The implementation and feasibility of Radiomics in an Aortic Stenosis patient cohort

Klaske Siegersma

January 26, 2017
Version: Final Version

University of Twente

UNIVERSITY OF TWENTE.

Faculty of Science and Technology
Technical Medicine

UMC Utrecht

Divisie Beeld
Radiologie

Master's Thesis

The implementation and feasibility of Radiomics in an Aortic Stenosis patient cohort

Klaske Siegersma

First Clinical Supervisor J.W.H. Verjans MD PhD
SAHMRI: South Australian Health & Medical Research Institute
University of Adelaide

Second Clinical Supervisor Prof. T. Leiner MD PhD
Department of Radiology
University Medical Center Utrecht

Technical Supervisor F. van der Heijden PhD
RAM: Robotics and Mechatronics
University of Twente

Process Supervisor B.J.C.C. Hessink-Sweep MSc
Technical Medicine
University of Twente

January 26, 2017

Klaske Siegersma
The implementation and feasibility of Radiomics in an Aortic Stenosis patient cohort
Master's Thesis, January 26, 2017

University of Twente

Technical Medicine
Faculty of Science and Technology
Drienerlolaan 5
7522 NB and Enschede

Abstract

Introduction Aortic Stenosis (AS) is a very common and deadly valvular disease, predominantly present in the elderly. Outflow obstruction of AS leads to hypertrophy of cardiac muscle cells, eventually leading to fibrosis. This non-ischemic scarring is visualized with (LGE)-images of cardiac magnetic resonance (CMR). Fibrosis can induce heart failure and sudden cardiac death. Since AS has relatively high mortality, timely and accurate risk-stratification of the patients that benefit from early valve surgery is important. Radiomics is a novel method for extraction of quantitative features from medical images, relating image features to phenotyping, diagnosis and treatment through predictive modelling. This study implements radiomics on an AS patient cohort for predicting risk of surgery. A second study utilizes radiomics for computer-aided diagnosis of myocardial fibrosis.

Methods Dataset-1 included 146 AS-patients for predictive modelling of aortic valve replacement (AVR). This cohort and additional controls were used for identification of fibrosis on CMR. A segmentation of the myocardium was performed for extraction of radiomic features. Cylindrical reconstruction of myocardium aided in extraction of case-specific features and texture feature analysis. Univariate analysis was performed on individual features. Multivariate analysis with temporal validation included a generalized linear model (GLM), random forest (RF) and support vector machine (SVM), with minimum redundancy, maximum relevance (mRMR) feature selection. A second feature set, comprised of clinical features, was used to determine the performance of these features in prediction of AVR and computer-aided diagnosis of LGE. Performance measures were concordance index (CI), respectively Area Under the Curve (AUC). A second dataset was implemented for external validation.

Results 5639 features were extracted from LGE-CMR images. Univariate analysis for AVR revealed 49 prognostic features (FDR q-value$<$0.05, CI$>$0.6). Multivariate clinical GLM, including peak aortic jet velocity, high-sensitivity troponin-I and electrocardiographic strain pattern showed higher CI (0.86) than models built with radiomic features (average CI: 0.55). Classification of fibrosis showed opposite performance; average AUC of models with radiomic features was 0.92, clinical modelling showed 0.78. External validation showed similar performance to temporal validation for prediction of AVR (average CI: 0.60), but lower for classification of LGE (AUC: 0.70).

Discussion This study was the first study to implement external validation for predictive modelling and computer-aided diagnosis in CMR. Although training data and external validation cohort were significantly different in patient characteristics, promising results were shown for classification of LGE. A larger dataset can aid in further analysis to determine the optimal timing of AVR and clinical pathway for patients with AS.

Acknowledgements

After approximately 8,5 years at the University of Twente, I have finally come to the point of defending my thesis for Technical Medicine and signing my diploma. Those 8,5 years have been a great and sometimes bumpy ride, with ups and down; personal challenges, doubts about my choice of education, a sidetrack towards the field of Health Sciences and many more.

This thesis would not be as it is, without the help of some people. First of all, Johan; Thanks for always being enthusiastic, promoting new ideas and helping me structure. Although you moved Down Under only 3 months into my graduation, I could always easily contact you through WhatsApp and Skype. You are a great, motivating supervisor, which became especially clear when you pushed me to go to Boston. I was in two minds at first, but I would have never want to miss this experience! Tim; thank you for the nice talks we had. You always made time for me and were able to make me feel welcome, although I know you have a super tight (multi-country) schedule. And thanks for the great coffee; much better than the regular machine coffee in the hospital. Majd, thank you for being critical at my work and machine learning pipelines. You taught me a lot about the world of machine learning. Ferdi, you always make time for the students that you supervise. It made me feel welcome and gave me the time to tell my progress and doubts. And my final supervisor, Bregje; I remember my first intervision with you during M2, when you were directly able to make a core quadrant for me, although I just met you. I sincerely appreciated your ability to read and mirror me, so I could discover new things about myself.

Furthermore, I have spent 4 months in the scientific capitol of the world; Boston. This would not have been possible without funding from the Dutch Heart Association. Thank you! I would also like to thank the principal investigator of the Computational and Bioinformatics Laboratory (CIBL) in Boston for having me; Hugo Aerts, and the postdocs, PhD's and other affiliates for their input, nice lunches and beers every once in a while.

There are many people that have made my 8,5 years during my education the best 8,5 years so far; de 'Zonnetjes van Noorderzon', my 'jaarclub' Extase, the most beautiful sorority of Enschede; 'Damesdispuut Pimpelle' and many, many others.

THANKS! Yet, during this last 1,5 year at the UMC Utrecht and in Boston, my appreciations go to Sander, never refusing an invitation for coffee and Diederik; our in-depth multi-hour analyses of El Giro, La Boucle and La Vuelta were a welcome shift from daily research. One day, we are going to develop a machine-learning algorithm for the prediction of winners. Until then, thanks for your advice so I can still secure my victory in the Giro-pool.

En dit lêste stikje giet fansels yn it Frysk. Leave Chris, bedankt foar it wêzen fan ús twadde mem, alle oeren foarlêzen en ús op bêd bringen. Leave Douwe, earder koe ik dij wol ris efter it behang plakke, mar ik bin grutsk op dij en op alles watsto al berikt hast. Oer ien jier bin ik bij dyn ôfstudearpraatsje! Leave Tyn, net allinich myn suske, mar ek myn bêste freundinne. Ik sjoch 'r sa nei út om mei dij op reis te gean. Dat wurdt wier in toptiid. Leave Gerben, do wiest noch bêst jong doe't ik út'e hûs gong, mar it ôfrûne jier thús ha ik in hiele leuke bân mei dij opboud. Ik hoopje datsto in moaie tiid yn Enschede temjitte giest. En as lêste, alderleafste heit en mem; Ik kin mij gjin bettere âlders winskje. Jimme hawwe my altyd steund, op 'e moaie mominten, mar ek op 'e minder goeie mominten. Tige tanke foar alle kânsen dy 't jimme my jûn en gund hawwe, foar alles dat ik dwaan koe, sa 't ik mij ûntwikkele ha oan 't de persoan dy 't ik no bin. En no, no wurdt it einliks tiid om op eigen fuorten te stean!

Contents

List of Abbreviations

AI	Artificial Intelligence.
AS	Aortic Stenosis.
AUC	Area Under the Curve.
AVA	Aortic Valve Area.
AVR	Aortic Valve Replacement.
CAD	Coronary Artery Disease.
CI	Concordance Index.
CMR	Cardiac Magnetic Resonance.
CT	Computed Tomography.
EKG	Electrocardiography.
FDG-PET	Fludeoxyglucose-Positron Emission Tomography.
FDR	False Detection Rate.
FLASH	fast low-angle shot technique.
FSE	Fast-spin echo.
GLCM	Grey-level Co-occurence Matrix.
GLM	Generalized Linear Model.
GLRLM	Grey-level Run-Length Matrix.
LGE	Late Gadolinium Enhancement.
LOF	Local Outlier Factor.
LV	Left Ventricle.
LVEF	Left Ventricular Ejection Fraction.
MRI	Magnetic Resonance Imaging.
mRMR	minimum redudancy - maximum relevance.
RF	Random Forest.
ROC	Receiver-Operating Characteristic.

SAVR	Surgical Aortic Valve Replacement.
SVM	Support Vector Machine.
TAVI	Transcatheter Aortic Valve Implantation.

General Introduction

1

1.1 Medical Background

1.1.1 Aortic Stenosis

Aortic Stenosis (AS) is a medical condition of the aortic valve. This is the valve that regulates the passage of blood between the Left Ventricle (LV) and the aorta during a heartbeat (figure 1.1). Normal aortic valves have three cusps; each <1 mm thick. Four different layers make up for the histologic structure of a cusp; the endothelium (facing the aorta inlet and continuous with the lining of the aorta), fibrosa, spongiosa and ventricularis (facing the ventricle). The outer layers consist of endothelial cells, covering the inner connective layers. The cusps are attached to the aortic root by the annulus; a collagenous network, which facilitates the transmission of forces enacted by the blood flows.[40]

In a subset of patients, the aortic valve becomes calcified; the cusps of the valve stiffen, caused by a process of thickening, calcification and fibrotization. Historically, this process has been dedicated to ageing of the valves and the constant force applied to the cusps. However, previous published research show that endothelial damage of the cusps causes a cascade of inflammatory processes.[94, 40] However, the exact underlying processes remain unclear and are under continuous investigation.

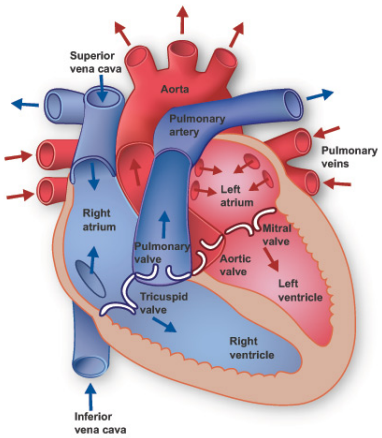


Fig. 1.1: Overview of the human heart. The aortic valve is located between the LV and the aorta and controls the flow of blood from the ventricle into the aorta.

The prevalence of clinically significant calcific AS is 1-3 % of all individuals >70. However, approximately 25% of all individuals >65 have a precursor of AS in the form of sclerotic aortic valve cusps.[157] If left untreated after the onset of symptoms, the 1-year survival is as low as 50%.[105] Previous research has shown that patients with severe asymptomatic AS can benefit from early surgery before the onset of symptoms, reducing their risk of cardiac mortality.[72]

There are two other diseases that can cause AS; the bicuspid aortic valve and rheumatic fever. The bicuspid valve is a heritable congenital malformation, where two cusps of the valve are fused into one. Early in life, a bicuspid aortic valve can be regarded as a benign lesion. However, when the patient becomes older, his morbidity and mortality increases. Due to the pathologic anatomy of the valve, the blood flow from the LV is different from a tricuspid valve. This leads to an increased risk of other cardiovascular pathologies, e.g. infective endocarditis and aortic dilatation and dissection.[54]

Rheumatic fever is a systemic inflammatory diseases, affecting the cardiovascular system and the pericardium as part of the mucosal membranes, but also the brain, skin and joints. The clinical manifestation of the disease in the cardiovascular system is mainly in the form of an infection of the heart. The aortic and mitral valve are the locations primarily affected. The infection of the aortic valve leads to AS and consequential therapy in this patient group.[31]

Bicuspid aortic valves are present in 0.6%-2% of the general population.[54] Rheumatic fever in developed countries is estimated at 0.01%. However, in developing countries, this number could triple.[147]

Diagnosis of Aortic Stenosis

Echocardiography is the most common imaging method used for diagnosis of AS. It enables assessment of the aortic valve, severity of stenosis and motion of the myocardial wall. Furthermore, measurements of the thickening of the left ventricular wall and the outflow are done. Doppler signals enable the measurement of the peak jet velocity over the valve, an important measure for severity of AS.[150, 13]

Echocardiography is a low-cost, noninvasive and fast method for the diagnosis of AS. However, it has some drawbacks. The image quality is operater-dependent and also varies with different patient characterics, e.g. BMI. There is a growing interest in the use of other imaging modalities for classification of AS. Cardiac Magnetic Resonance (CMR) gives an accurate assessment of the LV volume, has the ability to monitor flow and enables a more detailed characterization of the myocardium with late-gadolinium enhancement imaging and T1-mapping. However, there are disadvantages of Magnetic Resonance Imaging (MRI); it is less comfortable for the patient due to breath-holding sequences, not all patients are eligible for MRI-scanning and it is more expensive and time-consuming than echocardiography. Computed Tomography (CT)-imaging enables an accurate quantification of calcification of the aortic valve and gives the opportunity to screen for coronary disease, accompanying AS. The main drawback of CT is the use of ionizing radiation.[134]

Consequences of Aortic Stenosis

Aortic valve stenosis usually does not account for symptoms in patients; the effects of the AS do. AS has a long latent asymptomatic period of multiple years, in which the calcification steadily worsens. The onset of symptoms is an indication for failure of the left ventricular (LV) compensatory mechanisms, that have been in effect during development of the AS and worsening function of the aortic valve. During the thickening of the aortic valve, the mobility of the cusps decreases and the valve inlet loses its ability to open and close properly. This creates a backflow of blood from the aorta into the LV, leading to a pressure overload that causes increased wall stress. The first compensatory mechanism is dilation of the ventricle, followed by graduate wall thickening through myocyte hypertrophy, which reduces the wall stress to standard ranges.[48] The thickened myocardial wall requires an increased amount of oxygen. However, coronary flow reserve is reduced in LV hypertrophy through numerous causes; microvascular dysfunction, low coronary perfusion pressure, increased extravascular compressive forces and reduced diastolic perfusion time. All these factors decrease the blood flow and thus the flow of oxygen into the cardiac muscle, leading to ischemia of the myocardial cells followed by interstitial fibrosis.[157, 90] When these mechanisms start to fail, patients will start to experience symptoms. A schematic overview of the mechanism is shown in figure 1.2.[122]

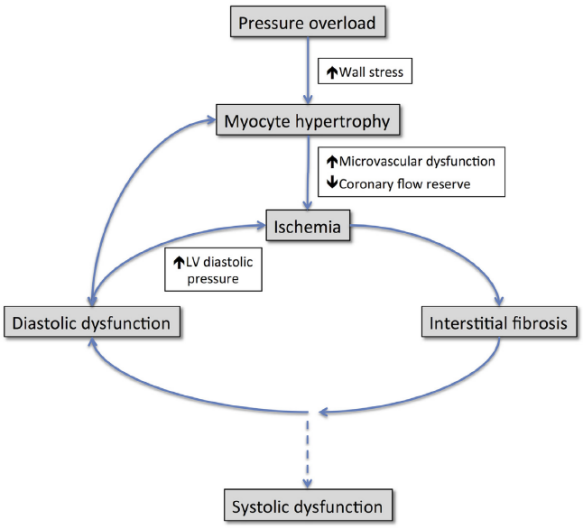


Fig. 1.2: Graphical representation of left ventricular remodelling occuring in patients with AS. *Image derived from Rassi et al., 2014 [122]*

Despite the clear mechanism causing failure of the LV in AS, research has indicated that the degree of left ventricular hypertrophy is only weakly correlated to the severity of valve obstruction. There are other factors that are assumed to have a higher influence on the degree of hypertrophy; age, male sex and obesity.[48, 74] Furthermore, Chin[29] has developed a clinical risk score, comprising five clinical variables, that predict adverse outcomes in AS. Table 1.1 shows the different clinical and imaging variables that predict outcome in patients with AS.

Variable	Description	Effect	Study
Age	Age of the patient	Association with midwall fibrosis Predicts adverse outcomes in aysmptomatic AS-patients.	Chin, 2016 [29]
Gender	Gender of the patient		
V_max	Peak aortic jet velocity		
High-sensitivity Troponin-I concentration	Blood concentration of cardiac troponin (ng/L)		
Electrocardiographic strain pattern	ST-depression and T-wave inversion in lateral leads		
Soluble urokinase plasminogen activator receptor (suPAR)	Inflammatory marker associated with subclinical cardiovascular damage and cardiovascular events.	Association with ischemic cardiovascular events and cardiovascular and all-cause mortality.	Hodges, 2016 [68]
Increase in aortic-jet velocity	Change of aortic jet velocity reflects the presence of moderate or severe valvular calcification.	Low 2-year survival rate.	Rosenhek, 2000 [126]
Midwall fibrosis	Enhancement on LGE-images in the middle section of the myocardium	Predictor of mortality in patients with moderate and severe AS	Dweck, 2011 [42]
Low-flow, low-gradient	Preserved ejection fraction (>50%) with a low-flow (<35 ml/m^2) and low gradient (<40mmHg) over the aortic valve.	Associated with an increased risk of cardiovascular mortality and hospital admission.	Gonzalez Gomez, 2017 [59]

Tab. 1.1: Variables that affect the prognosis of AS-patients.

Aortic Stenosis Therapy

No medicinal therapies exist for AS. Treatment of AS is done with balloon dilatation, Surgical Aortic Valve Replacement (SAVR) or Transcatheter Aortic Valve Implantation (TAVI). The latter two are the preferred choices of treatment.

Balloon valvuloplasty is the widening of the aortic valve inlet by increasing the size of a balloon. This method is used to relieve patients that are not eligible for surgery or to bridge the time towards surgery. The benefit is only temporarily, because the pathologic condition of the aortic valve does not change. Therefore, survival rate remains unchanged.[21]

In SAVR and TAVI an aortic valve is implanted. In the surgical procedure this is done by removal of the old valve and implantation of a mechanical or bioprosthetic valve. In TAVI the inborn aortic valve remains in place, but a replacement valve is positioned transcutaneously at the inlet of the aortic valve and dilated to move aside the stenosed aortic valve.[71]

TAVI is a less invasive intervention than SAVR. It is therefore recommended for patients at intermediate- or high-risk for surgery.[150, 92] Both interventions have different patterns of complications. SAVR has a higher risk of atrial fibrillation, major bleeding, transfusion requirements and acute kidney injury. TAVI complications were primarily the need for pacemaker implantation, higher rates of oartic regurgitation and vascular complications.[135, 27, 92, 123].

With respect to death or disabling stroke, both TAVI and SAVR have a similar risk; ranging from 13% [123] to 20% [92] at a 2-year follow-up in patients at an intermediate-risk for surgery. Patients at high-risk for surgery have an increased risk of death from a cardiac cause, approximately 13% at 1-year follow-up.[135] Currently, TAVI is more widely implemented for intermediate- and high-risk patients. It has shown that mortality is similar and sometimes lower than in /glssavr.[145, 79]

1.2 Imaging in Aortic Stenosis

In the previous section, a short overview was given of the use of imaging modalities for diagnosing AS. This section gives a more elaborate overview of the different imaging modalities and the publications that studied the use of the different imaging methods in AS research. The last part of this section elaborates on CMR imaging and specifically how late enhancement images are acquired; the sequence used for imaging analysis in the consecutive studies of this thesis.

1.2.1 Imaging modalities in Aortic Stenosis: Echocardiography, CT and PET

As previously mentioned, echocardiography is the primary imaging modality for the diagnosis of AS. Beside imaging opportunities, it also gives the opportunity to evaluate flow and gradient across the aortic valve. Flow refers to the ejection of blood from the heart. The pressure gradient gives the difference in pressure between the LV and the aorta, caused by obstruction of the outlet through the calcified aortic valve. [139] This information can be used to determine the severity of the AS and the timing of therapeutic intervention. Analysis of flow and pressure gradient have been a topic of interest lately.

A healthy aortic valve has a low pressure gradient and a low jet velocity (flow) with a normal Aortic Valve Area (AVA) (2.5-4.5 cm^2). After the onset of calcification of the aortic valve, the aortic valve area decreases, leading to an increase in pressure gradient and aortic jet velocity. However, there is a subgroup of severe AS-patients that has a preserved Left Ventricular Ejection Fraction (LVEF), a low flow and pressure gradient. Only the reduced aortic valve area ($<1cm^2$) indicates severe AS. This indicates no need for valve surgery. However, it is this group that might benefit from early valve replacement due to an increased mortality risk. Research has shown that in these patients the analysis of flow and gradient is important. Bavishi et al. [14] reviewed the available literature in a meta-analysis to determine the influence of gradient and flow in patients with severe AS. This analysis concluded that patients with a low flow and a low gradient AS have a higher mortality than patients with a normal flow and a low gradient. This was confirmed by Gonzalez-Gomez.[59] It is challenging to make a distinction between severe and non-severe AS in low-flow low-gradient AS with a preserved LVEF and to determine the right clinical pathway.[13] Aortic Valve Replacement (AVR) shows improved all-cause mortality in the low-flow low-gradient patient group, but no clear improvement after AVR was shown in the normal-flow low-gradient patient group.[14] This is an important example of how details from echocardiography can be used in patient management of AS.

CT uses the degree of attenuation of electromagnetic radiation to image different structures of the human body. It has a high spatial resolution and has therefore gained increasing interest in the assessment of the severity of AS, e.g. for the measurement of aortic valve area in comparison to the reference standard of measurement with echocardiography.[30] CT also enables the examination of Coronary Artery Disease (CAD). It has been demonstrated that CAD increases the procedural risk of AVR.[111] At last, CT is increasingly used as a work-up tool prior to AVR.[18]

PET-imaging of the heart in patients with aortic valve stenosis can be used in combination with CT to provide insight in the inflammation status of the aortic valve. The combination with CT gives an accurate overview of the specific anatomic

location of biochemical processes.[43] Research showed that the activity in the aortic valve increases with the severity of the AS confirming the hypothesis that inflammation processes are present in AS.[45]

1.2.2 The use of Cardiac Magnetic Resonance in Aortic Stenosis

The use of magnetic resonance imaging in the evaluation of AS has gained increasing interest. The application of CMR is four-fold; it enables the assessment of valve morphology (1), valve function (2), left ventricular function (3) and aortic disease (4).[127] CMR research in patients with AS is used for the reproducible measurements of cardiac volumes, function and myocardial mass.[137]

Cardiac Magnetic Resonance

CMR imaging contains all the magnetic resonance imaging of the heart. In this research we have focussed on the enhancement (LGE) images from the CMR-imaging sequences, obtained in patients with AS. Enhancement images are made with the help of the intravenous agent Gadolinium. This agent shortens T1-time and distributes differently within viable and non-viable myocardial tissue. Sequences to display late enhancement are a Fast-spin echo (FSE) technique, or a gradient-echo, also called a turbo-fast low-angle shot technique (FLASH). The latter group of sequences is used for the images in this research.

Gadolinium shortens both T1- and T2-time of the tissue. It is administered as an intravenous bolus. After distribution and accumulation in the myocardial tissue, the heart can be imagined with a T1-sequence. Especially at low dose gadolinium, the shortening in T1-time is dominant over the T2-shortening. This gives an hyperintense area on the acquired image.[20]

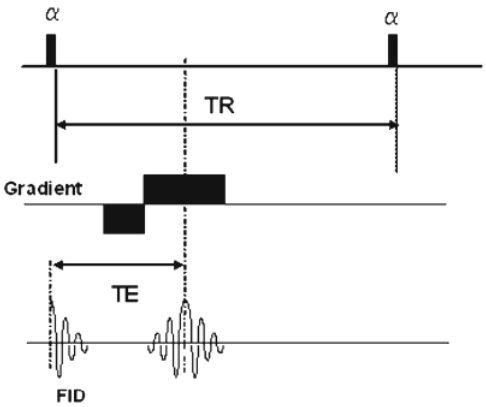


Fig. 1.3: Example of a Gradient-Echo sequence with signal intensities in the different directions. The definition of the T_e and T_r is shown in this image. *FID* is free induction decay; the signal after excitation.

Gradient-Echo Sequences are one of the large families in the sequences used for MRI. Gradient-echo is currently the most used form in CMR imaging, being less prone to

motion-artefacts than spin-echo. Figure 1.3 displays schematically how a gradient-echo sequence works. It starts with a slice-selective RF-pulse. This pulse produces a 10°to 90°rotation angle to align protons. Secondly, a phase-encoding gradient is applied simultaneously with a dephasing frequency-encoding gradient. This causes an accelerated dephasing gradient in the imaged object. The last step is a frequency-encoding gradient for the acquisition of the signal, which causes an echo in the signal. This gradient refocuses the accelerated dephasing, enabling signal acquisition at T_e; the echotime (figure 1.3). T_r represent the repetition time at which this sequence is repeated.[20] The last time variable in gradient-echo is the inversion time (T_i), describing the time before the inversion-recovery prepulse. This pulse ensures the hypointensity of the normal myocardium and increases the contrast between normal and pathologic myocardium. The duration of the inversion time is determined by the radiologic technician, based upon images of the myocardium with different T_i-times.[42]

1.3 Data Analysis

1.3.1 Radiomics

Radiomics has developed over recent years as a "high-througput extraction of large amounts of image features from radiographic images" (cited from [88]). This development has been derived from the hypothesis that medical images contain complementary and interchangeable information with respect to other sources of patient information; histologic evaluation of biopsies, analysis of blood samples, but also demographic and genetic characteristics.[88] Statistics of large radiomics datasets facilitate clinical decision-making.[57]

The Radiomics approach consists of four consecutive steps; imaging (1), segmentation of the region of interest (2), feature extraction (3) and analysis (4).[88, 86] Feature extraction includes the extraction of multiple groups of features from the region of interest. Four groups of features were identified in the Radiomics approach; first order intensity statistics (1), shape and size based features (2), texture features (3) and multi-scale wavelet features(4).[113, 1] These features have been derived from other fields, e.g. computer vision.[63] An explanation on the feature groups and the corresponding mathematical functions are found in the supplements.

Currently, the number of publications that have used radiomic features and machine learning analysis is exploding, predominantly in the field of oncology. Some examples of Radiomics research include; prediction of the risk of distant metastasis in lung cancer [35], discrimination between benign and malignent lesions in the pleura [115] and therapy response prediction in breast cancer [22]. By many clinicians and

data scientists, the use of machine learning in image analysis is regarded as a very promising, growing field.[142]

1.3.2 Statistical Analysis

In this section, we give an overview of feature selection and the used machine learning models. This is the fourth step in the Radiomics workflow. The machine learning pipeline of this research is displayed in the supplements (figure 6.5). For an overview of all the functions used in the R programming language [120], see table 6.1.

Machine learning

The term 'Machine Learning' has been introduced by Arthur Samuel in 1959, when he demonstrated a program, that could play checkers. He described machine learning as *the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning*.[129] To date, there are many different definitions of machine learning. The application in medicine is focussing on the medical diagnosis [77] and the automatic analysis of medical imaging [96], e.g. automatic segmentation of regions of interest, detection of pathology and registration of different imaging modalities.

Machine learning comes in many different forms. In the currently available algorithms, a main distinction can be made between the supervised and the unsupervised learning models. Supervised models are trained with labelled data or samples that have an outcome value. The goal of this type of machine learning is to predict the label or value of a new sample, when given to the model. Unsupervised learning contains raw data points, without a particular label that must be predicted. This type of machine learning seeks to cluster that data points to find general patterns in the data.[25]

Another important aspect in machine learning is the choice of a type of model. Every model has a certain level of complexity. The goal of a good and well-trained machine learning model is its ability to be generalizable and perform well on unseen data. Over- and underfitting are two key aspects that are leading to non-generalizability.

A machine learning model is trained to recognize specific patterns in the data that can be related to the outcome of the sample. Yet, data also contains random noise. This noise is not directly related ot the outcome of the data, but the machine learning model cannot discriminate between noise and data. If a model is trained too well on the data, the model has also learned the characteristics of the noise of each sample, reducing the generalizability of the model to other datasamples. This is called overfitting.[85] This especially occurs in more complex models that are exposed to

a limited amount of data or are too exhaustively trained on a dataset. Overfitting leads to great results in the trainingset, but a poor performance in during validation or testing. Overfitting can sometimes be hard to spot due to a limited availability of data for a testset.

The opposite of overfitting is underfitting. This occurs when a machine learning model is not specific enough and is unable to find patterns in the training data. This will result in a poor performance in both the training and validation. Therefore, it is easier to identify underfitting than overfitting models.

To train a machine learning model, a large dataset is required. In this project, data is obtained from LGE-images from CMR. Two different studies have been performed in this thesis. First, it was assessed if a model trained with radiomic features is able to predict the necessity for AVR during the patient's follow-up period. Secondly, a computer-aided diagnosis model is trained to determine whether it is possible to discrimate patients with and without fibrosis in the LGE-images from each other, based upon radiomic feature values.

As described in the paragraphs above, model selection can be a difficult process, involving trial-and-error. It can be hard to select a model for a specific problem. This research therefore implements three different models; a mathematically simple Generalized Linear Model (GLM), the widely-used Random Forest (RF) and the Support Vector Machine (SVM) with a linear kernel. Implementation of different types of machine learning models enables comparison between the performance of these models.

