiac motion.

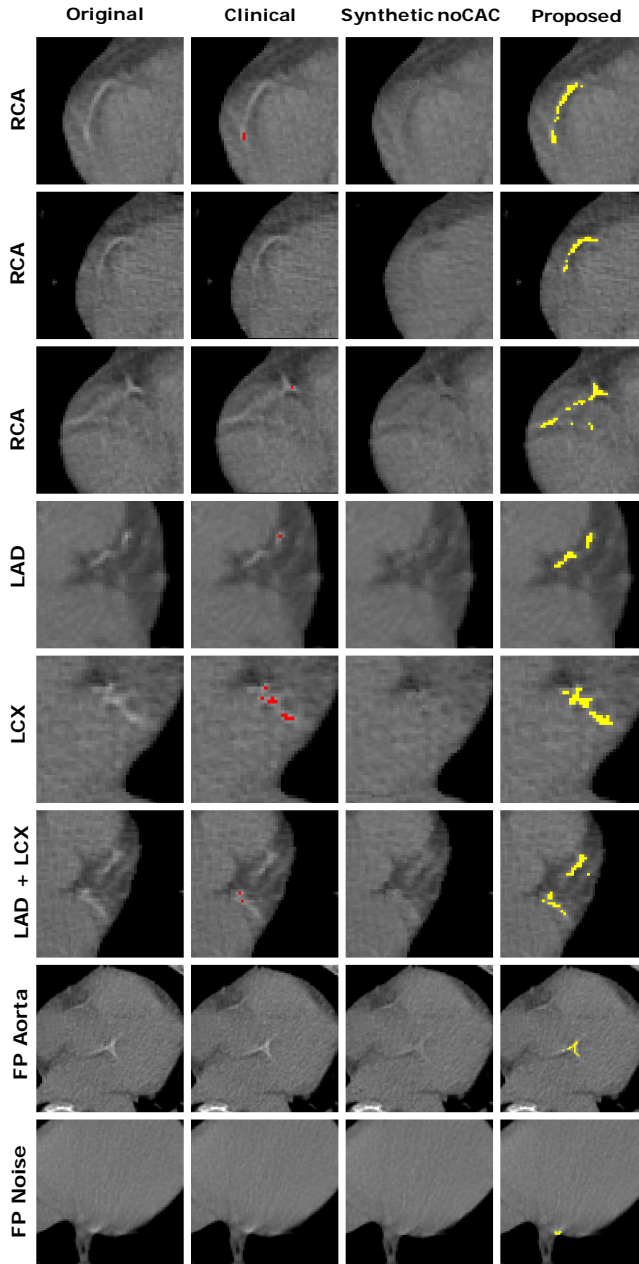


Figure 5.4: Examples of lesions, corresponding synthetic images without CAC and segmentation of CAC with the proposed method and the clinical protocol using the 130 HU threshold.

Table 5.3: Reproducibility of CAC pseudo mass and Agatston scores for automatic clinical calcium scoring and the proposed method in Test-set 2.

1543 pairs	n (%)	CAC mass		AG score
		Δ_R	ICC	ICC
All pairs				
Proposed	1543 (100%)	0.88	0.97 (0.96-0.97)	0.96 (0.96-0.96)
Clinical	1543 (100%)	0.92	0.93 (0.92-0.94)	0.91 (0.91-0.92)
Concordant positive pairs				
Proposed	515 (34%)	0.44	0.96 (0.96-0.97)	0.95 (0.95-0.96)
Clinical	375 (24%)	0.58	0.92 (0.90-0.93)	0.90 (0.88-0.92)

5.5.3 Interscan reproducibility of CAC quantification

To make CAC quantification in non-ECG-synchronized CT scans useful for clinical application the interscan reproducibility ideally should be high. Therefore, the interscan reproducibility of CAC pseudo mass was evaluated in Test-set 1. First, we evaluated the absolute relative difference in CAC pseudo mass in 51 artery pairs using per artery annotations of lesions described in Section 5.2. The interscan difference was 95% on average in all artery pairs (Table 5.2B, left). In concordant positive artery pairs, i.e. pairs in which the method detected CAC in both arteries, the interscan difference was 55% using the proposed method (Table 5.2B, right). Second, we evaluated the reproducibility of total CAC pseudo mass in scan pairs. The interscan difference was 94% on average in all scan pairs (Figure 5.3B, left). For concordant positive scan pairs the absolute relative difference was 47% (Figure 5.3B, right). The higher absolute relative difference in CAC pseudo mass (i.e. lower reproducibility) in all scan pairs compared to positive pairs, is partly explained by the presence of 21 discordant pairs. In these pairs the method shows a high absolute relative difference, since a CAC lesion is only visible in one of the two scans. Additionally, we compared our results to clinical calcium scoring. Both clinical methods show a lower interscan reproducibility than our proposed method in artery pairs as well as in scan pairs. For manual clinical calcium scoring the interscan difference was 77% in concordant positive artery pairs and 89% in concordant positive scan pairs. For automatic clinical calcium scoring the interscan difference was 76% and 114% in artery pairs and scan pairs, respectively. For similar reasons as for our method, the relative interscan difference is high when all scan pairs are considered.

Moreover, the reproducibility of the proposed method is evaluated in Test-set 2, in which manual annotations are not available. Hence, the results of our method are compared with automatic clinical calcium scoring. The results are listed in Table 5.3. The relative difference in CAC mass for positive pairs with detected CAC in both scans was 44%. For all scan pairs this was 88%. Moreover, the interscan agreement of CAC mass was high with an ICC of 0.96 (95% CI 0.96-0.97) in positive pairs and 0.97 (95% CI 0.96-0.97) in all pairs. In comparison, the relative difference was 58% in concordant positive pairs and 92% in all pairs when using the automatic clinical method. Moreover,

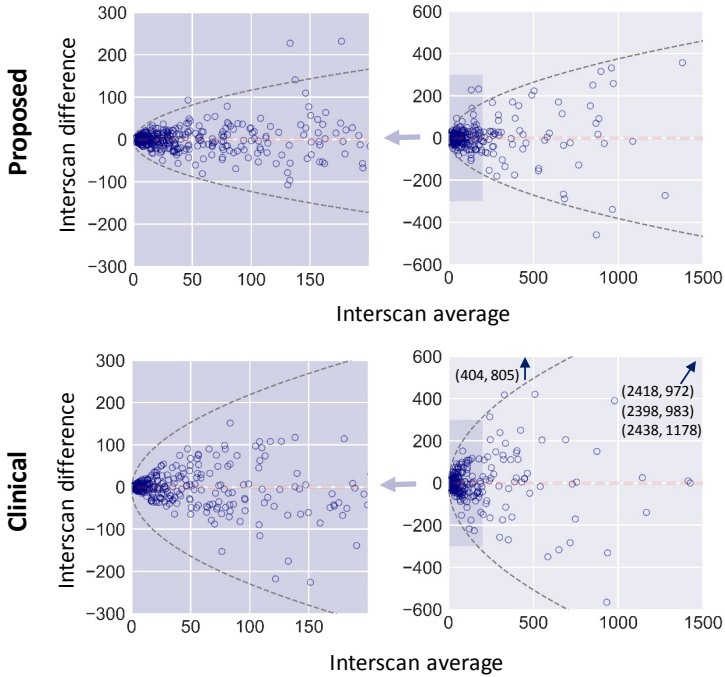


Figure 5.5: Bland-Altman plots for adjusted Agatston scores of the proposed and automatic clinical method. For comparison purposes the adjusted Agatston scores of the proposed method are scaled to the range of scores of the clinical method. Please note that, since the same linear scaling is used for all scores, this does not influence the agreement. 95% limits of agreement are indicated by dashed lines. Because the errors tend to increase with increasing CAC, regression for nonuniform differences was used to model the variation of the absolute differences in scan pairs. To calculate the 95% limits of agreement, the predicted absolute differences were multiplied by $1.96 \times (\pi/2)^{1/2}$, because the absolute differences have a half-normal distribution.⁴⁸

the ICC for the quantified CAC mass was significantly lower for the clinical automatic method than for the proposed method (Table 5.3).

The proposed method achieved an interscan agreement for adjusted Agatston scores with an ICC of 0.96 (95% CI 0.96-0.96). This is significantly better than for the automatic clinical method that achieved an ICC of 0.91 (95% CI 0.91-0.92). This improvement is clearly visible in the Bland-Altman plots (Figure 5.5), where the 95% limits of agreement are narrower for the proposed method than for automatic clinical calcium scoring.

Table 5.4: Interscan agreement of risk categorization based on (adjusted) Agatston score. The risk categories are I: 0-10, II: 11-100, III: 101-400, IV: ≥ 400 .

	Proposed [†]	Clinical _M	Clinical _A
Test-set 1			
κ	0.82 (0.71-0.94)	0.63 (0.42-0.84)	0.64 (0.41-0.87)
Accuracy	0.92	0.92	0.95
Test-set 2			
All pairs			
κ	0.84 (0.82-0.87)	-	0.81 (0.78-0.84)
Accuracy	0.90	-	0.92
Concordant positive pairs [‡]			
κ	0.77 (0.72-0.82)	-	0.67 (0.61-0.72)
Accuracy	0.79	-	0.68

[†] For comparison purposes, the adjusted Agatston scores obtained with the proposed method were linearly scaled to the range of the scores obtained with automatic clinical (Clinical_A) calcium scoring, using the mean score of both methods.

[‡] Pairs in which the Clinical_A method found CAC in both scans.

5.5.4 Ablation Study

To investigate the benefits of prior heart segmentation and slice classification, we evaluated performance using the following settings: 1 - using the settings as proposed, 2 - the proposed method without slice classification, where all CT slices of the heart were analysed, 3 - the proposed method without heart segmentation, 4 - the proposed method without heart segmentation and without slice classification. When heart segmentation was not used, we standardized the field of view over the datasets to include the heart and chest wall. CAC detection, FPR and absolute relative interscan difference of the quantified CAC mass was evaluated on Test-set 1. Figure 5.3 shows the results. The proposed method achieved the best performance, with an F1-score of 0.80 and an absolute relative difference in CAC mass of 47%. While both approaches without slice classification show a high sensitivity, the false positive rate is high: with settings 2 the method found false positive lesions in all negative scans and with settings 4 in 57% of the negative scans. The effect of defining the region of interest by heart segmentation is not only reflected in a better detection, but also in a higher interscan reproducibility of CAC mass, i.e. lower absolute relative difference.

5.5.5 Risk category assignment

Because CVD risk stratification based on the Agatston score is clinically relevant, we evaluate the interscan reproducibility of risk categorization. The interscan agreement of CVD risk categories is shown in Table 5.4. The agreement in Test-set 1 was higher using scores derived with the proposed ($\kappa = 0.82$) method than with manual or automatic clinical calcium scoring, with κ of 0.63 and 0.64 respectively. In concordant positive

pairs of Test-set 2 the κ was 0.77 for the proposed method. In contrast, for automatic clinical calcium scoring this was 0.67.

5.6 Discussion

We presented a method for calcium scoring that does not depend on the clinically accepted intensity value threshold of 130 HU for segmentation of CAC. Hence, the method can identify CAC that may remain undetected by the clinical threshold due to e.g. partial volume effect in small lesions or blurring due to cardiac motion. We achieve this by separating an image containing CAC into an image without CAC and an image only containing CAC. For this we exploit a CycleGAN that translates images between the CAC and *noCAC* domain. We have shown that segmentation of lesions using our method enables increased interscan reproducibility of CAC quantification, compared to clinically used manual and automatic CAC scoring, that both use intensity level thresholding for CAC segmentation.

Previous research shows that interscan reproducibility in non-ECG synchronized CT scans is lower than in dedicated cardiac CT.^{114–117} Because of limitations in the acquisition of radiotherapy treatment planning CTs, our method does not exceed the in literature reported reproducibility in dedicated cardiac CT. However, CAC scoring is increasingly performed in non-dedicated CT scans, hence, the need for more reproducible CAC quantification is growing, especially for quantification in challenging scans like the ones used in this work. Our proposed method shows potential for increasing the interscan reproducibility in non-ECG synchronized scans to the level of dedicated cardiac CT.

In contrast to clinical methods, the proposed method is able to detect visible CAC lesions that remain below the threshold. Owing to a more accurate segmentation that includes pairs of lesions below the threshold, our method outperforms clinical CAC scoring methods in interscan reproducibility of CAC quantification. The results of per artery evaluation show that this improvement was especially pronounced for CAC in the RCA. This is probably due to the ability of the proposed method to detect parts of lesions that are heavily affected by motion artefacts, which occurs more for the lesions in the RCA than in the LAD and LCX.

Errors were made in a number of scans. False positive detections mostly consisted of noise in the proximity of an artery or other types of calcifications than CAC, such as calcifications in the aorta or cardiac valves. Despite noise augmentation during training and post processing, the CAC map contained noise in a few cases. This was often located close to the apex and near the RCA, where the noise level was most severe. Noise reduction strategies described in e.g. Wolterink et al.¹³⁸ could offer a solution. False positives in the aorta or valves were often close to the ostia of the coronaries or occurred in the mitral valve. Since in scans without ECG triggering it is often difficult to distinguish CAC from calcifications in the thoracic aorta or mitral valve calcifications,

these type of errors are also not uncommon for human observers¹³⁹ and are also present in other automatic methods.^{32,104,120} False negative lesions were typically small with very low voxel intensities, making them difficult to distinguish from noise and soft tissue, also for experts.

Similarly to our preliminary work,¹³⁵ the method without heart segmentation and slice classification suffered from a large amount of false positives in the CAC map, which made fully automatic quantification of CAC infeasible. In our ablation experiment, we showed that adding slice classification with respect to CAC presence and heart segmentation solves this issue. Using classification to identify the slices containing CAC, prevents false positive errors in slices that do not contain CAC. By adding heart segmentation, we prevent anatomical structures outside the heart in the data from influencing the CAC segmentation, resulting in more reproducible quantification. Likewise, in previous methods for CAC quantification in scans with high noise levels, a false positive reduction step was used to decrease the number of false positive findings.^{32,104} Possibly, if a very large amount of training data would be available, the method would be able to learn to avoid these errors. However, in the medical field this is often challenging to obtain. In this work, heart segmentation and slice classification are performed independently. Although they are relatively straightforward, future work will investigate merging these steps into a single network.

Other approaches for increasing the reproducibility of CAC quantification have previously been proposed.^{121–125} Since the true amount of CAC is not available for patients, often a phantom is used for development and evaluation of the method. This makes direct comparison of our method infeasible. Šprem et al.¹²⁴ additionally evaluated their approach in a set of 293 subject CT scans and showed an improved reproducibility of CAC volume between different CT images. However, different reconstructions of the same acquisition were used, excluding interscan differences, in contrast to two separate acquisitions used in our study. Moreover, in contrast to our approach a major drawback of previously proposed methods is the use of the standard 130 HU intensity threshold for detection of CAC prior to partial volume correction, inevitably missing lesions below the threshold.

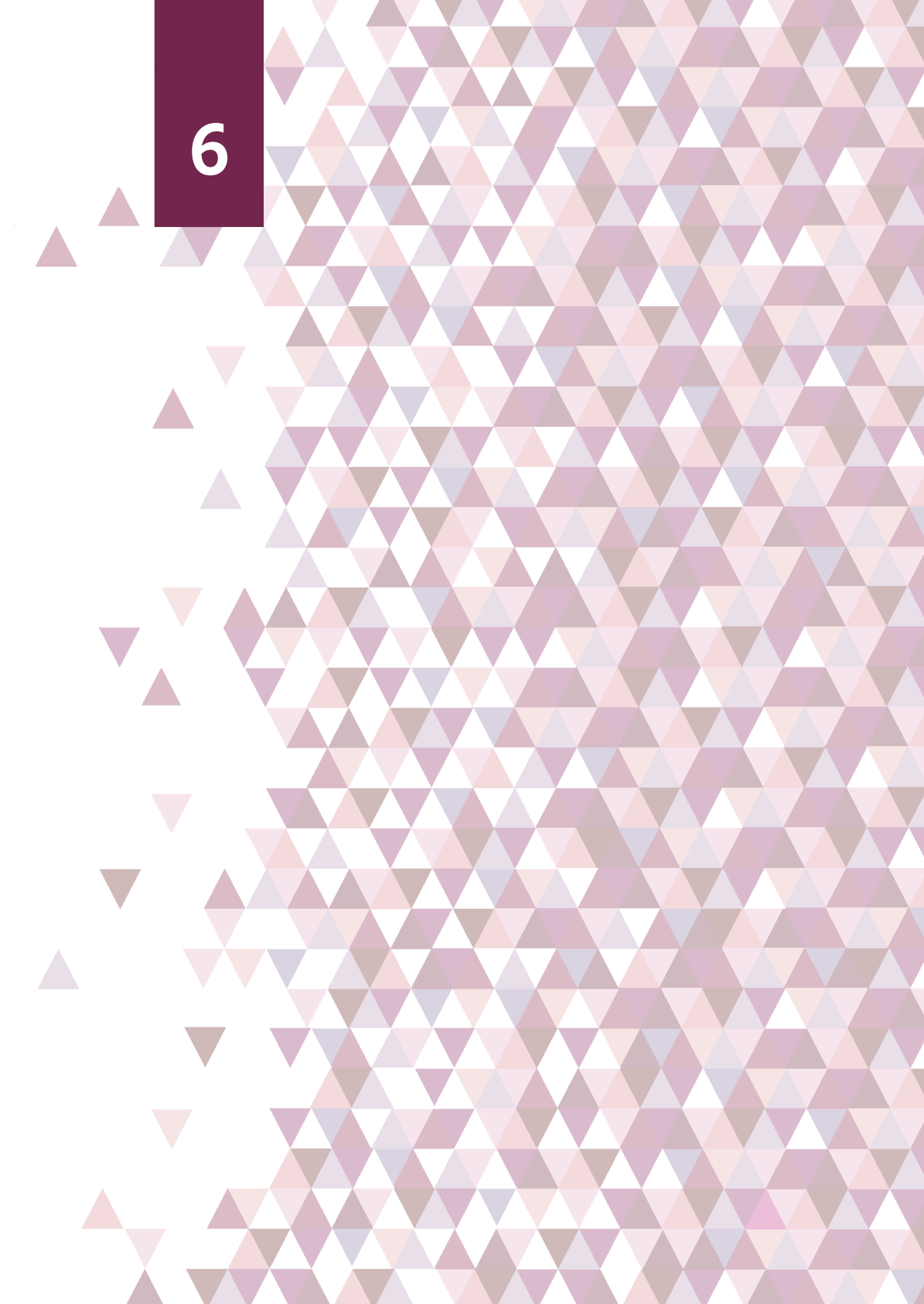
Risk categorization of subjects is typically done using total CAC Agatston scores. However, the proposed method also includes parts of lesions below the clinical threshold, and thus, the Agatston score tends to be (substantially) higher than for the clinical methods. Therefore, assigning risk categories based on the Agatston score derived with the proposed method would necessarily lead to different definitions of CVD risk categories. In order to estimate the performance of risk categorization of the proposed method, the obtained Agatston scores were scaled to the range of clinical Agatston scores. The results showed an improved agreement for the proposed method, compared to automatic clinical calcium scoring. However, to establish the clinical relevance, further research relating the adjusted Agatston scores with CHD events is warranted.

5.7 Conclusion

We have proposed a method for detection and quantification of CAC in non ECG-synchronized chest CT scans that does not require the application of the clinically used intensity level threshold. Hence, the method allows detection of visible CAC that may not exceed the standardly used 130 HU threshold due to e.g. partial volume effect or blurring due to cardiac motion. This leads to more reliable quantification of CAC and increased interscan reproducibility of risk categorization compared to the current clinical standard.

5.8 Acknowledgment

The authors gratefully acknowledge the Dutch Cancer Society for the financial support (NCT03206333). The authors thank the National Cancer Institute for access to NCI's data collected by the National Lung Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.



CHAPTER 6

Direct prediction of cardiovascular mortality from low-dose chest CT

BASED ON: S. G. M. van Velzen, M. Zreik, N. Lessmann, M. A. Viergever, P. A. de Jong, H. M. Verkooijen, and I. Išgum. "Direct prediction of cardiovascular mortality from low-dose chest CT using deep learning", *Medical Imaging*, vol. 10949 of *Proceedings of SPIE* (2019), 1094932.

Abstract

Cardiovascular disease (CVD) is a leading cause of death in the lung cancer screening population. Chest CT scans made in lung cancer screening are suitable for identification of participants at risk of CVD. Existing methods analyzing CT images from lung cancer screening for prediction of CVD events or mortality use engineered features extracted from the images combined with patient information. In this work we propose a method that automatically predicts 5-year cardiovascular mortality directly from chest CT scans without the need for hand-crafting image features.

A set of 1,583 participants of the National Lung Screening Trial was included (1,188 survivors, 395 non-survivors). Low-dose chest CT images acquired at baseline were analyzed and the follow-up time was 5 years. To limit the analysis to the heart region, the heart was first localized by our previously developed algorithm for organ localization exploiting convolutional neural networks. Thereafter, a convolutional autoencoder was used to encode the identified heart region. Finally, based on the extracted encodings subjects were classified into survivors or non-survivors using a neural network. The performance of the method was assessed in eight cross-validation experiments with 1,433 images used for training, 50 for validation and 100 for testing. The method achieved a performance with an area under the ROC curve of 0.73.

The results demonstrate that prediction of cardiovascular mortality directly from low-dose screening chest CT scans, without hand-crafted features, is feasible, allowing identification of subjects at risk of fatal CVD events.

6.1 Purpose

Besides lung cancer, cardiovascular disease is a leading cause of death in the lung cancer screening population⁸². Moreover, it has been shown that chest CT scans used for lung cancer screening are suitable for identification of participants at risk of cardiovascular disease (CVD)¹⁴⁰⁻¹⁴³. Previous methods that investigated prediction of CVD events and all-cause mortality used known quantitative CVD image markers and combined them with subject data. Using weighted Cox proportional hazards regression, Chiles et al.¹⁴¹ showed that quantitative as well as visually assessed coronary artery calcium (CAC) scores extracted from screening low-dose chest CT are predictive for CVD and all-cause mortality in the National Lung Screening Trial (NLST). Similarly, Mets et al.¹⁴² used Cox regression based on semi-automatically detected CAC scores and thoracic aorta calcium (TAC) volume as well as subject data, to perform prediction of CVD events and all-cause mortality in the Dutch-Belgian lung cancer screening trial (NELSON). De Vos et al.¹⁴³ performed prediction of CVD events in the same population using a support vector machine (SVM) classifier that employed automatically extracted CAC and TAC scores as features.

These approaches relied on hand-crafted features that are already established as CVD biomarkers. However, besides these known biomarkers, chest CT scans may contain yet unknown features predictive of CVD mortality. Hence, we propose a method based on unsupervised feature learning which is able to automatically predict CVD mortality directly from chest CT scans and is therefore not limited to known quantitative image markers related to CVD.

6.2 Data

This study included 1,583 participants of the National Lung Screening Trial. NLST included current and former heavy smokers between the age of 50 and 74⁸². All 395 participants who died of CVD within 5 years from acquisition of the baseline CT scan (non-survivors) were included. In addition, 1,188 participants who were still alive after this period (survivors) were randomly selected.

For each subject a CT scan acquired at baseline was analyzed. Low-dose chest CT scans were made with breath-hold, without contrast enhancement and without ECG synchronization. Scans were acquired using helical scanning mode and a tube voltage of 120 kVp or 140 kVp, depending on the subject's weight. In-plane resolution ranged from 0.49 mm to 0.98 mm with a slice thickness between 1.0 mm and 2.5 mm. The scans were acquired in 32 different medical centers with 13 different scanner models.