Generalized Linear Models are a larger class of models that include, among others, logistic and linear regression models. The generalized linear model as implemented in this research utilizes a linear function to calculate a score per sample, based upon the included feature values and a start value. Therefore, an intersection of the linear function with the y-axis is calculated, followed by coefficients for every included feature. With these values a linear function is fitted to the data.[83] Based upon the sample values for specific features, every sample obtains a certain score. This score can be related to the classification into a specific group. Generalized linear models are advantageous, because they have a clear mathematical structure in comparison to the somewhat black-box of other machine learning models.[149]

Random Forests are models that consists of a large sample of decision trees. A decision tree is a tree structure. At every node a decision is made for a certain classification, based on the evaluation of a feature value. Therefore, a threshold for the feature value is defined at each node, determining the pathway through the decision tree. These trees aim to partition the data into smaller and more homogeneous groups of samples. The goal in decision trees is to minimize misclassification at each node. The terminal node of a decision tree produces a vector of class probabilities for

a test sample. These probabilities are based upon the training data. As mentioned, an RF is a collection of decision trees. After training, each tree in the forests predicts a class for a presented sample. The proportion of votes for each class is the predicted probability for that sample.[81] The used RF in this research is the `rf`-model from the `caret`-package.[80] The tuning parameter that is optimized during the tuning process of the model is the number of features that are randomly tried at every node (`mtry`).

The RF model has several advantages, that are applicable to the presented data. First of all, RFs are very suitable for classification problems, due to the voting system of multiple trees. Secondly, an RF is not markedly activated by outliers in the predictor data. This reduces the burden of normalization and data preprocessing. Furthermore, the used RF function has a built-in feature selection tool, that applies weight to every feature during training process. Extreme feature reduction is therefore less required.

Support Vector Machines are a group of models that try to find multidimensional planes in multidimensional data. This type of models views every data point as a point in a d-dimensional feature space, where d represents the number of included features in the model. The goal of an SVM is to find a hyperplane that separates two classes in the data. SVMs can be modelled with linear classification boundaries, but also other kernel functions for non-linear separation can be applied.[84] The main tuning parameter in this model is the 'cost'. This is the penalty on the performance of the model when a sample is wrongly classified. If the cost parameter is high, the model will be trained to avoid wrong classifications. This might result in overfitting. If the cost is set to a smaller number, the model will be less strict. Underfitting is going to occur if the cost is too small.[85] The current study implements a linear SVM from the e1071-package.[99] In this model the only tuning parameter is cost.

One of the problems in working with SVMs is when you have as many features as data samples. In this case, a hyperplane will definitely be found, but it will result in a large overfit on the data. Dimensionality reduction is thus of great importance in SVMs.[24] Therefore, feature selection is implemented before training the linear SVM to include relevant features.

The advantage of SVMs is that these models are intuitively simple and the mathematical fundamentals are easy to understand. However, the linear separability can be problematic for higher-dimensional data. Furthermore, SVMs are fundamentally binary with a hyperplane splitting the data into two sets. It can therefore be hard to analyse samples that are at the edge of two classes.

SVMs are prone to outliers and differing orders of feature value size. It is therefore important to scale the data and to correct the outliers before using feature values as input in the model. Outlier correction is done with the Local Outlier Factor

(LOF).[23] An elaborate explanation of the modification of data for input in the SVM is found in the supplements.

Feature Selection

Many different feature selection algorithms are currently available. A way to group the available models is by how a selector is implemented in the model. There are filters, wrappers or embedded methods. Wrapper methods evaluate multiple subsets of features using procedures that add or remove features to find the optimal combination that maximizes the model performance. Therefore, every subset of features is evaluated in the model. In embedded methods, the training of the model and the feature selection cannot be separated. The feature selection is an incorporated part of the machine learning model.[61] This research employs two methods from the filter-category.

Filters Filter methods are methods that rank features according to their scoring criterion. Within the filter based feature selection methods, two subgroups can be distinguished; univariate and multivariate filters. In univariate methods, the scoring criterion is only based upon the feature's association in the outcome of the model. These filters only look at the relevancy of the feature. Multivariate methods also take into account the redundancy between features. Redundancy is a measure that describes the correlation between features. Multivariate methods determine the scoring criterion according to a weighted sum of feature relevance and redundancy. [112]

The implemented filter in this study is minimum redudancy - maximum relevance (mRMR). *mRMR* combines two important characteristics of features in a feature selection algorithm; maximum relevancy and minimal redundancy. Reducing redundancy extends to the removal of highly correlated features. Maximum relevancy relates to the features that have the largest dependency on the target class.[116]

Filter methods are preferred due to their efficiency. Furthermore, these methods are less prone to overfitting than wrappers and embedded methods.[61]

1.4 Thesis Overview

1.4.1 Motivation

AS is a widely occuring condition in the ageing population in first-world countries; 25% of the population >65 years has a precursor of aortic stenosis and clinically significant aortic stenosis occurs in 1-3% the individuals >70 years of age.[157]

AS leads to left ventricular hypertrophy. If left untreated, the left ventricular hypertrophy leads to ischemia of the myocardial cells followed by interstitial fibrosis.[157, 90] This increases the risk of heart failure, all-cause and cardiac mortality.[42, 122] Without intervention, the 2-year mortality and complications can be as high as 90%.[132] An appropriate risk-analysis of patients that require aortic valve replacement in the near future is therefore required.

Currently, LVEF >50% has been used as the main criterion for AVR in patients with severe asymptomatic AS. The decision tree for the clinical pathway of AS is shown in figure 1.4. However, previous research has shown that LVEF is not an appropriate measure for risk-stratification [114, 126] and that there are other stronger predictors for adverse events in patients with AS; left ventricular fibrosis, identified by enhancement on LGE-images, is a strong predictors of all-cause and cardiovascular disease-related mortality.[12, 42] However, the current guidelines for AS do not include any form of medical imaging in risk-stratification of patients.

Radiomics has shown to enable texture analysis in cardiac images [11]. Furthermore, its prognostic and predictive capabilities are currently widely reviewed[8] in oncology with promising outcomes in the prediction of therapy response [34, 69] and mortality [108].

The presented research has been set-up to broaden the scope of radiomics outside oncology. The studies in this thesis try to find a relation between feature values of Radiomic features and the risk of AVR or adverse events in patients with AS. Radiomic features can be added to the guidelines in place of the LVEF<50%-criterion to improve risk-stratification of AS-patients. Within this study, the clinical risk score, as described by Chin, 2016 [29], is also taken into account in the analysis to determine the performance of this score. Secondly, the relation between features and feature values and the qualitative analysis of fibrosis of the radiologists is studied to improve computer-aided diagnosis of fibrosis in patients with AS. This increases the inter- and intra-patient comparability of the status of the myocardium.

1.4.2 Objectives

This research focusses on the extraction of radiomic features from a patient cohort with AS. These feature values are used to determine if there is a difference in patients with and without AVR during follow-up. The ultimate goal is to enable risk-analysis of patients with AS to improve timing of AVR. This includes the identification of features that are able build a risk signature; a set of features which values can be used to determine the risk score of a patient, in this case for AVR or cardiac events. These features can then be integrated in the flowchart of figure 1.4 for appropriate identification of patients that benefit from early AVR. To enhance the

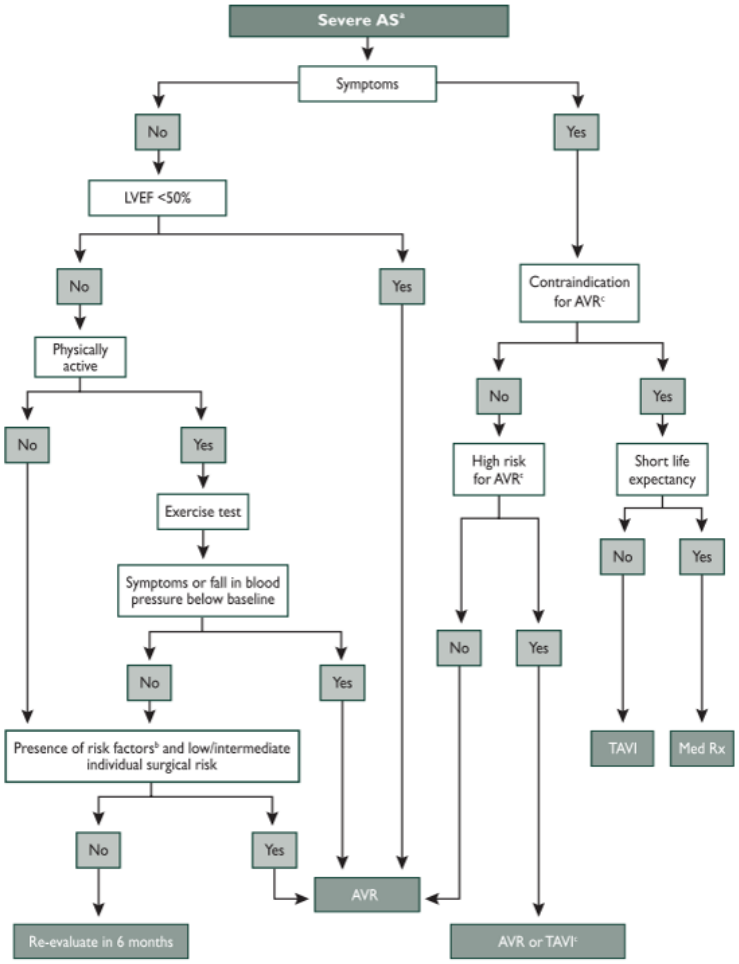


Fig. 1.4: Current management of Aortic Stenosis. BSA = body surface are, LVEDD = left ventricular end-diastolic diameter, Med Rx = medical therapy. *Image derived from the ESC/EACTS Guidelines on management of valvular heart disease, 2012 [150]*

extraction of features, a new method has been developed to rearrange the myocardial segmentation which follows the delineation of myocardial cells.

1.4.3 Research Questions

Main Research Question: What is the added value of the use of radiomic features in analysis of LGE-CMR images of patients with AS?

Sub-questions:

- Do radiomic features have the ability to predict the risk of AVR in the follow-up time? ∼ *Chapter 2*
- Do radiomic features have an added value on top of the current predictors and diagnostic tools of patient outcome in AS? ∼ *Chapter 2*
- Can Radiomic features discriminate subjects with fibrosis from subjects without fibrosis? ∼ *Chapter 3*

- What is the performance of a prognostic model with Radiomic features for AVR on an independent validation cohort? ∼ *Chapter 4*
- What is the performance of a computer-aided diagnosis model with Radiomic features for the identification of fibrosis on an independent validation cohort? ∼ *Chapter 4*

1.4.4 Outline Thesis

The purpose of this chapter (*chapter 1*) was to present the basic background of the different subjects and methods used in this thesis; the pathologic aortic valve, cardiac magnetic resonance, radiomics and the methods used for machine learning and statistical analysis.

The next two chapters (*chapter 2 and 3*) are written as scientific articles. These chapters focus on the performance of different models for classification. *Chapter 2* shows the performance of predictive models in the classification of patients eligible for AVR in the follow-up period. Both radiomics and clinical features are used as the predictors. *Chapter 3* focussed on computer-aided diagnosis. This chapter uses Radiomic features for the automatic classification of LGE-images in fibrosis and non-fibrosis.

Chapter 4 integrates the outcome of the univariate models of chapter 2 and 3. Furthermore, this chapter performs an external validation of the multivariate models in the previous chapters.

Chapter 5 includes an overall conclusion and discussion of the used methods and choices made in this research. It evaluates the effects of different choices on the outcome. Furthermore, a future personal perspective is included on the implementation of machine learning models in modern health care.

2

Radiomics features perform moderately in prediction of the risk of valve surgery in patients with Aortic Stenosis

2.1 Abstract

Introduction Timing of surgery is an important factor to avoid irreversible myocardial damage in patients with AS. Therefore, there is a need for new imaging biomarkers to improve risk stratification and predict outcome. Recently, it has been postulated that AS could be considered a disease of the myocardium, with LGE on CMR-imaging being a predictor of mortality. Radiomics is a novel method for extraction of quantitative features from medical images, relating image features to phenotyping, diagnosis and treatment through predictive modelling. We hypothesized that characterizing fibrosis patterns using a radiomics approach could lead to valuable prognostic information in patients with AS. This could be used as a clinical tool for risk stratification.

Methods 146 patients (age: 68.5, 29% female) with different degrees of AS were included in this analysis; 81 (55%) of patients underwent AVR during the follow-up period. The segmented myocardium was used for calculation of radiomic features. A cylindrical reconstruction was used to calculate texture features and case-specific features. Univariate analysis was done with Concordance Index (CI) as the performance measure. Multivariate analysis included mRMR feature selection and tested 3 models; a cox regression, RF and SVM with a temporal validation and a random permutation (n=1000). A model with clinical features was included to determine performance on prediction of glsavr.

Results 5639 features were extracted from the images. 49 features were found to be prognostic (False Detection Rate (FDR) q-value <0.05, CI>0.6) in univariate analysis. Performance of the models in the temporal validation was close and varying from CI: 0.53 (SVM) until 0.58 (cox regression). The clinical cox regression-model had the best performance (CI: 0.86). The combination of the clinical and radiomic model showed no improvement with relation to the clinical model performance (CI: 0.60).

Discussion To the best of our knowledge, this is the first study to demonstrate the application of radiomics in CMR for prediction of patient outcome. The radiomic features underperformed in comparison to a model with clinical features. However, previous studies showed a correlation between presence of enhancement in CMR and the risk of adverse events, and therefore AVR. It is therefore assumed that the dataset used in this study was too small to reveal this correlation.

2.2 Introduction

AS is the most common valve disease[97] and is a condition that occurs widely in the elderly population. Prevalence of AS in the elderly (>75 years) is approximately 1 in 9, and is expected to increase in the coming decades as the population ages. [109, 106] AS leads to progressive narrowing of the aortic valve inlet, which increases the afterload on the left ventricle. To normalize the wall stress, the left ventricle is remodeled, causing its wall to thicken.[44, 122] When left untreated, the excessive remodeling results in worsening performance of the left ventricle, and a considerable risk of adverse outcome.[52]

During the onset of AS, many patients remain asymptomatic.[114] Even in patients with severe AS, approximately 25% have no symptoms of heart failure.[109] This leads to high levels of mortality in this patient group. Current clinical guidelines prescribe a decreased LVEF as an indication for (AVR) or TAVI in asymptomatic patients.[150] However, previous research shows that this requirement is insufficient to determine which asymptomic patients are at risk for cardiac failure.[114, 126, 29] Patients with AS who have a preserved ejection fraction (LVEF>50%) do benefit in some cases from AVR/TAVI.[47] Nevertheless, measurable changes in LVEF occur only very late in the transition from hypertrophy to heart failure.[29] This points to an urgent need for the clinical guidelines to be reviewed and revised.

Given the above findings, it is important to make a clear selection of patients with AS who will benefit from early AVR/TAVI before their cardiac condition gets worse and results in sudden cardiac death or heart failure. Consequences of left ventricular remodelling due to increased afterload are detectable with the use of CMR. One of the methods in a developmental stadium is T1-mapping. This method assesses the extracellular volume expansion associated with diffuse fibrosis. Following an increased extracellular volume, LGE images are able to detect replacement fibrosis.[29] Enhancement on LGE-images is one of the strongest predictors of all-cause and cardiovascular disease- related mortality.[12] But current guidelines do not include the routine use of delayed contrast-enhanced CMR for risk-stratification in patients with CMR, due to a lack of prospective studies on the prognostic value of CMR imaging in patients with AS.[17]

Our goal is to realize the first steps in establishing the feasibility and added value of including CMR to risk-stratify patients with AS. Recently, the radiomics method has gained ground in the field of pattern recognition in medical images. This method aims to predict patient outcome by the mathematical quantification of distribution of voxel intensities, shape and size of the region of interest, and texture analysis.[58] Radiomics has been very promising in the analysis of tumour tissue on computed tomography (CT) images.[1, 22] Also in the field of cardiology, feature analysis has been based on the Radiomics feature groups, using CMR cine-images [11] and 2D

echocardiography [141]. Yet, these studies have centered primarily on the diagnosis of cardiac disease, not primarily on patient outcome.

In the current study, we have analyzed a CMR-dataset to identify AS-patients at risk for cardiac failure or for AVR in the follow-up. The onset of cardiac failure is usually a precursor for valve replacement. The need for AVR/TAVIis determined by the nature of relative symptoms or when the LVE<50%, as described in the guidelines of the European Cardiac Society.[150] As a result, the patient with AS is eligible for AVR/TAVI, surgery that dramatically improves the survival rate.[110]

The onset of symptoms or a decrease in LVEF, is preceded by, among others, left ventricular hypertrophy and fibrosis formation, which is visible on CMR.[90, 56] We test the hypothesis that radiomic features can discriminate between patients eligible for AVR within the follow-up time of the study, and those who are not. Radiomic features are derived from the LGE-images. This research is done to improve the guidelines for the management of valve disease, as shown in figure 1.4. Therefore, patients can be selected at an early stage for AVR/TAVI to avoid further deterioration of cardiac function.

2.3 Methods

2.3.1 Study Population

The study population used in this research has been obtained from the Edinburgh Heart Centre (clinicaltrials.gov, identifier: NCT01755936). The selected patient cohort used for this research consists of 166 patients with AS, ranging from mild to severe, based upon the peak aortic jet velocity (V_{max}) and the AVA (AHA/ACC Guidelines[105], ESC Guidelines[150, 13]). Start of the enrolment was January, 2012. Patients were excluded from the trial if they had other significant valvular heart disease, heart failure or infection, significant comorbidities, cardiomyopathies or contraindications to CMR imaging.

The obtained CMR imaging data had to include an LGE short-axis imaging sequence. Furthermore, a check for the availability of other clinical parameters was performed. After review of the data, 20 subject were excluded from the study population. Figure 2.1 shows the distribution of the included subjects and the final study population.

2.3.2 Clinical Outcome

The primary clinical outcome in this study was AVR. This included either a trans-catheter or a surgical procedure. Time to AVR was defined as the number of days between the baseline CMR-scan and the date of surgery. Secondary outcome was overall survival of patients. This outcome was defined as the days between the

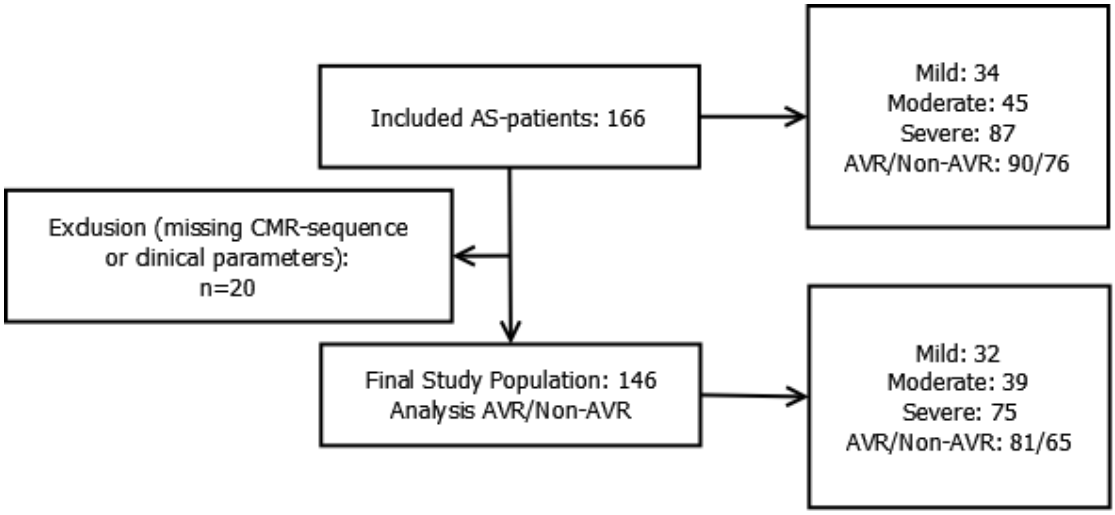


Fig. 2.1: Overview of the study population in this research

baseline scan and the last day of follow-up, the day of death or the pull-out date. Patients were followed up until 5 years after the start date of the study. This resulted in two groups of patients; patients with and without an outcome in the follow-up time combined with a time-frame of a definite number of days of follow-up.

2.3.3 MR Image Acquisition, Segmentation and Reconstruction

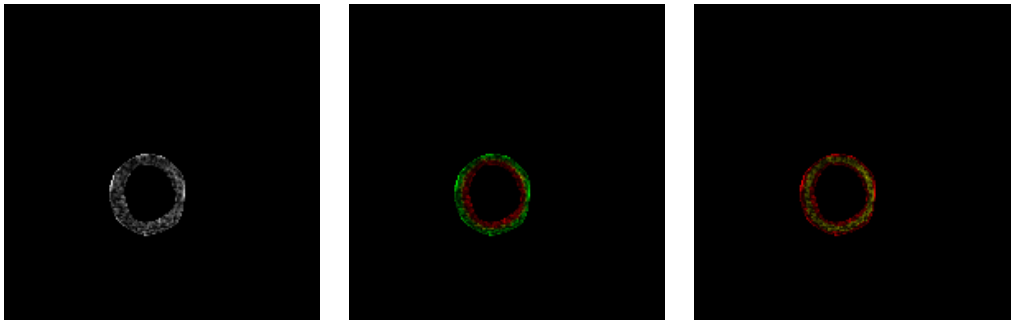
MRI of the subjects was acquired with a 3T scanner (MAGNETOM Verio, Siemens AG, Healthcare Sector, Germany). The LGE images are obtained 15 min following a 0.1 mmol/kg infusion of gadobutrol (Gadovist/Gadavist, Bayer Pharma AG, Berlin, Germany). Two approaches were used for the images; an inversion recovery fast gradient-echo sequence and a phase-sensitivity inversion recovery sequence. Both were obtained in two phase-encoding directions to differentiate true LGE from artefacts. The inversion time was optimized to achieve satisfactory nulling of the myocardium for the inversion-recovery images. Pixel size varies between 1.36*1.36mm and 1.95*1.95mm, with a slice thickness of 8mm and 2mm gap.[29]

The images were reconstructed along the short-axis. After selection of the correct slices, these were stacked in the right order with Matlab (Matlab R2016b, The Mathworks, Inc., Natick, Massachusetts, United States). Finally, the stacked slices were converted into an .mhd-file in Mevislab (Mevislab Version 2.8, MeVis Medical Solutions AG, Bremen, Germany) to facilitate segmentation of the myocardial wall of the left ventricle in this software package by the researcher. The outer borders of the segmentation were removed to minimize partial volume effects. This segmentation was checked and discussed with a cardiologist and a professor in cardioradiology.

Methods for Image Reconstruction

Different image reconstructions were used in this study to create a large variety of modified copies of the original image. Features were extracted from these images. Table 2.1 shows the different methods of reconstruction and the corresponding variations for this method of reconstruction. This section gives an overview of this table in a textual form.

Segmentation Previous research has described a higher probability of events when midwall enhancement[42] or subendocardial enhancement[152] is present. Therefore, the myocardium was split into 2 different images for analysis; a stack of medical images with a separate endo- and epicardium and a stack where the midmycoardium was extracted from the image (figure 2.2). Feature calculation is performed separately on every part of the myocardium.



(a) Original segmentation **(b)** Epi- and Endomyocardium. **(c)** Midmyocardium

Fig. 2.2: Different segmentations and reconstructions of the myocardium: Original segmentation (a), segmentation of the endo- (red) and epimyocardium (green) (b) and segmentation of the midmyocardium (yellow) (c).

Voxel selection Enhancement classification is a subjective judgement by the radiologist. It does not have a truly objective measure. This complicates inter- and intrapatient comparison.[87] Different measures have been proposed to quantify enhancement. One group focusses on the selection of voxels within a specific range of gray values.[51] The analysis presented in this thesis implemented this method. Four different voxel selections were made; ranges from the mean voxel value + $[\frac{1}{4}\sigma, \frac{1}{2}\sigma, 1\sigma, 2\sigma]$ until the maximum voxel value were used to select voxels for an image reconstruction.[115]

Reconstruction of segmentation For this research the first order statistics and the shape- and size-based features were calculated from the segmented myocardium.[2] To integrate the shape of the myocardium into the acquisition of texture features, a cylindrical reconstruction of the myocardium was executed in Matlab. This reconstruction represents the myocardium with the radius on the horizontal axis and the circular angle on the vertical axis, resulting in a horizontal band across the

image. An elaborate explanation of the cylindrical reconstruction is found in the supplements. An example of this reconstruction is shown in figure 2.3.

Filters Different image filters have been implemented in the PyRadiomics-package; the python package used for extraction of feature values from the images. The available image filters are applied to the original segmentation and subsegmentations, but not to the cylindrical reconstruction. An elaborate explanation of the different image filters can be found in the Supplements.

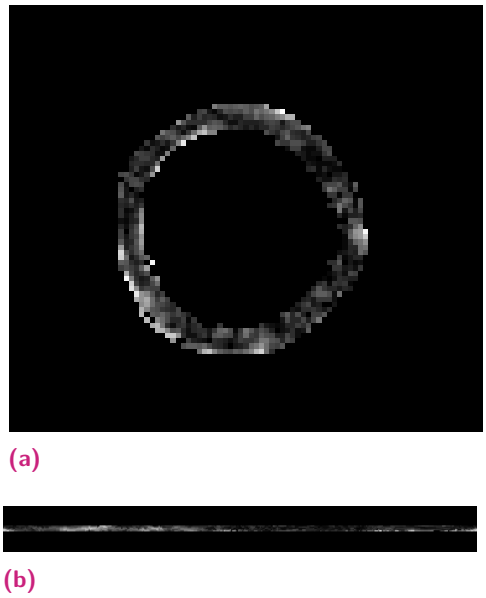


Fig. 2.3: Original segmentation (a) and the cylindrical reconstruction (b) of one slice of the LGE-image sequence.

Normalization Images have been normalized according to the histogram where the mean gray level (μ) is set to 0 with a standard deviation (σ) of 1. All voxel values are then changed to fit within a μ-gray level$\pm 3\sigma$-range. Therefore, all voxels with a gray-value outside this range are changed to the minimum or maximum value of this range. This image is handled as the original image and features were calculated from the original image and the cylindrical reconstruction.

Group	**Classes**				
Segmentation	Original	Epi	Mid	Endo	
Voxel Selection	$> \mu + \frac{1}{4}\sigma$	$> \mu + \frac{1}{2}\sigma$	$> \mu + 1\sigma$	$> \mu + 2\sigma$	
Reconstruction	Original	Cylindrical Reconstruction			
Image Filters	Square	Square Root	Logarithm	Exponential	Wavelet
Normalization	Original	+/- 3 σ			

Tab. 2.1: Image variables that are used and adjusted in this study as explained in section 2.3.3. (μ is the mean gray value of scan of one subject, σ is the standard deviation

2.3.4 Radiomics and Case-specific features

Radiomic features have the ability to quantitatively analyze a region of interest by calculating a large number of features. The feature calculation was performed in PyRadiomics, an open source python package for the extraction of Radiomics

features.[60] Segmentations of the myocardium were converted to an nrrd-fileformat in Matlab, before being imported into the python environment for feature extraction. An elaborate description of the obtained radiomics features is described in the supplements.

Case-specific features were designed for the purpose of this study. The six features in this set are described in table 2.2. These features were calculated in Matlab. Texture features on the cylindrical reconstruction of the image were also calculated in Matlab. In the calculation of texture features on the cylindrical reconstruction, the original location of a voxel was taken into account. Therefore, multiple counts of the same voxel, and thus overestimation of feature values, is avoided.

Features	**Description**
Mean Thickness + StD	Mean thickness of the myocardium in all slices + the standard deviation from this thickness.
Mean Difference to Midline + StD	Mean difference to the calculated midline of the myocardium + the standard deviation. (*e.g. An ellipsoid heart has a larger mean difference to the midline than a cylindrical left ventricle*)
Minimal Thickness	The minimal thickness of the myocardium in one subject.
Maximum Thickness	The maximum thickness of the myocardium in one subject.

Tab. 2.2: A description of the case-specific features of the myocardium added to the Radiomics analysis.

2.3.5 Clinical features

The current clinical indicator for AVR is LVEF.[150] Previous research showed that this criterion is not sufficient for classifying patients that might benefit from early AVR.[114, 126, 29] To test if LVEF at baseline has predictive value in this classification problem, LVEF is taken into account in a model with clinical variables.

Consequently, other studies have tried to find better predictors for AVR. Chin et al. [29] proposed a clinical risk score for myocardial fibrosis. This risk score predicts adverse outcomes in AS. It includes 5 variables as displayed in table 1.1; age, gender, peak aortic jet velocity, high-sensitivity troponin-I concentration and electrocardiographic strain pattern. With these clinical variables, patients at high-risk for adverse events can be identified. To evaluate the performance of these variables, they were implemented in this study in a clinical model.

2.3.6 Feature Ranking and Selection

Feature ranking for the selection of the most important features was performed with mRMR algorithm implemented in the mRMRe package (v2.0.5) [38, 37] in R. mRMR is a filter selection method, applied before modelling of the learning method.