6.3 Methods

To investigate whether analysis of the heart visualized in chest CT enables prediction of CVD mortality, the method first extracts a bounding box around the heart. This is done with our previously designed and trained algorithm⁸¹ that employs a CNN to determine the presence of the heart in axial, coronal and sagittal slices of the chest CT image and subsequently combines these to define a 3D bounding box around the heart. Thereafter, to ensure equal image resolution in our data set, we resample all extracted heart volumes to isotropic resolution of 1.0 mm. Moreover, to enhance differentiation among soft tissues in the heart (e.g. fat, muscle) and to preserve influence of high density structures (e.g. CAC, TAC), extracted volumes are clipped between $[-160, 840]$ HU.

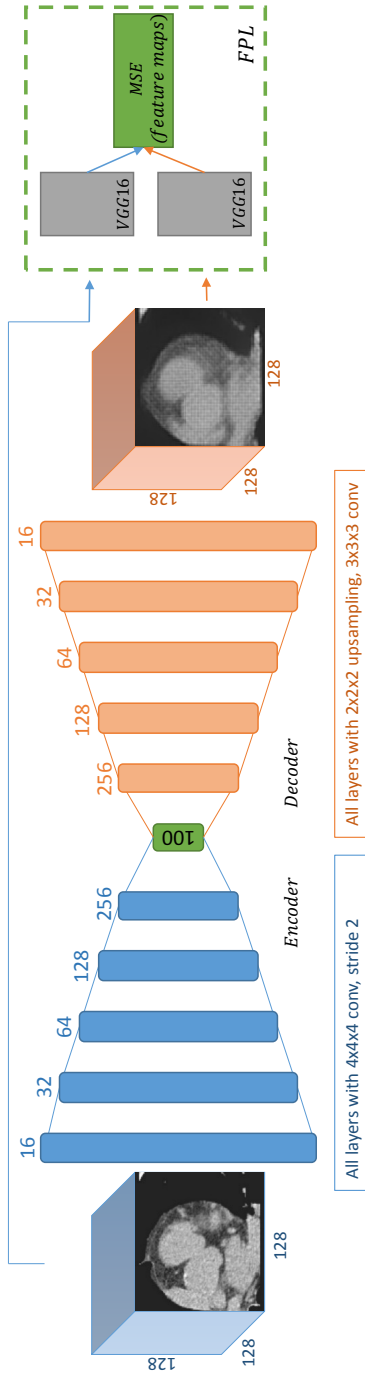


Figure 6.1: Architecture of the CAE with feature perceptual loss (FPL): Input images are encoded into 100 encodings and reconstructed by the decoder. The encoder consists of 5 convolutional layers with $4 \times 4 \times 4$ kernels and a stride of 2. The decoder consists of 5 upsampling layers, each followed by a $3 \times 3 \times 3$ convolution layer. Both the input and reconstructed images are fed into a fixed VGG16 network, where the FPL is determined by the MSE over the feature maps derived from the input image and the reconstructed image.

Because of the relatively low number of the available samples, prediction using e.g. a CNN that would extract features and perform classification of subjects into survivors and non-survivors was not feasible. Therefore, similar to the work of Zreik et al.¹⁴⁴, a convolutional autoencoder (CAE) is used to encode the volumes containing the heart in an unsupervised fashion. Thereafter, a conventional machine learning classifier exploiting the extracted encodings is employed to classify subjects into survivors and non-survivors.

The CAE used in this work (Figure 6.1) consists of an encoder that compresses the images into representative encodings and a decoder that reconstructs the images during training. The encoder analyzes images, cropped to the heart volume, zero-padded to $128 \times 128 \times 128$ voxels and consists of 5 convolutional layers with $4 \times 4 \times 4$ kernels. A stride of 2 was applied to achieve spatial downsampling without the need for deterministic spatial functions such as max-pooling. The encoder ends in a dense layer with 100 units, which represents the encodings vector. The decoder consists of 5 upsampling layers and 5 convolutional layers with $3 \times 3 \times 3$ kernels and a stride of 1. All convolutional layers are followed by batch normalization and LeakyReLU activation ($\alpha = 0.3$).

Typically, in training the CAE, the mean squared error (MSE) between the reconstructed image and the original image is used as a loss function. However, to capture the contrast among soft tissues in low-dose CT without intravenous contrast enhancement, in this work the loss function of the CAE is defined employing the feature perceptual loss (FPL), which captures perceptual differences and spatial correlations better¹⁴⁵. To compute the FPL, both the input image and reconstructed image are separately fed into a fixed VGG16¹⁴⁶ network pretrained on ImageNet¹⁴⁷ (Figure 6.1). The FPL is then defined as the MSE between the feature maps in this network derived from the input image and the reconstructed image. Because VGG16 is a network designed for 2D images, the loss was calculated over 2D axial slices of the 3D image volumes.

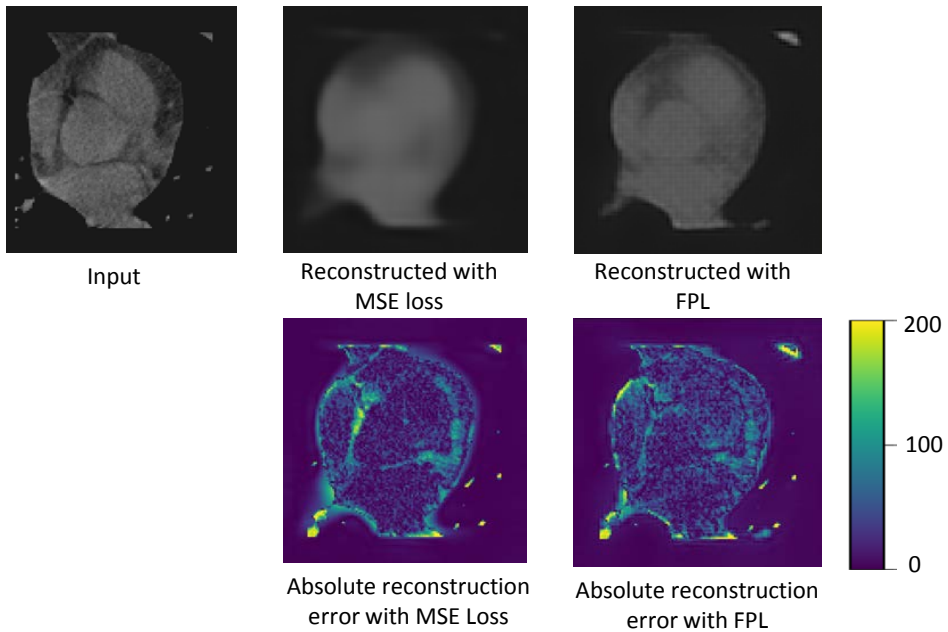
Thereafter, three different classifiers are trained using the encodings obtained with the CAE: a neural network (NN), a random forest classifier (RFC) and a support vector machine (SVM) classifier. For the RFC and SVM a grid search on the validation set was performed to find the optimal parameter settings.

6.4 Experiments and results

To assess the performance, eight cross-validation experiments were performed. In each experiment, 100 images were selected as test set, 50 images were selected as validation set and the remaining 1,433 images were used as training set. Test and validation sets were balanced with respect to classes. In the test sets, images of the non-survivors were sampled such that each non-survivor was included once in the test sets. Images of the survivors were randomly selected from the available set.

To augment the training set for the CAE, random rotations around all three image

Figure 6.2: Example of an input image and reconstructed images generated by two different CAEs: one trained with a standard MSE loss and one trained with the FPL calculated with VGG16 pre-trained with ImageNet (top). Corresponding error maps depict the absolute reconstruction error between the reconstructed image and the input image (bottom).

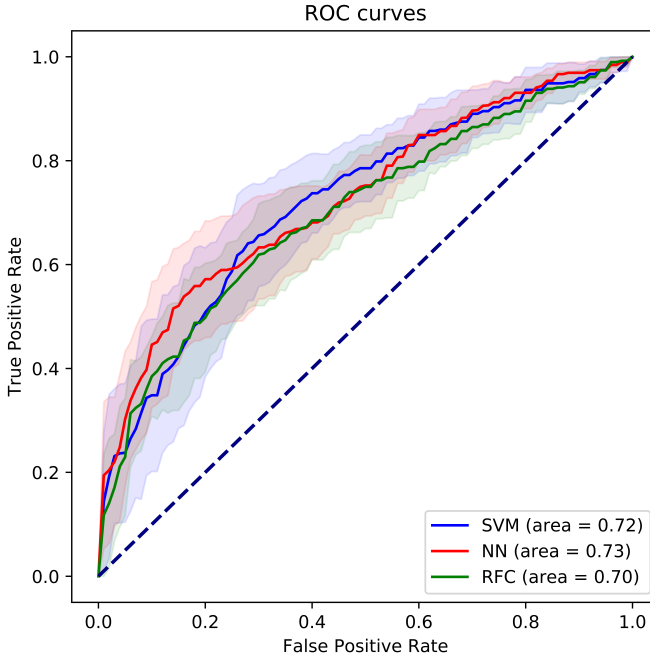


axes were used. The angle of rotation was randomly chosen from a normal distribution with a mean of 0 and a standard deviation of 10 degrees. The CAE was trained in 100,000 iterations, with a batch size of 2 images zero-padded to the required input size. The Adam optimization algorithm was used with a learning rate of 0.001.

To classify subjects into survivors and non-survivors, the classifiers were trained using the encodings obtained with the CAE. The NN (6 units dense layer and 2 units output, dropout $p=0.5$, categorical crossentropy loss, learning rate=0.0001) was trained in 25,000 iterations with balanced batches of 100 examples, using the Adam optimization algorithm. The RFC consisted of 75 trees and for the SVM the γ and c were set to 0.0001 and 100, respectively.

To determine the influence of FPL on the reconstructions, we trained an additional CAE with standard MSE as loss function. Using MSE as a loss function resulted in a mean absolute error of $19 (\pm 5)$ HU and training with FPL resulted in mean absolute error of $20 (\pm 6)$ HU. While the mean absolute reconstruction errors are similar, the

Figure 6.3: Average ROC curves and their standard deviation for prediction of cardiovascular mortality with three different classifiers: SVM, NN and RFC, over 8 cross-validation experiments.



CAE trained with FPL learned sharper contrast between structures, as shown in Figure 6.2.

In each cross-validation experiment a CAE was trained and the classifiers were trained and the performance of the method evaluated. The performance of the method was assessed with a receiver operating characteristic (ROC) curve. The neural network achieved the best performance with an area under the curve (AUC) of 0.73 (± 0.07 standard deviation, Figure 6.3). The ROC of the SVM and the RFC had an AUC of 0.72 (± 0.07) and 0.70 (± 0.06), respectively.

6.5 Discussion and conclusion

In this work, a method for prediction of cardiovascular mortality from lung screening chest CT scans has been proposed. Unlike previous predictive models, the proposed method does not use hand-crafted image features, but performs prediction directly from the images containing the heart.

The experiments show that the CAE using FPL preserves structures likely containing

important information for prediction of CVD mortality, such as the coronary arteries, aorta, and fat around the heart (Figure 6.2). The CAE was trained to reconstruct heart images and was agnostic to the subsequent classification task. It would be interesting to investigate end-to-end training where the CAE is trained while optimizing for a subsequent classification task. Besides likely improvement in the performance by end-to-end training, such an approach might allow identification of image areas important for prediction and thereby allow the confirmation of known and possibly the discovery of novel image markers of CVD mortality. Alternatively, a CNN could be employed to perform classification directly. However, our preliminary experiments showed that end-to-end training and direct classification require a larger data set, which is often not available.

The three evaluated classifiers show a similar performance for prediction of CVD mortality based on the extracted encodings. Furthermore, the presented method shows a similar performance to previous methods proposed by Mets et al. and De Vos et al. (AUC = 0.71),^{142,143} that describe prediction of CVD events. This is remarkable since the CAE is unsupervised and does not incorporate prior knowledge about the image, the subsequent classification task or relevant biomarkers. Moreover, the proposed method achieved similar performance to work by Mets et al. and De Vos et al.^{142,143}, without adding subject data like age, smoking history or sex, as they proposed. However, comparison with these methods is somewhat limited by different outcome definitions and populations. Although image analysis without additional subject data may be simpler in application, future work could investigate whether incorporating subject data in the proposed method would improve the performance.

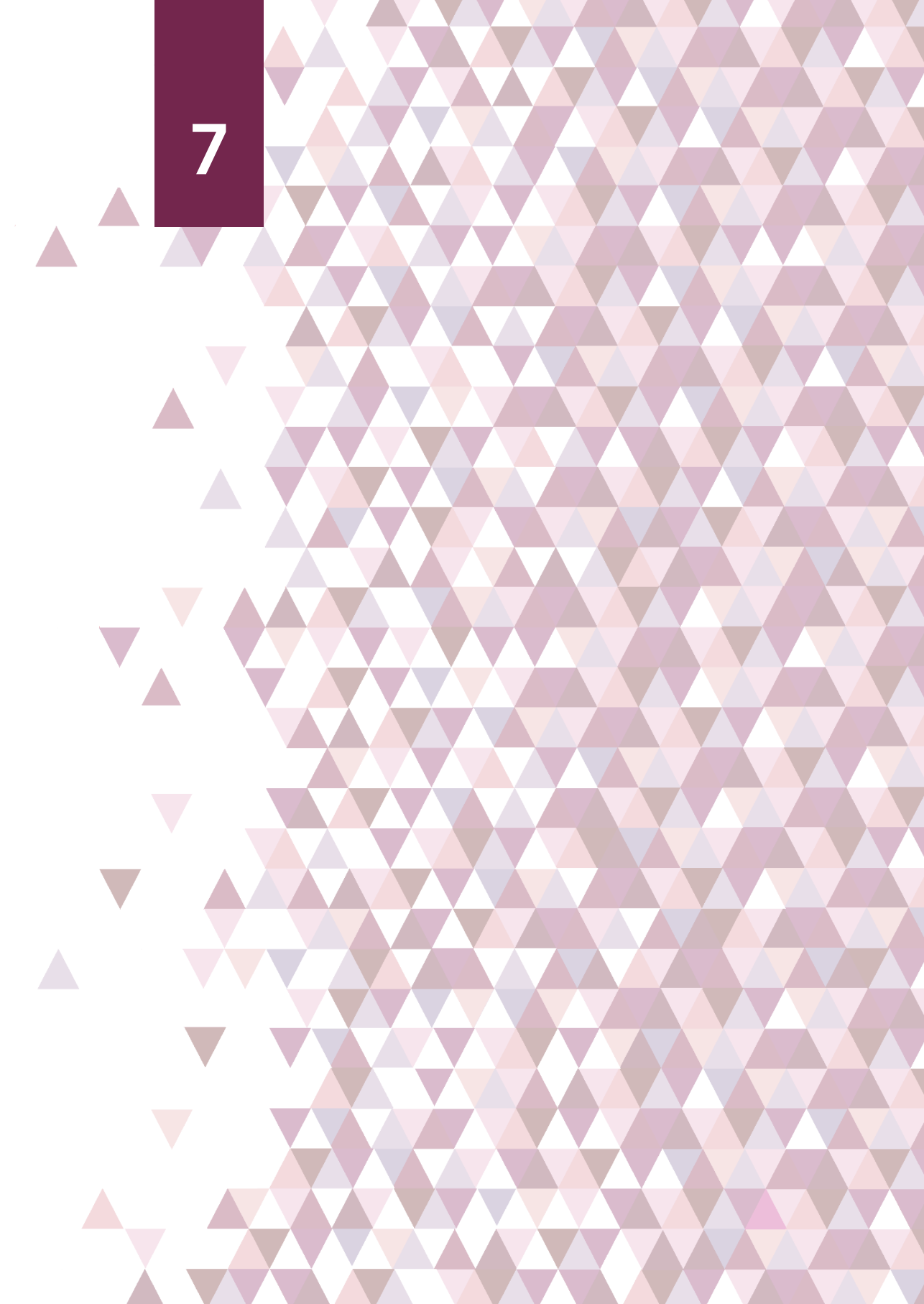
In conclusion, this work demonstrates that the prediction of cardiovascular mortality directly from low-dose screening chest CT scans is feasible. This might allow identification of subjects undergoing lung screening with CT who are at risk of fatal CVD events and who might benefit from preventive treatment.

6.6 New or breakthrough work

A machine learning system is presented that is able to predict cardiovascular death within five years from a low-dose chest CT scan, without prior information about the image or subjects.

Acknowledgments

The authors gratefully acknowledge the Dutch Cancer Society for the financial support. The authors thank the National Cancer Institute for access to NCI's data collected by the National Lung Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.



CHAPTER 7

AI for estimation of radiotherapy-related risk of ischemic heart disease in breast cancer survivors

BASED ON: S. G. M. van Velzen, R. Gal, S. Bruns, J. M. Wolterink, T. Leiner, A. J. Teske, F. van der Leij, D. H. van den Bongard, M. A. Viergever, H. M. Verkooijen, and I. Išgum. "AI for estimation of radiotherapy-related risk of ischemic heart disease in breast cancer survivors". Submitted.

Abstract

PURPOSE To investigate whether dose planned to cardiac structures is associated with the risk of ischemic heart disease (IHD) in breast cancer patients treated with radiotherapy, and whether this association is modified by presence of coronary artery calcification (CAC).

MATERIALS AND METHODS Radiotherapy planning CT scans and corresponding dose distribution maps of 5,561 patients were collected, of whom 1,899 patients received their CT scan before 2011 (which ensures long follow-up). CAC was detected automatically. An artificial-intelligence-based method to automatically segment cardiac structures (heart, cardiac chambers, large arteries, three main coronary arteries) was developed. Segmentations allowed determining the planned radiation dose to each structure separately, as well as the mean dose planned to the heart. Patients were assigned to a low-, medium-, or high-dose group based on the dose to the respective heart structure. Per patient, information on IHD hospitalization and mortality was obtained. The association of planned radiation dose to cardiac structures with risk of IHD was investigated in patients with and without CAC using Cox proportional hazard analysis in the long follow-up population. Tests for interaction were performed.

RESULTS After a median follow-up of 96.0 months (IQR 84.2-110.4) in the long follow-up group, 135 patients were hospitalized for IHD or died from IHD. For every 1 Gy increase on cardiac structures, the IHD risk increased by 3% -11%. The absolute increase in IHD risk was substantially higher in patients with CAC (event-rate_{low-dose}=14-15 vs. event-rate_{high-dose}=15-34 per 1,000 person-years) than in patients without CAC (event-rate_{low-dose}=6-8 vs. event-rate_{high-dose}=5-17 per 1,000 person-years). No interaction between CAC and radiation-dose was found.

CONCLUSIONS Radiation exposure of cardiac structures is associated with increased risk of IHD. Automatic segmentation of cardiac structures enables spatially localized dose estimation, which may aid in prevention of radiotherapy-induced cardiac damage. This may be especially valuable in breast cancer patients with CAC.

7.1 Introduction

Increasing incidence of breast cancer and advances in its treatment lead to increasing number of women living with the treatment consequences.^{61,148} Cardiovascular disease is among the leading causes of death in breast cancer survivors and in most older women with breast cancer it is the primary cause of death.³ The majority of patients with breast cancer is treated with adjuvant radiotherapy.¹⁴⁹ Radiation exposure of the heart is known to increase the risk of ischemic heart disease (IHD).¹⁹

Multiple studies have reported an increased risk of IHD with increasing mean heart dose (MHD).¹⁹ However, the dose distribution within the heart is not homogeneous. A study by Jacob et al. showed that the dose delivered to the left ventricle and the three main coronary arteries may differ substantially from the MHD.¹⁵⁰ Van den Bogaard et al. showed that replacing the MHD in a prediction model by the dose to the left ventricle improved prediction of acute coronary events.¹⁵¹ Moreover, previous research showed that a high dose to the coronary arteries can lead to an increased risk of stenosis in these arteries.^{152,153}

Taking into account the irradiation dose to specific heart structures during treatment planning may reduce the risk of IHD.¹⁵⁰ However, CT scans used for radiotherapy treatment planning are acquired without contrast agent and without ECG-triggering. Hence, visualization of cardiac structures (including the coronary arteries) is suboptimal, which hampers manual delineation of those structures. On the other hand, cardiac structures can be visualized with high accuracy in coronary CT angiography (CCTA) scans. Jacob et al. used registration to transfer delineations made in CCTA scans to the planning CT.¹⁵⁰ However, CCTA is not part of routine clinical care for breast cancer patients. Accordingly, other methods have been proposed that use atlas registration to delineate cardiac structures directly on the planning CT scan.^{154–156} Although several automatic methods for delineation of cardiac structures for dosimetry have been proposed, a large-scale study investigating the doses on these structures and relating these to IHD risk has not been conducted.

The risk of cardiotoxicity, including IHD, is higher in patients with pre-existing risk factors.¹⁵⁷ The presence of coronary artery calcification (CAC), which indicates the presence of atherosclerosis, is a strong and independent predictor of IHD events, even in individuals that have no traditional IHD risk factors.¹³ CAC is visible and can be quantified on radiotherapy planning CT scans and, hence, can be used to identify breast cancer patients at risk of IHD.^{12,33,104}

In this study we build on a deep learning method described by Bruns et al.¹⁵⁸ to develop a method for segmenting cardiac chambers and large arteries (the left ventricular myocardium, right ventricle (RV), left atrium (LA), right atrium (RA), ascending aorta, and pulmonary artery trunk) and localizing the trajectories of the three major coronary arteries in planning CT scans. Next, we investigate the association between radiation exposure of cardiac structures with risk of IHD in patients, and evaluate whether this

association is modified by the presence of CAC.

7.2 Material and Methods

7.2.1 Study population and data collection

This study included patients in the Utrecht Region who received radiotherapy treatment between 2005 and 2016 for primary breast cancer at the University Medical Center Utrecht.⁹⁷ For each patient, a standard CT scan (no contrast-enhancement, no ECG-triggering, free breathing, 120 kVp, 0.78-1.37 mm in-plane resolution, 3 mm slice thickness and 3 mm slice spacing) of the chest used for treatment planning for irradiation of the breast was collected. Dose distribution maps were obtained from the radiotherapy treatment planning software used in clinic. 3D-CT radiotherapy treatment planning was performed using Plato-RTS (Nucletron B.V.; 2005-2012), XiO (Elekta; 2012-2016) or Monaco (Elekta; 2016). If patients had two or more dose distribution maps, e.g. because of irradiation for recurrent breast cancer, the first map was used for analysis. Patients were excluded when they had metastasis at presentation or when planning CT scans were obtained ≥ 1 year after breast cancer diagnosis. Since all data of breast cancer patients were retrospectively and anonymously collected, the need for informed consent was waived by the local institutional review board.

For development of the segmentation method, 18 contrast-enhanced cardiac CT scans from the CLARITY study were included.¹⁵⁹ Scans were acquired on a dual-layer-detector CT scanner (IQon, Philips Healthcare, Best, The Netherlands). By using the dual-energy information, a CCTA and a virtual non-contrast (VNC) image, resembling a non-contrast-enhanced CT, could be reconstructed from the contrast-enhanced acquisition. To obtain the reference for training, the cardiac structures and the three main coronary arteries were segmented in the CCTA scans. Reference segmentations of the cardiac chambers and large arteries were made in a semi-automatic manner as described by Bruns et al.¹⁵⁸ and included the left ventricular (LV) myocardium, LV blood cavity, right ventricle (RV), left atrium (LA), right atrium (RA), ascending aorta, and pulmonary artery trunk. For reference segmentations of the coronary arteries, we employed an automatic artery tracking method in the CCTA scans.¹⁶⁰ Subsequently, an expert selected the main branches of the left anterior descending (LAD), left circumflex (LCX) and right coronary artery (RCA), verified the segmentations, and if needed manually corrected them.