This filter uses the output variable to rank the features by maximizing the mutual information with the outcome variable (= Maximum Relevance). Simultaneously, it is checking the mutual information with the already ranked features, thus minimizing the redundancy.[39]

Before feature ranking and selection, feature space was reduced by removing features with a near zero variance, followed by a removal of features that were highly correlated using Spearman's correlation factor (>0.9). Both functions come from the caret-packag.[70]

2.3.7 Statistical Analysis

The statistical analysis consists of a univariate and a multivariate analysis. Subgroups of the patient population were made according to the severity of AS. A temporal split accounted for a training and validation cohort in each subgroup. The combination of these groups gave the final training (n=109, 75%) and validation (n = 37, 25%) dataset.

Univariate analysis evaluated the prognostic performance of the predictors with the CI-measure. This index is equal to the Area Under the Curve (AUC), derived from the Receiver-Operating Characteristic (ROC).[62] The CI estimates the probability that the subject with the higher prognostic score will have an event in comparison with the subject with the lower prognostic score. It is a measure to quantify the discriminatory power of a distinction in prediction.[98] The CI takes survival times into account.[64] For the univariate analysis the survcomp package (v1.26.0)[101] was used. To correct for the multiple comparisons made in the analysis, the p-value was adjusted to the q-value by the FDR.[15] The univariate analysis was performed on the complete dataset of training and validation.

The multivariate analysis was performed by the comparison of a cox regression model from the survcomp package (v1.26.0) [101], an RF-model from the randomForest-package (v2.41-2) [93] and an SVM from the e1071-package(v1.6-8) [99]. Based upon the mRMR-ranking, features were added one-by-one to the cox regression model to identify the optimal number of features for this model. A random permutation (n=200) was done on the training set to determine the optimal number of included features.. In this random permutation 80% of the training data was used for training of the model and 20% for validation of the included number of features. Once the average CI dropped, the included set of features was retained and implemented in the final regression model. TheRF and SVM used the 50 highest-ranking features from the mRMR evaluation as predictors in the models.

To evaluate the validity of the defined test set, a montecarlo analysis (n=1000) is performed on the complete dataset (n=146). In this random sampling the full dataset (training and validation) was taken into consideration. Each iteration 75%

	All (n=146)	Training (n=109)	Test (n=37)	P-value
Age [years]	68.5 (28 - 92)	69 (31 - 92)	66 (28 - 86)	0.89
Gender [M/F]	103/43 (71%/29%)	76/33	27/10	0.87
Severity AS (mild/moderate/severe)	32/40/74	24/29/56	8/11/18	0.93
Follow-up [months]	28.7 (1-62.8)	28.5 (1-62.8)	25.1 (1 - 47.1)	0.24
AVR	81 (55%)	65	16	0.123
V_max [m/s ^ 2]	3.83 (1.9 - 8)	3.84 (1.9 - 8)	3.80 (2.37 - 5.1)	0.48
High Sensitivity Troponin 1	13.65 (0.4 - 126.1)	13.31 (0.4 - 109.3)	14.64 (1 - 126.1)	0.23
ECG Strain	21 (14%)	17 (15%)	4 (11%)	0.65
LVEF	57.83% (38.64 - 84.62)	57.16 % (38.64 - 77.42)	59.8 (41.18 - 84.62)	0.51

Tab. 2.3: This table shows the patient details. P-values have been obtained with a Wilcoxon rank sum test for continuous variables. A Chi-square test has been performed to calculate the P-values for categorical variables.

of the samples were used for model training and 25% for model evaluation. All statistical analyses were performed in R (v3.4.2).[120]

2.4 Results

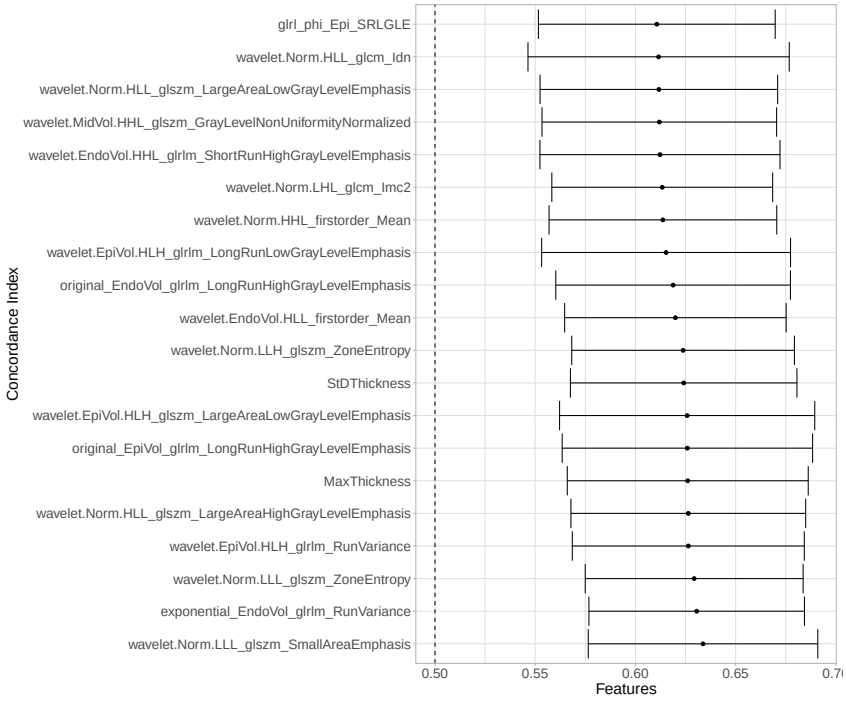
2.4.1 Patient Selection

Table 2.3 displays the patient details of the full population and the split between training and validation set. The majority of the patients was male (71%) with a mean age of 68.5 years. The follow-up time was on average 28.7 months, yet ranged between 1 and 1884 days. No significant different variables were found comparing the stratified random split between training and validation set.

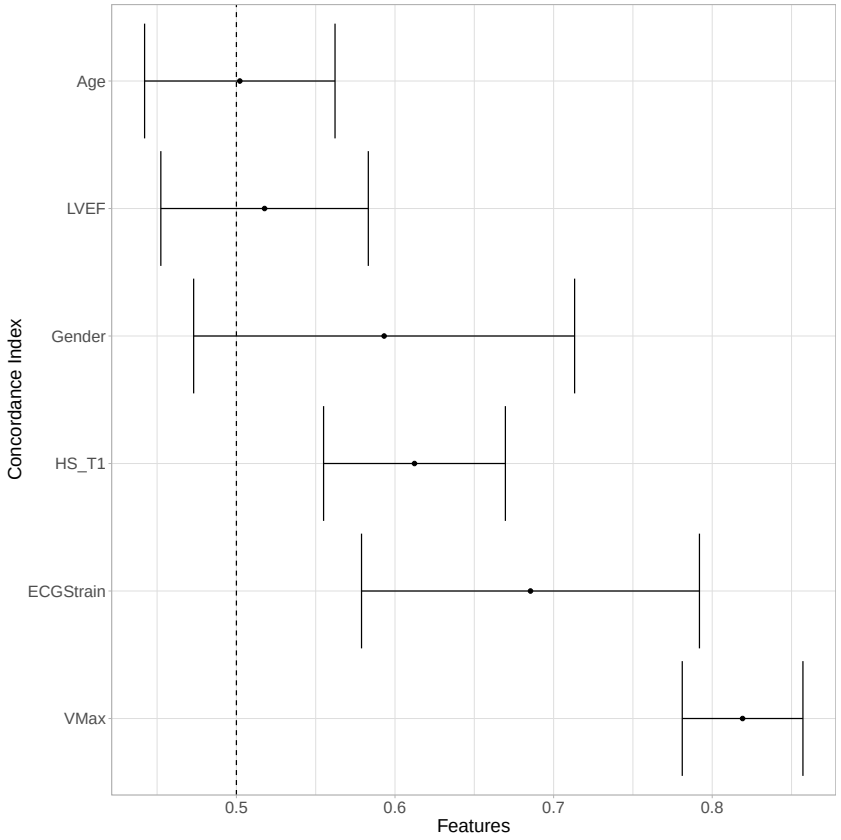
2.4.2 Univariate Analysis

5639 Radiomic features were calculated from the medical images. After removal of features with a near zero variance and highly correlated features, 1193 features remained. In the univariate analysis, 49 features were somewhat related to the risk of AVR (CI>0.6 and FDR p-value<0.05). More information on these features is found in table 6.2 in the supplements. These features were primarily derived from the wavelet-filtered images and consist of mainly texture features.

Figure 2.4 shows the CI of the 20 Radiomics features with the highest CI. This figure also includes the clinical variables as found by Chin [29] that are used for the clinical analysis. LVEF is currently the decisive variable in patients with AS and is therefore also analysed (figure 1.4. Yet, this clinical feature showed to have no prognostic value in the univariate analysis (q-value>0.05 and CI=0.49). The clinical features that have a significant (q-value<0.05) prognostic (CI>0.5) value are peak-aortic jet velocity (V_{max}), EKG-strain and High-sensitivity Troponin-I concentration.



(a) CI and confidence interval of 20 highest scoring radiomic features.



(b) CI and confidence interval of clinical features and LVEF.

Fig. 2.4: Forest plots of the CI and the confidence interval of the 20 best performing radiomic features in univariate analysis (a) and the clinical features (b).

2.4.3 Multivariate Analysis

mRMR feature selection was performed to make a ranking of features for the multivariate cox regression analysis. Based upon the mRMR-ranking, the 50 features with the highest scores were iteratively added to the model to determine the optimal number of features. This optimum is determined by calculation of the mean CI in a 200-fold Monte Carlo analysis on the training set. A graphical review of this result is displayed in figure 6.6 in the supplements.

Two features were selected to make the radiomic signature for risk-analysis of AVR. These features are listed in table 2.4. As can be seen, the features are derived from filtered images; wavelet and normalized square.

For the evaluation of the clinical signature, V_{max} (CI = 0.82), High-sensitivity Troponin-I (CI = 0.59) and electrocardiographic-strain (CI = 0.64) were selected. These features have a significant (FDR q-value$<$0.05) prognostic power for the risk analysis of AVR in the univariate analysis (figure 2.4).

To determine if the test set extracted from the dataset, according to a temporal validation principle, is representative for the complete dataset, a 1000-fold Monte Carlo analysis was performed on all subjects (training and test). This analysis followed a similar pipeline as described for the test set. The results of this random permutation and the final evaluation of the test data are shown in figure 2.5.

As a final step in the analysis of model performance, we compared the model with the clinical features, as identified by Chin [29] with the radiomics model to determine if these models perform better when combined. For this model, the outcome of the models on training set was combined in a new multivariate model. This model was then tested on the test set. This model showed no improvement with respect to the clinical model, but performed better than the radiomics model (CI: 0.60). However, the discrimination between the AVR and non-AVR group was not significant in this classification. This was also the case for the cox regression model with only radiomic features.

2.5 Discussion

In this study, we tested the hypothesis that radiomic features can discriminate between patients eligible for AVR/TAVI within the follow-up time and those who are

Feature	Imagefilter	Power	p-value	q-value (FDR)
Midvol GLCM Inverse Difference	HLH Wavelet	0.55	0.059	0.25
GLSZM Large Area High Gray Level Emphasis	Normalized Square	0.60	0.001	0.027

Tab. 2.4: The selected features for the radiomic signature. The power, p- and q-value were derived from the univariate analysis.

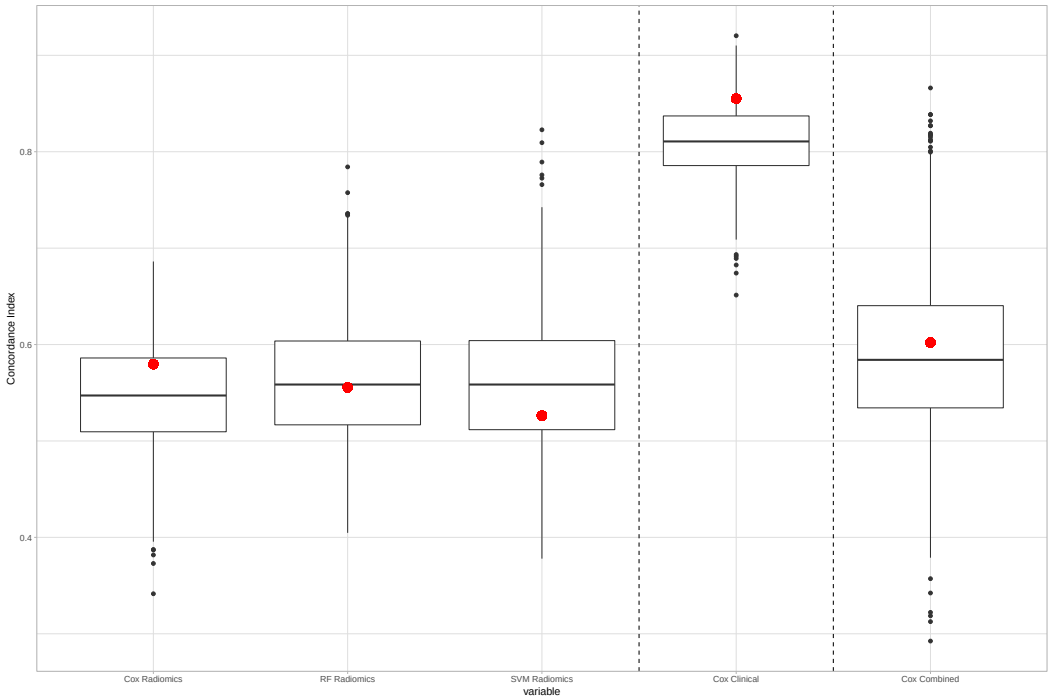


Fig. 2.5: Boxplots showing the performance of the clinical model and the radiomics model in a random permutation (n=1000) on the full dataset. The red dot indicates the performance of the model on the test data, based on training with the training set (Radiomics - Cox: 0.58, RF: 0.55, SVM: 0.53, Clinical - Cox: 0.86, Radiomics + Clinical - Cox: 0.60).

not. The results showed that the discriminative power of the radiomic features in LGE-images is very limited for the identification of patients at risk for valve surgery during 5-year follow-up in the test set. This indicates that radiomics does not have the power to predict adverse events in an AS patient cohort.

The machine learning model including the features identified by Chin [29] for a clinical risk score in patients with AS performs better than inclusion of solitary radiomic features. This result confirms the clinical risk score proposed by Chin. However, also radiomic features represent information about the status of the myocardium of the patient, albeit submediocre. The larger share of features selected in the univariate analysis belongs to the group of texture features. Texture is related to the gray values of the image. This confirms studies showing an adverse prognosis in patients with AS with midwall fibrosis and gadolinium enhancement.[42, 152, 12, 10] However, the CI of the radiomic features is not as high as the CI from the clinical features, indicating a worse performance in classification of subjects in high- and low-risk.

Currently, the selection criterion for AVR is based upon the LVEF in asymptomatic patients, which must be less than 50%. The LVEF at baseline was taken into account in the univariate analysis, to determine if this variable also has prognostic power. Yet, our analysis showed that LVEF at baseline has no value in the risk analysis for

AVR, with a CI of ± 0.5. However, this is not a good representation of the current clinical pathway. In clinical care, patients are monitored during the progression of their AS. As soon as their LVEF dips under 50%, patients will be evaluated for AVR and the option to undergo surgery will be discussed with patients in order to reduce the risk of adverse cardiac events.

Another notable result of this study is that the combined GLM with radiomic features and the GLM with the clinical variables from Chin [29] results in a worse performance than the clinical variables only. It is hard to explain where this result comes from. It is the way that the model interprets the results and puts a larger weight on the score in the radiomics model.

Furthermore, previous research regarding texture analysis in CMR and other imaging modalities have focused on the diagnostics and phenotyping of patients. Also these studies did not always publish promising results. Baessler et al. [11] showed that it is possible to discriminate between patients with and without scar tissue from myocardial infarction in cine-images. However, it was not possible to distinguish large and small infarct areas. Other studies focus only on the analysis of myocardial scar after infarction and not on the myocardium as a whole. [46] Kolossvary et al. [76] focusses on cardiac CT and cites drawbacks to the use of radiomics in cardiology, e.g. the small nature of coronary arteries for radiomic feature extraction. As far as known, this study was one of the first studies in cardiology to use prognostic phenotyping as an outcome for radiomic feature analysis.

Several causes can help to explain the non-discriminative power of radiomic features in the identification of patients vulnerable for AS. First of all, the patient group is relatively small and heterogeneous for the machine learning purposes presented in this study. Secondly, the patients have different degrees of AS and from the data we have obtained it is unclear which patients were symptomatic and asymptomatic. The current guidelines prescribe that symptomatic patients always have an indication for TAVI or AVR, when this will reduce their symptoms. Therefore, this group of patients will always have surgery, although their heart can be in a relatively good condition compared to other asymptomatic patients. To this point, it still remains uncertain to what extent asymptomatic patients can benefit from an early surgery. Furthermore, baseline scans were used in this analysis. It is hypothesized that in progression of the AS, also fibrosis becomes more evident and is better able to predict the need for AVR. Thirdly, the use of CMR imaging, and specifically LGE, can induce uncertainty in the analysis of radiomics features. A certain value for T_r is determined by the radiographer, based upon the relative blackness of the myocardium in comparison with the bloodpool and possible fibrotic tissue. This induces variability among patients. Another variability between patients is their voxel seize. This ranged from 1.37mm until 1.95mm. Segmentations of patients with a smaller voxel size include more voxels. The effect on feature values can be that

some samples have overestimated feature values, when having a small voxel size and others have underestimated feature values with a larger voxel size. A more in-depth analysis to the effect of a differing voxel size can be done with the PyRadiomics-package, that has an option for resampling voxels. The lower spatial resolution of CMR is also a disadvantage in comparison to regularly used CT for Radiomics analysis, which has a higher spatial resolution [95]. However, CT currently lacks the ability to identify fibrosis in the myocardium. Therefore, it is a less suitable method for the purpose proposed in this study.

This study has shown that radiomic features do not have a significant discriminative power in the determination of patients eligible for AVR. Yet, as previously mentioned, fibrosis and enhancement on CMR is a risk factor for patients with AS.[42, 152, 12, 10] In the light of this knowledge, future research should focus on the identification of enhancement with the help of radiomic features. This has several advantages. First of all, it gives the opportunity to characterize the myocardium using quantitative radiological features, instead of the currently used eye-balling and qualitative description performed by radiologists. Secondly, it opens doors towards the comparison of quantitative degree of enhancement in the follow-up of one patient, but also between patients. The quantification of enhancement with radiomics is therefore going to be the focus of the next study in this thesis.

3 Radiomics features discriminate fibrosis from healthy myocardium.

3.1 Abstract

Introduction Fibrosis, specifically in the midwall, is an independent predictor of mortality in patients with AS. It shows up as hyperenhancement on gadolinium-enhanced CMR-images. Fibrosis is currently examined and quantified by a radiologist and does not have a truly objective measure. Radiomics gives the opportunity to characterize tissue and with this predict patient prognosis. This study will focus on the computer-aided diagnosis of LGE-enhanement in a heterogeneous group of patients with AS and control subjects to perform a quantitative assessment of the myocardium in order to discriminate between patients with and without fibrosis.

Methods 186 subjects (age: 65.9, 70% male) were included in this analysis, 57 (31%) had some form of fibrotic enhancement on CMR. The myocardium was manually segmented from the images. This segmentation was used for radiomic feature extraction. Feature selection was done with mRMR feature selection, before modelling with a GLM, RF and an SVM. A temporal validation was done that included 142 subjects (75%) for training and 47 subjects for testing. Clinical features were implemented in a separate clinical model.

Results 5639 features were extracted from the images. 344 features showed a fair to good correlation in univariate analysis (FDR q-value$<$0.05, 0.6$<$AUC$<$1) with the presence of LGE. The models trained with radiomic features (GLM AUC: 0.91, RF AUC: 0.94, SVM AUC: 0.90, showed a higher performance than the model trained with clinical features (AUC: 0.78).

Discussion This study demonstrated that radiomic features have the ability to discriminate between subjects with and without enhancement. Multivariate analysis of LGE leads to a better classification of LGE than univariate analysis of features. Although this research has been applied to patients with AS and a control group with a number of subjects with myocardial infarction, the demonstration of enhancement analysis with radiomics features might be generalizable to other cardiac pathologies. Therefore, this study can be contributing to the ongoing investigations to the role of fibrosis in cardiac disease.

3.2 Introduction

Fibrosis, specifically in the midwall, is an independent predictor of mortality in patients with AS.[42, 121, 152] Fibrosis shows up as hyperenhancement in gadolinium-enhanced CMR images. Hyperenhancement images is an indicator for viability of myocardium in ischemic heart disease and for infiltration of fibrosis in non-ischemic heart disease.[75, 104] The enhancement of this patient group is caused by the pressure overload of the heart, which causes left ventricular hypertrophy.[122] This reduces the coronary flow reserve and therefore causes a decreased blood flow to the myocardial cells leading to ischemia.[90] Several mechanisms have been proposed that account for the increased risk in mortality, among those the magnitude of hypertrophy. The magnitude of left ventricular hypertrophy is associated with an increased risk of sudden cardiac death.[138] However, the mechanisms causing this increased risks remain unclear.[40, 19]

Fibrosis is currently examined and quantified by a radiologist and does not have a truly objective measure.[87] Therefore, quantitative assessment of non-viable myocardial tissue on enhancement images is subjective and raises difficulties for comparison within and between patients. As a result, measures have been proposed for the quantification of infarct and fibrotic tissue, yet focussing primarily on focal enhancement.[32, 51, 4] None of them have found their way yet into standard multi-centre clinical practice.

Radiomics gives the opportunity to characterize human tissue and predict prognosis in patients. The method extracts quantitative image features from a region of interest in a disease of interest.[1] Multiple studies applying radiomics have been published in recent years; among others, these predominantly oncological studies focus on differentiating malignant and benign lesions [115, 16], the staging of cancer [158, 33], phenotyping tumors [156] prediction of therapy response [34, 69] and mortality [108].

Textural image analysis is also a field that has gained interest in cardiology in recent years. Baessler et al. studied the use of cine-images in determining the presence of infarct tissue with textural feature analysis.[11] Sudarshan et al. reviewed the use of feature analysis in the diagnosis of myocardial infarction in echocardiography [141] and another study classified patients with this pathology into high- and low-risk groups.[78] Furthermore, Shah [131] analyzed the heart failure patients with a preserved ejection fraction, resulting in a novel classification of this disease. Oakden-Rayner [108] predicted 5-year mortality on the basis of feature analysis of chest-CT imaging. Feasibility for cardiac CT radiomics analysis was done by Hinzpeter et al.[67] A review on the application of radiomics in cardiac CT has been done by Kolossvary et al. [76] This review comprises the application and drawbacks of Radiomics in cardiac CT scans. This enumeration of publications shows the

upcoming field of investigation in the application of radiomics in cardiology for phenotyping cardiac disease.

The previous chapter of this thesis showed that prognostic evaluation with radiomics features of patients with AS is not accurate enough to discriminate between patients that are eligible for AVR in the follow-up time. The results did not meet the performance of the clinical risk score that has recently been developed by Chin et al.[29] in this relatively small patient group. However, previous studies have shown that fibrosis is an independent predictor of mortality in patients with AS.[42, 121, 152] Therefore, the current study will focus on the detection of LGE-enhanement in a heterogeneous group of patients with AS and control subjects. This classification is categorized as computer-aided diagnosis. It is hypothesized that radiomics features are able to perform a quantitative assessment of the myocardium in LGE CMR in order to discriminate between patients with and without fibrosis.

3.3 Methods

3.3.1 Study Population

The study population of this research has been obtained from the Edinbourgh Heart Centre (clinicaltrials.gov, identifier: NCT01755936). The enrolment of patients started in January, 2012 and consisted of 166 patients with different severities of AS and 48 control subjects. Patients were excluded from the trial if they had other significant valvular heart disease, heart failure or infection, significant comorbidities, cardiomyopathies or contraindications to CMR imaging.

The obtained CMR imaging data had to include an enhancement short-axis sequence. Furthermore, a check for the availability of clinical parameters was performed. These clinical parameters were taken into account in this research,. After review of the data, 25 subject were excluded from the study population. Figure 3.1 shows the distribution of the enrolled study population and the final distribution of subjects in this study.

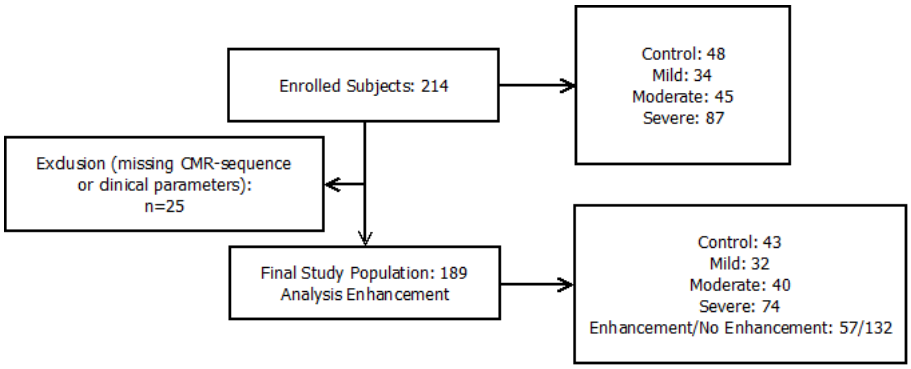


Fig. 3.1: Overview of the study population in this study

3.3.2 Clinical variables and Outcome

The outcome used in this research was the presence of enhancement in the CMR images. This presence has been determined by two independent and experienced operaters.[29] Focal fibrosis, as depicted in enhancement imaging, can be divided into different patterns; infarct and non-infarct. The latter group can again be subdivided into patchy and focal fibrosis.[148] A second group in classification of enhancement is diffuse fibrosis. This type of fibrosis hard to discriminate from normal myocardium with enhancement imaging, because they have similar gray values on CMR. Therefore, this type of fibrosis was not taken into account in this analysis.

Previous studies developed a clinical risk score for the prediction of midwall fibrosis. This clinical risk score was implemented in this study to test the performance in the prediction of all types of enhancement[29]. The clinical risk score consists of five variables; age, gender, peak aortic jet velocity, high-sensitivity Troponin-I concentration and electrocardiograhpic strain pattern. For more information on this risk score, see table 1.1.

3.3.3 MR Image Acquisition and Segmentation

For the details on the image acquisition of the CMR-images, I refer to the similar section in chapter 2. The same image acquisition, reconstruction and segmentation has been used for the data in both chapters.

3.3.4 Radiomics and Case-specific Features

Radiomic features have the ability to quantitatively analyze a region of interest by calculating a large number of features. Segmentations of the myocardium were converted to an nrrd-fileformat in Matlab (Matlab R2016b, The Mathworks, Inc., Natick, Massachusetts, United States). After this conversion, the segmenations were imported into the Python environment for feature extraction. Feature extraction was performed in PyRadiomics, an open source python package for the extraction of Radiomics features.[60] An elaborate description of the obtained radiomic features is described in the supplements.

Furthermore, case-specific features were designed for the purpose of this study. The six features in this set are described in table 2.2. These feature values were calculated from the segmentations in Matlab. Feature extraction from the cylindrical reconstruction was also performed in Matlab, with software designed by the researcher.

3.3.5 Feature Ranking and Selection

Feature ranking for the selection of the most important features was performed with an mRMR algorithm implemented in R from the mRMRe package [38]. mRMR is a filter selection method, applied before modelling of the learning method. This filter utilizes the output variable to rank the features by maximizing the mutual information with the outcome variable (= Maximum Relevance). Similarly, it is checking the mutual information with the already ranked features, thus minimizing the redundancy.[39]

Before feature ranking, selection and data analysis, feature space was reduced by removal of the features with a near zero variance among its sample values and removal of features that were highly correlated using Spearman's correlation factor ($>$0.9).

3.3.6 Statistical Analysis

Univariate and multivariate analysis were performed on the training set of patients. To create a similar training and test set, the data was divided into four categories, based upon the type of subject; patients with mild, moderate and severe AS and a group of controls. Per group a temporal split was performed, appointing 75% (142) of all subjects to training and the remaining 25% (47) to the validation cohort. This method ensured equally heterogeneous groups in training and validation.

An univariate analysis was performed, with the AUC from the ROC as the output measure. For the univariate analysis the survcomp package[101] was used for the analysis of radiomic features. For the analysis of clinical features, the ROC-function from the pROC-package [125] was used. To correct for the multiple comparisons made in the analysis, p-value adjustment was made according to Benjamini's FDR.[15]

A GLM, an RF and an SVM were trained as multivariate models. For the GLM the optimal number of features was determined by forward-selection on the training set. In this selection the features were added one-by-one to the model with a maximum of 50 features. The optimal number of features was determined with a random permutation (n=200) with 80% of the training data used for training and 20% for validation of the included number of features. The mean AUC per number of selected features was calculated over the number of iterations. If the mean AUC of the growing model dropped, the corresponding feature set was retained and used for the final model. The RF and the SVM included the 50 highest-ranking mRMR-features to create the model. To determine the representativeness of the validation set and the overall performance of the model, a random sampling (n=1000) was done with the complete dataset (n=189). In this random sampling, 75% was appointed as training data and 25% as test data.