7.2.2 Coronary artery calcium scores

Presence of atherosclerosis was measured by automatic CAC detection on radiotherapy treatment planning CT scans of all breast cancer patients using a validated deep learning method.¹⁰⁴ Patients were categorized as having CAC and not having CAC.¹⁶¹

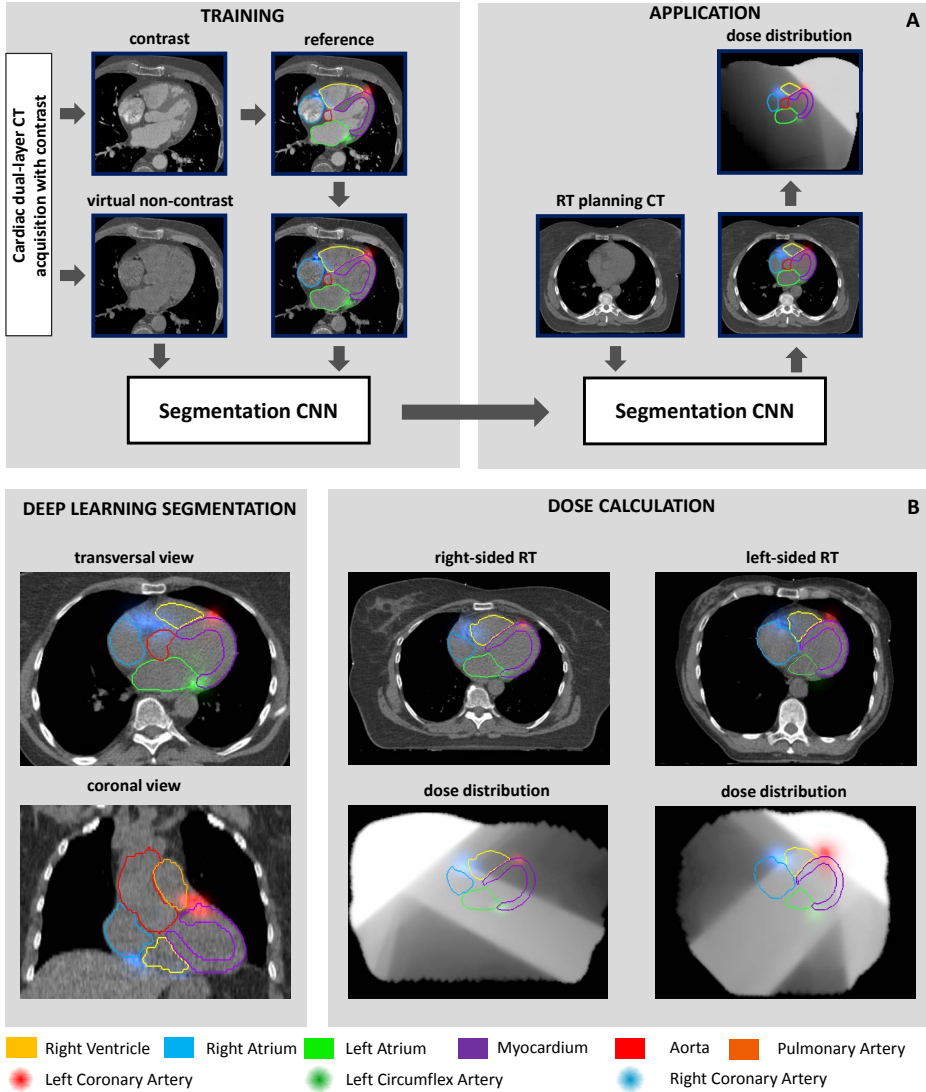
7.2.3 Segmentation of cardiac structures

We build on a segmentation method, that was proposed for segmentation of cardiac chambers and large arteries in non-contrast-enhanced ECG-synchronized cardiac CT.¹⁵⁸ Cardiac structures are generally poorly visible in non-contrast-enhanced CT scans, which renders it practically infeasible to segment them manually. On the other hand, CT scans with contrast clearly visualize cardiac chambers, large arteries and coronary arteries, and allow their segmentation. Hence, to segment cardiac chambers and large arteries in non-contrast-enhanced CT, Bruns et al.¹⁵⁸ proposed to leverage the dual-energy information provided by a dual-layer-detector CT scanner, which allows reconstruction of a virtual non-contrast (VNC) CT image from the contrast-enhanced acquisition. The VNC mimics a non-contrast-enhanced image and is perfectly aligned with the contrast-enhanced image, because both scans are reconstructed from the same acquisition. Therefore, segmentations made on the contrast-enhanced image can be transferred with voxel accuracy to the VNC image. The method is trained using VNC images with reference segmentations (Figure 7.1A, Training) and, thereafter, it can be applied to planning CT images (Figure 7.1A, Application).

Inspired by the method of Bruns et al.,¹⁵⁸ our method employs a 3D convolutional neural network (CNN). The coronary arteries are small structures compared with the other heart structures and are barely visible in non-contrast-enhanced CT, especially in non-ECG-triggered scans. Therefore, simply adding the arteries to the segmentation maps would not provide optimal results. Hence, our network architecture has two output branches: a branch for structure segmentation and a branch for localizing the coronary arteries. Contrary to the branch that segments cardiac structures, the artery localization branch does not predict voxel labels, but predicts per voxel the inverse distance to the target artery segmentation using regression. This renders a heatmap-like prediction of the artery trajectories. For training, we converted the reference segmentations of the coronary arteries to a heatmap by applying an inverse Euclidean distance transform.

The CNN architecture used in this work consists of an encoding path, which is connected to the output branches by six ResNet¹³⁶ blocks that exploit skip connections for optimal information propagation (Figure 7.2). Because an ensemble of CNNs provides more robust predictions than a single CNN,¹⁶² the final segmentation method consists of an ensemble of 18 CNNs.¹⁵⁸ The dataset of VNC images is randomly split into six folds with 15 scans for training, and 3 for validation. In each fold, three CNNs are trained with different random initializations, totalling 18 CNNs. The input of the CNNs consists of slabs, consisting of 5 stacked axial image slices of 256x256 voxels. For optimal performance in the target CT images, the VNC images are resampled to the average resolution of the radiotherapy planning CTs, namely 1.0 mm in-plane resolution and 3 mm slice spacing. All 18 CNNs are trained for 10,000 iterations with mini-batches of 32 randomly sampled image slabs, using Adam¹⁶³ for optimization.

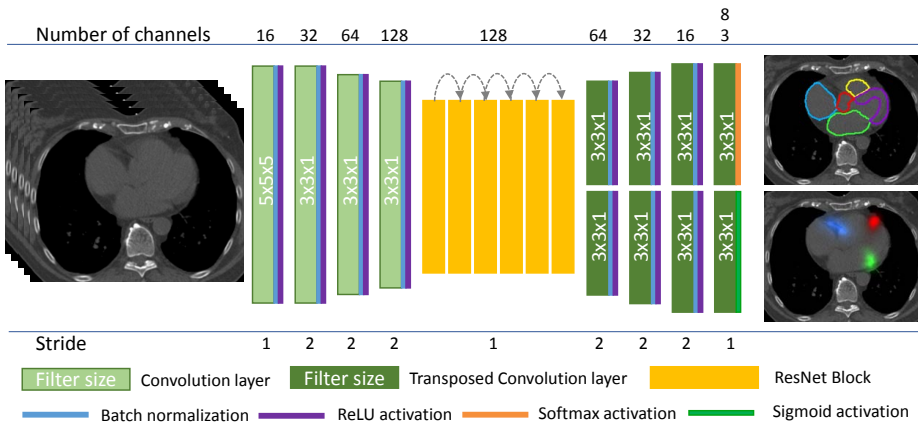
Figure 7.1: A) Schematic overview of the deep learning segmentation method. B) Examples of segmented cardiac structures and localized coronaries. Segmentations overlaid on dose distributions are shown for right-sided and left-sided radiotherapy. Please note that the dose distribution maps are log-transformed for visualization purposes



The loss function for segmentation of the cardiac chambers and large arteries is defined as the negative sum of soft Dice similarity coefficients over all classes. Additionally, for the coronary artery localization an L1 loss of the predicted heatmap with the reference heatmap is calculated. For more details regarding the training, we refer to Bruns et al.¹⁵⁸ To obtain the final prediction, predictions of all 18 CNNs are averaged. Coronary artery heatmaps are thresholded to obtain a segmentation.

After training, the segmentation method was applied to all available radiotherapy treatment planning CT scans (Figure 7.1A, Application).

Figure 7.2: CNN Architecture



7.2.4 Dose calculations

By combining the automatically obtained segmentations of the different anatomies with the dose distribution maps, derived from the same CT scan, we calculated the planned dose to these structures (Figure 7.1B). For each patient, the average dose was calculated for each segmented heart structure (including the whole heart), and expressed in Gray (Gy).

7.2.5 Outcome assessment

The primary outcome was hospitalization for, or death from ischemic heart disease (IHD), as defined by the International Classification of Disease 10th revision (ICD-10)¹⁶⁴ codes I20-25, e.g. myocardial infarction and angina pectoris.

Data on hospitalization for IHD was obtained by linkage with the National Basic Registration Hospital Care (LBZ) from Dutch Hospital Data (DHD). Before 2013, DHD used ICD-9 for registration and from 2013 onwards ICD-10. Therefore, all diagnoses classified according to ICD-9 were converted to ICD-10 codes using the conversion table CVZ'80-ICD-10 2006 of the Dutch WHO-FIC Collaborating Centre.

Data on cancer hospitalization were collected from the LBZ registry and the Netherlands Cancer Registry (NCR). Cancer, including metastases, was defined by ICD-10 codes C00-99; non-melanoma skin cancer (ICD-10 code C44) was excluded. Recurrence of breast cancer was defined by the NCR or as a hospitalization for breast cancer more than 1 year after the radiotherapy treatment planning CT by the LBZ.

Data on death were obtained from the National Cause of Death Register that registers causes of death from all diseased persons in The Netherlands. In this register, the causes of death are classified according to ICD-10.

7.2.6 Statistical analysis

Continuous baseline variables are presented as mean values with standard deviations or medians with interquartile ranges. Categorical baseline variables are presented as total numbers with percentages.

The relation between the MHD and the mean dose on the structures was modelled using linear regression. To decrease the influence of outliers, patients with an MHD above the 95th percentile were excluded. To investigate the predictive value of the MHD for the dose on different structures in a linear model, R squared (R^2) provided by the regression analysis was examined. R^2 indicates the proportion of variance in the dose on the structure that is predictable from MHD. $R^2 < 0.70$ indicates an inhomogeneous dose distribution within the heart, making the MHD not suitable for surrogate parameter purposes.¹⁵⁰ Pearson's correlation coefficients were used for correlation analyses between MHD and mean doses to the different structures.

Follow-up time started at the date of the planning CT scan and ended at the date of the event or censoring. Censoring observations include death from other causes than IHD, diagnosis of breast cancer recurrence, diagnosis of other cancers, or end of follow-up (December 31, 2017).

Since radiotherapy-induced cardiac damage is found to have a long-term effect,⁶ analysis of this damage was performed in a subgroup of patients with a CT scan acquired before 2011, which we refer to as the long follow-up population. Patients were assigned to a low-, medium-, or high-dose group based on the dose on the respective heart structure. For the MHD the dose groups were 0-4 Gy, 4-8 Gy and >8 Gy. The dose groups of cardiac structures were determined based on the dose distributions to the respective structure. Cox proportional hazard models were used to examine the association between dose on cardiac structures and IHD events. The analysis was performed for dose groups as well as for continuous dose variables. All models were adjusted for age at time of the planning CT scan and for treatment with anthracyclines (yes/no). Results were expressed as hazard ratios (HRs) with corresponding 95% confidence intervals (CIs).

To examine whether the association between dose to the cardiac structures and risk of IHD was modified by presence of CAC, a test for interaction was performed by

comparing models with an interaction term of dose with CAC to the model without the interaction term using the likelihood ratio test. Also, the association between dose and IHD was examined in patients without CAC and in patients with CAC.

7.3 Results

7.3.1 Population characteristics

Dose distribution maps and corresponding planning CT scans of in total 5,561 patients were collected. After removing duplicates, linking with NCR and Statistics Netherlands, and exclusion of ineligible patients, 5,301 patients were available for analysis. The long follow-up population consisted of 1,871 patients that received the planning CT scan before 2011.

For the long follow-up population mean age at time of planning CT was 58.7 years (SD = 10.7) and 52.9% received treatment for left-sided breast cancer (Table 7.1). In the long follow-up population 35.4% had CAC and in the full cohort 30.7% had CAC.

7.3.2 Planning dose on heart structures

Figure 7.4 shows the planned radiation dose for each evaluated anatomy together with the MHD. In right-sided radiotherapy, the MHD was on average 1.13 Gy and the mean dose to each cardiac structure was lower than 2 Gy (Table 7.2, Figure 7.3). The most exposed structures were the right atrium and the ascending aorta, with a mean dose of 1.70 Gy and 1.50 Gy, respectively. Regression analysis showed that for every 1 Gy increase in MHD the dose on the aorta on average increased by 1.19 Gy and the dose to the right atrium increased by 1.39 Gy. In left-sided radiotherapy the average MHD was 2.96 Gy and three out of nine structures received more than 2 Gy. The myocardium and LAD were the most exposed structures in left-sided radiotherapy with a mean dose of 4.87 Gy and 7.42 Gy, respectively. Regression analysis showed that for every 1 Gy increase in MHD the dose to the myocardium on average increased by 1.90 Gy and the dose to the LAD increased by 2.77 Gy.

The strongest correlation between planned dose to the structure and the MHD was found for the myocardium ($r = 0.93$) and the right ventricle ($r = 0.88$) in left-sided radiotherapy (Table 7.2). For left-sided radiotherapy, the proportion of the variance (R^2) in dose to the cardiac structure that could be explained by the MHD was particularly low for the left atrium (1%), right atrium (6%), pulmonary artery (2%) and LCX (1%) (Figure 7.4). For the myocardium and right ventricle the R^2 was 87% and 77%, respectively. In right-sided radiotherapy the correlation was strongest for the right ventricle ($r = 0.80$) and between 0.76-0.58 for other structures. The R^2 was lower than 70% for all evaluated anatomies, which indicates an inhomogeneous dose distribution within the heart, making the MHD not suitable to describe the dose in different structures.

Moreover, in patients with a planned MHD < 4 Gy, the radiation exposure of other structures could be substantially higher (Figure 7.5).

Figure 7.3: Histogram of the mean heart dose of all patients

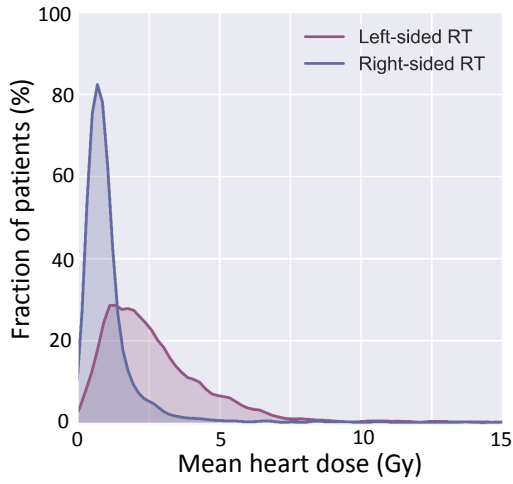


Table 7.1: Baseline characteristics

	Full cohort	Selection <2011
N	5300	1871
Age, mean (SD)	59.5 (10.8)	58.7 (10.7)
Follow-up, months, median (IQR)	53.3 (33.3-86.1)	96.0 (84.2-110.4)
Cases	265 (5.0)	135 (7.2)
Laterality, left (%)	2736 (51.6)	989 (52.9)
MHD, Gy, median (IQR)	1.34 (0.78-2.77)	1.75 (0.92-3.90)
Radiotherapy		
Local	1277 (24.1)	-
Local with boost	1395 (26.3)	-
Locoregional with or without boost	609 (11.5)	-
other	23 (0.4)	-
Unknown	1997 (37.7)	1871 (100.0) [†]
CAC		
No	3664 (69.3)	1208 (64.6)
Yes	1636 (30.7)	663 (35.4)
Pathological tumor stage, n (%)		
DCIS/T0	719 (13.6)	228 (12.2)
T1	3198 (60.3)	1161 (62.1)
T2	1118 (21.1)	420 (22.4)
≥T3	160 (3.0)	43 (2.3)
Unknown	106 (2.0)	19 (1.0)
Pathological node stage, n (%)		
N0	3662 (69.1)	1333 (71.2)
N1m	332 (6.3)	106 (5.7)
N1	808 (15.2)	271 (14.5)
N2	242 (4.6)	90 (4.8)
N3	121 (2.3)	50 (2.7)
Unknown	136 (2.6)	21 (1.1)
Chemotherapy		
None	3314 (62.5)	1196 (63.9)
With anthracyclines	1310 (24.7)	93 (5.0)
Without anthracyclines	456 (8.6)	376 (20.1)
Type of chemotherapy unknown	221 (4.2)	206 (11.0)
Hormonal therapy		
None	3972 (56.1)	1101 (58.8)
Aromatase inhibitor	129 (2.4)	19 (1.0)
Tamoxifen	1535 (29.0)	138 (7.4)
Other or unknown	665 (12.5)	613 (32.8)

[†] Type of radiotherapy was not recorded by NCI before 2011

Figure 7.4: Relation between the dose planned for different structures and the mean heart dose (MHD).

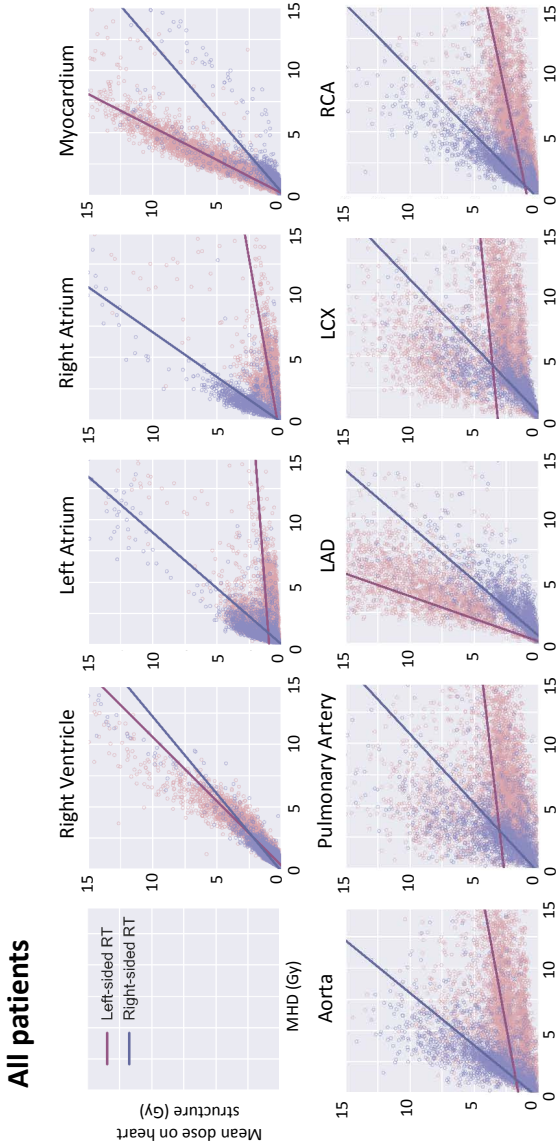


Figure 7.5: Histograms of dose per structure for patients with MHD < 4 Gy

Patients with MHD < 4 Gy

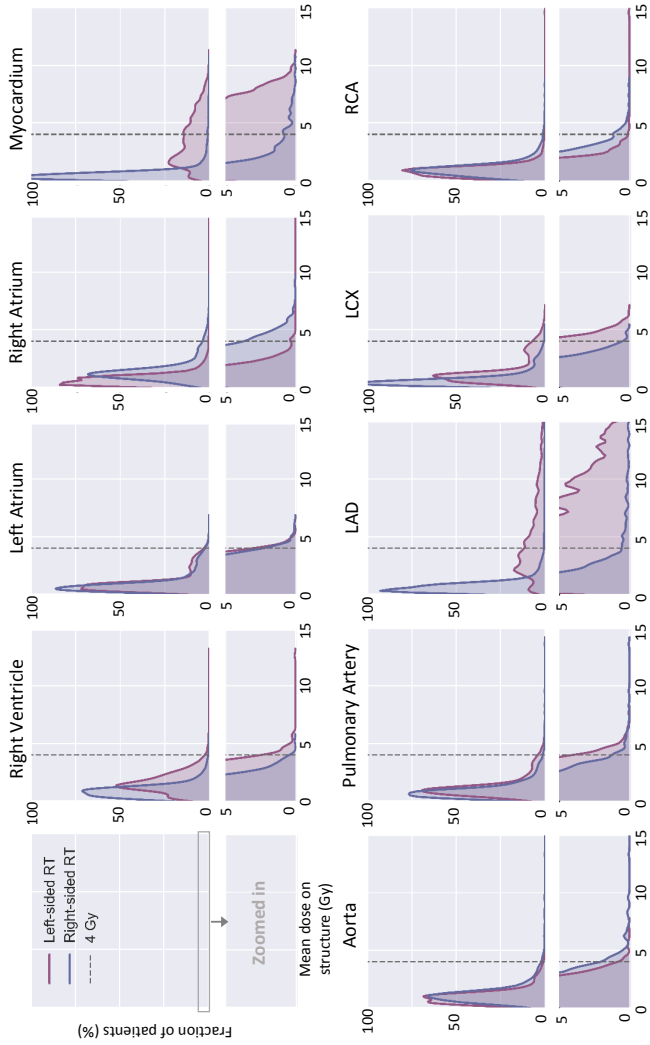


Table 7.2: Relation between mean dose on cardiac structures and MHD

	MHD	Right Ventricle	Left Atrium	Right Atrium	Myocardium	Aorta	Pulmonary Artery	LAD	LCX	RCA
Left										
Mean dose (Gy, SD)	2.96 (2.76)	2.74 (3.56)	1.35 (1.84)	0.98 (1.94)	4.87 (3.88)	1.40 (2.29)	1.71 (2.50)	7.42 (6.23)	1.73 (1.92)	1.14 (1.87)
β	-	0.97	0.07	0.17	1.90	0.18	0.11	2.77	0.09	0.21
α	-	-0.30	0.94	0.31	-0.37	0.63	1.08	-0.21	1.27	0.37
R^2	-	0.77	0.01	0.06	0.87	0.10	0.02	0.62	0.01	0.20
Pearsons' r	-	0.88	0.11	0.25	0.93	0.31	0.15	0.79	0.12	0.45
Right										
Mean dose (Gy, SD)	1.13 (1.47)	1.08 (1.57)	1.13 (1.63)	1.70 (2.01)	0.76 (1.58)	1.50 (2.10)	1.13 (1.76)	1.06 (2.33)	0.83 (1.45)	1.21 (1.76)
β	-	0.81	1.12	1.39	0.84	1.19	0.90	1.13	0.94	0.96
α	-	0.12	-0.05	0.23	-0.24	0.19	0.13	-0.29	-0.18	0.14
R^2	-	0.64	0.45	0.56	0.52	0.40	0.34	0.39	0.58	0.55
Pearsons' r	-	0.80	0.67	0.75	0.72	0.63	0.58	0.62	0.76	0.74

Linear regression of the form Dose-structure = $\alpha + \beta \cdot \text{MHD}$ was used. Patients with an $\text{MHD} > \text{MHD}_{95\text{percentile}}$ were not included for regression analysis, to decrease the influence of possible outliers. For left-sided radiotherapy the $\text{MHD}_{95\text{percentile}}$ was 6.8 Gy and for right-sided radiotherapy the $\text{MHD}_{95\text{percentile}}$ was 2.8 Gy. R^2 indicates the proportion of variance in the dose on the substructure that is predictable from MHD. The correlation between the dose on the structure and MHD is measured with Pearsons' correlation coefficient (r).

7.3.3 Radiotherapy dose and ischemic heart disease

After a median follow-up of 96.0 months (IQR 84.2-110.4) in the long follow-up group, 135 patients were hospitalized for IHD or died from IHD.