	All (n=189)	Training (n=142)	Test (n=47)	p-value
Age [years]	65.9 (19-92)	65.6 (19-92)	66.8 (28-86)	0.84
Gender [M]	130 (70%)	94 (66%)	36 (77%)	0.25
Severity AS (mild/moderate/severe)	43/32/40/74	33/24/29/56	10/8/11/18	0.97
LGE	57 (31%)	44 (31%)	13 (28%)	0.80
Infarct	15 (8%)	11 (7%)	4 (8%)	1
V_max [m/s^2]	3.28 (1.01-8)	3.28 (1.1-8)	3.31 (1.01-5.1)	0.76
High Sensitivity Troponin-I [ng/L]	11.47 (0.3-126.1)	11.08 (0.3-109.3)	12.64 (1-126.1)	0.98
ECG Strain	21 (11%)	17 (12%)	4 (8%)	0.70
LVEF [%]	54.68 (25.42-84.62)	56.92(25.42-81.43)	59.96 (41.18-84.62)	0.02

Tab. 3.1: Patient details of included subjects in analysis of LGE-patterns with radiomic features.

All analyses were executed in R.[120]

3.4 Results

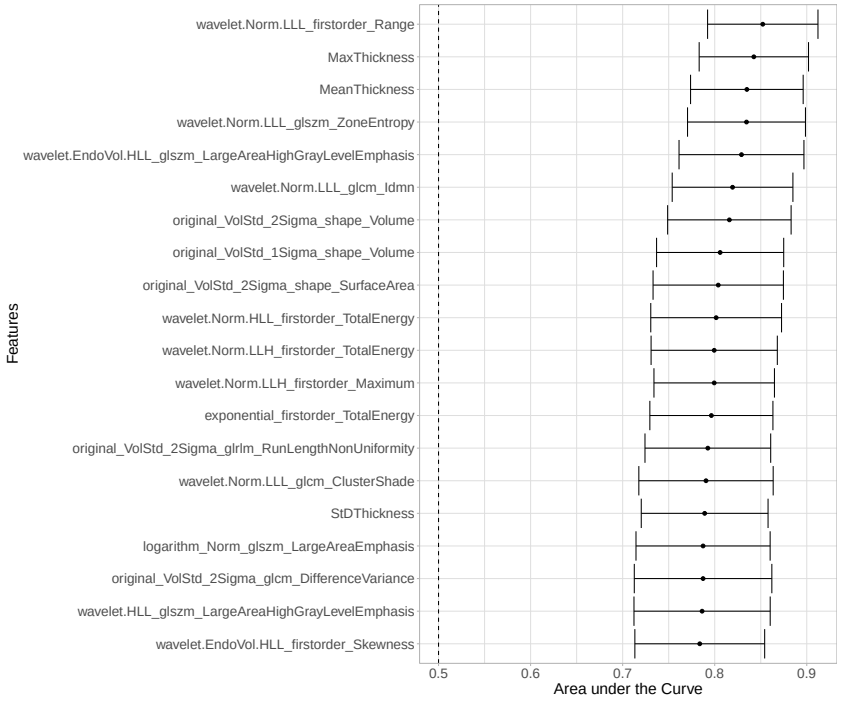
3.4.1 Patient Selection

Table 3.1 shows the patient details of the cohort used in this study. The mean age of the population was 65.9 years and the majority (70%) was male. In the training and validation cohort, respectively 31% and 28% of the subjects had signs of enhancement. LVEF is significantly different between the training and validation cohort.

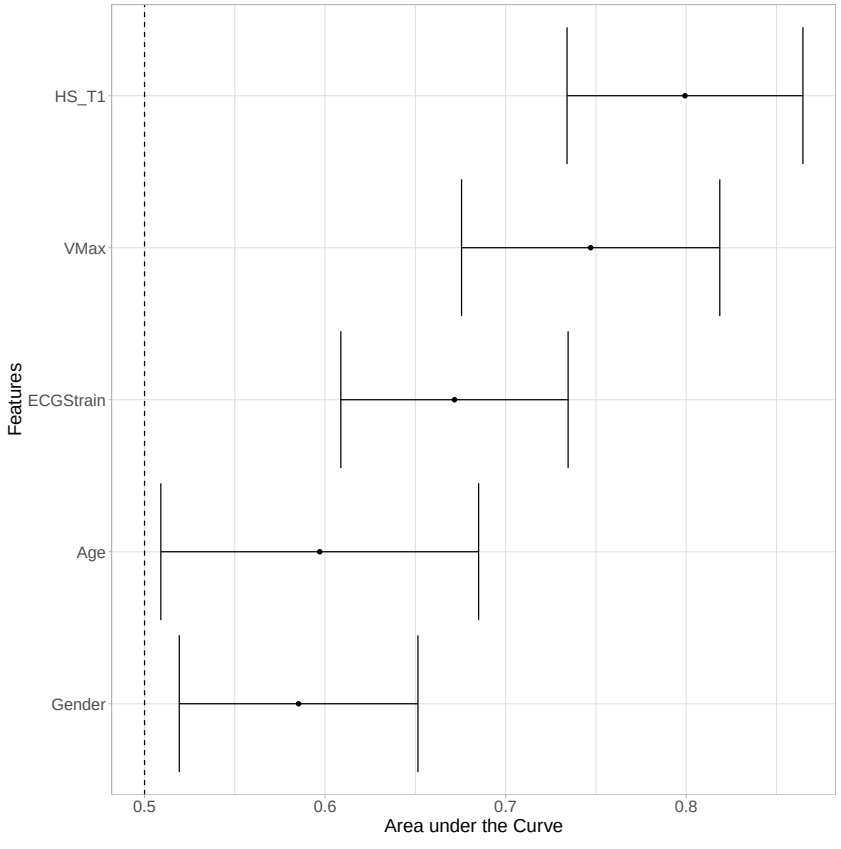
3.4.2 Univariate Analysis

5639 features were calculated from the segmentations of the LGE/CMR images. After removal of the features with a near-zero-variance (168 removed features) and a selection of highly correlated features (4332 features removed according to Spearman's ρ >0.9), 1139 features remained for the final analysis. 344 features showed a fair to good correlation (0.6<AUC<1 and FDR q-value<0.05) with presence of enhancement. The 20 highest scoring features are shown in figure 3.2. This list includes 3 case-specific features, with only 6 (0.1%) case-specific features in the initial dataset. It is noteworthy that wavelet features with three low-passing filters derived from normalized images show a high performance in comparison to the other groups. Table 6.3 in the supplements includes all the features that had an AUC>0.75 and a q-value<0.05 in univariate analysis.

Figure 3.2 also shows the AUC of the clinical features. The feature with the highest score is the high-sensitivity troponin-I. This feature has an AUC of 0.79. All features have a q-value of <0.05, indicating statistical significance for the evaluation of LGE. The univariate analysis was performed on the complete dataset. This included training and test set.



(a)



(b)

Fig. 3.2: Forest plots of the AUC and the confidence interval of the 20 best performing radiomic features (a) and all clinical features (b) in univariate analysis.

3.4.3 Multivariate Analysis

An GLM, an RF and an SVM were used for the classification of myocardial segmentations into LGE and non-LGE groups. The performance measure used for evaluation was AUC. mRMR feature selection was performed to make a ranking for the multivariate analysis, based upon feature values from the training set. The 50 highest-ranking features of this list were iteratively added to the GLM in a repeated random sub-sampling (n=200) set-up with the samples from the training set. The validation set was left out of this procedure to ensure independent validation. A graphical review of the optimal number of features is displayed in figure 6.7 in the supplemental figures. Table 3.2 in the supplements shows the included features and their p- and q-value from univariate analysis. The RF- and SVM-model implemented the 50 highest-ranking features in the model on the training set.

Figure 3.3 shows the performance of the different models on the test set, displayed as the red indication in the graph. Six radiomic features were taken into account for modelling of the GLM. The radiomics RF-model showed the highest performance with an AUC of 0.94 on the test data. A 1000-fold stratified random permutation was performed (75% training- and 25% test-set) on the complete patient cohort. The result of this analysis is also shown in figure 3.3 as the boxplot. This analysis was done to evaluate the representativeness of the temporal validation data as part of the whole dataset.

The GLM with clinical features has a lower performance than the models with radiomic features. The AUC of this model is 0.78.

The final evaluation was the performance of a combined model based upon the the GLM with radiomic features and clinical features to determine if these are complimentary. This model performed slightly (AUC = 0.92) better than the GLM with solely radiomic features. Yet, this performance is not as good as the RF with radiomic features.

Feature	Imagefilter	AUC	p-value	q-value (FDR)
First Order Range	Normalized Wavelet LLL	0.85	1.63E-14	1.86E-11
Mean Thickness	Case-specific	0.83	2.88E-13	8.76E-11
Zone entropy glszm	Normalized Wavelet LLL	0.83	3.08E-13	8.76E-11
Large Area High Gray Level Emphasis	Wavelet HLL Endocard	0.83	7.38E-13	1.68E-10
IDMN glcm	Normalized Wavelet LLL	0.82	3.39E-12	6.43E-10
First Order Maximum	Normalized Wavelet LLH	0.80	6.75E-11	6.41E-9

Tab. 3.2: The selected features for the radiomic signature. The power, p- and q-value were derived from the univariate analysis.

3.5 Discussion

This study has shown that radiomic features and features derived from the thickness of the myocardium have the ability to discriminate between non-enhanced myocardium and enhanced, thus fibrotic, myocardium. When combining multiple features in a machine learning model, the performance is better than evaluating single features in univariate analysis. When comparing the different models and their performance, a better classification is seen in the more complex RF than in the GLM. This result confirms the hypothesis that radiomic features have the ability to give a quantitative assessment of LGE. To our knowledge, this is the first research that evaluates myocardium with a large set of radiomics and radiomics-derived features.

This study contributes to the field of quantitative big data analysis of medical images, specifically focussing on MRI. Although machine learning in medical imaging has long been available in medicine,[77, 142] it has recently gained popularity due to faster computational hardware and availability of digitally stored datasets of medical images.

Texture analysis of cine-CMR has previously been performed and described by Baessler, 2017 [11]. This study showed good discrimination between non-infarct and infarct myocardium on cine-MRI. However, the current study selected other

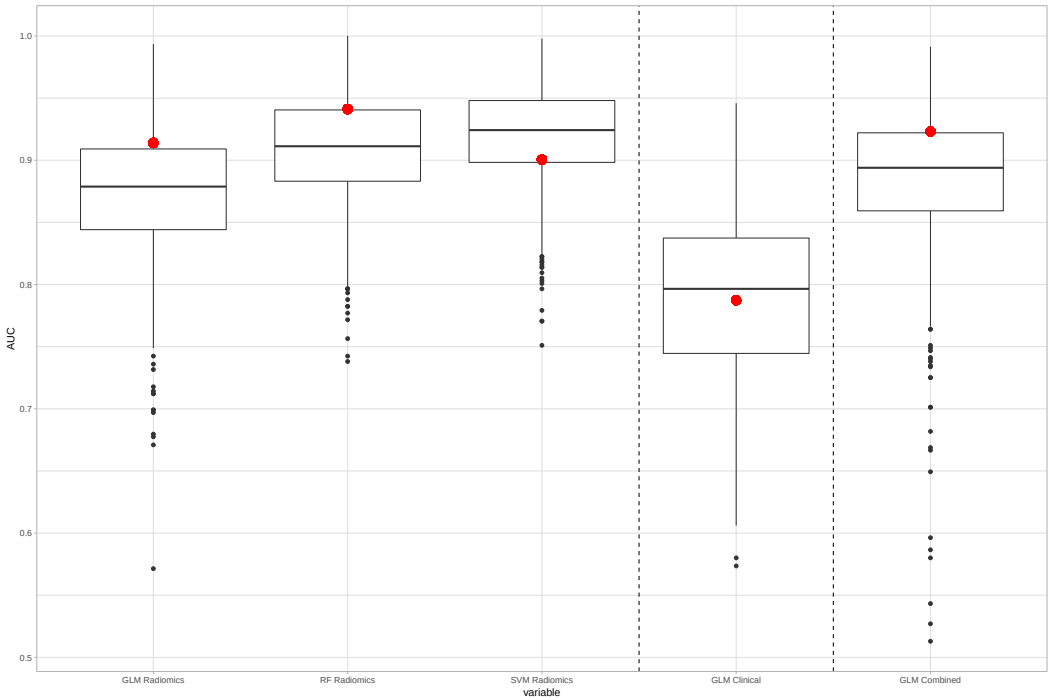


Fig. 3.3: Boxplots showing the performance of the clinical model and the radiomics model in a random permutation (n=1000) on the full dataset. The red dot indicates the performance of the model on the internal validation data.

features than Baessler as important, which can be explained by the use of a different imaging sequence for analysis. Furthermore, this study also confirms the use of the clinical features by Chin.[29] These features are integrated into a clinical risk-score that predicts the risk of midwall fibrosis. These features are not directly calculated from the medical images. This can be appointed as the reason that these perform worse than the radiomic features, which are directly calculated from medical images. When the clinical and the radiomic GLM are combined, a slightly better performance can be seen than in both models individually. Yet, the best model performance is seen in the more sophisticated RF and SVM trained with radiomic features.

This research showed the feasibility of multiple models for the analysis of LGE. A large feature set was used, giving a wide selection of different types of features. Features were eliminated during the process. This resulted in a reduced feature set for final analysis. Feature reduction is an important action with a limited dataset of samples, as was the case in this study. A poor feature reduction can lead to overfitting on the training set of data and a poor model performance on the test data. This study showed a good performance on the test set with the inclusion of 50 features in the RF and SVM. This does not indicate overfitting.

Furthermore, the dataset was heterogeneous with respect to the included subjects; there were different severities of AS and controls in the cohort. However, this was not taken into account in the analysis and the only focus was on enhancement or not. Therefore, the effects of this heterogeneity are assumed to be minimal. Yet, all scans were obtained from the same scanner with a similar sequence. Thus, this does not reveal anything about generalizability of features and applicability of the model to a validation set from a different institute.

Image manipulation also shows some of the strengths and weaknesses of this research. A disadvantage of this research is the use of manual segmentation. Segmentation of enhanced cardiac images is still a difficult process due to the similar voxel value of enhancement and the blood pool. Manual segmentation of the myocardium was performed by an inexperienced researcher and therefore prone to incorrect segmentation. However, the interobserver variability for manual segmentation of the myocardium is relatively small.[124, 117] Advantageously, this research showed the opportunities for image normalization according to a $\mu \pm 3\sigma$-method (for more information on this type of normalization, see section 2.3.3). A large share of the top-20 features in this analysis was derived from these normalized images. MR-normalization gives way to quantitative inter-patient comparability of MR-scans.

A large variety of features accounted for the presence of enhancement in univariate and multivariate analysis. It was expected that features that are directly related to the voxel intensities in the myocardium scored high in the feature analysis. These are the first order statistics features and the texture features derived from the Grey-level

Co-occurence Matrix (GLCM) and Grey-level Run-Length Matrix (GLRLM). However, less expected was the large share of shape- and size-based features and especially the presence of 3 case-specific features in the top-20. This confirms the presence of myocardial fibrosis in a hypertrophic heart.[133, 100] Another feature group that was present in the top-20 are the features derived from the >1 and >2 σ-images. This implies that late gadolinium enhancement has feature values higher than 1 and 2 standard deviations than the mean of the whole myocardium. This is in contrast with the results from Harrigan [65], where >6 sd threshold had the strongest correlation with LGE. However, the applied method in both studies is different; Harrigan et al. applied an extra segmentation of healthy myocardium to determine the mean signal intensity and standard deviation. Therefore, no enhancement is taken into account for evaluation of the mean, which explains the difference in standard deviation value. In univariate analysis of clinical features, the highest score was appointed to cardiac troponin-I, which is expected because the concentration of this substrate in blood is directly related to myocardial damage.[6]

This study explored the clinical opportunity of analysis of a cardiac disease dataset to find enhancement patterns that are inducing a higher risk of sudden cardiac death.[42, 121, 152] The presented research showed that it is possible to capture enhancement in texture features. It therefore confirms that AS is not only a disease of the valve, but also of the myocardium.[40] Although this research has been applied to patients with AS and a control group with a number of subjects with myocardial infarction, the demonstration of enhancement analysis with radiomics features might be generalizable to other cardiac pathologies. Therefore, this study can be contributing to the ongoing investigations towards the role of enhancement in e.g. ventricular arrhythmic events in coronary artery disease and non-ischaemic cardiomyopathy [130], mortality in pulmonary hypertension [143] and response to cardiac resynchronization therapy in patients with heart failure[118]. It gives the opportunity to monitor disease and changes in the myocardial wall, especially in a prone patient group with a high risk of sudden cardiac death, as AS patients.

This study demonstrated that radiomic features have the ability to discriminate between subjects with and without enhancement. Myocardial enhancement has many diagnostic and prognostic implications in different cardiac pathologies. Future research should focus on improvement of the models and implementation of other outcome measures to define high- and low-risk patients in cardiac pathologies.

4 External validation of a predictive model for AVR and a computer-aided diagnosis of LGE

4.1 Abstract

Introduction Predictive models and computer-aided diagnosis are becoming increasingly popular in the clinical field. Previous research of this thesis used a temporal validation for testing the performance of a predictive model for AVR and a computer-aided diagnosis of LGE. The performance of these models is now tested with an external validation cohort.

Methods Two datasets were included in this study. Dataset-1 included 199 subjects for computer-aided diagnosis of LGE and 156 subjects for prediction of AVR. Dataset-2 included 99 severe AS-patients for external validation. The myocardium was manually segmented from the images for extraction of features. Univariate and multivariate analysis was performed to evaluate the performance of radiomic features in the prediction of AVR and the computer-aided diagnosis of LGE. Performance measures were CI, respectively AUC.

Results 5639 features were extracted from the images. A correlation of 0.97 was shown in the univariate analysis between the significant (FDR q-value$<$0.05) features for AVR and LGE, with a strong linear relation (R^2: 0.94, correlation coefficient: 0.97). Multivariate analysis of the prediction of AVR showed similar performance on external validation as on temporal validation (average CI: 0.6). Performance on computer-aided diagnosis (average AUC: 0.70) was lower than in temporal validation.

Discussion To this date, this is the first study that tested the performance of classification models, trained with CMR-data, in an external validation cohort. Cardiac MRI contains valuable information on the state of the heart that is currently not analysed and considered in patient characterization and evaluation. Radiomics is able to extract this information from LGE-imaging.

4.2 Introduction

Prediction research and computer-aided diagnosis is becoming increasingly popular in the current clinical settings. These models are able to predict patient outcome or can aid a clinician in determining the diagnosis of a patient[153] However, to give these models clinical value, they must show to work in other samples than the data that has been used to train the model. Evaluation of the performance on a part of the original dataset, that has been used for training of the model, increases the risk of an overly optimistic outcome.[3] This is called internal validation and this method usually yields optimistic result, because training and test data is very similar. Therefore, external validation is of importance.

External validation is the use of another dataset for validation of the model. This dataset should be acquired in similar patients, but in a different setting of centre.[153] The goal of an external validation is to confirm the prognostic value of the model and to avoid optimism based on false positive results that could have been obtained during internal validation.[155]

The previous chapters have used a temporal validation set-up to test the models for the prediction of AVR and the computer-aided diagnosis of LGE. In the previous chapters the dataset was divided into 75% for training and 25% for testing of the model by using a temporal split. The current chapter implements an extra dataset from a different centre for the external validation of the previously developed models. An external validation requires data on all the variables of the model, in this case; radiomic features.[3]

This chapter tests the performance of the models in an external validation cohort. Training of the models is done with the complete dataset from the previous chapters. The classifiers are used for the prediction of the risk of AVR and for the computer-aided diagnosis of LGE. This chapter reviews the performance with the performance measures CI/AUC to support a decision for clinical implementation of one or both models.

4.3 Methods

4.3.1 Study Population

Two datasets were used in the analysis for validation of the models. Dataset-1 is obtained from the Edinburgh Heart Centre (clinicaltrials.gov, identifier: NCT01755936). The selected patient cohort consists of 166 patients with AS and 48 control subjects. The patient cohort is used as training data in the classifier for AVR (n = 156). The full dataset, patients and controls is, used as training data in the classifier for LGE (n = 199). Start of the enrolment was January, 2012. Patients were excluded from the

trial if they had other significant valvular heart disease, heart failure or infection, significant comorbidities, cardiomyopathies or contraindications to CMR imaging.

Dataset-2 has been obtained from the University College London Hospital NHS Trust (clinicaltrials.gov, identifier: NCT02174471). The patient cohort was obtained between January 2012 and January 2015. Patients were selected with severe, symptomatic AS and were selected as they were prepared for AVR. Exclusion criteria were other severe valvular disease, previous valve surgery, renal impairment, pregnancy and contraindications for CMR. This dataset existed of 167 patients. Figure 4.1 shows the distribution of the included subjects and the final cohort that remained. This patient cohort has been used in the model for the prediction of AVR and for the computer-aided diagnosis of LGE.

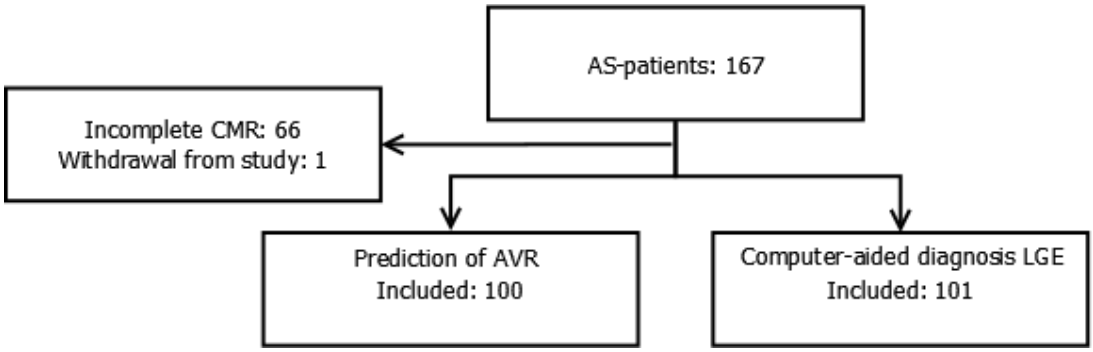


Fig. 4.1: Overview of the study population in this research

4.3.2 Clinical Outcome

The models from the previous chapter were validated in this chapter with an independent dataset. The primary outcome is AVR in the first model. Time to AVR was defined as the number of days between the baseline CMR-scan and the date of surgery. The secondary outcome is the overall survival, defined as the number of days between the baseline CMR-scan and the last day of follow-up, day of death or the pull-out date.

The outcome of the computer-aided-diagnosis model, designed in chapter 3 was based on the presence or absence of LGE. The presence of LGE was previously defined in both datasets by the qualitative assessment of experienced radiologists. This judgement was used as the outcome in the classification models.

In chapter 2 and 3, also clinical variables have been used for the prediction of AVR and the computer-aided diagnosis of LGE. In the current chapter only the radiomics models were used due to limited availability of clinical variables in dataset-2.

4.3.3 MR Image Acquisition and Segmentation

For details on the image acquisition of the CMR-images of dataset-1, I refer to the similar section in chapter 2. Image acquisition of dataset-2 was performed at a 1.5 Tesla MRI-scanner (MAGNETOM Avanto, Siemens AG, Healthcare sector, Germany) with 32 channel cardiac coil arrays. The LGE-images were obtained after a bolus (0.1 mmol/kg) of gadolinium contrast (Gadoterate meglumine-gadolinium-DOTA, marketed as Dotarem, Guerbet S.A., Paris, France). Post-contrast LGE-imaging was obtained 10 minutes after infusion.[148]

Further image modification and segmentation of dataset-2 has been performed in a similar way as for dataset-1. For a description of this process, I refer to the similar sections in chapter 2.

4.3.4 Radiomics and Case-specific features

The radiomic features used in this analysis are mathematically described in the supplements. Feature extraction was performed with PyRadiomics, an open source python package for the automated extraction of Radiomic features. [60] Furthermore, the case-specific features, as described in table 2.2, were calculated in Matlab (Matlab R2016b, The Mathworks, Inc., Natick, Massachusetts, United States). Features from the cylindrical reconstruction were also calculated with an in-house software tool coded by the researcher.

4.3.5 Feature Ranking and Selection

Feature ranking and selection was performed only on dataset-1 to leave dataset-2 as an independent validation cohort for the evaluation of both models. Feature selection was performed with an mRMR algorithm [37]. Before feature ranking and selection, feature space was reduced by removing features with a near zero variance and features that were highly correlated (Spearman's ρ >0.9).

4.3.6 Statistical Analysis

In the model for the prediction of AVR performance was evaluated with the CI, as explained in chapter 2. The model for classification of LGE uses the AUC from the ROC as the performance parameter. CIs and AUCs were directly computed from the continuous radiomic variables for univariate analysis. This analysis used dataset-1 for calculation.

Multivariate analysis includes a GLM or cox regression, RF-model and an SVM. For the GLM and cox regression a forward selection method was implemented, where features were added to the model one-by-one based upon the ranking of the mRMR-algorithm on dataset-1. Intermediate models were validated with random

subsampling (n=1000) on dataset-1 (75%-training, 25%-validation). When the mean CI (cox regression AVR-model) or AUC (GLM LGE-model) dropped, the corresponding feature set was selected for implementation in the final model. The RF- and SVM-model used the 50 highest-ranking features from the mRMR-feature selection for modelling.

All statistical analysis, feature ranking and feature selection was performed in RStudio, coded in R.[120]

4.4 Results

4.4.1 Study Population

The study population only included AS-patients from dataset-1 and -2 for the validation of the models for prediction of AVR. In the computer-aided diagnosis of LGE, controls from dataset-1 were also taken into account in the training set. The patient characteristics are shown in table 4.1 and 4.2 for AVR, respectively LGE. All features, except age, are significantly different between dataset-1 and dataset-2. This is caused by the different inclusion criteria. Average age of the complete dataset for AVR is 69.2, for LGE 67.44.

	Total (n=255)	**Dataset-1 (n=155)**	**Dataset-2 (n=100)**	**p-value**
Age [years]	69.2 (28 - 92)	68.5 (28 -92)	70.3 (34 -87)	0.28
Gender [M/F]	159/96 (62%/38%)	109/46 (70%/30%)	50/50 (50%/50%)	0.0017
Severity AS [mild/moderate/severe]	33/43/179	33/43/79	0/0/100	<0.01
Follow-up [months]	17.8 (0	27.5 (1 - 62.8)	2.91 (0 - 48.8)	<0.01
Primary Endpoint	195 (76.4%)	98 (63.2%)	97 (97%)	<0.01

Tab. 4.1: Patient details for the datasets used in the validation of the AVR-models. The Wilcoxon rank sum test was used for p-value calculation of the continuous variables, chi-squared test for p-value evaluation of categorical variables.

	Total (n=300)	**Dataset-1 (n=199)**	**Dataset-2 (n=101)**	**p-value**
Age [years]	67.44 (19 - 92)	66.4 (19 - 92)	70.2 (34 - 87)	0.04
Gender [M/F]	187/113 (62%/38%)	136/63 (68%/32%)	51/50 (50%/50%)	0.005
Severity AS [control/mild/moderate/severe]	44/33/43/180	44/33/43/79	0/0/0/101	>0.01
Primary Endpoint (AVR + Adverse events)	196 (65%)	99 (50%)	97 (96)	>0.01
Enhancement	132 (44%)	59 (30%)	73 (72%)	>0.01

Tab. 4.2: Patient details for the datasets used in the validation of the LGE-models. The Wilcoxon rank-sum test was used for p-value calculation of the continuous variables, chi-squared test for p-value evaluation of categorical variables.

4.4.2 Univariate Analysis

5639 Radiomic features were calculated from the medical images. In a comparison between the univariate analysis of LGE and AVR, no features were removed. The

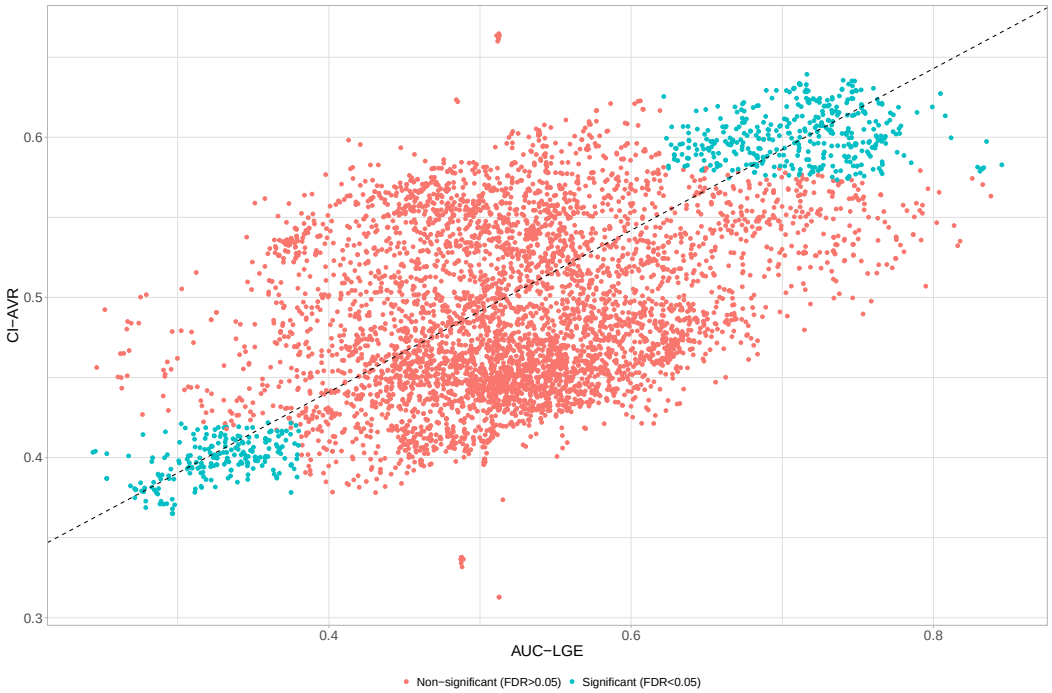


Fig. 4.2: Univariate Performances of Radiomic features for AVR and LGE. Each point refers to the CI and the AUC, predicting the risk of AVR, respectively indicating that there is enhancement present in the LGE-image. This figure shows the univariate performance measures of 5635 features, of which 622 are significant (FDR<0.05) for both AVR and LGE. The linear regression for the significant couples resulted in an R^2-value of 0.94 (F-test, p-value$<2.2 * 10^{-16}$).

AUC, respectively the CI, were calculated to determine the relationship between both values. This analysis only included patients from dataset-1. Controls were excluded in the univariate analysis of LGE, to ensure the same population in both analyses.

The correlation between the CI and AUC in the univariate analysis of the radiomic features of AVR, respectively LGE was investigated in more depth. Figure 4.2 shows the correlation between the CI and AUC of each individual feature. 1039 (18.4%) features had a significant (FDR<0.05) CI for the prediction of AVR. This number was 1541 (27.3%) for the computer-aided diagnosis of LGE. When these analyses were combined, 622 performance measures were significant from random (FDR<0.05) for both AVR and LGE. In this group of features, a strong linear relationship was observed between the AUC for the presence of LGE and the CI for the prediction of AVR (R^2: 0.94) and a significant correlation (P-value<0.01) of 0.97. The correlation was 0.51 for all data points in this figure. This correlation was included all the features, with significant and non-significant AUC and CI.

4.4.3 Multivariate Analysis

For multivariate analysis, features with a near zero variance and features that were highly correlated were removed from the feature set. In the analysis of the prediction

Feature	Imagefilter	Power	p-value	q-value (FDR)
Mean Thickness	Case-specific	0.62	0.15E-4	0.0025
Run Variance glrl	Wavelet HLH Epicard	0.63	0.83E-5	0.0022

Tab. 4.3: The selected features for the radiomic signature of the GLM for the prediction of AVR. The power, p- and q-value were derived from the univariate analysis.

of AVR, 1174 features were retained. For the identification of LGE the feature set included 1125 features.

Three models were designed that used radiomic features as predictors and prediction of AVR or computer-aided diagnosis of LGE as outcome measures. After forward selection, two features were included in the cox regression for prediction of AVR and five features made up for the GLM for computer-aided diagnosis of LGE. These features are described in table 4.3 and 4.4. Figure 4.3 shows the performance of the multivariate models in the prediction of AVR and the computer-aided diagnosis of LGE on the training and test data. Figure 6.9 shows the forward selection process for both analyses.

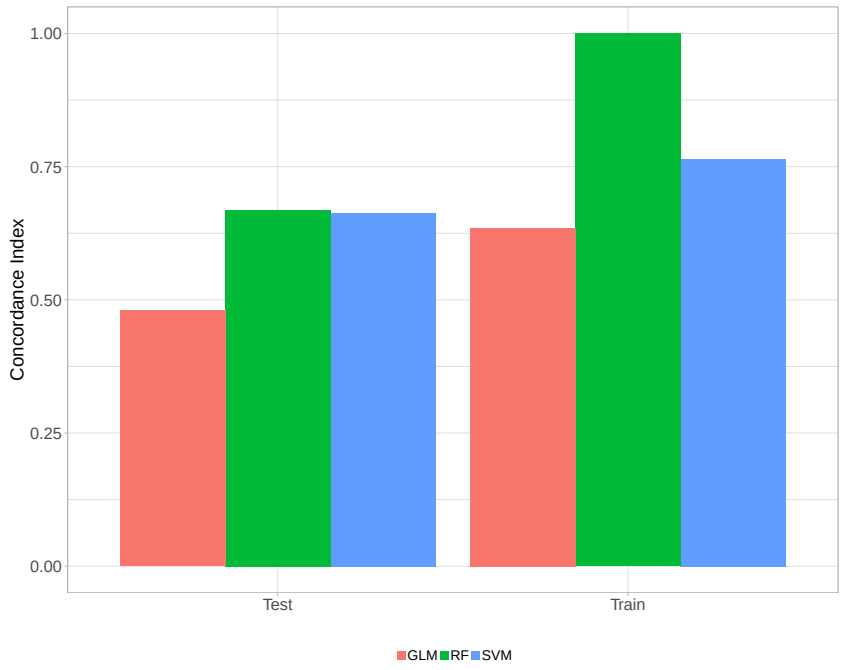
Feature	Imagefilter	Power	p-value	q-value (FDR)
First Order Range	Normalized Wavelet LLL	0.85	0.82E-31	0.46E-28
Mean Thickness	Case-specific	0.84	0.12E-28	0.46E-26
Zone Entropy glszm	Normalized Wavelet LLL	0.83	0.65E-25	0.18E-22
Large Area High Gray Level Emphasis glszm	Wavelet HLL Endocard	0.83	0.20E-22	0.46E-20
Shape	2σ-volume	0.81	0.87E-21	0.16E-18

Tab. 4.4: The selected features for the radiomic signature of the GLM for the computer-aided diagnosis of LGE. The power, p- and q-value were derived from the univariate analysis.

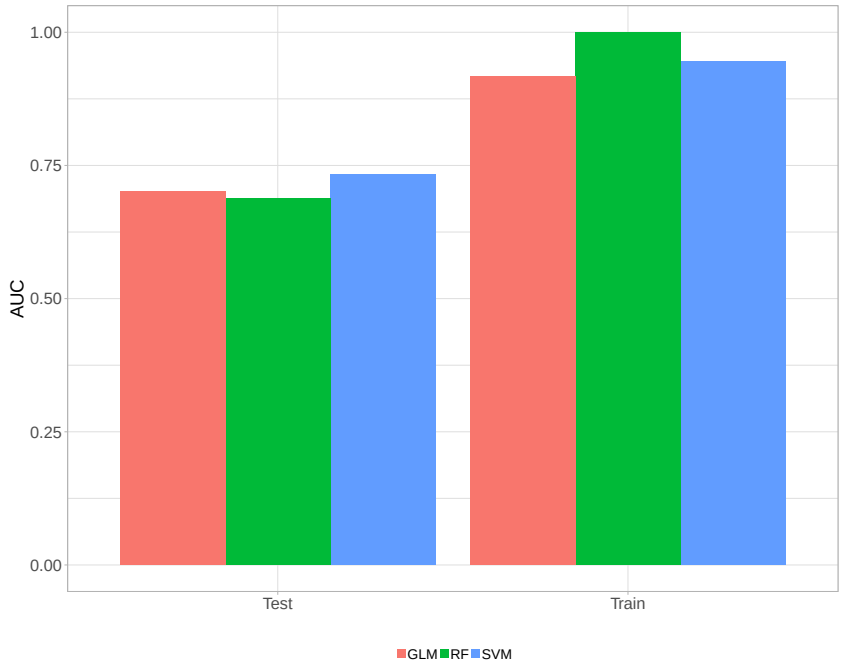
4.5 Discussion

Cardiac MRI contains valuable information on the state of the heart that is currently not analysed and considered in patient characterization and evaluation. Radiomics is able to extract this information in different imaging features. LGE-imaging is suitable for this type of evaluation, because the intensity differences of the myocardium are related to the vitality of the myocardial cells.[119, 49] In this chapter, the models of two classification tasks, defined in the previous chapters, were validated with an external validation patient cohort. The classification tasks included the prediction of AVR or adverse events during follow-up and the computer-aided diagnosis of enhancement.

5639 features were extracted from the medical images; 622 features had a significant CI, respectively AUC for both classification tasks. Validation of the models resulted in a poor performance in the prediction of AVR (cox regression: 0.48, RF: 0.66, SVM: 0.66). This result is comparable to the internal validation performed in chapter 2. Computer-aided diagnosis of LGE led to a better, moderate, performance (glsglm:



(a)



(b)

Fig. 4.3: Bar plots showing the performance of the models on the training- (dataset-1) and test set (dataset-2) for the prediction of AVR (a) and the computer-aided diagnosis of LGE (b).

0.70, RF: 0.68, SVM: 0.73). However, these AUCs are not as high when compared to the internal validation done in chapter 3.

To date, this is the first study that tests the performance of classification models, trained with CMR-data, in an external validation cohort. Previous research has mainly focussed on the extraction of features from cardiac MRI and showing their performance in a single dataset. Publications are relatively scarce within the combined field of cardiac MRI, feature analysis and machine learning models. Baessler [11] uses cine-CMR for the quantification of myocardial infarction in a single dataset from one institution.

This research also pointed towards the high ranking of the maximum myocardial thickness and the mean myocardial thickness. This confirms the importance of a thickened myocardium for the risk of adverse events [138]

As previously mentioned, the use of an external validation for determining performance of models on a different dataset is a strength of this study; univariate analysis, feature selection and model training were done on dataset-1. This left dataset-2 as an independent validation dataset. MRI is a relatively new field for the application of radiomics, because of the large interpatient and -scanner differences. With the presented work, we have shown that also CMR is suitable for radiomics analysis with several limitations.

One of the limitations of this research is the large difference in characteristics between dataset-1 and dataset-2. Firstly, all patient characteristics, except age, are significantly different between dataset-1 and dataset-2. Important to mention is the significant difference in the severity of AS and the primary endpoint in both groups. Dataset-1 has all severities included in its cohort. In this cohort also the choice for therapy and the clinical pathway of the patients is undetermined. Dataset-2 has only included patients with severe AS, which were already planned for AVR. These differences in inclusion criteria resulted in significant differences in follow-up time, number of patients with AVR and severity of AS.

Another difference between dataset-1 and -2 was the imaging time after gadolinium infusion; 15 minutes, respectively 10 minutes. The wash-out of gadolinium from the healthy myocardial cells takes time. If this time is too short, there is an incomplete removal of gadolinium from healthy tissue. This results in a white band-pattern over the myocardium. This increases the homogeneity of the obtained feature values. This effect can lead to higher voxel values in dataset-2 with an effect on the feature values. Yet, this assumption has not been researched in more depth.

For clinical validation of the performance a threshold-performance of the models in external validation should be defined before evaluation. If this performance is met, the model could be suitable for clinical implementation. However, it can be difficult to determine what value is acceptable and clinical evaluation of the prediction of the

model is still necessary.[3] In this exploratory research, with no purpose of clinical implementation, no performance threshold was determined before analysis.

The aggregation of the univariate analysis of features on LGE and AVR resulted in a strong linear relationship between all features that had both a significant (FDR q-value <0.05) AUC for LGE-classification, as a CI for AVR-classification (R^2 = 0.94 and correlation coefficient = 0.97). This high linear correlation can be expected, because previous studies have shown the importance of LGE in the prediction of adverse events.[42, 121, 152]

The performance of the AVR-models on the external validation cohort resembles the performance in the internal validation in chapter 2. The model is underperforming in comparison to the models that used the clinical feature values from Chin [29]. The CI of the models for the classification of AVR is close to 0.5. This means that the classification is close to random.

In contrast with the models for AVR, the model for computer-aided diagnosis of LGE had a worse performance in the external validation than in the internal validation in chapter 3. An explanation for this is overfitting. Presumably, the model is overfitting on the characteristics of the images in dataset-1. Therefore, these models are still performing well in the test set from dataset-1. Yet, when testing on a different dataset with different characteristics, the performance drops. However, with an AUC of approximately 0.7, the RF and SVM are still appropriate options for the classification of LGE.

The external validation of both models showed that the LGE model has the power to be implemented in the clinical practice. Before implementation, an appropriate threshold value has to be determined to classify myocardium with and without fibrosis. The current outcome score on GLM revealed a very small low range for non-LGE myocardium and a broad spectrum of model outcome values that can be regarded as LGE. The clinical added value should also be evaluated. This model aids the radiologist in the analysis of the myocardium. Yet, there are additional modifications (segmentation, cylindrical reconstruction) that have to be done before the scan can be analyzed. The extra burden on radiologists or on assistants of radiology has to weigh up against the additional information that these segmentations will provide. Currently, there is also ongoing research for the segmentation of myocardium from CT and CMR images.[9, 160] As these algorithms develop, the automatic analysis of medical images is a step closer.

This research was the first research that examined the performance of machine-learning models, trained with CMR-enhancement data, in an external validation set. Although the performance was worse in the external validation than in the internal validation presented in previous chapters, especially in the computer-aided diagnosis of LGE shows promising results for the future. With the inclusion of more

and larger datasets, the computer-aided diagnosis of LGE can be improved. These models can also be applied to other cardiac pathologies. It is expected that during this process also the correlation between LGE and adverse events, shown in other studies, [42, 121, 152] becomes more outstanding in machine learning models that are trained with radiomic features. These models have the opportunity to aid in clinical patient-centred decision-making.

5 Discussion

A radiomics approach for the analysis of myocardium in LGE-images of the heart is feasible. A predictive machine-learning model with quantitative radiomic features, derived from the myocardial enhancement image, can aid in phenotyping patients with AS. This phenotyping comprises the classification of patients with and without fibrosis. However, radiomic features showed only limited prognostic value for risk-evaluation of cardiac events or AVR in the follow-up time. The clinical risk score proposed by Chin, 2016 [29] performed better in this classification. This research is as far known the first study published to broaden to scope of radiomics outside oncology. Yet, research using groups of radiomic features in cardiology have been published before.[11, 141]

The results of this study confirm previous results in AS-patients. The general linear model including features of the clinical risk score published by Chin [29] performed better in the prediction of AVR than inclusion of solitary radiomic features. This result confirms the features included in the clinical risk score, relating clinical variables to midwall fibrosis, which on itself is a predictor for adverse events.[42] However, also the radiomic features contain information about the status of the myocardium of the patient. This was shown in the second study, where radiomics features outperformed the clinical risk score for the presence of LGE in cardiac MRI. Based upon the combined conclusions of both studies, it can be carefully concluded that left ventricular enhancement does not always describe the need for AVR in patients with AS. However, this is in contrast with literature published.[48]

The last chapter of this thesis will focus on the overall discussion of this research. It uses part of the headings of the methods section to discuss the methods used and the forthcoming results. Furthermore, a future perspective from the researcher's point of view on the use of machine learning in health care will be given.

5.1 Patient Selection and Aortic Stenosis

Dataset-1 was obtained from a research group in Edinburgh. Inclusion was not performed by the researcher herself and therefore, no influence could be practised on this process. This lead to the exclusion of some patients with scans of poor quality or missing data of clinical features. Dataset-2 was obtained from London and had the same disadvantage; inclusion was done before this study started, leading to exclusion of some patients.

There is a large heterogeneity in the included subjects in this research. First of all, there are different degrees of AS and control subjects with other cardiac pathologies in the cohort of dataset-1. This difference is minimized by making the train and validation group in chapter 2 and 3 consist of equal proportions of different severities. This stratified train- and test-set identification was not done in the permutation analysis, creating possibly significantly different training- and test-sets during this analysis. This probably decreased the CI or AUC. Influences on the results have been minimized by performing stratified random permutation analyses 1000 times. A similar distribution of subjects belong to the different groups is ensured in training and test.

The current guidelines also prescribe that patients with symptoms should always be treated.[150] With the results of this research, I can carefully conclude that the fibrosis of the myocardium is not always the reason to develop symptoms. Therefore, other causes can be proposed that cause symptoms, e.g. a decreasing LVEF due to backflow or coronary microvascular dysfunction due to coronary artery disease, which is a precursor of fibrosis and thus enhancement.[26] Heterogeneity can also be found in the inclusion of patients with a bicuspid aortic valve. These patients are also included in the analysis, although they have a higher risk on other cardiac pathologies.[54]

The analysis on dataset-1 can be improved by the use of follow-up scans of the included patients which had not had AVR in the follow-up time. These scans can give a better idea of the possible enhancement patterns that can be identified before AVR, because they are made when AS progresses. The patient is thus closer to the primary endpoint of AVR.

In comparison to dataset-1, dataset-2 was more homogeneous. It only included severe AS patients. All patients were worked up for AVR. This probably has not been the best choice for an external validation cohort, due to the significant differences demonstrated in dataset-1 and -2. A possibility was to combine both datasets into one and use temporal validation for a new split in training- and test set. However, this reduces the effect of external validation, which is an important aspect of machine learning models.

5.2 Image Acquisition, Segmentation and Reconstruction

The use of MRI for radiomics analysis has limitations. First of all, the spatial resolution of MRI is not as good as CT, increasing the partial volume effect of the surrounding tissue on the myocardium. This effect has been reduced by removal of the outer borders of the segmentation. However, also subtle intensity changes in

the myocardium are influenced by this effect. This is the case in patchy and diffuse fibrosis, where the affected areas are diffusely distributed in the myocardium.[66] In this case a voxel is made up of fibrotic and healthy myocardial cells, underestimating the amount of fibrosis and affecting the feature values.

Furthermore, MRI is depending on the scanner's manufacturer, field strength and field inhomogeneities. It is therefore more difficult to compare MRI with each other. Normalization of the imaging to a specific value, e.g. the grey value of the blood pool or air in the image, can be a solution. Yet, it gives increasing difficulties in the LGE images, because this sequence already utilizes the relative blackness of the myocardium to surrounding tissue and myocardial enhancement to construct images. The presented studies have used $\mu \pm 3\sigma$ (see section 2.3.3) as the normalization factor. This methods showed promising results in the second study; many of the top-20 features in univariate analysis were derived from the normalized images. This is an indication that feature values, and thus images, are comparable between different subjects. Further research has to determine if this method of normalization indeed increases the inter- and intra-patient comparability of MR-scans.

Segmentation and reconstruction of the myocardial image included some factors of variance. First of all, segmentation of the left ventricle has been done by an inexperienced researcher. However, previous research has described a low inter-observer variability in the segmentation of the myocardium.[124, 117] Therefore, it is assumed that the segmentation is not differing much from what an expert would segment.

Another factor of variance was the use of a cylindrical reconstruction of the segmented myocardium. It is a laborious proces to create the cylindrical reconstruction. However, features from the cylindrical reconstruction were not selected in the univariate or multivariate analysis. It was hypothesized beforehand that the histologic delineation of myocardial cells [5] should be followed in texture feature analysis to

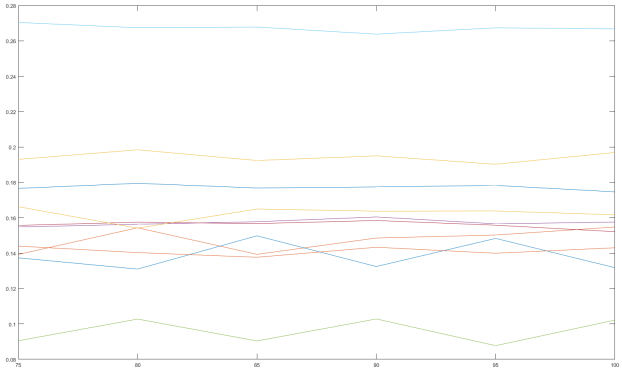


Fig. 5.1: The influence of the chosen diameter for reconstruction (x-axis) on the feature value (y-axis) in different subjects.

enable a proper quantification of texture features. The process was time-consuming, because every segmentation had to be manually checked for the right selection of the midpoint of the left ventricle. This manual checking is prone to mistakes. Other methods have been explored in this research for reconstruction of the myocardium. Yet, none were found suitable, e.g. due to interpolation and creation of extra data. Feature values were also influenced by the chosen values for variables of the reconstruction, e.g. the diameter of the circle used to make the reconstruction. The diameter against the change in feature values is shown in figure 5.1. This variability was reduced with fixed values for the diameter for every subject.

For a better risk analysis in patients with AS, the myocardium should be further subdivided to identify relevant regions and features determining the risk of AVR. In this research the full myocardium and splits of the myocardium in epi-, mid- and endocardium were used. Yet, in current clinicaly analysis the 17-segments model [28] is used to appoint the exact location of the fibrosis in. Implementation of the 17-segments model in this research can help to identify locations of fibrosis with a higher risk. It avoids masking of fibrotic segments by surrounding healthy tissue. However, manual assignment of the segments is time-consuming and induces another factor of variability.

At last, it is debatable whether an LGE-sequence is the optimal sequence for identifying enhancement patterns in AS. AS tends to have diffuse fibrosis patterns, which are hard to identify due to nulling of the myocardium relative to the bloodpool. Research showed that T1-mapping is able to identify diffuse myocardial fibrosis in cardiomyopathy.[133, 151, 73] This type of fibrosis is also expected to be present in AS, due to the left ventricular remodelling that occurs in patients with AS.[154] Therefore, a promising step is addition of T1-maps to the analysis.

5.3 Feature Extraction

No research has been done regarding the test-retest stability and reproducibility of LGE-CMR radiomic imaging features. This is an important disadvantage of this study. Test-retest stability of radiomic features has only been determined in Fludeoxyglucose-Positron Emission Tomography (FDG-PET)[91] and CT[1] with the RIDER-dataset.[159] To determine the test-retest stability of features in CMR, a scan has to be made twice within a short period of time, where the subject is moved from the scanner and repositioned in the scanner. This gives the ability to determine the intra-patient stability of features with the correlation coefficient.

During the process of this study multiple software packages have been used for analysis. A start was made in Matlab for the extraction of features in a self-made set-up. This set-up was prone to mistakes and not widely available to other researchers, working in the field of Radiomics. Therefore, a switch was made to the extraction

of features with the PyRadiomics toolbox, a free toolbox for the extraction of features.[60] This increased the reliability of the feature values. A comparison was made between the feature values extracted from Matlab and PyRadiomics and these were the same in the larger part of cases, indicating that the formulas were correctly coded.

5.4 Data Analysis

The choices for different machine learning models used in this research are based upon the characteristics of the models. A generalized linear model and a cox regression model have been implemented, because these models are mathematically simple models. An RF has the advantage that it does not require considerate preproccessing of input data; it is not prone to outliers and varying magnitudes of size between features. SVMs have the advantage that these are relatively simple to understand and widely used. However, these choices can be argued. As can be seen from recent literature published, every scientist makes a decision on the implemented model based on its experience with this model or on characteristics of the dataset.[77] Besides personal preference, it is also a matter of trial-and-error; "which model works best?", "Should the features be normalized?" and "which model can be smoothly implemented and debugged?". These questions arise when deciding upon a machine learning model. Likewise, studies invesigate the performance of different classifiers on generic datasets.[50]

For feature selection, the same statement holds as for the chosen machine learning models; in the currently presented research, the feature selector was chosen according to its performance in previous research.[112, 22] Yet, many more methods are available, suitable for different types of data or feature ranking methods[128], e.g. LASSO [158].

Returning to machine learning models used in this research, these models and their implementation have limitations. First of all, the number of included subjects is critically low for sophisticated machine learning purposes. There is a high probability of overfitting, due to the large number of features. Therefore, feature selection is extremely important. Secondly, some data was missing in patients for the selection of AVR; LVEF and V_{max} on echo or HS-T1. This resulted in the removal of these subject from the dataset, further decreasing the dataset.

5.5 Radiomics and Clinical Features and Outcome

For this research, different models were trained with different groups of features. The clinical features were implemented to test the clinical risk score that has been proposed by Chin.[29] In both studies there was one clinical feature that scored high in the univariate analysis. In the prediction of AVR, the V_{max} was the highest-scoring feature. This can be clinically explained, because this is a direct measure of the condition of the aortic valve; flow speed increases with severity of AS.[45] In the computer-aided diagnosis of LGE, the high-sensitivity troponin-I performed best in the univariate model. Also this feature has a clinical explanation; high-sensitivity troponin-I is a cardiac marker, released when cardiac damage occurs. It is released in high amounts during a cardiac infarction, but in relatively small concentrations in AS or other cardiac pathologies.[136, 41, 148] Lastly, LVEF measured on echocardiography was included in the univariate analysis, but showed to have no power in the univariate analysis in both studies. In this study, LVEF determined on cardiac echo was included instead of LVEF on MRI. The first is the current criterion used to select asymptomatic patients for AVR, although there are limitations on measuring variables on cardiac echo.[53] From this analysis, it can be concluded that a single LVEF-value, measured at baseline, is not sufficient to have predictive power. Patients with AS are usually monitored over multiple years. The change of LVEF over years is a feature that presumably has more prognostic power.

It was expected that more features in the univariate analysis of the AVR-analysis would relate to the shape and size of the left ventricle, as concluded from the research by Steadman [140]. However, this was not the case, indicating that fibrosis and the development of fibrosis might play a more important role in the development of symptoms than the size of the myocardium. Due to the numerous shape- and size-based features in the LGE-research, it can be hypothesized that shape and size of the myocardium are related to the development of fibrosis, which on its turn is causing cardiac failure. Yet, regarding the fair performance of the AVR-prediction models, symptom development cannot be fully dedicated to the development of fibrosis. However, a positive note on the analysis is that the computer-aided diagnosis of fibrosis showed good results, with an AUC of 0.9. Fibrosis of the heart has proven to be a predictor of adverse events in AS [42, 121, 152] and other pathologies [107, 7, 17]. Therefore, it is hypothesized that the machine learning models using adverse events or AVR as outcome are going to perform better with a larger included dataset.

To conclude; the assessment and prognosis of AS is a complicating task. There are many factors that play a role in the development of AS and many quantitative measurements can be derived from the heart, that individually or combined have

their own implication on the patient.[13] It was optimistic to think that the quantitative analysis of the myocardium had all the answers. The radiomics analysis of the myocardium should not be handled as an independent measurement, but must be taken into account with all the patient characteristics that are already available to make an adequate estimation of the risk of adverse events and the need for AVR. This is were radiomics turns into cardiomics; besides CMR-imaging, other patient characteristics are available, e.g. Electrocardiography (EKG), multiple measurement of cardiac echo and other cardiac measurements. Machine learning models trained with complete datasets of all available data on patients with AS can aid in risk determination.

5.6 Future Perspectives

Machine learning, deep learning, Artificial Intelligence (AI) and computer-aided image analysis are subjects that have gained infinite attention in the medical field in recent years. Avanzo published an overview of all the radiomics papers at the beginning of 2017 [8], which at the time of publishing was already outdated. Limiting to the field of cardiac pathologies, papers have been published that include, among others, machine learning of the right ventricular motion to predict outcome in pulmonary hypertension [36], machine-learning models to discriminate between physiological and pathological hypertrophic remodelling [102], machine learning to identify infarct tissue on cine-images [11] and machine learning for the selection of patients with significant coronary artery disease[146]. What does this increasing interest mean for health care? What are the possible advantages and disadvantages? And how will data-driven science develop in the coming years? This last section is an incorporation of scientific research and reports and my own view on the future of data science in medicine.

AI has important implications for our modern health care. Medicine has long been evidence-based; clinical research always focussed on specific groups of patients. Classical studies used one (or multiple) independent predictors and only one dependent outcome variable in their statistical analysis. Studies were designed to determine which independent predictors had the most influence on the outcome variable. Based upon the results of these studies, guidelines are written and many patients are treated. However, not every patient is the same; other factors might play a role, factors that we currently overlook. These factors have not yet been identified or have not been taken into consideration until now. AI creates the opportunity to find sophisticated patterns in data. This can aid clinicians in patient-specific decision-making.

There are large amounts of clinical data available in patient files, both diagnostic images as patient outcome. This information is currently not used. Machine learning

has the ability to use the existing data and find new patterns in this data, without the need for extra clinical research. However, before we can use all the available data, the data must be modified and processed. This has been elaborately described by Lawrence as 'Data Readiness Levels'.[89]. The conclusion from this paper; the current structure used in patient files is inappropriate for application in machine learning analysis. Before we can use the large amount of data available in patient files, a generic way of saving information (e.g. annotation of images, filing of patient consults and patient outcome) must be developed and implemented on a large scale.

Beside the preprocessing of the data, the models with the best performances must be identified and trained to process all the data samples. This can be a laborious coding process, where trial-and-error is a key feature. Being laborious, it is also time-consuming. Computers are getting faster and better at calculation tasks. However, with a large amount of data, including missing data and noise, calculations are time-consuming, even for fast computers. The use of patient files and the information in these patients files also has disadvantages. This information must be handled anonymously, so it cannot be tracked back to one individual. Firstly, wide use of this data also increases the risk of data leaks. Secondly, awareness must be created among researchers, that they are working with patient information. This information should be handled with care and respect. Unless possible drawbacks, clinicians have a growing fidelity in AI; opportunities are being seen and challenges are tackled with AI.