The incidence rate of IHD increased from 9.4 events per 1000 person-years in patients with an MHD <4 Gy to 16.0 in patients with MHD >8 Gy (Table 7.3). Likewise, for all other cardiac structures the incidence rate was higher in patients with a high dose to that respective structure than in patient with a low dose. For the left atrium, right atrium, aorta, pulmonary artery, RCA and LCX, the lowest incidence rate was found in the medium-dose group.

Adjusted Cox regression analysis showed a similar trend for all structures (Table 7.3 and 7.4): the high-dose group had an increased risk of IHD compared with the low-dose group. Patients that had a high dose planned to the myocardium or the left atrium had a statistically significantly higher risk of IHD (HR = 2.0, 95%CI 1.1-3.4 and HR = 2.1, 95%CI 1.1-4.1, respectively). No association was found for the medium-dose groups. When the dose was added as a continuous variable to the model instead of stratified into dose groups, a significant association was found for all cardiac chambers and large vessels except the right atrium. No association between dose and IHD was found in a subgroup of patients that received < 4 Gy MHD (Table 7.5).

For the LAD, LCX, and RCA, the high-dose group had a respectively 2.0 (95%CI 1.2-3.6), 3.0 (95%CI 1.7-5.5), and 2.0 (95%CI 1.1-3.7) higher risk of IHD compared with the low-dose group (Table 7.4). No association was found for the medium-dose groups. The association between continuous dose and IHD was significant for all three coronary arteries; for every 1 Gy increase on the LAD, LCX and RCA, the risk of IHD increased by 3% (95%CI 1%-5%), 11% (95%CI 4%-19%), and 8% (95%CI 1%-15%), respectively.

Table 7.3: Number of events, incidence rates and risk of heart disease by planned dose on different structures

Long follow up selection	All n=1871			
	N (%)	Events (%)	Rate	HR (95% CI)
MHD				
Low (0-4 Gy)	1424 (76)	99 (7.0)	9.4	1.00
Med (4-8 Gy)	359 (19)	28 (7.8)	10.6	1.15 (0.76-1.75)
High (>8 Gy)	88 (5)	8 (9.1)	16.0	2.29 (1.11-4.72)
Continuous				1.08 (1.03-1.14)
Myocardium				
Low (0-7 Gy)	1377 (74)	93 (6.8)	9.1	1.00
Med (7-11 Gy)	348 (19)	27 (7.8)	10.3	1.12 (0.73-1.72)
High (>11 Gy)	146 (8)	15 (10.3)	16.3	1.89 (1.10-3.27)
Continuous				1.04 (1.01-1.08)
RV				
Low (0-3 Gy)	1357 (73)	96 (7.1)	9.6	1.00
Med (3-6 Gy)	335 (18)	26 (7.8)	10.3	1.17 (0.76-1.81)
High (>6 Gy)	179 (10)	13 (7.3)	11.1	1.35 (0.76-2.42)
Continuous				1.06 (1.02-1.11)
LA				
Low (0-1 Gy)	1394 (75)	103 (7.4)	9.8	1.00
Med (1-2 Gy)	350 (19)	21 (6.0)	8.4	1.11 (0.69-1.78)
High (>2 Gy)	127 (7)	11 (8.7)	14.9	2.22 (1.18-4.16)
Continuous				1.10 (1.03-1.18)
RA				
Low (0-1.5 Gy)	1515 (81)	118 (7.8)	10.5	1.00
Med (1.5-2 Gy)	201 (11)	5 (2.5)	3.2	0.45 (0.18-1.10)
High (>2 Gy)	155 (8)	12 (7.7)	13.0	1.76 (0.97-3.20)
Continuous				1.06 (0.99-1.14)
AO				
Low (0-2 Gy)	1483 (79)	112 (7.6)	10.1	1.00
Med (2-3 Gy)	267 (14)	15 (5.6)	7.8	0.97 (0.56-1.67)
High (>3 Gy)	121 (7)	8 (6.6)	11.3	1.50 (0.73-3.09)
Continuous				1.07 (1.01-1.13)
PuIA				
Low (0-2 Gy)	1570 (84)	114 (7.3)	9.7	1.00
Med (2-4 Gy)	199 (11)	13 (6.5)	9.6	1.27 (0.71-2.25)
High (>4 Gy)	102 (6)	8 (7.8)	13.8	1.84 (0.90-3.79)
Continuous				1.09 (1.02-1.15)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio;
 Rate = incidence rate per 1000 person-years
 Models are adjusted for age at time of CT scan and anthracyclines (yes/no).

continued...

CAC n=663				No CAC n=1208			
N (%)	Events (%)	Rate	HR (95% CI)	N (%)	Events (%)	Rate	HR (95% CI)
527 (80)	54 (10.2)	14.2	1.00	897 (74)	45 (5.0)	6.7	1.00
121 (18)	12 (9.9)	13.9	0.98 (0.52-1.83)	238 (20)	16 (6.7)	8.9	1.34 (0.76-2.38)
15 (2)	3 (20.0)	33.3	2.12 (0.66-6.81)	73 (6)	5 (6.8)	12.2	2.63 (1.03-6.71)
			1.05 (0.96-1.14)				1.11 (1.04-1.18)
498 (75)	51 (10.2)	14.2	1.00	879 (73)	42 (4.8)	6.4	1.00
131 (19)	13 (9.9)	13.6	0.92 (0.50-1.69)	217 (18)	14 (6.5)	8.4	1.39 (0.76-2.55)
34 (5)	5 (14.7)	23.3	1.37 (0.54-3.45)	112 (9)	10 (8.9)	14.2	2.57 (1.29-5.14)
			1.01 (0.96-1.08)				1.07(1.02-1.12)
513 (77)	50 (9.7)	13.6	1.00	844 (70)	46 (5.5)	7.2	1.00
105 (16)	14 (13.3)	17.3	1.26 (0.70-2.28)	230 (19)	12 (5.2)	7.0	1.11 (0.59-2.10)
45 (7)	5 (11.1)	17.2	1.20 (0.48-3.01)	134 (11)	8 (6.0)	9.0	1.52 (0.72-3.13)
			1.04 (0.96-1.12)				1.08 (1.03-1.13)
542 (82)	55 (10.1)	13.9	1.00	852 (71)	48 (5.6)	7.3	1.00
89 (13)	10 (11.2)	16.3	1.21 (0.61-2.37)	261 (22)	11 (4.2)	5.9	1.08 (0.56-2.09)
32 (5)	4 (12.5)	21.0	2.02 (0.73-5.63)	95 (8)	7 (7.4)	12.8	2.35 (1.05-5.23)
			1.11 (0.99-1.25)				1.10 (1.01-1.20)
580 (88)	63 (10.9)	15.1	1.00	935 (77)	55 (5.9)	7.7	1.00
44 (7)	0 (0.0)	0.0	-	157 (13)	5 (3.2)	4.3	0.73 (0.29-1.84)
39 (6)	6 (15.4)	15.4	2.29 (0.98-5.36)	116 (10)	6 (5.2)	5.2	1.45 (0.62-3.38)
			1.05 (0.93-1.19)				1.07 (0.98-1.16)
565 (85)	57 (10.1)	13.9	1.00	918 (76)	55 (6.0)	7.9	1.00
67 (10)	9 (13.4)	19.2	1.54 (0.76-3.12)	200 (17)	6 (3.0)	4.1	0.63 (0.27-1.46)
31 (5)	3 (9.7)	15.7	1.36 (0.43-4.38)	90 (8)	5 (5.6)	9.6	1.59 (0.63-4.00)
			1.07 (0.97-1.18)				1.07 (1.00-1.15)
582 (88)	61 (10.5)	14.3	1.00	988 (82)	53 (5.4)	7.0	1.00
55 (8)	4 (7.3)	11.4	0.86 (0.31-2.37)	144 (12)	9 (6.3)	8.9	1.63 (0.80-3.33)
26 (4)	4 (15.4)	27.6	1.98 (0.72-5.47)	76 (6)	4 (5.3)	9.2	1.80 (0.65-5.02)
			1.09 (0.98-1.20)				1.09 (1.01-1.17)

Table 7.4: Number of events, incidence rates and risk of heart disease by planned dose on the three main coronary arteries

Long follow-up selection	All n=1871			CAC n=663			No CAC n=1208			
	N (%)	Events (%)	Rate	N (%)	Events (%)	Rate	N (%)	Events (%)	Rate	HR (95% CI)
LAD										
Low (0-10 Gy)	1377 (74)	94 (6.8)	9.2	496 (74)	51 (10.3)	14.1	881 (73)	43 (4.9)	6.5	1.00
Med (10-20 Gy)	378 (20)	27 (7.1)	9.7	129 (20)	11 (8.5)	12.0	249 (21)	16 (6.4)	8.5	1.34 (0.76-2.39)
High (>20 Gy)	116 (6)	14 (12.1)	20.3	38 (6)	7 (18.4)	30.4	78 (7)	7 (9.0)	15.2	2.37 (1.07-5.28)
Continuous			1.03 (1.01-1.05)			1.01 (0.98-1.04)				1.04 (1.01-1.08)
RCA										
Low (0-2 Gy)	1571 (84)	116 (7.4)	9.9	588 (89)	59 (10.0)	13.9	983 (81)	57 (5.8)	7.6	1.00
Med (2-3 Gy)	172 (9)	8 (4.7)	6.4	46 (7)	6 (13.0)	18.1	126 (10)	2 (1.6)	2.2	0.39 (0.10-1.61)
High (>3 Gy)	128 (7)	11 (8.6)	14.4	29 (4)	4 (13.8)	21.5	99 (8)	7 (7.1)	12.1	2.07 (0.94-4.57)
Continuous			1.08 (1.01-1.15)			1.05 (0.94-1.19)				1.09 (1.01-1.19)
LCX										
Low (0-2 Gy)	1612 (86)	115 (7.1)	9.5	593 (89)	61 (10.3)	14.1	1019 (84)	54 (5.3)	7.0	1.00
Med (2-4 Gy)	159 (9)	8 (5.0)	7.1	45 (7)	3 (6.7)	9.9	114 (9)	5 (4.4)	6.1	1.34 (0.53-3.39)
High (>4 Gy)	100 (5)	12 (12.0)	21.8	25 (4)	5 (20.0)	33.9	75 (6)	7 (9.3)	17.3	3.26 (1.47-7.24)
Continuous			1.11 (1.04-1.19)			1.12 (1.02-1.25)				1.11 (1.02-1.21)

CAC = coronary artery calcium; CI = confidence interval; HR = hazard ratio; Rate = incidence rate per 1000 person-years
 Models are adjusted for age at time of CT scan and anthracyclines (yes/no).

7.3.4 CAC vs no CAC

The association between planned radiation dose to any cardiac structure and IHD was the same for patients with and without CAC, i.e. the interaction between planned radiation dose and CAC was not significant. A higher dose is associated with IHD in both patients with CAC and without CAC. However, incidence rates of IHD events were higher in the group with CAC than in the group without CAC (Table 7.3 and 7.4). While the increase in relative risk due to dose may be similar in patients with and without CAC, the absolute increase in event rate is larger in patients with CAC.

Table 7.5: Risk of heart disease in patients with MHD < 4 Gy by planned dose on cardiac structures.

Patients[†] with MHD <4Gy	
N=1424	
Model with mean dose continuous	HR (95% CI)
MHD	1.04 (0.87-1.26)
Cardiac chambers and large vessels	
Myocardium	1.03 (0.96-1.11)
RV	1.05 (0.83-1.33)
LA	1.24 (0.85-1.81)
RA	0.77 (0.49-1.21)
AO	0.88 (0.55-1.41)
PulA	1.13 (0.72-1.77)
Coronary arteries	
LAD	1.02 (0.98-1.07)
RCA	0.80 (0.45-1.42)
LCX	1.36 (1.01-1.83)

[†] Long follow-up population

7.4 Discussion

The majority of breast cancer patients are treated with radiotherapy and it is well known that radiation exposure of the heart increases the risk of future heart disease.^{6,19} Our main finding is that a higher radiation dose to all individual cardiac structures except the right atrium is associated with an increased risk of IHD events. Even though the relative increase in risk due to radiotherapy is similar for patients with and without CAC, the absolute increase in risk is substantially higher in patients with CAC. Therefore, automatic segmentation of the cardiac structures could be especially helpful in patients with CAC for adjustment of radiotherapy treatment planning to minimize radiation-induced cardiac damage.

Previous research by Jacob et al.¹⁵⁰ indicated that the MHD is not an adequate surrogate parameter for dose on the myocardium and coronary arteries. In agreement with this, our results show low correlations and R^2 between planned radiation dose for different cardiac structures and MHD. Moreover, similar to our results, Jacob et

al. found an increase of 1.9 Gy on the left ventricle (incl. myocardium) for each 1 Gy increase in MHD in left-sided radiotherapy. For the LAD, an increase of 3.7 Gy per 1 Gy increase in MHD was reported, which is higher than our finding (2.8 Gy per 1 Gy increase MHD). This difference in findings could be due to the small sample size (n=89) in the study of Jacob et al. and uncertainty in the dose to the LAD because the LAD is a relatively small structure.

For every 1 Gy increase in MHD, we found an 8% increase in the risk of IHD. This is in line with previous research by Darby et al.,¹⁹ who reported a 7% (95% CI 3-15) increase in the risk of acute coronary events for every Gy increase MHD, and with Jacobse et al.¹⁶⁵ who reported a 6% (95% CI 1-16) risk increase per Gy. Studies by van den Bogaard et al.¹⁵¹ and Lorenzen et al.¹⁶⁶ found a stronger increase in risk of 17% (95% CI 1-35) per Gy and 19% (95% CI 1-63) per Gy. These studies, as well as in ours, report broad confidence intervals, which indicates that uncertainty in the data may explain the difference in results.

In this study, no statistically significant increase of IHD risk was found between the low and medium dose groups for any structure, including MHD. Similarly, a prediction model designed by van den Bogaard et al.¹⁵¹ predicted a very minimal increase in risk (<6%) of acute coronary events for up to 8 Gy MHD for patients with similar characteristics as the cohort used in this paper. The minimal increase in risk for radiation exposure below 8 Gy MHD might explain why we only found an effect in the high dose groups of cardiac structures. Additionally, this might explain why no association between dose and IHD was found in a subgroup of patients that received < 4 Gy MHD. The limited number of events in our cohort has likely contributed to this, as well.

Our results show that the group with a medium planned dose for the left atrium, right atrium, aorta, LAD, LCX, and RCA had a lower incidence rate than the group with a low dose planned for these structures. This might be explained by the dose distribution in these structures. Figure 7.4 shows that the low dose group for these structures contains patients that received left-sided radiotherapy with a relatively high MHD. For instance, a substantial number of patients with a planned dose to the left atrium of <1 Gy receive >4 Gy on the whole heart. The medium-dose group of these structures does not contain patients with a relatively high MHD to the same extent, which might explain the lower incidence rates.

Our results show that a higher dose to cardiac structures is significantly associated with (non-)fatal IHD. The association between planned radiation dose and IHD was the same for patients with- and without CAC. Nevertheless, the absolute risk of IHD was substantially higher in patients with CAC than in patients without CAC. Moreover, analysis of the dose to different cardiac structures has shown that there is a group of patients that, despite having a low MHD, receives a relatively high dose on several structures. Because the MHD does not accurately represent the dose delivered to the different evaluated anatomies, a more spatially localized dose calculation might

be especially valuable for treatment planning in patients with CAC. In case it is not possible to lower radiation exposure of the cardiac structures to a desirable level without compromising tumor coverage, other treatment options might be considered e.g. more accurate irradiation using protons¹⁶⁷ or opting for treatment without radiation. Moreover, the increased risk of IHD could be normalized in these patients by referring to a cardiologist for screening, lifestyle intervention or cardio protective medication.

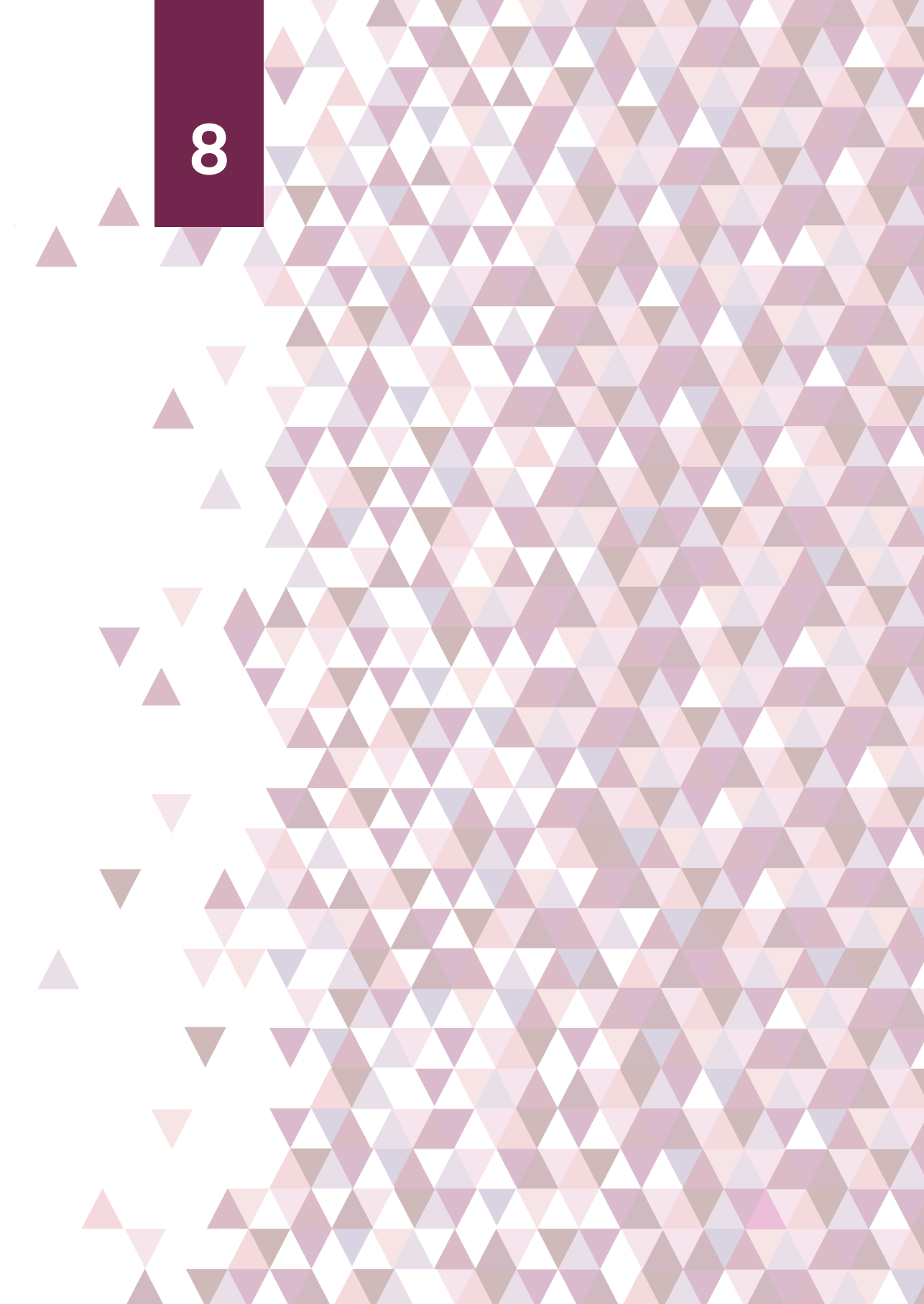
The amount of CAC quantified on CT scans and the commonly used corresponding CAC risk categories are strongly associated with IHD.¹⁶ In this study the number of events did not allow for subgroup analysis per CAC risk category and, therefore, the presence of CAC was used for stratification. Investigating the radiotherapy-related IHD risk per CAC risk category would be an interesting topic for future work with a larger data set.

The risk of IHD was different for different structures. This suggests that dose planning using specific constraints for each cardiac structure might be beneficial for mitigating radiation-related risk. Unfortunately, we were unable to clearly separate the effect of irradiation for different cardiac structures from the effects of MHD due to the co-linearity of the MHD and the dose to the respective structures. We found a higher increase of risk per Gy for structures with a smaller range in mean dose. This needs to be interpreted with caution, because of co-linearity between dose to different structures. For instance, in left-sided radiotherapy, a 1 Gy increase in the MHD means an average increase of 0.1 Gy on the left atrium and an increase of 1.9 Gy on the myocardium. In line with this, 1 Gy increase in dose to the left atrium gives a 10% increase in the risk of IHD (HR=1.10 95%CI 1.03-1.18), while 1 Gy increase to the myocardium leads to a 4% increase (HR=1.04 95%CI 1.01-1.08). A population-based case-control study as performed by Darby et al.¹⁹ might give more insight into separate risks per cardiac substructure.

In this study, we fully automatically determined the location of the heart chambers, large arteries and coronary arteries in >5000 radiotherapy planning CT scans. For this, we used an ensemble of CNNs that was trained with virtual non-contrast CT images for which reference segmentations were obtained in perfectly aligned CCTA images, similarly to Bruns et al.¹⁵⁸ This approach enabled us to train the method with CT exams from another set of patients. Allowing the use of different data sets is advantageous in comparison with existing methods^{150,168} that require patients to have both a radiotherapy planning CT and a CCTA scan to segment cardiac structures through intra-patient registration. A different approach performs automatic atlas-based contouring of cardiac structures directly on the planning CT, without using a CCTA scan for contouring in between.^{155,156,169} However, cardiac structures are barely differentiable on non-contrast-enhanced CT, making the atlas subject to inter-observer variability. Also, inter-patient registration may be difficult, because structure boundaries are poorly visible. Moreover, due to the lack of contrast, the performance in small structures like the coronary arteries is often suboptimal.^{155,156,170}

A limitation of this study is the inclusion of only patients who received radiotherapy. Hence, we could not compare the risk of IHD at certain radiation exposure with patients who did not receive radiotherapy. Second, the registry of hospital admissions used for defining events only recorded IHD events that required hospitalization with overnight stay. Therefore, mild IHD events that did not require hospitalization were not included, which might have caused underestimation of the effect of dose on IHD risk. Nevertheless, our findings provide clear data on a hard clinical endpoint during long term follow up. Third, the exact accuracy of the automatic segmentation in planning CT is not evaluated, because the lack of contrast-enhanced CT scans in our population makes this evaluation infeasible. Nevertheless, previous work that used different data showed accurate results.¹⁵⁸ Fourth, during the follow-up time of this study, clinical protocols for radiotherapy changed, e.g. the breath-hold technique was implemented and different treatment planning software was used. Because we calculated dose parameters per patient, we implicitly correct for inconsistencies in the data and we expect the influence to be marginal. Fifth, information on cardiac risk factor like smoking, diabetes or IHD history was not available for analysis. Finally, due to cardiac motion and imperfect alignment of planning CT to the patient position during irradiation, the planned dose may deviate from the actually delivered dose.¹⁷¹ This may be especially pronounced for smaller structures such as the coronary arteries. Future work could investigate the extent and influence of these deviations.

To conclude, we showed that for all cardiac structures a higher dose leads to an increased risk of IHD. Because the absolute increase in IHD event risk due to radiotherapy is larger in patients with CAC than without CAC, precise dose planning might be especially important in patients with CAC. Since cardiac structures are not manually delineated in clinical practice, the automatic segmentation can offer a solution for obtaining dose estimates for each cardiac chamber, large artery and the coronary artery. This study presents a step towards more detailed and spatially distributed dose planning and management of radiotherapy-induced adverse effects on the heart in breast cancer patients.



CHAPTER 8

Summary and discussion

In the clinical workflow for irradiation for breast cancer, patients undergo a CT scan for planning of the radiotherapy treatment. These CT scans provide an excellent opportunity for automatic identification of breast cancer patients at increased risk of CVD. The previous chapters describe and evaluate automatic methods for CVD risk prediction in CT scans. Moreover, the influence of radiotherapy on the risk of CVD in breast cancer patients is evaluated. This chapter summarizes the methods and findings presented in these chapters and discusses implications and possible directions for future research.

8.1 Summary

The amount of coronary artery calcification (CAC) and the amount of thoracic aorta calcification (TAC) are strong predictors of future cardiovascular disease events like heart attacks and stroke. CAC is usually quantified in cardiac CT scans. Given the clinical relevance, quantification of CAC is increasingly performed in other types of CT scans visualizing the heart and aorta, including radiotherapy treatment planning CTs.

Owing to the clinical relevance, several machine learning methods have been developed to automate calcium scoring. Typically, these methods are developed and evaluated in CT scans with specific acquisition characteristics. Consequently, the performance is often suboptimal in CT scans acquired with different acquisition parameters, which limits the clinical applicability. In our group an automatic deep learning calcium scoring method was developed for application in low-dose lung screening CT scans.³² Training was originally done with a diverse set of lung screening CT scans that originated from 31 different hospitals. However, this does not guarantee good performance in other types of CT that allow calcium scoring, because of differences in e.g. field of view, noise levels and population. In **CHAPTER 2** we evaluated the performance of the calcium scoring method. The un-adapted method, i.e. as trained for low-dose chest CT from lung cancer screening, showed a good performance in a large set of CT scans with diverse acquisition protocols. In general, supervised methods perform best when they are trained on data from the same distribution as the target data. However, collecting annotations for each new type of CT scan would require extensive manual labor from medical experts. To limit the need for new training data, we investigated retraining where the original training dataset was utilized. When a relatively low number of scans with a specific scan protocol was added to the training set for protocol-specific training, the performance of the method in that type of CT was close or equal to the level of human observers. Moreover, for implementation in the clinic, ideally a single method would be applicable to all available CT scans. Training with a combination of all available types of CT scans lead to a performance comparable to protocol-specific training. This type of generic method eliminates the need for a specialized network for every type of CT that shows the heart, which is practical for wide application in clinical practice.

The automatic calcium scoring method described in Chapter 2 was used in **CHAPTER 3** to investigate the difference in prevalence and amount of CAC and TAC between Western and Asian breast cancer patients. The prevalence of CAC and TAC in both patient populations is considerable and increases with age. Up to one third of breast cancer patients aged under 70 years has CAC. More than half of patients aged over 70 years has CAC and the prevalence in Asian patients is higher than the Netherlands. TAC is more prevalent than CAC in both Asian and Western breast cancer patients. Moreover, we showed that, despite differences in scanning parameters, the automatic scoring method trained with a protocol-specific training set is a reliable method to measure CAC and TAC on radiotherapy planning CT scans of both populations.

Although previous research suggests that there is a relation between CAC quantified on planning CT scans and CVD, the exact association between the planning CT calcium score and CVD was still unknown. To investigate this, we evaluated the association between CAC quantified on radiotherapy planning CT scans and risk of CVD events in breast cancer patients and the interaction with certain breast cancer treatments in **CHAPTER 4**. Automatic calcium scoring (Chapter 2) enabled evaluation in a large cohort of 15,915 breast cancer patients. Based on the scores, patients were classified into five commonly used CVD risk categories, ranging from 'low risk' for patients with a zero score to 'very high risk' for patients with a score >400. Survival analysis was used to evaluate the association between risk categories and CVD. A strong association was found between CAC quantified on radiotherapy planning CT scans and CVD risk. For patients that were treated with anthracyclines, a type of chemotherapy with cardiotoxic side effects, the association was even stronger. These results show that, because of the strong association, automatic calcium scoring provides a solution for identification of breast cancer patients at increased CVD risk.

Given the clinical relevance, calcium scoring is increasingly performed in CT scans made for other purposes, including non-ECG synchronized scans such as radiotherapy treatment planning CTs. However the interscan reproducibility of calcium scores is compromised in CT scans without ECG-synchronization and with low image resolution. This means that if two scans of the same patient are made within a short time interval, there can be a large difference in calcium scores between these scans. Besides image resolution and lack of ECG-triggering, a fixed intensity level threshold used for segmentation of calcium lesions causes low scoring reproducibility. To address this, **CHAPTER 5** describes a method for CAC quantification that does not require a segmentation threshold. The method uses a CycleGAN to decompose an image with CAC into an image without CAC and a CAC-map, i.e. an image showing only CAC. Subsequently, the CAC-map is used to quantify the amount of calcium. The method was evaluated in radiotherapy treatment planning CT scans of almost 1,700 breast cancer patients that each had two CT scans. This set with scan pairs made it possible to assess the interscan reproducibility of the method. Results showed that the method was able to accurately detect calcium lesions, including those below the clinical threshold.

Moreover, our method achieved a higher interscan reproducibility, compared with the clinical calcium scoring procedure.

Besides CAC and other known image biomarkers, more information is present in CT images that is possibly predictive for CVD events. Therefore, **CHAPTER 6** describes an automatic method for the prediction of CVD risk that does not need hand-crafted image features such as calcium scores. Instead, 5-year CVD mortality is directly predicted from the image itself. This method uses an unsupervised learning strategy to extract image features without designing them. First, the region of interest was defined by automatically localizing the heart and cropping the image to a bounding box around the heart. Next, an autoencoder was used to compress the identified heart region in an unsupervised manner into a lower dimensional space. The loss function of the autoencoder is defined to capture perceptual differences and spatial correlations in the heart image. Finally, the image encodings were used to classify subjects into survivors or non-survivors with a support vector machine classifier. The performance of the method was assessed using cross-validation experiments and the method achieved a performance comparable with other approaches that use calcium scores and patient data. We showed that predicting CVD mortality directly from CT scans, without hand-crafting image features, might be feasible.

Finally, in **CHAPTER 7** we investigated the effect of radiation exposure of the cardiac structures and coronary arteries on the risk of ischemic heart disease (IHD) in breast cancer patients. Additionally, we investigated whether the association between radiation dose and IHD is different for patients with and without CAC. Cardiac structures and coronary arteries are not clearly distinguishable in planning CT. Therefore, we used contrast-enhanced images acquired on a dual-layer detector CT scanner to create manual reference segmentations of cardiac structures and coronary arteries. Using the dual-energy properties of this acquisition a second, perfectly aligned virtual non-contrast image (VNC) can be reconstructed that resembles a non-contrast image. Because of voxel-wise correspondence, the manual reference annotations can be transferred to the VNC image, which we used to train an ensemble of CNNs for segmentation of cardiac structures (cardiac chambers, large arteries and coronary arteries). Thereafter, the method was applied to the radiotherapy planning CT scans and the segmentations were used to determine the radiation exposure per structure. Subsequently, survival analysis was used to evaluate the association between dose and IHD in breast cancer patients with and without CAC. The results showed that a high dose to the whole heart is associated with an increased risk of CVD. This is also true for all cardiac chambers, large arteries and coronary arteries. Moreover, we found that the dose on the whole heart is not an accurate surrogate parameter for the dose on the different structures. The relative increase in risk due to radiation exposure is the same in patients with and patients without CAC. However, the radiation-related absolute increase in IHD risk is substantially higher in patients with CAC than in patients without CAC. Therefore, spatially distributed dose calculations taking cardiac structures and coronary arteries

into consideration may be especially helpful in patients with CAC.

8.2 Discussion

Recent advances in machine learning and artificial intelligence research have led to a multitude of automatic methods that are able to perform medical image analysis tasks with human-level performance. Typically, these methods are developed for a relatively uniform set of input images where the variability with respect to pathology and acquisition is limited. However, for application in clinical routine, software tools need to be robust to variation in image acquisition parameters and variation in the imaged population. The methods described in this thesis are mostly supervised methods, and, therefore, their performance and robustness to variations in input data is largely dependent on the training data. For example, in Chapter 2 we have evaluated the performance of the automatic calcium scoring method in CT scans with multiple different acquisition protocols and populations. The performance of the calcium scoring method varied over the different types of CT scans. For instance, the methods showed nearly perfect agreement with experts in cardiac CT, while the agreement was lower in radiotherapy treatment planning CT and diagnostic chest CT. This is related to the supervised training of the method and the quality of the annotations in the training data. In challenging tasks, manual annotations may be subject to intra- and interobserver variability, causing noise in the labels that affects the performance of the method. This explains why the performance was excellent in cardiac CT, in which calcium scoring is considered relatively easy, because the image acquisition parameters are specifically optimized for visualizing coronary calcium. In radiotherapy treatment planning CT and diagnostic chest CT, where calcium scoring is more challenging, the interobserver variability was higher and likewise the performance was worse. Cardiac events or mortality, like the outcome labels used in Chapter 6, may be more reliable labels. These labels are defined according to strict clinical protocol and definitions,¹⁶⁴ which may make them less subject to interobserver variability. However, this type of data is often obtained in retrospective settings and difficult to obtain prospectively as long follow-up may be needed. In situations where noisy labels are inevitable, application of methods for decreasing the impact of noise in the reference during training could be considered.¹⁷²

Related to the dependency of supervised methods on training data, one of the main challenges in the field of medical image analysis is collecting data for development of automatic methods. More specifically, labels for supervised training are typically difficult to obtain since they generally require a substantial amount of manual input of a medical expert. For development of the methods described in this thesis, different types of labels are used for different levels of supervision during training, each with their own advantages and challenges. First, we used voxel-level supervision with manual annotations per voxel to train the calcium scoring method (Chapter 2) and the

cardiac segmentation method (Chapter 7). This manner of annotating is typically labor intensive, because an observer has to indicate a label class per voxel. However, by training on image patches or sub-images, a large amount of examples can be generated from a limited amount of images for training complex CNNs with a large number of parameters. Moreover, supervised methods often lead to the best results. Second, we used a weaker supervision with labels per image slice to train the CycleGAN for calcium segmentation in Chapter 5. This manner of annotating images is less labor intensive than voxel-painting. However, per image only a limited number of examples can be generated for training. Moreover, because the labels are sparse compared to the information of thousands of voxels in the image, training a method with good performance may be challenging. For training a complex CNN with this type of labels, a large number of scans may need to be annotated. Because the labels were defined per image slice, the calcium segmentation CycleGAN we used had a 2D architecture. A downside of this approach is that the method is not able to utilize all information from the full 3D structures that are visualized in the image. Third, for training the method for direct prediction of 5-year mortality described in Chapter 6 we used mortality outcome labels that are defined per image volume. Typically, such an approach would require a large amount of annotated scans, since one image volume only represents one training example. Because the amount of training data was limited, training the prediction method end-to-end was not feasible. Hence, we trained the method in two stages: a complex 3D CNN was trained in an unsupervised manner for feature extraction. The extracted features were substantially reduced in dimensionality, allowing for training a classifier with outcome labels in a supervised manner. A limitation of this approach is the fact that the unsupervised training is agnostic to the subsequent task, and therefore may not learn to extract all information that is relevant for prediction.

The research described in this thesis is largely focused on breast cancer patients. In addition, several other sets of CT scans of different populations are used. While often training on data similar to the target data leads to good results, in some cases inclusion of non-target data can be advantageous for achieving robustness of methods. For instance, by combining different sets comprising diverse CT protocols and different populations, we were able to create a generic calcium scoring method (Chapter 2). The combined training led to a performance that was comparable or better than training that was tuned to one specific type of CT scan. Subsequently, we used this generic method to perform calcium scoring in a large multi-center cohort of breast cancer patients. To develop the direct CVD prediction method (Chapter 6) we used scans from lung cancer screening participants, of which outcome data was available. Moreover, for development of the cardiac segmentation method in Chapter 6, training with non-target data was an explicit advantage. Due to lack of contrast in the planning CT images (target), defining the reference in planning CT images was not feasible. Because the non-target training data was acquired on a dual-layer detector CT scanner, an accurate segmentation reference could be determined on the contrast-enhanced image,

and transferred to the virtual non-contrast image. This way we were able to create a method for radiotherapy dose estimation per cardiac structure in planning CT scans.

In this thesis two methods for automatic quantification of CAC are described. Chapter 2 describes the evaluation of a deep learning method that follows the clinical definition of CAC. Using clinical calcium (Agatston) scores derived from radiotherapy treatment planning CT a CVD risk profile can be determined for breast cancer patients (Chapter 4). In Chapter 6 we describe a different method for CAC quantification in non-ECG-synchronized scans that is able to more accurately segment lesions with motion artefacts and quantify CAC with higher interscan reproducibility. However, because this method does not use the clinical definition of CAC, the quantified amount does not correspond to clinical calcium scores. Therefore, further research is needed to determine the association with CVD events and evaluate the predictive value of the quantified amount of CAC using this method.

The calcium scoring methods described in this thesis are developed for non contrast chest CT scans and cardiac CT scans. These types of CT scans make up roughly 14% of all CT scans that are acquired annually.¹⁷³ Several other CT scans, made for a diversity of clinical indications, visualize calcifications in large arteries, e.g. the aorta (abdomen CT), carotid arteries (head and neck CT) or all large arteries (full body CT). Future research could extend automatic calcium scoring methods to this broad range of scans to investigate opportunities for CVD risk screening in these scans.

In Chapter 6 we presented a method for mortality prediction directly from low-dose chest CT scans of lung screening participants. This method may be promising for application in other screening or patient populations, e.g. breast cancer patients. Although we show that direct prediction of CVD mortality is feasible, the method is not accurate enough for application on individual patient level at this time. Future research could focus on improving the performance, for instance by combining the unsupervised image features with established image biomarkers and non-image patient data.¹⁷⁴ Moreover, advanced augmentation techniques like GAN-based augmentation¹⁷⁵ may make end-to-end training with a limited dataset feasible.

For acceptance and application in clinic, the explainability of a deep learning method is of key importance. This means that an expert has insight in the decision making process of the method. The methods described in Chapters 2, 4 and 7 all produce a segmentation map, i.e. an image indicating per voxel whether it belongs to a target structure or tissue. Subsequently, the segmentation map is used for further quantification. Although the mechanism for prediction of the segmentation map itself is hidden in the complex network, a clinician can view the segmentation map. Based on inspection of the segmentation map, a clinician can judge the plausibility and accuracy of the quantified parameters. Although this is not an explicit explanation of how the segmentation was produced, the segmentation map provides interpretability of the subsequently quantified parameters. On the other hand, the method described in Chapter 6 only produces an output probability per class and, as such, acts to a greater

extent as a black box. In this method, visualization techniques for CNNs^{176–178} could be used to visualize which parts of the image contributed to the decision.

In Chapter 7 we showed that in breast cancer patients the absolute radiotherapy-related risk increase of IHD is largest in patients with CAC. Therefore, spatially distributed dose estimates may be most helpful in radiotherapy planning of these patients. Using the automatic segmentation method described in this chapter, it is possible to obtain estimates of planned radiotherapy dose to cardiac chambers, large vessels and coronary arteries. However, it is yet unknown whether implementation of radiation dose estimates per cardiac structure will benefit prevention of ischemic heart disease in breast cancer patients. Therefore, a next step would be conducting a clinical study that investigates whether structure- and coronary artery-specific dose estimates lead to a lower radiation dose administered to the cardiac structures, while maintaining effective tumor coverage. Implementation followed by evaluation in a prospective setting could thereafter determine the long-term impact on IHD risk.

We have shown that accurate quantification of CAC in planning CT is possible using an automatic calcium scoring method. Furthermore, we have shown that the amount of CAC automatically quantified on planning CT is strongly associated with risk of CVD events in breast cancer patients. Therefore, automatic calcium scoring in planning CTs provides an excellent way to routinely identify breast cancer patients with an increased risk of CVD. However, it is still unknown whether implementation of automatic calcium screening, followed by targeted prevention and management of CVD risk will lower the CVD burden in breast cancer survivors. Furthermore, potential disadvantages of screening, such as overdiagnosis, overtreatment and patient anxiety, need to be taken into consideration. A prospective clinical trial evaluating the value of calcium screening in breast cancer patients could bring clarification in this matter.

The artificial-intelligence-based methods presented in this thesis allow for quantification of disease- or treatment-related parameters in breast cancer patients that would otherwise be cumbersome or even infeasible to quantify. These parameters may aid medical experts in lifting the practice of personalized medicine to a new level, by more precisely measuring a broad set of CVD parameters and monitoring treatment consequences. After careful evaluation of clinical implications, the methods and evaluations in this thesis may pave the way towards more personalized treatment of breast cancer patients by enabling identification of patients at risk of CVD and by providing tools for estimating treatment induced CVD risk.

Nederlandse samenvatting

Hart- en vaatziekten bij borstkankerpatiënten

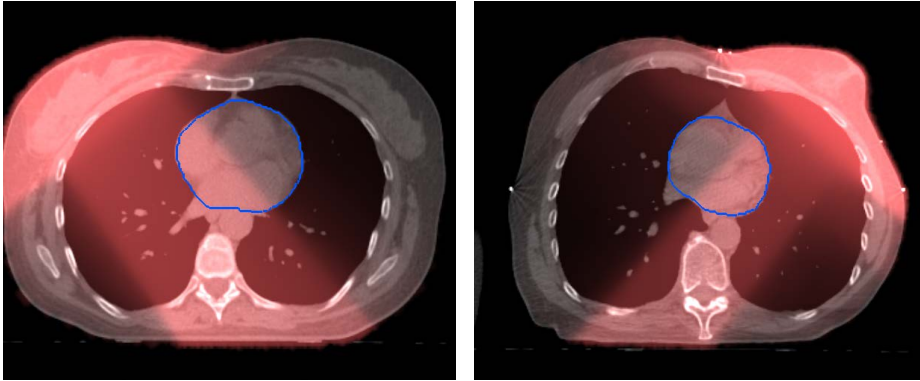
Door toename van het aantal borstkankerpatiënten, eerdere diagnose door screening en betere behandelingen is het aantal vrouwen dat borstkanker overleeft (survivors) toegenomen. Door deze toename is er ook meer aandacht voor de gezondheid en kwaliteit van leven na borstkanker. Een veel voorkomend ziektebeeld in borstkanker survivors is dat van hart- en vaatziekten. Daarnaast kunnen behandelingen voor borstkanker, zoals bijvoorbeeld bestraling, het risico op hart- en vaatziekten vergroten. Tevens is het zo dat van alle borstkankerpatiënten, degene met een verhoogd risico vóór de behandeling ook het hoogste risico hebben op cardiovasculaire bijwerkingen. Tijdig vaststellen van een verhoogd risico op hart- en vaatziekten bij patiënten biedt de mogelijkheid om preventief maatregelen te nemen om latere hart- en vaatziekten te voorkomen.

Atherosclerose is veelal de onderliggende oorzaak van hart- en vaatziekten. Atherosclerose is een aandoening die gekenmerkt wordt door de vorming van plaque in de vaatwanden. Plaque is een ophoping van vetten, cellen en kalk. Atherosclerose is een sluipmoordenaar, die jarenlang kan sluimeren zonder duidelijke klinische symptomen. Daarentegen kan de plaque plotseling instabiel worden en een trombose veroorzaken, wat kan resulteren in een levensbedreigende hartaanval of beroerte.

Normaal gesproken wordt een cardiovasculair risicoprofiel opgesteld aan de hand van traditionele risicofactoren, zoals bijvoorbeeld leeftijd, bloeddruk, cholesterol, diabetes mellitus en roken. Echter heeft één op de vijf borstkankerpatiënten met sub-klinische atherosclerose géén enkele risicofactor en zal daarom niet herkend worden als 'hoog risico'. Het meten van (sub-klinische) atherosclerose kan een uitkomst bieden om met meer zekerheid een verhoogd risico in individuele patiënten op te sporen.

Verkalkte atherosclerotische plaque, ook wel vaatkalk genoemd, is zichtbaar op Röntgen-gebaseerde beeldvormende modaliteiten. Met computed tomography (CT) wordt een 3D weergave van het lichaam gemaakt, waarbij vaatkalk goed zichtbaar is. Deze beelden kunnen gebruikt worden om vaatkalk te kwantificeren en zo patiënten met een verhoogd risico te identificeren. Voor het kwantificeren van vaatkalk, ook wel calcium score genoemd, worden eerst laesies met een hoge dichtheid in de kransslagaderen (of andere relevante slagaderen) gedetecteerd. Deze laesies kunnen vervolgens worden gekwantificeerd in volume, massa of Agatston score. De Agatston score wordt vaak gebruikt omdat deze direct te vertalen is naar een risicocategorie. In de kliniek wordt calcium score met de hand gedaan. Hoewel dit geen moeilijke taak is voor

een medisch specialist, is het repetitief en tijdrovend, vooral wanneer er sprake is van grote hoeveelheden scans. Met name voor grootschalige analyse, zoals bijvoorbeeld in klinische studies, is een manier om automatisch calcium te scoren dus gewenst.



(a) Rechtszijdige radiotherapie

(b) Linkszijdige radiotherapie

Figuur 8.1: Voorbeelden van planning CT scans met een illustratieve heatmap van de geplande dosis van een patiënt met een tumor in de rechter borst (a) en van een patiënt met een tumor in de linker borst (b). Het hart is ingetekend (in blauw) om stralingsschade aan gezond weefsel te voorkomen.

Planning van de radiotherapie behandeling

De meerderheid van de borstkankerpatiënten krijgt radiotherapie tijdens de behandeling. Om de 3D radiotherapie dosisverdeling voor iedere patiënt te bepalen wordt een planning CT scan gemaakt. Op deze CT scan wordt de tumor ingetekend om optimale bestraling van de tumor te garanderen. Ook wordt er een aantal organen ingetekend die in het stralingsveld kunnen liggen. Zo kan de potentiële stralingsschade aan gezond weefsel gemonitord worden. Voor iedere individuele patiënt wordt vervolgens een bestralingsplan gemaakt, wat een zorgvuldige balans is tussen optimale bestraling van de tumor en minimale schade aan gezonde organen.

Omdat blootstelling van hartweefsel aan straling kan leiden tot een verhoogd risico op hartziekten, worden verscheidene maatregelen genomen. Zo wordt het hart ingetekend op de planning CT scan om de hoeveelheid straling op het hart zo klein mogelijk te houden. Ook worden patiënten met een tumor in de linker borst bestraald terwijl ze hun adem inhouden. Op deze manier wordt de afstand tussen de tumor en het hart zo groot mogelijk gemaakt. Patiënten met linkszijdige borstkanker krijgen daarom een tweede CT scan waarbij ze hun adem inhouden.

De planning CT scans bieden de mogelijkheid om de stralingsblootstelling van het hart te monitoren. Omdat ze ook vaatcalcium weergeven, bieden de planning CT scans bovendien een kans om het cardiovasculair risico vast te stellen door middel van

calcium scoren. Naast vaatcalcium bevatten de CT scans nog meer informatie die mogelijk het risico op hart- en vaatziekten van een patiënt mogelijk kan voorspellen.

Automatische beeldanalyse

Voor het automatiseren van beeldanalyse hebben we in deze thesis gebruik gemaakt van verschillende machine learning technieken. Machine learning is een onderdeel van kunstmatige intelligentie (AI) waarbij een computer algoritme leert om een bepaalde taak uit te voeren op basis van voorbeelden (data). Bij conventionele machine learning wordt de data beschreven door bepaalde kenmerken die van tevoren ontworpen zijn en waarvan we denken dat ze relevant zijn voor de taak. Het ontwerpen van deze kenmerken vereist input van een expert met behoorlijke kennis van de taak. In deep learning wordt het ontwerpen van de kenmerken niet door een mens gedaan, maar overgelaten aan een zeer complex en krachtig computermodel. Dit model is in staat om op basis van de data te bepalen wat de relevante kenmerken zijn voor de taak.

Een veelgebruikt model in deep learning zijn kunstmatige neurale netwerken (KNN). Deze modellen simuleren de manier waarop neuronen in het menselijk brein informatie verwerken. Door vele kunstmatige neuronen met elkaar te verbinden in verschillende lagen kan een complex model gemaakt worden dat in staat is extreem non-lineaire verbanden in de data te modelleren. Een speciale soort KNN die in deze thesis (en in de beeldanalyse in het algemeen) veel wordt toegepast, is het convolutionele neurale netwerk (CNN).