In conclusion, this research has shown a field of Radiomics and machine learning in cardioradiology. Yet, the process of training, validating and testing models is complex. In the end, results must be good enough for clinical use. Then, the critical leap must be made towards implementation into clinical practice. This is the point that some clinicians are fearing; "Is AI going to take our jobs?" My answer would be "No, definitely not". I have been working for a year with clinicians, data scientists and other medical professionals and I have seen different stakeholders in health care. My experience with machine learning in radiology is that it is going to create opportunities for fast diagnosis and automatic analysis of medical images. Secondly, it has the chance to aid in the the evaluation of prognosis, prediction of patient outcome and determination of the best-suitable care path. 'Going to', because before machine learning will be available in clinical settings, years, probably decades, have passed. AI is going to enhance the work of clinicians. However, someone has to be responsible for the final decision and, in my opinion, this cannot be a model or a computer. This has to be a human, who also takes responsibility. In the end, I would rather talk to a empathic human being than to a robot or a computer.

Bibliography

[1] H J Aerts, E R Velazquez, R T Leijenaar, et al. „Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach“. In: *Nature Communications* 5.4006 (2014) (cit. on pp. 8, 19, 34, 60).

[2] Hugo J W L Aerts, Emmanuel Rios Velazquez, Ralph T.H. Leijenaar, et al. „Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach - Supplementary Information“. In: *Nature Communications* 5.4006 (2014), pp. 2–4 (cit. on pp. 22, 82, 92, 94).

[3] Douglas G Altman, Yvonne Vergouwe, Patrick Royston, and Karel G M Moons. „Prognosis and prognostic research: validating a prognostic model“. In: *the BMJ* 338 (2009) (cit. on pp. 46, 54).

[4] Luciano C. Amado, Bernhard L. Gerber, Sandeep N. Gupta, et al. „Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model“. In: *Journal of the American College of Cardiology* 44.12 (2004), pp. 2383–2389 (cit. on p. 34).

[5] Robert H. Anderson, Yen Ho Siew, Klaus Redmann, Damian Sanchez-Quintana, and Paul P. Lunkenheimer. „The anatomical arrangement of the myocardial cells making up the ventricular mass“. In: *European Journal of Cardio-thoracic Surgery* 28.4 (2005), pp. 517–525 (cit. on p. 59).

[6] Elliot M. Antman. „Decision making with cardiac troponin tests“. In: *The New England journal of medicine* 346 (2002), pp. 2079–2082 (cit. on p. 43).

[7] Ravi G. Assomull, Sanjay K. Prasad, Jonathan Lyne, et al. „Cardiovascular Magnetic Resonance, Fibrosis, and Prognosis in Dilated Cardiomyopathy“. In: *Journal of the American College of Cardiology* 48.10 (2006), pp. 1977–1985 (cit. on p. 62).

[8] Michele Avanzo, Joseph Stancanello, and Issam El Naqa. „Beyond imaging: The promise of radiomics“. In: *Physica Medica* 38 (2017), pp. 122–139 (cit. on pp. 13, 63).

[9] M R Avendi, Arash Kheradvar, and Hamid Jafarkhani. „A Combined Deep-Learning and Deformable-Model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI“. In: 30 (2015), pp. 1–34. arXiv: `arXiv:1512.07951v1` (cit. on p. 54).

[10] Clerio F. Azevedo, Marcelo Nigri, Maria L. Higuchi, et al. „Prognostic significance of myocardial fibrosis quantification by histopathology and magnetic resonance imaging in patients with severe aortic valve disease“. In: *Journal of the American College of Cardiology* 56.4 (2010), pp. 278–287 (cit. on pp. 29, 31).

[11] Bettina Baessler, Manoj Mannil, Sabrina Oebel, et al. „Subacute and Chronic Left Ventricular Myocardial Scar: Accuracy of Texture Analysis on Nonenhanced Cine MR Images.“ In: *Radiology* 000.0 (2017), p. 170213 (cit. on pp. 13, 19, 30, 34, 41, 53, 57, 63).

[12] Gilles Barone-Rochette, Sophie Piérard, Christophe De Meester De Ravenstein, et al. „Prognostic significance of LGE by CMR in aortic stenosis patients undergoing valve replacement“. In: *Journal of the American College of Cardiology* 64.2 (2014), pp. 144–154 (cit. on pp. 13, 19, 29, 31).

[13] Helmut Baumgartner, Judy Hung, Javier Bermejo, et al. „Recommendations on the echocardiographic assessment of aortic valve stenosis: A focused update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography“. In: *European Heart Journal Cardiovascular Imaging* 18.3 (2017), pp. 254–275. arXiv: 1611.06654 (cit. on pp. 2, 6, 20, 63).

[14] Chirag Bavishi, Kiruthika Balasundaram, and Edgar Argulian. „Integration of Flow-Gradient Patterns Into Clinical Decision Making for Patients With Suspected Severe Aortic Stenosis and Preserved LVEF: A Systematic Review of Evidence and Meta-Analysis.“ In: *JACC. Cardiovascular imaging* 9.11 (2016) (cit. on p. 6).

[15] Yoav Benjamini and Yosef Hochberg. „Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing“. In: *Journal of the Royal Statistical Society* 57.1 (1995), pp. 289–300 (cit. on pp. 25, 37, 87, 89).

[16] Sebastian Bickelhaupt, Daniel Paech, Philipp Kickingereder, et al. „Prediction of malignancy by a radiomic signature from contrast agent-free diffusion MRI in suspicious breast lesions found on screening mammography.“ In: *Journal of Magnetic Resonance Imaging* (2017) (cit. on p. 34).

[17] Pieter van der Bijl, Victoria Delgado, and Jeroen J. Bax. „Noninvasive imaging markers associated with sudden cardiac death“. In: *Trends in Cardiovascular Medicine* (2015), pp. 1–13 (cit. on pp. 19, 62).

[18] Philipp Blanke, U. Joseph Schoepf, and Jonathon A. Leipsic. „CT in Transcatheter Aortic Valve Replacement“. In: *Radiology* 269.3 (2013), pp. 650–669 (cit. on p. 6).

[19] David A. Bluemke, Richard A. Kronmal, João A C Lima, et al. „The Relationship of Left Ventricular Mass and Geometry to Incident Cardiovascular Events. The MESA (Multi-Ethnic Study of Atherosclerosis) Study“. In: *Journal of the American College of Cardiology* 52.25 (2008), pp. 2148–2155. arXiv: NIHMS150003 (cit. on p. 34).

[20] Jan. Bogaert, Steven Dymarkowski, and Andrew M. Taylor. *Clinical Cardiac MRI*. Ed. by A.L. Baerts and K. Sartor. New York: Springer-Verlag Berlin Heidelberg, 2005, p. 553 (cit. on pp. 7, 8).

[21] Robert O Bonow, Martin B Leon, Darshan Doshi, and Neil Moat. „Management strategies and future challenges for aortic valve disease.“ In: *Lancet (London, England)* 387.10025 (2016), pp. 1312–23 (cit. on p. 5).

[22] Nathaniel M Braman, Maryam Etesami, Prateek Prasanna, et al. „Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI.“ In: *Breast cancer research : BCR* 19.1 (2017), p. 57 (cit. on pp. 8, 19, 61).

[23] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. „LOF: Identifying Density-Based Local Outliers“. In: *Proceedings of the 2000 Acm Sigmod International Conference on Management of Data* (2000), pp. 1–12 (cit. on pp. 12, 83).

[24] Field Cady. „Machine Learning Classification“. In: *The Data Science Handbook*. Hoboken, N. Hoboken, New Jersey: John Wiley & Sons, Inc., 2017, pp. 97–120 (cit. on p. 11).

[25] Field Cady. „Machine Learning Overview“. In: *The Data Science Handbook*. First Edit. Hoboken, New Jersey: John Wiley & Sons, Inc., 2017. Chap. 6, pp. 87–91 (cit. on p. 9).

[26] Paolo G Camici and Filippo Crea. „Coronary microvascular dysfunction“. In: *New England Journal of Medicine* 356.8 (2007), pp. 830–840 (cit. on p. 58).

[27] Christopher Cao, Su C Ang, Praveen Indraratna, et al. „Systematic review and meta-analysis of transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis." In: *Annals of cardiothoracic surgery* 2.1 (2013), pp. 10–23 (cit. on p. 5).

[28] Manuel D. Cerqueira. „Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart". In: *Childhood A Global Journal Of Child Research* (2003), pp. 1448–1453 (cit. on p. 60).

[29] Calvin W L Chin, David Messika-Zeitoun, Anoop S V Shah, et al. „A clinical risk score of myocardial fibrosis predicts adverse outcomes in aortic stenosis". In: *European Heart Journal* 37.8 (2016), pp. 713–723 (cit. on pp. 3, 4, 13, 19, 21, 24, 26, 28–30, 35, 36, 42, 54, 57, 62).

[30] Marie Annick Clavel, Joseph Malouf, David Messika-Zeitoun, et al. „Aortic valve area calculation in aortic stenosis by CT and doppler echocardiography". In: *JACC: Cardiovascular Imaging* 8.3 (2015), pp. 248–257 (cit. on p. 6).

[31] Dennis V. Cokkinos and Stavros Chryssanthopoulos. „Rheumatic fever". In: *Diagnosis and Management of Adult Congenital Heart Disease*. Ed. by Michael A. Gatzoulis, Gary D. Webb, and Piers E.F. Daubeney. Second Edi. Saint Louis: Churchill Livingstone, 2011. Chap. 62, pp. 449–452 (cit. on p. 2).

[32] Alexandre Comte, Alain Lalande, Paul M. Walker, et al. „Visual estimation of the global myocardial extent of hyperenhancement on delayed contrast-enhanced MRI". In: *European Radiology* 14.12 (2004), pp. 2182–2187 (cit. on p. 34).

[33] Valentina D.A. Corino, Eros Montin, Antonella Messina, et al. „Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions". In: *Journal of Magnetic Resonance Imaging* (2017), pp. 1–12 (cit. on p. 34).

[34] Thibaud P. Coroller, Vishesh Agrawal, Elizabeth Huynh, et al. „Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC". In: *Journal of Thoracic Oncology* 12.3 (2017), pp. 467–476 (cit. on pp. 13, 34).

[35] Thibaud P. Coroller, Patrick Grossmann, Ying Hou, et al. „CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma". In: *Radiotherapy and Oncology* 114.3 (2015), pp. 345–350. arXiv: 15334406 (cit. on p. 8).

[36] Timothy J. W. Dawes, Antonio de Marvao, Wenzhe Shi, et al. „Machine Learning of Three-dimensional Right Ventricular Motion Enables Outcome Prediction in Pulmonary Hypertension: A Cardiac MR Imaging Study". In: *Radiology* 000.0 (2017), p. 161315 (cit. on p. 63).

[37] N De Jay, S Papillon-Cavanagh, C Olsen, G Bontempi, and B Haibe-Kains. „mRMRe: an R package for parallelized mRMR ensemble feature selection". In: *Submitted* (2012), p. . (Cit. on pp. 24, 48).

[38] Nicolas De Jay, Simon Papillon-Cavanagh, Catharina Olsen, et al. „MRMRe: An R package for parallelized mRMR ensemble feature selection". In: *Bioinformatics* 29.18 (2013), pp. 2365–2368 (cit. on pp. 24, 37).

[39] Chris Ding and Hanchuan Peng. „Minimum redundancy feature selection from microarray gene expression data". In: 3.2 (2003), pp. 523–528 (cit. on pp. 25, 37).

[40] Marc R. Dweck, Nicholas A. Boon, and David E. Newby. „Calcific aortic stenosis: A disease of the valve and the myocardium". In: *Journal of the American College of Cardiology* 60.19 (2012), pp. 1854–1863 (cit. on pp. 1, 34, 43).

[41] Marc R. Dweck and Russel J. Everett. „Multibiomarker Strategies in Aortic Stenosis". In: *JACC: Cardiovascular Imaging* Article in (2017) (cit. on p. 62).

[42] Marc R. Dweck, Sanjiv Joshi, Timothy Murigu, et al. „Midwall fibrosis is an independent predictor of mortality in patients with aortic stenosis“. In: *Journal of the American College of Cardiology* 58.12 (2011), pp. 1271–1279 (cit. on pp. 4, 8, 13, 22, 29, 31, 34, 35, 43, 54, 55, 57, 62).

[43] Marc R Dweck, Nikhil V Joshi, James H F Rudd, and David E Newby. „Imaging of inflammation and calcification in aortic stenosis.“ In: *Curr Cardiol Rep* 15.1 (2013), p. 320 (cit. on p. 7).

[44] Marc R Dweck, Sanjiv Joshi, Tim Murigu, et al. „Left ventricular remodeling and hypertrophy in patients with aortic stenosis: insights from cardiovascular magnetic resonance“. In: *J Cardiovasc Magn Reson* 14 (2012), p. 50 (cit. on p. 19).

[45] Marc Richard Dweck, Charlotte Jones, Nikhil V. Joshi, et al. „Assessment of valvular calcification and inflammation by positron emission tomography in patients with aortic stenosis“. In: *Circulation* 125.1 (2012), pp. 76–86 (cit. on pp. 7, 62).

[46] Trygve Eftestøl, Leik Woie, Kjersti Engan, et al. „Texture Analysis to Assess Risk of Serious Arrhythmias after Myocardial Infarction“. In: *Computing in Cardiology* 39 (2012), pp. 365–368 (cit. on p. 30).

[47] MacKram F. Eleid, Paul Sorajja, Hector I. Michelena, et al. „Flow-gradient patterns in severe aortic stenosis with preserved ejection fraction: Clinical characteristics and predictors of survival“. In: *Circulation* 128.16 (2013), pp. 1781–1789. arXiv: NIHMS150003 (cit. on p. 19).

[48] Sammy Elmariah. „Patterns of Left Ventricular Remodeling in Aortic Stenosis: Therapeutic Implications“. In: *Current Treatment Options in Cardiovascular Medicine* 17.7 (2015), pp. 1–15 (cit. on pp. 3, 57).

[49] R. J. Everett, C. G. Stirrat, S. I R Semple, et al. „Assessment of myocardial fibrosis with T1 mapping MRI“. In: *Clinical Radiology* 71.8 (2016), pp. 768–778 (cit. on p. 51).

[50] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, Dinani Amorim, and Dinani Amorim Fernández-Delgado. „Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?“ In: *Journal of Machine Learning Research* 15 (2014), pp. 3133–3181 (cit. on p. 61).

[51] Andrew S. Flett, Jonathan Hasleton, Christopher Cook, et al. „Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance“. In: *JACC: Cardiovascular Imaging* 4.2 (2011), pp. 150–156 (cit. on pp. 22, 34).

[52] Ana Flores-Marín, Juan José Gómez-Doblas, Juan Caballero-Borrego, et al. „Long-Term Predictors of Mortality and Functional Recovery After Aortic Valve Replacement for Severe Aortic Stenosis With Left Ventricular Dysfunction“. In: *Rev Esp Cardiol* 63.1 (2010), pp. 36–45 (cit. on p. 19).

[53] Thomas A. Foley, Sunil V. Mankad, Nandan S. Anavekar, et al. „Measuring Left Ventricular Ejection Fraction - Techniques and Potential Pitfalls“. In: *European Cardiology* 8.2 (2012), pp. 108–114 (cit. on p. 62).

[54] Hiroshi Furukawa and Kazuo Tanemoto. „Current topics on bicuspid aortic valve: Clinical aspects and surgical management“. In: *Annals of Thoracic and Cardiovascular Surgery* 21.4 (2015), pp. 314–321 (cit. on pp. 2, 58).

[55] Mary M. Galloway. „Texture analysis using gray level run lengths“. In: *Computer Graphics and Image Processing* 4.2 (1975), pp. 172–179 (cit. on p. 94).

[56] Philippe Genereux, Gregg W Stone, Patrick T O'Gara, et al. „Natural History, Diagnostic Approaches, and Therapeutic Strategies for Patients With Asymptomatic Severe Aortic Stenosis.“ In: *Journal of the American College of Cardiology* 67.19 (2016), pp. 2263–2288 (cit. on p. 20).

[57]Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. „Radiomics: Images Are More than Pictures, They Are Data.“ In: *Radiology* 278.2 (2015), p. 151169 (cit. on p. 8).

[58]Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. „Radiomics: Images Are More than Pictures, They Are Data.“ In: *Radiology* 278.2 (2015), p. 151169 (cit. on p. 19).

[59]Ariana González Gómez, Covadonga Fernández-Golfín, Juan Manuel Monteagudo, et al. „Severe aortic stenosis patients with preserved ejection fraction according to flow and gradient classification: Prevalence and outcomes“. In: *International Journal of Cardiology* 248 (2017), pp. 211–215 (cit. on pp. 4, 6).

[60]Joost J M van Griethuysen, Andriy Fedorov, Chintan Parmar, et al. „Computational Radiomics System to Decode the Radiographic Phenotype“. In: *Cancer Research* (2017) (cit. on pp. 24, 36, 48, 61).

[61]Isabelle Guyon, André Elisseeff, and Andre@tuebingen Mpg De. „An Introduction to Variable and Feature Selection“. In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182. arXiv: `1111.6189v1` (cit. on p. 12).

[62]James A. Hanley and Barbara J. McNeil. „The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve“. In: *Radiology* 143 (1982), pp. 29–36 (cit. on p. 25).

[63]Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. „Textural Features for Image Classification“. In: *IEEE Transactions on Systems, Man, and Cybernetics* 3.6 (1973), pp. 610–621 (cit. on pp. 8, 94).

[64]Frank E. Harrell, R M Califf, D B Pryor, K L Lee, and R A Rosati. „Evaluating the yield of medical tests.“ In: *Jama* 247.18 (1982), pp. 2543–6 (cit. on p. 25).

[65]Caitlin J. Harrigan, Dana C. Peters, C. Michael Gibson, et al. „Hypertrophic Cardiomyopathy: Quantification of Late Gadolinium Enhancement with Contrast-enhanced Cardiovascular MR Imaging“. In: *Radiology* 258.1 (2011), pp. 128–133 (cit. on p. 43).

[66]Einar Heiberg, Martin Ugander, Henrik Engblom, et al. „Automated quantification of myocardial infarction from MR images by accounting for partial volume effects: animal, phantom, and human study“. In: *Radiology* 246.2 (2008), pp. 581–588 (cit. on p. 59).

[67]Ricarda Hinzpeter, Matthias W. Wagner, Moritz C. Wurnig, et al. „Texture analysis of acute myocardial infarction with CT: First experience study“. In: *PLoS ONE* 12.11 (2017), pp. 1–16 (cit. on p. 34).

[68]Gethin W Hodges, Casper N Bang, Jesper Eugen-Olsen, et al. „SuPAR Predicts Cardiovascular Events and Mortality in Patients With Asymptomatic Aortic Stenosis.“ In: *The Canadian journal of cardiology* 32.12 (2016), pp. 1462–1469 (cit. on p. 4).

[69]Elizabeth Huynh, Thibaud P. Coroller, Vivek Narayan, et al. „CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer“. In: *Radiotherapy and Oncology* 120.2 (2016), pp. 258–266 (cit. on pp. 13, 34).

[70]Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, et al. *caret: Classification and Regression Training*. R package version 6.0-77.9000 (cit. on p. 25).

[71]Jessica Joseph, Syed Yaseen Naqvi, Jay Giri, and Sheldon Goldberg. „Aortic stenosis: pathophysiology, diagnosis and therapy“. In: *The American Journal of Medicine* (2016) (cit. on p. 5).

[72]Duk Hyun Kang, Sung Ji Park, Ji Hye Rim, et al. „Early surgery versus conventional treatment in asymptomatic very severe aortic stenosis“. In: *Circulation* 121.13 (2010), pp. 1502–1509 (cit. on p. 1).

[73]Peter Kellman and Andrew E. Arai. „Cardiac imaging techniques for physicians: Late enhancement“. In: *Journal of Magnetic Resonance Imaging* 36.3 (2012), pp. 529–542 (cit. on p. 60).

[74] Takeshi Kitai, Tomohiko Taniguchi, Takeshi Morimoto, et al. „Different clinical outcomes in patients with asymptomatic severe aortic stenosis according to the stage classification: Does the aortic valve area matter?“ In: *International Journal of Cardiology* 228 (2017), pp. 244–252 (cit. on p. 3).

[75] Christoph Klein, Stephan G. Nekolla, Frank M. Bengel, et al. „Assessment of myocardial viability with contrast-enhanced magnetic resonance imaging comparison with positron emission tomography“. In: *Circulation* 105.2 (2002), pp. 162–167 (cit. on p. 34).

[76] Márton Kolossváry, Miklós Kellermayer, Béla Merkely, and Pál Maurovich-Horvat. „Cardiac Computed Tomography Radiomics“. In: *Journal of Thoracic Imaging* 00.00 (2017), p. 1 (cit. on pp. 30, 34).

[77] Igor Kononenko. „Machine learning for medical diagnosis: History, state of the art and perspective“. In: *Artificial Intelligence in Medicine* 23.1 (2001), pp. 89–109 (cit. on pp. 9, 41, 61).

[78] Lasya Priya Kotu, Kjersti Engan, Reza Borhani, et al. „Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients“. In: *Artificial Intelligence in Medicine* 64.3 (2015), pp. 205–215 (cit. on p. 34).

[79] G. Krasopoulos, F. Falconieri, U. Benedetto, et al. „European real world trans-catheter aortic valve implantation: Systematic review and meta-analysis of European national registries“. In: *Journal of Cardiothoracic Surgery* 11.1 (2016), pp. 1–9 (cit. on p. 5).

[80] Max Kuhn, Contributions from: Jed Wing, et al. *caret: Classification and Regression Training. R package*. 2017 (cit. on p. 11).

[81] Max Kuhn and Kjell Johnson. „Classification Trees and Rule-Based Models“. In: *Applied Predictive Modeling*. New York: Springer Science+Business, 2013. Chap. 14, pp. 369–413 (cit. on p. 11).

[82] Max Kuhn and Kjell Johnson. „Data Pre-processing“. In: *Applied Predictive Modeling*. New York: Springer Science+Business, 2013. Chap. 3, pp. 27–59 (cit. on p. 83).

[83] Max Kuhn and Kjell Johnson. „Discriminant Analysis and Other Linear Classification Models“. In: *Applied Predictive Modeling*. New York: Springer Science+Business Media, 2013. Chap. 12, pp. 275–328 (cit. on p. 10).

[84] Max Kuhn and Kjell Johnson. „Nonlinear Classification Models“. In: *Applied Predictive Modeling*. New York: Springer Science+Business, 2013. Chap. 13, pp. 329–367 (cit. on p. 11).

[85] Max Kuhn and Kjell Johnson. „Over-fitting and Model Tuning“. In: *Applied Predictive Modeling*. New York: Springer Science+Business, 2013. Chap. 4, pp. 61–92 (cit. on pp. 9, 11).

[86] Virendra Kumar, Yuhua Gu, Satrajit Basu, et al. „Radiomics: The process and the challenges“. In: *Magnetic Resonance Imaging* 30.9 (2012), pp. 1234–1248. arXiv: NIHMS150003 (cit. on p. 8).

[87] Raymond Y. Kwong and Afshin Farzaneh-Far. „Measuring myocardial scar by CMR“. In: *JACC: Cardiovascular Imaging* 4.2 (2011), pp. 157–160 (cit. on pp. 22, 34).

[88] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, et al. „Radiomics: Extracting more information from medical images using advanced feature analysis“. In: *European Journal of Cancer* 48.4 (2012), pp. 441–446 (cit. on p. 8).

[89] Neil D. Lawrence. „Data Readiness Levels“. 2017 (cit. on p. 64).

[90] Davide Lazzeroni, Ornella Rimoldi, and Paolo G Camici. „From Left Ventricular Hypertrophy to Dysfunction and Failure.“ In: *Circulation journal : official journal of the Japanese Circulation Society* 80.3 (2016), pp. 555–564 (cit. on pp. 3, 13, 20, 34).

[91] Ralph T. H. Leijenaar, Sara Carvalho, Emmanuel Rios Velazquez, et al. „Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability“. In: *Acta Oncologica* 52.7 (2013), pp. 1391–1397. arXiv: 15334406 (cit. on p. 60).

[92] Martin B. Leon, Craig R. Smith, Michael J. Mack, et al. „Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients“. In: *New England Journal of Medicine* 374.17 (2016), pp. 1609–1620 (cit. on p. 5).

[93] Andy Liaw and Matthew Wiener. „Classification and Regression by randomForest“. In: *R News* 2.3 (2002), pp. 18–22 (cit. on p. 25).

[94] Jina Lim, Arshia Ehsanipour, Jeffrey J. Hsu, et al. „Inflammation Drives Retraction, Stiffening, and Nodule Formation via Cytoskeletal Machinery in a Three-Dimensional Culture Model of Aortic Stenosis“. In: *American Journal of Pathology* 186.9 (2016), pp. 2378–2389 (cit. on p. 1).

[95] Eugene Lin and Adam Alessio. „What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?“ In: *Journal of Cardiovascular Computed Tomography2* 3.6 (2009), pp. 403–408 (cit. on p. 31).

[96] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al. „A Survey on Deep Learning in Medical Image Analysis“. In: 1995 (2017). arXiv: 1702.05747 (cit. on p. 9).

[97] Patrick Mathieu and Marie-Chloé Boulanger. „Basic mechanisms of calcific aortic valve disease.“ In: *The Canadian journal of cardiology* 30.9 (2014), pp. 982–93 (cit. on p. 19).

[98] Andreas Mayr and Matthias Schmid. „Boosting the concordance index for survival data - A unified framework to derive and evaluate biomarker combinations“. In: *PLoS ONE* 9.1 (2014). arXiv: arXiv:1307.6417v1 (cit. on p. 25).

[99] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.6-8. 2017 (cit. on pp. 11, 25).

[100] James C C Moon, Emma Reed, Mary N. Sheppard, et al. „The histologic basis of late gadolinium enhancement cardiovascular magnetic resonance in hypertrophic cardiomyopathy“. In: *Journal of the American College of Cardiology* 43.12 (2004), pp. 2260–2264 (cit. on p. 43).

[101] Schroeder MS, Culhane AC, Quackenbush J, and Haibe-Kains B. „survcomp: an R/Bioconductor package for performance assessment and comparison of survival models“. In: *Bioinformatics* 27(22) (2011), pp. 3206–3208. (Cit. on pp. 25, 37).

[102] Sukrit Narula, Khader Shameer, Alaa Mabrouk Salem Omar, Joel T. Dudley, and Partho P. Sengupta. „Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography“. In: *Journal of the American College of Cardiology* 68.21 (2016), pp. 2287–2295 (cit. on p. 63).

[103] Franck Neycenssac. „Contrast Enhancement Using the Laplacian-of-a-Gaussian Filter“. In: *CVGIP: Graphical Models and Image Processing* 55.6 (1993), pp. 447–463 (cit. on p. 83).

[104] Marcelo Nigri, Clerio F. Azevedo, Carlos Eduardo Rochitte, et al. „Contrast-enhanced magnetic resonance imaging identifies focal regions of intramyocardial fibrosis in patients with severe aortic valve disease: Correlation with quantitative histopathology“. In: *American Heart Journal* 157.2 (2009), pp. 361–368 (cit. on p. 34).

[105] Rick A. Nishimura, Catherine M. Otto, Robert O. Bonow, et al. „2014 AHA/ACC guideline for the management of patients with valvular heart disease: A report of the American college of cardiology/American heart association task force on practice guidelines“. In: *Journal of the American College of Cardiology* 63.22 (2014). arXiv: NIHMS150003 (cit. on pp. 1, 20).

[106]Vuyisile T. Nkomo, Julius M. Gardin, Thomas N. Skelton, et al. „Burden of valvular heart diseases: a population-based study“. In: *Lancet* 368.9540 (2006), pp. 1005–1011 (cit. on p. 19).

[107]Rory O'Hanlon, Agata Grasso, Michael Roughton, et al. „Prognostic significance of myocardial fibrosis in hypertrophic cardiomyopathy“. In: *Journal of the American College of Cardiology* 56.11 (2010), pp. 867–874 (cit. on p. 62).

[108]Luke Oakden-Rayner, Gustavo Carneiro, Taryn Bessen, et al. „Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework“. In: *Scientific Reports* 7.1 (2017), p. 1648 (cit. on pp. 13, 34).

[109]Ruben L.J. Osnabrugge, Darren Mylotte, Stuart J. Head, et al. „Aortic Stenosis in the Elderly“. In: *Journal of the American College of Cardiology* 62.11 (2013), pp. 1002–1012 (cit. on p. 19).

[110]Ramdas G. Pai, Nikhil Kapoor, Ramesh C. Bansal, and Padmini Varadarajan. „Malignant Natural History of Asymptomatic Severe Aortic Stenosis: Benefit of Aortic Valve Replacement“. In: *Annals of Thoracic Surgery* 82.6 (2006), pp. 2116–2122 (cit. on p. 20).