Automatische bepaling van cardiovasculair risico

Op basis van de hoeveelheid vaatcalcium in de kransslagaderen (CAC) en in het thoracale gedeelte van de aorta (TAC) kan nauwkeurig een risicoprofiel voor hart- en vaatziekten worden bepaald. Om CAC en TAC te kwantificeren worden normaal gesproken hart CT scans gebruikt, maar ook andere soorten CT scans kunnen worden gebruikt. In **HOOFDSTUK 2** presenteren we een uitgebreide evaluatie van een deep learning methode voor calcium scoren. De methode is in eerste instantie ontworpen voor toepassing in CT scans die zijn gemaakt voor screening op longkanker. In dit hoofdstuk laten we zien dat de methode ook goed werkt in vijf andere typen CT scans. Verder leidt opnieuw trainen van de deep learning methode met representatieve voorbeelden van nieuwe typen CT scans tot een verbetering van het calcium scoren. Idealiter zou een methode voor calcium scoren in de kliniek op alle beschikbare CT scans kunnen worden toegepast. Wanneer alle verschillende typen CT scans gecombineerd worden tijdens het trainen, werkt de methode even goed als gespecialiseerde modellen. Door dit soort generieke methodes is het niet meer nodig om een gespecialiseerd model te ontwikkelen voor ieder type CT scan.

In **HOOFDSTUK 3** en **HOOFDSTUK 4** hebben we de deep learning methode voor calcium scoren gebruikt om onderzoek te doen naar de prevalentie van vaatcalcium en de associatie

met hart- en vaatziekten. Door automatisch calcium scoren was het mogelijk om in deze studies grote groepen borstkankerpatiënten te includeren. In Hoofdstuk 3 beschrijven we de prevalentie van vaatkalk in een groep westerse borstkankerpatiënten en in een groep Aziatische borstkankerpatiënten. In beide groepen hadden patiënten met een hogere leeftijd vaker CAC dan jongere patiënten. Verder kwam TAC in alle leeftijds groepen vaker voor dan CAC.

Verschillende onderzoeken suggereren dat er een relatie bestaat tussen CAC gemeten in planning CT scans en het risico op hart- en vaatziekten. Echter, de exacte associatie tussen CAC gemeten in planning CT scans en hart- en vaatziekten in borstkankerpatiënten is nog onbekend. Om dit verband te onderzoeken, hebben we in Hoofdstuk 4 deze relatie geanalyseerd in een grote populatie van meer dan 18,000 borstkankerpatiënten. De resultaten laten zien dat er een sterke associatie is tussen CAC gemeten op planning CT scans en hart- en vaatziekten. Dit maakt automatisch calcium scoren in planning CT een geschikte tool om borstkankerpatiënten met een verhoogd risico op hart- en vaatziekten te identificeren. Bovendien is de associatie nóg sterker in patiënten die behandeld zijn met antracyclines, een bepaalde vorm van chemotherapie.

In de kliniek wordt calcium scoren gedaan in hart CT scans. Dit zijn scans die gefocust zijn op het hart en gesynchroniseerd zijn met de hartslag door middel van een electrocardiografie (ECG) opname. Hierdoor is de beweging van het hart minimaal tijdens het maken van de CT scan, waardoor het hart en de kransslagaderen optimaal worden afgebeeld. In scans die voor andere doeleinden gemaakt worden wordt vaak geen gebruik gemaakt van ECG-synchronisatie, waardoor bewegingsartefacten in de scan kunnen ontstaan. Hierdoor kunnen vaatkalk laesies in het beeld vervagen. Dit is een probleem omdat de gebruikelijke methodes voor calcium scoren gebruik maken van een afkapwaarde voor het detecteren van vaatkalk. Vervaagde laesies kunnen hierdoor gedeeltelijk of volledig gemist worden. Als direct gevolg hiervan is de reproduceerbaarheid van de calcium score in niet-ECG-gesynchroniseerde scans laag, d.w.z dat er een groot verschil kan zijn tussen calcium scores van twee scans van dezelfde patiënt. In **HOOFDSTUK 5** presenteren we een methode die geen gebruik maakt van een threshold voor segmentatie van vaatkalk laesies. Hierdoor kan de methode met grotere reproduceerbaarheid calcium scoren, vergeleken met methodes die wel een afkapwaarde gebruiken. De methode bestaat uit een zeer complexe CNN die in staat is een beeld met vaatkalk te ontleden in twee onderdelen: een beeld waarin gezond weefsel zonder vaatkalk is afgebeeld en een beeld waarin alléén vaatkalk is afgebeeld. Dit laatste beeld kan vervolgens gebruikt worden om vaatkalk te kwantificeren.

HOOFDSTUK 6 beschrijft een methode die het calcium scoren of een andere manier van vooraf ontworpen informatie extractie overslaat. In plaats daarvan voorspelt de methode cardiovasculaire mortaliteit binnen 5 jaar direct vanaf de CT scan. Zonder kennis van de uiteindelijke taak wordt door een CNN informatie uit de scans geëxtraheerd. Vervolgens wordt deze informatie gebruikt om een conventionele machine

learning classifier te trainen om mortaliteit te voorspellen. Hiermee laten we zien dat het mogelijk is om direct vanaf een CT scan risico op hart- en vaatziekten te voorspellen.

Invloed van radiotherapie op cardiovasculair risico

In **HOOFDSTUK 7** hebben we het effect van bestraling van verschillende hart structuren en kransslagaderen op het risico op ischemische hartziekten onderzocht. Daarnaast hebben we onderzocht of het effect van bestraling anders is in patiënten met CAC dan in patiënten zonder CAC. Op planning CT scans, waarmee het bestralingsplan gemaakt wordt, zijn de hartstructuren en kransslagaderen niet goed te onderscheiden. Dit lukt echter wel goed op hart CT scans met contrast vloeistof. Daarom hebben we referentie segmentaties gemaakt in CT scans met contrast, die gemaakt zijn met een dual-layer detector CT scanner. Met deze speciale scanner kunnen we van een scan met contrast een virtueel non-contrast beeld maken: een beeld dat eruit ziet alsof het zonder contrast gemaakt is. Omdat het contrast beeld en het virtuele non-contrast beeld van dezelfde opname komen, liggen ze perfect over elkaar. Hierdoor kunnen we de segmentaties van het contrast beeld een-op-een vertalen naar het virtuele non-contrast beeld. Deze speciale eigenschappen hebben we gebruikt om een CNN te trainen voor segmentatie van hartstructuren en kransslagaderen in CT scans zonder contrast, zoals de planning CT scans. Met behulp van deze methode konden we de stralingsblootstelling van de verschillende hartstructuren en kransslagaderen berekenen per patiënt. Vervolgens hebben we met survival analyse de relatie tussen stralingsblootstelling en ischemische hartziekten onderzocht. De resultaten laten zien dat patiënten die een hoge stralingsdosis op hartstructuren of kransslagaderen hebben gehad een hoger risico hebben op ischemische hartziekten. Bovendien is de toename van absoluut risico op ischemische hartziekten groter voor patiënten die CAC hebben dan voor patiënten die dat niet hebben.

Toekomstperspectief

De AI methoden die in dit proefschrift gepresenteerd zijn maken het mogelijk om ziekte- en behandeling-gerelateerde parameters te kwantificeren, die anders moeilijk of zelfs onmogelijk te kwantificeren zouden zijn. In de klinische praktijk zouden deze parameters waardevol kunnen zijn om de fysieke toestand van de patiënt beter in kaart te brengen of om de bijwerkingen van behandeling in te schatten. De methoden en evaluaties in dit proefschrift zijn voornamelijk gefocust op toepassing in borstkankerpatiënten. De methoden zijn echter niet beperkt tot deze doelgroep en zullen breder toepasbaar zijn. In sommige gevallen kan dit zonder verdere aanpassingen en in andere zal het nodig zijn om de methode opnieuw te trainen of te "fine-tunen". Bovendien hebben we laten zien dat met behulp van automatische beeldanalyse studies in grote populaties mogelijk worden gemaakt. Verder zouden de gepresenteerde methoden

medisch specialisten kunnen helpen om diagnostiek en behandeling zo goed mogelijk af te stemmen op het individu.

Bibliography

1. P. T. Bradshaw, J. Stevens, N. Khankari, S. L. Teitelbaum, A. I. Neugut, and M. D. Gammon. "Cardiovascular disease mortality among breast cancer survivors", *Epidemiology (Cambridge, Mass.)* Vol. 27 (2016), p. 6 (cited on pp. 10, 55).
2. S. H. Armenian, C. Lacchetti, A. Barac, J. Carver, L. S. Constine, N. Denduluri, S. Dent, P. S. Douglas, J.-B. Durand, M. Ewer, et al. "Prevention and monitoring of cardiac dysfunction in survivors of adult cancers: american society of clinical oncology clinical practice guideline", *Journal of Clinical Oncology*, vol. 35 (2017), pp. 893–911 (cited on pp. 10, 41, 49, 55).
3. J. L. Patnaik, T. Byers, C. DiGuseppi, D. Dabelea, and T. D. Denberg. "Cardiovascular disease competes with breast cancer as the leading cause of death for older females diagnosed with breast cancer: a retrospective cohort study", *Breast Cancer Research*, vol. 13 (2011), R64 (cited on pp. 10, 55, 107).
4. L. A. Smith, V. R. Cornelius, C. J. Plummer, G. Levitt, M. Verrill, P. Canney, and A. Jones. "Cardiotoxicity of anthracycline agents for the treatment of cancer: systematic review and meta-analysis of randomised controlled trials", *BMC cancer*, vol. 10 (2010), pp. 1–14 (cited on pp. 10, 41, 55).
5. Y.-J. Cheng, X.-Y. Nie, C.-C. Ji, X.-X. Lin, L.-J. Liu, X.-M. Chen, H. Yao, and S.-H. Wu. "Long-term cardiovascular risk after radiotherapy in women with breast cancer", *Journal of the American Heart Association*, vol. 6 (2017), p. e005633 (cited on pp. 10, 55, 69).
6. M. J. Hoening, A. Botma, B. M. Aleman, M. H. Baaijens, H. Bartelink, J. G. Klijn, C. W. Taylor, and F. E. Van Leeuwen. "Long-term risk of cardiovascular disease in 10-year survivors of breast cancer", *Journal of the National Cancer Institute*, vol. 99 (2007), pp. 365–375 (cited on pp. 10, 55, 112, 123).
7. D. L. Hershman, C. Till, S. Shen, J. D. Wright, S. D. Ramsey, W. E. Barlow, and J. M. Unger. "Association of cardiovascular risk factors with cardiac events and survival outcomes among patients with breast cancer enrolled in swog clinical trials", *Journal of Clinical Oncology*, vol. 36 (2018), p. 2710 (cited on pp. 10, 55).
8. E. Falk. "Pathogenesis of atherosclerosis", *Journal of the American College of Cardiology*, vol. 47 (2006), pp. C7–C12 (cited on p. 10).
9. M. Zaid, A. Fujiyoshi, A. Kadota, R. D. Abbott, and K. Miura. "Coronary artery calcium and carotid artery intima media thickness and plaque: clinical use in need of clarification", *Journal of atherosclerosis and thrombosis* (2016), RV16005 (cited on p. 10).

10. P. W. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel. "Prediction of coronary heart disease using risk factor categories", *Circulation*, vol. 97 (1998), pp. 1837–1847 (cited on p. 10).
11. R. M. Conroy, K. Pyörälä, A. e. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetiere, P. Jousilahti, U. Keil, et al. "Estimation of ten-year risk of fatal cardiovascular disease in europe: the score project", *European heart journal*, vol. 24 (2003), pp. 987–1003 (cited on p. 10).
12. S. A. Gernaat, I. Išgum, B. D. De Vos, R. A. Takx, D. A. Young-Afat, N. Rijnberg, D. E. Grobbee, Y. van der Graaf, P. A. de Jong, T. Leiner, et al. "Automatic coronary artery calcium scoring on radiotherapy planning ct scans of breast cancer patients: reproducibility and association with traditional cardiovascular risk factors", *PLoS One*, vol. 11 (2016), p. e0167925 (cited on pp. 10, 13, 19, 22, 35, 41, 49, 70, 107).
13. H. S. Hecht. "Coronary artery calcium scanning: past, present, and future", *JACC: Cardiovascular Imaging*, vol. 8 (2015), pp. 579–596 (cited on pp. 11, 23, 75, 107).
14. M. J. Budoff, K. Nasir, R. L. McClelland, R. Detrano, N. Wong, R. S. Blumenthal, G. Kondos, and R. A. Kronmal. "Coronary calcium predicts events better with absolute calcium scores than age-sex-race/ethnicity percentiles: MESA (Multi-Ethnic Study of Atherosclerosis)", *Journal of the American College of Cardiology*, vol. 53 (2009), pp. 345–352 (cited on pp. 11, 75).
15. A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano. "Quantification of coronary artery calcium using ultrafast computed tomography", *Journal of the American College of Cardiology*, vol. 15 (1990), pp. 827–832 (cited on pp. 11, 55, 56, 75).
16. R. Detrano, A. D. Guerci, J. J. Carr, D. E. Bild, G. Burke, A. R. Folsom, K. Liu, S. Shea, M. Szklo, D. A. Bluemke, et al. "Coronary calcium as a predictor of coronary events in four racial or ethnic groups", *New England Journal of Medicine*, vol. 358 (2008), pp. 1336–1345 (cited on pp. 11, 125).
17. J. A. Rumberger, B. H. Brundage, D. J. Rader, and G. Kondos. "Electron beam computed tomographic coronary calcium scanning: a review and guidelines for use in asymptomatic persons", *Mayo Clinic Proceedings*, vol. 74 Elsevier. (1999), pp. 243–252 (cited on p. 11).
18. H. S. Hecht, P. Cronin, M. J. Blaha, M. J. Budoff, E. A. Kazerooni, J. Narula, D. Yankelevitz, and S. Abbara. "2016 SCCT/STR guidelines for coronary artery calcium scoring of non-contrast noncardiac chest CT scans: a report of the society of cardiovascular computed tomography and society of thoracic radiology", *Journal of Cardiovascular Computed Tomography*, vol. 11 (2017), pp. 74–84 (cited on pp. 11, 75).
19. S. C. Darby, M. Ewertz, P. McGale, A. M. Bennet, U. Blom-Goldman, D. Brønnum, C. Correa, D. Cutter, G. Gagliardi, B. Gigante, et al. "Risk of ischemic heart disease in women after radiotherapy for breast cancer", *New England Journal of Medicine*, vol. 368 (2013), pp. 987–998 (cited on pp. 12, 69, 107, 123–125).

20. Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”, *Nature*, vol. 521 (2015), pp. 436–444 (cited on p. 13).
21. G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. “A survey on deep learning in medical image analysis”, *Medical Image Analysis*, vol. 42 (2017), pp. 60–88 (cited on p. 13).
22. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. “Generative adversarial nets”, *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680 (cited on pp. 14, 81).
23. S. M. Grundy, N. J. Stone, A. L. Bailey, C. Beam, K. K. Birtcher, R. S. Blumenthal, L. T. Braun, S. de Ferranti, J. Faiella-Tommasino, D. E. Forman, et al. “2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines”, *Journal of the American College of Cardiology*, vol. 73 (2019), pp. e285–e350 (cited on pp. 19, 36).
24. H. S. Hecht. “Coronary artery calcium scanning: past, present, and future”, *JACC: Cardiovascular Imaging*, vol. 8 (2015), pp. 579–596 (cited on p. 19).
25. I. Išgum, A. Rutten, M. Prokop, and B. van Ginneken. “Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease”, *Medical Physics*, vol. 34 (2007), pp. 1450–61 (cited on p. 19).
26. U. Kurkure, D. R. Chittajallu, G. Brunner, Y. H. Le, and I. A. Kakadiaris. “A supervised classification-based method for coronary calcium detection in non-contrast CT”, *International Journal of Cardiovascular Imaging*, vol. 26 (2010), pp. 817–828 (cited on p. 19).
27. R. Shahzad, T. van Walsum, M. Schaap, A. Rossi, S. Klein, A. C. Weustink, P. J. de Feyter, L. J. van Vliet, and W. J. Niessen. “Vessel specific coronary artery calcium scoring: an automatic system”, *Academic Radiology*, vol. 20 (2013), pp. 1–9 (cited on p. 19).
28. J. M. Wolterink, T. Leiner, R. A. P. Takx, M. A. Viergever, and I. Išgum. “Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection”, *IEEE Transactions on Medical Imaging*, vol. 34 (2015), pp. 1867–78 (cited on p. 19).
29. J. Shemesh, C. I. Henschke, D. Shaham, R. Yip, A. O. Farooqi, M. D. Cham, D. I. McCauley, M. Chen, J. P. Smith, D. M. Libby, M. W. Pasmantier, and D. F. Yankelevitz. “Ordinal scoring of coronary artery calcifications on low-dose CT scans of the chest is predictive of death from cardiovascular disease”, *Radiology*, vol. 257 (2010), pp. 541–548 (cited on p. 19).
30. P. C. Jacobs, M. J. A. Gondrie, Y. van der Graaf, H. J. de Koning, I. Išgum, B. van Ginneken, and W. P. Th. M. Mali. “Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer”, *American Journal of Roentgenology*, vol. 198 (2012), pp. 505–511 (cited on p. 19).

31. C. Chiles, F. Duan, G. W. Gladish, J. G. Ravenel, S. G. Baginski, B. S. Snyder, S. DeMello, S. S. Desjardins, and R. F. Munden. "Association of coronary artery calcification and mortality in the National Lung Screening Trial: a comparison of three scoring methods", *Radiology*, vol. 276 (2015), pp. 82–90 (cited on p. 19).
32. N. Lessmann, B. van Ginneken, M. Zreik, P. A. de Jong, B. D. de Vos, M. A. Viergever, and I. Išgum. "Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions", *IEEE Transactions on Medical Imaging*, vol. 37 (2018), pp. 615–625 (cited on pp. 19, 22, 23, 36, 41, 43, 50, 51, 56, 75, 77, 91, 130).
33. S. A. Gernaat, S. G. M. van Velzen, V. Koh, M. J. Emaus, I. Išgum, N. Lessmann, S. Moes, A. Jacobson, P. W. Tan, D. E. Grobbee, D. H. J. G. van den Bongard, J. I. Tang, and H. M. Verkooijen. "Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of western and asian breast cancer patients", *Radiotherapy and Oncology*, vol. 127 (2018), pp. 487–492 (cited on pp. 19, 55, 107).
34. I. Išgum, B. D. de Vos, J. M. Wolterink, D. Dey, D. S. Berman, M. Rubeaux, T. Leiner, and P. J. Slomka. "Automatic determination of cardiovascular risk by CT attenuation correction maps in rb-82 PET/CT", *Journal of Nuclear Cardiology*, vol. 25 (2018), pp. 2133–2142 (cited on pp. 19, 35).
35. J. Takasu, R. Katz, K. Nasir, J. J. Carr, N. Wong, R. Detrano, and M. J. Budoff. "Relationships of thoracic aortic wall calcification to cardiovascular risk factors: the Multi-Ethnic Study of Atherosclerosis (MESA)", *American Heart Journal*, vol. 155 (2008), pp. 765–771 (cited on p. 19).
36. I. Išgum, A. Rutten, M. Prokop, M. Staring, S. Klein, J. P. W. Pluim, M. A. Viergever, and B. van Ginneken. "Automated aortic calcium scoring on low-dose chest computed tomography", *Medical Physics*, vol. 37 (2010), pp. 714–723 (cited on p. 19).
37. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. "The multimodal brain tumor image segmentation benchmark (brats)", *IEEE transactions on medical imaging*, vol. 34 (2014), pp. 1993–2024 (cited on p. 19).
38. A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, et al. "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge", *Medical image analysis*, vol. 42 (2017), pp. 1–13 (cited on p. 19).
39. S. H. Park and K. Han. "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction", *Radiology*, vol. 286 (2018), pp. 800–809 (cited on p. 19).
40. I. Bank. "Ischaemic heart disease: early recognition and risk disparities", PhD thesis. Utrecht University, 2017 (cited on p. 19).

41. H. A. Taylor Jr, J. G. Wilson, D. W. Jones, D. F. Sarpong, A. Srinivasan, R. J. Garrison, C. Nelson, and S. B. Wyatt. "Toward resolution of cardiovascular health disparities in african americans: design and methods of the jackson heart study", *Ethn Dis*, vol. 15 (2005), pp. S6–4 (cited on p. 20).
42. J. H. Sung, J. Yeboah, J. E. Lee, C. L. Smith, J. G. Terry, M. Sims, T. Samdarshi, S. Musani, E. Fox, Y. Ge, et al. "Diagnostic value of coronary artery calcium score for cardiovascular disease in african americans: the jackson heart study", *British journal of medicine and medical research*, vol. 11 (2016) (cited on pp. 20, 36).
43. The National Lung Screening Trial Research Team. "Reduced lung-cancer mortality with low-dose computed tomographic screening", *New England Journal of Medicine*, vol. 365 (2011), pp. 395–409 (cited on p. 20).
44. J. M. Wolterink, T. Leiner, R. A. Takx, M. A. Viergever, and I. Išgum. "Automatic coronary calcium scoring in non-contrast-enhanced ecg-triggered cardiac CT with ambiguity detection", *IEEE Transactions on Medical Imaging*, vol. 34 (2015), pp. 1867–1878 (cited on pp. 22, 35).
45. B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum. "Convnet-based localization of anatomical structures in 3-d medical images", *IEEE transactions on medical imaging*, vol. 36 (2017), pp. 1470–1481 (cited on p. 23).
46. A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte Jr, and R. Detrano. "Quantification of coronary artery calcium using ultrafast computed tomography", *Journal of the American College of Cardiology*, vol. 15 (1990), pp. 827–832 (cited on pp. 24, 44).
47. M. J. Budoff, L. J. Shaw, S. T. Liu, S. R. Weinstein, P. H. Tseng, F. R. Flores, T. Q. Callister, P. Raggi, D. S. Berman, and T. P. Mosler. "Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients", *Journal of the American College of Cardiology*, vol. 49 (2007), pp. 1860–1870 (cited on pp. 24, 44).
48. A. B. Sevrukov, J. M. Bland, and G. T. Kondos. "Serial electron beam CT measurements of coronary artery calcium: has your patient's calcium score actually changed?", *American Journal of Roentgenology*, vol. 185 (2005), pp. 1546–1553 (cited on pp. 24, 82, 88).
49. R. Lowry. "Kappa as a measure of concordance in categorical sorting [internet]", *VassarStats.net* (2017) (cited on pp. 25, 45).
50. M. J. Blaha, M. Cainzos-Achirica, P. Greenland, J. W. McEvoy, R. Blankstein, M. J. Budoff, Z. Dardari, C. T. Sibley, G. L. Burke, R. A. Kronmal, et al. "Role of coronary artery calcium score of zero and other negative risk markers for cardiovascular disease: the multi-ethnic study of atherosclerosis (mesa)", *Circulation*, vol. 133 (2016), pp. 849–858 (cited on p. 29).
51. J. J. Carr, D. R. Jacobs, J. G. Terry, C. M. Shay, S. Sidney, K. Liu, P. J. Schreiner, C. E. Lewis, J. M. Shikany, J. P. Reis, et al. "Association of coronary artery calcium in adults aged 32 to 46 years with incident coronary heart disease and death", *JAMA cardiology*, vol. 2 (2017), pp. 391–399 (cited on pp. 29, 35).

52. I. Mylonas, M. Kazmi, L. Fuller, R. A. Dekemp, Y. Yam, L. Chen, R. S. Beanlands, and B. J. Chow. "Measuring coronary artery calcification using positron emission tomography-computed tomography attenuation correction images", *European Heart Journal-Cardiovascular Imaging*, vol. 13 (2012), pp. 786–792 (cited on p. 35).
53. R. A. P. Takx, P. A. de Jong, T. Leiner, M. Oudkerk, H. J. de Koning, C. P. Mol, M. A. Viergever, and I. Išgum. "Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability", *PLOS ONE*, vol. 9 (2014), e91239 (cited on pp. 35, 44).
54. C. Cano-Espinosa, G. González, G. R. Washko, M. Cazorla, and R. S. J. Estépar. "Automated agatston score computation in non-ecg gated CT scans using deep learning", *Medical Imaging 2018: Image Processing*, vol. 10574 International Society for Optics and Photonics. (2018), 105742K (cited on pp. 35, 75).
55. N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. "Convolutional neural networks for medical image analysis: full training or fine tuning?", *IEEE transactions on medical imaging*, vol. 35 (2016), pp. 1299–1312 (cited on p. 36).
56. H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. "Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning", *IEEE transactions on medical imaging*, vol. 35 (2016), pp. 1285–1298 (cited on p. 36).
57. H. S. Hecht, P. Cronin, M. J. Blaha, M. J. Budoff, E. A. Kazerooni, J. Narula, D. Yankelevitz, and S. Abbara. "2016 SCCT/STR guidelines for coronary artery calcium scoring of non-contrast noncardiac chest CT scans: a report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology", *Journal of Cardiovascular Computed Tomography*, vol. 11 (2017), pp. 74–84 (cited on p. 36).
58. E. Colzani, A. Liljegren, A. L. Johansson, J. Adolfsson, H. Hellborg, P. F. Hall, and K. Czene. "Prognosis of patients with breast cancer: causes of death and effects of time since diagnosis, age, and tumor characteristics", *Journal of Clinical Oncology*, vol. 29 (2011), pp. 4014–4021 (cited on p. 41).
59. S. Gernaat, P. Ho, N. Rijnberg, M. Emaus, L. Baak, M. Hartman, D. Grobbee, and H. Verkoijen. "Risk of death from cardiovascular disease following breast cancer: a systematic review", *Breast cancer research and treatment*, vol. 164 (2017), pp. 537–555 (cited on p. 41).
60. P. McGale, C. Taylor, C. Correa, D. Cutter, F. Duane, M. Ewertz, R. Gray, G. Mannu, R. Peto, T. Whelan, et al. "Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials." *Lancet (London, England)*, vol. 383 (2014) (cited on p. 41).
61. E. B. C. T. C. Group et al. "Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials", *The Lancet*, vol. 366 (2005), pp. 2087–2106 (cited on pp. 41, 107).

62. C. Taylor and A. Kirby. "Cardiac side-effects from breast cancer radiotherapy", *Clinical Oncology*, vol. 27 (2015), pp. 621–629 (cited on p. 41).
63. T. Suter, N. Cook-Bruns, and C. Barton. "Cardiotoxicity associated with trastuzumab (herceptin) therapy in the treatment of metastatic breast cancer", *The Breast*, vol. 13 (2004), pp. 173–183 (cited on p. 41).
64. T. M. Suter, M. Procter, D. J. Van Veldhuisen, M. Muscholl, J. Bergh, C. Carlomagno, T. Perren, R. Passalacqua, C. Bighin, J. G. Klijn, et al. "Trastuzumab-associated cardiac adverse effects in the herceptin adjuvant trial", *Journal of clinical oncology*, vol. 25 (2007), pp. 3859–3865 (cited on p. 41).
65. R. B. D'agostino, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel. "General cardiovascular risk profile for use in primary care", *Circulation*, vol. 117 (2008), pp. 743–753 (cited on p. 41).
66. H. Kälsch, N. Lehmann, S. Möhlenkamp, C. Hammer, A. A. Mahabadi, S. Moebus, A. Schmermund, A. Stang, M. Bauer, K.-H. Jöckel, et al. "Prevalence of thoracic aortic calcification and its relationship to cardiovascular risk factors and coronary calcification in an unselected population-based cohort: the heinz nixdorf recall study", *The international journal of cardiovascular imaging*, vol. 29 (2013), pp. 207–216 (cited on p. 41).
67. M. J. Budoff, K. Nasir, G. L. Kinney, J. E. Hokanson, R. G. Barr, R. Steiner, H. Nath, C. Lopez-Garcia, J. Black-Shinn, and R. Casaburi. "Coronary artery and thoracic calcium on noncontrast thoracic ct scans: comparison of ungated and gated examinations in patients from the copd gene cohort", *Journal of cardiovascular computed tomography*, vol. 5 (2011), pp. 113–118 (cited on p. 41).
68. K. Nasir, R. Katz, M. Al-Mallah, J. Takasu, D. M. Shavelle, J. J. Carr, R. Kronmal, R. S. Blumenthal, K. O'Brien, and M. J. Budoff. "Relationship of aortic valve calcification with coronary artery calcium severity: the multi-ethnic study of atherosclerosis (mesa)", *Journal of cardiovascular computed tomography*, vol. 4 (2010), pp. 41–46 (cited on p. 41).
69. S. Kohsaka, Z. Jin, T. Rundek, B. Boden-Albala, S. Homma, R. L. Sacco, and M. R. Di Tullio. "Impact of mitral annular calcification on cardiovascular events in a multiethnic community: the northern manhattan study", *JACC: Cardiovascular Imaging*, vol. 1 (2008), pp. 617–623 (cited on p. 41).
70. D. E. Bild, D. A. Bluemke, G. L. Burke, R. Detrano, A. V. Diez Roux, A. R. Folsom, P. Greenland, D. R. Jacobs Jr, R. Kronmal, K. Liu, et al. "Multi-ethnic study of atherosclerosis: objectives and design", *American journal of epidemiology*, vol. 156 (2002), pp. 871–881 (cited on pp. 41, 50).
71. J. M. Hughes-Austin, A. Dominguez, M. A. Allison, C. L. Wassel, D. E. Rifkin, C. G. Morgan, M. R. Daniels, U. Ikram, J. B. Knox, C. M. Wright, et al. "Relationship of coronary calcium on standard chest ct scans with mortality", *JACC: Cardiovascular Imaging*, vol. 9 (2016), pp. 152–159 (cited on p. 41).

72. R. Detrano, A. D. Guerci, J. J. Carr, D. E. Bild, G. Burke, A. R. Folsom, K. Liu, S. Shea, M. Szklo, D. A. Bluemke, D. H. O'Leary, R. Tracy, K. Watson, N. D. Wong, and R. A. Kronmal. "Coronary calcium as a predictor of coronary events in four racial and ethnic groups", *New England Journal of Medicine*, vol. 358 (2008), pp. 1336–45 (cited on p. 41).
73. M. J. Blaha, M. J. Budoff, A. P. DeFilippis, R. Blankstein, J. J. Rivera, A. Agatston, D. H. O'Leary, J. Lima, R. S. Blumenthal, and K. Nasir. "Associations between c-reactive protein, coronary artery calcium, and cardiovascular events: implications for the jupiter population from mesa, a population-based cohort study", *The Lancet*, vol. 378 (2011), pp. 684–692 (cited on p. 41).
74. M. J. Budoff, K. Nasir, R. Katz, J. Takasu, J. J. Carr, N. D. Wong, M. Allison, J. A. Lima, R. Detrano, R. S. Blumenthal, et al. "Thoracic aortic calcification and coronary heart disease events: the multi-ethnic study of atherosclerosis (mesa)", *Atherosclerosis*, vol. 215 (2011), pp. 196–202 (cited on p. 41).
75. C. Iribarren, S. Sidney, B. Sternfeld, and W. S. Browner. "Calcification of the aortic arch: risk factors and association with coronary heart disease, stroke, and peripheral vascular disease", *Jama*, vol. 283 (2000), pp. 2810–2815 (cited on p. 41).
76. J. L. Zamorano, P. Lancellotti, D. Rodriguez Muñoz, V. Aboyans, R. Asteggiano, M. Galderisi, G. Habib, D. J. Lenihan, G. Y. Lip, A. R. Lyon, et al. "2016 esc position paper on cancer treatments and cardiovascular toxicity developed under the auspices of the esc committee for practice guidelines: the task force for cancer treatments and cardiovascular toxicity of the european society of cardiology (esc)", *European heart journal*, vol. 37 (2016), pp. 2768–2801 (cited on pp. 41, 49).
77. P. R. Villadsen, S. E. Petersen, D. Dey, L. Zou, S. Patel, H. Naderi, K. Gruszczynska, J. Baron, L. C. Davies, A. Wragg, et al. "Coronary atherosclerotic plaque burden and composition by ct angiography in caucasian and south asian patients with stable chest pain", *European Heart Journal–Cardiovascular Imaging*, vol. 18 (2017), pp. 556–567 (cited on p. 41).
78. D. E. Bild, R. Detrano, D. Peterson, A. Guerci, K. Liu, E. Shahar, P. Ouyang, S. Jackson, and M. F. Saad. "Ethnic differences in coronary calcification: the multi-ethnic study of atherosclerosis (mesa)", *Circulation*, vol. 111 (2005), pp. 1313–1320 (cited on pp. 41, 50, 51).
79. S. Yusuf, S. Reddy, S. Ôunpuu, and S. Anand. "Global burden of cardiovascular diseases: part ii: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies", *Circulation*, vol. 104 (2001), pp. 2855–2864 (cited on p. 41).
80. D. Young-Afat, C. Van Gils, H. Van Den Bongard, H. Verkooijen, U. S. Group, et al. "The utrecht cohort for multiple breast cancer intervention studies and long-term evaluation (umbrella): objectives, design, and baseline results", *Breast cancer research and treatment*, vol. 164 (2017), pp. 445–450 (cited on p. 41).

81. B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum. “Convnet-based localization of anatomical structures in 3-d medical images”, *IEEE transactions on medical imaging*, vol. 36 (2017), pp. 1470–1481 (cited on pp. 43, 98).
82. National Lung Screening Trial Research Team. “Reduced lung-cancer mortality with low-dose computed tomographic screening”, *New England Journal of Medicine*, vol. 365 (2011), pp. 395–409 (cited on pp. 43, 77, 97).
83. L. J. Shaw, P. Raggi, T. Q. Callister, and D. S. Berman. “Prognostic value of coronary artery calcium screening in asymptomatic smokers and non-smokers”, *European heart journal*, vol. 27 (2006), pp. 968–975 (cited on p. 44).
84. H. Brenner and U. Kliebsch. “Dependence of weighted kappa coefficients on the number of categories”, *Epidemiology* (1996), pp. 199–202 (cited on p. 44).
85. F. Cardoso, L. J. van’t Veer, J. Bogaerts, L. Slaets, G. Viale, S. Delaloge, J.-Y. Pierga, E. Brain, S. Causeret, M. DeLorenzi, et al. “70-gene signature as an aid to treatment decisions in early-stage breast cancer”, *New England Journal of Medicine*, vol. 375 (2016), pp. 717–729 (cited on p. 49).
86. I. H. Kunkler, L. J. Williams, W. J. Jack, D. A. Cameron, J. M. Dixon, et al. “Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (prime ii): a randomised controlled trial”, *The Lancet Oncology*, vol. 16 (2015), pp. 266–273 (cited on p. 49).
87. G. Mühlenbruch, C. Thomas, J. E. Wildberger, R. Koos, M. Das, C. Hohl, M. Katoh, R. W. Günther, and A. H. Mahnken. “Effect of varying slice thickness on coronary calcium scoring with multislice computed tomography in vitro and in vivo”, *Investigative radiology*, vol. 40 (2005), pp. 695–699 (cited on p. 49).
88. J. M. Groen, M. J. Greuter, B. Schmidt, C. Suess, R. Vliegenthart, and M. Oudkerk. “The influence of heart rate, slice thickness, and calcification density on calcium scores using 64-slice multidetector computed tomography: a systematic phantom study”, *Investigative radiology*, vol. 42 (2007), pp. 848–855 (cited on p. 49).
89. F. Bartlett, E. Donovan, H. McNair, L. Corsini, R. Colgan, P. Evans, L. Maynard, C. Griffin, J. Haviland, J. Yarnold, et al. “The uk heartspare study (stage ii): multicentre evaluation of a voluntary breath-hold technique in patients receiving breast radiotherapy”, *Clinical Oncology*, vol. 29 (2017), pp. e51–e56 (cited on p. 50).
90. I. Išgum, M. Prokop, M. Niemeijer, M. A. Viergever, and B. van Ginneken. “Automatic coronary calcium scoring in low-dose chest computed tomography”, *IEEE Transactions on Medical Imaging*, vol. 31 (2012), pp. 2322–34 (cited on p. 50).
91. M. E. Mast, M. W. Heijnenbroek, A. L. Petoukhova, A. N. Scholten, J. H. Schreur, and H. Struikmans. “Preradiotherapy calcium scores of the coronary arteries in a cohort of women with early-stage breast cancer: a comparison with a cohort of healthy women”, *International Journal of Radiation Oncology* Biology* Physics*, vol. 83 (2012), pp. 853–858 (cited on p. 50).

92. A. M. Kanaya, N. R. Kandula, S. K. Ewing, D. Herrington, K. Liu, M. J. Blaha, S. Srivastava, S. S. Dave, and M. J. Budoff. "Comparing coronary artery calcium among us south asians with four racial/ethnic groups: the masala and mesa studies", *Atherosclerosis*, vol. 234 (2014), pp. 102–107 (cited on p. 51).
93. K. Sutton-Tyrrell, L. H. Kuller, D. Edmundowicz, A. Feldman, R. Holubkov, L. Givens, and K. A. Matthews. "Usefulness of electron beam tomography to detect progression of coronary and aortic calcium in middle-aged women", *The American journal of cardiology*, vol. 87 (2001), pp. 560–564 (cited on p. 51).
94. M. J. Budoff, K. Nasir, R. Katz, J. Takasu, J. J. Carr, N. D. Wong, M. Allison, J. A. C. Lima, R. Detrano, R. S. Blumenthal, and R. Kronmal. "Thoracic aortic calcification and coronary heart disease events: the Multi-Ethnic Study of Atherosclerosis (MESA)", *Atherosclerosis*, vol. 215 (2011), pp. 196–202 (cited on p. 51).
95. S. M. Kim, M. J. Chung, K. S. Lee, Y. H. Choe, C. A. Yi, and B.-K. Choe. "Coronary calcium screening using low-dose lung cancer screening: effectiveness of mdct with retrospective reconstruction", *American Journal of Roentgenology*, vol. 190 (2008), pp. 917–922 (cited on p. 51).
96. P. C. Jacobs, M. Prokop, Y. van der Graaf, M. J. Gondrie, K. J. Janssen, H. J. de Koning, I. Išgum, R. J. van Klaveren, M. Oudkerk, B. van Ginneken, et al. "Comparing coronary artery calcium and thoracic aorta calcium for prediction of all-cause mortality and cardiovascular events on low-dose non-gated computed tomography in a high-risk population of heavy smokers", *Atherosclerosis*, vol. 209 (2010), pp. 455–462 (cited on p. 51).
97. M. J. Emaus, I. Išgum, S. G. M. van Velzen, H. J. G. D. van den Bongard, S. A. M. Gernaat, N. Lessmann, M. G. A. Sattler, A. J. Teske, J. Penninkhof, H. Meijer, J.-P. Pignol, and H. M. Verkooijen. "Bragatston study protocol: a multicentre cohort study on automated quantification of cardiovascular calcifications on radiotherapy planning ct scans for cardiovascular risk prediction in patients with breast cancer", *BMJ open*, vol. 9 (2019), p. e028752 (cited on pp. 51, 55, 77, 108).
98. K. M. Sturgeon, L. Deng, S. M. Bluethmann, S. Zhou, D. M. Trifiletti, C. Jiang, S. P. Kelly, and N. G. Zaorsky. "A population-based study of cardiovascular disease mortality risk in us cancer patients", *European heart journal*, vol. 40 (2019), pp. 3889–3897 (cited on p. 55).
99. N. B. Boekel, J. N. Jacobse, M. Schaapveld, M. J. Hooning, J. A. Gietema, F. K. Duane, C. W. Taylor, S. C. Darby, M. Hauptmann, C. M. Seynaeve, et al. "Cardiovascular disease incidence after internal mammary chain irradiation and anthracycline-based chemotherapy for breast cancer", *British journal of cancer*, vol. 119 (2018), pp. 408–418 (cited on p. 55).
100. W. Naaktgeboren, M. Linschoten, A. de Graeff, A. v Rhenen, M. Cramer, F. Asselbergs, A. H. Maas, and A. Teske. "Long-term cardiovascular health in adult cancer survivors", *Maturitas*, vol. 105 (2017), pp. 37–45 (cited on p. 55).
101. P. Greenland, M. J. Blaha, M. J. Budoff, R. Erbel, and K. E. Watson. "Coronary calcium score and cardiovascular risk", *Journal of the American College of Cardiology*, vol. 72 (2018), pp. 434–447 (cited on p. 55).

102. M. Kavousi, C. S. Desai, C. Ayers, R. S. Blumenthal, M. J. Budoff, A.-A. Mahabadi, M. A. Ikram, A. Van Der Lugt, A. Hofman, R. Erbel, et al. "Prevalence and prognostic implications of coronary artery calcification in low-risk women: a meta-analysis", *Jama*, vol. 316 (2016), pp. 2126–2134 (cited on pp. 55, 69).
103. C. T. Roos, V. A. van den Bogaard, M. J. Greuter, R. Vliegenthart, E. Schuit, J. A. Langendijk, A. Van Der Schaaf, A. P. Crijns, and J. H. Maduro. "Is the coronary artery calcium score associated with acute coronary events in breast cancer patients treated with radiotherapy?", *Radiotherapy and Oncology*, vol. 126 (2018), pp. 170–176 (cited on pp. 55, 69).
104. S. G. M. van Velzen, N. Lessmann, B. K. Velthuis, I. E. Bank, D. H. van den Bongard, T. Leiner, P. A. de Jong, W. B. Veldhuis, A. Correa, J. G. Terry, J. J. Carr, M. A. Viergever, H. M. Verkooijen, and I. Išgum. "Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols", *Radiology*, vol. 295 (2020), pp. 66–79 (cited on pp. 55, 56, 75, 78, 83, 91, 107, 108).
105. "Netherlands comprehensive cancer organisation", <https://www.iknl.nl/>, (cited on p. 56).
106. World Health Organization. "Global status report on noncommunicable diseases", 2014 (cited on pp. 56, 57).
107. C. C. voor de WHO-Familie van Internationale Classificaties NL. "Conversie cvz'80 - icd-10 2006", (Accessed April 2020.) (cited on p. 57).
108. J. P. Fine and R. J. Gray. "A proportional hazards model for the subdistribution of a competing risk", *Journal of the American statistical association*, vol. 94 (1999), pp. 496–509 (cited on p. 60).
109. F. Cai, M. A. F. Luis, X. Lin, M. Wang, L. Cai, C. Cen, and E. Biskup. "Anthracycline-induced cardiotoxicity in the chemotherapy treatment of breast cancer: preventive strategies and treatment", *Molecular and Clinical Oncology*, vol. 11 (2019), pp. 15–23 (cited on p. 69).
110. M. Mast, J. Pekelharing, M. Heijenbrok, D. van Klaveren, M. van Kempen-Harteveld, A. Petoukhova, A. Verbeek-de Kanter, J. Schreur, and H. Struikmans. "Reduced increase of calcium scores using breath-hold in left-sided whole breast irradiation", *Radiotherapy and Oncology* (2020) (cited on p. 69).
111. C. Simonetto, M. Eidemüller, A. Gaasch, M. Pazos, S. Schönecker, D. Reitz, S. Käab, M. Braun, N. Harbeck, M. Niyazi, et al. "Does deep inspiration breath-hold prolong life? individual risk estimates of ischaemic heart disease after breast cancer radiotherapy", *Radiotherapy and Oncology*, vol. 131 (2019), pp. 202–207 (cited on p. 69).
112. C. Taylor, P. McGale, D. Brønnum, C. Correa, D. Cutter, F. K. Duane, B. Gigante, M.-B. Jensen, E. Lorenzen, K. Rahimi, et al. "Cardiac structure injury after radiotherapy for breast cancer: cross-sectional study with individual patient data", *Journal of Clinical Oncology*, vol. 36 (2018), p. 2288 (cited on p. 69).

113. H. Wang, M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, et al. "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015", *The Lancet*, vol. 388 (2016), pp. 1459–1544 (cited on p. 75).
114. R. C. Detrano, M. Anderson, J. Nelson, N. D. Wong, J. J. Carr, M. McNitt-Gray, and D. E. Bild. "Coronary calcium measurements: effect of CT scanner type and calcium measure on rescan reproducibility: MESA study", *Radiology*, vol. 236 (2005), pp. 477–484 (cited on pp. 75, 90).
115. S. Mao, H. Bakhsheshi, B. Lu, S. C. Liu, R. J. Oudiz, and M. J. Budoff. "Effect of electrocardiogram triggering on reproducibility of coronary artery calcium scoring", *Radiology*, vol. 220 (2001), pp. 707–711 (cited on pp. 75, 90).
116. U. Hoffmann, U. Siebert, A. Bull-Stewart, S. Achenbach, M. Ferencik, F. Moselewski, T. J. Brady, J. M. Massaro, and C. J. O'Donnell. "Evidence for lower variability of coronary artery calcium mineral mass measurements by multi-detector computed tomography in a community-based cohort—consequences for progression studies", *European Journal of Radiology*, vol. 57 (2006), pp. 396–402 (cited on pp. 75, 90).
117. L. R. Van Hoe, K. G. De Meerleer, P. P. Leyman, and P. K. Vanhoenacker. "Coronary artery calcium scoring using ecg-gated multidetector CT: effect of individually optimized image-reconstruction windows on image quality and measurement reproducibility", *American Journal of Roentgenology*, vol. 181 (2003), pp. 1093–1100 (cited on pp. 75, 90).
118. P. C. Jacobs, I. Işgum, M. J. Gondrie, W. P. T. M. Mali, B. van Ginneken, M. Prokop, and Y. van der Graaf. "Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement", *American Journal of Roentgenology*, vol. 194 (2010), pp. 1244–1249 (cited on p. 75).
119. C. Hong, K. T. Bae, and T. K. Pilgram. "Coronary artery calcium: accuracy and reproducibility of measurements with multi-detector row CT—assessment of effects of different thresholds and quantification methods", *Radiology*, vol. 227 (2003), pp. 795–801 (cited on p. 75).
120. B. D. de Vos, J. M. Wolterink, T. Leiner, P. A. de Jong, N. Lessmann, and I. Işgum. "Direct automatic coronary calcium scoring in cardiac and chest CT", *IEEE Transactions on Medical Imaging* (2019) (cited on pp. 75, 91).
121. J. Groen, H. Dijkstra, M. Greuter, and M. Oudkerk. "Threshold adjusted calcium scoring using CT is less susceptible to cardiac motion and more accurate", *Medical Physics*, vol. 36 (2009), pp. 438–446 (cited on pp. 75, 91).
122. Y. Song, B. L. Eck, J. Levi, and D. L. Wilson. "Improved reproducibility of calcium mass score using deconvolution and partial volume correction", *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10953 International Society for Optics and Photonics. (2019), 109531O (cited on pp. 75, 91).