[111]Jean Michel Paradis, Justin Fried, Tamim Nazif, et al. „Aortic stenosis and coronary artery disease: What do we know?What don't we know? A comprehensive review of the literature with proposed treatment algorithms“. In: *European Heart Journal* 35.31 (2014), pp. 2069–2082 (cit. on p. 6).

[112]Chintan Parmar, Patrick Grossmann, Johan Bussink, Philippe Lambin, and Hugo J W L Aerts. „Machine Learning methods for Quantitative Radiomic Biomarkers“. In: *Scientific reports* 5 (2015), p. 13087 (cit. on pp. 12, 61).

[113]Chintan Parmar, Emmanuel Rios Velazquez, Ralph Leijenaar, et al. „Robust radiomics feature quantification using semiautomatic volumetric segmentation“. In: *PLoS ONE* 9.7 (2014) (cit. on p. 8).

[114]Patricia A. Pellikka, Maurice E. Sarano, Rick A. Nishimura, et al. „Outcome of 622 adults with asymptomatic, hemodynamically significant aortic stenosis during prolonged follow-up“. In: *Circulation* 111.24 (2005), pp. 3290–3295 (cit. on pp. 13, 19, 24).

[115]Elena Pena, MacArinze Ojiaku, Joao R. Inacio, et al. „Can CT and MR Shape and Textural Features Differentiate Benign Versus Malignant Pleural Lesions?“ In: *Academic Radiology* 11 (2017), pp. 1–11 (cit. on pp. 8, 22, 34).

[116]Hanchuan Peng. „Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance and Min-Redundancy“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238 (cit. on p. 12).

[117]Steffen E. Petersen, Nay Aung, Mihir M. Sanghvi, et al. „Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort“. In: *Journal of Cardiovascular Magnetic Resonance* 19.1 (2017), p. 18 (cit. on pp. 42, 59).

[118]Joanna Petryka, Jolanta Mi??ko, Andrzej Przybylski, et al. „Magnetic resonance imaging assessment of intraventricular dyssynchrony and delayed enhancement as predictors of response to cardiac resynchronization therapy in patients with heart failure of ischaemic and non-ischaemic etiologies“. In: *European Journal of Radiology* 81.10 (2012), pp. 2639–2647 (cit. on p. 43).

[119]Michael P. Pfeiffer and Robert W.W. Biederman. „Cardiac MRI: A general Overview with Emphasis on Current Use and Indications“. In: *Medical Clinics of North America* 99.4 (2015), pp. 849–861 (cit. on p. 51).

[120]R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017 (cit. on pp. 9, 26, 38, 49).

[121] Gopalan Nair Rajesh, Julian Johny Thottian, Gomathy Subramaniam, et al. „Prevalence and prognostic significance of left ventricular myocardial late gadolinium enhancement in severe aortic stenosis“. In: *Indian Heart Journal* (2017) (cit. on pp. 34, 35, 43, 54, 55, 62).

[122] Andrew N Rassi, Philippe Pibarot, and Sammy Elmariah. „Left ventricular remodelling in aortic stenosis“. In: *Can J Cardiol* 30.9 (2014), pp. 1004–1011 (cit. on pp. 3, 13, 19, 34).

[123] Michael J. Reardon, Nicolas M. Van Mieghem, Jeffrey J. Popma, et al. „Surgical or Transcatheter Aortic-Valve Replacement in Intermediate-Risk Patients“. In: *New England Journal of Medicine* 376.14 (2017), pp. 1321–1331 (cit. on p. 5).

[124] Johannes Riegler, King K. Cheung, Yiu Fung Man, et al. „Comparison of segmentation methods for MRI measurement of cardiac function in rats“. In: *Journal of Magnetic Resonance Imaging* 32.4 (2010), pp. 869–877 (cit. on pp. 42, 59).

[125] Xavier Robin, Natacha Turck, Alexandre Hainard, et al. „pROC: an open-source package for R and S+ to analyze and compare ROC curves“. In: *BMC Bioinformatics* 12 (2011), p. 77 (cit. on p. 37).

[126] Raphael Rosenhek, Thomas Binder, Gerold Porenta, et al. „Predictors of outcome in severe, asymptomatic aortic stenosis“. In: *N Engl J Med* 343.9 (2000), pp. 611–617 (cit. on pp. 4, 13, 19, 24).

[127] Raphael Rosenhek, Julia Mascherbauer, David C. Maintz, and Helmut Baumgartner. „Valvular Stenosis“. In: *The ESC Textbook of Cardiovascular Imaging*. London: Spring-Verlag, 2010. Chap. Chapter 7, pp. 123–147 (cit. on p. 7).

[128] Yvan Saeys, I??aki Inza, and Pedro Larra??aga. „A review of feature selection techniques in bioinformatics“. In: *Bioinformatics* 23.19 (2007), pp. 2507–2517 (cit. on p. 61).

[129] AL Samuel. „Some studies in machine learning using the game of checkers“. In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229 (cit. on p. 9).

[130] Paul A. Scott, James A. Rosengarten, Nick P. Curzen, and John M. Morgan. „Late gadolinium enhancement cardiac magnetic resonance imaging for the prediction of ventricular tachyarrhythmic events: A meta-analysis“. In: *European Journal of Heart Failure* 15.9 (2013), pp. 1019–1027 (cit. on p. 43).

[131] Sanjiv J. Shah, Daniel H. Katz, Senthil Selvaraj, et al. „Phenomapping for novel classification of heart failure with preserved ejection fraction“. In: *Circulation* 131.3 (2015), pp. 269–279. arXiv: 15334406 (cit. on p. 34).

[132] Umesh C Sharma, Paul Barenbrug, Saraswati Pokharel, et al. „Systematic review of the outcome of aortic valve replacement in patients with aortic stenosis.“ In: *The Annals of thoracic surgery* 78.1 (2004), pp. 90–5 (cit. on p. 13).

[133] Christopher T Sibley, Radwa A Noureldin, Neville Gai, et al. „T1 Mapping in cardiomyopathy at cardiac MR: comparison with endomyocardial biopsy.“ In: *Radiology* 265.3 (2012), pp. 724–732 (cit. on pp. 43, 60).

[134] Anvesha Singh, Christopher D. Steadman, and Gerry P. McCann. „Advances in the understanding of the pathophysiology and management of aortic stenosis: Role of novel imaging techniques“. In: *Canadian Journal of Cardiology* 30.9 (2014), pp. 994–1003 (cit. on p. 2).

[135] Craig R. Smith, Martin B. Leon, Michael J. Mack, et al. „Transcatheter versus Surgical Aortic-Valve Replacement in High-Risk Patients“. In: *New England Journal of Medicine* 364.23 (2011), pp. 2187–2198 (cit. on p. 5).

[136] Ole Geir Solberg, Thor Ueland, Ragnhild Wergeland, et al. „High-sensitive troponin T and N-terminal-brain-natriuretic-peptide predict outcome in symptomatic aortic stenosis." In: *Scandinavian cardiovascular journal : SCJ* 46.5 (2012), pp. 278–85 (cit. on p. 62).

[137] Gregor Sommer, Jens Bremerich, and Gunnar Lund. „Magnetic resonance imaging in valvular heart disease: Clinical application and current role for patient management". In: *Journal of Magnetic Resonance Imaging* 35.6 (2012), pp. 1241–1252 (cit. on p. 7).

[138] Paolo Spirito, Pietro Bellone, Kevin M. Harris, et al. „Magnitude of Left Ventricular Hypertrophy and Risk of Sudden Death in Hypertrophic Cardiomyopathy". In: *The New England Journal of Medicine* 342 (2000), pp. 1778–1785 (cit. on pp. 34, 53).

[139] Stanford University. *Aortic Stenosis Assessment*. 2009 (cit. on p. 6).

[140] Christopher D. Steadman, Michael Jerosch-Herold, Benjamin Grundy, et al. „Determinants and functional significance of myocardial perfusion reserve in severe aortic stenosis". In: *JACC: Cardiovascular Imaging* 5.2 (2012), pp. 182–189 (cit. on p. 62).

[141] Vidya Sudarshan, E.Y.K Ng, U. Rajendra Acharya, et al. „Computer-aided diagnosis of Myocardial Infarction using ultrasound images with DWT, GLCM and HOS methods: A comparative study". In: *Computers in Biology and Medicine* 62 (2015), pp. 86–93 (cit. on pp. 20, 34, 57).

[142] Kenji Suzuki. „Overview of deep learning in medical imaging". In: *Radiological Physics and Technology* June (2017) (cit. on pp. 9, 41).

[143] Andrew Swift, Smitha Rajaram, Dave Capener, et al. „Use of late gadolinium enhancement cardiac magnetic resonance for prediction of mortality in pulmonary hypertension". In: *The Lancet* 383 (2014), S99 (cit. on p. 43).

[144] Xiaoou Tang. „Texture information in run-length matrices". In: *IEEE Transactions on Image Processing* 7.11 (1998), pp. 1602–1609 (cit. on p. 94).

[145] Juan A. Terré, Isaac George, and Craig R. Smith. „Pros and cons of transcatheter aortic valve implantation (TAVI)". In: *Annals of Cardiothoracic Surgery* 6.5 (2017), pp. 444–452 (cit. on p. 5).

[146] Christian Tesche, Rozemarijn Vliegenthart, Taylor M. Duguay, et al. „Coronary Computed Tomographic Angiography-Derived Fractional Flow Reserve for Therapeutic Decision Making". In: *The American Journal of Cardiology* (2017) (cit. on p. 63).

[147] K B Tibazarwa, J A Volmink, and B M Mayosi. „Incidence of acute rheumatic fever in the world: a systematic review of population-based studies". In: *Heart* 94.12 (2008), pp. 1534–1540 (cit. on p. 2).

[148] Thomas A. Treibel, Begona Lopez, Arantxa Gonzalez, et al. „Reappraisingmyocardial fibrosis in severe aortic stenosis: an invasive and non-invasive study in 133 patients". In: *European Heart Journal* (2017), pp. 1–12 (cit. on pp. 36, 48, 62).

[149] Jack V. Tu. „Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes". In: *Journal of Clinical Epidemiology* 49.11 (1996), pp. 1225–1231 (cit. on p. 10).

[150] Alec Vahanian, Ottavio Alfieri, Felicita Andreotti, et al. „Guidelines on the management of valvular heart disease (version 2012)". In: *European Heart Journal* 33.19 (2012), pp. 2451–2496 (cit. on pp. 2, 5, 14, 19, 20, 24, 58).

[151] Joep W.M. Van Oorschot, Johannes M.I.H. Gho, Gerardus P.J. Van Hout, et al. „Endogenous contrast MRI of cardiac fibrosis: Beyond late gadolinium enhancement". In: *Journal of Magnetic Resonance Imaging* 41.5 (2015), pp. 1181–1189 (cit. on p. 60).

[152]Vassilios S. Vassiliou, Aris Perperoglou, Claire E. Raphael, et al. „Midwall Fibrosis and 5-Year Outcome in Moderate and Severe Aortic Stenosis“. In: *Journal of the American College of Cardiology* 69.13 (2017), pp. 1755–1756 (cit. on pp. 22, 29, 31, 34, 35, 43, 54, 55, 62).

[153]Akbar K. Waljee, Peter D R Higgins, and Amit G. Singal. „A primer on predictive models“. In: *Clinical and Translational Gastroenterology* 5.June (2014), pp. 1–4 (cit. on p. 46).

[154]Frank Weidemann, Sebastian Herrmann, Stefan Störk, et al. „Impact of myocardial fibrosis in patients with symptomatic severe aortic stenosis“. In: *Circulation* 120.7 (2009), pp. 577–584 (cit. on p. 60).

[155]Stephen S F Yip and Hugo J W L Aerts. „Applications and limitations of radiomics“. In: *Physics in Medicine and Biology* 61.13 (2016), R150–R166 (cit. on p. 46).

[156]Stephen S.F. Yip, John Kim, Thibaud P. Coroller, et al. „Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non–Small Cell Lung Cancer“. In: *Journal of Nuclear Medicine* 58.4 (2017), pp. 569–576 (cit. on p. 34).

[157]M. Zakkar, A.J. Bryan, and G.D. Angelini. „Aortic stenosis: diagnosis and management“. In: *The BMJ* 5425.October (2016), pp. 1–9 (cit. on pp. 1, 3, 12, 13).

[158]Shuixing Zhang, Bin Zhang, Jie Tian, et al. „Radiomics features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma“. In: *Clinical Cancer Research* (2017), clincanres.2910.2016 (cit. on pp. 34, 61).

[159]Binsheng Zhao, Leonard P. James, Chaya S. Moskowitz, et al. „Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer“. In: *Radiology* 252.1 (2009), pp. 263–272 (cit. on p. 60).

[160]Majd Zreik, Tim Leiner, Bob D De Vos, et al. „Automatic Segmentation of the Left Ventricle in Cardiac CT Angiography Using Convolutional Neural Netwerks“. In: *IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016*. Prague, Czech Republci, 2016, pp. 40–43 (cit. on p. 54).

[161]Alex Zwanenburg, Stefan Leger, Martin Vallières, Steffen Löck, and for the Image Biomarker Standardisation Initiative. „Image biomarker standardisation initiative“. In: November (2017). arXiv: 1612.07003 (cit. on p. 92).

List of Figures

List of Tables

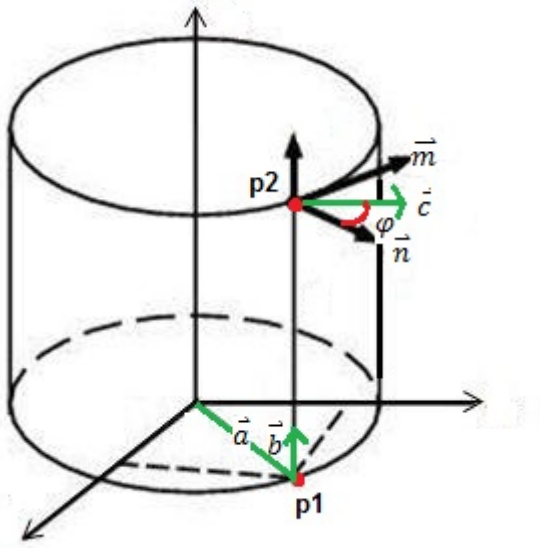
Supplements

6

6.1 Methods

6.1.1 CylindricalReconstruction

For the analysis of texture features, the segmented myocardium is reconstructed in a cylindrical coordinate system. This is done in Matlab 2016b (MathWorks, Natick, Massachusetts, United States). The cylindrical reconstruction is performed according to the following formula, which is graphically displayed in figure 6.1;



$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \vec{a} + \vec{b} * x + r(\vec{n} cos\phi \times \vec{m} sin\phi)$$

Fig. 6.1: The reconstruction from a cartesian coordinate system into a cylindrical coordinate system.

In the reconstruction of the image, every slice is reconstructed individually. The midpoint of the left ventricle is automatically detected for every slice and manually checked. The steps of circle identification is displayed in figure 6.2. These centres are used as the starting position for the reconstruction. The reconstruction process has been automatized in Matlab 2016b (MathWorks, Natick, United States). In this reconstruction, the steps taken on the radius (r) are displayed on the vertical axis. The angular steps (ϕ) are displayed on the horizontal axis.

6.1.2 Image Filters

The PyRadiomics package comes with different filters that can be applied to the input scans. These filters use different mathematical operations on the feature values. This section gives an outline of the available image filters, their mathematical

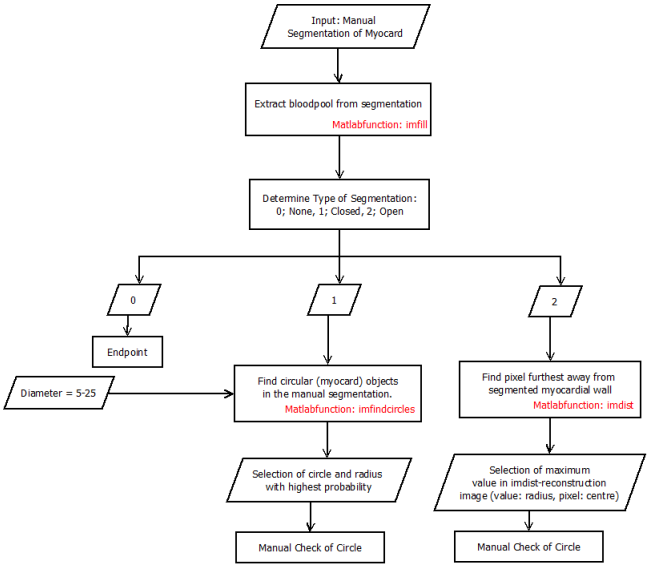


Fig. 6.2: Flowchart describing the different steps programmed in identification of the centres of the myocardium.

operations and their implication on LGE-CMR. To illustrate the different filters, 4 scans of subjects extracted from the dataset are shown. Every subject has a different LGE-pattern; none, patchy, diffuse and focal/infarct pattern.

Squared Filter This filter computes the square of the image intensities. Afterwards, it rescales the image intensities on the image intensity range of the original image.

$$I_{new} = (\frac{1}{\sqrt{I_{max}}} * I_{original})^2 \tag{6.1}$$

Squared Root Filter This filter computes the squared root of the original image intensity values and rescales it to the original range of the image intensities through the following function:

$$I_{new} = \sqrt{I_{max} * I_{original}} \tag{6.2}$$

Log Filter This filter uses the Laplacian of the gaussian filter. It yield a derived image for each σ specified. LoG-filters decrease the amount of noise in an image by the application of a filter. It aids in the detection

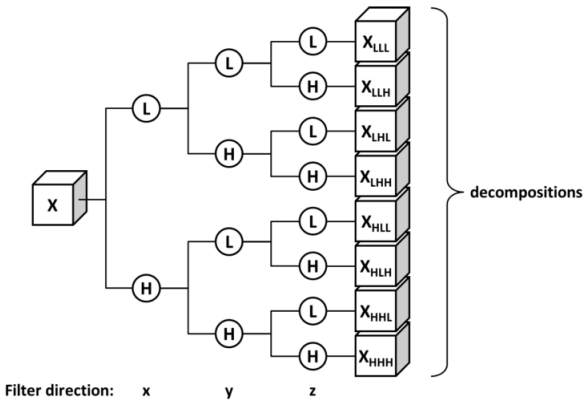


Fig. 6.3: Graphical overview how 8 different wavelet images can be constructed from one original scan. Image derived from Aerts, 2014a [2]

of edges and changes
in contrast.[103]

Wavelet Filter The wavelet transforms of images are calculated with the application of low- and high-pass functions in different directions of the image. This is comparable to the decomposition applied in Fourier analysis. Every 3D-scan is decomposed in 8 wavelet images through the application of the functions. Figure 6.3 describes the different images that can be generated.

Logarithmic Filter As described in the name of this filter, application of this filter to the image intensities generates the natural logarithmic value $+1$ of the original image intensities as the outcome image.

$$I_{new} = (\log I_{original} + 1) * \frac{I_{max}}{max(\log I_{original} + 1} \tag{6.3}$$

Exponential Filter The exponential filter computes the exponential of the original image intensities.

$$I_{new} = e^{\frac{lnI_{max}}{I_{max}} * I_{original}} \tag{6.4}$$

6.1.3 Processing of Feature Values for SVM

Normalization of feature values First of all, the data is centered and scaled according to the z-score. Centering is done by subtracting the average feature value from all the feature values. Scaling includes the division of every feature value by the feature's standard deviation. This method is done to improve the numerical stability of calculations. However, it reduces the interpretability of the individual values, because the original units are removed.[82]

Outlier Factor The other preprocessing step of the data is the detection of outliers by the LOF. The LOF-principle of outlier detection works with a K-nearest neighbours (K-NN) algorithm.[23] The LOF is based on K, which is the number of nearest neighbours used in the definition of the neighbourhood of the object. Objects within a cluster have an LOF close to 1. To determine an appropriate value for K, the LOF for every sample point is calculated for a range of K-values (15:50). To determine the final value of K for a certain feature, all LOFs from every sample over a range of K are taken into account. The minimum maximum LOF for a certain feature over all those K's is the final number of nearest neighbours used for determining the LOF for every sample of that feature. Thus, for every feature a different K is used to determine outliers. Figure 6.4 shows a plot of the feature 'Surface Area' and the different maximum values for LOF over a range of K.

After determining the number of K-NN, the upper boundary for LOF has to be determined. A sample is an outlier of the dataset if the LOF is significantly higher than one. Therefore, the upper boundary for the LOF is arbitrarily determined to be 1.8. All samples with an LOF> 1.8 are determined outliers and are set to the minimum or maximum non-outlier value, depending on which side of the spectrum the samples are. During the analysis it was shown that the LOF was sometimes > 1.8. However, the original value was within the minimum and maximum boundary of the corrected feature values. Therefore, it is checked if the original value of the feature, assumed to be an outlier, is higher than the maximum or lower than the minimum feature value. If this is the case, the sample feature value is set to the maximum or minimum. If the original feature value lies within the boundaries, the original value is retained.

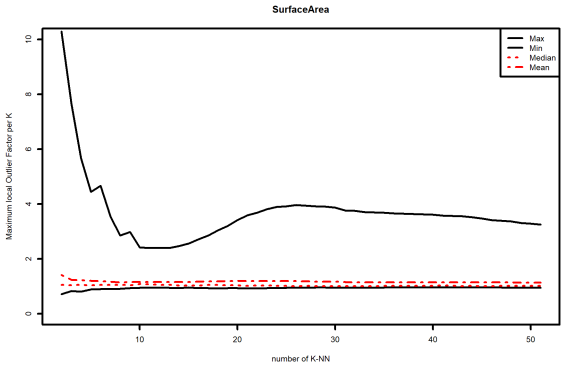


Fig. 6.4: The maximum, mean, median and minimum LOF values per K-nearest neighbours for the feature *Surface Area*

6.1.4 Pipeline of Data Analysis

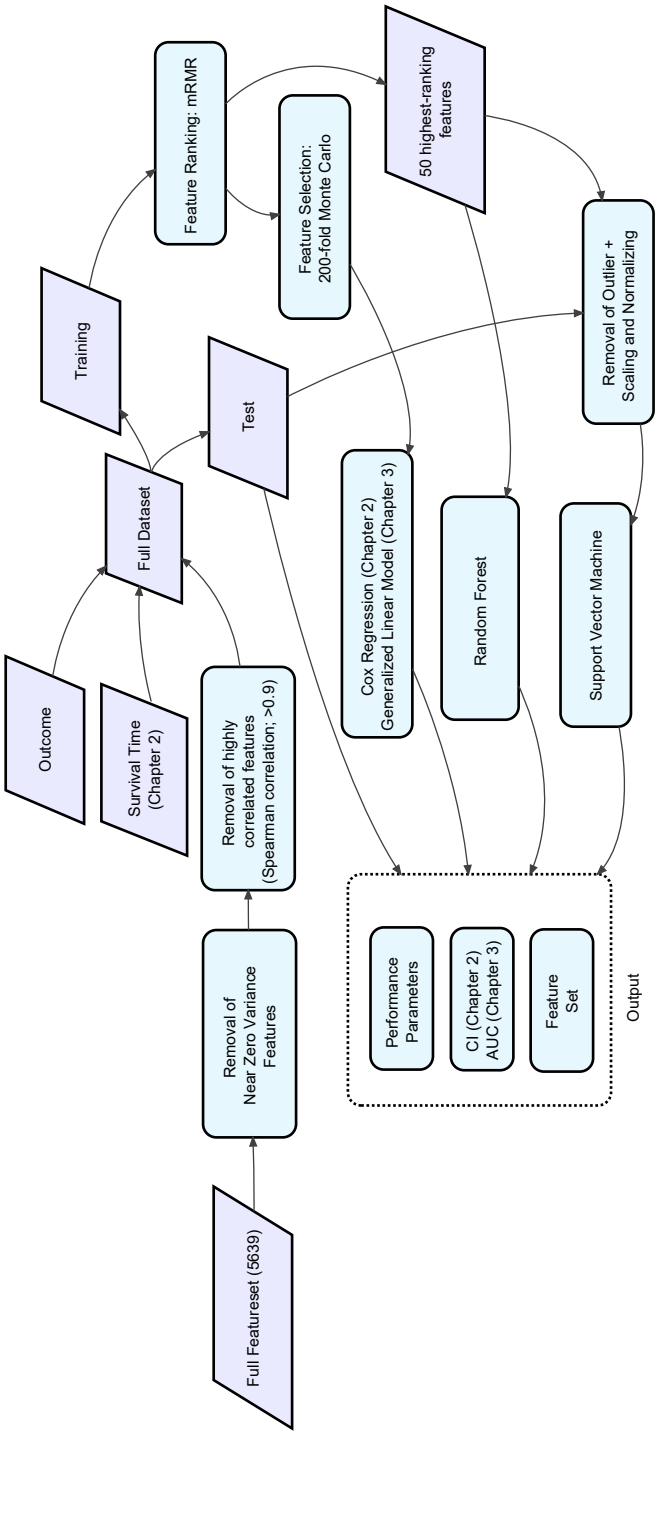


Fig. 6.5: Data analysis pipeline of the studies performed in this thesis

6.1.5 Used Functions R

Function	Reason	Package	Version
expand.grid	Define a grid for grid search for tuning parameters of machine learning models.	base	
scale	Centering and scaling the feature values.	base	
train	Building and training of classification models.	caret	6.0-77
nearZeroVar	Eliminate features with a near zero variance.	caret	
createDataPartition	Create a stratified random split for Monte Carlo analysis.	caret	
findCorrelation	Determine highly correlated variables for removal.	caret	
trainControl	Modification of the resampling method for model validation of the tuning parameters.	caret	
lofactor	Obtaining local outlier factors based upon the LOF-algorithm	DMwR	0.4.1
svmLinear2	Linear SVM-model.	e1071	1.6-8
ggplot	Plot with the ggplot2-package	ggplot2	2.2.1
mRMR.classic	Perform mRMR feature ranking.	mRMRe	2.0.7
roc	Determining the roc-curve from a set of observations and the true outcome	pROC	1.10.0.
coords	Determining the values of performance parameters at the optimal threshold of an ROC-curve.	pROC	
ci.auc	Calculation of 95%-confidence interval around AUC.	pROC	
rf	Random forest for model input in 'train'-function.	randomForest	4.6-12
prediction	Transform the input data to a standardized format for further evaluation.	ROCR	1.0-7
performance	Calculation of predictor evaluations.	ROCR	
cor	Determine Spearman's correlation matrix between features.	stats	
glm	Fitting a generalized linear model	stats	
lm	Fit linear model to data points.	stats	
p.adjust	P-value correction	stats	
wilcox.test	Wilcoxon-tests on vector(s) of continuous data comparison.	stats	
chisq.test	Chi-squared table test on categorical data comparison.	stats	
cor	Find correlation between two sets of datapoints.	stats	
concordance.index	Calculate the concordance index of a feature with relation to outcome and endpoint	survcomp	1.26

Tab. 6.1: Main used functions for the statistical analysis in R.