123. S. C. Saur, H. Alkadhi, L. Desbiolles, G. Székely, and P. C. Cattin. “ACCURATUM: improved calcium volume scoring using a mesh-based algorithm—a phantom study”, *European Radiology*, vol. 19 (2009), pp. 591–598 (cited on pp. 75, 91).
124. J. Šprem, B. D. De Vos, N. Lessmann, R. W. Van Hamersvelt, M. J. Greuter, P. A. De Jong, T. Leiner, M. A. Viergever, and I. Išgum. “Coronary calcium scoring with partial volume correction in anthropomorphic thorax phantom and screening chest CT images”, *PLoS One*, vol. 13 (2018), p. e0209318 (cited on pp. 75, 91).
125. J. Dehmehski, X. Ye, H. Amin, M. Abaei, X. Lin, and S. D. Qanadli. “Volumetric quantification of atherosclerotic plaque in CT considering partial volume effect”, *IEEE Transactions on Medical Imaging*, vol. 26 (2007), pp. 273–282 (cited on pp. 76, 91).
126. J. Šprem, B. D. de Vos, N. Lessmann, P. A. de Jong, M. A. Viergever, and I. Išgum. “Impact of automatically detected motion artifacts on coronary calcium scoring in chest computed tomography”, *Journal of Medical Imaging*, vol. 5 (2018), p. 044007 (cited on p. 76).
127. X. Chen and E. Konukoglu. “Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders”, *International Conference on Medical Imaging with Deep Learning*, 2018 (cited on p. 76).
128. D. Sato, S. Hanaoka, Y. Nomura, T. Takenaga, S. Miki, T. Yoshikawa, N. Hayashi, and O. Abe. “A primitive study on unsupervised anomaly detection with an autoencoder in emergency head CT volumes”, *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575 International Society for Optics and Photonics. (2018), 105751P (cited on p. 76).
129. N. Pawlowski, M. C. Lee, M. Rajchl, S. McDonagh, E. Ferrante, K. Kamnitsas, S. Cooke, S. Stevenson, A. Khetani, T. Newman, et al. “Unsupervised lesion detection in brain CT using bayesian convolutional autoencoders”, *International Conference on Medical Imaging with Deep Learning*, 2018 (cited on p. 76).
130. C. Baur, B. Wiestler, S. Albarqouni, and N. Navab. “Deep autoencoding models for unsupervised anomaly segmentation in brain MR images”, *International MICCAI Brainlesion Workshop*, Springer. 2018, pp. 161–169 (cited on p. 76).
131. J. C. Seah, J. S. Tang, A. Kitchen, F. Gaillard, and A. F. Dixon. “Chest radiographs in congestive heart failure: visualizing neural network learning”, *Radiology*, vol. 290 (2019), pp. 514–522 (cited on p. 76).
132. C. F. Baumgartner, L. M. Koch, K. Can Tezcan, J. Xi Ang, and E. Konukoglu. “Visual feature attribution using wasserstein gans”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8309–8319 (cited on p. 76).
133. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. “Unpaired image-to-image translation using cycle-consistent adversarial networks”, *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232 (cited on pp. 76, 81, 82).
134. L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan, and J. Paisley. “An adversarial learning approach to medical image synthesis for lesion detection”, *IEEE Journal of Biomedical and Health Informatics* (2020) (cited on p. 76).

135. S. G. M. van Velzen, B. D. de Vos, H. M. Verkooijen, T. Leiner, M. A. Viergever, and I. Išgum. "Coronary artery calcium scoring: can we do better?", *Medical Imaging*, vol. 11313 of *Proceedings of SPIE* (2020), 113130 (cited on pp. 77, 91).
136. K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778 (cited on pp. 81, 82, 109).
137. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. "Image-to-image translation with conditional adversarial networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134 (cited on p. 82).
138. J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum. "Generative adversarial networks for noise reduction in low-dose CT", *IEEE Transactions on Medical Imaging*, vol. 36 (2017), pp. 2536–2545 (cited on p. 90).
139. R. W. van Hamersvelt, M. J. Willeminck, R. A. Takx, A. L. Eikendal, R. P. Budde, T. Leiner, C. P. Mol, I. Isgum, and P. A. de Jong. "Cardiac valve calcifications on low-dose unenhanced ungated chest computed tomography: inter-observer and inter-examination reliability, agreement and variability", *European Radiology*, vol. 24 (2014), pp. 1557–1564 (cited on p. 91).
140. P. C. Jacobs, M. J. Gondrie, Y. van der Graaf, H. J. de Koning, I. Isgum, B. van Ginneken, and W. P. Mali. "Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer", *American Journal of Roentgenology*, vol. 198 (2012), pp. 505–511 (cited on p. 97).
141. C. Chiles, F. Duan, G. W. Gladish, J. G. Ravenel, S. G. Baginski, B. S. Snyder, S. DeMello, S. S. Desjardins, R. F. Munden, and N. S. Team. "Association of coronary artery calcification and mortality in the national lung screening trial: a comparison of three scoring methods", *Radiology*, vol. 276 (2015), pp. 82–90 (cited on p. 97).
142. O. M. Mets, R. Vliegenthart, M. J. Gondrie, M. A. Viergever, M. Oudkerk, H. J. de Koning, P. T. M. Willem, M. Prokop, R. J. van Klaveren, Y. van der Graaf, et al. "Lung cancer screening CT-based prediction of cardiovascular events", *JACC: Cardiovascular Imaging*, vol. 6 (2013), pp. 899–907 (cited on pp. 97, 103).
143. B. D. de Vos, P. A. de Jong, J. M. Wolterink, R. Vliegenthart, G. V. Wielingen, M. A. Viergever, and I. Išgum. "Automatic machine learning based prediction of cardiovascular events in lung cancer screening data", *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414 International Society for Optics and Photonics. (2015), p. 94140D (cited on pp. 97, 103).
144. M. Zreik, N. Lessmann, R. W. van Hamersvelt, J. M. Wolterink, M. Voskuil, M. A. Viergever, T. Leiner, and I. Išgum. "Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis", *Medical Image Analysis*, vol. 44 (2018), pp. 72–85 (cited on p. 100).
145. X. Hou, L. Shen, K. Sun, and G. Qiu. "Deep feature consistent variational autoencoder", *IEEE*. 2017, pp. 1133–1141 (cited on p. 100).

146. K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556* (2014) (cited on p. 100).
147. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision (IJCV)*, vol. 115 (2015), pp. 211–252 (cited on p. 100).
148. E. B. C. T. C. Group et al. “Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials”, *The Lancet*, vol. 378 (2011), pp. 1707–1716 (cited on p. 107).
149. J. Ragaz, I. A. Olivotto, J. J. Spinelli, N. Phillips, S. M. Jackson, K. S. Wilson, M. A. Knowling, C. M. Coppin, L. Weir, K. Gelmon, et al. “Locoregional radiation therapy in patients with high-risk breast cancer receiving adjuvant chemotherapy: 20-year results of the british columbia randomized trial”, *Journal of the National Cancer Institute*, vol. 97 (2005), pp. 116–126 (cited on p. 107).
150. S. Jacob, J. Camilleri, S. Derreumaux, V. Walker, O. Lairez, M. Lapeyre, E. Bruguière, A. Pathak, M.-O. Bernier, D. Laurier, et al. “Is mean heart dose a relevant surrogate parameter of left ventricle and coronary arteries exposure during breast cancer radiotherapy: a dosimetric evaluation based on individually-determined radiation dose (baccarat study)”, *Radiation Oncology*, vol. 14 (2019), p. 29 (cited on pp. 107, 112, 123, 125).
151. V. A. van den Bogaard, B. D. Ta, A. van der Schaaf, A. B. Bouma, A. M. Middag, E. J. Bantema-Joppe, L. V. van Dijk, F. B. van Dijk-Peters, L. A. Marteiijn, G. H. de Bock, et al. “Validation and modification of a prediction model for acute cardiac events in patients with breast cancer treated with radiotherapy based on three-dimensional dose distributions to cardiac substructures”, *Journal of Clinical Oncology*, vol. 35 (2017), p. 1171 (cited on pp. 107, 124).
152. G. Nilsson, L. Holmberg, H. Garmo, O. Duvernoy, I. Sjögren, B. Lagerqvist, and C. Blomqvist. “Distribution of coronary artery stenosis after radiation for breast cancer”, *Journal of clinical oncology*, vol. 30 (2012), pp. 380–386 (cited on p. 107).
153. C. R. Correa, H. I. Litt, W.-T. Hwang, V. A. Ferrari, L. J. Solin, and E. E. Harris. “Coronary artery findings after left-sided compared with right-sided radiation treatment for early-stage breast cancer”, *Journal of clinical oncology*, vol. 25 (2007), pp. 3031–3037 (cited on p. 107).
154. M. Feng, J. M. Moran, T. Koelling, A. Chughtai, J. L. Chan, L. Freedman, J. A. Hayman, R. Jagsi, S. Jolly, J. Larouere, et al. “Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer”, *International Journal of Radiation Oncology* Biology* Physics*, vol. 79 (2011), pp. 10–18 (cited on p. 107).
155. N. Maffei, L. Fiorini, G. Aluisio, E. D’Angelo, P. Ferrazza, V. Vanoni, F. Lohr, B. Meduri, and G. Guidi. “Hierarchical clustering applied to automatic atlas based segmentation of 25 cardiac sub-structures”, *Physica Medica*, vol. 69 (2020), pp. 70–80 (cited on pp. 107, 125).

156. F. Duane, M. C. Aznar, F. Bartlett, D. J. Cutter, S. C. Darby, R. Jagsi, E. L. Lorenzen, O. McArdle, P. McGale, S. Myerson, et al. "A cardiac contouring atlas for radiotherapy", *Radiotherapy and Oncology*, vol. 122 (2017), pp. 416–422 (cited on pp. 107, 125).
157. S. A. Gernaat, J. M. Boer, D. H. van den Bongard, A. H. Maas, C. C. van der Pol, R. M. Bijlsma, D. E. Grobbee, H. M. Verkooyen, and P. H. Peeters. "The risk of cardiovascular disease following breast cancer by framingham risk score", *Breast cancer research and treatment*, vol. 170 (2018), pp. 119–127 (cited on p. 107).
158. S. Bruns, J. M. Wolterink, R. A. Takx, R. W. van Hamersvelt, D. Suchá, M. A. Viergever, T. Leiner, and I. Išgum. "Deep learning from dual-energy information for whole-heart segmentation in dual-energy and single-energy non-contrast-enhanced cardiac ct", *Medical Physics*, vol. 47 (2020), pp. 5048–5060 (cited on pp. 107–109, 111, 125, 126).
159. R. W. van Hamersvelt, I. Išgum, P. A. de Jong, M. J. M. Cramer, G. E. Leenders, M. J. Willemink, M. Voskuil, and T. Leiner. "Application of spectral computed tomography to improve specificity of cardiac computed tomography (clarity study): rationale and design", *BMJ open*, vol. 9 (2019), p. e025793 (cited on p. 108).
160. J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum. "Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier", *Medical image analysis*, vol. 51 (2019), pp. 46–60 (cited on p. 108).
161. M. J. Blaha, M. J. Budoff, R. Tota-Maharaj, Z. A. Dardari, N. D. Wong, R. A. Kronmal, J. Eng, W. S. Post, R. S. Blumenthal, and K. Nasir. "Improving the CAC score by addition of regional measures of calcium distribution: Multi-Ethnic Study of Atherosclerosis", *JACC: Cardiovascular Imaging*, vol. 9 (2016), pp. 1407–16 (cited on p. 108).
162. T. G. Dietterich. "Ensemble methods in machine learning", *International workshop on multiple classifier systems*, Springer. 2000, pp. 1–15 (cited on p. 109).
163. D. P. Kingma and J. Ba. "Adam: a method for stochastic optimization", *arXiv preprint arXiv:1412.6980* (2014) (cited on p. 109).
164. W. H. Organization et al. "The icd-10 classification of mental and behavioural disorders: diagnostic criteria for research", vol. 2 (World Health Organization, 1993) (cited on pp. 111, 133).
165. J. N. Jacobse, F. K. Duane, N. B. Boekel, M. Schaapveld, M. Hauptmann, M. J. Hooning, C. M. Seynaeve, M. H. Baaijens, J. A. Gietema, S. C. Darby, et al. "Radiation dose-response for risk of myocardial infarction in breast cancer survivors", *International Journal of Radiation Oncology* Biology* Physics*, vol. 103 (2019), pp. 595–604 (cited on p. 124).
166. E. L. Lorenzen, J. C. Rehammar, M.-B. Jensen, M. Ewertz, and C. Brink. "Radiation-induced risk of ischemic heart disease following breast cancer radiotherapy in denmark, 1977–2005", *Radiotherapy and Oncology* (2020) (cited on p. 124).

167. C. Ares, S. Khan, A. M. MacArtain, J. Heuberger, G. Goitein, G. Gruber, G. Lutters, E. B. Hug, S. Bodis, and A. J. Lomax. "Postoperative proton radiotherapy for localized and locoregional breast cancer: potential for clinically relevant improvements?", *International Journal of Radiation Oncology* Biology* Physics*, vol. 76 (2010), pp. 685–697 (cited on p. 125).
168. A. Moignier, D. Broggio, S. Derreumaux, A. Beaudré, T. Girinsky, J.-F. Paul, D. Drubay, D. Lefkopoulos, D. Franck, B. Aubert, et al. "Coronary stenosis risk analysis following hodgkin lymphoma radiotherapy: a study based on patient specific artery segments dose calculation", *Radiotherapy and Oncology*, vol. 117 (2015), pp. 467–472 (cited on p. 125).
169. R. Kaderka, E. F. Gillespie, R. C. Mundt, A. K. Bryant, C. B. Sanudo-Thomas, A. L. Harrison, E. L. Wouters, V. Moiseenko, K. L. Moore, T. F. Atwood, et al. "Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients", *Radiotherapy and Oncology*, vol. 131 (2019), pp. 215–220 (cited on p. 125).
170. F. Van Dijk-Peters, N. Sijtsema, R. Kierkels, R. Vliegthart, J. Langendijk, J. Maduro, and A. Crijs. "Oc-0259: validation of a multi-atlas based auto-segmentation of the heart in breast cancer patients", *Radiotherapy and Oncology*, vol. 115 (2015), pp. S132–S133 (cited on p. 125).
171. R. Topolnjak, G. R. Borst, J. Nijkamp, and J.-J. Sonke. "Image-guided radiotherapy for left-sided breast cancer patients: geometrical uncertainty of the heart", *International Journal of Radiation Oncology* Biology* Physics*, vol. 82 (2012), pp. e647–e655 (cited on p. 126).
172. B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama. "Co-teaching: robust training of deep neural networks with extremely noisy labels", *Advances in neural information processing systems*, 2018, pp. 8527–8537 (cited on p. 133).
173. A. B. De González, M. Mahesh, K.-P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land. "Projected cancer risks from computed tomographic scans performed in the united states in 2007", *Archives of internal medicine*, vol. 169 (2009), pp. 2071–2077 (cited on p. 135).
174. M. A. Badgeley, J. R. Zech, L. Oakden-Rayner, B. S. Glicksberg, M. Liu, W. Gale, M. V. McConnell, B. Percha, T. M. Snyder, and J. T. Dudley. "Deep learning predicts hip fracture using confounding patient and healthcare variables", *NPJ digital medicine*, vol. 2 (2019), pp. 1–10 (cited on p. 135).
175. M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. "Synthetic data augmentation using gan for improved liver lesion classification", *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, 2018, pp. 289–293 (cited on p. 135).
176. M. D. Zeiler and R. Fergus. "Visualizing and understanding convolutional networks", *European conference on computer vision*, Springer, 2014, pp. 818–833 (cited on p. 136).
177. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. "Grad-cam: visual explanations from deep networks via gradient-based localization", *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626 (cited on p. 136).

178. M. Sundararajan, A. Taly, and Q. Yan. “Axiomatic attribution for deep networks”, *arXiv preprint arXiv:1703.01365* (2017) (cited on p. 136).

Acknowledgments

This thesis represents the conclusion of four wonderful years. I would have never reached this milestone without the support of a great number of people. I am grateful to all those who have contributed to this thesis and my PhD experience as a whole, and would like to thank a few people in particular.

First and foremost, I would like to thank Ivana, my supervisor. Thank you for all the support, advice and guidance. I could always trust that you wanted the best for me and acted in my best interest. Thank you for always being available, even during holidays and working so hard for my benefit. I have learned a great deal about research, writing and academia from you. I think I can still learn a lot from you, and am grateful for the opportunity to do so.

Dear Lenny, thank you for bringing a clear clinical perspective to my thesis. Your practicality and purposefulness have made the BRAGATSTON study what it is: a beautiful study of which I am proud to have played a part.

Dear Max, thank you for the constructive feedback you gave me for all of my articles. I am also grateful for the encouraging words in stressful times and the compliments on achievements along the way.

This thesis is the product of a collaborative effort of many people. I would like to thank all co-authors for the good collaborations, fruitful discussions and the helpful feedback over the years. I would like to thank the members of the reading committee, Prof. dr. van Gils, Prof. dr. Terhaard, Prof. dr. Alderliesten, Prof. dr. Sánchez Gutiérrez and Prof. dr. van der Harst, for their time and effort spent on evaluating this thesis.

Dear Roxanne, Marleen and Sofie, I enjoyed working together on the BRAGATSTON project very much. Thank you for the productive meetings, all the help with statistics and guidance in the jungle of CBS. I would also like to thank all people that contributed to the BRAGATSTON study.

I am grateful to the many wonderful colleagues that had a part in this journey. First, of course, Julia, I enjoyed sharing an office, a radio, office-grown plants and many jars of peanut butter. Thank you for being a good friend, a supportive colleague and my paranymp. Dear Steffen, thank you for the countless handstands and conversations about food. I am happy to call you my friend and paranymp. Matthijs, thank you for the deep conversations and making me laugh so many times. Nadieh, thank you for being 'my lovely colleague' and for making MICCAI so much fun. Dear MD. Max, you have stolen my phone many times to teach me a lesson I have still not learned. I had a lot of fun with you and Julia at RSNA. Bea, Mariëlle and Kim, I enjoyed our (girly)

dinner and evenings. Majd, thank you for the good advice and welcome ice-cream breaks. Nikolas, I have learned a lot from you and I am thankful for all the patience and help, especially in the beginning of my PhD. Bob and Jelmer, thank you for the feedback and ideas over the years. Former and current members of the QIA group, thank you for making it a nice group to work in, for the productive meetings and for the nice social events. My fellow PhD students in the OIO-steeg, thank you for many welcome coffee breaks and for the good atmosphere in our hallway. Everyone at ISI, thank you for sharing many lunch breaks, cakes, Christmas parties, the Ardennen weekend and more.

Dear Mom and Dad, you have raised me to be curious and to not give up easily, which both helped me enormously in the past years. I could always count on you for help and advice and you being proud of me was incredibly motivating. Thank you for that. Merel and Roos, thank you for being supporting sisters and friends.

Dear Sibren, we have built something beautiful together the past years. Thank you for guarding my boundaries when I forget to do that. Thank you for sharing the good and exciting moments with me and supporting me during the stressful times.

Publications

Journal publications

S. G. M. van Velzen, R. Gal, S. Bruns, J. M. Wolterink, T. Leiner, A. J. Teske, F. van der Leij, D. H. van den Bongard, M. A. Viergeever, H. M. Verkooijen, and I. Išgum. “AI for estimation of radiotherapy-related risk of ischemic heart disease in breast cancer survivors”. Submitted.

S. G. M. van Velzen, B. D. de Vos, J. M. Noothout, H. M. Verkooijen, M. A. Viergeever, and I. Išgum. “Generative models for reproducible coronary calcium scoring”. Submitted.

R. Gal, **S. G. M. van Velzen**, M. J. Hooning, M. J. Emaus, F. van der Leij, M. L. Gregorowitsch, E. L. Blezer, S. A. Gernaat, N. Lessmann, M. G. Sattler, T. Leiner, P. A. de Jong, A. J. Teske, J. Verloop, J. J. Penninkhof, V. Ilonca, H. Meijer, J. J. van Tol-Geerdink, J.-P. Pignol, D. H. van den Bongard, I. Išgum, and H. M. Verkooijen. “Identification of breast cancer patients at risk of cardiovascular disease by automatic quantification of coronary artery calcifications on radiotherapy planning CT scans”. Submitted.

S. G. M. van Velzen, N. Lessmann, B. K. Velthuis, I. E. Bank, D. H. van den Bongard, T. Leiner, P. A. de Jong, W. B. Veldhuis, A. Correa, J. G. Terry, J. J. Carr, M. A. Viergeever, H. M. Verkooijen, and I. Išgum. “Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols”, *Radiology*, vol. 295 (2020), pp. 66–79.

M. J. Emaus, I. Išgum, **S. G. M. van Velzen**, H. J. G. D. van den Bongard, S. A. M. Gernaat, N. Lessmann, M. G. A. Sattler, A. J. Teske, J. Penninkhof, H. Meijer, J.-P. Pignol, and H. M. Verkooijen. “Bragatston study protocol: a multicentre cohort study on automated quantification of cardiovascular calcifications on radiotherapy planning ct scans for cardiovascular risk prediction in patients with breast cancer”, *BMJ open*, vol. 9 (2019), p. e028752.

N. Hampe, J. M. Wolterink, **S. G. M. Van Velzen**, T. Leiner, and I. Išgum. “Machine learning for assessment of coronary artery disease in cardiac ct: a survey”, *Frontiers in cardiovascular medicine*, vol. 6 (2019), p. 172.

S. A. M. Gernaat*, **S. G. M. van Velzen***, V. Koh, M. J. Emaus, I. Išgum, N. Lessmann, S. Moes, A. Jacobson, P. W. Tan, D. E. Grobbee, D. H. J. G. van den Bongard, J. I. Tang, and H. M. Verkooijen. “Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of western and asian breast cancer patients”, *Radiotherapy and Oncology*, vol. 127 (2018), 487–492 *Both authors contributed equally.

Conference proceedings

S. G. M. van Velzen, B. D. de Vos, H. M. Verkooijen, T. Leiner, M. A. Viergever, and I. Išgum. “Coronary artery calcium scoring: can we do better?”, *Medical Imaging*, vol. 11313 of *Proceedings of SPIE* (2020), 113130.

S. G. M. van Velzen, M. Zreik, N. Lessmann, M. A. Viergever, P. A. de Jong, H. M. Verkooijen, and I. Išgum. “Direct prediction of cardiovascular mortality from low-dose chest CT using deep learning”, *Medical Imaging*, vol. 10949 of *Proceedings of SPIE* (2019), 1094932.

Conference abstracts

R. Gal, **S. G. M. van Velzen**, M. Emaus, M. Emaus, D. van den Bongard, M. Gregorowitsch, E. Blezer, S. Gernaat, N. Lessmann, M. Sattler, M. Hooning, A. Teske, J. Penninkhof, H. Meijer, J. Pignol, J. Verloop, I. Išgum, and H. Verkooijen. “The risk of cardiovascular disease in irradiated breast cancer patients: the role of cardiac calcifications and adjuvant treatment”, *European Journal of Cancer*, 2020.

S. G. M. van Velzen, J. Terry, B. de Vos, N. Lessmann, S. Nair, A. Correa, H. Verkooijen, J. Carr, and I. Išgum. “Automatic prediction of coronary heart disease events using coronary and thoracic aorta calcium among African Americans in the Jackson Heart study”, *Radiological Society of North America, 105th Annual Meeting*, 2019.

S. G. M. van Velzen, N. Lessmann, M. Emaus, D. van den Bongard, H. Verkooijen, and I. Išgum. “Deep learning for calcium scoring in radiotherapy treatment planning CT scans in breast cancer patients”, *Radiological Society of North America, 105th Annual Meeting*, 2019.

Biography



Sanne G.M. van Velzen was born on December 29, 1992 in Castricum, the Netherlands. After completing secondary school in 2011, she started her bachelors in medical natural sciences at the Vrije Universiteit in Amsterdam, the Netherlands. In 2014 she continued on this path by doing her masters in medical natural sciences with a specialization towards medical imaging. In her final year, she worked on analysis of positron emission tomography images for detection of coronary plaque and cardiac perfusion deficits at the department of nuclear medicine in the Vrije Universiteit Medical Center in Amsterdam, the Netherlands. Because medical image analysis and cardiovas-

cular disease sparked her interest, she continued to work on this topic at the Image Sciences Institute at the University Medical Center Utrecht, the Netherlands. Under supervision of prof. Ivana Išgum, prof. Helena M. Verkooijen and prof. Max A. Viergever, she worked on automated quantification of cardiovascular disease risk using artificial intelligence techniques in computed tomography images of breast cancer patients. The results are described in this thesis. She continues her research as postdoctoral researcher at the Quantitative Medical Image Analysis Group at the Amsterdam University Medical Center.