6.2 Results

6.2.1 Chapter 2

Features	Feature Class	CI-value	p-value	FDR q-value
wavelet.Norm.LLL glszm SmallAreaEmphasis	Texture	0.63	4.70E-06	0.002
exponential EndoVol glrlm RunVariance	Texture	0.63	1.94E-06	0.002
wavelet.Norm.LLL glszm ZoneEntropy	Texture	0.63	3.10E-06	0.002
wavelet.EpiVol.HLH glrlm RunVariance	Texture	0.63	1.87E-05	0.003
wavelet.Norm.HLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.63	2.34E-05	0.003
MaxThickness	Other	0.63	3.92E-05	0.005
original EpiVol glrlm LongRunHighGrayLevelEmphasis	Texture	0.63	7.78E-05	0.007
wavelet.EpiVol.HLH glszm LargeAreaLowGrayLevelEmphasis	Texture	0.63	0.00011	0.008
StDThickness	Case-specific	0.62	1.68E-05	0.003
wavelet.Norm.LLH glszm ZoneEntropy	Texture	0.62	1.25E-05	0.003
wavelet.EndoVol.HLL firstorder Mean	Statistics	0.62	2.09E-05	0.003
original EndoVol glrlm LongRunHighGrayLevelEmphasis	Texture	0.62	6.94E-05	0.007
wavelet.EpiVol.HLH glrlm LongRunLowGrayLevelEmphasis	Texture	0.62	0.00027	0.010
wavelet.Norm.HHL firstorder Mean	Statistics	0.61	8.70E-05	0.007
wavelet.Norm.LHL glcm Imc2	Texture	0.61	5.47E-05	0.006
wavelet.EndoVol.HHL glrlm ShortRunHighGrayLevelEmphasis	Texture	0.61	0.00024	0.010
wavelet.MidVol.HHL glszm GrayLevelNonUniformityNormalized	Texture	0.61	0.00018	0.009
wavelet.Norm.HLL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.61	0.00022	0.009
wavelet.Norm.HLL glcm Idn	Texture	0.61	0.00079	0.019
glrl phi Epi SRLGLE	Cylindrical	0.61	0.00024	0.010
wavelet.EpiVol.LHH glszm LargeAreaLowGrayLevelEmphasis	Texture	0.61	0.00052	0.014
wavelet.Norm.HHL glcm Idmn	Texture	0.61	3.43E-05	0.005
original Norm glszm LargeAreaHighGrayLevelEmphasis	Texture	0.61	0.00032	0.011
wavelet.EpiVol.HLL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.61	0.00062	0.016
wavelet.EndoVol.HLH glrlm LowGrayLevelRunEmphasis	Texture	0.61	0.00055	0.015
glrl phi Epi RLN	Cylindrical	0.61	8.79E-05	0.007
exponential Norm firstorder Energy	Statistics	0.61	0.00010	0.008
wavelet.EndoVol.HLH glrlm RunVariance	Texture	0.61	0.00017	0.009
wavelet.Norm.LLH glszm SmallAreaEmphasis	Texture	0.61	0.00014	0.009
wavelet.EpiVol.LHL firstorder Mean	Statistics	0.61	0.00035	0.011
exponential Norm glszm LargeAreaHighGrayLevelEmphasis	Texture	0.61	0.00022	0.009
wavelet.HLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.61	0.00020	0.009
wavelet.EndoVol.HLL firstorder Skewness	Statistics	0.61	0.00019	0.009
wavelet.LHL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.61	0.00016	0.009
wavelet.EndoVol.HLH glrlm LongRunLowGrayLevelEmphasis	Texture	0.61	0.00086	0.018
wavelet.EndoVol.LHL firstorder Mean	Statistics	0.61	0.00032	0.011
wavelet.EndoVol.HLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.61	0.00041	0.013
wavelet.EpiVol.HHL glrlm ShortRunHighGrayLevelEmphasis	Texture	0.61	0.00074	0.018
wavelet.Norm.HHH glcm Imc2	Texture	0.61	0.00032	0.011
original VolStd 2Sigma glrlm RunLengthNonUniformity	Texture	0.60	0.00021	0.009
wavelet.EndoVol.LHH glrlm RunVariance	Texture	0.60	0.00044	0.013
wavelet.EndoVol.HLL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.60	0.00117	0.024
wavelet.Norm.HLH glcm Idn	Texture	0.60	0.00043	0.013
wavelet.Norm.LLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.60	0.00055	0.015
glrl r Epi SRE	Cylindrical	0.60	0.00018	0.009
wavelet.EndoVol.HLL glcm Imc1	Texture	0.60	0.00046	0.013
wavelet.EndoVol.LHH firstorder Mean	Statistics	0.60	0.00114	0.024
wavelet.Norm.LHL glcm Idn	Texture	0.60	0.00022	0.009
wavelet.Norm.LHL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.60	0.00118	0.024

Tab. 6.2: This table describes all radiomics features that are somewhat related to the risk of AVR. P-value analysis has been done with the Wilcoxon-test, correction of the q-value was done with the correction from Benjamini.[15]

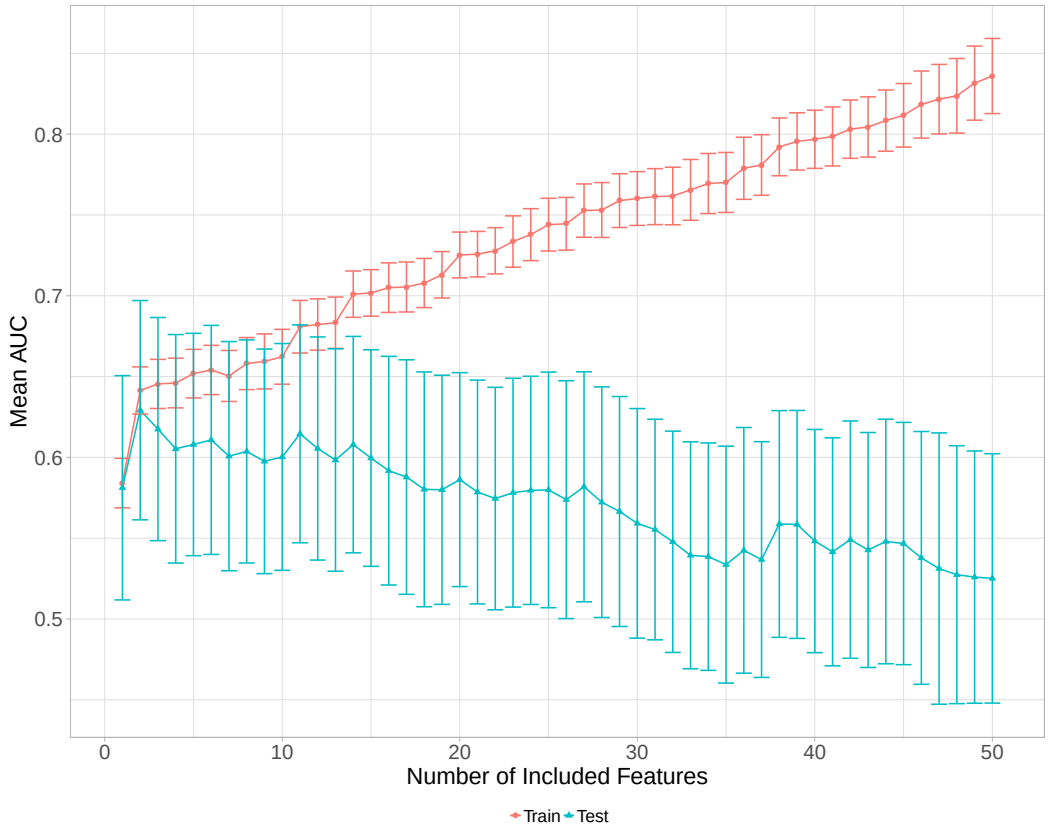


Fig. 6.6: 200-Fold Monte Carlo analysis on the training set to determine the optimal number of features that should be included in the multivariate model for the prediction of AVR.

6.2.2 Chapter 3

Features	Feature Class	AUC	p-value	FDR q-value
wavelet.Norm.LLL firstorder Range	Statistics	0.85	1.64E-14	1.86E-11
MaxThickness	Case-specific	0.84	6.78E-14	3.86E-11
MeanThickness	Case-specific	0.83	2.88E-13	8.76E-11
wavelet.Norm.LLL glszm ZoneEntropy	Texture	0.83	3.08E-13	8.76E-11
wavelet.EndoVol.HLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.83	7.38E-13	1.68E-10
wavelet.Norm.LLL glcm Idmn	Texture	0.82	3.39E-12	6.43E-10
original VolStd 2Sigma shape Volume	Shape	0.82	5.76E-12	9.37E-10
original VolStd 1Sigma shape Volume	Shape	0.81	2.62E-11	3.73E-09
original VolStd 2Sigma shape SurfaceArea	Shape	0.80	3.55E-11	4.49E-09
wavelet.Norm.HLL firstorder TotalEnergy	Statistics	0.80	4.95E-11	5.64E-09
wavelet.Norm.LLH firstorder Maximum	Statistics	0.80	6.75E-11	6.41E-09
wavelet.Norm.LLH firstorder TotalEnergy	Statistics	0.80	6.75E-11	6.41E-09
exponential firstorder TotalEnergy	Statistics	0.80	1.05E-10	9.21E-09
original VolStd 2Sigma glrlm RunLengthNonUniformity	Texture	0.79	1.82E-10	1.48E-08
wavelet.Norm.LLL glcm ClusterShade	Texture	0.79	2.42E-10	1.84E-08
StDThickness	Case-specific	0.79	2.97E-10	2.12E-08
original VolStd 2Sigma glcm DifferenceVariance	Texture	0.79	3.79E-10	2.40E-08
logarithm Norm glszm LargeAreaEmphasis	Texture	0.79	3.79E-10	2.40E-08
wavelet.HLL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.79	4.40E-10	2.64E-08
wavelet.EndoVol.HLL firstorder Skewness	Statistics	0.78	6.24E-10	3.55E-08
wavelet.Norm.LHH glszm LargeAreaLowGrayLevelEmphasis	Texture	0.78	9.67E-10	5.25E-08
wavelet.LHL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.78	1.49E-09	7.73E-08
wavelet.Norm.LLL glcm Correlation	Texture	0.78	1.75E-09	8.69E-08
original EndoVol shape SurfaceArea	Shape	0.77	2.06E-09	9.78E-08
wavelet.Norm.LLH glszm ZoneEntropy	Texture	0.77	2.17E-09	9.90E-08
wavelet.Norm.LLL firstorder Kurtosis	Statistics	0.77	2.74E-09	1.20E-07
wavelet.LHH glszm LargeAreaHighGrayLevelEmphasis	Texture	0.77	4.40E-09	1.79E-07
wavelet.Norm.LHL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.77	4.40E-09	1.79E-07
wavelet.Norm.LHH glcm_Imc2	Texture	0.77	6.12E-09	2.40E-07
original VolStd 2Sigma glcm JointEnergy	Texture	0.77	6.45E-09	2.45E-07
original VolStd HalfSigma glrlm RunLengthNonUniformity	Texture	0.76	9.10E-09	3.34E-07
wavelet.Norm.LHL glcm Idn	Texture	0.76	9.58E-09	3.41E-07
wavelet.HHL glszm LargeAreaHighGrayLevelEmphasis	Texture	0.76	1.32E-08	4.51E-07
original VolStd 2Sigma glszm SizeZoneNonUniformity	Texture	0.76	1.35E-08	4.51E-07
square Norm glcm SumSquares	Texture	0.76	1.59E-08	5.19E-07
wavelet.Norm.HHL glcm Idn	Texture	0.76	2.02E-08	6.38E-07
wavelet.HHH glszm LargeAreaHighGrayLevelEmphasis	Texture	0.76	2.27E-08	6.98E-07
logarithm Norm glszm LargeAreaHighGrayLevelEmphasis	Texture	0.75	3.79E-08	1.12E-06
square Norm glcm ClusterTendency	Texture	0.75	3.85E-08	1.12E-06
wavelet.Norm.HLL glszm LargeAreaLowGrayLevelEmphasis	Texture	0.75	4.18E-08	1.18E-06
original VolStd 1Sigma glcm DifferenceEntropy	Texture	0.75	4.25E-08	1.18E-06
wavelet.EndoVol.HLH glrlm RunVariance	Texture	0.75	4.76E-08	1.29E-06

Tab. 6.3: This table describes all radiomics features that are fairly related (AUC>0.75) to the risk of LGE. P-value analysis has been done with the Wilcoxon-test, correction of the q-value was done with the correction from Benjamini.[15]

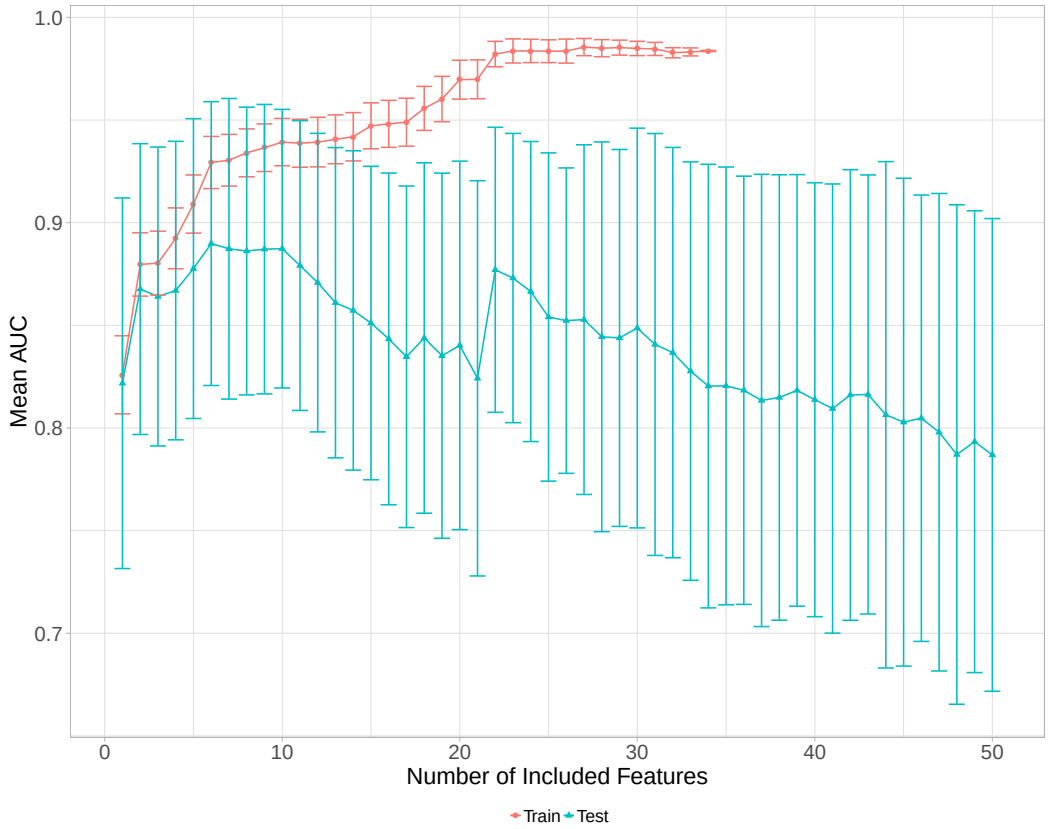


Fig. 6.7: 200-Fold Monte Carlo analysis on the training set to determine the optimal number of features that should be included in the GLM for computer-aided diagnosis of LGE.

6.2.3 Chapter 4

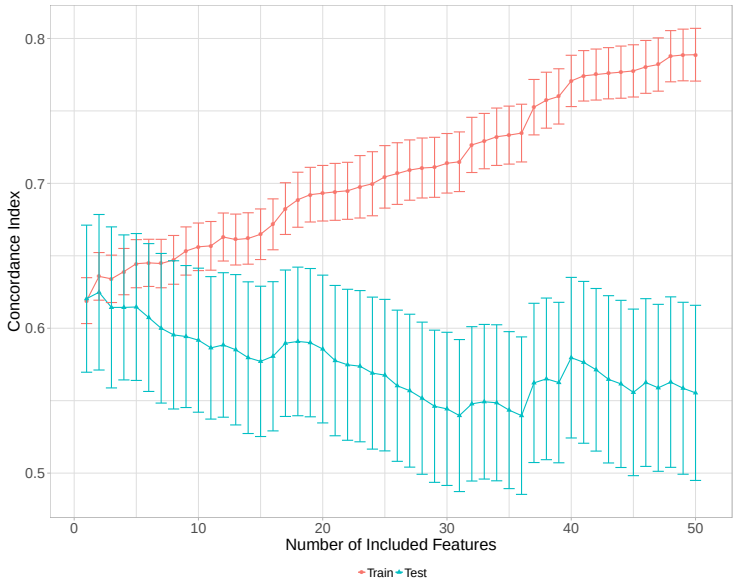


Fig. 6.8: 1000-fold random subsampling on the trainingset (dataset-1) to determine the optimal number of included features for the cox regression for classification of AVR.

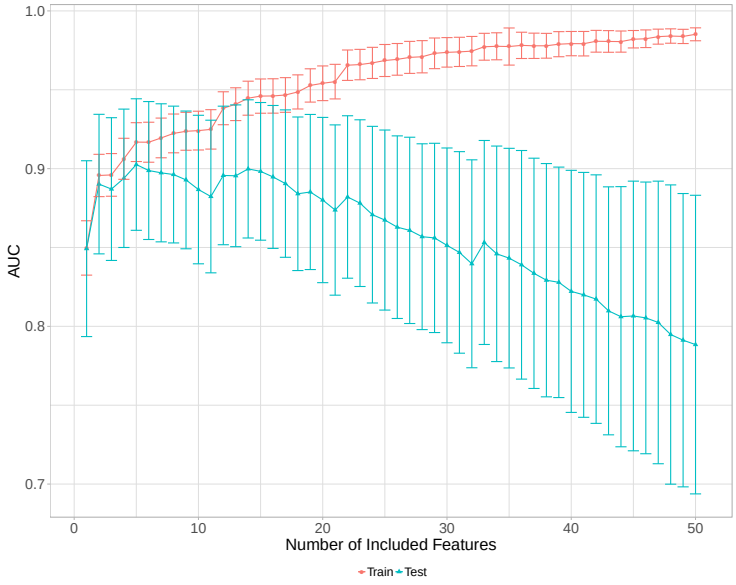


Fig. 6.9: 1000-fold random subsampling on the trainingset (dataset-1) to determine the optimal number of included features for the generalized linear model for computer-aided diagnosis of late gadolinium enhancement.

6.3 Radiomics Features

The features used in radiomics have been derived from different sources. To secure stability and a similar use of radiomics, the image biomarker standardisation initiative has been started. The features in the presented analysis have been derived from this source [161] and from the supplements from Aerts's article[2]

6.3.1 First Order Statistics

First order statistics are derived from the histogram with the voxel values of the segmentation. The following variables are defined: $\mathbf{X}$ = 3D-image matrix, $\overline{X}$ = the mean value of $\mathbf{X}$, N = number of voxels in this matrix and $\mathbf{P}$ = first order histogram with N_l = number of discrete intensity levels.

$$\text{energy} = \sum_{i}^{N} \mathbf{X}(i)^2 \tag{6.5}$$

$$\text{entropy} = \sum_{i=1}^{N_l} \mathbf{P}(i) \log_2 \mathbf{P}(i) \tag{6.6}$$

$$\text{kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X}^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X})^2}\right)^2} \tag{6.7}$$

$$\text{maximum} = \text{maximum intensity value of } \mathbf{X} \tag{6.8}$$

$$\text{mean} = \frac{1}{N}\sum_{i}^{N} \mathbf{X}(i) \tag{6.9}$$

$$\begin{aligned}\text{mean absolute deviation} &= \text{mean absolute deviation} \\ &\text{of all voxel intensities around the mean intensity value.}\end{aligned} \tag{6.10}$$

$$\text{median} = \text{median intensity value of } \mathbf{X}. \tag{6.11}$$

$$\text{minimum} = \text{minimum intensity value of } \mathbf{X} \tag{6.12}$$

$$\text{range} = \text{range of intensity values of } \mathbf{X} \tag{6.13}$$

$$\text{root mean square (RMS)} = \sqrt{\frac{\sum_i^N \mathbf{X}(i)^2}{N}} \tag{6.14}$$

$$\text{skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X})^2}\right)^3} \tag{6.15}$$

$$\text{standard deviation} = \left(\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X})^2\right)^{\frac{1}{2}} \tag{6.16}$$

$$\text{uniformity} = \sum_{i=1}^{N_l}\mathbf{P}(i)^2 \tag{6.17}$$

$$\text{variance} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}(i) - \overline{X})^2 \tag{6.18}$$

6.3.2 Shape- and size-based features

This feature group includes features that are related to the 3D size and shape of the segmented region. In these definitions V refers to the volume and A is the surface area of the segmented volume.

$$\text{compactness 1} = \frac{V}{\sqrt{\pi}A^{\frac{2}{3}}} \tag{6.19}$$

$$\text{compactness 2} = 36\pi\frac{V^2}{A^3} \tag{6.20}$$

$$\text{Maximum 3D diameter} = \text{largest pairwise Euclidean distance between voxels on the surface of the segmentational volume} \tag{6.21}$$

$$\text{spherical disproportion} = \frac{A}{4\pi R^2} \tag{6.22}$$

$$\text{sphericity} = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A} \tag{6.23}$$

$$\text{surface area} = \sum_{i=1}^{N} \frac{1}{2} |\mathbf{a}_i \mathbf{b}_i \times \mathbf{a}_i \mathbf{c}_i| \tag{6.24}$$

$$\text{surface to volume ratio} = \frac{A}{V} \tag{6.25}$$

$$\text{volume} = \text{counting the number of pixels in the segmentation and multiplying this value with the voxel size.} \tag{6.26}$$

6.3.3 Texture Features

This group of features includes descriptors for the patters and position in the gray levels in the medical images. The features are derived from two texture matrices;

- *Gray Level Co-occurence Matrix (GLCM)*: This is a matrix the size of $N_g \times N_g$, with N_g being the number of discrete intensity levels. The element at position i, j represents the number of times the combination i and j occur in two pixels next to each other. The matrix is described as $\mathbf{P}(i, j; \delta, \alpha)$, with δ the distance between the pixels in direction α.

- *Gray Level Run Length Matrix (GLRLM)*: This matrix type shows the length of the runs of pixels with similar gray levels in the image. A gray level run length matrix is described as $\mathbf{P}(i, j|\theta)$, where the (i, j)th element represents the number of times (j) a gray level (i) appears consecutively in the direction θ.

- *Gray level Distance Zone Based Features (GLSZM)*: This matrix counts the number of zones of linked voxels that share a similar bin or voxel value. These formulas have been derived from Thibault, 2014. citeThibault2014 N_g is the number of gray levels and N_z is the maximum size of zone. The element at position $S_{i,j}$ is the number of zone with gray value i and size j.

  For more information and examples on the texture matrices, see Aerts, [2]. Many of the gray level cooccurence texture features have been first described by Haralick, [63]. The features based on the gray level run length matrix have been described as first by Galloway, [55]. A complete overview of all the features that have been defined throughout the years has been given by Tang, [144].

Gray-Level Co-Occurence matrix based features Define:

$\mu =$ mean of $\mathbf{P}(i,j)$

$p_x(i) = \sum_{j=1}^{N_g}$ are the marginal row probabilities

$p_x(i) = \sum_{i=1}^{N_g}$ are the marginal column probabilities

$\mu_x =$ mean of p_x

$\mu_y =$ mean of p_y

$\sigma_x =$ standard deviation of p_x

$\sigma_y =$ standard deviation of p_y

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \mathbf{P}(i,j), i+j = k, k = 2,3,...,2N_g$

$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \mathbf{P}(i,j), |i-j| = k, k = 0,1,...,N_g - 1$

$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2[p_x(i)]$is the entropy of p_x

$HY = -\sum_{i=1}^{N_g} p_y(i) \log_2[p_y(i)]$is the entropy of p_y

$H = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \mathbf{P}(i,j) \log_2[\mathbf{P}(i,j)]$is the entropy of $\mathbf{P}(i,j)$

$HXY1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \mathbf{P}(i,j) log(p_x(i)p_y(j))$

$HXY1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) log(p_x(i)p_y(j))$

$$\text{autocorrelation} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij\mathbf{P}(i,j) \tag{6.27}$$

$$\text{cluster prominence} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i+j-\mu_x(i)-\mu_y(j)]^4 \mathbf{P}(i,j) \tag{6.28}$$

$$\text{cluster shade} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i+j-\mu_x(i)-\mu_y(j)]^3 \mathbf{P}(i,j) \tag{6.29}$$

$$\text{cluster tendency} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i+j-\mu_x(i)-\mu_y(j)]^2 \mathbf{P}(i,j) \tag{6.30}$$

$$\text{contrast} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j|^2 \mathbf{P}(i,j) \tag{6.31}$$

$$\text{correlation} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij\mathbf{P}(i,j) - \mu_i(i)\mu_j(j)}{\sigma_x(i)\sigma_y(j)} \tag{6.32}$$

$$\text{difference entropy} = \sum_{i=0}^{N_g-1} \mathbf{P}_{x-y}(i) \log_2 \left[\mathbf{P}_{x-y}(i)\right] \tag{6.33}$$

$$\text{dissimilarity} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j| \mathbf{P}(i,j) \tag{6.34}$$

$$\text{energy} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [\mathbf{P}(i,j)[^2 \tag{6.35}$$

$$\text{entropy (H)} = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \mathbf{P}(i,j) \log_2 [\mathbf{P}(i,j)] \tag{6.36}$$

$$\text{homogeneity 1} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{\mathbf{P}(i,j)}{1+|i-j|} \tag{6.37}$$

$$\text{homogeneity 2} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{\mathbf{P}(i,j)}{1+|i-j|^2} \tag{6.38}$$

$$\text{Informational measure of correlation 1 (IMC1)} = \frac{HXY - HXY1}{\max\{HX, HY\}} \tag{6.39}$$

$$\text{Informational measure of correlation 2 (IMC2)} = \sqrt{1 - e^{-1(HXY2-HXY}} \tag{6.40}$$

$$\text{Inverse Difference Moment Normalized (IDMN)} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{\mathbf{P}(i,j)}{1+\left(\frac{|i-j|^2}{N^2}\right)} \tag{6.41}$$

$$\text{Inverse Difference Normalized} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{\mathbf{P}(i,j)}{1+\left(\frac{|i-j|}{N^2}\right)} \tag{6.42}$$

$$\text{inverse variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{\mathbf{P}(i,j)}{|i-j|^2}, i \neq j \tag{6.43}$$

$$\text{maximum probability} = max\left\{\mathbf{P}(i,j)\right\} \tag{6.44}$$

$$\text{sum average} = \sum_{i=2}^{2N_g} [i\mathbf{P}_{x+y}(i)] \tag{6.45}$$

$$\text{sum entropy} = -\sum_{i=2}^{2N_g} \mathbf{P}_{x+y}(i) \log_2 [\mathbf{P}_{x+y}(i)] \tag{6.46}$$

$$\text{sum variance} = \sum_{i=2}^{2N_g} (i - SE)^2 \mathbf{P}_{x+y}(i) \tag{6.47}$$

$$\text{variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 \mathbf{P}(i, j) \tag{6.48}$$

Gray-Level Run-Length matrix based features Define:

$N_r =$ the number of different run lengths

$N_p =$ the number of voxels in the image

$$\text{Short Run Emphasis (SRE)} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i,j)|\theta)}{j^2}\right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.49}$$

$$\text{Long Run Emphasis (LRE)} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j|\theta)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.50}$$

$$\text{Gray Length Non-Uniformity (GLN)} = \frac{\sum_{i=1}^{N_g} \left[\sum_{j=1}^{N_r} p(i, j|\theta\right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.51}$$

$$\text{Run Length Non-Uniformity (RLN)} = \frac{\sum_{j=1}^{N_g} \left[\sum_{i=1}^{N_r} p(i, j|\theta\right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.52}$$

$$\text{Run Percentage (RP)} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j|\theta)}{N_p} \tag{6.53}$$

$$\text{Low Gray Level Run Emphasis(LGLRE)} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[\frac{p(i,j|\theta)}{i^2}\right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.54}$$

$$\text{High Gray Level Run Emphasis (HGLRE)} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 p(i, j|\theta)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)} \tag{6.55}$$

$$\text{Short Run Low Gray Level Run Emphasis(SRLGLE)} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)}{i^2 j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)} \tag{6.56}$$

$$\text{Short Run High Gray Level Run Emphasis(SRHGLE)} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)i^2}{j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)} \tag{6.57}$$

$$\text{Long Run Low Gray Level Run Emphasis(LRLGLE)} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)j^2}{i^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)} \tag{6.58}$$

$$\text{Long Run High Gray Level Run Emphasis(LRHGLE)} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)i^2 j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j|\theta)} \tag{6.59}$$

Gray-Level Size Zone Based Features Define: $p_{ij} = \frac{s_{ij}}{N_s}$

$\mu_i = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} i p_{ij}$

$\mu_i = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} j p_{ij}$

$$\text{Small Zone Emphasis} = \frac{1}{N_s}\sum_{j=1}^{N_z}\frac{s_{\cdot j}}{j^2} \tag{6.60}$$

$$\text{Large Zone Emphasis} = \frac{1}{N_s}\sum_{j=1}^{N_z} j^2 s_{\cdot j} \tag{6.61}$$

$$\text{Low Gray Level Zone Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g}\frac{s_{i\cdot}}{i^2} \tag{6.62}$$

$$\text{High Gray Level Zone Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g} i^2 s_{i\cdot} \tag{6.63}$$

$$\text{Small Zone Low Gray Level Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_z}\frac{s_{ij}}{i^2 j^2} \tag{6.64}$$

$$\text{Small Zone High Gray Level Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_z}\frac{i^2 s_{ij}}{j^2} \tag{6.65}$$

$$\text{Large Zone Low Gray Level Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_z}\frac{j^2 s_{ij}}{i^2} \tag{6.66}$$

$$\text{Large Zone High Gray Level Emphasis} = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_z} i^2 j^2 s_{ij} \tag{6.67}$$

$$\text{Gray Level Non-uniformity} = \frac{1}{N_s}\sum_{i=1}^{N_g} s_{i.}^2 \tag{6.68}$$

$$\text{Gray Level Non-uniformity normalized} = \frac{1}{N_s^2}\sum_{i=1}^{N_g} s_{i.}^2 \tag{6.69}$$

$$\text{Zone size non-uniformity} = \frac{1}{N_s}\sum_{j=1}^{N_z} s_{.j}^2 \tag{6.70}$$

$$\text{Zone size non-uniformity normalized} = \frac{1}{N_s}\sum_{j=1}^{N_z} s_{.j}^2 \tag{6.71}$$

$$\text{Zone Percentage} = \frac{N_s}{N + v} \tag{6.72}$$

$$\text{Gray level Variance} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z}(i - \mu_i)^2 p_{ij} \tag{6.73}$$

$$\text{Zone size Variance} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z}(j - \mu_j)^2 p_{ij} \tag{6.74}$$

$$\text{Zone size Entropy} = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_z} p_{ij}\log_2 p_{ij} \tag{6.75}$$